Assessing the value of warehousing as a service:
A case study in a fashion e-tailer

Paulo Jorge Fernandes Cardoso

Master's thesis

Supervisor: Prof. Dr. Mário Amorim Lopes

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Integrated Master in Engineering and Industrial Management

July 2018

Abstract

As the world economy becomes more globalised, increasingly efficient supply chains constitute a key advantage for the competitiveness of modern organisations. The objective of this work is to provide a multidisciplinary approach to assess the value of a potential investment project within a Supply Chain Management (SCM) initiative based on a concrete case study set in a luxury fashion e-tailer. In particular, the goal is to preliminary assess the value brought to the company by providing an outsourced warehousing service to key stakeholders of the network in order to bring supply closer to demand and ultimately gain operational control.

A preparatory analysis is first provided through a broad overview of the existing literature, the project background and the strategic rationale behind the initiative. Key costs and value drivers are identified from a qualitative standpoint.

A second stage illustrates the use of mathematical and computational tools in order to model and ultimately enable the quantification of the distinct revenue and cost items that drive the project's value. Emphasis was placed on the use of machine learning techniques, whose integration in simulation solutions has been largely unexplored to date. Besides that, this work also contributes to the literature with an analytical retention model.

A third stage describes a hybrid agent-based and discrete event simulation solution that was conceived in order to leverage the models and intelligence previously developed and generate results for hypothetical scenarios. The main objective of the simulation was to provide a rough estimate of the project's overall value and to quantify the relative importance and risk of the different drivers that may potentially intervene along its lifetime.

Finally, a fourth stage presents the results and sensitivity analysis iteratively obtained throughout the simulation phase and concludes with a review on the methodology employed, key recommendations concerning the value of the investment project, its strategic setting and the major risks.

Resumo

À medida que a economia mundial se torna cada vez mais globalizada, a existência de cadeias de abastecimento mais eficientes constitui uma vantagem primordial para a competitividade das organizações modernas. O objetivo do presente trabalho é desenvolver uma metodologia multidisciplinar capaz de aferir o valor de um potencial projeto de investimento inserido numa iniciativa de Gestão da Cadeia de Abastecimento e baseado, em particular, num caso de estudo relativo a um retalhista online atuando na indústria da moda de luxo. Em concreto, procura-se estimar de um modo preliminar o valor criado e capturado pela empresa ao providenciar serviços de armazenamento subcontratados a parceiros da rede de modo a aproximar a oferta da procura e, eventualmente, aumentar o controlo operacional.

Uma análise preparatória é efetuada em primeiro lugar através de uma revisão abrangente da literatura existente, do contexto do projeto e do raciocínio estratégico subjacente à iniciativa em causa. Os fatores determinantes na criação de valor e de custos foram identificados de um ponto de vista qualitativo.

Numa segunda fase, ilustra-se o uso de ferramentas matemáticas e computacionais a fim de permitir a quantificação das diversas rubricas de custo e de rendimento que contribuem para o valor do projeto. Enfatizou-se o uso de técnicas de *machine learning*, cuja integração em ambientes de simulação foi relativamente pouco explorada até à data. Além disso, este trabalho contribui para a literatura com o desenvolvimento de um modelo de retenção analítico.

Numa terceira fase, descreve-se a conceção de uma simulação híbrida — de eventos discretos e baseada em agentes — para alavancar os modelos e o conhecimento previamente desenvolvidos e, subsequentemente, gerar resultados para cenários hipotéticos. O objetivo principal da simulação foi o de obter uma estimativa preliminar do valor do projeto ao quantificar a relativa importância e o risco dos diferentes fatores que poderão, eventualmente, surgir ao longo da execução do projeto.

Finalmente, apresentam-se os resultados e as análises de sensibilidade iterativamente obtidas através do estudo de simulação. Conclui-se com uma avaliação da metodologia desenvolvida e com as recomendações principais relativamente ao valor do projeto de investimento, ao enquadramento estratégico e aos riscos subjacentes.

Acknowledgements

Throughout this project, many people somehow contributed to support my work. I would like to thank:

- Farfetch Operations Strategy team for their outstanding availability, support and friendship. In particular: Daniel Fernandes (my supervisor at Farfetch), as well as Mariana Guerreiro, Augusto Gonçalves, Carlos Gomes and João Ascenção for their particular contributions;
- The team from L. , who were open, extremely supportive and technically proficient. In particular, P.S., H.N. and J.A.;
- My supervisor, Professor Mário Amorim Lopes, who was available at all times, provided great revision and allowed me with great freedom and autonomy;
- My family and friends.

Contents

Acronyms and Symbols

CBP	U.S. Customs and Border Protection
CS	Customer Service
DES	Discrete-event simulation
KPI	Key Performance Indicator
MAE	Mean Average Error
RMSE	Root Mean Squared Error
SCM	Supply Chain Managgement
SD	System Dynamics
SOP	Standard Operating Procedures
SOS	Speed of Sending

List of Figures

List of Tables

Chapter 1

Introduction

According to M. Christopher (2004) and John Fernie (2009), the fashion industry is characterised by life cycles that have been getting shorter along the years, volatile trends, high impulse purchasing and sales difficult to predict. Yet, commercial success in the fashion industry, as referred by Martin Christopher (1992), is dependent on flexibility and responsiveness. These seemingly incompatible constraints and requirements give rise to some of the supply chain challenges faced today by modern organisations.

This thesis is based on a case study set in a fashion e-tailer and addresses an investment project within a supply chain management (SCM) initiative. The main question that is being answered is the following:

> *What is the value of providing outsourced warehousing services to key partners of the supply chain and what are the impacts regarding each one of the different stakeholders?*

The problem is complex, because it requires a strategic framing and assessment of the initiative within the company's internal and external environments, a qualitative understanding of both current and planned business processes and, finally, a financial quantification of the project's potential impact relative to an expected baseline performance. This work aims at delivering a preliminary answer to these questions by employing a holistic and multidisciplinary approach with the use of advanced modelling techniques. Even though the analysis is deeply focused on a particular case study, the main purpose is to illustrate the power and effectiveness of a combined use of such different tools in a practical context. The framework employed, though, can easily be generalised to other SCM problems.

1.1 Objectives

The objective of this project is to illustrate the use of a multidisciplinary approach to support decision making by assessing the risks, strategic adequacy and value captured by an e-tailer by providing warehousing services to partners — in particular, this work combines tools drawn from strategic thinking, mathematical modelling, machine learning and simulation, among others. Even though the methodology was applied with a specific case study in mind, the goal is for it to be easily extensible to other investment projects.

1.2 Methodology employed

Since the problem discussed is based on a Supply Chain Management issue that involves the quantification of cost and value drivers, as well as specific dynamics (such as the interaction between demand and distinct sources of supply with limited capacity and inventory), the problem itself is a complex one. As referred by Caroline Thierry (2006):

> Simulation seems the only recourse to model and analyze performances for such large-scale cases.

Moreover, Ingalls (1998) identified two major cases where simulation is adequate for supply chain problems and which presently apply: cases where analytical optimisation is too complex or where variance is an important driver. Hence, simulation was the main technique employed.

In order to represent with loyalty the distinct processes and agent behaviours, though, a modelling stage had to be undertaken prior to the development of the actual simulation solution. For the matter, distinct techniques were used when judged appropriate (including mathematical and machine learning tools, as previously referred) with the ultimate goal of gathering relevant business intelligence into a single place and subsequently replicate real world — or hypothesised — behaviours. As a final note, this methodology also has the benefit of generating actionable knowledge for other areas of the business, possibly combining collaborative work, due to the recurrent necessity of granular modelling and understanding of operating procedures.

1.3 Structure

The thesis is organised as follows:

Chapter 1 — Introduction briefly contextualises the problem, as well as the case study upon which this thesis is based on;

Chapter 2 — Literature review provides an overview of current practices and lack thereof regarding the different stages of the methodology employed;

Chapter 3 — Farfetch case study dives into the actual case study by providing an overview of existing processes, of the scenario under investigation, of potential value and cost driver involved and of the strategic motivation;

Chapter 4 — Modelling illustrates the use of different techniques to model cost and value items for the project, from a quantitative standpoint, and which will integrate the simulation solution;

Chapter 5 — Simulation describes the different components of the simulation and how they leverage the models developed to generate actionable results;

Chapter 6 — Results and discussion exposes the outputs obtained from the simulation and what they reveal regarding the case study;

Chapter 7 — Conclusion summarises the main insights gathered along the project regarding the techniques employed and formulates final recommendations about the investment project.

Chapter 2

Literature review

2.1 Supply chain management

There are many references to the concept of supply chain management (SCM) in the literature, with slightly different definitions or interpretations, according to John T. Mentze (2001). In an article entitled *Defining Supply Chain Management*, the same authors provided the following definition, after a thorough review of the literature:

> "(...) supply chain management is defined as the systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole."

Regarding in particular the purpose of SCM, Stevens (1989) provided a complementary definition which is linked to the customer:

> "The objective of managing the supply chain is to synchronize the requirements of the customer with the flow of materials from suppliers in order to effect a balance between what are often seen as conflicting goals of high customer service, low inventory management, and low unit cost."

The conflicting goals are known as the *supply chain triangle*, depicted in Figure 2.1:

In SCM, service usually measures the performance of inventory replenishment policies, lead times, payment options, and, overall, metrics somehow related to the customer. The lower the stockout rate is, for example, the higher the service level is. Stockouts can be reduced by increasing inventory levels; this increase in service level resulting from an investment in working capital (cash) constitutes one of the trade-offs. On the other hand, increasing service level by improving the efficiency of the process may result in higher unitary costs. Traditionally, these are some of the challenges that modern organisations face when managing the supply chain.

Michael Treacy and Fred Wiersema proposed what they called *"three paths to market leadership"* in (Michael Treacy, 1992) and which consist in the following strategies: customer intimacy,

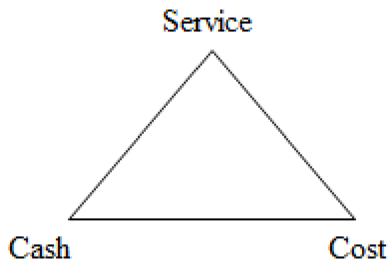

Figure 2.1: Supply Chain triangle

operational excellence and product leadership. As treated in (Desmet, 2018), these paths can be mapped somewhere on the supply chain triangle depending on each case; for example: customer intimacy is a strategy that favours high service, and thus usually requires more expensive processes to guarantee adequate quality and a higher working capital for product availability; operational excellence leads to efficient processes, and thus is a strategy linked to lower unitary costs, but also to relatively lower service (such as customisation potential due to less flexible processes) and higher working capital (for example due to make-to-stock (MTS) production); finally, product leadership focuses on delivering the best version in the market, and thus may require more expensive processes and may omit service features valued by the customer, such as low lead times or the best customer service.

According to Matthew Liberatore (2010), the supply chain strategy should depend on the characteristics of the product and market served. Hence the ideal supply chain strategy to follow with an innovative type of product likely differs from the one to follow with a functional one, given the classification found in (Fisher, 1997). An example of this principle, deeply related to the case study treated in this thesis, lies in the differences between luxury and low-cost fashion, both in terms of product and market. Even though they can both be seen, arguably, as innovative products up to a certain extent, their differing degrees and markets targeted result in different strategies: luxury fashion may require a customer intimacy strategy (with potentially higher inventories and more value-oriented processes) as opposed to low-cost fashion, where operational excellence may be more adequate in order to guarantee low prices.

2.2 Distribution decisions

The selection of an adequate distribution network is deeply linked to the supply chain strategy and decisive for success. Chopra (2003) reported the following:

> "Distribution is a key driver of the overall profitability of a firm because it directly impacts both the supply chain cost and the customer experience. Good distribution can be used to achieve a variety of supply chain objectives ranging from low cost to high responsiveness. As a result, companies in the same industry often select very different distribution networks."

Therefore, the configuration of warehousing facilities are of extreme importance both in terms of supply chain cost and to the customer experience — which, as previously seen, may be conflicting goals. According to E. Feldman (1966): "optimal sizing and locating of facilities are very sensitive to the shapes of the warehousing cost functions", where the cost functions depend on factors such as capacity or number of facilities under economies of scale, as referred by Chopra (2003). On the other hand, Erdem Eskigun (2005) reported that cost minimisation is often seen as the main focus, instead of solutions which explicitly take customer metrics into account, such as lead times and service level. Besides that, vehicle routing and inventory decisions are often secondary, despite their potential impact on customer experience. According to Mark S. Daskin (2005), this usually happens because these decisions are more flexible and less expensive than actual sizing and location of facilities, despite empirical evidence that suggests the superiority of a holistic optimisation approach. Hence, one key takeaway is that customer-oriented metrics should be important decision drivers when assessing distribution strategies within a supply chain management context, especially if the market and product favour a customer intimacy strategy.

2.3 Value assessment models

As referred by Beamon (1998), models for supply chain design and analysis can be divided into four categories, namely: deterministic analytical models, stochastic analytical models, economic models and simulation models.

Deterministic analytical models are usually employed to represent systems which do not possess considerable uncertainty associated. They constitute a simplified abstraction of real world processes. Usually, as stated by Buckley (2005), these kinds of model are limited due to simplifying assumptions and rigid structuring requirements. On the other hand, they have the advantage of providing explicit information regarding the effect of different variables and parameters, as opposed to other types of models that may be perceived as black boxes.

If uncertainty plays a role in the problem at hand, M.T. Melo (2009) suggests that stochastic models may be more adequate. These models can be obtained by introducing random variables or scenario with different probabilities.

Economic models are more rare in the context of supply chain management and, usually, relate to strategic and high scale macro-applications, such as in (D.P. Christy, 1994). These models may involve the analysis of distinct scenarios in a game-theoretic approach.

Finally, simulation models may more easily represent real world processes without the required simplifications and limiting assumptions of analytical models, as reported by Virginia L. M. Spiegler (2016). However, such flexibility usually requires more computational power, especially for iterative optimisation purposes.

2.4 Simulation and supply chain management

According to Y. Chang (2001):

> "Discrete-event simulation allows the evaluation of operating performance prior to the implementation of a system since: (a) it enables companies to perform powerful what-if analyses leading them to better planning decisions; (b) it permits the comparison of various operational alternatives without interrupting the real system and (c) it permits time compression so that timely policy decisions can be made."

These capabilities generally make simulation an adequate choice for large SCM problems, especially when combined with other techniques for modelling or optimisation.

As an example of simulation combined with other tools, June Young Jung (2004) applied simulation to model safety stock levels in the chemical process industry. According to the authors, the use of simulation enables the introduction of the uncertainty and complexity that is not adequately tackled by traditional inventory theory. Concerning the actual solution architecture, the authors used the following components for the simulation: a computation control module, stochastic models that represented the simulated processes and a data module. Moreover, simulation was used with an optimisation task in mind, based on a methodology exposed in (Subramanian, 2001) which tackled a computational solution for the Research and Development Pipeline management problem. The approach for the latter was referred to as *Sim-Opt* and was based on a two-layer framework which combined an analytical and a simulation model to optimise a given output variable (net present value); the first layer was a discrete event simulation, whereas the second was a gradient based heuristic that optimised the simulation's input parameters so as to maximise the output. This approach, at the cost of computational power, delivers the optimisation potential of anlytical tools without sacrificing the flexibility, uncertainty considerations and powerful modelling capabilities of simulation tools.

In the literature, hybrid approaches combining simulation with other techniques have been proposed. in (L. Rabelo, 2005), for example, System Dynamics (SD) and Discrete Event Simulation (DES) were combined, with SD being used as an overall enterprise system model and DES employed to represent manufacturing and more operational tasks. This approach is also discussed in (Kirandeep Chahal, 2008), with focus on UK healthcare. Later on, Luis Rabelo (2006) extended this framework by introducing an analytical hierarchy process (AHP) component. Z. Xu (2001) suggested the use of learning agents in agent-based models (ABM). Machine learning can also be used outside of the simulation itself, for parameter optimisation or efficiency purposes, as in (Walid Budgaga, 2015). These examples show that simulation techniques are in fact flexible and can be adapted into powerful modelling methodologies by combining them with the adequate tools.

The term *hybrid* can also refer to a solution that combines distinct simulation frameworks. As an example, Qi Hao (2007) developed a simulation that combined DES with ABM to represent material handling processes in an assembly line. The literature suggests that it is possible to build hybrid models by integrating different types of simulation frameworks and possibly combining them with distinct tools borrowed other fields — namely: statistics, machine learning, decision theory, and so on.

2.5 Simulation and machine learning

Sharma (2015) identified the following steps to guide a simulation project:

1. problem formulation;
2. set objectives and plan;
3. build a conceptual model;
4. collect data;
5. create the simulation model;
6. experimental design;
7. production runs and analysis;
8. documentation and report.

One of the most important steps of a simulation project is data collection, which will eventually determine the values taken by the simulation variables, such as a cost or a probability of occurrence of a given event for example. These values can be either deterministic or stochastic: in the former case, the variable will always have the same value under identical circumstances, whereas in the latter, the variable will have a random value obtained from a given probability distribution. In both cases, though, the value is a prediction of what would happen in reality if the circumstances at the moment of the simulation were to occur. The data that serves as basis is obtained from this data collection step.

When it comes to predicting reality under specific conditions, there are different techniques available. Sometimes, it is just a matter of considering an exact value — such as a fixed cost obtained from a third-party logistics provider —, whereas, other times, statistical tools may be used instead — such as regression analysis or using empirical distributions. With the recent advances in technology, though, machine learning may prove to be a great alternative to predict reality, especially when large amounts of data are available and when performance measurements suggest superiority over other models.

Nevertheless, despite the existing motivation, there seems to be a gap in the literature regarding the applications and use of machine learning techniques to interface the data collected and the simulation model — in fact, no reference was found during research for the thesis. There are, however, examples where machine learning techniques have been used to generate insights from simulation outputs or to emulate artificial intelligence agents within the simulation's universe; for instance: in an article called *Simulation assisted machine learning* by Timo Deist (2018), the authors refer the following:

> "In addition to biology and medicine, this approach [using machine learning on simulation generated data] should be applicable to other disciplines, such as weather forecasting, financial markets, and agricultural management, where predictive models are sought and informative yet approximate simulations are available."

The article itself is very recent and, according to the same authors, this somewhat related approach also lacks treatment in the literature:

> "(...) we propose a method to integrate simulations, a strong form of prior knowledge, into machine learning, a combination which to date has been largely unexplored."

In this thesis, the use of machine learning techniques to learn from the data and support the simulation with more accurate predictions will be explored given the modelling potential.

Chapter 3

Farfetch case study

In this chapter, the company will be introduced, as well as the case study that serves as a basis for the thesis. A preliminary and qualitative assessment is also provided in order to frame the problem and guide subsequent modelling.

3.1 Business and processes overview

Farfetch, funded in 2008 by Portuguese entrepreneur José Neves, is a leading online marketplace that operates in the luxury fashion industry. With more than 500 boutiques distributed over 190 countries, the e-tailer links final customers to real brick and mortar stores through multiple channels, namely: portal, mobile and phone.

3.1.1 The order cycle

The order cycle begins when the customer places an order and, granted that no exception occurs, ends when the item is received or returned. Actual order fulfilment encompasses six major steps, which will be explained further on:

1. Stock and fraud validation;
2. Payment approval;
3. Packaging;
4. Labelling;
5. Dispatch;
6. Transit;

Different agents play a role during the order cycle, namely: the customer who places an order; the partner who fulfils the order; multiple departments within Farfetch's organization who support the process and, finally, the courier who is responsible for delivering the product.

Internally, an order is divided into two major parts: sending and transit. The former is measured by Speed of Sending (SoS) — which should fall somewhere between two business days, as stated publicly in Farfetch's portal (Farfetch.com, 2018) — whereas the latter depends on the customer's shipping choice and will be addressed later on.

3.1.1.1 Order placement

Available products are displayed in Farfetch's portal. The customer can place an order either through the portal itself, through a mobile device or, alternatively, by phone. The product mix, prices and delivery options displayed in Farfetch's portal depend on the customer's geographical location. Additionally, certain products follow geographical pricing policies imposed by brands. Due to this, different effects may arise, for instance:

- A customer may witness a specific price for a given item and size, but when selecting another size from the available range can suddenly see a distinct price appear. This may happen because Farfetch's system allocates the item to another boutique which is selling the same product at a different price;
- For a given item, a specific price is displayed to a customer in a certain region, but to another customer in a distinct location another price is shown instead. This effect can occur due to geographical pricing policies or because Farfetch's system allocates the item to stores selling at different prices in order to minimise costs;
- Some products are not displayed to customers in specific regions due to legal constraints – such as fur products, for instance. The customer's location is estimated based on information provided by the browser or by explicitly selecting a region and currency through Farfetch's portal. Nevertheless, actual location is obtained from the address written by the customer during the final checkout phase.

A customer can order simultaneously more than one product, which may, or may not, come from the same boutique. In these cases, this gives origin to a single portal order and to multiple boutique orders (Figure 3.1 illustrates the process). Farfetch does not consolidate orders, which means that these will most likely be delivered in different dates. In fact, boutique orders are processed independently and, due to it being so, shipping costs are also charged separately for each of the former orders —- yet, up to a certain threshold, shipping costs may be shared between the customer and Farfetch (referred to as flat rate).

Finally, it is during this step that the customer chooses one shipping option, which also depends on location. Most Farfetch orders are delivered Monday to Friday from 9am to 5pm. The cost invoiced by the courier to Farfetch is a function of service type, shipped size, shipped weight and shipping route, whereas to the customer it is an estimate calculated by Farfetch (which accounts for flat rates, for example). Available services are:

DHL® Express Delivers to most of Europe and USA in 2 to 4 days and to the rest of the world in 3 to 7 days;

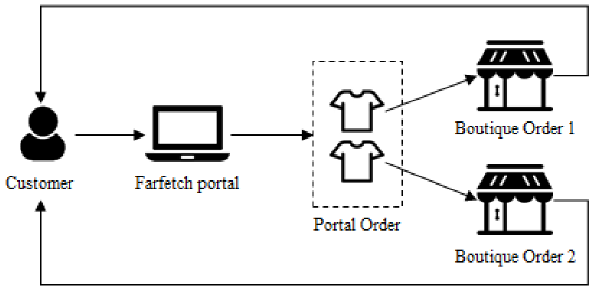

Figure 3.1: Example: one portal order giving rise to two boutique orders

UPS® Standard Delivers to Europe only in 5 to 7 days;

UPS® Express Saver Delivers to most of Europe and USA in 2 to 4 days and to the rest of the world in 3 to 7 days;

Sameday Orders submitted until 11am are delivered before 7pm. Available in Berlin, London, Paris, Los Angeles, New York, Miami, Milan, Rome, Barcelona, Madrid, Dubai and Hong Kong;

F90 Orders are delivered in 90 minutes. Available for some products in Berlin, Dubai, Hong Kong, LA, London, Madrid, Miami, Milan, New York, Paris, São Paulo and Tokyo;

Sameday and F90 services are extremely exigent and highly dependent on boutique performance and location, which are highly uncontrollable factors. Hence, these services would highly benefit from a higher operational control by Farfetch and more efficient fulfilment.

3.1.1.2 Step 1: check stock

Only products which have stock available are displayed in Farfetch's portal. Stock information is provided by the boutiques on products which are being sold through Farfetch and can happen by two different means: either in an automated and synchronous way, or manually. Due to the potential lag inherently existent in the process of having to manually update stock entries, it may happen that a product which is accepted as having stock in store is, in fact, not available anymore. This situation is detected in step 1 of the order fulfilment process, which is aimed at partners checking stock availability.

When a stock-out exception arises, an item-swap sub-process is then triggered and consists in the following: the original boutique order is cancelled and a child order for the same item is created, but in a different boutique where stock is shown as available in the system. In these cases, Farfetch covers potential price differences up to a certain level. If there is no boutique with available stock, or a stock-out is triggered by the second boutique, the order is cancelled and the customer is notified, refunded, and compensated.

Stock-outs are undesirable exceptions for Farfetch, since they increase costs due to Customer Service (CS) efforts, price differences, customer compensation initiatives and, last but not least,

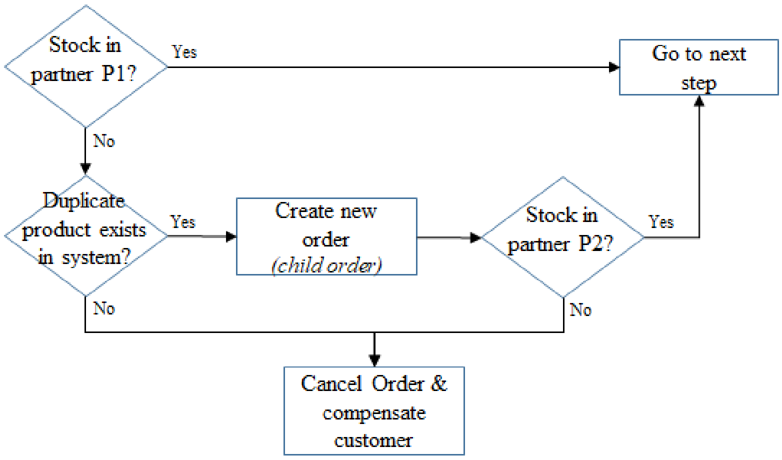

Figure 3.2: Step 1 No-stock exception

undermine the company's perceived value. Because of that, Farfetch favours synchronous exchanges of inventory information and, overall, initiatives which positively contribute to reducing stock-outs.

3.1.1.3 Step 2: payment approval

Step 2, which aims at approving payment, happens simultaneously with step 1. It encompasses two major stages: the first one responsible for fraud validation and the second one responsible for actually approving payment and issuing a proof of billing. Once the order goes through both stages, it then enters the next step.

3.1.1.4 Step 3: packaging

At the time of writing, Farfetch has an available assortment of nine distinct boxes to carry products sent by boutiques to the final customer. Actual packaging is performed by boutiques, and for each product Farfetch recommends a specific box. Sometimes, these recommendations are not followed by boutiques and, due to this, the amount charged by the courier — which depends on the volume or weight being shipped — is higher than expected, and thus becomes an increased cost to Farfetch. Incorrect box usage can be explained because of boutique negligence, box stock-outs or inadequate box recommendations provided by Farfetch; nevertheless, concerning the former, Farfetch offers financial incentives to boutiques with positive performance when it comes to the rate of adequate box usage.

3.1.1.5 Step 4: labelling

During this step, the parcel to be delivered is labelled with the air waybill. Even though the process is mainly automatic, some addresses trigger printing errors — for example due to character incompatibilities — and require manual correction. On the other hand, legal constraints, such as compulsory importation limits, may also block the process. This kind of exceptions may result in higher lead times and could be avoided, up to some extent, with a more efficientt and centralised process.

3.1.1.6 Step 5: Dispatch

Once the final package is ready to be shipped, it then awaits courier collection. The customer is notified as soon as the order is dispatched. Most cross-border deliveries are carried out by DHL®, whereas UPS® takes care of most U.S. domestic orders. Customers are able to track their orders through Farfetch's portal or by reaching the Customer Service (CS) department for a delivery status request. As a general rule, the more detailed and accurate the information provided on the portal is, the less CS tickets are created by the customer, thus reducing costs. However, this requires great visibility over the order status, as well as process and information integration.

During transit, some exceptions can occur — such as goods being held in customs or address changes, for example — which ultimately result in increased lead times or even failed delivery. These effects, in addition to bringing in costs, may also undermine customer retention.

3.1.2 Pricing

Because pricing is a sensitive topic, only a broad overview will be provided in the following section. In practice, however, a detailed mapping of existing fluxes had to be carried out.

Farfetch receives a commission for every order not returned. This is one of the two main sources of income — the other one is geomargin, which will be explained down below.

Because Farfetch serves different regions around the globe, the business is subject to currency fluctuations. Therefore, Farfetch also inflates prices charged to customers with a small margin in order to compensate for adverse fluctuations, in this way hedging against exchange rate validity.

Besides that, because Farfetch has to display a price to the customer before the order is actually fulfilled, there are differences between estimated costs (with shipping or taxes, for example) when compared the actual costs incurred by Farfetch.

Farfetch gathers supply from around the world into a single platform which serves a global market. Yet, because of local factors such as consumer buying power, tax incidence and so on, retail prices for an identical product can vary from region to region. In other words: a brand in the United States could be selling a product at a given price and, simultaneously, a boutique in Italy could be selling the exact same item, but at a lower local price — or vice-versa. By bringing every boutique and brand to the same global market place, such price divergences could create competition between brands and boutiques due to the existence of different retail prices for identical articles. Due to this, many brands impose a geographical pricing policy. Geographical pricing (referred to as geopricing at Farfetch) is when the final customer only perceives one single price according to his or her location.

Since geopriced products can only have one price for a given region, costs paid by the customer — except shipping costs, which are not directly related to the product itself and the value perceived — are included in this final price. A simplified diagram is depicted in Figure 3.3.

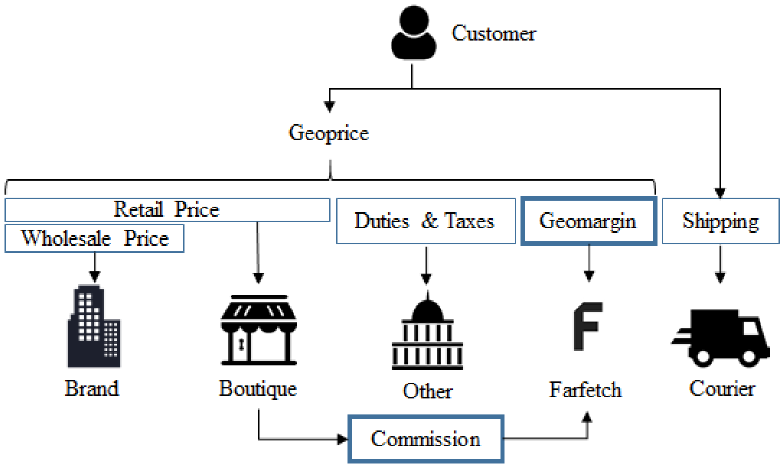

Figure 3.3: Geopricing simplified overview (box size non-representative)

Because geoprices tend to be higher than retail prices, even after deducting the costs with duties, Farfetch is entitled to the margin (geomargin) that results.

The boutique can practice markdowns — in which case both the retail price and geoprice lower by the markdown value — and Farfetch can offer discounts — in which case it can be seen as a cost which shrinks the geomargin.

Finally, non-geopriced products follow an identical pricing scheme with the only difference that no geomargin exists.

3.1.3 Routing algorithm

Because the same exact item can exist in different boutiques around the world, there is a routing algorithm which is responsible for allocating an order to a given boutique. The criteria followed, by decreasing order of priority, are the following (if there are ties, the algorithm advances to the next consecutive criteria):

1. Select the boutique(s) with lowest retail price (5% tolerance range);
2. Select the boutique(s) from the same country;
3. Select the brand(s) from the same country;
4. Select the boutique(s) from the same region (continent);
5. Select the boutique(s) with the highest performance;
6. Select the boutique(s) with the oldest product;
7. Select the oldest boutique;

The algorithm also allows the use of specific rules such as: orders from the United States which can be fulfilled in a certain boutique are always allocated to the latter. In the same fashion, the algorithm also allows blocking certain routes if required.

3.1.4 Reverse logistics

Because the customer cannot see or try the product before acquiring it, returns are free at Farfetch. This means that the customer will not pay for the cost of shipping back and will receive a full refund.

Depending on the reason for the return, the shipping cost may either be supported by Farfetch or, in alternative, by the partner. For instance, if the return is due to the partner's mistake — such as wrong item or defective product shipped — the partner is responsible for the refund, whereas if the return is due to reasons related to fitting, customer preference and alike, Farfetch is responsible.

In practice, however, the partner may not accept the returned product and refuse to refund the customer. From these cases result negotiations between Farfetch and the partner. The final result is a recovery rate, which measures the ratio of what Farfetch recovers from partners on returned orders. The more partners have control and authority over returns, the inferior Farfetch's recovery rate tends to be.

3.1.5 Packaging

Farfetch outsources package manufacturing. Partners are responsible for ordering boxes from Farfetch and managing their packaging inventory.

There are two major costs seen by Farfetch concerning packaging: acquisition and shipping. Regarding partners, Farfetch charges a fee per order which is associated to packaging costs and weights both acquisition and shipping costs on average.

In what concerns packaging efficiency — which is directly related to shipping costs —, Farfetch recommends an ideal box for each product, yet partners do not always follow the recommendations. Because the customer pays a shipping rate estimated by Farfetch based on factors that consider the recommended packaging volume and Farfetch pays the shipping costs invoiced by the courier based on the package actually shipped, this usually leads to financial losses. Arguably, a warehouse would allow greater operational control, and ultimately minimise these divergences.

3.1.6 Net promoter score

In order to assess customer satisfaction, Farfetch keeps track of a particular indicator: the net promoter score (NPS). Based on a survey, customers are classified either as being promoters, passives or detractors based on their answers and, finally, NPS is obtained through Equation 3.1.

$$NPS = \frac{promoters - detractors}{\sum(promoters + passives + detractors)} \tag{3.1}$$

Classification in one of the three groups is obtained through weighting answers to a survey which assesses factors such as packaging quality and speed. Despite measuring the proportion of promoters or detractors, this indicator can also be leveraged to get a grasp of the performance of a given partner, the sensitivity that certain customer segments have regarding specific factors, and finally establish relationships with retention. These relationships are extremely important and will be used in conjunction with other factors to model the effect of the investment project on retention.

3.1.7 Financial support and incentives

Because fashion is highly seasonal, the business is subject to peak periods. This is an important effect for boutiques that operate traditionally with an offline channel; however, it becomes much more relevant for boutiques that partner with Farfetch. In fact, some of the former have seen such a tremendous growth during the last years due to Farfetch that they struggle to maintain the pace. Due to this, and because partners are essential for Farfetch, financial labour support incentives are provided to key boutiques in order to aid with peaks.

Besides that, Farfetch also offers incentives in order to promote higher policy compliance. For instance, partners who have low no-stock rates, fast Speed of Sending scores or good packaging ratings receive financial incentives.

3.2 Warehousing as a service at Farfetch

Farfetch is a marketplace that links final customers to brands or boutiques, and thus operates based on the principle of no inventory. Due to this, inventory risks, which are relevant in the luxury fashion industry due to the high volumes, relative values and depreciation rates of products, are incurred by Farfetch's partners. In fact, the difference between Farfetch's and other more traditional e-commerce retailers' business models is that the latter incur in inventory holding costs and are systematically subject to obsoletion, whereas Farfetch dilutes these costs and risks among partners. For instance, because partners are the stakeholders which actually hold the inventory — and subsequent costs —, Farfetch is able to market a massive product assortment which would be excessively risky, costly and challenging to hold and manage by a single party, especially given the relatively high seasonality effects. Besides that, this same business model still leaves the boutiques' offline channels open. At first, this might seem to be irrelevant for Farfetch, or for that matter, counterproductive; yet, it helps to keep partners in the business. By allowing an offline channel, it minimises their inventory risks by increasing sell-through, especially of items which do not show adequate online performance, minimises treasury risks by providing cash inflows during the season, and finally constitutes a solid negotiation argument for Farfetch, since it gives the assurance to partners that if the online business was not to perform as expected, the offline alternative is still available, and thus business with Farfetch is always a dominant strategy with a backdoor available.

Despite this, Farfetch is now considering the scenario of providing warehousing services to partners. This scenario is explained in the upcoming subsection.

3.2.1 Main scenario

First off, Farfetch would start a negotiation process in order to determine and establish which partners would join the initiative. This process is delicate, because contractual terms and conditions should be clearly set prior to the undertaking of such enterprise, and given the uncertainty frame, it could result in critical moments along the way. For instance: because there are fixed costs, it is important that partners remain on-board throughout the season in order to guarantee at least break-even, but if sell-through does not meet one of the partners' expectations along-way, this same

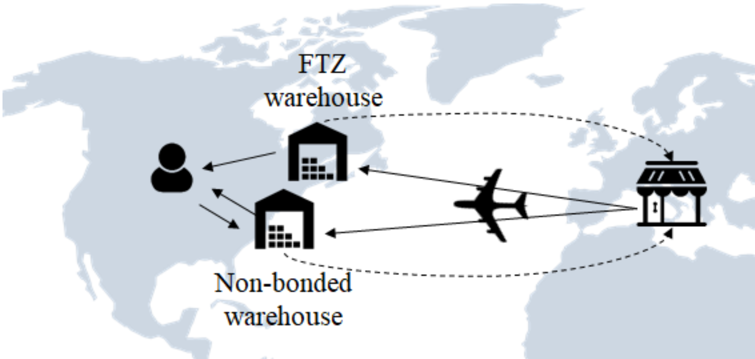

Figure 3.4: Merchandise fluxes

partner might try to reclaim the inventory back, possibly resulting in a subsequent dissemination of negative information through the network, perhaps giving rise to further of such incidences with other partners, undermining Farfetch's reputation and existing partnerships. Such possible occurrences should be anticipated, and that is why timely and conclusive negotiations about dedicated stock periods, minimum sell-through and other terms are crucial. This issue will be addressed later on.

At first, according to the scenario outlined by Farfetch, warehousing would not be charged to partners so as to increase adherence and eventually help prove the concept. On later stages, though, warehousing could potentially become a service charged by Farfetch. One of the goals of assessing the value of the project is, precisely, to understand the potential for pricing.

Once partners are selected and terms negotiated, a product selection phase begins. This second stage is a cooperative effort which involves both Farfetch and the partners. The reasons why it is a cooperative effort is due to the synergy that arises: on the one hand, partners hold the inventory — and, thus, the ultimate decision —, whereas, on the other hand, Farfetch possesses key knowledge about the market — Farfetch offers a massive product assortment worldwide, which inevitably provides a much higher visibility over the market than the one that singular partners possess by themselves. Finally, an adequate product selection is critical in order to guarantee a minimum sell-through, and eventually build business cases that help roll-out and scale the project subsequent to a proof-of-concept pilot.

After products have been selected and the beginning of the season has been planned, products are shipped overseas via air freight. Not all products might be shipped right away; in fact, contractual terms establish how many weekly shipments Farfetch offers per boutique. Besides that, some products may only be produced (made available online) after the season has already begun, and thus become available online later on — this may be an undesirable scenario, since it lowers the chance of selling at a higher full-price mix.

One important factor to consider is that the inventory is owned by partners at all times and not by Farfetch, since the latter only provides the warehousing service. Furthermore, the actual infrastructure, management and operation of the warehouses is outsourced to a third-party logistics provider (3PL).

The scenario being analysed at Farfetch and which serves as the basis for this study considers two warehouses that would potentially be located in New Jersey, United States, to be used in conjunction, and would mainly serve the North American Free Trade Agreement region (NAFTA) — even though they might be opened to worldwide orders if technically possible and justified from an economical standpoint. One of the warehouses would be in a free-trade zone, whereas the other one would be a non-bonded warehouse. The main differences between both types essentially lie on the moment in which dutiable goods are cleared and the costs inherently associated. A more detailed overview about the differences will be tackled in upcoming sections.

During early phases, only geographically priced products would be selected due to technical constraints. Moreover, because of tax reasons, products with a selling price below the United States' *de minimis* exemption threshold — $800 at the moment of writing, according to U.S. Customs and Border Protection (2018) — would be shipped to the non-bonded warehouse, whereas products priced above the threshold would be allocated to the free-trade zone warehouse. This will also be discussed later on.

Concerning reverse logistics, orders being returned (which would have already been duty and tax cleared) would be assigned to the non-bonded warehouse, even if originally dispatched from the free-trade zone warehouse.

Finally, there would also be a contractually predefined period after which partners may reclaim their stock back. Alternatively, if reclaimed during the aforementioned period, they would be required to pay a fee.

3.2.2 Free-trade zone warehouse

Over the years, perishability in the fashion industry has become increasingly important, as mentioned by Sang-Eun Byu (2008); yet, on the other hand, Pashigian (1988) remarked thirty years ago that markdowns have a growing tendency as product variety increases. Nowadays, these phenomenons are clearly evidenced, in particular within the present context: around 75% of apparel purchases were made at discounted prices in 2016, as reported by NPD (2016). Due to these rising inventory risks, it is important to minimise as much as possible the negative effect of perishability.

Farfetch does not directly possess inventory risks, since it is the partner that owns the inventory at all times. As long as the current business model is followed, Farfetch's main preoccupations concerning the possibility of unsold — or salvaged — products is the negative impact that it has on partners and sunk production costs (production costs refer to costs related to making a product available online).

The situation changes when considering a warehouse in the United States. If Farfetch moves stock to a potential warehouse in the U.S. by following current procedures, the U.S. Customs and Border Protection (CBP henceforth) requires duty payment on products imported. This means that unsold products at the end of the season, besides being highly depreciated, also have duty costs associated. In this case, Farfetch could either try to salvage products in the domestic market — which requires a special agreement with the partner, since the latter owns the inventory and might require it back — or, in alternative, open the warehouse to the rest of the world but risk additional costs with duties for cross-border demand.

According to U.S. Customs and Border Protection (2018a), free-trade zones (also known as foreign-trade zones in the United States) are areas in which foreign and domestic merchandise can be moved without the requirement of the usual formal CBP entry procedures and duty payments; actually, these can take place only when merchandise enters domestic territory for consumption. This is why Farfetch is considering a scenario with a warehouse in a free-trade zone: because products that are not sold or that are open to the global market do not have a sunk duty cost. Besides that, this can also be seen as a duty deferral mechanism that Farfetch can leverage on and capitalise as working capital, if required (this possibility, though, is not as relevant, in particular because of Farfetch's business model and reach that generates a healthy stream of cash-flows).

Regarding other benefits, CBP also offers theft protection — which might materialise in important savings on insurance premiums according to Teifenbrun (2015). Finally, there are the weekly entry procedures which consist in the following: CBP does not process an entry for each import shipment — which is the case if products were to be imported into a traditional domestic warehouse — but, rather, processes the accounting of imported products into a single entry filing per week. According to FTZC (2018), because a Merchandise Processing Fee (MPF) is charged for every formal customs entry filed, weekly entries contribute to savings due to the reduction in frequency. At the moment of writing, and as stated by U.S. Customs and Border Protection (2018b), MPF is an *ad valorem* fee of 0,3464% which is capped at $MPF_{max} = \$485$. Thus, if $r_{MPF} = 0.3464\%$ and $v_n \geq 0, n \in N$ represents the value of each order n imported, then the total MPF charged for the baseline scenario is given by equation 3.2:

$$MPF_{baseline} = r_{MPF} \sum_n \min\{v_n, \frac{MPF_{max}}{r_{MPF}}\} \tag{3.2}$$

Yet, this is only charged on formal entries, which is a rare — or even nonexistent — phenomenon in the baseline scenario due to the fact that Farfetch ships separated individual orders. In practice, $MPF_{baseline} = 0$ for the vast majority of times.

When it comes to MPF charged through weekly entry procedures, the calculation is given by Equation 3.3:

$$MPF = r_{MPF} \min\{\sum_n v_n, \frac{MPF_{max}}{r_{MPF}}\} \tag{3.3}$$

Since $\min\{\sum_n v_n, \frac{MPF_{max}}{r_{MPF}}\} \leq \sum_n \min\{v_n, \frac{MPF_{max}}{r_{MPF}}\}$, the free-trade zone scenario is a dominant strategy concerning MPF.

On the other hand, the United States — and many other countries — also have a *de minimis* threshold. According to GEA (2016), *de minimis* refers to the minimum value of goods being imported below which no duties and taxes are collected (including MPF). Weekly entry filings, in this case, have a negative impact. For instance, if r_n is the duty rate applicable to order n, duties paid in the baseline scenario are given by Equation 3.4:

$$duties_{baseline} = \sum_n d_n \quad \text{with} \quad d_n = r_n \begin{cases} 0 & v_n < minimis \\ v_n & minimis \leq v_n \end{cases} \tag{3.4}$$

Whereas duties with weekly entries is given by Equation 3.5:

$$duties = \begin{cases} 0 & \sum_n v_n < minimis \\ \sum_n r_n v_n & minimis \leq \sum_n v_n \end{cases} \tag{3.5}$$

In this case, the baseline situation dominates, since $\sum_n v_n \geq v_k, \forall k \in N$.

Therefore, because of weekly entry filings, the free-trade zone warehouse provides benefits when orders are valued above the MPF cap, but have a negative impact when orders are valued below the *de minimis* threshold — which, at the time of writing, is set at $800 for the United States, according to U.S. Customs and Border Protection (2018).

Finally, because products are imported directly for domestic consumption, the former are valued at the final selling price, which serves as basis for duty calculation.

3.2.3 Non-bonded warehouse

The non-bonded warehouse would be operated and managed by the same third-party logistics provider. The difference when compared to the free-trade zone warehouse is that the non-bonded warehouse is within CBP border, and thus goods that are stored need to be customs cleared. Besides that, there is no weekly entry procedure. These constitute disadvantages in comparison to the free-trade zone warehouse.

On the other hand, because the goods are being imported to be stored in the United States and sold opportunely, goods are valued at the wholesale price, which serves as basis for duty calculation and is lower than the final selling price. This provides an advantage to the non-bonded warehouse in comparison to the free-trade zone warehouse.

Concerning the calculation, duties depend, in this case, on the value of the cargo shipment. If $n \in \mathcal{N}$ designates an item, $i \in \mathcal{I}$ a bulk shipment and $w_{i,n}$ the wholesale price of item n in shipment i, duties for shipment i are given by Equation 3.6:

$$duties_i = \begin{cases} 0 & \sum_n w_{i,n} < minimis \\ \sum_n r_n w_{i,n} & minimis \leq \sum_n w_{i,n} \end{cases} \quad \forall i \in \mathcal{I} \tag{3.6}$$

MPF calculation is analogous to the free-trade zone warehouse, but takes as basis the wholesale prices (refer to Equation 3.7). Likewise, it is only charged if the bulk shipment is valued above the *de minimis* threshold.

$$MPF_i = r_{MPF} \min\{\sum_n w_{i,n}, \frac{MPF_{max}}{r_{MPF}}\} \quad \forall i \in \mathcal{I} \tag{3.7}$$

3.2.4 Reverse logistics

Items that are returned from the United States have already been imported into the domestic market. If returned to another country, there is a risk of repaying taxes while entering — unless special protocols are followed, such as a return via the United Kingdom for Italian partners, in detriment of transit times. Even then, around 40% of historical U.S. orders that had been returned to Europe ended up being shipped again to the U.S., and thus suffered retaxation.

Due to these reasons, and besides supporting operations, the non-bonded warehouse is also targeted at receiving and keeping domestic returns until new orders appear. This flux was portrayed in section 3.4.

If the non-bonded warehouse is open to the rest of the world, though, duties may be charged upon entry at the destination. The same principle applies — excluding few exceptions such as the previously mentioned U.K. detour — if partners reclaim their inventory back, which is expected at the end of season.

3.3 Qualitative assessment

This section provides a brief description of the main value and cost rubrics that need to be modelled during this project.

The main drivers result from increased operational control and proximity to the customer, conveyed by the warehouses. Due to these factors, drivers such as higher economies of scale, lower lead times, lower shipping distances and a higher efficiency might result in increased value and lower costs for the supply chain.

3.3.1 Cost drivers

Most of the changes in costs result from a transfer of ownership between different stakeholders of the supply chain. No pricing policy concerning the warehousing service is considered, since this is very dependent on Management decision — in fact, supply chain savings might be equally distributed between partner and Farfetch, and thus not justify the need for any pricing. This project will allow to understand how the value is distributed between the different stakeholders. Table 3.1 contains a summary of the analysis.

Packaging Packaging used to enclose the products are outsourced and originally financed by Farfetch. On demand, these boxes are shipped to partners — which store them — and a fee is charged by Farfetch on the ones that are used. The fee is an estimate of the average cost of box acquisition, handling and shipping;

Within the warehouse scenario, boxes would be stored in the warehouse. Consequently, there would be a transference of cost from partners to Farfetch — in terms of space and fee per box;

Storage Originally, partners hold the inventory, and therefore incur in costs associated to storage. Within the warehouse scenario, these costs would be assumed by Farfetch;

Bulk handling The cost of shipping and receiving products in great quantities is a new cost that would arise from having a warehouse and would apply to both partner and Farfetch: partners would have to prepare the merchandise in order to be collected and shipped to the warehouse, whereas Farfetch would be charged by the third-party logistics provider for the shipment inbound. An analogous principle applies to inventory that is reclaimed by partners;

Bulk shipping The flux of merchandise from partners to warehouses is new, and thus constitutes additional costs that did not exist before. Farfetch's policy concerning the warehousing scenario establishes that costs would be assumed by Farfetch for up to two weekly shipments to the warehouse and a return shipment at the end of the season with the products that did not sell. Shipments that constitute exceptions to this policy would be financed by partners;

Order handling This cost would be transferred from partners to Farfetch, since the former would not handle the orders for products that are in the warehouse;

Returns handling Returns handling is the reverse logistics analogous to order handling costs;

Performance incentives Since partners receive performance incentives for orders that, for example, are shipped on time, these incentives would not take place for orders fulfilled in the warehouses. Therefore, this opportunity cost for partners constitutes a saving for Farfetch;

Labour support This is analogous to performance incentive costs: partners would not receive financial support from Farfetch for seasonal peak-time labour;

Management fees These new costs are related to the warehouses' third-party logistics provider and are incurred by Farfetch;

Shipping costs It is expected of shipping costs from the warehouse to the customer to be lower than from the partner to the customer — this will be verified throughout the project. This cost reduction would result in a positive impact for both Farfetch and the customer. This also applies to returns;

Duties Products that are returned to partners are subject to duty drawback. At the time of writing, Farfetch is not capable of claiming duty drawbacks from products imported through weekly entry procedures from the free-trade zone warehouse to the NAFTA region, mainly due to technical reasons which do not have a solution in the near future. Customers that return an imported product are subject to a full refund, including the duties which Farfetch does not recover; hence this constitutes an added cost. This cost is cancelled if duty-cleared and geographically priced products that have been returned are resold;

Customer and partner service Because performance per order is expected to change — for example in terms of lead times, wrong item rates, packaging accuracy, and so on —, it is expected that the rate and type of contacts received at Farfetch from customers and partners will also change. In particular, if performance improves, a freeing up of resources is expected;

Recoveries When orders are returned by customers, partners may refuse to accept the product back — if, for example, the return is considered unreasonable or in inappropriate condition. In these cases, Farfetch will keep the item for resale, still refund the full price to the customer, but won't receive the retail price or other costs from the partner, and thus may be in a losing position. In some occasions, partners may only refund partially. This leads to a recovery rate, which is a metric that Farfetch keeps track of and translates the portion of value that Farfetch recovers from partners on returned items.

Because Farfetch would have control over which returns are accepted at the warehouse, the recovery rate is expected to increase;

Personnel Farfetch's employees that are allocated to the warehouse development or operation constitute an opportunity cost that has to be considered;

Infrastructure These costs are incurred by Farfetch and are related to software acquisitions, platform integrations and, overall, the infrastructure needed to support the warehouses' operations;

Marketing Farfetch strongly invests in marketing in order to bring in and retain customers. Warehousing is expected to have an impact in demand acquisition and retention, and thus lead to a higher volume and purchase frequency. At first, it might seem reasonable to believe that warehousing would result in marketing savings since it already generates the demand that marketing should, but, in practice, costs would in fact increase in the short-to-medium term. The reason for this is that Farfetch's marketing channels are highly associated to a per-order (or per-view) basis — retention emails and campaigns, online videos, and so on —, and therefore costs also increase accordingly. In the long-run, though, the positive impact of warehousing on demand, combined with the effect of marketing, can result in a synergistic and multiplicative outcome in comparison to the baseline. This effect can be perceived, equivalently, as a higher overall effectiveness of marketing investment — since for each dollar invested in marketing, a higher result in demand acquisition and retention is obtained due to the effect of warehousing —, and hence the savings are thus perceived.

As far as the scope of this project is concerned, the first perspective will be taken into account: it is assumed that Farfetch would maintain its marketing investment on a per order basis, and thus a growth in demand results in higher costs. This viewpoint is considered because it is a more conservative (and less riskier) approach and portrays more loyally the behaviour that is to be expected during launch and early stages of the project.

Insurance Products sold by Farfetch fall within the luxury segment, and are thus highly valued. Due to this, there are insurance costs that are to be considered for each additional order kindled by the warehouse.

One-off costs Finally, there are one-off costs related to launching the project, such as contracting with the third-party logistics provider or with consultancy services;

Table 3.1: Expected impacts (includes opportunity costs and ignores pricing of warehousing services)

Item	Costs	Savings
Packaging	F	🏪
Storage	F	🏪
Bulk Handling	🏪 F	
Bulk Shipping	F	🏪
Order Handling	F	🏪
Returns Handling	F	🏪
Performance Incentives	🏪	F
Labour support	🏪	F
Management Fees	F	
Shipping Costs		👤 F 🏪
Duties	F	
Customer & Partner service		F
Recoveries	🏪	F
Personnel	F	
Infrastructure	F	
One-off	F	
Marketing	F	
Insurance	F	

F Farfetch 🏪 Partner 👤 Customer

Some of these cost drivers are relatively straightforward to assess, whereas others require rough estimations and many inputs from different Farfetch internal stakeholders.

3.3.2 Value drivers

Because the warehouses will be located closer to customers, a better performance from the customers' viewpoint is expected in terms of lead times and shipping costs, for example. Besides that, higher operational control allows lower wrong item and no-stock rates. Because value is expected to arise from a higher control of the customer experience, this has two major consequences:

Retention Customers who receive orders from the warehouses have a better experience than if their orders had been fulfilled by partners. Therefore, these customers will more likely be retained, and hence purchase again or more frequently at Farfetch;

Acquisition Because Farfetch can offer a better service, and therefore promote it accordingly, it is also expected that new customers start purchasing at Farfetch. This effect may be harder to quantify;

The previous factors are related to demand. However, there are also factors which bring value and are related to supply:

Faster creation Products that are created (put online) earlier in the season can have a higher online exposure time, and thus are more likely to be sold at full price. This effect can be accomplished either by integrating a production centre in the warehouse — as soon as products are received, they can be created — or by leveraging the earlier planning and preparation with boutiques in order to create the products selected before the season starts;

Increased depth Assuming that warehousing saves a portion of the partner's costs with Farfetch and increases sell-through, it is reasonable to believe that these effects constitute an incentive for partners to bring in more stock depth. This would directly increase Farfetch's turnover;

3.3.3 Strategic rationale

In the upcoming subsections, a brief strategic framing will be provided in order to support the final recommendations and contribute to guiding the modelling stage.

3.3.3.1 Opportunities

The original motivation for Farfetch's warehousing as a service project is based on opportunities that exist mainly due to market needs, side-to-side with the current response offered by Farfetch and competitors, which has potential for improvement.

One key opportunity results from the benefits of moving supply closer to the country with the highest weight on demand. In fact, the United States has a minor weight over Farfetch's supply, yet accounts for a significant proportion of demand; on the other hand, Italy — which is where most of the United States orders are processed — provides most of the global supply. Besides that, almost half of the orders returned from the United States to Italy end up being shipped back again to the United States. These provide not only opportunities directly related to the customer's experience but also related to supply chain efficiency.

By moving inventory to the United States, Farfetch would reduce lead times by means of two levers: proximity and faster throughput enabled by the warehouse. This may directly lead to an overall greater customer experience.

Still due to proximity, shipping costs are also expected to diminish — and this applies not only to the United States, but to the whole influence area of the warehouse. Moreover, Farfetch can leverage the closeness of the warehouse's inventory to its biggest customer base by providing (or extending) new (or existing) delivery services which ultimately bring more value; examples include same-day delivery or F90. The warehouses may actually enable an overall development of Farfetch's route-to-market by capitalising on arising capabilities, customer reach and greater intelligence gathering potential. Finally, proximity also enables the acquisition of new customers, in particular of the VIP tier, who are very sensitive to the value proposition and, as a matter of fact, have been growing during the last years.

By providing warehousing services to partners, there is a shift of responsibility from partners to Farfetch's third party logistics provider in terms of order fulfilment which, ultimately, enables greater operational control. By exploiting this opportunity, Farfetch can reduce lead times, provide more accurate feedback to customers about order status, improve their shopping experience by reducing exceptions (wrong-items or no-stocks) and lead to savings supply-chain wide. This greater control also allows Farfetch to provide new services, such as consolidation of orders from the same customer. Lastly, this service would be charged to partners in the long-term, thus constituting an opportunity for a potential new revenue channel.

As was previously mentioned, a significant amount of orders returned to Italy end up being shipped back again to the United States. By handling returns at the warehouse, these wasteful fluxes are spared. In fact, many times partners refuse to accept returned products, and by transferring this decision to Farfetch the company can save money and products can be put online faster, leading to an increase in sell-through. Following a similar line of thought, Farfetch can also leverage the warehousing infrastructure and specialisation in order to produce items at the warehouse as soon as they arrive, instead of shipping samples to a dedicated production centre as happens today. This also enables stock to be put online faster.

By having a warehouse in a free-trade zone and by means of mechanisms such as weekly entry filings, Farfetch can optimise duty costs in many cases by reducing the amount paid per order. Besides that, returns to Europe that end up being shipped back again to the United States suffer retaxation, and by eliminating this flux with the warehouse the duty cost is also eliminated.

Finally, these factors combined lead to supply chain savings and to an increase in demand kindled by two levers: acquisition and retention. The former is induced essentially by providing new services, whereas the latter, which is expected to be much more significant, is due to an overall better experience provided to the customer and enabled by proximity and greater operational control. Altogether, these constitute opportunities on which Farfetch can capitalise by providing warehousing services to partners.

3.3.3.2 Threats

Farfetch's business model has proven to be successful along the last years, especially when compared to competitors. One of the most important reasons behind such success is the benefits conveyed by not possessing any inventory: Farfetch is not subject to inventory holding costs and depreciation. By offering warehousing services, though, Farfetch is slightly changing the paradigm and assuming part of these costs. One threat that arises is the fragmented vision in which such an initiative may result: Farfetch would no longer be a marketplace business that links partners directly to customers, but it would neither be a pure retailer; for certain products, Farfetch would ship directly from the partner, whereas for other products Farfetch would actually manage the inventory and incur in holding costs. The original vision was based on the premise that a given customer in the United States, for example, could purchase a luxury fashion product that would come directly from a boutique in Italy; the item would be picked, carefully wrapped and packaged by the hands of a human being working at the other side of the globe — at the hub of the luxury fashion industry. The warehousing model constitutes an alienation of this vision: orders are no longer fulfilled manually at the boutique, but rather come from a semi-automated warehouse. Arguably, the customer does not immediately perceive this because Farfetch does not inform during the order cycle about the origin of the product; nonetheless, this may still undermine Farfetch's vision in the long-term both for customer and employees if change is not adequately managed.

Concerning more practical and shorter term threats, partners play an important role. Because many of the latter are traditional boutiques that operate preponderantly offline, moving inventory to a warehouse overseas might seem like a very risky decision: partners would, up to a certain degree, lose visibility and control over their stock and they would not be able to use the offline channel. In fact, they would have to fully trust Farfetch and, if by any chance sales do not go as expected, at the end of the season partners would have to receive all the unsold and highly depreciated inventory at the door of their boutiques. This might result in an important threat to Farfetch: low partner commitment. If sales of products at the warehouses do not reach the expected sell-through, partners might put pressure on Farfetch in order to receive their inventory back before the season ends, so that they can, for example, use the offline channel. Moreover, fear from a partner could trigger a small scale chain reaction — somewhat similar to a bank run — and result in more than one partner reclaiming their inventory back.

Another threat that can be highly aggravated by the previous one is the existence of fixed costs and the inability to break-even. This is a project that requires a minimum scale in order to become profitable; however, under such frame of uncertainty — especially given the risk of low partner commitment — there is an existing threat that Farfetch would not be able to dilute the fixed costs in the long-run.

Finally, there is the threat of internal competition. By centralising inventory from different boutiques at the same place, brands, who supply products to boutiques in the first place, might feel that they are starting to compete against Farfetch. Because they are upstream of the supply chain, brands have negotiation power and can threat Farfetch's supply. Besides that, boutiques in the United States that see some identical products — referred to as duplicated — being brought from Italian boutiques to the warehouse might actually start loosing sales that they would have had if there had been no warehouse, because they would potentially be more favourable in the allocation algorithm. This threat of internal competition is not so obvious and can start showing itself during the long term, once the project reaches a minimum scale, thus disrupting the existing equilibrium.

3.3.3.3 Strengths

Concerning strengths, Farfetch has proven to be a solid and growing business. This means that the company already has recognition, especially in the United States, and that the warehouse would serve a market that has already been developed and where customer expectations are already well established.

Furthermore, Farfetch has a business model that minimises one very important risk present in the fashion industry: perishability. Because Farfetch would not actually own the inventory but would only provide the warehousing service, this strength would still apply. Actually, since Farfetch is not subject to buying inventory — especially at the beginning of a season — and because Farfetch receives payments right when an order is placed but only pays to partners, couriers and other stakeholders later in time, Farfetch is a cash-flow generating business. This may be an important strength during the roll-out phase of the warehousing project due to the plausibility of unaccounted events, needs and costs — such as with inventory or headcount.

Finally, the company possesses a lot of knowledge about the market, in particular when compared to competitors (Farfetch highly invests in gathering analytic insights about customers and offers a vast supply due to not needing to manage the inventory, whereas competitors have limited supply due to the opposite reason). By leveraging on this knowledge, Farfetch can minimise the risk of project failure by adequately selecting the partners, products and marketing tactics to be employed within the warehousing context so as to maximise full price sell-through.

3.3.3.4 Weaknesses

One of the most relevant of Farfetch's weaknesses related to this project is the company's lack of key knowledge about specific domains — like legal requirements —, in particular due to the fact that this is the first time that Farfetch ventures into such initiative and to the lack of literature and transparent examples that exist in the market.

Because Farfetch has been growing at an accelerated rate, many processes may not be adequately documented and lack of standard operating procedures (SOPs) at times. This complicates the task of planning the warehouse's operation because there are many inter-dependencies with other processes that may not be, themselves, properly standardised — or even documented at all. During peak times, this weakness may put Farfetch in a vulnerable position and even undermine the base business due to negative synergies.

Finally, and somewhat related to the previous weakness, Farfetch may have temporary technological barriers that play a role when implementing support for the warehouse into the system. Due to the fast growth and potential lack of standardisation and documentation at times, many technical solutions have been implemented in order to fix specific problems in the past and at the moment cannot be easily extended to accommodate the change required to support the warehouse's full scale roll-out (an example of this lies at the duty calculation algorithm, which was originally not designed to allow for free-trade zones). Nevertheless, Farfetch is technological company, therefore these do not constitute long-term weaknesses.

3.3.3.5 Preliminary Assessment

Providing warehousing services is a strategic move that aims at bringing Farfetch's supply closer to demand, and simultaneously gaining rights and subsequent control over the partners' inventory without actually owning it. This allows Farfetch to provide a better customer experience, to offer new services, to develop new and existing routes-to-market and to gain customer insights. In the long-term, these factors lead to operational savings and to a growth in demand, especially due to higher retention.

Nonetheless, partners adherence is key to guarantee success, especially due to the existence of fixed costs. If Farfetch does not guarantee a minimum sell-through, partners may lose confidence, reclaim their inventory back, undermine the relationship with Farfetch or event trigger a withdrawal chain effect. Due to this, a pilot phase with a rigorous product selection that maximises full-price sell-through should be devised in order to prove the concept and facilitate escalation — besides that, this phase also guarantees legal compliance and allows standardisation of operating procedures. Negotiation is a fundamental driver for success: in the shorter-to-medium term it keeps partner adherence under control and in the longer term it targets the potential threat of competition between Farfetch and brands.

Finally, efficiency gains can be induced by optimising technology and current processes so as to fully and efficiently support warehousing. This is a requirement that gains greater importance as the project scales.

Chapter 4

Modelling

This chapter aims at documenting the major steps that were taken in order to model and ultimately help quantify the most important cost and value drivers. These models constitute the intelligence behind the simulation exposed in the next chapter. For the sake of clarity and conciseness, only the modelling of the most complex or important drivers is reported next, since the goal of this chapter is to provide a possible work-flow for a bottom-up approach and illustrate each step with the use of analytical techniques. Finally, items that are not treated in this chapter are to be assumed, unless otherwise specified, as having been directly provided by Farfetch.

4.1 Assumptions and notation

Concerning overall assumptions, it was generally considered that:

Time constancy Costs do not change over time unless otherwise specified;

Labour costs are variable No data is available concerning the number of workers that partners or the warehouses dedicate to Farfetch and, in fact, fluctuations in the workforce are expected due to the seasonal nature of the industry. One alternative is to consider hourly rates and treat labour as a variable cost, which has two advantages: first, given the project's scope, it conveys a notion of resource utilisation — in particular during peaks — and, secondly, it provides a safer estimate rather than considering a fixed workforce without any supporting data. Finally, Farfetch already possesses estimates of costs per order for certain partners;

Since the project aims at assessing the value of an investment project, there are two scenarios that have to be taken into consideration at all times: a baseline and a warehousing scenario. The former represents either historical performance or simulated future performance if the warehousing project is not executed, whereas the latter represents simulated performance under an hypothetical warehousing scenario.

In terms of notation, items that relate to the warehousing scenario may have a prime ($'$) superindex to avoid confusion. Furthermore, $\mathcal{B}$ designates the set of bulk shipments, $\mathcal{O}$ the set of orders, $\mathcal{R}$ the set of returns, $\mathcal{W}$ the set of warehouses and $\mathcal{J}$ the set of locations.

4.2 Packaging

Concerning the modelling of packaging costs and incomes, there are three items that should be considered: acquisition paid by Farfetch to the supplier, shipping cost to the stock-point paid by Farfetch to the courier and a packaging fee charged to partners on a per order basis. In the warehousing scenario, partners would not pay the packaging fee, as opposed to the baseline. Packaging acquisition costs depend on the box type. At the moment of writing, Farfetch has an assortment of nine different boxes. Since boxes can be folded and shipped compactly, assessing shipping costs based on factors such as box volume is not possible. Hence, where data exists, average unitary costs per route were considered. If i represents the box type, j the destination (only one supplier is considered, and hence one single origin), $S_{i,j}$ the historical sum of shipping costs per box type and destination, $N_{i,j}$ the total number of boxes shipped and a_i the unitary (acquisition) cost per box type, the mean cost per box for destination j and box type i is given by $c_{i,j}$ in Equation 4.1:

$$c_{i,j} = \frac{S_{i,j}}{N_{i,j}} + a_i \quad \forall i \in \{1, \ldots, 9\}, j \in \mathcal{J} \tag{4.1}$$

When no data was available, the global mean given by Equation 4.2 was used instead ($N = \sum_i \sum_j N_{i,j}$):

$$\bar{c} = \sum_i \sum_j \frac{S_{i,j} + a_i N_{i,j}}{N} \tag{4.2}$$

Thus, box costs are given by 4.3:

$$c_{i,j} = \begin{cases} \frac{S_{i,j}}{N_{i,j}} + a_i & \text{if data for } \{i,j\} \text{ exists} \\ \bar{c} & \text{if data for } \{i,j\} \text{ does not exist} \end{cases} \quad \forall i \in \{1, \ldots, 9\}, j \in \mathcal{J} \tag{4.3}$$

The fee paid by partners to Farfetch is constant (f_p) and would not be charged on orders that are processed in the warehouse, thus constituting a saving to partners.

4.3 Bulk handling

Costs related to bulk handling are difficult to model due to the lack of historical data. Concerning partner costs, data was obtained based on Farfetch's estimates which roughly take into account the overall average hourly cost and an appraisal of the average time to prepare or receive a bulk shipment (for the sake of simplicity, no distinction was made between inbound and outbound). This approach theoretically assumes that all partners join the warehousing initiative in the same proportion, which is not realistic. On the other hand, because the great majority of supply comes from Italy — in fact, at the time of writing all partners that are currently involved in the project are Italian — it can be reasonably assumed that hourly costs likely balance out and that the main source of uncertainty is, actually, the time estimate. Hence, for each bulk shipment $i \in \mathcal{B}$ dispatched to a warehouse, the partners' bulk handling costs c_i are assumed constant and given by $c_i^p = \bar{c^p}$

Regarding bulk handling at the warehouse side, estimates were provided by the third-party logistics for inbound and outbound productivities (in units per hour) and a quote for hourly labour costs. An analogous analysis was replicated. For each bulk shipment i with N_i units shipped to or from warehouse w, taking $y = 1$ if it is an inbound or $y = 0$ if it is an outbound shipment, considering a productivity of $r_{y,w}$ units per hour and a hourly labour cost of h_w for warehouse w, the bulk handling cost is given by 4.4:

$$c_{i,w}^{FF} = \frac{N_i}{r_{y,w}} h_w \quad \forall i \in \mathcal{B}, w \in \mathcal{W} \tag{4.4}$$

In the baseline scenario, no bulk handling cost is incurred by partners nor Farfetch; these are new cost items that arise because of warehousing.

4.4 Bulk shipping

Bulk shipment costs were obtained from a quote for air freight provided by the carrier for the route Italy to the United States and under different hypothetical scenarios which relate costs to units shipped. These have been plotted in the left scatter plot of Figure 4.1.

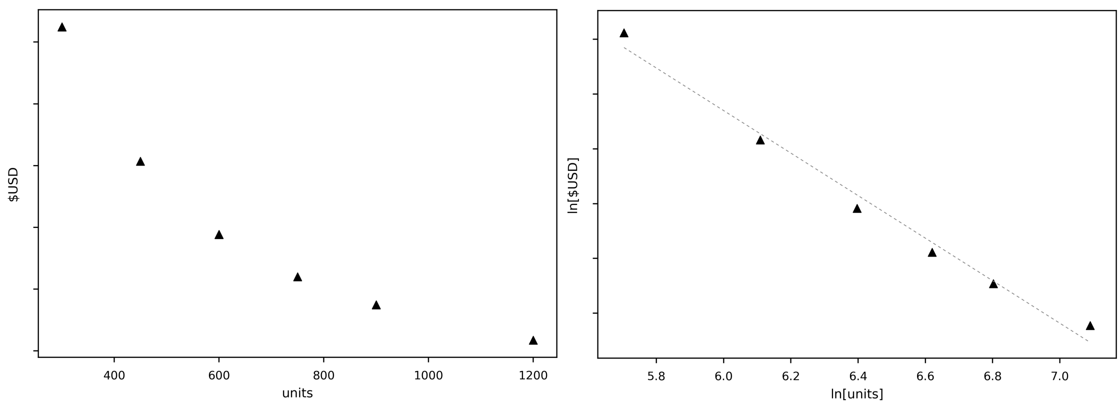

Figure 4.1: Bulk shipping costs natural and logarithmic scale (values omitted for confidentiality)

The shape described by the data points was expected due to the effects of scale; besides that, they suggest a power law. By plotting the same observations with logarithmic axis (Figure 4.1 right), the data indeed suggests the behaviour in question, evidenced by the resulting apparent linear relationship.

Thus, if y_i is the shipping cost quoted by the carrier for quantity x_i, the hypothetical relationship is given by 4.5 (where α and β are regression coefficients):

$$\ln \hat{y}_i = \hat{\alpha} + \hat{\beta} \ln x_i \quad \forall i \in \mathcal{B} \tag{4.5}$$

The parameters were estimated through least squares and brought back from logarithmic to a natural definition through Equation 4.6:

$$\hat{y}_i = e^{\hat{\alpha}+\hat{\beta}\ln x_i} \iff \hat{y}_i = \hat{\gamma} x_i^{\hat{\beta}} \quad \forall i \in \mathcal{B} \quad \text{with} \quad \hat{\gamma} = e^{\hat{\alpha}} \tag{4.6}$$

Finally, the cost for Farfetch of bulk shipment i with N_i units is estimated by direct application of the model:

$$c_i^{FF} = \hat{\gamma} N_i^{\hat{\beta}} \quad \forall i \in \mathcal{B} \tag{4.7}$$

4.5 Order handling

Regarding order handling, Farfetch possesses estimates on a per order basis for certain partners. Therefore, these were averaged and generalised, leading to the assumption that they apply to any partner. Thus, the cost of processing an order was taken as constant:

$$c_i^P = \bar{c} \quad \forall i \in O \tag{4.8}$$

When it comes to the warehouses, the labour hourly costs quoted by the third-party logistics provider were considered in order to determine the cost per order. If the hourly labour cost for warehouse w is given by h_w, and the time it takes, in hours, to process order i is equal to t_i, the cost of handling order i is given by $c_{i,w}^{FF}$ in Equation 4.9:

$$c_i^{FF} = h_w t_i \quad \forall i \in O \tag{4.9}$$

Note that index i implies w, because each order is allocated to one and only one warehouse. Finally, t_i is a variable predicted by the model, which is tackled in the speed of sending subsection.

4.6 Return handling

Concerning return handling, modelling is analogous to order handling. The cost to partners of processing return i is given by the average of the benchmark available:

$$c_{return_i}^P = \bar{c}_{return} \quad \forall i \in \mathcal{R} \tag{4.10}$$

Regarding the warehouses, it was assumed, for the sake of simplicity and due to the lack of data, that the distribution of the time it takes to process a return is the same as the distribution of the time it takes to process an order — both in shape and scale. Hence, if the hourly labour cost for warehouse w is given by h_w, and the time it takes, in hours, to process order i is equal to t_i, the cost to Farfetch of handling the return of order i in warehouse w is given by $c_{i,w}^{FF}$ in equation 4.11:

$$c_{return_i}^{FF} = h_w t_i \quad \forall i \in \mathcal{R} \tag{4.11}$$

4.7 Labour incentives

The only data that was made available by Farfetch about labour incentives is the total amount that was distributed among the partners selected for the warehousing pilot project, during the year of 2017. Since these incentives are distributed to support partners during peak times, it is reasonable to assume they are correlated with the number of orders they process. It was thus assumed that these incentives can be assessed on a per order basis. Hence, the incentive per order (c_{order}) is assumed to be constant and given by the average incentive paid in 2017, as expressed by equation 4.12:

$$c_{\text{order}} = \frac{C_{2017}}{N_{2017}} \tag{4.12}$$

Partners whose orders are processed in the warehouses do not receive the incentive that they would if they had processed the order.

4.8 Performance incentives

Performance incentives were assessed in an analogous manner to labour incentives due to the fact that data was provided in the same fashion and, in lack of further information, the average cost per order was the most secure estimate. In practice, though, it is reasonable to believe that by fulfilling orders from the warehouse partners free up capacity, and thus can actually increase their performance and receive higher incentives per order. Alternatively, partners could instead leverage their increase in capacity in order to bring in more supply — or prioritise the offline channel, for example —, thus leading to a scenario where performance does not increase but in fact remains constant — or even decreases. Due to the complexity inherently present in predicting these behaviours, especially linked to the lack of data, it was assumed that these effects cancel out each other so that the partners' performance remains constant at all times and incentives can be assessed on an average per order basis, as with labour incentives.

4.9 Shipping Costs

Shipping costs can refer to the costs invoiced by the courier, to the rates paid by the customer to Farfetch on checkout or to the costs charged by the courier on returns. Even though they differ in nature, the analysis performed on each type was very similar. Therefore, the following discussion refers to the modelling of any of these costs interchangeably (unless otherwise specified).

From the baseline to the warehousing scenario, there are two major factors that can potentially change and which, at first sight, may have an influence on shipping costs: route and volume shipped. The former is obvious: orders have the same destination but distinct origin due to the existence of the warehouse; whereas the latter occurs because warehouses will tend to have a higher rate of optimum packaging (in this context, optimum means more efficient in terms of volume). It was assumed that the service type — such as express or standard — does not change for a given order when compared to the baseline, since it is a customer decision, independent of where the order is processed.

Since there are so many combinations of route, box and service type and, furthermore, because there is variability present in historical data, machine learning techniques were used to learn from the existing observations and enable more accurate predictions. Analytical alternatives, such as average cost per route, would not capture, for example, the effect of a higher packaging accuracy; concerning classical regression, in order to be effective, it would require, for example, an instance per route — which is unpractical. Thus, machine learning was seen as an adequate alternative, given the dimension of the data and scope.

Because it is a regression problem with mixed-type features, a Random Forest was used. Random Forests have the advantage of being more resistant to over-fitting than, for example, single decision trees, handle well mixed-type predictors and are resistant to missing values and do not require complex data preparation — such as normalising. These advantages, at a cost of computational power, gain further importance within the present context: a simulation solution requires a robust and reliable method that can be incorporated and automated for future use, without the need to process the data every time it changes.

The data used consisted of Farfetch's 2016 orders extracted from the company's servers (2016 was used in order to function as a standalone analysis and to avoid any chance of over-fitting, since the simulation would be based on both existing and modified data from 2017). In order to first assess model performance, the dataset was split by implementing a nested K-fold cross-validation technique (as depicted in Figure 4.2). The reason for this is that, according to Gavin C. Cawley (2010), "variance in the cross-validation statistic introduces the potential for over-fitting in model selection as well as in training". This variance is expected to increase with the number of folds (and is very likely to be positive for K>1), thus, if the performance of the model with hyperparameter optimisation is to be estimated without the risk of incurring in an optimistic bias, Gavin C. Cawley (2010) suggests a nesting k-fold cross-validation technique — for an alternative technique, refer to Gavin C. Cawley (2007):

> Thus it makes sense to [...] fit the model iteratively using a pair of nested loops, with the hyperparameters adjusted to optimise a model selection criterion in the outer loop (model selection) and the parameters set to optimise a training criterion in the inner loop (model fitting/training)

Note that nested-cross validation was used in the context to reduce the bias of the performance estimate, potentially caused by the statistic variance expected with a single layer k-fold cross-validation. This technique could also be used, for instance, to compare the performance between different alternative models.

For the sake of simplicity and efficiency, the model hyperparameters would only be tuned during the current modelling phase and would be used as such on future instances of the model in the simulation, if required by the user — for example if data had changed — and thus, it was important to get an estimate of the performance of the model with minimum bias.

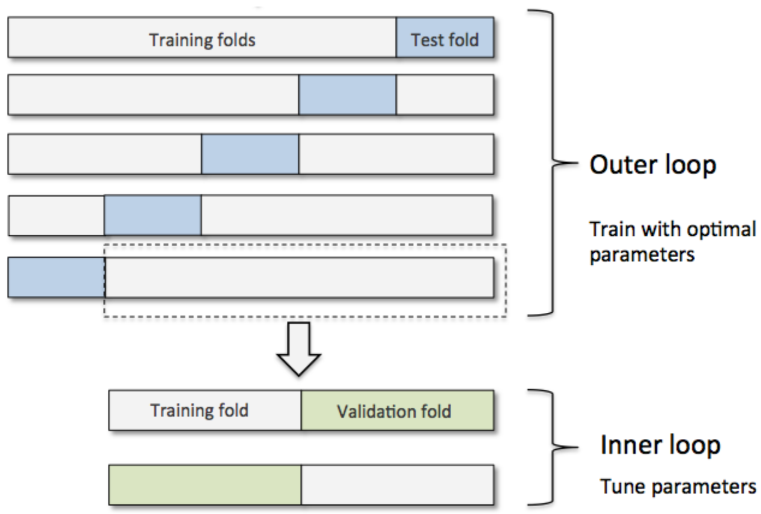

Figure 4.2: Nested k-fold cross-validation. Source: Raschka (2018)

Features selected were: route origin, route destination, service type, weight and box. For shipping and return costs charged by the courier, the actual box used was considered when available, whereas for shipping rates charged to the customer, the recommended box was used as feature instead.

In practice, shipping costs may actually reach relatively high values, and thus a right skew of the distribution is expected. Nevertheless, extreme and very unlikely values (empirically defined as accounting for less than 1%) were removed due to the likelihood of manual or system errors, as suggested by Farfetch Delivery Support analysts. Concerning the actual skew, the distributions of shipping costs were log-transformed in each model (except for shipping rates paid by user, where no skew exists). If the shipping cost of observation i is given by y_i, the transformation is given by eEuation 4.13 (1 is summed to y_i to avoid a negative z_i).

$$z_i = \ln(1 + y_i) \tag{4.13}$$

The result is shown in Figure 4.3:

Random Forests were initialised and trained on the subsets. Hyperparameter tuning was accomplished through a random search on an empirically bounded domain defined by the number of estimators, on whether to use bootstrap samples or not, the maximum depth of each tree, maximum number of features, maximum number of samples per leaf, maximum number of samples required to split a node and split criterion. A random search was used — instead of a grid search, for example — for efficiency purposes (in potential detriment of effectiveness).

Lastly, performance was evaluated through the following metrics: Mean Average Error (MAE) and Root Mean Squared Error (RMSE). The final performance assessment is shown in Table 4.1:

The models displayed adequate performance for the purpose and were thus validated. In order to generate the definitive models to be used during the simulation, all the orders from 2016 were used as the training dataset (with outlier removal and label log-transformation). Regarding the hyperparameters used to train the final models, these were obtained from an intermediate stage in

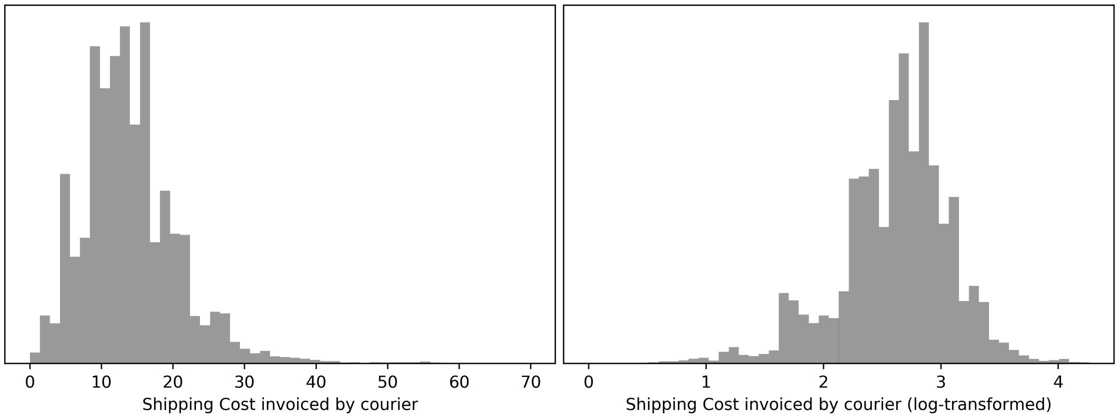

Figure 4.3: Shipping costs distribution (different scales used for clarity) — skewness reduction through log-transform

which the same random search procedures were performed on a single layer k-fold partition of the dataset.

Finally, for a given observation, if predictors for model $i \in \{1,2,3\}$ are given by $\boldsymbol{X}_i = \{$origin, destination, service,...$\}$ and $z_i(\boldsymbol{X}_i)$ is the output of model i, the respective cost prediction is given by Equation 4.14:

$$\hat{y}_i = e^{z_i(\boldsymbol{X}_i)} - 1 \quad \forall i \in \{1,2,3\} \tag{4.14}$$

Table 4.1: Shipping costs models — performance

Model	Target	MAE	RMSE
1	Order shipping (Farfetch)	0.016	7.4
2	Return shipping (Farfetch)	0.072	8.1
3	Shipping (Customer)	0.022	4.7

4.10 Transit time

Transit time was modelled by following the same steps as with shipping costs, that is, Random Forests were trained after preparing the subsets, hyperparameters tuned and performance assessed with nested k-fold cross-validation based on orders from 2016. After verifying that performance was adequate, the definitive model was trained on the whole dataset. Since the target variable was also found to be skewed, the same log-transformation was applied.

Nevertheless, the modelling of transit time differed from the modelling of shipping costs in two cases: first, additional features were found to be relevant, namely: the day of the week and dispatch hour (in partner local time) and the tariff code associated to the product ordered; secondly, the model itself, which was engineered as a two stage model.

The reason why the model was devised with two stages is the following: order transit time tends to take, on average, 0.92 extra days per Sunday that falls between dispatch and delivery dates, when compared to orders that have no Sundays between (p-value < 0.001). In practice, this is due to some couriers not operating on Sundays, or weekends in general. The problem is that the delivery date is not known, since it depends on transit time. Hence, the model first estimates a transit time (*TransitTime_0* in Figure 4.4), then calculates the number of Sundays that fall between the order dispatch date and the predicted delivery date (which is calculated from the dispatch date and the predicted transit time), adds this feature to the dataset and then feeds it to a second model which will update the old prediction with a new one (*TransitTime_1*). As long as the predictions do not converge (or up to 3 iterations, arbitrarily defined), the algorithm will keep calculating Sundays and estimating transit times. Most of the times, though (89% on average for 10 folds), the predictions *TransitTime_0* and *TransitTime_1* outcome with equal values right at the first iteration, and thus no further iterations are required.

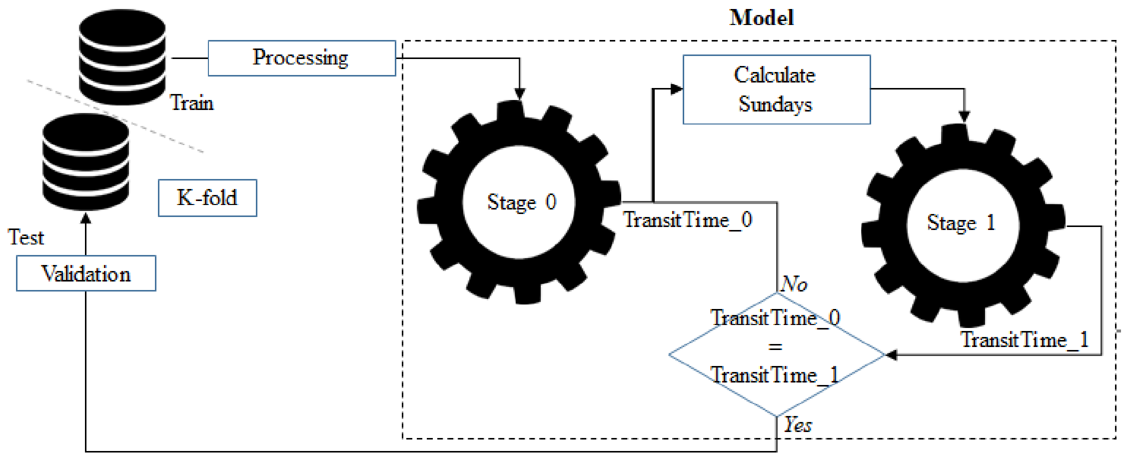

Figure 4.4: Two stage model for transit time prediction

Results obtained are displayed in Table 4.2:

Table 4.2: Transit time models — performance

Num. stages	Training time [s]	MAE	RMSE
1	311	-0.3	3.8
2	613	-0.27	2.6

Despite taking more time to train, the two-stage model was selected due to higher performance and trained on the whole dataset by following an analogous procedure as with the shipping cost models.

4.11 Speed of sending

The third-party logistics provider handed over an estimate with productivity rates, including an outbound rate in units per hour (not to be confounded with the bulk outbound previously used).

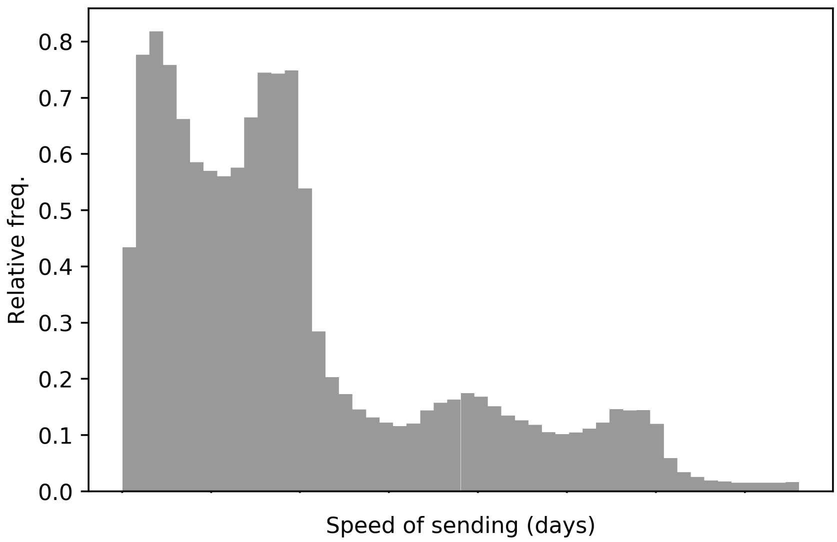

Figure 4.5: Speed of sending histogram of a sample of boutiques with at least 1000 orders in 2017

Let the outbound rate for warehouse w be represented as r_w. One possible approach would be to assume a constant speed of sending t_i per order i (again, each order i implies a unique warehouse w):

$$t_i = \frac{1}{r_w} \quad \forall i \in O \tag{4.15}$$

Yet, this is unrealistic since it does not take into account variability which, for example, is due to different capacity usage along the seasonal cycle.

However, a comparison with boutiques is also not possible because the former do not operate during night time or, some, during weekends. This means that, even though a different scale is expected and can be corrected in order to model the warehouse, variability cannot be replicated due to the incompatible behaviour (see Figure 4.5).

Because no data about the warehouse is available — and no variability benchmark was found in the literature —, modelling warehouse speed of sending was accomplished by using a three-point estimation technique, originally created by Clark (1962). The technique is based on the principle of gathering expert insights concerning optimistic (with a right-tail probability of less than 1%), most likely (based on historical insights) and pessimistic (with a left-tail probability of less than 1%) times and fitting a beta distribution such that these conditions are met. This technique is traditionally used in a project management context, but, according to D. Greenberg (2010):

> [The technique] can be applied to semi-automated activities, where the presence of man-machine influence under random disturbances is, indeed, very essential.

This matches the problem in discussion. A triangular distribution, for instance, could arguably be used, but the semi-automated nature of the warehousing fulfilment process, combined with the weights idealised by Clark (1962) — where the most likely time is given more importance, as expected with semi-automated processes —, likely result in a more accurate distribution rather than using a non-weighted one.

Let a be the optimistic time, m the most likely time ($m = \frac{1}{r_w}$) and b the pessimistic time. Under the technique's assumptions, estimates of mean and variance, as explained by Byung-cheol Kim (2009), are given respectively by Equations 4.16 and 4.17:

$$\mu = \frac{1}{6}(a + 4m + b) \tag{4.16}$$

$$\sigma^2 = \frac{1}{36}(b - a)^2 \tag{4.17}$$

The probability density function is given by Equation 4.18 (Γ designates the gamma function):

$$f(x) = \frac{\Gamma(p+q+2)}{\Gamma(p+1)\Gamma(q+1)} x^p (1-x)^{p(1/m_x - 1)} \quad \text{with} \quad m_x = \frac{p}{p+q} \tag{4.18}$$

Under Clark (1962) model, $p + q = 4$. Finally, the last parameter can be found from the condition:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{4.19}$$

The problem was solved by substituting $q = 4 - p$ and numerically solving for p^* in 4.20 such that, for ϵ small enough ($\epsilon = 0{,}001$ used), equation 4.20 holds:

$$|\int_{-\infty}^{+\infty} \frac{\Gamma(6)}{\Gamma(p^*+1)\Gamma(5-p^*)} x^{p^*} (1-x)^{p^*(4/p^*)} dx - 1| \leq \epsilon \tag{4.20}$$

The cumulative distribution function (to be used for generating random observations during simulation) was discretised and subsequently interpolated for simple access. Thus, if $\mathcal{P}$ is a partition of the domain, a pair $\{F_i, x_i\}$ can be obtained for each subset from 4.21:

$$F_i = F_{p=p^*}(x_i) = \int_{-\infty}^{x_i} f_{p=p^*}(\bar{x})d\bar{x} \quad \forall x_i \in \mathcal{P} \tag{4.21}$$

Finally, a spline interpolation (z) was fitted through each pair $\{F_i, x_i\}$ so that by providing a pseudo-random number, a value for speed of sending is obtained. In order to guarantee that randomly generated observations always fall within an adequate solution space for practical purposes — interpolations may be dangerous, especially on extremes —, the function's range has been truncated within the domain $[0, 1]$ at the maximum and minimum observed values in historical data. Figure 4.6 shows the pairs $\{F_i, x_i\}$ and the spline interpolation for one of the warehouses.

$$SoS = \begin{cases} \min(x_i) & \text{if} \quad z_{x_i \to x_i}(r) < \min(x_i) \\ z_{F_i \to x_i}(r) & \text{if} \quad \min(x_i) \leq z_{F_i \to x_i}(r) \leq \max(x_i) \\ \max(x_i) & \text{if} \quad z_{x_i \to x_i}(r) > \max(x_i) \end{cases} \quad \text{with} \quad r \in [0, 1] \tag{4.22}$$

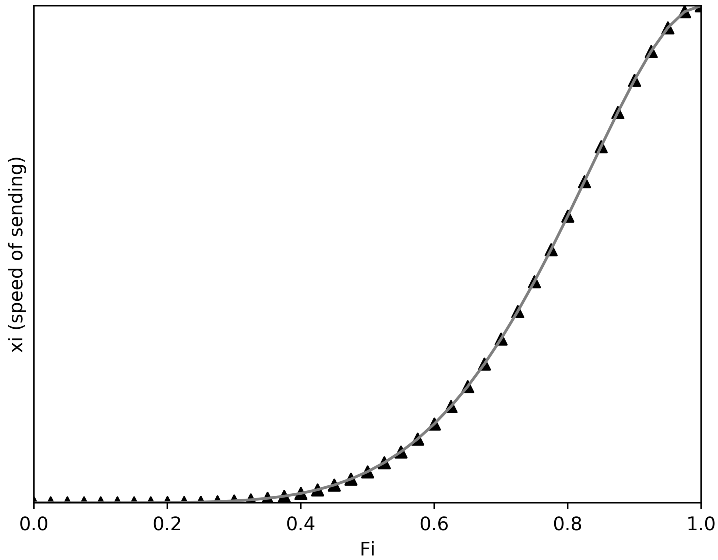

Figure 4.6: Spline interpolation of a discretised cumulative distribution beta-PERT function for one warehouse — speed of sending values omitted for confidentiality.

4.12 Customer Service

When a customer reaches Customer Service, a ticket is created. Farfetch possesses estimates about the average cost per ticket, and therefore by predicting the reduction in tickets enabled by the warehouse, it is also possible to estimate the resulting savings. Since the vast majority of tickets that are related to factors influenced by the warehouse are of type *order status*, only these were taken into consideration.

Due to data availability constraints, it was only possible to obtain an estimate of the average proportion of orders that have had at least one *order status* ticket for each bin that is part of a discretised range of speed of sending data. After applying a logarithmic transformation to the relative frequency data points, a linear relationship seemed to arise. Given the empirical evidence (the longer it takes to notify a customer that the order has been shipped, the more he tends to reach Customer Service), the relationship is considered reasonable and represented by Equation 4.23 — the behaviour was (arguably) extrapolated by assuming that the relationship holds for any speed of sending.

$$\hat{p}_{ticket} = \begin{cases} 0 & \text{if} \quad \alpha e^{\beta SoS} < 0 \\ \alpha e^{\beta SoS} & \text{if} \quad 0 \leq \alpha e^{\beta SoS} \leq 1 \qquad \text{with} \quad r \in [0, 1] \\ 1 & \text{if} \quad \alpha e^{\beta SoS} > 1 \end{cases} \tag{4.23}$$

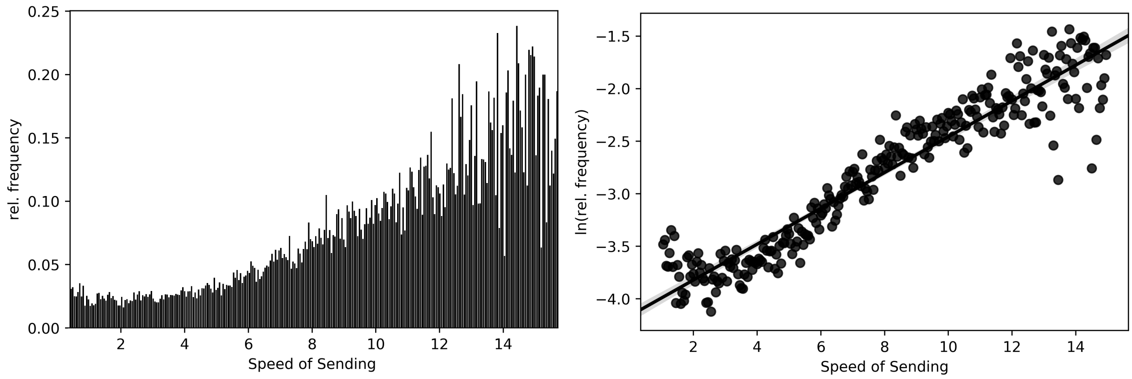

Figure 4.7: Proportion of tickets per order (0.05 days per bin) — original and transformed data

4.13 Other Costs

Data regarding costs such as management fees, infrastructure, insurance and marketing were directly provided by Farfetch or obtained in a way that results in straightforward modelling (such as monthly costs for a given warehouse or costs given on a per order basis) and, therefore, are not going to be presently reported for the sake of pertinence and scope. Concerning recoveries, it was assumed that Farfetch would recover from returns the amount related to merchandising (either the retail or wholesale price) but not the shipping costs. Finally, some other cost items were estimated based on Farfetch's internal insights — such as personnel required to support the warehouses operation.

4.14 Retention

By improving the service perceived by the customer through factors such as lower lead times, lower shipping costs, more adequate packaging, lower wrong-items or lower no-stock rates, retention is expected to increase. This, however, is an effect difficult to quantify and, in particular, difficult to translate into financial indicators capable of guiding decision making. In fact, not much is addressed in the literature concerning the task of marginally assessing the potential impact in income leveraged by the effect in retention that results from the footprint of an investment project. There are, however, instances where a model has been applied to measure retention or churn in terms of time or number of incidences for a given population. Survival models are commonly used; Andy Sadler (2006), for example, tried to predict retention rates by using survival analysis and DirkVan den Poel (2004) analysed attrition based on a similar principle. Yet, the challenge which is seldom treated in the literature is establishing the relationship between these insights and the long-term financial impact that they convey. Due to this necessity, a theoretical model was developed for the project and is described subsequently.

4.14.1 Model derivation

One possible way to assess the impact of retention is to translate the latter into a shift of the customer's lifetime value. Thus, if CLV_w is a customer's average lifetime value under the warehousing scenario — with expected higher retention — and CLV_b is the customer's average lifetime value under the baseline scenario, the average additional profit generated per customer due to the effect of retention (assuming that the variations in lifetime value result from shifts in retention) is given by:

$$\Delta = CLV_w - CLV_b \tag{4.24}$$

Assuming that the average net profit per order $\bar{v}$ remains constant over time and does not vary among scenarios, ignoring inflation and taking r as Farfetch's weighted average cost of capital, the customer lifetime value is given by the net present value of the sum of expected future net profit streams generated by the customer through his orders (t_i stands for the order time relative to the present date):

$$CLV = \sum_{i=1}^{\infty} \frac{\bar{v}}{(1+r)^{t_i}} \tag{4.25}$$

Customer attrition occurs when $t_i \to \infty$.

If a customer purchases, on average, at the same repurchase rate $\frac{1}{\Delta t}$ (orders per unit of time), the CLV is given by:

$$CLV = \sum_{i=1}^{\infty} \frac{\bar{v}}{(1+r)^{i\Delta t}} \tag{4.26}$$

The impact of warehousing on retention can be translated as a variation in the average time until next purchase. For instance, customers who have their orders fulfilled at a warehouse have, on average, a time until next purchase (Δt_w) different from customers who don't have their orders fulfilled at a warehouse (Δt). Yet, Equation 4.26 assumes a constant repurchase rate which under a warehousing scenario is wrong. In fact, it would mean that customers who have their next order fulfilled in a warehouse will always have future orders fulfilled there, but this hardly happens in practice because the customer does not know where the order comes from, warehouses would have a limited amount of supply and, finally, an algorithm takes care of the rooting process.

Let p_w be the probability that a customer's order is fulfilled at a warehouse — this can be obtained from relative frequencies, after simulating. The probability of not ordering at a warehouse is complementary $(1 - p_w)$. The expected net present value generated with the first order is given by:

$$NPV_1 = p_w \frac{\bar{v}}{(1+r)^{\Delta t_w}} + (1 - p_w) \frac{\bar{v}}{(1+r)^{\Delta t_{\bar{w}}}} \tag{4.27}$$

Assuming that the probabilities remain constant along time — which is equivalent to assuming that, even though the customer base may grow, the proportion that gets orders fulfilled from the warehouse remains constant —, the expected net present value generated with the second order is given by equation 4.28.

$$NPV_2 = p_w^2 \frac{\bar{v}}{(1+r)^{2\Delta t_w}} + 2p_w(1-p_w)\frac{\bar{v}}{(1+r)^{\Delta t_w+\Delta t}} + (1-p_w)^2 \frac{\bar{v}}{(1+r)^{2\Delta t}} \tag{4.28}$$

Note that the total time until the second purchase depends not only on whether the order is fulfilled or not in the warehouse, but also on where it happened for the first order: this leads to conditional probabilities. Figure 4.8 depicts the rationale:

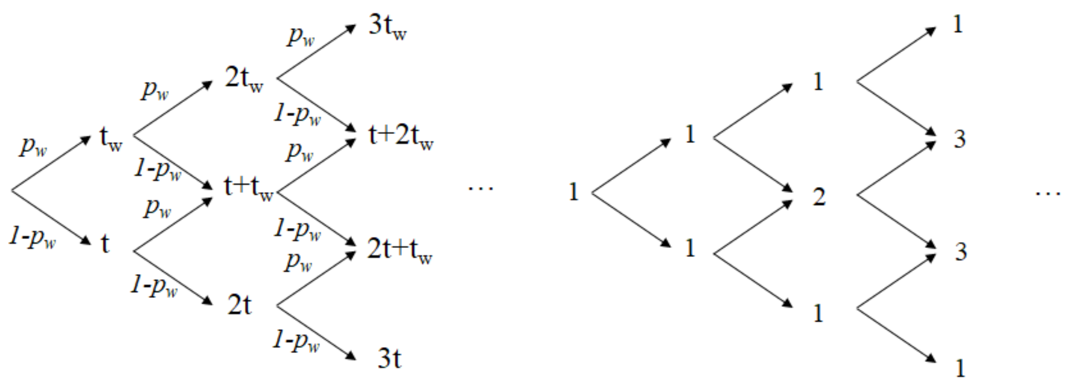

Figure 4.8: Total time until second purchase — dependency on previous times — and Pascal's triangle

For each order that was fulfilled in the warehouse (with probability p_w), it took a time of t_w until next purchase. The same analogous principle follows for orders not fulfilled in the warehouse. Therefore, a function $f(a,b)$ can be defined so that if a orders have been fulfilled in the warehouse and b orders in boutiques, the corresponding term in the expected net present value equation is given by $f(a,b)$ in equation 4.29:

$$f(a,b) = p_w^a(1-p_w)^b \frac{\bar{v}}{(1+r)^{a\Delta t_w + b\Delta t}} \tag{4.29}$$

Thus, Equation 4.28 can be rewritten as:

$$NPV_2 = f(2,0) + 2f(1,1) + f(0,2) \tag{4.30}$$

Extending the analysis for the third and fourth orders leads to:

$$NPV_3 = f(3,0) + 3f(2,1) + 3f(1,2) + f(0,3) \tag{4.31}$$

$$NPV_4 = f(4,0) + 4f(3,1) + 6f(2,2) + 4f(1,3) + f(0,4) \tag{4.32}$$

The pattern that arises is given by:

$$NPV_k = \sum_{i=0}^{k} \alpha_i f(k-i,i) \tag{4.33}$$

Concerning the coefficients α_i, and given the analogy of the left pattern in Figure 4.8 with Pascal's triangle, it is possible to observe that the coefficients follow the Pascal's triangle sequence of terms and are therefore given by Equation 4.34:

$$\alpha_i = \binom{k}{i} \tag{4.34}$$

Therefore, CLV is given by the net present value of the sum of net profits obtained for each and every future order from a given customer:

$$CLV_w = \lim_{n\to\infty} \sum_{n=1}^{n} \sum_{i=0}^{k} \binom{k}{i} f(k-i,i) \tag{4.35}$$

Rewriting the expanded form:

$$CLV_w = \bar{v} \lim_{n\to\infty} \sum_{k=1}^{n} \sum_{i=0}^{k} \binom{k}{i} \frac{p_w^{k-i}(1-p_w)^i}{(1+r)^{(k-i)\Delta t_w + i\Delta t}} \tag{4.36}$$

Which is equivalent to:

$$CLV_w = \bar{v} \lim_{n\to\infty} \sum_{k=1}^{n} \sum_{i=0}^{k} \binom{k}{i} \left(\frac{p_w}{(1+r)^{\Delta t_w}}\right)^{k-i} \left(\frac{1-p_w}{(1+r)^{\Delta t}}\right)^{i} \tag{4.37}$$

Let $\beta = \left(\frac{p_w}{(1+r)^{\Delta t_w}}\right)$ and $\lambda\beta = \left(\frac{1-p_w}{(1+r)^{\Delta t}}\right)$. Equation 4.37 can be written as:

$$CLV_w = \bar{v} \lim_{n\to\infty} \sum_{k=1}^{n} \beta^k \sum_{i=0}^{k} \binom{k}{i} \lambda^i \tag{4.38}$$

The binomial theorem implies that:

$$\sum_{i=0}^{k} \binom{k}{i} \lambda^i = (1+\lambda)^k \tag{4.39}$$

Thus:

$$CLV_w = \bar{v} \lim_{n\to\infty} \sum_{k=1}^{n} [\beta(1+\lambda)]^k \tag{4.40}$$

Finally, let $\eta = \beta(1+\lambda)$. Since $|\eta| < 1$ because $p_w \in [0,1]$ and $r > 0$, and since the polylogarithm's convergence is given by $Li_0(\eta) = \sum_{k=1}^{\infty} \eta^k = \frac{\eta}{1-\eta}$ (granted that $|\eta| < 1$), the customer's lifetime value converges to:

$$CLV_w = \bar{v} \frac{\eta}{1-\eta} \tag{4.41}$$

With $\eta = \beta(1+\lambda)$, $\beta = \left(\frac{p_w}{(1+r)^{\Delta t_w}}\right)$ and $\lambda\beta = \left(\frac{1-p_w}{(1+r)^{\Delta t}}\right)$.

Concerning the baseline scenario, no warehouse exists and, therefore, an average repurchase rate can be assumed to be constant and equal to Δt. Letting $\theta = \frac{1}{(1+r)^{\Delta t}} < 1$:

$$CLV = \sum_{k=1}^{\infty} \frac{\bar{v}}{(1+r)^{k\Delta t}} = \bar{v} \sum_{i=k}^{\infty} \theta^k = \bar{v} \frac{\theta}{1-\theta} \tag{4.42}$$

Finally, the expected increase in lifetime value per customer arising from a warehousing scenario in comparison to the baseline (and within the model's assumptions) is given by:

$$\Delta CLV = CLV_w - CLV = \bar{v}\left(\frac{\eta}{1-\eta} - \frac{\theta}{1-\theta}\right) = \bar{v} \frac{\eta-\theta}{(1-\theta)(1-\eta)} \tag{4.43}$$

Yet, the previous analysis only applies for customers that already exist, and thus does not capture the effect of retention of customers that are to be acquired in the future. Assuming that the previous behaviour and assumptions hold for a given customer that is acquired at time t, the expected increase in lifetime value at the moment would be given by Equation 4.44:

$$\Delta CLV = \frac{\Delta CLV_t}{(1+r)^t} = \frac{1}{(1+r)^t} \bar{v} \frac{\eta-\theta}{(1-\theta)(1-\eta)} \tag{4.44}$$

Let F_t be the number of customers in year t (or any given period). The net present value generated by all customers (existent and acquired) from period a to b can be approximated by:

$$\Delta CLV|_a^b \approx \sum_{t=a}^{b} \frac{F_t - F_{t-1}}{(1+r)^t} \bar{v} \frac{\eta-\theta}{(1-\theta)(1-\eta)} \tag{4.45}$$

This assumes that customers are only acquired exactly at the beginning of each period. However, customers can be acquired continuously throughout the year. Therefore, time t can be assumed as being continuous and, likewise, $F(t)$ could be a continuous function in $t \in (a,b)$ which represents the number of customers at any instant t. This continuous approach can be obtained by partitioning the domain into an infinite number of n infinitesimal intervals k of length $\frac{b-a}{n}$ while using the equivalent effective discount rate for the resulting elementary period and summing the value at each interval. The formulation thus becomes:

$$\Delta CLV|_a^b = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{b-a}{n} \frac{F(\frac{b-a}{n}(k+1)) - F(\frac{b-a}{n}k)}{\frac{b-a}{n}} \frac{1}{\left[(1+\frac{r}{n})^n\right]^{k\frac{b-a}{n}}} \bar{v} \frac{\eta-\theta}{(1-\theta)(1-\eta)} \tag{4.46}$$

Since $\frac{b-a}{n} \to 0$, the second fraction of the summation is the derivative $f(t) = \frac{dF(t)}{dt}$. The denominator in the third fraction of the summation is a case of Euler's limit formula and converges to $e^{rk\frac{b-a}{n}}$. Thus, 4.46 can be rewritten as:

$$\Delta CLV|_a^b = \lim_{n\to\infty} \frac{b-a}{n} \sum_{k=1}^{n} \frac{f(k\frac{b-a}{n})}{e^{rk\frac{b-a}{n}}} \bar{v} \frac{\eta-\theta}{(1-\theta)(1-\eta)} \tag{4.47}$$

Let the content of the summation be given by $g(t)$:

$$\Delta CLV|_a^b = \lim_{n\to\infty} \frac{b-a}{n} \sum_{k=1}^{n} g(a + k\frac{b-a}{n}) \tag{4.48}$$

Through the formal definition of the Riemann integral, equation 4.48 can be rewritten as:

$$\Delta CLV|_a^b = \int_a^b g(t)dt = \bar{v}\frac{\eta-\theta}{(1-\theta)(1-\eta)} \int_a^b \frac{f(t)}{e^{rt}} dt \tag{4.49}$$

The data provided by Farfetch consisted of a set of discrete data points with the projected growth in sales for the next years. By assuming that the average order value remains constant along-time and that it is acceptable to consider that a polynomial interpolation of these data points represents well enough the number of customers at any time within — and only within — the available domain, it is possible to substitute $f(t)$ in 4.49 by a general expression and evaluate the definite integral. This has the benefit of facilitating both the automation of the interpolation and subsequent calculation phase. Therefore, let $F(t)$ be a polynomial of order $p+1$ and $f(t) = \frac{dF(t)}{dt}$ be the resulting polynomial of order p; Equation 4.49 takes the following form:

$$\Delta CLV|_a^b = \bar{v}\frac{\eta-\theta}{(1-\theta)(1-\eta)} \int_a^b \sum_{k=0}^{p} \frac{\alpha_k t^k}{e^{rt}} dt = \bar{v}\frac{\eta-\theta}{(1-\theta)(1-\eta)} \sum_{k=0}^{p} \int_a^b \frac{\alpha_k t^k}{e^{rt}} dt \tag{4.50}$$

The last integral is a known identity. By substituting in equation 4.50, the final solution is given by:

$$\Delta CLV|_a^b = \bar{v}\frac{\eta-\theta}{(1-\theta)(1-\eta)} \sum_{k=0}^{p} \frac{(-1)^k}{(-r)^{k+1}} \left[\Gamma(1+k, rb) - \Gamma(1+k, ra)\right] \quad \text{where} \quad \Gamma(r,t) = \int_t^{\infty} t^{r-1} e^{-x} dx \tag{4.51}$$

The integral in $\Gamma(r,t)$ has no analytical solution and was evaluated numerically through a fixed-order Gaussian quadrature by using Python's *scipy.integrate.fixed_quad* package.

4.14.2 Estimation of parameters

Data provided by Farfetch concerning the future projection of sales growth considered a time horizon of three years, and so $a = 0$ and $b = 3$. Figure 4.9 shows the interpolation of the data points with a polynomial of degree 2 (Note that a degree 2 is equivalent to assuming that Farfetch's projected growth rate increases at a steady rate).

By assuming that the average number of orders per client and value per order remain constant over time, it is possible (and straightforward) to directly extract $f(t)$ from the curves shown in Figure 4.9.

In order to obtain the values for the remaining parameters of the retention model previously derived, an auxiliary model had to be first implemented to predict the average time until next purchase both for the warehousing (Δt_w) scenario and for the baseline (Δt). Parameters p_w and

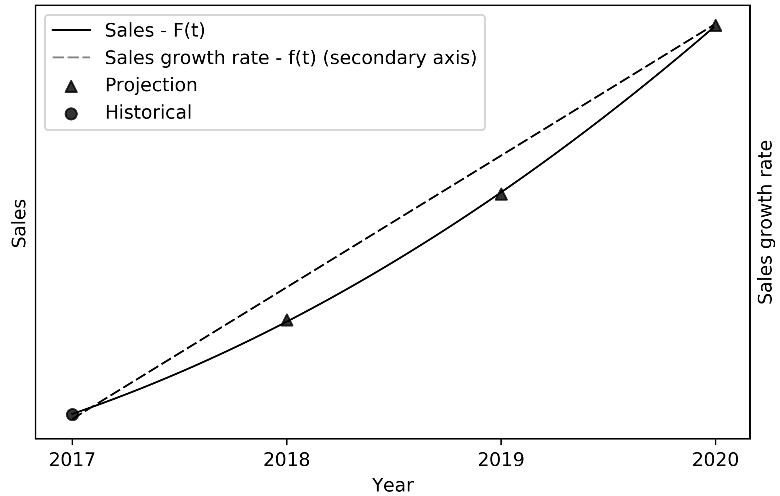

Figure 4.9: Interpolation of sales projection (p=2) and growth rate (p=1) — sales values omitted for confidentiality

$\bar{v}$ were obtained directly from historical data or from the simulation results (see Figure 4.10). Farfetch's Weighted Average Cost of Capital (WACC), was not provided due to confidentiality constraints, and thus the industry average from 2016, as reported by KPMG (2016), was used.

There may be many features involved in such prediction task (lead times, shipping costs, box accuracy, destination, warehouse, etc), hence, following the same rationale as with shipping costs and transit times, machine learning was used. The idea is that, for every order (historical and simulated), a time until next purchase (Δt_i and Δt_{w_i}) is predicted in order to ultimately calculate an average time per scenario (Δt and Δt_w); these averages are then fed into the model previously derived so as to obtain an estimate of the average CLV for each scenario and allow for a final comparison.

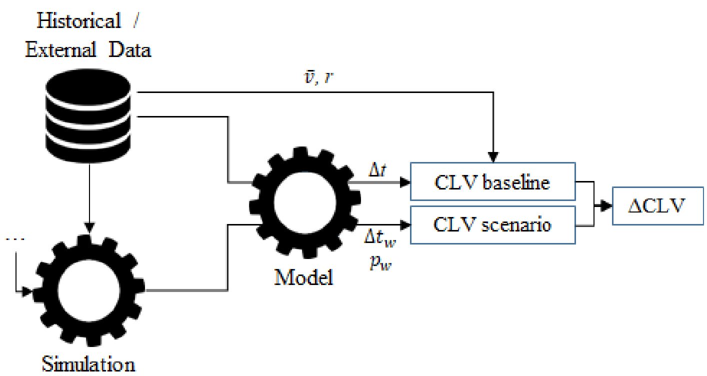

Figure 4.10: Impact of warehousing on retention — parameter estimation.

Time until next purchase was calculated from historical data of the last three years (a larger time horizon was used because of the nature of Δ_t, which in practice ranges from days to years). A Random Forest was trained and feature selection (based on mean decrease accuracy criteria) led to

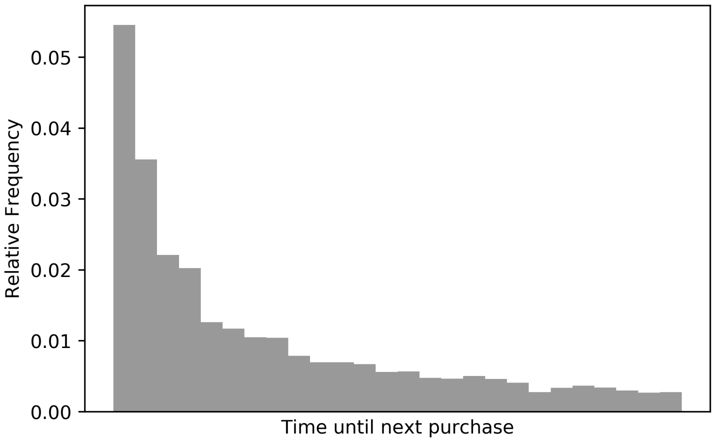

Figure 4.11: Histogram of historical Δ_t — highly skewed (values omitted for confidentiality)

considering the following features in the model (by decreasing order of importance): destination country, speed of sending, transit time, recommended box used (binary variable) and markdown. Due to the target variable being highly skewed (see Figure 4.11), a log-transformation was applied — as with shipping costs and transit times.

Finally, performance was assessed through k-fold cross-validation (refer to Table 4.3). Due to time constraints, no nested cross-validation was performed, and thus the performance after hyperparameter tuning risks being slightly optimistic.

The results show that even though the model possesses predictive power when compared to a naive approach, it may still incur in significant errors for practical purposes — especially when assessing impacts on retention by computing the present value of infinite sums discounted at a dubious rate and within a highly simplistic framework. Thus, the value estimated by retention drivers was not incorporated into the main simulation analysis but was rather addressed separately so as to provide a complementary, but highly uncertain assessment.

4.15 Acquisition

Modelling acquisition is complex due to the difficulty inherently present in establishing a relationship between performance improvements enabled by the warehouse and consequent acquisition of new customers. Besides that, this effect is not expected to be as powerful as retention, since

Table 4.3: Time until next purchase model (days) — mean average error and Root Mean Squared Error

Model	MAE	RMSE
Random Forest	9.61	293
Naive (mean historical Δ_t)	24.21	987

upgrades would not be easily perceived by non-customers — unless advertised, as may happen with new services such as same-day delivery — as opposed to existing customers, that would be directly subject to an improved experience. Hence, a conservative approach was adopted by neglecting the expected positive effect of acquisition resulting from the warehouse and assuming that additional demand would essentially be kindled by the retention lever. Finally, note that baseline acquisition — and the effect of retention on acquired customers — was considered in the model based on Farfetch's growth projections, either through formal inclusion of the effect in the model (as with the continuous acquisition component in the retention model) or through historical demand re-sampling — which aimed at generating artificial orders for simulation purposes.

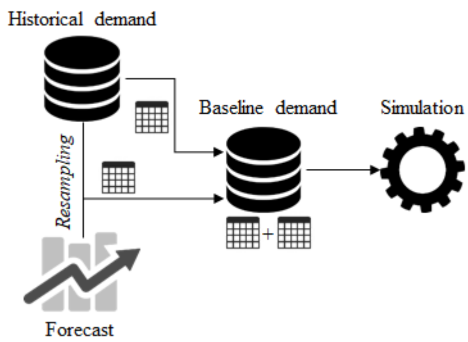

Figure 4.12: Growth projections were emulated through historical data re-sampling

Chapter 5

Simulation

This chapter's objective is to provide an overview of the simulation solution developed during the project in order to illustrate how it can constitute an intuitive, rational and powerful tool that fully leverages all the intelligence so far gathered and ultimately enables dynamic and actionable results. In order to avoid an overly detailed treatment, a high-level description of the simulation's components can be found in Appendix A.

5.1 Workflow overview

In this section, an overview of the simulation workflow will be provided so as to convey a broad idea of the steps that need to be taken to go from scenario specification until the outcome of results. This also allows foreseeing the purpose and interdependence of the distinct modules and how, ultimately, the latter enable the extraction of key business insights when the designed simulation workflow is followed throughout. Figure 5.1 shows a high level depiction of the major simulation stages.

The simulation consists of a standalone computer application capable of receiving inputs from the user, obtaining the necessary data, simulating a specific scenario and, finally, displaying the results.

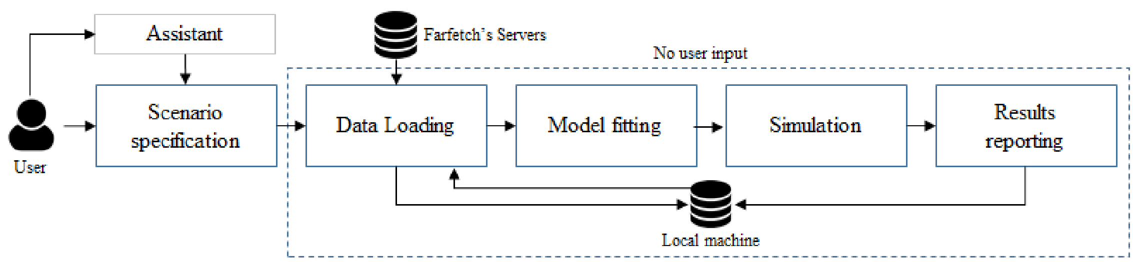

Figure 5.1: Simulation flow: from beginning to end

5.1.1 Scenario specification

The first step consists in defining the scenario to be simulated. In particular, the following information is specified:

Scenario The scenario is named and the simulation time window set by specifying a starting and ending date;

Warehouses Each warehouse to be simulated — in simultaneous — is instantiated, given a designation, a location, a service area and operational performance metrics. For each warehouse, it is also possible to specify, for example, in which days of the week it will operate, whether it is located in a free-trade-zone, what is the courier collection schedule, its priority in the order allocation algorithm and whether it accepts returns or not;

Product Selection Products and quantities to be shipped and stored at the warehouses are explicitly listed. In order to ease the task, an assistant program has been developed so as to find and filter products by partner and historical full price sell-through rate, for example — which are criteria actually used for product selection in Farfetch's business context;

Bulk shipments After the selected products and their quantities are listed, bulk shipments are created, origins and destinations set, products are allocated to each shipment and dispatch dates of the latter are scheduled. Each bulk shipment has a certain capacity (which can be manually specified for each instance, so as to increase realism and cost assessment accuracy). The reason for such granularity and for explicitly setting dispatch dates lies on the fact that timing is important in fashion retailing due to it being closely related to full price sell-through. As with product selection, this step can be assisted by the previously mentioned auxiliary program.

The scenario specification is self-contained in an object which carries all the previous information and can be saved locally so as to allow future retrieval.

5.1.2 Data loading

Once a scenario is specified, historical data is retrieved in order to be fed into the actual simulation procedure. For the sake of efficiency and resource utilisation, only the necessary data is obtained. Furthermore, once the data is retrieved from Farfetch's databases, it can be saved in a *csv* file; this way, the simulation can be run offline or without the need to query Farfetch's databases more than once. Below is listed the data required:

Products This data set contains information about the products selected, such as the category, ideal package volume, season and so on;

Orders This is the most important — but also the most resource consuming — data set. It contains information about all historical orders within the scenario time frame and locations subspace which will ultimately serve as an input to the predictive models. Besides that, a *demand* data set with the orders to be triggered in the simulation will be extracted from the former data set;

Stores The *stores* data set contains information such as the location, average commissions and average packaging cost for each store that plays a role in the simulation;

Miscellaneous Finally, auxiliary data sets are also obtained either from database queries, or alternatively from static and local *csv* files. Examples of such data sets are tables with duty rates or geographical data;

5.1.3 Model fitting

During this stage, data previously obtained is leveraged in order to instantiate all the different entities, such as locations, routes, orders, warehouses and other objects, and build or load predictive models which will subsequently be used during simulation run-time. Concerning the latter, key variables, such as shipping costs and lead times, are modelled either by fitting their empirical distributions or by using machine learning techniques (refer to the previous chapter for further details).

5.1.4 Simulation

After all entities and models have been instantiated, the simulation's time is set to the scenario's starting date and the actual simulation phase begins. Scheduled bulk shipments are dispatched, historical orders are initiated and the routing algorithm allocates orders sequentially to warehouses or existing stores. From the orders which are allocated to warehouses, order fulfilment performance is calculated based on their operational metrics, and costs and lead times are estimated by leveraging the predictive models. Finally, orders may be returned or not based on the new performance.

It is possible to specify a verbose level which will dictate how much information is displayed during simulation run-time. Besides that, it is also possible to display a synchronous chart with the different stakeholders' balance and a map which displays the distinct flows geographically. This way, it is possible to understand what the program is currently doing without having to go through the code. Finally, a log is saved after each iteration, mainly for debugging purposes. In Figure 5.2 is depicted an overview of the interface during run-time.

5.1.5 Results reporting

During the simulation phase, the engine keeps track of financial and other performance indicators while entities interact with each other. A baseline record is kept with the historical metrics so that a comparison basis exists and performance is assessed on a relative measure.

5.2 Programming the simulation

The simulation was programmed in Python 3.6 from the ground up due to the following reasons:

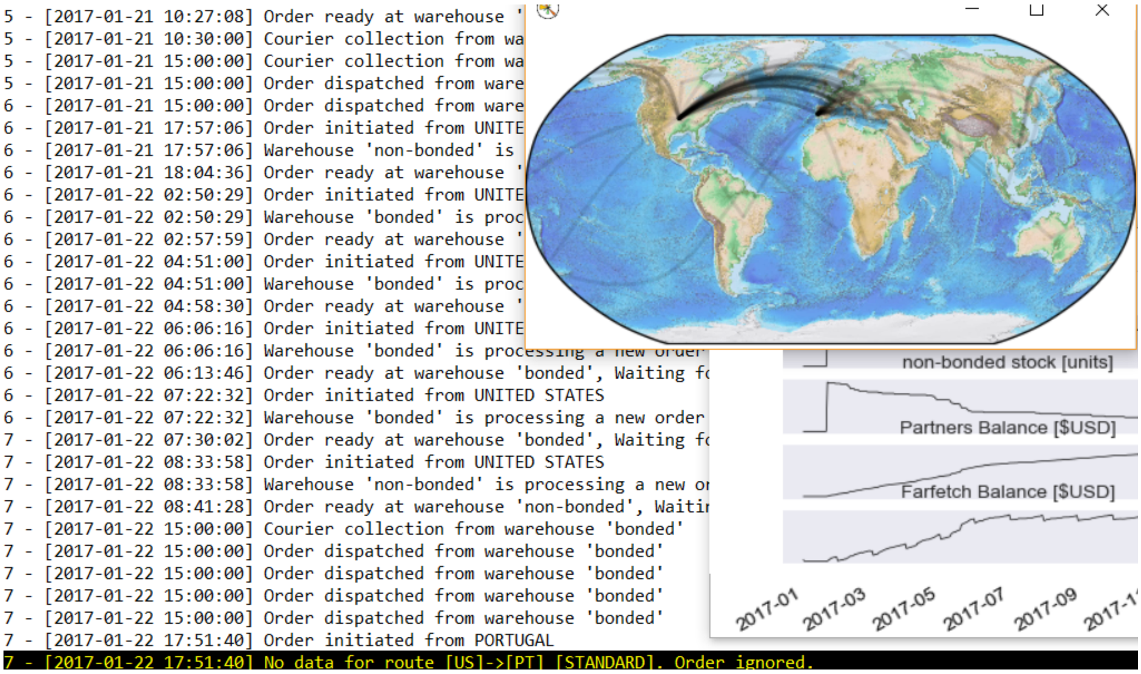

Figure 5.2: During run-time, a map, a chart and a log with events as they occur can be displayed and synchronously updated (iteration shown is non-representative and chart units have been hidden for confidentiality)

- First off, Python is a programming language adequate for prototyping purposes. Due to the complexity inherently present in the project and, in particular, to the resulting propensity that the latter has to change during its lifetime — for example, taxes and customs clearance rules suffered changes in terms of planned operating procedures for the warehouses due to legal and technical reasons —, a programming language capable of allowing fast and simple prototyping constitutes a clear advantage;

- Programming the simulation tool arguably results in more flexibility than using already existing simulation software solutions;

- By programming the simulation, it is easier to scale it and share it within the company due to its modularity, compatibility and lack of licensing requirements. Moreover, collaborative work is simpler due to the fact that different teams can work simultaneously in distinct modules or, similarly, distinct algorithms that were already developed within Farfetch's operations context can easily be replicated into the simulation framework;

- Concerning scientific computing, Python has many libraries available which allow a simple, fast, yet robust and reliable instantiation of advanced analytic modelling tools. In fact, machine learning and distribution fitting algorithms were used and instantiated during simulation run-time.

- Finally, commercially available simulation software needs licensing and has limited sharing capabilities. Regarding the project itself, and in particular given its scope and the need from

Farfetch's representative team for a modular and mutable decision support tool, the use of commercial simulation software lacks support in this context;

One of the main disadvantages of using Python is its slower speed when compared to other programming languages available, such as C or C++. Nevertheless, given the superior prototyping capability and scientific packages available regarding the former, it still constitutes an adequate use for the purpose. Other tools could potentially be used if this simulation solution was to be scaled and used with efficiency in mind, such as if it was to be part of an optimisation tool requiring repetitive instantiation of a value assessment component; however, this is outside the scope of this work.

5.3 Validation

In order to validate the solution developed, besides the individual validation tests performed for each model included in the simulation, some additional and more general tests were performed by simulating random samples of historical orders and comparing the results with actual performance. The simulation's logs were set to the highest verbose setting and the different events carefully monitored in order to guarantee that they followed logical and realistic sequences that adequately represented the actual processes.

Figure 5.3 summarises the validation of the simulation as a whole. For the purpose, 150 samples with 50 orders each were simulated exactly as they happened (hence, without the warehouses) and the mean gross margins per order calculated, both for the simulated and the historical results. This validation does not take into account effects such as retention or duty savings due to the free-trade-zone, for example, because these components do not exist in historical data.

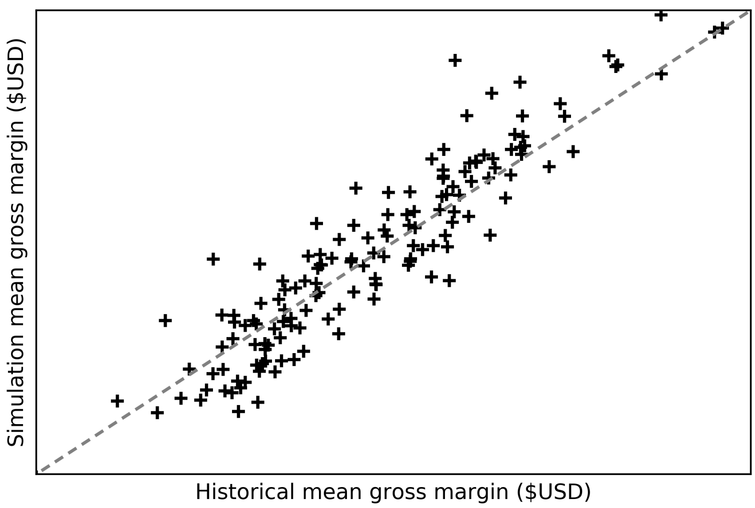

Figure 5.3: 50 orders per observation and results reported on a per order basis (horizontal and vertical axis have the same scale)

According to Figure 5.3, the simulation does not seem to have any bias, because observations appear to be equally scattered around the diagonal. There are however, some prediction errors, which are expected. The extent up until which errors are considered acceptable is a somewhat subjective criteria and, due to this, accuracy had to be acknowledged by Farfetch analysts (Mean Average Error and Root Mean Squared Error were used). It is also important to understand that the errors measured may not correspond to the real errors since the simulation results are being compared to historical data and at least two effects may occur: data itself may be inexact or the calculation method may be wrong and omit specific steps or cost items. The latter case is particularly dangerous since the whole model and comparison basis are being built based on wrong assumptions.

5.3.1 Machine learning *versus* empirical distributions

As found during research, machine learning combined with simulation is a largely unexplored topic. In order to understand its contribution to the present case without excessively drifting away from the main discussion, the same previous test was replicated by substituting machine learning models in the simulation with calls to functions that return a random value from the empirical distribution function of the specific variable (actually, from a computational standpoint, the truncated inverse cumulative empirical distribution function was used). The outcome is shown in Figure 5.4.

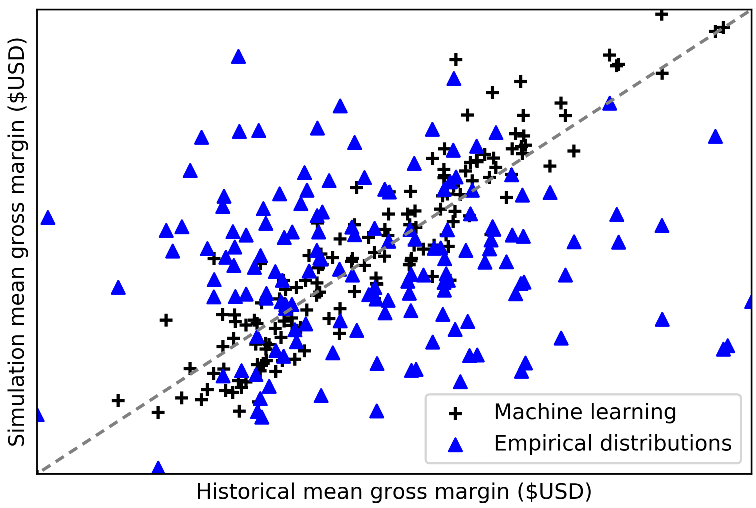

Figure 5.4: Simulation with machine learning models v.s. empirical distributions — 50 orders per observation and results reported on a per order basis (horizontal and vertical axis have the same scale)

Overall, both methods seem to result in unbiased results, yet using empirical distributions clearly leads to a higher dispersion. This is somewhat expected because, during the modelling stage, machine learning models showed superior predictive performances when compared to baseline comparison basis (using the average or sampling from the empirical distribution, for example).

One evident advantage, though, shows itself when the simulation logs are compared for orders simulated with different methods. The shipping costs obtained for a given route by using empirical distributions did not make much practical sense: heavier products occasionally had lower shipping costs than lighter products. This happened because shipping costs were generated randomly from the route shipping costs' empirical distribution function. On the other hand, results obtained by using machine learning models in the simulation made much more sense because they incorporated this knowledge. Therefore, using machine learning can also provide further insights by simulating elementary fluxes with higher realism — or, in other words, it is dangerous to analyse elementary fluxes when variables are random variables. Besides that, machine learning is a powerful technique when it comes to finding complex patterns in data.

Regarding disadvantages, machine learning models do not only need large datsets to be trained on, but they also require that the data itself be rich enough. Whereas empirical distributions based on historical data enable the generation of extreme tail values, this may not happen with machine learning models. For instance: there are orders which have had, historically, extreme lead times due to exceptional reasons. Sampling from empirical distributions would somehow capture these cases, whereas machine learning might fail at detecting such occurrences. This is a disadvantage when factors such as supply chain robustness or response to exceptional events, for example, are being studied.

Chapter 6

Results and discussion

In this chapter, the major results and knowledge obtained from the simulation analysis will be reported. Since certain business insights and metrics were flagged as confidential by Farfetch, some of the findings were omitted and values are reported relatively to the company's historical performance in 2017 (unless otherwise specified). Moreover, due to computational constraints, iterations were only simulated one single year at a time — for example: results for 2020 were based on inflated (through re-sampling) historical demand from 2017. Finally, financial results were discounted based on the estimate obtained for Farfetch's Weighted Average Cost of Capital and reported to the beginning of 2018's autumn-winter season (AW18).

6.1 Main scenario

The main scenario represented the investment project as originally projected by Farfetch, consisting of both a warehouse in a free-trade zone and a non-bonded warehouse used in conjunction for products priced under the United States *de minimis* threshold and for handling returns.

Figure 6.1 depicts the relationship between orders and markdowns obtained for AW18. The earlier products are online, the more likely these will be sold at their full price — this has always been important both for partners and for Farfetch, since both parties are affected by markdowns. However, subsequently inferring that it is economically rational to ship products to the warehouse as soon in the season as possible — and even before it begins — might be fallacious. The reason for this lies in an important difference between the baseline business model and the warehousing scenario: inventory holding costs that depend on time. It is indeed true that the overall probability of ending the season having sold a product at full price increases the earlier it has been put online, but Figure 6.1 suggests that the instantaneous probability of selling the given product increases, on average, the other way around (until the first markdown peak). In other words: efforts to sell at full price might be counterproductive if they increase holding costs up to a threshold where this scenario performs worse than the baseline with no warehouses.

Figure 6.1 also suggests a possible tactic from which partners can benefit: because most boutiques are sensitive to peaks due to limited capacity, partners can anticipate them by bulk shipping certain products to the warehouse before peak times, which are somewhat predictable.

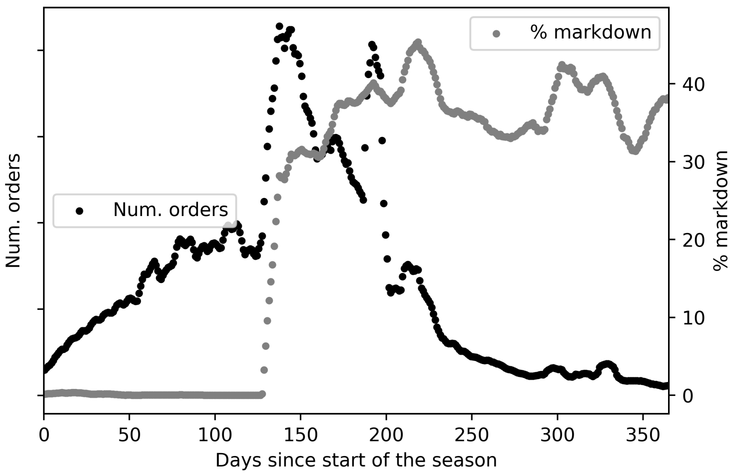

Figure 6.1: 7-day moving average of volume and markdown — example from simulated AW18.

This is something that they could not do in the past but that can potentially become a selling argument for Farfetch.

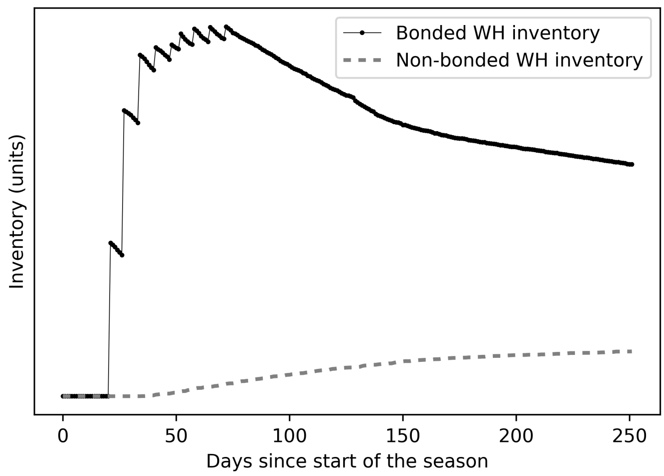

Figure 6.2: Time evolution of warehouses' inventories — example from simulated AW18.

Figure 6.2 shows the inventory of both warehouses along time. Whereas inventory decreases in the bonded warehouse, the non-bonded warehouse receives returned products and can possess an increasing inventory over time. This is not an undesirable effect when compared to the baseline scenario if the average total holding costs are inferior to the costs involved in returning the product to the boutique overseas.

Concerning the performance along a time horizon of three years, according to a roll-out plan suggested by Farfetch, Figure 6.3 shows that the project would likely result, on an absolute basis, in a positive operating income, both for Farfetch and partners, on average — even with the inclusion of one-off costs, which were omitted in Figure 6.3. This is a solid insight, because it means that

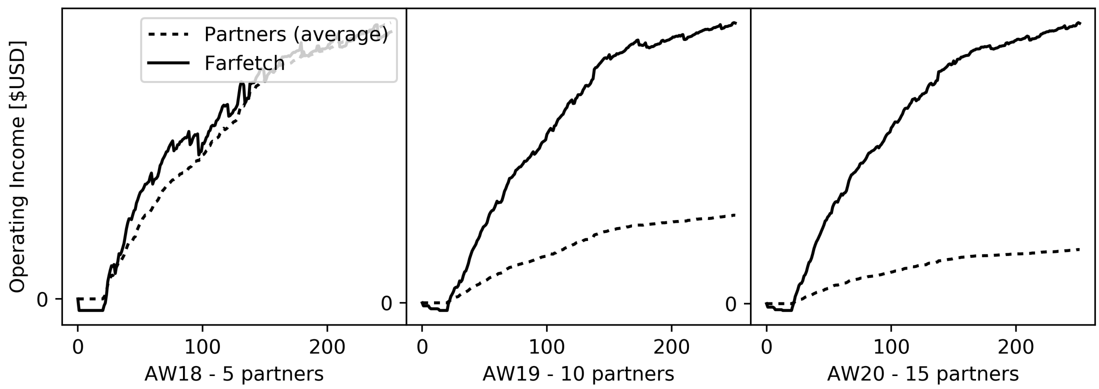

Figure 6.3: Main scenario absolute operating income for AW homologous seasons (vertical axis with different scales between charts and one-off costs omitted)

the project will tend to generate positive income by itself — independently of whether it is in fact superior to the baseline scenario or not — and, thus, alleviates the potential risk of losses.

Nevertheless, when assessed on a relative basis, the warehousing project results in losses for Farfetch during the first year when compared to the baseline (see Figure 6.4). The reason for this is evidenced by the accentuated dips present in Farfetch's curve from AW18 and the lack thereof in the curves from AW19 and AW20; whereas scale is sufficient to dilute the fixed costs in AW19 and AW20, the potential volume brought by five partners is not sufficient to result in a relatively more profitable scenario when compared to the baseline. However, as more partners venture into the initiative, the sum of individual contributions per order exceeds the fixed costs and warehousing as a service becomes a profitable project in the long-term.

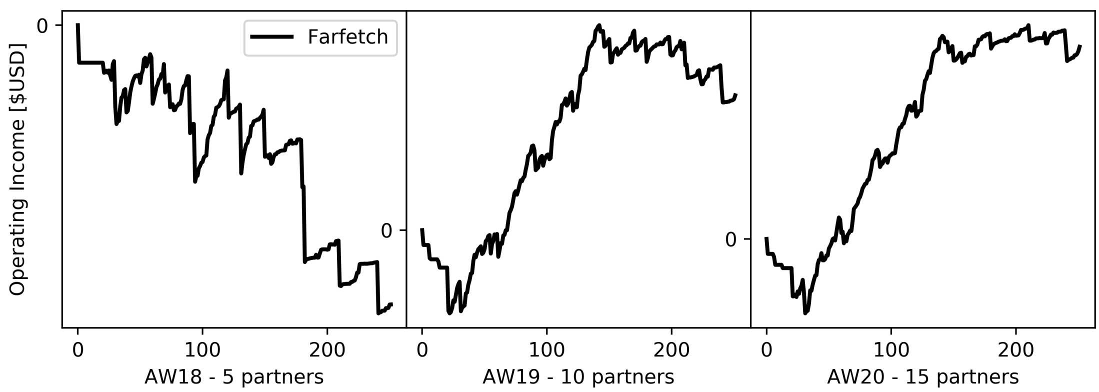

Figure 6.4: Main scenario relative to baseline operating income for AW homologous seasons (vertical axis with different scales between charts and one-off costs omitted)

Regarding partners, results for the comparative analysis have been omitted due to their confidential nature. The numerical outputs obtained from the simulation, though, confirm the qualitative assessments: partners have many savings, especially related to packaging, storage, handling and shipping and, despite the contrary belief that supported the original rationale for the project, they also have some opportunity costs associated to each order that goes through the warehouse — *i.e.:* incentives and recoveries. Concerning incentives, partners would not receive both labour and performance incentives for orders that would be fulfilled in a warehouse, as opposed to the baseline scenario. Regarding recoveries, some partners do not fully refund returned orders, giving rise to a positive surplus at the end of the season. On the other hand, within the warehousing scenario, because Farfetch would have greater control over the inventory and return acceptance policies, partners would lose this additional revenue.

Table 6.1 summarises the variation potentially induced by the warehouses on the main cost and revenue items, from Farfetch's perspective. Due to the free-trade zone warehouse, linked to possibility of handling returns in the domestic market through a non-bonded warehouse, Farfetch would considerably save on taxes. Regarding Customer Service *order status* tickets, even though these have a relatively low weight on Farfetch's income statement, a noticeable reduction is also evident for orders fulfilled in a warehouse.

Table 6.1: Warehousing *versus* baseline scenario — Farfetch's perspective (average from 10 simulation runs)

Item	AW18	AW19	AW20
Merchandising revenue ($USD)	0.0 (%)	0.0 (%)	0.0 (%)
Commissions charged	0.0	0.0	0.0
Recoveries from partners	+39.9	+39.9	+40.8
Others received	+4.1	+4.9	+6.9
Duties & taxes received from customers	0.0	0.0	0.0
Commissions lost	-2.6	-3.2	-2.8
Discounts offered	0.0	0.0	0.0
Duties & taxes paid	-12.0	-10.5	-9.7
Refunds issued to customers	-8.2	-9.4	-9.1
Retail price paid to partners	0.0	0.0	0.0
CS tickets costs	-41.0	-54.2	-43.4
Shipping costs paid	+2.4	+1.2	+2.7
Other costs	+1.7	+1.7	+1.6
One-off costs	new	-	-
Bulk Shipment costs	new	new	new
Labour Support incentives paid	none	none	none
Performance incentives paid	none	none	none
Monthly costs	new	new	new
Packaging costs	+18	+18	+18
Others paid	+1.0	+1.1	+1.3

Concerning performance metrics related to the customer, Table 6.2 shows solid results. For orders fulfilled at a warehouse, the speed of sending would be drastically increased due to higher efficiency and weekend operation, for example; returns would also diminish due to lower wrong-item rates and retention would increase — the model predicted that retained customers during the three seasons simulated would bring a net profit of $284\bar{v}$ (with $\bar{v}$ being the mean net profit per order) during their lifetime.

Table 6.2: Warehousing *versus* baseline scenario — performance metrics (average from 10 simulation runs)

Metric	AW18	AW19	AW20
Returns (% orders)	-4 (%)	-11 (%)	-13 (%)
Speed of sending (% days)	-89	-87	-86
Transit time (% days)	-7	-3	-4
Retained customers (CLV)	284x historical net profit per order		

6.2 Sensitivity Analysis

The goal of this sensitivity analysis is to understand how much results change when one parameter of the simulation changes. This provides an idea of the relative importance that certain factors may have in the outcome. In fact, it is highly unlikely that actual execution will follow the exact roll-out projection and, thus, it is important to get a grasp of the existing risk.

6.2.1 Bulk shipment timing

In order to understand the potential impact that bulk shipment timings have on performance, three scenarios were conceived in which bulk shipments were spaced exactly as in the main scenario simulated, but with differing dates for the dispatch of the first shipment. Results are shown in Figure 6.5.

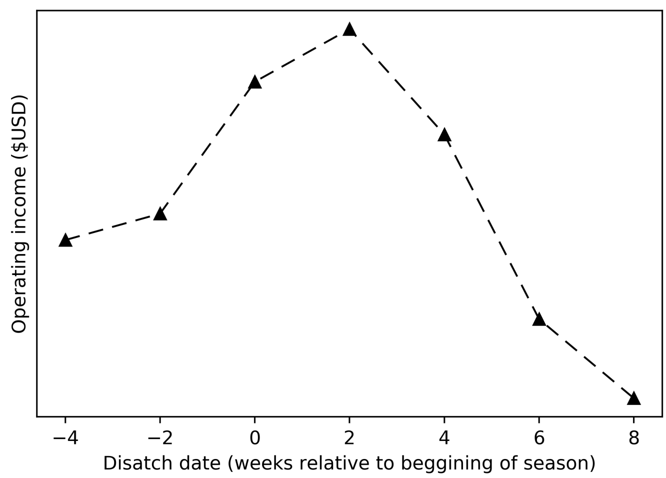

Figure 6.5: The effect of bulk shipment timings

As suggested by the results, there seems to be an optimum timing for bulk shipments. The reason for this is likely due to the trade-off between either having a higher sell-through and full price mix in spite of inventory holding costs *versus* lower sell-through, higher markdowns but lower holding costs. Even though shipment timing can be optimised, the key insight to retain from

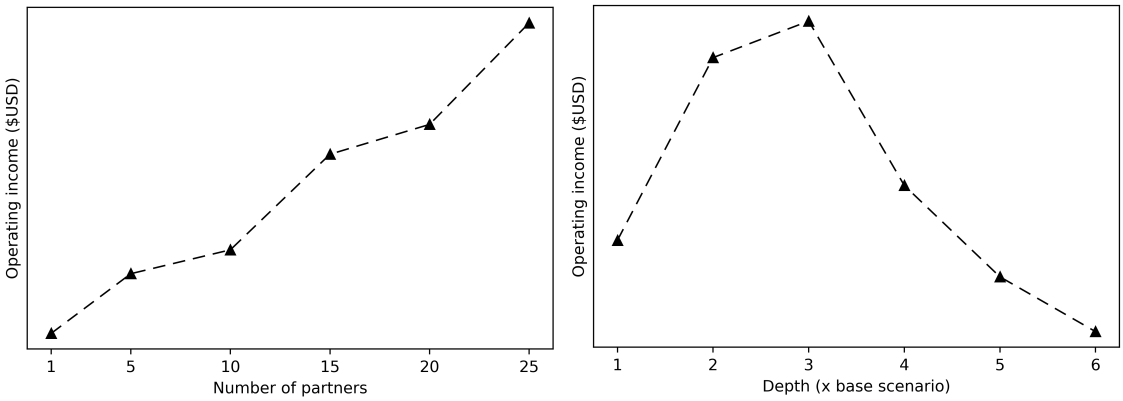

Figure 6.6: The effect of breadth and depth

the analysis is that, contrary to first belief, having items available as early as possible only makes sense for Farfetch when not in a warehousing context.

6.2.2 Breadth and depth

As suggested by the results so far, the project requires a minimum scale in order to help dilute fixed costs and become profitable. Concerning supply, there are two important levers: increasing breadth, which is deeply related to the number of partners that join the initiative; and increasing depth per product, which is related to how much supply is provided by each of the partners already in.

The first analysis was performed by simulating a common scenario with different number of partners but exact same depth in total, so as to simulate divergences in breadth. The second analysis was performed by simulating a common scenario with supply originating from the exact same number of partners, but with differing stock depth in total. Results are summarise in Figure 6.6.

The main insight that can be extracted from the analysis is that increasing depth is effective only up to a certain point for a given level of breadth. Scaling seems to be viable only by increasing breadth through higher partner adherence. Thus, during the early stages, Farfetch should not only focus on maintaining existing partners and negotiating optimum depth, but, more importantly, convince more partners to join the project. This can be more easily achieved by offering entry incentives, for example.

6.2.3 Product selection criteria

The goal of this analysis is to understand the effect that a given product selection criteria may have on performance. Because product selection is a decision highly dependent on negotiation with partners, it is unlikely that a single criteria adequately represents reality. One interesting proxy to consider, though, is the minimum historical sell-through enabled by Farfetch; in other words: select products from a given partner that have had a historical minimum of sales through Farfetch over a specific relative threshold

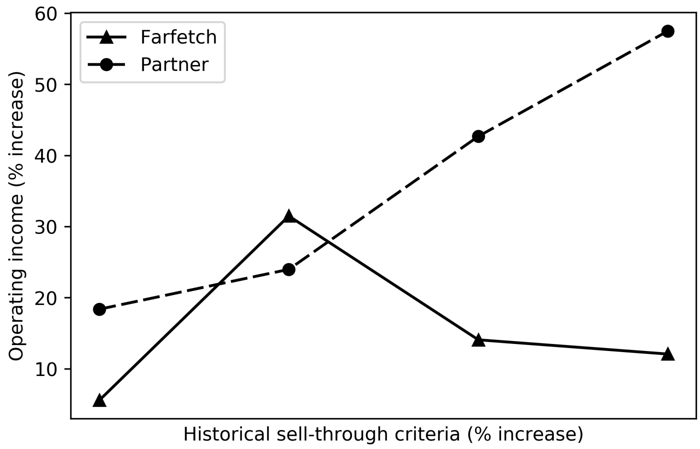

Figure 6.7: Operating income *versus* historical sell-through of supply; results relative to AW18 simulated with a historical sell-through of 50%

Figure 6.7 shows that products with higher sell-through scores lead to more profitable scenarios for partners. However, there appears to be an optimum configuration for Farfetch. This is due to the fact that the higher the sell-through requirement is, the less products are available. This means that, even though a high sell-through criteria may result in almost all the inventory having been sold at the end of the season, the actual quantity may not have been enough to dilute the fixed costs. On the other hand, a low sell-through criteria may lead to a higher inventory size, yet, because a relatively smaller percentage of items end up being sold, holding costs per unit are higher. Figure 6.7 suggests that there may exists an optimum sell-through — in practice, though, this point is not known before the season and has to be predicted, for example through historical sell through for related products.

Finally, the curve shown in Figure 6.7 for partners may be misleading: a higher sell-through criteria for warehouse product selection results in a more profitable outcome only when compared to not warehousing these products but having them available at Farfetch anyways. Because in practice sell-through for a given product is not known beforehand and can only be estimated through forecasts, partners have in fact three choices available for the product: selling through Farfetch and offline, selling through Farfetch's warehousing services only or selling offline only. Figure 6.7 considers the first two options. Since partners have greater profits per order when selling a product offline (with lower or equal probability), introducing this opportunity cost may result in a different curve, because a higher sell-through with Farfetch may be correlated with a higher offline sell-through, and thus partners may not have incentives to allocate their best offline performing products to Farfetch. This means that partners may also have an optimum point due to the selling offline-only opportunity cost and, thus, actual negotiations for product selection may turn out to be more complex than suggested by Figure 6.7. The key point to retain, though, is that, for partners selling a product through Farfetch, results suggest that it is more profitable to sell it through Farfetch's warehousing services rather than through boutique fulfilment.

Chapter 7

Conclusions

This final chapter aims at concluding the project by sharing the main insights regarding the methodology employed and the actual case study addressed.

7.1 On the methodology

In order to tackle the problem, a multidisciplinary approach was employed. Regarding the preliminary qualitative analysis, both the current business processes and the hypothetical scenario were explored in order to help build a conceptual model and, overall, a thorough understanding of processes and key cost and value drivers. This step was crucial, since the actual modelling and simulation stages were deeply biased by the insights and maps extracted during this phase.

A brief strategic analysis was also performed in order to help understand the project's rationale and pertinence within the organisation and other stakeholders' competitive frame. This stage was also found to be essential for addressing supply chain management problems, because it allows to uncover potential risks or opportunities not captured by the quantitative model. As an example related to the case study analysed, the simulation showed solid results for a given level of partner adherence; yet, the strategic analysis suggests that partner adherence should by no means be taken for granted and requires, in fact, careful negotiation.

Concerning the modelling stage, different techniques were used when considered appropriate. When large amounts of data were available for rather complex sub-problems, modelling was accomplished by using machine learning tools. This approach has seldom (or not at all) been described in the literature and, thus, this constituted a somewhat novel and experimental approach. Overall, machine learning tools proved to be powerful to capture essential knowledge in the data while, simultaneously, providing a simple way to access it directly from the simulation program. The main disadvantage found is that they require, in fact, large amounts of data and may fail at predicting outliers, which may play an important role when it comes to quantifying risk. When using empirical distributions, however, the simulation may lose realism in more granular fluxes due to generalisation and thus fail at capturing specific behaviours and, up to a certain extent, causal relationships, when applicable. As a conclusion from this work, it is recommended that machine learning be used generally when large amounts of data are available, when historical data does not generalise well to the simulation's scenario but may be used to learn from and issue better predictions or when alternative techniques are limited or lack capabilities to model complex structures and potential relationships, especially when a subsequent analysis of more elementary fluxes is important as a result.

Despite the numerical and computational components of such techniques, a mathematical model for retention had to be deduced due to the necessity and lack of adequate solutions. The model itself allowed to gain relevant insights regarding the potential impact of the investment project on retention and quantify the latter. Besides that, the anlytical model combined with a discrete-event simulation environment brought forward further evidence for the superiority of multidisciplinary approaches when assessing the value of projects within a supply chain management context.

The complex business processes were then simulated in a dedicated program fully developed for the purpose. Programming the solution was found to be adequate because of the flexibility — especially when instantiating machine learning or other advanced models —, the modular approach, the platform independence and the final solution being license-free. This does not mean, however, that commercial solutions are poorer; in fact, the latter have powerful and optimised modelling capabilities that allow, for example, much faster development and execution or interactive visualisations. Therefore, the platform is a case dependent decision.

Finally, the outputs of the simulation were studied and compiled into meaningful results. A sensitivity analysis was also performed and allowed to obtain further insights regarding the risk of the project and the potential for optimisation.

7.2 Case study recommendations

Concerning the case study analysed, the project would bring supply chain wide savings by improving the efficiency of the system as a whole. Within the supply chain itself, there would be a rearrangement of cost and value drivers among different stakeholders but, overall, the configuration would bring solid and positive results in the long term to every player. When appropriately scaled, this constitutes a profitable investment project.

Nevertheless, issues may emerge during the deployment phase and compromise the success of project escalation. The key factor found to be decisive was partner adherence, which somewhat constituted a model assumption required for such positive results — especially from Farfetch's standpoint —, but which in practice may require careful planning and negotiation. Partners may have opportunity costs not captured by the model and, besides that, the financial benefits brought by the project may be difficult to communicate in an effective way, despite the evidence for their existence. Shipping products to a warehouse may be perceived as an alienation of the boutiques' core strategy. Due to these reasons, it is recommended that Farfetch clearly communicates to partners the potential for long term benefits, both through financial and strategical reasonings, and that adherence and retention incentives are provided — for example, with sell-through guarantees, competitive pricing and collaborative work. It may also be essential to negotiate specific terms regarding withdrawal timings and penalties, especially during roll-out.

Finally, if this particular project succeeds, it may be the first of many more to come. In fact, this might just be a proof-of-concept. There is a strategic rationale that supports an integrated warehousing solution, with such infrastructure optimally scattered around the world. The potential is huge. For the time being, this constitutes a first step towards attaining Farfetch's vision:

> "Our vision is to be the global technology platform for the luxury fashion industry, with our customer at its heart." — Farfetch.com

7.3 Future work

The main objective of this work was to assess the value of a supply chain management project. Yet, as suggested by the results, there is potential for optimisation. As found during research, using machine learning tools on simulation generated data to learn from and ultimately find optimal solutions might be a powerful technique, yet has been largely unexplored. Therefore, a possible extension of this work could be the combined use of machine learning and simulation tools for optimisation purposes.

Bibliography

Andy Sadler, L. (2006). Using survival analysis to predict sample retention rates. *U.S. Bureau of Labor Statistics*.

Beamon, B. M. (1998). Supply chain design and analysis: Models and methods. *International Journal of Production Economics*.

Buckley, S. (2005). Supply chain management on demand. *Springer*.

Byung-cheol Kim, K. (2009). Probabilistic forecasting of project duration using bayesian inference and the beta distribution. *Journal of Construction Engineering and Management*, 135(3).

Caroline Thierry, André Thomas, G. (2006). Simulation for supply chain management. *Wiley-ISTE*.

Chopra, S. (2003). Designing the distribution network in a supply chain. *Transportation Research Part E*.

Clark, C. E. (1962). The pert model for the distribution of an activity. *Operations Research*, 10.

D. Greenberg, A. (2010). Beta-distribution models in stochastic project management. *Computer Modelling and New Technologies*, 14(4).

Desmet, B. (2018). *Supply Chain Strategy and Financial Metrics: The Supply Chain Triangle Of Service, Cost and Cash*. Kogan Page Publishers.

DirkVan den Poel, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1).

D.P. Christy, J. (1994). Safeguarding supply chain relationships. *International Journal of Production Economics*.

E. Feldman, F. A. Lehrer, T. (1966). Warehouse location under continuous economies of scale. *Management Science*.

Erdem Eskigun, Reha Uzsoy, P. G. S. J. (2005). Outbound supply chain network design with mode selection, lead times and capacitated vehicle distribution centers. *European Journal of Operational Research*, 165.

Fisher, M. L. (1997). What is the right supply chain for your product. *Havard Business Review*.

FTZC (2018). Weekly entry in a foreign-trade zone offers reduction in merchandise processing fees (mpf) and brokerage fees.

Gavin C. Cawley, N. (2007). Preventing over-fitting during model selection via bayesian regularisation of the hyper-parameters. *Journal of Machine Learning Research*, 8.

Gavin C. Cawley, N. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11.

GEA (2016). Overview of de minimis value regimes open to express shipments world wide.

Ingalls, R. G. (1998). The value of simulation in modeling supply chains. *Proceeding of the '98 Winter Simulation Conference*.

John Fernie, L. (2009). *Logistics and Retail Management*. Kogan Page Publishers.

John T. Mentze, William DeWitt, J. S. N. C. Z. (2001). Defining supply chain management. *Journal of Business Logistics*, 22.

June Young Jung, Gary Blau, J. G. D. (2004). A simulation based optimization approach to supply chain management under demand uncertainty. *Computers and Chemical Engineering*.

Kirandeep Chahal, T. (2008). Applicability of hybrid simulation to different modes of governance in uk healthcare. *Proceedings of the 2008 Winter Simulation Conference*.

KPMG (2016). Cost of capital study.

L. Rabelo, M. Helal, A. J. (2005). Enterprise simulation: a hybrid system approach. *International Journal of Computer Integrated Manufacturing*.

Luis Rabelo, Hamidreza Eskandari, T. M. (2006). Value chain analysis using hybrid simulation and ahp. *International Journal of Production Economics*.

M. Christopher, R. Lowson, H. (2004). Creating agile supply chains in the fashion industry. *International Journal of Retail and Distribution Management*.

Mark S. Daskin, Lawrence V. Snyder, R. (2005). Facility location in supply chain design. *Logistics Systems: Design and Optimization*.

Martin Christopher, Robert Lowson, H. (1992). Creating agile supply chains in the fashion industry. *International Journal of Retail & Distribution Management*, 32.

Matthew Liberatore, T. (2010). A supply chain strategic planning framework. *Northeast Decision Sciences Conference*.

Michael Treacy, F. (1992). Customer intimacy and other value disciplines. *Harvard Business Review*.

M.T. Melo, S. Nickel, F. (2009). Facility location and supply chain management – a review. *European Journal of Operational Research*.

NPD (2016). Off-price consumers account for 75 percent of all apparel purchases across retail channels.

Pashigian, B. P. (1988). Demand uncertainty and sales: A study of fashion and markdown pricing. *The American Economic Review*, 78(5).

Qi Hao, W. (2007). Implementing a hybrid simulation model for a kanban-based material handling system. *Robotics and Computer-Integrated Manufacturing*.

Raschka, S. (2018). https://sebastianraschka.com/faq/docs/evaluate-a-model.html.

Sang-Eun Byu, B. (2008). The antecedents of in-store hoarding: measurement and application in the fast fashion retail environment. *The International Review of Retail, Distribution and Consumer Research*, 18(2).

Sharma, P. (2015). Discrete-event simulation. *International Journal of Scientific & Technology Research*, 4.

Stevens, G. (1989). Integrating the supply chain. *International Journal of Physical Distribution*.

Subramanian, Pekny, R. (2001). A simulation-optimization framework for research and development pipeline management. *Process Systems Engineering*.

Teifenbrun, S. (2015). U.s. foreign trade zones and chinese free trade zones: A comparative analysis. *Journal of International Business and Law*, 14(2).

Timo Deist, Andrew Patti, Z. D. T. D. (2018). Simulation assisted machine learning. *Harvard Medical School*.

U.S. Customs and Border Protection (2018a). About foreign-trade zones and contact info.

U.S. Customs and Border Protection (2018b). User fee - merchandise processing fees.

Virginia L. M. Spiegler, Mohamed M. Naim, D. J. (2016). A technique to develop simplified and linearised models of complex dynamic supply chain systems. *European Journal of Operational Research*.

Walid Budgaga, Matthew Malenseka, S. N. F. S. (2015). Predictive analytics using statistical, learning, and ensemble methods to support real-time exploration of discrete event simulations. *Future Generation Computer Systems*.

Y. Chang, H. (2001). Supply chain modelling using simulation. *International Journal of Simulation*.

Z. Xu, E. . E. . (2001). Simulation with learning agents. *Proceedings of the 2001 IEEE*.

Appendix A

Simulation components

This appendix contains the distinct classes and modules that make up the simulation program as a whole. As depicted in Figure A.1, the simulation developed has 12 modules; yet, only the most important ones will be detailed in dedicated subsections and, the remaining, briefly described down below:

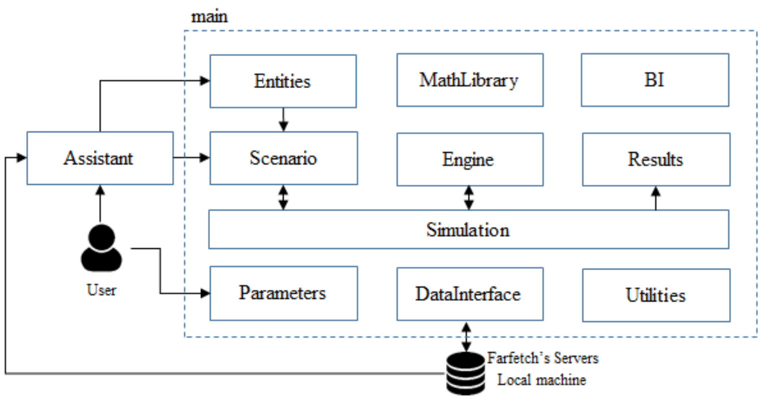

Figure A.1: Simulation modules (only the most important fluxes of information were represented)

Assistant The *Assistant* is a standalone and somewhat independent module that interfaces between the user and the simulation program. The objective is to ease the task of inserting information by providing a more user friendly interface and, simultaneously, automating the task of generating data, for example by automating product selection by following user-defined criteria, allocate products to warehouses, schedule bulk shipments, etc. Figure A.2 shows an example.

Scenario The *Scenario* module, as the name implies, stores information about the specific context that is being simulated and contains functions that instantiate objects and populate the datasets required for the simulation based on raw input obtained from the user, from Farfetch's server or from local files. Instances of *Scenario* can be saved for future use;

Engine The *Engine* module contains all the functions responsible for preparing the simulation — such as scheduling events or training predictive models — and actually running and coordinating the actual simulation;

MathLibrary This module contains functions that are mathematical in nature, such as interpolation or regression tools;

```
Season (leave empty to end list):  AW17
Season (leave empty to end list):  SS17
Season (leave empty to end list):
Select geopriced products only? (Y/N) [True]:  N
What is the minimum historical full price rate per product? [0.5]:  0.7
What is the minimum historical number of orders per product? [0]:
Select only from a specific list of stores? (Y/N) [True]:  Y
Please insert stores to select, one by one:
Store name (leave empty to end list):  STORE_NAME1
Store name (leave empty to end list):  STORE_NAME2
Store name (leave empty to end list):
Extracting products from Database...
```

Figure A.2: Example of the Assistant

BI This module contains models that carry information about the business, such as shipping costs or transit times predictors. It is arguable whether this component deserves a separate module, yet it was decided to follow this procedure for the sake of clarity and organisation;

Results The results module contains a class (*Results*) which contains simulation variables and the transaction account of each stakeholder. As the name implies, this module is used to gather and report results;

Simulation The simulation module serves a linking purpose: it interfaces *Scenario*, *Engine* and *Results* so that they are able to communicate throughout the whole simulation;

Parameters This component carries default parameters that the user can set prior to running the simulation. Among the former, there are parameters that relate to the scenarios being simulated (such as default cost rates), to the simulation models (such as default p-value acceptance threshold) and to the program's configuration (such as whether it should display maps or charts);

DataInterface This module contains functions that automate data related procedures, such as saving or loading a *Simualtion* instance, or extracting online data;

Utilities The *Utilities* module contains miscellaneous functions that support the program's operations, such as format conversion procedures, a logging function, and more;

main Finally, *main* is where the simulation is initialised and run;

A.1 Entities

The entities module contains both classes which are directly related to real-world concepts or tangible agents and classes which represent somewhat functional entities aimed at supporting the simulation's operation.

Entities from the first group will be first described:

Location The class *Location* contains attributes which store information about geographies, for instance: country, country code, city, region, latitude and longitude. The class also contains auxiliary methods which, for example, help find specific location objects on iterables.

The reason why it was decided to represent locations as objects and not as string vectors, for example, is that the former convey idiosyncrasy to location instances and, by keeping them unique simulation-wide, it is possible to store performance metrics — or, for that matter,

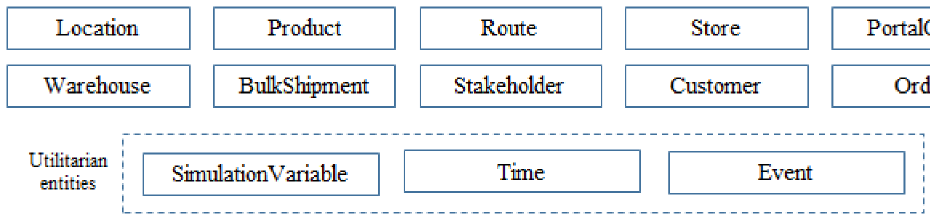

Figure A.3: Classes of the *Entities* module:

add or change any other attribute — for each location and guarantee that changes are applied globally, because location objects are used by reference to the same entity. Additionally, code becomes much more intuitive and easier to understand.

Route This class contains an origin and destination attributes, which are *Location* objects, as well as a service type attribute that relates to whether the route is express, standard, F90 or same day delivery. Besides that, *Route* also has attributes which store references to map objects, used for geographical visualisation in a dedicated map.

In the same fashion as *Location*, this class also has some auxiliary methods which help find object instances or, for example, allow plotting, highlighting or changing the transparency of their visual representations.

Finally, the *Route* class also contains reference holders for particular functions which call instances of regression models that ultimately predict shipping costs (for customers, for Farfetch or for returns) and lead times (deliveries to customer or returns). The reason why references to these predictive models are stored in *Route* objects is because route origin, destination and service are important features and, therefore, are passed directly to the model when called from *Route*. Other features — such as weight for Farfetch shipping costs — may also be passed explicitly.

Warehouse This class contains some attributes which are specified by the user, such as: name, location (which is a *Location* object), whether the warehouse is in a free-trade-zone, which are the locations served, whether it accepts returns, its priority in the rooting algorithm, whether it follows a Weekly Entry procedure for customs clearance, in which days of the week it operates, what is the carrier pick up schedule during the week and what are the operational and financial performance indicators such as: average recovery rate per return, costs per area, inbound rate, processing rate and outbound rate, no-stock and wrong-item rates, and so on.

Among the aforementioned attributes, *Warehouse* also has a *Stock* matrix with references to product objects which are currently in inventory, and vectors with products awaiting pick up or customs clearance.

Concerning its methods, *Warehouse* has functions triggered regularly which aim at accounting for periodic expenses. Furthermore, it includes methods linked to operational events, such as (not exhaustively): *receiveBulkShipment*, *processOrder* — which is called when an order is allocated to the warehouse —, *checkStock*, *pick* — which returns a specific product from inventory —, *WeeklyEntry* — which, when applicable, triggers a customs clearance event for products dispatched during the previous week —, *collection* — which is an event triggered when the courier picks up products awaiting dispatch —, *receiveReturn* — which is triggered when a return reaches the warehouse — and *dispatch*.

Finally, the class *Warehouse* also includes auxiliary methods, among which are functions aimed at handling or creating events; for example: *scheduleWeeklyEntryEvent*, as the name suggests, is called to add an event to the time-line so that, at the date passed as argument, the simulation engine triggers an event which calls the warehouse's *WeeklyEntry* method.

Product This entity contains attributes which describe the product, such as its ID, category, season, box and store (*Store* object). Besides that, it possesses an attribute called, *location* which is used for tax assessment, and *returnTo* which tells whether the product should be returned (if applicable) to its original boutique, to a warehouse accepting returns or to any one of them (the optimum cost-wise). In the beginning, this attribute is set to a user-defined default value for every product — which portrays the general allocation rule —, but can be changed during run-time for products following certain criteria — such as the ones that go through a specific warehouse, for example.

Store This class has attributes which identify name and location of the store and some financial indicators, such as the average cost per area, average labour and performance incentives, order handling costs and rate of recovery from returned products. These attributes are set to a user-defined default, but can be changed for specific stores. Finally, this class also has a *processOrder* method which is only used for reporting purposes — orders allocated to the original historical stores are expected to follow baseline behaviour and, therefore, do not need to be simulated further on.

BulkShipment This class contains attributes which identify the bulk shipment ID, its route, the warehouse at the destination, the number of pallets, units, dispatch and delivery dates (delivery date is predicted). Products being shipped (*Product* objects) are stored in a list. Concerning methods, this class has auxiliary functions targeted at creating or handling events — such as dispatching and generating a delivery event.

Order This entity contains attributes identifying the portal order of which the order is part of, the route, date, product and store. Besides that, it also contains all the historical information linked to the order, such as shipping costs, lead times, duties, commissions, whether it was a wrong-item or no-stock exception, whether it was returned, and so on. These same attributes are replicated to be replaced by the performance in the scenario being simulated. Ultimately, each *Order* object has attributes which describe baseline performance, and attributes which describe the predicted performance within the scenario frame. Results are assessed based on the comparison between both sets of homologous indicators.

PortalOrder This class has an attribute which identify the portal order and a vector containing all boutique orders associated. The main use for instances of this class is to easily assess performance on a portal order basis — which better relates to customer perception.

Customer The *Customer* does not represent specific customers but, rather, the concept. Therefore, it does not possess any identification attributes. Concerning methods, it has a *receiveOrder* function — which is triggered after the order is dispatched from the warehouse and the predicted transit time elapsed —, *assessReturn* — which is responsible for predicting whether the customer is going to return the product (based on whether it is a wrong item, for example) — and *returnOrder* — which, as the name implies, triggers the customer and partner refunds, allocates the return to either a warehouse or the original store (based on the previously mentioned *returnTo* attribute of the *Product* object) and predicts the reverse logistics lead time.

Stakeholder This class is conceived to represent Farfetch or partners. It contains an attribute which identifies the stakeholder, an *account* dataframe — which contains an entry for every financial movement relating to the stakeholder that includes a designation, the amount, date, possible comments and the account balance at the time — and a *baseline* dataframe — which it analogous to the *account* one, but contains historical movements and is intended for comparison purposes.

Concerning methods, there is a *movement* function — which is aimed an recording movements either in the account or baseline dataframes — and, finally, *balance* and *netBalance* which, as the name suggests, retrieve account balances.

Still within the entities module, classes which are more detached from the real-world but which support the simulation's functioning have been created. These are described next:

Time This class keeps track of the simulation's time. Among its attributes are the simulation's *date* — that is updated as events occur — and *timeline* — which is a dataframe that stores upcoming events and, therefore, works as an event execution pipeline.

The class has a *schedule* method which inserts events into the timeline, *nextEvent* which returns the next event and advances the simulation's time and, finally, *addTime* which is an auxiliary function that adds a given number of days, hours, minutes and / or seconds to a date provided and returns the result adequately formatted.

Event As the name implies, the *Event* class stores information about simulation events — such as a designation, the event date and a reference to the function that should be triggered upon occurrence. This class is instantiated by other entities and stored in the *timeline* of a *Time* object until it is called.

SimulationVariable instances of this class are mainly intended to be used for visualisation purposes. The class contains attributes which identify the variable, the corresponding graphical objects pertaining to a chart optionally displayed during run-time and a reference holder for a function which returns the variable's value. Besides that, the class also contains methods which update the variable's value, return the latter, generate update events and update the chart.

SimulationVariable objects are created before the simulation starts, together with events that trigger the function that updates the variable's value.

A.2 Scenario

This module contains a single class which has the same name and whose instances contain all the necessary information required to be fed into the simulation; therefore, it is possible to save *Scenario* instances locally so that they can be retrieved later on. This feature can be useful for collaborative work or for easily accommodating changes to the project or its required inputs along-way.

Among their attributes, *Scenario* objects have a designation field, start and end dates, a list of *Warehouse* objects and a list with *BulkShipment* instances. These are the only necessary inputs obtained from the user. As depicted in A.4, this data will serve as basis for the obtainment of other fields.

Scenario objects contain methods aimed at generating the remaining data from the user inputs and by querying Farfetch's servers. First off, *BulkShipment* objects are instantiated from the user's inputs, as well as a *Warehouse* list. Then, products are obtained from Farfetch's databases

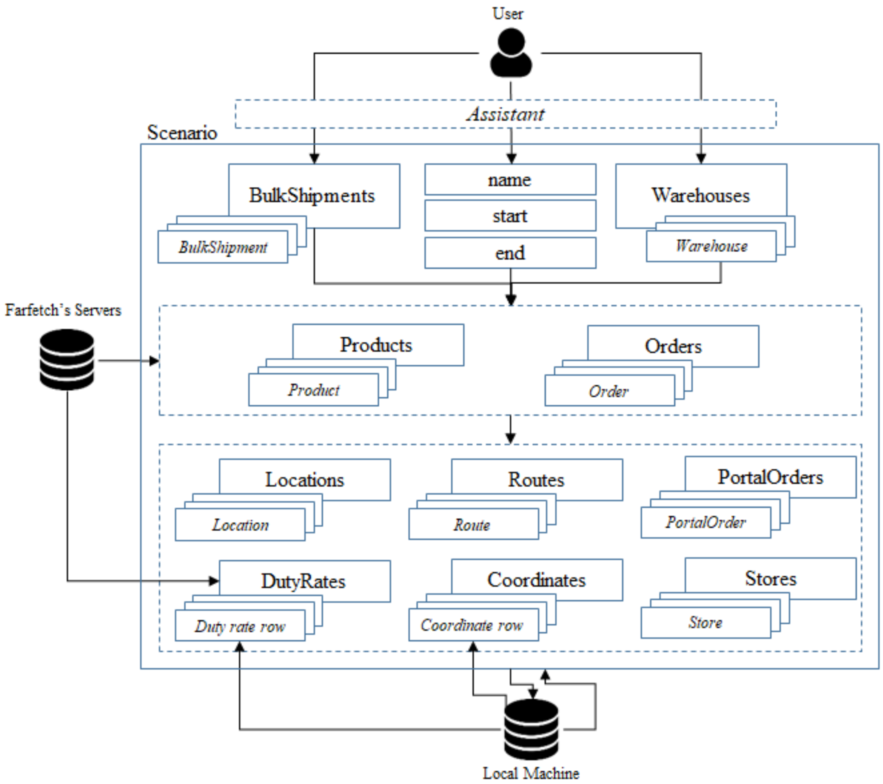

Figure A.4: *Scenario* data dependency diagram

according to the product selection implicitly established by the planned bulk shipments and orders are obtained by extracting records that occurred within the warehouses locations' and service boundaries. These orders, as they are, serve the purpose of later on enabling estimations of shipping costs and transit times, for example, whereas a subset of the former data set will be subsequently obtained by filtering exclusively by the product selection specified by the user so as to portray historical demand.

On a second stage, *PortalOrder* classes are instantiated by combining orders; *Route* and *Location* objects are created by extracting unique combinations of geographical data from the *orders* subset; *Store* instances are obtained by querying Farfetch's databases taking into consideration the product selection and, finally, other auxiliary datasets are obtained using both local files or online queries.

A.3 Engine

The simulation's engine has two main purposes (or distinct phases): preparation and execution. The former is run first and is responsible for leveraging the information that is contained within the *Scenario* instance and generating events, whereas the latter is responsible for running and updating the events in pipeline as they occur.

During the preparation phase, demand events are created and placed into the timeline. Likewise, bulk shipments, accounting events and other auxiliary ones are also scheduled. Figure A.5 depicts a visual representation of a simulation's *timeline*.

During the execution phase, events are run sequentially until no more events remain. During run-time, more events can be created and inserted into the timeline, allowing to emulate causal

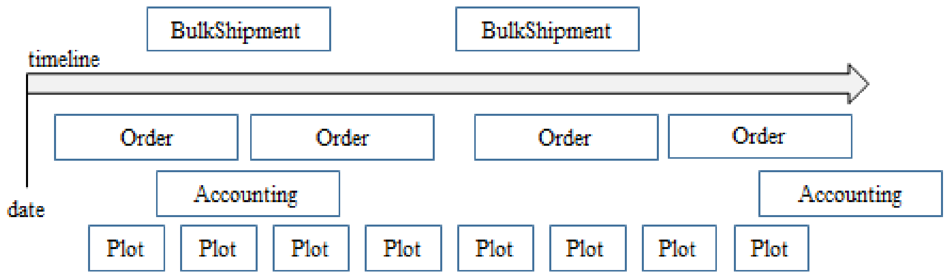

Figure A.5: Illustrative example of a timeline

relationships, such as when an order is dispatched, a delivery event is created after predicting the delivery time. Algorithm 1 illustrates this execution procedure.

```
while timeline still has events do
    event = timeline.pop() ;
    date = event.date ;
    event.triggerEvent() ;
    timeline = sortByDate(timeline) ;
end
```

Algorithm 1: Engine's execution procedure

In order to avoid pitfalls while modelling causality relationships, the thorough process mapping that had been undertaken prior to the simulation's development had to be leveraged so as to ascertain the aforementioned relationships. As an illustrative example, for each order that is being processed in a warehouse, there should be a method that first retrieves the product from the warehouse's stock and only then can the actual processing step take place with that given resource blocked. This may appear to be obvious from a practical viewpoint, but if not modelled as such, two identical orders could be processed simultaneously only to find out, later on, that there is only stock available for one of them — which is something that, assumptively, would not happen in practice within the warehouse context. In summary, the more loyal to reality the modelling of processes is, the more unlikely it is for such mistakes to occur.