

***Business Intelligence* para caracterização de clientes com base no consumo de televisão**

Mariana da Costa Fong Vieira Gomes

Dissertação de Mestrado

Orientador na FEUP: Prof. Bernardo Almada Lobo

Orientador na NOS: João Filipe Gomes



Mestrado Integrado em Engenharia e Gestão Industrial

2021-06-28

Resumo

Este projeto foi desenvolvido na NOS, na equipa de *Reporting* e Análise de Negócio do departamento de *Customer Relationship Management*, com o objetivo de monitorizar o consumo de conteúdos de televisão de forma a facilitar a caracterização e segmentação dos clientes deste serviço, através da análise do universo dos dados de audimetria. Para tal, foram desenvolvidas duas ferramentas.

Em primeiro lugar, foi criado um *dashboard* em Microsoft Power BI que integra e apresenta, de forma clara, toda a informação de audimetria, permitindo a monitorização diária do consumo de conteúdos televisivos e do serviço prestado pela NOS. O *dashboard* permite, ainda, a extração de conhecimento, de forma rápida e independente, sobre o consumo de televisão, informação que é frequentemente necessária para a construção de análises de negócio de várias equipas.

Em segundo lugar, fez-se a segmentação dos clientes do serviço de televisão da NOS em vários grupos, com base no seu consumo de conteúdos, através de um algoritmo de *clustering*. Nesta fase, foram selecionadas as variáveis que mais influenciam o comportamento de visualização dos clientes. No final, caracterizaram-se cada um dos grupos criados, para os quais foram tiradas conclusões.

Este projeto constitui uma base sólida de onde poderão surgir diversas ações de marketing relacional e direcionado a grupos específicos de clientes, com vantagens para a empresa tanto ao nível financeiro, como ao nível da satisfação dos seus clientes.

Abstract

This project was developed at NOS, in the Reporting and Business Analysis team of the Customer Relationship Management department, with the goal of monitoring the consumption of television contents in order to characterize and support the characterization and segmentation of this service's customers, through the analysis of television audience data. To this end, two tools were developed.

First, a *dashboard* was created in Microsoft Power BI that integrates and presents, in a clear and clean way, all the television audience data, allowing the daily monitoring of customers' television consumption. The *dashboard* also allows for the fast and independent knowledge extraction on television consumption. This information is often necessary for business analysis of various teams.

Secondly, NOS' television customers were segmented into different groups, based on their content consumption, through a clustering algorithm. In this step, the variables that mostly influence consumption were found and selected. At the end, each group was characterized and conclusions were drawn.

This project represents a solid base from which various relational marketing actions may arise, aimed at specific customer groups, with advantages for the company both financially and in terms of customer satisfaction.

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao Prof. Bernardo Almada Lobo, pelo conhecimento transmitido e pela orientação prestada ao longo do desenvolvimento desta dissertação.

Em segundo, ao João Filipe Gomes, o meu orientador na NOS, assim como a toda a equipa CRM-RAN, principalmente à Patrícia, à Raquel e à Catarina, por me acolherem neste projeto. Por tudo o que aprendi convosco e por tornarem o trabalho uma parte boa dos meus dias.

Em terceiro lugar, às minhas amigas, pelos desabafos, pelos conselhos e pela partilha de pequenas e grandes vitórias. Obrigada, Cat, JLo, Leo e Sara. Por me lembrarem, sempre, de que não estamos sozinhas.

Por último, à minha família, o meu maior apoio. Por me incentivarem, mesmo sem saberem, a querer deixar-vos orgulhosos. Porque partilharam as dores comigo, quero, agora, partilhar as conquistas convosco. Agradeço, com todo o meu coração, à minha mãe, Paula, ao meu pai, Paulo, às minhas irmãs, Sofi e Clara, aos meus avós, Chico e Nela. Pensando bem, nem todo o meu coração chega.

Índice

1. Introdução	1
1.1. A importância do Marketing Relacional	2
1.2. A NOS no mercado das telecomunicações	3
1.3. Alcance e Objetivo	3
1.4. Metodologia	4
1.5. Estrutura da dissertação.....	5
2. Revisão da Literatura	7
2.1. Business Intelligence.....	7
2.2. Business Intelligence nas telecomunicações	8
2.3. Visualização de dados	10
2.4. <i>Dashboards</i>	11
2.5. <i>Data mining e clustering</i>	12
2.6. Considerações Finais.....	15
3. Análise da situação inicial	16
3.1. Produto TV.....	16
3.2. Estrutura dos dados de cliente e de audimetria	18
3.3. Oportunidades	19
3.4. Objetivos e requisitos	20
4. Metodologia e Execução.....	22
4.1. Etapa 1 – Compreensão do negócio e enquadramento do projeto	22
4.2. Etapa 2 – Levantamento dos requisitos.....	23
4.3. Etapa 3 – Elaboração do design conceptual do <i>dashboard</i>	25

4.4.	Etapa 4 – Identificação e recolha dos dados	26
4.5.	Etapa 5 – Construção do <i>dashboard</i>	27
4.6.	Etapa 6 – Validação e avaliação do <i>dashboard</i>	28
4.7.	Etapa 7 – Caracterização dos clientes	30
4.8.	Etapa 8 – Conclusão do <i>dashboard</i>	34
5.	Resultados	36
5.1.	<i>Dashboard</i> “Universo de Audimetria”	36
5.1.1.	Estrutura e organização.....	36
5.1.2.	Elementos visuais.....	40
5.1.3.	Filtros	45
5.1.4.	Avaliação do utilizador	46
5.2.	Caracterização de clientes através de um algoritmo de <i>clustering</i>	47
5.2.1.	Famílias Tradicionais.....	48
5.2.2.	Famílias Digitais	48
5.2.3.	Espectadores em Direto	49
5.2.4.	Cinéfilos.....	49
5.2.5.	Cientes Cor-de-rosa	49
5.2.6.	Famílias com Crianças.....	49
5.2.7.	Fãs de Informação.....	50
5.2.8.	Adeptos de Desporto.....	50
6.	Conclusões e Trabalho Futuro	52
	Bibliografia	55
	Anexo A: Páginas da versão final do <i>dashboard</i>	58

Siglas e Acrónimos

B2B	Business-to-Business
B2C	Business-to-Consumer
BI	Business Intelligence
CRM	Customer Relationship Management
DTH	Direct To Home
DW	Data Warehouse
ETL	Extract, Transform, Load
FTTH	Fiber To The Home
HDR	High Dynamic Range
HFC	Hybrid Fiber Coax
IT	Information Technology
KPI	Key Performance Indicator
RAN	Reporting e Análise de Negócio
SA	Service Account
TAM	Television Audience Measurement
TV	Televisão
TVS	Televisão por Subscrição

Índice de Figuras

Figura 1: O papel de BI em cada etapa de CRM	9
Figura 2: Procedimento de <i>clustering</i>	13
Figura 3: Lógica das abordagens aglomerante e divisiva dos algoritmos de <i>hierarquical clustering</i>	15
Figura 6: Página de comparação da versão do <i>dashboard</i>	30
Figura 7: Curva de aplicação do <i>Elbow Method</i> para a definição do número ótimo de <i>clusters</i>	33
Figura 8: Cabeçalhos utilizados nas várias páginas do <i>dashboard</i>	37
Figura 9: Layout das páginas das secções 1 e 2 do <i>dashboard</i>	38
Figura 10: Layout das páginas das secções 3 e 4 do <i>dashboard</i>	38
Figura 11: Layout das páginas de comparação do <i>dashboard</i>	39
Figura 12: Gráfico de colunas empilhadas simples – Evolução do número de utilizadores por funcionalidade.....	41
Figura 13: Gráfico de colunas empilhadas 100% - Distribuição do tempo de consumo por categoria	41
Figura 14: Gráfico de linhas - Evolução da duração média das sessões por funcionalidade.....	42
Figura 15: Gráfico de área - Evolução diária do consumo total dos programas selecionados	42
Figura 16: Gráfico de barras empilhadas 100% - Tempo de consumo em cada horário por categoria	43
Figura 17: Gráfico de pontos - Duração média das sessões em cada funcionalidade por categoria.....	43
Figura 18: Gráfico circular - Distribuição do consumo dos programas selecionados por horário	44
Figura 19: Tabela - Top de canais com maior consumo.....	44
Figura 20: Critérios de distribuição dos clientes pelos <i>clusters</i>	51
Figura 21: Glossário da versão final do <i>dashboard</i>	58

Figura 22: Página 1.1.a da versão final do <i>dashboard</i>	59
Figura 23: Página 1.1.b da versão final do <i>dashboard</i>	59
Figura 24: Página 1.2.a da versão final do <i>dashboard</i>	60
Figura 25: Página 1.2.b da versão final do <i>dashboard</i>	60
Figura 26: Página 2.1.a da versão final do <i>dashboard</i>	61
Figura 27: Página 2.1.b da versão final do <i>dashboard</i>	61
Figura 28: Página 2.2.a da versão final do <i>dashboard</i>	62
Figura 29: Página 2.2.b da versão final do <i>dashboard</i>	62
Figura 30: Página 3.1 da versão final do <i>dashboard</i>	63
Figura 31: Página 4.1 da versão final do <i>dashboard</i>	63

Índice de Tabelas

Tabela 1: Procedimento geral do algoritmo de <i>clustering</i> – <i>K-means</i>	14
Tabela 2: Procedimento geral dos algoritmos de <i>hierarquical clustering</i>	14
Tabela 3: Tabelas da estrutura de dados utilizadas para a construção do <i>dashboard</i>	27
Tabela 4: Diferentes versões do <i>dashboard</i>	28
Tabela 5: Variáveis utilizadas na primeira e na última iteração do processo de <i>clustering</i>	31
Tabela 6: Grupos de clientes resultantes do algoritmo de <i>clustering</i>	34

Capítulo 1

Introdução

A crescente facilidade na recolha e no armazenamento de dados exige processos de tratamento, processamento e análise da informação contida nos mesmos. Desta forma, ferramentas e técnicas de *Business Intelligence* são cada vez mais uma componente fundamental de qualquer empresa ou negócio.

A indústria das telecomunicações é um dos maiores criadores e consumidores de grandes quantidades de dados. Com as tecnologias em constante desenvolvimento, tanto ao nível de *software* como de *hardware*, é cada vez mais fácil fazer o registo de dados neste setor. Assim, a vantagem competitiva neste mercado passa pela extração e interpretação eficaz da informação contida nos mesmos.

Em Portugal, este panorama das telecomunicações não é exceção, com grandes *players* bem estabelecidos no mercado a tentar destacar-se diariamente no seio da elevada competitividade desta indústria. Para tal, é fundamental tirar o máximo partido do conhecimento contido na imensidão dos dados gerados, sendo este um dos ativos mais valiosos neste negócio. Sendo a distribuição de conteúdos televisivos um dos principais produtos das empresas de telecomunicações, a compreensão do comportamento dos clientes relativamente ao seu consumo de conteúdos de televisão é decisiva para o sucesso de qualquer empresa, constituindo um apoio fundamental para várias decisões cruciais para o negócio, como a negociação de contratos com os fornecedores de conteúdos, criação de pacotes de produtos e distribuição da oferta aos consumidores.

Assim, a existência de uma estratégia de Business Intelligence clara e incisiva é fundamental para a sobrevivência nesta indústria. É essencial definir os processos de recolha e análise de informação e perceber como é que o conhecimento extraído pode

ser usado para estabelecer preços e ofertas de produtos e serviços, não só a novos, mas também à base atual de clientes, passando também pela avaliação dos melhores meios para entregar esses produtos e serviços. Adicionalmente, o rápido acesso a este conhecimento é um grande potenciador da eficiência no dia-a-dia de uma empresa de telecomunicações, já que a flexibilidade e a rapidez são fatores decisivos no sucesso dentro de qualquer indústria.

É neste contexto que surge o presente projeto de dissertação, uma colaboração entre a Faculdade de Engenharia da Universidade do Porto e a “NOS SGPS” (daqui em diante referida como NOS), uma empresa portuguesa de telecomunicações. O objetivo principal do projeto é responder à necessidade emergente de compreender o comportamento dos clientes da NOS, no que diz respeito aos seus hábitos de consumo de conteúdos de televisão. Com processos já bem implementados ao nível do registo de dados, a oportunidade no seio desta empresa encontra-se no tratamento dos mesmos, na extração de informação e na apresentação do conhecimento de forma clara e dinâmica, não só possibilitando uma resposta rápida a questões que surgem no dia-a-dia de várias equipas da empresa, mas também constituindo um apoio a diversos projetos correntes.

1.1. A importância do Marketing Relacional

As empresas estão, mais do que nunca, centradas no cliente. A NOS não é exceção, e a comunicação e a relação com os seus clientes é vital, já que estes têm cada vez mais poder de decisão e de influência, podendo determinar a forma como a marca é percebida. Para além da importância de conquistar novos clientes, indo de encontro aos seus interesses e resolvendo os seus problemas, a manutenção e o fortalecimento dessa relação é fundamental. É neste contexto que se destaca o papel chave do marketing relacional, devido às limitações do marketing de massas.

Segundo Philip Kotler, “o Marketing Relacional visa construir relações de longo prazo mutuamente satisfatórias com parceiros-chave a fim de conquistar e reter negócios”. Cada vez mais é necessário compreender as necessidades, objetivos e desejos dos clientes, tarefa para a qual se implementam estratégias de *Customer Relationship Management* (CRM). Conhecer os clientes, segmentar o público-alvo, criar uma estratégia de CRM sólida e monitorizar os resultados são tarefas de extrema importância para o sucesso das empresas. Num contexto empresarial cada vez mais tecnológico, as ferramentas de BI desempenham um papel fundamental, tornando todas estas tarefas mais eficientes e eficazes.

1.2. A NOS no mercado das telecomunicações

A NOS é o maior grupo de comunicações e entretenimento em Portugal. Surgiu em 2014, através da fusão da ZON e da Optimus, duas das maiores empresas de comunicações do país. Esta fusão possibilitou a integração, numa só estrutura, de todas as comunicações, fixas e móveis, com oferta direcionada a vários segmentos e mercados: Comunicações Particulares e Empresarial, Cinema e Audiovisuais.

Em Portugal, o mercado das telecomunicações é liderado por quatro empresas, sendo elas a NOS, a MEO, a Vodafone e a NOWO. Para além destas quatro, a competição é praticamente inexistente, resultado de fortes regulações e barreiras à entrada. No final de 2020, este mercado atingiu os 4,2 milhões de subscritores de serviços em pacote, representando receitas no valor de 1,7 mil milhões de euros. Neste contexto, a NOS apresenta-se como o segundo maior prestador destes serviços (36,5% de quota de mercado). No entanto, ao nível das receitas a empresa foi, no mesmo ano, o prestador com maior quota (41,0%) (ANACOM, 2021).

No que toca ao serviço de distribuição de sinais de televisão por subscrição (TVS), que no final de 2020 chegou a 93,2% das famílias portuguesas, a NOS é, mais uma vez, o segundo maior prestador, com 39,0% da quota de assinantes do serviço, o que corresponde a 1,6 milhões de clientes TVS. No entanto, considerando apenas o segmento Residencial (NOS Particulares), a NOS apresenta a quota mais elevada (39,8%) (ANACOM, 2021).

Para além do serviço de televisão por subscrição, a NOS é líder no mercado da banda larga de última geração e na distribuição e exibição de cinema. A sua estratégia atual passa por aumentar a quota de mercado em todos os segmentos, consolidar a posição de líder de telecomunicações e entretenimento do país e expandir internacionalmente.

1.3. Alcance e Objetivo

Devido à sua grande dimensão, com quase 2 mil colaboradores, a NOS apresenta uma estrutura hierárquica, com vários pelouros, departamentos e equipas. Apesar de ter uma vasta oferta para clientes particulares (*Business-to-Consumer* ou B2C) e clientes empresariais (*Business-to-Business* ou B2B), este projeto foca-se apenas nos clientes B2C, tendo sido desenvolvido dentro do pelouro de Marketing Particulares, no departamento de CRM.

Na NOS, o departamento de CRM tem diversas funções, distribuídas por várias equipas. O projeto desta dissertação foi elaborado no seio da equipa de Reporting e

Análise de Negócio (CRM-RAN), responsável por duas tarefas fundamentais: a elaboração de relatórios periódicos sobre diversos temas do interesse e responsabilidade do pelouro de Marketing Particulares e a criação de análises de negócio exaustivas que servem de apoio à tomada de decisão, como por exemplo a descontinuação de produtos ou serviços existentes ou a segmentação de uma campanha promocional. Este projeto foi desenvolvido dentro da equipa CRM-RAN, já que o seu objetivo principal assenta na construção de um relatório com informação atualizada periodicamente, que organize o grande volume de dados de audimetria recolhidos pela empresa, permitindo uma visualização clara e uma rápida extração de conhecimento acerca do comportamento de visualização de conteúdos televisivos dos clientes. Este relatório, com informação de um carácter sensível devido à sua extrema importância e influência em ações decisivas para a estratégia da NOS, será direcionado ao departamento de Gestão de Conteúdos, sendo também utilizado pelo departamento de Gestão de Produto.

O departamento de Gestão de Conteúdos é responsável pela angariação e gestão de conteúdos televisivos, lineares e não-lineares, base e *premium*, com o objetivo de disponibilizar aos clientes da NOS o melhor entretenimento. Algumas das tarefas realizadas pelos membros das equipas que o constituem são a aquisição e negociação de conteúdos, a organização de campanhas comerciais para canais e serviços premium e ainda a comunicação dos principais destaques de programação aos clientes. Por sua vez, o departamento de Gestão de Produto tem como missão desenvolver o Produto TV, garantindo a avaliação de novos conceitos e funcionalidades, tendo em conta o conhecimento do cliente e do mercado. Para isso, é necessário um acompanhamento constante das tendências do mercado do consumo de conteúdos televisivos e a construção de um plano de ação para o produto, alinhado com o plano de ação de marketing global para o segmento. São também funções deste departamento a coordenação das equipas de engenharia, conteúdos e marketing para manter o Produto TV alinhado com a estratégia global da empresa, e ainda a gestão de parcerias de produto.

1.4. Metodologia

De forma a alcançar com sucesso os seus objetivos, a elaboração desta dissertação seguiu um planeamento constituído por oito etapas.

As primeiras semanas foram dedicadas à apresentação da equipa CRM-RAN e à compreensão do negócio da empresa. Este objetivo foi conseguido através de conversas individuais com cada um dos membros da equipa, da elaboração de um relatório simples, onde foi possível compreender as ferramentas e processos adotados na elaboração de relatórios e análises, e ainda de reuniões com membros de outras

equipas e departamentos. Depois de se conhecer a base onde este projeto iria assentar, procedeu-se ao levantamento de requisitos e à definição dos *Key Performance Indicators* (KPI). O objetivo desta fase foi identificar as métricas desejadas no relatório, a periodicidade da informação e os parâmetros para filtrar a informação de forma útil.

A terceira fase consistiu na conceptualização do relatório, onde foi definida a melhor forma de apresentar a informação, suportada por conceitos teóricos e práticos. Estabeleceu-se a estrutura do relatório, na forma de um *dashboard*, ficando definidas as páginas principais e o seu conteúdo, a disposição e forma de apresentação das métricas e dos parâmetros de filtro. O passo seguinte consistiu no levantamento das variáveis necessárias e na recolha e tratamento dos dados, de forma a possibilitar a sua apresentação e atualização periódica. Em quinto lugar, foi feita a construção do *dashboard* na ferramenta Microsoft Power BI, com os dados tratados previamente e respeitando os requisitos definidos na segunda fase.

Após a construção, foi essencial fazer a validação do *dashboard* por parte do utilizador, de forma a garantir que a informação apresentada responde de forma clara às questões existentes. O grande objetivo desta fase foi a recolha de *feedback* do utilizador, tornando possível e eficaz a realização de ajustes ao relatório. É importante realçar que as etapas dois a seis foram executadas de forma iterativa, possibilitando uma evolução progressiva do *dashboard*, até à sua versão final. A sétima etapa consistiu na caracterização da base de clientes da NOS, através de um método de *clustering*, com base no seu comportamento de visualização de conteúdos de televisão. Finalmente, procedeu-se à conclusão do projeto, onde foi organizado todo o trabalho finalizado de forma a poder ser mantido e melhorado ao longo do tempo, tarefa que ficará a cargo da equipa CRM-RAN.

Esta metodologia será apresentada em detalhe no Capítulo 4 desta dissertação.

1.5. Estrutura da dissertação

A presente dissertação está organizada em seis capítulos, apresentados de seguida.

O Capítulo 1 introduz o projeto desenvolvido, onde é feita uma contextualização do problema, apresentadas as motivações e definidos os objetivos principais.

No Capítulo 2 é apresentada uma revisão da literatura acerca tópicos relevantes à elaboração deste trabalho, nomeadamente as aplicações de estratégias e ferramentas de Business Intelligence na indústria das telecomunicações. De seguida, é apresentado o estado da arte no que diz respeito à construção e ao desenvolvimento de *dashboards*.

Por último, é abordada a teoria referente às técnicas mais utilizadas para a criação de *clusters* e para a caracterização e análise dos mesmos.

O Capítulo 3 descreve detalhadamente o estado inicial do assunto tratado nesta dissertação, isto é, a forma como são recolhidos e tratados os dados de audimetria, e como é feita a sua apresentação, revelando as suas limitações atuais. Aqui, serão clarificados as oportunidades e os objetivos deste projeto.

O Capítulo 4 apresenta, de forma exaustiva, a metodologia seguida ao longo da elaboração do projeto da dissertação, referindo os processos e as técnicas utilizadas em cada uma das etapas.

No Capítulo 5 são apresentados os resultados, o que engloba a versão final do *dashboard* e a caracterização dos clientes de acordo com o seu consumo de conteúdos de televisão.

O Capítulo 6 finaliza a dissertação, expondo as principais conclusões do projeto e sugerindo algumas direções para trabalho futuro.

Capítulo 2

Revisão da Literatura

2.1. Business Intelligence

Em qualquer negócio, de forma a potenciar a sua eficiência e eficácia, processos de BI podem ser usados para extrair informação e conhecimento de enormes volumes de dados, através de ferramentas de *data mining* e outras técnicas analíticas (Aruldoss et al., 2014). De uma forma mais clara, e como o mesmo autor explicita, Business Intelligence é o conjunto integrado de ferramentas utilizadas para apoiar a transformação dos dados em informação que apoie os processos de tomada de decisão. Sendo uma ferramenta versátil, com aplicabilidade nos mais diversos domínios, Business Intelligence é largamente referida como um sistema de apoio que recolhe, armazena e analisa dados provenientes de diversas fontes, sendo o seu resultado fundamental para os processos de tomada de decisão (Aruldoss et al., 2014).

Ao longo do tempo, foram desenvolvidas várias aplicações de BI, solucionando problemas em diversas áreas, nomeadamente Marketing, Vendas e CRM, Deteção de fraude e Network management (Ashraf & Khan, 2015), e apoiando a tomada de melhores decisões no âmbito de vários negócios. A colaboração de vários autores e profissionais da área possibilitou o desenvolvimento das técnicas de BI, não só ao nível da sua aplicação, mas também no âmbito do melhoramento das estratégias de recolha de dados, através do desenvolvimento de técnicas de extração de informação mais eficientes.

Desta forma, no universo de Business Intelligence podem ser consideradas diversas perspetivas. Enquanto uns veem BI como um sistema de apoio à decisão orientado por dados (Power, 2013), outros alargam a definição de BI a um sistema de informação estratégico capaz de fornecer informações relevantes através de um

repositório centralizado de dados, provenientes de diversas fontes, transformados em conhecimento por ferramentas analíticas, levando a decisões bem sustentadas (Rus & Toader, 2008). Os sistemas de informação desta última perspectiva podem ainda ser classificados em *Operational Support Systems*, que suportam as decisões operacionais do dia-a-dia, ou em *Decision Support Systems*, sistemas que apoiam os processos de tomada de decisão (Ishaya & Folarin, 2012).

Uma ferramenta típica de Business Intelligence é constituída por vários componentes, sendo os mais comuns a extração de dados, o armazenamento de dados, técnicas de filtragem de atributos, análise através de processos analíticos e apresentação da informação (Aruldoss et al., 2014). É essencial uma boa estrutura de Tecnologia de Informação (IT), bem como uma integração eficaz com a mesma, de forma a potenciar a eficiência e utilidade da ferramenta de BI.

Pode concluir-se que, através da transformação de dados brutos em conhecimento útil ao negócio ou problema em questão, as ferramentas de BI sustentam grande parte do processo de tomada de decisão, bem como possibilitam uma monitorização próxima da performance nas várias operações de um negócio, facilitando a definição de uma estratégia.

2.2. Business Intelligence nas telecomunicações

A indústria das telecomunicações é um dos grandes negócios onde a troca de informação entre duas entidades ocorre a uma grande escala e com grande velocidade. Até hoje, verificou-se uma evolução surpreendente nesta área, sendo esta indústria uma das pioneiras no investimento em aplicações de BI ágeis e escaláveis, com capacidade de evoluir lado a lado dos recursos tecnológicos.

Começando com tecnologias limitadas, onde apenas eram registadas transações de dados simples e a sua análise tinha como objetivo principal a eficiência operacional, os fornecedores de serviços de telecomunicações deparam-se, nos anos 90, com recursos suficientes para realizar operações intensivas de dados, como a automatização da criação e atualização de relatórios e *dashboards* (Kastouni & Ait Lahcen, 2020).

A partir de 2000, foi possível passar de uma análise centrada em ativos para uma análise focada no comportamento dos clientes, devido a tecnologias de informação mais desenvolvidas e estruturas de dados mais robustas, que permitiram a geração de quantidades massivas de dados. Aqui, tornou-se essencial a implementação de aplicações de Business Intelligence, sendo estas um fator diferenciador na estratégia de qualquer empresa na indústria das telecomunicações.

Atualmente, a aplicação de BI na indústria das telecomunicações foca-se, principalmente, na gestão dos consumidores, gestão de fraude, *knowledge discovery* e análise de indicadores de performance (Ishaya & Folarin, 2012). O típico processo de BI numa empresa de telecomunicações começa com a recolha e o processamento de dados de várias fontes para um armazém de dados (DW), através do método ETL (Extract-Transform-Load). Depois deste processo, que ocorre numa base diária, são definidas as intenções e os objetivos de uma determinada análise, do ponto de vista do negócio. Ferramentas e técnicas analíticas são então utilizadas para transformar os dados em informações úteis e acionáveis, posteriormente apresentadas de uma forma conveniente.

Tipicamente, uma empresa prestadora de serviços de telecomunicações tem uma enorme base de clientes e uma larga oferta de produtos e serviços. A heterogeneidade que daqui deriva dificulta consideravelmente a interação com os clientes e a criação de ofertas adequadas. É então crucial que a estratégia de CRM assente em soluções e aplicações de BI. O conhecimento obtido da imensidão dos dados, através destas aplicações, deve ser usado para o desenvolvimento de novos produtos e serviços, adaptando a empresa à mudança constante das necessidades dos clientes, quer existentes quer novos.

Segundo Negash (2004), o processo de CRM numa empresa de telecomunicações é constituído por três etapas:

1. Identificar os clientes mais lucrativos ou potencialmente mais lucrativos para interação futura.
2. Compreender as suas necessidades e padrões de compra.
3. Interagir com eles de forma a ir de encontro a todas as suas expectativas.

A Figura 1 ilustra o papel de Business Intelligence em cada etapa do processo CRM.

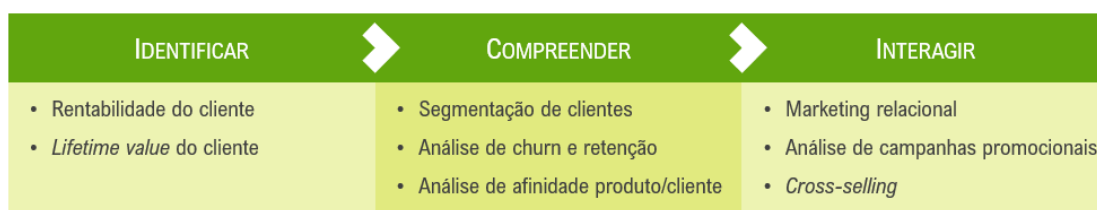


Figura 1: O papel de BI em cada etapa de CRM

Através de ferramentas de BI, é possível segmentar e caracterizar a base de atuais ou potenciais clientes, agrupando os que apresentem características semelhantes, possibilitando o desenvolvimento de ofertas personalizadas e a criação de iniciativas de

marketing direcionadas a grupos específicos. Este processo reduz significativamente os custos incorridos pela empresa, ao mesmo tempo que aumenta a satisfação dos clientes.

2.3. Visualização de dados

Devido do grande volume de dados que estão na base das análises de negócio, é extremamente difícil extrair informação diretamente dos mesmos. Por este motivo, a correta apresentação dos dados, através de ferramentas e tecnologias de visualização, é fundamental para uma interpretação eficiente da informação.

Visualização é a representação gráfica da informação. O principal objetivo da visualização de dados é fornecer ao utilizador uma compreensão qualitativa e quantitativa do conhecimento contido nos dados de forma fácil e rápida (Ashraf & Khan, 2015). Para facilitar este processo, foram criadas várias ferramentas de BI para visualização de dados, tornando o seu utilizador capaz de perceber padrões e relações de dados relevantes sem depender de especialistas de TI (Lousa et al., 2019).

Estas ferramentas, através de elementos como tabelas, gráficos e mapas, oferecem uma maneira eficiente e eficaz de compreender tendências, exceções e outros padrões nos dados. A visualização com base em imagens e gráficos adiciona uma tremenda vantagem relativamente à apresentação de dados através de números e texto. Para além de fornecer uma perspetiva diferente em como os dados são analisados, é uma forma mais cativante e menos cansativa para o utilizador, já que o cérebro humano tem uma maior facilidade em compreender com rapidez e qualidade informação contida em imagens (Kirk, 2021). A capacidade de interação e manipulação é também uma das vantagens das ferramentas de visualização de dados, já que dá ao utilizador a capacidade de obter informação e conhecimento personalizados, de acordo com as suas necessidades, mesmo não tendo uma conhecimento inicial profundo dos dados (Hu & Cleland, 2019).

Atualmente, o mercado das ferramentas de BI para visualização de dados apresenta uma oferta abundante. Assim, a escolha da ferramenta ideal pode ser baseada nos critérios seguintes:

- Facilidade de utilização;
- Capacidade de processar um grande volume de dados;
- Oferta de vários elementos de visualização, como tabelas, gráficos e mapas;
- Custo de utilização e conectividade.

De entre as mais utilizadas, nomeadamente, Microsoft Power BI, Tableau, Sisense e QlikView, é difícil escolher a melhor ferramenta, já que os seus diferentes

pontos fortes podem ser uma mais-valia para diferentes tipos de projetos. Segundo a revisão feita por Lousa et al. (2019), o Tableau é considerado o melhor ao nível da oferta de elementos visuais, apesar do Power BI ter também uma vasta biblioteca de visualizações padrão. O Sisense destaca-se pelo seu uso eficiente de RAM e memória do processador, sendo o mais aconselhado para processar volumes de dados excepcionalmente grandes. Por fim, o QlikView é o mais adequado para análise de dados, contando com análises sofisticadas que permitem a descoberta de informação nos dados através da tecnologia *in-memory*.

2.4. Dashboards

Hoje em dia, é frequente os managers sentirem-se sobrecarregados com o volume de dados disponível, agravado pela fragmentação dos meios de comunicação, gestão multi-canal, proliferação da oferta de produtos e serviços e personalização em massa (Hyde et al., 2004). Os *dashboards* oferecem uma resposta a esta crescente complexidade e diversidade de dados e informação num negócio, com a os quais as equipas têm de lidar diariamente.

Segundo LaPointe (2005), são vários os fatores que justificam a necessidade de utilização de *dashboards*, destacando-se a má organização do grande volume de dados relevantes para o processo de tomada de decisão, os preconceitos de gestão no processamento de informação e na tomada de decisões, e a necessidade de integração entre departamentos durante os processos de apresentação de informação.

Os *dashboards* surgiram na década de 1980 com o objetivo de apresentar alguns indicadores chave através de uma interface simples, de forma a serem compreendidos “até por um executivo” (Few, 2006). Desde aí, as funções de um *dashboard* são recolher, sintetizar e apresentar informação de múltiplas fontes de forma clara. De facto, um *dashboard* é definido por Pauwels et al. (2009) como um conjunto relativamente pequeno de indicadores de performance chave e motivadores de desempenho subjacentes que refletem os interesses e objetivos a curto e a longo prazo comuns a uma organização.

Ainda a mesma fonte refere que a integração é uma característica importante num *dashboard*, devendo existir de três formas:

- Integração de dados: reunir dados de diferentes fontes, com níveis diferentes de agregação e relativos a vários períodos de tempo.
- Integração de processos: relacionar inputs, performance e resultados das diferentes atividades do negócio.

- Integração de pontos de vista: permitir que pessoas de diferentes departamentos ou áreas de atuação tenham acesso ao mesmo conhecimento.

Na criação de um *dashboard* é necessário considerar dois tipos de atributos. Por um lado, os atributos funcionais, como as opções de *drill-down*, a criação automática de alertas, a frequência de atualização e a flexibilidade. Por outro lado, os atributos visuais, que dizem respeito à forma como a informação é apresentada, tendo em vista a máxima eficiência e eficácia (Few, 2006), como a seleção de cores e do tipo de elementos visuais.

Para além de recolher, sintetizar e apresentar informação, um bom *dashboard* deve ser capaz de:

- Monitorizar a performance, de forma a ser claro que atividades ou operações tiveram maior ou menor sucesso;
- Reforçar a consistência na avaliação de processos e apresentação de indicadores e métricas;
- Planear, isto é, perceber a situação atual da empresa e quais deverão ser os objetivos e as estratégias para o futuro;
- Comunicar, não só resultados operacionais ou financeiros, mas também as prioridades e valores da empresa, através das métricas e indicadores selecionados.

2.5. *Data mining e clustering*

Data mining é um processo de *unsupervised machine learning* usado para descobrir padrões e extrair conhecimento de grandes bases de dados ou repositórios de informação (Han & Kamber, 2006). *Clustering* é uma técnica de *data mining* que agrupa os dados de forma a que a similaridade entre objetos do mesmo grupo seja maior do que a similaridade entre objetos de grupos diferentes. É considerada uma das mais importantes técnicas de *unsupervised machine learning* porque permite dividir um conjunto finito de dados sem caracterização conhecida num conjunto finito de estruturas de dados com características semelhantes (R. Xu & Wunsch, 2005).

Por ser uma área de estudos relativamente recente, a definição completa de *clustering* ainda não chegou a um consenso. No entanto, Xu & Tian (2015) apresentam a seguinte definição clássica:

- Pontos do mesmo *cluster* devem ser o mais semelhantes possível;
- Pontos de *clusters* diferentes devem ser o menos semelhantes possível;

- As métricas para o cálculo da semelhança entre pontos devem ser claras e ter um significado prático.

Os mesmos autores ditam que o processo de *clustering* pode ser dividido em quatro fases (Figura 2), intimamente relacionadas entre si:

1. Extração e seleção dos atributos que melhor representam o conjunto de dados original;
2. Desenvolvimento e aplicação do algoritmo de *clustering* de acordo com as características do problema;
3. Avaliação dos resultados do algoritmo e validação da sua aplicação;
4. Explicação dos resultados, atribuindo um significado prático a cada *cluster*.

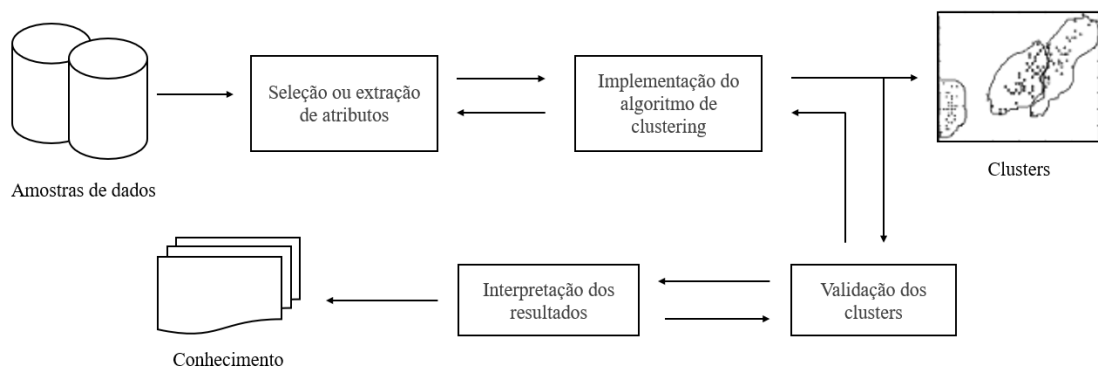


Figura 2: Procedimento de clustering

Distância e semelhança são a base para a construção de algoritmos de *clustering*, sendo que é recomendado o cálculo de distâncias quando se lida com atributos quantitativos. Por outro lado, a semelhança é recomendada para dados com atributos qualitativos (R. Xu & Wunsch, 2005).

O processo de *clustering*, por si só, não é um algoritmo, mas sim uma tarefa geral que pode ser realizada recorrendo a diversos algoritmos que variam nas suas abordagens e eficiência na formação de *clusters*. Uma teoria aproximada, mas amplamente aceita, é a classificação dos algoritmos em *hierarquical clustering* ou *partitional clustering*. O primeiro tipo de algoritmos recorre à divisão ou união sequencial dos objetos, indo de um único *cluster* inicial até *clusters* com um único objeto, ou vice-versa. Por outro lado, os algoritmos de *partitional clustering* dividem o conjunto de dados diretamente num número predefinido de *clusters*, sem construir uma estrutura hierárquica (R. Xu & Wunsch, 2005).

Algoritmos de *partitional clustering*

Um dos algoritmos frequentemente utilizados é o *K-means*. A ideia central deste algoritmo é dividir o conjunto de observações em *k clusters*, em que *k* é previamente definido pelo utilizador, através de um processo iterativo que considera a distância do

centroide de cada *cluster* a cada uma das observações (Jumadi et al., 2019). Os passos do algoritmo *K-means* são apresentados na Tabela 1 (Du et al., 2019).

Tabela 1: Procedimento geral do algoritmo de clustering – K-means

-
1. Selecionar aleatoriamente os centroides de cada *cluster*;
 2. Calcular a distância entre cada objeto e cada centroide;
 3. Atribuir cada objeto ao *cluster* cuja distância ao seu centroide é menor, comparada à distância aos centroides dos restantes *clusters*;
 4. Recalcular o centroide de cada *cluster*;
 5. Repetir os passos 2 a 4 até ser atingido o critério de convergência (se definido) ou até não existir qualquer mudança nos *clusters* de uma iteração para a seguinte.
-

Os algoritmos de *partitionial clustering* apresentam uma complexidade relativamente baixa e, no geral, uma elevada eficiência de computação. No entanto, são dos mais sensíveis a *outliers*, rapidamente convergem para o ótimo local, e o seu resultado depende do número de *clusters*, argumento que tem de ser definido pelo utilizador (D. Xu & Tian, 2015).

Outros algoritmos deste tipo são o *K-medoids*, PAM, CLARA e CLARANS.

Algoritmos de hierarquical clustering

A abordagem deste tipo de algoritmos é construir uma estrutura de relações hierárquicas entre os objetos do conjunto de dados, de acordo com a sua matriz de proximidade. Inicialmente, cada objeto representa um *cluster*, sendo que os dois *clusters* mais próximos são agrupados, até existir apenas um único *cluster*. Esta abordagem pode também ser feita no sentido inverso (D. Xu & Tian, 2015). O dendrograma da Figura 3 ilustra a lógica das abordagens aglomerante e divisiva dos algoritmos de *clustering* hierárquico. O procedimento geral dos algoritmos aglomerantes de *hierarquical clustering* é descrito na (R. Xu & Wunsch, 2005).

Tabela 2: Procedimento geral dos algoritmos de hierarquical clustering

-
1. Selecionar N *clusters* de um único objeto, sendo N o número total de objetos do conjunto de dados.
 2. Calcular a matriz de proximidade entre cada *cluster*.
 3. Juntar os dois *clusters* cuja distância é menor.
 4. Repetir os passos 2 e 3 até todos os objetos pertencerem ao mesmo *cluster*.
-

Estes algoritmos apresentam uma boa performance para conjuntos de dados de qualquer forma, com atributos de qualquer tipo. Além disso, são facilmente escaláveis.

As desvantagens associadas a este tipo de algoritmos são a sua elevada complexidade e a necessidade de predefinir o número de *clusters* desejado (D. Xu & Tian, 2015).

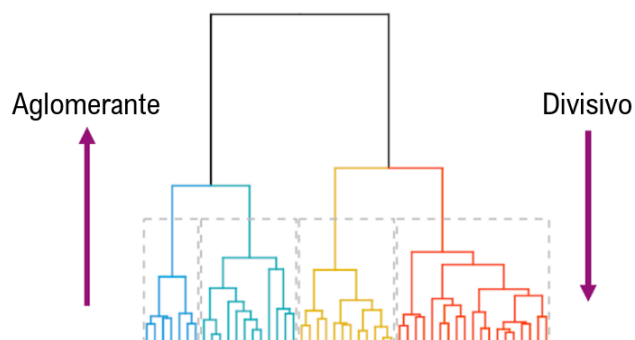


Figura 3: Lógica das abordagens aglomerante e divisiva dos algoritmos de hierarquical clustering

2.6. Considerações Finais

Os temas explorados neste capítulo constituem o ponto de partida para restante dissertação. A compreensão do estado da arte e da importância das aplicações de BI para qualquer negócio, dentro ou fora do mercado das telecomunicações, é o que motiva a elaboração deste projeto, já que fornece as ferramentas necessárias para a resolução dos problemas e para a exploração das oportunidades identificadas no seio da empresa onde o projeto foi desenvolvido.

O primeiro objetivo estabelecido para a presente dissertação irá focar-se na adoção das melhores estratégias de visualização de dados e construção de *dashboards*. Por sua vez, o segundo objetivo será alcançado através da aplicação de técnicas de *data mining*, mais especificamente o *clustering*. Esta estratégia permitirá obter conhecimento dos clientes, extraído do grande volume de dados de audimetria dos mesmos.

Capítulo 3

Análise da situação inicial

3.1. Produto TV

A NOS, uma das maiores empresas de telecomunicações em Portugal, apresenta uma vasta oferta de serviços, sendo um deles o fornecimento de conteúdos de televisão. Para além de uma interface premiada pelos consumidores e uma qualidade de transmissão de imagem que pode chegar ao 4K HDR, a NOS oferece aos seus clientes de televisão canais exclusivos e milhares de conteúdos para ver, a possibilidade de ver televisão em qualquer ecrã, através da NOS TV, disponibilizando uma experiência personalizada para cada perfil.

Os pacotes de conteúdos de televisão disponíveis dependem da tecnologia da rede de telecomunicações fixa na localização do cliente (DTH, HFC ou FTTH). Os diferentes pacotes de conteúdos variam no seu preço, estando alguns dependentes da presença de uma Box TV. Além de conjuntos de canais base que podem ir dos 20 aos 180 canais, a NOS disponibiliza ainda aos seus clientes uma lista de canais Premium que podem ser adquiridos por um preço adicional, bem como diversos *packs* de conteúdos de filmes, séries, desenhos animados, música e entretenimento.

Para além do produto TV, a NOS tem disponíveis mais quatro tipos de produtos: Internet Móvel, Internet Fixa, Voz Móvel e Voz Fixa. O serviço prestado a cada cliente é uma combinação dos produtos disponíveis, caracterizando o Perfil do Cliente, que pode ir de 1P (clientes com um único produto) a 5P (cliente com os cinco produtos). Todo o serviço prestado ao cliente é representado por um Service Account (SA). A ele estão vinculadas todas as características do serviço, bem como os dados de identificação do cliente.

Relativamente ao consumo de televisão, é necessário apresentar e clarificar alguns conceitos, utilizados pela empresa para caracterizar as sessões televisivas e abordados ao longo desta dissertação. São eles:

- Categoria de conteúdo: caracterização do programa transmitido pela atribuição de uma das categorias a seguir apresentadas.
 - Adultos;
 - Desporto;
 - Documentários;
 - Entretenimento;
 - Filmes;
 - Infantil;
 - Informação;
 - Séries.
- Plataforma de transmissão: serviço NOS utilizado para o consumo de conteúdos.
 - Box IRIS;
 - Box UMA;
 - NOS TV;
 - NOS KIDS;
 - NOS PLAY.
- Dispositivo de visualização: tipo de dispositivo utilizado para a transmissão dos conteúdos.
 - Box;
 - Smartphone ou Tablet (Android, iOS);
 - Web;
 - Apple TV;
 - Xbox.
- Horário: período do dia onde ocorre a sessão.
 - Início da manhã – das 07h às 10h;
 - Final da manhã – das 10h às 13h;
 - Tarde – das 13h às 18h;
 - Pré prime-time – das 18h às 20h;
 - Prime-time – das 20h às 24h;
 - Late night – das 00h às 03h;
 - Madrugada – das 03h às 07h.
- Funcionalidade (ou Função TV): função utilizada para a visualização dos conteúdos.
 - Live TV – o cliente vê o programa no momento em que está a decorrer;

- NPVR/PVR – o cliente vê o programa através de gravação previamente definida por ele;
- Timewarp – o cliente vê o programa após término do mesmo, estando ele disponível por sete dias após a emissão;
- Restart – o cliente reinicia o programa no período em que o mesmo está a decorrer;
- VOD – o cliente utiliza Video On Demand na box da NOS.

É importante salientar que, relativamente à funcionalidade, é designado por Consumo Linear todo aquele que é feito através de Live TV, sendo o restante designado por Consumo Não Linear.

3.2. Estrutura dos dados de cliente e de audimetria

Como foi anteriormente observado, os dados são, atualmente, um dos ativos mais valiosos de qualquer empresa ou negócio. Este facto é ainda mais relevante numa empresa de telecomunicações. Por este motivo, uma das prioridades da NOS é a recolha, o registo e o armazenamento de todo o tipo de dados relevantes para o seu negócio. Dentro deste universo, os dados de identificação dos clientes e os dados de audimetria ganham destaque no âmbito desta dissertação.

Os dados de cliente relevantes para o presente projeto são os que caracterizam o produto de cada cliente, bem como as suas informações pessoais. Recolhidos mensalmente e armazenados numa base de dados integrada, estes dados contêm informação sobre todos os produtos no portfólio de cada cliente, televisão, voz e internet, num determinado mês. A estrutura dos dados inclui também a caracterização do cliente ao nível do segmento, empresarial ou particular, informação pessoal do cliente e informação sobre o processo de venda do produto em questão. A informação pessoal inclui o número de contribuinte e a localização do cliente. Já os dados relativos ao processo de venda transmitem informação sobre a existência de alguma campanha promocional ou oferta, a identificação do colaborador NOS que efetuou a venda, bem como o canal onde esta foi efetuada. A recolha dos dados é feita para todos os clientes do parque da NOS num determinado mês, isto é, todos os clientes com um ou mais produtos ativos nesse mesmo mês.

Os dados de audimetria constituem um enorme universo de informação, devido à frequência e ao detalhe com que são recolhidos. É, no entanto, importante referir que a recolha deste tipo de dados não é possível em clientes onde a transmissão de conteúdos televisivos é feita através da tecnologia DTH, existindo dados apenas para clientes Cabo (HFC e FTTH).

Sempre que um cliente da NOS consome conteúdos televisivos, em qualquer dispositivo e plataforma, são registados dados sobre o seu consumo, a cada segundo. De forma a diminuir a complexidade que daqui resulta, existem processos internos implementados para agrupar a informação em sessões, podendo estas ser limitadas, do mais específico para o mais geral, pelo programa assistido, pelo canal transmitido, ou pela funcionalidade utilizada. Neste projeto, utilizou-se a definição de sessão limitada pelo canal, isto é, que começa quando um utilizador entra num canal e termina quando o utilizador sai desse mesmo canal.

Nesta estrutura de dados TAM (*television audience measurement*), o utilizador é identificado pelo código do seu SA. Os dados da sessão contêm informação sobre:

- A data da sessão;
- O minuto inicial, final e a duração da sessão;
- O código do dispositivo utilizado;
- O tipo do dispositivo de visualização;
- A plataforma de transmissão;
- A funcionalidade utilizada;
- O canal de televisão;
- O programa visualizado.

A atualização das bases de dados de audimetria é feita diariamente, tendo a informação um atraso de apenas três dias. Isto é, no dia D estão disponíveis os dados de audimetria até ao dia D-3.

Os dados utilizados ao longo do projeto, bem como a sua manipulação, serão abordados com mais detalhes no Capítulo 4 desta dissertação.

3.3. Oportunidades

Dada a complexidade presente na recolha e no armazenamento dos dados, acima descrita, a sua manipulação e análise em bruto exigem um conhecimento da sua estrutura. Este conhecimento, que proporciona um tratamento dos dados mais eficiente e eficaz, está limitado a um número reduzido de equipas, das quais se destaca a equipa de Estratégia de Dados. Por este motivo, qualquer equipa que precise de analisar a informação contida nestes dados, vê-se obrigada a fazer um pedido de informação à equipa de Estratégia de Dados, levando a uma sobrecarga da equipa com tarefas demoradas e sem valor acrescentado. Aqui surge a primeira oportunidade identificada, a inexistência de um processo pouco dispendioso, tanto a nível de tempo como a nível de recursos humanos, que possibilite o acesso aos dados de audimetria, bem como ao conhecimento contido nos mesmos. Este processo, para além de aliviar as tarefas da

equipa de Estratégia de Dados, tornaria todas as equipas que necessitassem desta informação mais independentes na elaboração das suas análises e relatórios.

Outro ponto a destacar é a enorme quantidade de dados relativos ao consumo de televisão dos clientes da NOS. Será possível identificar comportamentos semelhantes no vasto universo de clientes, no que diz respeito ao seu consumo de televisão? Será possível caracterizar esses mesmos grupos de clientes com o mesmo padrão de comportamento? A segmentação dos clientes com base no seu consumo de televisão que surgiria como resposta a estas perguntas, possibilitaria um vasto conjunto de ações de marketing e gestão de oferta com vantagens para a empresa, não só ao nível de eficiência de operações, como também ao nível da satisfação dos clientes. Surge, então, a segunda oportunidade de desenvolvimento nesta dissertação.

3.4. Objetivos e requisitos

De forma a atacar as oportunidades identificadas, foram definidos dois objetivos para este projeto. Em primeiro lugar, a construção de um *dashboard* de atualização permanente, com toda a informação relativa ao consumo de conteúdos de televisão. Por outro lado, a criação de *clusters* de clientes e a sua caracterização prática, com base no seu consumo de televisão.

Relativamente ao *dashboard*, espera-se que apresente, de forma simples e clara, informação sobre vários aspetos relativos ao consumo de conteúdos de televisão do parque de clientes da NOS. Estes parâmetros foram definidos nas reuniões de estabelecimento de requisitos, com os departamentos de Gestão de Conteúdos e Gestão do Produto, e são:

- Evolução do número de utilizadores e do tempo de consumo;
- Distribuição do número de utilizadores e do tempo de consumo por horário do dia, funcionalidade e categoria de conteúdo;
- Canais mais vistos num determinado período de tempo;
- Programas mais vistos num determinado período de tempo.

De forma a se conseguir obter um conhecimento mais detalhado, através de consultas mais flexíveis, foram também definidos os atributos relevantes para filtrar a informação, sendo eles o período de tempo considerado, o dispositivo e a plataforma utilizados, horário e funcionalidade de visualização, o perfil do cliente, e ainda a existência de subscrição dos canais premium mais procurados (Sport TV, TVCine e BTV).

Relativamente à caracterização dos clientes, o objetivo foi desenvolver um algoritmo que faça a alocação automática dos clientes ao respetivo *cluster*. Para que este parâmetro possa ser atualizado numa base semestral ou anual, o algoritmo deve ser de fácil utilização e manutenção. De forma a obter os grupos que melhor caracterizam o parque de clientes da NOS, foram gerados *clusters* através de um algoritmo de *clustering*, tendo por base o parque de clientes atual da NOS e os atributos que melhor definem o seu comportamento de consumo de televisão e o seu perfil como cliente. Após a criação teórica dos grupos, será necessário atribuir e validar um significado prático a cada grupo, sendo que o número de grupos criados deve possibilitar a sua fácil gestão e utilização.

Olhando para a metodologia proposta no Capítulo 1, apesar de ambos os objetivos se basearem na análise e na compreensão dos dados de audimetria, a construção do *dashboard* está incluída nas etapas 1 a 6. Já a criação de *clusters* de clientes e caracterização dos mesmos foi abordada na etapa 7.

Capítulo 4

Metodologia e Execução

Neste capítulo, será abordada e apresentada em detalhe cada etapa da execução do projeto da presente dissertação.

4.1. Etapa 1 – Compreensão do negócio e enquadramento do projeto

No contexto deste projeto, a compreensão do negócio envolve o estudo das dinâmicas de trabalho e das infraestruturas de informação da empresa, que deverão ser conhecidas de forma a atacar eficaz e eficientemente as oportunidades identificadas previamente, para atingir com sucesso os objetivos propostos. Para tal, foram realizadas duas tarefas:

- Elaboração de um relatório simples sobre a utilização da plataforma NOS KIDS;
- Exploração e compreensão dos produtos e dos serviços NOS.

A primeira tarefa foi realizada com o objetivo de introduzir a autora da tese às dinâmicas de *reporting* da equipa CRM-RAN, ao mesmo tempo que respondia a uma necessidade da equipa de Oferta e Segmentação. Por não ser um dos resultados deste projeto, não será explicado na presente dissertação. Em vez disso, são apresentadas as áreas nas quais esta tarefa contribui para o projeto final.

Em primeiro lugar, introduziram-se aspetos práticos dos conceitos de parque, utilização e consumo. Além disso, foi nesta tarefa inicial que foi realizado o primeiro contacto com as bases de dados de cliente e de audimetria, também necessárias para o desenvolvimento do *dashboard* e para a caracterização de clientes. Este contacto permitiu identificar as infraestruturas, a disponibilidade e atualização dos dados.

O objetivo da segunda tarefa foi compreender o contexto onde se insere o projeto, bem como todas as dinâmicas inerentes ao negócio e aos produtos e serviços

da NOS, necessários ao desenvolvimento da dissertação. Aqui, foram obtidas respostas a várias perguntas sobre a oferta da empresa: Quais os produtos e serviços disponibilizados pela NOS? Quais as diferentes formas disponíveis para a entrega destes mesmos produtos e serviços (pacotes, segmentação base/premium, campanhas, ofertas)? Quais os equipamentos e dispositivos necessários para que o cliente possa usufruir dos vários serviços da NOS? Depois deste enquadramento, foi importante definir os limites de atuação do projeto, identificando a informação útil e não útil ao desenvolvimento do mesmo. Presentes os objetivos do projeto, ficou definido que o foco seria o produto TV e os meios de entrega de conteúdos televisivos, bem como as características relevantes dos clientes desta área do negócio.

Esta tarefa foi conseguida através de várias reuniões individuais com membros da equipa CRM-RAN e ainda com colaboradores da NOS de outros departamentos e equipas. Estas reuniões consistiram numa apresentação do projeto, feita pela autora da tese, seguida de uma discussão dos conceitos e das dinâmicas inerentes ao dia-a-dia de cada colaborador, dentro do alcance do projeto. Fora da equipa CRM-RAN, uma das reuniões com maior destaque foi com um membro do departamento de Oferta e Segmentação, onde foram apresentados os produtos e serviços da oferta da NOS, com uma exploração mais detalhada daqueles que incluem o produto TV, o universo dos conteúdos disponíveis e os diferentes meios e plataformas de acesso a estes mesmos conteúdos.

4.2. Etapa 2 – Levantamento dos requisitos

Uma das etapas fundamentais no desenvolvimento de um *dashboard* é o levantamento de todos os requisitos do utilizador e, se possível, a sua prioridade. É necessário definir as métricas que se quer apresentar, bem como a sua viabilidade de inclusão no relatório. Assim, o levantamento inicial dos requisitos e a definição das métricas e KPIs a incluir no *dashboard* resultou de uma reunião entre dois dos principais *stakeholders* do projeto:

- Departamento de Gestão do Produto – sendo um dos principais utilizadores finais do *dashboard*, tem um papel essencial na definição dos requisitos do mesmo.
- Equipa CRM-RAN – sendo o responsável pela construção e manutenção do *dashboard* e conhecendo a estrutura das fontes de informação, tem um papel fundamental na validação da viabilidade de cada requisito proposto.

É importante referir que os requisitos que surgiram como resultado desta reunião foram posteriormente apresentados ao departamento de Gestão de Conteúdos,

outro importante utilizador final, que aprovou e validou a relevância dos requisitos e métricas definidas.

Durante esta etapa, tornaram-se claras as vantagens do *dashboard* ser construído por uma equipa não só com conhecimento do negócio, mas também com conhecimento e prática na manipulação e utilização das fontes e estruturas de dados da empresa. Uma das tarefas mais importantes da equipa CRM-RAN foi, quando confrontada com uma grande lista de requisitos propostos, a seleção rápida daqueles que são ou não viáveis para o *dashboard*. As principais razões para a inviabilidade de alguns requisitos são:

- A elevada complexidade dos dados necessários, que tem um grande impacto nocivo ao nível de performance do *dashboard*;
- A inexistência da informação exigida e impossibilidade de concretização dentro do tempo proposto para o projeto;
- A inexistência de dados atualizados num determinado assunto.

Um exemplo destas situações é a métrica de *Program Engagement* (percentagem de tempo de permanência num determinado programa, relativamente à sua duração total), que exige cálculos complexos, com um grande impacto na fluência de utilização do *dashboard*. A caracterização dos canais premium consoante o seu tema é também um indicador não viável, dado que esta informação não é atualmente incluída nas estruturas de dados.

Em síntese, a seleção dos requisitos a considerar para o *dashboard* passou por um processo de validação centrado em três características: simplicidade, utilidade e adequação, garantindo coerência e consistência de informação do início ao fim do relatório. Assim, os requisitos definidos para a construção do *dashboard* passam pela apresentação dos KPI indicados de seguida, com a possibilidade de filtrar toda a informação por datas, dispositivo de visualização, plataforma de transmissão, funcionalidade, horário, perfil do cliente e categoria do conteúdo.

1. Número total de utilizadores;
2. Tempo total de consumo;
3. Duração média das sessões;
4. Distribuição do tempo de consumo por categoria do conteúdo;
5. Número de utilizadores, tempo de consumo e duração média das sessões dos canais com maior consumo num determinado período de tempo;
6. Tempo total de consumo por canal;

7. Canal de transmissão, número de utilizadores, tempo de consumo e duração média das sessões dos programas com maior consumo num determinado período de tempo;

Relativamente aos requisitos apresentados, é importante clarificar algumas situações. O número de utilizadores é uma métrica que representa o número distinto de SA com qualquer tipo de consumo de conteúdos de televisão durante o período considerado. Dada a impossibilidade, anteriormente referida, de classificar os canais de conteúdos premium por tema, decidiu-se adicionar uma funcionalidade que permite filtrar os clientes com ou sem subscrição dos canais premium com maior aderência e impacto na NOS, os canais Sport TV, TVCine e BTV.

4.3. Etapa 3 – Elaboração do design conceptual do *dashboard*

Depois de definidas as métricas a apresentar no *dashboard*, foi necessário criar uma estrutura consistente e clara, que possibilite uma navegação intuitiva pelo *dashboard*. Este aspeto é muito importante, já que um dos objetivos do projeto é aumentar a independência no acesso a dados e informação de audimetria.

As duas características consideradas nesta fase são a estrutura e o design do *dashboard*. No que toca à estrutura, depois de analisados os requisitos, concluiu-se que a informação se pode dividir em quatro secções: Clientes, Categorias, Canais e Programas. A secção de Clientes pode ainda dividir-se em Utilização e Consumo. Já a secção de Categorias subdivide-se em Categorias por Funcionalidade e Categorias por Horário. Assim, nesta etapa, prevê-se a construção de um *dashboard* com seis páginas. Esta estrutura bem definida não só facilita a interpretação dos dados do *dashboard*, mas também agiliza o processo de recolha de informação.

Relativamente ao design do *dashboard*, optou-se pela clareza e simplicidade. De forma a tornar clara a estrutura, a cada secção é atribuída uma cor, visível no fundo de cada página e ainda na barra do cabeçalho. É nesta barra que está a informação da secção e, caso exista, subsecção da página apresentada. O outro elemento constante em todas as páginas do *dashboard* é a barra de filtros, dado que os atributos através dos quais será possível filtrar a informação se mantêm para todas as métricas apresentadas nos requisitos. Esta barra está, em todas as páginas, do lado esquerdo da janela, enquanto que os elementos visuais de apresentação dos dados estão dispostos ao centro e à direita.

Depois de construída, a estrutura do *dashboard* foi apresentada ao departamento de Gestão do Produto, que expôs algumas sugestões e pontos de melhoria. No final, depois de realizadas as respetivas alterações e ajustes, a proposta foi validada.

4.4. Etapa 4 – Identificação e recolha dos dados

Sabendo quais as métricas e parâmetros de filtro a incluir no *dashboard*, especificados nos requisitos, foi necessário identificar a fonte dos dados de forma a reunir a informação necessária. A elevada complexidade e dimensão dos dados de audimetria levou à necessidade de eleger uma de duas opções, onde se confrontou a performance do *dashboard* com a agilidade na manipulação dos dados do mesmo, o que alargaria as opções de design e dinamismo do *dashboard*.

A primeira opção, que privilegiava a agilidade na manipulação dos dados, consistia em recolher e preparar os dados da base de dados em ambiente Oracle SQL Developer, sendo depois importados para o Microsoft Power BI. Esta estratégia, a mais frequentemente utilizada na criação de *dashboard* na NOS, permite estabelecer várias conexões a diferentes bases de dados, tornando viável a adição de dados de outras fontes, de forma a enriquecer a informação contida no *dashboard*. Além disso, possibilita a manipulação da estrutura de dados através da Vista de Dados do Microsoft Power BI, permitindo ao utilizador a criação de elementos auxiliares, com benefícios para a estrutura do *dashboard*. No entanto, esta estratégia de importação de dados não se mostrou viável para o presente projeto, devido ao enorme volume de dados necessários, mais de 85GB por mês. Esta opção levaria à sobrecarga do *dashboard*, resultando em graves problemas de performance, tanto na utilização como na atualização.

A segunda opção, e aquela que foi escolhida para o projeto, é a utilização de uma estrutura de dados criada e gerida pela equipa de Engenharia de Dados do CRM – o cubo de audimetria. Um cubo, ou *OLAP (online analytical processing) cube*, é uma estrutura de dados que possibilita análises eficientes, envolvendo as múltiplas dimensões que definem o problema de um negócio. Um cubo organiza os dados de um armazém, apresentando as métricas em tabelas facto e as dimensões em tabelas de dimensão, destacando-se pela fiabilidade dos cálculos efetuados, velocidade de análise e possibilidade de *self-service reporting*.

Dado que o cubo de audimetria agrupa todos os dados necessários à elaboração do *dashboard*, optou-se por alimentar o *dashboard* através de uma ligação direta do Microsoft Power BI a esta estrutura de dados, através de uma ligação em direto do tipo *SQL Server Analysis Services database*. As vantagens deste tipo de ligação são a permanência dos dados numa localização independente ao ficheiro, o que se traduz num aumento significativo da performance do *dashboard*. Além disso, a atualização dos dados é feita automaticamente, paralelamente à atualização da base de dados, o que significa que os dados do *dashboard* estarão sempre atualizados para o dia D-3, em relação ao dia atual. No entanto, este tipo de conexão tem algumas limitações. Em

primeiro lugar, impossibilita a importação de dados de outras fontes, bloqueando a possibilidade de aumentar a extensão da informação do *dashboard*. Além disso, impede a manipulação da estrutura dos dados, como a criação de elementos auxiliares (tabelas e colunas), o que aumentaria o leque de opções ao nível do design e do dinamismo do *dashboard*.

Tabela 3: Tabelas da estrutura de dados utilizadas para a construção do dashboard

Tabelas Facto	Tabelas de Dimensão
• Sessões por funcionalidade • Sessões por canal	• Calendário • Funcionalidade • Horário • Programa • Dispositivo • Plataforma • Cliente

Apesar da ligação em direto permitir o acesso a todas as métricas e dimensões presentes no cubo, nem todas foram necessárias para a construção do *dashboard*, respeitando os requisitos previamente estabelecidos. As tabelas usadas apresentam-se na Tabela 3.

4.5. Etapa 5 – Construção do *dashboard*

Depois de definida a estrutura e o design e de identificada a fonte dos dados, a construção do *dashboard* passa pela sua implementação na ferramenta designada. No presente projeto, esta ferramenta foi o Microsoft Power BI, o software de *reporting* mais utilizado pela empresa na construção dos seus relatórios e análises de negócio. A outra ferramenta com elevada utilização na NOS é o Microsoft Excel. No entanto, manteve-se a escolha do Microsoft Power BI devido à sua maior disponibilidade de recursos e elementos visuais e à maior facilidade de manipulação de elementos e filtros. Para além destas duas ferramentas, mais nenhuma foi considerada porque introduzir a gestão de um novo *software* nas dinâmicas dos *stakeholders* do projeto não seria benéfico nem eficiente.

Graças à existência de uma estrutura bem definida e validada, a construção propriamente dita do *dashboard* em Power BI foi uma etapa que demorou relativamente pouco tempo. No entanto, este processo passou pela construção várias versões, frequentemente apresentadas à equipa CRM-RAN e aos departamentos de Gestão do

Produto e Gestão de Conteúdos, até à sua versão final. Estas versões são apresentadas na Tabela 4. A versão final do *dashboard* será apresentada ao detalhe no Capítulo 5.

Tabela 4: Diferentes versões do dashboard

Versão	Semana	Detalhes
0.1	6	Primeira versão interna, apresentada ao orientador do projeto.
0.2	7	Versão apresentada à equipa CRM-RAN.
0.3	7	Versão apresentada a membros dos departamentos de Gestão de Conteúdos e Gestão de Produto.
1.0	7	Versão apresentada ao diretor do departamento CRM.
1.1	10	Versão final do <i>dashboard</i> , apresentada aos diretores dos departamentos de Gestão de Conteúdos, Gestão de Produto, Oferta e Segmentação, e Experiência e Valor.

É importante referir que, devido à ligação em direto estabelecida entre o Microsoft Power BI e a base de dados, a atualização dos dados do *dashboard* é feita de forma autónoma, paralelamente à atualização dos dados na base de dados. Por este motivo, após construída a versão final do *dashboard*, não é necessário fazer a sua atualização periódica manualmente, sendo que alterações serão feitas apenas a pedido dos utilizadores, após validação da equipa CRM-RAN segundo os critérios: simplicidade, utilidade e adequação.

4.6. Etapa 6 – Validação e avaliação do *dashboard*

A construção do *dashboard* consistiu num processo iterativo, onde a adição, eliminação ou modificação de métricas, parâmetros de filtro e características dos elementos visuais foram baseadas no feedback recebido nas várias apresentações, representado o fim de cada iteração. A versão final do *dashboard* foi considerada aquela a partir da qual não foram realizadas quaisquer modificações, a versão 1.1, concluída e apresentada na semana 10 de elaboração do projeto. A Tabela 4 mostra as várias versões de construção do *dashboard*. A abordagem aplicada nesta etapa de validação e avaliação do *dashboard* consistiu em alargar progressivamente o alcance do mesmo, à medida que as melhorias iam sendo aplicadas. Isto permitiu garantir que após a apresentação a um novo *stakeholder*, o *dashboard* não fosse rejeitado.

As versões 0.1 a 0.3 são as versões de construção do *dashboard*, apresentadas e discutidas apenas com membros das equipas CRM-RAN, Gestão de Conteúdos e Gestão de Produto que participaram na elaboração do *dashboard*. Nenhuma destas versões foi apresentada a diretores. As três principais oportunidades de validação e

ajuste do *dashboard* foram, por ordem cronológica, a apresentação ao orientador do projeto, a toda a equipa CRM-RAN, e às equipas de Gestão de Conteúdos e Gestão de Produto, marcando, cada uma, uma nova versão do *dashboard*.

Destas oportunidades de melhoria resultou a versão 1.0 do *dashboard*, versão que, após todo o feedback recebido, garantia qualidade suficiente para ser apresentada ao diretor do departamento de CRM. Finalmente, após atuar sobre pequenas sugestões recebidas, a versão final, 1.1, do *dashboard* ficou concluída. Esta versão foi então apresentada aos diretores de quatro departamentos, Gestão de Conteúdos, Gestão de Produto, Oferta e Segmentação, e Experiência e Valor, que receberam o *dashboard* com interesse e entusiasmo, deixando um feedback bastante positivo sobre a organização e a utilidade da informação apresentada. No entanto, um fator destacado no final da apresentação foi a sensibilidade da informação contida no *dashboard*, o que levantou questões sobre partilha e segurança. Por este motivo, ficou definido que o *dashboard* não ficará público, mas sim disponível às equipas que necessitam frequentemente da sua informação, consoante a aprovação do diretor do departamento onde a equipa se insere.

As principais modificações que surgiram ao longo desta etapa estão apresentadas a seguir.

- Introdução dos seguintes parâmetros de filtro: plataforma de transmissão, perfil do cliente e categoria do conteúdo;
- Criação de uma página inicial, o glossário, onde são apresentadas as restantes páginas do *dashboard*, bem como as métricas disponíveis em cada uma;
- Exclusão de todas as sessões com duração inferior a 3 minutos, de forma a descartar os efeitos do *zapping* (relembra-se que se considera uma sessão o tempo de permanência de um utilizador no mesmo canal, através da mesma funcionalidade de visualização);
- Implementação da funcionalidade de comparação de métricas, segundo diferentes parâmetros de filtro.

A versão final será analisada em detalhe no Capítulo 5.

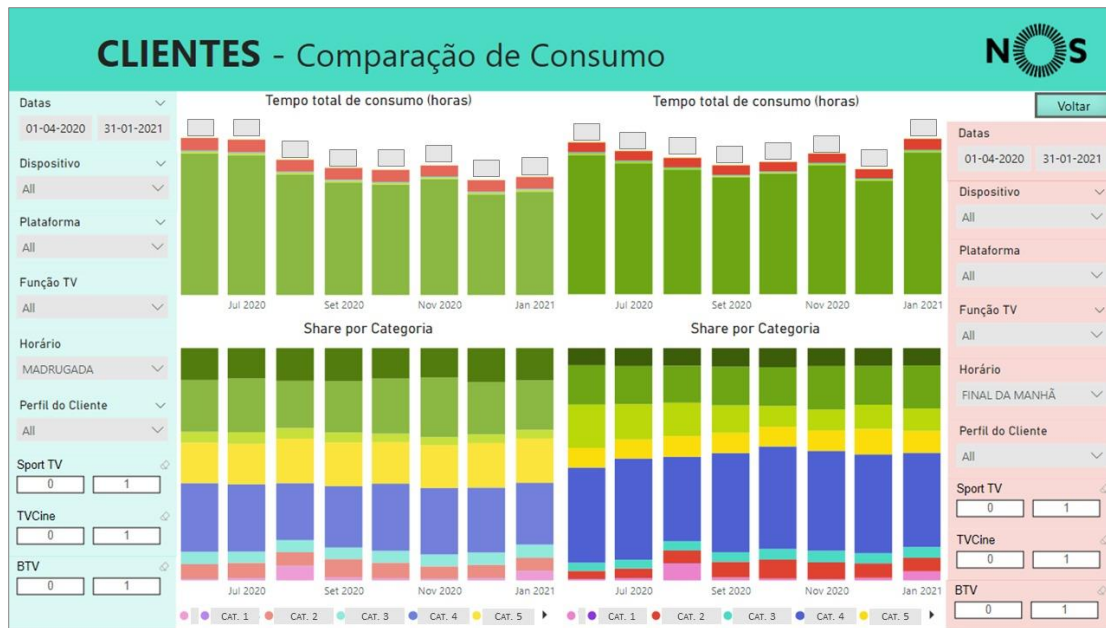


Figura 4: Página de comparação da versão do dashboard

4.7. Etapa 7 – Caracterização dos clientes

Após finalizado o *dashboard*, abordou-se a segunda oportunidade identificada, a exploração dos dados de audimetria com o objetivo de identificar padrões de consumo e características de cada tipo de clientes. A concretização deste processo pode dividir-se em cinco fases:

1. Definição dos dados e das variáveis;
2. Preparação dos dados e aplicação do algoritmo de *clustering*;
3. Definição dos *clusters*;
4. Distribuição do parque de clientes pelos *clusters*;
5. Caracterização do parque de clientes com base nos *clusters*.

Em primeiro lugar, foi necessário definir o alcance temporal dos dados considerados para a análise. Devido às limitações de performance e à complexidade dos dados de audimetria, considerou-se que dados de um mês proporcionam um bom equilíbrio entre representatividade de informação e eficiência dos recursos. Assim, os dados utilizados para a caracterização dos clientes foram os dados de audimetria e de cliente de janeiro de 2021.

Tabela 5: Variáveis utilizadas na primeira e na última iteração do processo de clustering

Primeira iteração	Última iteração
• Tempo total de consumo (semana) • Tempo total de consumo (fim de semana) • Tempo médio de consumo diário (semana) • Tempo médio de consumo diário (fim de semana) • Dias com consumo (semana) • Dias com consumo (fim de semana) • Número de sessões em zapping • Plataforma mais utilizada • Dispositivo mais utilizado • Número de dispositivos • Número de cartões móveis • Tempo de consumo por categoria e por funcionalidade • Tempo de consumo por funcionalidade e por horário • Tempo de consumo por horário e por categoria	• Tempo total de consumo • Tempo médio de consumo diário • Plataforma mais utilizada • Número de dispositivos • Número de cartões móveis • Percentagem de consumo linear • Percentagem de cada categoria em consumo linear • Percentagem de cada categoria em consumo não linear • Percentagem do consumo em cada horário em consumo linear • Percentagem do consumo em cada horário em consumo não linear

Relativamente à seleção das variáveis, esta tarefa foi integrada num processo iterativo que consistiu na extração e preparação dos dados, aplicação do algoritmo de *clustering*, validação e interpretação dos resultados e revisão das variáveis utilizadas. Esta metodologia permitiu chegar às variáveis que mais influência exercem na criação de grupos de clientes, através de eliminação de variáveis redundantes ou com pouco impacto. A eliminação de variáveis, de iteração para iteração, baseou-se na análise da correlação entre as variáveis, através do coeficiente de correlação de Pearson, e na semelhança entre os valores das variáveis nos diferentes *clusters*, através do valor médio e do desvio padrão, isto é, o impacto da variável na criação dos *clusters*. Ao longo do processo, foram também adicionadas algumas variáveis, resultado das várias reuniões de acompanhamento.

Também nesta etapa do projeto se sentiram dificuldades ao nível da performance. Dado que um grande número de atributos tornaria impraticável o

processamento dos *clusters* com os recursos disponíveis, foi necessário ter esse aspeto em conta na definição dos atributos iniciais. Os atributos considerados na primeira e na última iteração estão apresentados na Tabela 5.

As principais modificações na base de atributos, até se chegar ao conjunto de atributos final, são explicadas em seguida.

- Foi eliminado o fator do fim de semana no tempo total de consumo e no tempo médio de consumo diário, já que os valores destas variáveis para os dias da semana e para o fim de semana se revelam altamente relacionados (coeficiente de correlação $> 0,99$).
- Foi eliminada a variável ‘dias com consumo’, por não apresentar qualquer impacto na construção dos *clusters*. Depois de se calcular o valor médio desta variável para cada *cluster* criado, observou-se a semelhança entre todos os valores (valor médio = 18; desvio padrão = 1,42), concluindo-se que o impacto desta variável na criação dos *clusters* é praticamente nulo.
- Foi eliminada a variável ‘número de sessões em zapping’, por apresentar uma elevada correlação com a variável ‘tempo total de consumo’ (coeficiente de correlação = 0,83).
- Para diminuir o número de variáveis existentes e otimizar a eficiência do processo, optou-se por agrupar as cinco funcionalidades em consumo linear (Live TV) e consumo não linear (NPVR/PVR, Timewarp, Restart e VOD).

A preparação dos dados para extração envolveu uma série de queries, em ambiente Oracle SQL Developer, com vários objetivos. Primeiro, recolher os dados de audimetria relevantes e transformar essa mesma tabela, com uma linha por sessão, numa tabela com uma linha por cliente, de forma a ser possível aplicar o algoritmo de *clustering* aos vários pontos, onde cada ponto representa um cliente. Em segundo, juntar os dados de cliente aos dados de audimetria. Lembra-se que se manteve a representação de cada cliente pelo código da sua SA.

Depois de reunidos todos os dados necessários, fez-se a sua importação para o ambiente Spyder, através da linguagem Python. Aqui, concluiu-se a preparação dos dados, o que incluiu o tratamento das variáveis qualitativas através da criação de variáveis *dummy*, e ainda a normalização dos dados através do método Min-Max Normalization, de forma a manter a mesma escala de transformação em todos os atributos.

Com os dados preparados, aplicou-se o algoritmo K-means para a criação dos *clusters*. A definição do número ótimo de *clusters* fez-se através do Elbow Method,

cujas curvas de aplicação na Figura 5, resultando numa divisão dos clientes em 16 *clusters* ($WSS = 183 \times 10^3$).

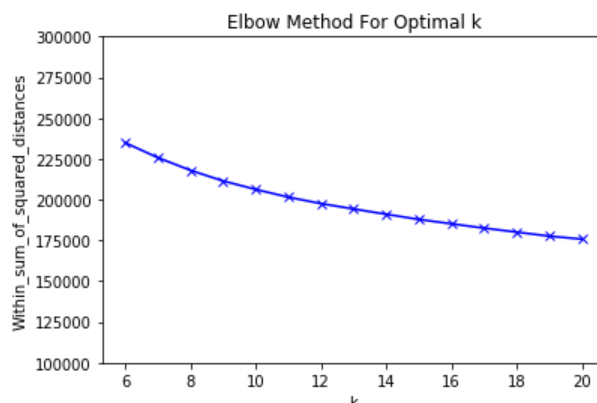


Figura 5: Curva de aplicação do Elbow Method para a definição do número ótimo de clusters

O passo seguinte consistiu na validação e na atribuição de um significado prático a cada *cluster*. Para tal, foram analisados, para cada *cluster*, os valores médios de cada variável considerada. Nesta fase, foi necessário recorrer ao conhecimento do negócio e a algumas pré-conceções relativamente ao comportamento de consumo de televisão dos clientes, de forma a se poder identificar o tipo de cliente representado por cada *cluster*. Depois da análise individual do estudante, achou-se benéfico fazer uma sessão de brainstorming com toda a equipa CRM-RAN, de forma a validar e atribuir as classificações finais. Durante esta fase, reconheceu-se a dificuldade em gerir 16 *clusters*, pelo que foram agrupados alguns deles, de acordo com a similaridade das suas características, dando origem a 8 grupos, alguns deles com subgrupos. A apresentação e caracterização ao detalhe dos grupos será feita no Capítulo 5.

Depois de serem identificados os atributos de definem cada *cluster* e que permitem a melhor divisão dos clientes pelos *clusters*, definiram-se os limites destes mesmos atributos para cada *cluster*, isto é, os critérios de distribuição dos clientes por cada *cluster*, disponível na Figura 18. O passo seguinte consistiu na distribuição prática dos clientes do parque de janeiro de 2021 pelos *clusters*, através de uma query em ambiente Oracle SQL Developer. Tal como a definição de variáveis para os *clusters*, também a definição dos critérios de distribuição pelos *clusters* se obteve como resultado de um processo iterativo. Deste processo fez parte a distribuição pelos *clusters* com posterior revisão dos critérios, onde o objetivo foi minimizar a diferença entre as características dos *clusters* inicialmente gerados e as características dos *clusters* criados pela query. Obtiveram-se, assim, os critérios de distribuição de uma base de clientes de televisão pelos 16 *clusters*. É importante referir que apenas é possível caracterizar os clientes com dados de audimetria disponíveis, isto é, os clientes de televisão com tecnologia de transmissão por Cabo e com consumo de conteúdos televisivos.

Tabela 6: Grupos de clientes resultantes do algoritmo de clustering

• Famílias Tradicionais ○ Heavy users ○ Tarde ○ Pré prime-time ○ Prime-time ○ Late night • Famílias Digitais ○ Normais ○ Light • Espectadores em Direto ○ Generalistas ○ Entretenimento	• Cinéfilos ○ Filmes ○ Séries • Clientes Cor-de-rosa ○ Novelas ○ Reality shows • Famílias com Crianças • Fãs de Informação • Adeptos de Desporto
--	--

4.8. Etapa 8 – Conclusão do *dashboard*

Esta etapa encerra a metodologia, marcando o fim do projeto. O objetivo desta etapa é assegurar que todo o trabalho realizado continuará funcional no futuro, mantendo-se uma ferramenta útil para toda a organização. A conclusão do projeto focou-se nos dois objetivos alcançados. Por um lado, foi necessário garantir que o *dashboard* se mantém operacional e com informação atualizada. Por outro, assegurar que a caracterização dos clientes TV se mantém atual.

A primeira tarefa foi facilmente cumprida. A ligação em direto estabelecida do Microsoft Power BI à fonte dos dados permite que a atualização dos dados do *dashboard* ocorra simultaneamente à atualização da fonte, responsabilidade assegurada por outras equipas. Por este motivo, e devido à estrutura dinâmica e às características dos elementos do *dashboard*, a sua manutenção e atualização não exigem qualquer procedimento especial. Apesar de ter sido o estudante a construir o *dashboard* de raiz, tornando-o o elemento mais apto para a manutenção do mesmo, o acompanhamento frequente do projeto pelo orientador na empresa permitiu-lhe adquirir o conhecimento necessário a esta tarefa, tornando-o, também, apto para dar continuidade à utilidade do *dashboard*.

Relativamente ao objetivo proposto na segunda tarefa, este foi atingido através de dois componentes. O primeiro é constituído por um documento com as instruções de trabalho do processo de caracterização do parque de clientes TV através da sua distribuição pelos *clusters*. Este documento apresenta ainda informação sobre a equipa responsável pelo processo, a frequência de execução e os programas necessários. Por

outro lado, o código SQL elaborado, que permite a execução do processo de distribuição pelos *clusters*, está enriquecido de grande volume de comentários, de forma a facilitar a leitura e compreensão do código em si. Assim, estes dois componentes asseguram que qualquer colaborador, devidamente familiarizado com os processos internos da equipa CRM-RAN, esteja habilitado não só a atualizar a caracterização do parque, mas também a efetuar qualquer correção ou adaptação necessária ao processo.

No capítulo seguinte, serão analisados os resultados do projeto da presente dissertação. Em primeiro lugar, o *dashboard* com os dados de audimetria e, em segundo, o processo de *clustering* e caracterização dos clientes TV.

Capítulo 5

Resultados

Neste capítulo, serão apresentados os resultados do projeto elaborado no âmbito desta dissertação. A primeira parte irá focar-se na estrutura e nos componentes da versão final do *dashboard*. Será também apresentada a opinião de um utilizador do *dashboard*, membro da equipa CRM-RAN. Por fim, serão expostos os resultados da caracterização dos clientes através do algoritmo de *clustering*.

5.1. *Dashboard* “Universo de Audimetria”

Como resultado do processo de construção iterativo adotado, baseado no feedback de vários utilizadores, chegou-se a uma versão final do *dashboard*, disponível no Anexo A. A sua estrutura, elementos visuais e parâmetros de filtros serão expostos nas próximas subsecções. No final, será apresentada a opinião de um utilizador sobre o impacto e os objetivos atingidos do *dashboard*.

5.1.1. Estrutura e organização

A estrutura do *dashboard* compõe-se por quatro secções que, por sua vez, são compostas por uma ou várias páginas. Esta organização é a seguinte:

Secção 1: Clientes

Página 1.1.a: Clientes – Utilização

Página 1.1.b: Clientes – Utilização – Comparação

Página 1.2.a: Clientes – Consumo

Página 1.2.b: Clientes – Consumo – Comparação

Secção 2: Categorias

Página 2.1.a: Categorias por Funcionalidade

Página 2.1.b: Categorias por Funcionalidade – Comparação

Página 2.2.a: Categorias por Horário

Página 2.2.b: Categorias por Horário – Comparação

Secção 3: Canais

Página 3.1: Canais

Secção 4: Programas

Página 4.1: Programas

A divisão por secções permite facilitar a abordagem inicial do utilizador, possibilitando uma navegação mais direta para a informação desejada. A primeira secção engloba as métricas que quantificam e qualificam o parque de clientes, com a sua evolução ao longo do tempo. A segunda secção agrupa a informação por categoria de conteúdo. Seguindo esta lógica de atuação, a terceira e quarta secções agrupam a informação por canal e programa, respetivamente.

Relativamente à identificação visual de cada secção, esta foi conseguida através da atribuição de uma cor diferente a cada uma. Estas cores estão presentes na página inicial do *dashboard*, o glossário, e ainda no cabeçalho e como fundo dos parâmetros de filtro das respetivas páginas. Os diferentes cabeçalhos do *dashboard* são apresentados na Figura 6.

CLIENTES Utilizadores	
CLIENTES Consumo	
CATEGORIAS por Função TV	
CATEGORIAS por Horário	
CANAIS	
PROGRAMAS	

Figura 6: Cabeçalhos utilizados nas várias páginas do dashboard

A organização da informação dentro de cada página também seguiu uma estrutura constante, ao longo de todo o *dashboard*, com uma ligeira alteração nas secções 3 e 4, devido à adição de parâmetros de filtro específicos nas páginas destas secções. As páginas de comparação, no entanto, por apresentarem os elementos visuais duplicados, apresentam uma organização diferente.

De forma a manter uma navegação intuitiva e consistente, os filtros foram colocados do lado esquerdo da página. Assim, os elementos visuais que apresentam a

informação ocupam o centro e o lado direito da página. A área ocupada pelos filtros foi construída com um fundo de cor diferente, de forma a haver um contorno claro da área designada à apresentação da informação. As páginas das secções 3 e 4, apresentam ainda parâmetros de filtro específicos na parte superior desta área.

As páginas de comparação apresentam uma simetria em relação a um eixo vertical. De cada lado, são apresentados todos os parâmetros de filtro e os gráficos afetados por estes. Assim, os gráficos afetados por filtros diferentes são apresentados no centro da página, lado a lado, permitindo uma comparação direta de informação.



Figura 7: Layout das páginas das secções 1 e 2 do dashboard

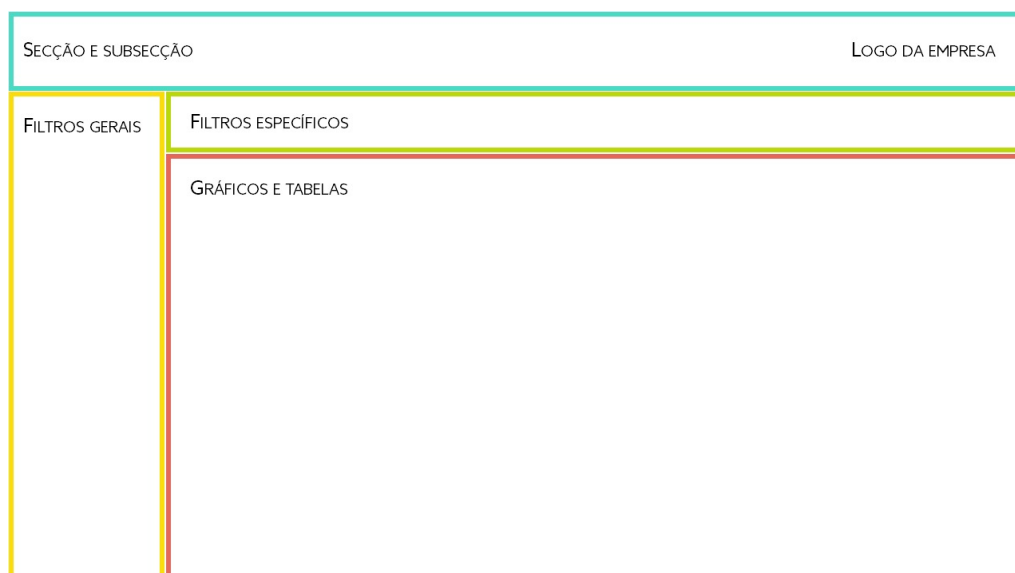


Figura 8: Layout das páginas das secções 3 e 4 do dashboard



Figura 9: Layout das páginas de comparação do dashboard

Relativamente à informação apresentada nas diferentes secções, a primeira secção contém, na primeira página, informação sobre a evolução do número de utilizadores diferentes, representados por SA distintas, do serviço de televisão da NOS, total, divididos por funcionalidade e divididos por horário. Na segunda página é apresentada informação sobre a evolução do tempo de consumo de televisão, total e por funcionalidade, a duração média das sessões, e ainda a distribuição do tempo de consumo total por categoria de conteúdo. A segunda secção apresenta informação sobre o número de utilizadores por categoria, a distribuição do tempo de consumo por categoria e a duração média das sessões de cada categoria. Os dados destas três métricas encontram-se segmentados por funcionalidade, na primeira página, e por horário, na segunda. A terceira secção apresenta a lista dos dez canais com maior tempo de consumo num determinado período de tempo, com os respetivos dados de share, número de utilizadores, tempo de consumo e duração média das sessões no canal. Existem ainda mais dois gráficos, um com a distribuição do consumo dos cinco canais mais vistos por horário, e outro com a evolução do tempo total de consumo destes canais. Finalmente, na última secção pode consultar-se uma lista com os programas com maior consumo num determinado período de tempo e informação sobre o número de utilizadores, o tempo total de consumo, a duração média das sessões e ainda o canal de transmissão do programa. Nesta página está também disponível informação sobre a distribuição do tempo de consumo dos programas dos canais selecionados, por funcionalidade e horário, e ainda a evolução do seu tempo total de consumo.

De forma resumida, as páginas do *dashboard* encontram-se agrupadas em secções, de acordo com o nível de profundidade da informação desejada: clientes, categoria, canal e programa, do mais abrangente para o mais específico. Qual a

informação que está presente em cada página é uma questão que pode ser respondida rapidamente através da consulta do glossário do *dashboard*.

A informação apresentada no *dashboard* foi selecionada de forma a responder aos requisitos propostos, respondendo à primeira oportunidade identificada, a necessidade sentida por várias equipas em recolher informação relativa aos dados de audimetria dos clientes da NOS de forma rápida, simples e independente.

Até este ponto, clarificou-se a estrutura e a organização do *dashboard*. Nas próximas subsecções serão analisados os elementos visuais e os filtros utilizados.

5.1.2. Elementos visuais

Os elementos visuais são os componentes do *dashboard* que possibilitam ao utilizador visualizar e interpretar os dados, facilitando a extração de informação e conhecimento. Por este motivo, são dos elementos mais importantes do *dashboard*, ocupando a maior parte da área de cada página.

Dependendo da agregação temporal da informação, os gráficos podem dividir-se em dois grupos: séries temporais e gráficos estáticos. As séries temporais apresentam a evolução de uma determinada métrica, ao longo do tempo. Presentes na primeira, terceira e quarta secções do *dashboard*, incluem: gráficos de colunas empilhadas (simples e 100%), gráficos de linhas e gráficos de área.

Os gráficos estáticos apresentam a informação agrupada para o período de tempo definido no filtro das datas. Este tipo de gráficos está presente na segunda, terceira e quarta secções do *dashboard*, e incluem: gráficos de barras e colunas empilhadas (simples e 100%), gráficos de pontos, gráficos circulares e tabelas.

É importante salientar que, devido à sensibilidade e teor confidencial da informação do *dashboard*, os dados dos elementos visuais apresentados nesta dissertação não são reais. No entanto, apenas os valores e algumas legendas foram censurados ou alterados.

5.1.2.1 Séries temporais: gráficos de colunas empilhadas (simples e 100%)

A maior parte das séries temporais encontram-se na primeira secção do *dashboard*. Nas páginas desta secção, predominam os gráficos de colunas empilhadas, onde cada coluna representa a informação de um determinado mês. As métricas apresentadas através deste tipo de elemento visual são o número de utilizadores e o tempo total de consumo, segmentados por funcionalidade e por horário. A vantagem deste tipo de gráficos é a de possibilitar uma rápida visualização da métrica apresentada, não só o seu valor total, mas também a sua segmentação de acordo com o parâmetro selecionado para a legenda.

A Figura 10 mostra um exemplo da utilização deste elemento visual, utilizado na página 1.1 do *dashboard*.

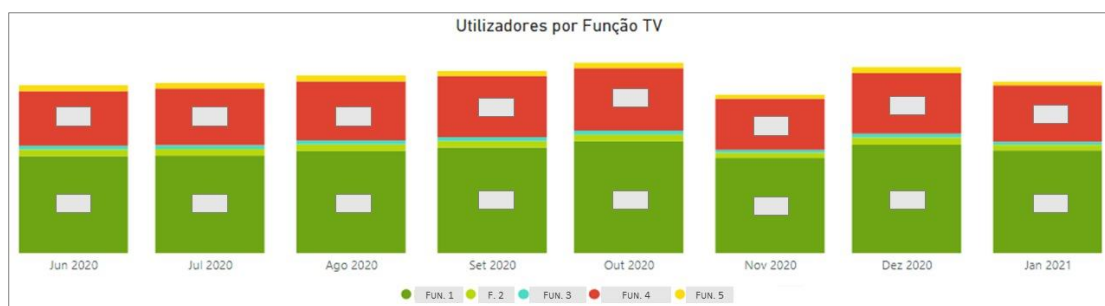


Figura 10: Gráfico de colunas empilhadas simples – Evolução do número de utilizadores por funcionalidade

Para além dos gráficos simples, utilizaram-se também os gráficos de colunas empilhadas 100%. Este tipo de gráficos omite informação relativa ao total da métrica apresentada, mas expõe de forma clara e direta a distribuição dos dados pelos diferentes parâmetros do atributo usado para segmentação. Um exemplo da utilização deste gráfico, mostrado na Figura 11, foi para a apresentação da distribuição do tempo de consumo dos clientes pelas diferentes categorias de conteúdo, ao longo do tempo.

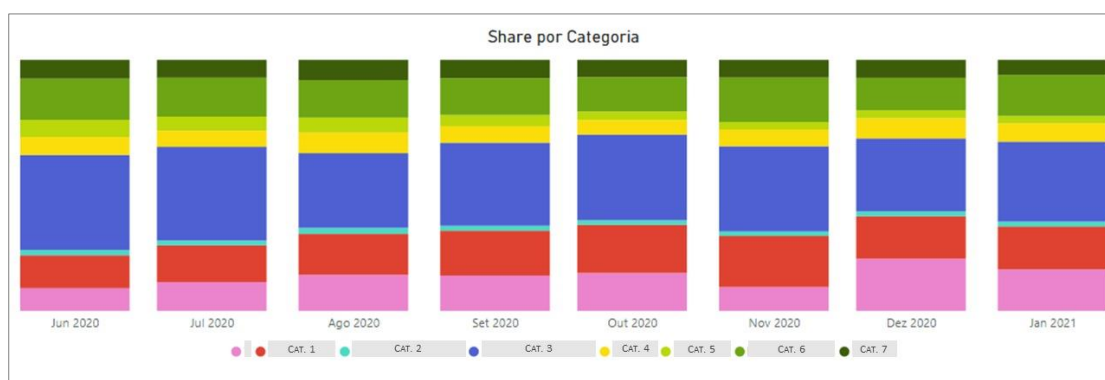


Figura 11: Gráfico de colunas empilhadas 100% - Distribuição do tempo de consumo por categoria

5.1.2.2 Séries temporais: gráficos de linhas e gráficos de área

Os gráficos de linhas enfatizam o formato geral da evolução temporal de uma série de valores. Por sua vez, os gráficos de área, baseados nos gráficos de linhas, realçam a magnitude da alteração ao longo do tempo.

No *dashboard*, estes elementos foram utilizados na primeira, terceira e quarta secção, apresentando, respetivamente, a evolução da duração média das sessões por funcionalidade, o tempo de consumo de vários canais e vários programas. Estes gráficos permitem compreender como evoluiu o comportamento de visualização dos clientes, ao nível da duração das sessões ou do canal e programa consumido, facilitando a identificação e compreensão de tendências.

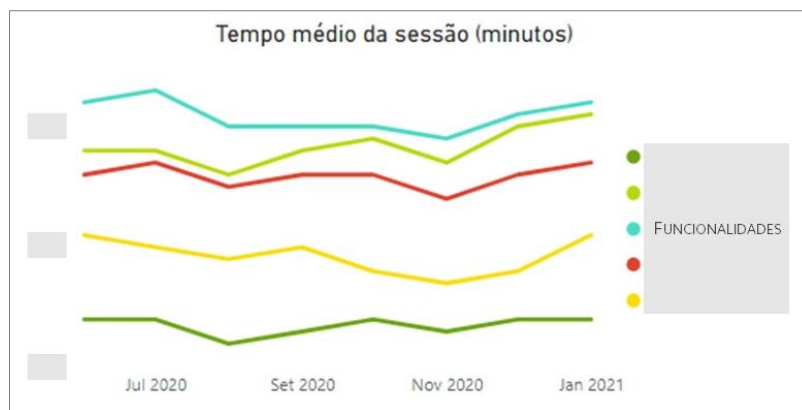


Figura 12: Gráfico de linhas - Evolução da duração média das sessões por funcionalidade

Em todas as séries temporais, é apresentada a evolução mensal dos dados, com exceção do gráfico da secção Programas, que expõe a evolução diária do consumo de um conjunto de programas. Este elemento visual permite, por exemplo, a identificação do efeito dos fins de semana e dos feriados no consumo de alguns tipos de programas. Por outro lado, o gráfico que apresenta a evolução mensal do consumo de vários canais facilita a identificação do impacto das férias escolares no tempo de visualização de canais com conteúdo maioritariamente infantil, por exemplo.

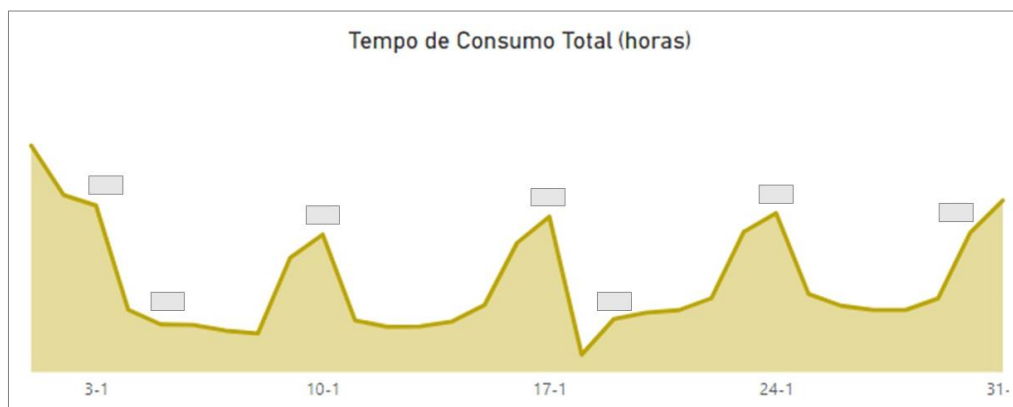


Figura 13: Gráfico de área - Evolução diária do consumo total dos programas selecionados

5.1.2.3 Estáticos: gráficos de barras e colunas empilhadas (simples e 100%)

Os gráficos estáticos de barras e colunas empilhadas são os elementos visuais que predominam na segunda secção do *dashboard*. Esta secção apresenta a informação agrupada por categoria de conteúdo, considerando o intervalo de tempo definido no filtro de datas. De forma a responder aos requisitos, foi necessário segmentar a informação por duas variáveis, funcionalidade e horário. Por este motivo, escolheram-se os gráficos de barras e colunas empilhadas para transmitir esta informação, onde cada barra ou coluna representa uma categoria, estando esta dividida nos vários componentes da variável de segmentação utilizada. A decisão entre gráfico de barras

ou colunas prendeu-se com a composição visual da página. A Figura 14 mostra um exemplo da utilização deste tipo de elemento visual.

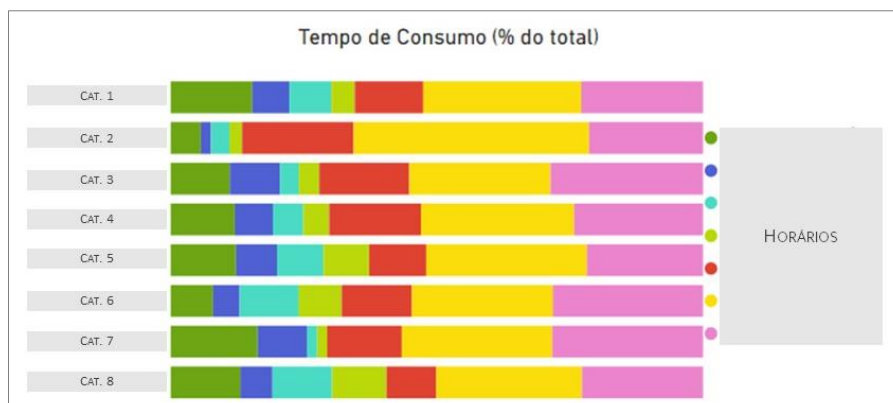


Figura 14: Gráfico de barras empilhadas 100% - Tempo de consumo em cada horário por categoria

5.1.2.4 Estáticos: gráficos de pontos

Os gráficos de pontos são idênticos aos gráficos de linhas, sendo que a sua utilização é recomendada quando não existe qualquer relação entre os atributos da variável independente. Este é o caso da duração média das sessões em cada categoria de conteúdo, de acordo com a funcionalidade (Figura 15) e o horário de visualização.

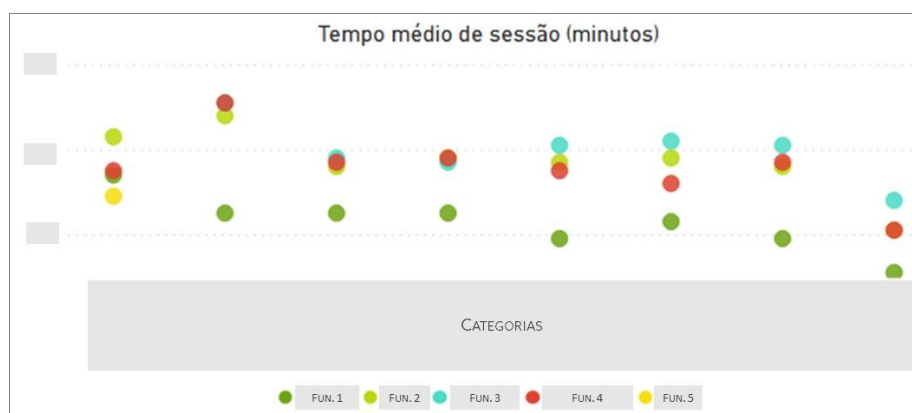


Figura 15: Gráfico de pontos - Duração média das sessões em cada funcionalidade por categoria

5.1.2.5 Estáticos: gráficos circulares

Os gráficos circulares permitem comparar as categorias de uma determinada variável entre elas ou em relação ao todo. Apenas foram utilizados na última secção do *dashboard* para compreender a distribuição do consumo de um ou vários programas pela várias funcionalidades e horários de visualização (Figura 16). A vantagem que se pretende obter com este tipo de elemento visual é uma compreensão rápida da categoria que se destaca para a variável em questão. Devido ao elevado número de categorias, e para permitir uma análise não só qualitativa, mas também quantitativa, um componente

essencial destes gráficos são as etiquetas, que mostram o valor da variável para cada categoria (valor absoluto ou relativo).

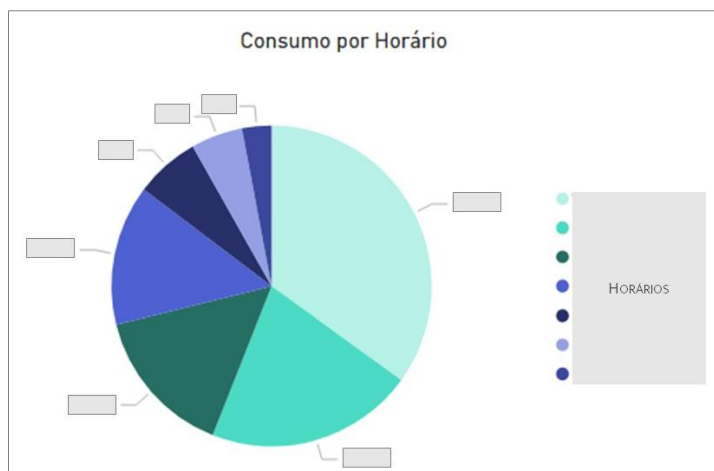


Figura 16: Gráfico circular - Distribuição do consumo dos programas selecionados por horário

5.1.2.6 Estáticos: tabelas

Finalmente, as tabelas foram utilizadas na terceira e na quarta secções do *dashboard*. O seu uso permite listar um grande número de objetos de forma clara, incluindo todas as variáveis desejadas. Na página 3.1 do *dashboard* introduziu-se uma tabela (figura X) que apresenta, para os 10 canais com maior tempo de consumo, por ordem decrescente, o seu nome, número de utilizadores, horas de consumo, duração média das sessões e share. Na página 4.1 utilizou-se uma tabela idêntica, mas com informação acerca dos programas. As variáveis seleccionadas são o nome do programa, o canal de transmissão, o número de utilizadores, as horas de consumo e, por fim, a duração média das sessões.

TOP Canais	Utilizadores	Horas	ATS (min)	Share
CANAL 1			25	14,14 %
CANAL 2			26	12,53 %
CANAL 3			27	11,79 %
CANAL 4			26	9,43 %
CANAL 5			21	6,83 %
CANAL 6			32	6,65 %
CANAL 7			16	6,60 %
CANAL 8			37	4,88 %
CANAL 9			24	3,70 %
CANAL 10			28	3,55 %

Figura 17: Tabela - Top de canais com maior consumo

Após a compreensão dos elementos visuais utilizados, serão apresentados os parâmetros de filtro seleccionados.

5.1.3. Filtros

Num *dashboard*, os filtros permitem restringir o alcance dos dados apresentados, num ou em vários parâmetros. Sem alterar a construção do *dashboard*, os filtros permitem não só seleccionar os dados expostos em cada página, como também alterar dinamicamente essa seleção. São, por isso, a componente do *dashboard* que mais diretamente afeta a sua flexibilidade.

Os filtros utilizados no presente *dashboard* podem ser divididos em três grupos: filtros temporais, de audimetria e de cliente. A caracterização dos grupos está diretamente relacionada com o tipo de variáveis afetadas por cada filtro.

Os filtros temporais, tal como o nome indica, permitem que seja definido o intervalo temporal da análise, restringindo os dados ao intervalo selecionado. Este filtro está presente em todas as páginas do *dashboard*.

Os filtros de audimetria afetam as variáveis que dizem respeito à forma como os conteúdos são consumidos e também à caracterização destes mesmos conteúdos. Os primeiros incluem o dispositivo de visualização, a plataforma de transmissão, a funcionalidade e o horário de visualização. Estes filtros estão presentes em todas as páginas do *dashboard*. Os filtros relativos à caracterização do conteúdo incluem a categoria do conteúdo transmitido, presente nas secções 3 e 4, e o canal de transmissão, apenas presente na quarta secção do *dashboard*. Todas as variáveis consideradas nos filtros deste grupo são variáveis categóricas, pelo que o tipo de filtro utilizado foi o de seleção múltipla.

Os filtros de cliente restringem a informação, não ao nível dos conteúdos consumidos pelos utilizadores, mas sim ao nível dos utilizadores em si. Estes filtros são o perfil do cliente e a existência (ou não) de subscrição em três canais premium com grande impacto, Sport TV, TVCine e BTV. O filtro do perfil do cliente, dado o carácter categórico da variável em causa, é do tipo de seleção múltipla. Por outro lado, as variáveis de subscrição dos canais são binárias, onde 1 marca a existência de subscrição e 0 marca a sua ausência. Por este motivo, o filtro utilizado foi o de seleção única, já que o estado pré-definido do filtro é o de nenhuma seleção, não efetuando qualquer corte nos dados.

É importante referir que os filtros afetam todas as páginas do *dashboard* onde se encontram, de forma a facilitar a consulta das métricas. Isto é, se, por exemplo, o filtro da plataforma for alterado na página 1.1 do *dashboard*, este será automaticamente alterado nas restantes páginas, de forma a eliminar essa responsabilidade do utilizador e tornar a análise mais eficiente.

5.1.4. Avaliação do utilizador

A análise técnica da versão final do *dashboard* foi discutida na subsecção anterior. Na prática, o sucesso de um *dashboard* depende de vários aspetos, avaliados pelos seus utilizadores, como a sua utilidade, facilidade de navegação, simplicidade e clareza, e eficácia na apresentação de informação. Para além dos comentários positivos após a apresentação do *dashboard* aos diretores de vários departamentos, especificados na Secção 4.5, para avaliar todos estes parâmetros, foi conduzida uma entrevista a um utilizador do *dashboard*, membro da equipa CRM-RAN, após um mês de utilização.

A entrevista começou com a apresentação do colaborador e da sua equipa, ao que se seguiu uma série de questões acerca do *dashboard* construído neste projeto, cujas respostas se apresentam reunidas de seguida.

Motivações e objetivos do *dashboard*

Para dar resposta aos pedidos frequentes de várias equipas da NOS, de informação relativa a dados de audimetria, a equipa CRM-RAN via-se obrigada a análises pontuais e demoradas, dependendo do tipo de pedido em questão, já que os dados de audimetria se encontravam disponíveis apenas em bruto. Assim, o objetivo principal do *dashboard* era reunir a informação de audimetria, bastante pesada e complexa, num formato de fácil acesso e de onde fosse possível extrair informação e conhecimento de forma rápida, sem a necessidade de processos ou recursos intermédios.

Sucesso do *dashboard* e impacto no dia-a-dia das equipas

Após a construção do *dashboard*, perante os vários pedidos de informação de outras equipas, os membros da equipa CRM-RAN têm agora uma ferramenta através da qual conseguem responder às perguntas colocadas de forma muito mais eficiente, o que aumenta a capacidade de resposta da própria equipa. Além disso, outras equipas com acesso ao *dashboard* ficam com capacidade para elaborar as suas análises de forma mais independente, já que a informação tratada lhes permite extrair o conhecimento necessário. Na opinião do entrevistado, o *dashboard* atingiu as expectativas estabelecidas, pois a sua estrutura permite responder às necessidades correntes das várias equipas interessadas, sendo também uma ferramenta extremamente útil na gestão do dia-a-dia de, por exemplo, um membro da equipa de Gestão de Produto.

Informação mais relevante do *dashboard*

Todas as informações incluídas no *dashboard* têm uma grande utilidade para a empresa. O que difere é a frequência com a qual cada informação é necessária. Por um lado, a informação das primeiras duas secções é importante para uma monitorização

mais regular, em alguns casos mesmo diária, da realidade do produto TV. Por outro lado, as duas últimas secções, com informação mais detalhada de Canal e Programa, apresentam informação que não é consultada tão frequentemente, mas que é de extrema importância para a preparação de algumas ações da empresa, principalmente na área da gestão de conteúdos.

Estrutura e experiência de navegação

A estrutura definida para o *dashboard* resulta numa interface *user-friendly*, de muito fácil navegação. Destaca-se também a página do Glossário, que explica de forma bastante clara a informação do restante *dashboard*, evitando uma situação em que se tem acesso a uma grande quantidade de informação, mas para a qual não se compreende o seu significado. A clarificação dos conceitos permite um alinhamento entre todas as equipas e análises, com ou sem base na informação do *dashboard*.

Oportunidades de melhoria

Até ao momento da entrevista, todas as necessidades de recolha de informação sentidas pelo colaborador entrevistado puderam ser respondidas recorrendo ao *dashboard*, o que é um indicador de uma boa definição e cumprimento dos requisitos. Também ao nível do design não foi apontado qualquer aspeto que devesse ser modificado. O único ponto mencionado foi o tempo do carregamento de informação que, devido à ligação em direto à fonte de dados, é, por vezes, longo. No entanto, este aspeto possibilita uma atualização diária da informação e é, considerando a alternativa exposta na Secção 4.4, a abordagem mais eficiente.

No geral, considerando a entrevista e o feedback recebido durante as apresentações, o *dashboard* atingiu com sucesso as expectativas estabelecidas. É a primeira ferramenta com toda a informação de audimetria tratada e com uma apresentação visual, pronta para a extração de conhecimento, pelo que será um apoio fundamental para análises e progressos futuros.

5.2. Caracterização de clientes através de um algoritmo de *clustering*

Como referido no Capítulo 4, a última etapa do projeto consistiu na caracterização do parque de clientes TV da NOS. Esta caracterização baseou-se na análise dos dados de audimetria de cada cliente, através dos quais foi possível identificar grupos de clientes com o mesmo comportamento de consumo de televisão, através da aplicação de um algoritmo de *clustering*.

Desta análise, resultaram 16 *clusters* que, após uma análise extensiva, qualitativa e quantitativa, dos seus atributos, foram agrupados em 8 grupos principais. Estes grupos e as suas características serão apresentados de seguida.

5.2.1. Famílias Tradicionais

O grupo das Famílias Tradicionais caracteriza-se por apresentar cerca de 50% do seu consumo de televisão em conteúdo das categorias de Informação, Filmes e Desporto. A média da percentagem de consumo não linear é ligeiramente inferior à média total do parque, indicando que estes consumidores, apesar de usufruírem das funcionalidades de consumo não linear, veem televisão principalmente em direto.

Dentro deste grupo, é ainda possível dividir os clientes em 5 subgrupos. Os clientes com um consumo médio diário superior a 4 horas e 45 minutos integram o subgrupo dos *Heavy Users*. Os restantes dividem-se de acordo com o período do dia no qual se verifica a maior parte do seu consumo, podendo este ser o período da Tarde, Pré prime-time, Prime-time ou *Late night*.

5.2.2. Famílias Digitais

As Famílias Digitais, tal como o nome indica, destacam-se pela plataforma de transmissão de conteúdos mais utilizada. Ao contrário dos restantes grupos, que consomem conteúdos maioritariamente através de boxes TV, as famílias digitais optam pela plataforma NOS TV. Paralelamente a esta característica, este grupo apresenta um número médio de dispositivos de visualização significativamente acima da média do parque, o que apoia a conclusão de que estes clientes visualizam conteúdos através de dispositivos móveis, como smartphones, tablets ou computadores. Relativamente ao consumo não linear, este valor apresenta-se ligeiramente acima da média do parque.

De acordo com o consumo médio diário, o grupo das famílias digitais pode ser dividido em *Light* ou Normais, caso este consumo seja inferior ou superior a 3 horas por dia, respetivamente. Em ambos os subgrupos se destaca o prime-time como o período com o maior consumo.

Relativamente às categorias de conteúdo, verifica-se que as Famílias Digitais consomem, em média, menos Informação e mais Desporto, comparativamente às Famílias Tradicionais. Este facto pode dever-se ao consumo de conteúdos de informação através de outros meios digitais e à visualização de eventos desportivos em dispositivos móveis, em simultâneo com outras atividades.

5.2.3. Espectadores em Direto

Este grupo de clientes destaca-se pelo seu consumo não linear ser praticamente nulo. Além disso, o número médio de cartões móveis e de dispositivos de visualização são significativamente inferiores à média do parque. Os Espectadores em Direto veem, em média, três horas e meia por dia, privilegiando o consumo durante o período da tarde.

Dentro deste grupo, ao nível das categorias com maior consumo, sobressaem dois pares: Informação e *Talkshows*, formando o subgrupo Generalista, e Filmes e Séries, formando o subgrupo de Entretenimento.

5.2.4. Cinéfilos

Os Cinéfilos ligam a televisão para ver, principalmente, filmes ou séries. Na maioria dos casos, o consumo numa destas categorias apresenta um destaque considerável em relação à outra, pelo que se criaram dois subgrupos, o de Filmes e o de Séries. Em ambos os casos, a segunda categoria com maior consumo é a categoria de Informação, sendo os períodos com maior consumo a tarde e o prime-time.

Os clientes deste grupo veem, em média, 4 horas de televisão por dia e são o segundo grupo que mais recorre ao consumo não linear, apresentando um valor desta variável consideravelmente superior à média do parque.

5.2.5. Clientes Cor-de-rosa

Os Clientes Cor-de-rosa, designação inspirada nas revistas cor-de-rosa, destacam-se pelo seu consumo de conteúdos das categorias de Novelas e *Reality shows*. Seguindo a lógica do grupo apresentado anteriormente, de acordo com a categoria com maior consumo, os Clientes Cor-de-rosa podem pertencer a um de dois subgrupos, Novelas ou *Reality shows*. Em ambos os casos, a maioria do consumo é feito nos períodos da tarde e prime-time.

Relativamente à quantidade e ao método de consumo, este grupo assiste, em média, a quatro horas e meia de televisão por dia, sendo o grupo que mais opta pelo consumo não linear. O número médio de cartões móveis e dispositivos de visualização apresenta-se ligeiramente superior aos valores médios do parque total de clientes.

5.2.6. Famílias com Crianças

Este grupo, tal como os dois grupos que serão apresentados de seguida, não apresentam subdivisão. As Famílias com Crianças são o terceiro grupo que mais recorre ao consumo não linear, apresentando esta variável um valor significativamente superior à

média do parque. Relativamente ao tempo de visualização, este é ligeiramente superior a duas horas por dia, em média.

A categoria com maior consumo é a categoria de conteúdo Infantil, representando cerca de duas horas diárias de visualização. Relativamente ao horário, o período com maior destaque é o período da tarde. Além disso, este grupo é o que apresenta a maior percentagem de consumo durante os dois períodos da manhã.

5.2.7. Fãs de Informação

Os Fãs de Informação veem, em média, quatro horas e meia de televisão por dia, sendo que mais de metade deste consumo é de conteúdos de Informação. Este consumo é feito entre os períodos da tarde e prime-time, inclusive. Relativamente ao tipo de consumo, este grupo apresenta uma percentagem de consumo não linear apenas ligeiramente inferior à média do parque. Este valor pode justificar-se pela grande recorrência à função de *Restart* em conteúdos de informação. As duas categorias que, juntamente com Informação, correspondem a cerca de dois terços do consumo são as categorias de Desporto e Filmes.

5.2.8. Adeptos de Desporto

Finalmente, os Adeptos de Desporto consomem, em média, mais de duas horas por dia de conteúdos da categoria de Desporto. O tempo restante, num total de quatro horas diárias, é ocupado, maioritariamente, com o consumo de conteúdos de Informação e Filmes. Este grupo de clientes privilegia o consumo linear, habitual em conteúdo desportivo, apresentando esta variável um valor significativamente inferior à média do parque total de clientes. Relativamente aos horários de consumo, os Adeptos de Desporto veem televisão quase exclusivamente nos períodos da tarde, pré prime-time e prime-time.

Com a caracterização do comportamento de consumo televisivo do parque de clientes, surge um universo de possibilidades para as ações de marketing. Estas ações podem agora adaptar-se ao cliente, tornando-as mais eficazes e menos dispendiosas.

1	FAMÍLIAS DIGITAIS - LIGHT	FAMÍLIAS DIGITAIS - NORMAIS
	• plataforma != 'BOX' OU dispositivos ≥ 6 • horas/dia < 3	• plataforma != 'BOX' OU dispositivos ≥ 6 • horas/dia ≥ 3
2	ADEPTOS DE DESPORTO	
	• desporto $\geq 37,5\%$	
3	ESPECTADORES EM DIRETO - GENERALISTAS	ESPECTADORES EM DIRETO - ENTRETENIMENTO
	• linear $\geq 99\%$ • dispositivos = 1 • cartões móveis ≤ 2 • informação $\geq$ filmes	• linear $\geq 99\%$ • dispositivos = 1 • cartões móveis ≤ 2 • informação $<$ filmes
4	FAMÍLIAS COM CRIANÇAS	
	• infantil $\geq 30\%$	
5	CINÉFILOS - SÉRIES	CINÉFILOS - FILMES
	• séries $\geq 20\%$	• filmes $\geq 32,5\%$
6	CLIENTES COR-DE-ROSA - NOVELAS	CLIENTES COR-DE-ROSA - REALITY SHOWS
	• novelas $\geq$ informação	• reality shows $\geq$ informação
7	FÃS DE INFORMAÇÃO	
	• informação $\geq 42,5\%$	
8	FAMÍLIAS TRADICIONAIS - HEAVY USERS	
	• linear $< 60\%$ OU horas/dia $\geq 4,75$	
9	FAMÍLIAS TRADICIONAIS - LATENIGHT	FAMÍLIAS TRADICIONAIS - PRÉ PRIME
	• latenight $>$ restantes horários	• pré prime $>$ restantes horários
	FAMÍLIAS TRADICIONAIS - PRIME	FAMÍLIAS TRADICIONAIS - TARDE
	• prime $>$ restantes horários	• tarde $>$ restantes horários

Figura 18: Critérios de distribuição dos clientes pelos clusters

Capítulo 6

Conclusões e Trabalho Futuro

O principal objetivo deste projeto foi delinear e desenvolver um conjunto de ferramentas de Business Intelligence que possibilitem analisar e caracterizar o consumo de conteúdos de televisão de um conjunto de clientes. A importância deste projeto não se limita à dissertação onde se insere, trazendo também vários benefícios para a empresa no seio da qual foi desenvolvido, a NOS.

Bem estabelecida como uma das maiores empresas de telecomunicações em Portugal, a NOS deve grande parte do seu valor aos processos e às estratégias de Business Intelligence que tem implementados e que oferecem informação às mais diversas áreas, ao nível do parque de clientes, receitas e custos proveniente dos vários serviços e produtos, eficácia de ações de marketing, entre outros. Estes processos começam com a recolha, registo e armazenamento de dados, pelas equipas de IT, que alimentam constantemente inúmeras análises de negócio, na forma de relatórios e *dashboards*. Uma das equipas responsáveis pela construção destes documentos é a equipa CRM-RAN, que prepara os dados em bruto e fornece informações claras e precisas a outras equipas, apoiando os seus processos de tomada de decisão.

Uma das áreas com uma grande disponibilidade de dados em bruto é a área de audimetria. No entanto, antes da elaboração deste projeto, não existia qualquer ferramenta que permitisse uma extração simples e clara de informação, quantitativa e qualitativa, sem ser necessário recorrer a intermediários com conhecimento da estrutura e capacidade de manipulação das bases de dados. Esta foi uma das necessidades atacadas pelo presente projeto. Além disso, a falta de tratamento deste universo de dados dificultava o conhecimento e a compreensão do comportamento dos clientes

deste serviço, o serviço de televisão, sendo esta a segunda oportunidade abordada neste projeto. Em ambos os sentidos, o projeto mostrou-se um sucesso.

A primeira oportunidade foi colmatada através da elaboração de um *dashboard*, dando aos seus utilizadores a capacidade de consultar informação de forma independente e com o nível de detalhe desejado. Para tal, foi essencial a identificação dos principais utilizadores do *dashboard* e a definição eficaz dos seus requisitos, bem como a construção de uma estrutura que garantisse flexibilidade, permitindo diferentes pontos e profundidades de análise. Assim, para além de disponibilizar informação para análises de negócio pontuais, o *dashboard* possibilita uma avaliação contínua das métricas de consumo de conteúdos de televisão dos clientes.

Relativamente à segunda oportunidade identificada, depois de identificados os atributos relevantes para a caracterização do comportamento de consumo de televisão dos clientes, foi aplicado um algoritmo de *clustering*, definindo-se, assim, os grupos de clientes com as características mais semelhantes entre si e que mais diferem relativamente aos restantes grupos. Criaram-se, então, 8 grupos: Famílias Tradicionais, Famílias Digitais, Espectadores em Direto, Cinéfilos, Clientes Cor-de-rosa, Famílias com Crianças, Fãs de Informação e Adeptos de Desporto. Estes grupos caracterizam o comportamento do parque de clientes do serviço de televisão da NOS, permitindo um conhecimento mais aprofundado do mesmo. Inúmeras vantagens podem surgir desta segmentação dos clientes, como ações de marketing direcionadas a grupos de clientes específicos, ajustes nas características do serviço disponibilizado e desenvolvimento de planos específicos de atuação.

Para além da manutenção e atualização das ferramentas, tarefa a cargo da equipa CRM-RAN, existem melhorias que podem ser implementadas em ambas. Para o *dashboard*, uma das propostas é a integração da caracterização dos clientes no mesmo, como parâmetro de filtro, de forma a reunir métricas e a identificar padrões específicos de comportamento dentro de cada grupo. Relativamente à ferramenta de segmentação dos clientes TV, um fator que aumentaria significativamente o conhecimento extraído dos dados é a introdução de variáveis sociodemográficas. No entanto, isto requer a implementação de outros processos, nomeadamente ao nível da recolha deste tipo de dados, tendo em conta o Regulamento Geral sobre a Proteção de Dados.

Finalmente, é importante avaliar o impacto na empresa das duas ferramentas implementadas, o *dashboard* e o *clustering* de clientes. Uma sugestão passa por analisar os KPI das operações e das áreas de negócio da NOS afetadas pelos resultados do presente projeto, antes e depois da elaboração do mesmo. Desta forma, os será possível quantificar as suas vantagens.

Bibliografia

- ANACOM. (2021). Pacotes de serviços de comunicações eletrônicas - 2020. [online] <https://www.anacom.pt/render.jsp?contentId=1603924>
- ANACOM. (2021). Serviço de distribuição de sinais de televisão por subscrição - ano de 2020. [online] <<https://www.anacom.pt/render.jsp?contentId=1603684>>
- Aruldoss, M., Travis, M. L., & Prasanna Venkatesan, V. (2014). A survey on recent research in business intelligence. *Journal of Enterprise Information Management*, 27(6), 831–866. <https://doi.org/10.1108/JEIM-06-2013-0029>
- Ashraf, S., & Khan, S. A. (2015). Visualizations-based analysis of Telco data for business intelligence. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, 2015-Novem*, 242–246. <https://doi.org/10.1109/ICSESS.2015.7339046>
- Du, X., Liu, B., & Zhang, J. (2019). Application of Business Intelligence Based on Big Data in E-commerce Data Analysis. *Journal of Physics: Conference Series*, 1395(1). <https://doi.org/10.1088/1742-6596/1395/1/012011>
- Few, S. (2006). Clarifying the Vision. *Information Dashboard Design The Effective Visual Communication of Data*, 223.
- Han, J. & Kamber, M. (2006). Data Mining: Concepts and Techniques, volume 54.
- Hu, M., & Cleland, S. (2019). A Pilot Study of Developing Introductory Course in Data Analytics and Business Intelligence. *Proceedings - Frontiers in Education Conference, FIE, 2019-Octob*. <https://doi.org/10.1109/FIE43999.2019.9028649>
- Hyde, P., Landry, E., & Tipping, A. (2004). Making the Perfect Marketer. *Booz Allen Hamilton*. <http://www.strategy-business.com/press/16635507/04405>

- Ishaya, T., & Folarin, M. (2012). A service oriented approach to Business Intelligence in Telecoms industry. *Telematics and Informatics*, 29(3), 273–285. <https://doi.org/10.1016/j.tele.2012.01.004>
- Jumadi Dehotman Sitompul, B., Salim Sitompul, O., & Sihombing, P. (2019). Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *Journal of Physics: Conference Series*, 1235(1). <https://doi.org/10.1088/1742-6596/1235/1/012015>
- Kastouni, M. Z., & Ait Lahcen, A. (2020). Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University - Computer and Information Sciences*, xxxx. <https://doi.org/10.1016/j.jksuci.2020.11.024>
- Kirk, A. (2021). A Recipe of Capabilities for Pursuing Expertise in Data Visualization: A Practitioner's Perspective. *IEEE Computer Graphics and Applications*, 41(1), 58–62. <https://doi.org/10.1109/MCG.2020.3034737>
- LaPointe, P. (2005). Marketing by the Dashboard Light. *Marketing NPV/Association of National Advertisers*
- Lousa, A., Pedrosa, I., & Bernardino, J. (2019). Avaliação e Análise de Ferramentas Business Intelligence para Visualização de Dados Evaluation and Analysis of Business Intelligence Data Visualization Tools. *Ieeexplore.Ieee.Org*, June, 19–22. <https://ieeexplore.ieee.org/abstract/document/8760677/>
- Negash, S. (2004). *Business Intelligence*. 13(February). <https://doi.org/10.17705/1CAIS.01315>
- Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a service: Why, what, how, and what research is needed? *Journal of Service Research*, 12(2), 175–189. <https://doi.org/10.1177/1094670509344213>
- Power, D. J. (2013). Mobile decision support and business intelligence: An overview. *Journal of Decision Systems*, 22(1), 4–9. <https://doi.org/10.1080/12460125.2012.760267>
- Rus, V. R., Toader, V., (2008). Business intelligence for hotels' management performance. *International Journal of Business Research*. http://findarticles.com/p/articles/mi_6773/is_4_8/ai_n31136506
- Xu, D., & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2(2), 165–193. <https://doi.org/10.1007/s40745-015-0040-1>

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678. <https://doi.org/10.1109/TNN.2005.845141>

Anexo A

Páginas da versão final do *dashboard*



Figura 19: Glossário da versão final do dashboard

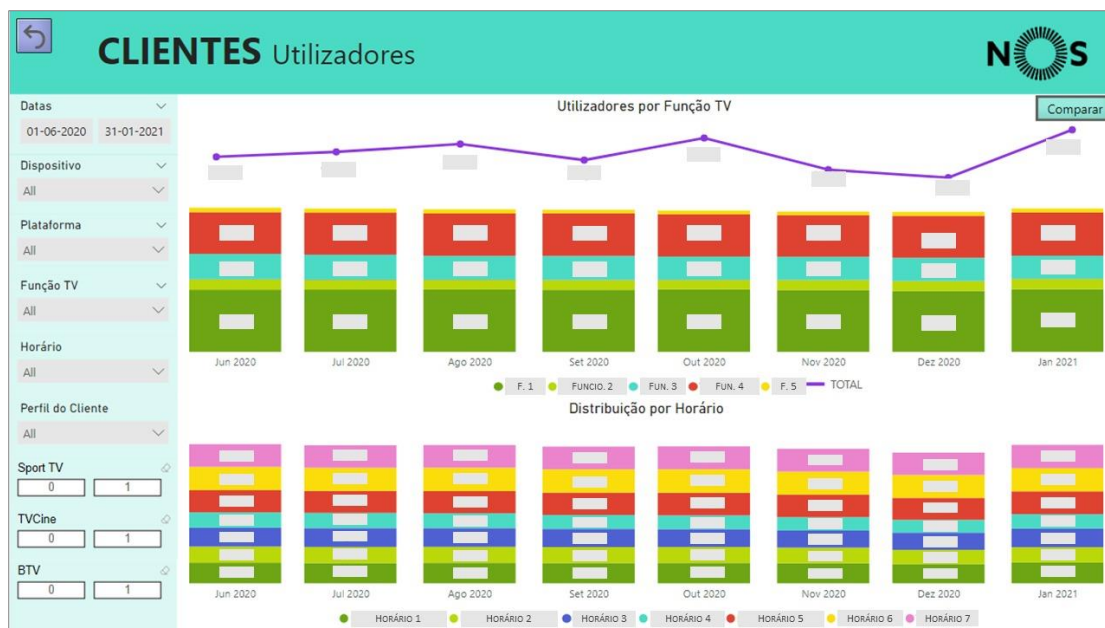


Figura 20: Página 1.1.a da versão final do dashboard

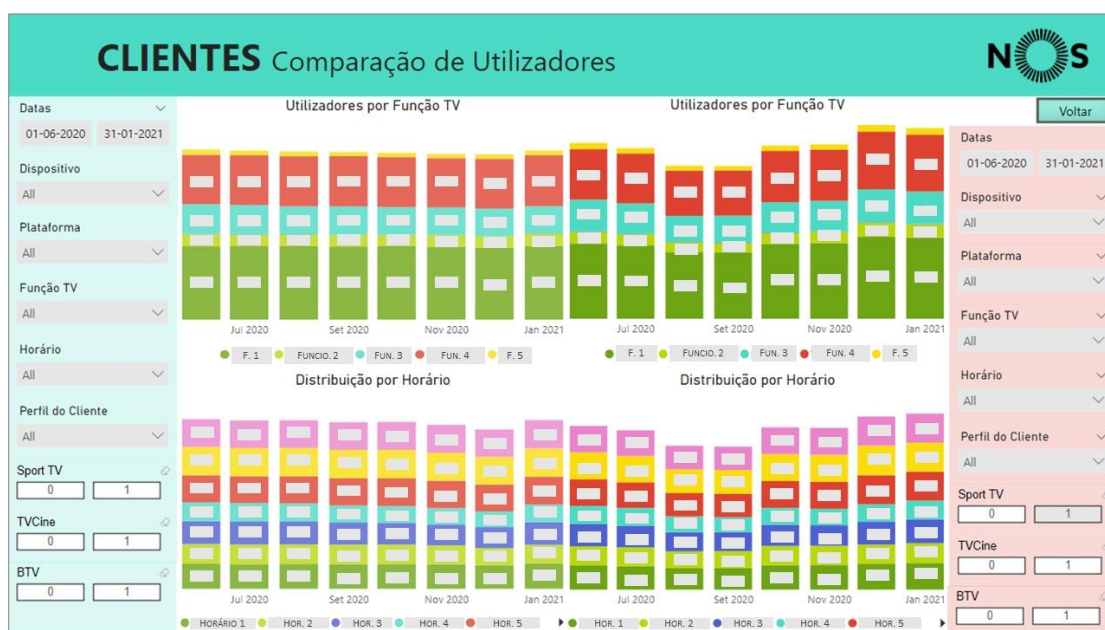


Figura 21: Página 1.1.b da versão final do dashboard



Figura 22: Página 1.2.a da versão final do dashboard

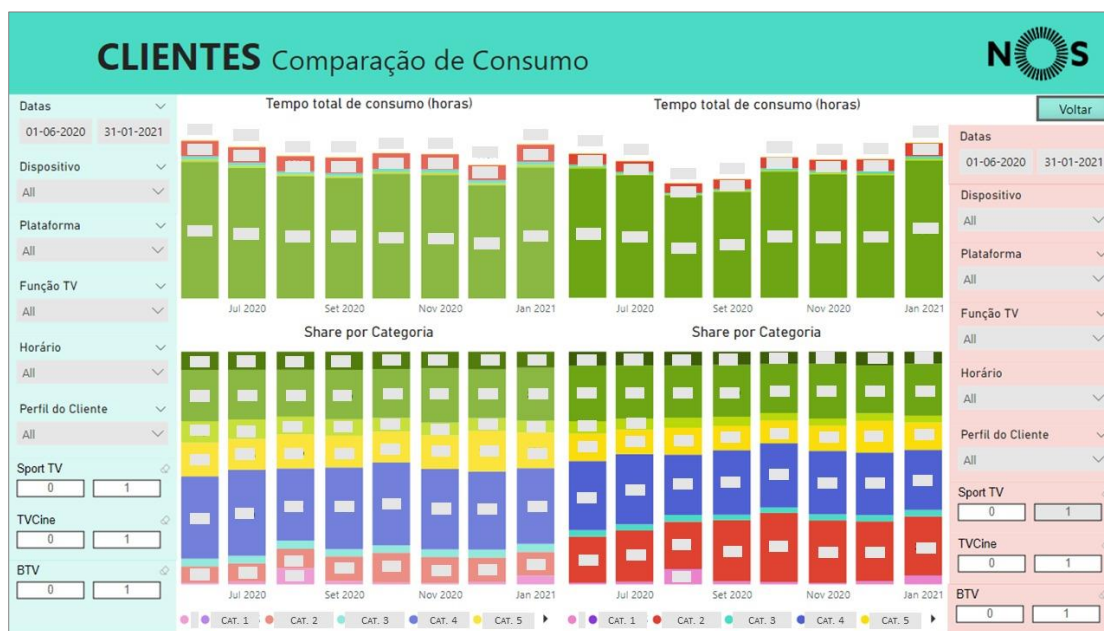


Figura 23: Página 1.2.b da versão final do dashboard

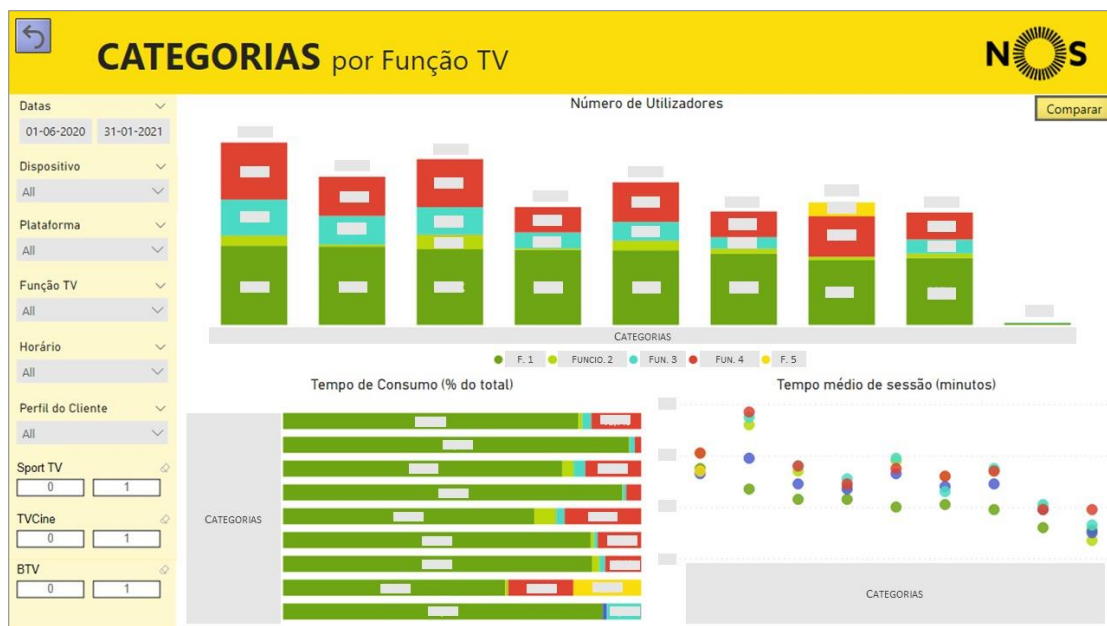


Figura 24: Página 2.1.a da versão final do dashboard

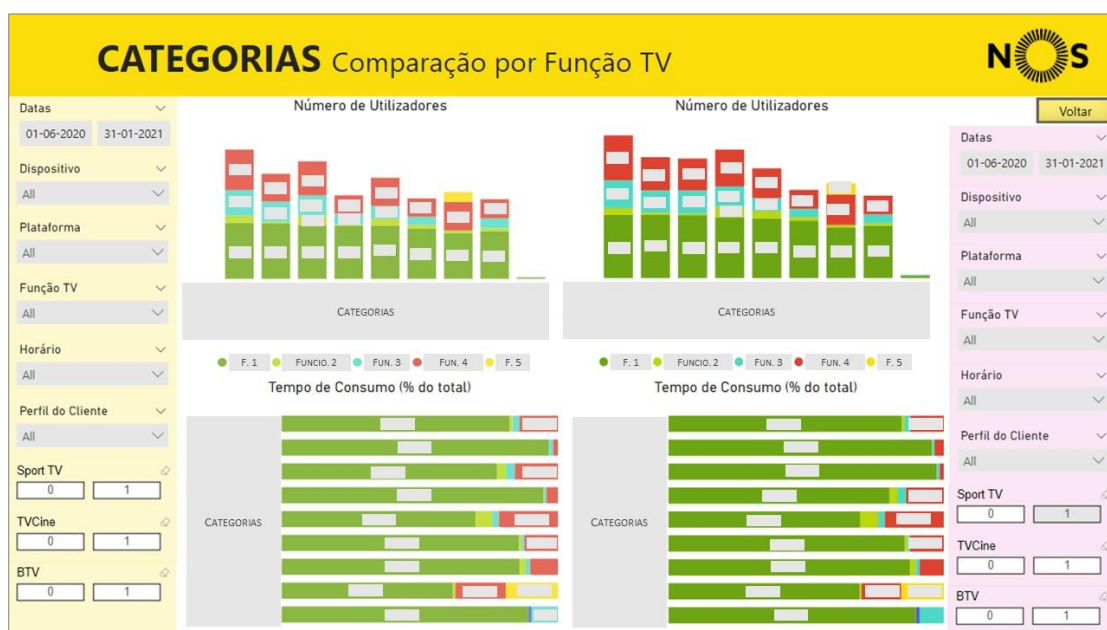


Figura 25: Página 2.1.b da versão final do dashboard



Figura 26: Página 2.2.a da versão final do dashboard

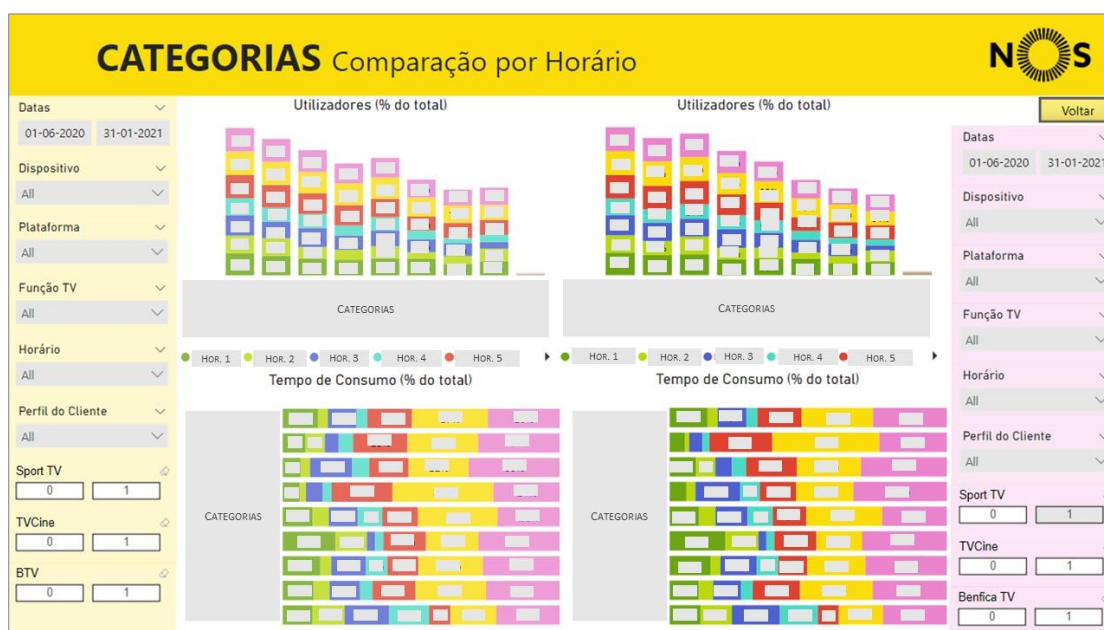


Figura 27: Página 2.2.b da versão final do dashboard



Figura 28: Página 3.1 da versão final do dashboard

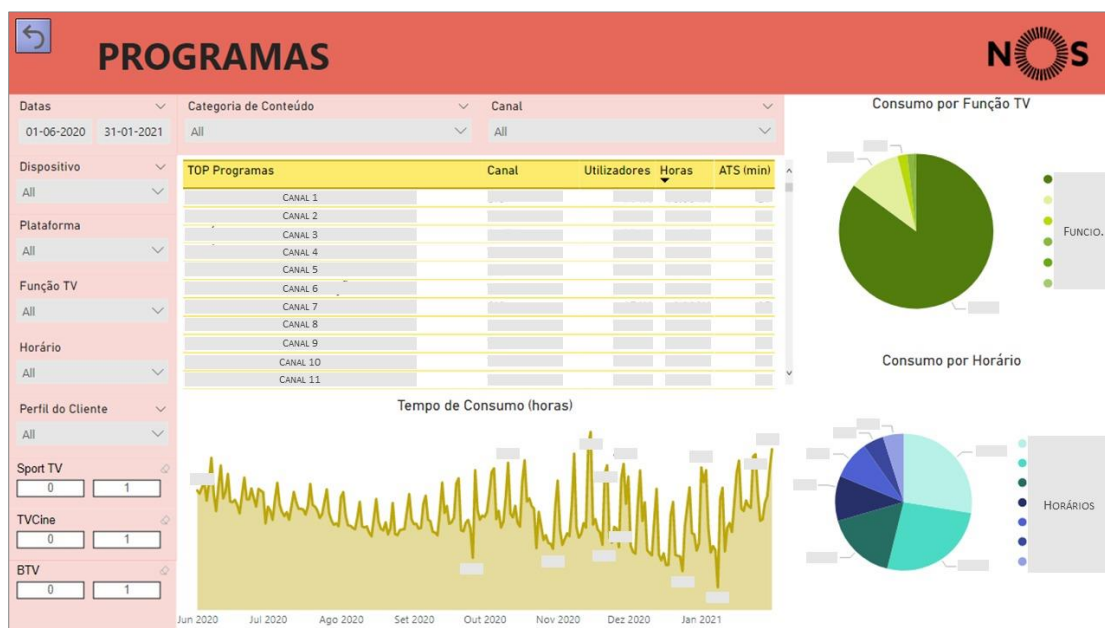


Figura 29: Página 4.1 da versão final do dashboard

