



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO

GESTÃO DE SISTEMAS DE INFORMAÇÃO

TRABALHO FINAL DE MESTRADO

RELATÓRIO DE ESTÁGIO

**PROJETO DE DATA MINING E MODELAÇÃO PREDITIVA
NA TAP PORTUGAL**

RAFAEL MENDONÇA VALÉRIO

SETEMBRO - 2014



Instituto Superior de Economia e Gestão

UNIVERSIDADE TÉCNICA DE LISBOA

DESDE 1911

MESTRADO EM GESTÃO DE SISTEMAS DE INFORMAÇÃO

TRABALHO FINAL DE MESTRADO RELATÓRIO DE ESTÁGIO

PROJETO DE DATA MINING E MODELAÇÃO PREDITIVA
NA TAP PORTUGAL

RAFAEL MENDONÇA VALÉRIO

ORIENTAÇÃO:

PROFESSOR ENGENHEIRO EDUARDO JORGE DIAS
RODRIGUES

SETEMBRO – 2014

Agradecimentos

Agradeço ao meu orientador pela gestão deste trabalho e pela oportunidade de realizar o estágio na Megasis/TAP.

À Paula Margaça e toda a equipa de BI & CRM da Megasis pela cooperação e atmosfera agradável de trabalho. Em especial ao Jorge Câncio pelo auxílio e ensinamentos durante todas as fases do projeto.

Ao Filipe Brandão pelo companheirismo e amizade que desenvolvemos no período de estágio.

Ao Guilherme, ao “David” e ao “Rogério” pelos serões de descompressão.

À Sara pela paciência e dedicação, hoje e no futuro.

Por fim, à família. Particularmente aos meus pais, cuja presença e apoio incondicional ao longo de toda a minha vida foram fundamentais para o meu sucesso académico e pessoal.

Lista de Abreviaturas

BI – *Business Intelligence*

CRM – *Customer Relationship Management*

DM – *Data Mining*

DQ – *Data Quality*

DW – *Data Warehouse*

ERP – *Enterprise Resource Planning*

FBTD – *Fare Basis Ticket Designator*

OLAP – *Online Analytical Processing*

PGA – *Portugália Airlines*

RBD – *Reservation Booking Designator*

SGPS – *Sociedade Gestora de Participações Sociais*

SI – *Sistemas de Informação*

SQL – *Structured Query Language*

TAP – *Transportes Aéreos Portugueses*

TI – *Tecnologias de Informação*

Resumo

Este relatório de estágio dá a conhecer a importância que as ferramentas de análise preditiva como o *data mining* e a modelação preditiva têm para o sistema de *Business Intelligence* de uma organização. No caso concreto, tem-se como objeto de trabalho o sistema de gestão de receitas de uma companhia aérea portuguesa que nele identificou uma oportunidade de otimização. O projeto realizado teve como foco a implementação de uma componente de *data mining* que permitisse antecipar a receita relativa a bilhetes comprados mas não utilizados, geradora de valores significativos para a companhia. Esta abordagem tem a finalidade de suportar a tomada de decisão em várias vertentes do negócio, como custos operacionais, calendarização de voos, ações de marketing, comunicação, de entre outros. Identificados alguns problemas ao nível da estruturação dos dados analisados, tecem-se considerações acerca do tratamento destes.

Palavras-chave: Análise de dados, *Business Intelligence*, *Data mining*, Modelação preditiva, Receitas de bilhetes de avião, Suporte à decisão

Abstract

This internship report was written in order to underline the importance of predictive analysis tools like data mining and predictive modelling to a Business Intelligence system in an organization. The object of this particular study is a Portuguese airline's revenue management system for which an optimization opportunity was identified: the implementation of a data mining component allowing to predict revenue regarding purchased, yet unused, plane tickets, which represents significant financial value to the airline. This approach aims to support decision making along several business areas, such as operational costs, flight scheduling, marketing practices and communication, among others. Since structural data problems were acknowledged during the project, some observations regarding data processing arise.

Keywords: Data analysis, Business Intelligence, Data mining, Predictive modelling, Airplane ticket sales revenue, Decision support

Índice

Agradecimentos	I
Lista de Abreviaturas	II
Resumo	III
Abstract.....	IV
Índice de Figuras	VI
Introdução	1
Objetivos e organização do trabalho.....	2
1. Revisão da literatura	3
1.1. <i>Business Intelligence</i>	3
1.2 Análise preditiva – <i>data mining</i> e modelos preditivos	5
1.3. A importância de BI e da análise preditiva para a indústria da aviação	7
2. Caracterização do estágio	10
3. Caracterização das empresas	11
3.1. A TAP Portugal	11
3.2. A Megasis	13
4. Projeto desenvolvido durante o estágio	15
4.1. SPSS Modeler	18
4.2. Resultados	26
5. Conclusões e sugestões de aplicação futura	27
5.1. Desenvolvimento pessoal e profissional.....	27
5.2. Problemas e limitações	28
5.3. Sugestões para o futuro.....	29
Glossário.....	32
Referências bibliográficas	35

Índice de Figuras

Figura 1 - Estrutura Acionista da TAP, SGPS, S.A.	2
Figura 2 – Exemplo de uma <i>stream</i> em SPSS Modeler	19
Figura 3 – <i>Query</i> utilizada como fonte de dados no SPSS Modeler	20
Figura 4 – Menu <i>Type</i> em SPSS Modeler	21
Figura 5 – Opções de <i>Measurement</i> no menu <i>Type</i>	22
Figura 6 – Opções de <i>Role</i> no menu <i>Type</i>	22
Figura 7 – Menu do nó de um modelo preditivo em SPSS Modeler	23
Figura 8 – Menu do nó <i>Analysis</i> em SPSS Modeler	24
Figura 9 – Menu do nó <i>Table</i> em SPSS Modeler	25
Figura 10 – Exemplo de uma <i>stream</i> concluída em SPSS Modeler	26

Introdução

A indústria da qual as companhias aéreas fazem parte insere-se num ambiente turbulento e complexo. Ao contrário de outras indústrias, a aviação está sujeita a rápidas alterações no que diz respeito às expectativas dos clientes, movimentos da concorrência, desenvolvimentos do fornecedor, regulamentações governamentais e dinâmicas dos funcionários (Riwo-Abudho et al, 2013).

Por esse motivo, o processo de tomada de decisão, por parte dos gestores, deve ser rigoroso e bem planeado, o que tem levado à emergência de sistemas dedicados de *Business Intelligence*, que têm como objetivo aperfeiçoar a eficiência das operações organizacionais e apoiar os decisores na tentativa de atingir vantagens competitivas (Ranjan, 2009).

Para uma companhia aérea, a prática da otimização de receitas através do controlo da disponibilidade e preço de lugares num voo é, habitualmente, designada de gestão de receitas. Sistemas sofisticados de gestão de receitas estão já em utilização no seio de todas as grandes companhias aéreas, e são vistos por muitos com uma componente crítica na infraestrutura tecnológica de uma companhia, como referem Lawrence et al (2003).

O *data mining* e a sequente modelação preditiva podem ser processos relevantes para um sistema de gestão de receitas, permitindo prever, com antecedência, quanto uma determinada empresa poderá faturar, com base em informação histórica contida nas bases de dados. Deste modo, os gestores poderão estar sempre prevenidos contra emergências e tomar decisões com menor risco, o que é uma mais-valia significativa na conjuntura atual das companhias aéreas.

O presente trabalho centrar-se-á nas técnicas de *data mining* e modelação preditiva aplicadas à realidade de uma companhia aérea portuguesa – a TAP Portugal – cuja leitura permite agilizar e alicerçar a tomada de decisão.

Objetivos e organização do trabalho

O objetivo deste relatório é descrever um projeto realizado no seio da TAP Portugal com a finalidade de obter previsões realistas de receitas oriundas de bilhetes vendidos e, efetivamente, não utilizados, assentes em informação histórica fornecida pelas reservas de passageiros.

No capítulo 1 far-se-á a revisão da literatura que subjaz ao desenvolvimento de sistemas de informação sustentados por ferramentas de *Business Intelligence*, mais concretamente, análises preditivas, tendo como base processos de *data mining*.

O capítulo 2 será dedicado à caracterização do estágio realizado, expondo os objetivos a atingir com o projeto a desenvolver.

No capítulo 3, será feita uma breve caracterização das empresas nas quais o estágio teve lugar, visando explicar o que leva à necessidade de um projeto desta natureza.

O capítulo 4 descreve detalhadamente as ações tomadas ao longo do projeto, apresentando, igualmente, os resultados daí provenientes e a sua análise.

Por fim, no capítulo 5, exibem-se algumas conclusões, sugestões de aplicação futura, bem como as limitações e problemas encontrados durante a execução do projeto.

1. Revisão da literatura

1.1. *Business Intelligence*

A realidade socioeconómica com que as empresas se deparam nos dias de hoje tem obrigado a que estas procurem soluções para manterem os seus negócios ou se destacarem da concorrência. Uma das principais razões pelas quais as empresas tendem a falir ou a perder terreno no mercado está diretamente relacionada com a falta de informação no momento da tomada de decisão por parte dos gestores. Já em 2009, a empresa de consultoria *Gartner* previa que, até 2012, devido à falta de informação, processos e ferramentas, mais de 35% das 5000 maiores empresas do mundo iriam falhar regularmente nas tomadas de decisão relativas a alterações significativas no mercado em que atuam. Nesse sentido, a implementação de técnicas e ferramentas que facilitem a aquisição, processamento e análise de dados provenientes de diversas fontes – como clientes, fornecedores, concorrentes, etc. – tem vindo a emergir como base para a obtenção de conhecimento (Olszak et al, 2007), surgindo assim o *Business Intelligence*, prenunciado, pela *Gartner*, vir a corresponder a cerca de 40% do orçamento das unidades de negócio em 2012.

Zhang et al (2011) definem *Business Intelligence* – doravante BI – como o uso da tecnologia de *data warehouse* para armazenar e gerir dados operacionais e, através de diversas ferramentas de análise estatística e técnicas de *data mining*, analisar esses mesmos dados, de forma a providenciar uma variedade de relatórios analíticos que, por sua vez, podem oferecer informação relevante no processo de apoio à decisão.

De uma forma mais simples, pode descrever-se BI, à semelhança de Azvine et al (2005), como os atos de capturar, aceder, entender e analisar um dos ativos mais valiosos de uma empresa – dados em bruto – e transformá-los em informação acionável, de modo a melhorar a *performance* do negócio.

Tendo em conta os benefícios encontrados nos processos de BI, é de estranhar que grande parte das empresas ainda não os utilize de forma proativa. Um inquérito realizado a executivos das indústrias de finanças e de energia pela *SunGard Financial Systems* em 2014 concluiu que apenas 20% das empresas inquiridas utilizava técnicas proativas e de investigação ou mostrava interesse em *scorecards*, *dashboards* e técnicas analíticas. Do mesmo inquérito concluiu-se também que somente 13% dos inquiridos utilizava técnicas avançadas de BI, como análises preditivas e alertas.

Segundo Ranjan (2009), é imperativo que as organizações tenham um conhecimento profundo acerca dos seus cliente, concorrentes e parceiros de negócio, do ambiente económico em que estão inseridos e das operações internas, de forma a tomar decisões acertadas. A existência de um sistema de *Business Intelligence* no seio da organização permite maior eficiência na execução destas decisões, tendo, também, como benefícios a eliminação de dúvidas, o aumento da comunicação entre departamentos no que diz respeito à coordenação de atividades e o facto de permitir respostas rápidas a alterações relativas a finanças, preferências dos clientes e operações da cadeia logística, melhorando a *performance* geral da empresa.

Gangadharan et al (2004) defendem que, embora o termo seja descrito por muitos como o resultado de análises profundas a dados detalhados do negócio, incluindo bases de dados, aplicações e práticas analíticas, BI é tecnicamente muito mais vasto do que isso, podendo englobar gestão do conhecimento, um sistema integrado de gestão empresarial

(ERP), sistemas de suporte à decisão e *data mining*, sendo neste último campo que este relatório se centra.

1.2 Análise preditiva – *data mining* e modelos preditivos

Nyce (2007) descreve a análise preditiva como um termo abrangente que engloba uma variedade de técnicas estatísticas e analíticas utilizadas para desenvolver modelos que sejam capazes de prever eventos ou comportamentos futuros. Um dos componentes da análise preditiva é o *data mining*, que Nyce refere como sendo um processo de análise de dados de modo a identificar tendências, padrões ou relações entre os dados, podendo esta informação ser posteriormente utilizada para desenvolver modelos preditivos.

Zhang et al (2011) definem *data mining* como o uso de matemática, estatística, inteligência artificial e aprendizagem automática para extração de conhecimento implícito e desconhecido mas com potencial utilidade, de uma grande quantidade de dados incompletos, confusos e aleatórios.

Rygielski et al (2002) explicam, inclusivamente, que o termo é uma analogia do ato de mineração de ouro ou carvão¹, dizendo respeito à extração de “pepitas” de conhecimento “enterradas” nos *data warehouses* das organizações. Acrescentam ainda, que uma abordagem de *data mining* deve servir para complementar outras técnicas de análise de dados, como o uso de estatística, abordagens OLAP, folhas de cálculo e o simples acesso a dados na sua forma mais básica.

¹ O termo “*data mining*”, quando traduzido à letra para português, significa “mineração de dados”.

Uma boa utilização de processos de *data mining* pode levar uma organização a concretizar alguns objetivos, como os identificados por Chopra et al (2011):

- *Cross-selling* de produtos;
- Diferenciação de clientes leais e não leais;
- Direcionamento do *marketing* para se focar em potenciais clientes;
- Prevenção de erros como incumprimentos ou maus empréstimos;
- Aumento de retenção de clientes.

Fayyad et al (1996) consideram que *data mining* é um passo no processo de descoberta de conhecimento, que consiste na aplicação de algoritmos específicos, para extração de padrões dos dados. Estes algoritmos são um conjunto de cálculos e processos heurísticos que, quando aplicados sobre os dados, dão origem a modelos que podem tomar várias formas, como sugere a *Microsoft*, na sua página “*Data Mining Algorithms*” (2014), incluindo:

- Conjuntos de *clusters* que descrevem como os casos num conjunto de dados se relacionam entre eles;
- Árvores de decisão que preveem um resultado e descrevem como diferentes critérios afetam esse resultado;
- Modelos matemáticos com prognósticos de vendas;
- Conjuntos de regras que descrevem como produtos se agrupam numa transação e a probabilidade de que determinados produtos sejam adquiridos em conjunto.

A criação destes modelos é designada por modelação preditiva, processo pelo qual um modelo é criado ou escolhido para tentar prever, da melhor forma, a probabilidade de um determinado resultado ocorrer (Geisser, 1993).

Porém, é de salientar que, tal como refere Nyce (2007), a validade de um modelo preditivo depende da qualidade e da quantidade de dados disponíveis utilizados para o seu desenvolvimento. Ou seja, mesmo que exista uma significativa quantidade de dados históricos com que trabalhar, o modelo poderá sofrer devido à falta de qualidade desses mesmos dados. Por sua vez, o mesmo se aplica à ausência de quantidades adequadas de dados.

No fundo, tal como refere Elkan (2013), no que diz respeito à modelação preditiva, não é possível haver progresso sem um conjunto de dados de qualidade e tamanho adequados.

1.3. A importância de BI e da análise preditiva para a indústria da aviação

A indústria da aviação tem vindo a experienciar cada vez mais dificuldades, especialmente desde os acontecimentos de 11 de setembro de 2001 e, mais recentemente, devido aos efeitos da recessão da economia mundial, situação agravada com o constante aumento dos preços do combustível (Miranda, 2012; Boland et al, 2002). Juntando-se a esta conjuntura a emergência das companhias *low cost* em cada vez mais segmentos de mercado, atraindo novos clientes e absorvendo quota de mercado das companhias já estabelecidas (Boland et al, 2002), as últimas vêm-se na obrigação de apostar em soluções assentes em melhores serviços, passando por aí a necessidade de gestão da informação.

É nesse sentido que os DW têm vindo a emergir no seio das companhias aéreas. O caso de sucesso da Continental Airlines é um exemplo que sublinha a importância de um

data warehouse. “À companhia faltava a infraestrutura de dados da organização que permitisse aos colaboradores aceder rapidamente à informação necessária para obter *insights* relevantes sobre o negócio”, dizem Anderson-Lehman et al (2004, tradução do autor). Acrescentam que, “no entanto, a visão dos gestores de topo foi a de juntar os dados numa única fonte, com informação distribuída a toda a organização, para que os funcionários de todo e qualquer departamento pudessem conduzir as suas próprias análises de negócio, de modo a gerir melhor a empresa, tornando-a mais lucrativa”, surgindo assim a tecnologia de DW na companhia aérea.

A possibilidade de obtenção de informação nas bases de dados da companhia por via mais rápida e eficiente é de extrema importância, especialmente devido aos vários desafios com que uma grande companhia aérea se depara no seu dia-a-dia, como os que referem Riwo-Abudho et al (2013):

- **Inovação nos produtos do fornecedor** – a crescente necessidade dos fornecedores de aeronaves, como a *Boeing* e a *Airbus*, de inovar nos seus produtos – de forma a produzir aviões seguros e modernos –, obriga as companhias a grandes investimentos em termos de aquisições, renovação de sistemas, manutenção, seguros, treino e formação, etc.;
- **Queda de reputação** – a má publicidade que pode afetar uma companhia aérea, devido a um simples erro, pode ser o suficiente para arruinar a reputação dessa do dia para a noite, causando prejuízo na casa das dezenas de milhões de dólares. Nesse sentido, é de extrema importância que os gestores tenham a possibilidade de aceder às probabilidades de determinada emergência ocorrer, protegendo assim os interesses do cliente e da própria companhia;

- **Poder dos clientes** – com o incremento do uso da internet para aquisição de bilhetes, os clientes têm muito maior facilidade em simplesmente selecionar o preço mais acessível a partir das suas casas. Dizem Riwo-Abudho et al “Os clientes apenas permanecerão leais se virem razão para tal” (2013, tradução do autor).
- **Concorrência intensa** – a criação de alianças entre companhias, aliada à proteção contra falência², tende a originar diferentes grupos entre as várias companhias da indústria, o que, por seu turno, leva a um aumento da rivalidade, pois as companhias, ao aderir a alianças estratégicas, passam a ter acesso a possibilidades que individualmente não teriam, ganhando potencial para incidir sobre novos segmentos de mercado com maior facilidade, por exemplo;
- **Aumento dos custos** – sobretudo desde os acontecimentos do 11 de setembro, existe a necessidade de um investimento significativo em medidas de segurança, não só para garantir o bem-estar dos passageiros, mas também para assegurar que estes não optam pela utilização de outros meios de transporte em detrimento do aéreo. Também os custos de *jet fuel* têm vindo a aumentar todos os anos, bem como os custos de mão-de-obra.

Tendo em conta todos os problemas capazes de afetar uma companhia aérea a qualquer momento, torna-se claro que o investimento em ferramentas e processos que permitam antecipar resultados não é apenas uma vantagem, mas sim fundamental para o negócio. É da tentativa de realizar previsões rigorosas que permitam antecipar resultados que este relatório se ocupa, focando-se, essencialmente, na modelação preditiva para gestão de receitas.

² Um direito a que algumas corporações têm nos Estados Unidos da América.

2. Caracterização do estágio

Este estágio foi realizado no âmbito do trabalho final do Mestrado em Gestão de Sistemas de Informação e teve a duração de 3 meses, havendo decorrido nas instalações da Megasis no *campus* da TAP Portugal, sito no Aeroporto de Lisboa.

O estágio consistiu no desenvolvimento de um modelo preditivo, utilizando processos de *data mining*, tendo como objetivo dotar as áreas de negócio da TAP Portugal da capacidade de realizar uma previsão do apuramento da receita vendida e não voada, com base na informação histórica relativa às reservas de passageiros.

Os principais objetivos curriculares, para o estagiário, consistiram na familiarização com os processos de negócio da indústria da aviação e na compreensão e aprendizagem de processos e ferramentas de SQL, *data mining*, modelos de *forecast* e *Business Intelligence* no geral, no seio de uma empresa, passando por:

- Perceber a importância da estruturação adequada da informação em bases de dados;
- Entender comandos básicos na linguagem SQL, de forma a obter facilmente a informação necessária para o projeto em causa;
- Identificar e analisar a relevância que os diferentes indicadores têm para diferentes situações de negócio;
- Criar tabelas e *views* em bases de dados, de forma a obter a informação pretendida;
- Conhecer detalhadamente um software de produção de modelos preditivos, por forma a saber identificar variáveis e regras de associação adequadas ao objetivo do negócio e, consequentemente, identificar o(s) modelo(s) mais adequado(s) para a obtenção de previsões rigorosas.

3. Caracterização das empresas

3.1. A TAP Portugal

A TAP Portugal é uma companhia aérea portuguesa fundada em 1945 com sede no Aeroporto de Lisboa, controlada inteiramente pelo Estado de Portugal. É, desde 2005, membro integrante da Star Alliance, “a primeira e maior aliança de companhias aéreas do mundo” (Wikipedia, 2014).

Tem como missão garantir “o serviço de Transporte Aéreo e atividades afins, aspirando a ser sempre a melhor opção para quem utilizar os seus serviços e uma de entre as melhores empresas para se trabalhar, atuando consciente do seu compromisso para com a sociedade e com o meio ambiente” (TAP, 2012). Para tal, o Grupo TAP utilizava – à data de 31 de dezembro de 2013 – uma frota de 71 aeronaves, 39 delas destinadas a voos de médio curso, 16 a voos de longo curso e 16 à frota regional da PGA, tendo, em 2013, transportado 10,7 milhões de passageiros e 82,4 milhares de toneladas de carga e/ou correio e operado para 198 destinos, distribuídos da seguinte forma:

- 104 na Europa (10 deles em Portugal);
- 65 nas Américas;
- 21 em África;
- 8 no Médio Oriente e Ásia.

O volume total de negócios do Grupo, em 2013, correspondeu a cerca de 2,6 mil milhões de euros, representando um ligeiro aumento (1,9%) em relação ao ano anterior, ainda que algumas áreas, como a “Manutenção e Assistência a Terceiros em Portugal” e

“Outras Atividades” tenham tido uma queda significativa de 35,2% e 54,4%,
respetivamente.

Volume de Negócios			
EUR milhões			
	2013	2012	var. (%)
Transporte Aéreo	2.346,5	2.255,9	4,0
Manutenção Assistência a Terceiros Portugal	74,2	114,5	(35,2)
Manutenção Assistência a Terceiros Brasil	72,6	70,8	2,5
Lojas Francas	162,5	154,4	5,2
<i>Catering</i>	7,0	6,0	17,3
Outras Atividades da TAP, SGPS, S.A.	8,7	19,1	(54,4)
TOTAL	2.671,5	2.620,7	1,9

Tabela I – Volume de Negócios da TAP, SGPS, S.A.

Fonte: TAP – Relatório Anual 2013, p. 25

Do Grupo TAP fazem parte diversas subsidiárias, nas quais se inclui a Megasis, como mostra o esquema abaixo, no qual está representada a estrutura acionista do grupo, bem como a respetiva percentagem que a TAP detém de cada uma delas.

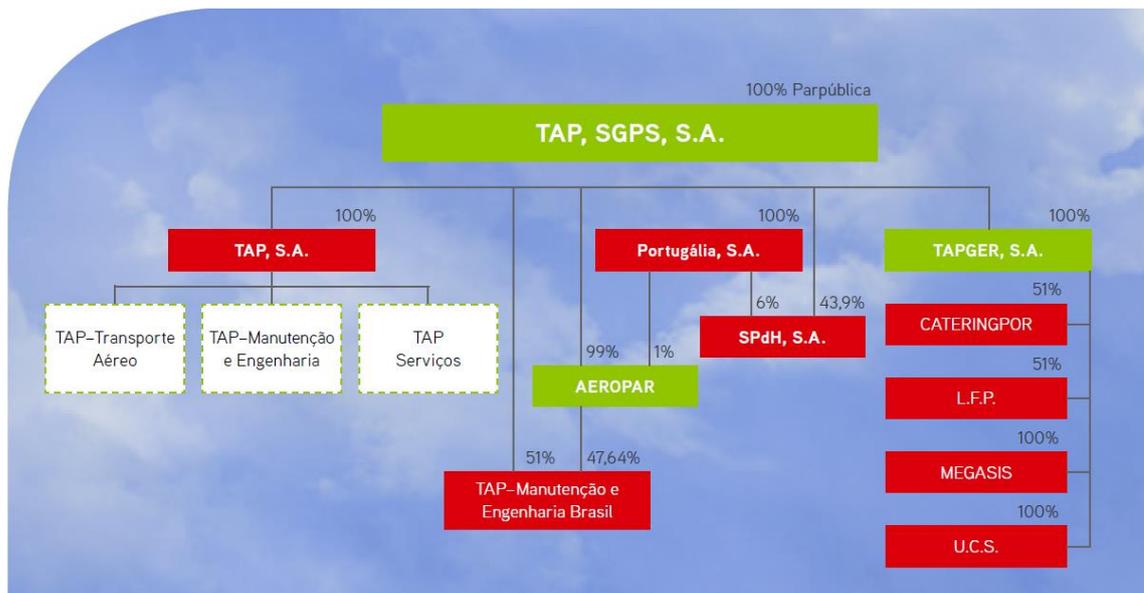


Figura 1 - Estrutura Acionista da TAP, SGPS, S.A.

Fonte: TAP – Relatório Anual 2013, p. 12

3.2. A Megasis

A Megasis é uma sociedade de serviços de engenharia informática constituída em 1989, na altura com capital repartido entre a TAP Portugal e outras empresas, tendo a companhia aérea adquirido a totalidade do seu capital social em 1994 e passando, desde então, a fazer parte do Grupo TAP.

As principais funções da Megasis para com o acionista e cliente, a TAP, consistem num conjunto de trabalhos desenvolvidos em diversas áreas de SI/TI, nomeadamente:

- Planeamento Estratégico de Sistemas de Informação;
- Gestão e Desenvolvimento de Projetos;

- Integração e Manutenção de Sistemas e Tecnologias;
- Suporte de Formação em Tecnologias de Informação.

3.2.1. A equipa de *Business Intelligence & Customer Relationship Management*

A equipa na qual o estagiário se integrou para propósitos de desenvolver o projeto pretendido foi a equipa de BI & CRM, departamento que tem a responsabilidade de consolidar e garantir a qualidade de informação dos diversos sistemas da TAP. Neste sentido, agrupa e constrói indicadores de áreas diversificadas, tais como as de finanças, vendas, operações de voo, operações de terra, manutenção, recursos humanos, entre outras, que visam dotar a TAP de um controlo efetivo das suas operações (atuais e passadas).

Esta equipa é ainda responsável pela construção de indicadores e relatórios para monitorização e análise de tendências, sendo que as informações daí decorrentes têm permitido aumentar a eficiência de algumas áreas de operação da TAP e, assim, reduzir efetivamente os custos decorrentes destas. Atente-se no exemplo da redução dos custos de *jet-fuel*, resultante das análises desenvolvidas pela equipa de BI & CRM.

Por fim, é importante referir que uma das mais importantes funções desta equipa passa por fornecer informação que apoie na tomada de decisões e na definição de estratégias e objetivos, surgindo, nesse âmbito, o projeto desenvolvido durante o estágio, tarefa que tinha como finalidade prover *insights* sobre as receitas de bilhetes não voados, melhorando, eventualmente, as operações da TAP.

4. Projeto desenvolvido durante o estágio

O estágio teve início a 16 de junho de 2014, prevendo-se desenvolver um projeto no prazo de 3 meses. Esse projeto – que viria a ter o constante acompanhamento de um membro da equipa de BI & CRM, Jorge Câncio – tal como referido anteriormente, objetivava dotar as áreas de negócio da capacidade de realizar uma previsão do apuramento da receita vendida e não voada, com base na informação histórica relativa às reservas de passageiros. Concretamente, através da criação de um modelo preditivo cujo algoritmo permitisse obter resultados realistas.

O resultado deste projeto permitiria à TAP identificar, com a devida antecedência, qual a receita que corresponde a bilhetes comprados por passageiros que, no prazo de 18 meses após a data prevista de voo: (1) não irão comparecer no voo previsto; (2) não irão efetuar a troca do bilhete; (3) não irão requisitar a sua devolução; (4) não irão efetuar outro voo posteriormente, utilizando esse bilhete (esta opção é rara, mas pode acontecer no caso de tarifas de preços mais elevados). Como exemplo, imagine-se um passageiro que compre um bilhete com data prevista de voo a 1 de janeiro de 2015. Nessa data, o passageiro não compareceu ao voo e nos 18 meses seguintes não lhe dá utilização³. Apenas a partir do dia 1 de julho de 2016 o preço pago por este bilhete passa a corresponder a receita para a TAP, dado que até esta data ele podia ser utilizado.

Assim, foi solicitada ao estagiário a construção de um modelo preditivo que, à data da compra do bilhete, permita saber se esse irá corresponder a receita voada – o que engloba voo, troca ou devolução – ou receita não voada, realidade que a equipa de BI & CRM designa por *purge*.

³ Por utilização entenda-se voo realizado, troca de bilhete ou devolução de bilhete.

A informação a fornecer por este projeto pode ter diversos objetivos, embora nenhum deles tenha sido explicitamente partilhado com o estagiário, dado que é um pedido que parte diretamente de um representante da TAP com o qual o estagiário nunca teve contacto ao longo do tempo que passou na empresa. No entanto, é sabido que a possibilidade de prever quais os bilhetes que resultarão em *purge*, poderá trazer benefícios ao nível de:

- Custos operacionais – tendo em conta que permite saber, com mais ou menos certeza, qual o número de passageiros que deverá efetuar um determinado voo;
- Calendarização de voos – ao perceber as combinações de datas e destinos que tendem a resultar em receita voada e não voada;
- *Marketing* – pois permite saber quais os voos e datas que mais facilmente podem ser alvo de promoções;
- Execução de *overbooking* – dado que existe um risco consideravelmente menor de sobrelotação de um voo, se houver uma previsão de que passageiros poderão não comparecer;
- Comunicação – ao permitir afirmar com mais rigor quanto a empresa irá gerar em receitas nesse ano ou no seguinte.

Inicialmente, para integração na equipa e entendimento de conceitos relevantes para a realização do projeto, o estagiário teve que atravessar uma fase de familiarização com os processos necessários e os *softwares* a utilizar. Aqui, apenas houve lugar a esclarecimento de dúvidas com o acompanhante do projeto e à leitura de documentação relacionada com o SPSS Modeler, a principal ferramenta a utilizar ao longo do estágio.

Para contextualização, o SPSS Modeler – da IBM – é uma ferramenta de *data mining* e outras tarefas analíticas, tendo sido desenhado com o intuito de desenvolver previsões,

de forma a auxiliar os indivíduos ou empresas na tomada de decisão, através da criação de modelos preditivos baseados num algoritmo de um vasto leque de opções. Por conseguinte, oferecendo ao utilizador sugestões automatizadas sobre os modelos ideais a utilizar dependendo da situação pretendida. Adiante, explicar-se-á, de forma mais detalhada, as funções desta ferramenta, por via de se entender o que se pretendeu atingir com a sua utilização.

Posteriormente, o estagiário teve a tarefa de criar *queries* simples em SQL, utilizando uma ferramenta de base de dados relacional – Microsoft SQL Server –, de modo a ganhar experiência na formulação dessas para, no futuro, saber criar tabelas com a informação pretendida para o projeto.

Após claro entendimento das ferramentas e processos necessários para o estágio, o estagiário teve a missão de seleccionar os indicadores que poderiam influenciar a utilização ou não de um bilhete. Mais tarde, alguns destes indicadores viriam a ser excluídos ou transformados. Contudo, numa primeira fase seleccionaram-se os seguintes:

- Data de emissão do bilhete;
- Agente emissor;
- País emissor;
- Data prevista de voo;
- Aeroporto de partida;
- Aeroporto de destino;
- Tarifa ou RBD;
- Preço do bilhete ou FBTD;
- Itinerário.

Adicionalmente, com estes indicadores, foi possível criar outros, personalizados, que poderiam vir a ter relevância para o projeto, nomeadamente:

- Diferença de dias entre datas de emissão e de voo;
- Mês de emissão;
- Mês de voo;
- Dia (da semana) de voo;
- País de partida;
- Setor⁴ de partida;
- País de destino;
- Setor de destino;

Após seleção das variáveis a utilizar, procedeu-se à criação de uma tabela com esta informação nas bases de dados da TAP, através das técnicas de SQL adquiridas ao longo dos primeiros dias do estágio, gerando-se, assim, condições para iniciar o trabalho com o SPSS Modeler.

4.1. SPSS Modeler

No primeiro contacto com a ferramenta, o estagiário e o acompanhante do projeto assistiram a um breve *webinar*, durante o qual foram transmitidos mais alguns *insights* acerca de como obter mais proveito do SPSS Modeler, particularmente algumas dicas práticas, não tão claras na documentação anteriormente lida pelo estagiário. Posto isto, procedeu-se à fase de experimentação da ferramenta, onde se colocou em prática a criação de *streams*.

⁴ Setores são: Portugal, Europa, África, Médio Oriente e Ásia, Oceânia e Américas.

As *streams* são a base do SPSS Modeler. *Stream* é o termo que, em inglês, corresponde a “corrente”, comumente associado às águas de um rio. Neste contexto, refere-se ao facto de a informação ser lida de acordo com a corrente criada pelo utilizador, sendo que qualquer ponto – ou nó – na corrente é dependente da informação presente nos nós anteriores. Para que o exposto se torne mais perceptível, tome-se como exemplo a figura abaixo na qual o nó B depende sempre da informação presente no nó A, o nó C depende do B e assim sucessivamente, sendo que qualquer alteração ao nó A irá automaticamente afetar os nós mais “abaixo” – facilmente identificados pela direção das setas – na corrente (neste caso, os nós B, C e D).

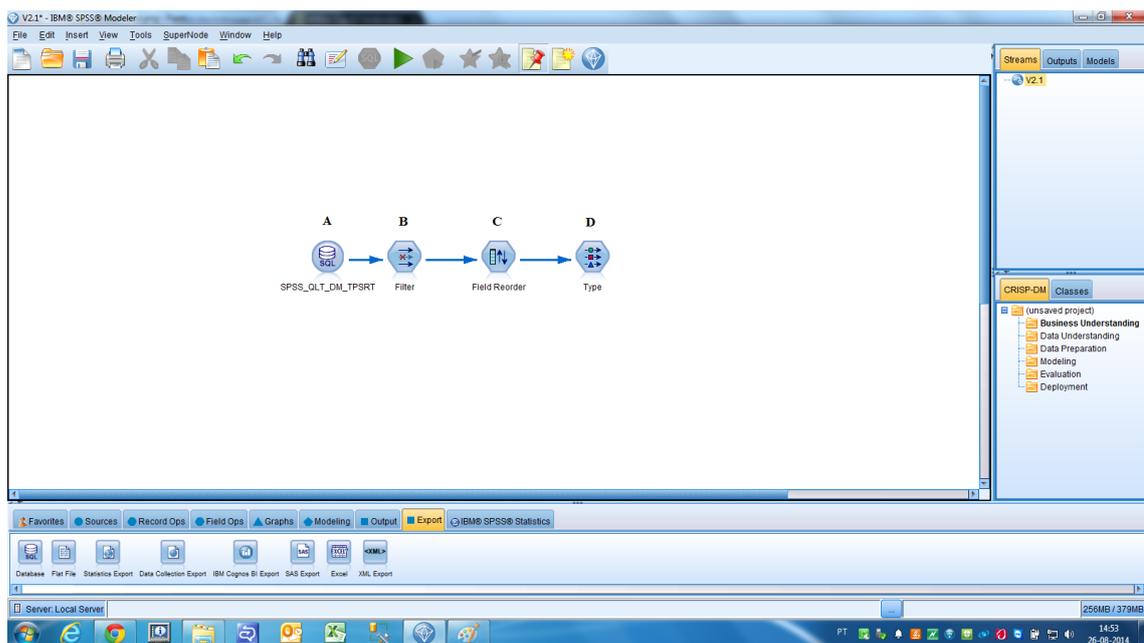


Figura 2 – Exemplo de uma *stream* em SPSS Modeler

De forma a construir uma *stream* que apresentasse os resultados pretendidos, começou-se por criar um nó *Database* de entre o grupo de nós *Source* – os nós que dizem

ao programa qual a fonte dos dados. O nó *Database* permite referenciar uma tabela ou uma *query* SQL e, dessa forma, dar os dados a conhecer ao SPSS Modeler. Neste caso, optou-se pela referência de uma *query*, tal como é visível na figura abaixo.

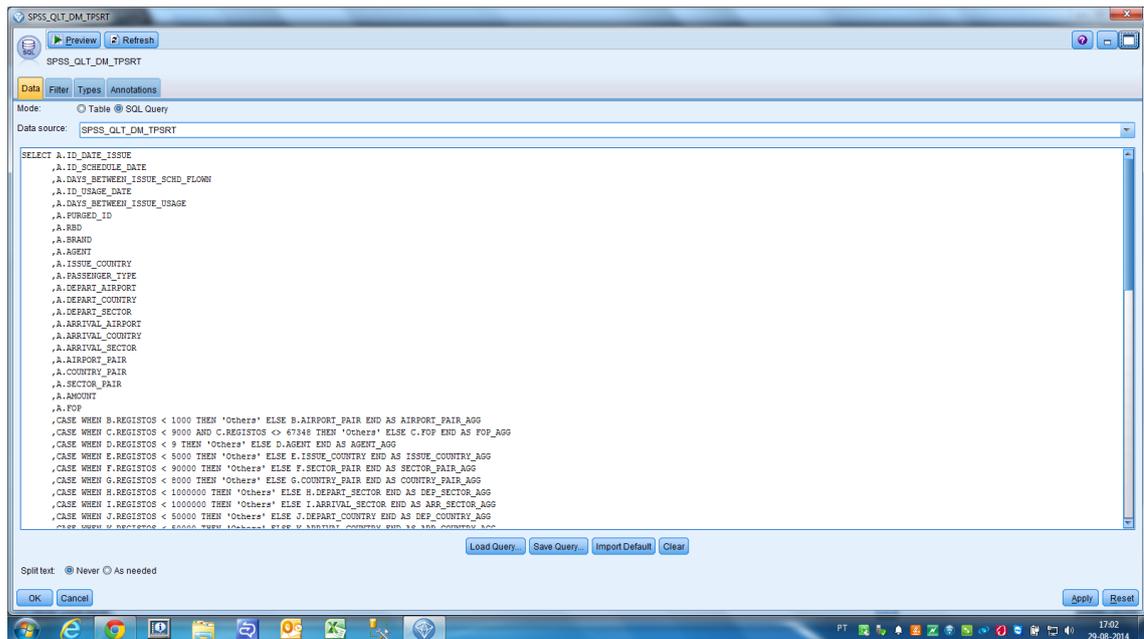


Figura 3 – Query utilizada como fonte de dados no SPSS Modeler

À frente, na *stream*, criou-se um nó *Filter*, que permite filtrar as colunas que se pretende que sejam visíveis nos nós seguintes. Deste modo, todas as colunas que viriam a tornar-se irrelevantes à medida que o modelo seria testado, seriam simplesmente colocadas de lado, consoante a necessidade.

Irrelevante para o projeto em si, mas de forma a facilitar a leitura da informação, seguiu-se um nó *Field Reorder*. Este nó tem a simples função de ordenar o posicionamento das variáveis, para que estas sejam visualizadas na ordem mais conveniente para o utilizador (e.g.: alfabética, por tipo, por data de adição, etc.).

O próximo nó presente na *stream* pretendida é talvez um dos mais importantes para qualquer *stream* na modelação preditiva. O nó *Type* permite especificar o tipo de medida das variáveis e o papel que cada uma delas terá para o modelo (ver figura 4). O tipo de medida afetará a forma como o Modeler vai interpretar a variável na fase de modelação. Tal como se pode verificar na figura 5, existem diversos tipos de medição – ou *Measurement*, no inglês do SPSS Modeler – de uma variável e o utilizador deve conhecê-la de forma a definir o tipo ideal, de acordo com o pretendido. O papel – ou *role* – diz respeito à função que a variável terá no processo de modelação. Dependendo do objetivo, há diversas opções de escolha (ver figura 6), mas para propósitos deste projeto, as únicas relevantes foram *Input* e *Target*, a primeira representando independência e a segunda representando dependência. Neste caso, pretendia-se assinalar todas as variáveis como *Inputs*, exceto a que nos diz se houve *purge* ou não, sendo essa o *Target*, ou seja, a variável que depende e é afetada pelo resultado das outras.

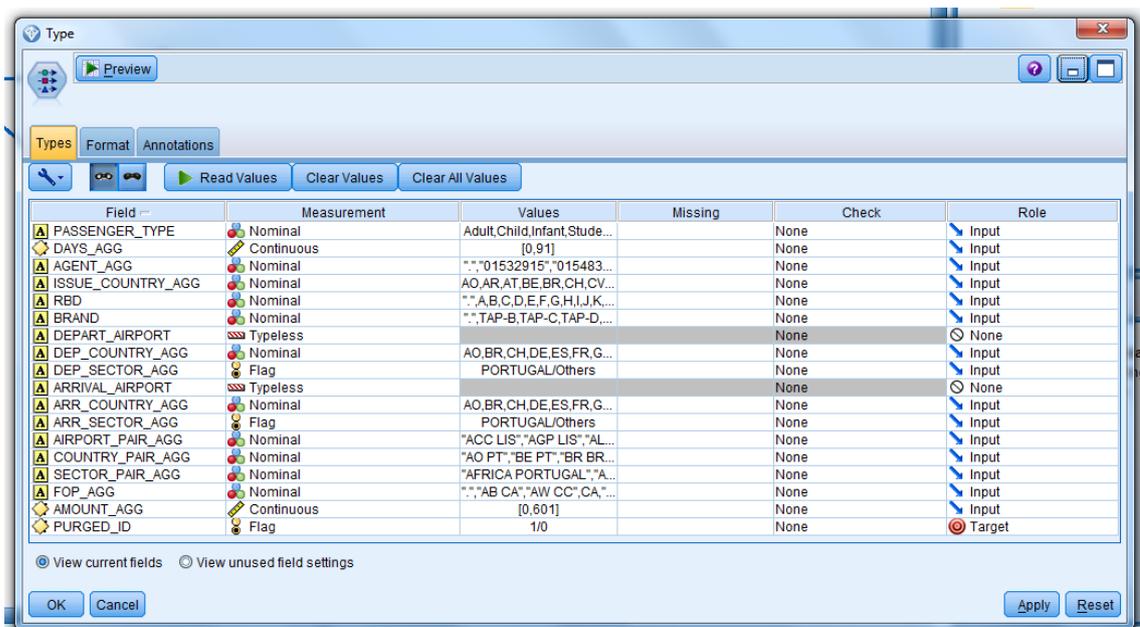


Figura 4 – Menu *Type* em SPSS Modeler

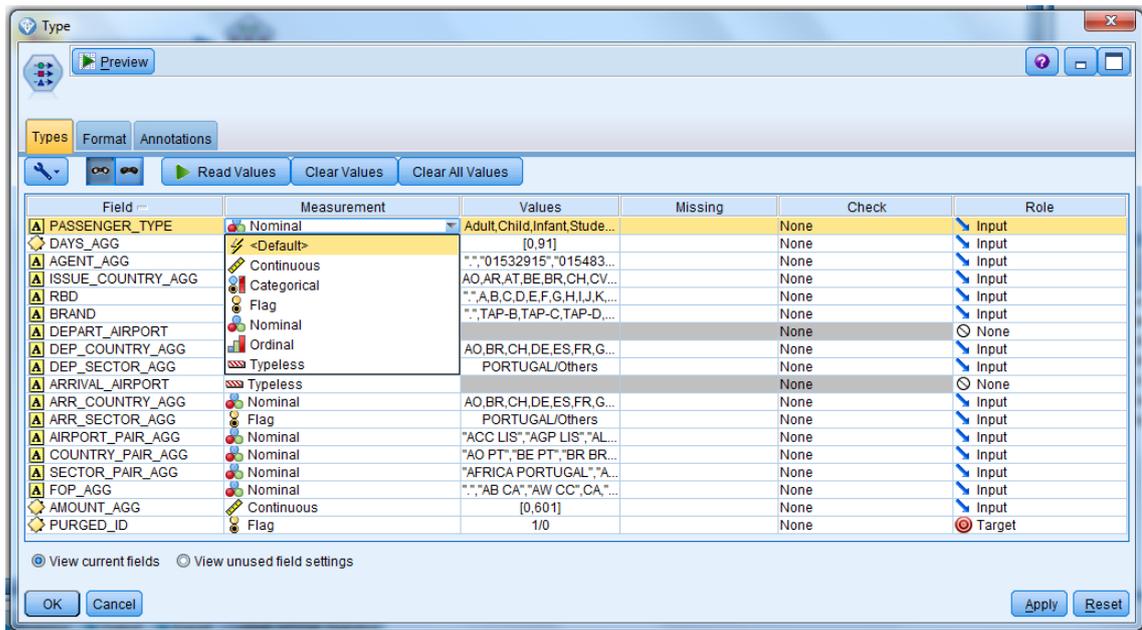


Figura 5 – Opções de *Measurement* no menu *Type*

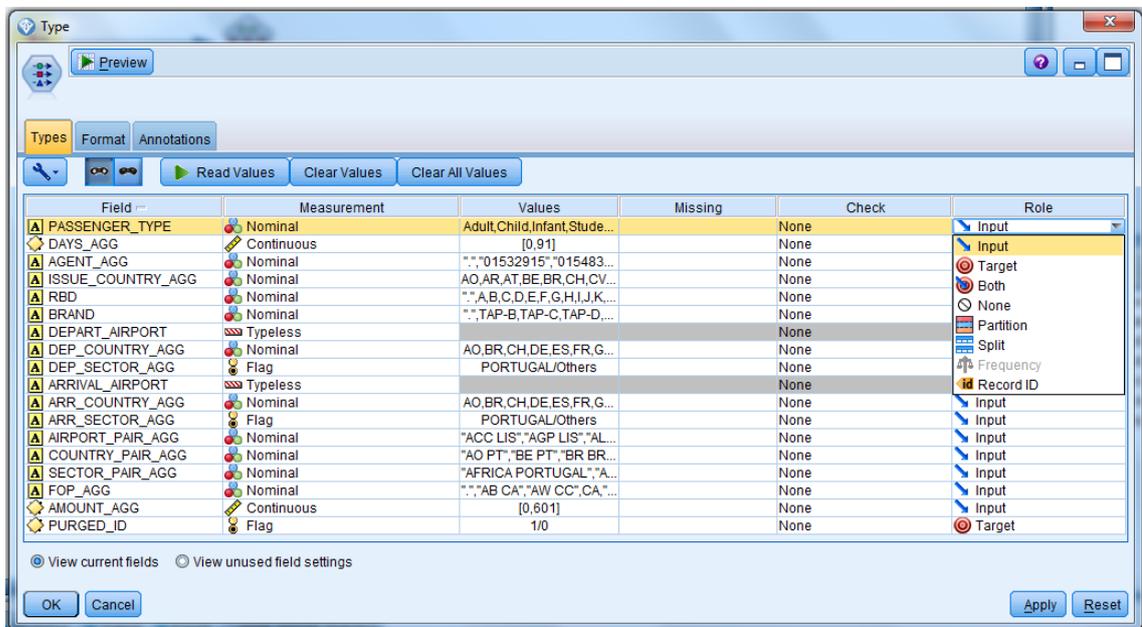


Figura 6 – Opções de *Role* no menu *Type*

Definidos, no menu *Type*, os parâmetros de cada variável, inicia-se a seleção do modelo e a sua construção. O Modeler oferece a possibilidade de utilização de diversos modelos, dando a conhecer ao utilizador, automaticamente, qual ou quais os melhores a utilizar para cada caso de entre inúmeras hipóteses.

Após escolher o modelo pretendido, adiciona-se um nó relativo a esse modelo, tendo este sido substituído várias vezes ao longo do projeto, na tentativa de encontrar o que desse mais garantias. Os menus dos nós de modelo permitem definir regras manualmente ou seguir as regras definidas anteriormente no nó *Type* (ver figura 7), sendo esta a opção de eleição em todos os casos.

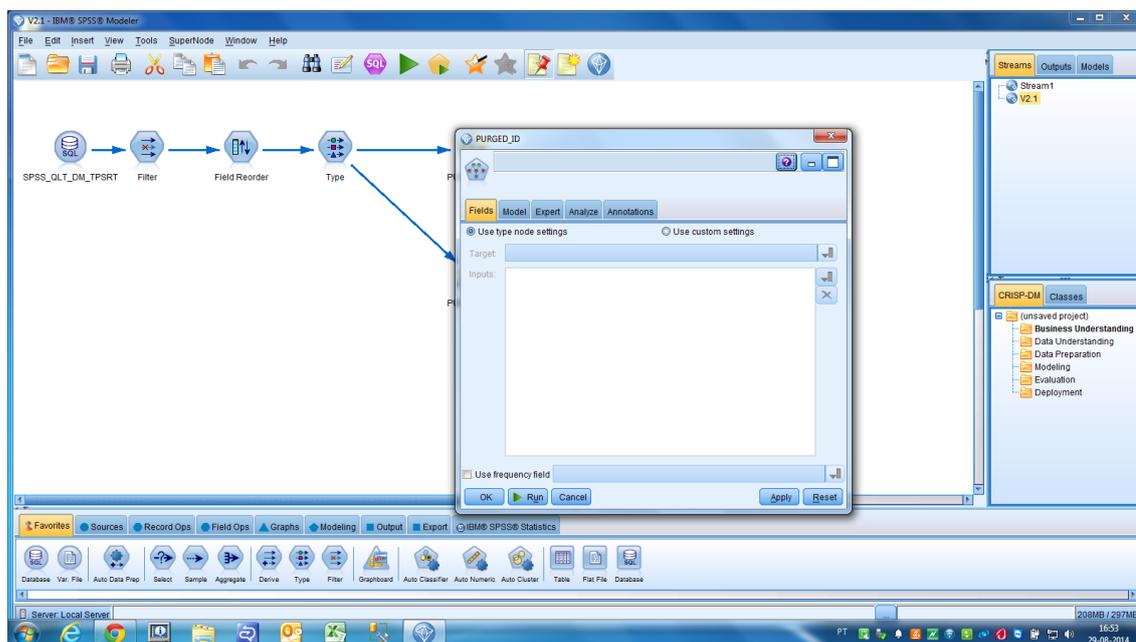


Figura 7 – Menu do nó de um modelo preditivo em SPSS Modeler

É nesta fase que o utilizador manda o programa executar o modelo (processo que, neste caso, podia demorar vários minutos, dependendo do modelo utilizado, devido aos milhões

de registos tidos em conta para a análise pretendida), gerando-se, posteriormente, um nó, em forma de diamante, com várias informações relativas ao modelo.

Após a execução do pedido, o programa cria e apresenta o modelo a utilizar para a previsão desejada. Nesta fase, o estagiário criava, habitualmente, dois novos nós, *Analysis* (ver figura 8) e *Table* (ver figura 9), sendo que o primeiro transmite informação relativa à percentagem de registos cujo valor acertou aquando da previsão de resultados com base no algoritmo do modelo; o segundo permite apresentar uma tabela com o resultado previsto para cada registo e a percentagem de certezas com a qual cada previsão é feita. Se o utilizador estiver satisfeito com o resultado, cria outro nó, de exportação, com o intuito de inserir a nova informação na base de dados inicial (ou numa nova, dependendo das necessidades).

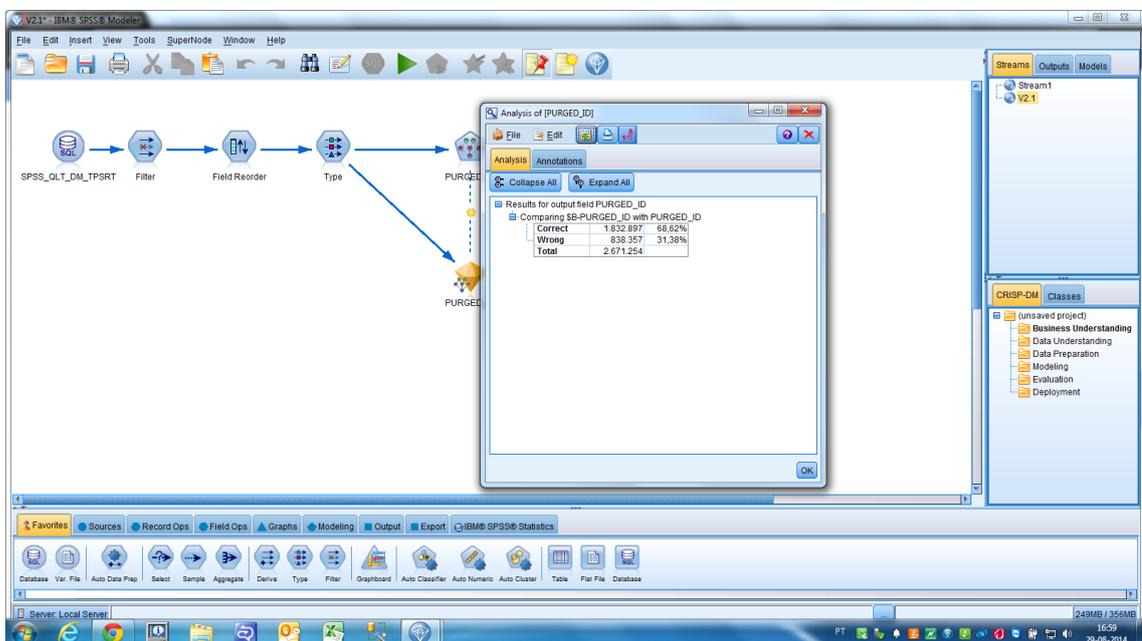


Figura 8 – Menu do nó *Analysis* em SPSS Modeler

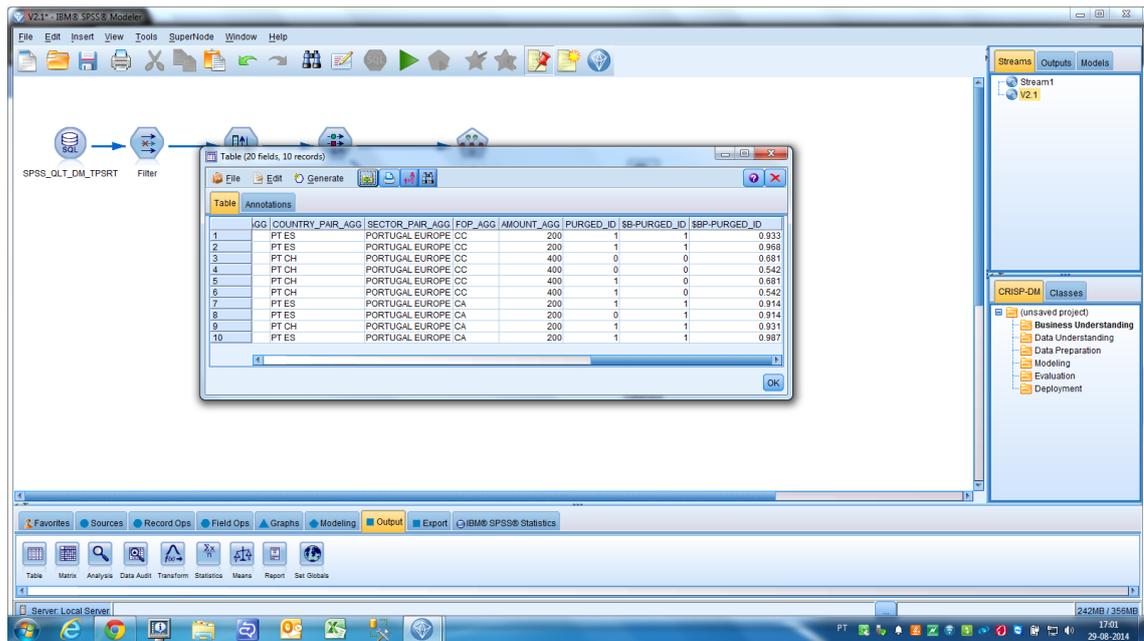


Figura 9 – Menu do nó *Table* em SPSS Modeler

No fundo, o trabalho do estagiário com o SPSS Modeler resumia-se a criar *streams* com esta estrutura (ver figura 10), para posterior análise de resultados. Este processo moroso deveu-se não só ao facto de se tratar de um projeto assente em tentativa e erro, com diversas variáveis a ter em causa, como também ao facto de um projeto com estas características ser pioneiro na empresa, não havendo qualquer membro da equipa BI & CRM com experiência em modelação preditiva ou na utilização de ferramentas SPSS, ficando o estagiário bastante limitado em termos de apoio.

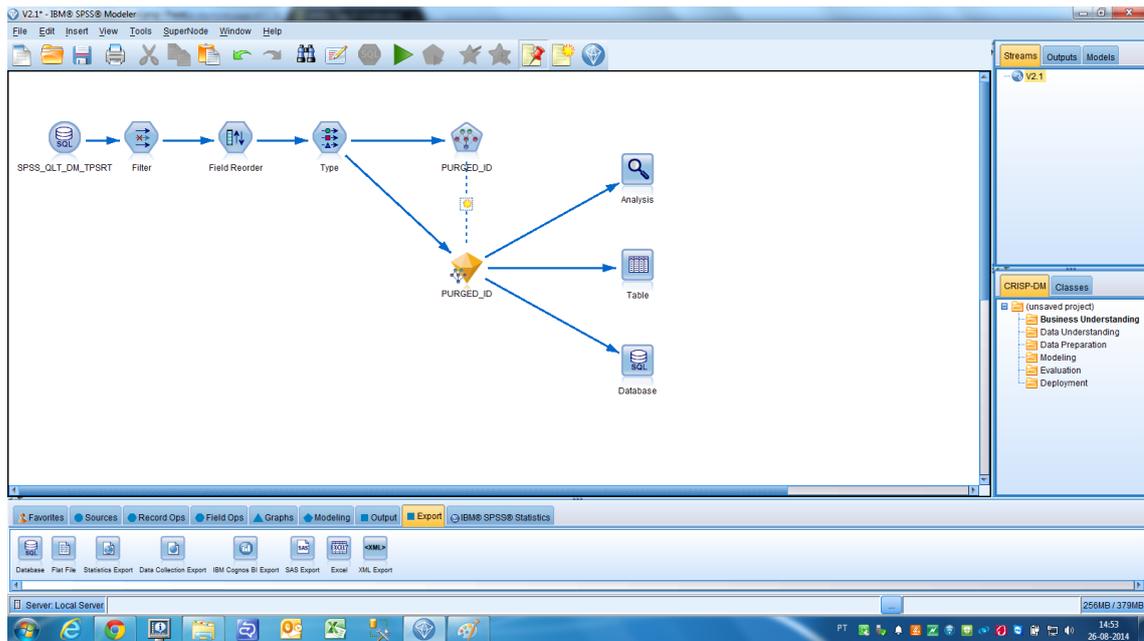


Figura 10 – Exemplo de uma *stream* concluída em SPSS Modeler

4.2. Resultados

Infelizmente, aquando do término do período de estágio, não foi possível atingir os resultados esperados, nomeadamente, um modelo rigoroso que pudesse prever com garantias – com um mínimo de 95% de acerto, tal como definido pelo estagiário e o seu acompanhante –, as receitas relativas ao *purge*. Os problemas foram revistos diversas vezes em reuniões com outros colegas com maior experiência em BI ou em análise de dados, bem como com um consultor externo, tendo-se adotado várias potenciais soluções ao longo do projeto, mas em nenhuma altura se foi capaz de aumentar a percentagem de sucesso na previsão de receitas acima dos 80%, um valor demasiado reduzido para as expetativas do estagiário e do seu acompanhante. Por conseguinte, não estando reunidas as condições para apresentação de resultados ao cliente.

Consequentemente, o projeto acabou por não se revelar bem-sucedido para a Megasis, não se colocando de parte a hipótese de este ser retomado no futuro, se novas necessidades o exigirem. É de salientar, porém, que grande parte do insucesso se deveu à carência de qualidade de informação, face a esta estar mal estruturada ou em falta, algo que foi notado pelos membros da equipa que tiveram acesso aos contornos do projeto, o que permitirá repensar alguns processos no futuro.

5. Conclusões e sugestões de aplicação futura

5.1. Desenvolvimento pessoal e profissional

A oportunidade de realizar este estágio na Megasis foi de extrema importância para o desenvolvimento do estagiário. Efetivamente, constituiu-se como a sua primeira experiência profissional e também como uma componente essencial da sua formação académica e como oportunidade de colocar em prática alguns dos conceitos apreendidos ao longo do curso. De destacar, ainda, o privilégio de acontecer numa empresa do Grupo TAP, uma das maiores e mais importantes organizações de Portugal, bem como uma companhia aérea de renome a nível internacional.

A possibilidade de trabalhar a nível curricular com os processos e ferramentas assimilados de forma teórica ao longo do mestrado permitiu ao estagiário estar, hoje, mais preparado para os desafios que enfrentará ao longo da sua vida profissional, contribuindo igualmente para uma maior abertura na tentativa de entrada no mercado de trabalho.

Trabalhar com ferramentas como o Microsoft SQL Server e o IBM SPSS Modeler foi também bastante proveitoso, pois permitiu conhecer as funcionalidades de cada uma, atribuindo ao estagiário competências práticas em linguagem SQL, *data mining* e modelação preditiva.

Para além da aprendizagem e da experiência profissional adquiridas, é ainda de destacar a importância que o ambiente proporcionado pela convivência com os colegas – quer em termos de trabalho como também de lazer – teve para o desenvolvimento pessoal do estagiário.

5.2. Problemas e limitações

Tal como referido anteriormente, um dos problemas identificados como responsáveis pelo insucesso do projeto surgiu da falta de informação disponível em relação aos bilhetes emitidos, o que, se não fosse o caso, poderia ter trazido grande benefício no que toca à obtenção de resultados.

No entanto, a maior limitação encontrada ao longo do estágio prende-se com a ferramenta em si. O facto de não haver membros na equipa de BI & CRM com experiência a trabalhar com o SPSS Modeler – excetuando a breve formação à qual três membros assistiram uns meses antes – revelou-se um problema para o desenvolvimento do projeto, especialmente devido ao facto de a IBM, empresa responsável pela ferramenta, não estar a oferecer apoio técnico no tempo durante o qual o estágio decorreu. Esta lacuna teve especial impacto, dados os frequentes erros com que o estagiário se deparou na tentativa de criação de análises e modelos, sendo que grande parte destes não conseguiu ser identificada por nenhum dos membros da equipa, o que, por sua vez, atrasou bastante o

desenvolvimento do projeto. A chamada de um consultor externo especializado em SPSS permitiu a resolução de alguns dos problemas com que o estagiário se deparou, não sendo essa opção suficiente para impedir atrasos relacionados com erros gerados pela ferramenta.

Também teria sido interessante chegar a uma fase de apresentação de resultados ao cliente, o que teria permitido a elaboração de um *report* com alguns *dashboards* em MicroStrategy, uma aplicação bastante utilizada pela equipa e que suscitou a curiosidade do estagiário, dado que nunca houve a possibilidade de o experienciar durante o desenvolvimento do projeto, mas tal não foi possível.

5.3. Sugestões para o futuro

Para o desenvolvimento de um ou mais modelos preditivos que permitam a identificação do *purge* com elevada probabilidade de sucesso, seria importante para a Megasis, no futuro, ter uma maior qualidade de dados no seu DW, assegurando uma melhor estruturação da informação e a garantia de que todos os campos necessários para a análise se encontram devidamente preenchidos. No entanto, é já conhecida a intenção da gestão de adquirir uma ferramenta de *Data Quality*, o que poderá resolver alguns problemas identificados pelo estagiário ao longo do projeto, como:

- **Registos duplicados** – uma ferramenta de DQ tem a capacidade de gerar alertas para evitar estas situações;
- **Registos em branco** – a ferramenta pode eliminar ou alertar para estes casos;

- **Perfil dos dados** – uma ferramenta desta natureza permite entender melhor os dados, revelando relações entre tabelas, bases de dados e fontes.

Um estudo, realizado pela *Gartner* (2011), concluiu que a principal razão pela qual 40% de todas as iniciativas de negócio falham na obtenção dos benefícios visados se deve a fraca qualidade dos dados. Percebe-se, assim, a importância que a resolução destes problemas terá para a TAP, em virtude de trazer benefícios não só para o projeto em causa ou para a equipa de BI & CRM, mas também para todas as áreas de negócio da companhia.

Seria igualmente interessante ver a TAP a obter mais informação acerca dos seus clientes. Para propósitos do projeto realizado, teria sido bastante benéfico que a organização tivesse mais dados relativos aos seus passageiros, como o género, faixa etária, históricos de destinos, etc., pois permitiria realizar análises preditivas de teor mais rigoroso. Todavia, uma abordagem mais centrada no consumidor poderia trazer, também, outros benefícios para a companhia, permitindo, por exemplo, executar uma segmentação dos seus clientes.

Um estudo da IBM mostrou que uma das maiores companhias aéreas dos Estados Unidos da América descobriu que o seu *top* de mil clientes por receita era responsável por 60% mais receitas do que o seu *top* de mil clientes por milhas, o que mostra que a análise do valor dos clientes deve ser feita por mais do que um fator (valor monetário, frequência de voo, etc.), e não apenas pelas milhas percorridas por cada um (Boland et al, 2002). Com uma boa segmentação, a companhia poderá conseguir não só oferecer determinados privilégios aos melhores clientes, mas também ter um melhor conhecimento dos mesmos e, assim, poder oferecer serviços personalizados a cada um, e

ainda estabelecer campanhas de *marketing* específicas a determinado grupo de clientes (Ang & Buttle, 2006; Ryals & Knox, 2001; Wilson et al, 2002).

Este tipo de abordagem, centrada no cliente, seria bastante interessante do ponto de vista do *data mining* e da modelação preditiva, que são componentes atualmente pouco aproveitadas pela equipa de BI & CRM e ferramentas essenciais na segmentação de clientes e análise de tendências.

Saber antecipar as circunstâncias que poderão afetar a indústria na qual uma organização se insere é, afinal, um fator chave para o sucesso da sua gestão:

“...é a presciência que permite ao soberano sábio e ao bom general atacar e conquistar, alcançando coisas que estão para lá das possibilidades do homem comum.”

– Sun Tzu, 544 A.C. – 496 A.C.

Glossário

Aprendizagem automática – Subcampo da Inteligência Artificial dedicado ao desenvolvimento de algoritmos e técnicas que permitam a um computador aperfeiçoar o seu desempenho em determinada tarefa.

Base de dados relacional – Base de dados que armazena informação acerca dos seus dados como estes se relacionam.

Business Intelligence – Conjunto de técnicas e ferramentas para a transformação de dados em bruto em informação útil e significativa para efeito de análise de negócios.

Cluster – Grupo formado por dados com semelhanças, dependendo dos critérios definidos num determinado algoritmo.

Companhia aérea low-cost – Companhia com tarifas de valores reduzidos e menor conforto em relação às companhias tradicionais.

Cross-selling – Ato ou prática de venda de um produto ou serviço adicional a um cliente já existente.

Customer Relationship Management – sistema de gestão das interações entre uma organização e os clientes, presentes ou futuros.

Dashboard – Apresentação gráfica dos indicadores de performance de uma organização, permitindo tomadas de decisão informadas e imediatas.

Data Mining – Processo de exploração de grandes quantidades de dados em busca de padrões consistentes.

Data Warehouse – sistema utilizado para análise de dados, integrando dados de uma ou mais fontes. Um repositório central de dados de uma organização.

Modelação Preditiva – Processo pelo qual um modelo é criado ou selecionado para melhor prever a probabilidade de um determinado resultado.

OLAP – Ato de manipular e analisar um grande volume de dados sob múltiplas perspetivas, com o intuito de aceder, visualizar e analisar os dados organizacionais com elevada flexibilidade e *performance*.

Overbooking – O ato de venda de lugares num voo que excede a lotação da aeronave.

Query – Forma, em linguagem de computação, utilizada para realizar consultas em bases de dados e sistemas de informação.

Scorecard – Ferramenta de gestão de *performance* utilizada por gestores para estes acompanharem a execução de atividades do negócio.

Sistema de Informação – Sistema automatizado ou manual que engloba pessoas, processos e tecnologias para o processamento de informação útil ao utilizador.

SPSS Modeler – Software de *data mining* e análise de texto utilizada para a construção de modelos preditivos e condução de outras tarefas analíticas.

SQL – Linguagem de consulta estruturada, assim designada devido à sua função de gerir dados incluídos numa base de dados relacional.

Tabela – Coleção de dados relacionais, numa base de dados, num formato devidamente estruturado, desenhada através de linhas e colunas.

View – Resultado de uma query em SQL que, ao contrário de uma tabela, não é materializada. Por analogia, tabela virtual.

Webinar – Conferência interativa cujos participantes estejam à distância, funcionando muitas vezes como um *workshop online*.

Referências bibliográficas

- Anderson-Lehman et al (2004). Continental Airlines Flies High With Real-Time Business Intelligence. *MIS Quarterly Executive*. 3 (4), 163-176.
- Ang & Buttle (2006). Customer Retention Management Processes: A Quantitative Study. *European Journal of Marketing*. 40 (1-2), 83 – 99.
- Azvine et al (2005). Towards Real-time Business Intelligence. *BT Technology Journal*. 23 (3), 214-225.
- Boland et al (2002). The Future Of CRM In The Airline Industry: A New Paradigm For Customer Management. *IBM Institute for Business Value*, 1-17.
- Chopra et al (2011). Implementation of Data Mining Techniques for Strategic CRM Issues. *International Journal of Computer Technology and Applications*. 2 (4), 879-883.
- Elkan, C (2013). *Predictive analytics and data mining* [em linha]. Disponível em: <http://cseweb.ucsd.edu/~elkan/255/dm.pdf> [acesso em 14/10/2014].
- Fayyad et al (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*. 17 (3) 37-54.
- Gangadharan et al (2004). Business Intelligence Systems: Design and Implementation Strategies. In Luzar-Stiffler & Dobric (Eds.) *Proceedings of 26th International Conference on Information Technology Interfaces*. IEEE: Cavtat/Dubrovnik.
- Gartner (2009). *Gartner Reveals Five Business Intelligence Predictions for 2009 and Beyond* [em linha]. Disponível em: <http://www.gartner.com/newsroom/id/856714> [acesso em 14/10/2014].
- Gartner (2011). *Measuring the Business Value of Data Quality* [em linha]. Disponível em: https://www.data.com/export/sites/data/common/assets/pdf/DS_Gartner.pdf [acesso em 14/10/2014].

- Geisser, S (1993). *Predictive Inference: An Introduction*. New York: Chapman & Hall.
- Lawrence et al (2003). Passenger-Based Predictive Modeling of Airline No-show Rates. In ACM (Orgs.) *KDD '03 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM: New York. 397-406.
- Microsoft (2014). *Data Mining Algorithms (Analysis Services - Data Mining)* [em linha]. Disponível em:
<http://msdn.microsoft.com/en-us/library/ms175595.aspx> [acesso em 14/10/2014].
- Miranda, H (2012). *Building Cluster For CRM Strategies By Mining Airlines Customer Data* [em linha]. Disponível em:
<http://run.unl.pt/handle/10362/9181> [acesso em 14/10/2014].
- Nyce, C (2007). *Predictive Analytics White Paper*. AICPCU/IAA: Malvern.
- Olszak et al (2007). Approach to Building and Implementing Business Intelligence Systems. *Interdisciplinary Journal of Information, Knowledge, and Management*. (2), 135-148.
- Ranjan, J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*. 1 (9), 60-70.
- Riwo-Abudho et al (2013). Key Success Factors in Airlines: Overcoming the Challenges. *European Journal of Business and Management*. 30 (5), 84-88.
- Ryals & Knox (2001). Cross-Functional Issues In The Implementation Of Relationship Marketing Through Customer Relationship Management (CRM). *European Management Journal*. 19 (5), 534-542.
- Rygielski et al (2002). Data mining techniques for customer relationship management. *Technology in Society*. (24), 483-502.
- Sungard Financial Systems (2014). *Assessing Your Firm's Business Intelligence Maturity* [em linha]. Disponível em:
http://financialsystems.sungard.com/resources/consulting-services/white-papers/assessing-your-firms-bi-maturity?prcamp=scs_bimaturitysurvey_073014
[acesso em 14/10/2014].

- TAP Portugal (2013). *Relatório Anual 2012* [em linha]. Disponível em:
http://new.flytap.com/prjdir/flytap/mediaRep/editors/Contentimages/PDFs/Institucional/Relatorios/TAP_Relatorio_Anual_2012_PT-site2.pdf [acesso em 14/10/2014].
- TAP Portugal (2014). *Relatório Anual 2013* [em linha]. Disponível em:
http://new.flytap.com/prjdir/flytap/mediaRep/editors/Contentimages/PDFs/Institucional/Relatorios/2014/TAP_Relatorio_Anual_2013_PT.pdf [acesso em 14/10/2014].
- Tzu, S (2009). *A Arte da Guerra*. Bertrand Editora: Lisboa.
- Wikipédia (2014). “Star Alliance”. Wikipédia, a enciclopédia livre [em linha]. Disponível em:
http://pt.wikipedia.org/wiki/Star_Alliance [acesso em 14/10/2014].
- Wilson et al (2002). Factors For Success In Customer Relationship Management (CRM) Systems. *Journal of Marketing Managements* 18 (1), 193-219.
- Zhang et al (2011). Applications of Business Intelligence Technology in the Airports and Airlines Companies. *International Journal of Applied Science and Technology*. 5 (1), 74-78.