ISCTE Business School
INSTITUTO UNIVERSITÁRIO DE LISBOA

DATA SCIENCE – THE STATE OF THE ART

João Manuel Garcia Faustino

Dissertação submetida como requisito parcial para obtenção do
grau de Mestre em Gestão de Empresas

Orientador:
Prof. Doutor José Dias Curto, Prof. Associado, ISCTE Business School, Departamento de Métodos Quantitativos para Gestão e Economia

Coorientador:
Dr. Ricardo Santos, Partner da WINNING

Outubro 2017

AGRADECIMENTOS

Agradeço ao meu orientador de tese José Dias Curto por me ter sugerido nesta dissertação e providenciado o necessário para que atingisse os meus objetivos. Agradeço ao meu coorientador Ricardo Santos, que me fez sempre elevar os meus objetivos, conduzindo-me a fazer as perguntas certas, bem como agradecer a forma incansável com que me acompanhou durante toda esta etapa.

Agradeço à minha família que sempre me apoiou e ajudou na conquista dos vários desafios académicos, profissionais e pessoais.

Por fim, agradeço aos meus amigos que sempre estiveram presentes em todas as fases da minha vida, quer pessoal quer profissional, para me apoiar.

Obrigado também a todas as pessoas que participaram neste estudo.

Muito Obrigado.

RESUMO

As organizações têm cada vez mais acesso a um maior volume de dados. A contribuir para este fenómeno está o desenvolvimento tecnológico e o conceito de internet das coisas, que permite cada vez mais interligar mecanismos e dispositivos, e consequentemente, diversificar as fontes de informação.

Esta evolução tecnológica permite que os dados sejam retirados das mais diversas formas e plataformas, quer qualitativa quer quantitativamente. Este fenómeno que designamos por *Big Data*, está a tornar disruptivas muitas empresas, alterando desta forma modelos de negócio, inovando o marketing, produtos e serviços e tornando ainda algumas organizações mais eficientes.

Sabe-se também que as capacidades analíticas das empresas têm de dar resposta a este crescimento de dados através de modelos mais avançados, orientados para tomadas de decisões mais acertadas, como a análise preditiva e prescritiva, e recorrendo a técnicas de *Data Mining* ou *Machine Learning* por forma a otimizar a gestão dos recursos e contribuindo para a eficácia e eficiência das organizações.

Esta condição obriga as empresas a aumentar a sua capacidade de adaptação e decisão, para que os dados e a sua compreensão se tornem fontes de vantagem competitiva.

Palavras-chave: Gestão, Análise de Dados, Métodos de Investigação Empresarial, Criação de Valor

Classificação JEL: M10, M15

ABSTRACT

Organizations increasingly have access to a growing volume of data. Contributing to this is the technological development and the concept of the internet of things, which allows increasingly interconnecting mechanisms and devices, and consequently diversify the sources of information. This technological evolution allows the data to be withdrawn in the most diverse forms and platforms, both qualitatively and quantitatively.

This phenomenon, which we call Big Data, is disrupting many companies. Changing this way, business models, innovating the marketing, products and services, still making some organizations more efficient.

It is also known that the analytical capabilities of companies have to respond to this increase in data through more advanced models oriented towards better decision making, such as predictive and prescriptive analysis and using Data Mining or Machine Learning techniques to optimize the management of resources and contributing to efficiency and effectiveness.

This condition forces companies to increase their ability to adapt and make decisions, so that data and their understanding become sources of competitive advantage.

Keywords: Management, Data Analysis, Business Research Methods, Value Creation

JEL Classification: M10, M15

SUMÁRIO EXECUTIVO

O tema dos Dados e a sua análise científica centra-se numa problemática fundamental para o desenvolvimento das empresas no presente e no futuro, relacionando temas como Gestão, Tecnologia e Ciência, pois são alguns dos pilares para o desenvolvimento sustentável das organizações.

Os desafios para as empresas, centram-se na forma como estas recolhem os dados internos e externos à organização, como os armazenam, como os analisam e com que objetivo. Estando todo este processo de gestão dos dados orientado para a criação de valor.

O estudo feito relaciona também a forma como os diferentes setores de atividade, mercados e a dimensão das empresas retiram conhecimento dos dados.

Tendo em conta que os dados têm um papel fundamental para a tomada de decisões nas organizações, por um lado há que considerar a capacidade destes descreverem o passado, por outro a capacidade de estes poderem prescrever as melhores soluções futuras.

A atualidade dos assuntos aqui discutidos deve trazer *insights* valiosos para o mundo empresarial, fazendo com que os conteúdos aqui desenvolvidos tenham aplicabilidade prática. Mais que uma resposta, deve trazer novas perguntas, contribuindo para a forma como são utilizados os dados nas organizações.

O estudo desenvolvido pretende, portanto, avaliar o grau de maturidade das empresas nestas temática de *Data Science*.

ÍNDICE

ÍNDICE DE TABELAS

ÍNDICE DE FIGURAS

PARTE 1 - INTRODUÇÃO

1.1 - Introdução

O facto do *Big Data* ser um fenómeno recente, oferece grandes desafios às organizações. É importante referir que os Dados podem trazer valor para as empresas, ao conhecerem melhor os seus clientes, o mercado, os seus produtos e as suas capacidades operacionais.

Estes dados podem ainda ser utilizados para desenvolvimento de novos produtos e serviços, como também, na implementação de processos mais eficientes, sendo desta forma crucial entender de que maneira é que se processa a sua gestão no sentido da criação de valor para a organização.

Se por um lado existem já inúmeras tecnologias que fornecem dados, não será menos importante também a capacidade de inovar em formas e tecnologias de os obter.

Por outro lado, o desenvolvimento tecnológico destas novas plataformas e mecanismos permite aceder a um conjunto de dados que podem ser quantitativos ou qualitativos e que podem estar organizados de forma estruturada ou não estruturada. Sendo que a par do seu surgimento e da sua dimensão, surge a necessidade de utlizar novos modelos estatísticos e analíticos por forma a dar resposta a este volume de dados.

O dado ao ser muitas vezes um elemento isolado requer a capacidade de o tornar em sabedoria, uma vez que a sua importância não se deve consignar a descrever o passado, mas antes em desenhar o futuro, sendo que, para isso é importante entender a problemática que o permite a tal.

A par destas questões não é alheio que muitas empresas disruptivas tenham visto na tecnologia e utilização de dados a sua vantagem competitiva, criando novos modelos de negócio e alterando as estruturas do mercado.

Desta forma, torna-se pertinente perceber como podemos transformar os dados e toda a informação num recurso a ser gerido de forma a criar a vantagem competitiva, tal como referido anteriormente.

Assim, todos estes fatores se tornam num assunto pertinente e numa problemática para as empresas, pois todas as questões em causam acarretam mudanças significativas na

forma como as empresas atuam para com todos os seus stakeholders e como se podem posicionar perante estes.

PARTE 2 - REVISÃO DE LITERATURA

2.1 – Informação

2.1.1 – Dos Dados à Sabedoria

Segundo Russel Ackoff, o conteúdo da mente Humana pode ser classificado em cinco categorias distintas: Dados, Informação, Conhecimento, Compreensão e Sabedoria. (Bellinger, Castro, & Mills, 2004) Sabemos, no entanto, que o dado por si só é um elemento cru, que pode existir de alguma forma, quer seja útil ou não, visto que de forma isolada não tem significado. (Jifa & Lingling, 2014) Importa também referir que, segundo Cooper (2014), um dado pode ser simplesmente um valor, uma medida, que só depois do devido enquadramento e contexto adquire significado.

Assim, e aquando da relação e conexão entre dados, surge a informação, atribuíndo-se então, significado aos dados, quer seja este significado útil ou não. (Jifa & Lingling, 2014). Cooper (2014) diz-nos ainda que a informação é estruturada seguindo um processo cognitivo e a sua validação vai gerar o conhecimento.

A partir daqui temos o conhecimento, que é a apropriação da informação com a intenção de ser útil, sendo o conhecimento um processo determinístico. Para Jifa e Lingling (2014) o conhecimento apesar de ter um significado útil, por si só não tem capacidade de gerar mais conhecimento. (Jifa & Lingling, 2014) No entanto, Bellinger, Castro e Mills (2004) consideram que o processo em causa está associado à memória da informação.

Cooper (2014) diz-nos também que o conhecimento se divide em dois tipos: as diretrizes escritas, prontamente disponíveis e facilmente passíveis aos outros e as diretrizes implícitas, internas e adquiridas pela experiência ou intuição.

A compreensão torna-se então num processo interpolativo e probabilístico, cognitivo e analítico, mecanismo pelo qual as pessoas podem adquirir novos conhecimentos. (Jifa & Lingling, 2014)

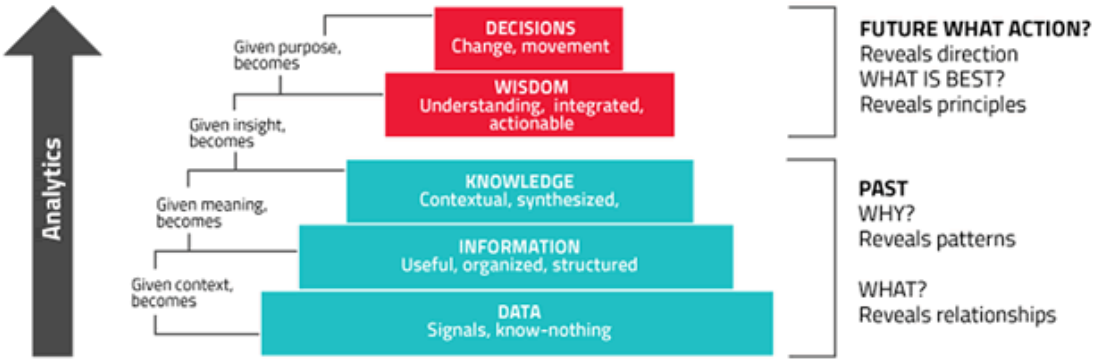
Bellinger, Castro e Mills (2004) acreditam que a diferença entre compreensão e conhecimento reside na diferença entre aprender e memorizar dados. Desta forma, consideramos que os sistemas de inteligência artificial estão aptos a sintetizar novos conhecimentos de informação e conhecimento previamente armazenados.

Por fim, temos a Sabedoria, que é um processo extrapolativo, não determinístico e não probabilístico. Processo este que invoca todos os níveis prévios de consciência e especificamente tipos de programação Humanos (moral, códigos éticos, etc). (Jifa & Lingling, 2014) Através do processo em causa, julgamos o certo e o errado, o bem e o mal, e são ainda feitas perguntas para quais as respostas não são facilmente alcançáveis. (Cooper, 2016)

Resumindo, para Zeleny (1987) o dado está associado ao não saber nada, a informação associada ao saber o quê, o conhecimento ao saber como e, a sabedoria ao saber porquê. Importa também referir que, segundo Awad e Ghazani (2004) a sabedoria é o mais alto nível de abstração, com visão, previsão e sabedoria que permitem ver para lá do horizonte.

É a partir dos conceitos supra-referidos que Mattheus (1998) propõe o modelo de Kennotion (junção de *Knowledge* e *Information*). Modelo este que relaciona todos os níveis, considerando assim, que os dados são a base para a informação, a informação a base para o conhecimento, o conhecimento a base para a sabedoria, a sabedoria a base para a criatividade e, por último, tem a criatividade como base para a inovação.

Figura 1- *Pirâmide DIKW (Fonte: i-scoop.eu)*



2.1.2 – Dados Estruturados e Não Estruturados

Atualmente a tecnologia permite retirar dados relativos ao comportamento dos consumidores, sendo alguns destes mais tradicionais e por isso designados como estruturados, outros mais comportamentais e por isso não estruturados. (Erevelles, Fukawa, & Swayne, 2016)

Como tal, as empresas têm uma crescente quantidade de dados disponíveis não organizados que se tornam difíceis de analisar. Por outro lado, têm também disponíveis dados que, supostamente, já estão organizados e especificados para as suas necessidades de decisão. A estes dois tipos de dados, designados como não estruturados e estruturados respetivamente, podem ser combinados de forma a criar um conjunto de informações mais relevantes e completas. (Kangas, Leskinen, & Kangas, 2007)

Sabemos também que, dados estruturados são considerados informação útil, agrupada, e classificada com capacidade para ser visualizada e de onde pode ser extraída informação. (Sukanya & Biruntha, 2012)

São portanto dados que estão alojados em campos fixos, como as bases de dados relacionadas ou folhas de cálculo. Por outro lado, os dados não estruturados não estão em campos fixos, como o caso de texto em livros, artigos e mensagens de e-mails ou como audio, imagens e vídeos. (Manyika, et al., 2011)

Estas informações, não estruturadas, carecem de regras e tipos de dados definidos para se poder impor onde são armazenados esses dados. Estando muitas vezes armazenados em diretórios definidos pelo *user*, para lá do alcance das regras da empresa. No entanto são muitas vez cruciais para a organização. (Mckendrick, 2011)

Ainda relativamente a dados não estruturados, Beach e Schiefelbein (2014) acreditam que estes não são monitorizados num amplo universo de dados, tal como nos já referidos e-mails mas também documentos do desktop, logs de internet, telefonemas, mensagens de texto e mensagens de redes sociais, e ainda em opiniões de clientes de produtos online. Apesar deste facto, os autores em causa consideram ainda que os dados não estruturados podem ajudar a detetar vários tipos de risco.

Com a recente explosão no acesso a dados digitais, as organizações podem usar dados não estruturados para melhorar as suas tomadas de decisões. (McAfee & Brynjolfsson, 2012)

2.1.3 – Características dos Dados

Atualmente, a tecnologia fornece enormes quantidades de dados, como já referido anteriormente, fenómeno ao qual chamamos *Big Data*. (George, Haas, & Pentland, 2014)

As três definições que nos ajudam a definir *Big Data* são frequentemente referidas como os 3 V's: Volume, Velocidade e Variedade. (Dijcks, 2013)

Posteriormente, foram adicionados dois V's pela importância na recolha e análise de dados para daí retirar conhecimento: Veracidade e Valor (Lycett, 2013)

2.1.3.1 – Volume de Dados

O volume de dados das organizações tem crescido de forma exponencial. De referir que, em 2013, o tamanho do universo digital foi estimado em 4.4 Zettabytes (International Business Management, 2013). Sabemos também que a internet das coisas contribui de forma significativa para o crescimento explosivo, uma vez que a computorização está incorporada em carros, brinquedos, aparelhos, turbinas e coleiras de cães. (Erevelles, Fukawa, & Swayne, 2016). Para além disso, há que referir que os dados também podem surgir de *machine-generated data*, *entreprise data e social data* (Dijcks, 2013). Outros exemplos de onde se podem retirar dados são os sensores de ambiente e de manufactura, *smart meters*, *smart cards*, *scanning equipment e machine to machine electronic tenders*. (Chan, 2013)

Para Das and Kumer (2013) dados não estruturados de texto, documentos, imagens, vídeos, entre outros, representarão 90% do total de dados na próxima década.

2.1.3.2 – Velocidade dos Dados

Lycett (2013) considera que a segunda dimensão da *Big Data* é a velocidade. Minelli (2013) acredita que a velocidade se define como a rapidez com que os dados são criados, acumulados, ingeridos e processados. Este facto é disruptivo em relação à gestão de bases de dados tradicionais.

A IBM (2013) dá como exemplo de velocidade o facto de se analisarem 5 milhões de eventos de mercado em cada dia, de forma a identificar o potencial de fraude ou como outro exemplo a análise de 500 milhões de chamadas diárias em tempo real que detetam a agitação dos clientes de forma mais rápida.

Importa referir, que a análise em tempo real e a resposta a toda a informação, são caraterísticas da gestão de *Big Data* em muitas situações. As informações em tempo real ou quase em tempo real tornam possível a maior agilidade de uma determinada empresa quando comparada com as suas concorrentes. (McAfee & Brynjolfsson, 2012)

Por exemplo, os *Marketers* usam as redes sociais e navegam na Web para a transação de dados, de forma a dar respostas em tempo real a determinados segmentos e targets.

Por outro lado, as empresas monitorizam operações e geram respostas em tempo real, aumentando desta forma a capacidade de resposta em tempo real. (Chan, 2013)

É de notar ainda, que os executivos de Marketing têm acesso a bases de dados completas e perspicazes permitindo-lhes tomar decisões mais acertadas com base em evidências em vez de intuições. (Erevelles, Fukawa, & Swayne, 2016)

2.1.3.3 – Variedade dos Dados

À medida que mais atividades empresariais são digitalizadas, assistimos à conjugação de novas fontes de informação e equipamentos mais baratos, conduzindo-nos desta forma, a uma nova era onde grandes quantidades de informação digital existem virtualmente em qualquer tópico de interesse para uma empresa. (McAfee & Brynjolfsson, 2012)

Sabemos que a pesquisa de dados se tem vindo a modificar, uma vez que, ao contrário dos processos tradicionais, passaram a usar-se plataformas que nos permitem aceder a dados como a Internet (*clickstream*, redes sociais), pesquisa de dados (*surveys* e reportes industriais), dados de localização (dados de dispositivos móveis e geo-espaciais), imagens, dados da cadeia de abastecimento (EDI) e dispositivos de dados (sensores e aparelhos RFID). (Minelli, Dhiraj, & Chambers, 2013)

Importa, no entanto, referir que muitos dados são não estruturados, como textos, documentos, e-mails, blogs, PDF, áudio, vídeo e imagens e essa variedade contribui

para a complexidade de capturar, armazenar, processar e analisar *Big Data*. (Chan, 2013)

A par disto está o desenvolvimento de software *Standard Generalized Mark-up Language* (SGML), que nos permite ver vídeos por forma a determinar elementos comuns que a organização quer capturar, como por exemplo nos vídeos do Youtube, onde se consegue perceber como os clientes reagem a um determinado produto. (Erevelles, Fukawa, & Swayne, 2016)

2.1.3.4 – Veracidade dos Dados

A veracidade é a precisão dos dados, pois estes devem ser adquiridos de forma segura a partir dos recursos certos, garantindo apenas que as pessoas autorizadas tenham permissão de acesso. (Ozkose, Ari, & Gencer, 2015)

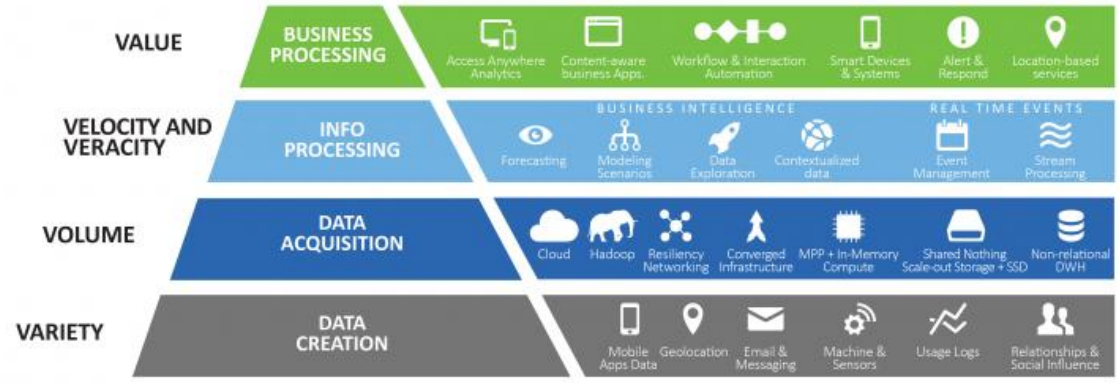
Estabelecer a confiança nos dados representa, no entanto, um enorme desafio com o aumento das fontes. (International Business Management, 2013)

2.1.3.5 – Valor dos Dados

O aumento da quantidade de dados a que as empresas têm acesso cada vez mais, introduz a questão sobre o valor que estes podem ter ou não para as empresas. Daqui advém a necessidade de eliminar os dados não importantes para manter aqueles que são úteis, de forma a retirar daí valor para a empresa. (Lycett, 2013)

O valor deve por fim ser um resultado gerado após todos os procedimentos e esse resultado deve por sua vez enriquecer o processo. (Ozkose, Ari, & Gencer, 2015)

Figura 2 *- 5 V's dos Dados (fonte: excelecom.com)*



2.1.4 – Dados Qualitativos e Quantitativos

As metodologias utilizadas numa pesquisa de dados podem ser indutivas ou dedutivas. Seguir uma abordagem indutiva, que é principalmente qualitativa, significa que a teoria é gerada a partir dos dados. Enquanto que a abordagem dedutiva, principalmente quantitativa, usa a teoria ou hipóteses para testá-la contra os dados. (Hesse-Biber & Patricia, 2011)

***Tabela 1** - Diferença entre dados quantitativos e qualitativos (fonte: Bryman Bell, 2011)*

Quantitative	Qualitative
Numbers	Words
Point of view of researcher	Points of view of participants
Researcher distant	Researcher close
Theory testing	Theory emergent
Static	Process
Structured	Unstructured
Generalisation	Contextual understanding
Hard, reliable data	Rich, deep data
Macro	Micro
Behaviour	Meaning
Artificial settings	Natural settings

Numa pesquisa quantitativa são atribuídos valores numéricos às respostas e podem ser analisados por métodos diretos e em programas de estatística. Em contrapartida, numa pesquisa qualitativa os dados são não-estruturados na maior parte dos casos. (Johnson, Dunlap, & Benoit, 2010)

Neste último caso, dos dados qualitativos, podemos incluir a análise de conteúdo, que estando sob a forma estruturada ou não estruturada, está relacionada com texto, gravações audio, comunicação visual, obras de arte, websites e artefatos culturais. (Mayer, 2015)

A contribuir para esta análise está o avanço na tecnologia computacional e software, que tem permitido gerir determinados conteúdos por forma a que o uso do método qualitativo, através de recolha, armazenamento, organização e análise, dê significado a este tipo de informação obtida, sendo que, a este conteúdo podem ser posteriormente

atribuídos números ou variáveis, que podem ser contados e analisados quantitativamente. (Johnson, Dunlap, & Benoit, 2010)

Enquanto a pesquisa de dados quantitativa é estruturada pelas preocupações do pesquisador, a pesquisa qualitativa é estruturada pelas preocupações do sujeito de pesquisa. (Bryman & Bell, 2011)

Por fim, para fazer a pesquisa de dados qualitativamente trabalha-se com perguntas abertas que podem mudar durante o curso da pesquisa, sendo por isso importante a flexibilidade da abordagem. A questão da pesquisa é obtida com base num ciclo empírico indutivo, estando esta indução centrada na geração da teoria a partir dos dados, contrariamente à abordagem dedutiva. Note-se que, na abordagem dedutiva apenas pequenas amostras podem ser apropriadas. (Mayer, 2015)

O *Data reduction*, *Data display* e o *Drawing and Verifying conclusions* são os três maiores componentes da análise qualitativa (Schutt, 2011)

2.2 – Métodos de Investigação

2.2.1 – Método Científico

O método cientifico é "um método de procedimento que caracterizou a ciência natural desde o século XVII. Esta técnica consiste na observação sistemática, medição, experimentação, formulação, teste e modificação de hipóteses" segundo o dicionário Oxford-English. (Sarma, 2015)

O método em causa inclui duas principais etapas. A primeira consiste em formular hipóteses e a segunda consiste em testá-las experimentalmente. O que diferencia a ciência de outro conhecimento é a segunda etapa: submeter hipóteses a testes empíricos observando se as previsões derivadas de uma hipótese são ou não o caso em observações e experiências relevantes. (Ayala, 2009)

Segundo Vining (2013), o método científico é um processo indutivo-dedutivo que envolve uma interação entre o concreto (universo físico e o contexto específico do problema) e o abstrato (teoria científica e matemática), e cada vez que o concreto e o abstrato interagem geram dados. É, portanto, uma investigação de um problema concreto onde de seguida se propõe uma explicação teórica a esse problema. A teoria orienta o pesquisador para possíveis soluções e, a partir daí, o pesquisador recolhe

dados apropriados, constrói modelos para analisar os dados e, de seguida, interpreta os resultados para testar essas possíveis soluções.

2.2.2 - Grounded Theory

A *Grounded Theory* é descrita como a teoria que deriva dos dados, sistematicamente recolhidos e analisados através do processo de pesquisa, estando neste caso a recolha de dados, a análise e eventual teoria em estreita relação. (Bryman & Bell, 2011)

Segundo Birks e Mills (2010) os métodos para a análise de dados devem ser os seguintes: codificação inicial e categorização, gerar ou recolher e analisar simultaneamente os dados, escrita de memorandos, amostra teórica, análise comparativa constante usando a lógica indutiva, sensibilidade teórica, codificação intermédia, seleção de uma categoria central, saturação teórica e integração teórica.

Easterby-Smith, Thorpe e Jackson (2012, p.58) identificaram as seguintes operações analíticas fundamentais ao aplicar a teoria fundamentada: ciclo de amostragem teórica e comparações constantes; evoluindo e levando à saturação teórica. A amostragem teórica representa um processo repetitivo. Ao usar a *Grounded Theory*, as amostras não são desenhadas de grupos particulares de indivíduos, mas sim em termos dos conceitos e das suas propriedades. A primeira amostra é baseada numa ideia geral do fenómeno em estudo. O investigador descobrirá então que são necessários mais dados para as categorias que emergiram dos estágios iniciais da análise e, portanto, continuará a apurar as categorias até que a saturação seja atingida (Corbin e Strauss, 1990). A análise comparativa constante descreve o processo de comparação contínua de incidentes com outros incidentes para semelhanças ou diferenças. A identificação das diferenças dentro das categorias leva à criação de subcategorias. Este processo continua até que uma *Grounded Theory* seja completamente integrada (Birks e Mills, 2010). Essa forma de comparação leva a uma maior precisão e consistência na pesquisa. (Corbin & Strauss, 1990)

O processo de recolha e análise de dados deve continuar até que a saturação teórica tenha sido realizada, ou seja, até que não surjam novas variações das categorias que saem. Uma teoria integrada fundamentada, que explica um processo relacionado a um fenómeno, é o produto final dessa abordagem. (Birks & Mills, 2010)

2.2.3 – Business Research Methods

Segundo Sue Greener (2008) *Business Research Methods* refere-se à atividade desenhada especificamente para gerar dados (ex: questionários, entrevistas, *focus group*, observações) e à metodologia de pesquisa, atitude e estratégia para o entendimento desta com a finalidade de responder a questões.

Por sua vez, a pesquisa traduz-se num processo de recolha, análise e interpretação de dados para entender um determinado fenómeno. E, por sua vez, esta pesquisa origina pelo menos uma questão sobre o fenómeno de interesse, sendo a abordagem a esta pesquisa feita de uma forma qualitativa, quantitativa ou mista. (Williams, 2007)

Nas pesquisas comerciais é promovido o aumento da consciência e compreensão sobre problemas e oportunidades empresariais. Por conseguinte, é importante desenvolver e executar planos alternativos e, finalmente, monitorizar o desempenho do negócio com dados factuais. (Zikmund, Babin, Carr, & Griffin, 2010)

O foco da pesquisa é que a informação possa facilitar a tomada de decisão e diminuir os riscos de modo a aumentar a probabilidade de sucesso das decisões. Para tal, é tão importante ter informação do que se passa dentro da empresa, como informação do que a rodeia. (Zikmund, Babin, Carr, & Griffin, 2010)

Para Cooper e Schindler (2013) a pesquisa deve estar aliada a boas práticas para que possa ser usada na tomada de decisão. Usando para tal, padrões presentes no método científico, tornando a análise confiável e racional.

Existem muitas ferramentas e técnicas que permitem apoiar o processo de obtenção de dados e análise, fornecendo a todo o processo a racionalidade e o rigor necessários. Alguns exemplos dessas ferramentas são: *Benchmarking*, *Focus Group*, *Interviews*, *Control Group*, Observação, Estudo de Mercado, Levantamento / Questionário, Julgamento de Perito, Histórico Relatórios / Relatórios (Coopler & Schindler, 2013), *Wisdom Crowd* (Yi, Steyvers, Lee, & Dry, 2012) e *Hall Test* (Dumas & Redish, 1999). Na tabela a seguir está descrita cada ferramenta / técnica enunciada antes:

***Tabela 2** - Métodos de investigação Empresarial (Adaptado de várias fontes)*

Ferramentas e técnicas de pesquisa	**Descrição**
Benchmarking	Processo de nos compararmos com os outros, e dessa comparação podermos extrair dados. Esta comparação, pode ser feita entre processos, pessoas, programas, etc (Coopler & Schindler, 2013)
Wisdom Crowd	Tem como objetivo obter dados / pensamentos de um grupo de indivíduos, em vez de obtê-lo de um indivíduo. É percebido que tomar uma decisão com base em dados recolhidos do uma "Multidão", tem melhores resultados do que as decisões básicas em dados fornecidos por um indivíduo (Yi, Steyvers, Lee, & Dry, 2012)
Focus Group	Envolvimento de um pequeno grupo de pessoas (8 a 10) que interagem entre si gerando dados sobre um tópico específico. Essa interação é moderada pelo pesquisador ou equipa de pesquisadores (Coopler & Schindler, 2013)
Interviews	Abordagem de comunicação para recolher dados, via telefone, pessoalmente, videoconferência. (Coopler & Schindler, 2013)
Observations	Monitorizar comportamentos, atividades e condições. Exemplos: análise linguística, análise linguística extra, análise condicional física, análise espacial. (Coopler & Schindler, 2013)
Control Group	Grupo de participantes que não estão expostos a variáveis em estudo, para usá-las como uma medida de comparação de base com grupos expostos à variável independente. (Coopler & Schindler, 2013)
Hall Test	Teste de usabilidade, onde pessoas aleatórias se reúnem com o propósito de testar um produto ou serviço. (Dumas & Redish, 1999)
Expert judjement	Reunir dados fornecidos por alguém conhecedor de um tópico reconhecido por outros a credibilidade necessária para expressar tais dados. (Coopler & Schindler, 2013)
Market Study	Tipo particular de pesquisa para recolha e avaliação dos dados relativos às preferências, comportamentos e ideias do consumidor. (Coopler & Schindler, 2013)
Survey/Questionnaire	Entrevista estruturada com o objetivo de recolher dados. Os levantamentos são compostos por ferramentas de medição, como questionários e instrumentos de medição. (Coopler & Schindler, 2013)

2.3 – Técnicas de Análise de Dados

2.3.1 – Data Mining

Data Mining é um termo utilizado para analisar dados e extrair relações e padrões potencialmente úteis, interessantes e previamente desconhecidos. (Cooper, 2016)

Sabemos que, os processos típicos de *Data Mining* consistem em várias etapas e o processo é inerentemente interativo e repetitivo, consistindo em cinco principais fases (Bhatt & Kankanhalli, 2011):

1- Compreensão do Domínio
2- Seleção de dados
3- Pré-processamento de Dados, limpeza e transformação
4- Descoberta de padrões
5- Interpretação
6- Informar e usar o conhecimento descoberto

Para Wu e Li (2013) a maior parte dos trabalhos de pesquisa na comunidade de *Data Mining* têm tido como foco principal o desenvolvimento de *Mining Algorithms* eficientes, descobrindo a variedade dos padrões de grandes recolhas de dados. As técnicas de minar dados incluem minar de regras de associação, minar itens frequentes, minar padrões sequenciais, minar padrões máximos e minar padrões fechados.

Por fim, é importante mencionar que é através de *Data Mining* que Copper (2014), pela análise de dados retira interessantes relações e padrões previamente desconhecidos.

2.3.2 – Modelos de Análise

Existe um interesse relevante em *Big Data*, caracterizado por alto volume, variedade e velocidade. Os dados em causa são transmitidos às organizações através de dispositivos de deteção de máquinas, sites, redes sociais, chips de RFID, Sistemas GPS e arquivos de voz, imagem e vídeo. (Watson, 2015)

A interação das pessoas nas lojas, permite que as estas lojas tenham a oportunidade de aprender algo sobre os seus clientes em tempo real pois essa informação permite prever preferências dos consumidores e se estes estão interessados numa próxima visita. Com o surgimento de bases de dados extremamente grandes e registos de transações cada vez

melhores, a relação entre o que compramos, onde vamos e o que vamos fazer em seguida, é cada vez mais clara. (Tucker, 2013)

Eric Siegel refere-se a essa semi-clarividência computorizada como "efeito de previsão". (Tucker, 2013)

Muitos dos valores dos dados podem vir, segundo Mayer-Schunberger e Cukier, de segundos usos e todas as bases de dados têm informação intrínseca e escondida com valor, sendo o objetivo descobri-la. (Hayashi, 2013)

No entanto, a abordagem estatística típica confia nos valores de p para estabelecer a significância de uma descoberta e é improvável que seja eficaz porque o imenso volume de dados significa que quase tudo é significativo. O desafio é mudar o foco nos valores de p para a focalização, e ao contrário disso, nos tamanhos de efeito e variação explicados. Mais do que em médias e agregados o importante é o foco nos *outliers*. Nesse universo, os outliers podem ser o mais interessante permitindo inovações críticas, tendências, interrupções ou reduções que estejam a acontecer fora das tendências médias. (George, Haas, & Pentland, 2014)

Para Watson (2015), o trabalho de um *Data Scientist* é descobrir padrões e relações nos dados que mais ninguém conseguiu interpretar ou questionar, e transformar essas descobertas em informações úteis e que criam valor para a organização. Para tal, há que ter um conjunto completo de compreensão de dados, habilidades analíticas e conhecimento empresarial.

A partir daqui, podemos dividir o *Business Analytics* em três maiores perspectivas como é comumente vista: Descritiva, Preditiva e Prescritiva. (Evans & Lindner, 2012)

2.3.2.1 – Análise Descritiva

Muitos negócios têm início numa análise descritiva, remetendo-nos para o uso de dados no sentido de perceber performances de negócio passadas e atuais, por forma a tomar decisões informadas, categorizando, caracterizando, consolidando e classificando os dados para converter em informação útil. Resumindo-os, posteriormente, em gráficos e relatórios sobre orçamentos, vendas, receitas, etc. Os relatórios, processamento analítico on-line (OLAP), painéis/*scorecards* e visualização de dados são exemplos desta análise descritiva. (Watson, 2015)

Esses dados históricos são o fundamento sobre o qual os algoritmos preditivos são desenvolvidos, a análise preditiva é identificada e, em último caso, a partir do qual a análise prescritiva deriva. Sabemos também que atualmente, para qualquer número de possíveis dispositivos conectados à Internet, é essencial uma ferramenta de gestão de dados. Em segundo lugar, os dados históricos precisam de ser preparados para garantir que eles refletem as tendências atuais. (Rowe, 2017)

2.3.2.2 – Análise Preditiva

Por sua vez, a Análise Preditiva avalia a performance passada, detetando padrões ou relações entre os dados e extrapolando-os para a frente no tempo. Se relacionarmos isto com a teoria DIKM (*Data*, *Information*, *Knowledge*, *Wisdom*), é através dos primeiros dois níveis (dados e informação) que usando vários métodos como o *Data Mining*, *Text Mining*, *Web Mining* e as ferramentas como bases de dados, *data warehouses* e *Management Information System* podemos tornar os dados úteis. Para seguir para o próximo nível, o conhecimento, é usado "KDD", *Knowledge Engineering and Management e Intelligence Knowledge*. (Jifa & Lingling, 2014)

As análises preditivas podem por tanto incluir estatísticas, *machine learning*, aprendizagem profunda, *data mining* e simulação. (Rowe, 2017)

Ainda de notar que os algoritmos e métodos para análises preditivas incluem análise de regressão, análise fatorial e redes neurais. (Watson, 2015)

A partir desses conjuntos de dados constroem-se por exemplo modelos preditivos que relacionem vendas com preços, condições climáticas e qualquer outra informação disponível e, em seguida, extrapolam os níveis de vendas subsequentes. Ao aplicar análises preditivas, não se sabe antecipadamente quais os dados importantes. A análise preditiva determina que dados são preditivos do resultado que se deseja prever. (Rowe, 2017)

Técnicas, como a média móvel tentam descobrir padrões históricos nas variáveis de resultados e extrapolá-los para o futuro. Por outro lado, a regressão linear também visa capturar interdependências entre variáveis explicativas tentando fazer previsão. Apesar de tudo, muitos fatores têm contribuído para o desenvolvimento de novos modelos estatísticos em *Big Data*, contrariamente aos métodos tradicionais que estão orientados para a significância estatística, pois uma pequena parte é obtida da população inteira,

deixando a significância estatística de ser tão importante em grandes volumes de dados. (Gandomi & Haider, 2015)

Como tal, usar as típicas ferramentas estatísticas para analisar *Big Data* pode facilmente levar ao erro. No entanto, não significa necessariamente que deveríamos passar para técnicas mais sofisticadas e complexas de lidar com o problema. As estatísticas Bayesianas básicas e os métodos de regressão escalonados podem ser apropriados. Para além destes, há uma série de técnicas especializadas para analisar *Big Data*, técnicas essas que se baseiam em várias disciplinas, incluindo estatística, ciências da computação, matemática aplicada e economia. As técnicas são testes A/B, processamento de sinais, análise espacial, simulação, análise de séries temporais e visualização (Manyika, et al., 2011)

Fan, Han e Liu (2014) afirmam que, o facto do *Big Data* ser caracterizado pelo tamanho massivo de exemplos, vai ter impacto para a heterogeneidade. Isto porque a enorme dimensão permite descobrir padrões escondidos que não são reconhecidos em populações mais pequenas, sendo necessário, por isso, novos modelos estatísticos mais sofisticados. Em segundo lugar, surgem fenómenos associados à grande dimensão como a acumulação de ruído, correlação falsa e endogeneidade incidental, tornando desta forma também inadequados os modelos estatísticos tradicionais.

Considerando que o *Big Data* para Baessens, Bapna, Maisden, Vanthienen e Zhao (2016) permite alavancar análises preditivas e causais, o *State-of-the-art* assenta numa variedade de técnicas como *Machine Learning*, Estatística Clássica, Econometria ao Design de experimentação, etc.

No entanto, para Shmueli e Koppius (2010), um modelo explicativo é melhor ajustado para testar hipóteses causais e um modelo preditivo empírico puro é melhor em termos de poder preditivo.

***Tabela 3**- Modelos preditivos (Adaptado de várias fontes)*

Modelo	Descrição
Neural Network	Grupo interconectado de elementos de processamento simples (neurónios artificiais ou nós) análogo à rede de neurónios do cérebro humano. Funciona como um processador paralelo que usa um modelo matemático para processamento de informações com base numa abordagem de conexão para computação. Este método pode modelar relações complexas não-lineares entre

	entradas e saídas desejadas através de testes para encontrar padrões em dados que não são facilmente analisados usando métodos convencionais. (Azari, Samani, & Mansoori, 2015)
Suppor Vector Machine	Método de *data mining* baseado na teoria da aprendizagem estatística e o seu objetivo é encontrar o hiperplano de separação ideal que pode atender aos requisitos da qualificação. (Ding, Huang, Yu, & Zhao, 2015)
Bayesian Networks	As redes bayesianas são modelos gráficos probabilísticos que permitem incorporar conhecimento e atualizar o grau de confiança sobre as variáveis target dando nova informação a outras variáveis. Bastante utilizadas para problemas com incerteza inerente, como classificação, diagnóstico e tomada de decisão. (Mrad, Delcroix, Piechowiak, Leicester, & Abid, 2015)
K-nearest Neighbours	Provavelmente um dos algoritmos mais estimados no *data mining*. Tem como base o facto de amostras desconhecidas serem colocadas em classes mais próximas. Este conceito consiste em detetar os vizinhos mais próximos e na escolha da classe mais frequente entre eles. Os vizinhos são classificados de acordo com a sua distância à amostra não classificada, sendo atribuída maior importância aos vizinhos mais próximos. (Geler, Kurbalija, Radovanović, & Ivanović, Comparison of different weighting schemes for the kNN classifier on time-series data, 2016)
Decision Trees	A árvore de decisão traduz-se num gráfico direcionado usado para classificar vários nós. Consiste em um nó da raiz (um nó no gráfico ao qual nenhum outro nó aponta), nós internos (nós que derivam de outros pontos e seguem para outros nós) e folhas (nós que não apontam para outros nós). Durante o processo de classificação, o nó classificado "viaja" da raiz para uma das folhas, onde uma classificação é feita. A classificação pode ser simplesmente uma das classes possíveis ou um conjunto de probabilidades (uma para cada um dos valores de classe possíveis). (Katz, Shabtai, Rokach, & Ofek, 2014)
Linear Regression Models	Traduz a relação entre uma variável dependente, y, e um conjunto de variáveis independentes, x, através de uma função linear. A regressão linear simples consiste em apenas uma variável independente, denotada como x, enquanto a regressão múltipla consiste em mais de uma variável independente. (Hu, 2011)
Logistic Regression Models	O modelo de regressão logística deve ser usado em vez de um modelo de regressão linear, tendo em conta uma variável dependente binária, mesmo que o tamanho da amostra seja pequeno. Este modelo pressupõe que a probabilidade do evento está vinculada a uma combinação linear das variáveis independentes ou preditoras no estudo (a função de regressão) por uma função logística de distribuição cumulativa (função não-linear). (Ge & Whitmore, Binary response and logistic regression in recent accounting research publications: a methodological note, 2010)

Time Series Models	Time Series é uma sequência de dados, que consiste em medidas ou observações sucessivas de variáveis quantificáveis, feitas num intervalo de tempo, regular ou irregular. (Fawumi, 2015)
Random Forests	Processo de agregação (várias árvores são combinadas para obter o estimador final) e diversidade das árvores que são agregadas. Podemos distinguir duas fontes de diversidade: - criar aleatoriedade na partição Pfinal, - criar aleatoriedade nos rótulos, isto é, do valor previsto em cada célula de Pfinal, dado Pfinal. Em florestas puramente aleatórias, as partições são construídas independentemente dos dados, de modo a que ao reescrever (se houver) só atue na randomização dos rótulos. (Arlot & Genuer, 2016)

2.3.2.3 – Análise Prescritiva

A Análise Prescritiva utiliza a otimização de forma a identificar as melhores alternativas e maximizar ou minimizar algum objetivo. (Gandomi & Haider, 2015)

Testa vários cenários, avaliando cada um deles para determinar que curso de ação produzirá o resultado mais desejável - um processo que geralmente envolve a aprendizagem automática. (Rowe, 2017)

Esta análise investiga assim o que deve ocorrer e, é usada para otimizar o desempenho do sistema. (Watson, 2015)

Segundo George, Haas e Pentlantd (2014) as metodologias para analisar os dados são tão importantes quanto as suas fontes, e os padrões de evidência que seriam aceitáveis para os estudiosos em questão.

Ainda de notar que o crescente uso de *machine learning* tem permitido o aumento da análises prescritivas, pois permite que os sistemas tenham a capacidade de aprender e reagir a um determinado cenário, pois estão constantemente a recolher dados e a gravar ações. (Rowe, 2017)

Tabela 4 *- Modelos prescritivos (Adaptado de várias fontes)*

Modelo	Descrição
Optimisation Methods	Os modelos de simulação tentam explicar as relações entre entrada e saída de sistemas complexos mas não fornecem a capacidade de encontrar o conjunto ótimo de variáveis de decisão em termos de função objetiva predefinida. Este é o

	objetivo dos modelos de otimização, que permitem aos decisores encontrar as melhores alternativas possíveis, enquanto o seu impacto no desempenho do sistema é avaliado usando modelos de simulação. Os métodos de otimização usam diferentes mecanismos para procurar a solução ideal. Isto está altamente dependente de muitos fatores, como a abordagem de modelagem, a complexidade do problema e os objetivos dos tomadores de decisão. A solução ideal é o vetor que dá o valor ótimo global (máximo/mínimo) da função objetivo e evita o local ótimo. (Abo-Hamad & Arisha, 2011)
Monte Carlo Simulation	Algoritmo computacional que usa aleatoriamente amostragem repetida, de forma a calcular um determinado resultado. (Ben-Assuli & Leshno, 2013)
Matrix Factorization	Dado um padrão específico ou matriz de dados de entrada, acredita-se que qualquer um deles resida num espaço de dados com dimensões muito menores. Entenda-se que, uma observação pode ser considerada como uma combinação linear ou não-linear de apenas algumas variáveis ocultas ou latentes. A técnica de factorização da matriz não negativa (NMF) geralmente produz uma representação de dados escassa, codificado por uma pequena fração de componentes. (Li, Bu, Zhang, & Chen, 2015)
Deep Learning Methods	Conceito que pode ser visto como um conjunto de técnicas de aprendizagem, que apresentam determinadas características em comum. Assim, consistem em várias camadas de neurónios, interligados de forma a identificar, extrair e processar certas características. O resultado produzido por cada camada intermediária é usado como uma entrada para o seguinte. Estas técnicas podem aplicar métodos supervisionados ou não supervisionados. (Petroşanu & Pîrjan, 2017)

2.4 – Criação de Valor

2.4.1 – Pereira's Diamond

Envolvidas num ambiente tecnológico, global e competitivo, as empresas têm cada vez mais escassez de recursos, quer sejam estes financeiros, humanos ou de tempo. Daqui nasce a necessidade de cada vez mais os racionalizar para que estas possam garantir o seu futuro. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

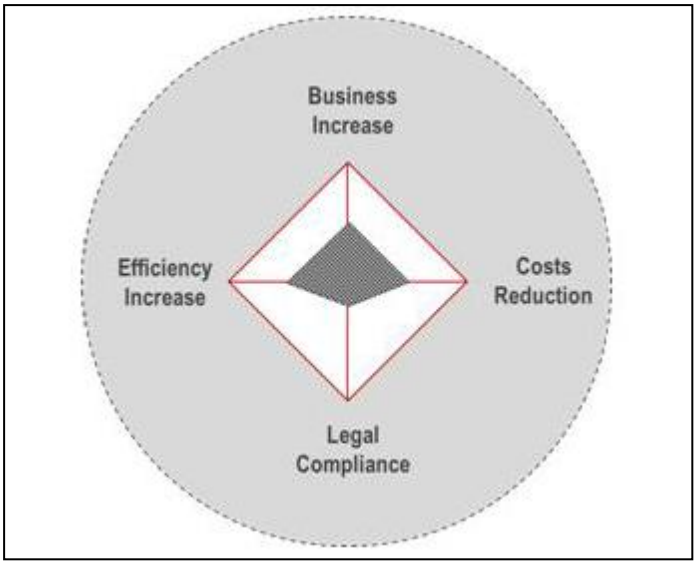
Teixeira e Pereira (2015) afirmam que segundo o BABOK (Business Analysis Body of Knowledge, uma referência na prática de Business Analysis) as necessidades das Organizações podem vir de quatro origens:

- *Top Down*: necessidade de alcançar um objetivo estratégico;
- *Bottom Up*: um problema com os processos, funções ou sistemas;
- *Middle Management*: necessidade de mais informação para apoiar decisões ou novas necessidades para atingir objetivos;
- *External Sources*: devido a terceiras partes, conformidades legais ou competição do mercado

Neste contexto de recursos escassos, as empresas devem investir em projetos que aumentem o valor ao seu negócio, fazendo para isso face às contínuas mudanças de necessidades dos clientes, gerindo bem os recursos internos, estando atentas ao mercado e convertendo a informação em projetos inovadores. Para serem racionais, os investimentos nestes projetos devem ser objetivos, imparciais, mais previsíveis, evitando desvios de alcance, orçamento e tempo. Tendo em conta esta missão na criação de valor, a origem dos investimentos nos projetos deve ser feita sobre quatro vetores que definem o Pereira's Diamond. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

No entanto, para Pereira e Teixeira (2015) é ainda importante perceber que o valor de algo deve ser medido pelo impacto que gera e não pelo seu custo. Tomando em consideração a criação de valor, o Pereira's Diamond define-se em quatro dimensões: Aumento do Negócio, Aumento da Eficiência, Redução de Custos e *Legal Compliance*. Quando aplicado num *Business Case*, como já referido anteriormente, o impacto deve ser medido pelo valor económico gerado e não na perspetiva financeira. Cada dimensão do Pereira's Diamond tem um objetivo que pode ser alcançado de diferentes formas.

Figura 3 - *Pereira's Diamond (fonte: Teixeira e Pereira, 2015)*



2.4.1.1 – Aumento de Negócio

No caso do aumento de negócio, o resultado pode ser alcançado através de quatro formas diferentes:

- Aumento da cota de mercado e do volume de vendas através de novos mercados ou aumento de portfólio de produtos.
- *Cross Selling* vendendo outros produtos para os mesmos clientes.
- *Up Selling* vendendo mais dos mesmos produtos aos mesmos clientes.
- Retenção dos clientes com a fidelização e consequente aumento do ciclo de vida do cliente. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

2.4.1.2 – Redução de Custos

A redução de custos é conseguida através da redução das despesas a nível financeiro e não em termos de horas de esforço. É quantificada em termos de redução de custo ou por evitar o aumento deste no futuro. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

2.4.1.3 – Aumento de Eficiência

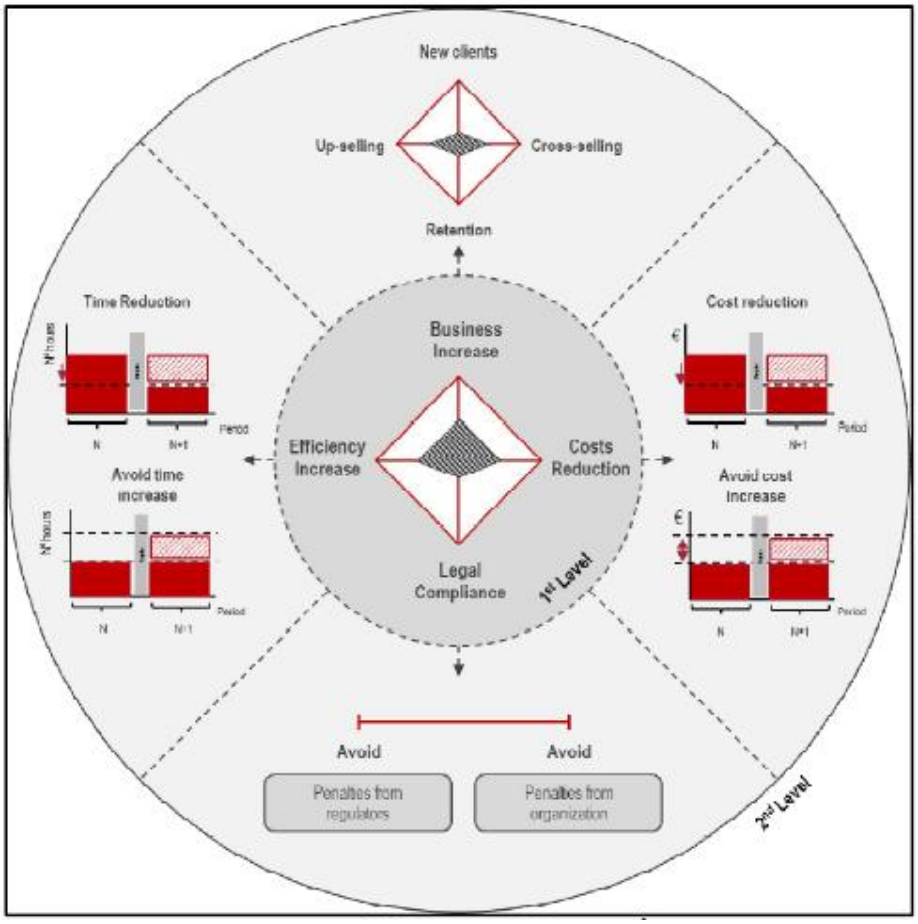
O aumento da eficiência pelo contrário não tem implicação financeira mas antes nas habilidades humanas para otimizar processos de redução de tempo ou que evitem o seu

aumento no futuro. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

2.4.1.4 – Legal Compliance

No caso dos projetos de conformidade legal, procura-se cumprir com as entidades reguladoras ou políticas de modo a que não advenham daí mais custos. (Teixeira & Pereira, Pereira Diamond: Benefits Management Framework, 2015)

Figura 4 *- Pereira's Diamond 2º nível (fonte: Teixeira e Pereira, 2015)*



PARTE 3 – METODOLOGIA

Este capítulo tem como objetivo descrever como é feita a investigação por forma a perceber a problemática da investigação e a relacionar os dados e os métodos de investigação empresarial dos vários setores com as técnicas utilizadas em cada um deles. A revisão de literatura está alinhada com a metodologia utilizada, uma vez que esta serve de enquadramento à investigação.

3.1 – Objetivos de Investigação

O objetivo da investigação é perceber de que forma as empresas estão a utilizar os dados por forma a tomarem decisões mais informadas na sua gestão. À medida que a tecnologia evolui e permite aceder a mais dados externos e internos, a capacidade de os utilizar em seu benefício requer a utilização de um conjunto de métodos e técnicas para que os dados se tornem em conhecimento útil para o futuro, incrementando a eficiência e eficácia, cobrindo riscos e reformulando as estratégias.

Resume-se esta investigação à pergunta: **Como e para quê as empresas analisam os dados?**

Por um lado, na investigação, devemos perceber quais os modelos de análise de dados as empresas utilizam para tomarem decisões. Sendo que, neste caso podemos dividir em métodos descritivos (que tentam perceber o passado), métodos preditivos (que tentam perceber o que vai acontecer) e os métodos prescritivos (que simulam várias hipóteses na procura da mais eficaz tendo em vista determinado objetivo).

Por outro lado, importa também perceber os métodos de investigação empresarial que complementam as técnicas de análise de dados e definem a forma como podemos retirar informação da envolvente empresarial e, por fim, perceber com que objetivo utilizamos estas técnicas e métodos nos mais variados setores.

3.2 – Questões da pesquisa

As técnicas e métodos definidos anteriormente deram origem a um conjunto de perguntas divididas em três grupos:

Q1: Quais as técnicas de análise descritiva de dados mais usadas?

Q1.1: Para que fim as empresas utilizam técnicas de análise descritiva?

Q2: Quais as técnicas de análise preditiva de dados mais usadas?

Q2.1: Para que fim as empresas utilizam técnicas de análise preditiva?

Q3: Quais as técnicas de análise prescritiva de dados mais usadas?

Q3.1: Para que fim as empresas utilizam técnicas de análise prescritiva?

O objetivo é perceber qual o nível de maturidade da recolha e da análise de dados que as empresas estão a fazer dos dados que são gerados todos os dias no ambiente empresarial.

3.3 – Metodologia de Pesquisa

A abordagem de pesquisa deve estar enquadrada com a problemática do tema de investigação. O que se pretende perceber com a investigação é a forma como as empresas estão a utilizar dados por forma a tomarem melhores decisões de gestão. Primeiramente, o estudo deve estar segmentado por diferentes setores empresariais, volume de faturação, geografia e dimensão orgânica, permitindo perceber quantitativamente a forma como cada um faz análise de dados e também quantificar as várias metodologias utilizadas. Esta análise quantitativa permitirá a aplicação da estatística descritiva para a sua interpretação.

Uma vez que o estudo é descritivo e amplo, e o objetivo é simplesmente recolher dados quantitativos, a escolha de questionário foi a melhor opção.

3.4 – Técnica de Pesquisa

Uma vez considerado o questionário como a técnica de pesquisa que permitiria uma maior amplitude para interpretação dos resultados, procurou-se desenvolvê-lo com base nas premissas anteriores, feito em Inglês por permitir maior abrangência e enquadrado no âmbito da temática de investigação, resumindo a tabela seguinte a estrutura principal do survey:

Tabela 5 - *Descrição do Questionário (Tabela Construída)*

Página	Método	Técnicas e Ferramentas	Pergunta
3	Métodos de Análise Descritiva	• Benchmarking • Wisdom Crowd • Focus Group • Interviews • Observation	Q1

		• Control Group • Hall Test • Descriptive Statistics • Expert Judgment • Market Study • Survey/Questionnaire	
3	Objetivos do Método utilizado	• New Clientes • Up-Selling • Cross-Selling • Clientes Retention • Cost Reduction • Cost Avoidance • Efficiency Increase • Other	Q1.1
4	Métodos de Análise Preditiva	• Neural Network • Support Vector Machine • Bayesian Networks • K-nearest Neighbours • Decision Trees • Linear Regression Model • Logistic Regression Model • Time Series Models • Random Forests	Q2
4	Objetivos do Método utilizado	• New Clientes • Up-Selling • Cross-Selling • Clientes Retention • Cost Reduction • Cost Avoidance • Efficiency Increase • Other	Q2.1
5	Métodos de Análise Prescritiva	• Neural Network • Optimisation Methods • Monte Carlo Simulation • Matrix Factorization • Deep Learning Methods	Q3
5	Objetivos do Método utilizado	• New Clientes • Up-Selling • Cross-Selling • Clientes Retention • Cost Reduction • Cost Avoidance • Efficiency Increase • Other	Q3.1

Para cada pergunta do questionário, existe uma escala de resposta de “Don’t Know” a 6, sendo que o 6 significa sempre, para quantificar as várias ferramentas dos vários

métodos de investigação e previsão. Para além destas páginas principais do survey, existem duas páginas anteriores, a primeira pretende fazer um enquadramento do estudo e do seu objetivo aos participantes, a segunda pretende, essencialmente, caracterizar a resposta de cada interveniente e segmentar por dimensão do mercado, país, dimensão, volume de negócios, setor, cargo e experiência.

Existe ainda uma pergunta que sucede a cada Método de investigação, previsão ou prescrição com o objetivo de perceber qual o fim para que é usado determinado método.

Antes de cada grande grupo de perguntas existe sempre uma pequena frase de enquadramento a explicar resumidamente o tema.

3.4.1 – Técnica de Pesquisa Pré-teste

Foi feito um pré-teste a um grupo restrito, importante para perceber alguns erros que possam existir no questionário, bem como incorporar sugestões que possam tornar o questionário mais impactante.

Neste sentido, o pré-teste foi feito a um conjunto de 5 pessoas, estando estes fortemente relacionados com os métodos científicos de investigação empresarial e também com os métodos de previsão, que permitiram estruturar o survey sob os vários modelos de análise de dados e para cada modelo questionar o objetivo da sua utilização com as questões sobre criação de valor. Tornando desta forma o questionário o mais sintético possível e com o objetivo claro de perceber o quê e para quê.

3.5 – Target da Investigação

O Target desta investigação é essencialmente gestores que no seu dia-a-dia tenham de tomar decisões com base em modelos analíticos.

3.6 – Considerações finais

O questionário foi enviado entre Junho e Setembro de 2017, com o target definido para Gestores com capacidade de decisão. O seu envio foi fundamentalmente através de Linkedin, onde foram apuradas 102 respostas.

PARTE 4 – RESULTADOS

Na parte 4 o objetivo é fazer a análise e retirar conclusões dos resultados e dados obtidos no survey. Pretende-se analisar e testar as hipóteses definidas na metodologia. Para simplificar a análise os dados serão arredondados.

4.1- Perfil da Amostra

Para começar a análise do estudo, e segundo o primeiro grupo de perguntas, é feito um enquadramento das empresas que responderam ao estudo e a caracterização do seu perfil. Isto inclui o setor da empresa, o mercado de atuação da empresa, o país da empresa, a dimensão da empresa, o volume de negócio do ano anterior, o cargo do colaborador e a sua experiência.

4.1.1- Setor

Para perceber se o uso do *Data Science* está influenciado pelo setor/indústria, foi perguntado aos participantes o setor onde trabalham.

***Tabela 6** - Setor (sumário de respostas %)*

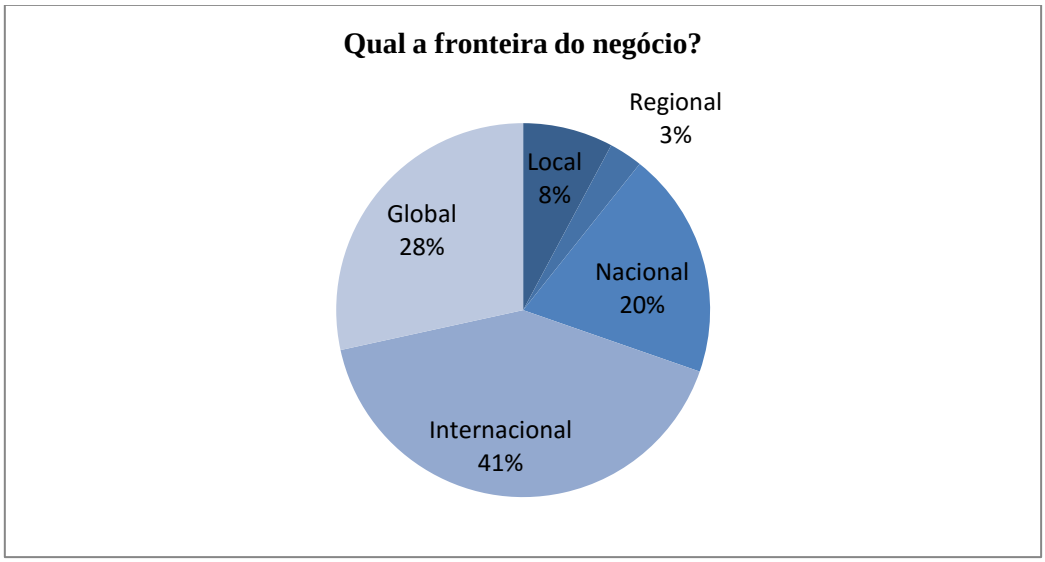
Indústria	Respostas
Serviços	29,4%
Tecnologia	18,6%
Banca e Serviços Financeiros	15,7%
Telecomunicações	7,8%
Saúde e Ciências da Vida	5,9%
Retalho	5,9%
Energia	4,9%
Logística e Transportes	4,9%
Produção Industrial	2,9%
Engenharia e Construção	2%
Hotelaria e Turismo	1%
Desporto	1%

O maior nível de respostas foi obtido no setor de Serviços, Tecnologia e Banca e Serviços Financeiros, respetivamente com 29,4%, 18,6% e 15,6% num total de 63,6%, o que faz com que deva ser estudado em detalhe. É seguido pelo setor das Telecomunicações, Saúde e Ciências da Vida e Retalho com respetivamente 7,8% e 5,9% para os dois últimos, num total de 19,6% no total do estudo.

4.1.2- Mercado de Atuação

Para perceber qual a dimensão do mercado onde atuam as empresas foi perguntado qual a fronteira do negócio desta.

***Figura 5** - Mercado de atuação (% respostas)*

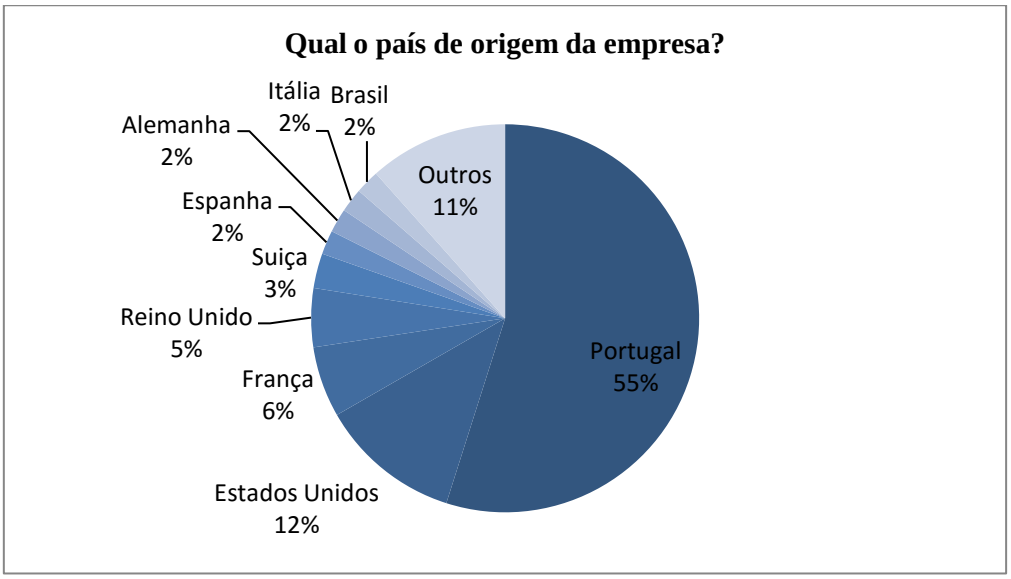


Como podemos ver na Figura 6, a maior parte das empresas atua no mercado Internacional (41%) e Global (28%), o que representa (69%) do total do estudo, ainda assim com 20% que atuam num mercado nacional.

4.1.3- País da Empresa

Para um enquadramento da origem das respostas e empresas foi perguntado no questionário qual o país de origem da empresa.

***Figura 6** - País da Empresa (% respostas)*

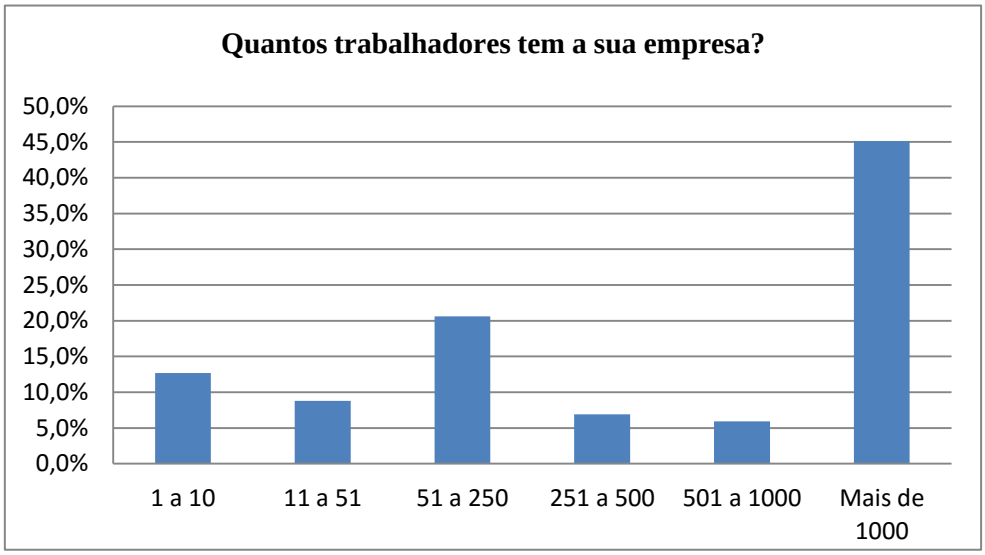


Neste caso, a maioria das empresas do estudo são Portuguesas (55%), sendo que seguidamente as empresas que maioritariamente responderam ao survey são dos Estados Unidos (12%) e da Europa Ocidental (20%).

4.1.4- Dimensão da Empresa

Para perceber a dimensão das empresas que representam o estudo foi perguntado qual o número de trabalhadores que estas têm.

***Figura 7** - Nº de trabalhadores da empresa (% respostas)*

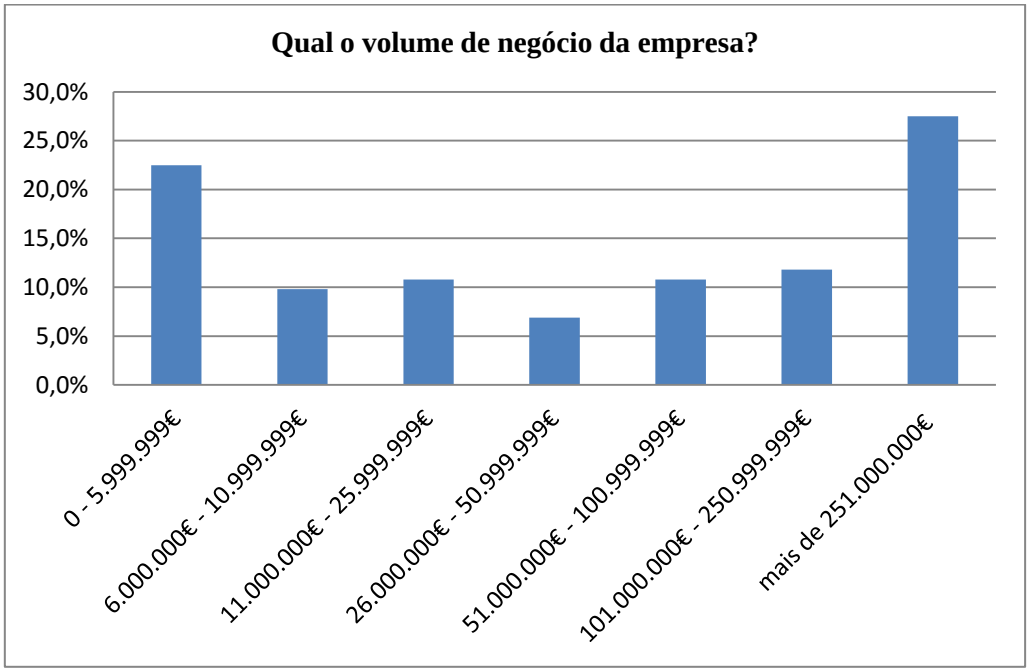


A maior parte das empresas são de grande escala, tendo mais de 1000 trabalhadores em 45% dos casos, no entanto, são seguidas por médias empresas, compostas entre 51 e 250 trabalhadores, representando estas 20,6% do total de respostas.

4.1.5- Volume de Negócio

Para melhor enquadrar a dimensão das empresas também foi perguntado o volume de negócio das mesmas.

Figura 8 *- Volume de negócio da empresa (% respostas)*

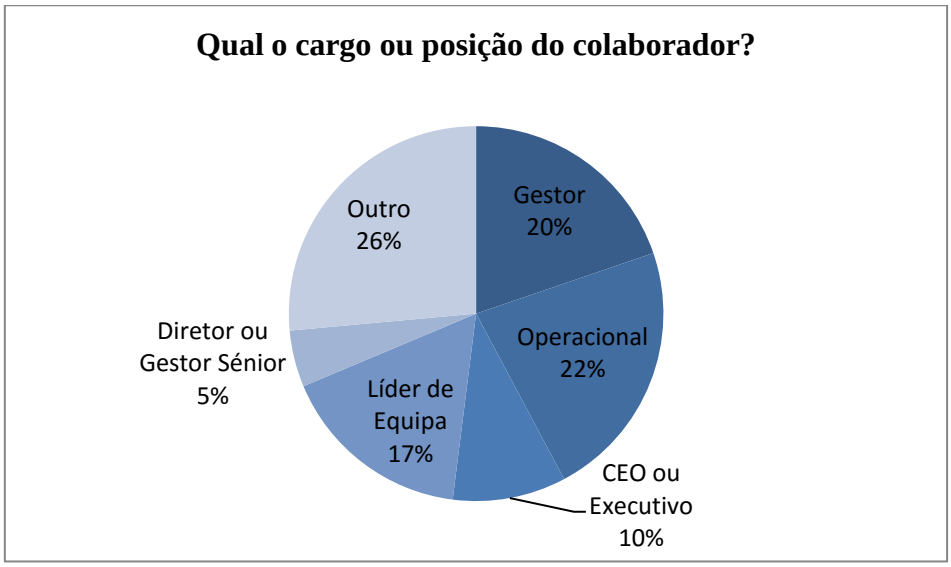


Neste caso as respostas situaram-se nos extremos, de certa forma com algum paralelismo com a quantidade de trabalhadores, 22,5% das empresas faturam até ao limite de 5.999.999€ e 27,5% faturam mais de 251.000.000€. É desta forma permitido um estudo amplo em termos do uso dos dados nas várias dimensões das empresas.

4.1.6- Cargo e Experiência

Para o entendimento das posições que utilizam os dados para influenciar o processo de tomada de decisão foi perguntado qual a posição ou cargo.

Figura 9 *- Posição do colaborador (% respostas)*



Verificando-se neste caso que muitos dos dados que influenciam o processo de tomada de decisão não são analisados por gestores ou altos quadros, uma vez que a resposta a Outro representa 26% e os Operacionais representam 22%. Uma parte significativa são no entanto potenciais decisores, o que engloba diretores ou gestores sénior, gestor, líder de equipa ou CEO ou Executivo (52%).

4.2- Análise de Resultados

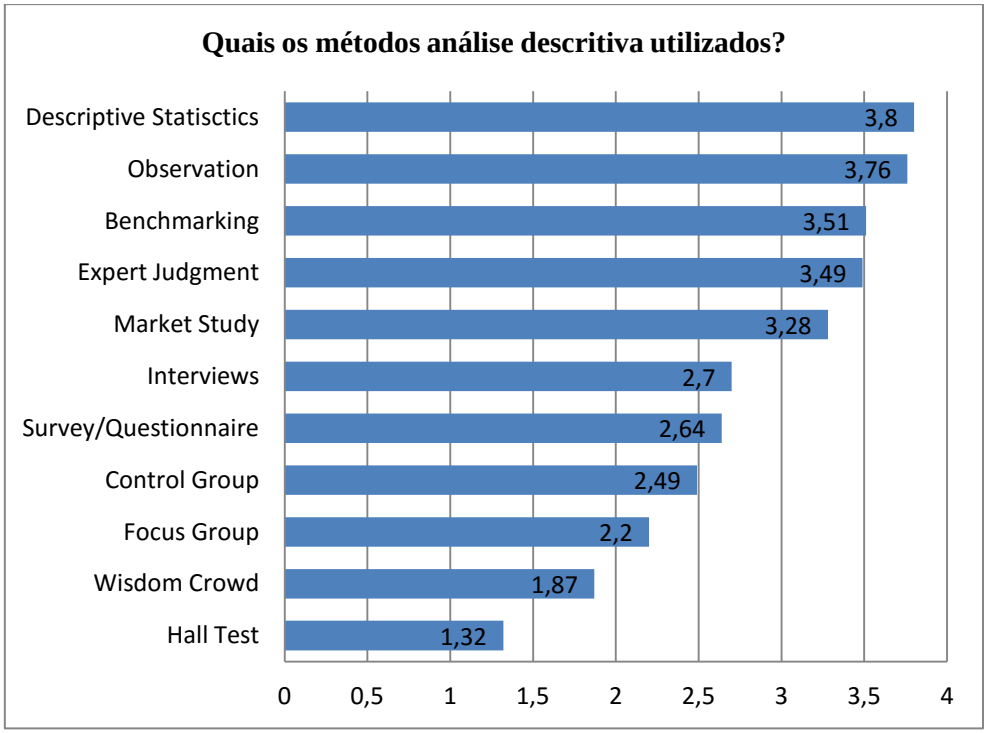
Depois da caracterização da amostra de empresas, o objetivo centra-se na análise dos resultados do questionário e no objetivo central do estudo. A análise será organizada em 3 partes e, em cada uma serão analisadas as questões fundamentais da investigação de uma forma mais detalhada. Os resultados serão analisados segundo a escala descrita e baixo:

- Resposta Nunca: Valor de 1
- Resposta Raramente: Valor de 2
- Resposta com pouca frequência: Valor 3
- Resposta frequentemente: Valor 4
- Resposta com muita frequência: Valor de 5
- Resposta Sempre: Valor de 6
- Resposta Não Sabe: “Don´t Know”

4.2.1- Análise Descritiva

O grupo da Análise Descritiva no questionário contempla 11 técnicas de análise de dados. O objetivo é perceber quais as técnicas mais usadas e menos usadas e perceber se há algum padrão no resultado.

Figura 10 - *Análise descritiva (análise da média)*



Analisando a Figura 10 podemos concluir que metades das técnicas de análise descritiva são usadas raramente, uma vez que os seus valores estão à volta de 2 e que na outra metade dos casos são usados com pouca frequência ou frequentemente.

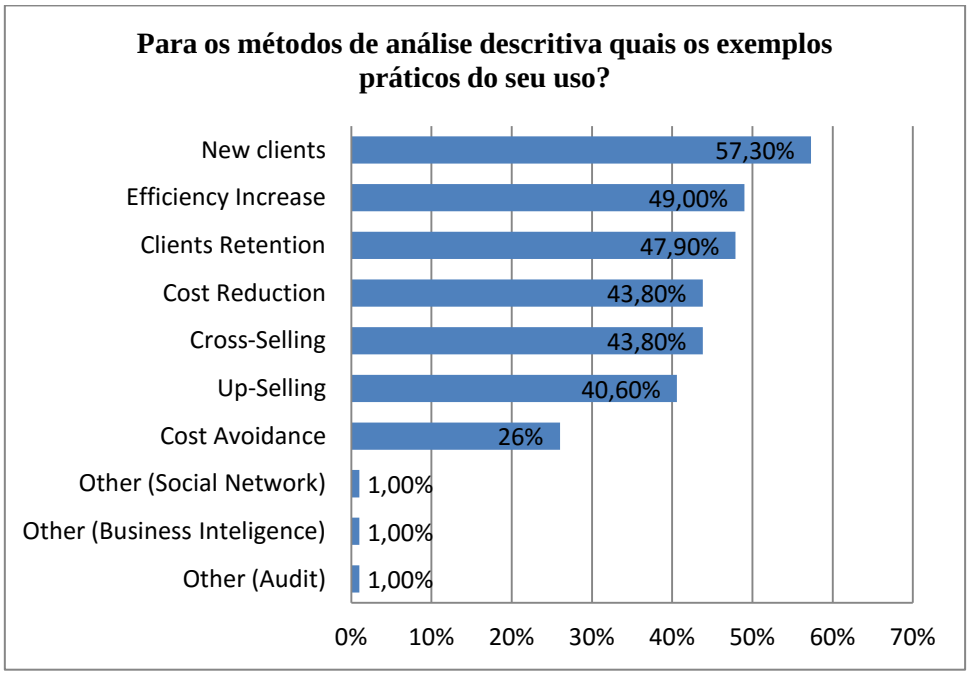
A Estatística Descritiva é, no entanto, a técnica mais usada, com uma pontuação média de 3,8 o que significa que é usada com frequência. É seguida imediatamente pelas Observações com uma pontuação média de 3,76 e pelo *Benchmarking* que é também uma técnica usada frequentemente. Com pouca frequência são ainda usadas as técnicas de *Expert Judgmente*, *Market Study*, *Interviews* e *Survey* com os valores de 3,49, 3,28, 2,7 e 2,64 respetivamente. Por outro lado, temos as técnicas que nunca são usadas ou raramente o são, como é o caso do *Control Group*, *Focus Group*, *Wisdom Crowd* ou *Hall Test* com valores de (2,49), (2,2), (1,87) e (1,32) respetivamente.

É interessante perceber desta análise que as técnicas mais utilizadas são aquelas que se baseiam em dados gerais com uma amostra maior, enquanto que quando a amostrada é mais restrita ou focada num grupo as técnicas são menos usadas. Por outro lado, os dados cingidos a um grupo mais restrito e mais focado são os menos utilizados na análise descritiva, talvez por não permitirem uma tão grande abrangência nem tão grande grau de comparação.

4.2.1.1- Objetivos dos métodos de análise descritiva utilizados

Para perceber o objetivo das técnicas de análise descritiva foi ainda feito um questionário com 8 respostas, sendo que a resposta *Other* podia ser definida pelo participante no estudo.

Figura 11 *- Objetivo dos métodos descritivos (% respostas)*



Analisando as respostas, podemos perceber claramente que mais de metade das pessoas usa as técnicas anteriores para conquistar Novos Clientes (57,3%) e desta forma criarem valor para a sua empresa, metade para Aumentar Eficiência ou Reter Clientes, 49% e 47,9% respetivamente, onde podem ser dados exemplos como de otimização de stock, otimização de rotas ou marketing direto e quase metade para Redução de Custos, fazer *Cross-Selling* ou *Up-Selling* (43,8%, 43,8% e 40,6%). Com menos expressão (26%) surge a Prevenção de Custos, sendo, por isso, as técnicas de análise descritiva são pouco usadas para criar por exemplo modelos de risco ou deteção de fraude.

Com pouco significado (1%) surgem as Redes Socias, o *Business Intelligence* e a Auditoria a utilizar as técnicas de análise descritiva.

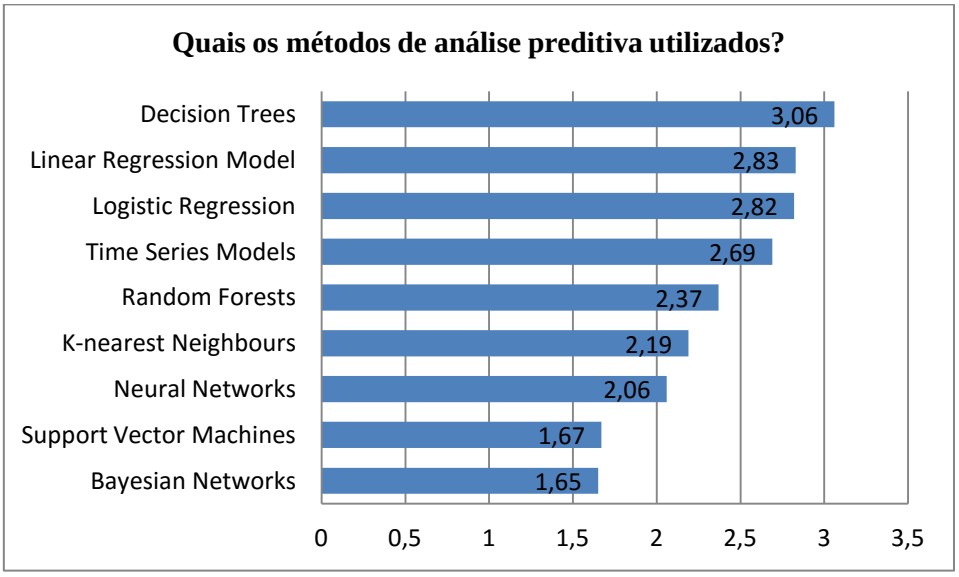
Neste ponto, é interessante perceber a importância de aumentar clientes para a criar valor ao negócio, e portanto os temas relativos a performance dos produtos e serviços são dados bastante importantes para as empresas e para os analisar recorrem às técnicas de análise descritiva de dados, bem como para as questões relacionadas com otimização e eficiência negócio, sendo portanto importante descrever os dados passados e presentes para tomar decisões futuras no negócio. Estas técnicas são também importantes para analisar meios de produção e fazer a sua mais correta manutenção bem como fazer recomendação de produtos para determinados segmentos de clientes.

Por outro lado, não são as técnicas mais usadas para criar modelos de risco e deteção de fraude, pois apenas descrevem acontecimentos presentes e passados que podem não indicar nada de relevante no futuro, o que leva a crer que os métodos descritivos na análise de dados históricos não permitem da melhor forma prever falhas no futuro.

4.2.2- Análise Preditiva

O grupo da Análise Preditiva no questionário contempla 9 técnicas de análise de dados, técnicas que são usadas não só para entender e descrever estes historicamente, como também para fazer previsão. O objetivo é perceber quais as técnicas mais usadas e menos usadas e perceber se há algum padrão no resultado.

Figura 12 - *Análise preditiva (análise da média)*



Analisando a Figura 12 podemos concluir que metade das técnicas de análise preditiva são usadas com pouca frequência, uma vez que os seus valores estão entre 2,5 e 3 e que na outra metade dos casos são usados raramente.

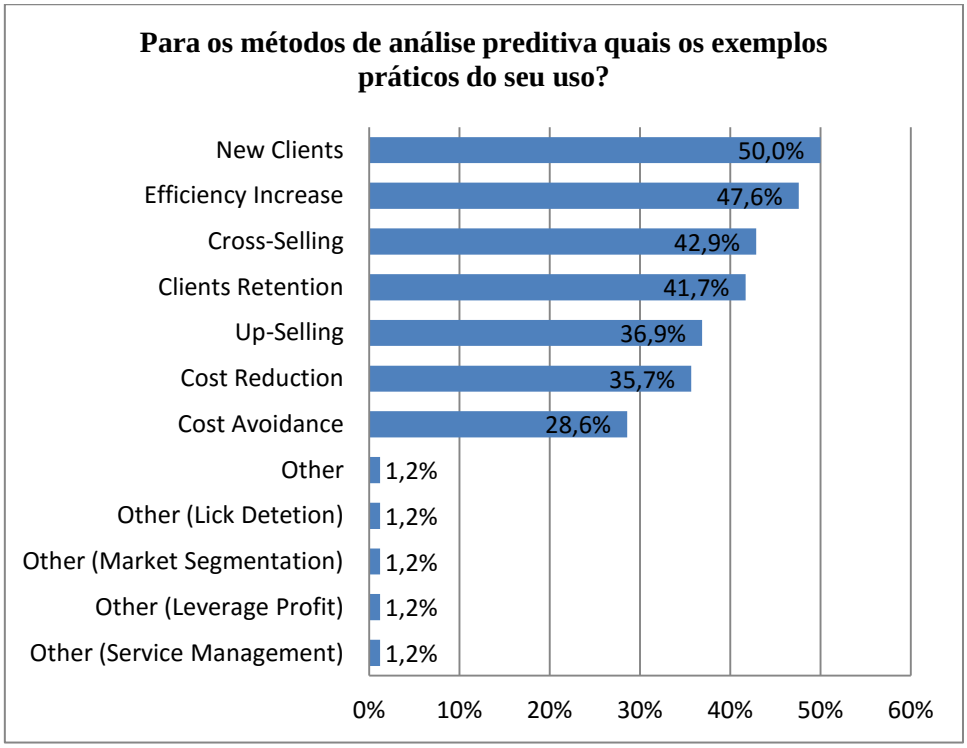
As Árvores de Decisão são, no entanto, a técnica mais usada, com uma pontuação média de 3,06 o que significa que é usada com pouca frequência. É seguida imediatamente pelos Modelos de Regressão Linear com uma pontuação média de 2,83 e pela Regressão Logística que é também uma técnica usada pouco frequentemente com 2,82. Com pouca frequência é ainda usada a técnica de *Time Series Model* com o valor de 2,69. Por outro lado, temos as técnicas que raramente são usadas, como o caso das *Random Forest*, *K-nearest Neighbours*, *Neural Networks*, *Support Vector Machine* e *Bayesian Networks* com respetivamente (2,37), (2,19), (2,06), (1,67) e (1,65) valores de média.

Desta análise podemos entender que as técnicas preditivas ainda são usadas pouco frequentemente ou mesmo raramente. Ainda assim podemos afirmar que as técnicas mais simples do ponto de vista da resolução e utilização são as mais utilizadas, como o caso das Árvores de Decisão ou os Modelos de Regressão Linear, cujos dados depois de recolhidos são mais fáceis de projetar para decisões futuras. Por outro lado, as técnicas que exigem o recurso computacional e maior tecnologia na recolha dos dados são as menos usadas.

4.2.2.1- Objetivo dos métodos de análise preditiva utilizados

Para as técnicas de análise preditiva foi ainda feito um questionário com 8 respostas, sendo que a resposta *Other* podia ser detalhada por quem responde, com o intuito de perceber o objetivo das técnicas usadas.

Figura 13 *- Objetivo dos métodos preditivos (% respostas)*



Analisando as respostas podemos perceber que exatamente metade das pessoas usa as técnicas anteriores para conquistar Novos Clientes (50,0%), sendo este o objetivo principal na criação de valor, e que pode ser conseguido por exemplo com otimização de preço ou previsão de receitas, mas não só. Quase metade faz previsão dos dados para Aumentar Eficiência (ex: otimização de stocks ou de rotas), fazer *Cross-Selling* (ex: sistemas de recomendação) ou Reter Clientes (também através de sistemas de recomendação) com os seguintes valores médios 47,6%, 42,9% e 41,7% respetivamente. Ainda com alguma expressão (36,9% e 35,7% respetivamente) surge com o intuito de fazer *Up-Selling* ou Redução de Custos, conseguindo também, através de por exemplo sistemas de recomendação para o primeiro e manutenção preventiva para o segundo caso. Com menos expressão e com um valor médio de 28,6% está a Prevenção de Custos através de modelos de risco ou deteção de fraude. Com pouca expressão (1,2%) foram dadas outras repostas pelos participantes no estudo que indicaram que usavam as técnicas de análise preditiva com o objetivo de *Lick Detetion*, *Market Segmentation*, *Leverage Profit* ou *Service Management*.

É de notar, mais uma vez, que as empresas procuram analisar dados relativos a preços e receitas e, que para o fazer usam técnicas preditivas, que tentam extrapolar para o futuro os acontecimentos presentes ou passados com o objetivo de tomar decisões de gestão mais acertadas. Procuram também analisar, desta forma, os dados relacionados com otimização e eficiência de negócio ou mesmo sistemas de recomendação para aumentar as margens.

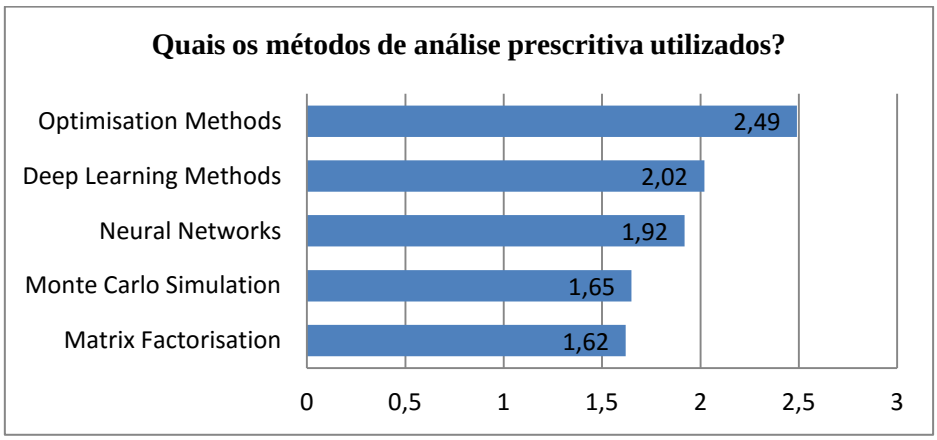
Por outro lado, e mais uma vez, não são as técnicas mais usadas para criar modelos de risco e deteção de fraude, neste caso também pode ser entendido que o conceito de Risco ou de Falha não é um ponto importante na análise de dados, pois neste caso o efeito de previsão permitido por esta técnica poderia trazer importantes insights para os negócios na sua gestão de risco e de falhas.

De certa forma, os analistas procuram no geral utilizar esta técnica no sentido de criar mais volume de negócio e melhorar as margens e só por fim estão preocupados com os riscos e falhas.

4.2.3- Análise Prescritiva

O grupo da Análise Prescritiva no questionário contempla 5 técnicas de análise de dados históricos não só para fazer previsão, mas também gerar múltiplos cenários e prescrever as melhores soluções. O objetivo é perceber quais as técnicas mais usadas e menos usadas e perceber se há algum padrão no resultado.

Figura 14 *- Análise prescritiva (análise da média)*



Analisando a Figura 15 podemos concluir que todas as técnicas de análise preditiva são usadas com pouca frequência em geral e os valores centram-se entre 1,62 e 2,49.

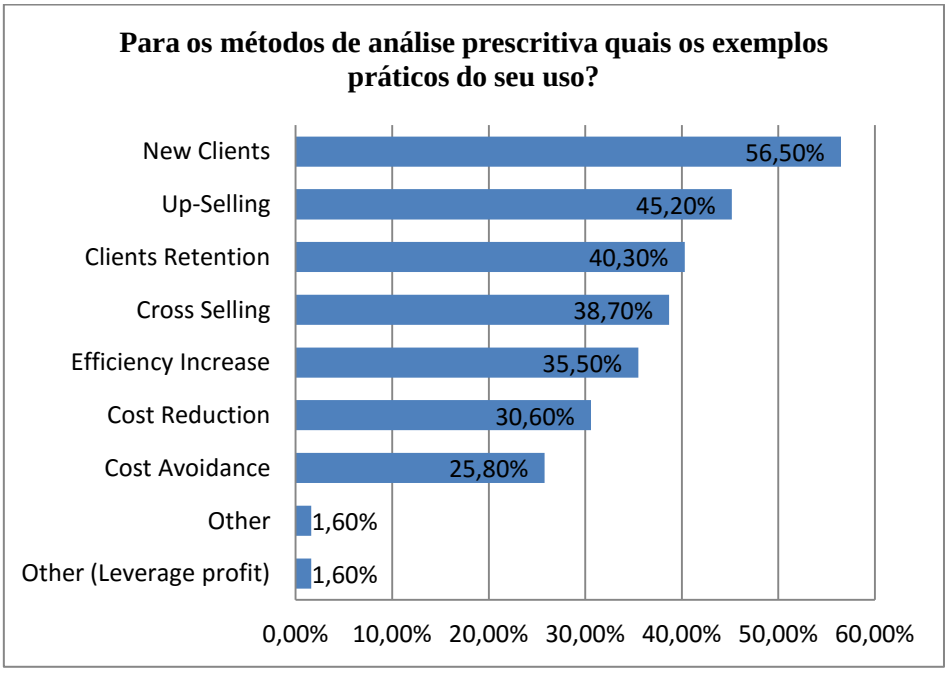
Os métodos de otimização são ainda assim a técnica mais usada com um valor médio de 2,49. É seguida pelos *Deep Learning Methods* com uma pontuação média de 2,02 e, por Neural Networks que é também uma técnica usada pouco frequentemente com 1,92. Por fim, e com menos uso ainda estão as *Monte Carlo Simulation* e *Matrix Factorization* com 1,65 e 1,62 respetivamente.

É importante analisar que o Método de Otimização, por talvez ser o método mais simples na prescrição de uma solução em função de várias simulações, tendo em conta um dado e problema específico, é utilizado mais amplamente. Nas outras técnicas os dados são mais amplos e as soluções prescritas tornam-se também mais vagas.

4.2.3.1- Objetivo dos métodos de análise prescritiva utilizados

Para as técnicas de análise prescritiva foi também feito um questionário com 8 respostas, à semelhança dos grupos anteriores com o intuito de perceber o objetivo das técnicas usadas.

Figura 15 *- Objetivo dos métodos prescritivos (% respostas)*



Podemos ver, mais uma vez, que a conquista de novos clientes é novamente o objetivo da utilização dos métodos anteriores, com 56,5% das respostas a afirmar esse facto. Seguindo-se ainda com algum impacto pelo objetivo de fazer *Up-Selling*, Retenção de Clientes, *Cross-Selling* e Aumento da Eficiência com respetivamente 45,2%, 40,3%, 38,7% e 35,5% das respostas a indicar isso. Por fim, a Redução de Custos ou a Prevenção de Custo não é um dos objetivos principais na utilização das técnicas já referenciadas e com muito pouca expressão (1,6%) vem como resposta Outro ou *Leverage Profit*.

À semelhança do que já foi analisado anteriormente, as técnicas centram-se essencialmente no aumento do negócio ou na margem deste e pelo contrário estão menos preocupados com as potências falhas ou riscos.

4.3- Análise Comparativa

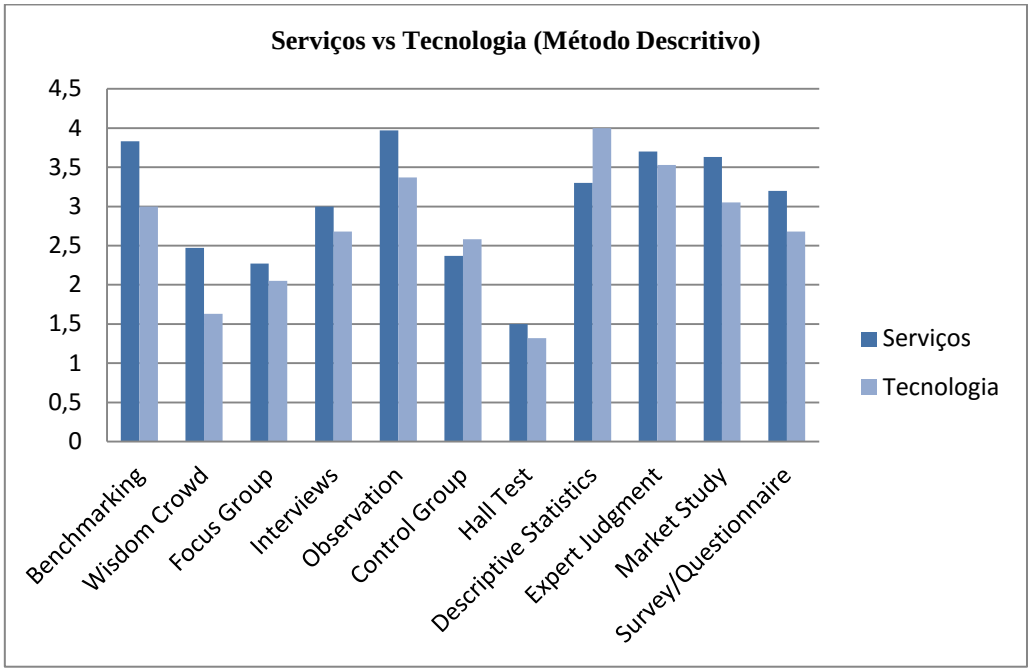
Para aumentar o detalhe da investigação, neste grupo será feita uma análise comparativa entre os setores com mais impacto no estudo. Esta análise comparativa será feita para os vários métodos de análise de dados (descritivos, preditivos e prescritivos).

4.3.1- Comparação de Setores

A análise comparativa será feita aos setores que tiveram mais impacto no estudo e que deram um maior número de respostas, neste caso o setor dos Serviços e o setor Tecnológico.

4.3.1.1- Método Descritivo

Figura 16 *- Serviços vs Tecnologia (Método Descritivo)*

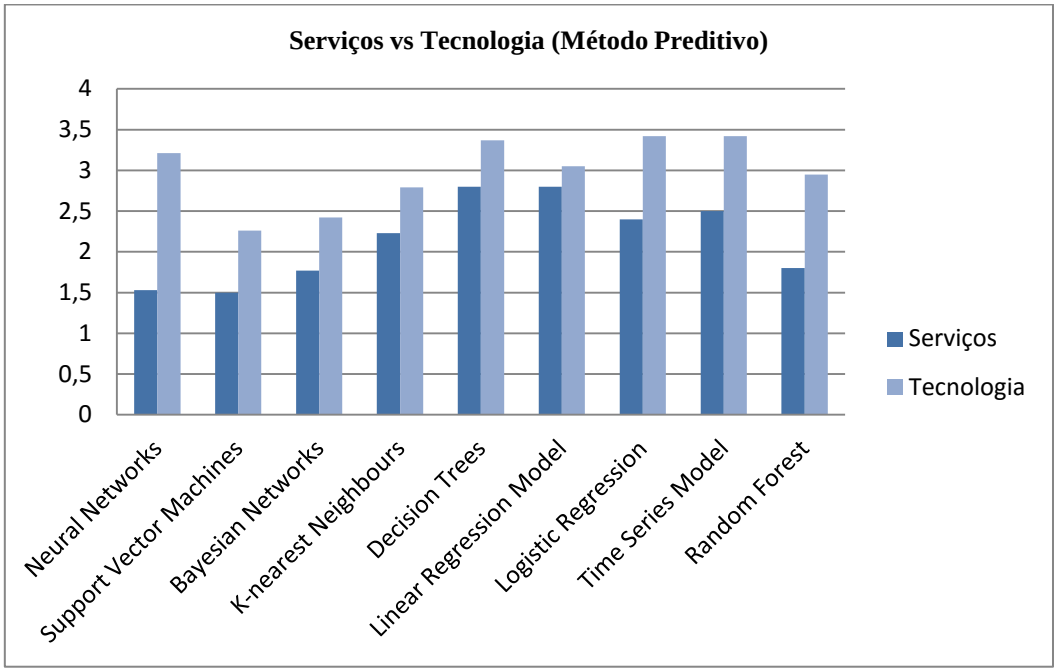


Comparando as duas indústrias podemos ver que, em geral existe uma relação nas técnicas usadas, com exceção para o *Benchmarking* e *Wisdom Crowd* que tem uma diferença mais significativa a tender para os Serviços e, a Estatística Descritiva a tender para o setor Tecnológico. No entanto, as técnicas que são usadas frequentemente em média são Expert Judgment, com médias acima de 3,5 para o setor dos Serviços e Tecnológico, e no o caso do *Benchmarking*, *Observation* e *Market Study* só para o setor dos Serviços; e Estatística Descritiva só para o setor Tecnológico.

O facto de o setor dos serviços dar mais enfase ao *Benchmarking* e *Wisdom Crowd* leva a crer que normalmente opta por modelos matemáticos mais simples e com menores recursos computacionais, baseando-se mais em evidências. Por outro lado, o setor Tecnológico ao destacar a Estatística Descritiva opta por modelos matemáticos mais sofisticados, bem como a necessidade de recorrer a recursos computacionais mais avançados para dar resposta.

4.3.1.2- Método Preditivo

Figura 17 - *Serviços vs Tecnologia (Método Preditivo)*



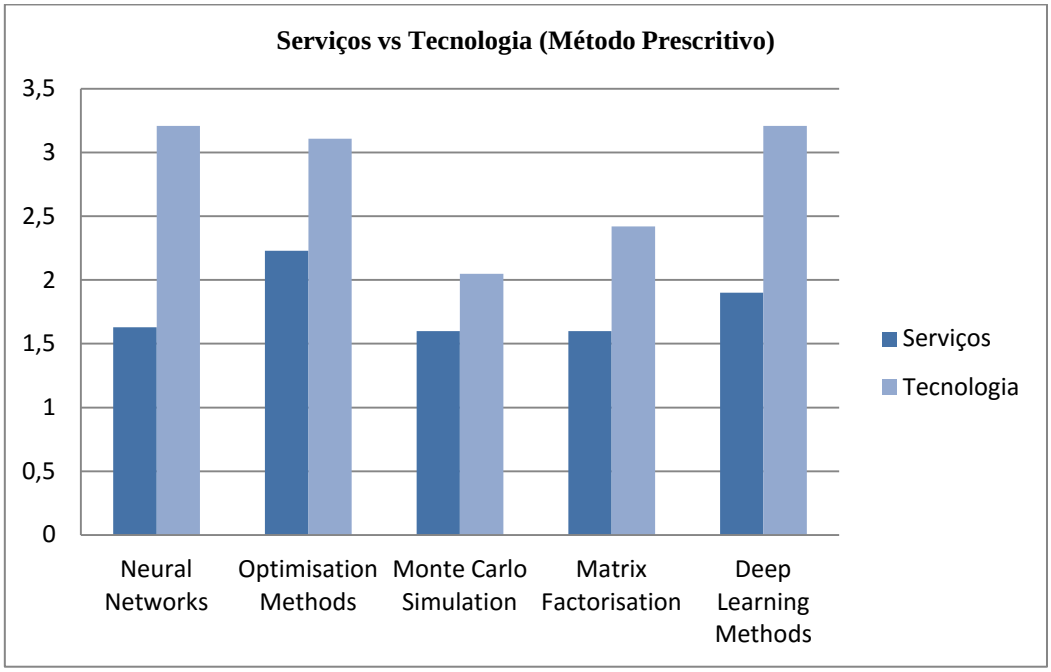
No caso do Método Preditivo denota-se em média uma maior utilização das técnicas por parte do setor tecnológico, com maior preponderância no caso de *Neural Networks*, *Logistic Regression*, *Time Series Model* e *Random Forest*.

No entanto, nenhuma das técnicas chega a ser usada frequentemente em média (maior que 3,5) e só mesmo algumas são usadas pouco frequentemente (mais de 2,5) como é o caso de *Neural Networks*, *K-nearest Neighbours*, *Logistic Regression* e *Random Forests* só no setor Tecnológico e *Decision Trees*, *Regression Model* e *Time Series Model* nos dois casos.

Aqui nota-se claramente que as técnicas preditivas, que recorrem a modelos matemáticos mais complexos bem como à necessidade de maior capacidade computacional, são usadas em média com mais frequência pelo setor Tecnológico. Ainda assim, as duas técnicas usadas em média com maior frequência pelo setor dos Serviços, como o caso das Árvores de Decisão e dos Modelos de Regressão, são as que dependem dos modelos matemáticos menos complexos e mais fáceis de cálculo.

4.3.1.3- Método Prescritivo

Figura 18 *- Serviços vs Tecnologia (Método Prescritivo)*



O método prescritivo apresenta também uma maior apropriação em média por parte do setor Tecnológico com maior amplitude no caso de *Neural Networks* e *Deep Learning Methods*.

Embora nenhum chegue em média a ser usado frequentemente (mais de 3,5), podemos considerar que pelo menos *Neural Networks, Optimisation Methods* e *Deep Learning Methods* são usados em média pouco frequentemente (mais de 2,5) pelo setor Tecnológico. Não podemos, no entanto, dizer que em média nunca é usado nenhum dos métodos por parte dos dois setores.

Estas técnicas, para além de recorrerem a modelos matemáticos, estatísticos e algorítmicos mais avançados, com forte componente computacional, o que torna o seu uso mais frequente no setor Tecnológico, também exige uma forte componente de engenharia. Este método destaca-se pela forte necessidade de recorrer a engenheiros para poder tirar partido da sua utilização nos vários dispositivos e máquinas, o que nos remete invariavelmente para importância de cada vez mais criar uma ponte entre a gestão e a engenharia. Só assim o recurso à tecnologia poderá trazer mais insights e garantir maior capacidade de prescrever decisões com base nas simulações criadas. Para além disso, é importante perceber que a grande quantidade de dados está a fazer com

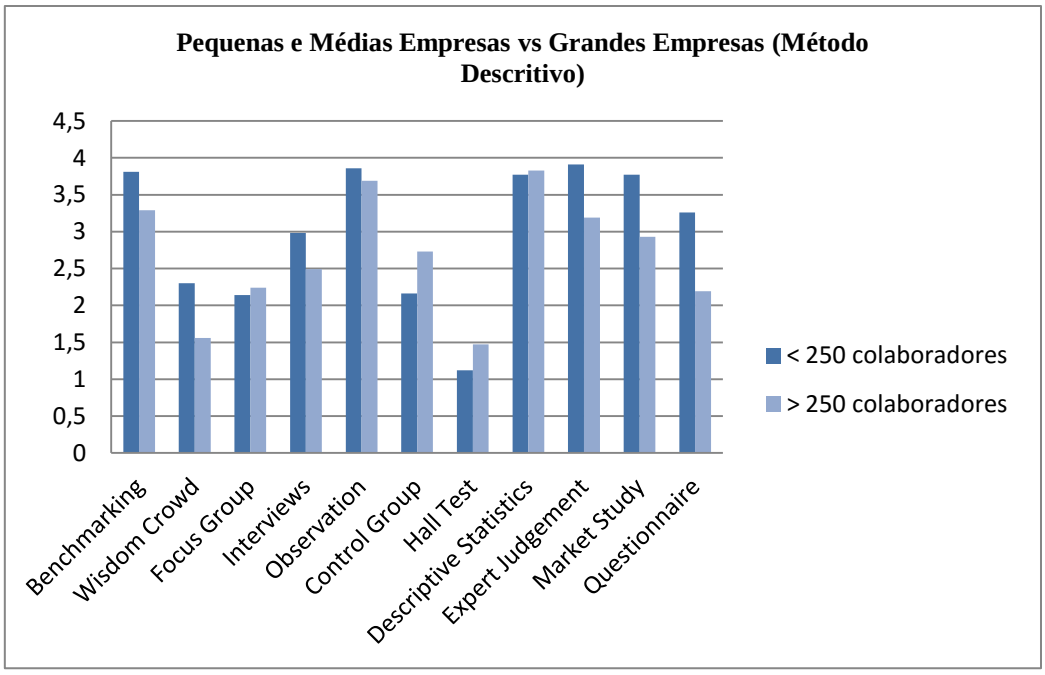
que surjam empresas de tecnologia que estão a desenvolver estratégias nesse nicho de mercado.

4.3.2- Comparação por Dimensão das Empresas

A análise comparativa pela dimensão das empresas será feita com base no número de colaboradores. Foram consideradas por um lado as Pequenas e Médias empresas que contam com até 250 colaboradores e as grandes empresas com mais de 250 colaboradores.

4.3.2.1- Método Descritivo

Figura 19 *- Pequenas e Médias Empresas vs Grandes Empresas (Método Descritivo)*

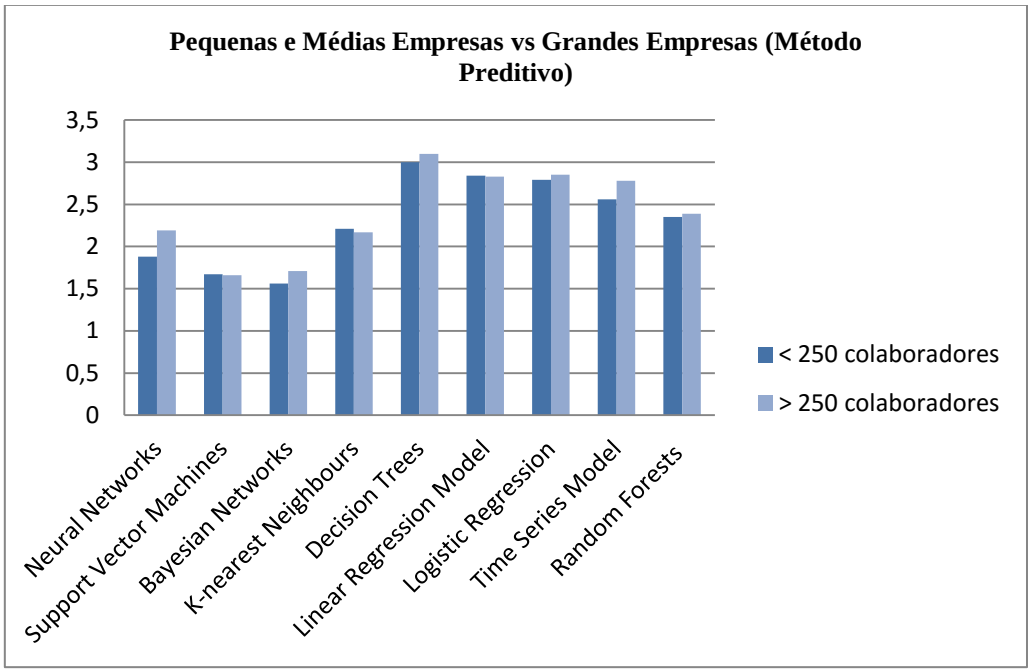


Podemos ver que as pequenas e médias empresas na maioria dos casos usam os métodos descritivos, em média mais frequentemente do que as empresas grandes, apenas nas técnicas de *Focus Group*, *Control Group*, *Hall Test* e *Descriptive Statistics* isso se inverte. O facto de ser um método com maior facilidade de acesso a dados e, tendo as empresas mais pequenas, por sua vez menores recursos financeiros, faz com que isto aconteça. Por outro lado, também se nota a maior necessidade de as empresas grandes usarem técnicas de análise descritiva mais focadas em pequenos grupos ou focos, talvez pela sua dimensão e por terem um contacto mais amplo com os clientes, necessitem de focar mais a sua atenção em pequenos nichos. No caso das pequenas empresas

conseguem do ponto de vista da comunicação estar mais próximas dos clientes e mais rapidamente tirar esses inputs.

4.3.2.2- Método Preditivo

Figura 20 *- Pequenas e Médias Empresas vs Grandes Empresas (Método Preditivo)*



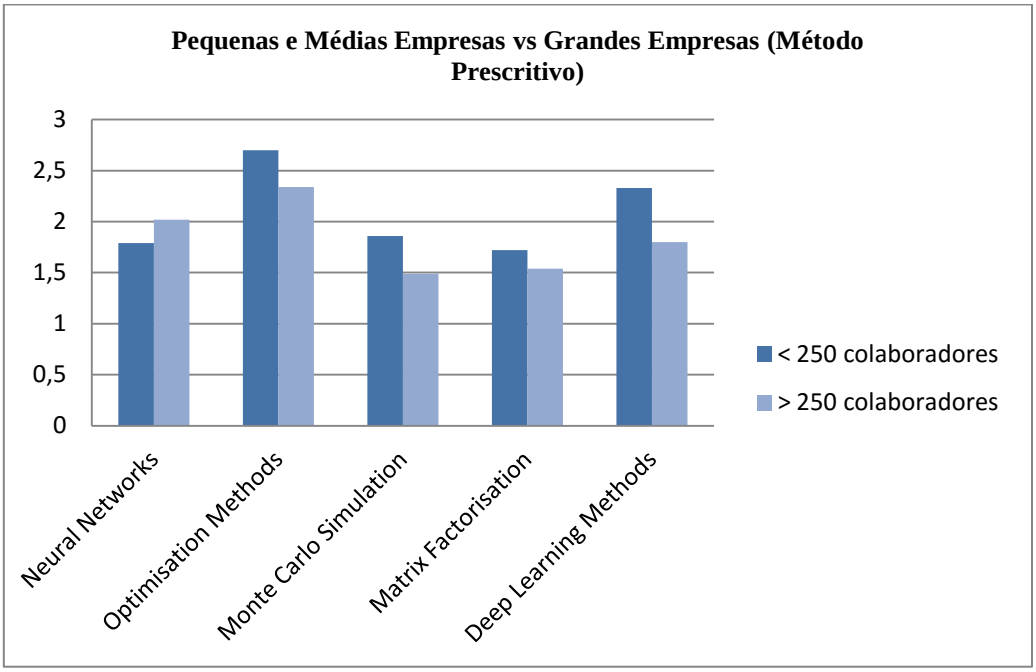
É interessante analisar que com o método descritivo a tendência inverte-se e a maior parte das grandes empresas tendem em média a analisar dados com técnicas preditivas mais frequentemente.

Talvez porque por um lado consigam ter maior capacidade de investimento financeiro e capital humano para o fazer e, por outro lado porque o facto de terem uma rede de clientes maior permite recolher mais dados e enriquecer as redes e dispositivos que permitem fazer previsão. A exceção é feita para o modelo de regressão linear e *K-Nearest Neighbours*, talvez por serem mais simples e não necessitarem de uma rede de informação tão ampla para serem usados.

No entanto, é de salientar que a diferença de utilização em média deste método não é significativo, o que leva a crer que independentemente da dimensão das empresas, existe a noção que a análise de dados é um tema importante para o processo de tomada de decisão das empresas.

4.3.2.3- Método Prescritivo

Figura 21 *- Pequenas e Médias Empresas vs Grandes Empresas (Método Prescritivo)*



No Método de análise Prescritivo constata-se que na maior dos casos, as técnicas prescritivas são usadas na maior parte dos casos por empresas mais pequenas. Isto demonstra que, por um lado as empresas mais pequenas estão atentas ao fenómeno Big Data e sabem que isso pode ser uma vantagem competitiva para elas, por outro lado, porque tendo em conta que a maior parte do estudo se centrou no setor dos serviços e tecnologia, estas empresas mais pequenas podem estar a apostar em nichos de mercado em análise de dados.

Não obstante é interessante constatar que o fenómeno *Big Data* e a utilização de Métodos Preditivos é transversal à dimensão das empresas.

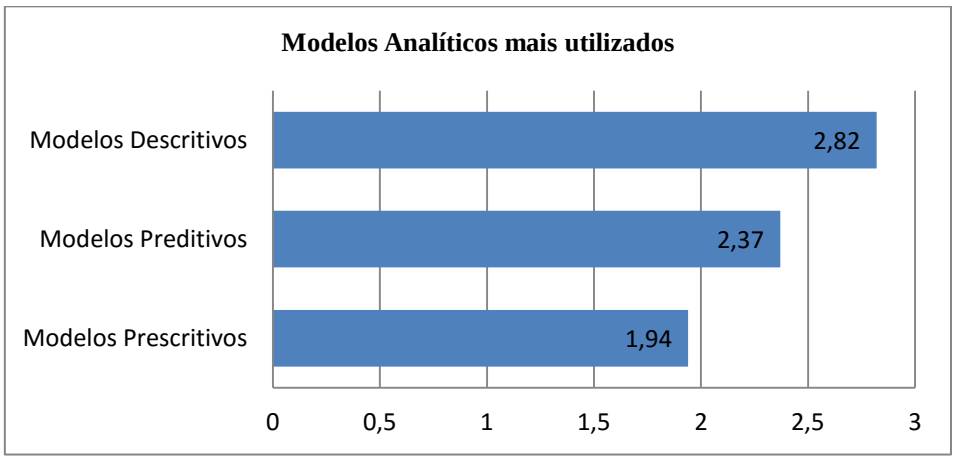
4.4- Visão Geral

Esta análise é feita de forma a ter uma visão geral dos resultados com objetivo de perceber quais os métodos (descritivos, preditivos ou prescritivos) e técnicas de análise de dados mais utilizados e os principais objetivos da sua análise.

4.4.1- Métodos de Análise de Dados Mais Utilizados

A partir da Figura 23 podemos perceber por ordem decrescente os modelos de análise de dados utilizados. Sendo que os Métodos Descritivos são usados frequentemente e os Métodos Preditivos e Prescritivos são usados pouco frequentemente.

***Figura 22** – Métodos analíticos mais usados (análise da média)*



O que pode ser importante perceber desta análise é que a dificuldade dos modelos preditivos e prescritivos bem como a necessidade de recursos mais avançados para a recolha e análise de dados faz com isto possa suceder.

4.4.2- Técnicas de Análise de Dados Mais Utilizadas

Para cada método (Descritivo, Preditivo e Prescritivo) existe uma técnica de análise de dados que é mais utilizada consoante descrito na tabela seguinte (Tabela 7).

***Tabela 7** - Técnica de análise mais utilizada por Modelo (análise da média)*

Método Analítico	Técnica	Média
Método Descritivo	Descriptive Statistics	3,8
Método Preditivo	Decision Trees	3,06
Método Prescritivo	Optimisation Methods	2,49

A Estatística Descritiva é a técnica de análise de dados mais utilizada no Modelo Descritivo e é usada frequentemente (3,8), já no Modelo Preditivo a técnica mais utilizada é a Árvore de Decisão e é usada com pouca frequência (3,06). Por fim, no Método Descritivo recorre-se com maior frequência aos Métodos de Otimização, ainda

assim utilizados raramente (2,49). Em geral, podemos ver que as técnicas utilizadas são talvez as mais acessíveis e de fácil utilização e que, possivelmente, continuam a garantir, apesar da sua simplicidade, um importante contributo para a análise do negócio e a sua gestão. Por outro lado, também pode sugerir que ainda não existem profissionais suficientemente formados que possam tirar partido das técnicas mais avançadas e na criação de redes mais complexas para a análise de dados.

4.4.3- Objetivo Principal das Técnicas de Análise de Dados

Para cada técnica de análise de dados existe um objetivo principal com descrito na tabela em baixo.

Tabela 8 - *Técnica de análise mais utilizada e seu objetivo (análise da média)*

Técnica	**Técnica**	**%**
Descriptive Statistics	New Clients	57,3%
Decision Trees	New Clients	50,0%
Optimisation Methods	New Clients	56,5%

O objetivo principal das técnicas utilizadas é a conquista de Novos Clientes, para todas, considerando para isso modelos de otimização de preço ou de previsão de receita. Para cada técnica maioritariamente utilizada (Estatística Descritiva, Árvores de Decisão e Métodos de Otimização) 57,3%, 50,0% e 56,5% respetivamente, responderam *New Clients*. Isto acontece porque, possivelmente, o maior interesse na gestão do negócio é fazê-lo crescer e ganhar cota de mercado e só depois controlar custo, eficiência e mitigar riscos.

4.5- Validação das Perguntas de Pesquisa

Neste ponto, já estamos em condições de responder às questões que orientaram este questionário e ao: **Como e para quê analisamos os dados?**

A primeira questão é: **Quais as técnicas de análise descritiva de dados mais usadas?** Se analisarmos em média qual técnica é mais utilizada, esta é a estatística descritiva, usada frequentemente. Daqui advém a pergunta: **Para que fim as empresas utilizam técnicas de análise descritiva?** Decorrente da análise temos que o objetivo principal, de acordo com a criação de valor no Pereira's Diamond, é aumentar o negócio, mais concretamente através de novos clientes.

A segunta questão é: **Quais as técnicas de análise preditiva de dados mais usadas?** Se analisarmos em média qual técnica é mais utilizada, esta é a Árvore de Decisão, usada ainda assim pouco frequentemente. Daqui advém a pergunta: **Para que fim as empresas utilizam técnicas de análise preditiva?** Decorrente da análise temos que o objetivo principal é também aumentar o negócio com novos clientes.

A terceira e última questão é: **Quais as técnicas de análise prescritiva de dados mais usadas?** Se analisarmos, em média, a técnica mais utilizada são os modelos de Otimização, usados ainda assim pouco frequentemente. Daqui podemos perguntar: **Para que fim as empresas utilizam técnicas de análise prescritiva?** Decorrente da análise temos que o objetivo principal é a conquista de novos clientes mais uma vez, cujo impacto se repercute no aumento do negócio.

PARTE 5 - CONCLUSÕES

5.1 – Conclusões Principais

O objetivo da dissertação é perceber a forma como as empresas analisam os dados, com que finalidade os utilizam e de que forma os dados estão a influenciar os processos de tomada de decisão.

De acordo com os métodos de análise que dispomos, descritivos, preditivos e prescritivos, e de acordo com os seus objetivos segundo o Pereira´s Diamond podemos concluir que:

- A maior parte das organizações recorre mais frequentemente às técnicas de análise de dados descritivas para a tomada de decisões;
- Pouco frequentemente, os analistas recorrem a técnicas de análise preditivas para poderem tomar decisões com base na previsão dos acontecimentos ou situações;
- A utilização de técnicas prescritiva, com capacidade de simular cenários e prescrever a solução com base no melhor cenário previsível é ainda pouco frequente;
- O principal objetivo na utilização dos vários métodos de análise de dados centra-se na conquista de novos clientes, e consequentemente o aumento do negócio, criando assim valor para a organização;
- O objetivo último na análise de dados é a prevenção de custos, e portanto estando menos preocupadas as empresas com a eficiência do negócio;
- As técnicas mais usadas pelos analistas de dados são a Estatística Descritiva, Observação e Benchmarking, todos eles pertencentes aos Métodos de Análise Descritiva;
- Para cada Método de Análise (Descritivo, Preditivo e Prescritivo) as técnicas mais usadas são a Estatística Descritiva, Árvores de Decisão e os Métodos de Otimização;
- As técnicas menos utilizadas, são por sua vez *Hall Test*, *Matrix Factorisation*, *Monte Carlo Simulation* e *Bayesian Networks*, as três últimas por necessidade de colaboradores mais qualificados e de recursos tecnológicos mais avançados;

- Nos setores que maior contribuiram para o estudo, o setor Tecnológico é o que utiliza com maior frequência os Métodos de Análise Preditiva e Prescritiva;
- Existe uma relação entre as técnicas que são mais utilizadas em cada método analítico, quer para as pequenas e médias empresas, quer para as grandes empresas;
- Os métodos analíticos descritivos e prescritivos são mais usados pelas pequenas e médias empresas, enquanto que apenas o método preditivo é mais usado pelas grandes empresas.

Para concluir, é importante mencionar que a análise de dados é um tema tranversal à dimensão das empresas, o que demonstra a sua importância para estas. No entanto o seu potencial de previsão e prescrição ainda não está a ser aproveitado, bem como o facto da sua utilização não englobar todos os tópicos da criação de valor para as organizações.

5.2 – Limitações ao estudo

É possível considerar que este estudo descritivo utilizou uma metodologia que permite retirar conclusões válidas. No entanto existem sempre algumas limitações que são importantes mencionar.

A primeira limitação é o facto da amostra usada estar limitada aos linkedin e à rede de contactos pessoais, o que não garante que o estudo seja mais abrangente e representantivo.

A segunda limitação está associada ao facto do número de respostas não permitir que se tomem considerações a uma maior escala sobre o tema do *Data Science*.

A terceira limitação está associada ao facto do estudo ser feito com base num resumo dos métodos e técnicas de análise de dados, que para além de estarem em constante evolução, é impossivel abranger na totalidade.

5.3 – Pesquisa Futura

Tendo em conta a amplitude de temas que a temática do *Data Science* engloba, gostaria de sugerir alguns tópicos interessantes para se explorarem no futuro e que podem completar este estudo:

- Uma vez que este estudo se centra nos métodos e objetivos da análise de dados, seria interassante analisar quais as plataformas tecnológicas e software utilizados pelas empresas para gerirem estes.
- Também seria interessante perceber quais as plataformas e dispositivos mais usados e valorizados para extrair dados dos consumidores e empresa.
- Tendo em conta uma escala global, seria interessante fazer um estudo que permitisse uma maior abrangência do tema em causa.

REFERÊNCIAS BIBLIOGRÁFICAS

Abo-Hamad, W., & Arisha, A. (2011). Simulation–Optimisation Methods in Supply Chain Applications: A Review. *Irish Journal of Management , 30* (2), 95-124.

Arlot, S., & Genuer, R. (2016). Comments on: A Random Forest Guided Tour. *TEST , 25* (2), 228-238.

Awad, E., & Ghaziri, H. (2004). *Knowledge Management* (1st edition ed.). Englewood Cliffs: Prentice Hall.

Ayala, F. (2009). Darwin and the Scientific method. *Proceedings of the National Academy of Sciences of the United States of America , 106*, 10033-10039.

Azari, T., Samani, N., & Mansoori, E. (2015). An artificial neural network model for the determination of leaky confined aquifer parameters: an accurate alternative to type curve matching methods. *Iranian Journal of Science & Technology , 39* (A4), 463-472.

Baars, H., & Kemper, H.-G. (2008). Management Support with Structured and Unstructured Data-An Integrated Business Intelligence Framework. *Information Systems Management , 25* (2), 132-148.

Baesens, B., Bapna, R., Marsden, J., & Vanthienen, J. (2016). Big Data & Analytics in Networked Business. *Management Information Systems Quarterly , 40* (4), 807-818.

Beach, C., & Schiefelbein, W. (2014). Unstructured data: How to implement an early warning system for hidden risks. *Journal of Accountancy* , 46-51.

Bellinger, G., Castro, D., & Mills, A. (2004). *Systems Thinking*. Obtido em 15 de Novembro de 2017, de www.systems-thinking.org/dikw/dikw.htm

Ben-Assuli, O., & Leshno, M. (2013). Implementing a Monte-Carlo simulation on admission decisions. *Journal of Enterprise Information Management , 26* (1/2), 154-164.

Bhatt, C., & Kankanhalli, M. (2011). Multimedia data mining: state of the art and challenges. *Multimedia Tools and Applications , 51* (1), 35-76.

Birks, M., & Mills, J. (2010). *Grounded Theory.* Thousand Oaks: SAGE.

Bryman, A., & Bell, E. (2011). *Business Research Methods* (3rd Edition ed.). Oxford: Oxford University Press.

Çaparlar, C., & Donmez, A. (2016). What is Scientific Research and How Can it be Done? *Turkish Journal of Anaesthesiology and Reanimation , 44* (4), 212-218.

Chan, J. (2013). An Architecture for Big Data Analytics. *Communications of the International Information Management , 13* (2), 1-14.

Cooper, P. (2016). Data, information, knowledge and wisdom. *Anaesthesia and intensive care medicine , 18* (1), 55-56.

Coopler, D., & Schindler, P. (2013). *Business Research Methods* (12th ed ed.). Nova Iorque: McGraw-Hill.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology , 13* (1), 3-21.

Das, T., & Kumar, M. (2013). BIG Data Analytics: A Framework for Unstructured Data Analysis. *International Journal of Engeneering and Technology , 5* (1), 153-156.

Dijcks, J.-P. (June de 2013). *Oracle.* Obtido em 13 de Dezembro de 2016, de www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

Ding, S., Huang, H., Yu, J., & Zhao, H. (2015). Research on the hybrid models of granular computing and support vector machine. *Artificial Intelligence Review , 43* (4), 565-577.

Dumas, J., & Redish, J. (1999). *A Practical Guide to Usability Testing.* Intellect Books.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing . *Journal of Business Research , 69* (2), 897-904.

Evans, J. R., & Lindner, C. H. (2012). Business Analytics: The Next Frontier for Decision Sciences. *Decision Line , 43* (2), pp. 4-6.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review , 1* (2), 293-314.

Fawumi, K. (2015). *Design of an Interactive and Web-based Software for the Management, Analysis and Transformation of Time Series.* Munique: Tese de Mestrado, Universidade Técnica de Munique.

Flick, U. (2013). *The SAGE Handbook of Qualitative Data Analysis* (1st ed.). Thousand Oaks: SAGE.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management , 35* (2), 137-144.

Ge, W., & Whitmore, G. (January de 2010). Binary response and logistic regression in recent accounting research publications: a methodological note. *Review of Quantitative Finance and Accounting* , pp. 81-93.

Geler, Z., Kurbalija, V., Radovanović, M., & Ivanović, M. (Agosto de 2016). Comparison of different weighting schemes for the kNN classifier on time-series data. *Knowledge and Information Systems* , pp. 331-378.

George, G., Haas, M., & Pentland, A. (2014). Big Data and Management. *Acadamy of Management Journal , 57* (2), 321-327.

Greener, S. (2008). *Business Research Methods.* Dinamarca: Ventus Publishing ApS.

Hayashi, A. (2013). Thriving in a Big Data World. *Mit Sloan Management Review , 55* (2), 35-39.

Hesse-Biber, S. N., & P. L. (2011). *The Practice of Qualitative Research* (2nd edition ed.). Thousand Oaks: SAGE.

Hooker, G., & Mentch, L. (2016). Comments on: A random forest guided tour. *TEST , 25* (2), 254-260.

Hu, Y. (2011). Linear Regression 101. *Journal of Validation technology , 17* (2), 15-22.

International Business Machines. (2012). Obtido em 24 de Maio de 2017, de www.ibm.com/https/www-01.ibm.com/software/in/data/bigdata/

International Business Management. (2013). Obtido em 24 de Maio de 2017, de www.01.ibm.com/software/data/bigdata/

Jifa, G., & Lingling, Z. (2014). Data, DIKW, Big data and Data science. *Procedia Computer Science , 31*, 814-821.

Johnson, B. D., Dunlap, E., & Benoit, E. (2010). Organizing "Mountains of Words" for Data Analysis, both Qualitative and Quantitative. *Substance Use and Misuse, 45* , 648-670.

Kangas, A., Leskinen, P., & Kangas, J. (2007). Comparison of Fuzzy and Statistical Approaches in Multicriteria Decisionmaking. *Forest Science , 53*, 37-44.

Katz, G., Shabtai, A., Rokach, L., & Ofek, N. (2014). ConfDTree: A Statistical Method for Improving Decision Trees. *JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY , 29* (3), 392-407.

Li, P., Bu, J., Zhang, L., & Chen, C. (2015). Graph-based local concept coordinate factorization. *Knowledge and information systems , 43* (1), 103-126.

Lycett, M. (2013). "Datafication": making sense of (big) data in a complex world. *European Journal of Information Systems , 22* (4), 381-386.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (May de 2011). *McKinsey & Company.* Obtido em 21 de Dezembro de 2016, de www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Matthews, P. (1998). What Lies Beyond Knowledge Management: Wisdom Creation and Versatility. *Journal of Knowledge Management , 1* (3), 207-214.

Mayer, I. (2015). Qualitative Research with a Focus on Qualitative Data Analysis. *International Journal of Sales, Retailing and Marketing* , 53-67.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution That will Transform How We Live, Work, and Think.* Boston: Eamon Dolan Book.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review* , 59-69.

Mckendrick, J. (2011). *The Post-Relational Realty Sets In: 2011 Survey on Unstructured Data.*

Minelli, M., Dhiraj, A., & Chambers, M. (2013). *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trend for Today's Business.* Hoboken, NJ: John Wiley & Sons, Inc.

Mrad, A. B., Delcroix, V., Piechowiak, S., Leicester, P., & Abid, M. (2015). An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence , 43* (4), 802-824.

Ozkose, H., Ari, E. S., & Gencer, C. (2015). Yesterday, Today and Tomorrow of Big Data. *Procedia Social and Behavioral Sciences , 195*, 1042-1050.

Petroşanu, D.-M., & Pîrjan, A. (2017). IMPLEMENTATION SOLUTIONS FOR DEEP LEARNING NEURAL NETWORKS TARGETING VARIOUS APPLICATION FIELDS. *JOURNAL OF INFORMATION SYSTEMS & OPERATIONS MANAGEMENT , 11*, 155.

Rowe, S. D. (2017). Leveraging the Three Stages of ANALYTICS. *CUSTOMER RELATIONSHIP MANAGEMENT* , 20-23.

Sarma, G. (2015). The Art of Memory and the Growth of the Scientific Method. *Interdisciplinary Description of Complex Systems , 13* (3), 373-396.

Schutt, R. (2011). *Investigating the Social World: The Process and Practice of Research* (7th edition ed.). Thousand Oaks: SAGE.

Shmueli, G., & Koppius, O. (2010). Predictive Analytics in Information Systems Research. *Robert H. Smith School of Business , RHS-06-138*, 1-47.

Sukanya, M., & Biruntha, S. (4 de October de 2012). *International Conference on Advanced Communications Control and Computing Technologies.* Obtido em 23 de Fevereiro de 2017, de www.ieeexplore.ieee.org/abstract/document/6320784/

Teixeira, C. S., & Pereira, L. L. (2015). Pereira Diamond: Benefits Management Framework. *The International Journal Of Business & Management , 3* (3), 47-56.

Tucker, P. (2013). Expanding the Predictable Universe. *The Futurist* , 56-57.

Vining, G. (2013). Technical Advice: Scientific Method and Approaches for Collecting Data. *Quality Engineering , 25*, 194-201.

Watson, H. (2015). Should You Pursue a Career in BI/Analytics? *BUSINESS INTELLIGENCE JOURNAL* , 4-8.

Williams, C. (2007). Research Methods. *Journal of Business & Economic Research* , 65-72.

Wu, S.-T., & Li, Y. (2013). Pattern-Based Web Mining Using Data Mining Techniques. *International Journal of e-Education, e-Business, e-Management and e-Learning , 3* (2), 163-167.

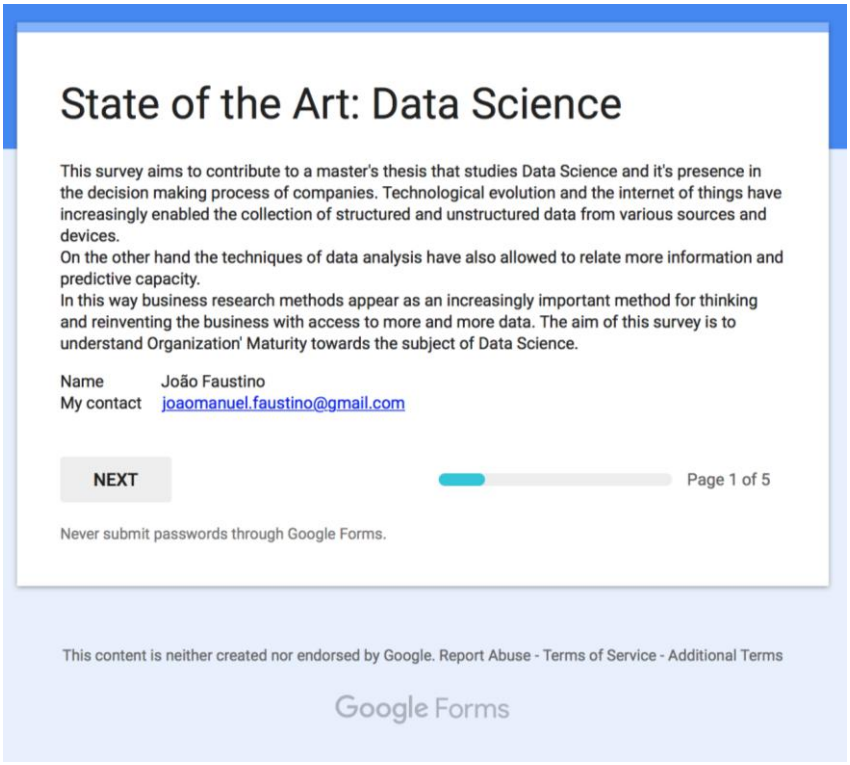
Yi, S. K., Steyvers, M., Lee, M., & Dry, M. (2012). The Wisdom of the Crowd in Combinatorial Problems. *Cognitive Science , 36* (3), 452-470.

Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. *Human Systems Management , 7* (1), 59-70.

Zikmund, W., Babin, B., Carr, J., & Griffin, M. (2010). *Business Research Methods.* South-Wester: Cengage Learning.

ANEXOS

ANEXO A – SURVEY



State of the Art: Data Science

This survey aims to contribute to a master's thesis that studies Data Science and it's presence in the decision making process of companies. Technological evolution and the internet of things have increasingly enabled the collection of structured and unstructured data from various sources and devices.
On the other hand the techniques of data analysis have also allowed to relate more information and predictive capacity.
In this way business research methods appear as an increasingly important method for thinking and reinventing the business with access to more and more data. The aim of this survey is to understand Organization' Maturity towards the subject of Data Science.

Name João Faustino
My contact joaomanuel.faustino@gmail.com

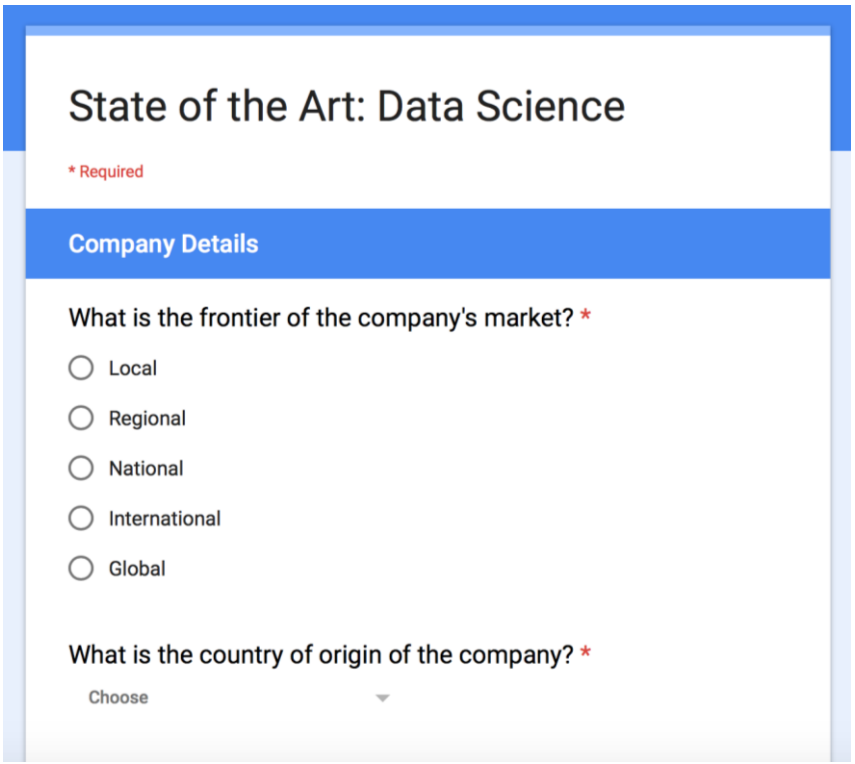
NEXT

Page 1 of 5

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



State of the Art: Data Science

* Required

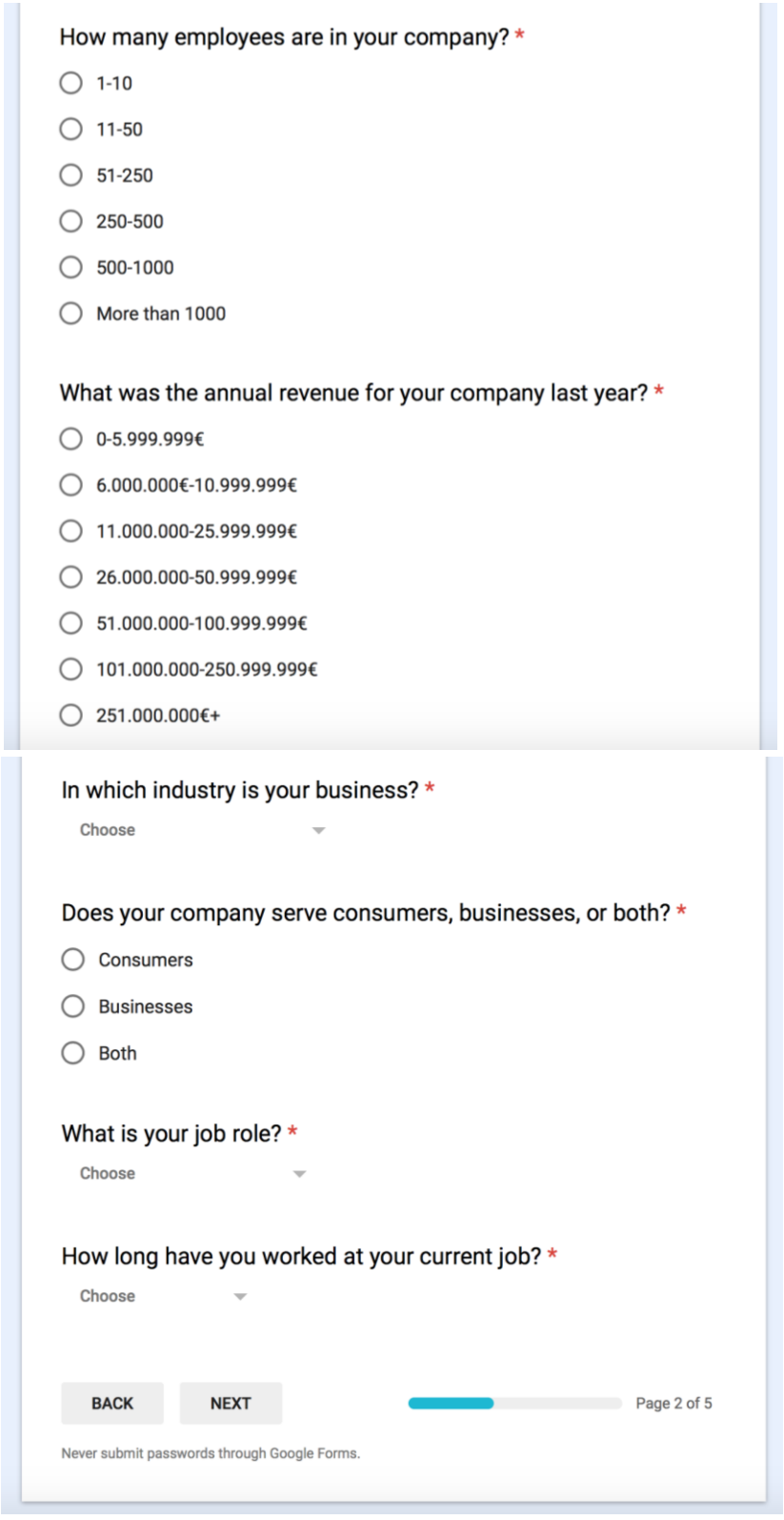
Company Details

What is the frontier of the company's market? *

- Local
- Regional
- National
- International
- Global

What is the country of origin of the company? *

Choose



How many employees are in your company? *

- 1-10
- 11-50
- 51-250
- 250-500
- 500-1000
- More than 1000

What was the annual revenue for your company last year? *

- 0-5.999.999€
- 6.000.000€-10.999.999€
- 11.000.000-25.999.999€
- 26.000.000-50.999.999€
- 51.000.000-100.999.999€
- 101.000.000-250.999.999€
- 251.000.000€+

In which industry is your business? *

Choose

Does your company serve consumers, businesses, or both? *

- Consumers
- Businesses
- Both

What is your job role? *

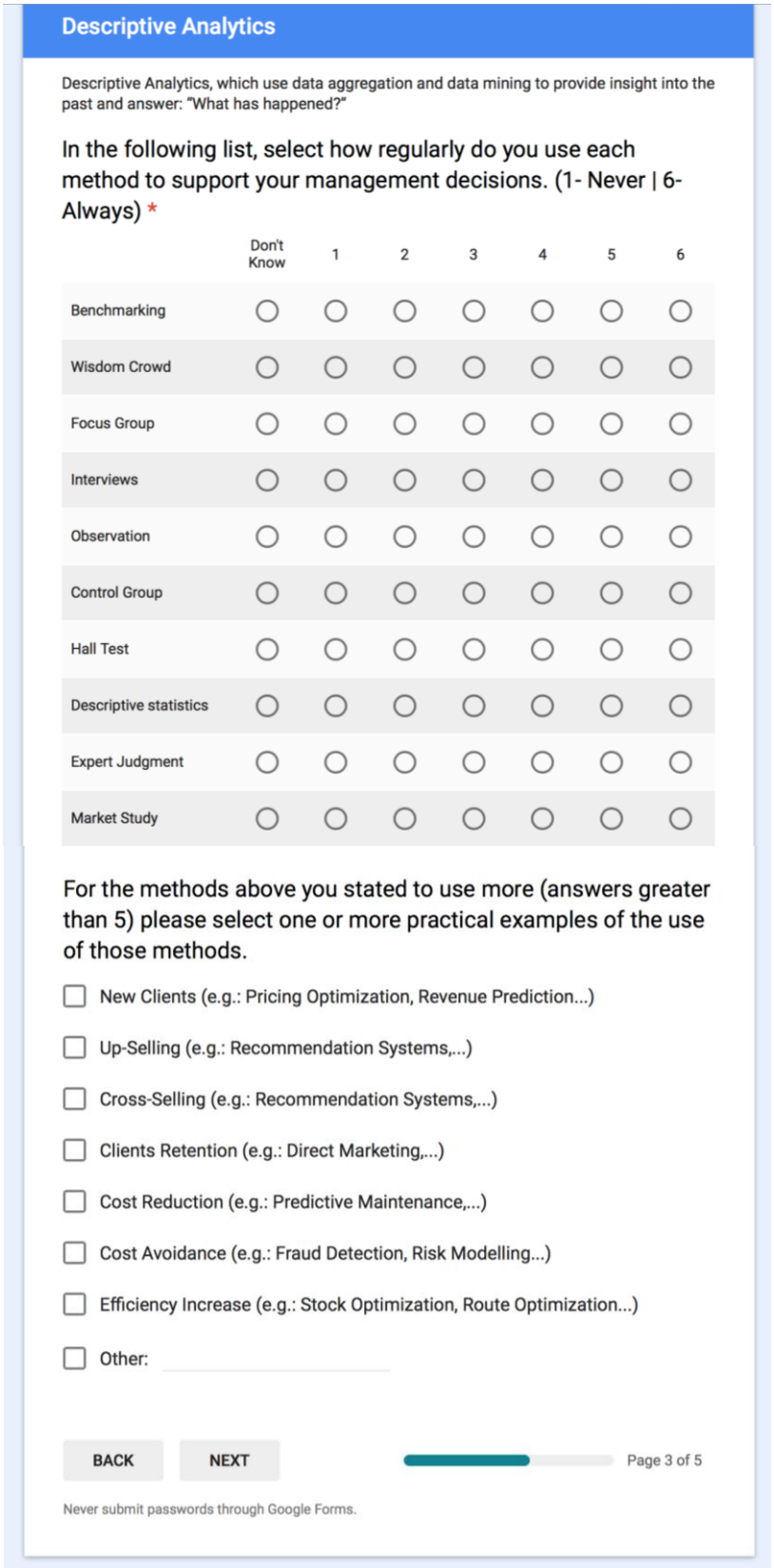
Choose

How long have you worked at your current job? *

Choose

BACK NEXT Page 2 of 5

Never submit passwords through Google Forms.



Descriptive Analytics

Descriptive Analytics, which use data aggregation and data mining to provide insight into the past and answer: "What has happened?"

In the following list, select how regularly do you use each method to support your management decisions. (1- Never | 6- Always) *

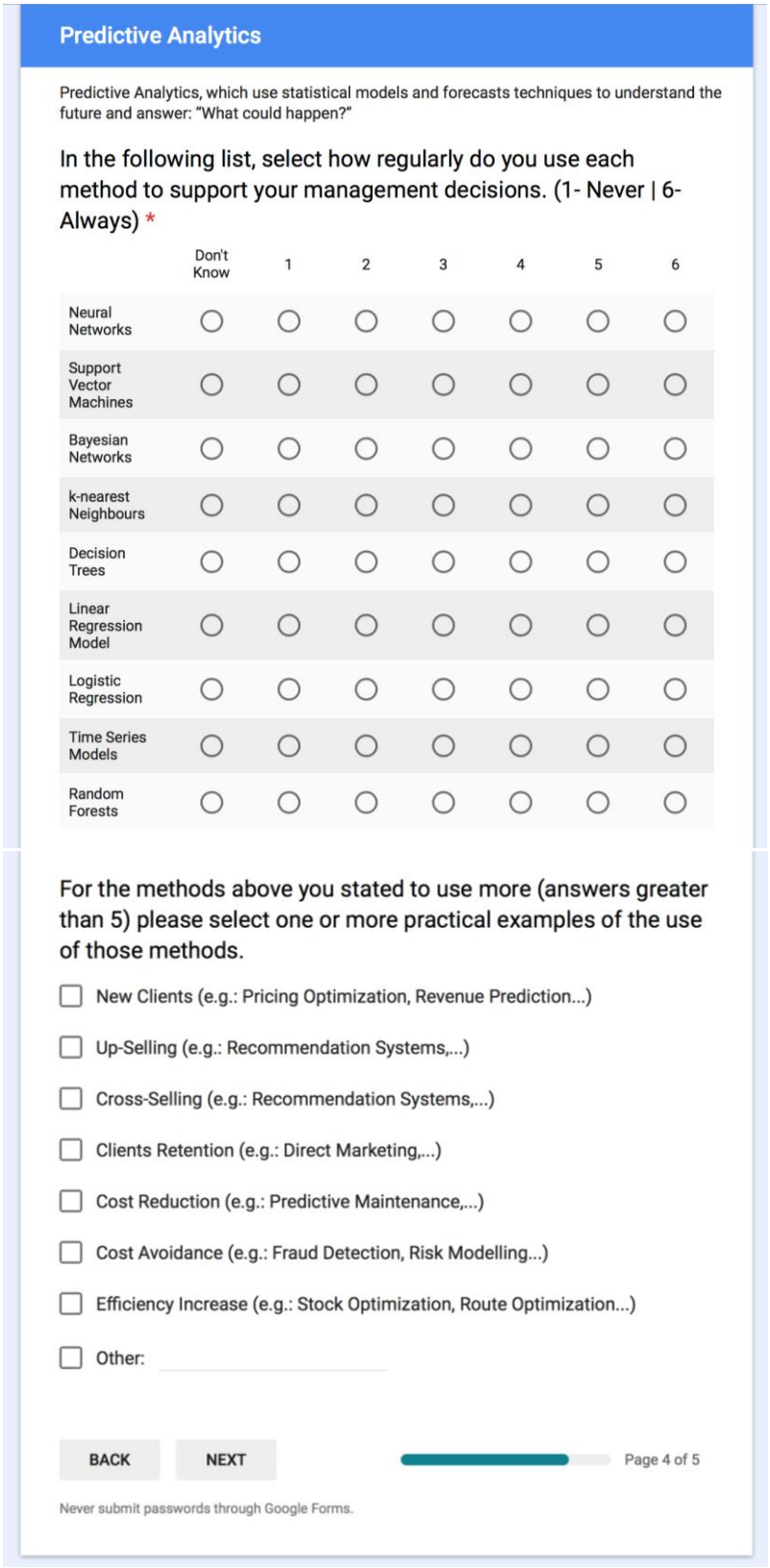
	Don't Know	1	2	3	4	5	6
Benchmarking	○	○	○	○	○	○	○
Wisdom Crowd	○	○	○	○	○	○	○
Focus Group	○	○	○	○	○	○	○
Interviews	○	○	○	○	○	○	○
Observation	○	○	○	○	○	○	○
Control Group	○	○	○	○	○	○	○
Hall Test	○	○	○	○	○	○	○
Descriptive statistics	○	○	○	○	○	○	○
Expert Judgment	○	○	○	○	○	○	○
Market Study	○	○	○	○	○	○	○

For the methods above you stated to use more (answers greater than 5) please select one or more practical examples of the use of those methods.

- ☐ New Clients (e.g.: Pricing Optimization, Revenue Prediction...)
- ☐ Up-Selling (e.g.: Recommendation Systems,...)
- ☐ Cross-Selling (e.g.: Recommendation Systems,...)
- ☐ Clients Retention (e.g.: Direct Marketing,...)
- ☐ Cost Reduction (e.g.: Predictive Maintenance,...)
- ☐ Cost Avoidance (e.g.: Fraud Detection, Risk Modelling...)
- ☐ Efficiency Increase (e.g.: Stock Optimization, Route Optimization...)
- ☐ Other: _____

BACK NEXT Page 3 of 5

Never submit passwords through Google Forms.



Predictive Analytics

Predictive Analytics, which use statistical models and forecasts techniques to understand the future and answer: "What could happen?"

In the following list, select how regularly do you use each method to support your management decisions. (1- Never | 6- Always) *

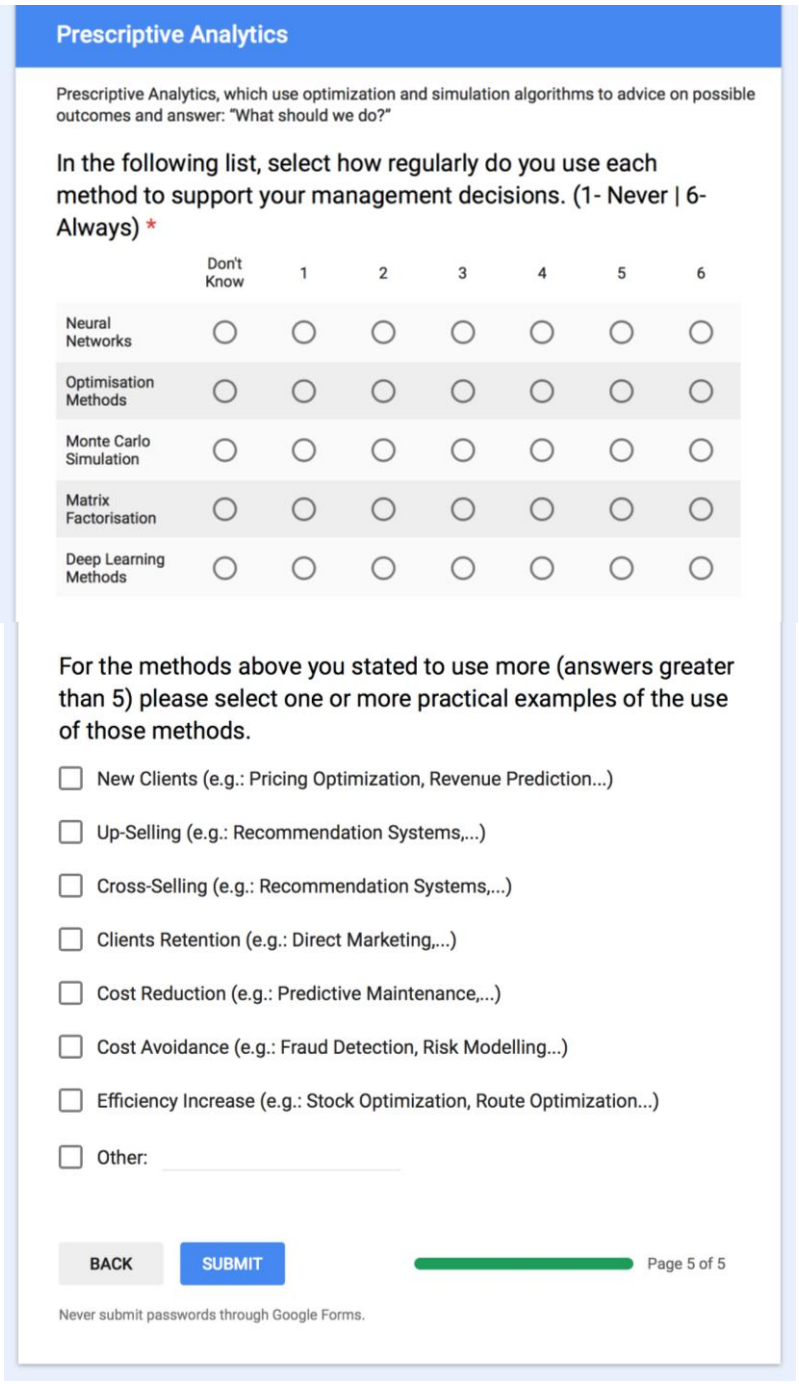
	Don't Know	1	2	3	4	5	6
Neural Networks	○	○	○	○	○	○	○
Support Vector Machines	○	○	○	○	○	○	○
Bayesian Networks	○	○	○	○	○	○	○
k-nearest Neighbours	○	○	○	○	○	○	○
Decision Trees	○	○	○	○	○	○	○
Linear Regression Model	○	○	○	○	○	○	○
Logistic Regression	○	○	○	○	○	○	○
Time Series Models	○	○	○	○	○	○	○
Random Forests	○	○	○	○	○	○	○

For the methods above you stated to use more (answers greater than 5) please select one or more practical examples of the use of those methods.

- ☐ New Clients (e.g.: Pricing Optimization, Revenue Prediction...)
- ☐ Up-Selling (e.g.: Recommendation Systems,...)
- ☐ Cross-Selling (e.g.: Recommendation Systems,...)
- ☐ Clients Retention (e.g.: Direct Marketing,...)
- ☐ Cost Reduction (e.g.: Predictive Maintenance,...)
- ☐ Cost Avoidance (e.g.: Fraud Detection, Risk Modelling...)
- ☐ Efficiency Increase (e.g.: Stock Optimization, Route Optimization...)
- ☐ Other: _____

BACK NEXT Page 4 of 5

Never submit passwords through Google Forms.



Prescriptive Analytics

Prescriptive Analytics, which use optimization and simulation algorithms to advice on possible outcomes and answer: "What should we do?"

In the following list, select how regularly do you use each method to support your management decisions. (1- Never | 6- Always) *

	Don't Know	1	2	3	4	5	6
Neural Networks	○	○	○	○	○	○	○
Optimisation Methods	○	○	○	○	○	○	○
Monte Carlo Simulation	○	○	○	○	○	○	○
Matrix Factorisation	○	○	○	○	○	○	○
Deep Learning Methods	○	○	○	○	○	○	○

For the methods above you stated to use more (answers greater than 5) please select one or more practical examples of the use of those methods.

- ☐ New Clients (e.g.: Pricing Optimization, Revenue Prediction...)
- ☐ Up-Selling (e.g.: Recommendation Systems,...)
- ☐ Cross-Selling (e.g.: Recommendation Systems,...)
- ☐ Clients Retention (e.g.: Direct Marketing,...)
- ☐ Cost Reduction (e.g.: Predictive Maintenance,...)
- ☐ Cost Avoidance (e.g.: Fraud Detection, Risk Modelling...)
- ☐ Efficiency Increase (e.g.: Stock Optimization, Route Optimization...)
- ☐ Other: _____

BACK SUBMIT Page 5 of 5

Never submit passwords through Google Forms.