

MESTRADO
ECONOMIA E ADMINISTRAÇÃO DE EMPRESAS

O valor da informação em empresas de comércio eletrónico. Uma avaliação.

Daniel Moreira Martins

M

2018



O valor da informação em empresas de comércio eletrónico. Uma avaliação.

Daniel Moreira Martins

Dissertação

Mestrado em Economia e Administração de Empresa

Orientado por

Pedro José Ramos Moreira de Campos

Maria Isabel Gonçalves da Mota Campos

2018

Agradecimentos

Esta dissertação é o culminar de dois anos de repletos de desafios, que exigiram muito esforço e dedicação e é o resultado de muitas horas de trabalho. Por isso, não posso deixar de agradecer de forma especial a algumas pessoas que tornaram tudo isto possível.

Ao Professor Doutor Pedro Campos e à Professora Doutora Isabel Campos, por terem aceite o desafio de me orientar nesta dissertação, pela disponibilidade sempre demonstrada, e pelo precioso auxílio prestado durante estes últimos meses.

Aos professores da Faculdade de Economia da Universidade do Porto, pelo conhecimento que me transmitiram, extremamente útil para ser um melhor profissional.

Aos meus colegas do mestrado, pelo espírito de interajuda.

À minha família, em particular à minha esposa, pelo seu esforço em tantas horas em que estive ausente, mesmo em algumas em que eu estava presente, pelo seu apoio incondicional e pela sua compreensão. E ao meu filho, por este primeiro ano em que me trouxe tanta alegria.

Resumo

O presente trabalho tem como objetivo avaliar o valor da informação enquanto ativo das empresas, considerando o caso particular das empresas de comércio eletrônico. De acordo com a literatura, a informação é, cada vez mais, um dos principais ativos das empresas. A informação privada das empresas é geradora de assimetrias, revelando-se uma vantagem nas suas operações enquanto fator gerador de benefícios económicos. Ainda assim, a sua gestão é ainda desconsiderada em muitas empresas.

Recorrendo a dados primários de empresas de comércio eletrônico, são utilizadas técnicas de análise exploratória de dados (Análise de Componentes Principais Categórica, Análise de *Clusters*), propondo-se a construção de um índice do valor da informação. Da análise dos resultados é possível concluir que, em termos gerais, as características que contribuem para um melhor posicionamento em termos do índice do valor da informação proposto, é o desenvolvimento da “Comunidade digital”, seguido do desenvolvimento da “Informação sobre produtos acessórios”, do desenvolvimento da “Informação sobre produtos comparáveis” e da “Atratividade do *site*”.

Palavras-chave: Valor da Informação, Economia da Informação, Monetização, Análise de Componentes Principais Categórica, Análise de *Clusters*.

Códigos JEL: C38, D82, M41.

Abstract

The present study aims to evaluate the value of the information as an asset of the companies, considering the particular case of e-commerce companies.

According to the literature, information is becoming more and more one of the main assets of the companies. Companies' private information generates asymmetries, revealing as an advantage on companies' operations, as a factor that generates economic benefits. Nevertheless, its management is still disregarded in many companies.

Using primary data of e-commerce companies, exploratory data analysis techniques are used (Categorical Principal Component Analysis, Cluster Analysis), and an index of information value is proposed. From the analysis of the results it is possible to conclude that, in general terms, the characteristics that contribute to a better positioning in terms of the proposed index of the value of the information, is the development of the "Digital community", followed by the development of the "Information about accessory products", the development of the "Information on comparable products", and of the "Site Attractiveness".

Keywords: Information Value, Information Economics, Monetization, Categorical Principal Component Analysis, Cluster analysis.

JEL codes: C38, D82, M41

Índice

Agradecimentos	i
Resumo	ii
Abstract.....	iii
Índice de quadros	vi
Índice de figuras.....	vii
Capítulo 1. Introdução.....	1
Capítulo 2. O valor da informação: uma revisão de literatura	3
2.1. Informação	3
2.2. Monetizar a informação.....	6
2.3. <i>Big data</i> e <i>Data Mining</i>	12
2.4. Informação de identificação pessoal.....	13
Capítulo 3. O valor da informação em empresas de comércio eletrónico: uma aplicação ...	16
3.1. Metodologia.....	16
3.2. Dados	17
3.3. Análise de Componentes principais.....	20
3.4. Análise de <i>clusters</i>	23
3.5. Índice do valor da informação.....	28
3.6. Características dos <i>sites</i> melhor posicionados no Índice comparativo.....	31
Capítulo 4. Índice financeiro do valor da informação.....	34
Capítulo 5. Conclusões	41
6. Referências bibliográficas.....	45
7. Anexos	47
Anexo 1 - Variáveis: Fonte, especificação e justificação da sua adoção	47
Anexo 2 - Empresas da amostra e dados recolhidos.....	50
Anexo 3 - Descrição das variáveis	52
Anexo 4 - Sumário do Modelo com 6 componentes.....	54
Anexo 5 - Sumário do Modelo com 4 componentes.....	55
Anexo 6 - Classificação dos <i>sites</i> em cada componente.....	56
Anexo 7 - Dendrograma da análise hierárquica de <i>clusters</i> aos <i>sites</i>	59
Anexo 8 - Identificação dos <i>clusters</i> agrupados em cada passo e respetivas distâncias.....	60
Anexo 9 - Identificação do <i>cluster</i> a que pertence cada <i>site</i> e distância ao respetivo centróide.....	63
Anexo 10 - <i>Sites</i> constituintes de cada <i>cluster</i>	66

Anexo 11 - Dendrograma da análise hierárquica de <i>clusters</i> às componentes (variáveis resultantes da CATPCA).....	68
Anexo 12 - Índice correspondente aos 100 <i>sites</i>	69
Anexo 13 - Posicionamento dos <i>sites</i> em termos de média de páginas visualizadas diariamente por visitante.....	72
Anexo 14 - Posicionamento dos <i>sites</i> em termos de média do tempo passado diariamente por visitante	75
Anexo 15 - Posicionamento dos <i>sites</i> em termos de percentagem de visitantes que se mantêm no <i>site</i> , visualizando para além de uma página.....	78
Anexo 16 - Mapa bidimensional com componentes 1 e 2.....	81
Anexo 17 - Mapa bidimensional com componentes 1 e 3.....	82
Anexo 18 - Mapa bidimensional com componentes 1 e 4.....	83
Anexo 19 - Mapa bidimensional com componentes 1 e 5.....	84
Anexo 20 - Mapa bidimensional com componentes 2 e 3.....	85
Anexo 21 - Mapa bidimensional com componentes 2 e 4.....	86
Anexo 22 - Mapa bidimensional com componentes 2 e 5.....	87
Anexo 23 - Mapa bidimensional com componentes 3 e 4.....	88
Anexo 24 - Mapa bidimensional com componentes 3 e 5.....	89
Anexo 25 - Mapa bidimensional com componentes 4 e 5.....	90
Anexo 26 - Dados relativos a 10 empresas tecnológicas que procederam a algumas das maiores IPOs dos últimos anos	91

Índice de quadros

Quadro 1 – Sumário do modelo com 5 componentes.....	21
Quadro 2 - Correlações entre variáveis e componentes	22
Quadro 3 - Denominação das componentes e resumo das respectivas variáveis com maior correlação.....	23
Quadro 4 - Centros dos <i>clusters</i> em cada componente	27
Quadro 5 - Variáveis a incluir no índice do valor da informação.....	28
Quadro 6 - 20 sites melhor classificados em termos do índice do valor da informação (Top 20)	29
Quadro 7 - Índice financeiro do valor da informação das dez empresas.....	37
Quadro 8 - Rácio Q das dez empresas	38
Quadro 9 - Índice financeiro do valor da informação das 5 empresas cuja base geradora de receitas é apenas informação.....	39
Quadro 10 - Rácio Q das 5 empresas cuja base geradora de receitas é apenas informação.....	40

Índice de figuras

Figura 1 - Número de <i>clusters</i> e suas distâncias.....	25
---	----

Capítulo 1. Introdução

Apesar de a informação contribuir para as receitas e custos das empresas, por vezes em maior escala que os ativos tradicionais, e ainda que a informação possa cumprir os requisitos das normas de internacionais de contabilidade para que seja reconhecida como ativo, as empresas não refletem o valor da informação no seu balanço (Garifova, 2015). Verifica-se ainda que a atribuição de valor económico mensurável à informação é uma questão muito pouco tratada na literatura, ainda que tenha uma relevância crescente nos tempos mais recentes. Por exemplo, desde os ataques terroristas de 11 de setembro de 2001, quando muitas empresas perderam informação vital para os seus negócios, as seguradoras indemnizaram bens materiais mas não a perda de informação, por esta alegadamente não ter valor.

O valor económico da informação será tanto maior quanto maior a redução de incerteza permitida por essa informação na tomada de decisões. Para Hubbard (2010), se há incerteza quanto ao resultado de uma decisão, há alguma probabilidade de se tomar a decisão errada, na medida em que a alternativa se venha a revelar como preferível por no momento da decisão se ter essa informação. No mesmo sentido, Mowshowitz (1992) refere que a informação consiste na habilidade de decidir e controlar tendo em vista atingir um objetivo, consistindo essa decisão na escolha de uma alternativa entre as várias que podem ser escolhidas tendo em vista atingir esse objetivo. O valor económico da informação poderá também traduzir-se na vantagem competitiva obtida com a assimetria de informação (Najjar & Kettinger, 2013).

A questão da atribuição de valor mensurável e monetizável à informação foi tornada célebre por Doug Laney (Laney, 2017), com a publicação do seu livro *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*. O autor analisa como medir a informação, como gerir-la e como monetizá-la, de forma a retirar desse ativo vantagens competitivas para as empresas. O termo “monetização”, da tradução de “*monetization*”, surge assim cada vez mais frequentemente associado às novas tecnologias com o sentido de “converter um ativo em dinheiro”.

O principal objetivo deste estudo é mensurar o valor económico da informação em empresas de comércio eletrónico. Este estudo propõe-se assim responder a duas questões principais: Como mensurar o valor da informação em empresas de comércio eletrónico? Quais as características das empresas de comércio eletrónico que, em termos comparativos, apresentam maior valor da informação?

Do ponto de vista social, as respostas a estas questões tornam-se úteis na medida em que, na sociedade da informação, as empresas podem melhorar a utilização da informação que detêm para promover os seus produtos ou serviços, para criar novos produtos ou serviços, para entrar em novos mercados, ou para partilhar em parcerias, podendo a sociedade em geral beneficiar de tais ações. Esta temática é também relevante do ponto de vista académico, uma vez que se trata de um tema ainda pouco explorado. A título de exemplo pode referir-se Najjar & Kettinger (2013), Buff et al., (2015), Garifova, (2015) e Laney (2017).

Para este efeito foi desenvolvido um estudo empírico onde foram recolhidas informações acerca de 100 empresas de comércio eletrónico. A partir da informação recolhida e da aplicação de técnicas de análise exploratória de dados (Análise de Componentes Principais Categórica, Análise de *Clusters*) será proposto um índice que permita medir o valor da informação. Por fim, serão analisadas as principais características comuns às empresas que obtiverem o melhor posicionamento nesse índice. Em alternativa, é também considerado um índice financeiro do valor da informação calculado com base em informação financeira de 10 empresas tecnológicas seguindo a metodologia de Laney (2017).

Este estudo socorre-se de dados de empresas de comércio eletrónico que são recolhidos e disponibilizados nos respetivos *sites* e não relativos a outro tipo de empresas ou outros dados, o que resulta essencialmente da exequibilidade do trabalho. A escolha de empresas de comércio eletrónico prende-se também com a possibilidade atual de rastreamento dos hábitos dos utilizadores nos *sites*, tornando-se esse rastreamento em informação que pode ser monetizada. No entanto, a definição da metodologia de análise e mensuração do valor da informação em termos comparativos é passível de ser aplicável a outros tipos de empresas e a outros tipos de informação detidos pelas empresas.

O presente trabalho encontra-se estruturado da seguinte forma: No capítulo 2 é realizada a revisão da literatura. No capítulo 3 são apresentados os dados e a metodologia usados no estudo, é efetuada uma análise de *clusters* e proposto um índice comparativo do valor da informação. Também no capítulo 3 são analisadas as principais características comuns às empresas que obtiverem melhor posicionamento no índice. No capítulo 4 é também analisado um índice comparativo com base em informação financeira de 10 empresas tecnológicas. No capítulo 5, por fim, são apresentadas as conclusões do trabalho e pistas de investigação futura.

Capítulo 2. O valor da informação: uma revisão de literatura

Tendo como objetivo a análise do valor da informação, este capítulo discute os principais conceitos relacionados com esta temática: informação, mensuração de ativos e monetização sendo também apresentada uma revisão e análise crítica da literatura.

2.1. Informação

A importância da informação para as empresas e a possibilidade de a utilizar para criar novos produtos ou serviços, para entrar em novos mercados, ou para partilhar em parcerias é, atualmente, reconhecida (Garifova, 2015). Interessa por isso definir os vários conceitos relacionados com informação.

De acordo com Stiglitz (2017), a maioria das teorias e modelos centrais em Economia baseiam-se na informação perfeita, ou seja, assumem que não existe incerteza relativamente às decisões dos rivais nem relativamente aos seus ganhos. A informação corresponderá assim ao conhecimento disponível para cada agente (empresa ou consumidor) relativamente às variáveis de decisão, num determinado momento.

Segundo Stiglitz (2017), alguns economistas da escola de Chicago trataram a informação no âmbito de um modelo da procura e da oferta de informação. No entanto, a informação é diferente de outros bens como o aço ou milho, apresentando muitas vezes características de bem público. No mesmo sentido, Arrow (1996) procurou estruturar as características da informação que permitem identificá-la como um bem económico. Para este autor, o tratamento da informação como uma variável e as suas implicações para o comportamento económico necessita de uma análise aprofundada, pelo impacto direto que tem na empresa, sendo que o valor da informação surge por vezes diretamente associado à escala das operações. Arrow (1996) refere ainda que os incentivos para a produção de informação podem ser insuficientes, uma vez que a informação quando obtida apresenta uma reduzida apropriabilidade e é facilmente transmissível.

Desde há cerca de meio século, vários autores se têm debruçado sobre a informação, por vezes referida como Economia da Informação, explorando até que ponto os mercados e outras instituições processam e transmitem informações (Stiglitz, 2017). Muitos dos problemas dos mercados resultam de falhas de informação, que estão associadas a inúmeras falhas do mercado, incluindo mercados de risco incompletos,

mercados de capitais imperfeitos e imperfeições na concorrência.

De acordo com Stigler (1961), há vários fatores a considerar quando se fala em informação de mercado, na medida em que os preços de alguns bens poderão ser distintos consoante o vendedor e, por vezes, consoante o comprador. Neste sentido, é variável e requer um esforço de procura tanto pelos compradores como pelos vendedores para o determinar com um grau de certeza adequado. Os compradores procuram tendo em vista encontrar o preço mínimo, enquanto os vendedores procuram conhecer os preços dos seus concorrentes e desse modo apresentar o seu melhor preço aos compradores.

De acordo com Stigler (1961), a dispersão de preços é uma característica dos mercados, mesmo perante bens homogêneos. Os mercados não são estáticos e a sua constante alteração causa perda de conhecimento. À medida que as condições de oferta e de procura variam, os compradores e os vendedores terão de procurar para conhecer o novo preço de mercado dos bens. Quanto maior esta instabilidade do mercado, maior é a dispersão de preços existente.

Vários autores em Economia tentaram avaliar o efeito da informação assimétrica sobre os mercados. Akerlof (1970) estuda a assimetria de informação quando os vendedores de um produto têm mais informação do que os compradores acerca da qualidade dos bens. Através do exemplo do mercado de automóveis usados, o autor conclui que esta assimetria pode originar seleção adversa de bens de baixa qualidade, como automóveis usados com defeitos, conhecidos nos EUA como “limões”. O autor procura demonstrar que a existência de diferentes qualidades de bens, associada à falta de informação dos compradores relativamente a esta variável, pode levar a uma diminuição da qualidade dos bens à disposição dos consumidores ou até à extinção do mercado. Assim, como os compradores desconhecem a verdadeira qualidade dos bens, vão atribuir-lhes um valor médio, ou seja, os bens de elevada qualidade serão subvalorizados e os de baixa qualidade sobrevalorizados. Tal facto levará a que os vendedores de bens de elevada qualidade se afastem deste mercado, enquanto os vendedores de bens de baixa qualidade serão atraídos para este mercado, o que levará a que este mercado transacione bens de baixa qualidade, ou “limões”.

Spence (1973) também analisa problemas de informação assimétrica, investigando de que modo os agentes económicos podem utilizar a sinalização para contrariar os efeitos da seleção adversa, considerando-se sinalização as ações observáveis dos agentes para convencer a parte contrária do valor ou da qualidade dos seus produtos. O autor considera

como exemplo o mercado de trabalho, onde a formação acadêmica é utilizada como sinal da produtividade do trabalhador. Se um empregador não consegue distinguir a produtividade de um candidato a um lugar de trabalho, a solução pode corresponder a um mercado de trabalho onde apenas os trabalhadores com baixa produtividade são contratados a salários baixos, de forma análoga à seleção adversa referida por Akerlof (1970). Spence (1973) propõe então que, ainda que os empregadores não consigam observar diretamente a produtividade dos candidatos, conseguem observar o seu nível de formação acadêmica que funciona como um sinal da sua produtividade. É assim possível atingir um equilíbrio em que as expectativas do empregador são confirmadas pela sinalização geradas pelos salários oferecidos e os candidatos com diferentes produtividades escolhem diferentes níveis de formação acadêmica.

Outros autores analisam problemas de informação assimétrica nos mercados dos seguros e bancário. Rothschild & Stiglitz (1976) estudam o mercado dos seguros em que as seguradoras não têm informação completa sobre o risco individual dos clientes. De forma análoga a Akerlof (1970), a solução dependerá da divisão dos seus clientes da seguradora por classes de risco, com contratos distintos para cada classe. Stiglitz & Weiss (1981) apresentam um modelo de racionamento de crédito no qual, entre diferentes candidatos com características observáveis idênticas, uns recebem o empréstimo e outros não. Os candidatos a quem é recusado o empréstimo não teriam acesso a este mesmo que estivessem dispostos a contratar uma taxa de juro superior ou a apresentar garantias adicionais às exigidas. Tal situação poderia levar ao aumento do risco do portfólio do banco, assim como ao afastamento de candidatos que pretendem fazer investimentos seguros ou à atração de candidatos que pretendem fazer investimentos mais arriscados. O racionamento origina um excesso de procura de crédito para a taxa de juro de equilíbrio. A razão para tal é a seleção adversa presente neste mercado, onde candidatos com diferentes projetos implicam diferentes níveis de risco sendo o tipo de candidato desconhecido pelo mutuante. Os candidatos mais seguros não serão tolerantes a taxas de juro superiores, uma vez que, com reduzida probabilidade de falhar o reembolso, acabariam por pagar valores superiores. Os candidatos mais arriscados já serão tolerantes a essas taxas, dado que têm uma menor probabilidade de serem bem sucedidos nos seus projetos, têm menor probabilidade de cumprir o reembolso.

Após a abordagem feita quanto à informação e às assimetrias por esta criadas, importa agora abordar como monetizar a informação detida.

2.2. Monetizar a informação

A monetização de informação surge quando o valor intangível da informação é convertido em valor real, por vezes através da venda dessa informação. (Najjar & Kettinger, 2013).

Najjar & Kettinger (2013) analisam a monetização de informação no caso de uma empresa de retalho líder nos EUA. No caso concreto, informações relacionadas com os pontos de venda, com a fidelização dos clientes, ou com os *stocks*, podem ser vendidas a fornecedores. Por sua vez, estes terão interesse em utilizar estas informações para melhorar o seu planeamento e melhor gerir o seu *stock*. No entanto, vender a informação a fornecedores pode retirar à empresa a vantagem competitiva obtida com a assimetria de informação. Nesse sentido, devem ser elaborados contratos que assegurem que a informação é usada em benefício de ambas as partes. Devem ainda ser definidos pacotes de dados a incluir, assim como modelos de preços a praticar, tendo em atenção o custo de tornar a informação disponível e o valor atribuído a esta pelo comprador.

Najjar & Kettinger (2013) apresentam ainda três opções possíveis a seguir na monetização da informação, dependendo da capacidade técnica (hardware, software e redes) e analítica (capacidade de análise das pessoas envolvidas) da empresa. A primeira opção a seguir consiste em criar ambas as capacidades, a técnica e a analítica, sendo consequentemente a mais arriscada. Trata-se aqui de desenvolver a infraestrutura técnica e de contratar e/ou formar os colaboradores. Apesar de ser a opção mais dispendiosa, é aquela que mais rapidamente torna a empresa pronta para monetizar a informação que detém. Outra das opções possíveis é começar por desenvolver a capacidade analítica. À medida que esta vai crescendo, a empresa pode alavancar a sua capacidade de análise através da obtenção de cada vez mais informação de fontes internas ou fontes externas. Ainda assim, o desenvolvimento da capacidade analítica pode não ser suficiente para a monetização, sendo necessário expandir seguidamente a capacidade técnica. Esta opção é a que exige maior investimento interno para desenvolver a capacidade analítica interna. A terceira opção será começar pelo desenvolvimento da infraestrutura técnica, escolhendo entre expandir esta a nível interno, ou efetuando o seu *outsourcing*, tendo em vista constituir um conjunto de dados atrativo que possa ser vendido a fornecedores. Através de plataformas digitais, como infraestruturas com base na nuvem que permitam o acesso em

tempo real a informação, a empresa pode mais rapidamente monetizar a informação que detém, evitando alguns custos de análise que incorreria se esta fosse produzida internamente, através da alavancagem da capacidade analítica de alguns dos seus fornecedores.

No caso estudado, Najjar & Kettinger (2013) dividem a monetização da informação em quatro passos. O primeiro refere-se ao desenvolvimento de capacidades técnicas e analíticas que deem resposta a necessidades internas da empresa. Os custos principais de obtenção de informação são de origem técnica, sendo principalmente os incorridos no armazenamento de dados e sua ligação às ferramentas de reporte, e os incorridos na análise da informação. O segundo passo refere-se à partilha de informação com os fornecedores através de plataformas digitais. Aqui, a confiança entre as organizações é uma questão importante, pelo são elaborados contratos de confidencialidade especificando qual a utilização que os fornecedores poderão dar à informação obtida. Os custos de obtenção de informação principais são de origem técnica e analítica, sendo principalmente os incorridos com o desenvolvimento e/ou utilização da plataforma digital de partilha de informação e os incorridos com as contratualizações. O terceiro passo diz respeito à venda de informação aos fornecedores, através da identificação da informação mais vendável e da elaboração de reportes personalizados constituindo diferentes pacotes de informação com diferentes detalhes e com diferentes preços. Aqui, a confiança entre as organizações é uma questão que vai além da questão contratual, baseando-se não só nos contratos escritos, mas também na convicção de que os fornecedores vão utilizar a informação em benefício de ambas as partes. Os custos de obtenção de informação principais, de origem técnica e analítica, são aqui cobertos, pelo menos parcialmente, com as receitas obtidas com a venda da informação. O quarto passo, por último, refere-se à alavancagem da capacidade analítica dos fornecedores, evitando alguns custos de análise que incorreriam se esta fosse feita internamente. A empresa passa a confiar ainda mais nas sugestões feitas pelos fornecedores, após a análise efetuada por estes à informação que adquiriram da empresa. Aqui, a confiança entre as organizações é uma questão que se centra na competência, isto é, a empresa acredita que os fornecedores têm a capacidade técnica e de gestão necessárias para a análise da informação cedida pela empresa. Significa que a empresa reconhece que os benefícios da partilha da informação com um conjunto restrito de fornecedores são superiores aos benefícios da não partilha desta. A competição entre fornecedores para serem incluídos no grupo restrito pode, deste modo, servir como incentivo a um maior

empenho destes na análise da informação da empresa e no fornecimento de sugestões pertinentes quanto às estratégias de vendas da empresa, de promoções e políticas de preço.

De acordo com Najjar & Kettinger (2013), a monetização da informação cria um novo modelo de negócio para a empresa, em que a estrutura de custos, geração de receitas e proposta de valor mudam. A informação da empresa torna-se um produto digital que a empresa pode usar para gerar receitas e cobrir custos associados com a criação e organização da informação. As empresas devem criar pacotes de informação para venda, que integrem informação relevante, respondendo a diferentes necessidades, capacidades de análise e predisposições para o preço, aumentando assim o valor da informação da empresa, devendo criar também uma estratégia de marketing para promover este valor. É comum os potenciais compradores, nomeadamente fornecedores, terem maior capacidade de análise da informação da empresa vendedora da informação. Essa capacidade deve traduzir-se numa maior qualidade das sugestões efetuadas pela compradora à vendedora, pelo que o desempenho associado deve ser constantemente monitorizado e devem ser atribuídos benefícios que sirvam de incentivo. A confiança entre as organizações pode diminuir os custos contratuais e o nível de conflitos. Nesse sentido, o nível de transparência na plataforma digital de colaboração desempenha um papel fulcral.

Para Laney (2017), as organizações podem monetizar a informação essencialmente de duas formas: através da troca de bens e serviços, ou através do uso da informação para aumentar receitas ou reduzir despesas/riscos. Segundo o autor, existem várias razões para monetizar a informação, nomeadamente: aumento da aquisição/retenção de clientes; criação de fluxo de receitas suplementares; introdução de novas áreas de negócio; entrada em novos mercados; diferenciação competitiva; troca por bens ou serviços, cobertura de custos de gestão e análise de informação; redução de custos de manutenção ou excesso de custos, e melhoria do bem-estar da sociedade.

Segundo Laney (2017), a maioria dos métodos para monetizar a informação baseiam-se no pressuposto de que a organização tem capacidade de análise, isto é, tem capacidade para encontrar padrões que permitam atribuir sentido à informação detida. À medida que o volume de transações, velocidade e variedade de informação crescem para níveis extraordinários, a capacidade de análise humana revela-se uma ferramenta insuficiente para retirar valor económico da informação, revelando-se o poder da análise computacional uma ferramenta essencial.

Para Hubbard (2010), se há incerteza quanto a uma decisão, há alguma

probabilidade de se tomar a decisão errada, na medida em que a alternativa se venha a revelar como preferível. O custo de errar é a diferença entre o custo da alternativa errada e o custo da melhor alternativa, sendo esta a que seria escolhida se se estivesse perante informação perfeita (ausência de incerteza). A existência deste risco e o desejo da sua mitigação é a razão pela qual o decisor necessita da informação.

Hubbard (2010) utiliza um modelo para estimar o valor da informação em casos concretos, concluindo que grande parte das vezes o valor económico de medir uma variável é normalmente inversamente proporcional à atenção dedicada na medição dessa variável. Como muitas das organizações não têm capacidade analítica para medir o valor de determinadas variáveis, há probabilidade cometerem erros. Saber como medir a incerteza é essencial para avaliar o risco, e saber avaliar o risco é essencial para medir o valor da informação. Ainda, conhecer valor da informação é essencial para saber o que medir e saber qual o nível de esforço a aplicar nessa medição. Será desaconselhável tentar medir exaustivamente algo cujo nível de incerteza é elevado, sendo preferível fazer essa medição apenas superficialmente, avaliando à *posteriori* se é necessária uma medição adicional.

De acordo com Copeland & Friedman (1992), o valor que um indivíduo racional atribui à informação pode ser aferido pelo excesso da utilidade esperada que obtém ao tomar uma decisão ótima informada em relação à utilidade esperada que obtém ao tomar uma decisão ótima não informada.

Segundo Mowshowitz (1992) a informação é usada pelos agentes económicos para tomar decisões e para controlar processos. Neste sentido, a informação consiste na habilidade de decidir e controlar tendo em vista atingir um objetivo, consistindo essa decisão na escolha de uma alternativa entre várias tendo em vista atingir esse dado objetivo.

Como referido até aqui, a informação detida pelas empresas pode ser convertida em valor real, monetizando-a (Najjar & Kettinger, 2013), tendo sido apontadas essencialmente duas formas de ocorrer essa monetização: através da troca de bens e serviços, ou através do uso da informação para aumentar receitas ou reduzir despesas/riscos (Laney, 2017), sendo essa informação um auxílio na tomada de decisão segundo Hubbard (2010) e Copeland & Friedman (1992). Para além do papel desempenhado nesse auxílio, a informação pode em algumas organizações, ser produzida e vendida (Buff *et al.*, 2015), como uma mercadoria (Mowshowitz 1992). Mowshowitz (1992) aponta dois fatores para algo poder ser considerado uma mercadoria: a possibilidade de ser apropriável e a possibilidade de ser valorizável. Se algo não pode ser objeto de propriedade,

não pode ser transacionável. Além disso, se algo não pode ser valorizado, não pode ser determinada nem quantificada a sua moeda de troca. O valor de mercado da informação deriva da sua capacidade de fornecer informação para tomar decisões e para controlar processos, como instrumento ou como elemento intermédio, podendo também ser objeto de consumo final. Neste sentido, a produção de informação pode ser usada para venda ou como um bem de consumo intermédio.

Segundo Mowshowitz (1992) a relação entre incerteza e o valor económico da informação é análoga a relação entre quantidade e valor económico presente nos bens tangíveis. Como o autor refere, um quilo de manteiga não tem o mesmo valor económico que um quilo de ouro. Da mesma forma, as especificações de um processo bancário e uma receita culinária provavelmente terão valor económico diferente, ainda que possam representar o mesmo nível de incerteza. No mundo da informação, tal como dos bens tangíveis é altamente diferenciável e engloba informações com diferentes valores económicos.

Buff *et al.*, (2015) defende que o papel desempenhado pela informação está a mudar em algumas organizações, deixando de ser utilizada apenas como um ativo secundário que suporta decisões e processos, passando a ser o ativo principal que alguns negócios produzem e vendem, monetizando essa informação.

Bütler & Birchler (2007) defendem que o conhecimento e a informação têm valor na medida em que ao reduzir a incerteza, a informação ajuda a tomar a ação mais adequada. Os autores vão mais longe e identificam um fenómeno conhecido como a “maldição do vencedor”, que pode ocorrer em leilões com informação incompleta em que o vencedor tende a pagar acima do valor real do bem. Tal facto pode dever-se ao fervor do leilão, no entanto este fenómeno verifica-se também em algumas operações das empresas e decorre de falhas de informação. Laney, (2017) aponta 5 métodos de determinação do valor da informação, são eles:

O método do valor intrínseco da informação (VII), que depende da Validade (VAL), Plenitude (PLE), Exclusividade (EXC) e Ciclo de vida (CDV) inerentes à informação. A Validade refere-se à percentagem de registos nos dados que se consideram corretos, a Plenitude à percentagem de registos em relação ao universo de registos ou registos supostos, a Exclusividade à percentagem do mercado ou dos concorrentes que tem probabilidade de ter os mesmos dados, por último, o Ciclo de vida refere-se ao período de tempo em que a informação se mantêm como útil (por exemplo um mês), sendo a sua

fórmula correspondente:

$$VII = \boxed{VAL * PLE * (1 - EXC) * CDV} \quad (2.2.1)$$

O método do valor da informação para o negócio (VIN), que depende da Relevância (REL), da Validade (VAL), Plenitude (PLE), e Tempestividade (TEM). A Relevância refere-se ao grau de importância que a informação representa para um ou mais processos de negócio (de 0 a 1), a Validade à percentagem de registos nos dados que se consideram corretos, a Plenitude à percentagem de registos em relação ao universo de registos ou registos supostos, por último, a Tempestividade refere-se à rapidez com que os dados são registados e disponíveis para consulta, sendo a sua fórmula correspondente:

$$VIN = \boxed{\sum_{p=1}^n (REL_p) * VAL * PLE * TEM} \quad (2.2.2)$$

O método do valor da informação do desempenho (VID), que depende dos KPI (*Key performance indicators*) do grupo informado dos (KPIi), dos KPI do grupo de controlo (KPIc), do tempo médio de “vida” de cada dado (T) e da duração da medição dos KPI, (t), sendo a sua fórmula correspondente:

$$VID = \boxed{\left(\frac{KPIi}{KPIc} - 1 \right) * \frac{T}{t}} \quad (2.2.3)$$

O método do custo do valor da informação (CVI), que depende dos custos anuais do(s) processo(s) envolvidos na recolha dos dados (ProRD), da percentagem das despesas desses processos que são atribuíveis à recolha dos dados (AtrRD), das receitas perdidas no período (RECP), do tempo médio de “vida” de cada dado (T) e da duração da medição das despesas do(s) processo(s), (t), sendo a sua fórmula correspondente:

$$CVI = \boxed{\frac{ProRD * AtrRD * T}{t} + \sum_{p=0}^n RECP_p} \quad (2.2.4)$$

O método do valor económico da informação (VEI), que depende das receitas geradas usando a informação, pelo grupo informado, (RECi), das receitas geradas não usando a informação, pelo grupo não informado ou grupo de controlo, (RECC), do total de despesas ocorridas na abrangência temporal da duração da medição do valor económico da informação, do tempo médio de “vida” de cada dado (T) e da abrangência temporal da duração da medição do valor económico da informação, (t), sendo a sua fórmula correspondente:

$$VEI = \boxed{(RECi - RECC - DESP) * \frac{T}{t}} \quad (2.2.5)$$

Para ser possível valorizar a informação, segundo Laney (2017) e tratá-la como uma

mercadoria, como apontado por Mowshowitz (1992), será necessário analisá-la e quantificá-la. Neste sentido, será necessário a extração de informação dos dados de grande dimensão, denominados *big data*, detidos pela empresa, através de técnicas como o *data mining* (Rajagopal 2011).

2.3. *Big data e Data Mining*

Segundo Rajagopal (2011), *data mining* é o processo de extração de informação potencialmente útil de dados em estado bruto. Na opinião de Rajagopal (2011), a identificação de padrões de dados de clientes é muito importante para a tomada de decisão. A identificação atempada das novas tendências é muito importante para as operações da empresa e, ainda que hoje em dia as empresas lidem com grandes volumes de dados, na maioria dos casos não estão a extrair o melhor conhecimento destes. Neste sentido, o *data mining* surge como uma técnica fulcral de análise de dados. Aplicando esta técnica ao caso de uma loja de retalho de equipamentos de *outdoor* Rajagopal (2011) consegue identificar os clientes de elevado retorno, elevado valor e baixo risco, permitindo a segregação de clientes ou *customer clustering*.

Segundo Buckinx et al., (2007) a informação retirada do *software* de gestão de uma empresa deve ser complementada com modelos que permitam prever mais adequadamente a fidelização dos clientes e melhor segmentá-los. Através da utilização de um modelo de regressão linear e de dois modelos de *machine learning*, nomeadamente “*random forests*” e “*automatic relevance determination*”, os autores concluem que a fidelização dos clientes, avaliada pela variedade dos produtos comprados anteriormente, pode ser estimada com um grau de confiança elevado. Quanto melhor a estimação da fidelização do cliente e quanto melhor for a segmentação, maior a possibilidade de efetuar ações de marketing direcionadas que possam monetizar essa informação.

Hayashi & Ohsawa (2017) referem que maior problema no estabelecimento de um mercado de informação prende-se com o facto de tanto os detentores da informação como os seus utilizadores não reconhecerem a informação como um bem transacionável, como uma mercadoria. Os detentores da informação desconhecem o nível de investimento adequado na produção de informação devido à falta de conhecimento que detêm acerca do valor da informação. Desta forma, os preços a serem praticados serão os preços pedidos pelos detentores dessa informação, estando os utilizadores desta disponíveis para oferecer

preços que dependem do seu uso. Segundo os autores, através da utilização de “data jackets”, pode observar-se que o valor da utilidade da informação varia com o contexto da utilização da informação. “Data jackets” corresponde a uma técnica de partilha de informação de base de dados que publica o resumo das bases de dados como metadados. Desta forma, conseguindo evitar a divulgação pública dos dados, é possível a identificação do valor inerente dos dados, por parte dos interessados nessa informação, através da informação descrita nos “data jackets”.

Einav & Levin (2014) abordam como a *big data* tem um impacto cada vez maior na economia. Bases de dados em grande escala podem melhorar a forma como medimos, controlamos e descrevemos a atividade económica, através da possibilidade da utilização de novos modelos de estimação, sem substituir o senso comum ou a teoria económica, mas sim complementando-os.

2.4. Informação de identificação pessoal

Segundo Krishnamurthy & Wills (2010), o conceito de “Informação de identificação pessoal (IIP)”, da tradução de “*Personally identifiable information (PII)*”, refere-se à informação usada para distinguir ou traçar a identidade dum indivíduo específico, podendo ser utilizada por si só (através de identificadores diretos, como o número do cartão de cidadão por exemplo) ou utilizando informação pública relativa a esse indivíduo (através da combinação de identificadores indiretos, “*quase-identifiers*”, como o código postal e data de nascimento por exemplo). A Internet tem vindo a facilitar a recolha de IIP, desenvolvendo um mercado que explora esta recolha e a sua monetização¹. No seguimento da popularidade das redes sociais online e na conseqüente maior disponibilidade de informações pessoais na Internet, os autores concluem neste estudo que é possível a terceiras entidades estabelecer ligações entre a IIP contida nas redes sociais com as ações do utilizador, tanto em *sites* de redes sociais como em *sites* que não de redes sociais. Estas terceiras entidades conseguem rastrear os hábitos dos utilizadores em diferentes *sites* e servidores muitas vezes através das designadas “*cookies*”².

¹Conceito idêntico em https://pt.wikipedia.org/wiki/Informação_pessoalmente_identificável, acedido em 14/08/2018

² De acordo com o *Knowledge base* da Universidade do Indiana, disponível em <https://kb.iu.edu/d/agwm>, acedido em 14/08/2018, o termo “*cookies*” refere-se a mensagens trocadas entre o servidor e o navegador (“*browser*”), quando se visita um sitio na Internet, que contém informações relativas ao histórico da atividade do utilizador na Internet e informações relativas ao próprio utilizador.

Narayanan & Shmatikov (2010), por sua vez, referem que dados tornados anónimos podem ser re-identificados, através de algoritmos que analisam informação que permite distinguir uma pessoa de outra, nomeadamente através características humanas como preferências de consumo, transações comerciais efetuadas, histórico de navegação e de procura na Internet, entre outros, uma vez que tais características se mantêm razoavelmente estáveis ao longo do tempo e a probabilidade destas serem iguais entre duas pessoas é baixa. Bettini *et al.* (2006) concluem também que é possível ocorrer fuga de informação, sendo possível a re-identificação com base em “*quasi-identifiers*”.

O estudo de Krishnamurthy & Wills (2010) permite concluir ainda que por vezes existem fugas de informação relativa à IPP via redes sociais e, ainda que possa haver contratos entre essas terceiras entidades e as redes sociais que limitem o uso por essas entidades de informação que possam obter como resultado da interação do utilizador com as redes sociais, a existência de tais contratos ou o seu conteúdo não pode ser confirmada publicamente. Segundo os autores, um dos casos em que essa fuga de informação pode ocorrer é através de aplicações externas, instaladas através das redes sociais. Pela enorme cobertura noticiosa dada e pelo enquadramento com este tema cumpre referir que terá sido este tipo de fuga de informação que terá alegadamente ocorrido com uma aplicação da rede social Facebook que originou o escândalo conhecido em 2018 que envolvia a rede social e a Cambridge Analytica, relacionando essa fuga de informação com o resultado das eleições presidenciais nos Estados Unidos da América, em 2016 e o resultado do referendo conhecido com “*Brexit*” no Reino Unido³.

Com o objetivo de diminuir alguns dos riscos associados à facilidade com que hoje se trocam dados pessoais de utilizadores de serviços, em 25 de Maio de 2018 entrou em vigor o Regulamento Geral sobre a Proteção de Dados (RGPD), estipulando um conjunto único de regras sobre a proteção de dados para todas as empresas ativas na União Europeia (UE), independentemente da sua localização⁴. O RGPD vem reforçar a proteção dos direitos dos utilizadores de serviços, até essa data consagrados pela Lei nº67/98 de 26 de Outubro, que transpôs para a ordem jurídica portuguesa a Diretiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de outubro de 1995. As principais mudanças

³ Para melhor enquadramento do tema consultar: <https://www.bbc.com/news/business-43983958> e <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram>

⁴ Para consulta do Regulamento e aprofundamento do tema consultar: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_pt

são relativas ao território alargado de aplicação, à capacidade de fiscalização do regulador, ao valor das coimas e ao reforço das condições de consentimento e utilização dos dados. A Legislação de proteção de dados para empresas ativas fora da EU tem também sido objeto de atenção dos reguladores. Nos EUA, por exemplo, foi aprovado o *California Consumer Privacy Act of 2018 (CCPA)*, que entrará em vigor em 2020⁵, sendo considerado inspirado no RGPD⁶

⁵ Teor da Lei consultável em:

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

⁶Consultar artigo escrito por Jeremy Arnold em:

<https://www.forbes.com/sites/quora/2018/07/12/california-just-passed-a-new-data-privacy-bill-heres-what-it-means/#130a1db87a1b>

Capítulo 3. O valor da informação em empresas de comércio eletrônico: uma aplicação

A literatura analisada no capítulo anterior permitiu identificar alguns fatores que potenciam o valor da informação detida pelas empresas, tornando-se agora importante analisar esta matéria a nível empírico. Em primeiro lugar, é identificada a metodologia e são caracterizados os dados recolhidos e as suas limitações. Seguidamente, são identificados os componentes principais resultantes da Análise de Componentes Principais Categórica e é efetuado o agrupamento dos sujeitos, assim como o agrupamento das variáveis, em grupos homogéneos quanto a uma ou mais características comuns recorrendo à Análise de *Clusters*. Posteriormente, é construído um índice que possibilita mensurar o valor da informação em termos comparativos e são caracterizados os *sites* melhor posicionados no índice proposto.

3.1. Metodologia

Este estudo é concretizado através da aplicação da metodologia Análise de Componentes Principais Categórica (“*Categorical Principal Components Analysis*” (CATPCA)). A Análise de Componentes Principais (ACP), introduzida por Pearson (1901), desenvolvida inicialmente por Hotelling (1933) e a partir anos 60, é uma das técnicas de análise exploratória mais comuns quando se pretende analisar dados multivariados. Esta técnica permite transformar um conjunto de variáveis correlacionadas num conjunto menor de variáveis independentes, as componentes principais, combinações lineares das variáveis originais. O objetivo principal é, então, resumir a informação contida no conjunto de variáveis originais (correlacionadas e, portanto, de alguma forma redundantes), num conjunto menor de variáveis (independentes) perdendo o mínimo possível da informação disponível nas variáveis originais. Estas componentes podem ser consideradas “índices” ou indicadores que resumem a informação presente nas variáveis originais (Maroco, 2007).

A ACP é uma técnica aplicável apenas a variáveis quantitativas. A CATPCA é aplicável a variáveis medidas em escalas diferentes, através do procedimento designado por “*optimal scaling*”, que atribui quantificações numéricas às categorias de cada uma das variáveis qualitativas (medidas numa escala nominal ou ordinal), possibilitando o recurso a métodos de análise numérica. A CATPCA consiste, então, na quantificação das categorias e

na realização da ACP (Maroco, 2007).

Após a realização da CATPCA, é realizado um procedimento de “*clustering*”, nomeadamente “*clustering*” hierárquico, com o objetivo de encontrar semelhanças entre diferentes empresas (sujeitos da amostra).

Por último é efetuada uma análise multivariada dos dados tendo em vista propor um índice que possibilite mensurar o valor da informação em termos comparativos e são caracterizados os *sites* que obtiverem as melhores classificações no índice proposto.

Na análise dos dados será utilizado o programa estatístico IBM SPSS Statistics, versão 25.

3.2. Dados

Uma vez que o tema em análise é um tema ainda pouco explorado, não existem bases de dados disponíveis que permitam identificar o valor da informação. Neste sentido, são utilizados dados primários, isto é, especificamente recolhidos para a concretização dos objetivos do estudo. Neste estudo são recolhidos e analisados dados relativos à informação disponibilizada em *sites* de empresas de comércio eletrónico, através de uma técnica de amostragem não-aleatória ou não-probabilística denominada por amostragem de conveniência. A amostragem de conveniência, acidental ou casual, é uma técnica em que os elementos são selecionados pela sua conveniência. Neste tipo de amostragem a probabilidade de um determinado elemento pertencer à amostra é diferente da dos restantes elementos, não obedecendo assim a um dos princípios básicos da teoria das probabilidades. Assim, este tipo de amostra pode, ou não, ser representativo da população em estudo, podendo, portanto, conter algumas limitações (Maroco, 2007).

Pela natureza do objeto de estudo em análise, e pelas limitações que lhe são inerentes, não seria possível analisar toda a informação detida por qualquer tipo de empresa ou utilizar uma técnica de amostragem aleatória ou probabilística.

O foco no tipo de informação privada detida pelas empresas que é recolhida e disponibilizada nos respetivos *sites* prende-se com o facto de ser de esta poder ser acedida através da consulta aos respetivos *sites*. Outro tipo de informação privada que as empresas detenham apenas poderia ser acedida se esta fosse especificamente disponibilizada para o estudo.

A escolha de empresas de comércio eletrónico prende-se com o enquadramento

teórico efetuado em que é referida a possibilidade atual de estabelecer ligações entre a IIP, por vezes contida nas redes sociais, com as ações dos utilizadores e a possibilidade de rastreamento dos hábitos desses utilizadores em diferentes *sites*. O facto desses utilizadores em *sites* de comércio eletrónico serem principalmente clientes ou potenciais clientes, pode permitir à empresa monetizar a informação de forma mais direta.

Foram recolhidos dados relativos aos *sites* de 100 empresas de comércio eletrónico, nacionais e estrangeiras, por se considerar que seria um número praticável e suficiente para uma análise preliminar ao objeto deste estudo (Anexo 2). Uma vez que na literatura explorada não se encontrou nenhum estudo semelhante, optou-se pela adoção de algumas variáveis que se consideraram teoricamente apropriadas para a valoração da informação (Anexo 1). Para além de informação recolhida no *sites* das empresas da amostra, foram utilizadas outras fontes de informação de modo a permitir a produção de níveis de informação de acordo com os objetivos pretendidos.

No Anexo 2 pode consultar-se a listagem das 100 empresas da amostra e dos dados recolhidos, constando do Anexo 3 a descrição detalhada das variáveis utilizadas. Algumas das empresas da amostra foram selecionadas por serem empresas de comércio eletrónico de grande dimensão, enquanto outras foram selecionadas seguindo a sugestão da Alexa. A Alexa é uma empresa de análise de *sites* especializada no estudo do seu conteúdo e do seu tráfego⁷. Esta empresa analisa a ligação entre *sites*, procedendo à sua identificação e à sua classificação, disponibilizando no seu próprio *site* estatísticas e informações de tráfego eletrónico⁸. Uma das principais métricas que a empresa disponibiliza é o “Posicionamento de Tráfego Alexa” (do inglês “*Alexa Traffic Rank*”), conhecida simplesmente por “Posicionamento Alexa”, também conhecida por “Posicionamento de Tráfego Global”, que consiste numa estimativa da popularidade do *site*, atualizada diariamente, calculada com base na média de visitantes diários ao respetivo *site* e o número de páginas visualizadas nesse mesmo *site* nos três meses anteriores. O “Posicionamento de Tráfego Global”, pode ser utilizado para medir a popularidade dum *site* e para comparar a popularidade entre diferentes *sites*. A estimativa do “Posicionamento de Tráfego Global” é baseada nos dados do “painel de tráfego global”, que consiste numa amostra de milhões de utilizadores da

⁷ Conforme informação disponível em: <https://support.alexa.com/hc/en-us/articles/200450254-Why-is-Alexa-crawling-my-site->

⁸ Disponível em: <https://www.alexa.com/siteinfo>.

Internet que usam uma das várias extensões de navegação disponíveis para o efeito.⁹ Essa mesma amostra origina também outras estimativas, da mesma fonte, utilizadas como variáveis deste estudo, tais como a “percentagem de permanência”, “visualização de páginas diária por visitante”, “tempo diário no *site*” e “número de *sites* fonte”, cujas descrições se encontram no Anexo 3. Reforça-se que o “Posicionamento de Tráfego Global” é medido pela Alexa com base nos utilizadores que utilizam a extensão no navegador, sendo esses dados, provenientes da amostra de milhões de utilizadores da Internet, extrapolados. Tal facto pode originar dados que podem ser considerados imprecisos, especialmente para *sites* com estimativa de posicionamento baixo.

É importante referir que os resultados deste estudo podem ser considerados não extrapoláveis para a população em geral. Apesar das imperfeições reconhecidas da amostra, é possível considerar que as informações obtidas são uma das melhores fontes para a sugestão de um índice que permita mensurar o valor da informação em termos comparativos. A definição da metodologia de análise e valoração da informação é passível de ser aplicável a outras empresas e a outros tipos de informação detidos pelas empresas.

Antes da fase da análise dos resultados propriamente dita, procedeu-se a uma fase preliminar que consiste na recolha, organização e classificação dos dados de forma a poderem ser transformados em informação passível de ser analisada e interpretada.

O Anexo 3 contem a síntese da recolha de informação utilizada nesta fase do estudo e a descrição detalhada das variáveis utilizadas. Atente-se que foram utilizadas variáveis de diferentes tipos e com diferentes escalas. McCall (1998), como referido em Maroco (2007), refere que as variáveis estatísticas podem ser classificadas como qualitativas ou quantitativas, se atendermos apenas à sua natureza. As primeiras são variáveis cuja escala de medida apenas indica a presença de atributos ou a classificação em categorias discretas. As segundas são variáveis cuja escala de medida permite a ordenação e a quantificação de diferenças entre elas. No caso de não existir uma ordem das categorias, estamos perante variáveis qualitativas medidas numa escala nominal. Caso contrário, se existir uma ordenação, a variável diz-se medida numa escala ordinal. Por sua vez, as variáveis quantitativas podem ser medidas numa escala intervalar, em que as variáveis assumem valores quantitativos numa escala numérica com intervalos iguais que, contudo, não possui um zero absoluto, isto é, não possui uma medida de ausência de atributo, ou

⁹ Conforme informação disponível em: <https://support.alexa.com/hc/en-us/articles/200080859-What-exactly-is-the-Alexa-Traffic-Panel->

podem ser medidas numa escala razão, em que as variáveis assumem valores quantitativos com origem fixa, isto é, numa escala que possui um zero absoluto. No sentido de reduzir a complexidade dos cálculos, as variáveis originais classificadas como quantitativas e medidas numa escala razão foram transformadas, tendo em vista a redução do número de valores do atributo contínuo, dividindo a amplitude do atributo em intervalos, passando os rótulos dos intervalos a substituir os valores. Deste modo, estas variáveis foram transformadas em variáveis qualitativas, medidas numa escala ordinal de 1 (um) a 10 (dez). Na linguagem do SPSS esta transformação de dados é denominada como “recodificação de variáveis¹⁰”, que consiste na reafecção de valores de variáveis originais, ou na agregação destes valores, em novos valores numa nova variável “codificada”.

3.3. Análise de Componentes principais

Tendo em vista o objetivo de resumir a informação contida nas variáveis originais em componentes principais, utilizou-se uma CATPCA com o método de normalização da variável principal. Segundo Maroco (2007), este método otimiza a associação entre variáveis, sendo as coordenadas das variáveis no espaço os “*component loadings*”.

Como critério de retenção das componentes, utilizou-se a regra do valor próprio¹¹ superior a 1, observado no Quadro 1, uma vez que se pretendia que cada componente principal selecionada explicasse pelo menos tanta variância quanto cada uma das variáveis originais e, uma vez que estamos perante uma ACP com variáveis estandardizadas, este valor é 1. Jonhson (1998), como referido em Maroco (2007), defende que se uma componente principal tem valor próprio inferior a 1, provavelmente não é importante, ainda que essa regra possa nem sempre ser de aplicabilidade geral e possa em alguns casos levar à seleção de um número de componentes diferente do realmente necessário, como defendido por Cliff (1988), como referido em Maroco (2007).

A consistência interna de cada componente foi medida com o “Alfa de Cronbach”, sendo este uma medida de fidelidade de cada dimensão e do modelo no geral que é tanto melhor quanto maior for este indicador (Maroco, 2007).

¹⁰ Como se pode consultar em:

https://www.ibm.com/support/knowledgecenter/en/SSLVMB_24.0.0/spss/base/idh_reci_main.html

¹¹ Valor Próprio é um escalar relacionado com sistemas lineares de equações (equações matriciais). Para mais informação consultar: <http://mathworld.wolfram.com/Eigenvalue.html>

Quadro 1 – Sumário do modelo com 5 componentes

Dimensão (Componente)	Alfa de Cronbach	Variância Explicada	
		Total (<i>Valor Próprio</i>)	% da Variância
1	0,795	3,758	28,911
2	0,636	2,421	18,627
3	0,376	1,531	11,780
4	0,182	1,202	9,247
5	0,023	1,022	7,862
Total	0,974	9,935	76,427

Fonte: Adaptação do SPSS

De acordo com a regra do valor próprio superior a 1, é possível resumir a informação relacional entre as variáveis em 5 componentes ortogonais que explicam 76,4% da variância total das variáveis originais. A decisão sobre o número de componentes a reter é crucial na ACP e depende primeiramente da quantidade de informação disponível nas variáveis originais que é possível “dispensar”. Em experiências laboratoriais pode conseguir-se reter 95% ou mais da variação total, com 2 ou 3 componentes, mas em estudos sociais a maioria das vezes é necessário recorrer a 5 ou mais componentes para explicar 70 a 75% da variação total (Johnson, 1998, como referido em Maroco, 2007), sendo o caso do presente estudo.

O Anexo 4 e o Anexo 5 sustentam a seleção de 5 componentes, apresentando os sumários dos modelos alternativos correspondentes à seleção de 6 e 4 componentes, respetivamente. Pode verificar-se que a seleção de 6 componentes representaria já um valor próprio inferior a 1 e um Alfa de Cronbach negativo, indicando claramente que a seleção de 6 componentes não era fidedigna apesar de uma maior percentagem de variância explicada total em relação à seleção de 5 componentes (82,9% ao invés de 76,4%), enquanto que a seleção de 4 componentes representaria ainda valores próprios superiores a 1 e Alfas de Cronbach positivos, mas representaria uma percentagem inferior de variância explicada total (68,7%), comparativamente à correspondente a 5 componentes (76,4%),

O Quadro 2 reproduz os “pesos” (“*component loadings*”) de cada variável em cada componente, possibilitando a nomeação de cada componente de acordo com esses pesos. Sabendo que, segundo Maroco (2007), geralmente aceita-se como variáveis determinantes aquelas que apresentam pesos (correlação) pelo menos superiores a 0,5 em valor absoluto, é possível verificar que:

- A componente 1 é determinada, por ordem de grandeza, pelas variáveis “Opinião”, “Avaliação”, “Classificação”, “Tráfego”, “Fontes”, “Nativa” e “Comprados”;
- A componente 2 é determinada, por ordem de grandeza, pelas variáveis “Tempo”, “Visualização” e “Permanência”;
- A componente 3 é determinada, por ordem de grandeza, pelas variáveis “Fontes” e “Comparação”;
- A componente 4 é determinada, por ordem de grandeza, pelas variáveis “Acessórios” e “Nativa”;
- A componente 5 é determinada, por ordem de grandeza, pelas variáveis “Similares” e “Acessórios”.

Repare-se que algumas variáveis estão correlacionadas com mais do que uma componente, nomeadamente a variável “Acessórios” (componentes 4 e 5), a variável “Fontes” componentes (1 e 3) e a variável “Nativa” (componentes 1 e 4).

Quadro 2 - Correlações entre variáveis e componentes

Variáveis	Componentes				
	1	2	3	4	5
Similares	0,482	-0,173	-0,015	-0,168	0,621
Acessórios	0,139	0,064	-0,080	0,613	0,615
Comprados	0,502	0,030	0,088	0,332	0,039
Comparação	0,383	-0,065	0,505	-0,040	-0,235
Classificação	0,755	-0,276	-0,500	-0,181	-0,059
Opinião	0,768	-0,189	-0,488	-0,140	-0,119
Avaliação	0,760	-0,254	-0,256	-0,060	-0,073
Tráfego	0,714	0,382	0,464	-0,170	0,087
Permanência	-0,374	0,571	-0,297	-0,454	0,197
Visualização	0,088	0,846	-0,328	0,129	-0,113
Tempo	0,168	0,920	-0,180	0,009	-0,006
Fontes	0,628	0,317	0,513	-0,329	0,118
Nativa	0,539	0,267	0,040	0,519	-0,326

Fonte: Adaptação do SPSS

No modelo, é possível interpretar o significado de cada uma das componentes. É de referir, contudo, que este tipo de interpretação pode ser sujeita a alguma subjetividade. A componente 1 pode ser interpretada como definindo a “Comunidade digital”, já que os pesos das variáveis “Opinião”, “Avaliação”, “Classificação”, “Tráfego”, “Fontes”, “Nativa” e “Comprados” são elevados nesta componente. A componente 2 define essencialmente a “Atratividade do site”, através da forte correlação com as variáveis “Tempo”,

“Visualização” e “Permanência”. A componente 3 pode ser interpretada como “Informação sobre produtos comparáveis”, devido ao peso da variável “Comparação” e a componente 4 interpretada como “Informação sobre produtos acessórios”, devido à representação da variável “Acessórios”. Por último, a componente 5 pode ser interpretada como “Informação sobre produtos similares”, pela forte correlação com a variável “Similares”. O Quadro 3 resume esta informação.

Quadro 3 - Denominação das componentes e resumo das respectivas variáveis com maior correlação

Componente	Denominação	Variáveis
1	Comunidade digital	Opinião
		Avaliação
		Classificação
		Tráfego
		Fontes
		Nativa
		Comprados
2	Atratividade do site	Tempo
		Visualização
		Permanência
3	Informação sobre produtos comparáveis	Comparação
4	Informação sobre produtos acessórios	Acessórios
5	Informação sobre produtos similares	Similares

Fonte: Elaboração própria

3.4. Análise de *clusters*

A análise de grupos ou de “*clusters*”, é uma técnica exploratória de análise multivariada que permite agrupar sujeitos (ou variáveis) em grupos homogêneos quanto a uma ou mais características comuns. Este agrupamento pode ser efetuado através de medidas de semelhança (ou proximidade) ou medidas de dissemelhança (ou distância), inicialmente entre dois sujeitos (ou variáveis) e seguidamente entre dois *clusters*, utilizando métodos hierárquicos ou não hierárquicos (Maroco, 2007). Existem numerosas medidas de semelhança e de dissemelhança, cuja utilização depende essencialmente das escalas de medidas das variáveis. Por norma, os sujeitos são agrupados de acordo com algum tipo de distância métrica enquanto as variáveis são agrupadas de acordo com medidas de

correlação (Jonhson & Wichern, 2002, como referido em Maroco, 2007). Neste estudo, ao agrupar sujeitos, foi utilizada a medida de dissemelhança métrica conhecida como “quadrado da distância euclidiana”, correspondendo esta ao quadrado do comprimento da reta que une duas observações num espaço p -dimensional, para p -variáveis. Para p -variáveis, a distância euclidiana entre os sujeitos i e j é dada por (Jonhson & Wichern, 2002, como referido em Maroco, 2007):

$$D_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (3.4.1)$$

Onde x_{ik} é o valor da variável k no sujeito i e x_{jk} é o valor da variável k no sujeito j . De igual modo, de forma matricial $D_{ij} = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)'(\mathbf{X}_i - \mathbf{X}_j)}$ onde \mathbf{X}_i e \mathbf{X}_j representam os vetores p -dimensionais das observações relativas aos indivíduos i e j . Segundo Maroco (2007), a distância euclidiana pode ser calculada a partir de variáveis originais, em que variáveis com maior amplitude apresentarão maiores distâncias euclidianas, ou a partir de variáveis estandardizadas, em que a distância euclidiana não é influenciada pela amplitude das variáveis.

Com o objetivo de agrupar os *sites* de acordo com as semelhanças que estes possuem nas componentes resultantes da CATPCA, e seguindo a sugestão de boa prática estatística referida por Maroco (2007), foi efetuada uma análise hierárquica (com o método da menor distância¹²) como técnica exploratória para a decisão do número de *clusters* a reter, seguida de uma análise não hierárquica (*k-means*, neste caso) para refinar e interpretar a solução de *clusters*. A probabilidade de se classificar erradamente determinado sujeito em determinado *cluster* é menor nos métodos não hierárquicos, sendo, portanto, mais rigorosa a classificação dos sujeitos em cada *cluster* nos métodos não hierárquicos (Maroco, 2007).

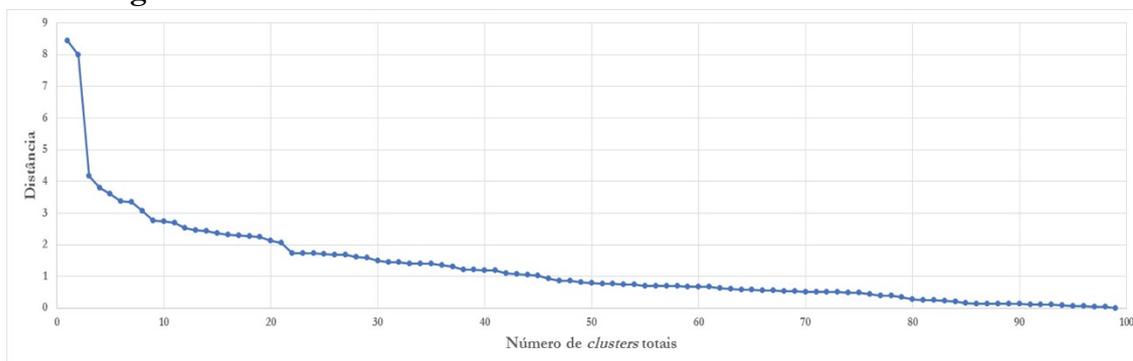
O agrupamento dos *sites* foi efetuado com base nas componentes resultantes da CATPCA e as correspondentes classificações (“*scores*”) de cada *site*, que podem ser consultadas no Anexo 6. Estas variáveis estavam já estandardizadas, apresentando amplitudes semelhantes e contribuindo de igual modo para a solução. De outro modo, as

¹² Método da Menor distância. “*Nearest Neighbor*” ou “*Single linkage*”: neste método agregativo, após a formação do primeiro *cluster* (sendo este no passo inicial apenas um objeto), a distância deste aos restantes objetos é a menor das distâncias de cada um dos objetos que constituem esse *cluster* a cada um dos restantes objetos (Maroco, 2007).

variáveis teriam ainda que ser transformadas, uma vez que distância euclidiana é sensível à amplitude destas.

Numa primeira fase, como referido anteriormente, foi efetuada uma análise hierárquica (com o método da menor distância) como técnica exploratória para a decisão do número de *clusters* a reter. O resultado deste procedimento está representado num dendrograma (Anexo 7). O Anexo 8 identifica os *clusters* que foram agrupados em cada passo e as respetivas distâncias entre cada par de *clusters*. Foi identificado o número de *clusters* totais no final de cada passo e foi calculada a diferença de distâncias entre cada passo tendo em vista qualificar essa diferença como pequena ou grande. Segundo Maroco (2007), se a distância entre dois *clusters* é pequena, estes devem ser agregados, se for grande, estes devem manter-se separados. Para efeitos deste estudo, foram consideradas como diferenças de distâncias grandes aquelas acima de 0,37 por ser um valor que se destaca no Anexo 8. Neste sentido, pode observar-se que tal critério apenas está presente nas diferenças de distâncias entre um e dois *clusters*, entre dois e três *clusters* e entre três e quatro *clusters*, sendo essas diferenças de 0,438; 3,829 e 0,370 respetivamente. Assim, o método da menor distância utilizado como técnica exploratória na análise hierárquica sugere quatro como o número de *clusters* a reter. Adicionalmente, ao observar-se a Figura 1, que representa o número de *clusters* em função das suas distâncias, nota-se que o declive da reta que une a distância entre quatro e cinco *clusters* é já relativamente pequeno em relação ao declive que une a distância entre um e dois, dois e três e três e quatro, pelo que, de acordo com esta observação, a sugestão do número de *clusters* a reter será igualmente de quatro.

Figura 1 - Número de *clusters* e suas distâncias



Fonte: Elaboração própria

Após a decisão do número de *clusters* a reter, procedeu-se a uma análise não hierárquica para refinar e interpretar a solução obtida de quatro *clusters* através do

procedimento *k-means*. A classificação dos sujeitos em cada *cluster* é mais rigorosa nos métodos não hierárquicos (Maroco, 2007). A classificação de cada um dos 100 *sites* na solução refinada com o método *k-means* com $k=4$, assim como a distância de cada *site* em relação ao centróide do *cluster* que permite avaliar quão semelhante (ou dissemelhante) aquele é, encontra-se no Anexo 9. O Anexo 10 representa os *sites* que constituem cada um dos *clusters*.

No Quadro 4 apresentam-se os centros dos *clusters* para cada variável onde é possível verificar que:

- O *cluster 1* corresponde ao grupo de *sites* cujos valores médios absolutos são altos nas componentes 1 e 4 e baixos nas componentes 2, 3 e 5. Assim, o *cluster 1*, é composto por *sites* que em média são muito desenvolvidos quanto à Comunidade digital e quanto à Informação sobre produtos acessórios e que se caracterizam como pouco desenvolvidos em média no que respeita à quanto à Atratividade do *site*, à Informação sobre produtos comparáveis e quanto à Informação sobre produtos similares. Pode-se, portanto, classificar este *cluster* como sendo respeitante a *sites*, que na sua maioria são “*Sites* pouco atrativos, com comunidade digital muito desenvolvida, com informação muito desenvolvida de produtos acessórios e com informação pouco desenvolvida de produtos comparáveis e similares”;

- O *cluster 2* é composto pelos *sites* cujos valores médios absolutos são intermédios nas componentes 2, 3 e 5 e baixos nas componentes 1 e 4. Deste modo, o *cluster 2* corresponde ao grupo de *sites* que em média são desenvolvidos quanto à Atratividade do *site*, quanto à Informação sobre produtos comparáveis e quanto à Informação sobre produtos similares e que se definem como pouco desenvolvidos em média no que concerne à Comunidade digital e à Informação sobre produtos acessórios. Este *cluster* pode, então, ser classificado como sendo correspondente a “*Sites* atrativos, com comunidade digital pouco desenvolvida, com informação desenvolvida de produtos comparáveis e similares e com informação pouco desenvolvida de produtos acessórios”;

- O *Cluster 3* é correspondente ao grupo de *sites* cujos valores médios absolutos são altos na componente 3, intermédios na componente 4 e baixos nas componentes 1, 2 e 5. Assim, o *cluster 3*, é composto por *sites* que em média são muito desenvolvidos quanto à Informação sobre produtos comparáveis, desenvolvidos quanto à Informação sobre produtos acessórios e pouco desenvolvidos quanto à Comunidade digital, à Atratividade do *site* e à Informação sobre produtos similares. Pode-se, portanto, classificar este *cluster* como sendo respeitante a “*Sites* pouco atrativos, com comunidade digital pouco desenvolvida,

com informação muito desenvolvida de produtos comparáveis, informação desenvolvida de produtos acessórios e com informação pouco desenvolvida de produtos similares”;

- O *cluster* 4, por último, é composto pelos *sites* cujos valores médios absolutos são altos nas componentes 2 e 5, intermédios na componente 1 e baixos nas componentes 3 e 4. Deste modo, o *cluster* 4 corresponde ao grupo de *sites* que em média são muito desenvolvidos quanto à Atratividade do *site* e quanto à Informação sobre produtos similares, desenvolvidos quanto à Comunidade digital e pouco desenvolvidos quanto à Informação sobre produtos comparáveis e quanto à Informação sobre produtos acessórios. O *cluster* 4 pode, então, ser classificado como sendo correspondente a “*Sites* muito atrativos, com comunidade digital desenvolvida, com informação muito desenvolvida de produtos similares, informação pouco desenvolvida de produtos comparáveis e acessórios”.

Quadro 4 - Centros dos *clusters* em cada componente

	Centro do <i>Cluster</i>			
	1	2	3	4
Componente 1	1,16719	0,07260	0,11861	-0,92017
Componente 2	-0,06000	-0,11849	0,03612	0,16716
Componente 3	0,31569	-0,65717	1,06197	-0,42051
Componente 4	1,48323	-0,24430	-0,65534	0,22831
Componente 5	-0,22523	0,72919	-0,02000	-0,90952

Fonte: Adaptação do SPSS

	Pouco desenvolvido
	Desenvolvido
	Muito desenvolvido

Com o objetivo de agrupar também as componentes resultantes da CATPCA de acordo com as semelhanças que estas possuem foi efetuada uma análise hierárquica de *clusters* a estas variáveis, através do método da menor distância e utilizando como medida de semelhança os coeficientes de correlação.

O resultado deste procedimento está representado num dendrograma (Anexo 11). Segundo este, as componentes 3 e 4 estão muito próximas (fortemente correlacionadas), assim como as componentes 1 e 2, sendo que estas duas componentes associam-se àquelas a uma maior distância. Pelo contrário, a componente 5 é muito “dissemelhante” de todas as outras. Este dendrograma sugere, assim, a existência de 2 *clusters* de variáveis: um *cluster* com as componentes 1, 2, 3 e 4; e outro *cluster* apenas com a componente 5.

É importante referir que a análise de *clusters* é uma técnica multivariada que procura

agrupar objetos (sujeitos ou variáveis) mais ou menos homogêneos segundo critérios mais ou menos heurísticos e que não possui sólidos fundamentos teóricos (Maroco, 2007).

3.5. Índice do valor da informação

Nesta fase do estudo, pretende-se contribuir com uma proposta de um índice do valor da informação em empresas de comércio eletrónico. Identificámos anteriormente as correspondentes componentes principais reveladas pela CATPCA. Espera-se agora que estas componentes contribuam de forma direta para a construção daquele índice, sendo este índice um dos objetivos principais deste estudo. O Quadro 5 apresenta as variáveis a incluir no índice.

Quadro 5 - Variáveis a incluir no índice do valor da informação

Variável (Componente)	Denominação	Descrição	% da Variância Explicada
1	Comunidade digital	Classificação estandardizada ("score") quanto à Informação sobre Opinião de clientes, Avaliação de opinião, Classificação de produtos, Tráfego do site, Sites fonte, Natividade digital e Informação sobre produtos comprados	28,911
2	Atratividade do site	Classificação estandardizada ("score") quanto ao Tempo no site, Visualização de páginas e Permanência	18,627
3	Informação sobre Produtos comparáveis	Classificação estandardizada ("score") quanto à Informação sobre produtos comparáveis	11,780
4	Informação sobre Produtos acessórios	Classificação estandardizada ("score") quanto à Informação sobre produtos acessórios	9,247
5	Informação sobre Produtos similares	Classificação estandardizada ("score") quanto à Informação sobre produtos similares	7,862
Total			76,427

Fonte: Elaboração própria

Na construção do índice, é atribuído como ponderador de cada variável a percentagem da variância explicada. O cálculo do índice comparativo, será, portanto, definido da seguinte forma:

$$\text{Índice do valor da informação} = \sum_{i=1}^p (\text{variância}_i * \text{componente}_{ij}) \quad (3.5.1)$$

Onde $variância_i$ é a percentagem da variância explicada na componente i , $componente_{ij}$ é a classificação (“score”) da componente i no sujeito j , e p é o número de componentes principais seleccionadas (variáveis). No caso específico deste estudo,

Índice do valor da informação = 0,28911 * Comunidade digital + 0,18627 * Atratividade do *site* + 0,1178 * Informação sobre Produtos comparáveis + 0,09247 * Informação sobre Produtos acessórios + 0,07862 * Informação sobre Produtos similares.

Aplicando à expressão acima as classificações (“scores”) correspondentes a cada *site*, apresentadas no Anexo 6, resultantes da CATPCA, pode identificar-se os 20 *sites* melhor classificados em termos do índice do valor da informação. Estes *sites* estão identificados no Quadro 6, podendo ser consultado o Índice relativo aos 100 *sites* no Anexo 12.

Quadro 6 - 20 sites melhor classificados em termos do índice do valor da informação (Top 20)

Site	Índice	Cluster
Amazon	1,172	1
Asos	0,847	2
Bang good	0,743	1
Deal Extreme	0,727	1
HM	0,721	3
Gearbest	0,703	1
Aliexpress	0,686	2
Ikea	0,666	2
DH Gate	0,660	1
Yoox	0,631	1
Ebay	0,626	3
Alibaba	0,601	3
Wayfair	0,592	1
Homedepot	0,589	3
Farfetch	0,505	1
Tomtop	0,471	1
Urbanoutfitters	0,470	2
Bestbuy	0,421	3
Light in the box	0,386	1
Zara	0,369	3

Fonte: Elaboração própria

Como se pode observar no Quadro 6, a empresa melhor classificada em termos de Índice do valor da informação é a Amazon. Trata-se de uma empresa pertencente ao *cluster* 1, como tal, pode dizer-se que a Amazon se caracteriza como tendo um *site* com comunidade digital muito desenvolvida, com informação muito desenvolvida de produtos acessórios e com informação ainda pouco desenvolvida de produtos comparáveis e similares, comparativamente aos outros *sites* englobados na amostra deste estudo. Em termos comparativos também, trata-se de um *site* ainda pouco atrativo, no sentido em que apresenta uma média de páginas visualizadas diariamente no *site* por visitante, uma média do tempo passado diariamente no *site* por visitante e uma percentagem de visitantes que se mantêm no *site* visualizando para além de uma página, comparativamente baixas, correspondendo à 9^a, 10^a e 41^a posição, respetivamente, conforme Anexos 13, 14 e 15.

No Quadro 6, correspondente aos 20 *sites* melhor classificados em termos de Índice do valor da informação, pode verificar-se que metade dos *sites* pertencem ao *cluster* 1, quatro *sites* pertencem ao *cluster* 2, seis *sites* ao *cluster* 3 e nenhum *site* pertence ao *cluster* 4. Isto poderá indicar que contribui para uma melhor classificação, em termos de índice do valor da informação, *sites* com comunidade digital muito desenvolvida, juntamente com Informação sobre produtos acessórios também muito desenvolvida. *Sites* com Informação sobre produtos comparáveis muito desenvolvida parecem conseguir uma melhor classificação comparativamente aos *sites* com atratividade desenvolvida, em termos de páginas visualizadas, tempo passado no *site* e percentagem de permanência. Informação muito desenvolvida relativamente a produtos similares, parece não contribuir tanto para uma melhor classificação em termos do índice do valor da informação. *Sites* com um desenvolvimento intermédio equilibrado no que respeita à sua atratividade, à Informação sobre produtos comparáveis e à Informação sobre produtos similares, parecem também obter uma classificação razoável em termos do índice do valor da informação.

Repare-se que alguns *sites* apresentam valor negativo do Índice do valor da informação (Anexo 12). Tal facto poderá ser interpretado como indicando que alguns *sites* têm um comportamento de alguma forma diferente no que respeita ao valor da informação, em relação aos *sites* melhor classificados, devido ao desenvolvimento menor em alguma(s) das cinco componentes.

3.6. Características dos *sites* melhor posicionados no Índice comparativo

Nesta fase do estudo pretende-se caracterizar os *sites* que obtiveram as melhores classificações no índice proposto, nomeadamente as características dos 20 *sites* melhor classificados em termos de Índice do valor da informação (denominados de “Top 20”), dando especial atenção ao *site* melhor classificado (Correspondente ao Sujeito 2, ou seja, Amazon).

Uma vez que neste estudo estamos perante cinco componentes principais identificadas, um mapa dimensional com as cinco componentes e as correspondentes classificações (“*scores*”) de cada *site* (Anexo 6) não seria visualmente facilmente interpretável. Neste sentido, foram elaborados mapas bidimensionais com pares de componentes, encontrando-se estes nos Anexos 16 a 25.

De acordo com o posicionamento de cada *site* (sujeito) em cada mapa bidimensional das componentes principais, e a sua posição relativa às variáveis originais, após a CATPCA, é possível caracterizar os *sites* quanto a cada uma das componentes e a cada uma das variáveis originais.

No Anexo 16 pode observar-se que, de acordo com o posicionamento dos *sites* no mapa dimensional e os pesos das variáveis nas componentes 1 e 2 é possível caracterizar a comunidade digital dos *sites* Top 20 como desenvolvida em termos de Opinião de clientes, Avaliação de opinião, Classificação de Produtos, Posicionamento de tráfego Global, Número de *sites* fonte, Digitais e em termos de Compras em conjunto, isto é, os Top 20 apresentam *scores* elevados na componente 1, à exceção de quatro. Os Top 20 representam mais de metade das 20 primeiras posições relativas à componente 1. Pode ainda caracterizar-se a atratividade dos *sites* Top 20 como desenvolvida em termos de Tempo diário no *site*, Visualização de páginas e em termos de Percentagem de permanência, uma vez que os Top 20 apresentam de um modo geral *scores* elevados na componente 2, ainda que cinco dos Top 20 apresentem *scores* baixos nesta componente e que possam, portanto, ser considerados pouco desenvolvidos no que respeita à atratividade do *site*. Os Top 20 representam menos de metade das 20 primeiras posições relativas à componente 2. É ainda observável que o *site* da Amazon apresenta um *score* de comunidade digital claramente superior aos restantes Top 20 e um *score* de atratividade do *site* que não é dos mais elevados, sendo ultrapassada por seis Top 20 e por dois *sites* não Top 20.

No Anexo 17, por sua vez, é possível observar o posicionamento dos *sites* no mapa dimensional e os pesos das variáveis na componente 1 novamente e, desta vez, na

componente 3. Pode-se, então, caracterizar a Informação sobre produtos comparáveis como desenvolvida, uma vez que os Top 20 apresentam de um modo geral *scores* elevados na componente 3, ainda que oito dos Top 20 possam ser considerados como pouco desenvolvidos no que respeita à Informação sobre produtos comparáveis, uma vez que apresentam *scores* baixos nesta componente. Os Top 20 representam menos de metade das 20 primeiras posições relativas à componente 3. É ainda possível observar que o *site* da Amazon apresenta um *score* de Informação sobre produtos comparáveis que não é dos mais elevados, sendo ultrapassada por cinco Top 20 e por onze *sites* não Top 20.

No que concerne ao Anexo 18, pode observar-se o posicionamento dos *sites* no mapa dimensional e os pesos das variáveis na componente 1 mais uma vez e, desta vez, na componente 4. É possível, então, caracterizar a Informação sobre produtos acessórios como desenvolvida, uma vez que os Top 20 apresentam de um modo geral *scores* elevados na componente 4, ainda que sete dos Top 20 apresentem *scores* baixos nesta componente e que possam, portanto, ser considerados pouco desenvolvidos no que respeita à Informação sobre produtos acessórios. Os Top 20 representam menos de metade das 20 primeiras posições relativas à componente 4. É ainda possível observar que o *site* da Amazon apresenta um *score* de Informação sobre produtos acessórios que ocupa a 9ª posição do Top 20, não sendo, portanto, dos *scores* mais elevados.

No Anexo 19 é possível observar o posicionamento dos *sites* no mapa dimensional e os pesos das variáveis na componente 1 novamente, juntamente com a componente 5, desta vez. Pode-se, então, caracterizar a Informação sobre produtos similares como desenvolvida, uma vez que os Top 20 apresentam de um modo geral *scores* elevados na componente 5, ainda que sete dos Top 20 possam ser considerados como pouco desenvolvidos no que respeita à Informação sobre produtos similares, uma vez que apresentam *scores* baixos nesta componente. Os Top 20 representam menos de um quarto das 20 primeiras posições relativas à componente 5. É ainda possível observar que o *site* da Amazon apresenta um *score* de Informação sobre produtos similares dos mais baixos, assumindo a 15ª posição dos Top 20 e sendo ultrapassada por vários *sites* não Top 20.

Nos Anexos 20 a 25, por último, é possível corroborar a interpretação feita quanto a cada uma das cinco componentes, observando o posicionamento dos *sites* nos mapas dimensionais relativos às conjugações restantes de cada par das cinco componentes.

Pelo exposto, as características que parecem mais contribuir para um melhor posicionamento em termos do índice do valor da informação proposto, observadas nos

sites Top 20, é o desenvolvimento respeitante à Comunidade digital, em termos de Opinião de clientes, Avaliação de opinião, Classificação de Produtos, Posicionamento de tráfego Global, Número de *sites* fonte, em termos Digitais e em termos de Compras em conjunto. Segue-se o desenvolvimento respeitante à Informação sobre produtos acessórios, o desenvolvimento respeitante à Informação sobre produtos comparáveis e à Atratividade do *site*, em termos de Tempo diário no *site*, Visualização de páginas e em termos de Percentagem de permanência. Por último, surge o desenvolvimento respeitante à Informação sobre produtos similares.

Conforme observado, o *site* melhor classificado, relativo à Amazon, apresenta-se, comparativamente aos outros *sites* da amostra, como um *site* com comunidade digital muito desenvolvida, com informação muito desenvolvida de produtos acessórios, com informação ainda pouco desenvolvida de produtos comparáveis e similares. Apresenta-se ainda como um *site* com pouca atratividade, no sentido em que apresenta uma média de páginas visualizadas diariamente no *site* por visitante, uma média do tempo passado diariamente no *site* por visitante e uma percentagem de visitantes que se mantêm no *site* visualizando para além de uma página, comparativamente baixas.

Capítulo 4. Índice financeiro do valor da informação

Neste capítulo, pretende-se analisar um índice comparativo com base em informação financeira de empresas, seguido o método de mensuração do valor da informação apresentado por Laney (2017). De acordo com Laney (2017), é possível observar o valor da informação das empresas, através da avaliação interna que estas fazem, juntamente com as instituições financeiras envolvidas, quando perante uma Oferta Pública Inicial, usualmente referida como IPO, da tradução de “*Initial Public Offer*”, em que as ações dessa empresa são vendidas ao público em geral pela primeira vez. Usando o exemplo da IPO do Facebook, Laney (2017) refere que o valor da informação será facilmente observável em IPO de empresas cujo principal bem é informação. Neste tipo de empresas, a diferença entre o valor da avaliação interna da empresa, isto é, entre o valor que a empresa prevê de capitalização bolsista na IPO, e o valor contabilístico dos ativos reportáveis, à data do último reporte prévio à IPO, corresponde essencialmente ao valor da informação que a empresa detém, uma vez que os únicos ativos não reportáveis que a empresa terá serão, na sua maioria, ativos correspondentes ao valor da informação.

O método de mensuração do valor da informação apresentado por Laney (2017), de certa forma, vem lembrar o Rácio Q, ou Rácio Tobin, que relaciona o valor de mercado de uma empresa com o custo de substituição dos ativos dessa empresa, apresentado por James Tobin (Tobin, 1977). Este rácio dá-nos a proporção do valor de mercado da empresa em relação ao valor total dos ativos (reportáveis), através da divisão daquele valor por este, indicando se empresa está ou não sobreavaliada. Algumas empresas da indústria das tecnologias de informação e comunicação baseadas na Internet podem não ter tantos ativos físicos ou tangíveis como empresas de outras indústrias. De acordo com o Rácio Tobin, estas empresas podem ter tendência para serem identificadas como estando sobreavaliadas, uma vez que os seus valores de mercado muitas vezes excedem os custos de reposição dos seus ativos, ou seja, a venda dos seus ativos não iria atingir as suas capitalizações bolsistas. Em muitas destas empresas o seu principal ativo é informação e, como este ativo não obedece às condições para ser considerado reportável, a análise deste rácio pode revelar-se insuficiente para avaliar se a empresa está ou não sobreavaliada, tornando-se importante, portanto, analisar adicionalmente outros indicadores financeiros relativos a estas empresas. Desde 1900, o máximo atingido em termos do Rácio Tobin no

mercado dos EUA, terá sido precisamente no final dos anos 90, com a forte alta das ações das empresas daquela indústria, numa bolha conhecida como “bolha da Internet” ou “bolha das dot com”¹³.

Laney (2017), por sua vez analisa o valor da informação, não em termos de proporção, mas em termos absolutos, através da diferença entre aquele valor de mercado (capitalização bolsista) na IPO e o valor daqueles ativos. Pode-se então dizer que uma empresa com algum valor de informação, terá obrigatoriamente um Rácio Q acima de 1 (um), sendo que este Rácio acima de 1 (um), não é interpretado como estando a empresa sobreavaliada, mas sim interpretado como detendo a empresa informação aquando da IPO cujo valor não está incluído no valor dos seus ativos, por não obedecerem aos requisitos para serem considerados ativos reportáveis.

O valor da informação das empresas com base em informação financeira, será melhor observado quando perante uma IPO, uma vez que é nesse momento que é efetuada uma avaliação da capitalização bolsista, com base em informação detida pelos gestores internos da empresa. De acordo com Grinblatt (2002), esta informação inclui nomeadamente o valor presente de fluxos de caixa (“*cash flows*”) futuros e, deste modo, o “*full information value*” da empresa, usualmente referido como o valor intrínseco da empresa. Segundo Grinblatt (2002), os investidores não conseguem observar o valor dos fluxos de caixa operacionais da empresa, assim como os detalhes dos gastos de investimento, pelo que não seria fácil efetuarem uma avaliação da capitalização bolsista. Por sua vez, após a IPO, a observação do valor da informação das empresas com base em informação financeira, utilizando a capitalização bolsista diária da empresa pode ser criticável quanto à sua eficácia, uma vez que esta capitalização oscila a cada momento e com base na lei da oferta e da procura, que grande parte das vezes pouco terá a ver com o valor da informação privada da empresa.

É importante referir que a avaliação interna que as empresas fazem quando perante uma IPO pode ser considerada como não completamente fidedigna para observar o valor da informação destas. Como se trata de uma situação em que se está perante assimetria de informação, o valor da avaliação interna pode ir ao encontro da teoria da sinalização, referida por Spence (1973), uma vez que a empresa pode utilizar esta avaliação interna para convencer os investidores do valor da empresa. Pode aqui observar-se também por vezes

¹³ Segundo informação consultada a 06/09/2018 em:
<https://www.advisorperspectives.com/dshort/updates/2018/08/02/the-q-ratio-and-market-valuation-july-update>

situações de “*underpricing*”, em que o preço inicial das ações definido para a IPO é abaixo do seu valor de mercado previsto para o final do primeiro dia de negociação, originando retornos médios positivos no primeiro dia. Segundo Grinblatt (2002), o “*underpricing*”, é uma situação recorrente em IPOs de todo o mundo, utilizada pelas empresas como incentivo à subscrição da IPO por parte de investidores, uma vez que estes vêm aqui a oportunidade de obterem retornos positivos ao adquirem as ações a um preço abaixo do seu valor de mercado previsto e venderem-nas posteriormente. Grinblatt (2002) refere ainda que o “*underpricing*” pode ser benéfico para a empresa se esta pretender fazer ofertas públicas de ações posteriores à IPO, uma vez que os investidores poderão ter maior tendência para subscrever as ofertas públicas posteriores à IPO se tiverem obtido retornos positivos nesta. Adicionalmente, a IPO pode ser motivada pelo acesso ao mercado de capitais tendo como objetivo o financiamento de projetos de investimento ou o financiamento da aquisição de outras empresas, pelo que a diferença entre o valor da capitalização bolsista prevista e o valor contabilístico dos ativos reportáveis, à data do último reporte prévio à IPO, pode não corresponder na sua totalidade ao valor da informação da empresa, podendo esta situação notar-se ainda mais em empresas cujo principal bem não é informação. Por último, convém referir que o preço inicial das ações na IPO, e consequentemente o valor da capitalização bolsista prevista, é muitas vezes definido com base no “*Price to Earnings Ratio*” (normalmente designado de PER, corresponde ao rácio entre o preço da ação e o resultado por ação) de empresas comparáveis na respetiva indústria.

Apesar das limitações reconhecidas da avaliação interna efetuada pelas empresas numa IPO, neste estudo, foram recolhidas informações relativas a dez das maiores IPOs de empresas tecnológicas dos últimos anos, apresentadas no Anexo 26, por ordem alfabética do nome da empresa. Foi calculado o valor da informação para cada uma destas empresas, seguindo o critério de Laney (2017), correspondente à diferença entre a capitalização bolsista prevista na IPO e o valor dos ativos. Foi, portanto, considerado que esta diferença corresponde na sua totalidade ao valor da informação da empresa, o que pode ser criticável, uma vez que as motivações da IPO podem ser distintas, tal como referido anteriormente, e como tal neste valor podem estar incluídos outros itens para além do valor da informação. Ainda relativamente ao Anexo 26, cabe referir que estão incluídas cinco empresas cuja base geradora de receitas é apenas informação, isto é, cujas receitas, na sua maioria em publicidade, derivam da utilização da informação como o seu principal ativo e

bem transacionado, e cinco empresas cuja base geradora de receitas incluí outros itens para além de informação, nomeadamente receitas com comissões e subscrições. Procedeu-se a uma análise conjunta das dez empresas e seguidamente a uma análise para apenas as cinco empresas cuja base geradora de receitas é apenas informação, uma vez que, de acordo com Laney (2017) o valor da informação das empresas será mais facilmente observável em IPO de empresas cujo principal bem é informação.

O Quadro 7 apresenta o valor de informação das dez empresas por ordem de grandeza, calculado segundo Laney (2017). Para uma melhor interpretação, foi criado um Índice comparativo, através da padronização, em que o valores de informação das dez empresas foram padronizados numa mesma escala, através da diferença entre cada um dos valores e a média, dividido pelo desvio padrão, ou seja:

Índice financeiro do valor da informação =

$$= \frac{\text{Valor da informação} - \text{Média do Valor da Informação na amostra}}{\text{Desvio Padrão do Valor da Informação na amostra}} \quad (4.1)$$

Um valor positivo indica que o valor é maior que a média, e um valor negativo indica que o valor é menor que a média.

Quadro 7 - Índice financeiro do valor da informação das dez empresas

Empresa	Valor da informação	Índice	Base geradora de receitas
Alibaba Group Holding Limited	133.366.000.000	2,2961	Informação e outros
Facebook, Inc	90.059.000.000	1,3208	Informação
Spotify Technology S.A.	26.298.000.000	-0,1152	Informação e outros
Google Inc	20.508.410.000	-0,2456	Informação
Snap Inc	19.888.072.000	-0,2595	Informação
Twitter, Inc	11.454.353.000	-0,4495	Informação
Dropbox, Inc	7.762.500.000	-0,5326	Informação e outros
LinkedIn Corporation	3.786.424.000	-0,6222	Informação e outros
Yelp Inc	759.934.000	-0,6903	Informação
Amazon.com, Inc	238.908.000	-0,7021	Informação e outros

Fonte: Elaboração própria. Valores em USD.

No Quadro 7, é possível observar que duas empresas, a Alibaba e o Facebook, se destacam com um valor de informação acima da média das dez empresas, liderando o topo da tabela. A empresa pior classificada no conjunto das dez, neste Índice é a Amazon.

Atente-se que a IPO da Amazon foi já em 1997, o que poderá influenciar negativamente o valor da informação da empresa, aquando da IPO, comparativamente às outras empresas cujas IPO ocorreram já nos últimos anos, uma vez que a tecnologia e os hábitos que permitem a obtenção de informação parece ter vindo a evoluir.

O Quadro 8 apresenta a ordenação das dez empresas em termos do Rácio Q e de Rácio Q após padronização numa mesma escala. Como se pode observar, de acordo com este rácio, as dez empresas podem ter tendência para serem identificadas como estando sobreavaliadas. Estas empresas podem não ter tantos ativos físicos ou tangíveis como empresas de outras indústrias. Nestas dez empresas, pelo menos um dos seus principais ativos é informação e, como tal, a análise deste rácio pode ser complementada com o Índice anterior para avaliar as empresas. Como se pode observar, três das quatro empresas que lideram esta tabela, e que apresentam um Rácio Q acima da média, estão incluídas nas quatro posições que lideram a tabela do Índice do valor da informação em termos comparativos. Tal situação pode representar que as empresas não estão sobreavaliadas, como induz o Rácio Q, mas sim que estas empresas têm informação com valor, que não obedece às condições para ser considerado como ativo reportável.

Quadro 8 - Rácio Q das dez empresas

Empresa	Rácio Q	Rácio Q padronizado	Base geradora de receitas
Google Inc	9,2311	1,3192	Informação
Spotify Technology S.A.	9,2130	1,3091	Informação e outros
LinkedIn Corporation	9,1679	1,2839	Informação e outros
Facebook, Inc	7,4600	0,3291	Informação
Yelp Inc	6,4255	-0,2491	Informação
Dropbox, Inc	6,4000	-0,2634	Informação e outros
Snap Inc	5,8367	-0,5783	Informação
Twitter, Inc	5,1718	-0,9500	Informação
Amazon.com, Inc	4,9106	-1,0960	Informação e outros
Alibaba Group Holding Limited	4,8957	-1,1044	Informação e outros

Fonte: Elaboração própria

A análise conjunta das dez empresas pode ser criticável, uma vez que, tal como referido, a comparação engloba empresas cuja base geradora de receitas é apenas informação e empresas cuja base geradora de receitas inclui outros itens para além de informação. Além disso é considerado que a diferença entre a capitalização bolsista prevista na IPO e o valor dos ativos corresponde na sua totalidade ao valor da informação da

empresa, o que pode não se verificar. Apesar destas limitações, é possível considerar que esta análise é uma das melhores formas de mensurar o valor da informação em termos comparativos entre as empresas.

O Quadro 9 apresenta também o valor de informação, calculado segundo Laney (2017), desta vez relativo apenas às cinco empresas cuja base geradora de receitas é apenas informação, uma vez que, de acordo com Laney (2017) o valor da informação das empresas será mais facilmente observável em IPO de empresas cujo principal bem é informação.

Quadro 9 - Índice financeiro do valor da informação das 5 empresas cuja base geradora de receitas é apenas informação

Empresa	Valor da informação	Índice
Facebook, Inc	90.059.000.000	1,7424
Google Inc	20.508.410.000	-0,2273
Snap Inc	19.888.072.000	-0,2448
Twitter, Inc	11.454.353.000	-0,4837
Yelp Inc	759.934.000	-0,7865

Fonte: Elaboração própria. Valores em USD.

No Quadro 9, é possível observar que uma empresa, o Facebook, se destaca com um valor de informação acima da média das cinco empresas, liderando o topo da tabela. Como visto anteriormente, esta empresa destaca-se também na segunda posição, com um valor de informação acima da média das dez empresas (Quadro 7). A empresa pior classificada no conjunto das cinco, neste Índice é a Yelp.

O Quadro 10 apresenta a ordenação em termos do Rácio Q e de Rácio Q após padronização numa mesma escala, desta vez relativo somente às cinco empresas cuja base geradora de receitas é apenas informação. De acordo com este rácio, como se observa, as cinco empresas podem ter tendência para serem identificadas como estando sobreavaliadas. Nestas cinco empresas, o seu principal ativo é informação e, como tal, a análise deste rácio pode ser complementada com o Índice anterior. Como se observa, as duas empresas que lideram esta tabela, e que apresentam um Rácio Q acima da média, são também as empresas que lideram a tabela do Índice do valor da informação em termos comparativos relativo às cinco empresas (Quadro 9), e relativo às dez empresas (Quadro 7), encontrando-se neste nas primeiras quatro posições. Tal como observado anteriormente no conjunto das

dez empresas, esta situação pode representar que as empresas não estão sobreavaliadas, como induz o Rácio Q, mas sim que estas empresas têm informação com valor, que não obedece às condições para ser considerado como ativo reportável.

Quadro 10 - Rácio Q das 5 empresas cuja base geradora de receitas é apenas informação

Empresa	Rácio Q	Rácio Q padronizado
Google Inc	9,2311	1,5170
Facebook, Inc	7,4600	0,4004
Yelp Inc	6,4255	-0,2519
Snap Inc	5,8367	-0,6232
Twitter, Inc	5,1718	-1,0424

Fonte: Elaboração própria.

Capítulo 5. Conclusões

O principal objetivo deste estudo foi mensurar o valor económico da informação em empresas de comércio eletrónico. Este estudo propôs-se a responder a duas questões principais: Como mensurar o valor da informação em empresas de comércio eletrónico? Quais as características das empresas de comércio eletrónico que, em termos comparativos, apresentam maior valor da informação?

Para este efeito foi desenvolvido um estudo empírico onde foram recolhidas informações acerca de 100 empresas de comércio eletrónico. A partir da informação recolhida e da aplicação de técnicas de análise exploratória de dados (Análise de Componentes Principais Categórica - CATPCA, Análise de *Clusters*) foi proposto um índice que permite medir o valor da informação. Por fim, foram analisadas as principais características comuns às empresas que obtiveram o melhor posicionamento nesse índice. Em alternativa, foi também considerado um índice financeiro do valor da informação calculado com base em informação financeira de 10 empresas tecnológicas.

A aplicação do método Análise de Componentes Principais Categórica permitiu resumir a informação contida nas variáveis originais em 5 componentes ortogonais que explicam 76,4% da variância total das variáveis originais, indo ao encontro da literatura (Johnson, 1998, como referido em Maroco, 2007). Foi possível interpretar o significado de cada uma das cinco componentes: a componente 1 foi interpretada como definindo a “Comunidade digital”, a componente 2 define essencialmente a “Atratividade do *site*”, a componente 3 foi interpretada como “Informação sobre produtos comparáveis”, e por último, a componente 5 foi interpretada como “Informação sobre produtos similares”.

A Análise de *Clusters* permitiu o agrupamento dos *sites* com base nas componentes resultantes da CATPCA e as correspondentes classificações (“*scores*”) de cada *site*. Numa primeira fase, foi efetuada uma análise hierárquica (com o método da menor distância) como técnica exploratória para a decisão do número de *clusters* a reter. Numa segunda fase, após a decisão do número de *clusters* a reter, procedeu-se a uma análise não hierárquica para refinar e interpretar a solução obtida de quatro *clusters* através do procedimento *k-means*, uma vez que a classificação dos sujeitos em cada *cluster* é mais rigorosa nos métodos não hierárquicos (Maroco, 2007).

O *cluster* 1 pôde ser classificado como sendo respeitante a *sites*, que na sua maioria são “*Sites* pouco atrativos, com comunidade digital muito desenvolvida, com informação muito desenvolvida de produtos acessórios e com informação pouco desenvolvida de produtos comparáveis e similares”. O *cluster* 2 pôde ser classificado como sendo correspondente a “*Sites* atrativos, com comunidade digital pouco desenvolvida, com informação desenvolvida de produtos comparáveis e similares e com informação pouco desenvolvida de produtos acessórios”. O *cluster* 3 pôde ser classificado como sendo constituído por “*Sites* pouco atrativos, com comunidade digital pouco desenvolvida, com informação muito desenvolvida de produtos comparáveis, informação desenvolvida de produtos acessórios e com informação pouco desenvolvida de produtos similares”. Por último, o *cluster* 4 pôde ser classificado como sendo respeitante a “*Sites* muito atrativos, com comunidade digital desenvolvida, com informação muito desenvolvida de produtos similares, informação pouco desenvolvida de produtos comparáveis e acessórios”.

Os resultados obtidos no índice proposto do valor da informação em empresas de comércio eletrónico, permitiram identificar os 20 *sites* melhor classificados em termos desse índice (denominados de “Top 20”). Observa-se que metade dos *sites* pertencem ao *cluster* 1, quatro *sites* pertencem ao *cluster* 2, seis *sites* ao *cluster* 3 e nenhum *site* pertence ao *cluster* 4, o que poderá indicar que contribui para uma melhor classificação, em termos de índice do valor da informação, *sites* com comunidade digital muito desenvolvida, juntamente com “Informação sobre produtos acessórios” também muito desenvolvida. *Sites* com “Informação sobre produtos comparáveis” muito desenvolvida parecem conseguir uma melhor classificação comparativamente aos *sites* com atratividade desenvolvida. *Sites* com um desenvolvimento intermédio equilibrado no que respeita à sua atratividade, à Informação sobre produtos comparáveis e à Informação sobre produtos similares, parecem também obter uma classificação razoável em termos do índice do valor da informação.

Em termos gerais, as características que parecem mais contribuir para um melhor posicionamento em termos do índice do valor da informação proposto, observadas nos *sites* Top 20, é o desenvolvimento respeitante à Comunidade digital, em termos de Opinião de clientes, Avaliação de opinião, Classificação de Produtos, Posicionamento de tráfego Global, Número de *sites* fonte, em termos Digitais e em termos de Compras em conjunto.

Por último, uma análise ao índice financeiro do valor da informação com base em informação financeira de 10 empresas tecnológicas, parece indicar que empresas tecnológicas podem por vezes não estar sobreavaliadas, como pode induzir o Rácio Q.

Assim, estas empresas podem não ter tantos ativos físicos ou tangíveis como empresas de outras indústrias, sendo o seu principal ativo a informação, pelo que, para avaliar estas empresas, a análise deste rácio pode ser complementada com outra informação, nomeadamente com o índice financeiro do valor da informação.

Os resultados obtidos ajudam a perceber as teorias existentes na literatura acerca do valor da informação e como monetizá-la. Najjar & Kettinger (2013) defende que com o acesso em tempo real a informação, a empresa pode mais rapidamente a monetizar, sendo uma das três opções possíveis a seguir na monetização o desenvolvimento da infraestrutura técnica. As empresas de comércio eletrónico estudadas parecem ter estas infraestruturas desenvolvidas, o que lhes poderá proceder a essa monetização, seja através duma melhor gestão de *stocks* ou da criação de pacotes de informação para venda a terceiros, apontadas por Najjar & Kettinger (2013), seja através do uso da informação para aumentar receitas, apontado por Laney (2017). Segundo Laney (2017), uma das formas de monetizar a informação, é através do aumento da aquisição/retenção de clientes. Tal parece ser possível nas empresas estudadas através do desenvolvimento da informação nas 5 componentes identificadas, que poderão permitir rastrear os hábitos dos utilizadores nos *sites* e, desta forma, trabalhar no sentido dessa aquisição/retenção. A facilidade de análise computacional de atualmente revela-se uma ferramenta essencial para tal. A identificação de padrões de dados de clientes através de técnicas como o *data mining*, revela-se também muito importante para a tomada de decisão (Rajagopal, 2011). A fidelização dos clientes, avaliada pela variedade dos produtos comprados anteriormente, poderá também ser estimada mais facilmente. Quanto melhor a estimativa da fidelização do cliente e quanto melhor for a segmentação, maior a possibilidade de efetuar ações de marketing direcionadas que possam monetizar essa informação (Buckinx et al., 2007). Adicionalmente, a informação recolhida nos *sites* estudados poderá ir ao encontro do conceito de valor intrínseco da informação e valor da informação para o negócio, devido a características como a validade, a plenitude, a exclusividade e a relevância.

A informação detida pelas empresas estudadas poderá permitir também reduzir alguma da incerteza presente em decisões, identificada por Hubbard (2010). Poderá permitir também a estas empresas tomar decisões e controlar processos, como referido por Mowshowitz (1992). Para além deste papel, a informação poderá ainda ser produzida e vendida (Buff et al., 2015), como uma mercadoria (Mowshowitz 1992). Como defendido por Buff et al., (2015) o papel desempenhado pela informação está a mudar em algumas

organizações, deixando de ser utilizada apenas como um ativo secundário que suporta decisões e processos, passando a ser o ativo principal que alguns negócios produzem e vendem, monetizando essa informação. Nas 10 empresas tecnológicas correspondentes ao estudo do índice financeiro do valor da informação, parece identificar-se algumas nesta situação.

Reconhece-se, ainda, a existência de algumas limitações ao trabalho desenvolvido. Uma vez que o tema em análise é um tema ainda pouco explorado, não existiam bases de dados disponíveis. Pela natureza do objeto de estudo em análise, e pelas suas limitações inerentes, não foi possível analisar toda a informação detida por qualquer tipo de empresa ou utilizar uma técnica de amostragem aleatória ou probabilística. Adicionalmente, a interpretação do significado de cada uma das cinco componentes identificadas foi sujeita à subjetividade inerente. É importante referir também que a análise de *clusters* efetuada é uma técnica que segue critérios mais ou menos heurísticos e que não possui sólidos fundamentos teóricos (Maroco, 2007). Por último, refere-se que a avaliação interna que as empresas fazem quando perante uma IPO pode ser considerada como não completamente fidedigna para observar o valor da informação destas.

Para investigação futura, considera-se que seria interessante efetuar o estudo do valor da informação com uma amostra maior, se possível. Além disso considera-se que seria interessante estudar este tema utilizando outro tipo de informação privada que as empresas possam deter através de informação especificamente disponibilizada por estas para esse estudo. A utilização de outros métodos que não a CATPCA, poderia também revelar-se interessante.

Por último, considera-se que o presente estudo se revela pertinente, procurando contribuir para colmatar algumas insuficiências científicas sentidas no tema do valor da informação, nomeadamente ao nível empírico. Espera-se que este abra caminhos para novas investigações e sirva de incentivo para o desenvolvimento de trabalhos futuros.

6. Referências bibliográficas

- Akerlof, G. A. (1970). "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". *The Quarterly Journal of Economics*, 84(3), 488-500.
- Arrow, K. J. (1996). "The economics of information: An exposition". *Empirica*, 23(2), 119-128.
- Bettini, C., Wang, X. S., & Jajodia, S. (2006). "The role of quasi-identifiers in k-anonymity revisited", *Technical Report RT-11-06*, Milan: DICO, University of Milan.
- Buckinx, W., Verstraeten, G., & Van den Poel, D. (2007). "Predicting customer loyalty using the internal transactional database". *Expert Systems With Applications*, 32(1), 125.
- Buff, A., Wixom, B. H., & Paul, T. (2015). "Foundations for Data Monetization". *MIT Sloan WP No 5213-15 and CISR WP No 402*.
- Bütler, M., & Birchler, U. (2007). *Information Economics*. Oxon, United Kingdom: Routledge Advanced Texts in Economics and Finance.
- Copeland, T. E., & Friedman, D. (1992). "The Market Value of Information: Some Experimental Results". *The Journal of Business*, 65(2), 241-266.
- Einav, L., & Levin, J. (2014) "The Data Revolution and Economic Analysis". *Innovation Policy and the Economy*, 14, 1-24.
- Garifova, L. F. (2015). "Infonomics and the Value of Information in the Digital Economy". *Procedia Economics and Finance*, 23(Supplement C), 738-743.
- Grinblatt, M. & Titman, S. (2002). *Financial Markets and Corporate Strategy*, Second Edition, The McGraw–Hill Companies.
- Hayashi, T., & Ohsawa, Y. (2017). "Preliminary Case Study on Value Determination of Datasets and Cross-disciplinary Data Collaboration Using Data Jackets". *Procedia Computer Science*, 112, 2175.
- Hotteling, H. (1993). "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 24, pp. 417-441.
- Hubbard, D. W. (2010). *How to measure anything: Finding the value of "intangibles" in business*, Second Edition, New Jersey: John Wiley & Sons, Inc.
- IBM Corp. (2017). IBM SPSS Statistics for Macintosh, Version 25.0. Armonk, NY: IBM Corp.
- Krishnamurthy, B., & Wills, C. E. (2010). "On the Leakage of Personally Identifiable Information Via Online Social Networks". *Acm Sigcomm Computer Communication*

Review, 40(1), 112-117.

Laney, D. B. (2017). *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*: New York: Gartner, Inc.

Maroco, J. (2007). *Análise Estatística Com Utilização do SPSS*, 3ª edição, Lisboa: Edições Sílabo.

Mowshowitz, A. (1992). "On the market value of information commodities. 1. The nature of information and information commodities". *Journal of the American Society for Information Science*, 43(3), 225-232.

Najjar, M. S., & Kettinger, W. J. (2013). "Data Monetization: Lessons from a Retailer's Journey". *MIS Quarterly Executive*, 12(4), 213.

Narayanan, A., & Shmatikov, V. (2010). "Myths and Fallacies of Personally Identifiable Information". *Communications of the Acm*, 53(6), 24-26.

Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space". *Philosophical Magazine Series*, 2, 559-572.

Rajagopal, D. S. (2011). "Customer Data Clustering using Data Mining Technique". *International Journal of Database Management Systems (IJDMS) Vol.3, No.4, November 2011*.

Rothschild, M., & Stiglitz, J. (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information". *The Quarterly Journal of Economics*, 90(4), 629-649.

Spence, M. (1973). "Job Market Signaling". *The Quarterly Journal of Economics*, 87(3), 355-374.

Stigler, G. J. (1961). "The Economics of Information". *Journal of Political Economy*, 69(3), 213-225.

Stiglitz, J. E., & Weiss, A. (1981). "Credit Rationing in Markets with Imperfect Information". *The American Economic Review*, 71(3), 393-410.

Stiglitz, J. E. (2017) "The revolution of information economics: the past and the future", *NBER working paper series 23780*, National Bureau of Economic Research.

Tobin, J. & Brainard, W. C. (1977). "Asset Markets and the Cost of Capital", In B. Belassa and R. Nelson (eds.), *Economic Progress, Private Values and Public Policies, Essays In Honor of William Fellner*, Amsterdam, 235-262.

7. Anexos

Anexo 1 - Variáveis: Fonte, especificação e justificação da sua adoção

Variável	Fonte, especificação e justificação da adoção como teoricamente apropriada
Produtos similares	Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i> , de informação relativa a produtos similares, quando o utilizador/cliente está a visualizar determinado produto. A generalidade dos <i>sites</i> denomina como produtos similares aqueles produtos que são idênticos, no sentido em que possam satisfazer as mesmas necessidades do cliente que produto que está a ser visualizado, diferindo apenas em algumas características. Considera-se que o facto de um <i>site</i> ter informação sobre produtos similares pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se de alguma forma direta ou indiretamente a obtiver ou a extrair previamente, nomeadamente através de técnicas como o <i>data mining</i> . Considera-se também que quanto mais informação tiver acerca de produtos similares, maior interação poderá obter no <i>site</i> , conseguindo melhor definir o perfil do utilizador/cliente e, desta forma, mais potencial poderá ter de monetizar essa informação.
Produtos acessórios	Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i> , de informação relativa a produtos acessórios, quando o utilizador/cliente está a visualizar determinado produto. A generalidade dos <i>sites</i> denomina como produtos acessórios aqueles produtos que são complementares, no sentido em que possam ajudar a satisfazer as necessidades do cliente de forma mais alargada se adquiridos em conjunto com o produto que está a ser visualizado. Considera-se que o facto de um <i>site</i> ter informação sobre produtos acessórios pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se de alguma forma direta ou indiretamente a obtiver ou a extrair previamente, nomeadamente através de técnicas como o <i>data mining</i> . Considera-se também que quanto mais informação tiver acerca de produtos acessórios, maior interação poderá obter no <i>site</i> , conseguindo melhor definir o perfil do utilizador/cliente e, desta forma, mais potencial poderá ter de monetizar essa informação, nomeadamente através do possível aumento das vendas.
Comprados em conjunto	Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i> , de informação relativa a produtos que usualmente são comprados em conjunto com o produto que o utilizador/cliente está a visualizar. Considera-se que o facto de um <i>site</i> ter informação sobre produtos comprados em conjunto pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se de alguma forma direta ou indiretamente a obtiver ou a extrair previamente, nomeadamente através de técnicas como o <i>data mining</i> . Considera-se também que quanto mais informação tiver acerca de produtos comprados em conjunto, maior pode ser a probabilidade de efetuar vendas cruzadas, assim como maior pode ser a interação no <i>site</i> , conseguindo melhor definir o perfil do utilizador/cliente e, desta forma, mais potencial poderá ter de monetizar essa informação, nomeadamente através do possível aumento das vendas.
Comparação Produtos	Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à possibilidade ou não, de se efetuar a comparação do produto que está a ser visualizado com outros produtos do <i>site</i> , identificando as características díspares entre eles. Considera-se que o facto de um <i>site</i> ter a possibilidade de se comparar produtos pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se de alguma forma direta ou indiretamente a obtiver ou a extrair previamente, nomeadamente através de técnicas como o <i>data mining</i> . Considera-se também que quanto mais informação tiver acerca de produtos comparáveis, maior pode ser a interação no <i>site</i> , conseguindo melhor definir o perfil do utilizador/cliente e, desta forma, mais potencial poderá ter de monetizar essa informação, nomeadamente através do possível aumento das vendas.

Classificação Produtos	<p>Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i>, de informação relativa à classificação atribuída, numa escala predefinida, por compradores anteriores ao produto que o utilizador/cliente está a visualizar.</p> <p>Considera-se que o facto de um <i>site</i> ter informação sobre a classificação de produtos pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se a obtiver através de compradores anteriores. Considera-se também que quanto mais informação tiver acerca da classificação de produtos, maior pode ser a capacidade de adaptar os produtos em função deste <i>feedback</i> recebido dos clientes anteriores, assim como maior pode ser a interação no <i>site</i>, podendo conseguir ir ao encontro dos requisitos do cliente de forma mais eficiente e podendo conseguir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>
Opinião Clientes	<p>Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i>, de informação relativa à opinião atribuída por compradores anteriores ao produto que o utilizador/cliente está a visualizar. Considera-se que o facto de um <i>site</i> ter informação sobre a opinião atribuída pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se a obtiver através de compradores anteriores. Considera-se também que quanto mais informação tiver acerca da opinião de produtos, maior pode ser a capacidade de adaptar os produtos em função deste <i>feedback</i> recebido dos clientes anteriores, assim como maior pode ser a interação no <i>site</i>, podendo conseguir ir ao encontro dos requisitos do cliente de forma mais eficiente e podendo conseguir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>
Avaliação de opinião	<p>Proveniente do <i>site</i> de comércio eletrónico da empresa e respeitante à disponibilização ou não, pelo <i>site</i>, de informação relativa à avaliação da utilidade para compradores atuais e posteriores da opinião atribuída por compradores anteriores ao produto que o utilizador/cliente está a visualizar. Considera-se que o facto de um <i>site</i> ter informação sobre a avaliação da opinião atribuída pode indicar que a empresa detém mais informação, uma vez que a empresa só a conseguirá disponibilizar se a obtiver através de compradores anteriores. Considera-se também que quanto mais informação tiver acerca da avaliação da opinião de produtos maior pode ser a interação no <i>site</i>, podendo conseguir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>
Posicionamento tráfego Global	<p>Proveniente do Alexa e respeitante à estimativa da popularidade do <i>site</i>, atualizada diariamente. Uma vez que a classificação é calculada com base na média de visitantes diários ao respetivo <i>site</i> e no número de páginas visualizadas nesse mesmo <i>site</i> nos três meses anteriores, considera-se que quanto maior for esta popularidade, maior pode ser a interação no <i>site</i>, podendo permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>
Percentagem de permanência (com base na <i>Bounce Rate</i>)	<p>Proveniente do cálculo da diferença entre 100% e a "<i>Bounce Rate</i>", sendo a percentagem de permanência correspondente à percentagem de visitantes que se mantêm no <i>site</i>, visualizando para além de uma página, e sendo a "<i>Bounce Rate</i>" proveniente do Alexa e correspondente à percentagem de visitantes que abandona o <i>site</i> depois de visualizar apenas uma página. Considera-se que quanto maior for a percentagem de permanência, maior pode ser a interação no <i>site</i>, podendo permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>
Visualização páginas diária por visitante	<p>Proveniente do Alexa e respeitante à média de páginas visualizadas diariamente no <i>site</i>, por visitante. Considera-se que quanto maior for esta média, maior será a interação no <i>site</i>, podendo permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.</p>

Tempo diário no site	Proveniente do Alexa e respeitante à média do tempo passado diariamente no <i>site</i> , por visitante. Considera-se que quanto maior for esta média, maior será a interação no <i>site</i> , podendo permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.
Número sites fonte	Proveniente do Alexa e respeitante ao número de <i>sites</i> com hiperligação para o respetivo <i>site</i> de comércio eletrónico. Considera-se que quanto maior for este número, maior será o número de utilizadores canalizados para o <i>site</i> e consequentemente maior poderá ser a interação no <i>site</i> , podendo permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.
Nativa digital	Proveniente da pesquisa de informação geral da empresa e respeitante ao facto da empresa utilizar ou não o canal de comércio eletrónico desde o momento em que iniciou a sua atividade. Considera-se que o facto de utilizar este canal desde o início da atividade pode indicar uma maior predisposição para a extração de informação relativa à interação no <i>site</i> , nomeadamente através de técnicas como o <i>data mining</i> , podendo levar a que a empresa detenha mais informação que possa permitir definir o perfil do utilizador/cliente mais facilmente aumentando, desta forma, o potencial de monetizar essa informação, nomeadamente através do possível aumento das vendas.

Fonte: Elaboração própria

Anexo 2 - Empresas da amostra e dados recolhidos

Sujeito	Empresa/ Site	Endereço eletrónico	Produtos similares	Produtos acessórios	Comprados em conjunto	Comparação Produtos	Classificação Produtos	Opinião Clientes	Avaliação de opinião	Posionamento tráfego Global	Bounce Rate	Percentagem de permanência	Visualização páginas diária por visitante	Tempo diário no site	Tempo diário no site (mm:ss)	Número sites fonte	Data de acesso	Nativa digital
1	Fnac	https://www.fnac.pt	1	1	0	0	1	1	1	10088	44,10%	55,90%	3,08	0,142	03:25	2366	18-04-2018	0
2	Amazon	https://www.amazon.com	1	1	1	1	1	1	1	11	30,20%	69,80%	7,7	0,319	07:40	857812	18-04-2018	1
3	Decathlon	https://www.decathlon.pt	1	0	1	0	1	1	1	41582	30,40%	69,60%	6,43	0,219	05:15	309	18-04-2018	0
4	Sport Zone	https://www.sportzone.pt	0	0	0	0	1	1	0	81092	26,40%	73,60%	5,7	0,240	05:46	165	18-04-2018	0
5	Bazar desportivo	https://www.bazardesportivo.com	1	0	0	0	0	0	0	210381	36,90%	63,10%	6,8	0,144	03:28	29	18-04-2018	0
6	Spartoo	https://www.spartoo.pt	0	1	0	0	0	1	0	90049	24,90%	75,10%	6,3	0,366	08:47	144	18-04-2018	1
7	Worten	https://www.worten.pt	1	1	1	1	1	1	1	13761	37,40%	62,60%	3,34	0,178	04:16	346	19-04-2018	0
8	Rádio Popular	https://www.radiopopular.pt	0	1	0	0	0	0	0	60070	30,10%	69,90%	2,9	0,161	03:52	143	19-04-2018	0
9	Media Markt	https://www.mediamarkt.pt	1	1	0	0	1	1	0	74573	24,70%	75,30%	3,7	0,153	03:41	115	19-04-2018	0
10	PC Diga	https://www.pcdiga.com	1	0	0	1	0	0	0	17367	32,40%	67,60%	5,11	0,194	04:40	1206	19-04-2018	0
11	Zara	https://www.zara.com	1	1	0	0	0	0	0	1035	46,20%	53,80%	4,36	0,213	05:06	16809	17-05-2018	0
12	HM	http://www2.hm.com	0	1	1	0	0	0	0	589	21,30%	78,70%	8,13	0,346	08:18	28560	17-05-2018	0
13	Mango	https://shop.mango.com	0	0	0	0	0	0	0	3684	36,90%	73,10%	5,59	0,258	06:11	7482	17-05-2018	0
14	Bershka	https://www.bershka.com	1	1	0	0	0	0	0	6531	22,40%	77,60%	3,78	0,173	04:09	2669	17-05-2018	0
15	Pullandbear	https://www.pullandbear.com	0	0	0	0	0	0	0	6957	15,60%	84,40%	6,65	0,240	05:46	2145	17-05-2018	0
16	Stradivarius	https://www.stradivarius.com	0	0	0	0	0	0	0	11844	19,40%	80,60%	4,02	0,186	04:28	2127	17-05-2018	0
17	Reserved	http://www.reserved.com	0	0	0	0	0	0	0	9695	21,20%	78,80%	7,01	0,311	07:28	1184	17-05-2018	0
18	Continente	https://www.continente.pt	0	1	0	0	0	0	0	27235	31,30%	68,70%	3,75	0,291	06:59	812	17-05-2018	0
19	Jumbo	https://www.jumbo.pt	0	0	0	0	0	0	0	53448	30,10%	69,90%	5,1	0,218	05:14	248	17-05-2018	0
20	Pingo doce	https://www.pingodoce.pt	0	0	0	0	0	0	0	49533	35,80%	64,20%	2,4	0,124	02:58	341	17-05-2018	0
21	Apple	https://www.apple.com	0	0	0	1	0	0	0	73	48,30%	51,70%	3,13	0,146	03:30	439586	17-05-2018	0
22	Walmart	https://www.walmart.com	1	0	0	0	1	1	1	200	39,90%	60,10%	3,7	0,184	04:25	35297	17-05-2018	0
23	Target	https://www.target.com	1	0	0	0	1	1	1	419	57,50%	42,50%	2,73	0,143	03:26	37067	17-05-2018	0
24	Wayfair	https://www.wayfair.com	1	0	1	1	1	1	1	958	35,80%	64,20%	5,07	0,276	06:38	3966	17-05-2018	1
25	Homedepot	https://www.homedepot.com	1	1	1	1	1	1	1	403	35,80%	64,20%	3,77	0,188	04:31	24694	17-05-2018	0
26	Sears	http://www.sears.com	1	0	0	0	0	0	0	2474	43,30%	56,70%	3,62	0,169	04:04	9166	17-05-2018	0
27	Lowes	https://www.lowes.com	1	1	0	0	1	1	1	820	27,30%	72,70%	4	0,200	04:48	12733	17-05-2018	0
28	Menards	https://www.menards.com	0	1	0	1	0	0	0	5739	25,40%	74,60%	5,18	0,238	05:42	1709	17-05-2018	0
29	Acehardware	http://www.acehardware.com	1	1	0	0	1	1	1	13853	42,20%	57,80%	3,3	0,101	02:26	4463	17-05-2018	0
30	Truevalue	https://www.truevalue.com	0	0	0	0	1	1	1	48123	47,00%	53,00%	3,22	0,197	04:44	2547	17-05-2018	0
31	Ikea	https://www.ikea.com	1	1	0	0	1	1	1	281	23,40%	76,60%	8,26	0,356	08:33	52685	17-05-2018	0
32	Leroy Merlin	https://www.leroymerlin.pt	1	1	0	0	0	0	0	39031	26,00%	74,00%	5,74	0,252	06:03	219	17-05-2018	0
33	Aki	https://www.aki.pt	1	0	0	0	1	1	1	60189	34,00%	66,00%	4,1	0,185	04:26	188	17-05-2018	0
34	Conforama	https://www.conforama.pt	1	1	0	0	0	0	0	97242	24,00%	76,00%	5,8	0,307	07:22	116	17-05-2018	0
35	La Redoute	https://www.laredoute.pt	0	0	0	0	1	1	0	61383	28,20%	71,80%	7,5	0,308	07:23	256	17-05-2018	0
36	Kiabi	https://www.kiabi.pt	0	0	0	0	0	0	0	355901	33,10%	66,90%	3,2	0,142	03:24	28	17-05-2018	0
37	Zippy	https://pt.zippykidstore.com	0	0	0	0	0	0	0	187776	34,20%	65,80%	5,3	0,167	04:00	93	17-05-2018	0
38	Chicco	https://www.chicco.pt	1	1	0	0	0	0	0	411808	23,10%	76,90%	4,8	0,158	03:48	120	17-05-2018	0
39	Bebitus	https://www.bebitus.pt	0	0	0	0	1	1	1	206567	23,70%	76,30%	9,9	0,292	07:00	38	17-05-2018	1
40	Bybebe	https://www.bybebe.pt	0	1	0	0	0	1	0	774410	40,00%	60,00%	7,5	0,118	02:50	2	17-05-2018	0
41	Bestbuy	https://www.bestbuy.com	1	1	0	1	1	1	1	345	34,00%	66,00%	3,81	0,159	03:49	15452	17-05-2018	0
42	Newegg	https://www.newegg.com	0	0	0	1	1	1	1	646	44,70%	55,30%	2,81	0,174	04:10	10886	17-05-2018	1
43	Tesco	https://www.tesco.com	0	0	1	0	0	0	0	2795	38,30%	61,70%	4,62	0,201	04:50	7994	17-05-2018	0
44	Sainsburys	https://sainsburys.co.uk	0	0	0	0	1	1	1	6125	33,70%	66,30%	7,89	0,306	07:20	3151	17-05-2018	0
45	Argos	http://www.argos.co.uk	1	0	1	0	1	1	1	2446	36,10%	63,90%	4,46	0,190	04:34	4714	17-05-2018	0
46	Currys	https://www.currys.co.uk	1	1	0	0	1	1	1	5840	36,50%	63,50%	4,66	0,198	04:45	1698	17-05-2018	0
47	Staples	https://www.staples.pt	1	1	1	0	1	1	1	80232	31,20%	68,80%	3	0,143	03:26	206	17-05-2018	0

48	Gearbest	https://www.gearbest.com	1	1	1	0	1	1	1	354	48,70%	51,30%	4,45	0,181	04:21	8134	17-05-2018	1
49	Geekbuying	https://www.geekbuying.com	1	0	0	0	1	1	1	5413	53,10%	46,90%	3,34	0,139	03:20	1576	17-05-2018	1
50	Tomtop	https://www.tomtop.com	1	1	1	0	1	1	1	4669	58,00%	42,00%	5,16	0,150	03:36	1250	17-05-2018	1
51	Aliexpress	https://www.aliexpress.com	1	0	0	0	1	1	1	52	21,30%	78,70%	10,97	0,565	13:33	45737	24-05-2018	1
52	Alibaba	http://www.alibaba.com	1	0	0	1	0	1	0	172	48,90%	51,10%	5,84	0,215	05:10	52900	24-05-2018	1
53	Ebay	https://www.ebay.com	1	0	0	0	0	1	0	42	28,40%	71,60%	7,15	0,394	09:27	177274	24-05-2018	1
54	Farfetch	https://www.farfetch.com	1	1	0	0	0	0	0	1389	41,60%	58,40%	6,25	0,251	06:02	4197	25-05-2018	1
55	Yoox	https://www.yoox.com	1	1	0	0	0	0	0	2751	33,30%	66,70%	8,02	0,365	08:46	3127	25-05-2018	1
56	Ssense	https://www.ssense.com	0	1	0	0	0	0	0	3537	64,50%	35,50%	3,31	0,122	02:56	1139	25-05-2018	1
57	Barneys	https://www.barneys.com	1	0	0	0	0	0	0	10496	49,10%	50,90%	2,81	0,112	02:41	5147	25-05-2018	0
58	Neiman Marcus	https://www.neimanmarcus.com	1	1	0	0	0	0	0	5582	44,60%	55,40%	3,61	0,153	03:40	9027	25-05-2018	0
59	Sports Direct	https://www.sportsdirect.com	1	0	1	0	0	0	0	3871	24,70%	75,30%	8,42	0,431	10:21	2063	25-05-2018	0
60	ID Sports	https://www.idsports.co.uk	1	1	0	0	1	1	0	8836	33,90%	66,10%	5,02	0,177	04:15	848	25-05-2018	0
61	Footasylum	https://www.footasylum.com	1	1	0	0	1	1	0	23668	33,30%	66,70%	4,7	0,181	04:21	279	25-05-2018	0
62	Overstock	https://www.overstock.com	1	0	0	0	1	1	1	1694	35,20%	64,80%	4,65	0,216	05:11	9186	25-05-2018	1
63	Light in the box	https://www.lightinthebox.com	1	1	0	0	1	1	1	2046	47,40%	52,60%	5,65	0,219	05:16	3584	25-05-2018	1
64	DH Gate	https://www.dhgate.com	1	1	1	0	1	1	1	1241	49,60%	50,40%	7,03	0,199	04:47	9689	25-05-2018	1
65	Bang good	https://www.banggood.com	1	1	1	0	1	1	1	528	47,00%	53,00%	5,3	0,233	05:35	7558	25-05-2018	1
66	Prozis	https://www.prozis.com	0	1	0	0	1	1	0	16878	36,90%	63,10%	7,39	0,299	07:11	449	25-05-2018	1
67	Nutritienda	https://www.nutritienda.com	1	0	1	0	1	1	0	93722	53,50%	46,50%	3,4	0,173	04:09	48	25-05-2018	1
68	Asos	http://www.asos.com	1	1	0	0	1	1	1	365	25,10%	74,90%	10,9	0,404	09:42	23222	25-05-2018	1
69	Urbanoutfitters	https://www.urbanoutfitters.com	1	1	1	0	1	1	0	1352	28,10%	71,90%	5,15	0,251	06:01	13497	25-05-2018	0
70	Pacsun	https://www.pacsun.com	1	0	0	0	1	1	0	8919	27,70%	72,30%	4,72	0,197	04:43	1401	25-05-2018	0
71	Zumiez	https://www.zumiez.com	1	1	0	0	0	0	0	8215	41,10%	58,90%	4,59	0,153	03:41	1253	25-05-2018	0
72	Tactics	https://www.tactics.com	1	0	0	1	1	1	0	33797	45,70%	54,30%	4,5	0,147	03:31	486	25-05-2018	0
73	Topshop	http://eu.topshop.com	1	0	0	0	0	0	0	6250	29,70%	70,30%	4,95	0,183	04:24	11910	25-05-2018	0
74	Parfois	https://www.parfois.com	1	1	0	0	0	0	0	59348	17,00%	83,00%	6,9	0,224	05:23	807	25-05-2018	0
75	Oysho	https://www.oysho.com	0	1	0	0	0	0	0	25743	25,20%	74,80%	5,46	0,228	05:29	1395	25-05-2018	0
76	New Balance	http://www.newbalance.com	1	0	0	0	1	1	0	12519	34,60%	65,40%	4,16	0,155	03:43	3071	25-05-2018	0
77	Adidas	http://www.adidas.com	1	0	1	0	1	1	0	1869	38,00%	62,00%	4,02	0,176	04:13	11652	25-05-2018	0
78	Forever21	https://www.forever21.com	1	0	0	0	1	1	0	1763	22,90%	77,10%	6,8	0,354	08:30	11971	25-05-2018	0
79	Sfera	https://www.sfera.com	0	0	0	0	0	0	0	94628	17,60%	82,40%	4,7	0,176	04:14	157	25-05-2018	0
80	Tiffosi	https://www.tiffosi.com	0	0	0	0	0	0	0	257020	22,70%	77,30%	5	0,181	04:21	72	25-05-2018	0
81	Salsa Jeans	https://www.salsajeans.com	1	1	0	0	1	1	0	132071	32,40%	67,60%	5,1	0,297	07:08	79	25-05-2018	0
82	Kiabi	https://www.kiabi.com	0	1	0	0	0	0	0	17252	27,80%	72,20%	6,71	0,269	06:27	854	25-05-2018	0
83	Swatch	https://shop.swatch.com	1	0	0	0	0	0	0	33732	29,10%	70,90%	4,56	0,199	04:47	2247	25-05-2018	0
84	Deal Extreme	http://www.dx.com	1	1	1	1	1	1	1	3505	65,60%	34,40%	2,72	0,102	02:27	29461	25-05-2018	1
85	Cortefiel	https://cortefiel.com	1	1	0	0	0	0	0	88607	27,50%	72,50%	4	0,174	04:11	356	25-05-2018	0
86	Springfield	https://myspringfield.com	1	0	0	0	0	0	0	48548	24,20%	75,80%	4,5	0,175	04:12	281	25-05-2018	0
87	Costco	https://www.costco.com	1	0	0	1	1	1	1	864	27,00%	73,00%	4,44	0,182	04:22	7762	25-05-2018	0
88	Nike	https://www.nike.com	1	0	0	0	1	1	0	470	28,10%	71,90%	4,99	0,201	04:49	21835	25-05-2018	0
89	Reebok	https://www.reebok.com	1	0	0	0	1	1	1	16935	40,60%	59,40%	3,78	0,142	03:25	3619	25-05-2018	0
90	Puma	https://eu.puma.com	1	0	0	0	0	0	0	6661	26,70%	73,30%	4,96	0,190	04:34	5872	25-05-2018	0
91	Diadora	https://www.diadora.com	1	0	0	0	1	1	0	97953	25,50%	74,50%	5,3	0,192	04:37	621	25-05-2018	0
92	Asics	https://www.asics.com	1	1	0	0	1	1	1	14726	36,40%	63,60%	3,88	0,144	03:27	2191	25-05-2018	0
93	Under Armour	https://www.underarmour.com	1	0	0	0	1	1	1	5973	36,80%	63,20%	4,25	0,201	04:49	2153	25-05-2018	0
94	Skechers	https://www.skechers.com	0	0	0	0	1	1	1	24226	31,90%	68,10%	4,27	0,176	04:14	1308	25-05-2018	0
95	Converse	http://www.converse.com	1	0	0	0	0	0	0	14883	49,80%	50,20%	2,73	0,075	01:48	3743	25-05-2018	0
96	Le coq sportif	https://www.lecoqsportif.com	0	1	0	0	0	0	0	175770	24,90%	75,10%	4,2	0,132	03:10	452	25-05-2018	0
97	Fila	https://www.fila.com	1	0	0	0	0	0	0	97038	49,80%	50,20%	3	0,083	02:00	650	25-05-2018	0
98	Champion	https://www.champion.com	1	1	0	0	1	1	1	31474	30,70%	69,30%	3,78	0,140	03:21	174	25-05-2018	0
99	Hanes	https://www.hanes.com	1	0	0	0	1	1	1	40108	35,60%	64,40%	3,97	0,164	03:56	1008	25-05-2018	0
100	Vibram	http://us.vibram.com	1	1	0	0	1	1	1	92294	32,60%	67,40%	4,5	0,150	03:36	752	25-05-2018	0

Fonte: Elaboração própria

Anexo 3 - Descrição das variáveis

Fonte	Variável original	Tipo	Escala	Descrição	Variável Codificada	Tipo	Escala	Descrição
Site de comércio eletrónico da empresa	Produtos similares	Qualitativa	Nominal	Variável dicotómica quanto à disponibilização, pelo site, de informação relativa a produtos similares. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Similares	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Produtos acessórios	Qualitativa	Nominal	Variável dicotómica quanto à disponibilização, pelo site, de informação relativa a produtos acessórios. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Acessórios	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Comprados em conjunto	Qualitativa	Nominal	Variável dicotómica quanto à disponibilização, pelo site, de informação relativa a produtos comprados em conjunto. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Comprados	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Comparação Produtos	Qualitativa	Nominal	Variável dicotómica quanto à possibilidade de efetuar comparação entre produtos comercializados no site. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Comparação	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Classificação Produtos	Qualitativa	Nominal	Variável dicotómica quanto à possibilidade de efetuar e consultar avaliação de produtos comercializados no site. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Classificação	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Opinião Clientes	Qualitativa	Nominal	Variável dicotómica quanto à possibilidade de efetuar e consultar opinião acerca de produtos comercializados no site. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Opinião	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Site de comércio eletrónico da empresa	Avaliação de opinião	Qualitativa	Nominal	Variável dicotómica quanto à possibilidade de avaliar a utilidade de opinião acerca de produtos comercializados no site. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Avaliação	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como "missing values" todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".
Alexa	Posicionamento tráfego Global	Quantitativa	Razão	Estimativa da popularidade do site, atualizada diariamente pela Alexa. A classificação é calculada com base na média de visitantes diários ao respetivo site e o número de páginas visualizadas nesse mesmo site nos três meses anteriores. O site com o número mais elevado de visitantes diário e de páginas visualizadas é classificado com o número 1.	Tráfego	Qualitativa	Ordinal	Variável codificada tendo em vista a alteração de variável quantitativa, razão, para qualitativa, ordinal. Atribuída escala de 1 (um) a 10 (dez).

Alexa e cálculo próprio	Porcentagem de permanência (com base na <i>Bounce Rate</i>)	Quantitativa	Razão	A "Bounce Rate" corresponde à percentagem de visitantes que abandona o site depois de visualizar apenas uma página. A percentagem de permanência corresponde à percentagem de visitantes que se mantêm no site, visualizando para além de uma página, correspondendo à diferença entre 100% e a "Bounce Rate".	Permanência	Qualitativa	Ordinal	Variável codificada tendo em vista a alteração de variável quantitativa, razão, para qualitativa, ordinal. Atribuída escala de 1 (um) a 10 (dez).
Alexa	Visualização páginas diária por visitante	Quantitativa	Razão	Média de páginas visualizadas diariamente no site, por visitante.	Visualização	Qualitativa	Ordinal	Variável codificada tendo em vista a alteração de variável quantitativa, razão, para qualitativa, ordinal. Atribuída escala de 1 (um) a 10 (dez).
Alexa	Tempo diário no site	Quantitativa	Razão	Média do tempo passado diariamente no site, por visitante.	Tempo	Qualitativa	Ordinal	Variável codificada tendo em vista a alteração de variável quantitativa, razão, para qualitativa, ordinal. Atribuída escala de 1 (um) a 10 (dez).
Alexa	Número sites fonte	Quantitativa	Razão	Número de sites com hiperligação para o respetivo site de comércio eletrónico	Fontes	Qualitativa	Ordinal	Variável codificada tendo em vista a alteração de variável quantitativa, razão, para qualitativa, ordinal. Atribuída escala de 1 (um) a 10 (dez).
Informação geral da empresa	Nativa digital	Qualitativa	Nominal	Variável dicotómica que avalia se a empresa, quando iniciou atividade, apenas utilizava o canal de comércio eletrónico ou se este surgiu no seguimento do canal de comércio físico. Atribuído o valor 0 (zero) quando resposta é "não" e o valor 1 (um) quando a resposta é "sim".	Nativa	Qualitativa	Nominal	Variável codificada de modo a que todas as categorias assumam valores superiores ou igual a 1 (um), uma vez que o SPSS interpreta como " <i>missing values</i> " todos os valores das variáveis qualitativas inferiores a 1 (um). Atribuído o valor 1 (um) quando resposta é "não" e o valor 2 (dois) quando a resposta é "sim".

Fonte: Elaboração própria

Anexo 4 - Sumário do Modelo com 6 componentes

Dimensão (Componente)	Alfa de Cronbach	Variância Explicada	
		Total (<i>Valor Próprio</i>)	% da Variância
1	0,794	3,745	28,805
2	0,623	2,353	18,097
3	0,406	1,599	12,299
4	0,175	1,193	9,175
5	0,013	1,012	7,782
6	-0,144	0,883	6,792
Total	0,983	10,783	82,949

Fonte: Adaptação do SPSS

Anexo 5 - Sumário do Modelo com 4 componentes

Dimensão (Componente)	Alfa de Cronbach	Variância Explicada	
		Total (<i>Valor Próprio</i>)	% da Variância
1	0,795	3,759	28,918
2	0,637	2,428	18,674
3	0,396	1,577	12,132
4	0,156	1,169	8,989
Total	0,962	8,933	68,714

Fonte: Adaptação do SPSS

Anexo 6 - Classificação dos *sites* em cada componente

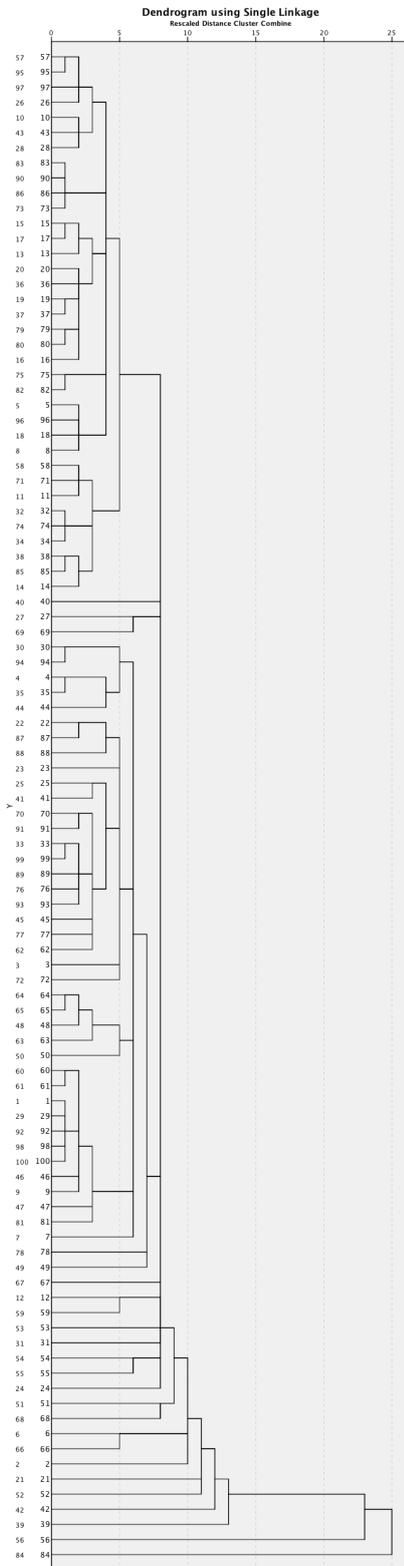
<i>Site</i>	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
1	0,481	-1,309	-0,720	0,055	0,954
2	2,383	1,634	0,869	0,955	-0,156
3	0,595	-0,323	-1,452	0,040	-0,395
4	-0,627	0,305	-1,550	-0,744	-1,394
5	-1,236	-0,361	-0,177	0,200	-0,106
6	-0,522	1,680	-1,256	1,698	-0,920
7	1,021	-1,042	0,030	0,687	0,373
8	-1,328	-0,692	0,454	1,116	0,108
9	-0,427	-0,805	-1,379	-0,062	1,233
10	-0,534	-0,014	1,510	-0,425	-0,513
11	-0,135	0,353	1,670	0,124	1,701
12	0,001	2,775	0,846	0,644	0,567
13	-0,885	1,141	0,768	-0,901	-0,718
14	-0,889	-0,070	0,541	0,029	1,790
15	-1,146	1,128	0,137	-0,488	-0,946
16	-1,236	0,265	0,521	-0,628	-0,798
17	-1,242	1,142	-0,126	-0,291	-1,047
18	-1,132	0,121	0,392	1,048	0,110
19	-1,326	0,160	0,098	0,282	-1,290
20	-1,377	-0,893	0,711	0,012	-1,061
21	-0,057	-0,077	3,354	-0,969	-1,351
22	1,146	-0,364	0,536	-1,969	0,240
23	1,143	-1,290	0,859	-1,361	-0,046
24	1,844	0,364	0,650	0,339	-1,486
25	1,682	-0,475	1,206	-0,110	0,751
26	-0,430	-0,247	1,624	-0,744	0,449
27	0,914	0,003	-0,216	-1,228	1,591
28	-0,671	0,987	1,317	0,288	-0,312
29	0,540	-1,302	-0,577	-0,051	0,999
30	0,057	-0,964	-0,871	-0,575	-1,630
31	1,278	2,074	-0,792	-1,101	1,345
32	-1,003	0,661	-0,239	0,447	1,524
33	0,151	-1,015	-1,360	-0,746	-0,377
34	-1,152	0,540	-0,503	0,615	1,442
35	-0,547	0,653	-1,622	-0,738	-1,452
36	-1,558	-0,831	0,229	0,313	-1,233
37	-1,505	-0,202	-0,094	0,446	-1,369
38	-1,245	-0,271	-0,170	0,477	1,543
39	0,042	0,910	-2,488	0,560	-2,600
40	-0,985	-0,360	-0,868	1,479	-0,573
41	1,423	-0,608	1,293	-0,914	0,704
42	1,262	-0,600	1,311	-0,266	-2,728
43	-0,450	0,371	1,405	0,234	-0,787
44	0,290	0,356	-1,178	-0,704	-1,684

45	1,036	-0,512	-0,286	-0,616	0,013
46	0,614	-0,612	-0,804	-0,142	1,035
47	0,504	-1,569	-1,038	0,894	1,033
48	1,803	-0,080	0,299	1,292	0,530
49	0,985	-1,301	-0,121	0,407	-1,180
50	1,456	-0,612	-0,392	2,279	-0,011
51	1,567	2,544	-0,782	-1,002	-0,710
52	1,064	0,745	2,353	-0,329	-1,170
53	0,707	2,387	0,586	-0,782	-0,257
54	-0,010	1,092	0,856	1,683	0,610
55	-0,014	2,201	0,182	1,776	0,497
56	-0,342	-1,078	1,654	3,357	-1,288
57	-0,779	-0,927	1,279	-0,487	0,347
58	-0,448	-0,334	1,406	0,339	1,617
59	-0,304	2,157	-0,019	-0,011	0,379
60	-0,004	-0,348	-0,990	0,290	0,979
61	-0,070	-0,648	-0,908	0,247	1,040
62	1,147	-0,084	-0,212	-0,448	-0,790
63	1,182	0,051	-0,590	0,974	0,183
64	1,634	0,344	-0,394	1,621	0,256
65	1,755	0,473	-0,156	1,461	0,392
66	0,114	0,750	-1,232	1,748	-1,223
67	0,487	-1,168	-0,526	1,764	-1,195
68	1,559	2,533	-1,052	0,156	0,437
69	0,799	0,817	-0,212	-0,199	1,655
70	-0,121	-0,228	-0,791	-1,480	0,182
71	-0,712	-0,394	0,804	0,738	1,404
72	0,166	-1,173	0,349	-0,721	-0,928
73	-0,699	0,311	1,010	-1,343	0,702
74	-1,016	0,844	-0,389	0,488	1,458
75	-1,134	0,934	0,082	0,507	0,302
76	-0,025	-0,745	-0,496	-0,972	-0,076
77	0,784	-0,278	0,455	-0,809	0,285
78	0,464	1,558	-0,575	-1,864	0,276
79	-1,516	-0,036	0,055	-0,295	-0,954
80	-1,645	-0,065	-0,218	-0,134	-1,033
81	-0,279	-0,223	-1,665	0,635	0,803
82	-1,222	1,061	-0,250	0,738	0,166
83	-0,986	0,119	0,432	-0,907	0,511
84	2,238	-1,724	2,082	2,273	-0,694
85	-1,125	-0,262	0,121	0,296	1,659
86	-1,177	-0,145	0,147	-0,686	0,409
87	1,053	-0,189	0,656	-2,171	-0,404
88	0,434	0,326	0,134	-2,098	0,482
89	0,471	-1,311	-0,486	-1,073	-0,206
90	-0,873	0,210	0,638	-1,048	0,578
91	-0,415	-0,223	-1,490	-1,078	-0,065
92	0,439	-1,209	-0,803	-0,105	1,035
93	0,442	-0,640	-0,914	-1,079	-0,221

94	0,085	-0,765	-0,895	-0,743	-1,565
95	-0,779	-0,927	1,279	-0,487	0,347
96	-1,500	-0,263	0,003	0,701	0,253
97	-1,067	-1,158	0,753	-0,124	0,178
98	0,254	-1,355	-1,158	0,166	0,916
99	0,170	-1,292	-1,120	-0,876	-0,291
100	0,243	-1,089	-1,342	0,205	0,857

Fonte: Adaptação do SPSS

Anexo 7 - Dendrograma da análise hierárquica de *clusters* aos *sites*



Fonte: SPSS

Anexo 8 - Identificação dos *clusters* agrupados em cada passo e respectivas distâncias

Passo	Clusters Agrupados		Distância	Diferença de distância	Nº de <i>clusters</i> totais
	Cluster 1	Cluster 2			
1	57	95	0,000	0,000	99
2	1	29	0,037	0,037	98
3	1	92	0,051	0,014	97
4	32	74	0,062	0,011	96
5	30	94	0,073	0,011	95
6	83	90	0,088	0,015	94
7	60	61	0,106	0,018	93
8	98	100	0,109	0,003	92
9	79	80	0,124	0,015	91
10	15	17	0,127	0,003	90
11	64	65	0,132	0,005	89
12	4	35	0,136	0,004	88
13	32	34	0,140	0,004	87
14	38	85	0,146	0,006	86
15	33	99	0,159	0,013	85
16	75	82	0,206	0,047	84
17	19	37	0,233	0,027	83
18	83	86	0,246	0,013	82
19	1	98	0,259	0,013	81
20	73	83	0,281	0,022	80
21	14	38	0,357	0,076	79
22	1	46	0,388	0,031	78
23	20	36	0,389	0,001	77
24	76	93	0,436	0,047	76
25	19	79	0,491	0,055	75
26	5	96	0,491	0,000	74
27	9	60	0,507	0,016	73
28	22	87	0,509	0,002	72
29	1	9	0,515	0,006	71
30	16	19	0,522	0,007	70
31	33	89	0,538	0,016	69
32	16	20	0,539	0,001	68
33	33	76	0,559	0,020	67
34	48	64	0,563	0,004	66
35	57	97	0,574	0,011	65
36	5	18	0,575	0,001	64
37	5	8	0,611	0,036	63
38	58	71	0,640	0,029	62
39	28	43	0,665	0,025	61
40	1	47	0,665	0,000	60
41	10	28	0,677	0,012	59

42	33	70	0,689	0,012	58
43	13	15	0,689	0,000	57
44	11	58	0,693	0,004	56
45	1	81	0,697	0,004	55
46	25	41	0,740	0,043	54
47	48	63	0,752	0,012	53
48	45	77	0,779	0,027	52
49	26	57	0,781	0,002	51
50	33	91	0,798	0,017	50
51	14	32	0,807	0,009	49
52	11	14	0,857	0,050	48
53	45	62	0,874	0,017	47
54	13	16	0,943	0,069	46
55	33	45	1,033	0,090	45
56	4	44	1,042	0,009	44
57	5	75	1,086	0,044	43
58	10	26	1,105	0,019	42
59	10	73	1,183	0,078	41
60	5	13	1,202	0,019	40
61	5	10	1,207	0,005	39
62	22	88	1,221	0,014	38
63	3	33	1,304	0,083	37
64	6	66	1,367	0,063	36
65	4	30	1,394	0,027	35
66	3	25	1,407	0,013	34
67	22	23	1,413	0,006	33
68	48	50	1,451	0,038	32
69	3	72	1,453	0,002	31
70	3	22	1,492	0,039	30
71	5	11	1,580	0,088	29
72	1	7	1,604	0,024	28
73	12	59	1,688	0,084	27
74	3	4	1,693	0,005	26
75	54	55	1,704	0,011	25
76	1	48	1,725	0,021	24
77	1	3	1,725	0,000	23
78	27	69	1,738	0,013	22
79	12	54	2,057	0,319	21
80	1	78	2,119	0,062	20
81	1	49	2,244	0,125	19
82	1	67	2,272	0,028	18
83	1	27	2,301	0,029	17
84	5	40	2,322	0,021	16
85	1	5	2,375	0,053	15
86	12	53	2,442	0,067	14
87	1	12	2,470	0,028	13
88	1	24	2,535	0,065	12
89	1	31	2,702	0,167	11
90	51	68	2,730	0,028	10

91	1	51	2,761	0,031	9
92	1	6	3,068	0,307	8
93	1	2	3,348	0,280	7
94	21	52	3,376	0,028	6
95	1	42	3,616	0,240	5
96	1	21	3,796	0,180	4
97	1	39	4,166	0,370	3
98	1	56	7,995	3,829	2
99	1	84	8,433	0,438	1

Fonte: Elaboração própria

Anexo 9 - Identificação do *cluster* a que pertence cada *site* e distância ao respectivo centróide

<i>Site</i>	<i>Cluster</i>	Distância
1	2	1,306
2	1	2,210
3	2	1,507
4	4	1,592
5	4	1,039
6	4	2,291
7	1	1,433
8	4	1,862
9	2	1,227
10	3	0,955
11	3	2,015
12	3	3,083
13	3	1,684
14	2	1,878
15	4	1,335
16	4	1,315
17	4	1,188
18	4	1,550
19	4	0,756
20	4	1,630
21	3	2,659
22	3	1,802
23	3	1,820
24	1	1,901
25	3	1,893
26	3	0,954
27	2	1,611
28	3	1,596
29	2	1,310
30	4	1,889
31	2	2,704
32	2	1,738
33	2	1,660
34	2	1,781
35	4	1,735
36	4	1,383
37	4	0,910
38	2	1,774
39	4	2,935
40	4	1,462
41	3	1,651
42	3	3,027

43	3	1,385
44	4	1,874
45	2	1,363
46	2	0,807
47	2	1,944
48	1	0,999
49	1	1,952
50	1	1,241
51	2	3,446
52	3	2,106
53	3	2,472
54	1	1,928
55	1	2,660
56	1	3,103
57	3	1,388
58	3	2,051
59	2	2,420
60	2	0,714
61	2	0,831
62	3	1,820
63	1	1,122
64	1	1,059
65	1	1,104
66	4	2,105
67	1	1,841
68	2	3,090
69	2	1,560
70	2	1,370
71	2	2,052
72	3	1,668
73	3	1,312
74	2	1,794
75	4	1,555
76	2	1,263
77	3	1,011
78	2	2,397
79	4	0,942
80	4	0,868
81	2	1,380
82	4	1,522
83	3	1,399
84	1	2,787
85	2	1,785
86	2	1,578
87	3	1,869
88	3	1,829
89	2	1,773
90	3	1,300
91	2	1,498

92	2	1,199
93	2	1,434
94	4	1,857
95	3	1,388
96	4	1,503
97	3	1,796
98	2	1,411
99	2	1,735
100	2	1,279

Fonte: Elaboração própria

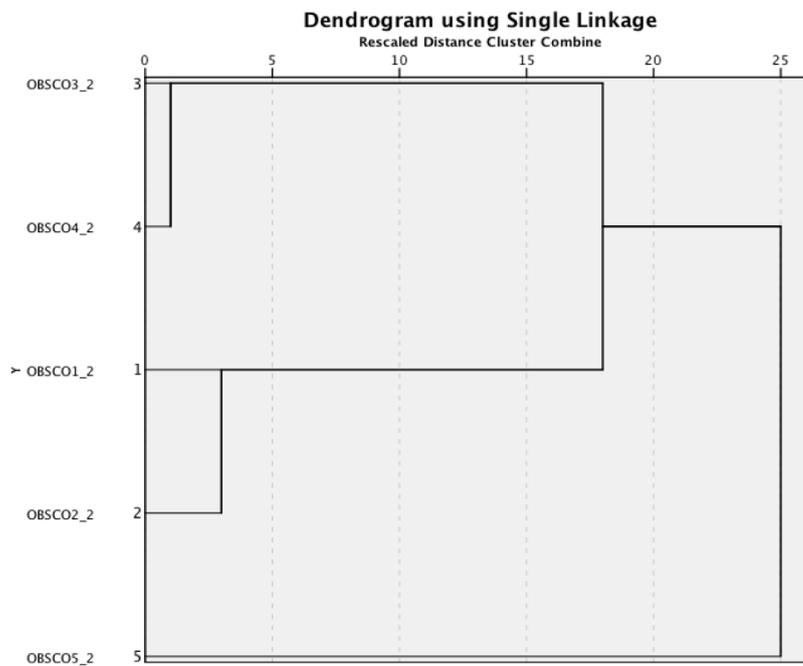
Anexo 10 - *Sites constituintes de cada cluster*

<i>Cluster1</i>	Amazon	Light in the box
	Bang good	Nutritienda
	Deal Extreme	Ssense
	DH Gate	Tomtop
	Farfetch	Wayfair
	Gearbest	Worten
	Geekbuying	Yoox
<i>Cluster2</i>	Acehardware	Ikea
	Aki	JD Sports
	Aliexpress	Leroy Merlin
	Argos	Lowe's
	Asics	Media Markt
	Asos	New Balance
	Bershka	Pacsun
	Champion	Parfois
	Chicco	Reebok
	Conforama	Salsa Jeans
	Cortefiel	Sports Direct
	Currys	Springfield
	Decathlon	Staples
	Diadora	Under Armour
	Fnac	Urbanoutfitters
	Footasylum	Vibram
	Forever21	Zumiez
Hanes		
<i>Cluster3</i>	Adidas	Newegg
	Alibaba	Nike
	Apple	Overstock
	Barneys	PC Diga
	Bestbuy	Puma
	Converse	Sears
	Costco	Swatch
	Ebay	Tactics
	Fila	Target
	HM	Tesco
	Homedepot	Topshop
	Mango	Walmart
	Menards	Zara
	Neiman Marcus	
<i>Cluster4</i>	Bazar desportivo	Pullandbear
	Bebitus	Rádio Popular
	Bybebe	Reserved

Continente	Sainsburys
Jumbo	Sfera
Kiabi	Skechers
Kiabi	Spartoo
La Redoute	Sport Zone
Le coq sportif	Stradivarius
Oysho	Tiffosi
Pingo doce	Truevalue
Prozis	Zippy

Fonte: Elaboração própria

Anexo 11 - Dendrograma da análise hierárquica de *clusters* às componentes
(variáveis resultantes da CATPCA)



Fonte: SPSS

- OBSC01_1: Componente 1, resultante da CATPCA
- OBSC02_1: Componente 2, resultante da CATPCA
- OBSC03_1: Componente 3, resultante da CATPCA
- OBSC04_1: Componente 4, resultante da CATPCA
- OBSC05_1: Componente 5, resultante da CATPCA

Anexo 12 - Índice correspondente aos 100 sites

Site	Índice	Cluster
Amazon	1,172	1
Asos	0,847	2
Bang good	0,743	1
Deal Extreme	0,727	1
HM	0,721	3
Gearbest	0,703	1
Aliexpress	0,686	2
Ikea	0,666	2
DH Gate	0,660	1
Yoox	0,631	1
Ebay	0,626	3
Alibaba	0,601	3
Wayfair	0,592	1
Homedepot	0,589	3
Farfetch	0,505	1
Tomtop	0,471	1
Urbanoutfitters	0,470	2
Bestbuy	0,421	3
Light in the box	0,386	1
Zara	0,369	3
Sports Direct	0,340	2
Lowe's	0,251	2
Forever21	0,206	2
Worten	0,197	1
Overstock	0,187	3
Adidas	0,176	3
Apple	0,168	3
Newegg	0,168	3
Walmart	0,163	3
Menards	0,147	3
Neiman Marcus	0,132	3
Argos	0,115	2
Costco	0,114	3
Ssense	0,104	1
Spartoo	0,099	4
Prozis	0,093	4
Tesco	0,064	3
Target	0,062	3
Nike	0,046	3
Currys	0,037	2
Zumiez	-0,006	2
Sears	-0,013	3
Parfois	-0,023	2
Geekbuying	-0,027	1

Leroy Merlin	-0,034	2
PC Diga	-0,059	3
Bershka	-0,063	2
Nutritienda	-0,070	1
Oysho	-0,074	4
JD Sports	-0,079	2
Acehardware	-0,081	2
Decathlon	-0,087	2
Mango	-0,093	3
Topshop	-0,094	3
Kiabi	-0,104	4
Staples	-0,105	2
Fnac	-0,109	2
Asics	-0,121	2
Conforama	-0,121	2
Footasylum	-0,143	2
Continente	-0,153	4
Sainsburys	-0,186	4
Puma	-0,190	3
Salsa Jeans	-0,196	2
Cortefiel	-0,202	2
Vibram	-0,204	2
Under Armour	-0,216	2
Pullandbear	-0,225	4
Champion	-0,228	2
Swatch	-0,256	3
Bebitus	-0,264	4
Barneys	-0,265	3
Converse	-0,265	3
Chicco	-0,265	2
Tactics	-0,269	3
Reserved	-0,270	4
Reebok	-0,281	2
Pacsun	-0,293	2
New Balance	-0,300	2
Media Markt	-0,345	2
Rádio Popular	-0,348	4
Bybebe	-0,362	4
Stradivarius	-0,367	4
Springfield	-0,381	2
Le coq sportif	-0,398	4
Aki	-0,404	2
La Redoute	-0,410	4
Skechers	-0,415	4
Jumbo	-0,417	4
Hanes	-0,427	2
Fila	-0,433	3
Bazar desportivo	-0,435	4
Diadora	-0,442	2

Truevalue	-0,447	4
Sport Zone	-0,485	4
Sfera	-0,541	4
Zippy	-0,550	4
Pingo doce	-0,563	4
Tiffosi	-0,607	4
Kiabi	-0,646	4

Fonte: Elaboração própria

Anexo 13 - Posicionamento dos *sites* em termos de média de páginas visualizadas diariamente por visitante

Posição	Sujeito	Site	Visualização diária de páginas
1	51	Aliexpress	10,97
2	68	Asos	10,90
3	39	Bebitus	9,90
4	59	Sports Direct	8,42
5	31	Ikea	8,26
6	12	HM	8,13
7	55	Yoox	8,02
8	44	Sainsburys	7,89
9	2	Amazon	7,70
10	35	La Redoute	7,50
11	40	Bybebe	7,50
12	66	Prozis	7,39
13	53	Ebay	7,15
14	64	DH Gate	7,03
15	17	Reserved	7,01
16	74	Parfois	6,90
17	5	Bazar desportivo	6,80
18	78	Forever21	6,80
19	82	Kiabi	6,71
20	15	Pullandbear	6,65
21	3	Decathlon	6,43
22	6	Spartoo	6,30
23	54	Farfetch	6,25
24	52	Alibaba	5,84
25	34	Conforama	5,80
26	32	Leroy Merlin	5,74
27	4	Sport Zone	5,70
28	63	Light in the box	5,65
29	13	Mango	5,59
30	75	Oysho	5,46
31	37	Zippy	5,30
32	65	Bang good	5,30
33	91	Diadora	5,30
34	28	Menards	5,18
35	50	Tomtop	5,16
36	69	Urbanoutfitters	5,15
37	10	PC Diga	5,11
38	19	Jumbo	5,10
39	81	Salsa Jeans	5,10
40	24	Wayfair	5,07
41	60	JD Sports	5,02
42	80	Tiffosi	5,00

43	88	Nike	4,99
44	90	Puma	4,96
45	73	Topshop	4,95
46	38	Chicco	4,80
47	70	Pacsun	4,72
48	61	Footasylum	4,70
49	79	Sfera	4,70
50	46	Currys	4,66
51	62	Overstock	4,65
52	43	Tesco	4,62
53	71	Zumiez	4,59
54	83	Swatch	4,56
55	72	Tactics	4,50
56	86	Springfield	4,50
57	100	Vibram	4,50
58	45	Argos	4,46
59	48	Gearbest	4,45
60	87	Costco	4,44
61	11	Zara	4,36
62	94	Skechers	4,27
63	93	Under Armour	4,25
64	96	Le coq sportif	4,20
65	76	New Balance	4,16
66	33	Aki	4,10
67	16	Stradivarius	4,02
68	77	Adidas	4,02
69	27	Lowes	4,00
70	85	Cortefiel	4,00
71	99	Hanes	3,97
72	92	Asics	3,88
73	41	Bestbuy	3,81
74	14	Bershka	3,78
75	89	Reebok	3,78
76	98	Champion	3,78
77	25	Homedepot	3,77
78	18	Continente	3,75
79	9	Media Markt	3,70
80	22	Walmart	3,70
81	26	Sears	3,62
82	58	Neiman Marcus	3,61
83	67	Nutritienda	3,40
84	7	Worten	3,34
85	49	Geekbuying	3,34
86	56	Ssense	3,31
87	29	Acehardware	3,30
88	30	Truevalue	3,22
89	36	Kiabi	3,20
90	21	Apple	3,13
91	1	Fnac	3,08

92	47	Staples	3,00
93	97	Fila	3,00
94	8	Rádio Popular	2,90
95	42	Newegg	2,81
96	57	Barneys	2,81
97	23	Target	2,73
98	95	Converse	2,73
99	84	Deal Extreme	2,72
100	20	Pingo doce	2,40

Fonte: Elaboração própria

**Anexo 14 - Posicionamento dos *sites* em termos de média do tempo passado
diariamente por visitante**

Posição	Sujeito	Site	Tempo diário no site (mm:ss)
1	51	Aliexpress	13:33
2	59	Sports Direct	10:21
3	68	Asos	09:42
4	53	Ebay	09:27
5	6	Spartoo	08:47
6	55	Yoox	08:46
7	31	Ikea	08:33
8	78	Forever21	08:30
9	12	HM	08:18
10	2	Amazon	07:40
11	17	Reserved	07:28
12	35	La Redoute	07:23
13	34	Conforama	07:22
14	44	Sainsburys	07:20
15	66	Prozis	07:11
16	81	Salsa Jeans	07:08
17	39	Bebitus	07:00
18	18	Continente	06:59
19	24	Wayfair	06:38
20	82	Kiabi	06:27
21	13	Mango	06:11
22	32	Leroy Merlin	06:03
23	54	Farfetch	06:02
24	69	Urbanoutfitters	06:01
25	15	Pullandbear	05:46
26	4	Sport Zone	05:46
27	28	Menards	05:42
28	65	Bang good	05:35
29	75	Oysho	05:29
30	74	Parfois	05:23
31	63	Light in the box	05:16
32	3	Decathlon	05:15
33	19	Jumbo	05:14
34	62	Overstock	05:11
35	52	Alibaba	05:10
36	11	Zara	05:06
37	43	Tesco	04:50
38	88	Nike	04:49
39	93	Under Armour	04:49
40	27	Lowes	04:48
41	64	DH Gate	04:47
42	83	Swatch	04:47

43	46	Currys	04:45
44	30	Truevalue	04:44
45	70	Pacsun	04:43
46	10	PC Diga	04:40
47	91	Diadora	04:37
48	90	Puma	04:34
49	45	Argos	04:34
50	25	Homedepot	04:31
51	16	Stradivarius	04:28
52	33	Aki	04:26
53	22	Walmart	04:25
54	73	Topshop	04:24
55	87	Costco	04:22
56	80	Tiffosi	04:21
57	61	Footasylum	04:21
58	48	Gearbest	04:21
59	7	Worten	04:16
60	60	JD Sports	04:15
61	79	Sfera	04:14
62	94	Skechers	04:14
63	77	Adidas	04:13
64	86	Springfield	04:12
65	85	Cortefiel	04:11
66	42	Newegg	04:10
67	14	Bershka	04:09
68	67	Nutritienda	04:09
69	26	Sears	04:04
70	37	Zippy	04:00
71	99	Hanes	03:56
72	8	Rádio Popular	03:52
73	41	Bestbuy	03:49
74	38	Chicco	03:48
75	76	New Balance	03:43
76	71	Zumiez	03:41
77	9	Media Markt	03:41
78	58	Neiman Marcus	03:40
79	50	Tomtop	03:36
80	100	Vibram	03:36
81	72	Tactics	03:31
82	21	Apple	03:30
83	5	Bazar desportivo	03:28
84	92	Asics	03:27
85	47	Staples	03:26
86	23	Target	03:26
87	89	Reebok	03:25
88	1	Fnac	03:25
89	36	Kiabi	03:24
90	98	Champion	03:21
91	49	Geekbuying	03:20

92	96	Le coq sportif	03:10
93	20	Pingo doce	02:58
94	56	Ssense	02:56
95	40	Bybebe	02:50
96	57	Barneys	02:41
97	84	Deal Extreme	02:27
98	29	Acchardware	02:26
99	97	Fila	02:00
100	95	Converse	01:48

Fonte: Elaboração própria

Anexo 15 - Posicionamento dos sites em termos de percentagem de visitantes que se mantêm no site, visualizando para além de uma página

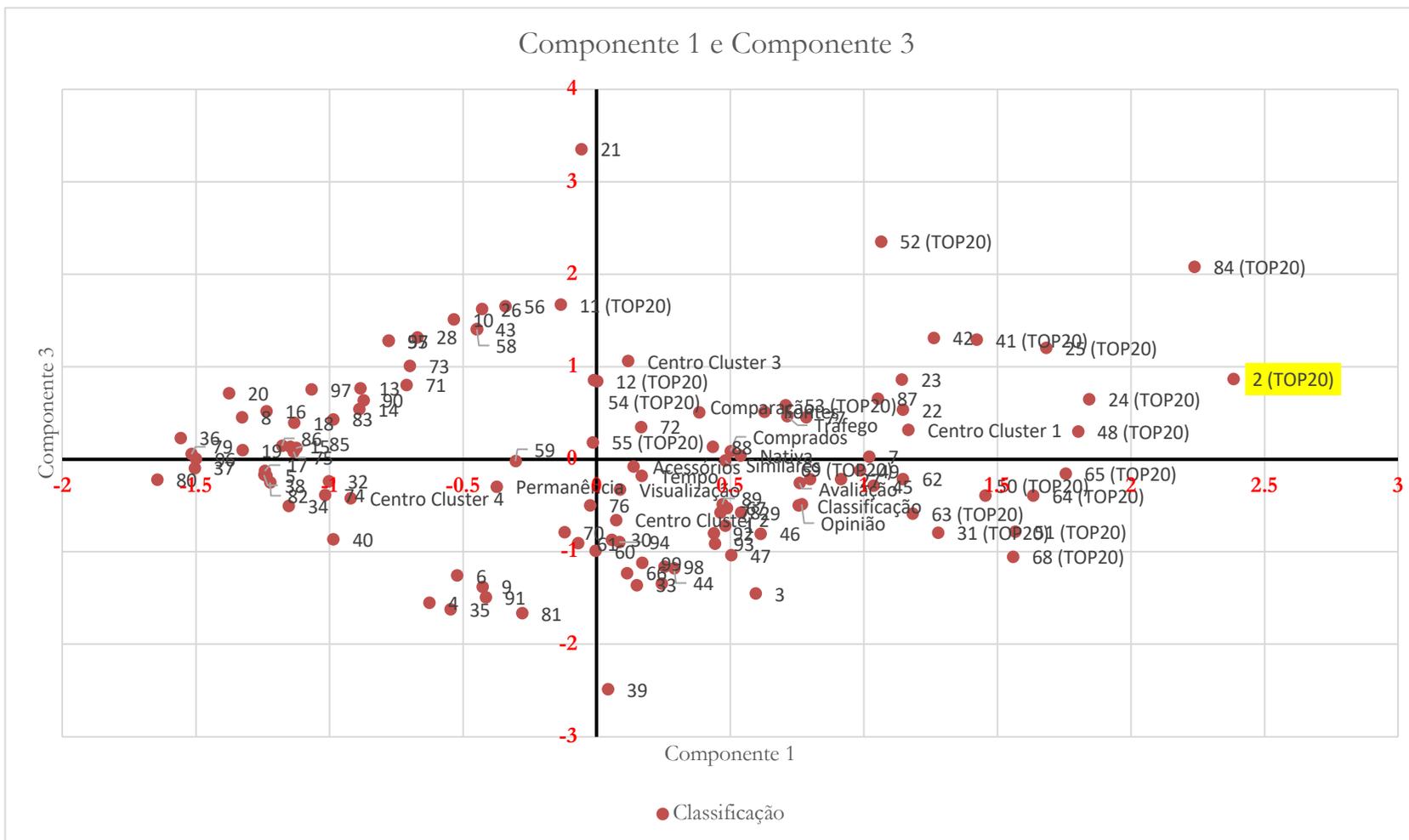
Posição	Sujeito	Site	Percentagem de permanência
1	15	Pullandbear	84,40%
2	74	Parfois	83,00%
3	79	Sfera	82,40%
4	16	Stradivarius	80,60%
5	17	Reserved	78,80%
6	51	Aliexpress	78,70%
7	12	HM	78,70%
8	14	Bershka	77,60%
9	80	Tiffosi	77,30%
10	78	Forever21	77,10%
11	38	Chicco	76,90%
12	31	Ikea	76,60%
13	39	Bebitus	76,30%
14	34	Conforama	76,00%
15	86	Springfield	75,80%
16	59	Sports Direct	75,30%
17	9	Media Markt	75,30%
18	6	Spartoo	75,10%
19	96	Le coq sportif	75,10%
20	68	Asos	74,90%
21	75	Oysho	74,80%
22	28	Menards	74,60%
23	91	Diadora	74,50%
24	32	Leroy Merlin	74,00%
25	4	Sport Zone	73,60%
26	90	Puma	73,30%
27	13	Mango	73,10%
28	87	Costco	73,00%
29	27	Lowes	72,70%
30	85	Cortefiel	72,50%
31	70	Pacsun	72,30%
32	82	Kiabi	72,20%
33	69	Urbanoutfitters	71,90%
34	88	Nike	71,90%
35	35	La Redoute	71,80%
36	53	Ebay	71,60%
37	83	Swatch	70,90%
38	73	Topshop	70,30%
39	19	Jumbo	69,90%
40	8	Rádio Popular	69,90%
41	2	Amazon	69,80%
42	3	Decathlon	69,60%

43	98	Champion	69,30%
44	47	Staples	68,80%
45	18	Continente	68,70%
46	94	Skechers	68,10%
47	10	PC Diga	67,60%
48	81	Salsa Jeans	67,60%
49	100	Vibram	67,40%
50	36	Kiabi	66,90%
51	55	Yoox	66,70%
52	61	Footasylum	66,70%
53	44	Sainsburys	66,30%
54	60	JD Sports	66,10%
55	33	Aki	66,00%
56	41	Bestbuy	66,00%
57	37	Zippy	65,80%
58	76	New Balance	65,40%
59	62	Overstock	64,80%
60	99	Hanes	64,40%
61	24	Wayfair	64,20%
62	25	Homedepot	64,20%
63	20	Pingo doce	64,20%
64	45	Argos	63,90%
65	92	Asics	63,60%
66	46	Currys	63,50%
67	93	Under Armour	63,20%
68	66	Prozis	63,10%
69	5	Bazar desportivo	63,10%
70	7	Worten	62,60%
71	77	Adidas	62,00%
72	43	Tesco	61,70%
73	22	Walmart	60,10%
74	40	Bybebe	60,00%
75	89	Reebok	59,40%
76	71	Zumiez	58,90%
77	54	Farfetch	58,40%
78	29	Acchardware	57,80%
79	26	Sears	56,70%
80	1	Fnac	55,90%
81	58	Neiman Marcus	55,40%
82	42	Newegg	55,30%
83	72	Tactics	54,30%
84	11	Zara	53,80%
85	65	Bang good	53,00%
86	30	Truevalue	53,00%
87	63	Light in the box	52,60%
88	21	Apple	51,70%
89	48	Gearbest	51,30%
90	52	Alibaba	51,10%
91	57	Barneys	50,90%

92	64	DH Gate	50,40%
93	97	Fila	50,20%
94	95	Converse	50,20%
95	49	Geekbuying	46,90%
96	67	Nutritienda	46,50%
97	23	Target	42,50%
98	50	Tomtop	42,00%
99	56	Ssense	35,50%
100	84	Deal Extreme	34,40%

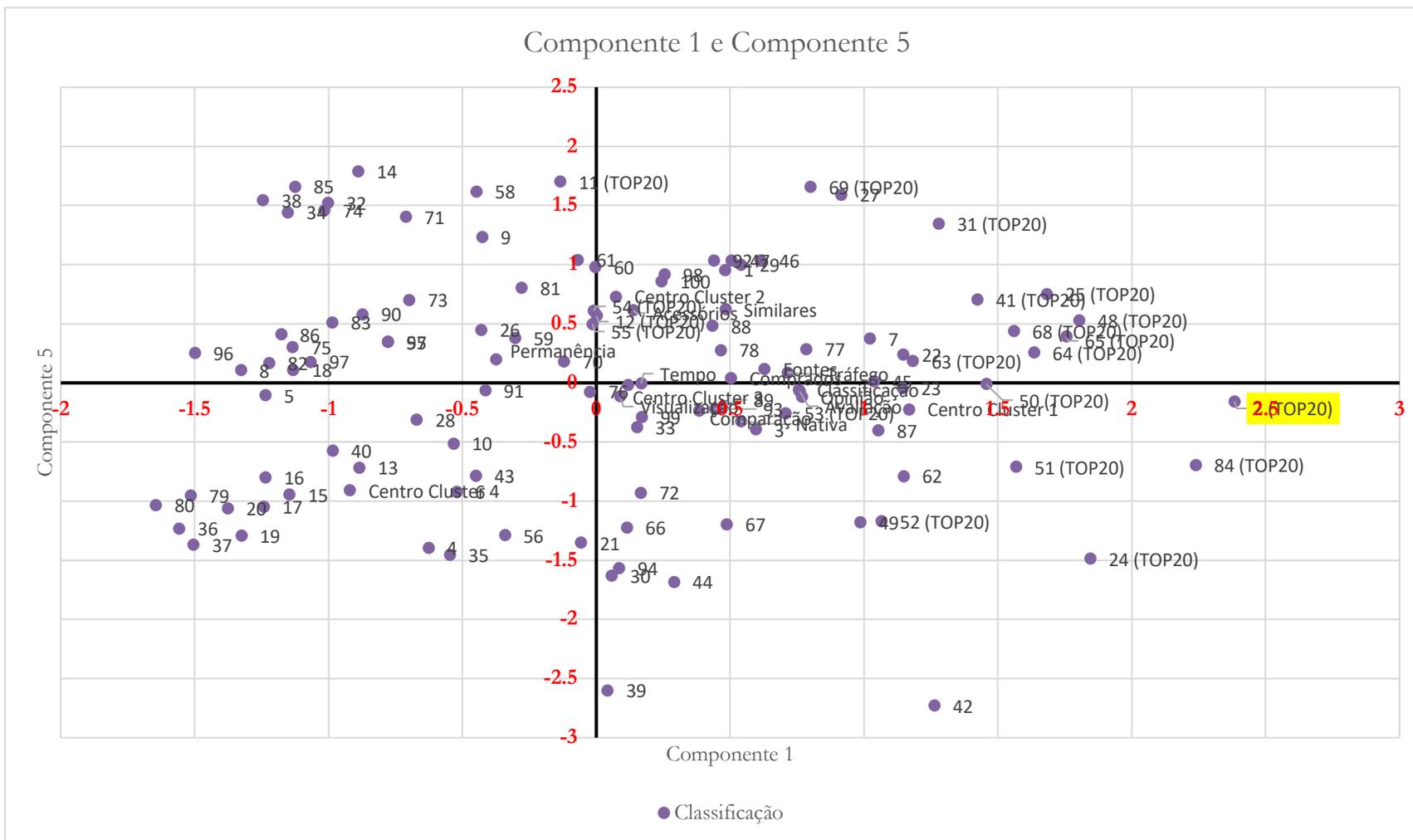
Fonte: Elaboração própria

Anexo 17 - Mapa bidimensional com componentes 1 e 3



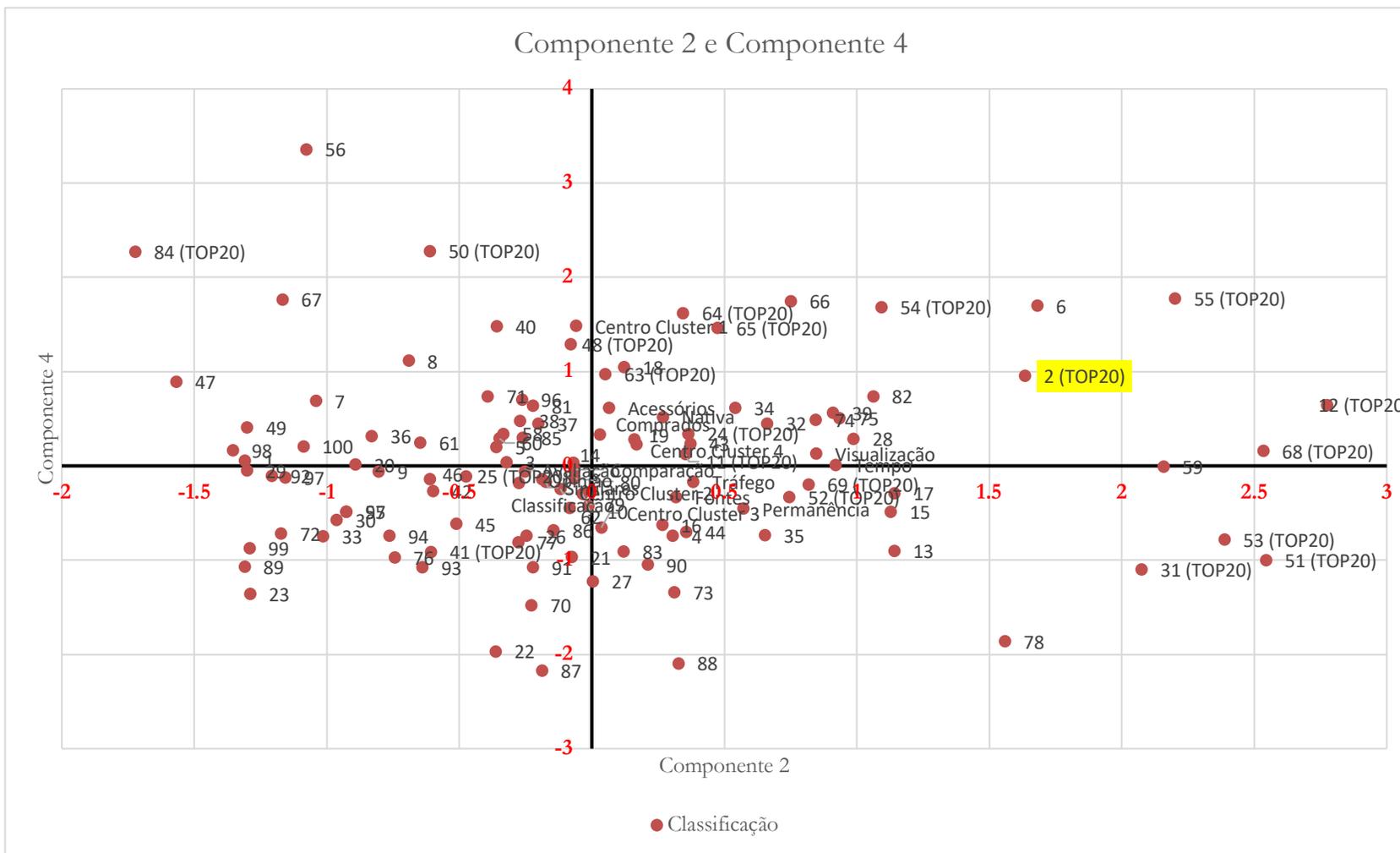
Fonte: Elaboração própria

Anexo 19 - Mapa bidimensional com componentes 1 e 5



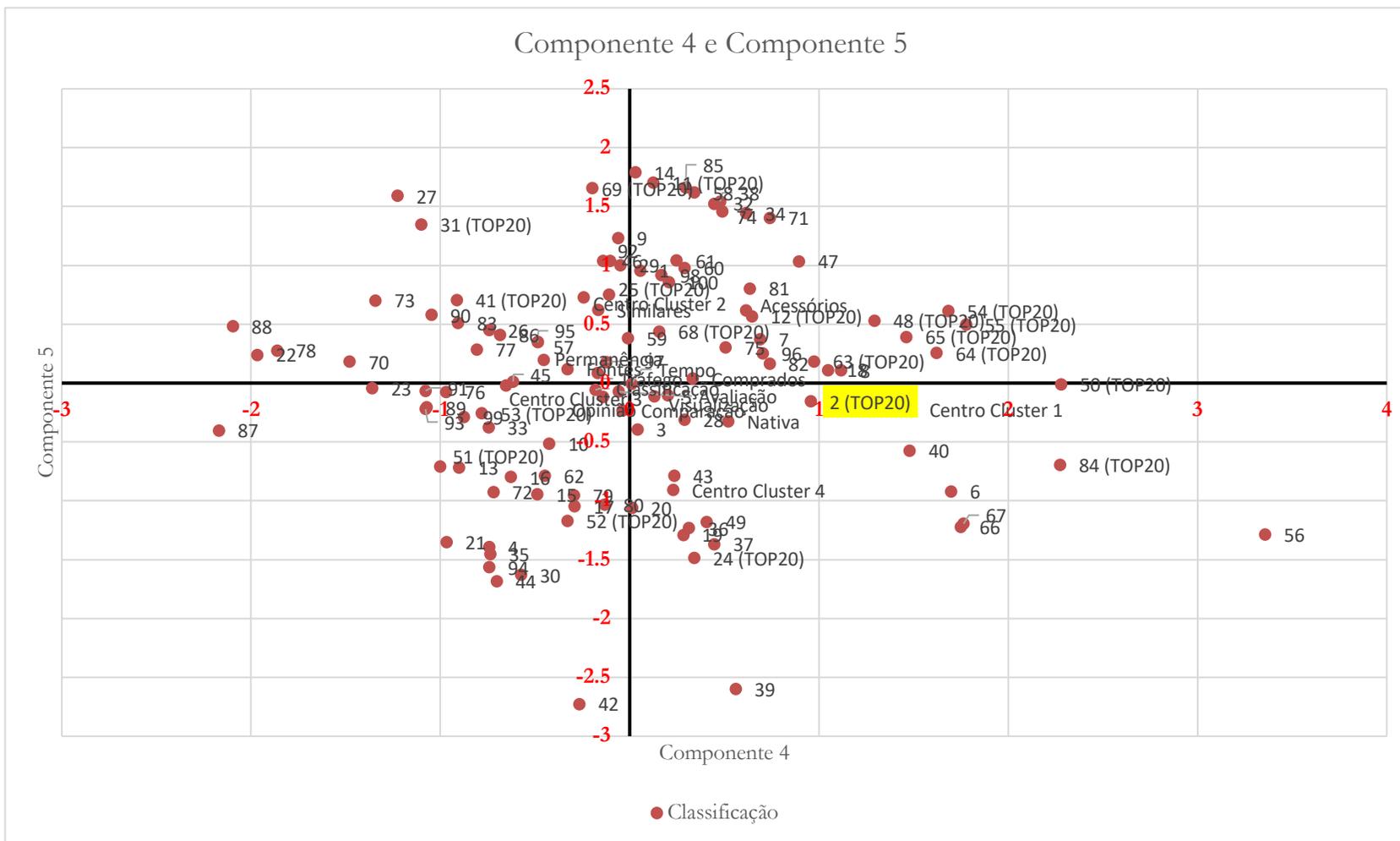
Fonte: Elaboração própria

Anexo 21 - Mapa bidimensional com componentes 2 e 4



Fonte: Elaboração própria

Anexo 25 - Mapa bidimensional com componentes 4 e 5



Fonte: Elaboração própria

Anexo 26 - Dados relativos a 10 empresas tecnológicas que procederam a algumas das maiores IPOs dos últimos anos

Empresa	Preço de cada Ação na IPO	Valor Total da IPO	Capitalização bolsista prevista na IPO	Data da IPO	Último Valor Contabilístico dos Ativos reportado	Data do reporte	Valor da informação	Fontes	Acedido
Alibaba Group Holding Limited	68,00	21.767.215.000	167.600.000.000	19-09-2014	34.234.000.000	30-06-2014	133.366.000.000	https://www.reuters.com/article/us-alibaba-ipo/alibaba-ipo-prices-at-top-of-range-raising-21-8-billion-idUSKBN0HD2CQ20140918	2-ago-18
								http://ipodatabase.net/database/IPO/alibaba-group-holding-limited	6-ago-18
								https://www.sec.gov/Archives/edgar/data/1577552/000119312514347620/d709111d424b4.htm	6-ago-18
Amazon.com, Inc	18,00	54.000.000	300.000.000	15-05-1997	61.092.000	31-03-1997	238.908.000	https://www.wsj.com/articles/SB859220492737069500	9-ago-18
								https://www.nasdaq.com/markets/ipo/filing.ashx?filingid=416090	9-ago-18
								https://www.nasdaq.com/markets/ipo/filing.ashx?filingid=416090	9-ago-18
Dropbox, Inc	21,00	756.000.000	9.200.000.000	23-03-2018	1.437.500.000	31-12-2017	7.762.500.000	https://www.nytimes.com/2018/03/22/business/dealbook/dropbox-ipo.html	10-ago-18
								http://www.nasdaq.com/markets/ipo/company/dropbox-inc-808000-86123	10-ago-18
								https://www.nasdaq.com/markets/ipo/filing.ashx?filingid=12643008	10-ago-18
Facebook, Inc	38,00	16.006.877.000	104.000.000.000	18-05-2012	13.941.000.000	31-03-2012	90.059.000.000	https://dealbook.nytimes.com/2012/05/17/facebook-raises-16-billion-in-i-p-o/	2-ago-18
								http://ipodatabase.net/database/IPO/facebook-inc	2-ago-18
								https://www.sec.gov/Archives/edgar/data/1326801/000119312512240111/d287954d424b4.htm	6-ago-18
Google Inc	85,00	1.666.429.420	23.000.000.000	19-08-2004	2.491.590.000	30-06-2004	20.508.410.000	https://www.economist.com/unknown/2004/08/20/googles-ipo-rollercoaster	9-ago-18
								https://www.nasdaq.com/markets/ipo/company/google-inc-633170-39207	9-ago-18
								https://www.sec.gov/Archives/edgar/data/1288776/000119312504143377/d424b4.htm	9-ago-18
LinkedIn Corporation	45,00	352.800.000	4.250.000.000	19-05-2011	463.576.000	31-03-2011	3.786.424.000	https://www.wsj.com/articles/SB10001424052748703421204576331610501421954	9-ago-18
								http://www.nasdaq.com/markets/ipo/company/linkedin-corp-606852-66240	9-ago-18
								https://www.sec.gov/Archives/edgar/data/1271024/000119312511145240/d424b4.htm	9-ago-18
Snap Inc	17,00	3.400.000.000	24.000.000.000	02-03-2017	4.111.928.000	31-12-2016	19.888.072.000	https://www.nytimes.com/2017/03/01/business/dealbook/snap-ipo-snapchat.html	10-ago-18
								https://www.nasdaq.com/markets/ipo/company/snap-inc-899497-82723	10-ago-18

								https://www.sec.gov/Archives/edgar/data/1564408/000119312517068848/d270216d424b4.htm	10-ago-18
Spotify Technology S.A.	165,90	9.245.852.532	29.500.000.000	03-04-2018	3.202.000.000	31-03-2018	26.298.000.000	https://techcrunch.com/2018/04/03/spotify-opens-at-165-90-valuation-company-at-30-billion/	10-ago-18
								https://www.nasdaq.com/markets/ipo/company/spotify-technology-sa-967774-86175	10-ago-18
								https://www.sec.gov/Archives/edgar/data/1639920/000119312518149451/d579509d424b3.htm	10-ago-18
Twitter, Inc	26,00	1.820.000.000	14.200.000.000	07-11-2013	2.745.647.000	30-09-2013	11.454.353.000	https://money.cnn.com/2013/11/06/technology/social/twitter-ipo-price/index.html	7-ago-18
								http://ipodatabase.net/database/IPO/twitter-inc	7-ago-18
								https://www.sec.gov/Archives/edgar/data/1418091/000119312513431301/d564001d424b4.htm	9-ago-18
Yelp Inc	15,00	107.250.000	900.000.000	02-03-2012	140.066.000	31-12-2011	759.934.000	https://www.reuters.com/article/yelp-idUSL4E8E20NB20120302	10-ago-18
								https://www.nasdaq.com/markets/ipo/company/yelp-inc-691649-68649	10-ago-18
								https://www.nasdaq.com/markets/ipo/filing.ashx?filingid=8457075	10-ago-18

Fonte: Elaboração própria. Valores em USD.

FACULDADE DE ECONOMIA

