

Estudo da imagem da região turística de Lisboa

João H. da Luz Cerdeira¹ e Adriana Fernandes²

jhca@iscte-iul.pt; adriana.fernandes@iscte-iul.pt

¹ Departamento de Marketing, Operações e Gestão Geral, Instituto Universitário de Lisboa ISCTE-IUL, 1649-026, Lisboa, Portugal.

² Departamento de Ciências e Tecnologias da Informação, Instituto Universitário de Lisboa ISCTE-IUL, 1649-026, Lisboa, Portugal.

DOI: 10.17013/risti.40.89–111

Resumo: A indústria turística mundial registou um crescimento anual constante e encontrava-se em franca expansão até março de 2020, quando foi fortemente afetada pela pandemia Covid-19. Acredita-se que para a reinvenção e reforma do setor se terá de investir fortemente na criação e gestão da perceção dos turistas sobre a imagem dos destinos turísticos. Tendo em vista este contexto, o presente estudo objetiva analisar a imagem da região de Lisboa através da utilização de *User Generated Content* (UGC). A metodologia utilizada tem como base de análise três componentes de interações do turista com o destino: imagem Designativa, Estimativa e Prescritiva. Foram utilizados 51 mil *reviews* do ano de 2019, em inglês, contendo opinião de turistas de 131 nacionalidades sobre a região turística de Lisboa. Os resultados revelam que Lisboa tem uma imagem caracterizada por *reviews* positivos e oferecem novos conhecimentos sobre a gestão da imagem do destino.

Palavras-chave: User Generated Content; Tourism Destination Image; Smart Tourism; Lisboa

Study of the image of the tourist region of Lisbon

Abstract: The tourism industry has consistently been flourishing until 2020 when the Covid-19 pandemic broke. It is believed that the knowledge and analysis of tourism destination images will be a crucial tool for the reform and reinvention of the activity and relaunch of tourism. Through the usage of User Generated Content (UGC) collected from *TripAdvisor*, this study aims to reveal the tourism destination image of Lisbon. To analyze the tourism image, a three-phase method was utilized bearing three distinct tourist interactions with the destination: designative image, appraisive image and prescriptive image. This study compiled 51.000 reviews written in English in 2019, with the opinion of 131 different nationalities about the region of Lisbon. Results recognize Lisbon with having several more positive reviews than negative ones and offer new untapped knowledge to manage the region's tourism.

Keywords: User Generated Content; Tourism Destination Image; Smart Tourism; Lisbon

1. Introdução

Até março de 2020 a indústria turística tinha registado um crescimento anual constante e encontrava-se em franca expansão (WTTC, 2019). A pandemia Covid-19 fez tremer os alicerces de toda a economia mundial e teve um impacto sem precedentes na indústria turística (Chakraborty & Maity, 2020). Perdas superiores a \$1,2 mil biliões em receitas de exportação turística e de cerca de 120 milhões em trabalhos diretos (Sigala, 2020) levaram os organismos que controlam as decisões no setor turístico a pensar como relançar o setor e atrair os turistas num cenário pós-pandémico.

Sigala (2020) afirma que as tecnologias vão estar no centro das soluções para a reabertura da economia e do turismo após o profundo impacto ao nível estrutural e transformacional na atividade turística causado pela pandemia. A análise de dados e uma correta gestão das informações sobre a perceção dos turistas disponível virtualmente, como é o caso do *User Generated Content* (UGC) (Kaosiri et al., 2017), serão fundamentais na compreensão de como formar uma *Tourist Destination Image* (TDI) e para a reinvenção desta indústria.

O UGC, aliado à informação turística nos meios de comunicação social, é aceite como informação de apoio aos turistas para desenvolver o seu plano de viagem e tomar uma decisão quanto ao destino turístico (Xiang & Gretzel 2010; Kaosiri et al., 2017; Marine-Roig, 2019). Além disso, é considerado um método altamente fiável para a procura de informação segura e instantânea (Yoo & Gretzel 2011; Kaosiri et al. 2017) uma vez que os utilizadores podem procurar informação e interagir com outros com ampla dispersão (Nezakati et al., 2015). O impacto em todo o processo de viagem torna o UGC importante, tanto para os turistas como para as empresas turísticas que podem recorrer-se dessas informações para atrair novos clientes (Kaosiri et al., 2017; Marine-Roig, 2019).

O presente estudo concentra-se na examinação da imagem de uma região turística, por meio da utilização de *User Generated Content* (UGC) sobre diferentes tipos de recursos turísticos (atrações, atividades relacionadas a viagens, hotéis e restaurantes). Para tal, foi necessário atentar-se para a construção desta imagem de uma perspetiva holística, derivando da imagem percebida do turista e da imagem projetada dos agentes e organizações turísticas. Utilizou-se o modelo proposto por Marine-Roig (2019) criado especificamente para analisar a TDI dando uso a conteúdo criado por visitantes.

Considera-se que os resultados, com base nas necessidades, preferências e opiniões dos visitantes, podem ser úteis para melhor compreensão do que mais atrai o turista na região, para que as organizações voltadas para o turismo possam investir os seus recursos de forma mais sustentada, criando assim uma experiência turística de alto valor para o visitante.

2. Revisão de Literatura

2.1. Imagem de um destino turístico

A soma de convicções, ideias, e impressões que um indivíduo manifesta por um destino, tem sido alvo de estudo pela comunidade científica ao longo das últimas 4 décadas (Crompton, 1979; Ekinci & Hosany, 2006; Marine-Roig, 2017; Shankar, 2018). A TDI

consiste na interpretação subjetiva da realidade feita pelo turista e, não só tem uma influência considerável na intenção de visitar e propensão de regressar/recomendar um destino, como tem ainda um papel fundamental na evolução do próprio destino (Chon, 1990; Gretzel & Yoo, 2008; Marine-Roig, 2017).

De uma forma geral a TDI é considerada uma construção dialética onde a imagem projetada (*projected*) – representada pela imagem criada por agentes e organizações turísticas sobre determinado destino – e a imagem percebida (*perceived*) – subjetiva e contextual, criada por cada indivíduo sobre um destino turístico – se influenciam mutuamente num fluxo constante (Marine-Roig, 2015), ou seja, uma está constantemente a depender da outra. A TDI criada por um indivíduo retrata a sua perceção e interação com a imagem projetada, o seu contexto individual (contexto social, variáveis demográficas, estereótipos, identidade, entre outros) e a sua experiência e relação com outras fontes de construção de informação (Marine-Roig, 2015), como os conteúdos de viagem criados na *internet*.

No intuito de decifrar a imagem de um destino turístico, por meio da utilização de recursos como o conteúdo criado por turistas, a comunidade científica tem recorrido a diferentes plataformas digitais. Nomeadamente, a análise de fóruns de conteúdo de viagens (Tamajón & Valiente, 2017; Gkritzali et al., 2018), *tweets* (Jabreel et al., 2017), *Facebook posts* (Lalicic et al., 2019), *photo posts* (Agustí, 2018; Deng et al., 2019) e *blogs* de viagens (Choi et al., 2007; Marine-Roig, 2010). Embora a grande maioria dos estudos existentes se centrem neste último, a verdade é que esse tipo de recurso está a tornar-se cada vez menos utilizado pelos turistas. Em contrapartida, o foco das pesquisas tem-se deslocado para plataformas digitais como o *TripAdvisor* (Gretzel & Yoo, 2008; Kladou & Mavragani, 2015; Marine-Roig, 2017; Marine-Roig, 2019), devido ao crescimento exponencial das avaliações e comentários de viagens feitos *online* (*Online Travel Reviews* - OTRs) (Marine-Roig, 2019).

O estudo da imagem de um destino foi realizado por diferentes autores e cada um deles avalia a imagem de um destino por meio de um conjunto de diferentes componentes (Apostolopoulou & Papadimitriou, 2015; Baloglu & McCleary, 1999; Kladou & Mavragani, 2015; Shankar, 2018), sendo a grande maioria dos estudos guiados por modelos herdados da psicologia (Prayag et al., 2017; Marine-Roig, 2019).

Neste estudo utilizou-se o modelo proposto recentemente por Marine-Roig (2019), criado especificamente para analisar a TDI utilizando conteúdo criado por turistas, correspondendo assim ao objetivo deste estudo. Este modelo divide a interação entre o turista e o ambiente em três áreas: conhecer algo (Cognitivo/designativo); sentir algo (Afetivo/estimativo) e fazer algo em relação a isso (Conativo/prescritivo) (Marine-Roig, 2019).

A componente designativa diz respeito aos elementos físicos e temporais da imagem e responde sucintamente às questões “O quê?”, “Onde?” e “Quando?” (Baloglu & McCleary, 1999; Martín & Bosque, 2008; Marine-Roig, 2019). Já a componente estimativa, por sua vez, incorpora a avaliação e preferência do indivíduo e relaciona por isso a resposta aos seus sentimentos, valor e significado atribuído à imagem (Pocock & Hudson, 2000; Kladou & Mavragani, 2015; Marine-Roig, 2019). E por fim a componente prescritiva, incluída recentemente em estudos sobre a imagem de um destino, corresponde à reação dos visitantes perante a imagem realizada de um destino. Neste estudo é medida como

o comportamento que sugere possível recomendação e/ou futura re-visita (Martín & Bosque, 2008; Kladou & Mavragani, 2015; Shankar, 2018; Marine-Roig, 2019).

2.2. User Generated Content

Uma das principais fontes de informação para o setor do turismo (Buhalis & Amaranggana, 2014) encontra-se na enorme quantidade de conteúdo criado por utilizadores (UGC) nos vários canais das mídias sociais (*social media*). Para o turista, o UGC é aceite como informação de apoio para desenvolverem os seus planos de viagem e auxiliá-los na tomada de decisão de viagem (Xiang & Gretzel 2010). Além disso, é considerado um método altamente confiável para obtenção de informação instantânea (Gretzel & Yoo, 2008; Yoo et al., 2009), uma vez que os utilizadores podem procurar informações e interagir com outras pessoas com ampla dispersão (Nezakati et al., 2015; Kaosiri et al., 2017).

Tendo em conta a importância deste conteúdo para os visitantes, as organizações e agentes do turismo devem ser capazes de recolher esta enorme quantidade de dados (*Big Data*) e aplicar técnicas de análise para promover o aumento da eficiência operacional, a inovação dos serviços, a dinamização dos modelos de negócio e a melhoria da experiência turística para os visitantes (Gretzel & Yoo, 2008; Buhalis & Amaranggana, 2014; Marine-Roig, 2019).

As *online travel reviews* - OTR's, e classificações de produtos, serviços e empreendimentos representam a fonte mais prevalente e acessível de UGC no turismo (Gretzel & Yoo, 2008). De acordo com Kaosiri et al. (2017) os comentários de utilizadores servem dois propósitos funcionando, simultaneamente, com um carácter informador, na medida em que providenciam informação na óptica do utilizador de um produto, e ainda como método de recomendação, ao fornecer uma plataforma que permite aos consumidores expressarem o seu grau de recomendação ou contentamento com um produto, estabelecimento, serviço, ou mesmo destino, sob a forma de *electronic word of mouth* (eWOM). Segundo Marine-Roig (2019) essa fonte de informação pode ser considerada o principal atributo de um destino, ultrapassando a crença anterior de que eram as atrações de um destino que constituíam o elemento decisivo na motivação de visita por parte dos turistas.

3. Metodologia e Região de estudo

3.1. Região de estudo

Segundo os dados fornecidos por World Travel & Tourism Council (WTTC, 2019) a contribuição total em 2018 do sector turístico para o produto interno bruto (PIB) português foi de 19,1%, ou seja, quase um quinto do PIB do país. O impacto causado por chegadas internacionais foi de 20,6 mil milhões de euros, ou cerca de 23,1% das exportações do país. A mesma organização estima que a indústria turística é ainda responsável por cerca de 21,8% da empregabilidade total de Portugal.

Lisboa por si só foi, em 2018, responsável por mais de 1/4 de todas as dormidas nacionais (25,9%) e quase 1/3 da totalidade de hóspedes (29,9%). Quando isolamos o *outbound tourism* (turismo internacional), a região foi responsável por 36% da totalidade de hóspedes e quase 29% das dormidas (INE, 2018). Lisboa assume-se como uma ferramenta

de combate à sazonalidade turística nacional uma vez que cerca de dois terços das dormidas em Lisboa ocorre entre os meses de Setembro e Junho, fora do período sazonal, e possui uma vulnerabilidade turística inferior à do resto do país (Nunes, 2017; Silva, et al. 2018).

A entidade responsável pelo turismo da região de Lisboa (Associação Turismo de Lisboa), divide a região por 5 centralidades (zonas), Lisboa, Sintra, Cascais, Arco do Tejo e Arrábida, as quais vão ser estudadas ao longo do artigo de forma isolada e em grupo.

3.2. Descrição geral

Para revelar a imagem turística de Lisboa, foi seguida a metodologia proposta por Marine-Roig & Clavé (2015). São por isso segmentadas as 4 etapas fulcrais que antecedem os resultados do estudo, sendo: Seleção dos websites e plataformas a utilizar; Recolha de dados e construção de base de dados primária; Fase pré-processamento, que vai englobar toda a filtragem, gestão e organização dos dados e vai resultar na construção de uma segunda base de dados (*corpus*); e Fase analítica, onde vão ser utilizados métodos quantitativos que vão potenciar o objetivo principal do estudo.

3.3. Seleção da plataforma

A plataforma escolhida para o estudo foi o *TripAdvisor*, considerada a plataforma de maior visibilidade, popularidade e tamanho (Baka, 2016; Marine-Roig, 2017;2019) do setor de viagens e turismo. Criada em 2000, está presente em 49 países e em 28 idiomas, tem cerca de 460 milhões de visitantes mensais e mais de 830 milhões de avaliações e opiniões sobre a indústria turística mundial (*TripAdvisor LLC*, 2017).

As OTRs do *TripAdvisor* estão organizadas consoante a sua natureza e localização. Neste estudo foram consideradas OTRs de Restaurantes, Atividades e Alojamentos. O *TripAdvisor* engloba ainda outras naturezas que, por não serem fontes de OTRs (*Airlines*, *Rent a car*, fóruns de pergunta e resposta) não foram considerados neste estudo.

3.4. Recolha de dados e construção de base de dados primária

Para extração de dados (*data scrapping*) foi utilizada a ferramenta *R software*, que permite a pré-filtragem do que é extraído. Como tal todo o conteúdo não criado pelo utilizador não é considerado, e seu *download* não é realizado.

Para construção da base de dados primária, com todos os URLs (*Uniform Resource Locator* - Endereço de rede) que continham *reviews* da região de estudo foi utilizado o pacote *RCrawler*. Todas as páginas do *TripAdvisor* têm o seu propósito indicado no URL no seguimento do domínio (“*www.TripAdvisor.com*”). As páginas que contém OTRs apresentam sempre o código “*ShowUserReviews*” que constituiu o filtro final para agregar todas as páginas com OTRs de todas as localizações através das ferramentas *linkextractor* e *rcrawler*.

Durante a construção da base de dados primária a região de estudo foi dividida por centralidade e microcentralidade e os dados foram processados de acordo com a região correspondente no *TripAdvisor*. Só foram consideradas localizações com número de OTRs superior a 100. Foram criados 3 vetores de listas de URLs de OTRs para cada natureza (OTRs de Restaurantes, de Alojamentos e de Atividades).

3.5. Pré-processamento

Pela sua natureza, o *TripAdvisor* agrega resultados de localizações não só da região em específico como também possui um algoritmo para dispor os resultados que a plataforma considera “perto” ou “sugestivos” ou ainda páginas promovidas. Assim, paralelamente à formação da base de dados primária foram eliminadas possíveis duplicações de URLs e feita uma correspondência exata da localização a considerar.

Para cada uma das naturezas descritas no ponto anterior foi feita nova pré filtragem para respeitar a metodologia proposta. Neste sentido, foram criados 3 parâmetros de filtragem, sendo:

1. Data: A amostra foi filtrada para devolver apenas *reviews* de 2019;
2. Linguagem: Por ser a língua estrangeira mais frequente da plataforma, apenas OTRs em inglês foram consideradas;
3. Tamanho: Para harmonizar a análise quantitativa as amostras foram reduzidas a um número comum às 3 naturezas. 17.000 *reviews* de cada natureza foram considerados.

Na fase de *data scrape* da página *web* foram consideradas três fontes de conteúdo relevante para o estudo em cada um dos resultados da base de dados primária:

- *HTML¹ metadata*: consiste na informação de uma *webpage* a ser lida pelo *browser*. A *metadata* de cada URL da base de dados primária, no *TripAdvisor* contém o título da *review* (inclui localização e título escrito); o *coding* da página (protocolo *http*², tipo *html*) e a linguagem da página. O próprio URL, que faz parte da *metadata*, revela também o código interno da localização; o código interno do alojamento, restaurante ou atividade; o código único da OTR; o nome interno do alojamento, restaurante ou atividade; a localização e o nome da região em que se enquadra.
- *Corpo de Texto*: a parte mais importante da OTR. Trata-se da componente escrita pelo visitante, a sua opinião e experiência expressa no título e comentário.
- *Elementos paratextuais*: Todos os elementos que, não fazendo parte do corpo de texto, são considerados para análise da imagem de um destino turístico. Nomeadamente a classificação dada; a localização do alojamento, restaurante ou atividade; a data da *review*; a linguagem ou os dados sociodemográficos do *reviewer*.

Foi feito um *webscrapping* dos elementos acima mencionados utilizando o pacote *rvest* e criada uma base de dados secundária, denominada *corpus* que constituiu uma tabela com os seguintes elementos: URL; nacionalidade e localidade do *reviewer* (quando possível); classificação da *review*; data da OTR; nome do recurso; categoria do recurso (para restaurantes e atividades); classificação geral do recurso (para alojamentos); localização do recurso; linguagem; título e corpo de texto.

¹ HTML (HyperText Markup Language) – linguagem de marcação utilizado nas páginas *web*

² HTTP (Hypertext Transfer Protocol) – protocolo seguido pelos servidores na comunicação *www*

Tal como sugerido por Welbers, et al. (2017) os elementos textuais da base de dados (Título e Corpo de Texto) foram ainda preparados para uma posterior análise quantitativa fundamentada no uso do pacote *quanteda*.

Foram realizadas operações de *string* (*String* é uma sequência de caracteres que representam texto) para retirar todo o conteúdo genérico que pudesse aparecer (ex. “*Este texto foi traduzido pelo Google*”); remover todo o *whitespace* (espaço que facilita a leitura como parágrafos ou avançamentos de texto); padronizar todas as palavras em minúsculas para que não houvessem erros de contagens; fazer um *stemming* do texto (*Stemming* é uma técnica que reduz as palavras a uma forma comum a todas as terminações morfológicas e sufixos); e remover todas as *stopwords* (conjunto de palavras que são apenas conectores ou separadores de palavras e pontuação). Por fim foi realizado o *Tokenization* dos dados – processo de separar o texto em *tokens* ou elementos individuais. Estes elementos são palavras simples ou palavras compostas tais como “*long-term*”.

3.6. Fase Analítica

Inicialmente foi criada uma tabela de frequências de todos os *tokens* através do pacote *quanteda*, nomeadamente a função *DTM* (*Document term matrix*), que produz uma matriz na qual as linhas são elementos, as colunas são termos e as células indicam o número de repetições por documento.

Sequencialmente, na análise estatística quantitativa, as palavras mais vezes repetidas foram consideradas. Estas foram categorizadas de forma exaustiva, mutuamente exclusiva e independente (Stepchenkova, 2012). As 5 categorias criadas, num momento *a priori*, de acordo com o modelo proposto por Marine-Roig (2019), foram:

1. Atrações, Estruturas e Serviços:
Foi efetuado uma análise quantitativa aos *reviews* e menções dos estabelecimentos e categorias em que se inserem, uma vez que, um maior número de OTRs num local indica uma maior popularidade desse estabelecimento ou atração/atividade dentro do panorama da região. Para esta categoria foi utilizado um método dedutivo de contagem relativa e absoluta dos termos únicos textuais das OTRs (Welbers, et al., 2017).
2. Dimensão Espaciotemporal:
Os recursos do *TripAdvisor* permitem recolher o local da atração, serviço, estabelecimento ou atividade que está a ser comentado. É por isso possível ver a sazonalidade das *reviews* por microcentralidade, natureza ou geral, e perceber ainda a evolução da eWOM que certas regiões tiveram ao longo do período de estudo ou quais regiões estão a ter um maior influxo de comentários ao longo do tempo.
3. Dimensão Avaliativa:
Todas as OTRs da plataforma *TripAdvisor* têm uma classificação entre “Terrível” e “Excelente” numa escala de 1 a 5, respetivamente. De acordo com o modelo proposto por Marine-Roig (2019) a análise irá utilizar dois tipos de métricas diferentes:

- *Score* médio: São consideradas avaliações positivas (*score* +) as avaliações de “excelente” (5) e “muito bom” (4) e avaliações negativas (*score* -) as avaliações “Fraco” (2) e “Terrível” (1). As avaliações de *score* médio (3) não alteram a escala. É analisada em valor percentual face à amostra.
- Média ponderada: a segunda métrica constitui uma média ponderada através da conversão das avaliações numa escala de 10 valores em que: “Excelente” = 10; “Muito bom” = 7,5; “Médio” = 5; “Fraco” = 2,5; “Terrível” = 0.

4. Dimensão afetiva:

A análise de sentimentos tenta deduzir o sentimento de uma *review* com base nas palavras e expressões que são utilizadas para descrever uma experiência. Classifica-se por isso como um mecanismo de avaliação dedutiva uma vez que a polaridade de uma palavra é ditada por um conjunto de regras pré-concebidas (Welbers, et al., 2017).

Para a análise de sentimentos foi utilizada a ferramenta *sentimentr* que, através da utilização de dicionários integrados no pacote, interpreta as frases do *corpus*, equacionando expressões que mudem a polaridade geral de uma frase, intensificando o seu significado ou alterando-o. Na língua inglesa, estas expressões são denominadas de *valence shifters*, e incluem, *negators*, *amplifiers* (*intensifiers*), *de-amplifiers* (*downtoners*), *adversative conjunctions* entre outros (Rinker, 2019).

Para um cálculo mais real da análise de sentimentos, foi utilizado uma ferramenta para ponderação das médias denominado *average_weighted_mixed_sentiment*. Esta função é aconselhada na análise de textos com várias frases curtas de carácter opinativo. Nestas condições é frequente assistir a um conjunto de frases positivas ou neutras precederem uma frase negativa no sentido de a atenuar. A fórmula equaciona esta possibilidade atribuindo um peso ligeiramente superior às frases negativas e baixando o peso das frases neutras (Rinker, 2019).

5. Dimensão de atitude e comportamento:

Tal como na Dimensão Afetiva (4) para analisar o grau de recomendação de uma *review* foi utilizado o pacote *quanteda* e *sentimentr*. Ao contrário da dimensão anterior foi criado um dicionário próprio para a análise, que contém listagens de *tokens* sugestivos de recomendações positivas (*do not miss*, *don't miss*, *worth a visit*, etc.) e negativas (*avoid*, *be careful*, *beware*, etc.). O dicionário utilizado surgiu da adaptação do proposto por Marine-Roig & Ferrer-Rosell (2018). A palavra “*recommend*” e derivados como “*recommended*” e “*recommendation*” foram analisadas isoladamente e cruzadas com a polaridade da frase dada pela ferramenta *sentimentr* para haver um escrutínio mais rigoroso da intenção de recomendação ou não, através da polaridade da frase.

4. Resultados

4.1. Amostra

Ao todo foram considerados 51.000 *reviews*, com uma distribuição em pelo menos 131 nacionalidades sendo que os 10 países mais intervenientes são, respetivamente:

Estados Unidos (9726); Reino Unido (8663); Portugal (2715); Canadá (1978); Irlanda (1363); Austrália (1246); Singapura (834); Países Baixos (819); Alemanha (819) e Espanha (567).

Os comentários em si refletem uma predominância nas palavras relacionadas com alojamento (Tabela 1) sendo que *hotel* (24129), *room* (21777) e *stay* (14928) encontram-se, respetivamente no *rank* 1, 3 e 9 das palavras mais mencionadas. Os comentários refletem maioritariamente palavras com entoação positiva, *great* (22323), *good* (19670), *friend*(ly) (13109), *nice* (13043), *love* (11462).

Keyword			Contagem			Keyword			Contagem		
	Palavras únicas: 55,084	Total palavras: 3,981 milhões	%		Palavras únicas: 55,084	Total palavras: 3,981 milhões	%		Palavras únicas: 55,084	Total palavras: 3,981 milhões	%
1	hotel	24129	0,606%	11	restaur	14572	0,366%				
2	great	22323	0,561%	12	recommend	14104	0,354%				
3	room	21777	0,547%	13	friend	13109	0,329%				
4	good	19670	0,494%	14	nice	13043	0,328%				
5	lisbon	17792	0,447%	15	time	12026	0,302%				
6	food	17209	0,432%	16	one	11467	0,288%				
7	staff	16745	0,421%	17	love	11462	0,288%				
8	place	15661	0,393%	18	servic	11338	0,285%				
9	stay	14928	0,375%	19	tour	11193	0,281%				
10	us	14851	0,373%	20	day	11013	0,277%				

Tabela 1 – Top 20 keywords dos comentários das OTRs

4.2. Imagem Designativa

Quanto às OTRs da natureza das Atividades, no Top 10 constam apenas atividades da microcentralidade Lisboa, referente a locais e monumentos da cidade (Tabela 2). Entretanto, foram também referenciadas duas empresas de fornecimento de *tours*, a *Discover Lisbon* (466) e *Withlocals* (321) e o serviço de transporte público *Tram 28* (534) como atividades a realizar na localidade. A atração *Time Out Market Lisboa* foi a mais comentada da amostra.

Nome da Atividade	Localização	Contagem	Tipo de Atividade / Produto
Time Out Market Lisboa	Lisbon	947	Cooking Classes, Points of Interest and Landmarks
Castelo de S. Jorge	Lisbon	650	Historic Sites, Castles
Torre de Belem	Lisbon	610	Points of Interest and Landmarks
Tram 28	Lisbon	534	Tramways
Mosteiro dos Jeronimos	Lisbon	498	Historic Sites, Architectural Buildings, Sacred and Religious Sites

Nome da Atividade	Localização	Contagem	Tipo de Atividade / Produto
Discover Lisbon	Lisbon	466	Walking Tours, City Tours, Night Tours, Bar, Club and Pub Tours, Sightseeing Tours More
Oceanario de Lisboa	Lisbon	459	Aquariums
Alfama	Lisbon	445	Neighborhoods
Praca do Comercio (Terreiro do Paco)	Lisbon	338	Points of Interest and Landmarks
Withlocals	Lisbon	321	Food Tours, Wine Tours and Tastings, Private Tours, City Tours, Sightseeing Tours More

Tabela 2 – Atividades mais comentadas em OTRs

Foram contabilizados 278 tipos de atividades diferentes na região. É importante realçar que cada recurso desta natureza no *TripAdvisor* tem associada, geralmente, múltiplos tipos de atividade. Espantosamente, na análise realizada, o tipo mais comum de atividade citada pelos visitantes são os *tours*, cujas variedades somam 12 dos 20 tipos de atividades mais comuns. Dentro do mesmo Top 20 incluem-se ainda atividades relacionadas com monumentos e património e atividades relacionadas com gastronomia e bares.

No que diz respeito aos restaurantes, 18 dos 20 mais mencionados estão localizados na centralidade de Lisboa, sendo os dois restantes na localidade Sintra. Sem grande surpresa os restaurantes com gastronomia *European* (9895), *Mediterranean* (9187) e *portuguese* (5706) completam as três categorias mais comentadas pelos utilizadores da plataforma (Tabela 3).

Região	País			Culinária				
European	9895	21,0%	Portuguese	5706	12,1%	Seafood	3194	6,8%
Mediterranean	9187	19,5%	Italian	1284	2,7%	Vegetarian	2509	5,3%
International	2150	4,6%	Indian	1241	2,6%	Contemporary	1135	2,4%
Asian	911	1,9%	American	504	1,1%	Healthy	894	1,9%
Middle Eastern	141	0,3%	Nepali	429	0,9%	Vegan	781	1,7%

Tabela 3 – Tipo de Restaurantes mais populares nas OTRs

Quanto aos OTRs referentes à natureza Alojamento todos os alojamentos do Top 20 são localizados em Lisboa. A totalidade da Amostra da natureza de Alojamentos tem 1066 alojamentos dos quais quase 50% tem classificação igual ou superior a 4,5 pontos no *TripAdvisor*. A média de classificação do *TripAdvisor* para esta natureza é de 4,05 pontos.

A amostra recolhida revela que os trimestres com maior incidência de *reviews* são o terceiro (15.937) e o segundo (14.313) que corresponde ao período entre abril e setembro, inclusivamente, com julho, outubro e setembro a serem os meses com maior afluência de OTRs, respectivamente. A região de estudo tem 58 localizações do *TripAdvisor* afetas.

Destas, a amostra aleatória recolheu OTRs de 54, com as seguintes localizações a não estarem representadas na amostra: Póvoa de Sta. Iria; Pegões; Samouco e Cajados. A metrópole, Lisboa, detém a grande maioria dos *reviews* sendo responsável por 67% de todos os comentários de restaurantes, e 86% de alojamentos e atividades.

4.3. Imagem Estimativa

A Tabela 4 revela a dimensão avaliativa da amostra. Regra geral, as classificações das naturezas são bastante positivas, especialmente no que diz respeito às OTRs das Atividades e dos Restaurantes.

Ao analisar a amostra foi possível constatar que embora a média ponderada e o *score* das atividades seja, no geral, a mais elevada da amostra, o *Top 5* de atividades mais comentadas não seguiram a tendência das restantes ao apresentar classificações abaixo da média (*Time out market*, *Castelo são Jorge*, *Torre de Belem*, *Tram 28* e *Mosteiro dos Jerónimos*, respectivamente), especialmente o recurso *Tram 28*, que apresenta uma média ponderada de apenas 7,36 e um *score* negativo superior a 11%.

	Score +	Neutro	Score -	% Score +	% Score -	Média Ponderada
<i>Restaurantes</i>	15115	752	1133	88,91%	6,66%	8,82
<i>Alojamentos</i>	14580	1244	1176	85,76%	6,92%	8,52
<i>Atividades</i>	15191	1072	737	89,36%	4,34%	8,84

Tabela 4 – Avaliação por natureza da amostra

A Tabela 5 representa a análise sentimental descrita em 3.6 por centralidade. É possível constatar que os comentários têm frases geralmente positivas, com destaque para a centralidade de Lisboa que tem a pontuação mais elevada de *sentimentr* geral (0,234). De realçar ainda que, embora tenha uma contagem de palavras relativamente reduzida, a centralidade do Arco do Tejo apresenta valores negativos na análise sentimental aos comentários relativos a Alojamentos e Atividades, representando as únicas médias negativas da amostra.

Centralidade	Natureza	Contagem palavras	Sentimentr	Sentimentr agregado
<i>Arco do Tejo</i>	Atividades	1436	-0,074	0,162
	Alojamentos	1129	-0,204	
	Restaurantes	7323	0,225	
<i>Arrábida</i>	Atividades	20131	0,283	0,230
	Alojamentos	25225	0,220	
	Restaurantes	41654	0,212	
<i>Cascais</i>	Atividades	48437	0,246	0,216
	Alojamentos	138191	0,201	
	Restaurantes	184336	0,219	

Centralidade	Natureza	Contagem palavras	Sentimentr	Sentimentr agregado
Lisboa	Atividades	975942	0,222	0,234
	Alojamentos	1522263	0,211	
	Restaurantes	695365	0,290	
Sintra	Atividades	101846	0,141	0,186
	Alojamentos	103958	0,187	
	Restaurantes	109258	0,219	

Tabela 5 – Valor de Sentimentr por centralidade e natureza

A ferramenta *Sentimentr* contabilizou, no geral, um total de 290.885 frases, das quais 210.119 têm uma polaridade positiva, 44.039 tiveram uma atribuição de score ou carácter neutro, e 36.727 têm uma polaridade negativa. Os restaurantes registam a melhor percentagem de frases com polaridade positiva, e, sem surpresa, são também a natureza que apresenta melhor índice de análise sentimental (Tabela 6).

	Frases positivas	%	Frases negativas	%	Sentimentr
<i>Restaurantes</i>	62 910	74%	9499	11%	0,2671
<i>Alojamentos</i>	88 828	71%	17635	14%	0,2086
<i>Atividades</i>	58 381	72%	9593	12%	0,2173

Tabela 6 – Análise Sentimentos por natureza

4.4. Imagem Prescritiva

Quanto à imagem prescritiva, que tem em consideração a possível recomendação e/ou futura re-visita de um destino turístico, foram contabilizadas um número de recomendações positivas muito superior às recomendações negativas. Na totalidade, quase 4% do total de frases da amostra contém recomendações positivas, e menos de 0,3% das frases tem recomendações negativas o que favorece a recomendação positiva. Quando analisamos ao nível das *reviews*, 1 em cada 5 (20,8%) contém recomendações positivas, contrastando com a percentagem de recomendações negativas de 1,4% (Tabela 7).

	Recomendações positivas	Recomendações negativas
Contagem	11217	844
% relativa ao n° frases	3,825%	0,274%
% relativa ao n° reviews	20,820%	1,424%

Tabela 7 – Análise prescritiva absoluta

Quanto à imagem prescritiva por natureza, tal como na imagem estimativa os restaurantes têm o melhor rácio de recomendações positivas por frase (5,037%) e por

review (22,518%) assim como têm o mais elevado número absoluto de recomendações positivas por *review* e total (Tabela 8).

	Restaurantes		Alojamentos		Atividades	
	Positivo	Negativo	Positivo	Negativo	Positivo	Negativo
Contagem	4099	300	3476	305	3642	239
% frases	5,037%%	0,353%	4,279%	0,357%%	4,470%	0,275%
% reviews	22,518%%	1,500%	19,565%%	1,582%%	20,382%%	1,188%%

Tabela 8 – Análise prescritiva por natureza

5. Discussão e Conclusões

A pandemia Covid-19 teve um impacto sem precedentes no setor turístico em Portugal, sendo responsável por perdas elevadas de receita e inúmeros postos de trabalho. Torna-se por isso indispensável que o país reaja e aplique recursos para melhorar a sua imagem e atratividade de forma sustentável, e que projete a reforma do setor, responsável por quase 1/3 do PIB nacional. Sigala (2020) afirma que o conhecimento e gestão de uma TDI vai ser essencial nas fases de resposta e recuperação à pandemia. Nesse sentido o presente estudo teve como objetivo revelar qual a imagem da região de Lisboa através da utilização de UGC.

Para concretizar este estudo foram analisados 51.000 OTRs da plataforma *TripAdvisor*, escritos na língua inglesa no ano de 2019, por mais de 130 nacionalidades. Para harmonizar a análise estatística foram recolhidos valores iguais de cada uma das naturezas dos recursos da plataforma (Atividades, Alojamentos e Restaurantes). O modelo de análise foi baseado no extenso trabalho de Marine-Roig, investigadora da universidade de Lleida na Catalunha e descrito no capítulo “Metodologia e Região de estudo”.

Considerando o objetivo geral deste estudo pode dizer-se que, na generalidade, a imagem turística da região de Lisboa de 2019 é bastante positiva em todas as naturezas e, especialmente, no que diz respeito aos restaurantes. As recomendações positivas são muito mais frequentes que as negativas.

Existe na imagem da região turística de Lisboa uma predominância no número de *reviews* na metrópole. Ao todo a centralidade de Lisboa detém 80% dos OTRs da Amostra, com as centralidades Cascais (10%), Sintra (8%), Arrábida (2%) e Arco do Tejo (0,3%) a seguirem-se.

No que diz respeito às Atividades, a Amostra conclui que, embora apresentem o melhor *score* (89,36% positivo) e resultado da média ponderada (8,84/10), têm o segundo melhor índice de *Sentimentr* (0,22) e não são a natureza com melhor índice de recomendação. Esta análise parece revelar que não existe uma relação direta clara entre estes elementos da imagem e demonstra ainda que, embora a satisfação com esta natureza é a mais evidente através da pontuação dada, a intensidade dos comentários parece ter uma polaridade inferior à das restantes naturezas. Da mesma forma, a natureza parece ter menos capacidade de cativar a lealdade dos visitantes que os restaurantes, e menos expressões de recomendação positiva são usadas. É também importante realçar

a prevalência dos *tours* na imagem designativa da Amostra, não só pelo número de comentários que lhe são afetos como também por serem o tipo de atividade mais comum e por parecerem ser o tipo de atividade que melhor pontua na imagem estimativa. É relevante por isso realçar o papel que as atividades desta natureza têm na ótica do visitante da região. As atividades mais populares parecem manifestar *scores* positivos e médias ponderadas abaixo da média, com todo o *top5* de recursos mais visitados a respeitar esta tendência.

A natureza dos Alojamentos é a natureza com maior volume de frases nas suas OTRs (125,221), o que indica que os visitantes parecem escrever maiores *reviews* nesta categoria. A imagem designativa demonstra a importância que o pequeno-almoço tem nestes estabelecimentos com a palavra *breakfast* a ser a quinta mais mencionada nos comentários (10,163). Embora tenham maiores comentários, o índice de recomendação (4,279% das frases) o valor de *sentimentr* (0,209), e a média ponderada (8,52/10) são os mais baixos da amostra, o que parece revelar que os comentários têm uma polaridade menos intensa face ao resto da amostra e são, provavelmente, na sua maioria, descritivos. Os 1066 alojamentos da amostra têm uma classificação do *tripadvisor* elevada com quase 50% (49,43%) a registar um valor igual ou superior a 4,5. A média em 2019 parece aumentar face à geral uma vez que se registou um aumento na média de classificação dada nos *reviews* (4,437 vs. 4,5). Comparativamente à classificação geral dos alojamentos, a amostra de 2019 apresenta 72,60% de classificações dadas pelos *reviewers* iguais ou superiores a 4,5, o que pode levar a um aumento da classificação geral.

Os Restaurantes são a natureza com maior dispersão de comentários por centralidade, com Lisboa a assumir 67,1%. Têm também a segunda melhor média ponderada, muito próxima das atividades (8,82/10), o melhor índice de recomendação (4,8% das frases e 22,5% das *reviews*) e a melhor classificação de *sentimentr* (0,27). A imagem designativa demonstra a importância do serviço nesta categoria com as palavras *friend(ly)* (4,934) e *staff* (4301) a serem mencionadas no *top10* e ainda o destaque para a palavra *wine* (4572) que parece ter também uma elevada importância nesta natureza. A gastronomia portuguesa tem um destaque na natureza (12%) e, olhando para a região do tipo de cozinha, sem surpresa, a região *european* (21%) e *mediterranean* (19,5%) são as mais representadas. No que diz respeito à culinária, parece haver uma clara predominância no número de *reviews* de restaurantes com opções de culinária de dieta e saudável, com os restaurantes com opções vegetarianas (5,3%), comida saudável (1,9%), opções *vegans* (1,7%) e *glúten free* (1,5%) a ocupar 4 dos 6 lugares cimeiros no *top6* de culinárias. A culinária *seafood* é a mais mencionada em *reviews*.

No que diz respeito às centralidades, a metrópole, Lisboa, detém a melhor marca de imagem prescritiva (21,24% de *reviews* com recomendações positivas), a melhor média ponderada de classificação (9,02), e o melhor valor de *sentimentr* (0,234). É seguro por isso dizer que a região é largamente dependente da sua metrópole na sua eWOM e, por conseguinte, constitui o elemento âncora da imagem turística da região. As centralidades apresentam níveis de recomendação muito semelhantes, que oscilam entre 3% (Arco do Tejo) e 4% (Lisboa) das frases. Em relação às *reviews*, Lisboa apresenta 21,24% das *reviews* com recomendações positivas e apenas 1,39% com *reviews* negativas. No que diz respeito à análise do sentimento, Lisboa apresenta o valor mais elevado (0,234) com o Arco do Tejo a apresentar o lado oposto (0,162). De realçar ainda que o Arco

do Tejo apresenta os únicos dois elementos negativos da amostra com as atividades e os alojamentos nesta centralidade a ter uma análise *sentimentr* de -0,07 e -0,20 respetivamente.

A amostra não parece apresentar uma sazonalidade muito elevada, visto que apresenta valores acima da média de *reviews* por mês (4250) entre abril e outubro, inclusive. Parece haver por isso um equilíbrio na criação de *Tourism Generated Content* (TGC) ao longo do ano com apenas o primeiro trimestre do ano a registar um valor inferior a 10,000 *reviews*.

O conhecimento desta imagem é essencial uma vez que permite combater a subjetividade da mesma dado o volume, variedade, velocidade, veracidade e valor que caracterizam o *big data*. O estudo do conteúdo criado por visitantes permite adaptações em tempo real por parte dos organismos regentes do turismo, gestores hoteleiros, turísticos, de museus ou de empreendimentos que dependam ou contem com a atividade, bem como o alargamento do conhecimento do seu mercado, da sua concorrência, e da sua própria atividade e evolução.

A imagem projetada por parte dos organismos turísticos, ou fontes inferidas, deve ter em consideração o presente estudo. Um futuro estudo por parte do Turismo de Portugal deve procurar cruzar a imagem conseguida em 2019 com o plano estratégico proposto para 2015 a 2019 (Turismo de Lisboa, n.d.). A organização pode medir a eficiência do seu plano e cruzar as medidas de desenvolvimento propostas por centralidade e geral com a eWOM na mesma. Também com o conhecimento criado pela exploração da imagem, um futuro estudo deve tentar perceber o que se passa com o *top5* de atividades mais mencionadas uma vez que as mesmas apresentam *scores* médios muito abaixo da média da amostra seguindo a metodologia proposta em Marine-Roig & Clavé (2015).

Como sugestão para futuros investigadores parte ainda o desafio de relacionar os resultados aqui projetados com a realidade estatística da região em 2019. Um estudo com o intuito de cruzar dados reais com a imagem aqui produzida permitirá perceber o nível de relação entre as variáveis e explorar mais a fundo o grau de dependência entre a eWOM e a chegada/dormida de visitantes ou receitas. Parece ainda evidente que, na evolução do turismo nacional para uma *smart destination*, devem ser realizados estudos da imagem turística nas restantes regiões do país. O conhecimento criado por estes estudos torna-se indispensável nas regiões com maior preponderância turística como o Algarve e a Madeira, sobretudo na resposta à pandemia Covid-19.

Embora seja expectável haver uma maior concentração de OTRs na metrópole, os *destination marketing organizations* (DMOs) podem utilizar a imagem da região como uma fonte para dinamizar a oferta fora da metrópole. Por exemplo compreender os recursos mais apreciados pelos visitantes fora de Lisboa e promover os mesmos, ou explorar os recursos com maior classificação e com margem de crescimento sustentável, que provavelmente estão a ser subvalorizados pelos operadores turísticos. Combinar elementos de diferentes naturezas valorizadas para aumentar a diversidade, como por exemplo o aproveitamento do tipo de atividade mais utilizado – *tours* – para promover a natureza com maior preponderância fora da metrópole – restaurantes – e criar, por exemplo, *tours* gastronómicos às centralidades menos visitadas.

As várias categorias que cada recurso apresenta no *TripAdvisor* (ex. Restaurante: Shish – Gastronomia: Libanês, Fast food, Mediterrâneo, Médio Oriente) podem constituir um enclave no valor da análise uma vez que existe demasiada diversidade e subjetividade na categoria a que o recurso pertence, podendo por isso produzir resultados que não refletem necessariamente a amostra (Marine-Roig, 2019).

Outra das limitações do estudo é a margem de erro que os pacotes utilizados para a análise em R podem ter. A medição da polaridade de uma frase bem como da intenção de recomendar é uma das tarefas mais difíceis na análise de dados. Sobretudo no que diz respeito à intensidade do comentário ou aplicação de expressões de carácter irónico, sarcástico ou retórica.

Por fim, é também necessário mencionar a volatilidade que a pandemia covid-19 trouxe a estudos turísticos. A indústria encontra-se em mudança e reinvenção e, embora o conhecimento da imagem turística de um destino seja considerado um dos elementos chave na resposta à pandemia e relançamento da indústria (Sigala, 2020), a verdade é que deve haver uma perceção geral de que as expectativas turísticas devem ser reajustadas.

Referências

- Agustí, D. P. I. (2018). Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides. *Annals of Tourism Research*, 73, 103 - 115. <https://doi.org/10.1016/j.annals.2018.09.001>
- Apostolopoulou, A., & Papadimitriou, D. (2015). The role of destination personality in predicting tourist behaviour: implications for branding midsized urban destinations. *Current Issues in Tourism*, 18 (12), p. 1132–1151. <https://doi.org/10.1080/13683500.2013.878319>
- Baka, V. (2016). The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector. *Tourism Management*, 53, 148 - 162. <https://doi.org/10.1016/j.tourman.2015.09.004>
- Baloglu, S., & McCleary, K. W. (1999). A Model of Destination Image Formation. *Annals of Tourism Research*, 26 (4), 868 - 897. [https://doi.org/10.1016/S0160-7383\(99\)00030-4](https://doi.org/10.1016/S0160-7383(99)00030-4)
- Buhalis, D., & Amaranggana, A. (2014). Smart Tourism Destinations. *Information and Communication Technologies in Tourism*, 553-564. https://doi.org/10.1007/978-3-319-03973-2_40
- Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment*, (728), 1 - 7.
- Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, 28, p. 118–129. <https://doi.org/10.1016/j.tourman.2006.03.002>
- Chon, K.-S. (1990). The Role of Destination Image in Tourism: A Review and Discussion. *The Tourist Review*, 2, 2 - 9. <https://doi.org/10.1108/eb058040>

- Crompton, J. (1979). An Assessment of the Image of Mexico as a Vacation Destination and the Influence of Geographical Location Upon That Image. *Journal of Travel Research*, 17, 17 - 23. <https://doi.org/10.1177/004728757901700404>
- Deng, N., Liu, J., Dai, Y., & Li, H. (2019). Different cultures, different photos: A comparison of Shanghai's pictorial destination image between East and West. *Tourism Management Perspectives*, 30, 182 - 192. <https://doi.org/10.1016/j.tmp.2019.02.016>
- Ekinçi, Y., & Hosany, S. (2006). Destination Personality: An Application of Brand Personality to Tourism Destinations. *Journal of Travel Research*, 45, 127 - 139. <https://doi.org/10.1177/0047287506291603>
- Gkritzali, A., Gritzalis, D., & Stavrou, V. (2018). Is Xenios Zeus Still Alive? Destination Image of Athens in the Years of Recession. *Journal of Travel Research*, 57 (4), 540 - 554. <https://doi.org/10.1177/0047287517705225>
- Gretzel, U., & Yoo, K. H. (2008). *Use and Impact of Online Travel Reviews*. Proceedings of the International Conference in Innsbruck, Austria, 2008.
- Instituto Nacional de Estatística (2018). *Estatísticas do Turismo - 2018*. [Online] Disponível em: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=358629548&PUBLICACOESStema=55581&PUBLICACOESmodo=2 [Acedido em Outubro 2019].
- Jabreel, M., Moreno, A., & Huertas, A. (2017). Semantic comparison of the emotional values communicated by destinations and tourists on social media. *Journal of Destination Marketing & Management*, 6, 170 - 183. <https://doi.org/10.1016/j.jdmm.2016.03.004>
- Kaosiri, Y. N. et al. (2019). User-Generated Content Sources in Social Media: A New Approach to Explore Tourist Satisfaction. *Journal of Travel Research*, 58 (2), 25-265.
- Kladou, S., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of *TripAdvisor*. *Journal of Destination Marketing & Management*, 4, 187 - 193. <https://doi.org/10.1016/j.jdmm.2015.04.003>
- Lalicic, L., Huertas, A., Moreno, A., & Jabreel, M. (2019). Which emotional brand values do my followers want to hear about? An investigation of popular European tourist destinations. *Information Technology & Tourism*, 21, 63 - 81. <https://doi.org/10.1007/s40558-018-0134-7>
- Marine-Roig, E. (2010). *Los "Travel Blogs" como objeto de estudio de la imagen percibida de un destino*. VIII Congreso "Turismo y Tecnologías de la Información y las Comunicaciones, Málaga, Spain.
- Marine-Roig, E. (2015). Identity and authenticity in destination image construction. *Anatolia – An International Journal of Tourism and Hospitality Research*, 26 (4), 574 - 587.

- Marine-Roig, E. (2017). Measuring Destination Image through Travel Reviews in Search Engines. *Sustainability*, 9 (8), 1425 - 1443. <https://doi.org/10.3390/su9081425>
- Marine-Roig, E. (2019). Destination Image Analytics Through Traveller-Generated Content. *Sustainability*, 11 (12), 1 - 23. <https://doi.org/10.3390/su10023392>
- Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4 (3), 162-172. <https://doi.org/10.1016/j.jdmm.2015.06.004>
- Marine-Roig, E., & Ferrer-Rosell, B. (2018). Measuring the gap between projected and perceived destination images of Catalonia using compositional analysis. *Tourism Management* 68, 236-249. <https://doi.org/10.1016/j.tourman.2018.03.020>
- Martín, H. S., & Bosque, I. A. R. d. (2008). Exploring the cognitive–affective nature of destination image and the role of psychological factors in its formation. *Tourism Management* 29, 263-277.
- Nezakati, H., Amidi, A., Jusoh, Y. Y., Moghadas, S., Aziz, Y. A., & Sohrabinezhadtalemi, R. (2015). Review of Social Media Potential on Knowledge Sharing and Collaboration in Tourism Industry. *Procedia—Social and Behavioral Sciences*, 172, 120-25.
- Nunes, M. G. B. (2017). *A concentração da procura turística internacional em Portugal*. Estoril: Escola Superior de Hotelaria e Turismo do Estoril.
- Pocock, D. C. D. (1978). *Images of the urban environment: Images of the Urban Environment*. London: Macmillan Press Ltd.
- Prayag, G., Hosany, S., Muskat, B., & Del Chiappa, G. (2015). Understanding the relationships between tourists' emotional experiences, perceived overall image, satisfaction, and intention to recommend. *Journal Travel Research*, 56, 41–54. <https://doi.org/10.1177/0047287515620567>
- Shankar, R. S. (2018). Destination Personality and Destination Image: A Literature Review. *The IUP Journal of Brand Management*, 15 (4), 47 - 60.
- Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research* (117), 312 - 321.
- Silva, F. B. E. et al. (2018). Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, 101-115.
- Stepchenkova, S. (2012). Content Analysis. Em: L. Dwyer, A. Gill & N. Seetaram (eds.) *Handbook of Research Methods in Tourism: Quantitative and Qualitative Approaches* (pp. 443 - 458). Edward Elgar Publishing.
- Tamajón, L. G. & Valiente, G. C. (2017). Barcelona seen through the eyes of *TripAdvisor*: actors, typologies and components of destination image in social media platforms. *Current Issues in Tourism*, 20 (1), p. 33 - 37.
- TripAdvisor* LLC (2017). *TripAdvisor*. <https://TripAdvisor.mediaroom.com/pt-about-us> [Acedido em 27 02 2020].

- Turismo de Lisboa (s.d). *Plano Estratégico para o Turismo da Região de Lisboa 2015-19*. www.visitlisboa.com
- Welbers, K., Atteveldt, W. V., & Benoit, K. (2017). Text analysis in R. *Communication Methods and Measures*, 11 (4), 245- 265.
- World Travel & Tourism Council (2019). *Travel & Tourism Economic Impact 2019 World*. <https://www.wttc.org/economic-impact/country-analysis/region-data/> [Acedido em 3 Outubro 2019].
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management* 31, 179 - 188. <https://doi.org/10.1016/j.tourman.2009.02.016>
- Yoo, K.-H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior* 27, 609 - 621. <https://doi.org/10.1016/j.chb.2010.05.002>
- Yoo, K.-H., Lee, Y., Gretzel, U., & Fesenmaier, D. R. (2009). Trust in Travel-Related Consumer Generated Media. *Information and communication technologies in tourism 2009*, 49-59. https://doi.org/10.1007/978-3-211-93971-0_5

Apêndice:

Nome do Pacote	Criador	Descrição
<i>rvest</i>	Hadley Wickham	Utiliza pacotes “xml2” e “httr” para efectuar download e manipulação de páginas HTML e XML.
<i>Rcrawler</i>	Salim Khalil	Efectua simultaneamente <i>web crawling</i> ³ e <i>web scrapping</i> de páginas web
<i>purrr</i>	Lionel Henry, Hadley Wickham	Cria ferramentas que permitem a manipulação de vectores e funções
<i>magrittr</i>	Stefan Milton Bache, Hadley Wickham	Proporciona a interligação de comandos com ordens consequentes
<i>tidyverse</i>	Hadley Wickham	Ferramenta importante no tratamento de dados
<i>lubridate</i>	Vitalie Spinu, Garrett Grolemond	Tratamento de dados que incluem datas
<i>splitstackshape</i>	Ananda Mahto	Permite o desemparcelamento de dados textuais em células
<i>dplyr</i>	Hadley Wickham, Romain François, Lionel Henry, Kirill Müller	Permite a criação de tabelas a partir de objetos, vetores e funções.
<i>httr</i>	Hadley Wickham	Proporciona ferramentas úteis para lidar com conteúdo HTTP
<i>sentimentr</i>	Tyler Rinker	Cálculo da polaridade sentimental de um texto
<i>quanteda</i>	Kenneth Benoit	Ferramenta rápida e flexível que permite a análise quantitativa de texto em R

Tabela A 1 – Pacotes utilizados em R

³ *Web crawling* – corresponde a um processo automatizado que procura e indexa páginas consoante uma orientação anterior.

Centralidades		Localizações <i>TripAdvisor</i>								
A. Lisboa 40631	Lisboa	38408	Carvoeira	75	Algés	32	Porto salvo	10	Sobreda	2
	Mafra	777	Paço de arcos	62	Carnaxide	23	Sacavem	8	Porto brandao	1
	Costa da caparica	356	Prior velho	52	Linda-a-velha	15	Caparica	6		
	Almada	307	Odivelas	45	Loures	13	Santa iria azoria	5		
	Oeiras	176	Charneca caparica	42	Barcarena	12	Corroios	4		
B. Sintra 3917	Belem	154	Amadora	34	Moscavide	10	Fernao ferro	2		
	Sintra municipality	3766	Praia das maçãs	26						
C. Cascais 5071	Linhó	124	Rio de mouro	1						
	Cascais	4711	Parede	89	Sao domingos de rana	1				
D. Arco do Tejo 167	Carcavelos	262	Alcabideche	8						
	Alcochete	44	Seixal	27	Vila f. Xira	16	Alverca do ribatejo	7	Alhandra	1
	Montijo	35	Barreiro	21	Amora	12	Moita	4		
E. Arrábida 1214	Sesimbra	539	Palmela	82	Aldeia do meco	15	Portinho da arrábida	13	Vila nogueira az.	4
	Setubal	481	Azeitão	57	Qta. Do conde	13	Pinhal novo	8	Qta. Do anjo	2

Tabela A 2 – Distribuição da amostra por localização

	Palavras únicas: 10 453	Contagem Total: 211 991	%		Palavras únicas: 10 453	Contagem Total: 211 991	%
1	great	7844	3,70016%	11	tour	2411	1,13731%
2	hotel	4336	2,04537%	12	servic	2403	1,13354%
3	food	3910	1,84442%	13	nice	2176	1,02646%
4	lisbon	3780	1,78309%	14	stay	1986	0,93683%
5	good	3189	1,50431%	15	love	1827	0,86183%
6	excel	2854	1,34628%	16	experi	1766	0,83305%
7	locat	2806	1,32364%	17	wonder	1516	0,71512%

	Palavras únicas: 10 453	Contagem Total: 211 991	%		Palavras únicas: 10 453	Contagem Total: 211 991	%
8	amaz	2571	1,21279%	18	beauti	1412	0,66607%
9	best	2550	1,20288%	19	perfect	1318	0,62172%
10	place	2527	1,19203%	20	staff	1226	0,57833%

Tabela A 3 – Top 20 Keywords dos títulos das OTRs

Atividades			Alojamento		Restaurantes	
Palavras únicas: 23 366	Total palavras: 1 149 524		Palavras únicas: 29 755	Total palavras: 1 791 865	Palavras únicas: 22 817	Total palavras: 1 040 099
Keyword	Contagem		Keyword	Contagem	Keyword	Contagem
1	tour	9630	hotel	22915	food	11574
2	lisbon	7318	room	20953	good	7791
3	great	6389	stay	13433	restaur	7727
4	us	5956	staff	11805	great	6789
5	time	5260	breakfast	10163	place	6470
6	visit	4821	great	9145	servic	6253
7	place	4650	locat	8870	recommend	5240
8	day	4644	good	8580	friend	4934
9	guid	4456	lisbon	7346	wine	4572
10	recommend	4345	nice	6525	staff	4301

Tabela A 4 – Top 10 Keywords dos comentários das OTRs por natureza

Tipos de actividades			
Private Tours	2810	Historical and Heritage Tours	1076
City Tours	2080	Points of Interest and Landmarks	1074
Day Trips	1634	Walking Tours	1059
Points of Interest and Landmarks	1605	Castles	1051
Sightseeing Tours	1597	Food Tours	1042
City Tours	1449	Sightseeing Tours	803
Historic Sites	1388	Cultural Tours	738
Cultural Tours	1274	Wine Tours and Tastings	730
Architectural Buildings	1177	Sightseeing Tours	654
Cooking Classes	1118	Neighborhoods	653

Tabela A 5 – Top 20 Tipos de Atividades mais comuns nas OTRs

Restaurantes				
Restaurante Figus	Lisbon	363	Restaurante Sacramento	Lisbon 110
Ola Nepal	Lisbon	277	Incomum by Luis Santos	Sintra 107
Zenith Brunch and Cocktails - Lisboa	Lisbon	276	Come Prima Restaurante Italiano	Lisbon 99
A Nossa Casa	Lisbon	166	Villa 6	Sintra 97
Byf Steakhouse	Lisbon	157	Allchiado	Lisbon 92
Royal Spice Restaurante	Lisbon	152	Augusto Lisboa	Lisbon 89
Floresta Das Escadinhas	Lisbon	141	O Tainadas Restaurante and Bar	Lisbon 89
Jam Club	Lisbon	140	M'arrecreeo Pizzeria	Lisbon 88
Sanskar Nepal	Lisbon	128	Populi	Lisbon 88
Meson Andaluz	Lisbon	114	Mar Ao Carmo	Lisbon 86

Tabela A 6 – Top 20 Restaurantes mais comentados na amostra

Alojamentos			
Corpo Santo Lisbon Historical Hotel	486	Pestana Palace Lisboa	162
Jupiter Lisboa Hotel	421	Santiago de Alfama - Boutique Hotel	152
Hotel Avenida Palace	305	memmo Principe Real	151
Hotel Real Palacio	280	Hotel Santa Justa	145
Tivoli Avenida Liberdade Lisboa	232	Lisbon Marriott Hotel	145
Hotel da Baixa	214	EPIC SANA Lisboa Hotel	144
Corinthia Lisbon	205	Pousada de Lisboa, Praca do Comercio	134
DoubleTree by Hilton Lisbon - Fontana Park	193	Brown's Central Hotel	132
Radisson Blu Hotel, Lisbon	189	Sofitel Lisbon Liberdade	131
Britania Hotel, a Lisbon Heritage Collection	167	1908 Lisboa Hotel	129

*Nota: Todos os alojamentos são localizados em Lisboa

Tabela A 7 – Top 20 Alojamentos mais comentados na amostra

Top 5 por categoria		Score + %	Score - %	Média ponderada
Restaurantes	Restaurante Figus	88,98%	5,51%	8,99
	Ola Nepal	94,95%	4,33%	9,36
	A Nossa Casa	98,80%	0,00%	9,71
	Byf Steakhouse	93,63%	2,55%	9,38
	Royal Spice Restaurante	97,37%	1,97%	9,59

	Top 5 por categoria	Score + %	Score - %	Média ponderada
Atividades	Time Out Market Lisboa	81,73%	4,01%	8,17
	Castelo de S. Jorge	86,00%	3,38%	8,30
	Torre de Belem	82,79%	3,61%	8,09
	Tram 28	68,35%	11,42%	7,36
	Mosteiro dos Jeronimos	88,55%	4,02%	8,54
Alojamentos	Corpo Santo Lisbon Historical Hotel	98,56%	0,41%	9,81
	Jupiter Lisboa Hotel	92,16%	2,61%	9,00
	Hotel Avenida Palace	97,38%	1,64%	9,60
	Hotel Real Palacio	80,36%	5,00%	8,13
	Tivoli Avenida Liberdade Lisboa	91,38%	3,45%	8,94

Tabela A 8 – Score e Média Ponderada por Top5 das Naturezas