

Análise de preços e tendências das hospedagens do Airbnb utilizando a ferramenta Apache Spark

Audrey Emmely Rodrigues Vasconcelos

Instituto de Computação

Universidade Federal de Alagoas

aerv@ic.ufal.br

Resumo—Neste estudo, realizamos uma análise exploratória distribuída de dados do Airbnb, utilizando a plataforma Apache Spark. Focamos na avaliação de características dos anúncios, como tipos de quarto, pontuações de revisão e distribuição geográfica. A análise dos tipos de quarto revelou a variedade e prevalência de diferentes acomodações. A avaliação das pontuações de revisão, por tipo de propriedade, nos permitiu entender o desempenho relativo das diferentes categorias de anúncios. A contagem de anúncios e bairros por cidade nos forneceu uma visão da atividade do Airbnb em diferentes regiões. Além disso, analisamos as tendências de revisão ao longo do tempo, bem como efetuamos análises específicas para cidades como Paris e Londres. As informações obtidas por meio dessa análise são valiosas para compreender tendências e padrões nos anúncios do Airbnb, podendo orientar hóspedes e anfitriões, além de informar estratégias futuras da plataforma. Com base nos gráficos analisados, podemos concluir que os viajantes têm preferência por reservar “Entire home/apt” (casa/apartamento inteiro) como tipo de quarto mais escolhido. As propriedades com maior pontuação média são “Heritage hotel”, “Entire floor” e “Casa particular”, indicando uma busca por acomodações de alta qualidade e personalizadas. Paris, Londres e Berlim são as cidades com o maior número de anúncios, refletindo sua popularidade como destinos turísticos. O ano de 2016 registrou o maior pico no número de avaliações, sugerindo um período movimentado. Além disso, os histogramas de preços revelam as faixas predominantes e picos específicos em cada cidade. Essas conclusões combinadas oferecem insights sobre as preferências dos usuários e as tendências do mercado de hospedagem nessas cidades.

Palavras-chave—análise de dados, mineração de dados, processamento de dados, apache spark, big data

I. INTRODUÇÃO

Nos últimos anos, o setor de hospedagem tem passado por transformações significativas impulsionadas pelo avanço da tecnologia e das plataformas digitais. Entre elas, destaca-se o aplicativo Airbnb, que revolucionou a forma como as pessoas procuram e oferecem acomodações ao redor do mundo. Com sua vasta base de dados e abrangência global, o Airbnb se tornou uma fonte valiosa de informações para análises de preços e tendências no setor hoteleiro [1].

Neste artigo, iremos explorar o uso do Apache Spark como uma ferramenta poderosa para realizar análises de preços e identificar tendências nas hospedagens do aplicativo Airbnb. O Apache Spark é um framework de processamento de dados distribuído e de alto desempenho, projetado para lidar com grandes volumes de dados em tempo real [2]. Uma das principais vantagens do Apache Spark é sua arquitetura

distribuída, que permite processar e analisar dados em clusters de computadores. Isso significa que é possível executar cálculos complexos e análises em paralelo, aproveitando o poder de processamento de vários nós em um cluster. Com isso, é possível lidar com grandes volumes de dados de forma eficiente e rápida. Ao usar o Apache Spark na análise de preços e tendências do Airbnb, é possível extrair insights valiosos sobre a dinâmica dos preços das hospedagens. O Spark oferece recursos para realizar análises estatísticas, como médias, desvios padrão e distribuições de preços, além de permitir a identificação de padrões e tendências ao longo do tempo. Com base nesses insights, os proprietários de imóveis podem ajustar suas estratégias de preços para maximizar o retorno sobre o investimento..

Com o objetivo de obter insights valiosos sobre o mercado de hospedagens do Airbnb, iremos aproveitar as capacidades do Apache Spark para processar, limpar e analisar um conjunto de dados extenso e diversificado, que abrange informações sobre preços, localização, avaliações e outros atributos relevantes. Através dessa análise, pretende-se fornecer insights acionáveis sobre as tendências de preços, fatores que influenciam a demanda e outras informações relevantes para auxiliar na tomada de decisões estratégicas. Além disso, o presente estudo também tem o potencial de contribuir para a compreensão mais ampla do mercado de hospedagens e das dinâmicas envolvidas, bem como destacar a importância do Apache Spark como uma ferramenta poderosa para análise de dados em tempo real.

II. FUNDAMENTAÇÃO TEÓRICA

A. Apache Spark

O Apache Spark é um poderoso framework de processamento de dados distribuído e escalável, que tem sido amplamente utilizado na indústria para lidar com grandes volumes de informações e realizar análises em tempo real. Sua arquitetura permite executar tarefas complexas de processamento em memória, o que contribui significativamente para a eficiência e capacidade de processamento do Spark [2]. A capacidade de acelerar o tempo de processamento possibilita a realização de análises interativas e iterativas em tempo real. Isso significa que os usuários podem explorar seus dados de maneira mais dinâmica, executando consultas e obtenção de resultados quase que instantaneamente. Essa característica é particularmente útil para empresas que necessitam tomar decisões rápidas com

base em informações atualizadas. Outro aspecto importante do Apache Spark é o módulo Spark SQL, que torna o Spark mais acessível a uma ampla gama de usuários. Com o Spark SQL, é possível executar consultas SQL tradicionais em dados distribuídos, facilitando a integração do Spark em ecossistemas de dados existentes. Isso significa que os profissionais que já estão familiarizados com SQL podem aproveitar seus conhecimentos para trabalhar com o Spark, eliminando a necessidade de aprender uma nova linguagem ou abordagem de programação [3]. O Apache Spark também se destaca no processamento de dados de streaming, ele fornece tolerância a falhas e um mecanismo de processamento eficiente para fluxos contínuos de dados. Essa capacidade é extremamente valiosa em cenários em que a análise de dados em tempo real é essencial, como em sistemas de monitoramento de sensores, análise de dados de mídias sociais em tempo real ou detecção de anomalias em transações financeiras [4].

O Apache Spark é um framework altamente eficiente e versátil, com suporte a consultas SQL, processamento em memória e processamento de dados de streaming. Sua capacidade de lidar com grandes volumes de informações, processá-las de forma distribuída e executar análises em tempo real o torna uma ferramenta poderosa para análises de big data em escala. Com suas características avançadas, o Spark tem se destacado como uma solução robusta e flexível para as demandas cada vez maiores de processamento e análise de dados nas empresas modernas. Sua popularidade crescente e adoção generalizada são testemunho do seu sucesso no campo do processamento de dados distribuído.

B. Análise de preços no Airbnb

A análise de preços no Airbnb é um tema de grande interesse e tem sido objeto de estudos que fornecem insights valiosos sobre os fatores que influenciam os preços no Airbnb e as estratégias utilizadas pelos anfitriões. Um estudo relevante examina os determinantes de preços no Airbnb, explorando fatores como localização, tipo de propriedade, disponibilidade, número de avaliações e características do anfitrião [1]. Esses fatores desempenham um papel fundamental na definição dos preços das acomodações e ajudam os anfitriões a tomar decisões informadas sobre como precificar suas propriedades.

Outro estudo investiga os estímulos econômicos por trás dos preços de aluguel de acomodações no Airbnb, analisando o impacto de fatores como demanda sazonal, disponibilidade de oferta, concorrência de hotéis e características dos anfitriões [5]. A demanda sazonal, por exemplo, pode levar a variações significativas nos preços das acomodações, com picos durante períodos de alta procura, como feriados ou eventos especiais. Compreender esses estímulos econômicos permite que os anfitriões ajustem seus preços de acordo com as flutuações da demanda e otimizem seus ganhos.

Além disso, há pesquisas que se concentram na análise dos padrões de preços dinâmicos no Airbnb, identificando os fatores que levam a mudanças nos preços das acomodações e como os anfitriões ajustam seus preços com base na demanda e na oferta [6]. Esses estudos têm como objetivo compreen-

der como os anfitriões utilizam estratégias de precificação dinâmica para otimizar a ocupação e maximizar seus lucros. A capacidade de ajustar os preços de acordo com a demanda e a concorrência é uma vantagem significativa para os anfitriões no mercado do Airbnb.

Esses estudos fornecem uma visão abrangente sobre a análise de preços no Airbnb, abordando fatores como localização, características da propriedade, demanda sazonal, concorrência de hotéis e estratégias de precificação dinâmica. Com base nesses conhecimentos, tanto os anfitriões quanto os hóspedes podem se beneficiar na busca por melhores ofertas. Os anfitriões podem tomar decisões mais informadas sobre a precificação de suas propriedades, levando em consideração os fatores que influenciam os preços. Por outro lado, os hóspedes podem usar essas informações para identificar os melhores momentos para reservar acomodações com preços mais favoráveis e entender como os preços podem variar de acordo com a demanda e a oferta.

A análise de preços no Airbnb é um campo de pesquisa importante que oferece insights valiosos para anfitriões e hóspedes. Compreender os fatores que influenciam os preços, como localização, características da propriedade, demanda sazonal e concorrência de hotéis, permite que os anfitriões tomem decisões estratégicas sobre a precificação de suas acomodações. Da mesma forma, os hóspedes podem aproveitar essas informações para obter as melhores.

C. Análise de tendências no Airbnb

A análise de tendências no Airbnb é uma área de pesquisa importante para entender os padrões de demanda, preferências dos hóspedes e mudanças no mercado de hospedagem [7].

Um estudo relevante utiliza análise de séries temporais para prever os preços de aluguel e entender as dinâmicas de demanda no Airbnb, considerando fatores sazonais, econômicos e eventos especiais [8].

Esses estudos fornecem uma visão abrangente sobre a análise de tendências no Airbnb, abordando aspectos como ocupação, preços, preferências dos hóspedes e expansão geográfica [7], [8]. Essas análises são fundamentais para auxiliar anfitriões, empresas e pesquisadores a entenderem as mudanças no mercado de hospedagem compartilhada e se adaptarem às demandas em constante evolução.

Os resultados dessas análises são especialmente relevantes para os anfitriões do Airbnb, que podem utilizar as informações obtidas para ajustar suas estratégias de precificação, identificar períodos de maior demanda e otimizar suas ofertas. Além disso, empresas do setor hoteleiro e pesquisadores também se beneficiam desses estudos, pois podem acompanhar as tendências e tomar decisões baseadas em dados sólidos.

À medida que o mercado de hospedagem compartilhada continua a evoluir e novos concorrentes surgem, a análise de tendências se torna ainda mais crucial para se manter competitivo. Compreender como as preferências dos hóspedes estão mudando, as flutuações sazonais e os impactos econômicos é

essencial para adaptar-se às demandas em constante evolução e garantir o sucesso nesse setor em crescimento.

Em suma, a análise de tendências no Airbnb desempenha um papel fundamental na compreensão do mercado de hospedagem compartilhada. Por meio de métodos analíticos, é possível obter insights valiosos que ajudam a tomar decisões informadas e se adaptar às mudanças do setor.

D. Limitações

A ferramenta Apache Spark apresenta algumas limitações ao realizar a análise de preços e tendências das hospedagens do aplicativo Airbnb. Primeiramente, o Spark é uma plataforma de processamento distribuído projetada para lidar com grandes volumes de dados em escala. No entanto, ao trabalhar com dados provenientes do Airbnb, é importante considerar que a disponibilidade desses dados pode ser limitada. O acesso às informações de preços e tendências das hospedagens pode estar sujeito a restrições impostas pelo Airbnb, como limitações de acesso à API ou restrições de uso dos dados. Além disso, o Spark, embora seja altamente escalável, pode enfrentar desafios de desempenho ao lidar com tarefas que envolvem processamento em tempo real [2]. A análise de preços e tendências do Airbnb requer a análise contínua de dados em constante evolução, o que pode exigir recursos computacionais significativos para garantir respostas rápidas. Por fim, o Spark é uma ferramenta poderosa para processamento de dados, mas sua eficácia depende da qualidade e estrutura dos dados disponíveis. Problemas como dados ausentes, inconsistências ou ruído nos dados podem afetar negativamente os resultados da análise. Portanto, é necessário realizar uma limpeza e preparação adequada dos dados antes de executar as análises no Spark.

III. METODOLOGIA

A primeira etapa envolveu a configuração de um ambiente no Google Colab para utilizar o Apache Spark. Esses passos são essenciais para a instalação do Java e do Apache Spark, bem como para a configuração das variáveis de ambiente necessárias.

Após a configuração do ambiente, algumas bibliotecas foram importadas para facilitar o processamento e a visualização dos dados. A biblioteca *findspark* foi usada para iniciar o Apache Spark, enquanto *pyspark.sql* foi usada para trabalhar com DataFrames Spark. Além disso, a biblioteca *os* foi usada para manipular variáveis de ambiente e *matplotlib.pyplot* foi usada para visualização dos dados.

A base de dados "airbnb-listings" é um conjunto de ferramentas e dados independentes e não comerciais que permitem explorar como o Airbnb está sendo realmente usado nas cidades ao redor do mundo. Esta base é fornecida pelo Inside Airbnb, que analisa informações publicamente disponíveis sobre os anúncios do Airbnb em diferentes cidades, proporcionando filtros e métricas-chave para entender como o Airbnb está sendo usado em concorrência com o mercado de habitação residencial [9].

Essa base pode ser utilizada para responder perguntas fundamentais sobre o Airbnb em qualquer bairro ou em toda a cidade como um todo. As perguntas podem abranger desde o número de anúncios em um determinado bairro, quantas casas e apartamentos estão sendo alugados frequentemente para turistas em vez de residentes de longo prazo, quanto os anfitriões estão ganhando com o aluguel para turistas, até quais anfitriões estão administrando um negócio com vários anúncios e onde estão localizados.

Também é possível usá-la para responder perguntas relacionadas a legislações ou regulamentos de cidades ou estados que tratam de habitação residencial, aluguéis de curto prazo ou férias e zoneamento, como quantas noites uma habitação é alugada por ano, mínimo de noites de estadia, se o anfitrião está presente, quantos quartos estão sendo alugados em um prédio e o número de ocupantes permitidos em um aluguel.

Os dados por trás do site Inside Airbnb são provenientes de informações publicamente disponíveis no site do Airbnb. Estes dados foram analisados, limpos e agregados onde era conveniente para facilitar a discussão pública [9].

A base de dados é disponibilizada sob uma licença Creative Commons CC0 1.0 Universal (CC0 1.0) "Public Domain Dedication", permitindo análises adicionais ou produção de visualizações alternativas dos dados [9].

A última data de processamento dos dados e o território alvo não são explicitamente mencionados na página de descrição da base de dados [9].

Uma sessão do Spark foi iniciada com o nome "AirbnbProcessing" para facilitar o processamento distribuído dos dados. Os dados foram carregados de um arquivo CSV em um DataFrame Spark. Durante o pré-processamento, os nomes das colunas foram convertidos para letras minúsculas e substituídos os espaços por sublinhados para facilitar a manipulação dos dados. Além disso, o DataFrame foi filtrado para incluir apenas as colunas de interesse e armazenado em cache para melhorar o desempenho. A coluna 'price' foi limpa, removendo o símbolo do dólar e as vírgulas, e convertida para um tipo numérico.

A análise dos dados do Airbnb foi realizada através de uma combinação de consultas SQL Spark e operações de DataFrame, abrangendo várias dimensões do conjunto de dados, como por exemplo:

- 1) **Contagem de Tipos de Quarto:** A primeira análise focou na distribuição dos tipos de quartos. Os anúncios do Airbnb incluem várias opções, como quartos inteiros, quartos privados e quartos compartilhados. A contagem destes diferentes tipos de quartos fornece um indicador da variedade de acomodações disponíveis e da prevalência de cada tipo.
- 2) **Média de Pontuações de Revisão por Tipo de Propriedade:** A pontuação de revisão é um indicador chave da satisfação dos hóspedes com os anúncios do Airbnb. A média dessas pontuações, agrupadas por tipo de propriedade, fornece insights sobre o desempenho relativo de diferentes tipos de propriedades. Por exemplo, pode-se descobrir se os hóspedes tendem a dar pontuações

mais altas para casas inteiras em comparação com quartos privados ou compartilhados.

- 3) **Análise de Tendências de Reviews ao Longo do Tempo:** Esta análise examina as mudanças nas pontuações de *reviews* ao longo do tempo. As tendências podem indicar melhorias ou declínios na qualidade dos anúncios do Airbnb, ou mudanças nas expectativas e avaliações dos hóspedes.
- 4) **Análises Específicas da Cidade:** Além das análises gerais, foram realizadas análises específicas para cidades particulares, como Paris e Londres. Estas análises permitem uma compreensão mais detalhada das características e tendências do Airbnb nessas cidades.

Os resultados das consultas foram convertidos para DataFrames Pandas para facilitar a visualização dos dados. Diversas técnicas de visualização foram empregadas com o uso da biblioteca matplotlib, cada uma escolhida por suas capacidades específicas de representar certos aspectos dos dados.

Os gráficos de barras foram usados para representar a contagem de categorias distintas em uma variável categórica. Eles fornecem uma representação visual clara da diferença de magnitude entre as categorias. No contexto deste estudo, gráficos de barras foram úteis para representar dados como a contagem de tipos de quartos e o número de anúncios em diferentes bairros.

Os histogramas foram utilizados para representar a distribuição de variáveis contínuas ou discretas. Eles permitem a visualização da forma geral da distribuição, identificando assim qualquer assimetria, picos ou outliers. Neste caso, os histogramas foram usados para visualizar a distribuição de variáveis como as pontuações de revisão.

Os gráficos de linhas foram utilizados para representar tendências ao longo do tempo. Eles são particularmente úteis para visualizar mudanças contínuas e tendências em um conjunto de dados ao longo de um período. Desse modo, os gráficos de linha foram empregados para analisar tendências nas pontuações de revisão ao longo do tempo.

IV. RESULTADOS

O Airbnb revolucionou a forma como as pessoas viajam e se hospedam ao redor do mundo. Com a plataforma, os viajantes têm acesso a uma ampla variedade de opções de acomodação, desde quartos privados até casas inteiras. Nesta análise, vamos explorar os dados disponíveis sobre os tipos de quartos mais selecionados pelos usuários do Airbnb.

Ao examinar os dados, observa-se que o tipo de quarto mais escolhido pelos usuários é "Entire home/apt" (casa/apartamento inteiro). Essa categoria representa a opção de reserva de um espaço completo, proporcionando aos hóspedes privacidade e conforto durante sua estadia. Na Figura 1 existe um gráfico de barras onde é possível visualizar essa preferência.

O gráfico de barras mostra a distribuição dos diferentes tipos de quartos oferecidos no Airbnb. A categoria "entire home/apt" se destaca, representando a maior parte das escolhas dos usuários. Isso indica que muitas pessoas optam por ter uma

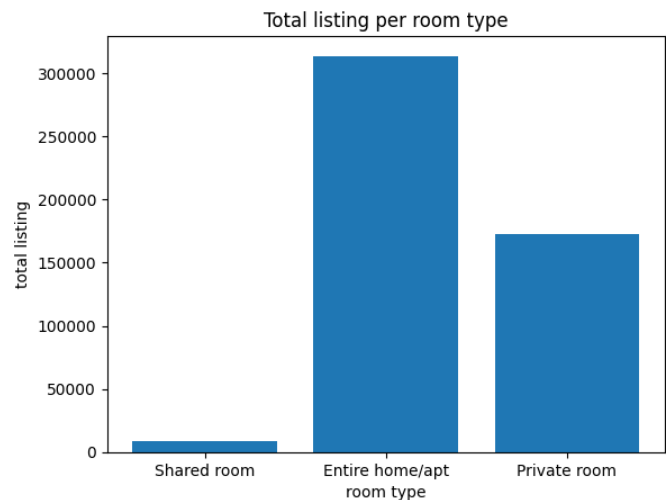


Fig. 1. Gráfico de barras com o total de quartos por tipo

casa ou apartamento inteiro só para si durante sua estadia, em vez de compartilhar o espaço com outras pessoas.

Existem várias razões pelas quais o tipo de quarto "entire home/apt" é tão popular. Em primeiro lugar, ele oferece privacidade aos hóspedes, permitindo-lhes desfrutar de seu próprio espaço sem a presença de estranhos. Além disso, essa opção é particularmente atraente para grupos de amigos ou famílias que desejam compartilhar uma experiência de viagem juntos, com espaço suficiente para todos se sentirem confortáveis.

Outro fator que contribui para a popularidade dos quartos inteiros é a possibilidade de cozinhar suas próprias refeições. Muitos imóveis do Airbnb possuem cozinhas totalmente equipadas, permitindo que os hóspedes preparem suas refeições e economizem dinheiro em restaurantes. Isso pode ser especialmente vantajoso para estadias mais longas, em que os hóspedes desejam ter mais controle sobre suas opções de alimentação.

Além disso, a popularidade dos quartos inteiros pode estar relacionada à demanda por maior segurança e distanciamento social. Durante a pandemia de COVID-19, muitos viajantes preferiram evitar aglomerações em hotéis e optaram por reservar acomodações privadas, onde pudessem manter o distanciamento social e reduzir os riscos de contágio.

Essas descobertas são importantes tanto para os viajantes que procuram acomodações quanto para os anfitriões do Airbnb, que podem considerar essa preferência ao anunciar suas propriedades.

O gráfico na Figura 2 mostra a média da pontuação de avaliações para diferentes tipos de propriedades. Os três primeiros tipos de propriedade com a maior pontuação média são "Heritage hotel", "Entire floor" e "Casa particular".

Isso sugere que os hóspedes que ficaram em "Heritage hotels", "Entire floor" e "Casas particulares" tiveram experiências positivas em relação a esses tipos de propriedades no Airbnb. Essas categorias específicas de propriedades rece-

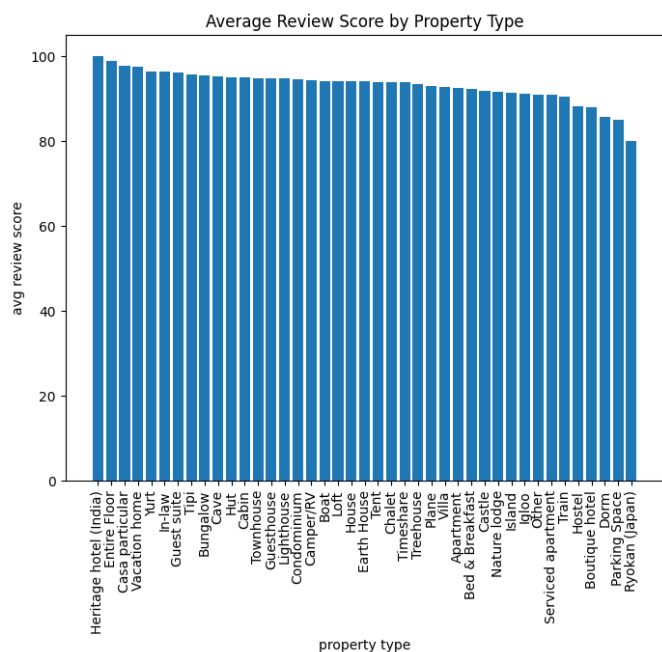


Fig. 2. Gráfico de barras com a pontuação média de avaliação por tipo de propriedade

beram avaliações consistentemente altas em comparação com outros tipos de propriedades.

Isso pode ser útil para os usuários do Airbnb, pois eles podem considerar esses tipos de propriedades como opções preferenciais ao procurar acomodações com pontuações de avaliação mais altas. Além disso, os anfitriões também podem se beneficiar dessas informações, pois podem identificar quais tipos de propriedades tendem a receber avaliações positivas e adaptar suas ofertas de acordo.

Uma análise interessante refere-se ao crescimento de anúncios mostrado na Figura 3, ao analisar a quantidade de novos anúncios ao longo do tempo observamos um aumento significativo no número de anúncios no Airbnb nos meses de julho de 2015, junho de 2015 e maio de 2015. Esses meses foram os de maior crescimento na plataforma, com 12.807 anúncios em julho, 12.153 anúncios em junho e 12.001 anúncios em maio.

Esse conhecimento pode ser válido para compreender o desenvolvimento da plataforma e reconhecer os momentos de intensa atividade e disponibilidade de anúncios. Esses dados podem ser utilizados para tomar decisões estratégicas, como elaborar campanhas de marketing direcionadas a potenciais anfitriões durante esses períodos de expansão ou adaptar políticas e recursos para suprir a crescente demanda.

Um outro gráfico interessante mostra o número total de anúncios nas 30 principais cidades, como podemos ver na Figura 4. Os três primeiros colocados nesse gráfico são Paris, Londres e Berlim. Isso sugere que essas três cidades possuem uma quantidade significativa de anúncios no Airbnb em comparação com outras cidades.

Isso pode ser interpretado como um indicativo de que

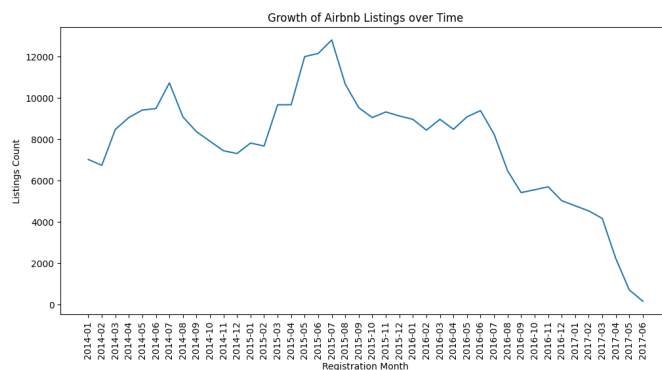


Fig. 3. Gráfico de linha mostrando o crescimento de anúncios do Airbnb ao longo do tempo

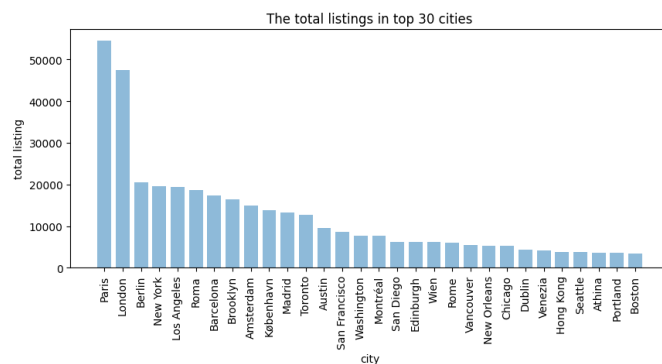


Fig. 4. Gráfico de barras com o total de anúncios nas 30 principais cidades

Paris, Londres e Berlim são destinos populares para turistas e viajantes, e há uma alta demanda por acomodações nesses locais. Essas cidades têm uma ampla variedade de opções de hospedagem disponíveis, o que pode atrair viajantes que procuram diferentes tipos de propriedades, como apartamentos, casas e quartos.

Além disso, essas informações também podem sugerir que Paris, Londres e Berlim possuem um mercado imobiliário ativo e uma economia turística bem estabelecida. A presença de um grande número de anúncios no Airbnb nessas cidades pode indicar que os proprietários de imóveis veem valor em alugar suas propriedades para viajantes, o que pode ser impulsionado pelo potencial de ganhos financeiros e pela demanda turística nessas regiões.

Falando sobre tendências, com base no gráfico da Figura 5, pode-se observar que o ano de 2016 teve o pico de avaliações. Isso pode indicar um aumento significativo na popularidade ou na quantidade de reservas realizadas através do Airbnb nesse período específico.

Isso pode ser útil para entender tendências e padrões de comportamento dos usuários ao longo do tempo. O aumento no número de avaliações em 2016 pode ser resultado de vários fatores, como um aumento geral na demanda por acomodações do Airbnb, melhorias na plataforma, expansão para novos mercados ou eventos especiais que ocorreram naquele ano.

Com essa informação, os provedores de hospedagem e a

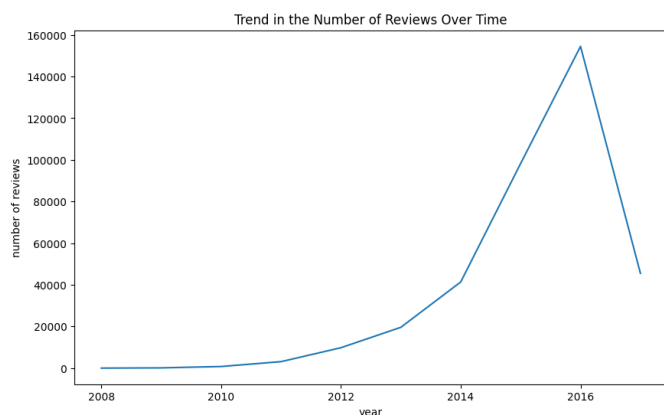


Fig. 5. Gráfico de linha com a tendência do número de avaliações ao longo do tempo

equipe do Airbnb podem se concentrar em analisar o que ocorreu especificamente em 2016 para entender melhor os fatores que contribuíram para o crescimento das avaliações. Isso pode ajudar a identificar estratégias eficazes para atrair mais reservas e melhorar a experiência do usuário em anos futuros.

Como Paris, Londres e Berlim foram as três cidades com mais anúncios focaremos em uma análise dos dados dessas cidades, respectivamente. Nessa análise, serão examinados histogramas que representam a distribuição de preços e gráficos de linha que mostram a tendência do número de avaliações ao longo do tempo para cada uma dessas cidades. Essa abordagem permitirá uma compreensão mais profunda dos padrões e das características específicas desses mercados de hospedagem, fornecendo entendimento sobre os preços praticados e a popularidade dessas cidades entre os usuários do Airbnb. Esses dados serão fundamentais para a tomada de decisões estratégicas e a formulação de políticas eficazes em relação à oferta de acomodações e ao turismo nessas localidades.

A distribuição de preços em Paris, representada pelo histograma na figura, revela algumas características interessantes. O eixo X abrange valores de 0 a 1000, enquanto o eixo Y ultrapassa 12000. O gráfico demonstra a contagem de ocorrências de diferentes faixas de preços.

Observando o histograma, é possível notar que a maioria dos preços está concentrada em faixas mais baixas, com uma queda gradual na frequência à medida que os preços aumentam. Isso indica que a maioria das propriedades em Paris tem preços relativamente acessíveis.

A presença de barras finas e altas no histograma sugere que existem alguns valores discrepantes (outliers) nos preços dos imóveis em Paris. Esses outliers representam propriedades com preços significativamente diferentes da média, o que pode ser devido a características exclusivas, localização privilegiada ou outros fatores.

A distribuição geral parece ser assimétrica e concentrada em valores mais baixos, mas com uma variação considerável na

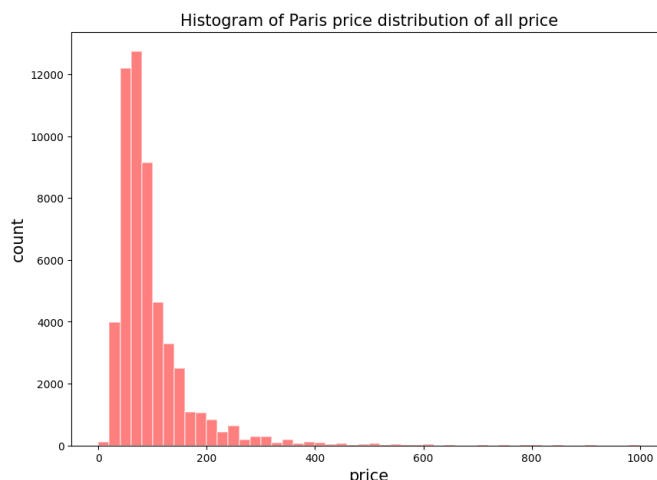


Fig. 6. Histograma da distribuição de preços de Paris

faixa de preços mais alta. Essa informação pode ser útil para quem procura imóveis em Paris, pois indica que há uma ampla gama de opções disponíveis, desde opções mais acessíveis até propriedades de luxo.

Em resumo, o histograma revela uma distribuição de preços em Paris que é predominantemente concentrada em faixas mais baixas, mas com algumas propriedades de alto valor. Essa informação pode auxiliar na compreensão do mercado imobiliário parisiense e orientar na tomada de decisões relacionadas à compra, venda ou locação de imóveis na cidade.

Olhando para o histograma da cidade de Paris abaixo do preço de 400 dólares, na Figura 7, podemos observar que a maioria dos preços das propriedades anunciadas com valores abaixo de 400 dólares, está concentrada entre 45 e 60 dólares. No entanto, o pico do histograma ocorre em torno de 50 dólares.

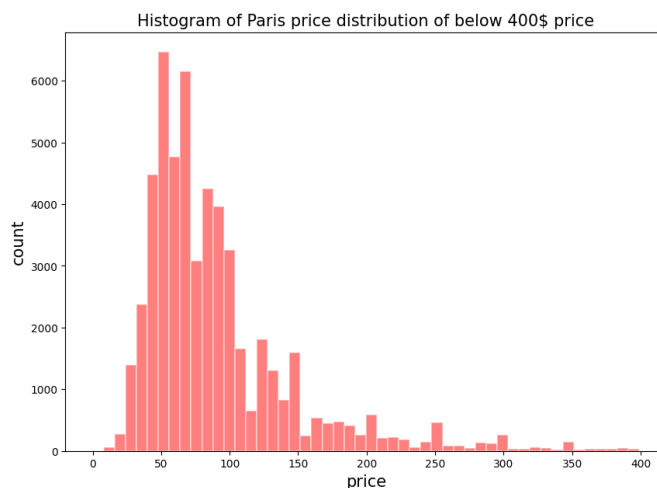


Fig. 7. Histograma da distribuição de preços de Paris abaixo do preço de 400 dólares

Com base no gráfico de tendência de *reviews* da cidade de Paris, Figura 8, podemos observar que o número de avaliações

de propriedades do Airbnb em Paris atingiu seu pico em 2016, de acordo com os dados disponíveis no conjunto de dados. Isso sugere que houve um aumento significativo na popularidade e no uso do Airbnb em Paris naquele ano.

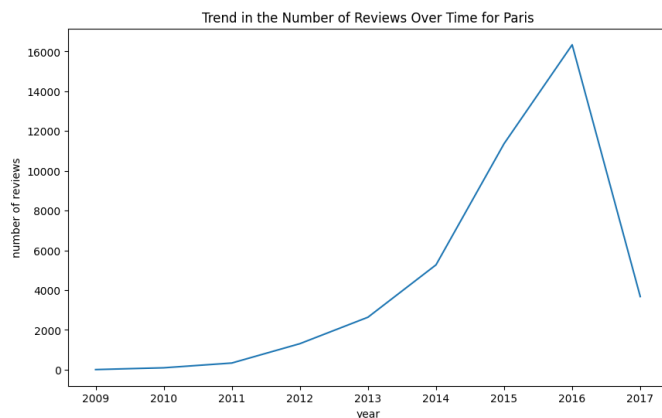


Fig. 8. Gráfico de linha com a tendência do número de avaliações ao longo do tempo na cidade de Paris

Essa informação é relevante porque indica um possível ponto de inflexão ou momento de maior interesse por parte dos usuários em relação às propriedades do Airbnb em Paris. Pode ter havido eventos específicos, como festivais, exposições ou outras circunstâncias que impulsionaram a demanda naquele ano.

Sendo assim, seria interessante investigar os fatores que contribuíram para o aumento do número de avaliações em 2016. Poderiam ser realizadas análises adicionais para identificar se houve mudanças na oferta de propriedades, políticas governamentais relacionadas ao compartilhamento de acomodações ou outros fatores que possam ter influenciado essa tendência.

Compreender os motivos por trás desse pico de avaliações em 2016 pode fornecer informações valiosas para proprietários de propriedades no Airbnb em Paris, gestores de políticas públicas e profissionais do setor de turismo, permitindo-lhes tomar decisões informadas e estratégicas com base nessa tendência observada.

Isso significa que há uma quantidade significativa de propriedades com preços próximos a 50 dólares, representando uma faixa de valor comum para as acomodações disponíveis em Paris. Essa informação pode ser útil para viajantes que desejam encontrar acomodações dentro dessa faixa de preço específica. Além disso, os anfitriões podem usar esse conhecimento para ajustar seus preços de acordo com a demanda e a concorrência nessa faixa de valor mais popular.

A segunda cidade é Londres, a distribuição de preços nessa cidade é representada pelo gráfico na Figura 9 e apresenta algumas características interessantes. O eixo x do gráfico abrange uma faixa de valores de 0 a 1000, indicando os diferentes intervalos de preços considerados. Já o eixo y, por sua vez, representa a contagem de ocorrências de cada faixa de preço, sendo que pode chegar a mais de 10000.

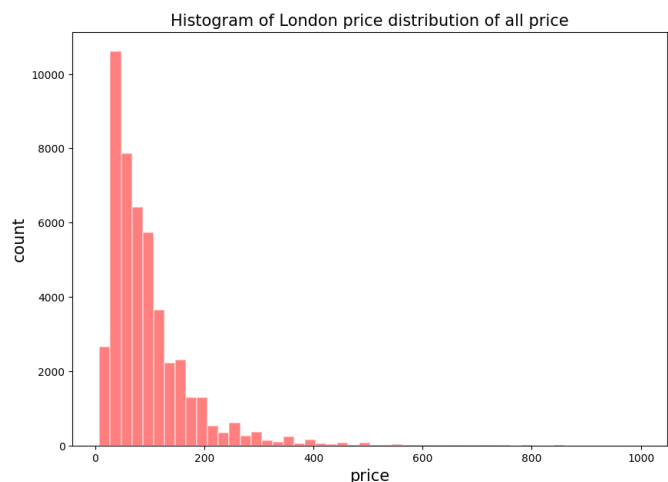


Fig. 9. Histograma da distribuição de preços de Londres

Pode-se inferir que a maioria dos imóveis em Londres tem preços relativamente baixos, concentrando-se em faixas de valores inferiores. Essa distribuição de preços pode ser influenciada por diversos fatores, como localização, tamanho e condição dos imóveis. Além disso, é possível que existam também algumas propriedades de luxo com preços significativamente mais altos, representando as ocorrências na extremidade direita do histograma.

No geral, a distribuição de preços em Londres, conforme representada pelo histograma, demonstra a existência de uma grande variedade de valores imobiliários na cidade. Através desse gráfico, é possível ter uma visão geral da distribuição de preços e identificar padrões e tendências que podem ser úteis para a compreensão do mercado imobiliário em Londres.

Olhando para o histograma que mostra a distribuição de preços dos alugueis de Airbnb em Londres, Figura 10, onde o valor está abaixo de 300 dólares. O gráfico revela que a maioria dos alugueis possui um preço na faixa de 25 a 50 dólares, conforme indicado pelo pico observado nessa faixa.

Isso indica que em Londres existem muitos alugueis de baixo custo, com preços concentrados em torno de 25 a 60 dólares. Isso pode indicar a presença de uma demanda significativa por opções de hospedagem acessíveis nessa faixa de preço na cidade.

Com base nisso, os usuários interessados em encontrar acomodações econômicas em Londres podem se concentrar na faixa de preço de 25 a 60 dólares para maximizar suas chances de encontrar opções dentro desse intervalo de preço. Além disso, os anfitriões que desejam atrair mais hóspedes podem considerar estabelecer preços competitivos nessa faixa popular, a fim de atender à demanda por acomodações de baixo custo em Londres.

Olhando o gráfico que mostra a tendência do número de avaliações ao longo do tempo para a cidade de Londres, na Figura 11, o insight gerado pelo gráfico é que o pico de número de avaliações ocorreu em 2016, assim como nos gráfico de tendências gerais e no gráfico de Paris.

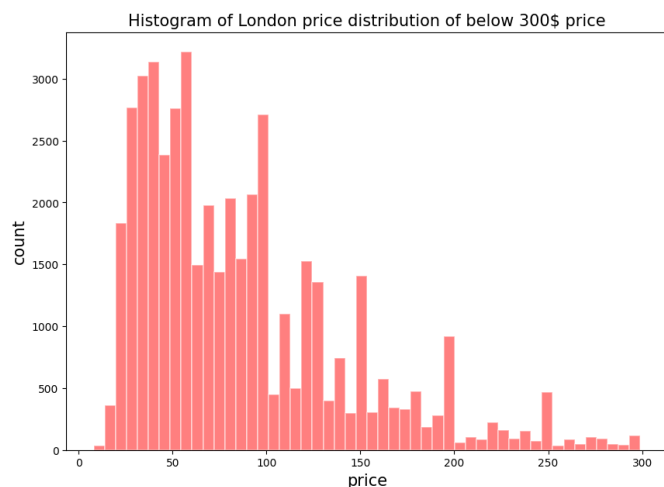


Fig. 10. Histograma da distribuição de preços de Londres abaixo do preço de 300 dólares

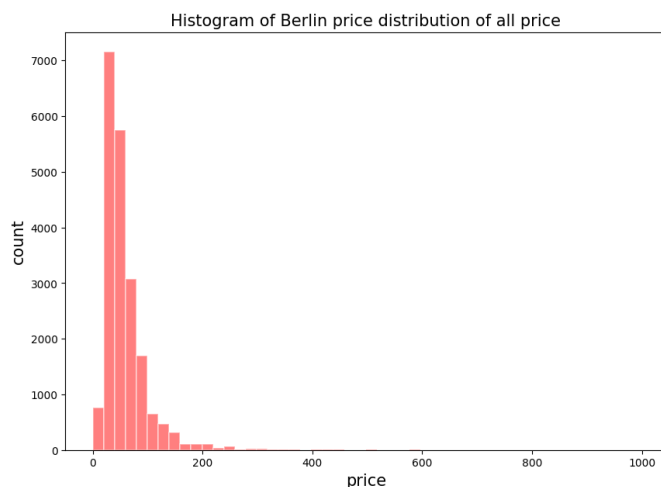


Fig. 12. Histograma da distribuição de preços de Berlim



Fig. 11. Gráfico de linha com a tendência do número de avaliações ao longo do tempo na cidade de Londres

A terceira cidade com mais anúncios é Berlim, através da visualização do histograma na Figura 11, é possível obter alguns insights sobre a distribuição dos preços das listagens em Berlim. Primeiramente, é importante notar que a maioria dos preços está concentrada em uma faixa relativamente estreita. Isso pode indicar que existe uma tendência para os preços se concentrarem em torno de um valor central, possivelmente refletindo características do mercado imobiliário de Berlim

Outro aspecto a ser observado é a presença de alguns outliers na distribuição dos preços em Berlim. Esses são valores extremos que se desviam significativamente do restante dos dados. No histograma, esses outliers seriam representados por barras isoladas ou muito distantes das demais. Essa observação sugere que há algumas listagens com preços muito acima da média ou da faixa normal de preços em Berlim.

O histograma mostrado na figura Figura 13 revela a distribuição dos preços em Berlim que são inferiores a 200 dólares. A maioria dos preços está concentrada em um intervalo entre 25 e 50 dólares, indicando que a faixa de preços

mais comum na cidade está nessa faixa específica. O gráfico mostra que há uma contagem significativa de propriedades com preços próximos a 25 e 50 dólares, sugerindo que esses valores são bastante representativos na oferta imobiliária em Berlim. Essa análise pode ser útil para pessoas interessadas em encontrar acomodações acessíveis na cidade, pois fornece uma visão geral da distribuição de preços e ajuda a entender o mercado imobiliário local.

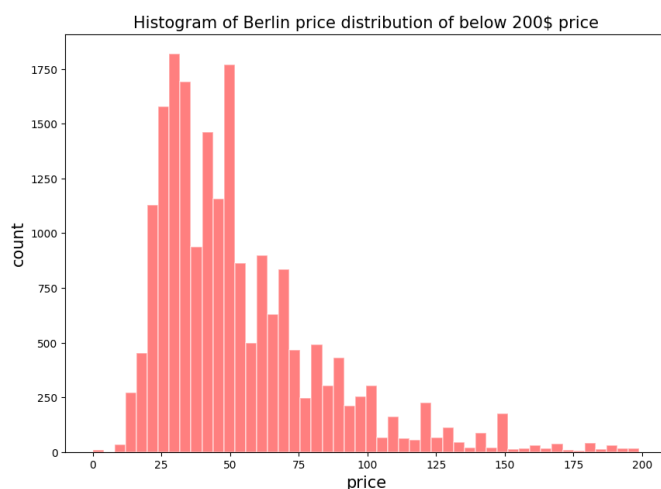


Fig. 13. Histograma da distribuição de preços de Berlim abaixo do preço de 200 dólares

Com base na Figura 14, podemos observar uma tendência interessante no número de avaliações ao longo do tempo para a cidade de Berlim. O gráfico mostra que 2016 foi o ano com o maior número de avaliações, indicando um possível aumento na popularidade dos locais anunciados nesse período. No entanto, a partir de 2017, podemos observar uma queda gradual no número de avaliações, sugerindo uma possível diminuição do interesse ou uma estabilização no número de novas avaliações. Essa tendência pode ser influenciada por di-

versos fatores, como mudanças nas preferências dos viajantes, concorrência com outras cidades ou até mesmo questões específicas relacionadas à cidade de Berlim. Acompanhar de perto essas tendências pode ajudar proprietários de imóveis e profissionais do setor de hospitalidade a tomar decisões informadas sobre seus negócios na cidade.



Fig. 14. Histograma da distribuição de preços de Berlim abaixo do preço de 200 dólares

V. CONCLUSÃO

Com base nas informações fornecidas nos diferentes gráficos, podemos chegar a algumas conclusões interessantes sobre o mercado de hospedagem em algumas cidades populares. Vamos analisar cada gráfico separadamente e, em seguida, combinar essas informações para tirar conclusões mais abrangentes.

No gráfico mostrado na Figura 1, observamos que o tipo de quarto mais escolhido pelos usuários é "Entire home/apt" (casa/apartamento inteiro). Isso indica uma preferência dos usuários por ter um espaço privativo completo durante sua estadia, em vez de compartilhar o espaço com outros viajantes. Isso pode ser atribuído ao desejo de privacidade, conforto e conveniência que um espaço privativo oferece.

No gráfico apresentado na Figura 2, os três primeiros tipos de propriedade com a maior pontuação média são "Heritage hotel", "Entire floor" e "Casa particular". Isso sugere que esses tipos de propriedade têm recebido feedback positivo dos usuários e são considerados opções de hospedagem de alta qualidade. Os hotéis de herança, em particular, podem ser valorizados por sua rica história e ambiente encantador, enquanto as opções de "Entire floor" e "Casa particular" podem oferecer uma experiência mais exclusiva e personalizada.

Na Figura 4, Paris, Londres e Berlim ocupam as três primeiras posições em termos de número total de anúncios. Essas cidades são destinos turísticos populares, conhecidas por sua cultura, história e atrações. O fato de terem um alto número de anúncios sugere que há uma demanda significativa por acomodações nessas cidades, o que pode ser impulsionado pelo turismo e pelo interesse geral em visitar esses locais famosos.

Na Figura 5, observamos que o ano com o maior pico de avaliações é 2016. Isso pode indicar que esse foi um ano especialmente movimentado ou significativo para a plataforma de hospedagem analisada. O aumento no número de avaliações pode refletir um aumento na popularidade das acomodações listadas, uma maior conscientização da plataforma ou um aumento geral no número de usuários e viajantes.

No gráfico apresentado na Figura 7, observamos que a faixa de preço concentrada está entre 45 e 60 dólares, com o pico do histograma ocorrendo em torno de 50 dólares. Isso sugere que a maioria das acomodações em Paris, com preços abaixo de 400 dólares, está concentrada nessa faixa de preço específica. Esse intervalo pode ser considerado acessível para muitos viajantes e pode refletir a oferta de acomodações de diferentes níveis de luxo e comodidades.

No gráfico mostrado na Figura 10, observamos que a faixa de preço concentrada está entre 25 e 60 dólares. No entanto, o pico do histograma ocorre em torno de 60 dólares. Isso sugere que, em Londres, a maioria das acomodações com preços abaixo de 300 dólares está concentrada em uma faixa de preço mais ampla, com um pico específico em torno de 60 dólares. Isso pode indicar uma variedade de opções de preço e níveis de luxo disponíveis para os viajantes em Londres.

No gráfico da Figura 13, observamos que a faixa de preço concentrada está entre 25 e 50 dólares. No entanto, o pico do histograma ocorre em torno de 30 dólares. Isso sugere que, em Berlim, a maioria das acomodações com preços abaixo de 200 dólares está concentrada em uma faixa de preço mais ampla, com um pico específico em torno de 30 dólares. Isso pode indicar uma variedade de opções de preço e níveis de luxo disponíveis para os viajantes em Berlim.

Em resumo, com base nos gráficos fornecidos, podemos concluir que os usuários preferem reservar "Entire home/apt" (casa/apartamento inteiro) como tipo de quarto mais escolhido. Os tipos de propriedade com a maior pontuação média são "Heritage hotel", "Entire floor" e "Casa particular", sugerindo uma preferência por acomodações de alta qualidade e personalizadas. Paris, Londres e Berlim são as três cidades com o maior número de anúncios, refletindo sua popularidade como destinos turísticos. O ano de 2016 apresentou o maior pico no número de avaliações, indicando um período movimentado ou significativo. Além disso, os histogramas de preços em Paris, Londres e Berlim revelam as faixas de preço predominantes e os picos específicos de cada cidade.

Essas conclusões combinadas fornecem insights valiosos sobre as preferências dos usuários e as tendências do mercado de hospedagem nessas cidades.

REFERÊNCIAS

- [1] D. Guttentag, "Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector," *Current Issues in Tourism*, vol. 18, pp. 1–26, 12 2013.
- [2] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, and I. Stoica, "Apache spark: A unified engine for big data processing," *Commun. ACM*, vol. 59, no. 11, p. 56–65, oct 2016. [Online]. Available: <https://doi.org/10.1145/2934664>

- [3] M. Armbrust, R. S. Xin, C. Lian, Y. Huai, D. Liu, J. K. Bradley, X. Meng, T. Kaftan, M. J. Franklin, A. Ghodsi, and M. Zaharia, "Spark sql: Relational data processing in spark," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1383–1394. [Online]. Available: <https://doi.org/10.1145/2723372.2742797>
- [4] M. Zaharia, T. Das, H. Li, S. Shenker, and I. Stoica, "Discretized streams: An efficient and fault-tolerant model for stream processing on large clusters," in *Proceedings of the 4th USENIX Conference on Hot Topics in Cloud Computing*, ser. HotCloud'12. USA: USENIX Association, 2012, p. 10.
- [5] G. Zervas, D. Proserpio, and J. W. Byers, "The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry," *Journal of Marketing Research*, vol. 54, no. 5, pp. 687–705, 2017. [Online]. Available: <https://doi.org/10.1509/jmr.15.0204>
- [6] C. Gibbs, D. Guttentag, U. Gretzel, L. Yao, and J. Morton, "Use of dynamic pricing strategies by airbnb hosts," *International Journal of Contemporary Hospitality Management*, vol. 30, no. 1, pp. 2–20, 2018.
- [7] J. Gutiérrez, J. C. García-Palomares, G. Romanillos, and M. H. Salas-Olmedo, "The eruption of airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in barcelona," *Tourism management*, vol. 62, pp. 278–291, 2017.
- [8] A. Ahuja, A. Lahiri, and A. Das, "Predicting airbnb rental prices using multiple feature modalities," *arXiv preprint arXiv:2112.06430*, 2021.
- [9] I. Airbnb, "Airbnb listings," 2023, accessed: 19 mai. 2023. [Online]. Available: https://public.opendatasoft.com/explore/dataset/airbnb-listings/table/?disjunctive=host_verifications&disjunctive=amenities&disjunctive=features