

Relatório de Descoberta de Dados

1. Fonte, Licença, Cidade, Frescor e Cobertura



Fonte e Licença

Os dados foram obtidos através do projeto **Inside Airbnb**, que coleta e disponibiliza publicamente informações do site Airbnb em várias cidades do mundo, com o objetivo de facilitar a análise do impacto do turismo de curta duração na habitação e nas comunidades.

- **Fonte:** Inside Airbnb.
- **Licença:** Os dados disponibilizados pelo Inside Airbnb são licenciados sob a **Creative Commons Attribution 4.0 International License (CC BY 4.0)**.
 - **Atribuição (BY):** A principal restrição é a **obrigação de atribuir o crédito** apropriadamente, fornecer um link para a licença e indicar se foram feitas alterações. Isso significa que, ao utilizar os dados em qualquer produto ou relatório, deve-se citar o Inside Airbnb como fonte.
 - **Restrições e Mitigação:** A natureza de licença aberta do tipo CC BY 4.0 é relativamente permissiva, permitindo o uso comercial e a modificação dos dados. A principal **restrição** é a **atribuição**. A **mitigação** para essa restrição é simplesmente garantir a **citação e o reconhecimento claros** do Inside Airbnb em todas as publicações ou produtos derivados. É crucial revisar a política de dados do Inside Airbnb para garantir total conformidade com quaisquer termos adicionais (embora a CC BY 4.0 seja o principal framework).



Por que Buenos Aires?

A escolha de **Buenos Aires** (BA) se justifica por dois fatores principais:

1. **Relevância de Mercado:** Buenos Aires é um dos destinos internacionais mais procurados por turistas brasileiros. A análise do mercado de aluguel por temporada na cidade oferece *insights*

valiosos sobre o comportamento de um nicho importante para o Brasil.

2. **Conexão Afetiva e Local:** A existência de **parentes (tio e primo)** na cidade permite o acesso a **conhecimento local** (*local knowledge*).

Frescor e Cobertura

- **Frescor:** O Inside Airbnb fornece dados trimestrais. Um bom frescor é vital para a análise de preços, pois o mercado de aluguel por temporada é **altamente dinâmico**. A **restrição** aqui é que os dados podem estar desatualizados em relação às últimas semanas, e a **mitigação** é a **comparação com benchmarks** e a consideração da sazonalidade no momento da coleta.
- **Cobertura:** O Inside Airbnb busca listar todas as propriedades ativas no Airbnb para a cidade. A cobertura é geralmente considerada **alta** para as listagens ativas. Contendo informações de diversas propriedades em bairros renomados da cidade de Buenos Aires, como Palermo, Recoleta, Belgrano, La Boca, Puerto Madero, Almagro, etc. No entanto, a **restrição** reside no fato de que **não inclui propriedades que não estão listadas no Airbnb** (ex: hotéis, aluguéis tradicionais de longo prazo), limitando a visão do mercado imobiliário total. A **mitigação** é aceitar que a análise se restringe ao mercado de **aluguéis de curta duração via Airbnb**.

2. Campos Usados e Mini Dicionário

Foram utilizadas duas bases principais:

Listings (listings.csv - Visão Estática da Propriedade)

Contém informações sobre a propriedade, o anfitrião e métricas resumidas.

| Campo Exemplo | Descrição |

id	O identificador único do anúncio (listing) no Airbnb.
listing_url	O link/URL direto para a página do anúncio no site do Airbnb.
scrape_id	O identificador único da coleta de dados (scrape) em que este registro foi obtido.
last_scraped	A última data em que as informações deste anúncio foram coletadas (extraídas) do site.
source	A fonte de onde o anúncio foi extraído (ex: city, previous sources).
name	O título ou nome do anúncio dado pelo anfitrião.
description	A descrição detalhada do imóvel fornecida pelo anfitrião.

neighborhood_overview	A descrição do bairro e da área circundante, fornecida pelo anfitrião.
picture_url	O link/URL da foto principal do anúncio.
host_id	O identificador único do anfitrião.
host_url	O link/URL direto para a página de perfil do anfitrião.
host_name	O nome do anfitrião.
host_since	A data em que o anfitrião se juntou ao Airbnb.
host_location	A localização de onde o anfitrião diz ser.
host_about	O texto de descrição/biografia do anfitrião.
host_response_time	A média de tempo que o anfitrião leva para responder às mensagens

	dos hóspedes (ex: "em poucas horas", "em um dia").
host_response_rate	A percentagem de mensagens às quais o anfitrião respondeu (geralmente nos últimos 30-60 dias).
host_acceptance_rate	A percentagem de pedidos de reserva que o anfitrião aceitou.
host_is_superhost	Indica se o anfitrião tem o status de Superhost (booleano: true ou false).
host_thumbnail_url	O link/URL da miniatura (thumbnail) da foto de perfil do anfitrião.
host_picture_url	O link/URL da foto de perfil do anfitrião.
host_neighbourhood	O bairro onde o anfitrião vive ou onde o anfitrião diz que o anúncio está localizado (pode ser

	diferente de neighbourhood).
host_listings_count	O número de anúncios ativos que o anfitrião possui na plataforma.
host_total_listings_count	O número total de anúncios do anfitrião (ativos e inativos).
host_verifications	Uma lista das verificações de identidade ou credenciais que o anfitrião forneceu.
host_has_profile_pic	Indica se o anfitrião possui uma foto de perfil (booleano: true ou false).
host_identity_verified	Indica se a identidade do anfitrião foi verificada pelo Airbnb (booleano: true ou false).
neighbourhood	O nome do bairro conforme fornecido pelo anfitrião.

neighbourhood_cleansed	O nome do bairro padronizado e limpo pelo Airbnb.
neighbourhood_group_cleansed	O grupo maior ao qual o bairro pertence (ex: distrito, região administrativa).
latitude	A coordenada geográfica de latitude do imóvel.
longitude	A coordenada geográfica de longitude do imóvel.
property_type	O tipo de propriedade (ex: Apartamento, Casa, Cabana, etc.).
room_type	O tipo de espaço que está sendo alugado (ex: Casa/Apto Inteiro, Quarto Privado).
accommodates	O número máximo de hóspedes que o imóvel pode acomodar.

bathrooms	O número de banheiros (pode estar descontinuado em bases mais recentes).
bathrooms_text	A descrição do número de banheiros em formato de texto (ex: "1.5 banheiros").
bedrooms	O número de quartos/dormitórios disponíveis.
beds	O número total de camas disponíveis.
amenities	Uma lista de comodidades/itens oferecidos no anúncio.
price	O preço noturno do anúncio (em valor monetário).
minimum_nights	O número mínimo de noites para reserva.
maximum_nights	O número máximo de noites para reserva.

minimum_minimum_nights	O menor valor de noites mínimas permitido (considerando as regras variáveis do calendário).
maximum_minimum_nights	O maior valor de noites mínimas permitido (considerando as regras variáveis do calendário).
minimum_maximum_nights	O menor valor de noites máximas permitido (considerando as regras variáveis do calendário).
maximum_maximum_nights	O maior valor de noites máximas permitido (considerando as regras variáveis do calendário).
minimum_nights_avg_ntm	A média ponderada do número mínimo de noites futuras (Next Twelve Months).
maximum_nights_avg_ntm	A média ponderada do número máximo de noites futuras (Next Twelve Months).

calendar_updated	O último momento em que o calendário do anúncio foi atualizado (pode estar descontinuado).
has_availability	Indica se o imóvel tem ou pode ter disponibilidade (booleano).
availability_30	O número de dias disponíveis nos próximos 30 dias .
availability_60	O número de dias disponíveis nos próximos 60 dias .
availability_90	O número de dias disponíveis nos próximos 90 dias .
availability_365	O número de dias disponíveis nos próximos 365 dias .
calendar_last_scraped	A última data em que o calendário foi extraído.

number_of_reviews	O número total de avaliações recebidas.
number_of_reviews_ltm	O número de avaliações recebidas nos últimos 12 meses .
number_of_reviews_l30d	O número de avaliações recebidas nos últimos 30 dias .
availability_eoy	A disponibilidade do anúncio até o final do ano (End of Year).
number_of_reviews_ly	O número de avaliações recebidas no último ano civil .
estimated_occupancy_l365d	A ocupação estimada nos últimos 365 dias .
estimated_revenue_l365d	A receita total estimada gerada nos últimos 365 dias .
first_review	A data da primeira avaliação recebida.

last_review	A data da avaliação mais recente recebida.
review_scores_rating	A pontuação geral de avaliação do anúncio.
review_scores_accuracy	A pontuação sobre a precisão/fidelidade da descrição.
review_scores_cleanliness	A pontuação sobre a limpeza do imóvel.
review_scores_checkin	A pontuação sobre a facilidade do check-in .
review_scores_communication	A pontuação sobre a comunicação do anfitrião.
review_scores_location	A pontuação sobre a localização do imóvel.
review_scores_value	A pontuação sobre o custo-benefício .

license	O número da licença ou registro governamental, se houver.
instant_bookable	Indica se o anúncio pode ser reservado instantaneamente (booleano).
calculated_host_listings_count	O número de anúncios do anfitrião, calculado pelo Airbnb.
calculated_host_listings_count_entire_homes	O número de anúncios do anfitrião do tipo Casa/Apto Inteiro .
calculated_host_listings_count_private_rooms	O número de anúncios do anfitrião do tipo Quarto Privado .
calculated_host_listings_count_shared_rooms	O número de anúncios do anfitrião do tipo Quarto Compartilhado .
reviews_per_month	A média de avaliações recebidas por mês.

Contém uma linha para cada dia por listagem, fornecendo a série temporal de preços.

| Campo Exemplo | Descrição |

listing_id	O identificador único do anúncio ao qual esta entrada de calendário se refere. Liga-se ao campo id da base <i>listings</i> .
date	A data específica (dia) para a qual as informações de disponibilidade e preço são válidas.
available	Indica se o anúncio está disponível para reserva nesta data (booleano: t para true/verdadeiro, f para false/falso).
price	O preço diário do anúncio nesta data, incluindo taxas ou descontos padrão. É o preço original listado.
adjusted_price	O preço diário final do anúncio nesta data, após a aplicação de quaisquer ajustes, como descontos sazonais ou regras de preço dinâmico do anfitrião. Este é o preço que o hóspede pagaria por uma única noite naquela data.
minimum_nights	O número mínimo de noites que deve ser reservado se a reserva começar nesta data específica. Este valor pode mudar dia a dia.

maximum_nights	O número máximo de noites que pode ser reservado se a reserva começar nesta data específica. Este valor também pode mudar dia a dia.
----------------	--

3. Limitações e Mitigação

🔴 Limitação Crucial: Calendário Indisponível

A principal limitação na base **calendar** é que um dia marcado como **available = 'f'** (False/indisponível) não significa necessariamente que houve uma reserva. Pode significar:

1. **Reserva:** O imóvel foi efetivamente alugado por um hóspede.
 2. **Bloqueio:** O anfitrião **bloqueou** o dia por motivos próprios (manutenção, uso pessoal, aluguel de longo prazo fora do Airbnb, etc.).
- **Restrição:** Não é possível diferenciar uma **reserva real** de um **bloqueio** do anfitrião. Isso dificulta a medição da **taxa de ocupação real** e da **demanda efetiva**.
 - **Mitigação:**
 - **Adoção de Métrica Conservadora:** Utilizar a métrica de "dias não disponíveis" como um **limite superior** da ocupação.
 - **Análise de Padrões:** Em análises mais aprofundadas, pode-se tentar identificar **padrões de bloqueio** (ex: bloqueios em períodos longos e contínuos) que são menos prováveis de serem reservas, tratando-os de forma diferente de bloqueios de 1-2 dias dispersos (mais prováveis de serem reservas ou manutenções pontuais).
 - **Foco na Demanda de Preço:** Priorizar a análise de como o **preço varia** nos dias em que o imóvel está **disponível** (**available = 't'**), relacionando essa variação a eventos e sazonalidade.