

Dashboard Acomodações de Airbnb – Rio de Janeiro

Introdução

Este projeto foi criado para o Curta a Metragem, uma empresa que presta serviços a anfitriões (*hosts*) de acomodações de Airbnb. O objetivo do dashboard é oferecer visibilidade acerca do mercado, com uma visão geral das acomodações na principal cidade em que a empresa atua (Rio de Janeiro), tanto para os sócios do Curta a Metragem, como para seus clientes, que também terão acesso ao dashboard interativo. O banco de dados encontrado é da época da virada do ano (Réveillon).

Metodologia

A metodologia utilizada para este projeto foi o CRISP-DM (sigla para *Cross Industry Standard Process for Data Mining*, ou Processo Padrão Inter-Indústrias para Mineração de Dados), que é bem difundida para todo tipo de projeto de dados, e consiste em 6 etapas fundamentais:

1. Entendimento do negócio
2. Entendimento dos dados
3. Preparação dos dados
4. Modelagem
5. Avaliação
6. Implementação

O projeto será explicado em maiores detalhes nas próximas páginas seguindo esta estrutura.

Entendimento do negócio

Esta é uma etapa fundamental do projeto, pois consiste em entender os objetivos do projeto, como ele será usado, a quem se destina (o público), qual o tempo hábil para desenvolvimento.

Neste momento, é importante obter respostas a algumas perguntas:

- Qual o objetivo deste projeto?

O objetivo é construir um dashboard que possa ser visualizado tanto por sócios da empresa como por seus clientes, e que traga insights sobre o mercado em geral, especificamente sobre o período de Réveillon, com foco no que pode ser feito para aumentar a taxa de ocupação do imóvel.

- Qual o objetivo de cada grupo para o qual este projeto se destina?

Como empresa, o Curta a Metragem quer entender mais sobre o mercado através deste dashboard, e assim ter maior capacidade de focar em algum tipo de negócio específico, com base em localização, preço e tipo de acomodação.

Por outro lado, a empresa também quer aumentar sua captação de leads, e este dashboard deve aumentar a autoridade que ela traz junto aos seus clientes.

Por fim, seus clientes também querem entender melhor o mercado, saber se estão no caminho correto e assim faturar mais com seu imóvel.

- O que queremos conhecer ou descobrir?

Queremos conhecer mais sobre o mercado na cidade, com mais informações sobre preço, ocupação média, importância de avaliações, se é melhor para o anfitrião focar em oferecer para acomodação a casa inteira ou quartos compartilhados.

- Supondo que consigamos atingir o objetivo, o que vem depois? Como este trabalho será usado?

Primeiro, este dashboard será disponibilizado para os sócios do Curta a Metragem, e depois será disponibilizado no site da empresa para seus clientes. Futuramente, ele poderá ser utilizado como base para desenvolvimento de textos no blog da empresa, tornando os insights presentes no dashboard mais claros para o público geral.

Entendimento e preparação dos dados

Os dados foram baixados em formato .csv da plataforma Kaggle, deste [endereço](#).

Variável	Descrição	Tipo
id	Identificador único da listagem	Qualitativa
name	Nome da listagem	Qualitativa
host_id	Identificador único para o anfitrião	Qualitativa
host_name	Nome do anfitrião	Qualitativa
neighbourhood	Bairro	Qualitativa
latitude	Latitude do imóvel	Quantitativa
longitude	Longitude do imóvel	Quantitativa
room_type	Tipo de acomodação	Qualitativa
price	Preço da diária na moeda local	Quantitativa
minimum_nights	Número mínimo de noites	Quantitativa
number_of_reviews	Número de avaliações	Quantitativa
reviews_per_month	Avaliações por mês	Quantitativa
calculated_host_listings_count	Número de listagens do anfitrião na região	Quantitativa
number_of_reviews_ltm	Número de avaliações dos últimos 12 meses	Quantitativa
availability_365	Disponibilidade nos próximos 365 dias	Quantitativa
last_review	Data da última avaliação	Quantitativa

Ao realizar a análise exploratória dos dados, algo que se percebe imediatamente é o valor exorbitante dos preços. Apesar de a coleta dos dados ter sido realizada em período festivo, época em que é esperado que os preços sejam significativamente maiores, não faz sentido algum uma diária ser R\$ 600 mil, que dirá para um quarto compartilhado. É pensado que houve erro na coleta da maioria desses dados, porém é quase impossível saber se o erro foi em ignorar a vírgula (nesse caso a solução seria simplesmente dividir por 100) ou não. Portanto, a solução mais conservadora foi excluir os *outliers*.

Foram excluídas acomodações com preços de diária acima de R\$ 1 mil para os tipos de acomodação quarto privado e quarto compartilhado. Da mesma forma, foram excluídas acomodações de quarto de hotel e imóvel inteiro com preços acima de R\$ 20 mil. Essas exclusões somaram 718 listagens, apenas 2,5% do número total de imóveis.

Também foi necessário realizar mais alguns ajustes latitude e longitude, que vieram sem vírgula e com casas decimais variadas. Ou seja, foi necessário usar uma fórmula específica para esse ajuste no excel:

=ESQUERDA([@latitude];3)&" "&EXT.TEXT([@latitude];4;10)

Uma fórmula análoga foi usada para longitude.

Com base nestas colunas nativas da base, foram criadas as seguintes colunas adicionais:

Variável adicional	Descrição	Tipo
lat1	Latitude ajustada	Quantitativa
long1	Longitude ajustada	Quantitativa
Occupancy rate	1-Disponibilidade nos próximos 365 dias/365	Quantitativa
Entire home/apt	Se é listagem de acomodação inteira	Qualitativa binária
Private room	Se é listagem de acomodação com quarto privativo	Qualitativa binária
Shared room	Se é listagem de acomodação com quarto compartilhado	Qualitativa binária

“Occupancy rate” é a taxa de ocupação, calculada com o percentual de dias dos próximos 365 dias que estão reservados.

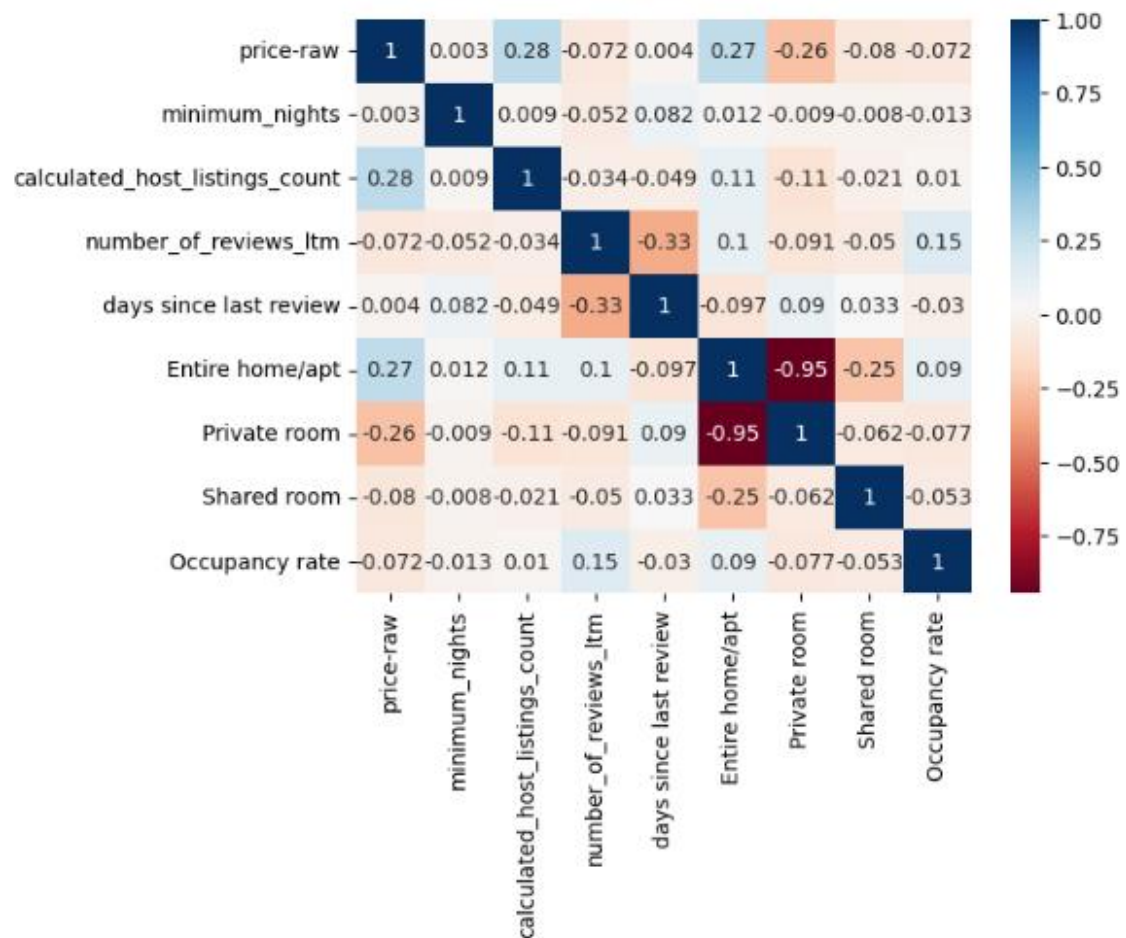
Utilizando a técnica ‘one hot encoding’, também foram criadas as colunas “Entire home/apt”, “Private room” e “Shared room”, com o objetivo de entender como elas impactam em preço e

Fazendo uma análise por meio de medidas resumo das principais variáveis, temos a tabela abaixo:

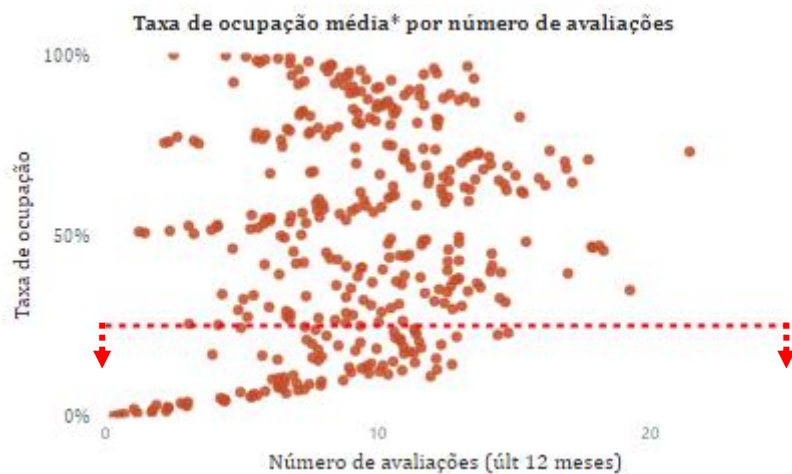
Medida resumo	price-raw	minimum_nights	calculated_host_listing_s_count	number_of_reviews_ltm	days since last review	Occupancy rate
Contagem	27.573,00	27573	27573	27573	19483	27573
Média	898,92	4,81	9,42	6,26	581,35	44%
Desvio padrão	1.081,51	23,58	27,84	10,68	503,47	37%
Mínimo	33,00	1	1	0	324	0%
25%	309,00	2	1	0	345	4%
50%	592,00	2	1	1	372	41%
75%	1.000,00	4	4	8	462	79%
Máximo	10.000,00	1125	228	122	4289	100%

Analisando os preços, agora a média é de R\$ 898,92 e a mediana é de R\$ 592,00, o que parece fazer sentido para os preços de diária no período de virada do ano no Rio de Janeiro, ao analisar o site.

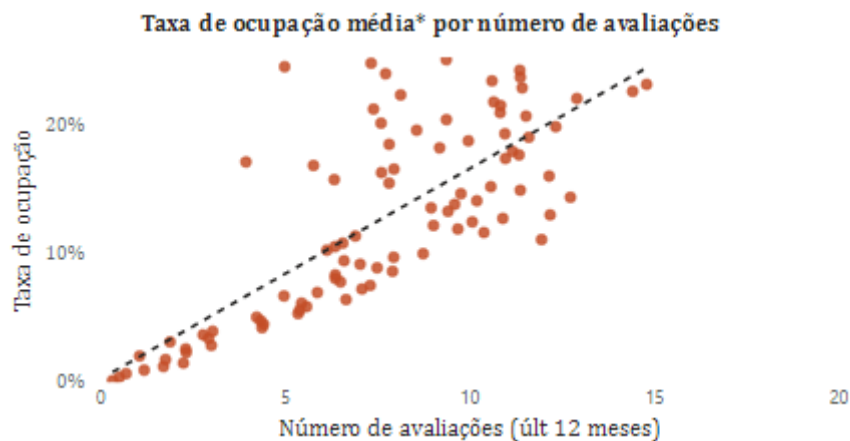
Também é importante conhecer como as variáveis se correlacionam:



Observando a dispersão de taxa de ocupação e número de avaliações, a princípio não parece haver correlação, como mostra o número na tabela acima, de 0,15. No entanto, quando observamos as taxas de ocupação menores, vemos como é importante ter avaliações no site.



Para taxas de ocupação inferiores a 25%, a correlação é de 0,40. Na figura abaixo, é fácil entender como avaliações podem ser importantes para tirar o imóvel da vacância.



Seleção das variáveis

As variáveis foram selecionadas para o dashboard com base nas informações mais importantes para os sócios da empresa e seus clientes.

Eles querem conhecer o mercado de Airbnb no Rio de Janeiro em número de acomodações, preços e faixas de preço, taxa de ocupação e localidade. Todas essas informações devem estar disponíveis por bairro e zona.

As correlações mais fortes para prever preço foram tipo de quarto e `calculated_host_listings_count`, ou seja, quantos imóveis o anfitrião tem listado no site. É considerado que colocar o número de imóveis que um anfitrião tem no dashboard traria um ganho pequeno, inclusive porque esta provavelmente é uma relação não-causal: a maioria dos anfitriões que têm mais de um imóvel listado é empresas, que realizam gestão profissional dos imóveis, por isso conseguem colocar preços maiores.

Outra variável importante é a taxa de ocupação, que tem as maiores correlações com o número de avaliações, principalmente nos últimos 12 meses.

Portanto, as variáveis escolhidas para integrar o dashboard foram:

- Número de acomodações;
- Preço;
- Taxa de ocupação;
- Tipo de acomodação;
- Localização geográfica (mapa).

Modelagem – Construção do dashboard

Foi escolhido para o dashboard principal um formato simples, de fácil visualização. Foram colocadas nos visuais 3 categorias, que podem ser visualizadas através de drill up/drill down:

- Zona: visualização pelas zonas da cidade: zona central, zona sul, zona norte e zona oeste;
- Top bairros: visualização simplificada, com os 10 bairros de com maior número de acomodações, e outros bairros aglutinados em "Outros";
- Bairros: visualização completa, com todos os bairros da cidade.

Em um dashboard secundário, é explicada a forma como o número de avaliações impacta na taxa de ocupação, principalmente para baixas taxas de ocupação – típicas de listagens novas.

Validação

O dashboard foi apresentado aos clientes, que ficaram satisfeitos com o trabalho.

Roadmap – melhorias futuras

Dadas as limitações da base de dados, o trabalho ficou restrito à época de Réveillon. Seria interessante no futuro ter uma análise e um dashboard para todas as épocas do ano. Também traria ganhos para o negócio dashboards de outras cidades. O Rio de Janeiro é o principal mercado onde a empresa presta serviços. No entanto, atualmente ela já presta serviços em outras cidades do Estado do Rio de Janeiro.

Algumas variáveis adicionais, que trouxessem maiores informações sobre o imóvel, poderiam ajudar a aprofundar as análises, como por exemplo: número de cômodos, principais comodidades, área do imóvel, número de quartos, número de banheiros.

Implantação

Os dashboards serão implementados no website do Curta a Metragem, e serão divulgados nas mídias sociais da empresa.

Fontes

Fonte dos dados: <https://www.kaggle.com/datasets/liamarguedas/rio-de-janeiro-airbnb-listings-2022>

Análises anteriores da base: <https://www.kaggle.com/code/liamarguedas/rio-de-janeiro-airbnb-listings-ead-walkthrough>

Pesquisa de acomodações no Airbnb: [Pesquisa Airbnb](#)