

Ocupação & Preço

Realizar o cálculo do preço dinâmico e previsões de ocupação ajustáveis, baseando-se em fatores temporais, representou um grande desafio. Este documento detalha todo o **raciocínio analítico e experimental** que levou ao desenvolvimento do modelo.

Análise e Extração de Características dos Dados

O processo iniciou-se com a análise de dois documentos extraídos do Airbnb (***listing*** e ***calendar***), a fim de compreender suas características e extrair atributos relevantes que influenciariam tanto a previsão de preço quanto a ocupação.

Documentos Utilizados:

- ***listing***: Lista os imóveis disponíveis na cidade de Buenos Aires de forma **estática**.
- ***calendar***: Traz uma **noção temporal** dos imóveis, exibindo características em datas que variam de janeiro de 2025 a fevereiro de 2026.

Características Selecionadas:

Embora os arquivos (principalmente o ***listing***) contivessem muitas características irrelevantes para o escopo (como características do proprietário, latitude e longitude), as seguintes foram selecionadas por sua influência no preço dinâmico:

- Tipo da sala (**privativa, compartilhada**, etc.).
- Número de pessoas que podem ser **acomodadas** (m).
- Número de **quartos** (q).
- Número de **banheiros** (b).
- Número de **camas** (c).
- A **data normalizada**.
- O **bairro** de localização do imóvel.
- A característica da **sazonalidade**.

Tratamento de Dados Ausentes (NaNs)

Durante a etapa de análise, notou-se a presença de muitos valores *NaN* (ausentes) nos dados. Para tratar especificamente os dados ausentes de cômodos (banheiros, quartos e camas), foi utilizada uma substituição baseada na variável: **número máximo de pessoas que o imóvel pode acomodar (m)**.

Estratégia de Imputação:

Os valores *Nan*s para as características de cômodos foram substituídos por um valor inteiro (*c*, *b*, *q*) conforme as seguintes equações, onde *m* é o número máximo de hóspedes:

$$c = m \text{ (número de camas);}$$

$$b = \frac{m}{3} \text{ (número de banheiros);}$$

$$q = \frac{m}{2} \text{ (número de quartos);}$$

Valores de ponto flutuante resultantes foram arredondados para cima utilizando a função **ceil()** (arredondamento para o inteiro superior).

Cálculo da Ocupação e Sazonalidade

A base de dados *calendar* possui a característica **available**, que indica se o imóvel estava ocupado ou disponível em uma data específica. Isso permitiu o cálculo da ocupação, uma das características mais importantes para a predição de preços dinâmicos.

Metodologia de Ocupação e Sazonalidade:

1. A ocupação foi calculada pelo agrupamento de imóveis disponíveis em uma **janela de 14 dias**.
2. Foi calculada a **média de ocupação** de imóveis nesta faixa de 14 dias.
3. A **sazonalidade** foi classificada com base no desvio dessa média:
 - **Alta Temporada:** Janelas com quantidade de imóveis ocupados **66% acima da média**.
 - **Média Temporada:** Janelas que estavam entre a faixa de **33% e 66%** acima da média.
 - **Baixa Temporada:** Janelas **abaixo de 33%** acima da média.

Pré-processamento e Treinamento do Modelo

Após o tratamento e limpeza dos dados, a etapa final de pré-processamento e o treinamento do modelo de regressão foram realizados.

Codificação de Variáveis Categóricas:

Características de natureza categórica, como os níveis de sazonalidade, foram codificadas numericamente utilizando a estratégia **LabelEncoder()**:

- Baixa Temporada -> 0
- Média Temporada -> 1
- Alta Temporada -> 2

Modelo de Regressão:

Para o treinamento do modelo, optou-se pelo algoritmo de **Descida do Gradiente Estocástico (SGD)** para realizar a regressão, devido à sua eficiência e leveza computacional para grandes volumes de dados.

- **Configurações:** 2000 iterações e tolerância de 0.001.
- **Vantagem:** A característica estocástica (inicialização aleatória dos pesos) ajuda o gradiente a convergir para um **mínimo global**, evitando falsos positivos (mínimos locais).

Próximas Etapas:

Com o modelo regressor e preditor concluído, a próxima fase é a **implementação de um endpoint via API Rest** para comunicação entre o aplicativo e o modelo preditor de preço dinâmico.