

Mini Projeto - Roof Imóveis

DEX04 - Grupo 21

Projeto 1, Slot 3

1.0 A Questão de Negócio

1.1 Meta do projeto

Fornecer consultoria estratégica para a empresa Roof Imóveis na tomada de suas decisões na expansão de sua área de atuação, sobre quais são os cinco melhores imóveis para realizar investimento no Condado de Country nos Estados Unidos, assim como, quais os cinco imóveis que não são recomendáveis para realizar investimento.

1.2 Metodologia

Através da análise dos dados e arquivos fornecidos pela empresa, por meio do uso de bibliotecas Python e metodologias específicas da Ciência de Dados, analisamos tais informações para identificar quais são os dados mais relevantes para gerar informações acerca do objetivo de forma mais precisa. Com base nisso, realizamos um alinhamento e um ranking dos imóveis com maior probabilidade de venda e os que não são propícios para o investimento. Para atingir nosso objetivo, verificaremos qual área se apresenta mais lucrativa, e se utilizando das métricas *grade*, *zipcode* e *price/living* encontraremos os possíveis 5 melhores investimentos e os piores também.

2.0 O Entendimento do Negócio

2.1 Dados disponíveis

O arquivo disponibilizado pela empresa em forma de CSV contém informações detalhadas sobre a venda de imóveis e suas características na região do Condado de King no período de maio de 2014 a maio de 2015. O arquivo possui 21.613 registros e 21 atributos com a seguinte descrição:

Variável	Descrição
id	Identificador único do imóvel
date	Data da Venda
price	Preço de venda
bedrooms	Nº de Quartos
bathrooms	Nº de Banheiros
sqft_liv	Tamanho de área habitável em ft²
sqft_lot	Tamanho do terreno em ft²
floors	Número de andares
waterfront	Indicativo se o imóvel é a beira-mar

Variável	Descrição
view	Grau de quão belo é a vista do imóvel (0 a 4)
condition	Condição da casa (1 a 5)
grade	Classificação por qualidade de material utilizado na construção
sqft_above	Área em cima do solo em ft²
sqft_basmt	Área em abaixo do solo em ft²
yr_built	Ano de construção
yr_renov	Ano de restauração, caso o contrário, 0.
zipcode	Zip Code 5 - Similar ao CEP
lat	Latitude
long	Longitude
sqft_liv15	Média da área habitável dos 15 imóveis mais próximos, em ft²
sqft_lot15	Média da área do lote dos 15 imóveis mais próximos, em ft²

2.2 Relevância dos dados para o nosso objetivo

Os dados fornecidos pela empresa se mostram relevantes visto que nos permite analisar o escopo do problema a ser resolvido. Por meio dessas informações fornecidas poderemos aplicar as metodologias existentes na Ciência de Dados de forma a nos gerar insights importantes para o negócio analisado além de ser possível criar mapas para auxiliar na localização dos imóveis, observando sua distribuição na cidade. Isso auxiliará o cliente na tomada de decisões, pois ele será capaz de visualizar onde os imóveis mais rentáveis estão.

2.3 Sobre a influência dos dados na solução

Os dados fornecidos não nos trazem informações diretas, então será necessário que tratemos esses dados e identifiquemos quais são os mais relevantes para o nosso objetivo.

2.4 Setor do negócio analisado

Setor imobiliário com expansão internacional.

3.0 A Coleta de Dados

3.1 Sobre a aceção dos dados

Os dados fornecidos fazem sentido com relação ao negócio analisado, porém, é necessário que façamos um tratamento bem elaborado para identificar quais dados serão mais importantes para a análise. Por exemplo os dados referentes a longitude e latitude, é necessário que façamos um tratamento específico para que dessa forma um endereço seja gerado e consigamos identificar questão como localidade exata dos imóveis, visto que apenas com o dado *zipcode* a informação não é tão precisa. Outro exemplo é com relação ao número de banheiros, onde é possível identificar que alguns possuem informações como 0.25 e 0.75. Conforme a descrição do atributo *bathrooms* o valor de 0.50 representa um quarto com vaso sanitário, mas sem chuveiro, logo, os valores 0.25 e 0.75 não fazem sentido para a nossa análise.

3.2 Sobre o formato dos dados fornecidos

Sobre os dados fornecidos, algumas colunas possuem dados em formatos e padrões diferentes uns dos outros, sendo necessário tratar tais ocorrências. Um exemplo é o campo *data* que está no formato *object*, e deverá ser alterado para o formato *date*. Outro exemplo, é a coluna *floors* que está no formato *float*, e deverá ser alterado para o formato *int*. Com essas alterações poderemos usar as funções existentes para a análise dos dados.

3.3 Informações importantes que são possíveis de se obter por meio dos dados

Por meio dos dados fornecidos podemos verificar, por exemplo, o ano de construção da casa, quantos banheiro e quartos ela possui, a qualidade do material utilizado na construção, o preço de venda, assim como a data de venda, além de verificarmos se a mesma já foi vendida mais de uma vez dentro de certo período, e se a mesma foi ou não reformada, entre outras informações importantes para formular nossos *insights* através da correlação de tais informações.

4.0 Limpeza de Dados

4.1 Remoção de dados outliers e valores ausentes

Ao usar a biblioteca *Geopy* a partir da latitude e longitude fornecida na base de dados, identificamos campos ausentes na coluna *country*. Concluimos que algumas colunas, geradas após a biblioteca, deveriam ser excluídas, como por exemplo, as colunas *number*, *address*, *street*, *city*, *county*, *state*, *number2*, *country*, *x*, *y*, *state*, *county*, *number2*.

5.0 Exploração de Dados

5.1 Visualização dos dados

Através da biblioteca *Seaborn* plotamos gráficos para analisar dados referentes aos *outlines*, onde verificamos os valores que deveriam ser ajustados. Também plotamos um gráfico para analisar a quantidade de imóveis vendidos por mês. Utilizando também a mesma biblioteca para encontrar as correlações nos dados e para identificarmos a região de maior lucro. Por meio desses resultados pudemos gerar novas tabelas relacionando algumas outras informações como *grade* e *price/living*.

5.2 Nossas métricas essenciais.

Nossas métricas essenciais para chegarmos aos nossos objetivos foram latitude e longitude, para gerarmos o endereço e dessa forma identificarmos a áreas de interesse, também utilizamos o *zipcode*, *price/living* (onde fizemos o cálculo por área) e *grade*, principalmente.

5.3 Decisão que tomamos com base nos dados

Com base no local que teve maior lucro usando como parâmetro o *zipcode*, definimos os cinco melhores e os cinco piores imóveis para se investir.

Análise final + Sugestão de imóveis

5 melhores:

Definindo a área mais lucrativa, verificamos os 10 imóveis de menor preço por área habitável para considerar a mais lucrativa

Com os imóveis mais lucrativos, pegamos os 5 imóveis que tinham a maior grade para definir os melhores investimentos

id: 2923049393

id: 2023049372

id: 2923049421

id: 2923049399

id: 6121800050

5 piores:

Definindo a área menos lucrativa, verificamos os 30 imóveis (pela amostra ser aproximadamente 6 vezes maior do que a amostra utilizada para calcular os melhores) de menor preço por área habitável para considerar as menos lucrativas

Com os imóveis menos lucrativos, pegamos os 5 imóveis que tinham menor grade para definir os piores investimentos

id: 3885801970

id: 3885803895

id: 3885804225

id: 4151800375

id: 4151800420