

Mini Projeto (Deliverable 01) - Roof Imóveis

Alexandre Nonato (RID: #28700)

1.0 A Questão de Negócio:

1.1 A Empresa

A Roof Imóveis é uma das maiores empresas do ramo imobiliário brasileiro e quer expandir sua área de atuação fazendo um investimento internacional. A empresa irá investir em imóveis no Condado de County, nos Estados Unidos.

1.2 A Região escolhida

O Condado de King é um dos 39 condados do estado americano de Washington. A sede e cidade mais populosa do condado é Seattle. Foi fundado em 1852.

A região tem mais de 2,2 milhões de habitantes, de acordo com o censo nacional de 2020, sendo o condado mais populoso do estado e o 12º mais populoso do país.

2.0 O Entendimento do Negócio:

2.1 A Meta do trabalho

Deste modo, fui contratado para uma consultoria estratégica. A Roof Imóveis deseja saber quais os 5 imóveis que a empresa deveria investir e o porquê e quais 5 imóveis que não seriam recomendados para o investimento, em hipótese alguma.

2.2 Como posso chegar?

Foi utilizado o dataset disponibilizado pela Escola DNC, a partir do website:

<https://www.kaggle.com/datasets/harlfoxem/housesalesprediction>

O Dataset apresenta a venda de imóveis e suas características na região de maio de 2014 à maio de 2015, tendo um total de 21.613 registros com um total de 21 atributos.

A proposta é aplicar os conceitos de CRISP-DM e Python, Bibliotecas aprendidos até o momento em um projeto prático, além de desenvolver um pensamento analítico, insights para tomada de decisões.

3.0 A Coleta de Dados:

3.1 Os dados disponíveis

Os dados disponíveis no website Kaggle destacam 21 atributos de imóveis espalhados pelo Condado de King.

- Identificador do imóvel;
- Data da venda;
- Preço da venda;
- Número de quartos;
- Número de banheiros;
- Tamanho da área habitável;
- Tamanho do terreno;
- Número de andares;
- Indicativo se o imóvel é beira-mar;
- Beleza da vista;
- Condição da casa;

- Qualidade do material usado na construção;
- Área acima do solo;
- Área abaixo do solo;
- Ano da construção;
- Ano da restauração;
- Zipcode;
- Latitude;
- Longitude;
- Média da área habitável dos 15 imóveis mais próximos;
- Média da área do lote dos 15 imóveis mais próximos.

3.2 O formato dos dados

O dados coletados no website estão em formato .csv e estão muito bem organizados, favorecendo o trabalho do cientista de dados nessa consultoria.

4.0 A Exploração de Dados:

4.1 Os *outliers*

Os outliers podem ser definidos como os dados que se diferenciam drasticamente de todos os outros no dataset. Ou seja, um outlier é um valor fora da curva, para melhor ou pior, e que podem gerar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.

No caso deste trabalho, o outlier não é um ponto negativo, já que procuramos justamente os melhores e piores imóveis da região. Nesse caso, alguns outliers encontrados podem ser exatamente o que estamos buscando.

4.2 Passo a passo

- Importamos as bibliotecas necessárias;
- Comandos iniciais para importar o arquivo .csv;
- A partir do zipcode, encontramos cada cidade e iniciamos a criação de nova coluna "Cidade";
- Conferimos a criação da coluna "Cidade";
- Conferimos quantas cidades únicas foram encontradas (Total: 24 cidades);
- Removemos os 'id's duplicados;
- Deletamos colunas que não seriam utilizadas: 'sqft_living15', 'sqft_lot15';
- Criamos coluna para descobrir o preço mediano por região;
- Utilizamos fórmula para identificar o quão acima os preços estão do preço mediano da região (%)?;
- Adicionamos a percentagem mediana como recomendação de preço de venda e calculando o lucro;
- Delimitamos as melhores casas pelos parâmetros condition e grade, reduzindo para 581 imóveis;
- Delimitamos as melhores casas pelos parâmetros recommendation e profit, reduzindo para 82 imóveis;
- Delimitamos as melhores casas pelos parâmetros bedrooms e bathrooms, reduzindo para 54 imóveis;
- Delimitamos as melhores casas pelo parâmetro view, reduzindo para 5 imóveis;
- Chegamos às 5 casas recomendadas;
- Delimitamos as piores casas pelos parâmetros condition e grade, reduzindo para 8 imóveis;
- Delimitamos as piores casas pelo parâmetro bathrooms, reduzindo para 5 imóveis;

- Delimitamos as melhores casas pelos parâmetros bedrooms e bathrooms, reduzindo para 54 imóveis;
- Chegamos às 5 casas não-recomendadas;

5.0 Análise final e Sugestão de imóveis:

Os 5 imóveis que recomendados para compra foram os seguintes:

- Id: 5072200040
- Id: 2045800006
- Id: 3365900175
- Id: 2024059059
- Id: 2817100570

Os 5 imóveis que não-recomendados para compra foram os seguintes:

- Id: 6303400395
- Id: 745000005
- Id: 5702500050
- Id: 3980300371
- Id: 8658300340