

ROOF IMÓVEIS

Relatório de recomendação de imóveis para King County – Washington – USA

Aluno: Luiz Paulo Zamprogno

RID: 32705

1 – Carregamento dos dados

Foi feito o download do arquivo csv, copiado para o drive e transformado em um pandas dataframe.

2 – Limpeza dos dados

Todos os tipos de dados foram analisados para verificar se estavam de acordo com o que é esperado para o nome de cada coluna.

Um ponto que chamou a atenção foi que as colunas, 'date', 'price', 'bathrooms', 'floors' estavam como tipo de dados object e float64.

Com isso, as seguintes alterações foram feitas nessas colunas:

- A coluna 'date' foi transformada para o tipo datetime;
- A coluna 'price' foi transformada para o tipo int64, pelo motivo de que não havia valores decimais, e que isso não afetaria nossa recomendação significativamente;
- A coluna 'bathrooms' apresentava vários dados com valores decimais e zerados, e como não é possível existir uma casa com 2,5 banheiros ou nenhum, por exemplo, foi considerado o arredondamento dos valores para baixo, exclusão dos valores zerados e a transformação para o tipo int64;
- Na coluna 'bedrooms', os valores zerados encontrados foram excluídos e a coluna foi transformada para o tipo int64;
- Na coluna 'floors', foi seguido a mesma lógica do que na coluna banheiros. Valores decimais foram arredondados para baixo e transformados para o tipo int64

- Por fim, foi verificado a existência de dados zerado em todo o dataframe, e constatado que não havia nenhum.

3 – Verificação de outliers

Foram utilizados gráficos de boxplot para todas as colunas, com biblioteca matplotlib, para tentar encontrar algum dado discrepante. A única coluna que chamou a atenção foi a coluna 'bedrooms', onde uma análise foi feita para as casas com quantidade de alta de quartos. Foi encontrado uma casa com 33 quartos, 1 banheiro, 1 piso e uma área de construção relativamente pequena para 33 quartos, então essa linha foi excluída do dataframe.

4 – Análise de dados

Foi criado uma coluna com o nome da cidade e população de cada linha do dataframe, de acordo com o seu zipcode, utilizando a biblioteca uszipcode.

A seguir, foi criado uma coluna com o valor por área do terreno.

Mais duas colunas foram adicionadas, uma para sabermos a idade de cada casa e uma para sabermos quantos anos se passaram desde a última reforma, se existiu.

Outro dataframe foi criado, agrupando a coluna 'id', para sabermos se houve mais de uma operação de venda por 'id', e retornando os valores de mínimo e máximo da coluna 'price', para sabermos se houve valorização, desvalorização ou não houve mais de uma venda. Com isso foi possível analisar que 172 casas tiveram mais de uma operação de venda e que todas valorizaram.

5 – Avaliação de critérios

A análise de recomendação de investimento foi baseada em:

- Encontrar as 5 casas que mais valorizaram na venda;
- Buscar a localização dessas 5 casas;
- Avaliar se são próximas;
- Encontrar outras 5 casas com atributos semelhantes, mas com menor valor.

Com a ajuda da biblioteca geopandas, foi possível plotar e confirmar que as 5 casas que mais valorizaram são próximas umas das outras.

6 – Recomendação

6.1 - Investimento

Confirmando a proximidade das 5 casas mais valorizadas, com o passo a passo descrito no passo 5 e aplicando filtros no dataframe, conseguimos recomendar outras 5 casas com atributos semelhantes às mais valorizadas, mas com menor valor, pensando em que elas têm grandes chances de também valorizar.

6.2 – Não investimento

Foi feita uma classificação descendente na coluna 'price', para encontrarmos as casas com os maiores valores do dataframe. Porque não recomendo o investimento, logo no início, em casas com valores muito altos, pois a venda dessas casas é mais difícil e demorada para acontecer, assim gerando muitas despesas para a imobiliária.