



Deliverable:

Projeto – Roof Imóveis

Willian Kakuta

RID: 29018

Turma: DS05

Sapucaia do Sul – Agosto de 2022

1.0 A Questão de Negócio:

1.1 Qual a minha meta?

Análise de banco de dados para identificar quais os 5 imóveis que apresentam uma melhor oportunidade de negócio para se investir e quais os 5 imóveis que seriam as piores oportunidades de negócio que nunca devem ser vistas como opção de investimento.

1.2 Como posso chegar?

Todos os imóveis apresentados recebiam a data e o valor em que foram vendidos, tendo conhecimento disto, podemos ver os dados da frequência de venda para encontrar um padrão de imóvel, assim tendo conhecimento do comportamento do consumidor. Com esta filtragem podemos comparar m² dos imóveis e com o zip code podemos encontrar o preço médio dos imóveis de cada região, sabendo se aquele imóvel está barateou ou caro.

2.0 O Entendimento do Negócio:

2.1 Quais dados eu tenho disponível?

Separei os dados em 4 categorias:

- Valor vendido
- Estrutura do imóvel (todos os dados relacionados a características estruturais do imóvel).
- Condição do imóvel (dados mais “subjetivos”, como a vista, a condição, se o imóvel é muito antigo, qualidade do material).
- Localidade (dados referentes ao local como, área habitável, terreno, zip code)

2.2 Eles são relevantes para o problema?

Seguindo a minha tese, esses dados são relevantes para o problema, porque a estrutura do imóvel e a condição do imóvel, apresentam o estilo de imóvel que existe uma maior venda. Já a localidade consegue nos mostrar de onde é esses imóveis e comparando com os imóveis locais quais são os maiores que a média. E por último o valor vendido, que com base neste valor você consegue analisar quais estão avaliados abaixo do valor de mercado no site <https://www.unitedstateszipcodes.org/>, assim encontrando as melhores oportunidades com o melhor custo/benefício do mercado.

2.3 Eles me trazem uma solução direta?

Não trazem, porque terei que correlacionar os dados de estrutura do imóvel com a condição do imóvel para ver a frequência de venda. Assim, encontrando o tipo de imóvel mais vendido. Depois, terei que ver a relação média de tamanho do imóvel e entrando no site do governo americano para localizar esses imóveis para ver o preço de mercado nas suas respectivas regiões. Por fim, fazendo um comparativo com o preço de venda para ver o melhor custo/benefício e oportunidade de compra dos imóveis.

2.4 Qual o meu setor ou range de negócio?

O setor é o imobiliário, empresa Roof Imóveis, gigante brasileira que tem o interesse de expansão para o Estados Unidos no Condado de King em Washington.

3.0 A Coleta de Dados:

3.1 Os dados que eu tenho fazem sentido?

Seguindo a minha tese os dados fazem sentido. Quando separei nas 4 categorias, consegui ter uma visualização e entendimento maior de onde eu queria chegar. Assim, montando uma estratégia de trabalho, onde fui

seguindo até chegar em uma conclusão, depois de 2 testes falhos anteriormente.

3.2 Estão no formato que eu gostaria?

Os dados utilizados foram “fáceis” de manipular. Tive algumas dificuldades e imprevistos, mas nada que não tenha conseguido resolver, nem que se fosse com uma solução mais trabalhosa, mas funcional.

3.3 O que mais eu consigo obter de informação destes dados?

As informações que obtive dos dados que utilizei, foram o padrão de casas que foram mais vendidas, a localidade do imóvel, comparativo de tamanho com a média dos imóveis da região, comparativo de preço com a média de preços da região. Mas se eu tivesse mais tempo e conhecimento, gostaria de ter trazidos dados externos para completar o relatório, como tamanho das famílias, faixa etária das pessoas, renda familiar, para deixar mais elaborado o padrão de consumo, e outros dados como criminalidade por exemplo, pois talvez uma casa esteja ‘barata’ em relação as outras, por fatores externos que não conseguimos ter apenas com o dataset.

4.0 A Limpeza de Dados:

4.1 Remoção de dados outliers e valores ausentes

O Dataset até onde analisei não tinha nenhum valor ausente, mas tinha valores iguais a 0. O que deixou em dúvida se era um valor nulo ou realmente fazia parte das informações. Por exemplo, casas com 0 quartos e 0 banheiros, para a maioria das pessoas isso é algo impossível, mas para mim que conheço a realidade do Japão, casas com nenhum quarto é uma realidade, e algumas não tem banheiro e é usado um banheiro comunitário. Então, resolvi deixar esses dados, pois como foi usado muito histograma para ver frequência, média, moda e mediana. O que fosse fora do padrão automaticamente seriam excluídos, como os outliers. E excluí algumas colunas que não

apresentaram resultados satisfatório nos gráficos e não estavam mais sendo usadas para a análise.

5.0 A Exploração de Dados:

5.1 Visualização dos dados

Na visualização dos dados, conseguimos chegar nas respostas que buscávamos que seriam os 5 imóveis para se investir e os 5 para não se investir, usando as métricas que foram importantes para solução deste desafio.

5.2 Quais as minhas métricas essenciais

Primeiramente fizemos um comparativo entre os tamanhos da área habitável e lote da média dos 15 imóveis mais próximos, para filtrarmos os imóveis maiores que a média.

Segundo, fizemos gráficos de histograma para encontrar frequência, média, moda e mediana para descobrir um padrão de consumo do consumidor daquele Condado. Onde foi apresentado as seguintes informações:

- 3 a 5 quartos
- 2 a 3 banheiros
- 2 andares ou menos
- 3 a 4 condição do imóvel
- 7 a 9 qualidade do material usado
- Ano de construção maior que 1980

Terceiro passo foi pegar o zip code para identificar o local do imóvel e ver a média de preços dos imóveis daquela região.

Quarto passo, calculamos com uma função qual é a % comparando o preço do imóvel com a média de preços dos imóveis vendidos na região.

Para encontrarmos os 5 imóveis que não se deve investir, fizemos exatamente os mesmos passos com as métricas ao inverso.

5.3 Baseado nos dados, qual decisão devo tomar.

Baseado nos dados, chegamos à conclusão dos 5 imóveis que a empresa deveria investir, com ótimo custo/benefício, tendo um imóvel com 29% abaixo do valor de mercado e o resto perto de 20%. E os 5 imóveis que não deveríamos investir, o primeiro colocado chegou a ser 74% acima do valor de mercado por uma casa em péssimas condições, teve outra sem quartos e banheiros, em condições precárias. Em seguida, nas conclusões mostraremos as sugestões de imóveis.

Análise final + Sugestão de imóveis

5 Melhores Imóveis para se investir

	id	price	zipcode	potencial_lucro%
17698	2215900930	225000.0	98038	29
2826	8100000090	256000.0	98010	20
2221	7137900960	235000.0	98092	18
9404	4233400490	262000.0	98010	18
5738	2025700080	265000.0	98038	17

Segundo todas as nossas análises de dados, os imóveis apresentados com o melhor custo/benefício são esses. Apresentam todas as características buscadas em um imóvel, seguindo o padrão de consumo, são imóveis maiores que a média da região, abaixo do valor de mercado e com ótimo potencial de lucro.

5 Piores Imóveis para se investir

	id	price	zipcode	potencial_lucro%
13629	2023049218	445000.0	98148	-74
11224	3468800310	425000.0	98108	-30
8623	6303400395	325000.0	98146	-5
12849	293000180	370000.0	98126	-3
19452	3980300371	142000.0	98024	71

Analisando todos os dados apresentados, esses são os 5 imóveis com o pior custo/benefício que em nenhuma hipótese deve-se investir. Porque além de serem imóveis supervalorizados, as condições dos imóveis são extremamente ruins e são imóveis menores que a média do local. O último imóvel ele é muito barato pois, não tem quarto, não tem banheiro, se encontra em um péssimo estado e foi feito com a pior qualidade de material, por isso entrou nesta lista, mesmo não sendo um imóvel supervalorizado.

Colab:

https://colab.research.google.com/drive/1DXvxATl_88XLKgmDyELvDgAc1zlcQroo?usp=sharing