

1. Questão de Negócio

A empresa 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 Country, nos Estados Unidos.

Eu tenho como objetivo, apontar as melhores oportunidades de imóveis, para que a empresa possa tomar decisões mais assertivas nesse início de ciclo de investimentos internacionais. Para a realização desse projeto utilizaremos a metodologia CRISP-DM, como é um projeto single estarei responsável por todas as tarefas.

2. Entendimento de Negócio

Como a empresa Roof está situada no setor de imobiliário os dados disponibilizados pela empresa nos dão acesso aos preços de venda dos imóveis, nº de: banheiros, quartos, andares; tamanho da área habitável e terreno, condição atual do imóvel, qualidade do material utilizado na construção, entre outros.

Existem uma grande variedade de dados na tabela, porém alguns deles não tem grande relevância na análise que foi realizada e foram retirados da tabela, os dados que possuem grande relevância somente por si não apresentam uma solução direta para o problema, mas através da junção e criação de novas variáveis conseguimos chegar ao nosso objetivo.

3. Coleta de Dados

3.1 Formato dos Dados

Os dados apresentados pela empresa têm muita sinergia, porém alguns dados estão em formatos que devem ser modificados para facilitar a análise, foram eles:

1. Formatação da data: Excluindo números e letras que não poderiam estar presentes em um formato de data comum.
2. Transformação dos andares: Transformando esses dados em float(Números quebrados) em int(Números inteiros).

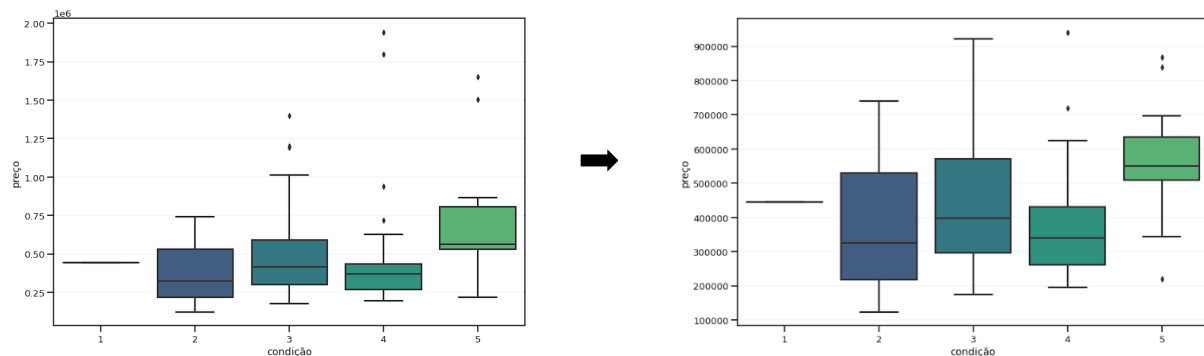
3.2 Informações Adicionais

Outras informações (Insights) que podem ser extraídos dessa tabela são:

1. Usando o zipcode(CEP) foi possível encontrar a cidade de cada imóvel
2. Valor por m² da Área Habitável e do Terreno
3. Valorização de Imóveis vendidos mais de uma vez naquele período

4. Limpeza de Dados

Foi visto que na tabela não existem valores nulos, porém existem alguns dados que são outliers de acordo com os filtros que seguiremos a seguir, esses dados outliers eu determinei que seriam os imóveis com valores acima de \$1,000,000 de dólares, segue o gráfico:



Como é visto no primeiro gráfico acima, existiam valores outliers (Esses pontos isolados fora dos candlesticks) que estão acima de \$1,000,000 de dólares, e nós trabalharemos somente com imóveis abaixo desse valor. E no segundo gráfico já é possível analisar após a exclusão desses valores.

5. Exploração de Dados

5.1 Início da Análise

A tabela disponibilizada pela empresa tem um total de 21,613 imóveis, para filtrar e diminuir consideravelmente esses números, foi levado em consideração somente os imóveis que já foram vendidos mais de uma vez naquele período (Maio de 2014 à Maio de 2015), com isso o número de imóveis que passava de 21 mil, agora é de somente 166 imóveis.

5.2 Métricas Essenciais

Então após a aplicação desse filtro macro, outras métricas importantes para a análise devem ser levadas em consideração daqui em diante, são elas:

- Condição: Condição atual em que o imóvel se encontra
- Ano_construção: O ano em que o imóvel foi construído
- Material_construção: A qualidade do material que o imóvel foi construído
- Percentual_valorização: O quanto o imóvel se valorizou em percentual (%)

5.3 Médias

Diante as métricas apresentadas anteriormente acredito que a condição atual do imóvel é uma das mais relevantes na análise, com isso tracei as médias das outras métricas de acordo com a condição, para avaliar se existe uma condição com melhores oportunidades. As medias analisadas foram: preço médio de venda, quantidade média de banheiros e quartos, preço por m² da área habitável e terreno, tempo médio que levou para o imóvel ser revendido e o percentual médio de valorização dos imóveis. Segue a tabela:

	preço	quartos	banheiros	preço_area	preço_terreno	tempo_revenda	percentual_valorização
condição							
1	445000.00	2.00	1.00	478.49	57.49	243.00	0.76
2	381118.75	3.00	1.31	314.77	54.09	149.12	0.36
3	443618.75	3.30	1.79	278.90	70.80	186.52	0.33
4	383995.84	3.03	1.61	266.74	43.15	166.26	0.31
5	564904.17	3.75	2.06	297.39	99.77	187.50	0.37

Após analisar a tabela cheguei as seguintes conclusões:

1°Condição: Além de ser a pior condição onde a reforma deve ser feita com mais urgência, ela tem as maiores médias em quase tudo, pois somente existe um único imóvel presente na condição 1.

2°Condição: Tem até valores interessantes e baixos, porém esses imóveis também deveriam ser reformados o mais rápido possível por estarem em condição 2.

3°Condição: Aqui é onde apresenta a maior concentração de imóveis de demais variados tipos, porém como é visto pelas suas médias os valores são bem altos.

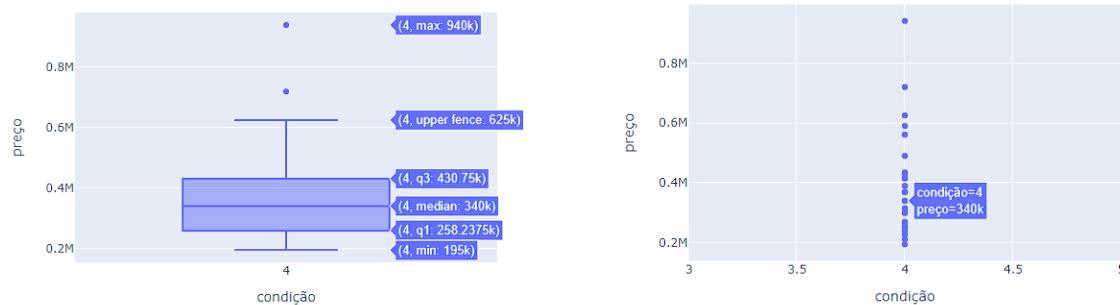
4°Condição: A condição 4 com certeza é a melhor escolha, tem médias abaixo até mesmo da condição 3 e além de estar em melhores condições.

5°Condição: A condição 5 como já era de se esperar tem médias muito altas, então não vale a pena.

Conclusão: Após a análise desses valores podemos levar em conta que os imóveis de condição 4 tem o melhor custo benefício quando levamos em consideração as médias das métricas, pois apresenta valores bem abaixo da condição 3, e também apresenta valores consideravelmente abaixo da condição 5. Então a partir daqui trabalharemos somente com imóveis de condição 4.

5.4 Visualizando Dados

Agora faremos uma exploratória em alguns gráficos, os dois apresentam os mesmos dados, porém de forma diferentes, segue os gráficos:



Visualizando esses dois gráficos podemos ver que o valor da mediana é de 340 mil dólares, e abaixo desse valor existem muitos imóveis disponíveis. Com isso vamos trabalhar com esses imóveis que estão com o valor abaixo da média.

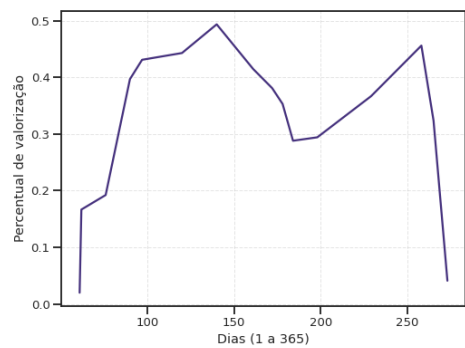
5.5 Observações Importantes

Daqui em diante todos os imóveis presentes na tabela são boas opções, porém agora após esses filtros, diferentes métricas podem ser traçadas de acordo com os requisitos que a empresa deseja. Ex: Imóveis que tenham mais de 2 quartos ou mais de 2 banheiros, preferência por alguma cidade, ano de construção mais recente, tamanho área habitável, etc.

Seguirei as métricas que me levam aos imóveis com maior potencial de valorização em torno do tempo. Pois dessa forma como a empresa está se adentrando em solo internacional, acredito que seja uma boa estratégia comprar imóvel que lhes darão retornos "rápidos", dessa forma o lucro pode ser usado para a compra de novos imóveis ou reinvestido no mesmo para maior valorização.

Análise Final

A seguir veremos um gráfico que representa o percentual de valorização dos imóveis, entorno dos dias levados para a revenda. Segue o gráfico:



Como é visto no gráfico acima, existem imóveis que até mesmo antes da metade do ano já se valorizaram cerca de 40% a 50%. Então iremos analisar os maiores percentuais de valorização antes da metade do ano. (Observação: Essa porcentagem de valorização não é 100% assertiva, é mera especulação de dados passados)

Sugestão de imóveis

Melhores Imóveis:

Esses são os 5 melhores imóveis na minha concepção e análise, apesar de nenhum deles terem recebido reforma todos estão em boas condições, além disso vale a ressalva de que eles já têm uma “idade” avançada, e mesmo assim ainda estão presentes na condição número 4, com isso certamente esses imóveis não se depreciam rápido e ainda terão muita longevidade.

Enfim, a empresa terá imóveis com certa variação de tamanho da área habitável e terreno, quantidade de quartos e banheiros, cidades em diferentes localizações que podem atender diferentes tipos de requisitos, e também todas eles têm grande potencial de lucro pela valorização ao longo do tempo como foi detalhado nas tabelas e gráficos anteriores. Segue a tabela:

	preço	quartos	banheiros	area_habitavel	tamanho_terreno	andares	condição	material_construção	ano_construção	ano_renovação	cidade	preço_area	preço_terreno	valorização	tempo_revenda	percentual_valorização
176	227000.0	1	2.00	1150	9812	1	4	7	1962	0	Auburn	197.39	23.13	112000.0	140.0	0.49
56	212000.0	2	1.00	790	7153	1	4	6	1944	0	Seattle	268.35	29.64	93875.0	120.0	0.44
190	267950.0	4	1.00	1730	7350	1	4	6	1970	0	Federal Way	154.88	36.46	115450.0	97.0	0.43
236	299000.0	3	1.75	1910	17003	1	4	8	1963	0	Auburn	156.54	17.59	124000.0	161.0	0.41
212	314950.0	3	1.75	1520	9600	1	4	7	1967	0	Renton	207.20	32.81	124950.0	90.0	0.40

Sugestão de imóveis

Piores Imóveis:

Esses são os 5 piores imóveis na minha concepção, todos eles têm valores muito acentuados pelas suas condições atuais, e algo que me chamou muita atenção é 4 deles tem suas datas de construção bem recentes, porém mesmo assim eles perderam valor de mercado em vez de se valorizarem, mesmo dois desses imóveis estando presentes em Seattle onde é natural que sempre os imóveis se valorizem e não que o contrário aconteça. Segue o gráfico:

	preço	quartos	banheiros	area_habitavel	tamanho_terreno	andares	condição	material_construção	ano_construção	ano_renovação	cidade	preço_area	preço_terreno	valorização	tempo_revenda
256	700000.0	4	3.00	3350	199253	2	3	10	2004	0	Black Diamond	208.96	3.51	-40000.0	204.0
124	489950.0	2	1.00	820	4206	1	3	5	1949	0	Seattle	597.50	116.49	-20050.0	118.0
128	489000.0	2	2.25	1290	1334	3	3	8	2007	0	Seattle	379.07	366.57	-11000.0	246.0
320	419950.0	2	1.75	1350	4003	1	3	8	2004	0	Redmond	311.07	104.91	-10050.0	299.0
118	765000.0	4	2.50	3230	4290	2	3	9	2004	0	Kirkland	236.84	178.32	-7500.0	79.0

Identificação

Nome: Samuel Nunes Ferreira

RID: 96623

COLAB:

https://colab.research.google.com/drive/1N6sZo4QhWWeud0qp0l8glMF2UprcBV_G?usp=sharing