



dnc

Mini Projeto - Roof Imóveis

Entrega 01 - Slot 03

Valdir Nascimento dos santos
RID - 97581

A Questão de Negócio:

O negócio refere-se à tomada de decisão para investimento em ativos imobiliários no Condado de King, um dos maiores dos estados unidos. A empresa Roof, precisa investir em 5 imóveis que possam ser lucrativos para eles. A Roof pretende expandir seus negócios imobiliários para outros países, e escolheu para iniciar, os Estados unidos.

A minha meta é além de fornecer os dados solicitados pela empresa Roof Imóveis, é verificar quais imóveis podem ser negociados no futuro. Buscar nos dados, informações que possam visualizar um padrão de negociação e assim entender se no futuro, outros imóveis na região podem dar um lucro maior

Para chegar ao resultado esperado, além de cruzar informações já disponibilizadas e buscando outras informações da região através de outras fontes se for necessário. Utilizarei neste processo, as bibliotecas do python como: pandas, seaborn, plotly e demais ferramentas estatísticas.

O Entendimento do Negócio:

Eu tenho um DataSet CSV contendo mais 21 mil linhas de dados, todas preenchidas corretamente, contendo informações relevantes sobre o negócio. Existem informações importantes que o dataset não traz, tais com: Tipo de residência se é bungalows e casas modernas, ou seja o estilo arquitetônico do imóvel, que é importante para os americanos. Outra informação que falta é se a residência possui ar-condicionado, internet e etc. esses dados foram colhidos na internet: **Dataset:**

<https://www.kaggle.com/harlfoxem/housesalesprediction>, **Dataset detalhado:**

<https://geodacenter.github.io/data-and-lab/KingCounty-HouseSales2015/>.

Os dados que tenho em mãos, são relevantes, poderiam ter mais informações para o atendimento de minhas metas, mais já dá uma boa dimensão para iniciar o trabalho.

O Entendimento do Negócio:

Algumas informações da empresa Roof Imóveis, seriam importante, tal como o valor estimado a ser investido, mas considerando alguns interesses dos norte-americanos na aquisição ou aluguel de imóveis, defini um valor máximo para a aquisição de uma residência. Falta também nos dados, informações de área que possa indicar um elemento de valorização, tais como aeroporto, cinema, segurança na região, parques e etc., desta forma me valho de dois indicadores, que é qualidade da vista do local e proximidade de beira-mar.

Os dados não trazem soluções diretas, será necessário processar e cruzar informações até chegar um ponto aceitável, legível e apresentável. A minha área de atuação é Análise de dados, não domino o ramo imobiliário, mas estou estudando informações úteis para desenvolver uma análise de excelência. Consegui fazer diversas pesquisas, inclusive o que os americanos exigem na hora de adquirir um imóvel. No arquivo do colabory, apresento mais detalhes.

A Preparação e Exploração dos Dados:

De posse dos dados do arquivo `kc_house_data.csv`, do qual traduzi seus atributos para português e nomeei o dataset gerado como `house.csv`, percebi a ausência de informações que melhorasse a qualidade da análise, resolvi pesquisar dados faltantes como as preferências dos norte-americanos na hora de se decidir por um imóvel. Os resultados obtidos são os seguintes:

1. **Localização:** A localização é muitas vezes considerada a característica mais importante na escolha de um imóvel nos Estados Unidos. As pessoas geralmente procuram por imóveis em bairros seguros e próximos a escolas, comércio, transporte público e outros serviços.
2. **Preço:** O preço é sempre um fator importante na compra de um imóvel. Os norte-americanos geralmente têm um orçamento definido para sua compra e procuram encontrar a melhor relação entre preço e qualidade.

A Preparação e Exploração dos Dados:

3. Tamanho: O tamanho da propriedade também é uma consideração importante. As necessidades variam dependendo do tamanho da família e da finalidade do imóvel (por exemplo, se será uma casa para morar ou uma propriedade de investimento).
4. Condição da propriedade: A condição da propriedade é um fator importante na hora de decidir comprar uma casa ou apartamento. As pessoas geralmente preferem imóveis em boas condições, com poucas ou nenhuma reforma necessária.
5. Comodidades: Os norte-americanos também valorizam as comodidades oferecidas pelo imóvel, como ar-condicionado, aquecedores, piscinas, áreas externas, entre outros.
6. História do imóvel: Para alguns compradores, a história do imóvel também pode ser um fator importante na decisão de compra. Isso pode incluir coisas como a idade da propriedade, sua história arquitetônica e sua história como um local importante na comunidade.

A Preparação e Exploração dos Dados:

7. A localização costeira pode oferecer vistas deslumbrantes, acesso a praias e atividades ao ar livre, além de um estilo de vida tranquilo. No entanto, a proximidade ao mar pode influenciar no preço do imóvel, então os compradores devem estar preparados para pagar um valor mais elevado nesses casos. É importante lembrar que a localização é um fator pessoal e depende das preferências de cada indivíduo ou família que está procurando comprar um imóvel.

Tendo estas informações, inicio o processo exploratório verificando a consistência dos dados do DataSet: Faço um IF para testar se existe valores nulos, aplicando `isna().any().any()` para o teste. O Resultado do teste foi negativo, não foram encontrados valores nulos.

A Preparação e Exploração dos Dados:

Na sequência busco os primeiros outliers observando o desvio padrão de cada atributo e comparo com média fornecida pelo `describe()`. Apesar de ter outliers, alguns deles podem ser justificados, embora não haja informações mais específicas sobre estas inconsistências. Verifiquei que todos os atributos estão preenchidos com valores corretos, como já citado, alguns com valor bastante acima da média (desvio padrão elevado), não vi necessidade de tratamento pois os mesmos acabam por não corresponder aos critérios estabelecidos. Além disto, mapeei as principais inconsistências e verifiquei através de filtros e gráficos se poderia ser ou não uma oportunidade, o resultado será visto mais adiante, neste relatório.

A Preparação e Exploração dos Dados:

1. Uso um método para buscar o nome da cidade pelo zipcode, para na sequencia, adicionar o campo cidade no dataframe house_data. Este método possibilita buscar uma cidade com maior vantagem para o cliente
2. Fiz a conversão do atributo date de objects para datetime, para ficar mais legível e poder operar com mais facilidade.
3. Coloquei o dataset house.cfg no meu repositório GIT, para ser chamado sempre que for usar o colab, evitando perder tempo fazendo upload do mesmo sempre que o colab parar. Caso desejem ver os dados com os atributos nomeado originalmente, deixei no google colaboratory, um link para meu git com o dataset original.
4. Com base na pesquisa que fiz sobre os interesses dos americanos na hora de comprar imóveis, decide procurar imóveis com o seguinte requisitos:
 - 1 - com vista para mar
 - 2 - com bela paisagem - 3 ou 4
 - 3 - Menor preço - entre 250000 e 1500000
 - 4 - Maior quantidade de quartos - entre 2 e 4

A Preparação e Exploração dos Dados:

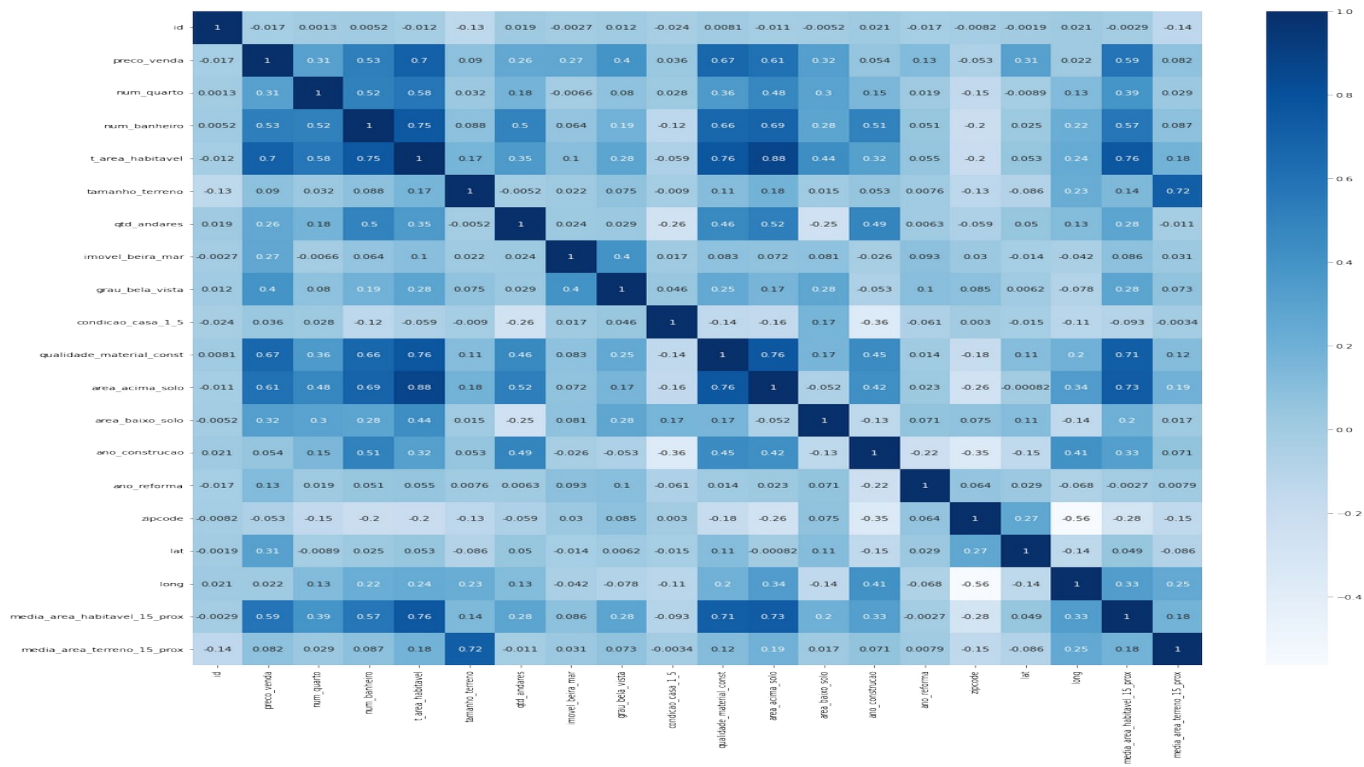
- 5 - maior quantidade de banheiro entr 1 e 3
- 6 - Maior área habitável
- 7 - Ano de Construção - preferência acima de 1980 se atender as outras condições
- 8 - Ano de Restauração* - Não é muito importante
- 9 - Quantidade de andares - entr 1 e 8
- 10- Estado de conservação - entre 3 e 5
- 11- Maior Avaliação de Material de construção - entre 10 e 13
- 12- Local com opções de lazer e segurança

Desta forma vou selecionar as principais colunas que devo buscar respostas para as minhas perguntas:

Quais os imóveis mais conservados, com quantidade confortável de quartos e banheiros com bela vista e próximo ao mar que representa o melhor custo benefício?

A Preparação e Exploração dos Dados:

5. Busco as principais correlações de variáveis que afetam o preço:



A Preparação e Exploração dos Dados:

Existe uma boa relação entre:

- preço de venda e a qualidade do material utilizado
- preço de venda e numero de banheiro
- preço de venda e tamanho de área habitável
- preço de venda e quantidade de quartos apresenta uma relação menor. - vou testar com o boxplot esta variável. Achei importante entender que não existe uma relação direta entre tamanho do terreno e o preço da venda.
- Existem outras ralações, mas não interessa para uma análise imediata do problema

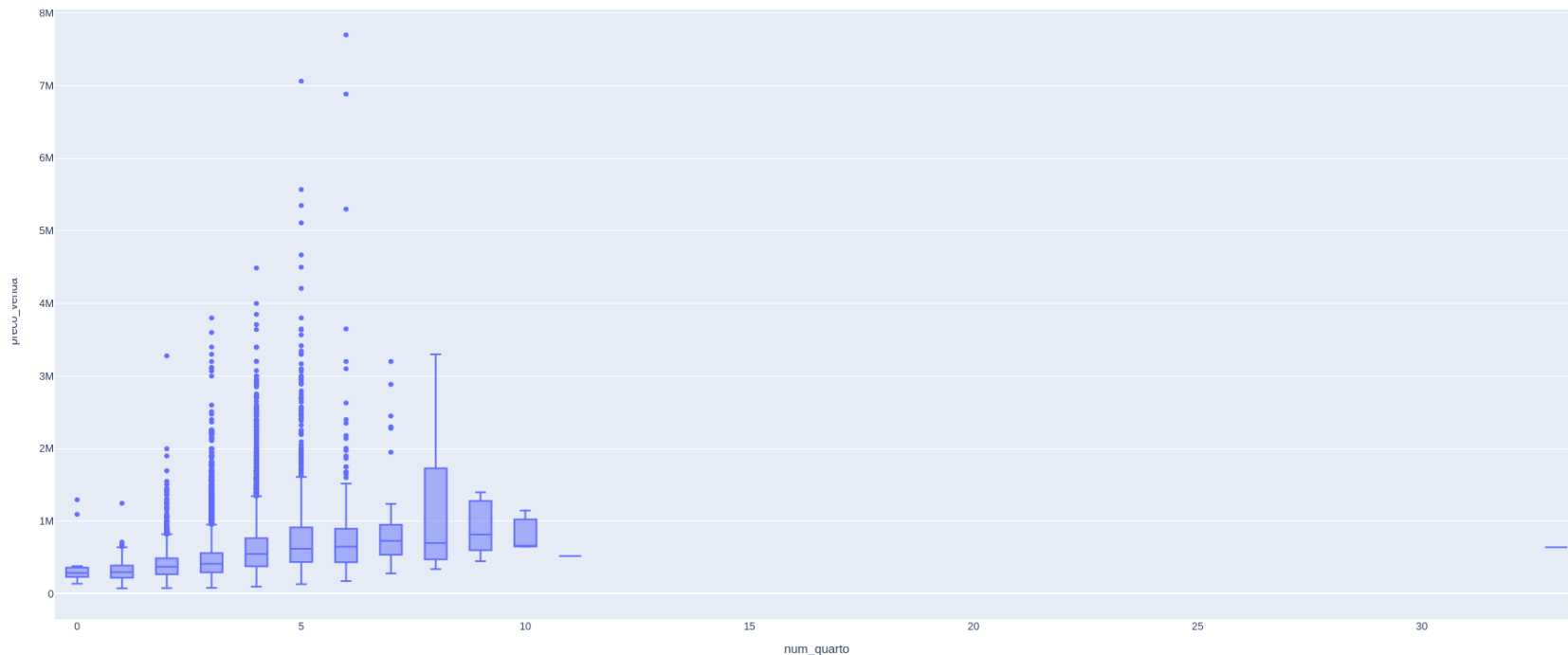
A Preparação e Exploração dos Dados:

6. Agrupei os dados utilizando as principais variáveis que afetam o preço de venda, e atribuo o valor a um novo dataframe chamado `agrupa_house`.

7. Busco compreender a relação entre a quantidade de quartos e preço tendo observado quantos banheiros possuem uma casa com determinada quantidade de quartos. Apesar da quantidade de quartos não dizer muito sobre o preço, este item é importante para quem buscar a comodidade de uma residência. No gráfico Box do plotly, é possível observar muitos pontos fora da curva, o que pode indicar outliers, mas o gráfico mostra muitos dados relevantes para a busca.

A Preparação e Exploração dos Dados:

Gráfico Boxplot Preço x Quantidade de quartos



A Preparação e Exploração dos Dados:

Neste Gráfico de caixas, fica claro que realmente não há uma relação direta entre preço de venda e quantidade de quartos. Há muita discrepância nestas duas variáveis. Tem casas com grande numero de quartos mais baratas do que outras que possuem um número bem menor, aparentemente pode representar alguma vantagem, mas precisei comparar com outra variáveis.

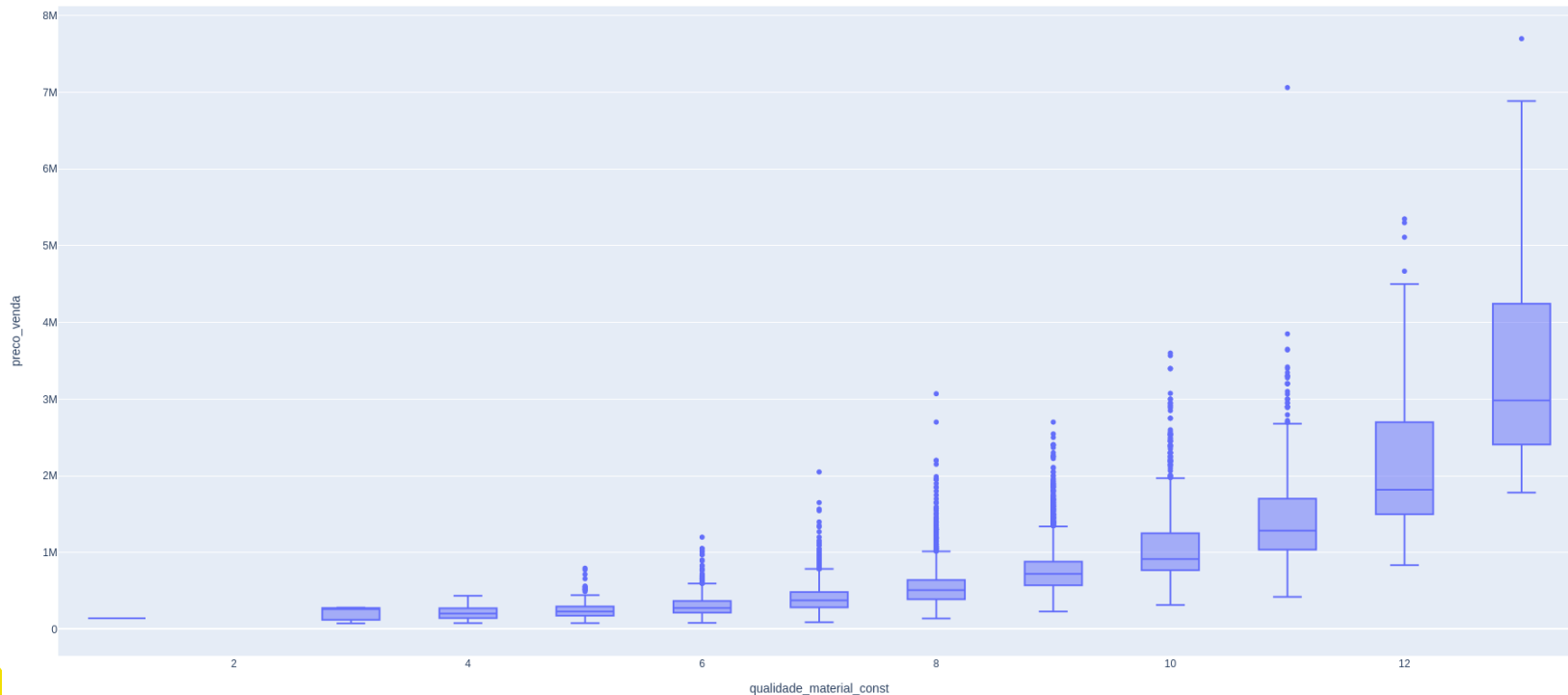
- Qualidade do Material usado na construção
- Número de Banheiros
- Quantidade de andar
- Proximidade com praias (Mar)
- Boa vista da casa
- Tamanho da área habitável
- Média de área habitável dos 15 imóveis mais próximos

A Preparação e Exploração dos Dados:

8. observando que uma casa com 33 quartos (descobri esta casa com auxílio do Box) com um tamanho habitável considerável, custando apenas 640000, chama atenção, por que foge astronomicamente da média, já que casa com área menor e outras variáveis parecidas com a referida entrada, custam muito mais que o valor observado. Na mesma observação percebe-se que a qualidade do material impacta no preço das casas, já que a qualidade do material é baixa (7 de 13), mas o estado da casa é 5, nos padrões americanos isso representa: Nível 5: A casa está em condições excepcionais e tem recursos de alto nível.
9. Após este achado, busquei fazer uma análise dos preços em relação à qualidade do material e o estado de conservação do imóvel, utilizando novamente o gráfico Box do plotly.

A Preparação e Exploração dos Dados:

Gráfico Boxplot Preço x Qualidade do Material usado na construção do Imóvel

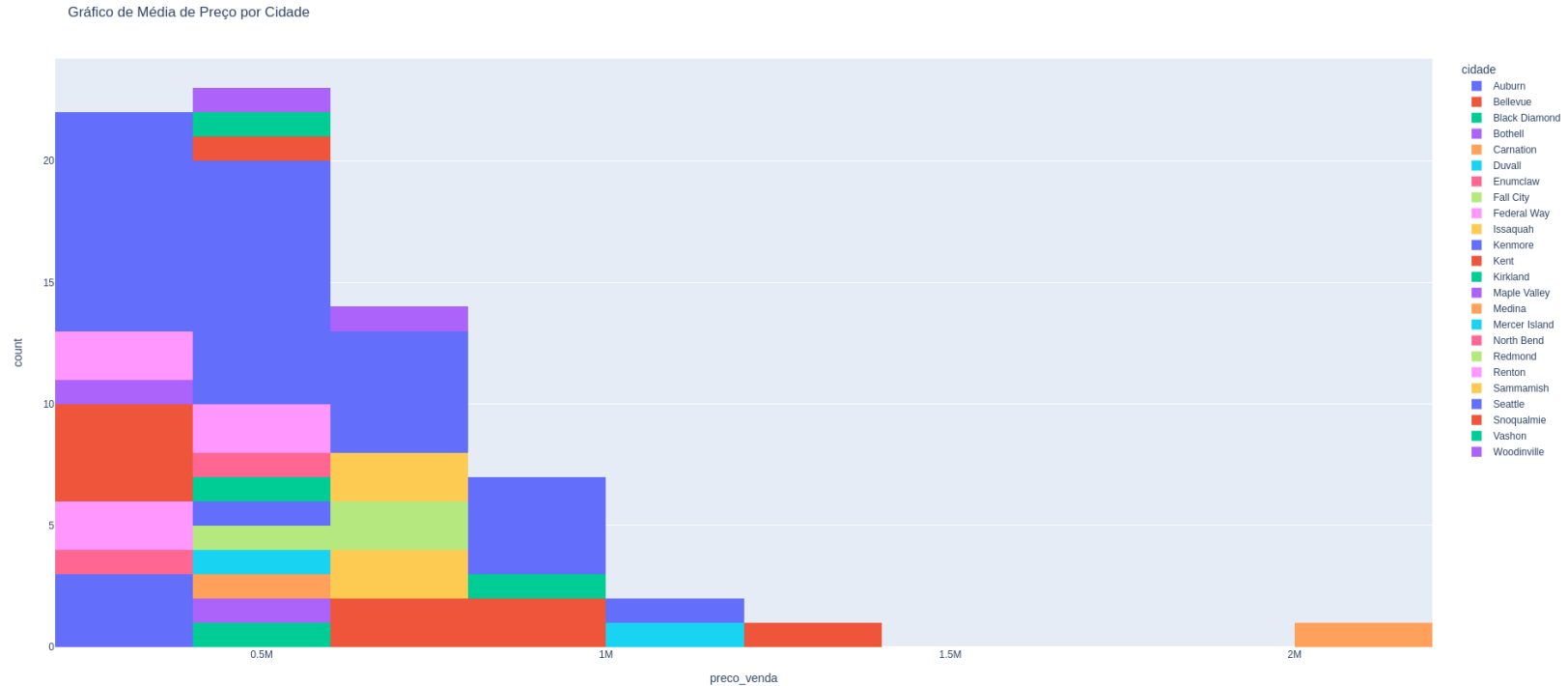


A Preparação e Exploração dos Dados:

Apesar de alguns valores estarem distantes da média geral (possui outliers), percebe-se que quanto maior a qualidade do material, maior é a avaliação do imóvel e maior o preço.

9. O histograma construído no plotly, facilitou a observação de concentração de imóveis pela média de preços em cada cidade do condado, optei pela média a princípio, pois me dá a possibilidade de comparar a média de preço por cidade.

A Preparação e Exploração dos Dados:



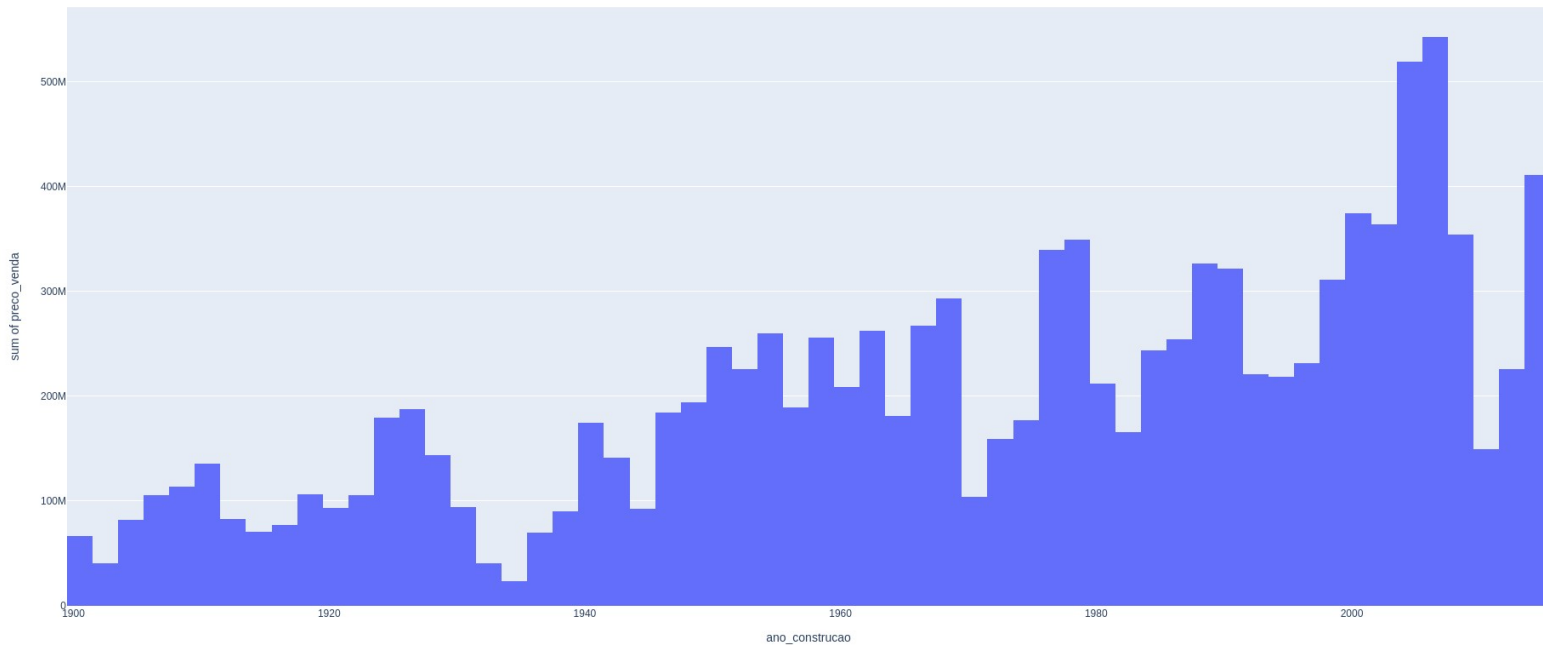
A Preparação e Exploração dos Dados:

10. Apartir deste histograma, tive uma melhor visualização da concentração do imóveis por cidade. Quais cidades oferecem melhores vantagens e possuem os imóveis dentro do critério de busca. Como pode ser visto no arquivo do colab, Iniciei a verificação por Seattle, mas depois inclui outras cidades de interesse baseado no mesmo critério.

11. Analisando preços com um histograma com base no ano de construção, com a finalidade de verificar se o ano de construção influencia nos preços dos imóveis.

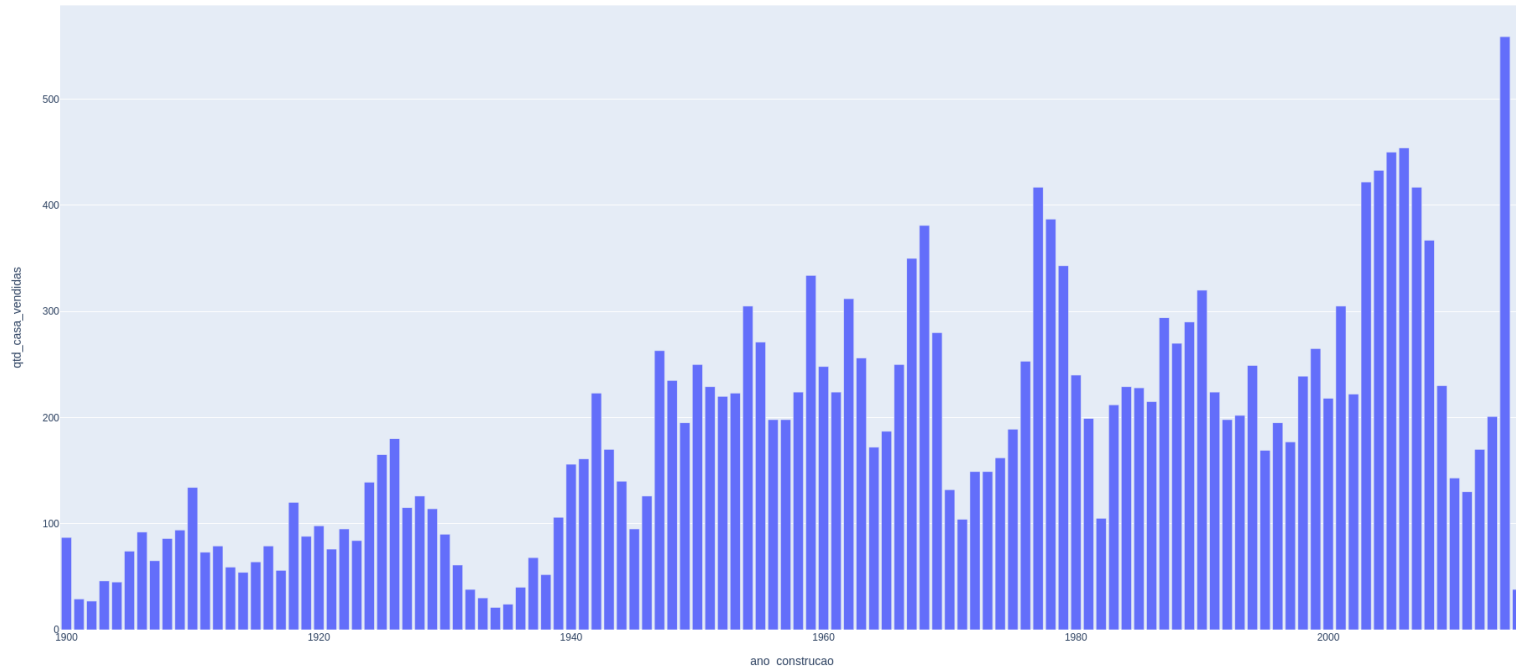
A Preparação e Exploração dos Dados:

Histograma de preços de Imóveis no Condado de Kings



A Preparação e Exploração dos Dados:

12. Agrupo os dados pelo ano de construção, buscando encontrar oportunidade em imóveis mais novos, mas sem descartar imóveis históricos com boa avaliação.



A Preparação e Exploração dos Dados:

O número de casas vendidas com ano de construção inferior a 1980 é relevante, desta forma, não coloquei restrição de busca por ano, mas observei com mais atenção os detalhes dos imóveis cujo o período de construção variam de 1920 a 1979. Criei duas variáveis para armazenar os valores.

Quantidade de casas vendidas com ano de construção até 1979: 12334

Quantidade de casas vendidas com ano de construção a partir de 1980: 9279

13. Com base em dados do interesse de americanos em imóveis, desenvolvi um filtro para buscar oportunidades localizadas nas análises acima.

- Media de área habitável < 6000
- Tamanho de área habitável < 6000
- Numero de banheiro > 1

A Preparação e Exploração dos Dados:

Conservação da casa > 3

Preço de venda < 2000000

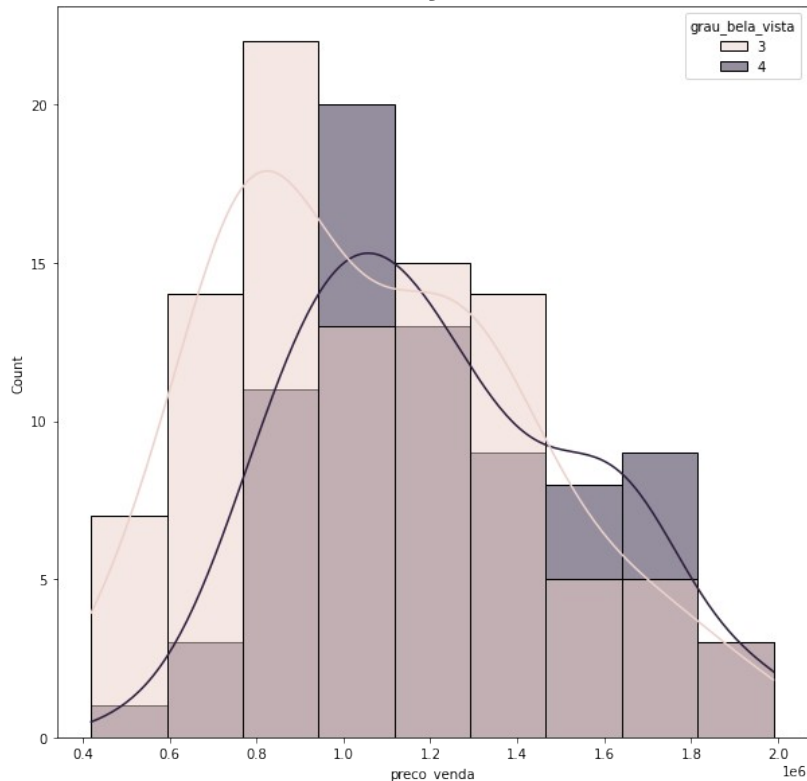
Bela vista do imóvel > 2

E a busca nas cidades de Seattle, Kirkland, Bellevue e Redmond. Outras cidades entraram no filtro porém, a localização e dificuldade acesso de algumas delas impediram que pudessem estar presente na valiação final dos imóveis.

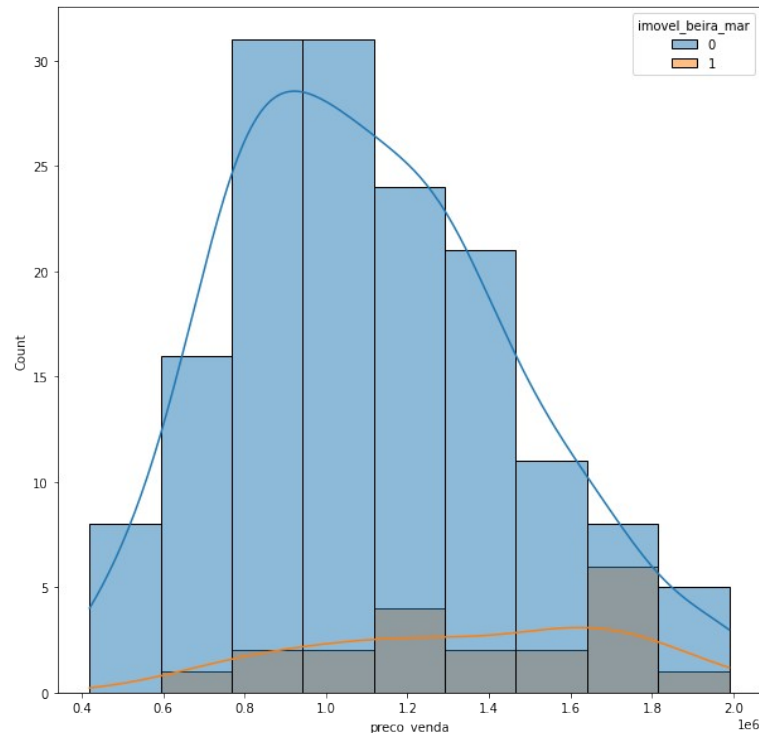
14. O filtro é atribuído à variável `house_filter`. Trabalho com esta variável no restante das análises. Uso o seaborn para construir histogramas mais refinado, buscando entre os imóveis filtrados, aqueles que estejam numa localização beira mar e outro com localização com bela vista, em cidades que tem muitos atrativos.

A Preparação e Exploração dos Dados:

Histograma Preço x Bela Vista

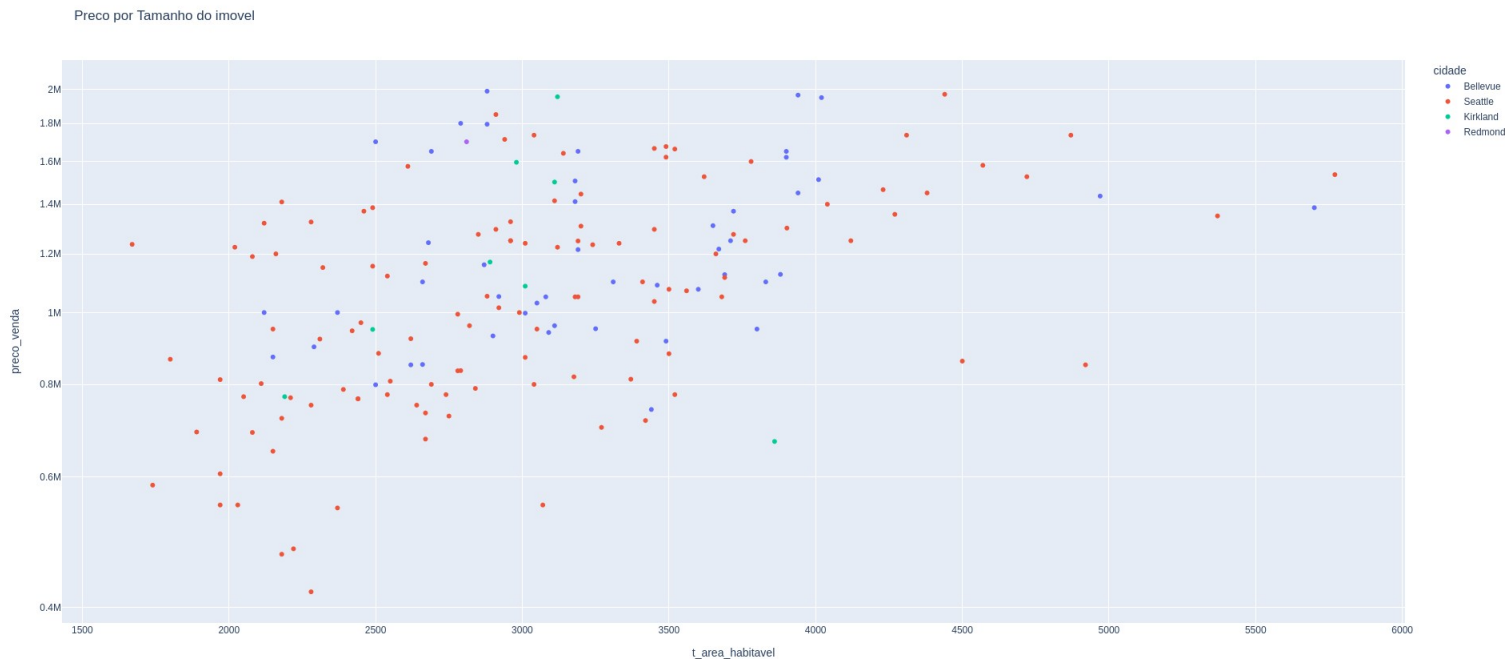


Histograma Preço x Beira Mar



A Preparação e Exploração dos Dados:

15. Criei um gráfico de dispersão para verificar a relação entre preço e tamanho do imóvel e suas tendências.



A Preparação e Exploração dos Dados:

O gráfico de dispersão só revelou o que já havia sido observado nos outros gráficos depois do filtro. Mas alguns pontos fora da curva, me ajudaram a encontrar boas oportunidades. Ainda analisei sob a ótica de mais dois gráficos que podem ser vistos no arquivo do colabore. Um que demonstra a quantidade de imóveis vendidos em cada cidade filtrada e um outro boxplot, para verificar a média e a mediana dentre as oportunidades correspondentes ao filtro.

Por fim eu analiso o preço por metro de área habitável e adiciono ao dataset dois novos campos a fim de testar os valores que encontrei nas análises anteriores para confirma ou refutar as vantagens percebidas por mim nos 5 imóveis escolhidos. Os dois campos são:

- `preco_metro`
- Duplicado

Esclarecendo que o campo se chama duplicado, mas não porque os dados estejam inconsistentes, pois os Ids das residências se repetem por ter havido vendas de alguns deles mais de uma vez em um dado período. Com isso conclui que realmente os imóveis que indico, são os mais vantajosos encontrados.

Análise final + Sugestão de imóveis

Ao final, com todas as variáveis devidamente exploradas e verificadas, conclui-se que o melhor investimento para a Roof, são casa mais populares e dentro das características de preferência dos estadunidenses. Tais características já foram elencadas, mas em resumo são: preço, qualidade do material empregada na construção, comodidade com quartos e banheiros, espaço para acomodar objetos e outros, inclusive algumas das residências possuem sótão, eles preferem regiões com opção de trabalho, segurança e lazer. Também há preferência por casas bem conservadas, que não precisem de reformas profundas ou outras intervenções. Neste sentido, os imóveis escolhidos não precisaram passar por obras que venham onerar a Roof, estão localizadas em regiões que podem valorizar os imóveis, pois existem outros imóveis com preços mais altos e com menor qualidade, além de se situar em áreas consideradas nobres. As duas cidades escolhidas são promissoras, a primeira, Seattle, é populosa e possui todos os requintes de uma cidade moderna e em desenvolvimento, já Bellevue está em crescimento, apresenta grandes possibilidades de valorização e é próxima a Seattle, perfeito para quem busca tranquilidade.

Análise final + Sugestão de imóveis

OS CINCO MELHORES IMÓVEIS PARA INVESTIR

ID	CIDADE	DATA/ VENDA	DATA CONSTRUÇÃO	PREÇO VENDA
7567600030	Seattle	17/01/2015	1954	750.000,00
2623039082	Seattle	18/02/2015	1930	770.000,00
7856640460	Bellevue	18/12/2014	1986	950.000,00
7805450560	Bellevue	20/08/2014	1984	960.000,00
4139430250	Bellevue	30/03/2015	1992	1.435.000,00

Análise final + Sugestão de imóveis

Já os imóveis listados abaixo, considero péssimo investimento devido serem imóveis de construção antiga, a condição das casas estão em media de 2 em uma escala que vai até 5, os imóveis não possuem uma boa vista e a qualidade do material para suas construções é ruim. Os preços desses imóveis não se justificam, estão acima do preço médio de construções melhores avaliadas. Para se ter uma ideia da importância do nível de condição do imóvel, tem que entender que cada nível da escala é geralmente definido da seguinte forma: **Nível 1:** A casa precisa de reparos significativos, reformas extensas e/ou tem problemas estruturais graves. **Nível 2:** A casa está em condições abaixo da média, com a necessidade de várias reformas e reparos, mas ainda habitável.

Nível 3: A casa está em boas condições, mas pode precisar de algumas atualizações e reparos menores. **Nível 4:** A casa está em excelentes condições e pronta para morar, sem necessidade de reparos ou atualizações significativas. **Nível 5:** A casa está em condições excepcionais e tem recursos de alto nível, como acabamentos de luxo, equipamentos modernos e/ou design personalizado.

Análise final + Sugestão de imóveis

OS CINCO PIORES IMÓVEIS PARA INVESTIR

ID	CIDADE	DATA/ VENDA	DATA CONSTRUÇÃO	PREÇO VENDA
2767604170	Seattle	17/01/2015	1900	975.000,00
3331001995	Seattle	18/02/2015	1921	509.990,00
6205500030	Bellevue	18/12/2014	1950	480.000,00
1498302783	Seattle	20/08/2014	1906	333.000,00
1900000060	Bellevue	30/03/2015	1925	313.000,00

Relatório elaborado por:

Valdir Nascimento dos Santos

RDI – 97851

03/03/2023

Link do colab:

<https://colab.research.google.com/drive/1z9T5zhL6yk9kiWDGIWCfDsB4csd90PQT?usp=sharing>