

## 1. A QUESTÃO DO NEGÓCIO

### 1.1 Qual a minha meta?

A empresa que me contratou, tem como objetivo ampliar sua atuação e se consolidar na concorrência internacional em seu ramo imobiliário.

Minha meta, com os dados fornecidos, é realizar uma avaliação de aplicação estratégica para iniciar essa nova jornada de extrema importância para a Roof Imóveis, e realizar o levantamento dos 5 imóveis que ela deveria investir, esclarecendo as razões pelas quais esses são os escolhidos, bem como os cinco que não deveria investir.

Para isso, desenvolvi a ideia inicial como **plano** de fundo, a qual se trata de buscar e detectar quais os imóveis com características valorizadas e explorar suas maiores chances **de valorização após uma revenda.**

### 1.2 Como posso chegar?

Utilizando-me do dataset empenhado, o caminho a ser encontrado é aquele onde haja a oportunidade de valorização sobre o imóvel comprado pela empresa em sua revenda. A lógica abordada foi, primeiramente, verificar os preços dos metros quadrados de cada imóvel, e posteriormente verificar o mesmo valor, porém agregado de cada **região(zipcode), buscando uma flutuabilidade no preço que pudesse indicar uma oportunidade de compra com uma cotação abaixo do mercado regional.**

**Entretanto, somente isto não seria o suficiente para segurança e maior credibilidade no investimento, então localizei os imóveis com base neste primeiro filtro, e logo após, desenvolvi um segundo retém, com o objetivo de verificar dentre esses, quais tinham valorizado mais com o tempo desde sua construção até sua venda, e quais seus padrões na construção, para que, quando comparado, verificasse as correlações de preços altos com os itens mais atrativos, e então indicasse os que mais valorizaram.**

(Ref. Colab: Grupamento 1 a 7)

## 2. O ENTENDIMENTO DO NEGÓCIO

### 2.1 Quais dados eu tenho disponível?

Os dados disponibilizados para a realização dessas análises são **os registros oficiais** de imóveis na região de King County, Estado de Washington, seu preço, data de venda, data de construção, condições, dentre outros.

Verifiquei seus formatos bem como observei as necessidades iniciais de tratamento dos dados.

(Ref. Colab: Grupamento 1)

### 2.2 Eles são relevantes para o problema?

**Alguns dados se tornam importantes de acordo com a estratégia escolhida, outros, porém, se tornam irrelevantes por não constituírem importância para o alvo do estudo.**

**Os mais importantes, na primeira análise, eram aqueles que demonstraram melhor correlação com os preços, indicando um possível motivo de valorização aprimorada do imóvel.**

(Ref. Colab: Grupamento 2)

### 2.3 Eles me trazem uma solução direta?

Não, é necessário um tratamento e filtragem dos dados escolhidos para que tragam as informações e levem ao conhecimento necessário para a tomada de decisão mais assertiva.

Neste caso, realizei a correção das datas, e corriji os banheiros inicialmente.  
(Ref. Colab: Grupamento 1)

## 2.4 Qual o meu setor ou range de negócio?

O setor imobiliário, e no caso deste serviço em específico, uma empresa nacional que está tomando as primeiras iniciativas visando a ampliação para concorrência internacional.

## 3. A COLETA DE DADOS

### 3. 1 Os dados que eu tenho fazem sentido?

Após uma avaliação criteriosa e separação dos dados mais relevantes, notei que algumas informações possuem sentidos conectados que podem trazer insights e demonstram correlações de importância para a análise proposta, inferindo com nível de 0.5 a 0.7 de correlação com os preços dos imóveis.

Sendo assim, alguns itens nitidamente interferem frontalmente com os preços, elevando o valor dos imóveis.

(Ref. Colab: Grupamento 2)

### 3. 2 Estão no formato que eu gostaria?

A maioria sim, entretanto, alguns não estão, os quais eu tratei para que ficassem conforme o necessário para os devidos cálculos e funções.

(Ref. Colab: Grupamento 1)

### 3. 3 O que mais eu consigo obter de informações destes dados?

Com os dados, obtive a correlação dos preços com as características, bem como sua valorização, verifiquei também que a maioria dos imóveis a beira-mar estão com preços altíssimos comparados com a média e a mediana da região, claramente por sua posição privilegiada.

(Ref. Colab: Grupamento 1)

## 4. A LIMPEZA DE DADOS

### 4. 1 Remoção de dados outliers e valores ausentes.

Foram utilizados diversos filtros e correções ao longo do projeto, para adequá-los ao resultado que eu pretendia chegar. Todos mais detalhados no próprio colab.

No 'Grupamento 1' (do colab) verifiquei se haviam valores nulos, e não haviam. Observei também a quantidade de cada cômodo, notei a estranha casa com '33 quartos', um possível erro de digitação, entretanto, irrelevante nos resultados.

De todos, logo consertei as datas que estavam em formato incorreto e incompreensível, colocando-as em sua forma correta e posteriormente filtrando somente para o ano, que é o que interessava para a minha análise. Alterei os banheiros para números inteiros, facilitando as comparações entre as características, verificando a correlação junto aos preços.

Como optei por seguir uma linha de indicação de imóveis baseada na valorização ano a ano de cada um, alguns imóveis que possuíam menos de um ano ou um ano de construção, representaram desvios que atrapalhavam as análises, os quais foram filtrados a fim de melhorar a percepção das oportunidades, desconsiderando objetos que não se enquadram nas medidas de comparação da valorização ao longo do tempo.

(Ref. Colab: Grupamento 1, Grupamento 3, Grupamento 4)

## 5. A EXPLORAÇÃO DE DADOS

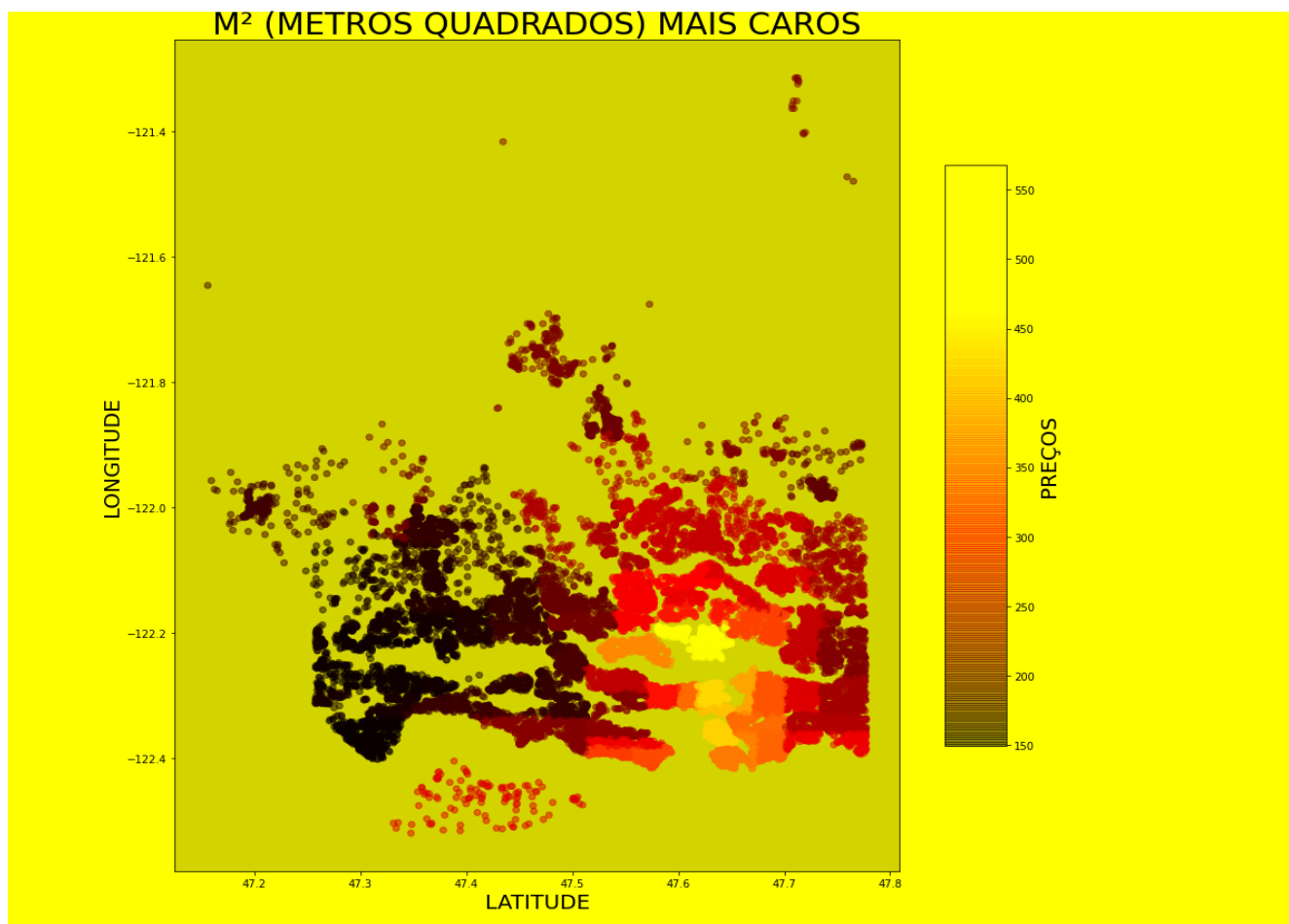
### 5. 1 Visualização dos dados.

Para apoiar as conclusões e visualizar a percepção coletada dos dados, realizei alguns gráficos, os quais estão no colab também.

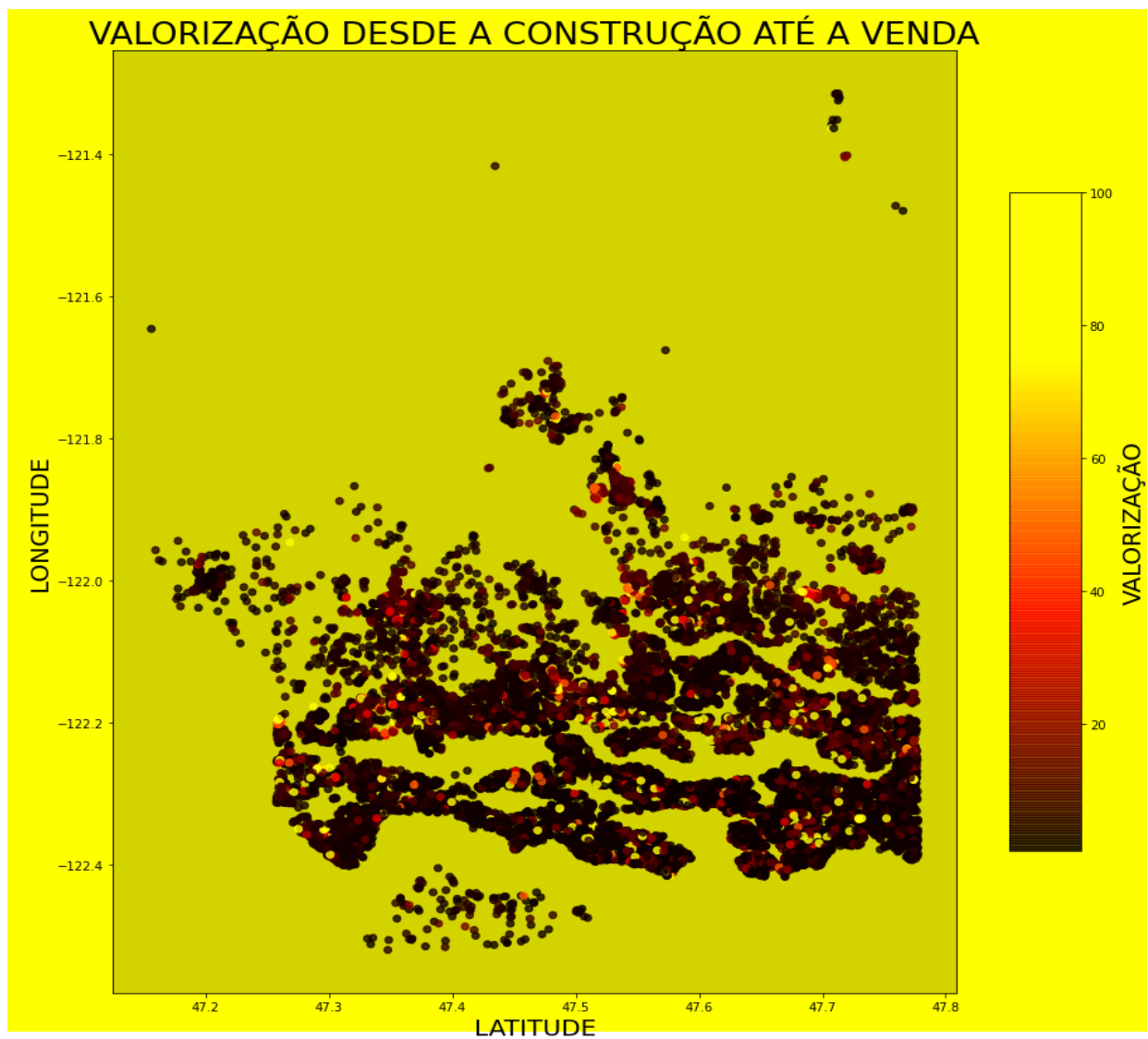
O primeiro mostra os imóveis mais caros com cores mais claras, regredindo para os mais baratos, notamos uma condensação dos imóveis caros, entretanto, não necessariamente os melhores a investir, pois muitos podem estar acima do preço da região, e se fossem indicados, poderiam gerar prejuízo na revenda.

O segundo gráfico mostra a valorização dos imóveis ao longo do tempo, e vemos que existem ótimas valorizações fora da zona mais cara, aparentando boas oportunidades espalhadas por toda região.

O terceiro destaca os imóveis que estão com seus preços acima do preço médio do metro quadrado da região. (Ref. Colab: Grupamento 5)

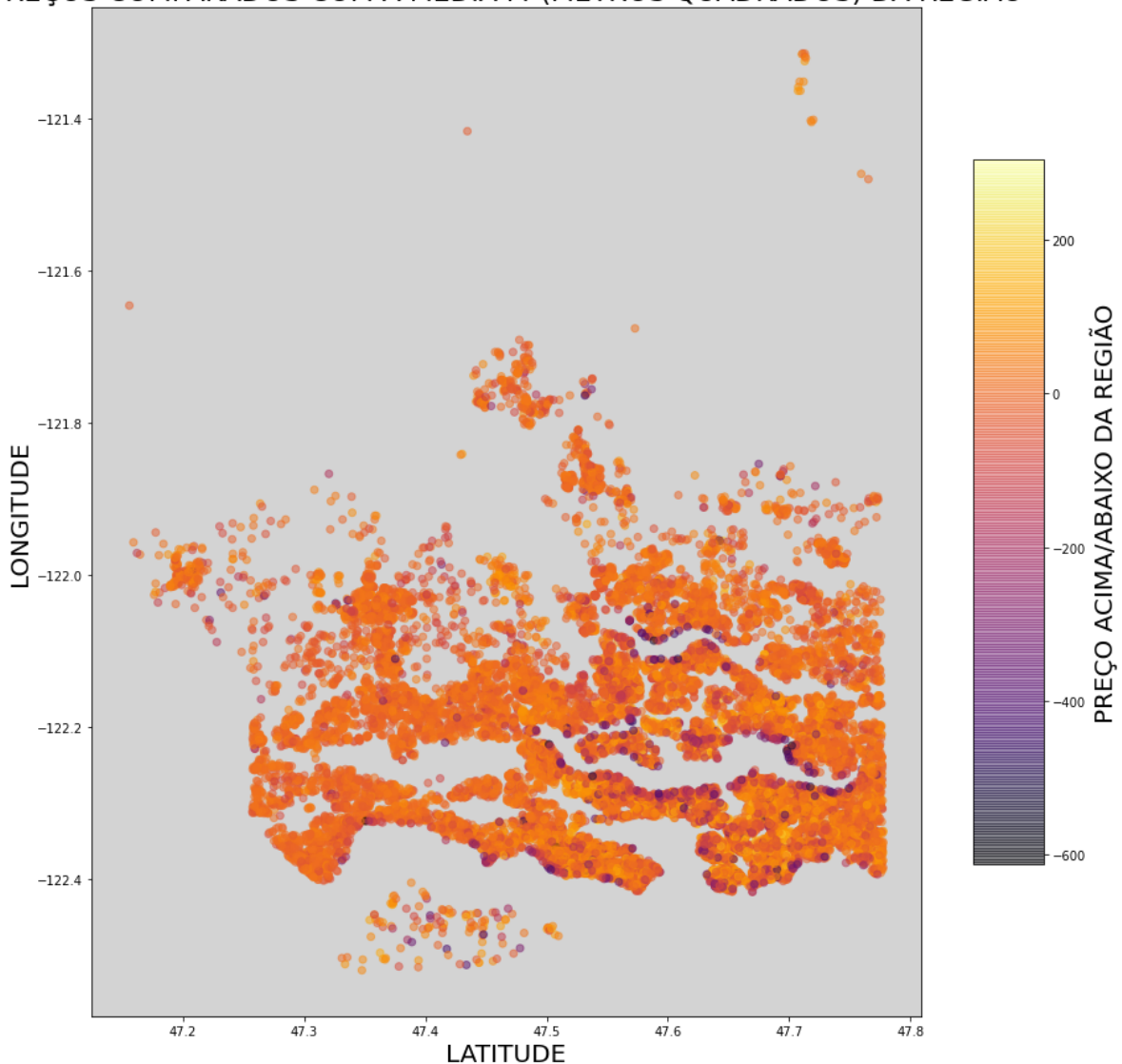


Os imóveis mais caros são destacados em amarelo, com efeito degradê para o preto, conforme os preços ficam menores.



Os imóveis que mais valorizaram ao ano, de sua construção até venda, notamos uma boa quantidade de valorização em torno dos 40% ao ano, espalhados por toda a região observada.

## PREÇOS COMPARADOS COM A MÉDIA M<sup>2</sup>(METROS QUADRADOS) DA REGIÃO



Os imóveis com os preços muito acima da média, destacados em roxo escuro, nitidamente são os que estão com visão para as águas.

### Análise Final + Sugestão de imóveis.

Com o objetivo definido, a linha de pensamento seguida foi buscar com os conhecimentos desenvolvidos até esse momento do curso, as seguintes informações:

- 1- Quais características mais possuem correlação com os preços mais altos, valorizando o imóvel.
- 2- Qual o preço do metro quadrado das regiões, agrupadas por zipcode.
- 3- Quais imóveis possuem o metro quadrado mais barato comparados ao preço de sua própria região.

Avaliando os dados, após filtrar todas essas informações com base no escopo estratégico mencionado, buscando atingir maior chance de valorização ao longo do tempo, os imóveis que se enquadram nesses requisitos filtrados são os destacados abaixo, todos com quantidades consideráveis das características que mais possuem correlação com os

preços mais altos e portanto valorizados, as quais são: área habitável, quantidade de banheiros e grade(qualidade do material utilizado), entretanto, com preço do metro habitável individual abaixo do custo regional, sugerindo a possibilidade de compra abaixo do preço do mercado, sendo então indicados como as melhores oportunidades para revenda com lucro.

Os imóveis selecionados são:

	sqft_living	sqft_living15	yr_built	bathrooms	grade	zipcode	lat	long	date	price	qtd_anos	valor_cresc	price_m2	mean_zip	Valoriz_y	dist_price_mean
14835	2370	1590	1987	2.0	9	98119	47.6503	-122.366	2015	300523.0	28	10732.964286	126.802954	432.256998	3.571429	305.454045
3844	1540	1220	2007	4.0	7	98109	47.5420	-122.322	2014	216650.0	7	30950.000000	140.681818	433.401195	14.285714	292.719377
11436	1480	1140	2007	4.0	7	98109	47.5421	-122.322	2014	216650.0	7	30950.000000	146.385135	433.401195	14.285714	287.016060
16931	3990	2090	1962	2.0	8	98112	47.6236	-122.299	2014	610000.0	52	11730.769231	152.882206	438.637872	1.923077	285.755667
9417	2240	1470	1954	2.0	7	98112	47.6191	-122.296	2015	375000.0	61	6147.540984	167.410714	438.637872	1.639344	271.227158

Os quais possuem as seguintes ID:

O imóvel na posição 14835 é o id: 3613600150  
O imóvel na posição 3844 é o id: 2734100734  
O imóvel na posição 11436 é o id: 2734100732  
O imóvel na posição 16931 é o id: 5016002275  
O imóvel na posição 9417 é o id: 9828701565

Estes são os imóveis que mais oferecem custo x benefício, com alta taxa de valorização anual, e preço abaixo da média do preço do metro quadrado de sua região, portanto, os melhores para um investimento.

Já os que eu não indicaria de forma alguma, são o oposto dos princípios dos selecionados, ou seja, os que possuem valorização anual baixa e custo do metro quadrado muito acima do preço médio da região. Estes são:

	sqft_living	sqft_living15	yr_built	bathrooms	grade	zipcode	lat	long	date	price	qtd_anos	valor_cresc	price_m2	mean_zip	Valoriz_y	dist_price_mean
18597	720	1120	1905	1.0	6	98198	47.3509	-122.323	2015	570000.0	110	5181.818182	791.666667	178.428610	0.909091	-613.238057
8623	410	1190	1953	1.0	4	98146	47.5077	-122.357	2015	325000.0	62	5241.935484	792.682927	225.491431	1.612903	-567.191496
4013	3000	3000	2001	2.0	11	98075	47.5884	-122.086	2014	2400000.0	13	184615.384615	800.000000	268.639846	7.692308	-531.360154
1081	1150	2440	1908	1.0	6	98178	47.5020	-122.222	2014	825000.0	106	7783.018868	717.391304	189.172528	0.943396	-528.218776
656	3930	2730	1957	2.0	8	98034	47.7022	-122.224	2014	3070000.0	57	53859.649123	781.170483	265.952248	1.754386	-515.218235

Suas IDs:

O imóvel na posição 18597 é o id: 3222049087  
O imóvel na posição 8623 é o id: 6303400395  
O imóvel na posição 4013 é o id: 724069059  
O imóvel na posição 1081 é o id: 7567600045  
O imóvel na posição 656 é o id: 3760500116

Esses não seriam bons negócios, por já possuírem seus preços acima da média da região, e sua valorização é baixa se comparada ao dos demais.

Segue link do colab:

<https://colab.research.google.com/drive/1loeIClfYkrE2O2dTJIBsvd9GNZv7xDIh?usp=sharing>