

Introdução

Nesse artigo temos a análise dos dados coletados pela API da Zillow. Como a Zillow não possui mais uma API oficial, foi utilizada uma API feita por terceiros para a coleta de dados. Zillow (<https://www.zillow.com>) é uma empresa focada na venda, compra e aluguel de propriedades (casas, terrenos, condomínios e etc) nos EUA com mais de 68 milhões de visitantes em 2021. Ela possui dados de milhares de propriedades espalhadas por todo território americano. Dentre esses dados, temos: quantidade de banheiros, quartos, área do terreno, preço do imóvel dentre outros.

Com os dados obtidos, vamos fazer uma análise de quais tipos de imóveis são mais anunciados, relação entre preço e os tipos de propriedades, preço com a quantidade de banheiros e quartos em cada tipo de propriedade e mais.

Toda análise foi feita na linguagem de programação Python.

Objetivo

O objetivo desse artigo é utilizar a API da Zillow para coletar os dados e realizar uma análise deles.

API

A API utilizada foi a Zillow (<https://rapidapi.com/s.mahmoud97/api/zillow56/>) do site Rapidapi. Nela podemos coletar dados como cidade, tipo de propriedade, preço, área do terreno e etc. Um ponto importante de ressaltar é que são disponibilizadas apenas 30 chamadas gratuitas da API.

Dados

Os dados coletados foram:

- Bathrooms: número de banheiros do imóvel
- Bedrooms: número de quartos do imóvel
- City: cidade
- Country: país
- homeStatus: se a propriedades está à venda, para aluguel
- homeType: se o imóvel é para uma única família (**SINGLE_FAMILY**), condomínio (**CONDO**), terreno (**LOT**), **townhouse** (casas separadas, mas que compartilham uma parede), multifamilia (**MULTI_FAMILY**), manufaturada (**MANUFACTURED**) (casas feitas em fabricas sob medida).
- isNonOwnerOccupied: se o morador está morando na propriedade anunciada
- isPre foreclosure Auction: quando algum pagamento do imóvel está atrasado (ex: hipoteca)
- livingArea: área construída
- lotAreaValue: área total do terreno

- lotAreaUnit: unidade de medida do terreno (acres, sqft)
- price: preço do imóvel
- rentZestimate: aluguel estimado pelo site
- zestimate: preço da propriedade estimado pelo site

Nem todos os dados coletados serão utilizados na análise, pois alguns contêm, por exemplo, apenas a mesma informação numa mesma coluna.

Uma informação que poderia agregar na análise, mas não possuímos é a idade da construção. Não temos esse dado, pois seria necessária uma chamada na API para cada imóvel e, como são disponibilizadas apenas 30 chamadas gratuitas, não foi possível obtê-la.

Análise

Figura 1 – Dados

bathrooms	bedrooms	city	country	homeStatus	homeType	isNonOwnerOccupied	isPreforeclosureAuction	livingArea	lotAreaUnit	lotAreaValue	price
0.0	0.0	Orlando	USA	FOR_SALE	LOT	True	False	0.0	acres	1.00	275000.0
1.0	2.0	Orlando	USA	FOR_SALE	CONDO	True	False	810.0	acres	0.29	170000.0
2.0	2.0	Orlando	USA	FOR_SALE	SINGLE_FAMILY	True	False	976.0	sqft	10454.40	305000.0
2.0	4.0	Orlando	USA	FOR_SALE	SINGLE_FAMILY	True	False	1200.0	sqft	8276.40	285000.0
2.0	2.0	Orlando	USA	FOR_SALE	SINGLE_FAMILY	True	False	1117.0	sqft	5227.20	270000.0
...
3.0	3.0	Orlando	USA	FOR_SALE	TOWNHOUSE	True	False	1594.0	sqft	2178.00	349990.0
3.0	3.0	Orlando	USA	FOR_SALE	CONDO	True	False	1581.0	sqft	871.20	369500.0
0.0	0.0	Orlando	USA	FOR_SALE	LOT	True	False	0.0	sqft	5662.80	499000.0
2.0	4.0	Orlando	USA	FOR_SALE	SINGLE_FAMILY	True	False	1868.0	acres	0.29	469900.0
4.0	6.0	Orlando	USA	FOR_SALE	SINGLE_FAMILY	True	False	3142.0	sqft	7840.80	789990.0

Na figura 1, temos todos os dados que foram coletados. Todos são de propriedades/imóveis na cidade da Florida nos Estados Unidos. As principais colunas que vamos usar na análise são: bathrooms, bedrooms, homeType, livingArea, lotAreaValue e price. As restantes foram descartadas, pois não possuíam informações relevantes, por exemplo, os valores de isNonOwnerOccupied são todos TRUE nas 600 linhas.

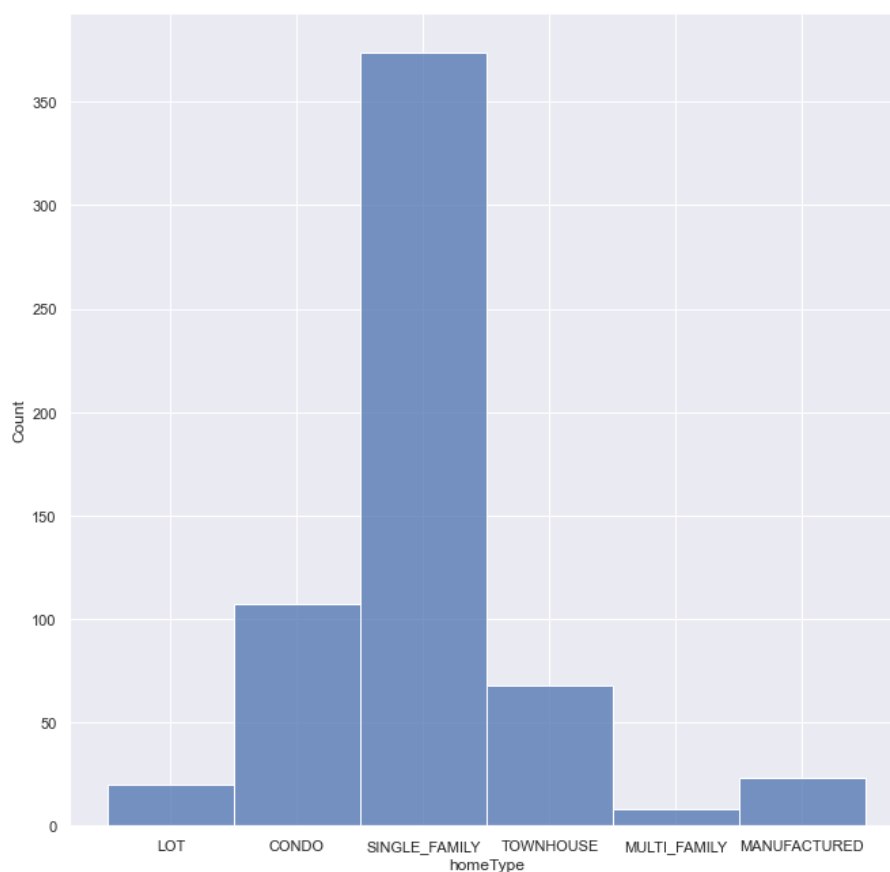
Como a unidade de medida para área nos dados estão em acres e sqft, vamos convertê-los para m² para termos uma ideia melhor do tamanho dos terrenos/imóveis. Vamos também fazer um pouco de feature engineering, calculando uma coluna para preço por metro quadrado (price/m²) e a somatória do número de banheiros e quartos (bathPlusBed). Essas duas novas colunas vão ajudar na análise.

Figura 2 – Dados limpos

	bathrooms	bedrooms	bathPlusBed	homeType	lotAreaValue_m2	livingArea_m2	rentZestimate	zestimate	price/m2	price
0	0.0	0.0	0.0	LOT	4046.8	0.0	1895.0	0.0	68.0	275000.0
1	2.0	2.0	4.0	SINGLE_FAMILY	971.2	90.7	1500.0	256000.0	314.0	305000.0
2	2.0	4.0	6.0	SINGLE_FAMILY	768.9	111.5	0.0	0.0	371.0	285000.0
3	2.0	2.0	4.0	SINGLE_FAMILY	485.6	103.8	1800.0	269800.0	556.0	270000.0
4	2.0	2.0	4.0	SINGLE_FAMILY	566.6	107.8	1900.0	245000.0	432.0	245000.0
...
489	2.0	2.0	4.0	CONDO	485.6	116.9	1898.0	226024.0	474.0	229990.0
490	3.0	3.0	6.0	TOWNHOUSE	202.3	148.1	0.0	0.0	1730.0	349990.0
491	3.0	3.0	6.0	CONDO	80.9	146.9	0.0	0.0	4567.0	369500.0
492	0.0	0.0	0.0	LOT	526.1	0.0	2499.0	0.0	948.0	499000.0
493	4.0	6.0	10.0	SINGLE_FAMILY	728.4	291.9	0.0	0.0	1085.0	789990.0

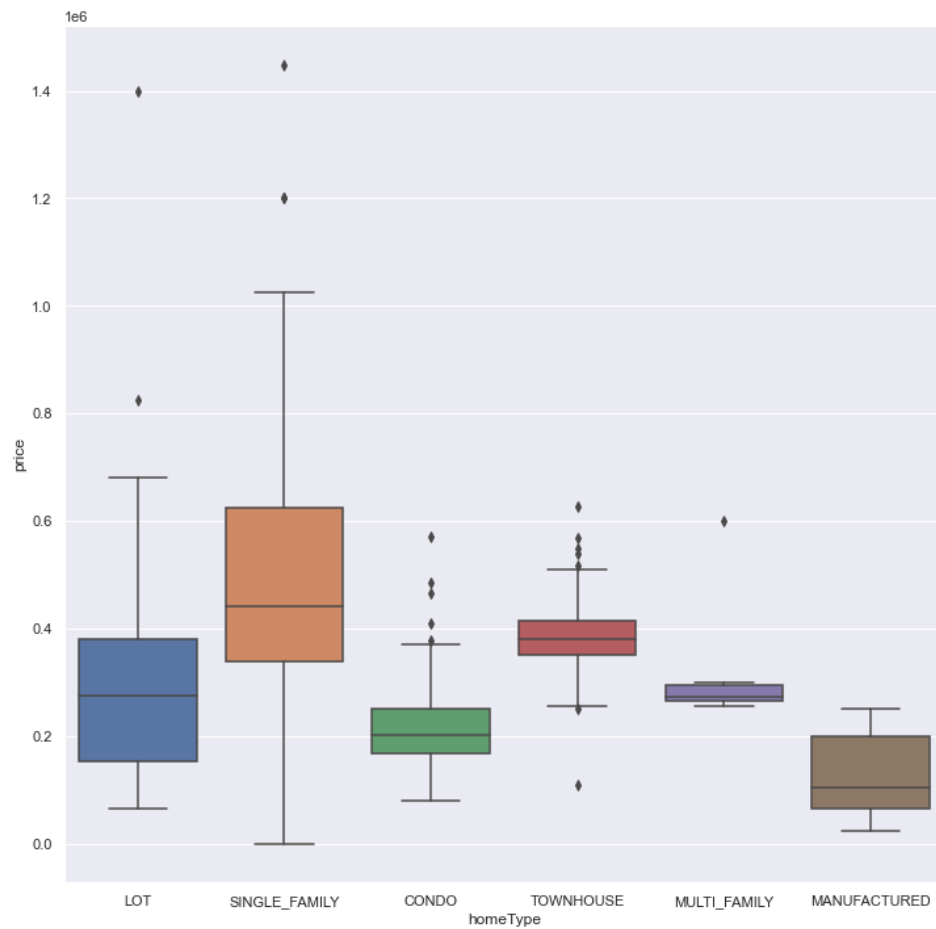
Com os dados limpos, vistos na figura 2, podemos começar o estudo.

Figura 3 – Histograma de homeType



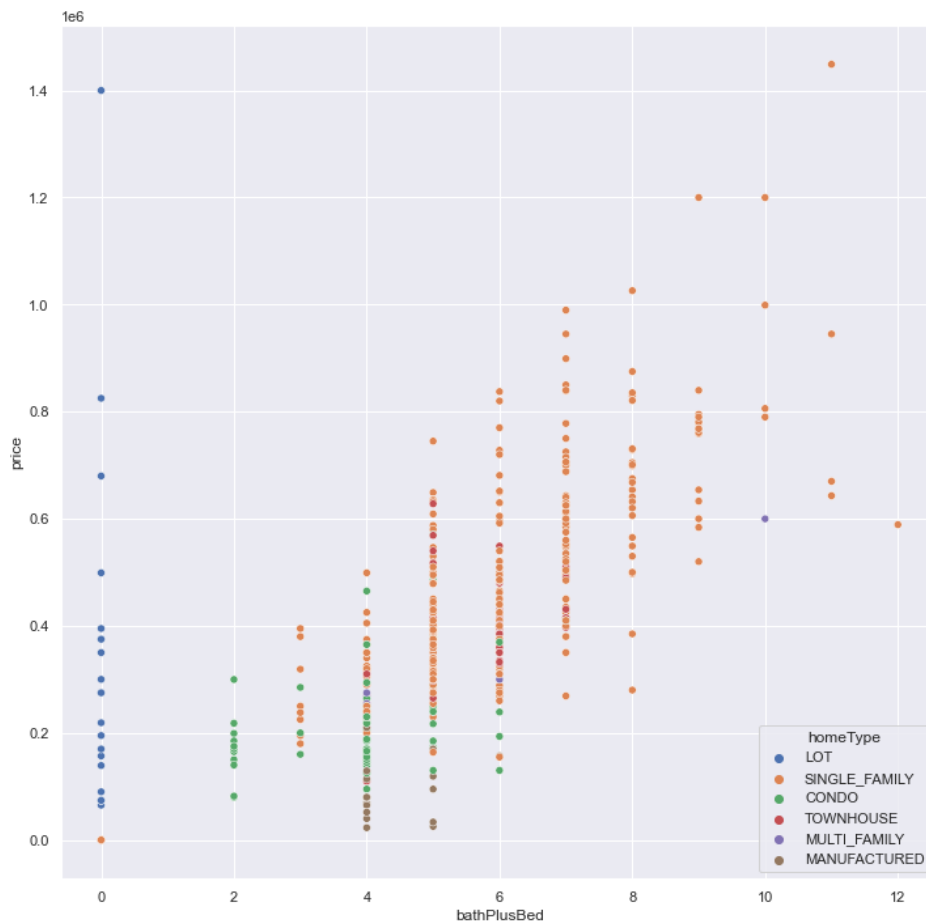
Na figura 3 temos o histograma de homeType e vemos que, aproximadamente, 62,5 % dos anúncios do site são de casas para uma única família, seguido por casas em condomínios. Casa para múltiplas famílias são as menos anunciadas nesse dataset.

Figura 4 – Boxplot de price e homeType



No boxplot da figura 4, temos que a maior variação de preços de imóveis são para os tipos de casa SINGLE_FAMILY, com 75 % dos preços variando até pouco mais de 600 mil dólares. No dos terrenos (LOT) também temos uma variação grande nos preços, entre, aproximadamente, 100 mil dólares até 700 mil, sem contar as outliers.

Figura 5 – Gráfico de dispersão de price x bathPlusBad



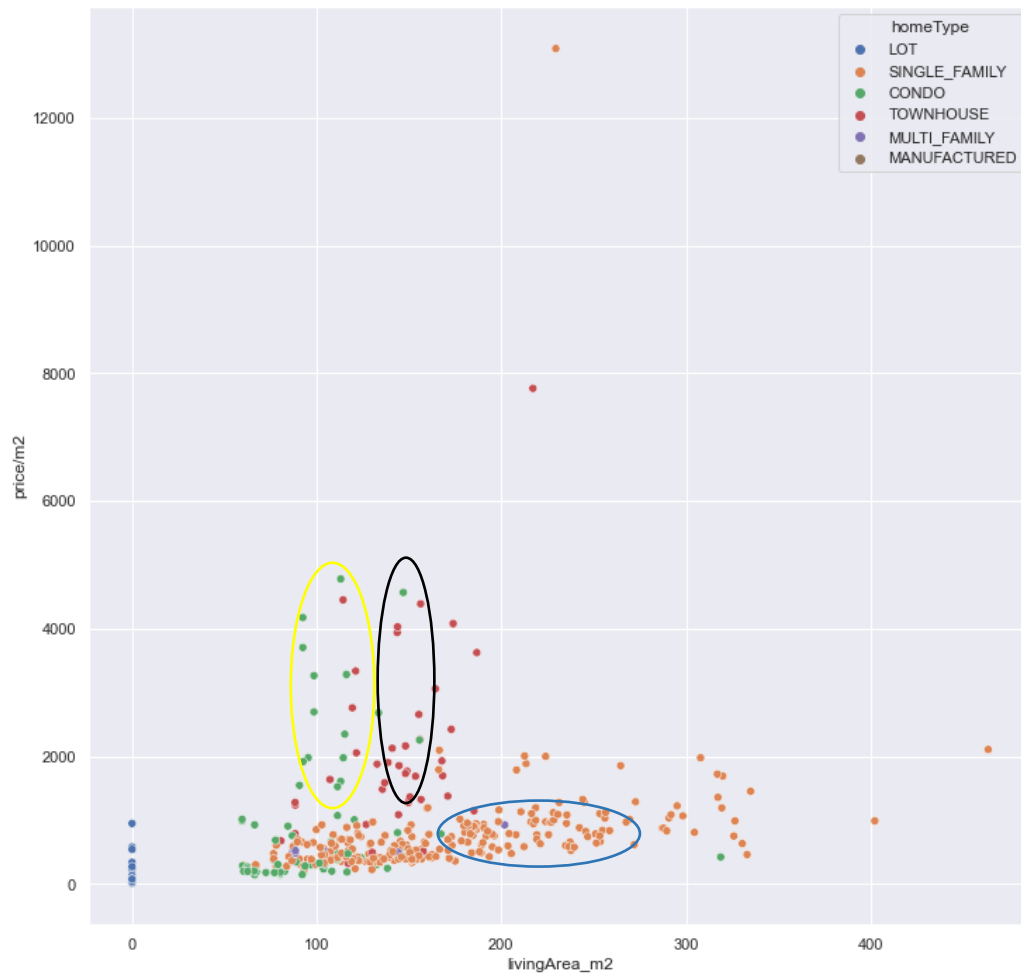
Temos a figura 5 para verificar como se comporta o preço com a variação no número de banheiros e quartos. Como LOT é terreno não há banheiros nem quartos.

Como as casas para uma única família (SINGLE_FAMILY) são as mais flexíveis em termos de construção, vemos uma grande distribuição no número de quartos e banheiros. Vamos que, geralmente, quanto maior o número desses 2 cômodos temos um aumento no preço, pois é necessária uma área maior para construção. No caso do aumento de preço com o mesmo número de banheiros e quartos, isso provavelmente, se deve por conta de uma área mais valorizada pelo mercado de imóveis.

Para condomínios, temos que a maioria desses imóveis possuem entre 2 e 4 banheiros e quartos somados.

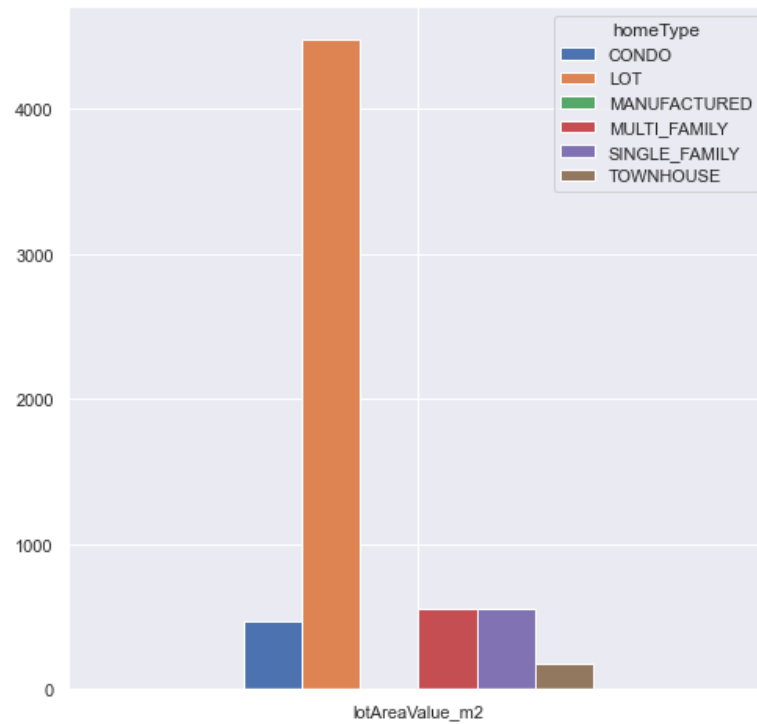
Já para casas pré-construídas (MANUFACTURED) os banheiros e quartos somados são 4, já que são feitas sob medidas em fábricas.

Figura 6 – Gráfico de dispersão de price/m2 x livingArea_m2



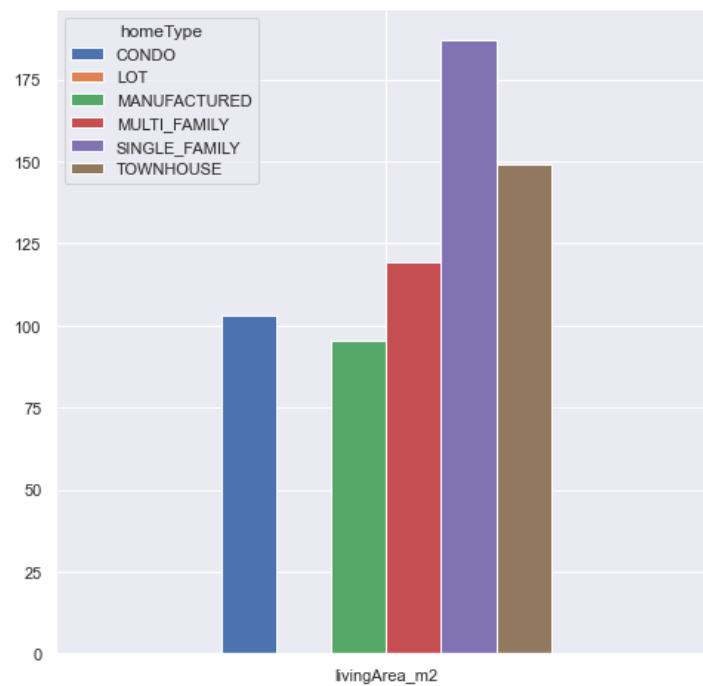
Podemos ver na figura 6 a valorização dos imóveis, para os casos de TOWNHOUSE e CONDO, pois, há imóveis com a mesma livingArea_m2, mas com preços cada vez mais elevados, visto nos círculos amarelo e preto. Esses tipos de imóveis, principalmente TOWNHOUSE, são construídos nas áreas mais urbanas da cidade, logo podendo sofrer esse aumento de preço por metro quadrado. Já as SINGLE_FAMILY, podem ser construídas em vários locais da cidade, muito valorizadas ou não, logo podem ter mais área com o mesmo preço por metro quadrado, visto no círculo azul.

Figura 7 – Gráfico de barras de lotAreaValue_m2



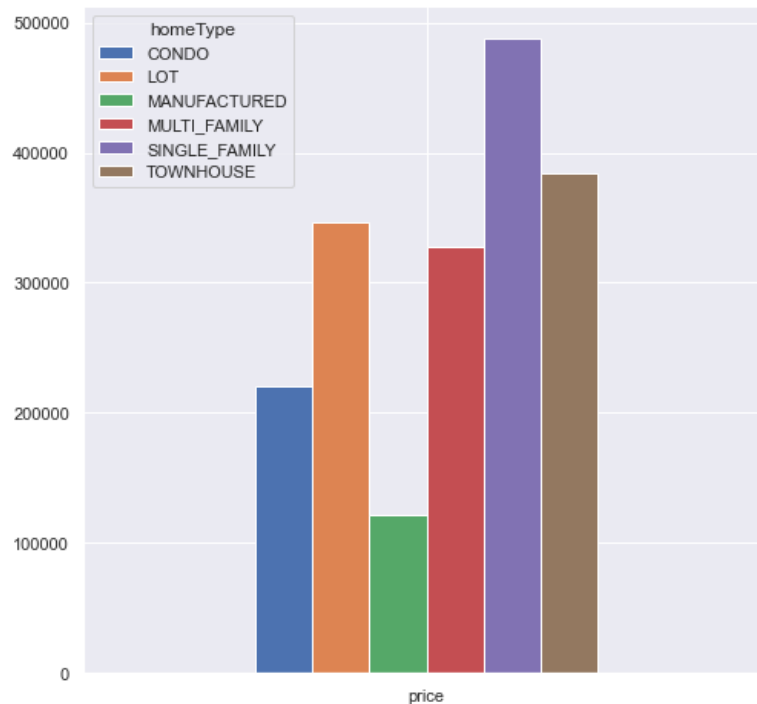
Na figura 7, temos a área média das propriedades são as mesmas, com exceção para os terrenos LOT que são maiores.

Figura 8 – Gráfico de barras de livingArea_m2



A figura 8 mostra que SINGLE_FAMILY é o tipo de imóvel com maior média de área construída. Podemos justificar esse fato com o que foi visto na figura 6, na qual temos uma grande quantidade de casas com grandes áreas.

Figura 9 – Gráfico de barras de price



Para o preço médio do imóvel, figura 9, temos que SINGLE_FAMILY possui o maior valor médio, já que, como visto na figura 4, temos uma grande variação de preços, com 50 % dos valores acima de, aproximadamente, 420 mil dólares. As TOWNHOUSES também apresentam um valor médio alto, pois, visto na figura 6, boa parte dos imóveis são bem valorizados e a variação entre os preços não é muito alta, confirmado pelo boxplot da figura 4.

Conclusões

Com o uso da API, vimos que a maior parte dos anúncios são de casas para uma única família, esse tipo de imóvel possui a maior variação de preço e de área construída. Temos dois tipos de imóveis, townhouse e condo, que apresentam uma alta valorização, com diversos imóveis com a mesma área, mas com preços maiores. Por fim, vimos que, geralmente, quanto mais banheiros e quartos, maior o preço do imóvel.