

## **Metodologia empregada no trabalho de Regras de Associação**

Durante a implementação do projeto de regras de associação, adotei a seguinte metodologia:

1° - Escolhi os atributos que usaria para rodar o algoritmo, sendo eles: bedrooms, bathrooms, sqft\_living, sqft\_lot, floors, waterfront, condition, grade, yr\_built e a social class.

2° - Discretizei esses atributos, tendo em vista que os algoritmos de regras de associação funcionam com base em atributos categóricos. Para realizar essa mudança, utilizei a matplotlib para ter uma visualização melhor dos gráficos, e usei a estratégia de quantis para tal ação, sobretudo nas variáveis com uma distribuição muito ampla, como as relativas à área.

3° - Após discretizar os atributos, separei o dataframe em listas, sendo uma lista para cada classe.

4° - Para cada classe, rodei o algoritmo apriori de regras de associação presente na biblioteca apyori. Foram realizadas 3 iterações para cada classe, sendo uma iteração com parâmetros altos de suporte e confiança para encontrar regras mais gerais para melhor entendimento das classes, uma iteração com parâmetros baixos/médios de suporte e confiança a fim de achar uma grande quantidade de regras e por fim uma iteração com um lift alto para achar regras que potencialmente são interessantes para objeto de estudo.

## **Resultados**

**Lower class:** ao se realizar as iterações, foram se obtendo regras relacionadas a não existência de frente para o mar, o que já se é esperado já que casas de classe baixa dificilmente teriam esse tipo de recurso, característica de casas mais caras. Algumas regras com o lift mais alto se destacam:

```
[frozenset({'good_condition', 'low_floors', 'middle_bed', 'middle_grade'}),  
frozenset({'okay'}),  
0.7148594377510039,  
2.1066320074819824]
```

Ou seja, com uma confiança de 70% e um lift de 2.1, nos indicando um suporte de aproximadamente 33%, casas com boas condições, com poucos andares, com uma quantidade média de quartos e com uma avaliação média apresentam uma idade okay (construídas entre 1960 e 1980) apresentam. Para um suporte de 33% e com atributos que não são tão óbvios a relação, é uma confiança alta ao meu ver. As regras com um lift interessante se resumem nas permutações desses atributos juntamente com a falta de uma frente para o mar citado anteriormente.

**Working class:** Essa classe possui um lift máximo bem maior do que as demais classes, muito provavelmente por ter mais casas dentro dessa classe, então como há mais 'transações', haverão mais regras. Uma regra interessante é a seguinte:

```
[frozenset({'middle_bath', 'okay_condition', 'very_low_sqft_lot'}),  
frozenset({'middle_grade', 'very_recent'}),  
0.8361244019138755,  
4.5927118933697875]
```

Com uma casa com quantidade média de banheiros, numa condição ok e num terreno muito pequeno, sua nota será média e ela foi construída após 2000. Essa regra possui uma confiança de mais de 83% e com um lift alto de 4.59, nos indicando um suporte de aproximadamente 18%. É uma quantidade alta se pararmos para pensar e levando em conta uma confiança muito alta e com dados que não são tão relacionáveis assim de cara.

**Low middle class:**

```
[frozenset({'lower middle',  
           'middle_bath',  
           'okay_condition',  
           'very_low_sqft_lot'}),  
frozenset({'very_recent'}),  
0.7707707707707707,
```

3.412708483131018]

Casas que apresentam quantidade média de banheiros, uma condição okay e um terreno bem pequeno, também são muito recentes (construídas após os anos 2000). Esses dados se relacionam bastante e há algumas permutações e variantes com a presença de vista para o mar, mostrando que a partir de casas de classe low middle, as casas com waterfront já estão mais presentes. Alguns dados mais gerais também são interessantes nessa classe. Como exemplo:

```
[frozenset({'low_floors', 'middle_grade', 'no_waterfront'}),  
  frozenset({'middle_bed'}),  
  0.8201913559881227,  
  0.971641558412788],
```

Casas com poucos andares (1), avaliações médias e sem vista para o mar, possuem uma quantidade média de quartos. A confiança é de aproximadamente 82% com um lift de 0.97, indicando assim um suporte de 84%.

**Upper middle class:** [frozenset({'middle\_bed', 'no\_waterfront', 'old'}),  
 frozenset({'low\_floors', 'middle\_grade'}),  
 0.7675,  
 2.7014568764568763]

Esses dados nos dizem que casas que tem quantidade média de quartos, sem vista para o mar e velhas, também possuem poucos andares (1) e avaliação média. Algumas relações mais gerais também podem ser válidas:

```
[frozenset({'middle_bath'}),  
  frozenset({'middle_bed'}),  
  0.9501452081316554,  
  1.0407829265714905],
```

Nos revelando que, com uma confiança de 95% e um lift de 1.04, na classe upper middle as casas com quantidade média de banheiros tem quantidade média de quartos. Apesar de parecer óbvio, pode ser um bom indício na mão de um especialista. Por exemplo, com esses dados, pode-se ter uma ideia da quantidade esperada de quartos e banheiros que uma casa nessa faixa de preço deve ter para ser atrativa.

**Upper class:** Alguns dados interessantes se mostram nas casas consideradas de upper class. Como exemplo, temos:

```
[frozenset({'low_bath'}),  
  frozenset({'middle_grade'}),  
  0.7225,  
  2.590875796178344]
```

Com 72% de confiança e 2.59 de lift, casas que tem uma baixa quantidade de banheiros, possui uma avaliação média.

```
[frozenset({'middle_bed', 'middle_high_sqft_living', 'recent'}),  
  frozenset({'middle_bath',  
            'middle_floors',  
            'middle_sqft_lot',  
            'no_waterfront',  
            'upper'}),  
  0.525305410122164,  
  2.506330049989647],
```

Casas com quantidade média de quartos, área construída média/alta, e recentes (1980-2000) possuem uma quantidade de banheiros média, quantidade média de andares, quantidade média de lote e sem vista para o mar, com uma confiança de 52% e suporte de 20% nessa classe, uma quantidade considerável se analisarmos.

**Toppest class:** Na análise geral, os termos “middle” se relacionam frequentemente, revelando que dificilmente se um dos quesitos da casa é mediano, os outros não serão. Nesse caso, praticamente todos os atributos estão beirando o máximo, e as exceções se localizam nessa classe por motivos extras, como localização ou outro significado.

```
[frozenset({'middle_bed',  
           'middle_sqft_lot',  
           'no_waterfront',  
           'very_recent'}),  
frozenset({'high_grade', 'high_sqft_living', 'okay_condition'}),  
0.6470588235294118,  
3.526470588235294]
```

Na maior classe desse dataframe, as casas com quantidade mediana de camas, de lote, sem vista para o mar e muito recentes (após os anos 2000) apresentam uma avaliação alta, uma área construída grande e uma condição ok, com 64% de confiança e um lift de 3.5, aproximadamente.