Machine Learning Presentation

Akmal Ariq - 1301174378





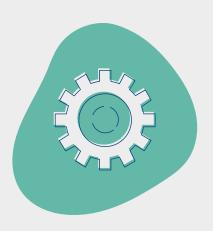
Problem

Create a clasification and clustering experiment on air_bnb.csv



Problem 1

Create classification models based on the given dataset

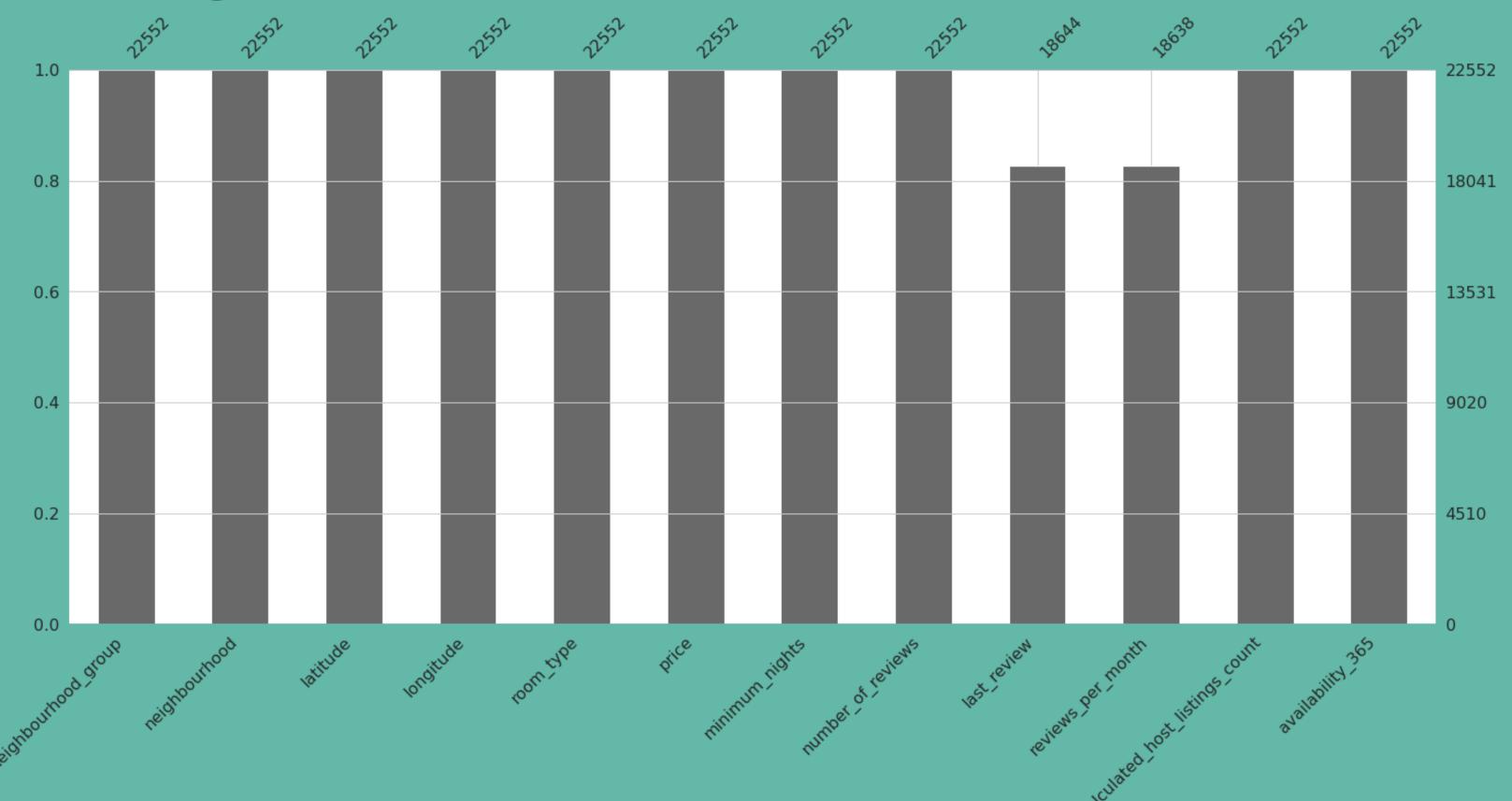


Problem 2

Create clustering model based on the given dataset

Exploratory Data Analysis

Missing Values



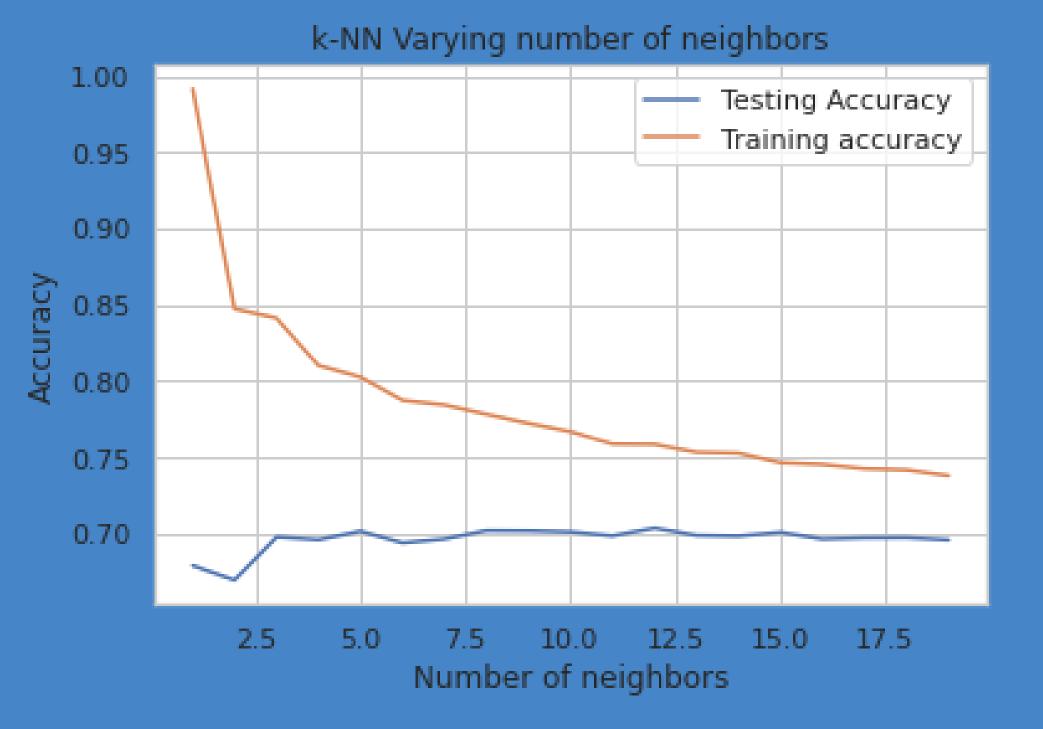
Missing Values



Correlation between features

| neighbourhood_group | 1 | 0.076 | -0.14 | -0.039 | -0.016 | -0.0051 | -0.0039 | -0.011 | 0.011 | -0.0078 | -0.0035 | 0.03 |
|--------------------------------|---------------------|---------------|----------|-----------|-----------|---------|----------------|-------------------|-------------|-------------------|------------------------|------------------|
| neighbourhood | 0.076 | 1 | -0.14 | -0.17 | 0.027 | -0.0014 | -0.0082 | -0.038 | -0.021 | -0.05 | -0.048 | -0.06 |
| latitude | -0.14 | -0.14 | 1 | -0.11 | -0.032 | 0.0022 | 0.012 | 0.038 | 0.018 | 0.039 | 0.04 | 0.012 |
| longitude | -0.039 | -0.17 | -0.11 | 1 | 0.02 | -0.043 | -0.026 | -0.021 | -0.015 | -0.033 | -0.044 | -0.087 |
| room_type | -0.016 | 0.027 | -0.032 | 0.02 | 1 | -0.068 | -0.076 | -0.035 | 0.011 | 0.019 | -0.019 | -0.086 |
| price | -0.0051 | -0.0014 | 0.0022 | -0.043 | -0.068 | 1 | 0.0036 | -0.0012 | -0.027 | -0.0077 | 0.073 | 0.11 |
| minimum_nights | -0.0039 | -0.0082 | 0.012 | -0.026 | -0.076 | 0.0036 | 1 | -0.022 | -0.053 | -0.054 | 0.027 | 0.099 |
| number_of_reviews | -0.011 | -0.038 | 0.038 | -0.021 | -0.035 | -0.0012 | -0.022 | 1 | 0.34 | 0.58 | 0.086 | 0.28 |
| last_review | 0.011 | -0.021 | 0.018 | -0.015 | 0.011 | -0.027 | -0.053 | 0.34 | 1 | 0.49 | 0.065 | 0.23 |
| reviews_per_month | -0.0078 | -0.05 | 0.039 | -0.033 | 0.019 | -0.0077 | -0.054 | 0.58 | 0.49 | 1 | 0.11 | 0.23 |
| calculated_host_listings_count | -0.0035 | -0.048 | 0.04 | -0.044 | -0.019 | 0.073 | 0.027 | 0.086 | 0.065 | 0.11 | 1 | 0.25 |
| availability_365 | 0.03 | -0.06 | 0.012 | -0.087 | -0.086 | 0.11 | 0.099 | 0.28 | 0.23 | 0.23 | 0.25 | 1 |
| | neighbourhood_group | neighbourhood | latitude | longitude | room_type | price | minimum_nights | number_of_reviews | last_review | reviews_per_month | ed_host_listings_count | availability_365 |

Classification K-Nearest Neighbours



Classification K-Nearest Neighbours

Classification K-Nearest Neighbours

| | precision | recall | f1-score | support |
|---------------------------------------|----------------------|----------------------|----------------------|----------------------|
| 0 1 2 | 0.70 0.71 0.77 | 0.70 0.72 0.22 | 0.70 0.72 0.35 | 3217 3460 89 |
| accuracy macro avg weighted avg | 0.73 0.70 | 0.55 0.70 | 0.70 0.59 0.70 | 6766 6766 6766 |

Feature Selection and Model Tuning

```
In this model the features are: 'price', 'minimum_nights', 'availability_365', 'number_of_reviews'
```

The parameters are:
Best leaf_size: 1
Best p: 1
Best n_neighbors: 29

Feature Selection and Model Tuning

| | precision | recall | f1-score | support |
|---------------------------------------|--------------|--------------|----------------------|----------------------|
| 0 1 | 0.79 0.77 | 0.76 0.82 | 0.77 0.79 | 3217 3460 |
| 2 | 0.00 | 0.00 | 0.00 | 89 |
| accuracy macro avg weighted avg | 0.52 0.77 | 0.52 0.78 | 0.78 0.52 0.77 | 6766 6766 6766 |

Classification Random Forest

Classification Random Forest

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.79 | 0.81 | 0.80 | 2145 |
| 1 | 0.81 | 0.80 | 0.81 | 2307 |
| 2 | 0.90 | 0.31 | 0.46 | 59 |
| | | | | |
| accuracy | | | 0.80 | 4511 |
| macro avg | 0.83 | 0.64 | 0.69 | 4511 |
| weighted avg | 0.80 | 0.80 | 0.80 | 4511 |

Feature Selection and Model Tuning

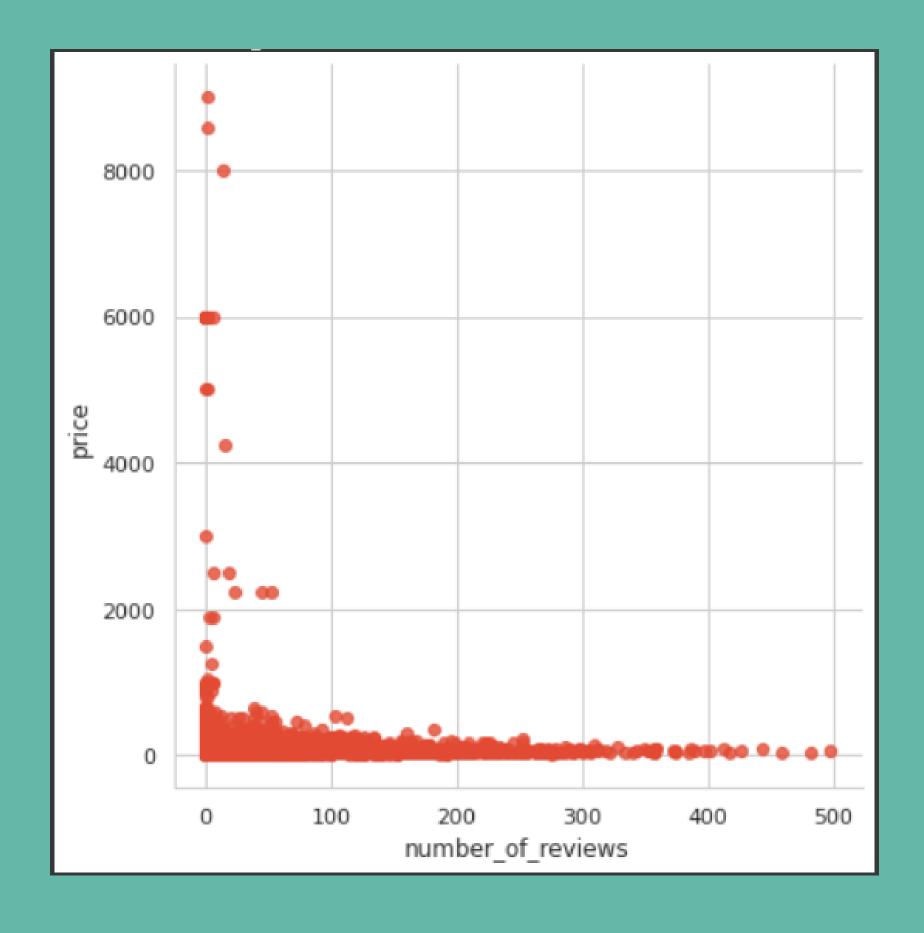
In this model the features are already optimal

```
The parameters are:
'criterion': 'entropy',
'min_samples_leaf': 5,
'min_samples_split': 16,
'n_estimators': 1000
```

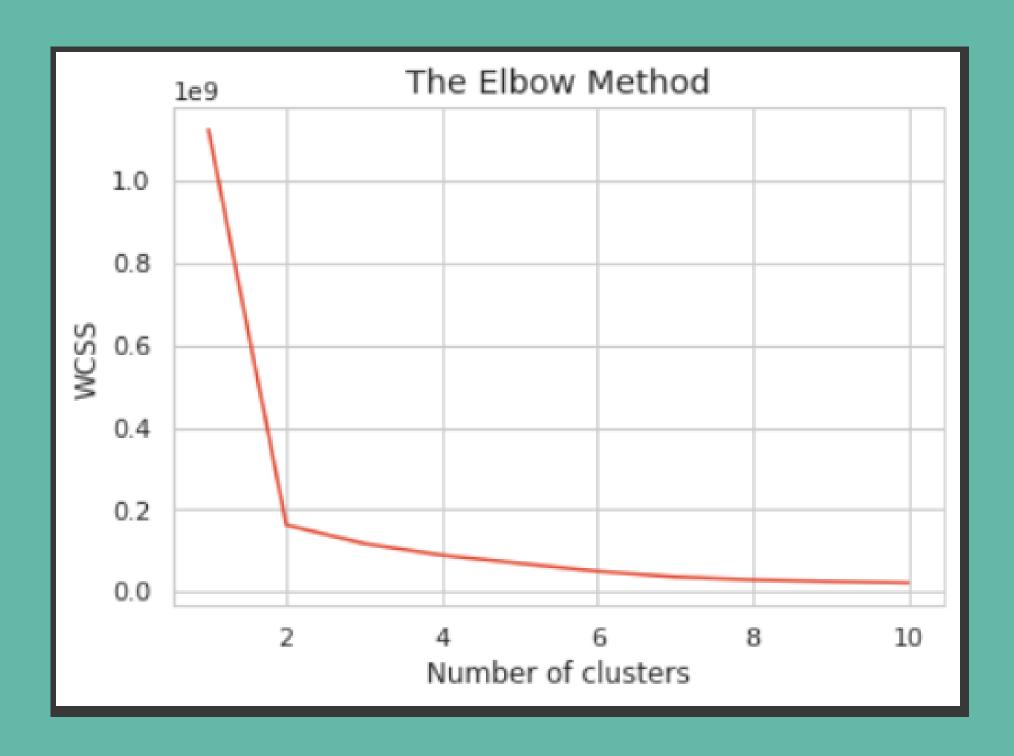
Feature Selection and Model Tuning

| | precision | recall | f1-score | support |
|--------------|--------------|--------------|--------------|------------|
| 0 | 0.78 | 0.80 | 0.79 | 3217 |
| 1 2 | 0.80 0.73 | 0.80 0.27 | 0.80 0.39 | 3460 89 |
| | 0.73 | 0.27 | 0.55 | 0,7 |
| accuracy | | | 0.79 | 6766 |
| macro avg | 0.77 | 0.62 | 0.66 | 6766 |
| weighted avg | 0.79 | 0.79 | 0.79 | 6766 |

Clustering Experiment 1

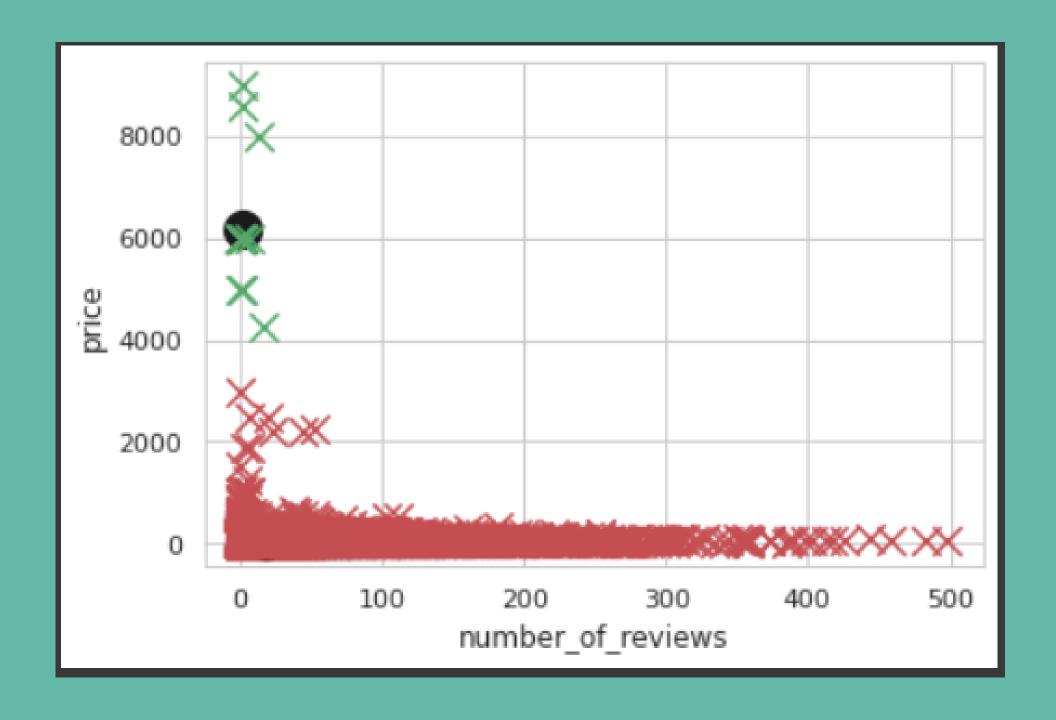


Elbow Method

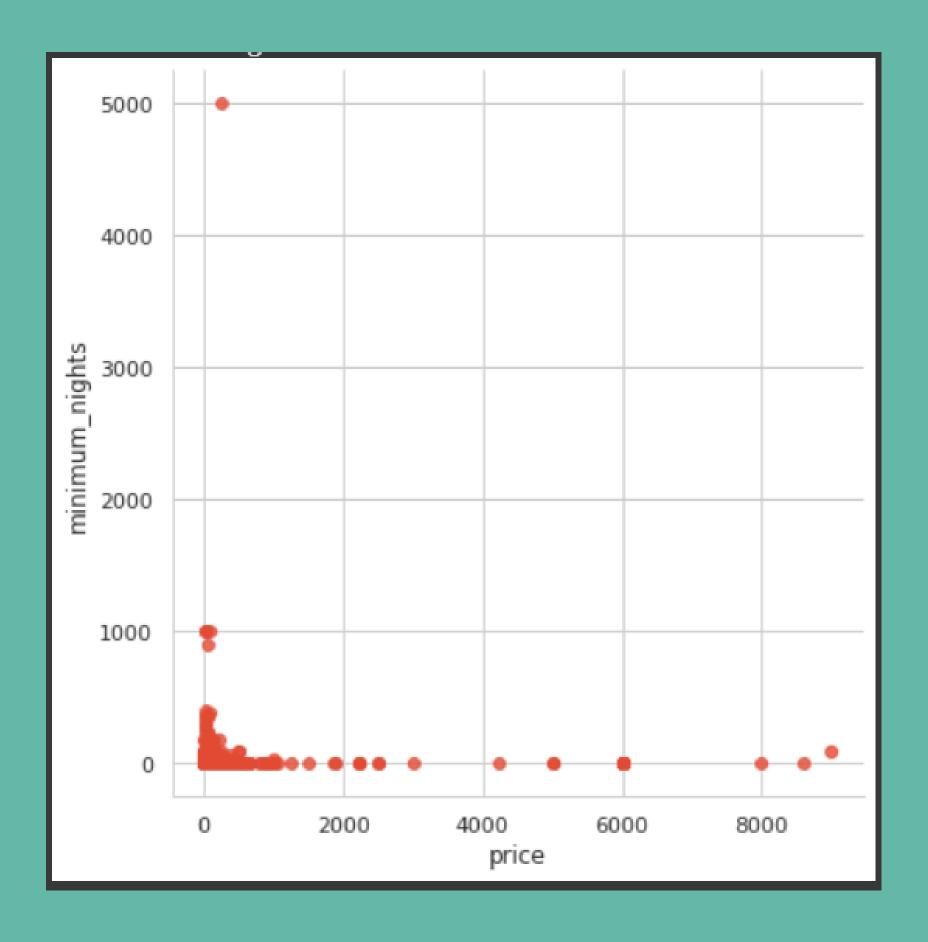


Clustering Experiment 1

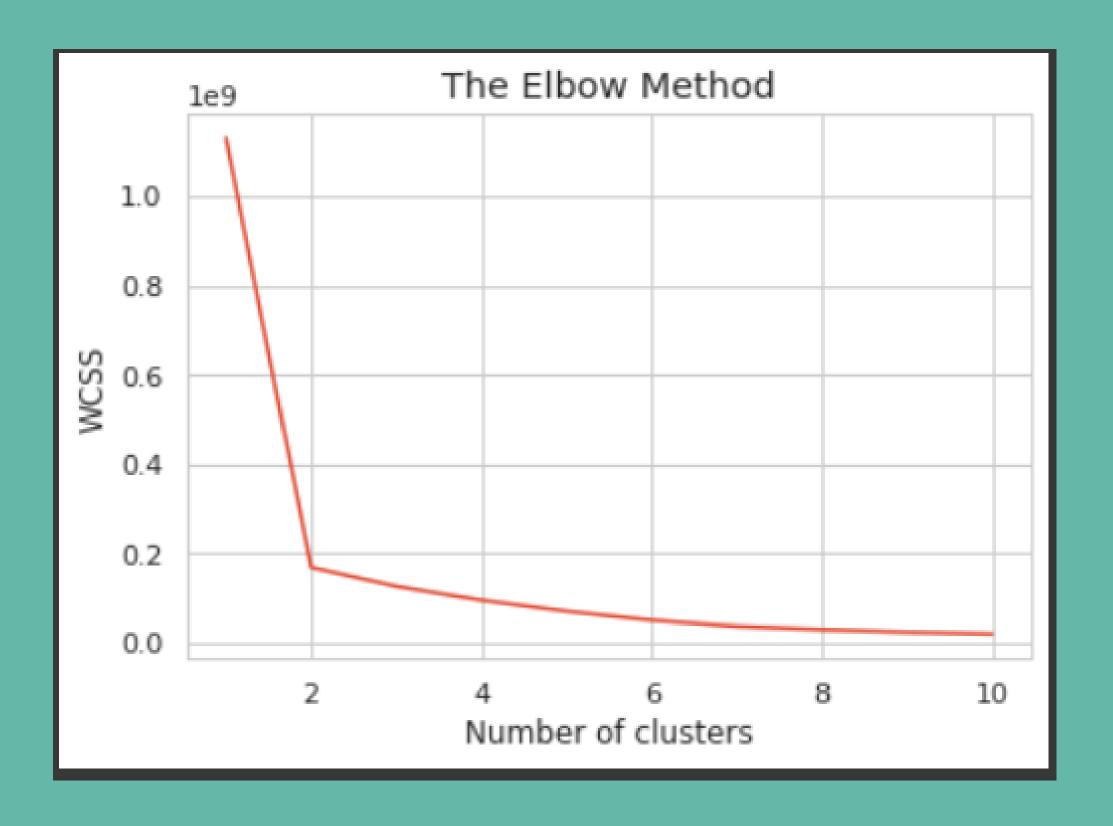
There are 2 clusters: cluster 1: low reviews and expensive cluster 2: high reviews and moderate price



Clustering Experiment 2

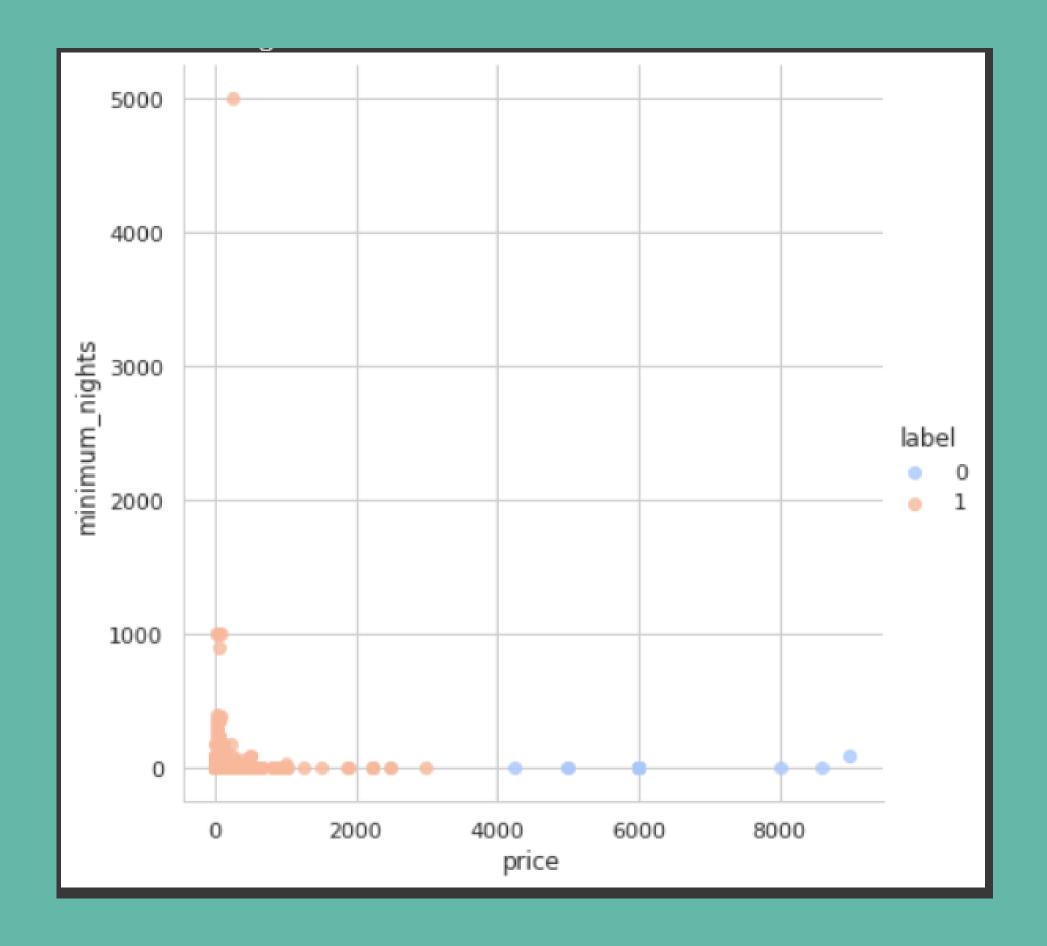


Elbow Method



Clustering Experiment 2

There are 2 clusters:
cluster 1:
lower minimum_nights and
expensive
cluster 2:
moderate to high minimum_nights
and moderate price



Thank You