Machine Learning Presentation

Akmal Ariq - 1301174378

Problem

Create a clasification and clustering experiment on air_bnb.csv

Problem 1

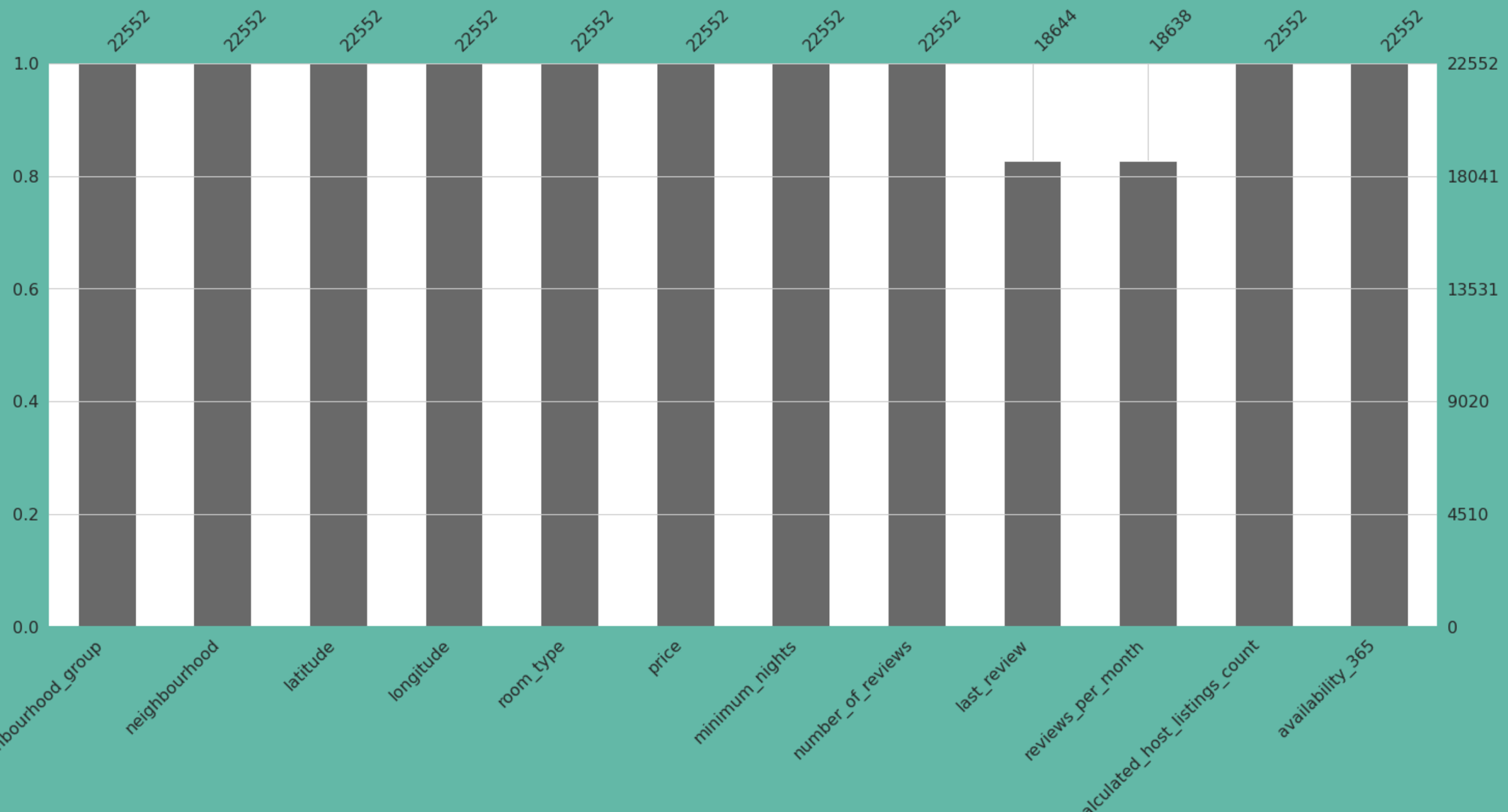
Create classification models based on the given dataset

Problem 2

Create clustering model based on the given dataset

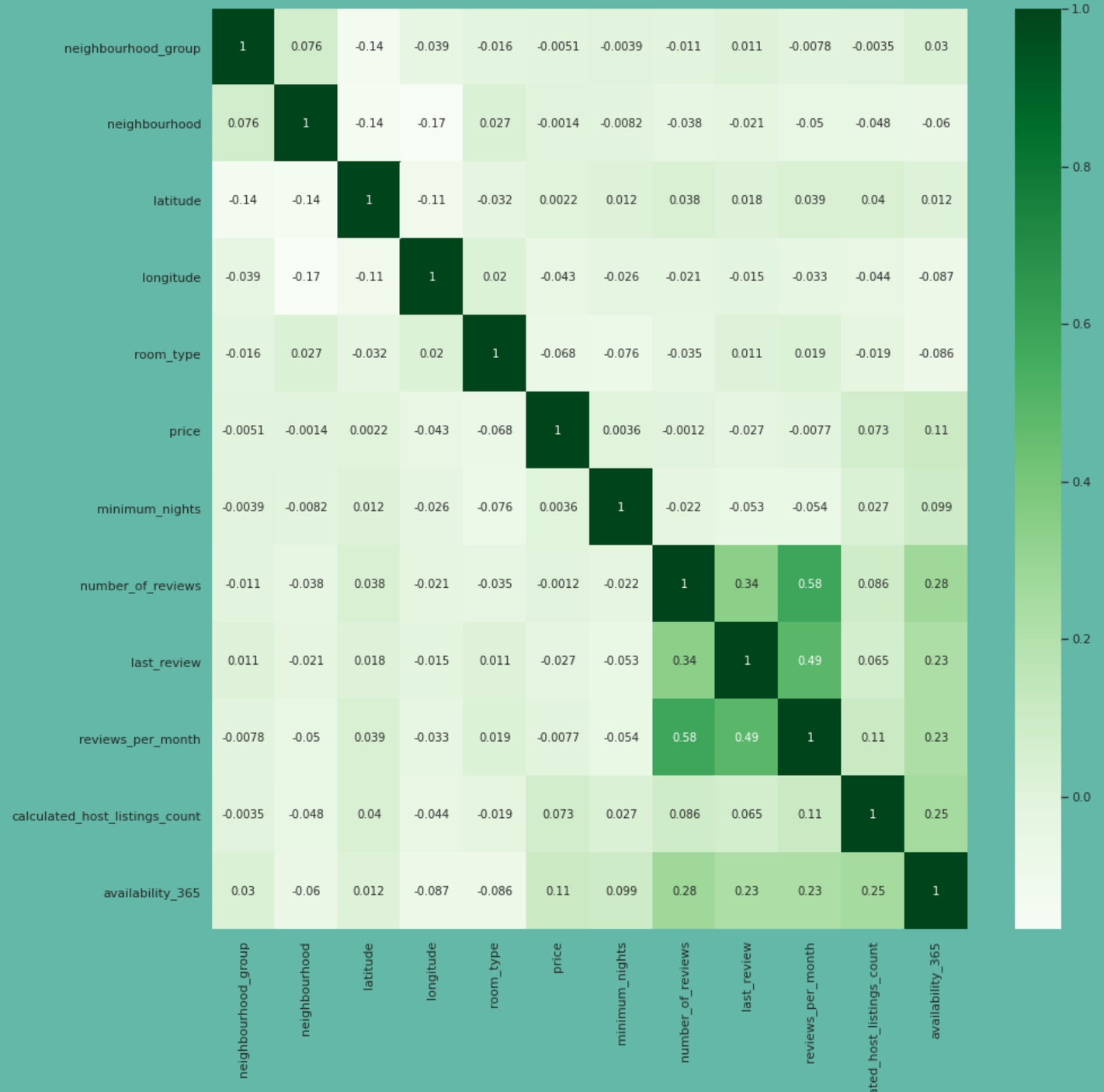
Exploratory Data Analysis

Missing Values



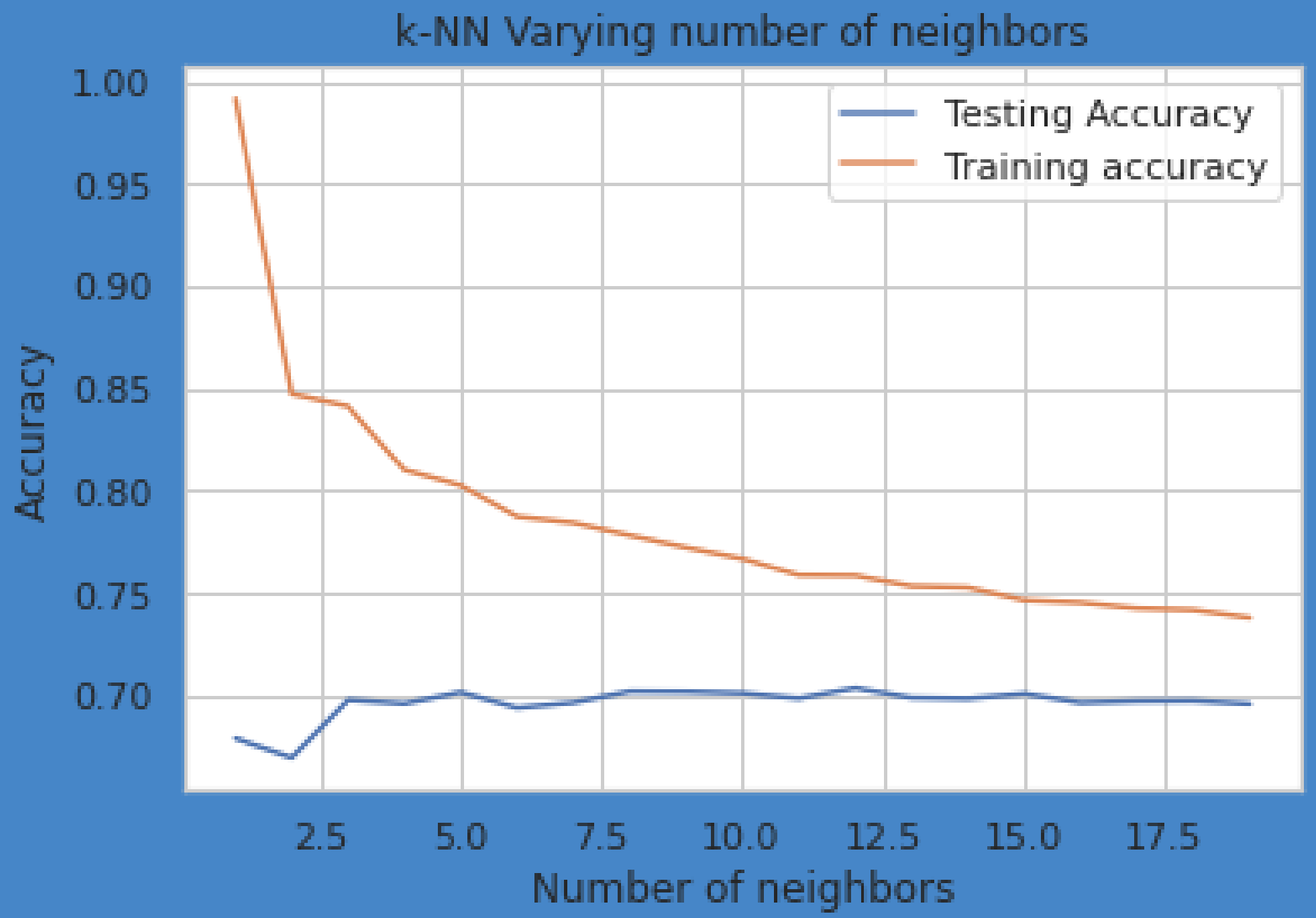
Missing Values

Correlation between features



Classification K-Nearest Neighbours

Classification K-Nearest Neighbours



Classification K-Nearest Neighbours

	precision	recall	f1-score	support
0	0.70	0.70	0.70	3217
1	0.71	0.72	0.72	3460
2	0.77	0.22	0.35	89
accuracy			0.70	6766
macro avg	0.73	0.55	0.59	6766
weighted avg	0.70	0.70	0.70	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	0.79	0.76	0.77	3217
1	0.77	0.82	0.79	3460
2	0.00	0.00	0.00	89
accuracy			0.78	6766
macro avg	0.52	0.52	0.52	6766
weighted avg	0.77	0.78	0.77	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

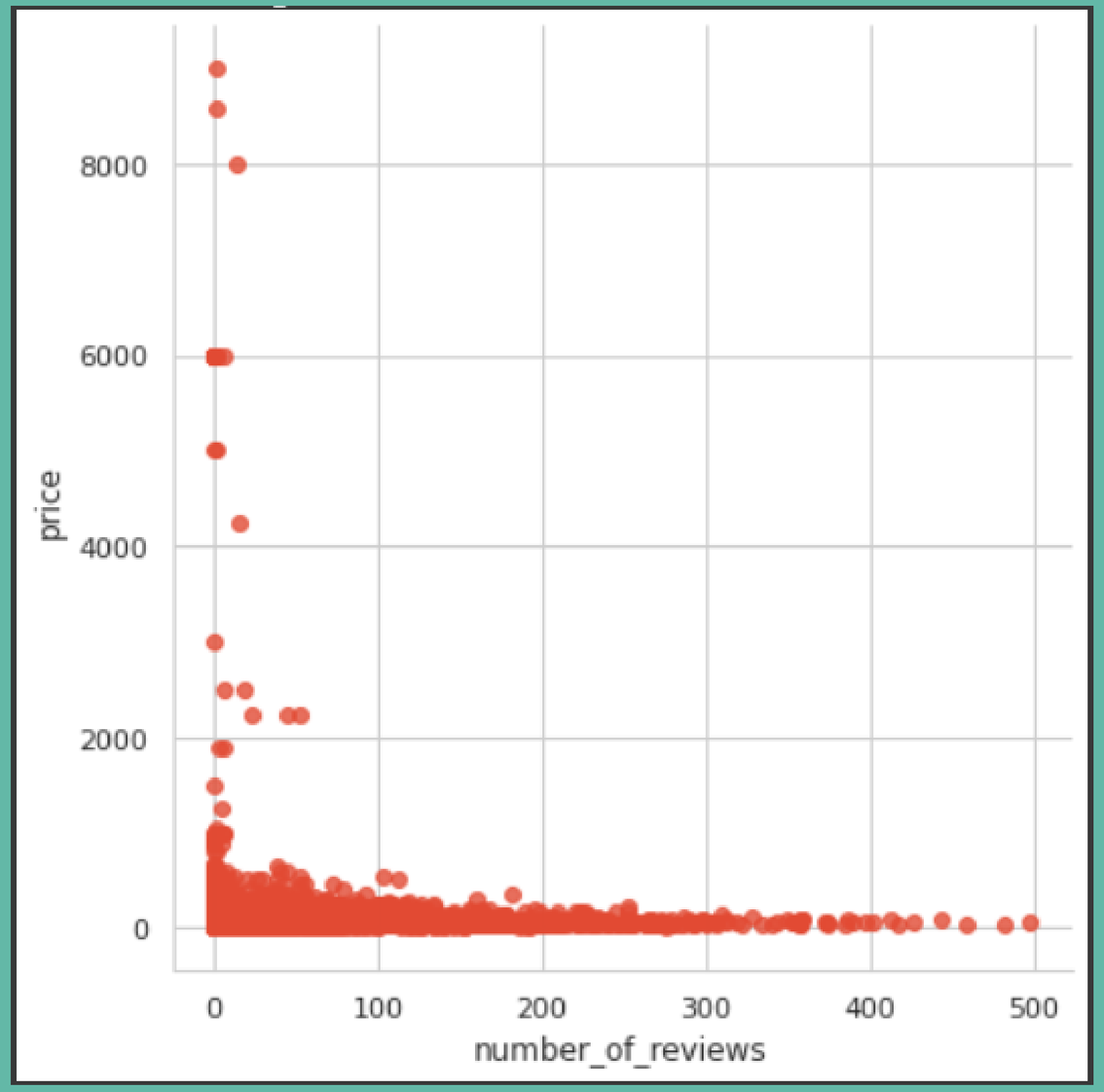
In this model the features are already optimal

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

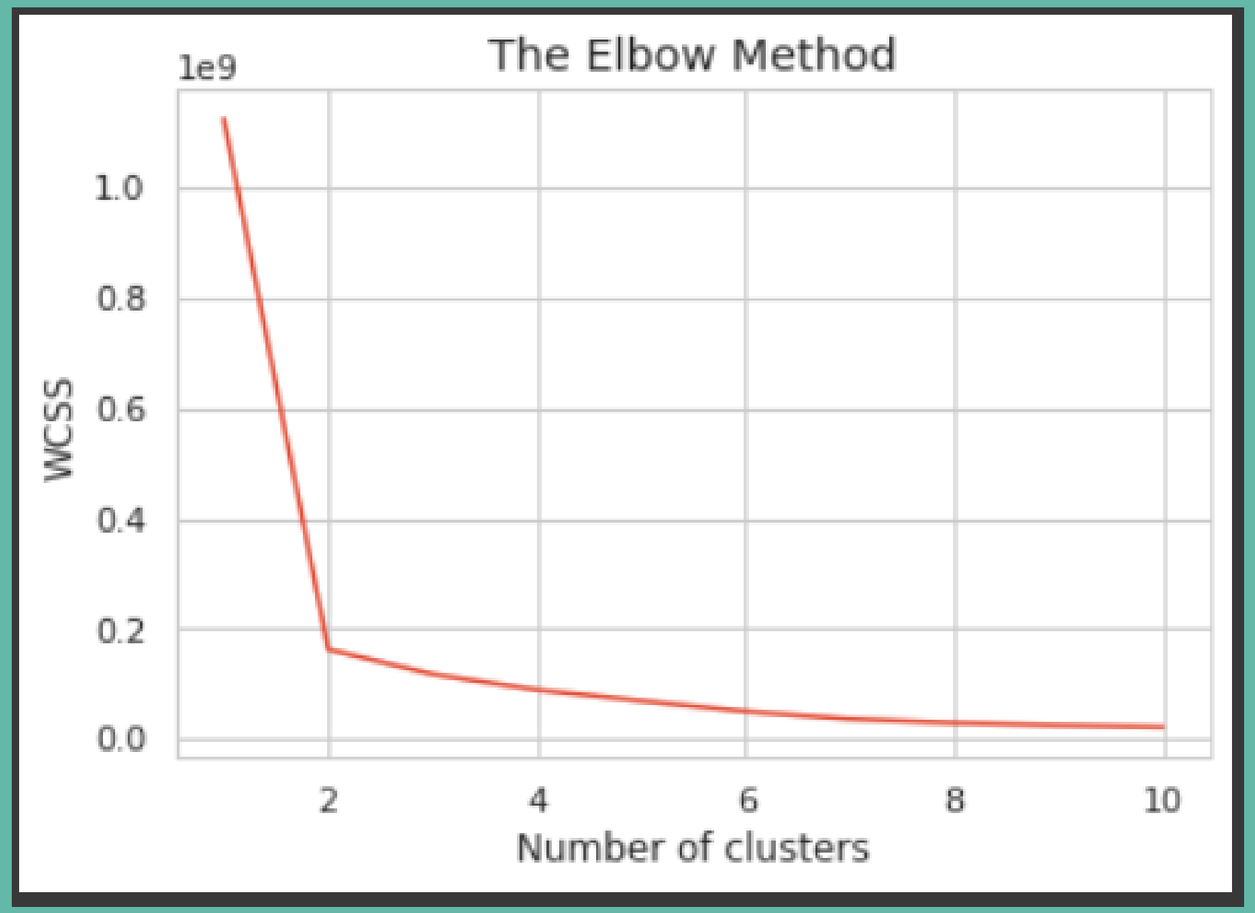
Feature Selection and Model Tuning

	precision	recall	f1-score	support
0	0.78	0.80	0.79	3217
1	0.80	0.80	0.80	3460
2	0.73	0.27	0.39	89
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

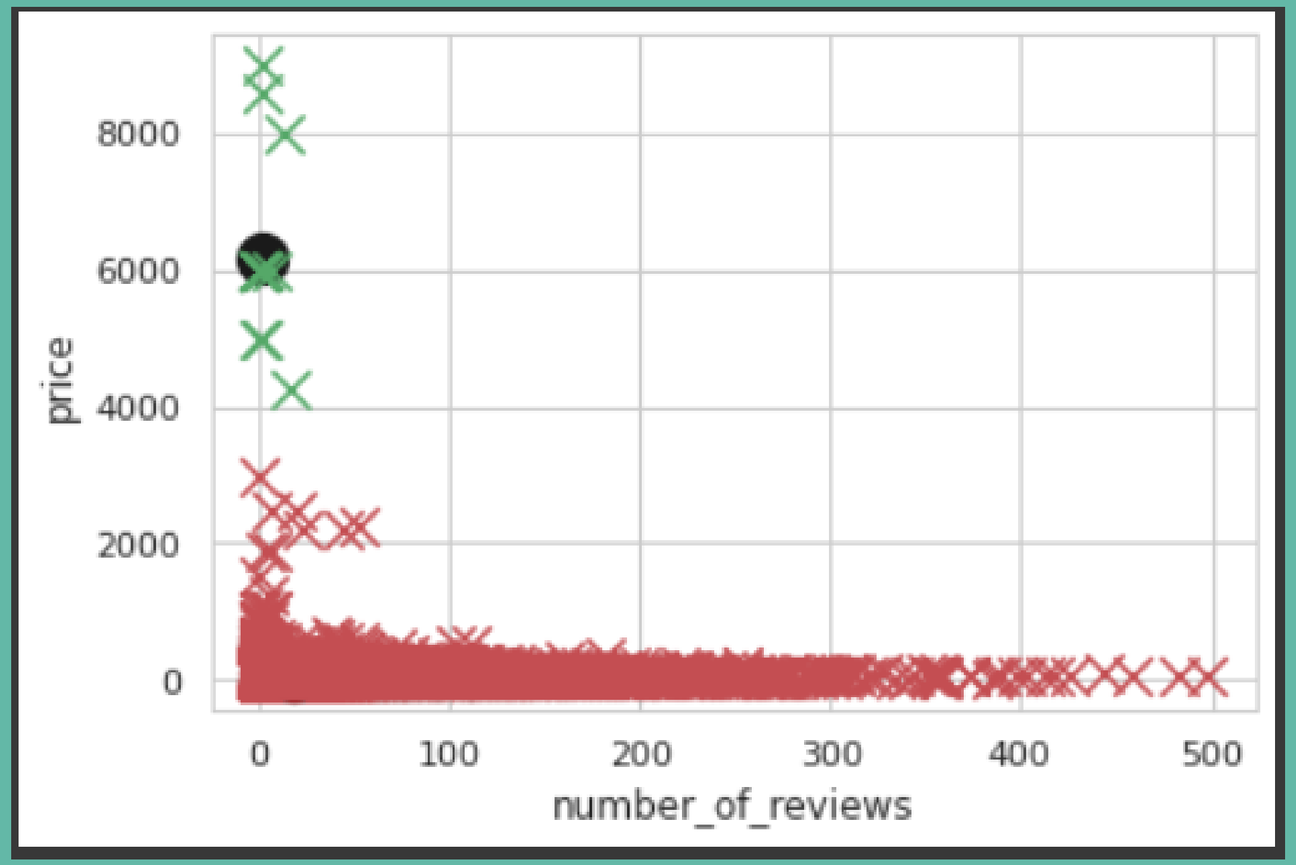


Elbow Method

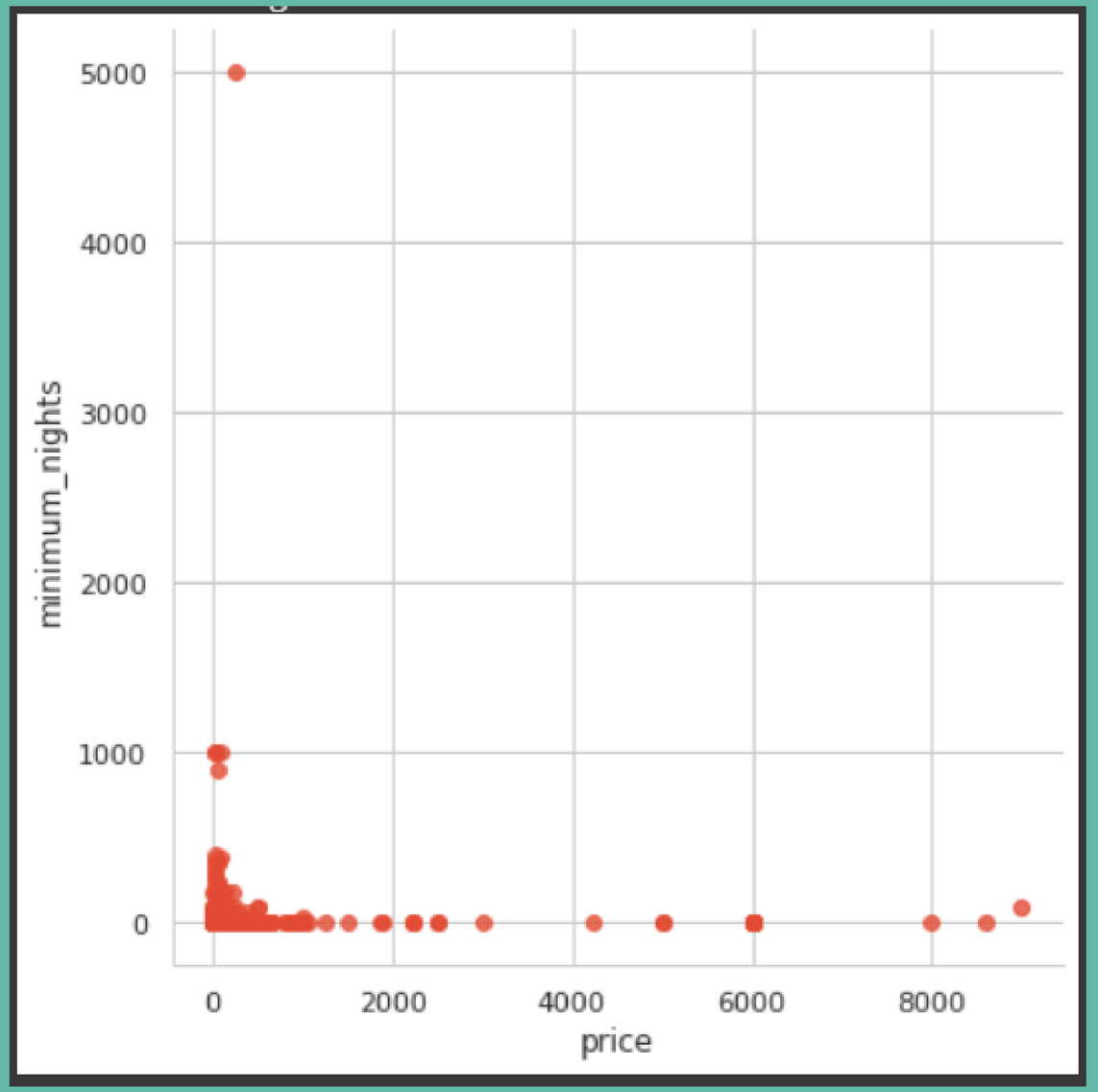


Clustering Experiment 1

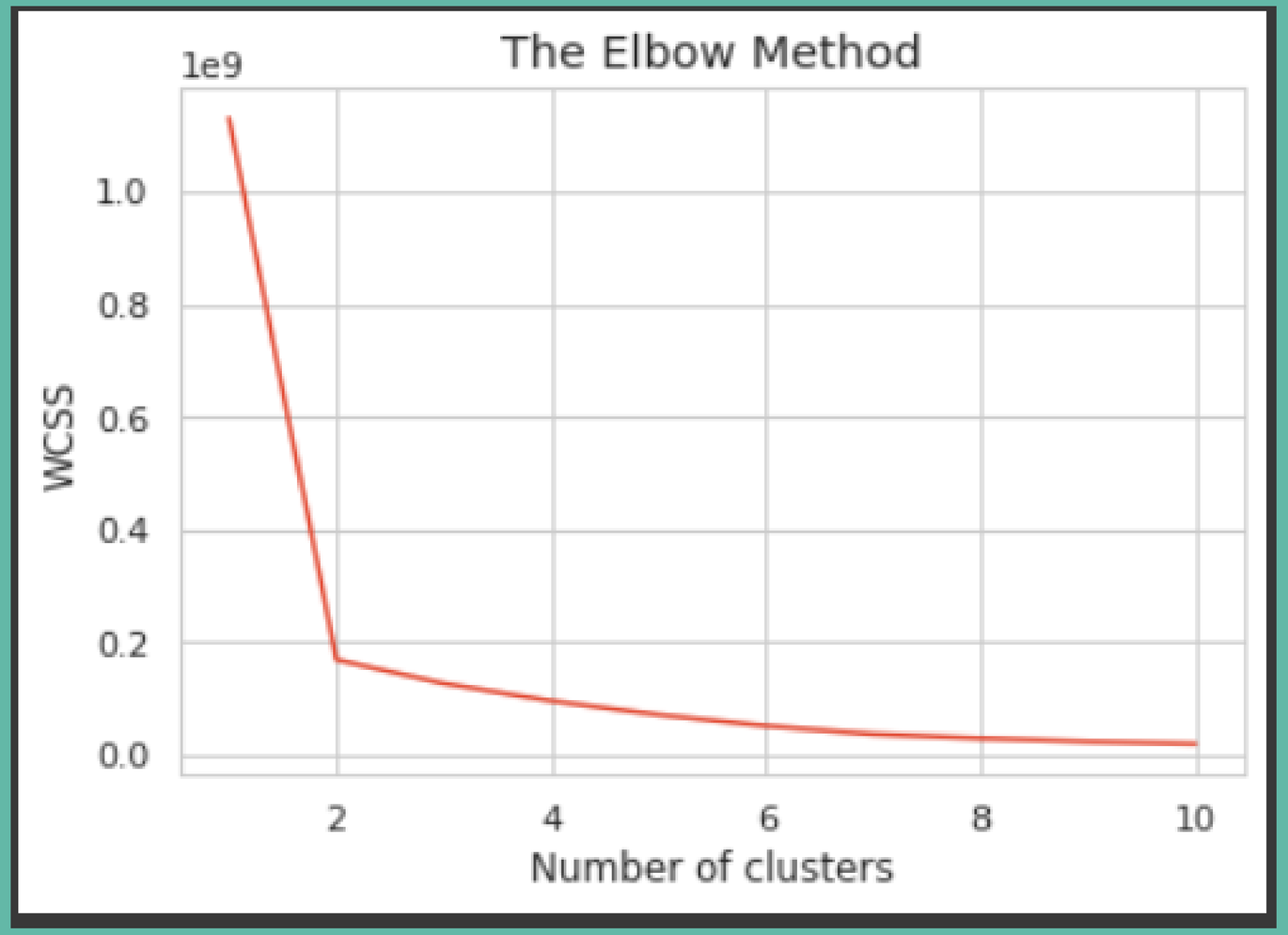
There are 2 clusters:
cluster 1:
low reviews and expensive
cluster 2:
high reviews and moderate price



Clustering Experiment 2

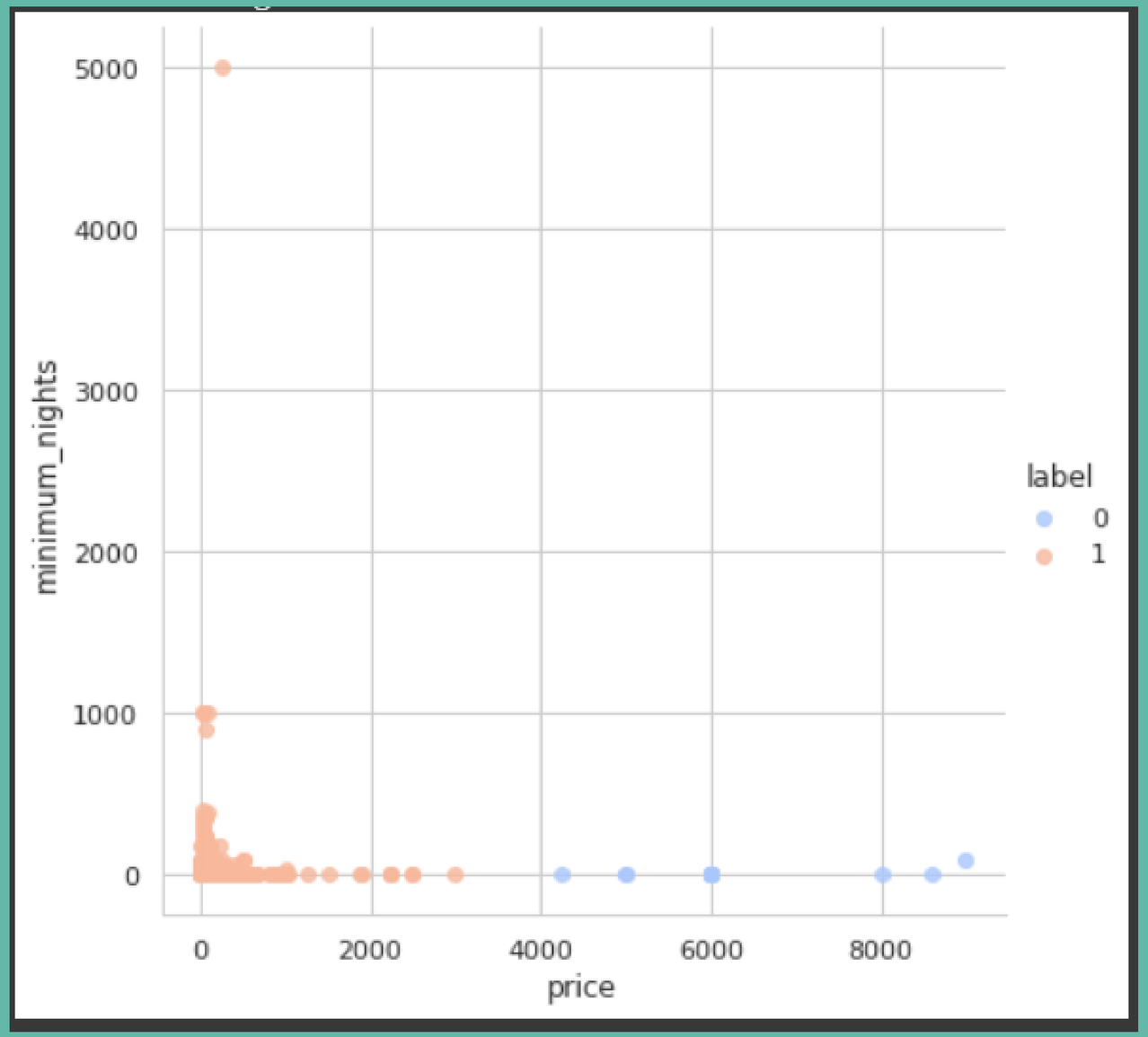


Elbow Method



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