

# Exploring Airbnb Listings in Lisbon

A data-driven Analysis

Group B7

Programming for Data Science  
2024



# Agenda

**1 | Abstract**

**2 | Introduction**

**3 | Data and Methods**

**4 | Results and Discussion**

**5 | Conclusions**

# Abstract

- This project analyzed Airbnb listings in Lisbon to provide insights for hosts, investors, and guests.
- After cleaning the data and addressing outliers and missing values, K-Means clustering identified five distinct groups: Luxury Rentals, Standard Rentals, Budget-Friendly Rentals, Corporate Housing, and Extended Stay Rentals.
- Patterns in pricing, rental frequency, and property types across neighborhoods were revealed.



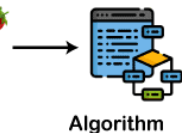
NumPy



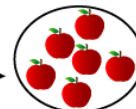
pandas



Raw Data



Algorithm



Output

# Introduction



- **Motivation**
  - Lisbon's growing popularity as a tourist destination. Understanding the market can benefit hosts and guests.
- **Objectives**
  - Analyze pricing trends based on property characteristics. Investigate geographic patterns to see how location impacts Airbnb listings.
- **Approach**
  - Apply machine learning techniques, specifically unsupervised learning, to segment the market. Visualize the results to uncover hidden patterns and trends and define a strategy for hosts and guests.



# Data and Methods

## Python Libraries:

Pandas

NumPy

Matplotlib

Seaborn

Scikit-learn

Plotly

### I. DATA COLLECTION AND EXPLORATION

AIRBNB  
LISTINGS  
DATASET

BASIC  
EXPLORATION  
(Columns, Rows,  
Types)

STATISTICAL  
EXPLORATION

VISUAL  
EXPLORATION  
(Outliers and  
Missing Values)

### II. PREPROCESSING

DATA CLEANING

REMOVING  
OUTLIERS

FILLING MISSING  
VALUES  
(KNN Imputer)

EXPLORING DATA  
AFTER CLEANING  
(Checking  
correlations)

### III. MODELLING

NORMALIZING  
DATA  
(StandardScaler)

ELBOW AND  
SILHOUETTE  
METHODS  
(Clustering)

K-MEANS, K-  
MEANS AFTER  
PCA AND  
DBSCAN

SELECTION OF  
THE BEST MODEL  
(Performance x  
Clusters)

### IV. DATA VISUALIZATION AND ANALYSIS

INTERPRETING  
THE 5 K-MEANS  
CLUSTERS

ANALYZING  
K-MEANS WITH  
CATEGORICAL  
VARIABLES

HOST\_INFO,  
NEIGHBORHOOD,  
NEIGHBORHOOD\_  
GROUP &  
ROOM\_TYPE

PLOTTING  
CLUSTERS BY  
PRICE RANGE

# Clustering

- Other algorithms were tested before concluding that K-Means was the best option.

## Advantages

## Results

## Comments

### DBSCAN

No need to pre-define the number of clusters, robust to outliers.

```
label_dbscan
0    22560
1      289
-1     23
2       19
3       15
Name: count, dtype: int64
```

- Based on the **Silhouette score**, which evaluates clustering performance metrics.

```
{'K-means': 0.4646801803341521,
 'K-means with pca ': 0.41394902739710115,
 'DBSCAN': 0.6908694531785264}
```

### K-Means after performing PCA

To reduce the number of variables by combining them into fewer, more meaningful features.

```
label_pca_kmeans
0    14872
3    6060
1    1270
4     658
2       46
Name: count, dtype: int64
```

- DBSCAN was ineffective in clustering, despite its strength in handling outliers.
- K-means produced balanced clusters and clearer insights while K-means after PCA was less interpretable and efficient.

### K-Means

Computationally efficient and can handle large datasets – performs the best on our dataset.

```
label_kmeans
1    15746
2     5231
3     1191
0       692
4        46
Name: count, dtype: int64
```

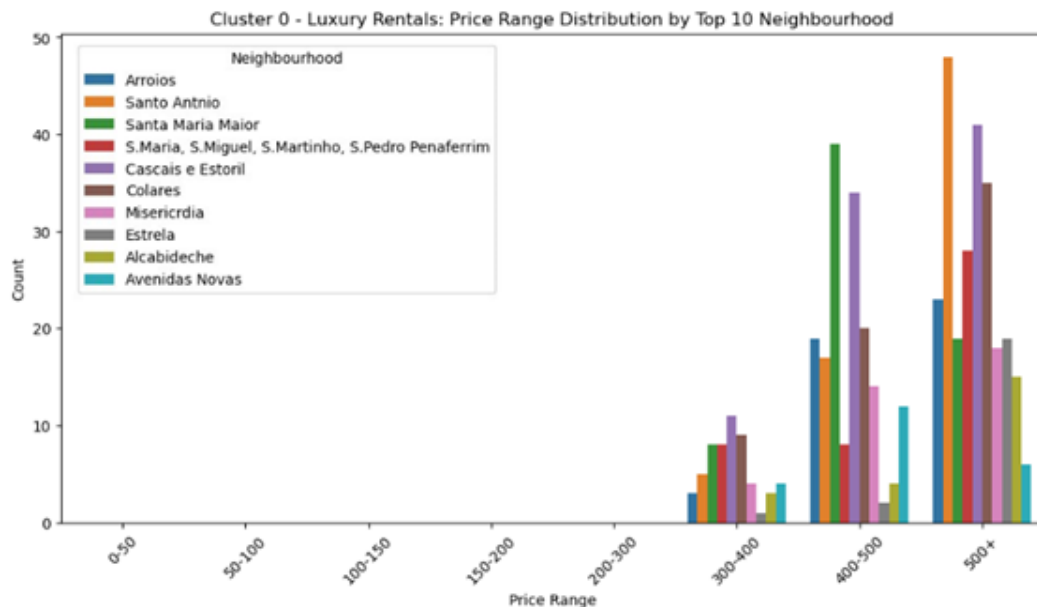
# Results and Discussion

The K-Means analysis revealed 5 clusters:

- Cluster 0: Luxury Rentals
- Cluster 1: Standard Rentals
- Cluster 2: Budget-Friendly Rentals
- Cluster 3: Corporate Housing
- Cluster 4: Extended Stay Rentals

# Results and Discussion

- **Cluster 0 - Luxury Rentals:** This cluster consists of 692 data points and features high-end properties

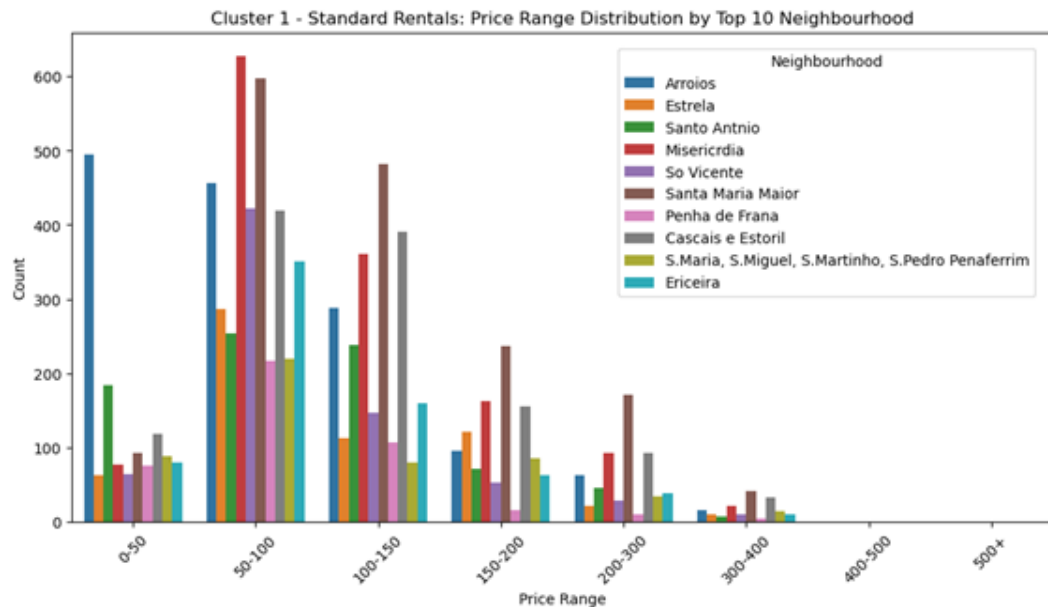


- **Average Price:** €652 per night, entire home/apartment as room types (87%), with moderate minimum night requirements averaging around 3 nights.
- **Rental Frequency:** Low, indicating service to a select group of luxury clientele, with reviews averaging 0.50 per month.
- **Leading Hosts:** 'RoomPicks' and 'OhMyGuest', likely managed by enterprises or rental agencies.
- **Locations:** Situated in exclusive neighborhoods like Cascais e Estoril, Santo António, Santa Maria Maior, and Colares.



# Results and Discussion

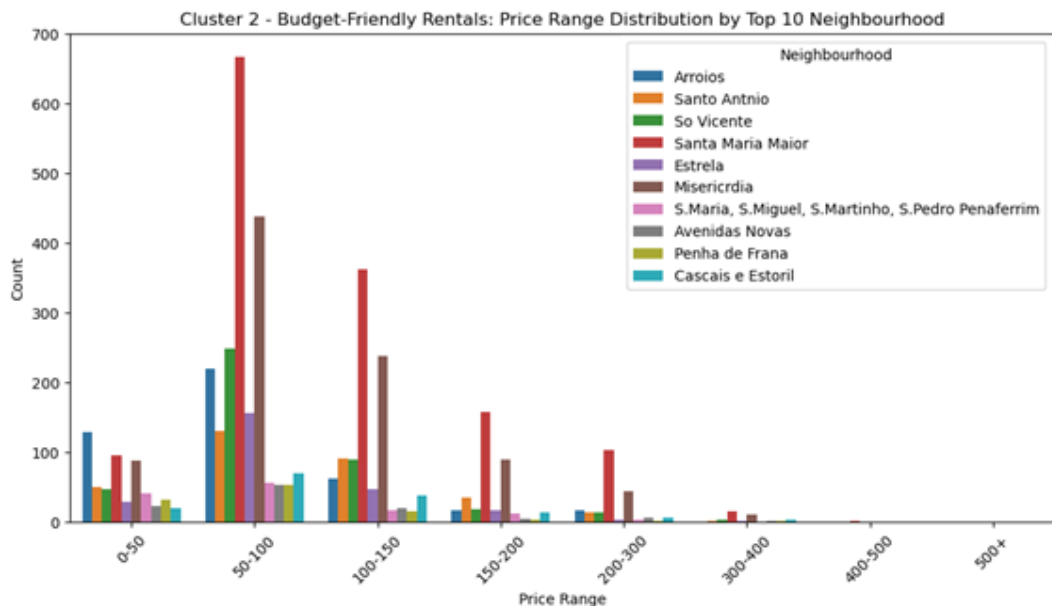
- **Cluster 1 - Standard Rentals:** Comprising 15,746 data points, this cluster represents the most common rental type



- **Average Price:** €106 per night, with listings mostly being mostly entire home/apartment (70%) and private rooms (27%), requiring around 4 nights minimum stay.
- **Target Audience:** Primarily families and business travelers.
- **Reviews per Month:** 0.63, making it the second most reviewed cluster.
- **Top Hosts:** Managed by both companies and individuals like 'Luís' and 'Albertino'.
- **Locations:** Popular neighborhoods including Santa Maria Maior, Arroios, Misericórdia, and Cascais e Estoril.

# Results and Discussion

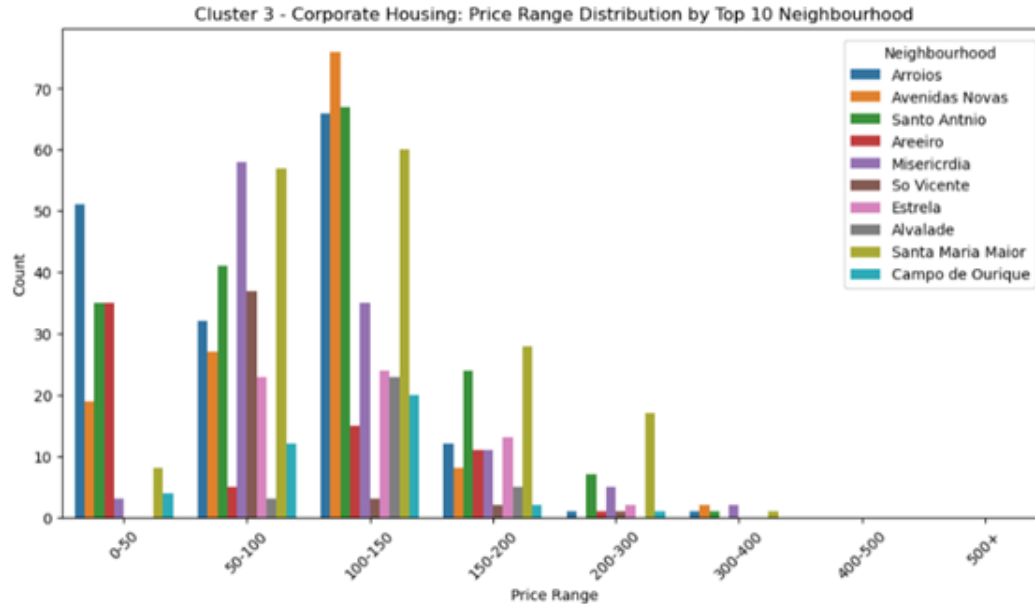
- **Cluster 2:** Budget-Friendly Rentals: Containing 5,231 data points, this cluster is characterized by affordability



- **Average Price:** €96 per night
- **Room Types:** Predominantly home/apartment (82%) and private rooms (17%).
- **Minimum Night Requirements:** Average around 2 nights.
- **Rental Frequency:** High, frequently booked, with 3.36 reviews per month, the highest among clusters.
- **Leading Hosts:** 'Alexandra Pedro and Team' and 'Sarah & Pedro', specializing in cost-effective accommodations.
- **Locations:** Accessible neighborhoods such as Santa Maria Maior, Misericórdia, Arroios, and São Vicente.

# Results and Discussion

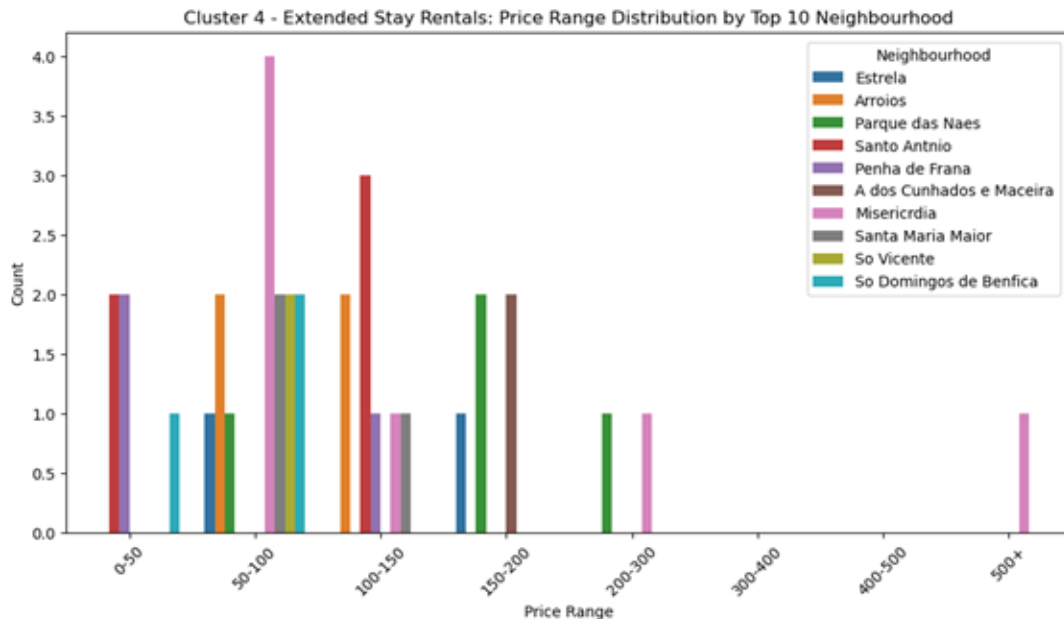
- **Cluster 3:** Corporate Housing: Including 1,191 data points, this cluster features listings with a moderate average



- **Average Price:** €107 per night.
- **Room Types:** Generally entire homes/apartments.
- **Minimum Night Requirements:** Average around 13 nights rental
- **Frequency:** Moderate, with 0.54 reviews per month.
- **Leading Hosts:** Managed by property companies like 'Feels Like Home' and 'Blueground'.
- **Locations:** Central neighborhoods including Santo António, Santa Maria Maior, Arroios, and Avenidas Novas.

# Results and Discussion

- **Cluster 4:** Extended Stay Rentals smallest cluster with 46 data points



- **Average Price:** €134 per night
- **Room Types:** Predominantly entire homes/apartments (93%).
- **Minimum Night Requirements:** Exceptionally high, averaging around 269 nights.
- **Rental Frequency:** Moderate, with 0.54 reviews per month
- **Leading Hosts:** Managed by companies 'Martinho Hotels' and hosts such as 'Paulo Eduardo', and 'Tiago'.
- **Locations:** Neighborhoods including Misericórdia, Santo António, Arroios, and Parque das Nações.

# Conclusions

We applied unsupervised machine learning algorithms, selecting K-Means for its balanced clusters, and identified five distinct Airbnb groups:

- **Luxury Rentals:** High-end properties in exclusive neighborhoods.
- **Standard Rentals:** Moderate pricing for families and business travelers.
- **Budget-Friendly Rentals:** Affordable options with high demand.
- **Corporate Housing:** Long-term stays in central areas for business travelers.
- **Extended Stay Rentals:** Cater to expatriates with high minimum stay requirements.

These clusters reveal the diverse nature of Lisbon's Airbnb market, offering insights for optimizing listings, making informed investment decisions, and finding suitable accommodations.

# Recommendations

- Based on our findings, we recommend implementing several actionable strategies:
  - **Cluster 0** (Luxury Rentals): Partnering with high-end travel agencies to reach wealthy travelers seeking luxury accommodations.
  - **Cluster 1** (Standard Rentals): Promoting listings in popular neighborhoods through local tourism websites and business travel platforms to increase visibility.
  - **Cluster 2** (Budget-Friendly Rentals): Encouraging guests to leave reviews by offering small incentives, as high review counts can attract more bookings. Promoting listings on budget travel websites to reach the target audience.

# Recommendations

- **Cluster 3** (Corporate Housings): Providing business travelers with essentials such as high-speed internet, workspaces, and early check-in/late check-out options. Partnering with local businesses and corporations to offer accommodation packages for relocating employees and business trips.
- **Cluster 4** (Extended Stays Rentals): Offering long-term stay discounts and flexible lease options to attract expatriates and long-term travelers.



# Thank you

Questions?