

# Booking.com and Airbnb Price Prediction Project

## - Amsterdam Edition -

Eppu Henriksson



# Content

1. Introduction
2. Web Scraping and Data Wrangling
3. Data Analysis and Predictive Model
4. Review Analysis
5. Dashboard
6. Conclusions



# 1. Introduction

- How much should you charge?
- What factors have the biggest impact on room rate?
- Is there an easy way to set optimal rate based on room features?



# 2. Web Scraping and Data Wrangling

Booking.com web scraping

- Selenium: Webdriver, popups, dynamic scrolling, ‘Load more’ button
- Random wait time between clicks
- Human-like scrolling
- Page slicing

Airbnb dataset, [link](#)

Booking.com reviews, [link](#)

Airbnb reviews, [link](#)

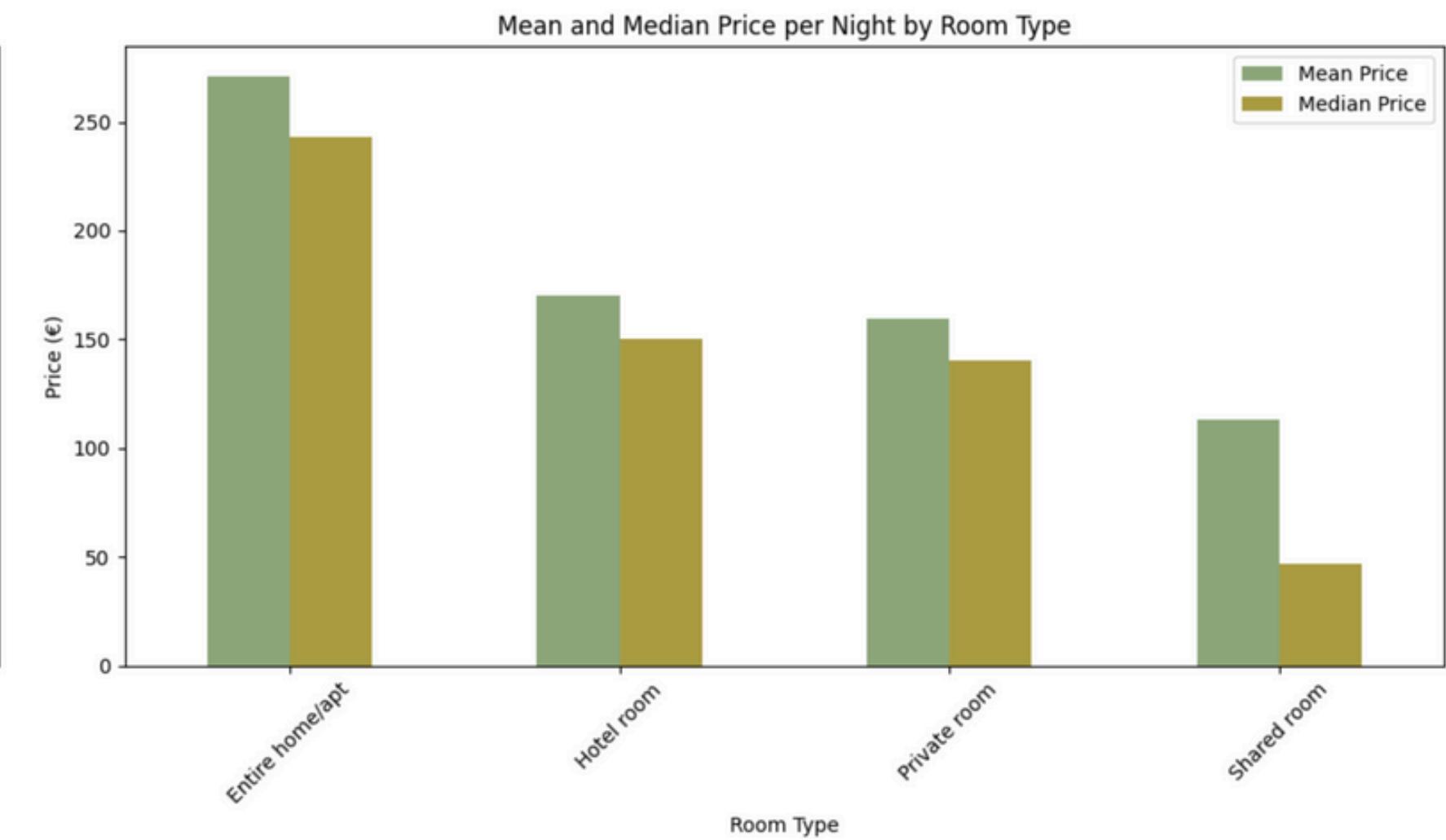
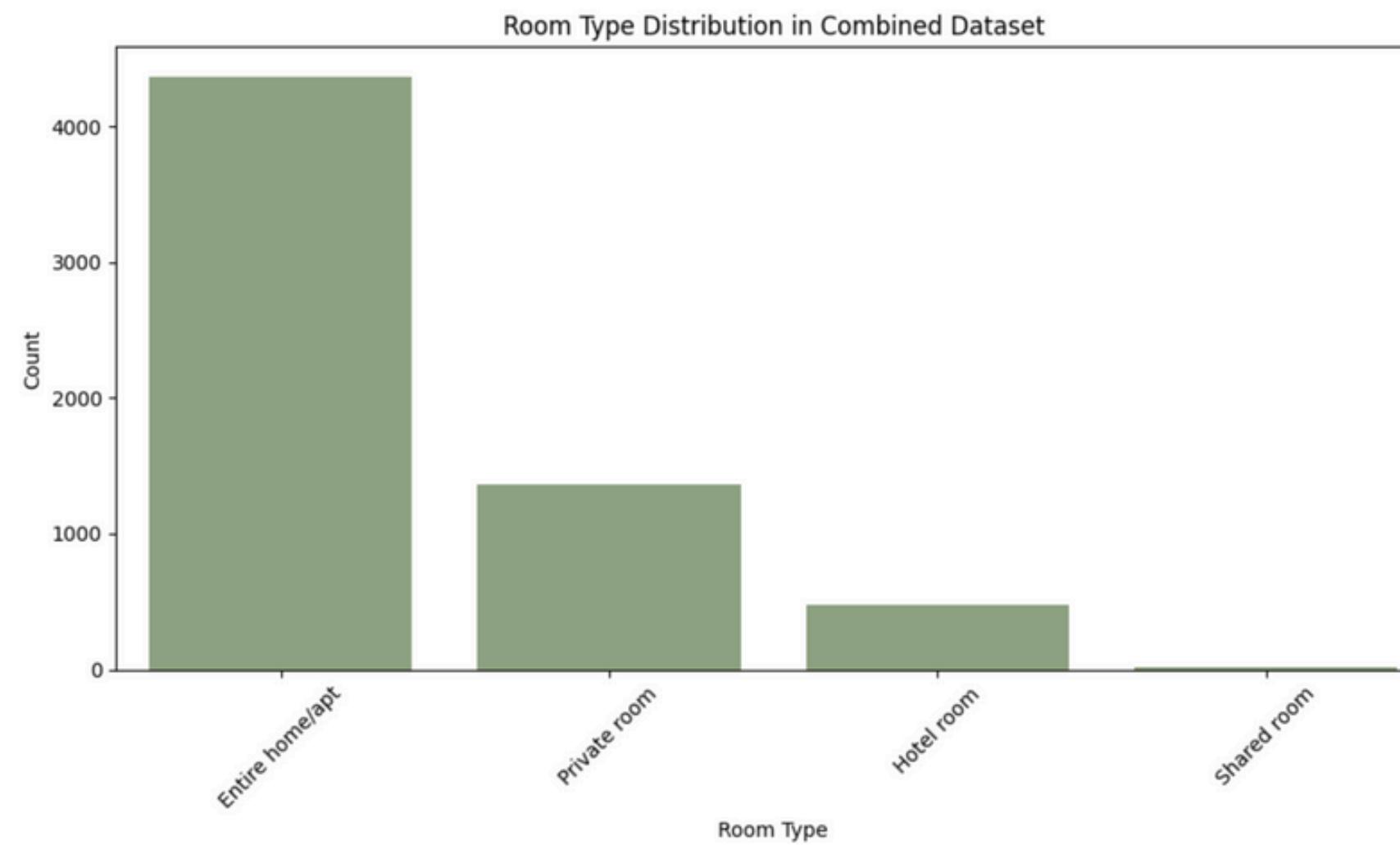


# 2. Web Scraping and Data Wrangling

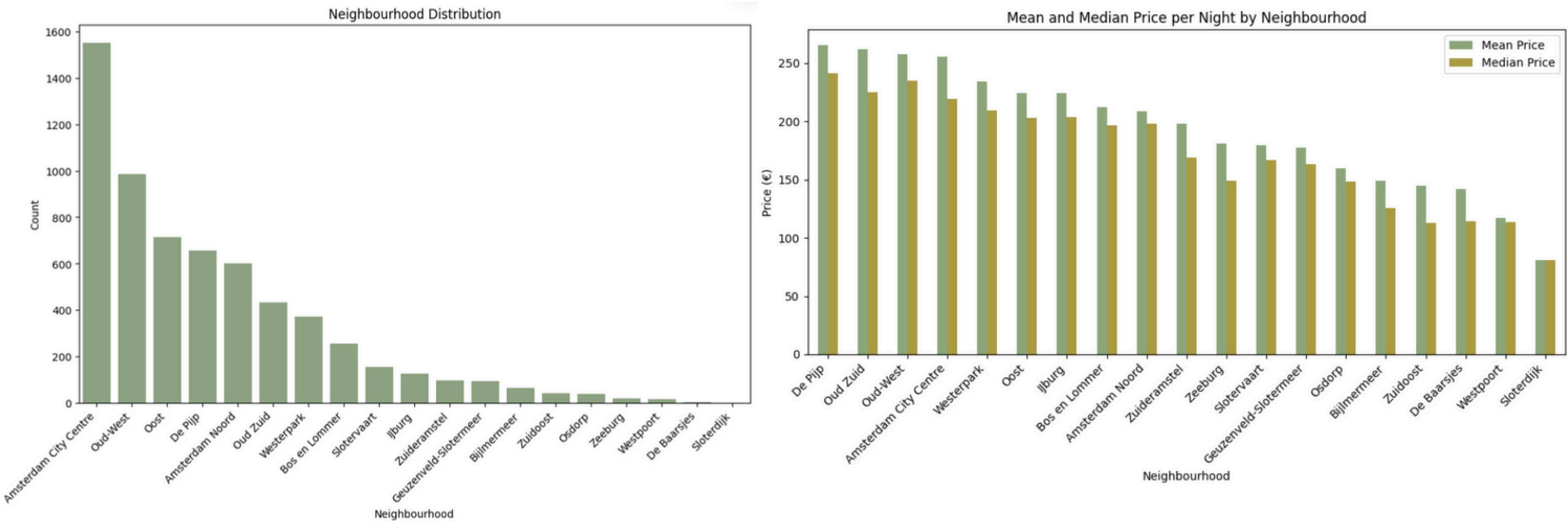
Data	Booking.com	Airbnb
Rows, pre wrangling	800	10 168
Rows, post wrangling	656	5 568
Price per night	Avg 15-22 Nov 2025	Avg Q2 2025
Star rating		
Review score	'1 - 10	'1 - 5
Room types	199	4
Part of city	19	22
Distance from centre		
Address		
Latitude/Longitude		
URL		

- Geopy to convert address to latitude and longitude
- Haversine formula to calculate distance from centre

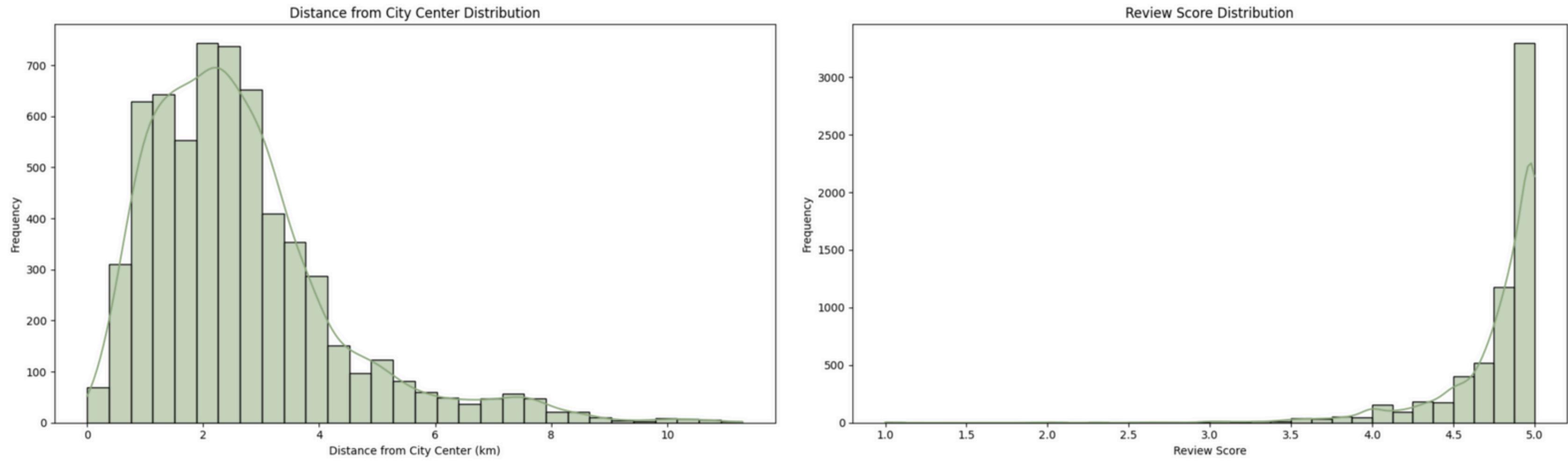
# 3. Data Analysis and Predictive model



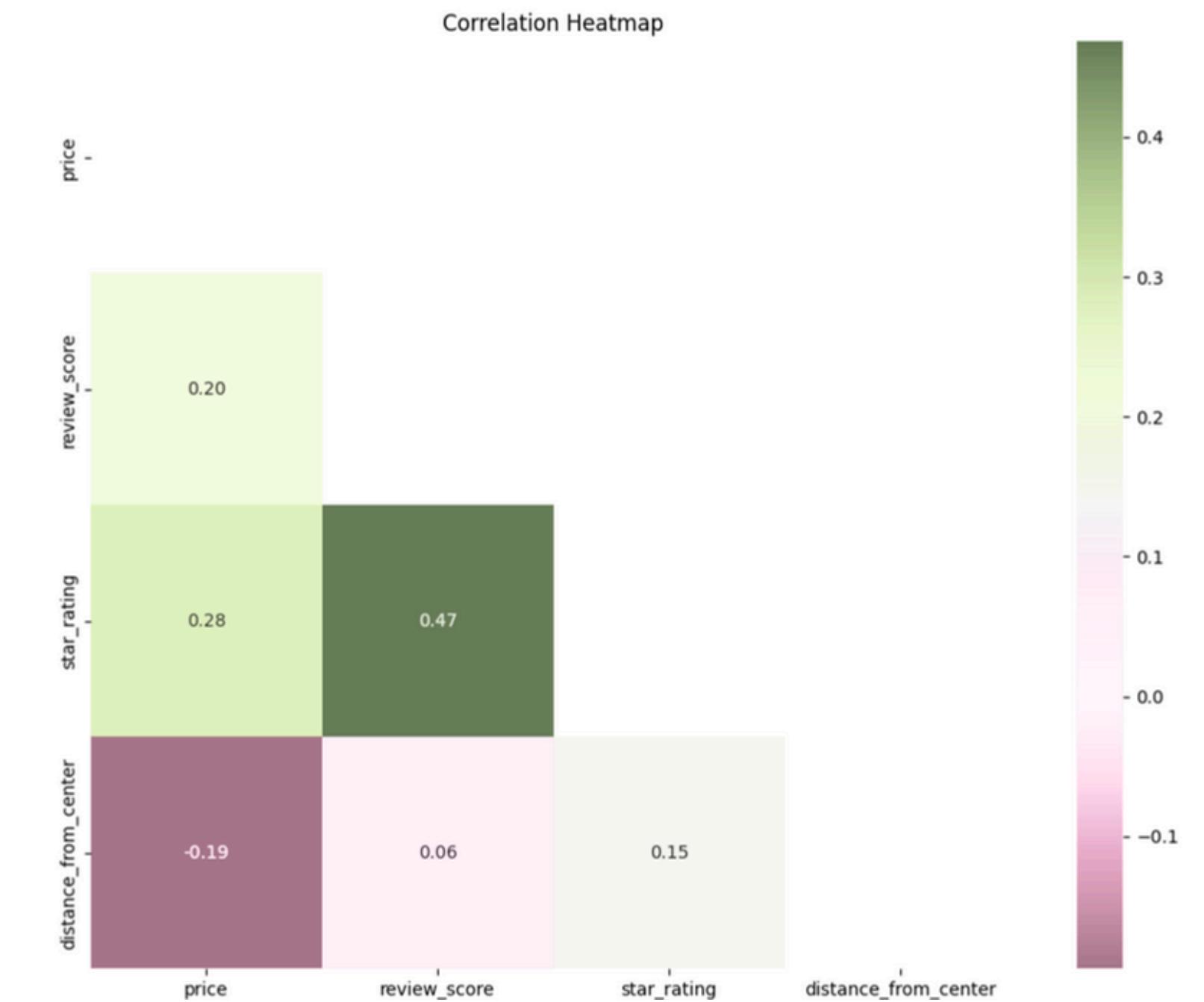
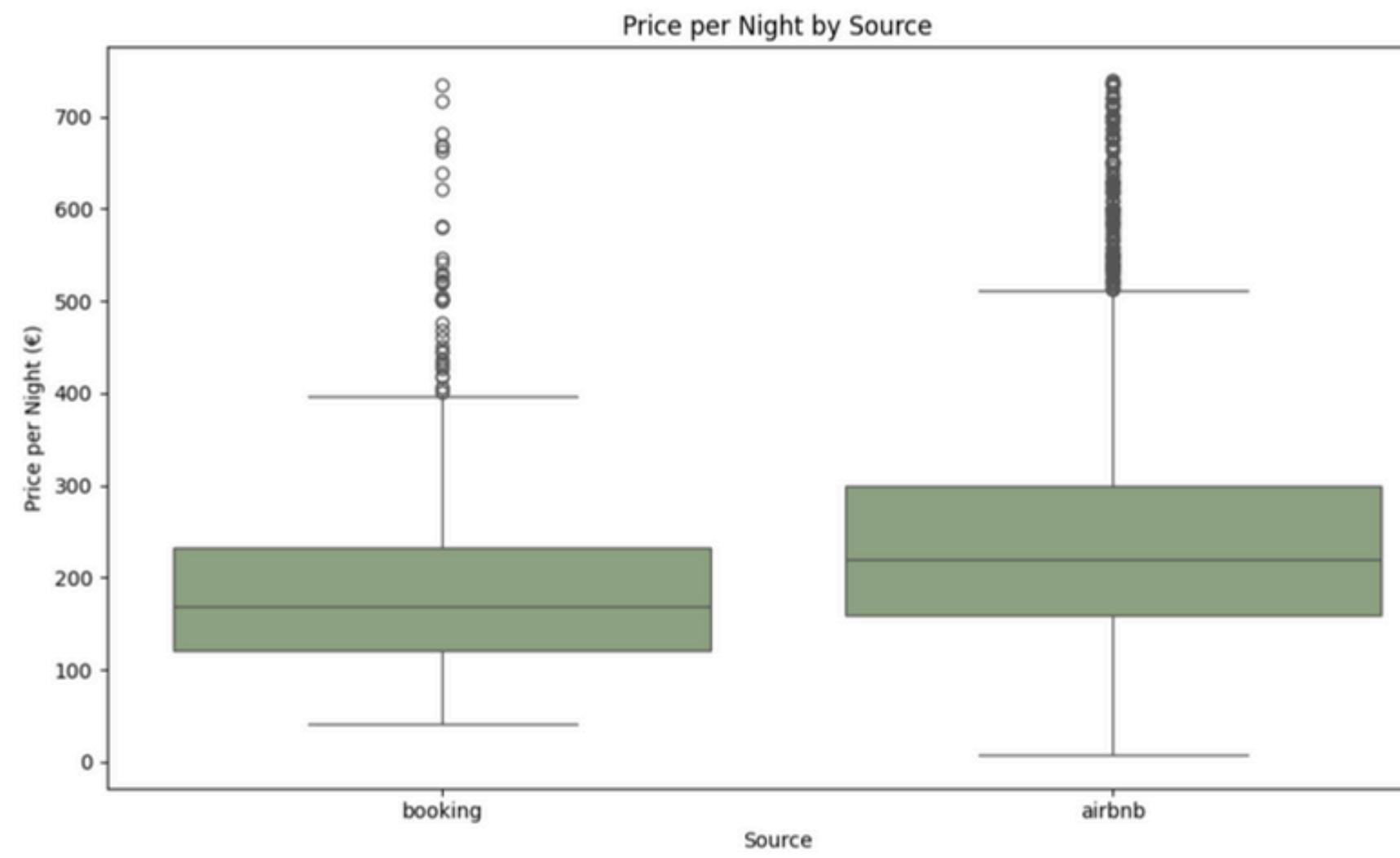
# 3. Data Analysis and Predictive model



# 3. Data Analysis and Predictive model



# 3. Data Analysis and Predictive model



# 3. Data Analysis and Predictive model

Feature selection:

- Review score
- Distance from center
- Latitude/Longitude or Neighborhood
- Room type

Target:

- Price

Trained models:

- Linear regression
- Lasso, Ridge and ElasticNet
- Decision tree
- Random forest
- XGBoost



# 3. Data Analysis and Predictive model

Least bad model was linear regression with neighbourhood encoded

Predictions	Test	Train
R2 score	0.24	0.25
MAE	78.3	77.19
RMSE	11248.59	10789.01

# 3. Data Analysis and Predictive model

review_score	167.635147
room_type_Entire home/apt	101.230791
neighbourhood_IJburg	62.827058
neighbourhood_Zuidoost	60.591528
neighbourhood_Oud Zuid	29.142238
neighbourhood_Bijlmermeer	27.793794
neighbourhood_Osdorp	23.116674
neighbourhood_Zuideramstel	17.684017
room_type_Hotel room	15.431922
neighbourhood_Geuzenveld-Slotermeer	14.495777
neighbourhood_Amsterdam City Centre	6.854909
neighbourhood_De Pijp	4.327289
neighbourhood_Oud-West	0.731967
neighbourhood_Zeeburg	-0.195229
neighbourhood_Amsterdam Noord	-3.214522
neighbourhood_Slotervaart	-6.281374
room_type_Private room	-13.459084
neighbourhood_Oost	-18.838849
neighbourhood_Westerpark	-30.503768
neighbourhood_Westpoort	-33.941793
neighbourhood_Bos en Lommer	-37.688727
neighbourhood_De Baarsjes	-50.063485
neighbourhood_Sloterdijk	-66.837505
room_type_Shared room	-103.203629
distance_from_center	-218.236544

# 4. Review Analysis

Reviews	Booking.com	Airbnb
Rows, pre wrangling	515 738	483 680
Rows, post wrangling	515 738	341 515
Language	en	mixed
Split between +/-		

- Language detection
- Vader Sentiment Analyzer
- Stopword removal with customised words
- Creation of unigram, bigram and trigram wordclouds

# 4. Review Analysis, Booking.com

## Positive Trigrams

staff friendly helpful  
staff good location within walking distance  
would definitely stay friendly staff good  
hotel great location nothing much trouble  
good value money friendly staff great  
location great staff  
staff helpful friendly staff great location  
hotel good location  
**friendly helpful staff**  
helpful friendly staff great location friendly  
staff really friendly location friendly staff  
great value money staff extremely helpful

## Negative Trigrams

room th floor making facilities room  
front desk staff tea coffee room  
**room little small**  
beds pushed together far city center  
**room bit small**  
four star hotel tea coffee facilities  
coffee facilities room  
**tea coffee making**  
room quite small two single beds  
good value money air conditioning room tea making facilities  
**coffee making facilities**  
breakfast included price breakfast could better

# 4. Review Analysis, Airbnb

Positive Trigrams

close public transport  
central train station  
feel right home 10 minute walk  
location easy access  
red light district  
20 minute walk  
close public transportation  
walk central station  
home away home felt like home  
anne frank house  
location walking distance  
easy walking distance 15 minute walk  
feel like home close city center

Negative Trigrams

right outside door far city center  
steep narrow stairs close tram stop  
central train station  
red light district  
walking distance central  
anne frank house  
walk central station water pressure shower  
bit noisy night minute walk away  
location walking distance 10 minute walk  
location location location  
worst airbnb experience  
close city center 20 minute walk  
close public transport flights steep stairs

# 5. Dashboard

[Link to Tableau](#)



# 6. Conclusions

- Benefit of combining data from two sources was questionable
- Not enough features and/or records to train a good predictive model
  - too few room types
  - star rating
  - no. of max guests
  - size of the room/home
- Randomness in pricing?
- Location is very important attribute for travellers
- Travellers appreciate good and friendly service, and can really drive good review score





A scenic view of a canal in Amsterdam, featuring traditional Dutch houses with red roofs and white-framed windows. The buildings are built along the water's edge, with some having multiple stories. In the foreground, there are several trees with autumn-colored leaves, ranging from yellow to green. A street lamp is visible on the right side of the frame.

Thank you!  
Eppu Henriksson  
Github: [eppuoskari](https://github.com/eppuoskari)