AIRBNB
AN EXPLORATORY DATA ANALYSIS

Made by:
Yash Patil
Shreesh Patil
Hrishikesh Patil
Adwait Patkhedkar

Before we begin...

Why did we choose Airbnb as the project and what are the benefits of the insights provided:

◦ We wanted to work with a real life dataset and focus on the cleaning and EDA (pure data science and data analysis part)

◦ We also wanted to test out a few models for price prediction

◦ The insights provided in this presentation will help new hosts to simplify decision making

◦ The insights also set the stage for more complex insights which are possible with more feature extraction of the data

Our Insights

◦ We have made many observations

◦ However we will present the best / non obvious observations which are made so far and also some data representations whichdon't directly give insights but form the base for more complex insights

◦ The observations include : East Coast, West Coast, Mid and South US

Insights from geospatial graph
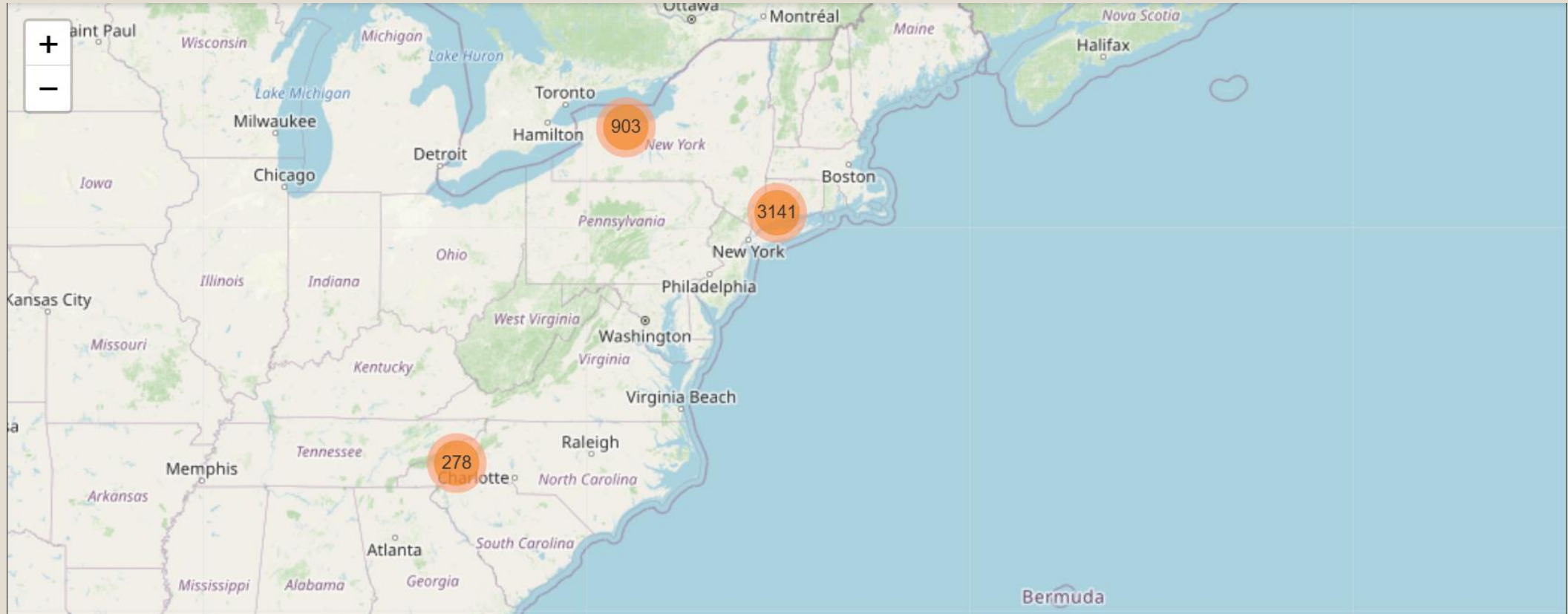
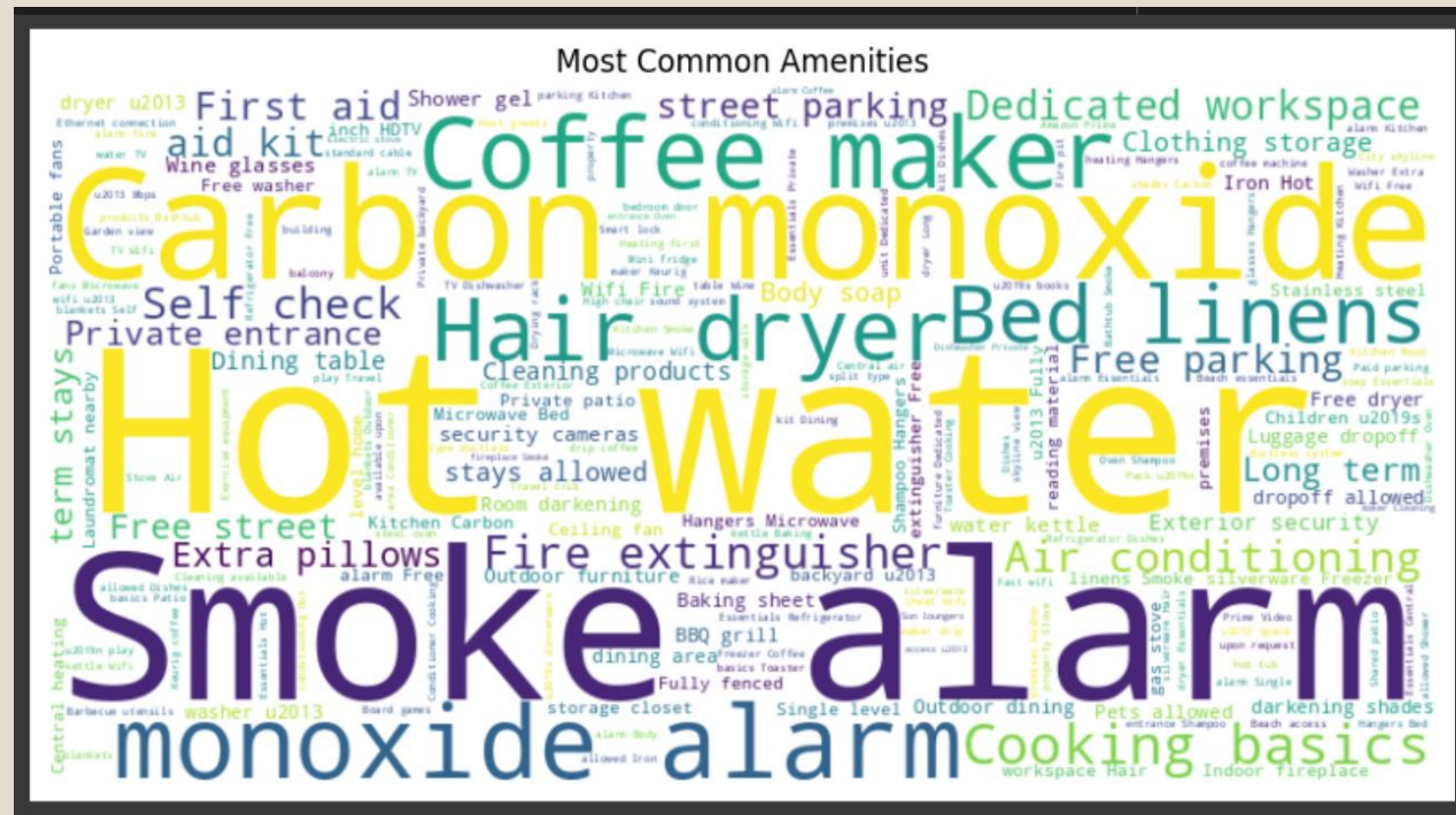A cluster map with neighbourhood overviews created using folium

Cntd…

◦ Rooms are located near coastlines, city centres and away from forest areas

◦ In Florida known for its beach resorts, amusement parks, warm and sunny climate, room clusters are more located near casinos, which indicates that operating rooms near casino areas more profitable. Florida also has less strict laws on operations of casinos. Fort Lauderdale is an emerging area competing with Miami and Hollywood

◦ The Ashville city in North Carolina has clusters of rooms near the forests, making it a good destination for relaxing vacations

◦ Rochester has rooms more near the defunct railways and particularly large number of rooms near Kodak's original headquarters

◦ Islands can host rooms if they are connected by airports (eg. Long Island, Newshoreham Island in NYC) with Staten Island an emerging island

◦ The summer tourism is higher in east coast (not with west coast) owing due to the highly cold winters
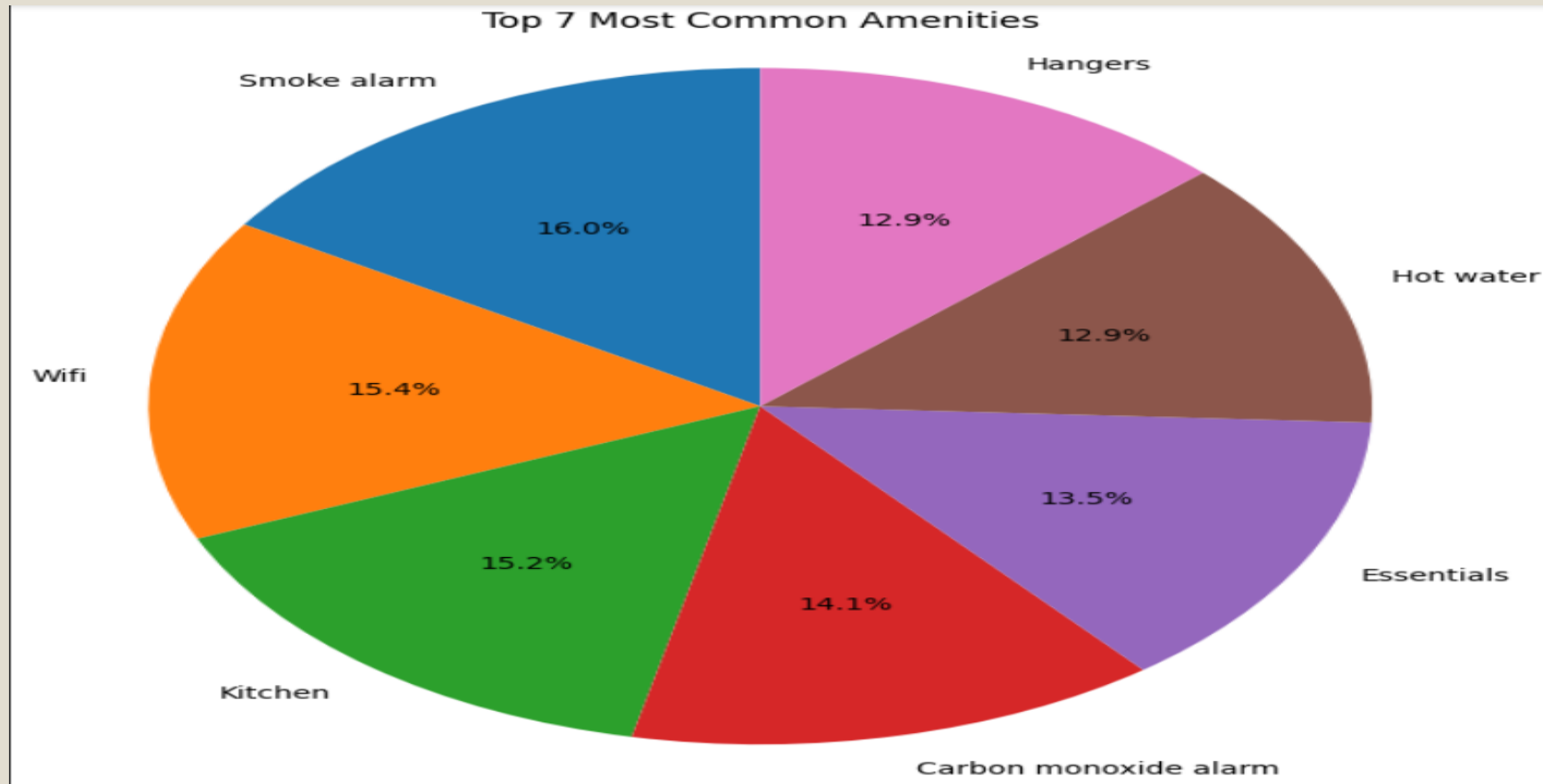
Cntd..

◦ California has a peculiarity: In emerging localities and small towns with large Asian populations

◦ Proximity to public transport in commercial cities like NYC, LA

◦ College students are more welcome in Cambridge and Boston with presence of premier institutes like Harvard , MIT , Boston University etc

◦ California has a peculiarity: In emerging localities and small towns with large Asian populations, large room clusters with positive descriptions are found
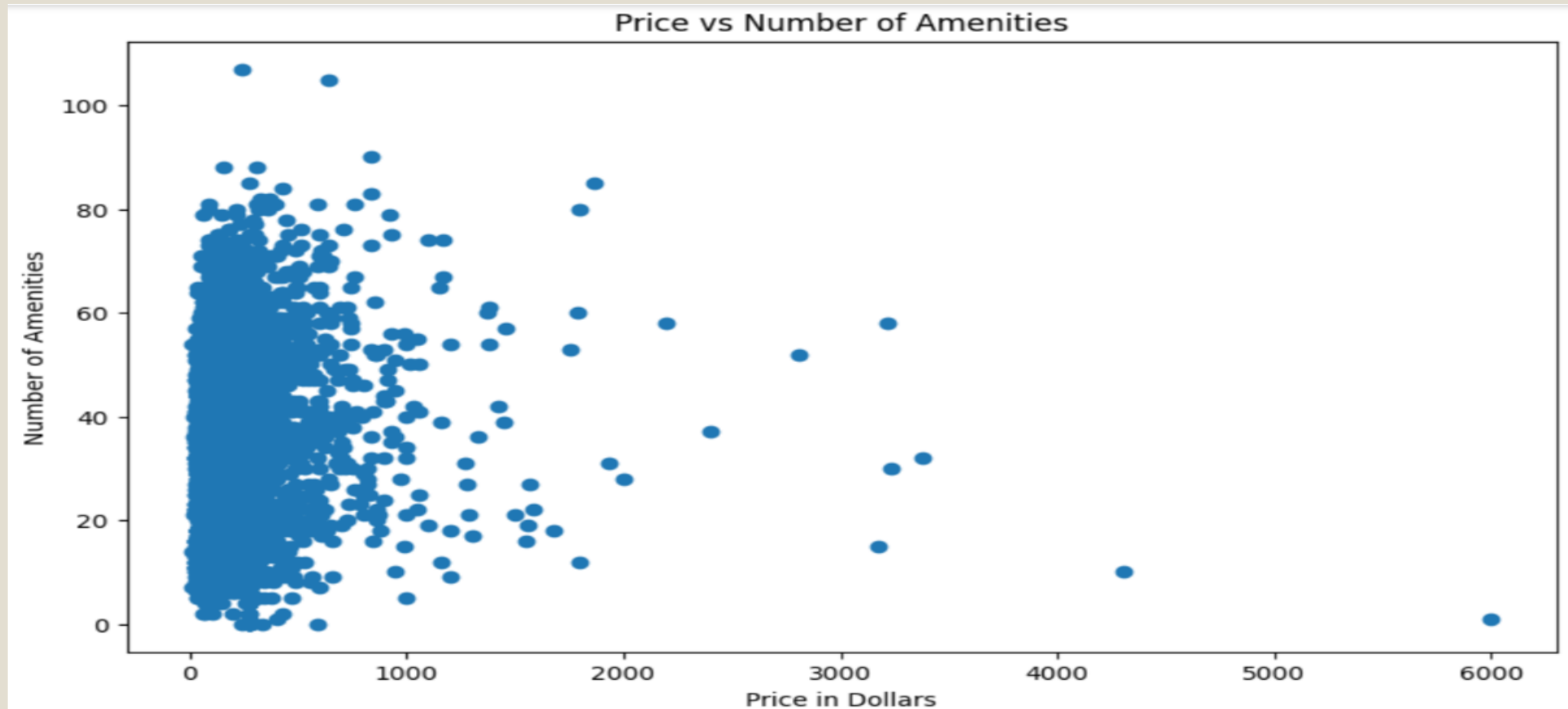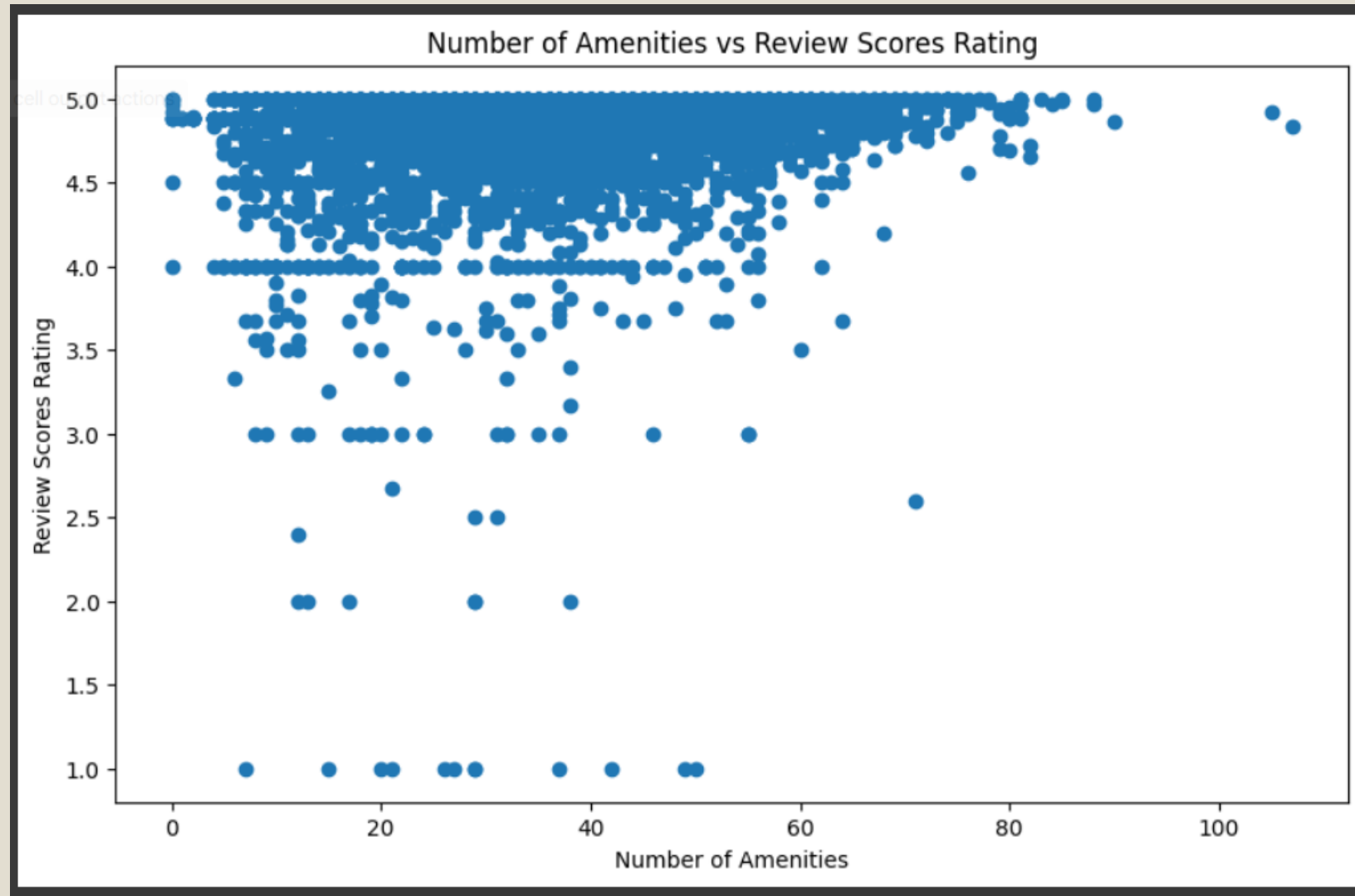
Insights from amenities

Word Cloud

Cntd….

Top 7 Most Common Amenities

- Hangers — 12.9%
- Smoke alarm — 16.0%
- Hot water — 12.9%
- Wifi — 15.4%
- Essentials — 13.5%
- Kitchen — 15.2%
- Carbon monoxide alarm — 14.1%

Pie chart of common amenities

Scatter Plot

Scatter Plot

Cntd..

- Carbon monoxide/smoke alarm are almost essentials

- Wi-fi is 2nd most popular amenities highlighting the need of the people to stay connected and other digital needs in the 21st century

- Hot water/geysers are necessary in winters and cities closer to the northern US

- Dishes/Silverware and kitchen(not in figure) are also observed frequently, even with the presence of restaurants nearby: Reason: As Airbnb allows anybody with space to rent, kitchen is present in almost every home, and a trend towards self-cooked meals can be seen. So in the future, hosts can include branded ready to eat food packets in the kitchen to further satisfy guests

- Guests prioritize essential comforts and safety features. Clearly number of amenities is not an important feature.
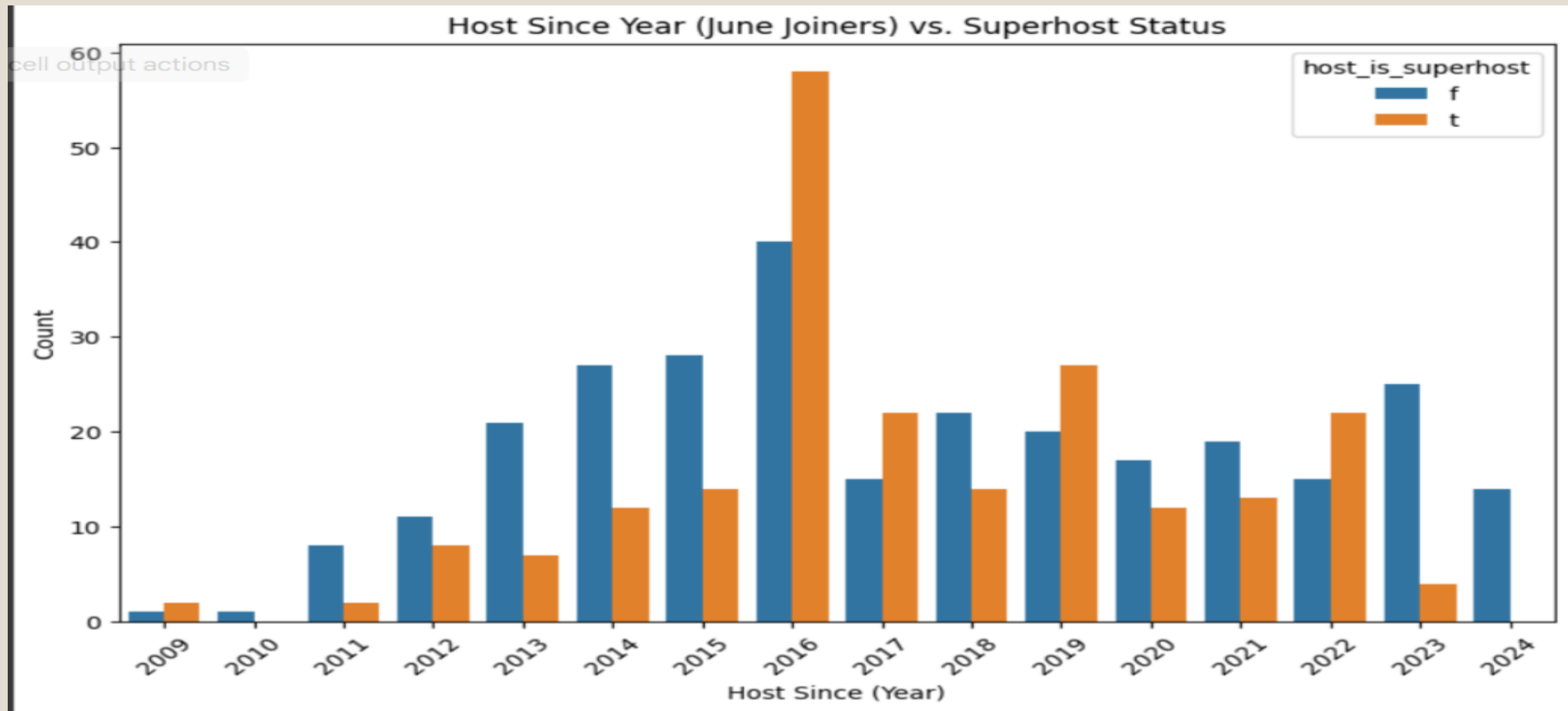
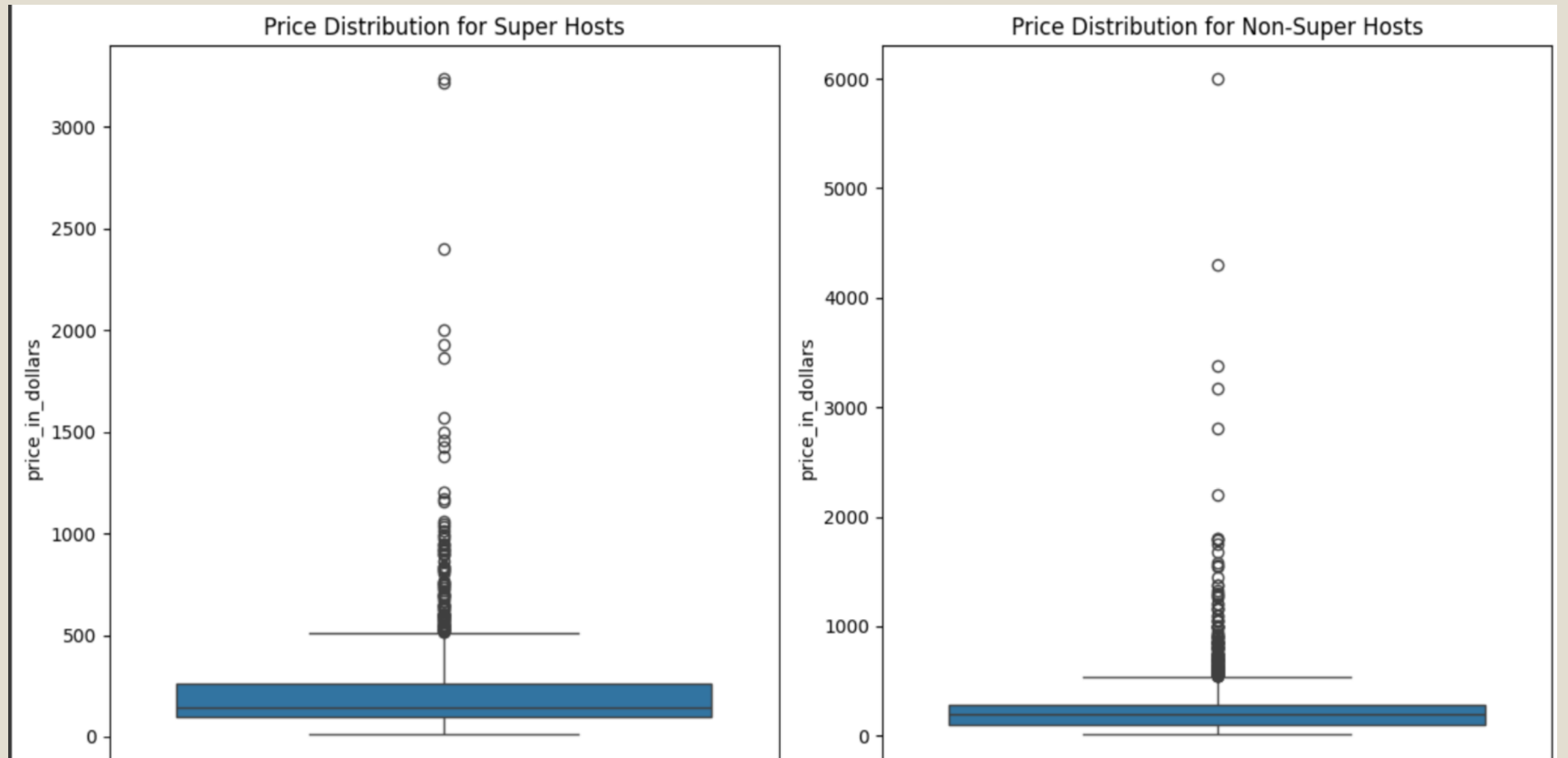# Best time to become a superhost
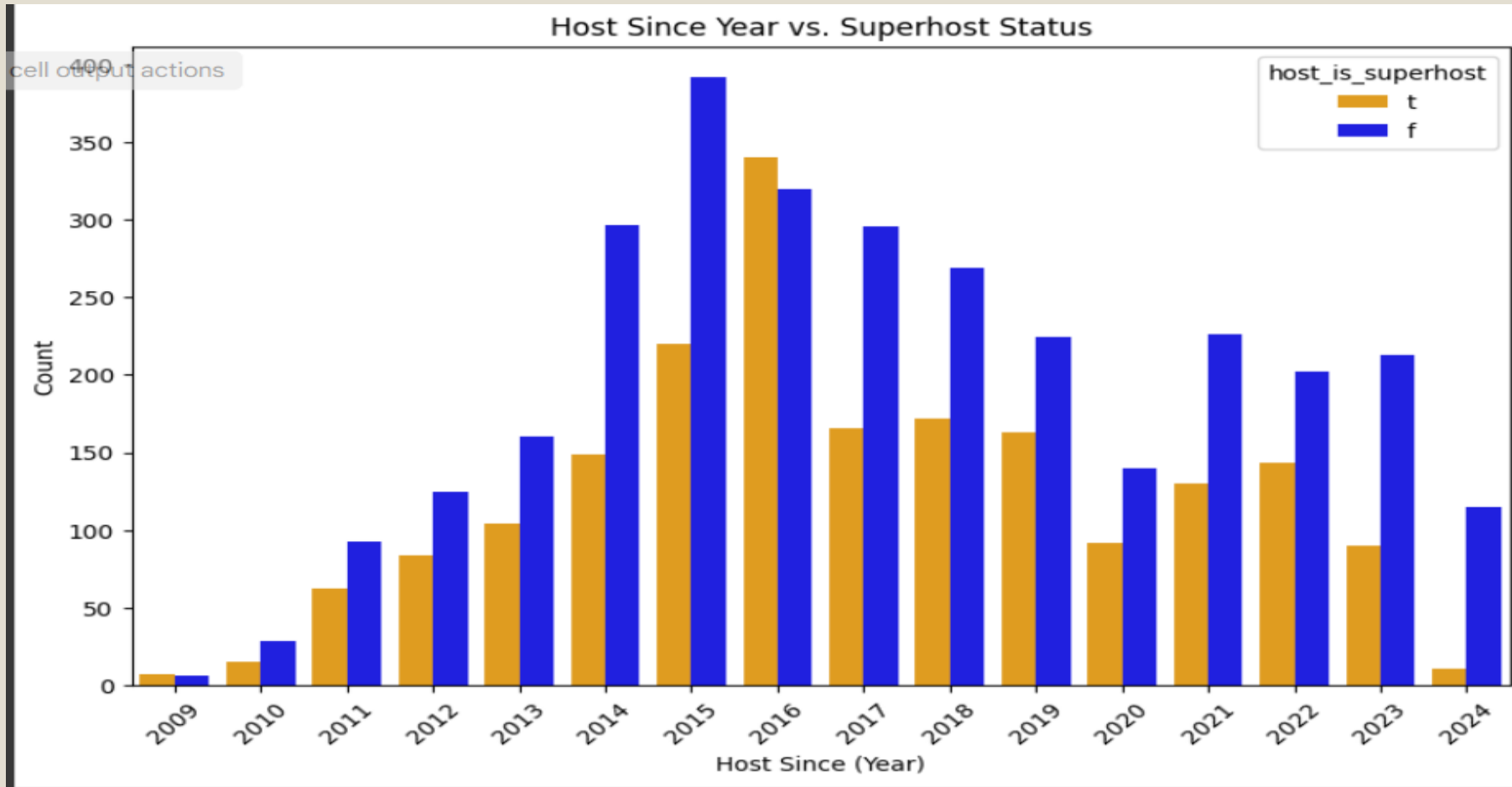
Bar chart

Cntd...
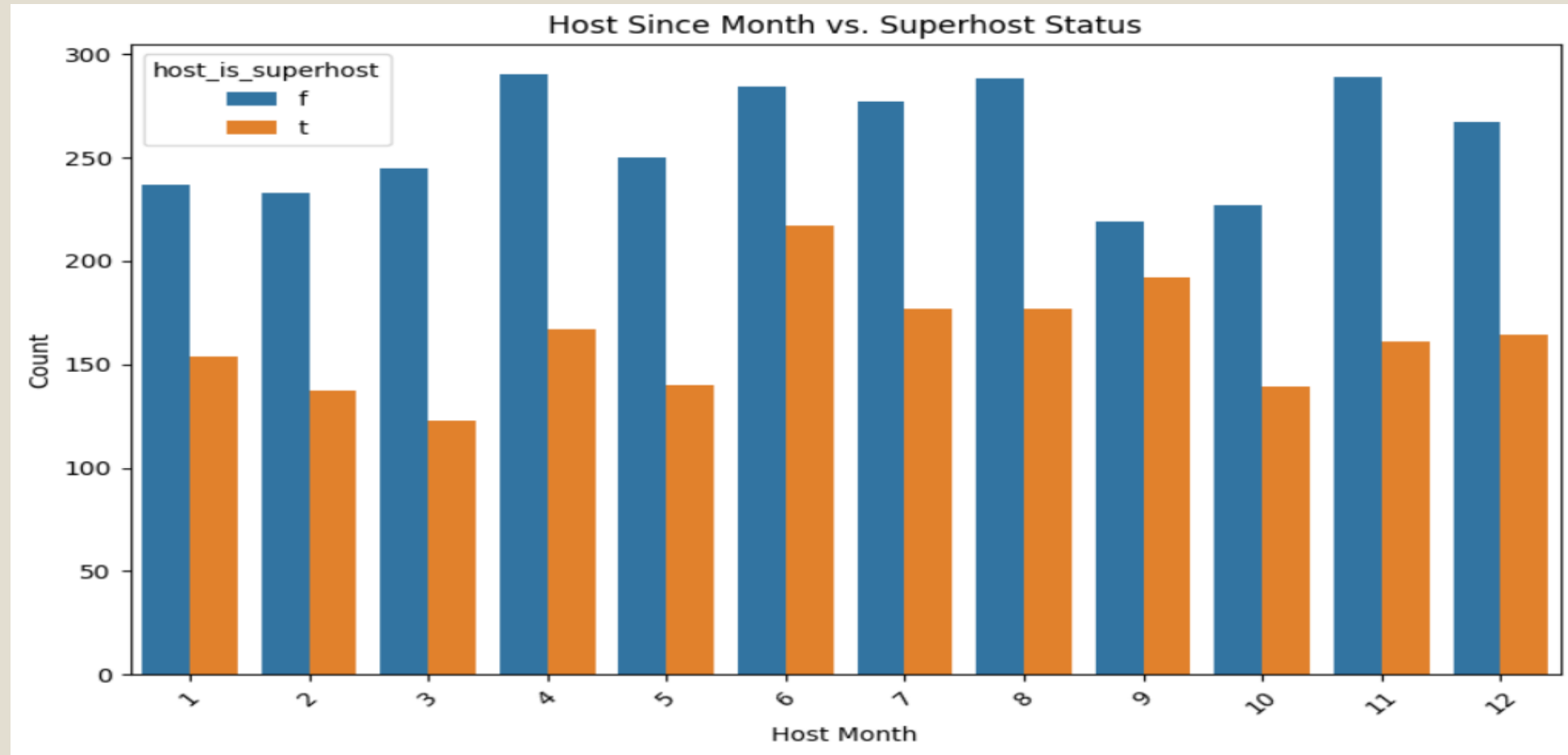
Bar chart
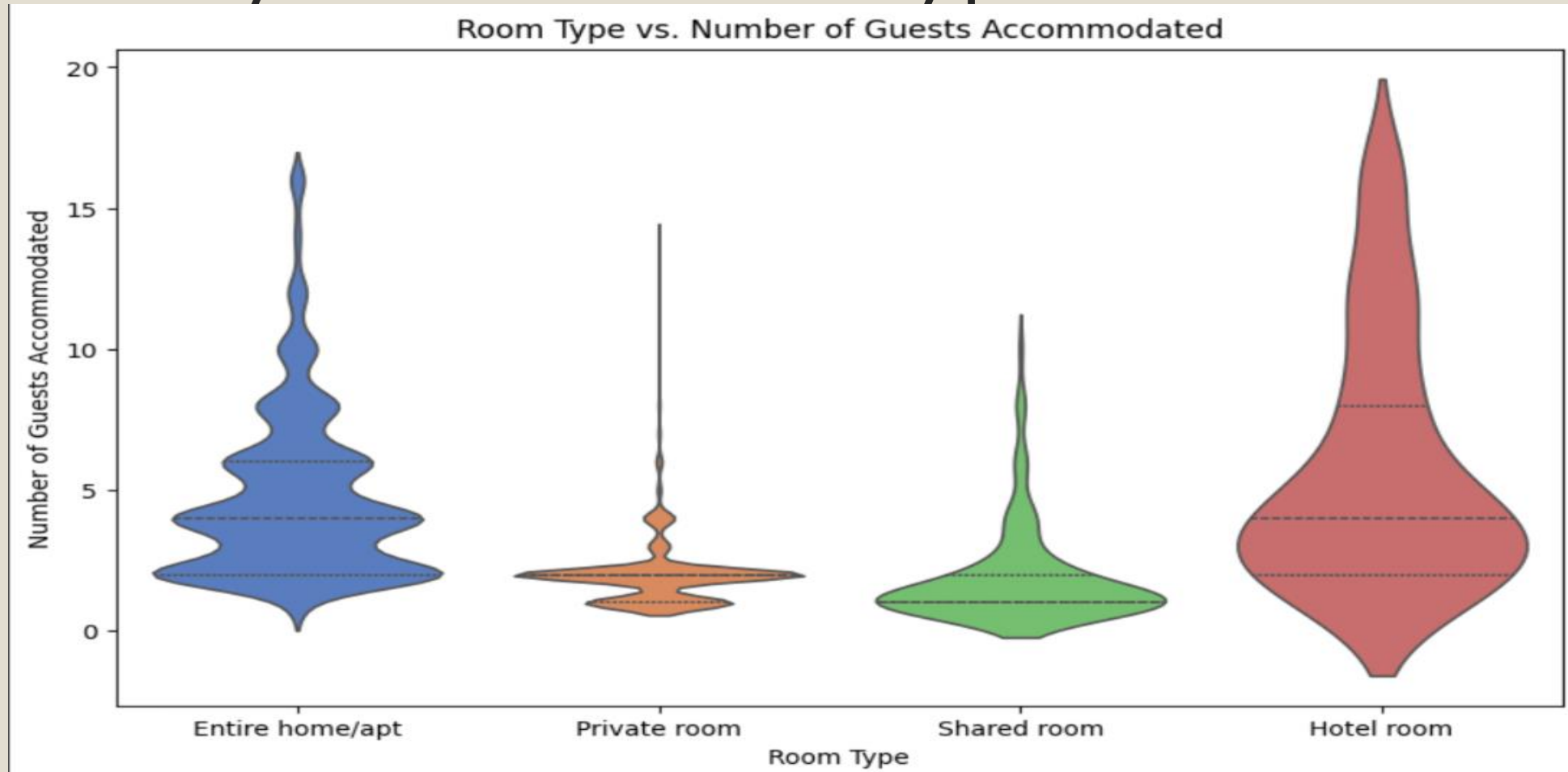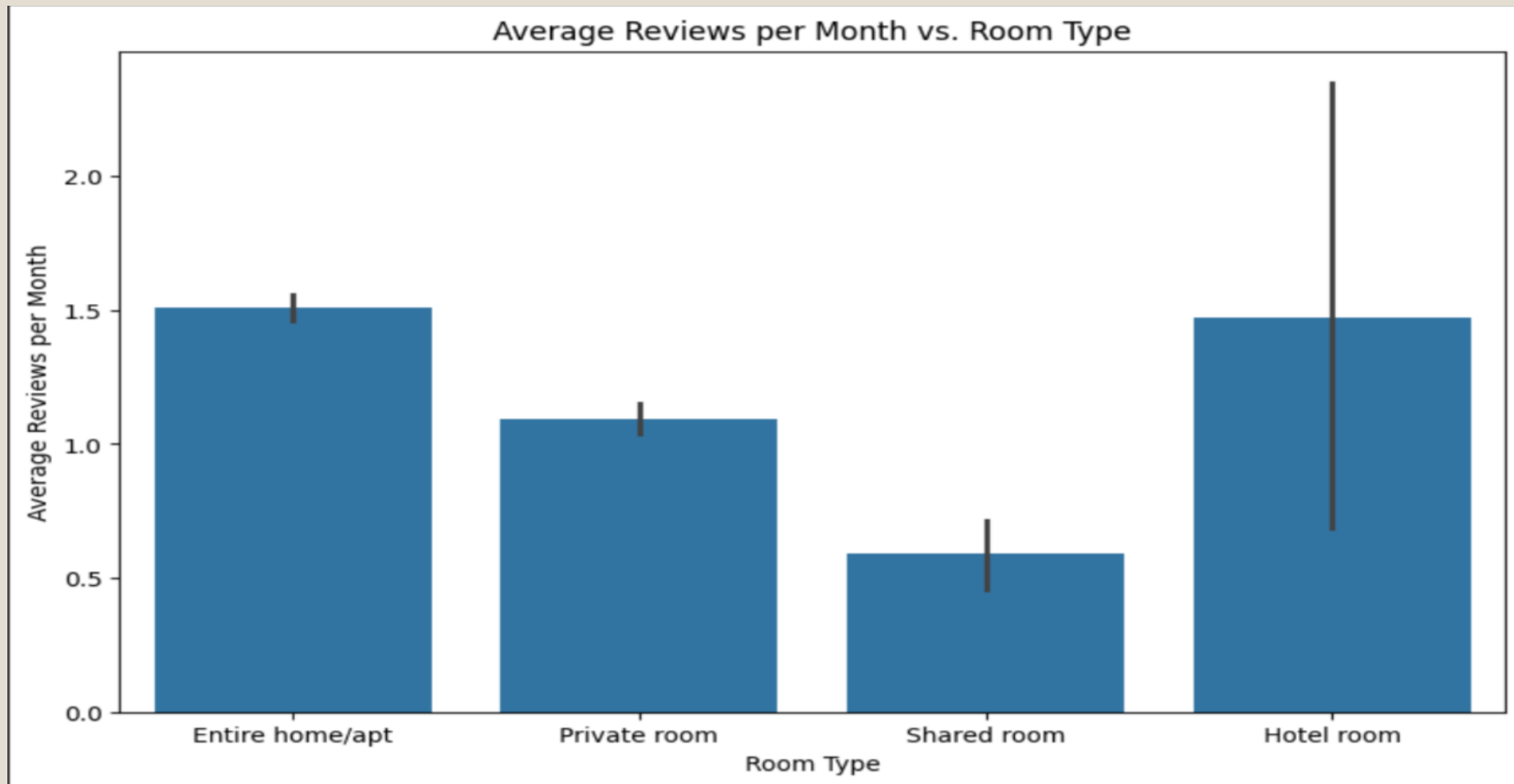
Cntd...

Bar chart

Box plots

Cntd…

◦ AirBNB grew quite in popularity between 2015-2017 due to various global sports events, new features, legal compliances and new feature called AirBNB experiences

◦ June is the best month to become a host and start hosting as summer vacations begin and for many people it is a new phase in their career and in general their life

◦ Hosts who begin in June are also more likely to be superhosts; superhosts have added benefits of higher visibility on the Airbnb app, badges showing signs of premium hospitality and other perks

◦ Also superhosts pricing is better and comparatively lesser than non-superhosts. We can also find the sweet spot for the general price of the room: somewhere between $10-$200

Analysis on room types


Violin Plot

Bar chart

Cntd...

◦ The most popular room type is entire home/apartment reflecting the guests emphasis on privacy and size.

◦ Market Segmentation:

◦ Entire Home/Apartment: Ideal for large groups or families, with a broad guest range. This segment can command premium prices.

◦ Private and Shared Rooms: Cater to solo travelers or couples, often budget-conscious, requiring competitive pricing.

◦ Hotel Rooms: Serve a range of guests, suitable for both solo travelers and small groups, allowing for dynamic pricing

◦ Also, shared room/ private rooms are the least preferred due to above reasons(like sharing bathrooms with other unknown guests , privacy concerns, lack of space etc)

◦ From the violin plot the average family size can be estimated: 0 to 5. This means that at least there is a possibility of a child present, so keeping child-friendly environment and amenities like toys can be helpful

Analysing the room ratings

Correlation Heatmap of Review Scores
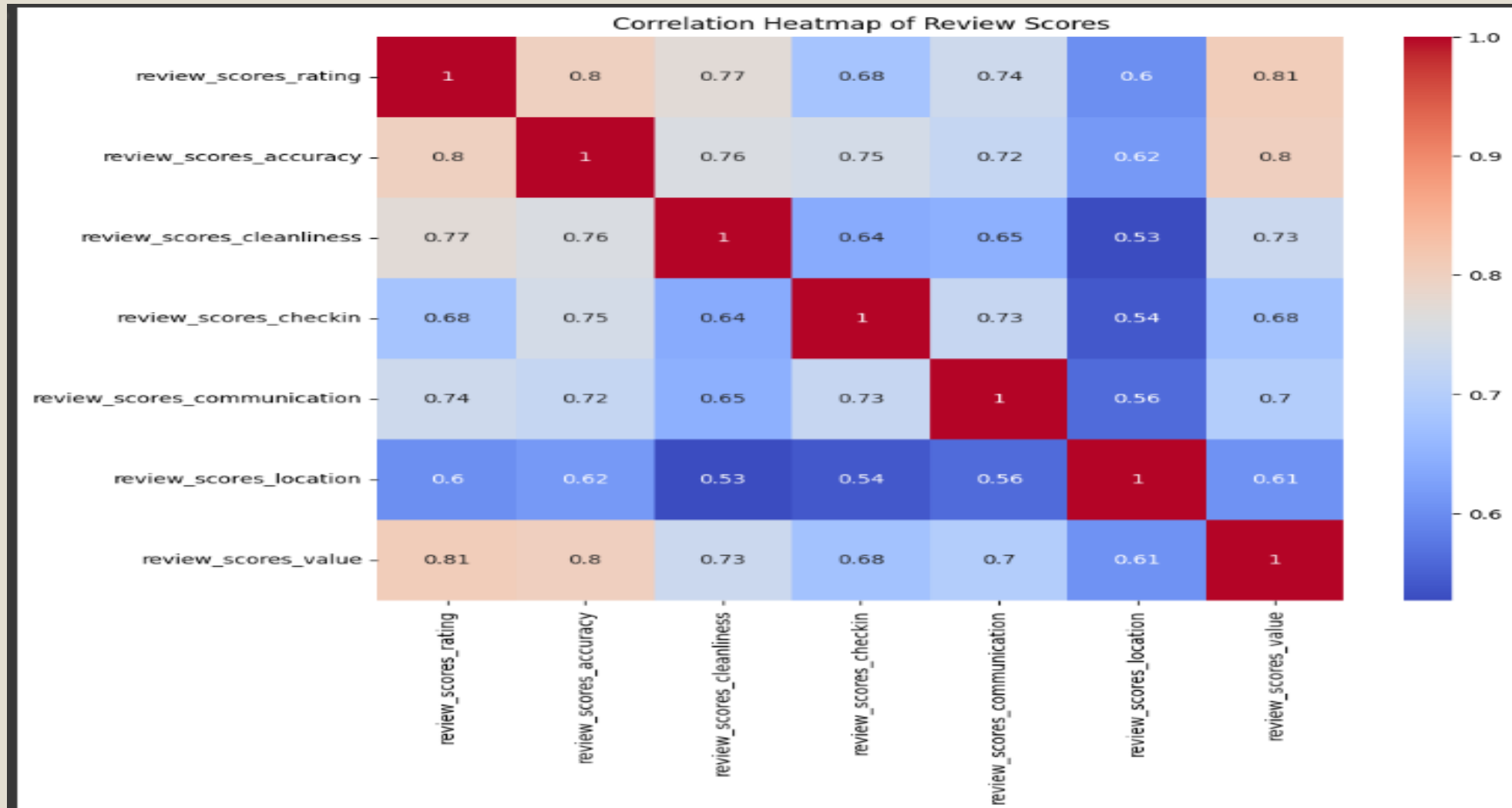
Average Review Scores Rating over Months
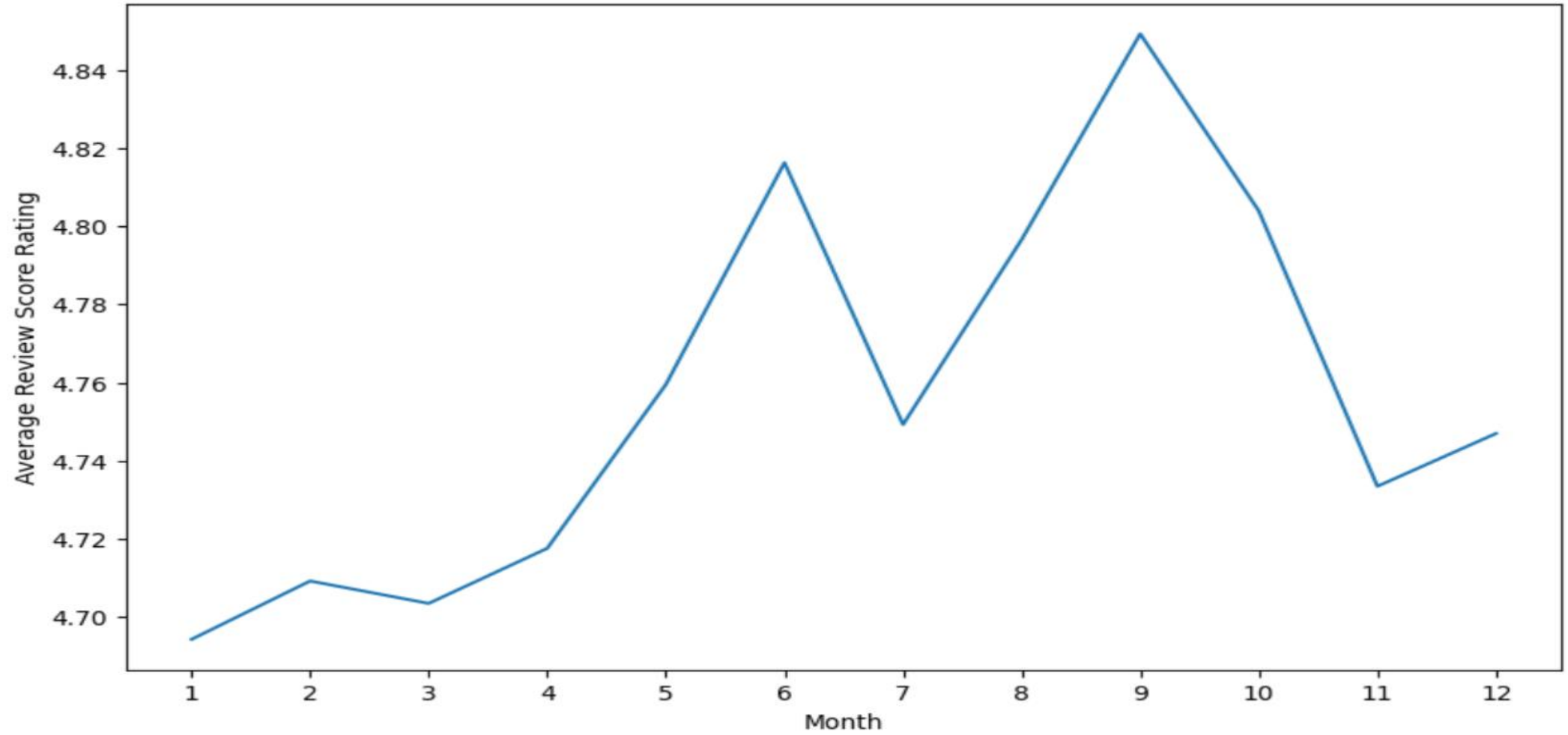
Cntd...

◦ Although all ratings are positively correlated which is obvious here are important insights:

◦ Review_scores_Rating (RSR) is the numerical average of the other 6 rating features

◦ RSR is a decent substitute for overall rating and is highly influenced by RS value and RS accuracy: this means that guests look out for authenticity and value for money the most

◦ Location and cleanliness are not as much positively correlated reason being that for locations near city centres, popular spots etc (prime locations) cleanliness can be compromised

◦ People are more likely to review positively in summers (month 6)and autumn(month 9) season: Americans are generally emotionally better during these months due to vacation periods, new beginnings, sales and discounts, good weather and a relaxed US work/study culture. Also, Jan and Dec have low ratings as they mark the beginning of the financial year, tough winters, stress and so on…

Treemaps for popular neighborhoods

Number of Listings per Neighborhood in Boston

Future Scope for EDA of our project

◦ This is just a subset of many possible observations (at most 2-3 variables analysed together in a graph)

◦ Many description type columns can be processed using NLP techniques and extract sentiments

◦ Profit and loss prediction for a given time duration is possible and many more…

Thank
You! ☺