

# comparebnb

A NEW WAY TO SEARCH AND COMPARE  
YOUR HOME AWAY FROM HOME

Miguel Gutierrez, Ashourrina Henshaw,  
Yue Shen, Clara Zhang

p. 01

## INTRODUCTION

# Agenda

p. 02-07

## EXPLORATORY DATA ANALYSIS

p. 08

## DEMO & SUPPORTING VISUALS

p. 09

## SUMMARY & NEXT STEPS



# Executive Summary

Airbnb is the world's largest online platform for short term rentals and tourism activities.

By the end of 2020, there were over **seven million** listings on Airbnb, run by **four million** hosts.

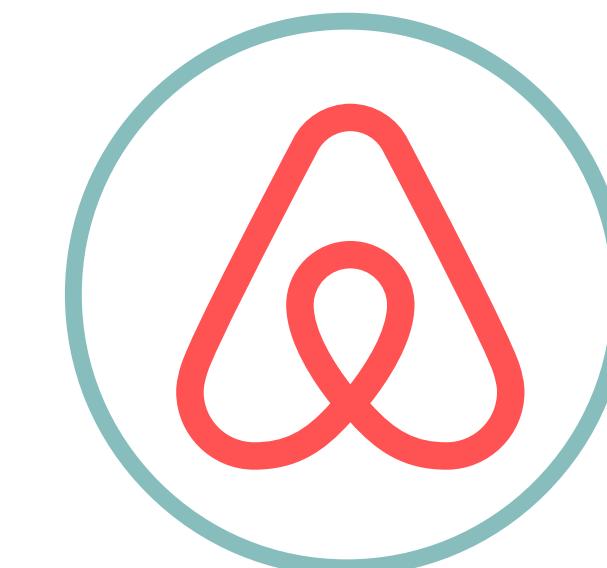
Comparebnb is here to help both users and hosts get true and transparent listing prices in a format allowing for easy comparisons.

## Our deliverables include:

- Two websites, including interactive rental searching
- Tableau dashboards
- Balsamiq wireframes
- Infographics
- Poster

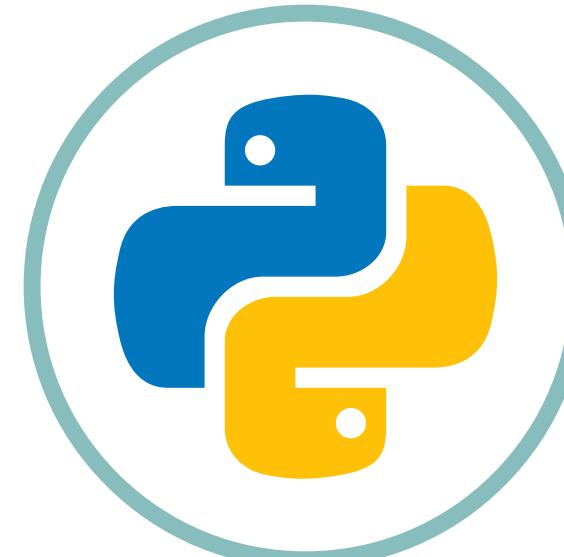


# Tools Used



Airbnb Data  
from Inside Airbnb

02



Python for  
Data Analysis



Balsamiq Wireframes  
for Initial Design

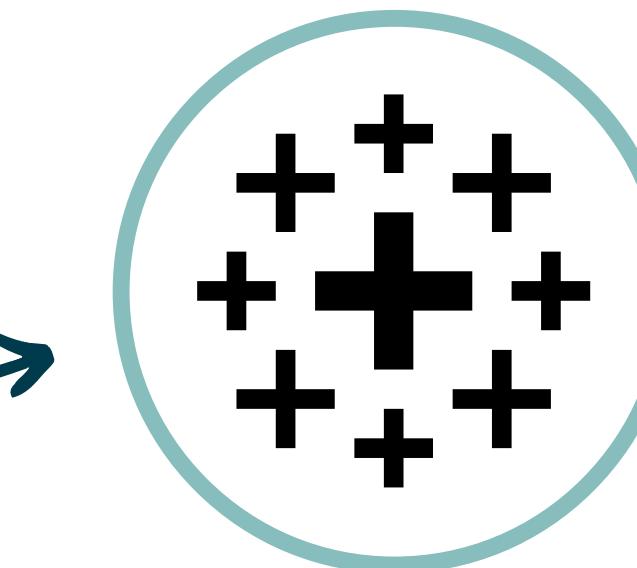
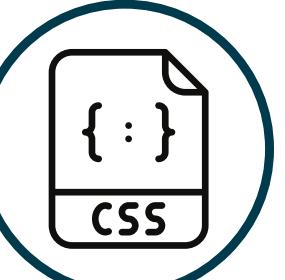


Tableau for  
Data Visualizations



Canva for collaborative  
presentation and visuals



Github for  
collaboration

EDA

# EDA Process

During the EDA process we performed an initial investigation on data to discover patterns, spot anomalies and check assumptions by using summary statistics and graphical representations.



## PHASE 1

### COMPILING/LOADING

We downloaded the data from the Inside Airbnb website. There were three files in csv format, which were: listings, calendar and reviews.

## PHASE 2

### CLEANING

We cleaned the data by removing null values and columns, detecting and dropping outliers, and transferring categorical variables into numeric format.

## PHASE 3

### ANALYZING

We created data visualizations to better understand the distribution and correlations of the variables within the dataset.

## PHASE 4

### RESULTS

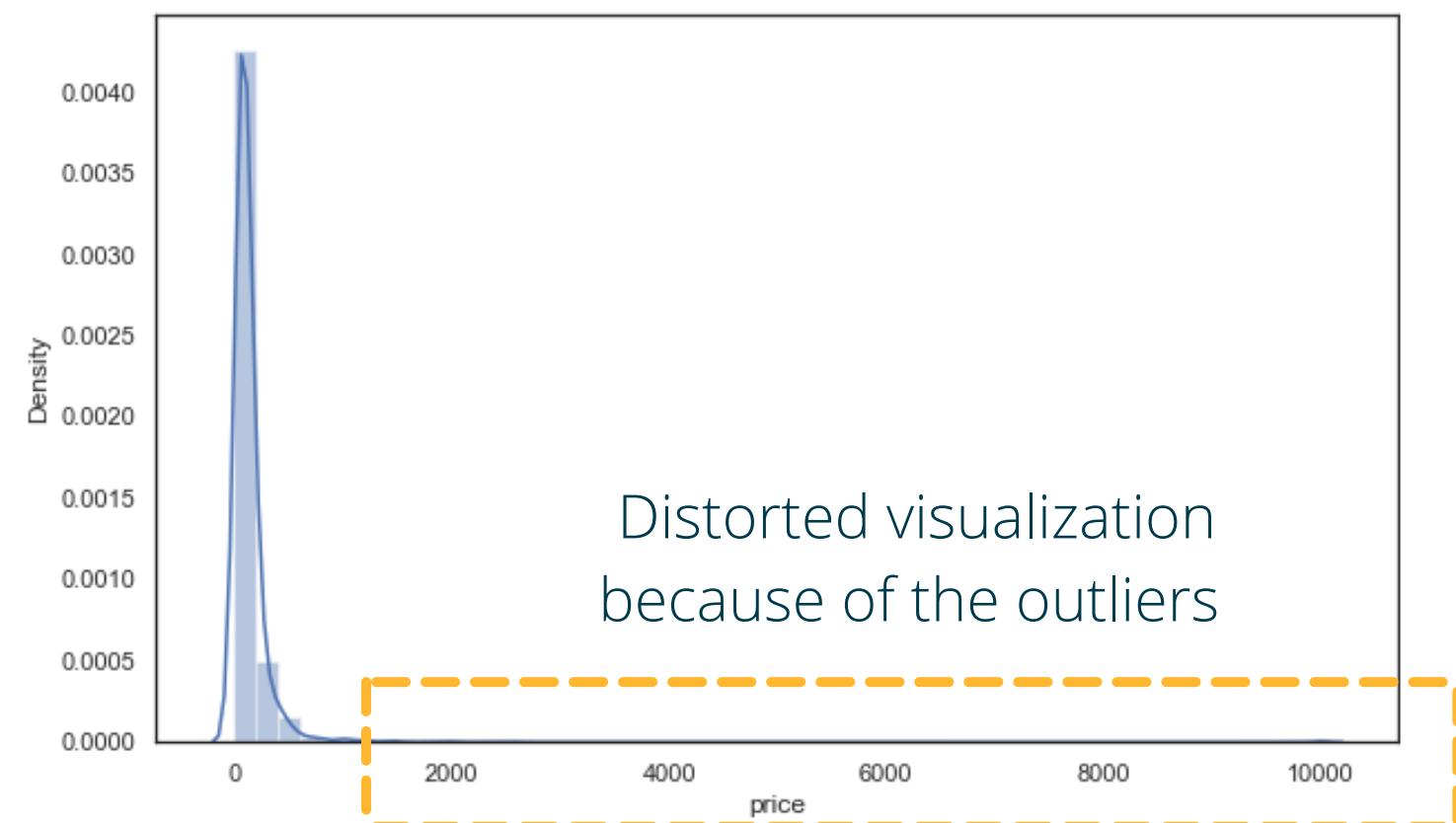
The EDA process provided a better understanding of the data and prepared us for building additional analytical models and visualizations.

# LOADING & CLEANING

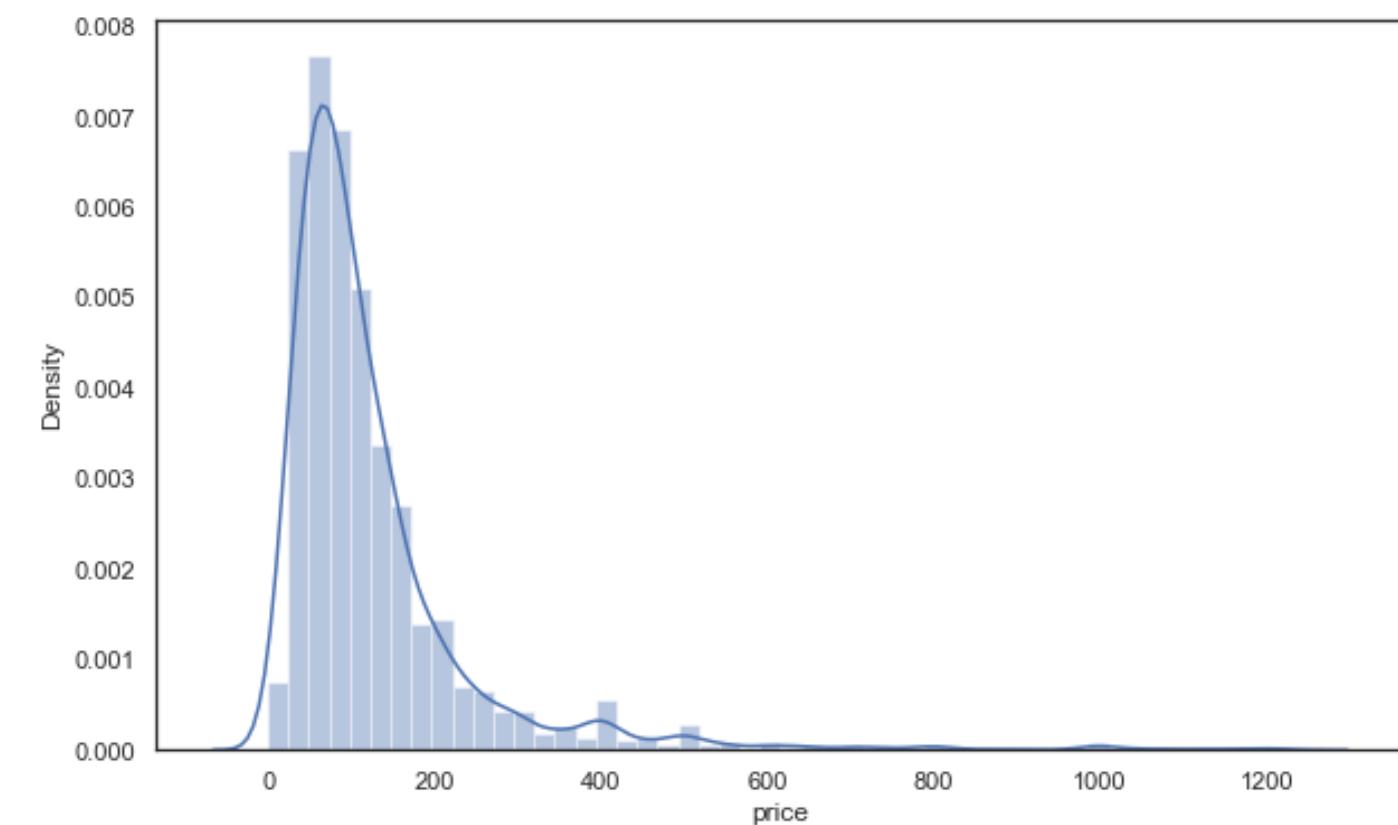
- The datasets were loaded from the Inside Airbnb website. The three files contain detailed **listing**, **calendar** and **review** data for Airbnb listings in Chicago.
- Dataset attributes:
  - Listings file: 16 columns, 6372 rows
  - Calendar file: 7 columns, one million+ rows
  - Review file: 6 columns, 267k+ rows
  - **Key output feature: listing price**
- An initial investigation shows that there is **no duplicated values** in the dataset, but there are both null columns and null values, so we **dropped the null columns** and **imputed the null values** in the numeric columns with the number zero (0).
- The investigation also indicates that there are outliers in the datasets, so we **dropped outliers with z-score greater than 3**.

04

Price Distribution before Dropping Outliers

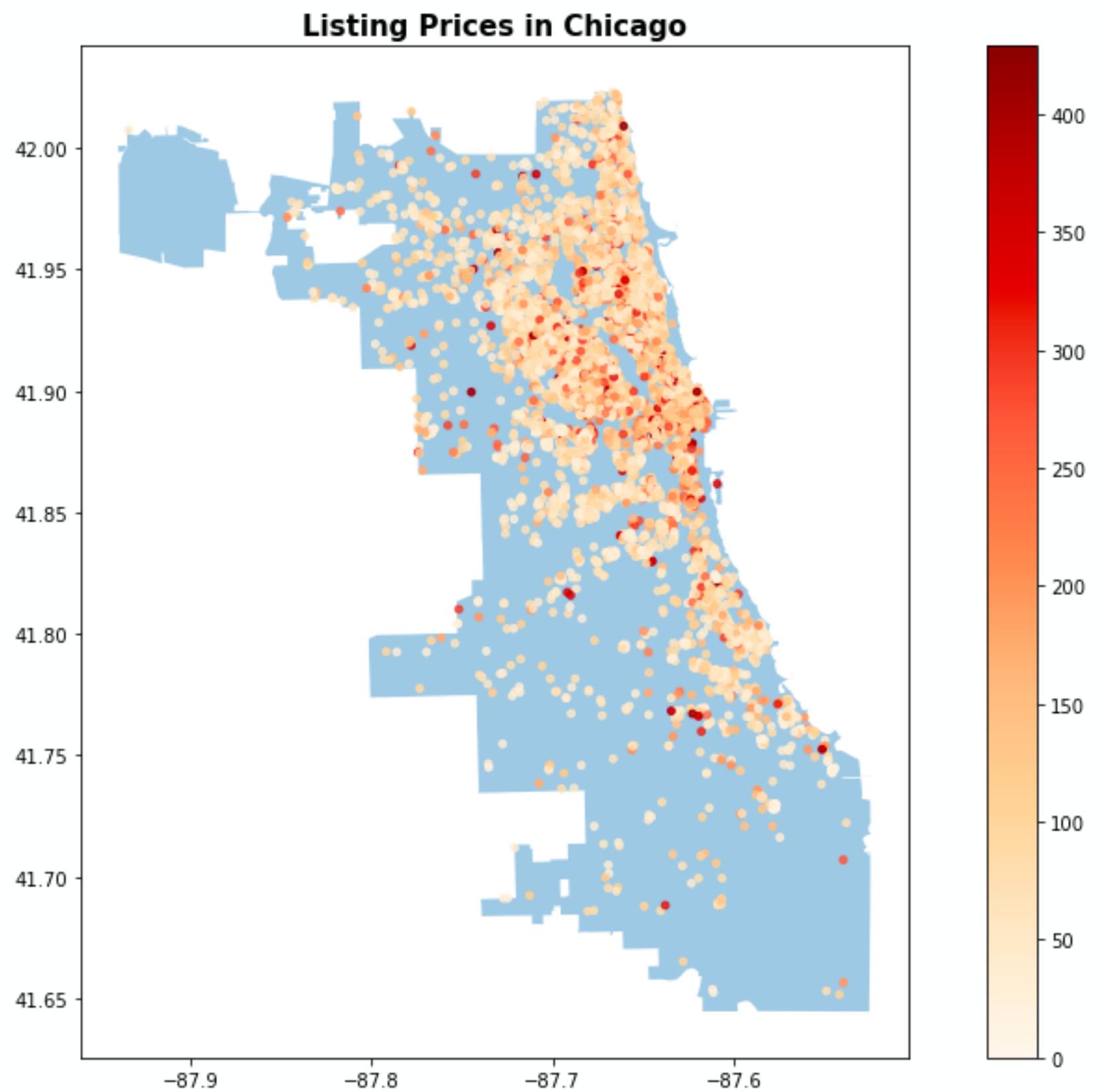


Price Distribution after Dropping Outliers



# ANALYZING

05



- **Step 1: Exploring Distributions**
  - Plot price distribution
  - Show statistic summary with `data.describe()`
  - Visualize the price distribution on the Chicago city map
    - The listing price on map shows that most of listings are located in the urban area of Chicago, and more listings with high prices are located in the Northern and Central areas(downtown) of Chicago.

# ANALYZING

## • Step 1: Exploring Distributions (Cont.)

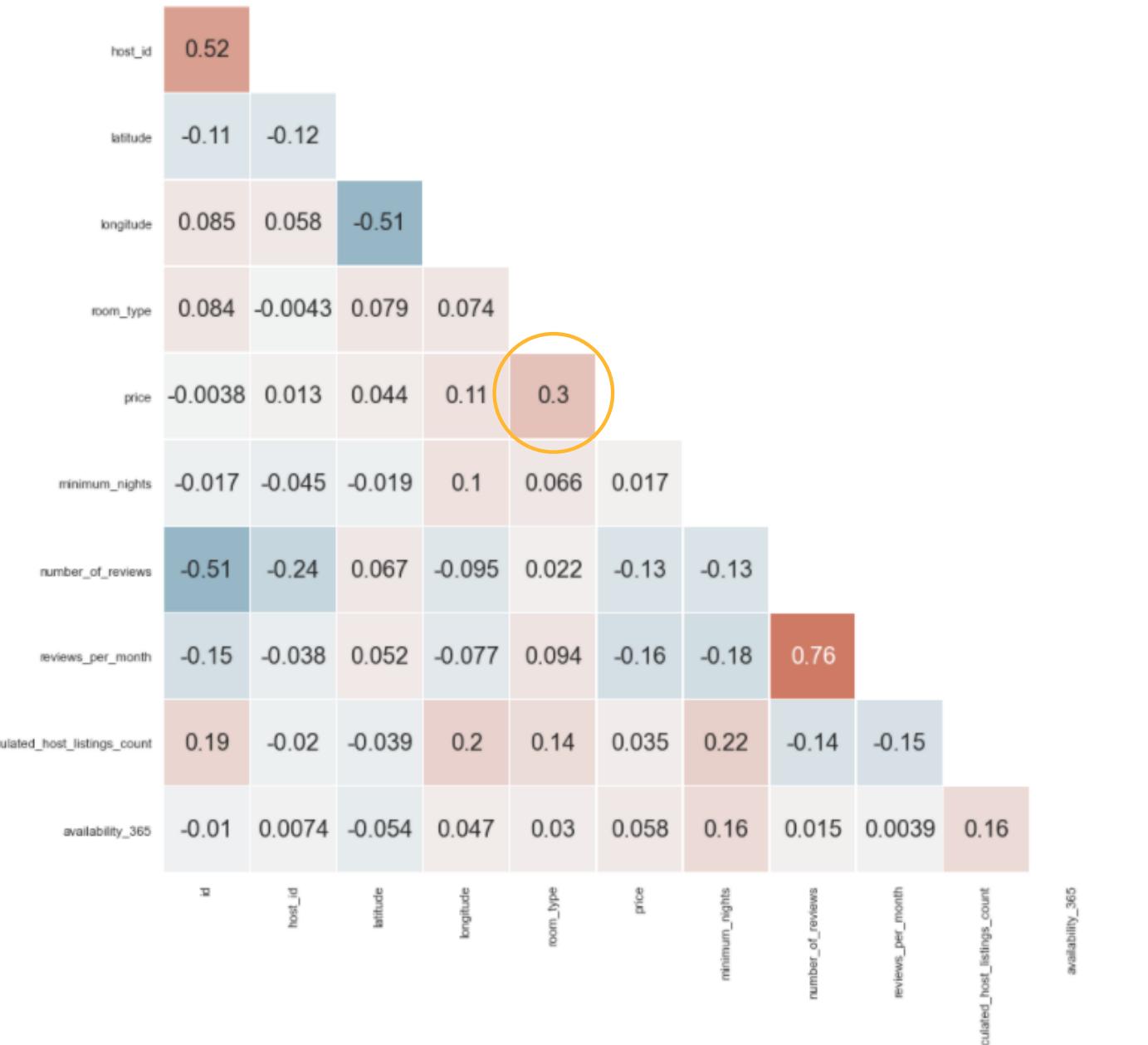
- Create a word-cloud to show the keywords appeared in the reviews comments
    - With the word cloud in comments of reviews, we can see that most reviews are positive, people care about the location, bed, and kitchen.

# WordCloud of Comments in Reviews

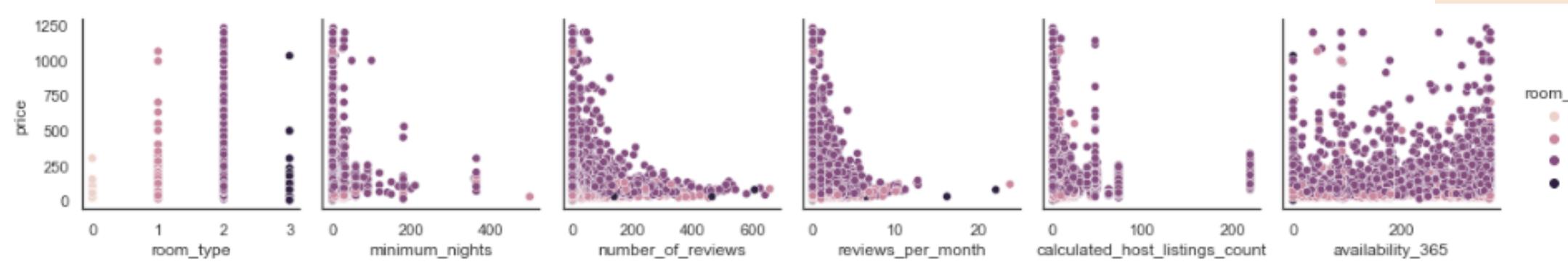


EDA

# Listings Data Variables Correlation Heatmap



# Listings Data Variables Pairplot



# ANALYZING

- **Step 2: Detecting correlations**

- Transfer key categorical variables into numeric format (i.e. room\_type)
- Draw correlation heat-map
- Draw pair-plots

- **Key Findings:**

- Price distribution is influenced by the room type. Generally, the price of shared rooms < private rooms < entire home/apt
- Interestingly, the number of reviews has a slight negative correlation with price

## Listing Data after EDA

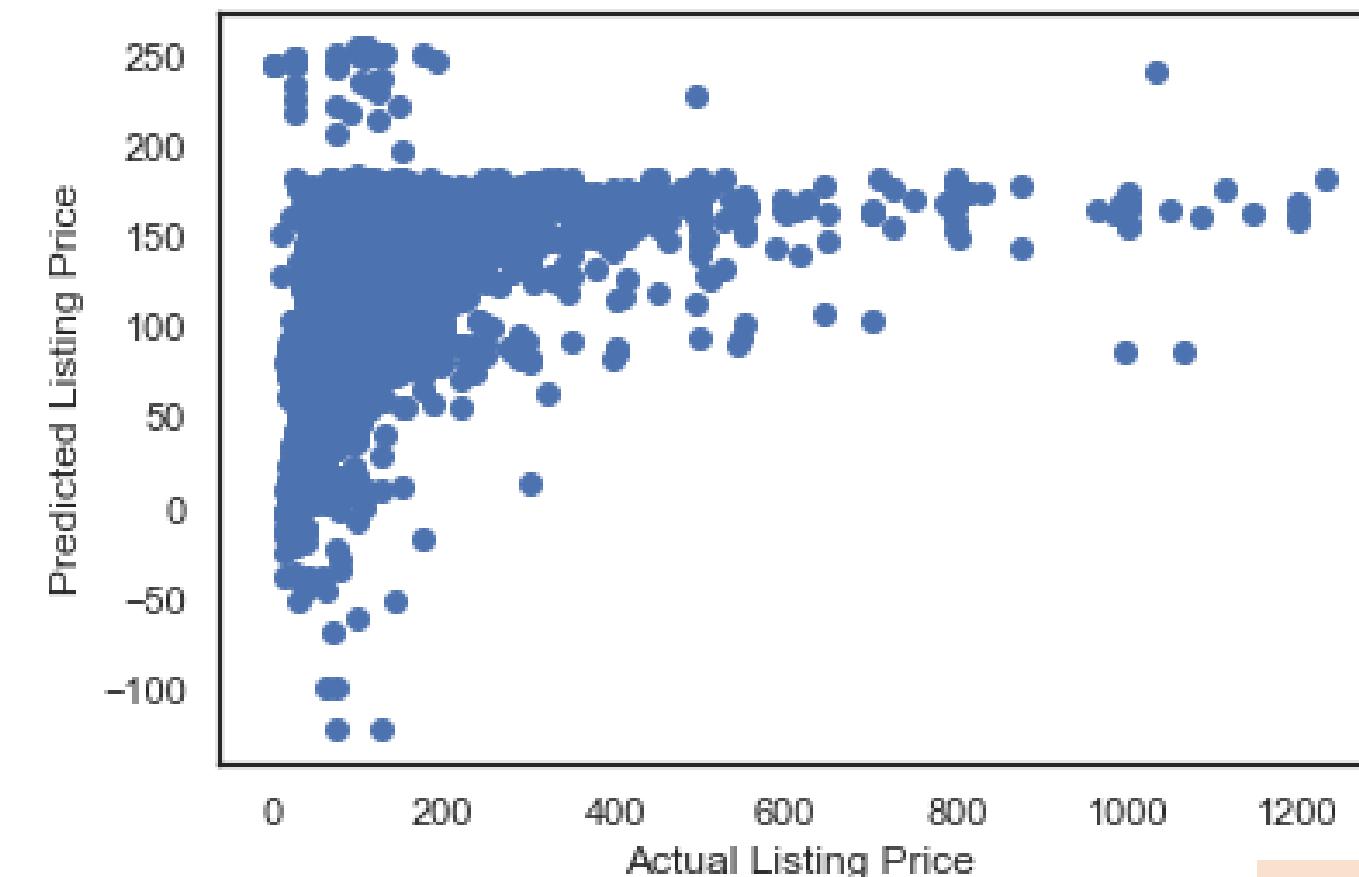
```
[Int64Index: 6337 entries, 0 to 6371
Data columns (total 15 columns):
 #   Column           Non-Null Count   Dtype  
 ---  -- 
 0   id               6337 non-null    int64  
 1   name              6337 non-null    object  
 2   host_id            6337 non-null    int64  
 3   host_name          6334 non-null    object  
 4   neighbourhood       6337 non-null    object  
 5   latitude           6337 non-null    float64 
 6   longitude          6337 non-null    float64 
 7   room_type          6337 non-null    int64  
 8   price              6337 non-null    int64  
 9   minimum_nights     6337 non-null    int64  
 10  number_of_reviews   6337 non-null    int64  
 11  last_review         5172 non-null    object  
 12  reviews_per_month   6337 non-null    float64 
 13  calculated_host_listings_count 6337 non-null    int64  
 14  availability_365    6337 non-null    int64  
dtypes: float64(3), int64(8), object(4)
memory usage: 792.1+ KB
```

# NEXT STEPS

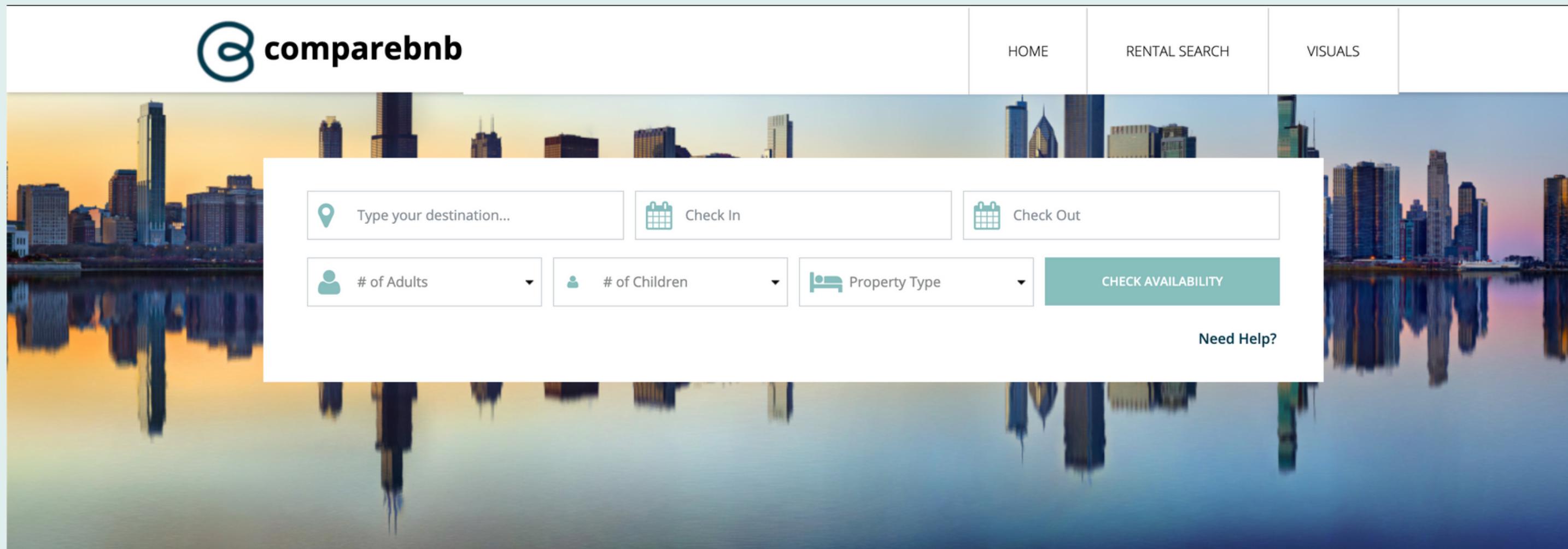
## 08

- **Ready for Building Predictive Models**
  - The dataset has been cleaned up and key input features are in numeric format
  - The first trial of linear regression model indicates that we need more input features
- **Outsource More Input Features**
  - Use the listing dataset of Amsterdam as reference and figure out what we can look for

## Sample Linear Regression



# EXPLORE COMPAREBNB!



The screenshot shows the Comparebnb homepage. At the top left is the Comparebnb logo. To the right are three navigation links: HOME, RENTAL SEARCH, and VISUALS. Below the navigation is a search interface. It features a large input field with a location pin icon and the placeholder "Type your destination...". To its right are two date selection fields: "Check In" and "Check Out", each with a calendar icon. Below these are dropdown menus for "# of Adults" and "# of Children", both with person icons. To the right of these dropdowns is another dropdown menu for "Property Type" with a bed icon. A large teal button labeled "CHECK AVAILABILITY" is positioned to the right of the property type dropdown. At the bottom right of the search interface is a link "Need Help?". The background of the page is a blurred image of a city skyline at sunset or sunrise, reflected in water.

Type your destination...

Check In

Check Out

# of Adults

# of Children

Property Type

CHECK AVAILABILITY

Need Help?

Find your next adventure

DEMO

09

## SUMMARY

According to the data analytical results, the price of listing is influenced by different factors, such as location, the day of week, the time of year, room type, etc. Thus, with taking these factors into consideration, users will make better decision for renting by using comparebnb.

## NEXT STEPS

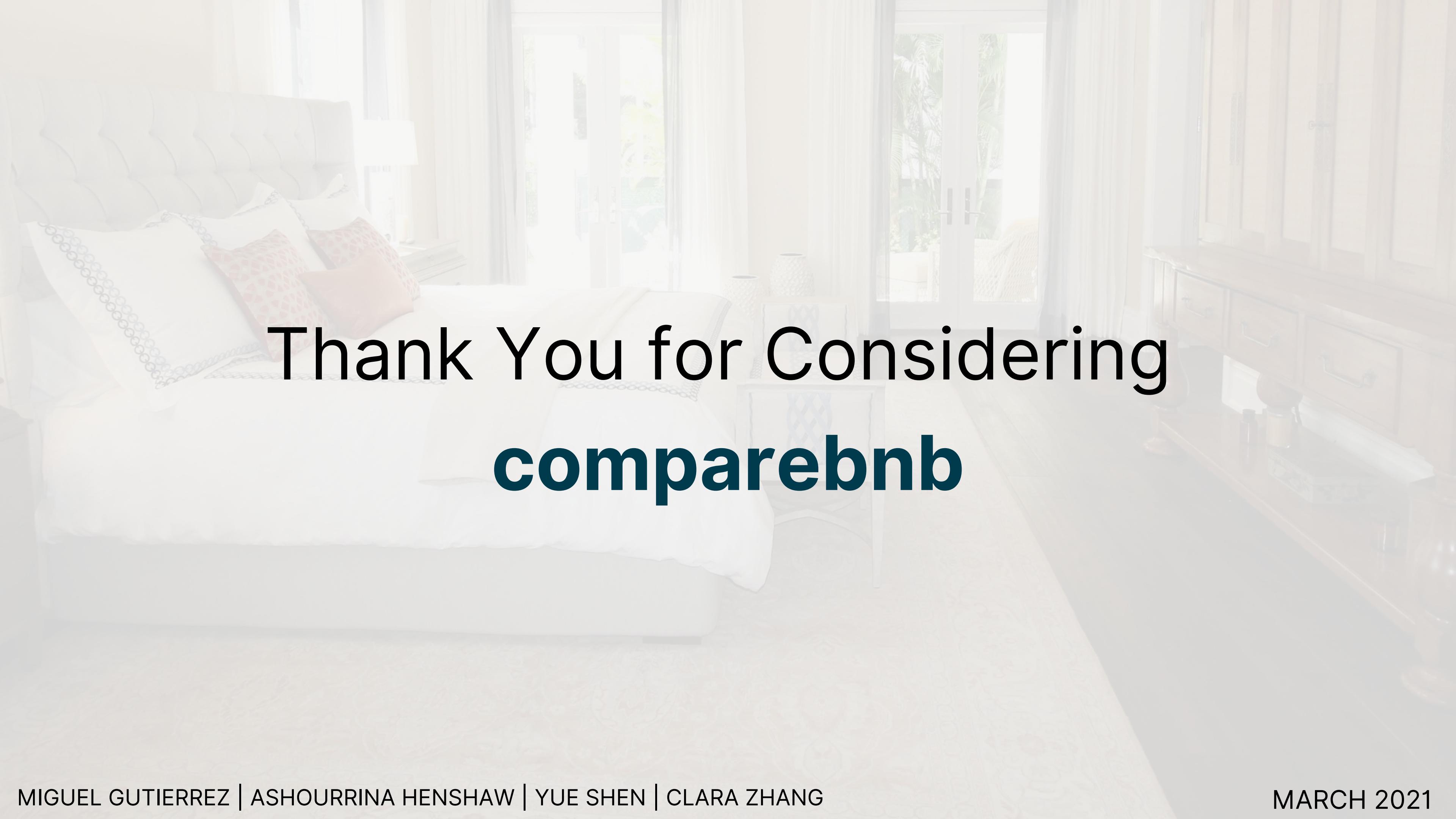
- Expanding Data
- Creating Models
- Generate Automated Recommendation System

# Summary



## LESSONS LEARNED

- How to build an interactive data visualization dashboard
- How to present data in a way that people can easily understand
- How to develop an eyecatching, modern website



Thank You for Considering  
**comparebnb**



## About Us

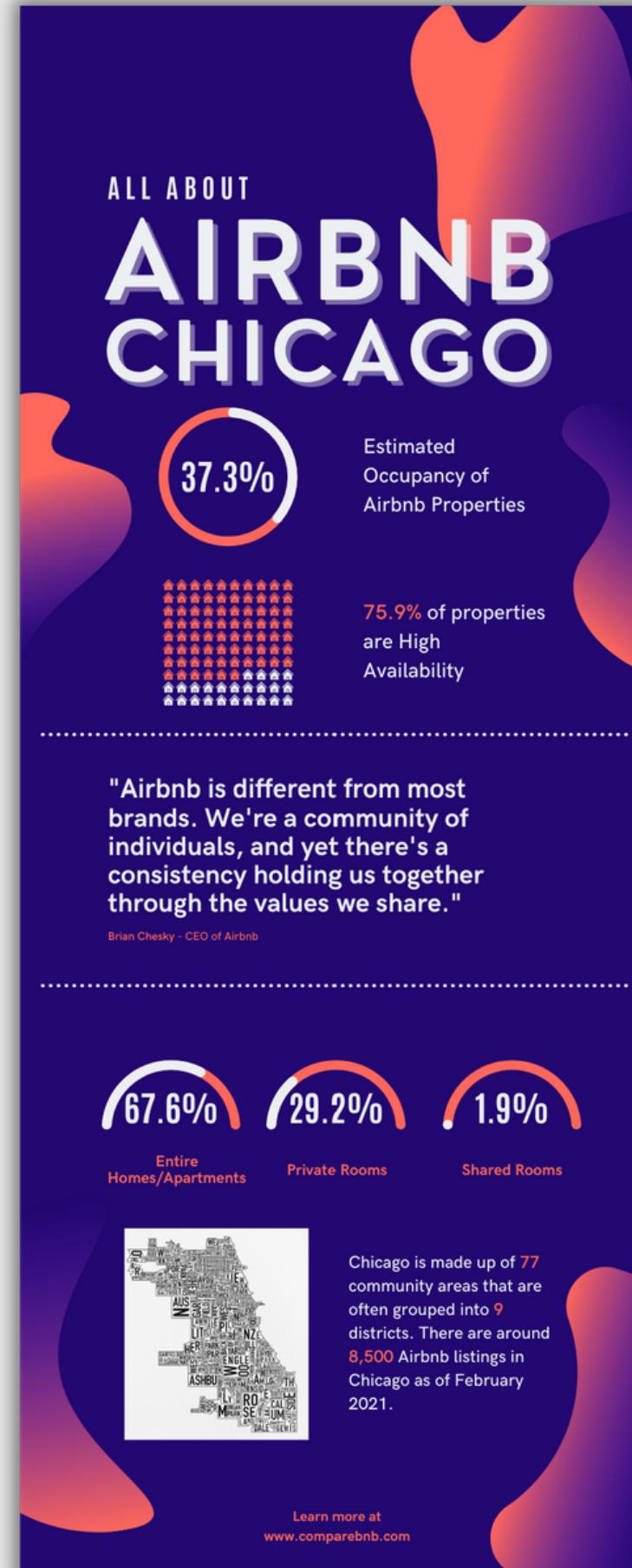
Alpha Analytics is dedicated to creating custom data science solutions to the client's specific needs. Interpreting data may be confusing – Alpha Analytics was created with this in mind. We strive to create understandable and actionable insights based on data you provide.

[MEET OUR TEAM](#)

## See What We're Up To!

Comparebnb is our newest project! We are bringing data and analytics direct to the you by bringing transparency in pricing among short term rentals

[CLICK TO LEARN MORE!](#)



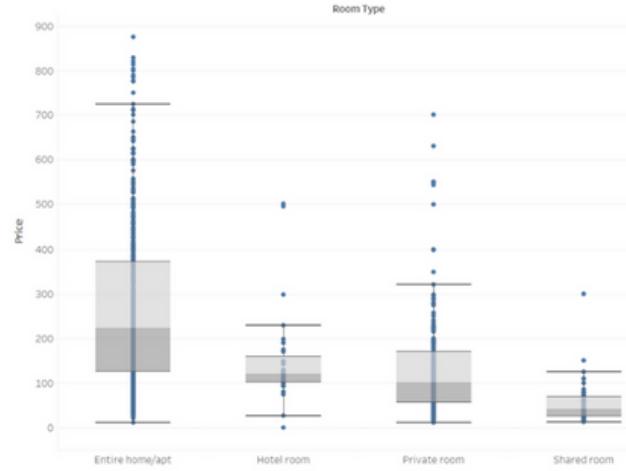
## Appendix - Infographics

# Analysis of Airbnb Data in Chicago

## Motivation

Airbnb prices vary greatly with different room types and neighborhoods. The time of year and day of the week also affects the prices. This short analysis focuses on visualizing the different factors that influence price for Airbnb properties in the Chicago area. The data is from Inside Airbnb and was updated February 17th, 2021 and contains listings for 365 days.

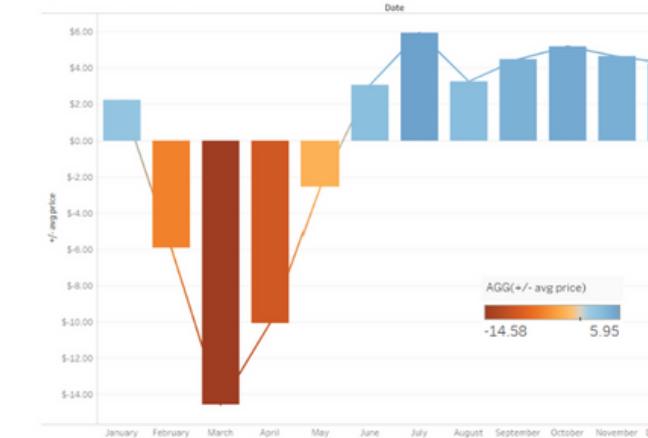
Room Type and Price Boxplot



## Room type

Inside Airbnb's data split the properties into four room types - Entire home/apartment, Hotel room, Private room, and Shared room. The distributions and spread of these room types are vastly different, with Hotel room and Shared room having the smallest spread and Entire home/apartment having the greatest spread. Every room type still has outliers (especially on the high end). The medians of the room types are as expected, with Shared rooms having the lowest median price and the highest median price being an entire home/apartment.

+/- Average Price per Month



## Time of Year

Trends are shown by plotting out the average prices for months and the average price for the year. February, March, April and May all have lower prices when compared to the yearly price average, with March having the lowest average cost at about \$14 less than the yearly average. Prices are lower in the Spring and higher in the other seasons, with the highest prices being in July, in the Summer. There is a monotonic decline in prices starting from October until March.

Price and Adjusted Price by Month



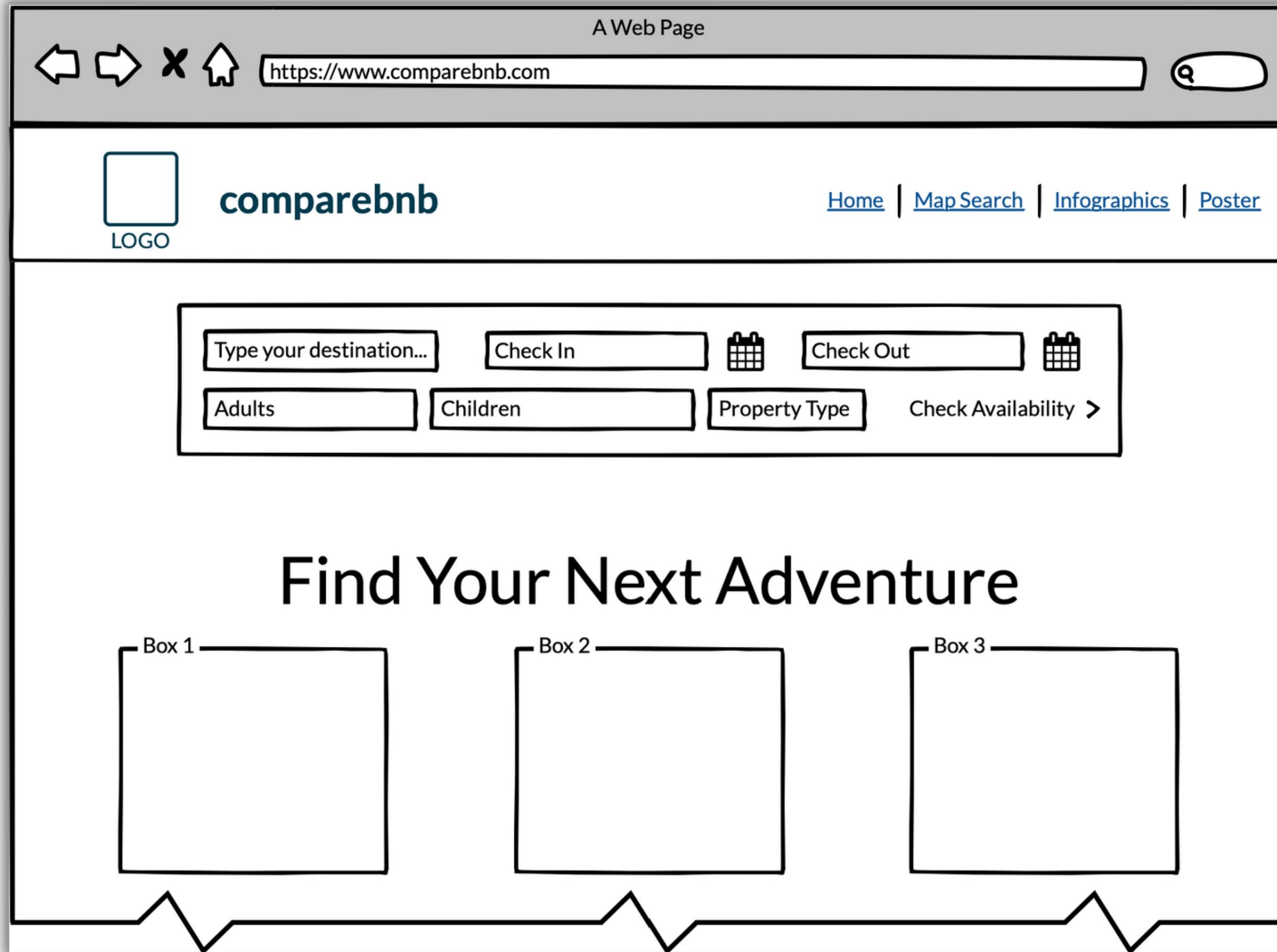
## Day of Week

As expected, the prices for Airbnb's Fridays and Saturdays are the highest amongst the days of the week. There is a slight dip in the prices going on Tuesdays, but it is almost negligible compared to the increase in prices on the weekends



## Conclusions

Airbnb's prices are influenced by many factors, including a specific host's pricing and neighborhood. However, room prices depend on the room type, with many hosts charging a higher price than the spread of the prices. Additionally, these prices tend to fluctuate in a more predictable manner over the week as well as over the months in this year. The consumer can make better decisions for renting an Airbnb based on these factors. A website, comparebnb, was created in order to help in this process.



## Appendix - Comparebnb Wireframe

Comparebnb

<https://alpha-analytics.ai/comparebnb>

# Alpha Analytics

Providing data solutions and developing custom reporting and dashboards for your business needs

## Comparebnb

City ▾ Neighborhood ▾ Type of Listing ▾ Rating ▾

Sample Listing  
Average Cost Per Night: \$88  
Latest Cost Per Night: \$76  
Best Time to Rent: October-December

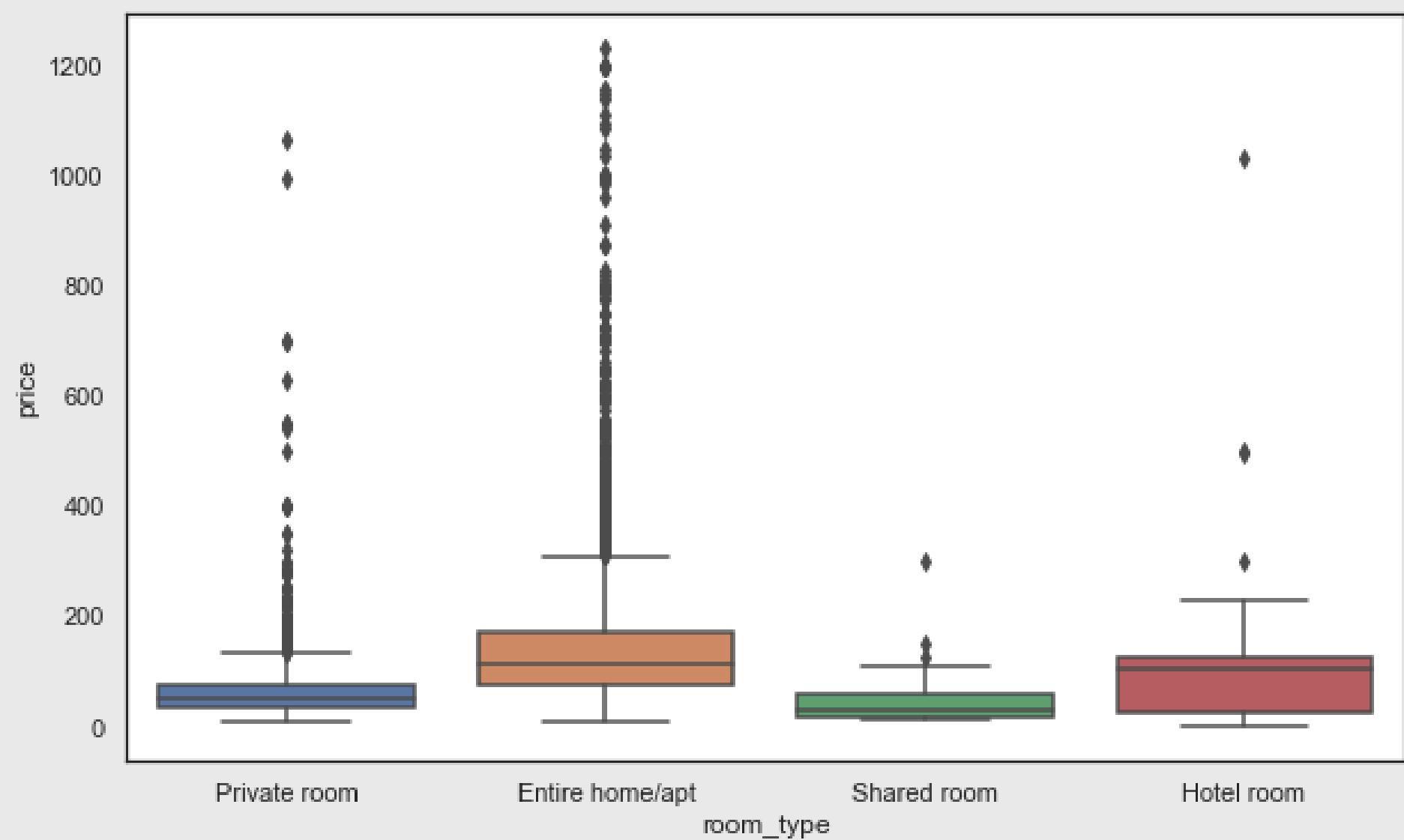
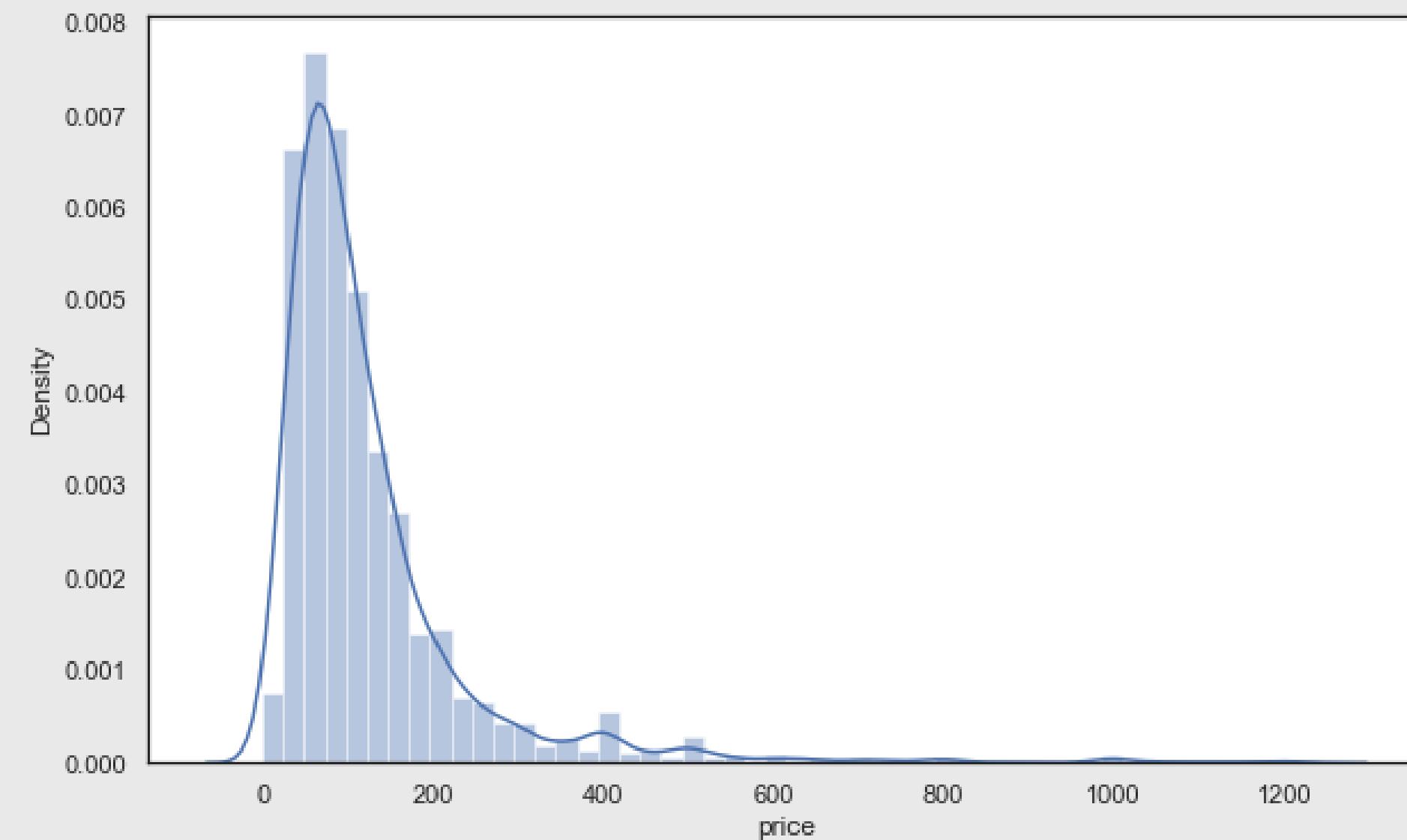
Cost of selected listing over time and average cost of listings in selected area

Visualization regarding reviews

lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

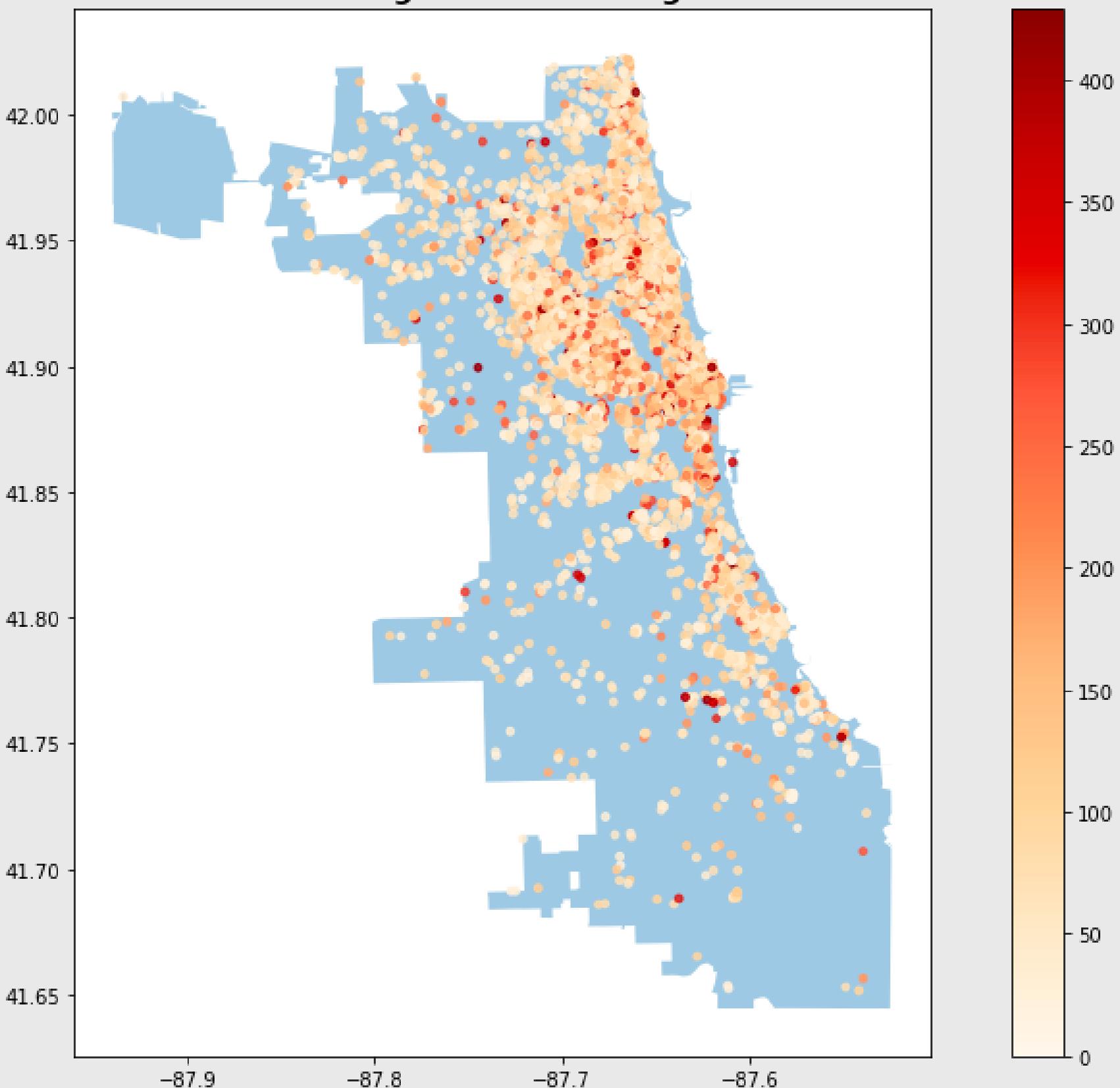
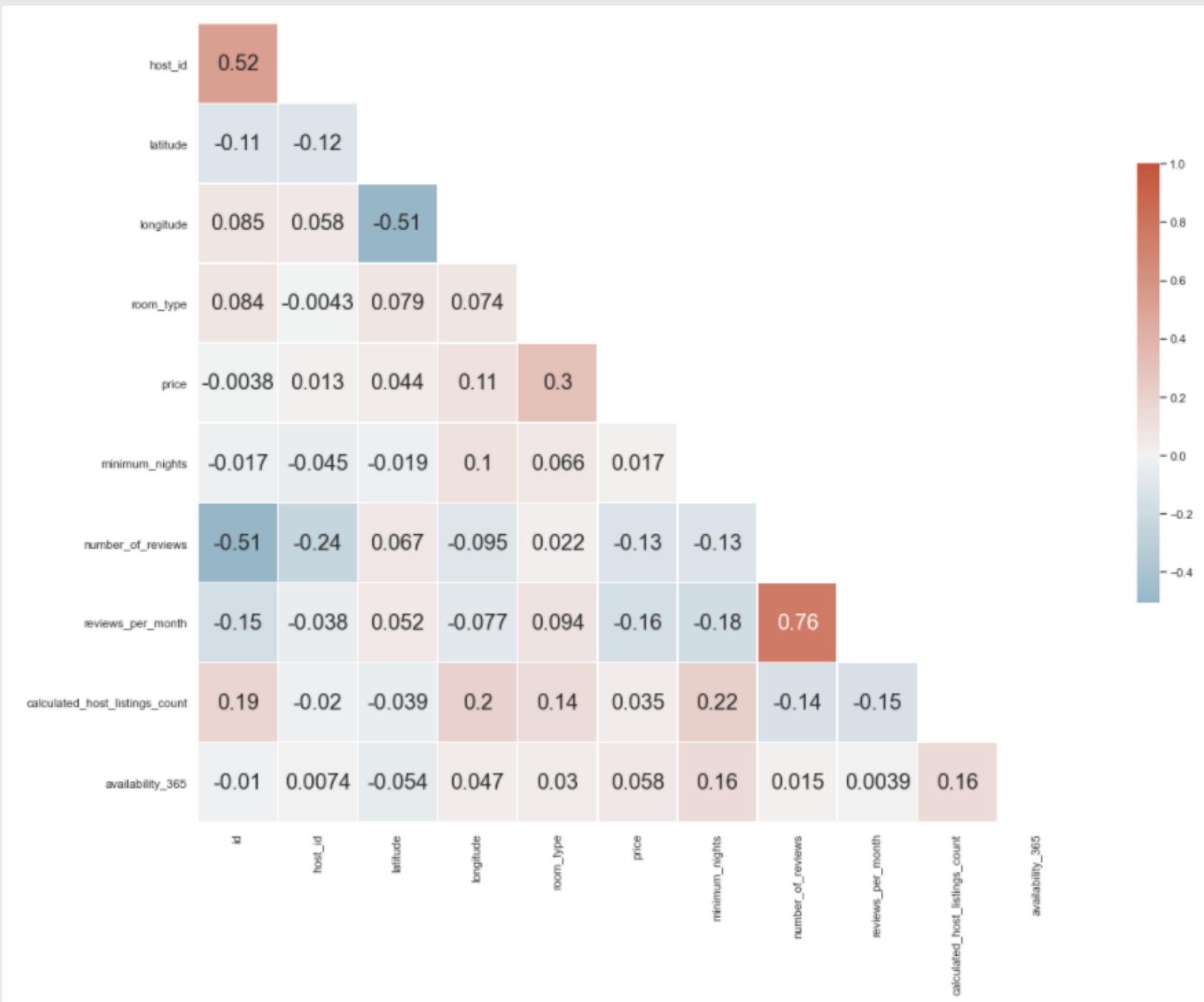
The wireframe illustrates a user interface for a real estate analysis tool. At the top, there's a header bar with navigation icons (back, forward, search) and the URL https://alpha-analytics.ai/comparebnb. Below the header is a blue navigation bar with the 'Alpha Analytics' logo and a tagline: 'Providing data solutions and developing custom reporting and dashboards for your business needs'. A three-line menu icon is on the right. The main content area has a title 'Comparebnb'. On the left is a map with several red and blue dots representing listings. A yellow callout box on the map provides details for a 'Sample Listing': Average Cost Per Night (\$88), Latest Cost Per Night (\$76), and the Best Time to Rent (October-December). Above the map are four dropdown filters: City, Neighborhood, Type of Listing, and Rating. To the right of the map are two data visualizations: a line graph showing the cost of the selected listing over time and the average cost of listings in the selected area, and a bar chart showing reviews. A large amount of placeholder text (lorem ipsum) is present at the bottom of the main content area.

## Appendix - Map Search Wireframe



## Appendix - Python or Tableau Visual

## Listing Prices in Chicago



## Appendix - Python or Tableau Visual



# Appendix - Python or Tableau Visual