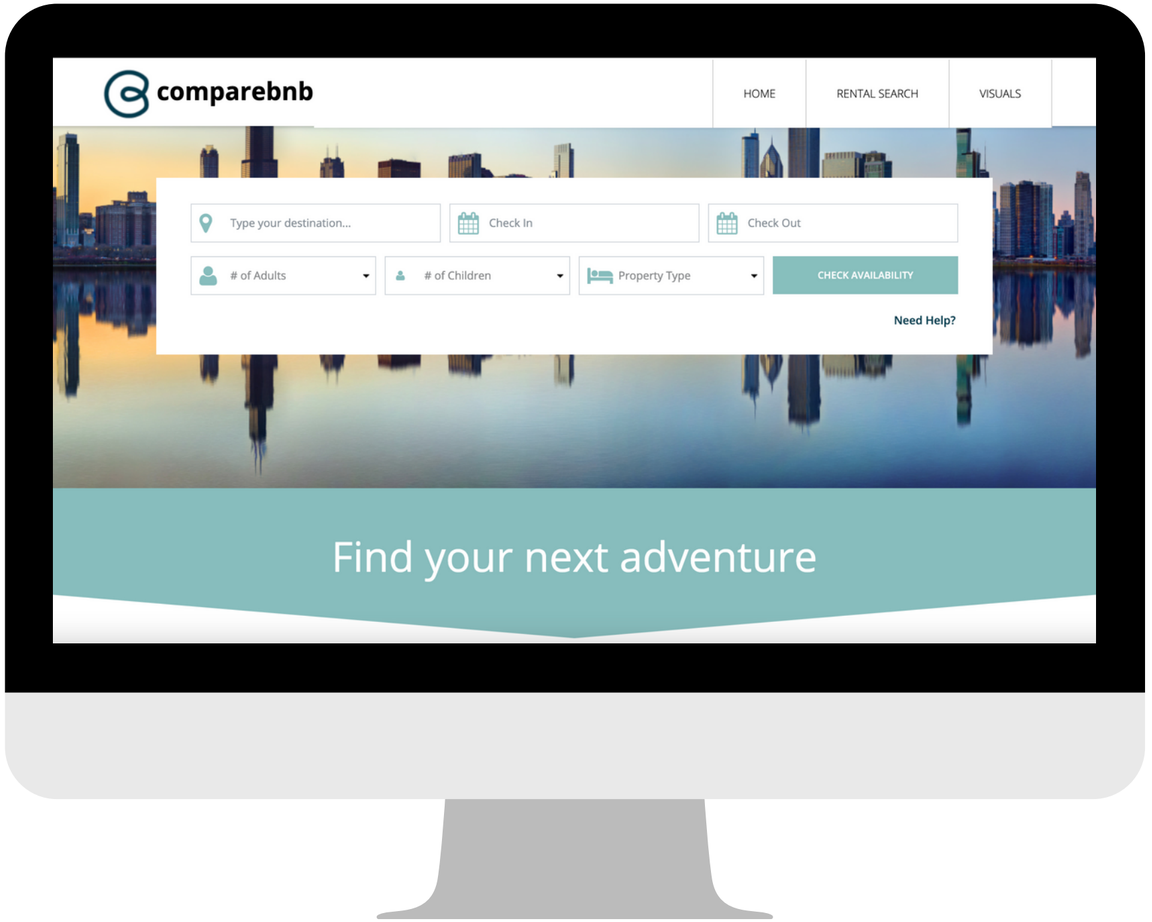


comparebnb

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

Miguel Gutierrez, Ashourrina Henshaw,
Yue Shen, Clara Zhang

Agenda

Executive Summary

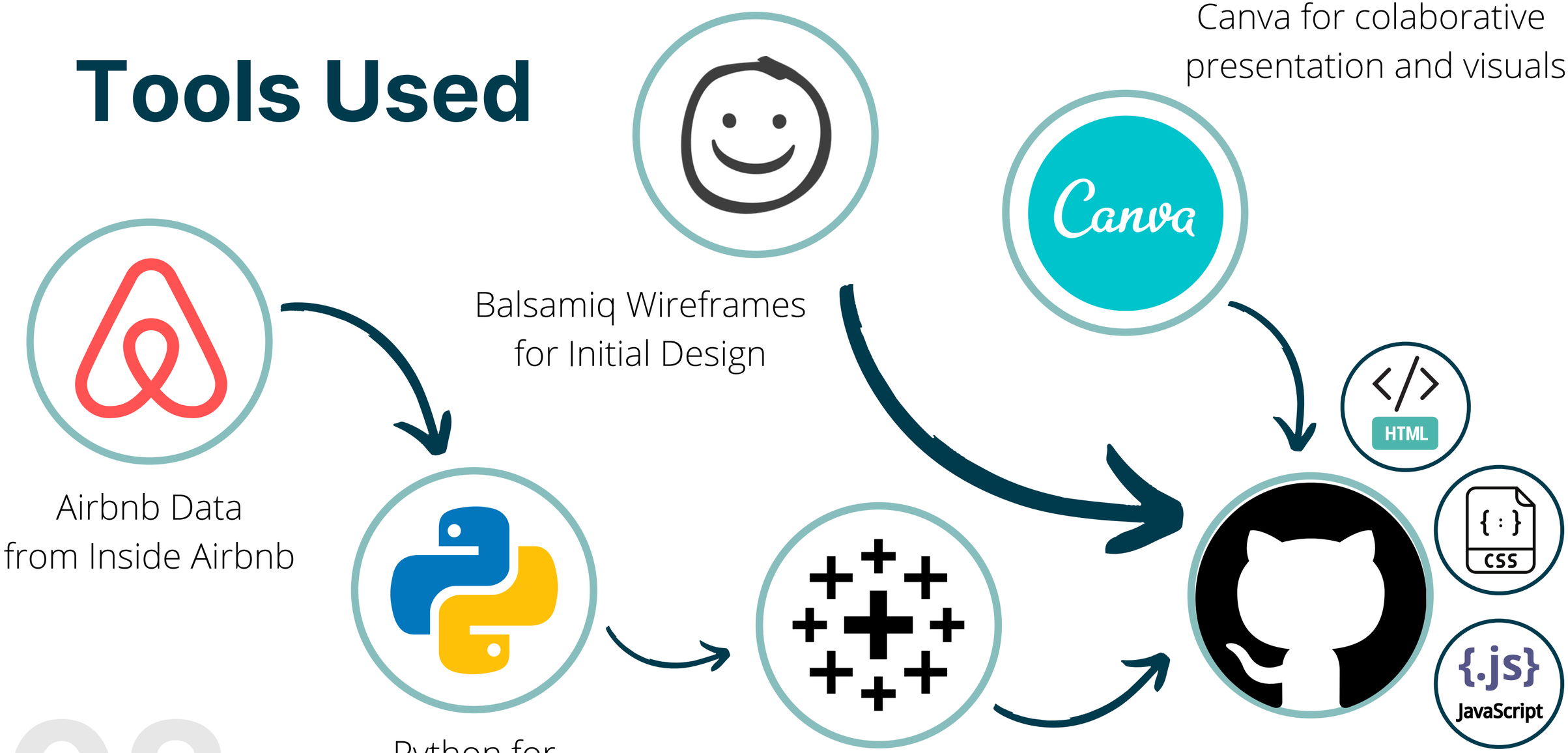
Airbnb is the world's largest online platform for short term rentals and tourism activities.
By the end of 2020, there were over **seven million** listings on Airbnb, run by **four million** hosts.

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



02

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

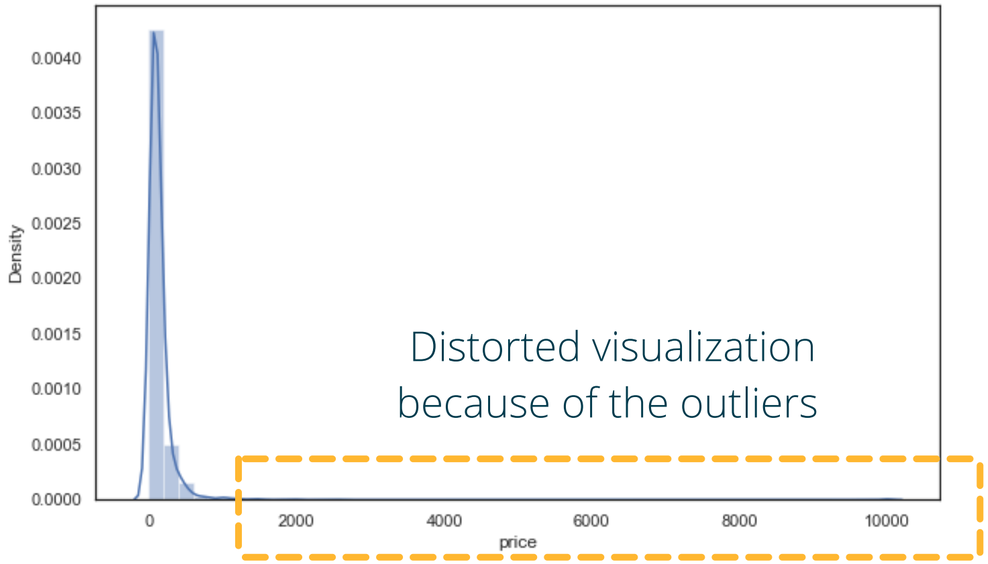
RESULTS

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

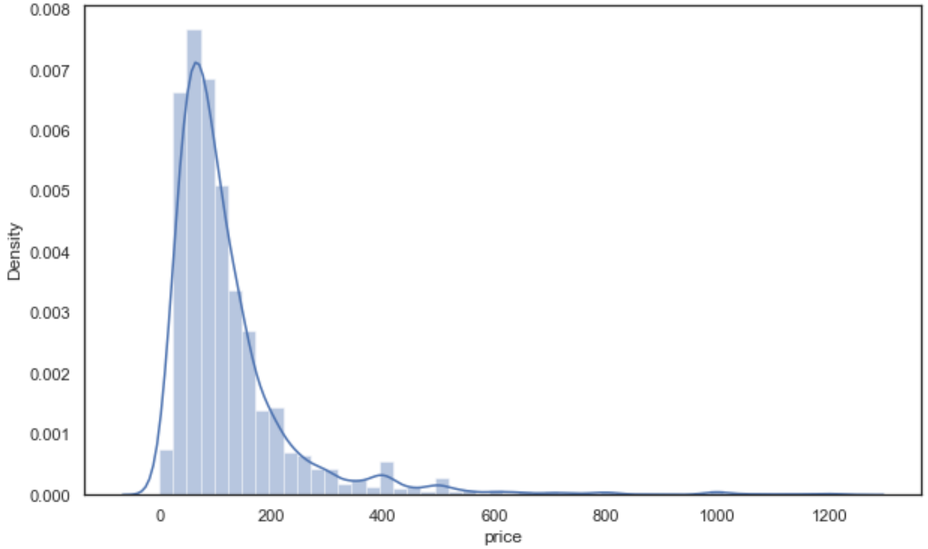
LOADING & CLEANING

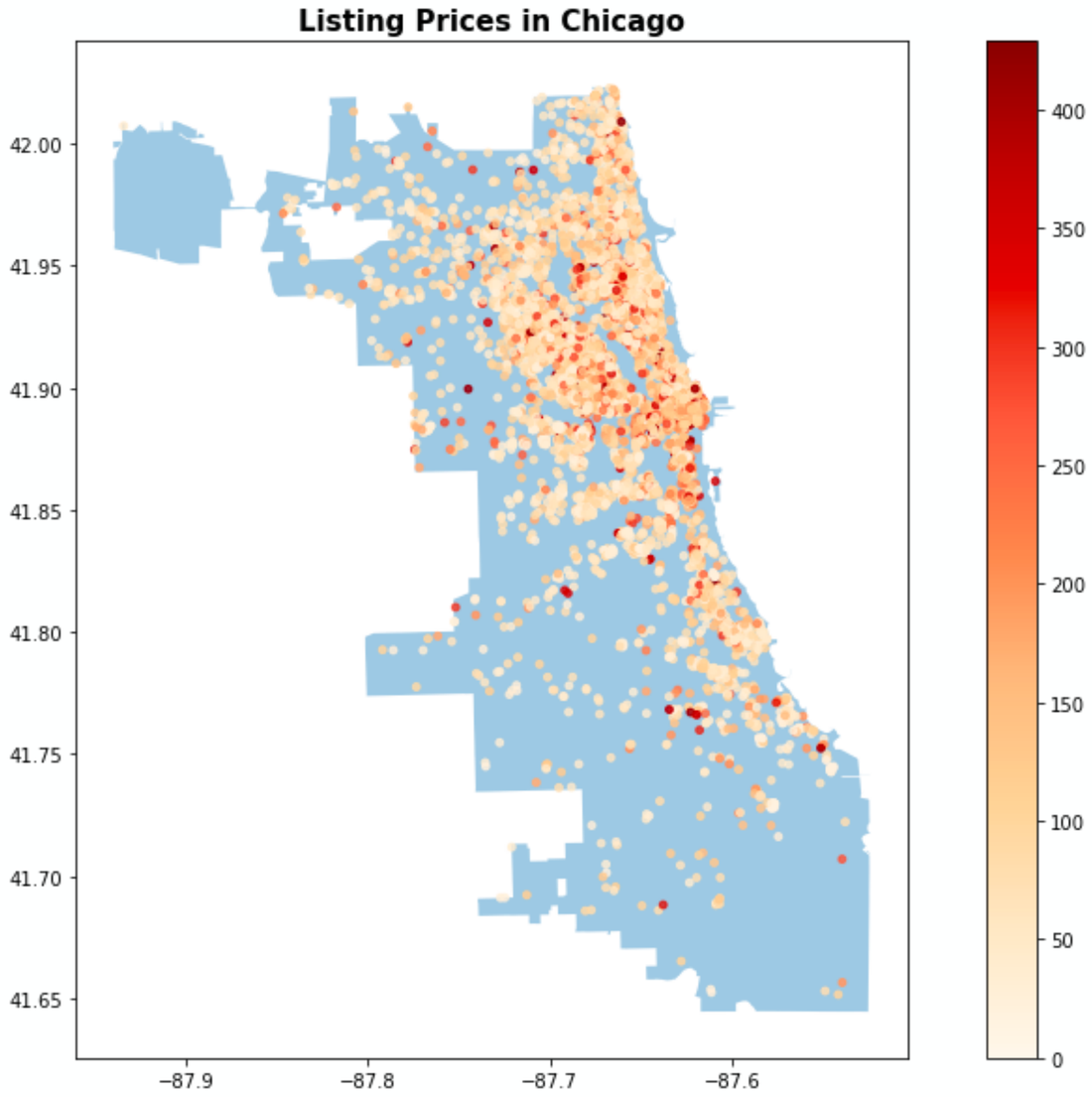
- The datasets were loaded from the Inside Airbnb website. The three files contain detailed **listing**, **calendar** and **review** data for Airbnb listings in Chicago.
- Dataset attributes:
  - Listings file: 16 columns, 6372 rows
  - Calendar file: 7 columns, one million+ rows
  - Review file: 6 columns, 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**.

Price Distribution before Dropping Outliers



Price Distribution after Dropping Outliers





ANALYZING

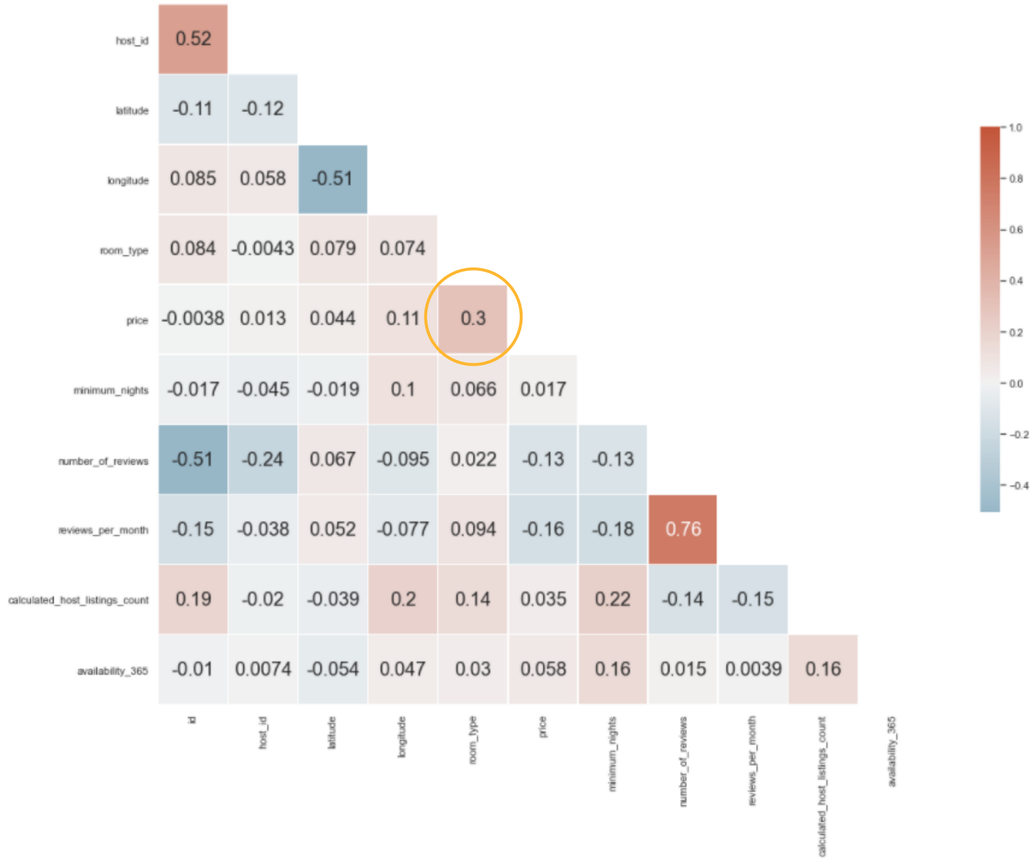
- **Step 1: Exploring Distributions**
  - Plot price distribution
  - Show statistic summary with data.describe()
  - Visualize the price distribution on the Chicago city map
    - The listing price on map shows that most of listings are located in the urban area of Chicago, and more listings with high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

ANALYZING

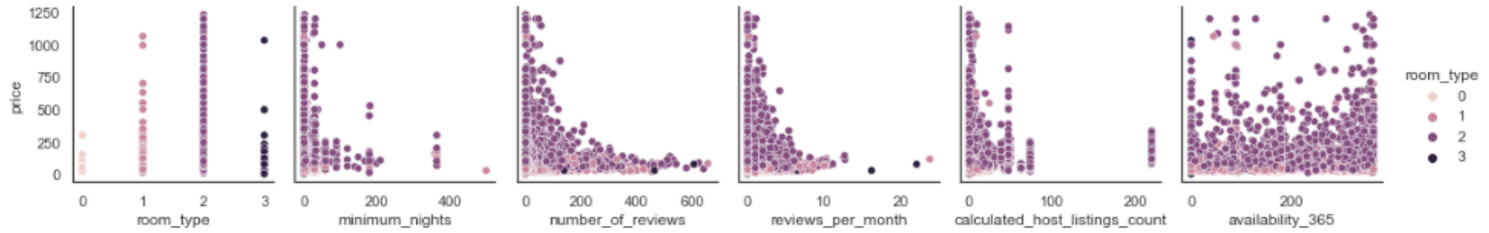
- **Step 1: Exploring Distributions (Cont.)**
  - Create a word-cloud to show the keywords appeared in the reviews comments
    - With the word cloud in comments of reviews, we can see that most reviews are positive, people care about the location, bed, and kitchen.

WordCloud of Comments in Reviews

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

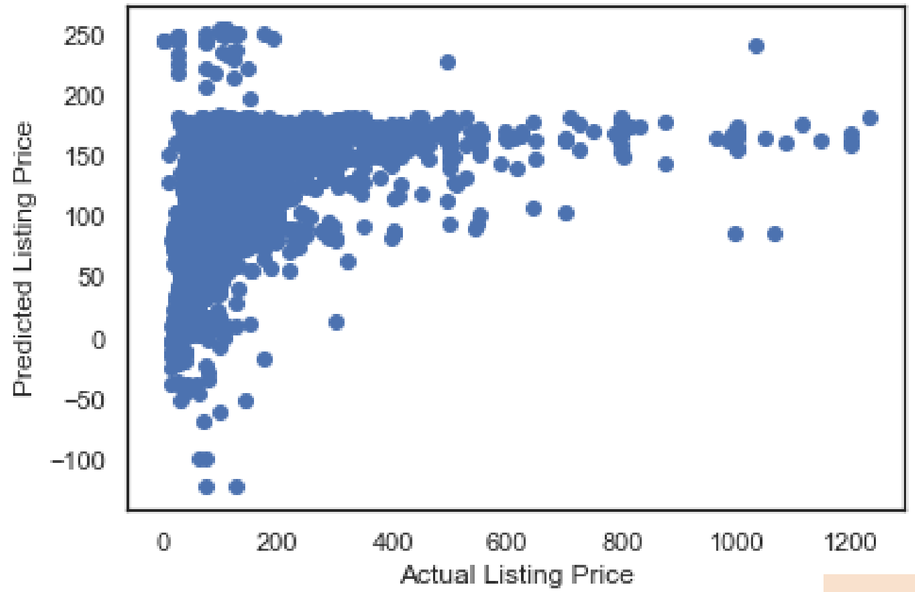
NEXT STEPS

- **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

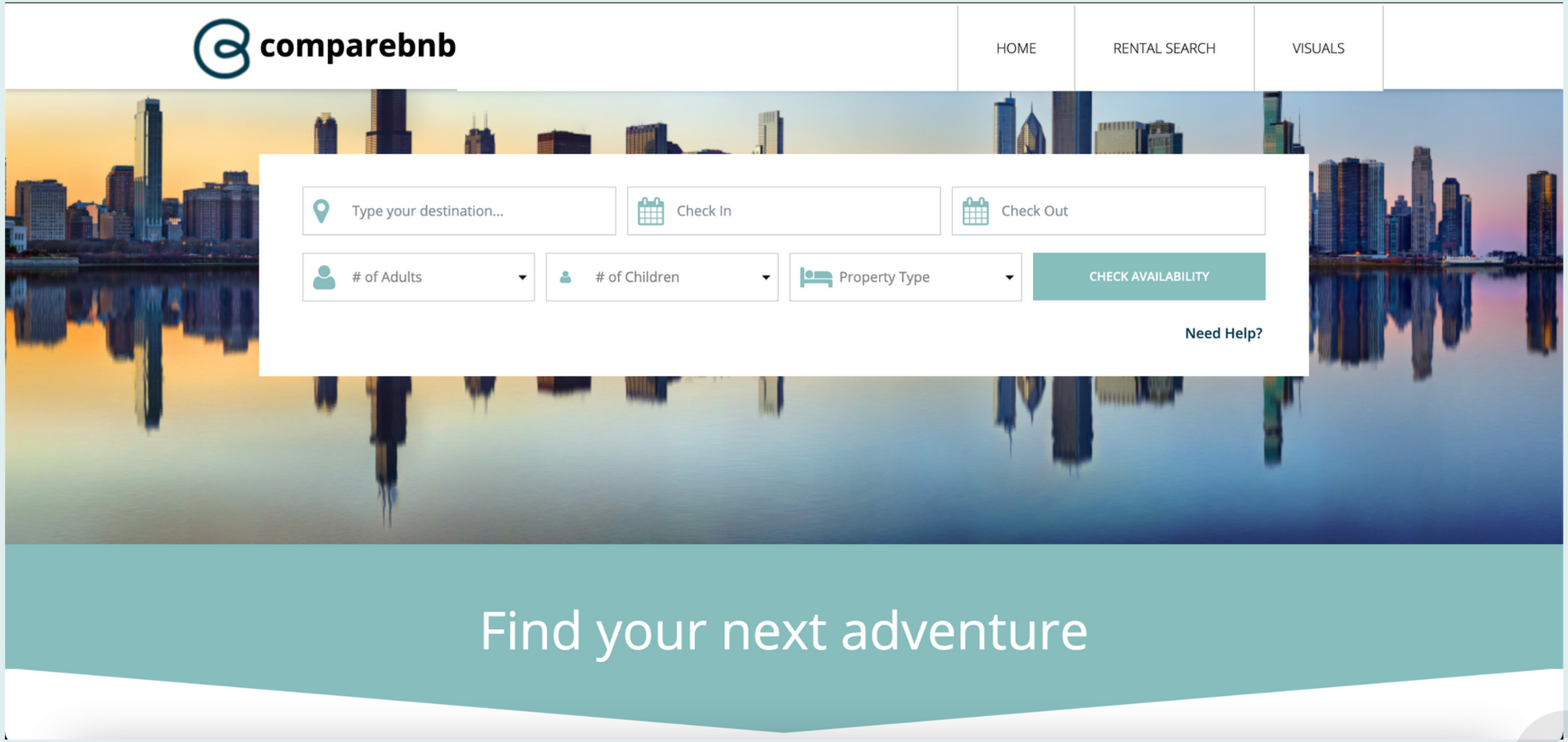
Listing Data after EDA

```
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
```

Sample Linear Regression



EXPLORE COMPAREBNB!



DEMO

Summary

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

LESSONS LEARNED

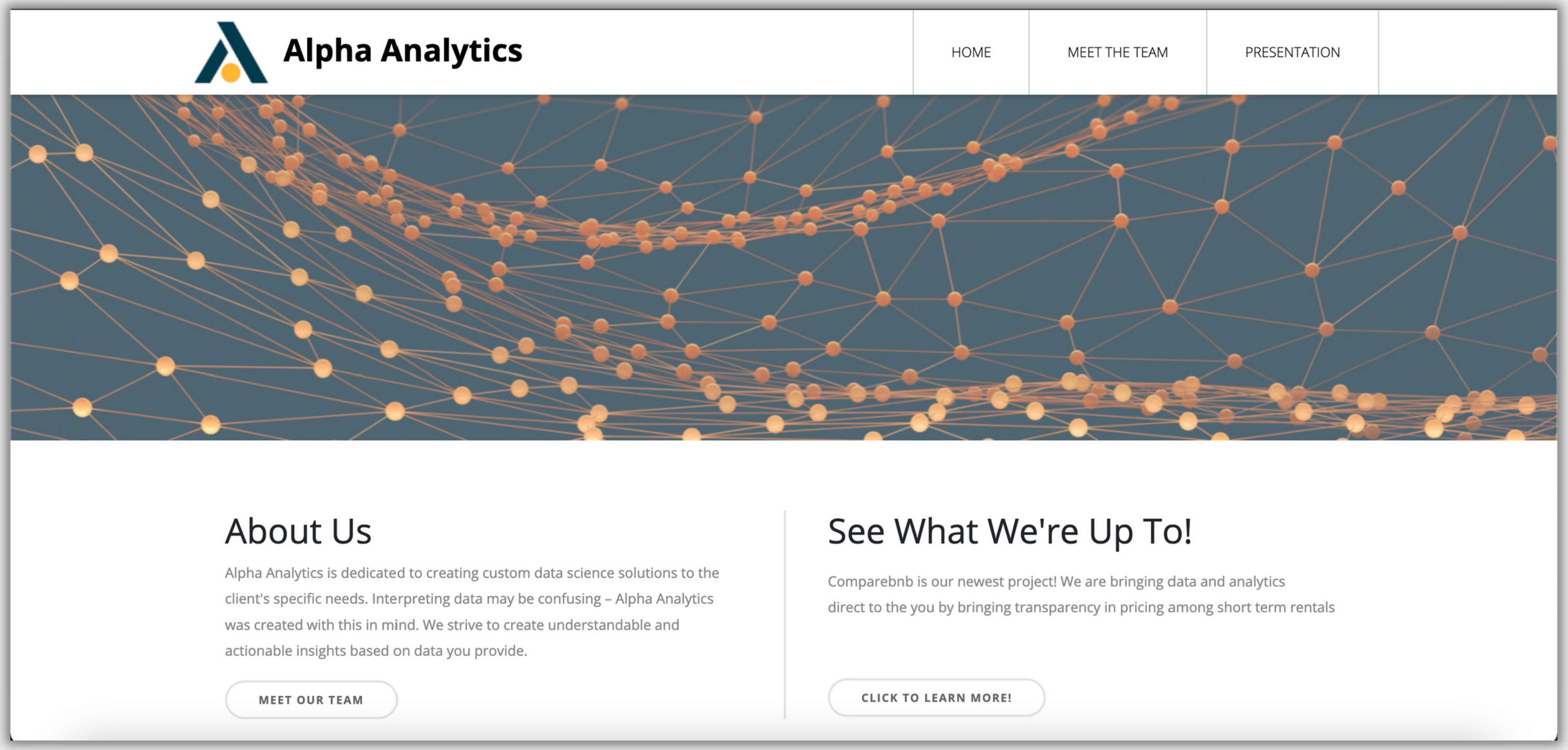
- How to build an interactive data visualization dashboard
- How to present data in a way that people can easily understand
- How to develop an eyecatching, modern website

Thank You for Considering

comparebnb

MIGUEL GUTIERREZ | ASHOURRINA HENSHAW | YUE SHEN | CLARA ZHANG

MARCH 2021



Appendix - Alpha Analytics Website

comparebnb

CHICAGO

CITY GUIDE

State	**Illinois**	Time Zone	**Central (UTC-6)**
Population	**2,722,389**	Climate	**Humid Continental**
Area	**234.0 sq. mi**	Mayor	**Lori Lightfoot**

HEADED TO CHICAGO?

comparebnb offers you a new way to search for your home away from home whether booking for business or leisure. Easily compare the best rates and all properties side by side

Architecture

Chicago's architecture is one of the most stunning in the world. Some landmarks you shouldn't miss are the Sears Tower, John Hancock Center, and Cloud Gate.

Beaches

Think that Chicago is just for urban-dwellers? Think again. Chicago has **26 miles** of shoreline, **26 beaches**, and an **18.5 mile** bike path along Lake Michigan.

Food

There are over **4,700** restaurants in Chicago. The city is home to some of the world's most elite chefs at the forefront of cutting-edge cuisine.

Green Spaces

There are over **500** gardens and parks to explore in Chicago. One of this is Millennium Park, where you will find the iconic Cloud Gate metal sculpture.

The Arts

Chicago is home to the finest art galleries. The Art Institute of Chicago has more than **300,000** works. Tripadvisor hails it as the world's **#1** museum.

Live Music Venues

This city is in love with jazz and blues, as well as music festivals and concerts of all kinds. If you're lucky, catch a show by the Chicago Symphony Orchestra.

FUN FACTS ABOUT CHICAGO

35M An average of **35 million** visitors and tourists flock to Chicago every year.

4 states are visible from the top of the Willis Tower: Indiana, Illinois, Michigan, and Wisconsin.

One of the largest collections of **Impressionist** paintings is found at the Art Institute of Chicago.

At **23.5 miles**, Chicago's Western Avenue is the world's longest continuous street.

Bring the fun back into traveling and let us worry about the details. Search our site now to check our your next deal in Chicago!

ALL ABOUT

AIRBNB CHICAGO

Estimated Occupancy of Airbnb Properties

75.9% of properties are High Availability

"Airbnb is different from most brands. We're a community of individuals, and yet there's a consistency holding us together through the values we share."

Brian Chesky - CEO of Airbnb

Entire Homes/Apartments | Private Rooms | Shared Rooms

Chicago is made up of 77 community areas that are often grouped into 9 districts. There are around 8,500 Airbnb listings in Chicago as of February 2021.

Learn more at www.comparebnb.com

Appendix - Infographics

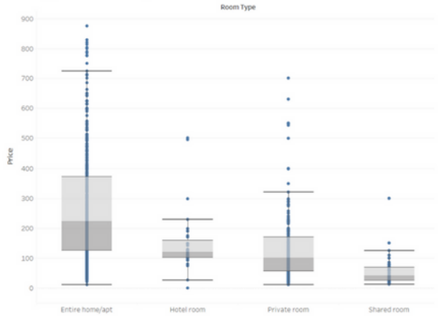
Analysis of Airbnb Data in Chicago

THE UNIVERSITY OF 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.

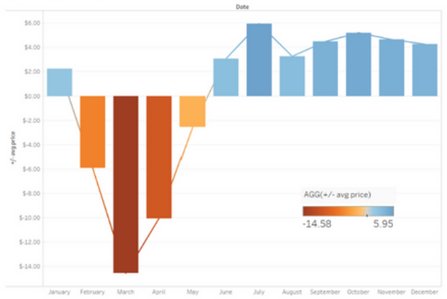
Room Type and Price Boxplot



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.

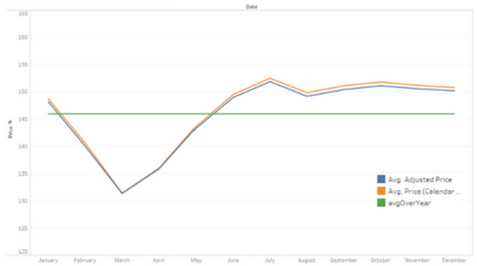
+/- Average Price per Month



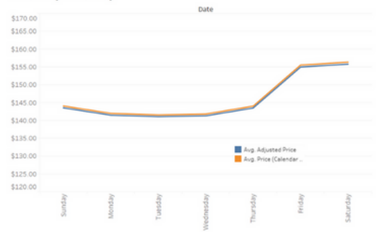
Time of Year

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



Prices by Weekday



Day of Week

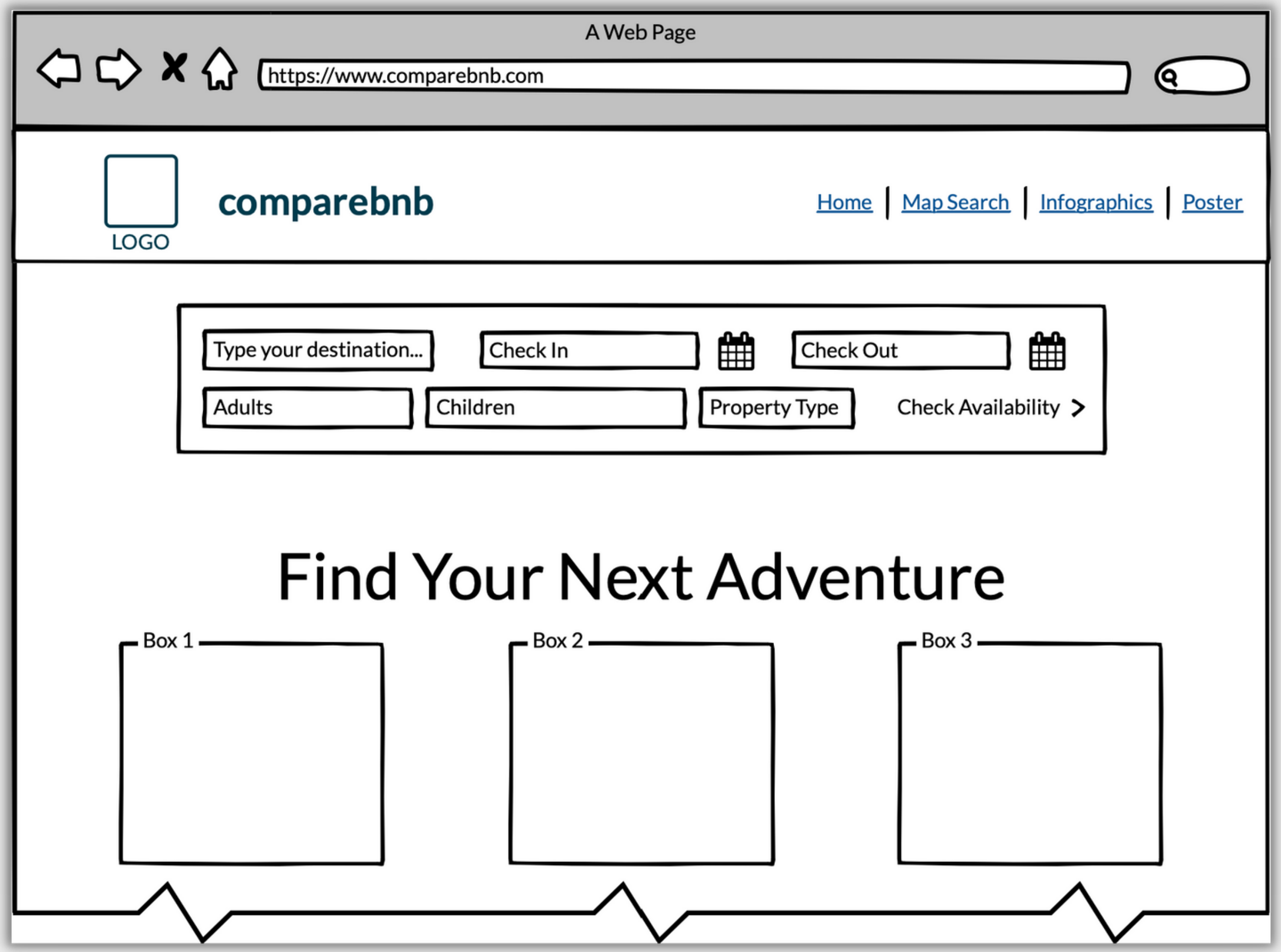
As expected, the prices for Airbnb's Fridays and Saturdays are the highest amongst the days of the week. There is a slight dip in the prices going on Tuesdays, but it is almost negligible compared to the increase in prices on the weekends

airbnb

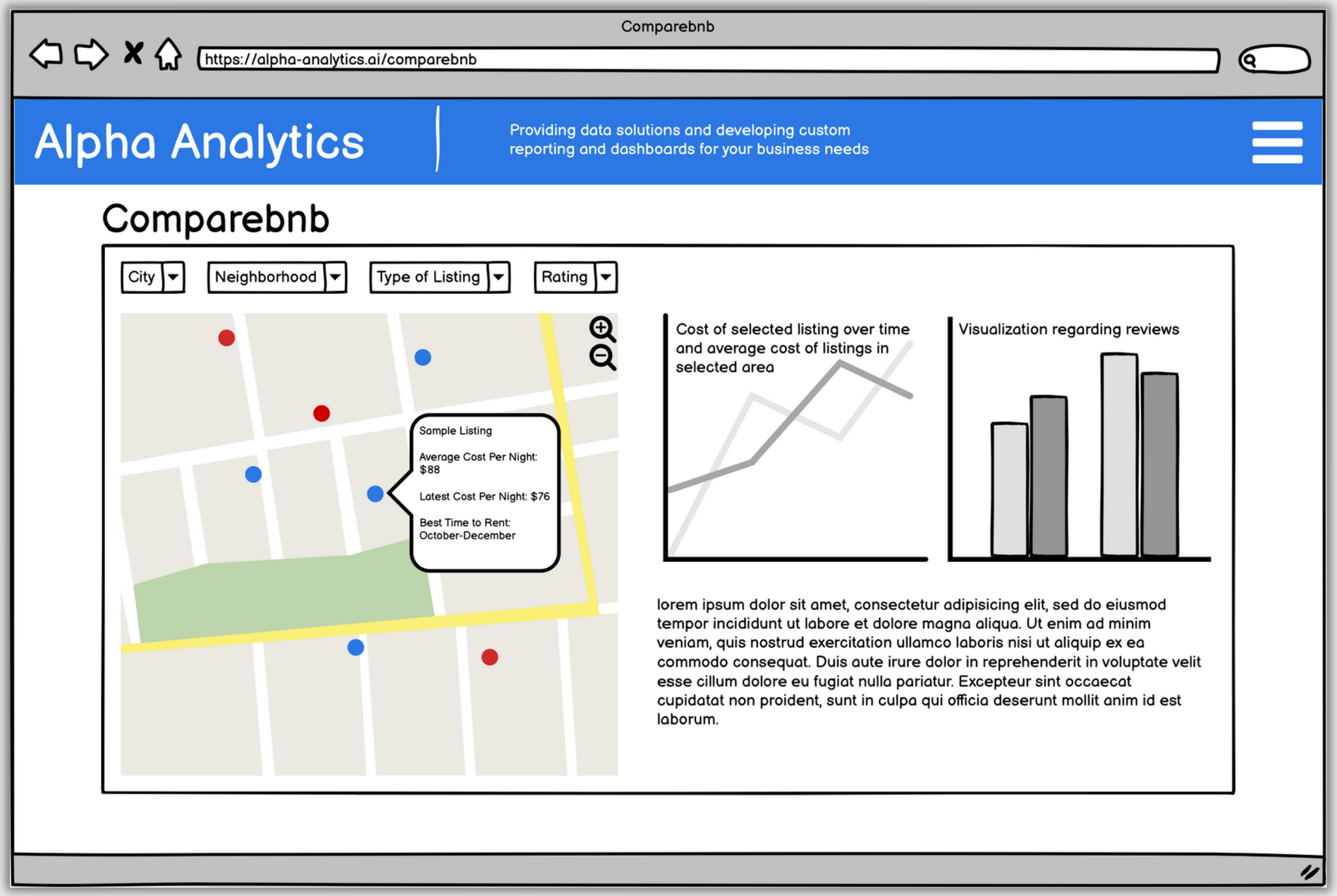
Conclusions

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.

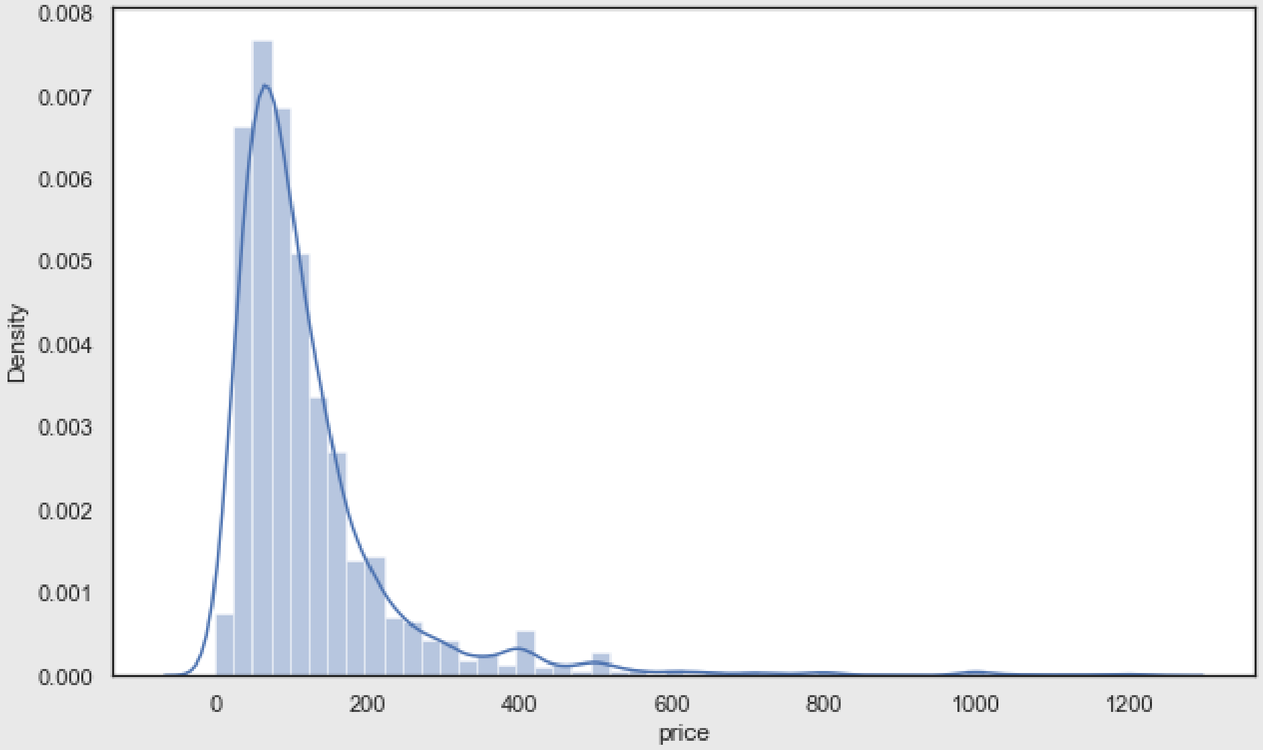
Appendix - Project Analysis Poster

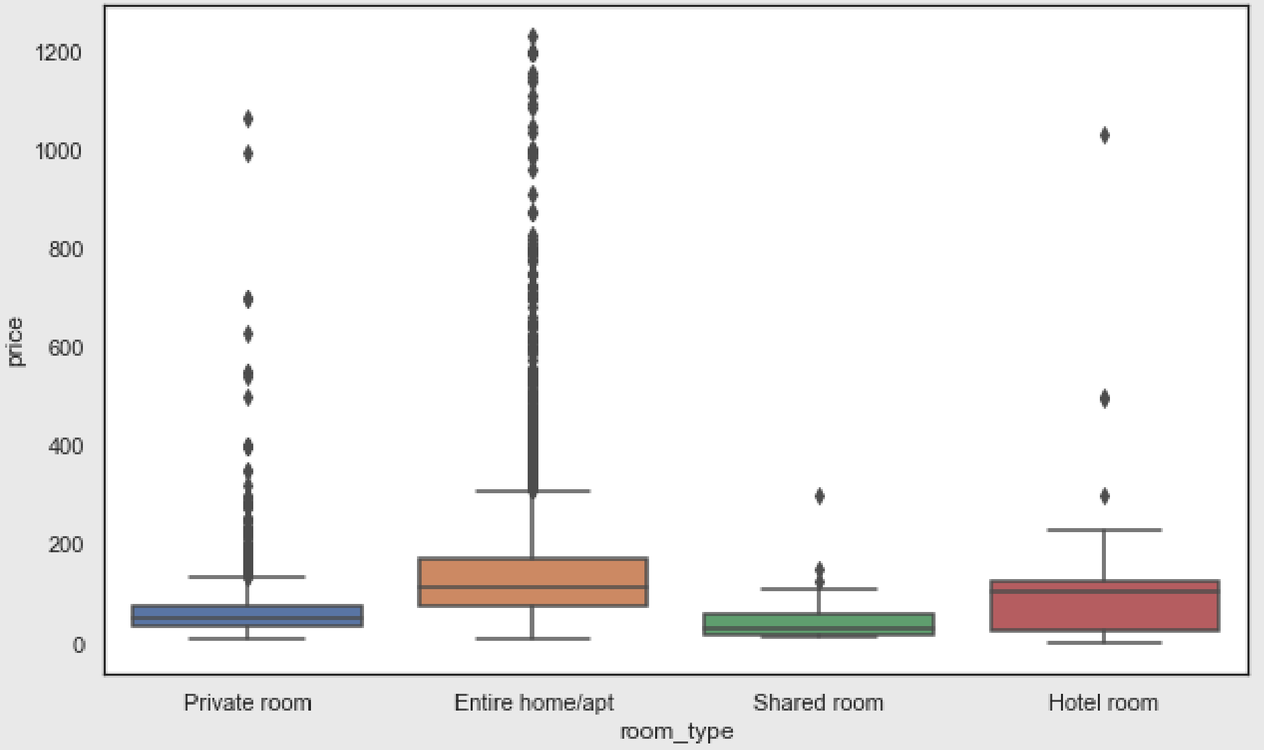


Appendix - Comparebnb Wireframe

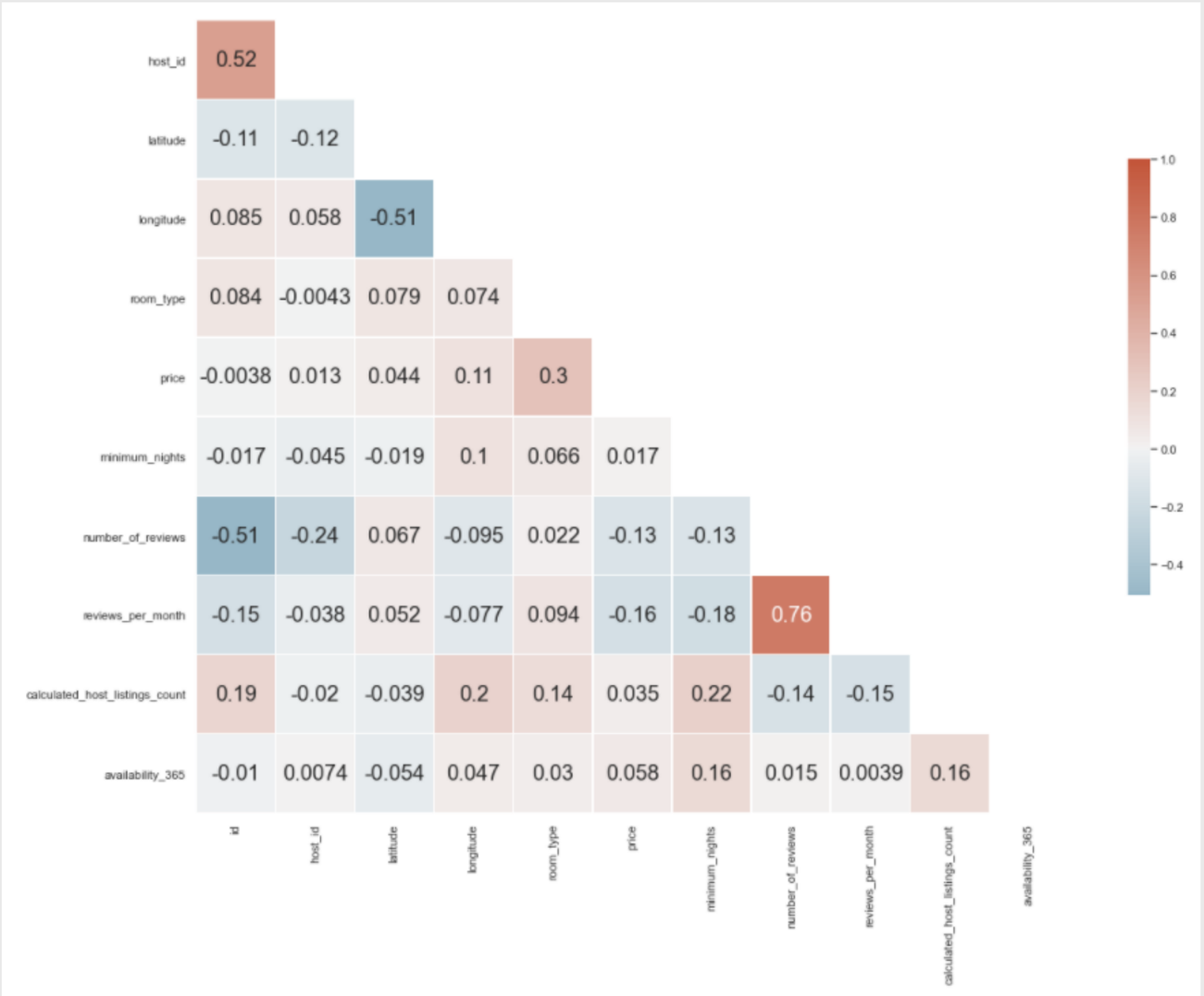


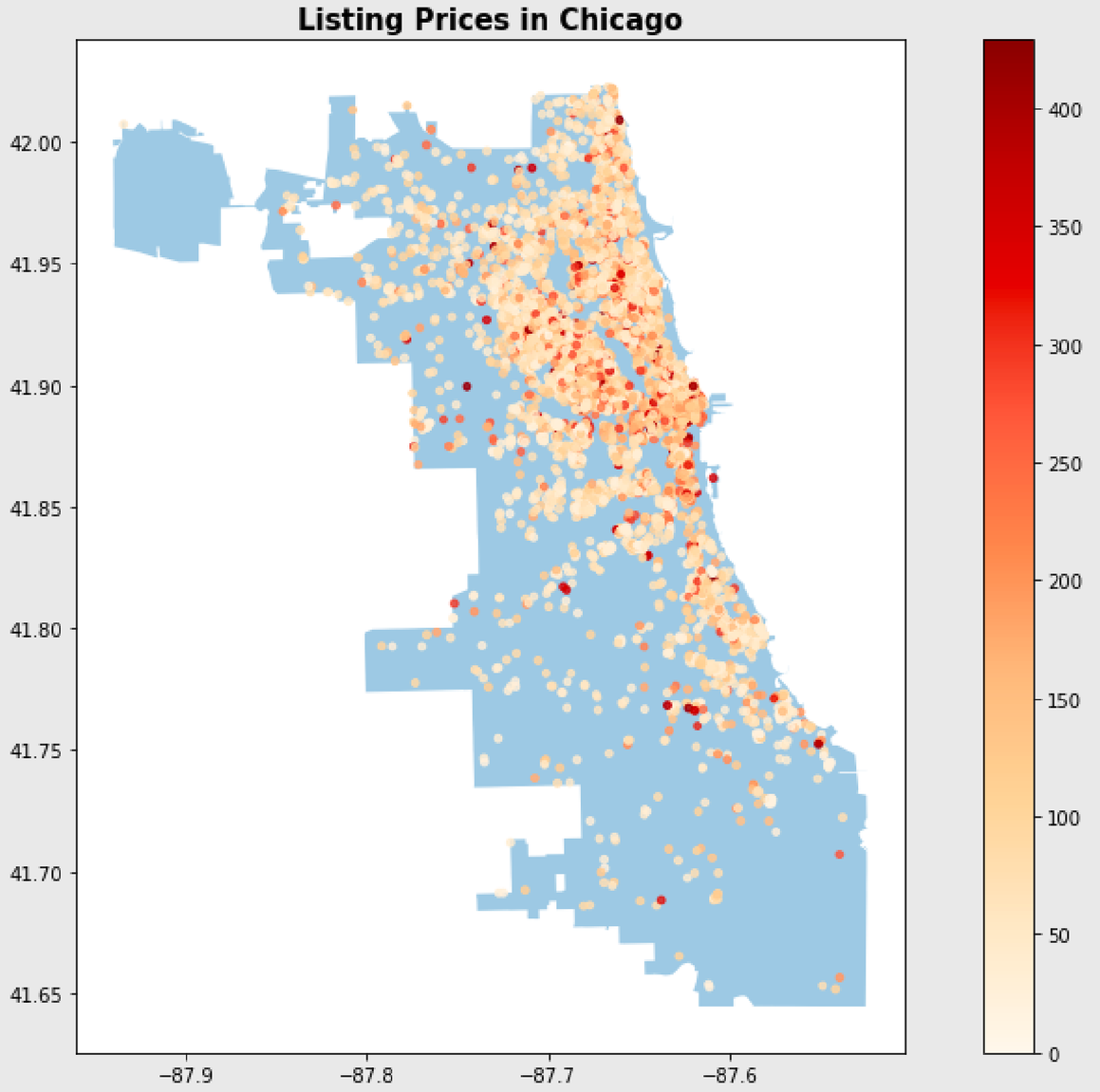
Appendix - Map Search Wireframe





Appendix - Python or Tableau Visual





Appendix - Python or Tableau Visual

Appendix - Python or Tableau Visual