Process Book

Neer Jariwala
Kenneth Morton

Introduction

Overview

Our projects used data from Kaggle's "New York City Airbnb Open Data" dataset. Find it here. The data consisted of about 40K rows with individual Airbnb listings. Some of the columns included neighborhood, price, ratings, and type of building. We created our website using vue.js and beufy, which was new to both of us.

Project Goals/Questions

Our goal was to find compare the price of listing to different aspects. We wanted to explore data that would let us and the people who would see our visualizations to be able to understand where in NYC would be an optimal place to stay. In order to organize ourselves, we created a few questions:

1. What are some of the most crazy and expensive airbnb listings in NYC?
2. How do airbnb's average prices changes depending on how close they are to popular landmarks?
3. Which neighborhoods have the most expensive airbnbs?

Motivation/Inspiration

With summer just around the corner, vacation tends to be on people's mind. Sometimes, trying to find a place to stay close to the places you want to visit can be the hardest part. As such, we wanted to create a set of visualizations that would help people in their search.

When choosing what kind of graphs to make, we drew some inspiration from our classes about small multiples and maps. We felt that using these methods would allow us to explore the data from certain angles that others graphs would not.

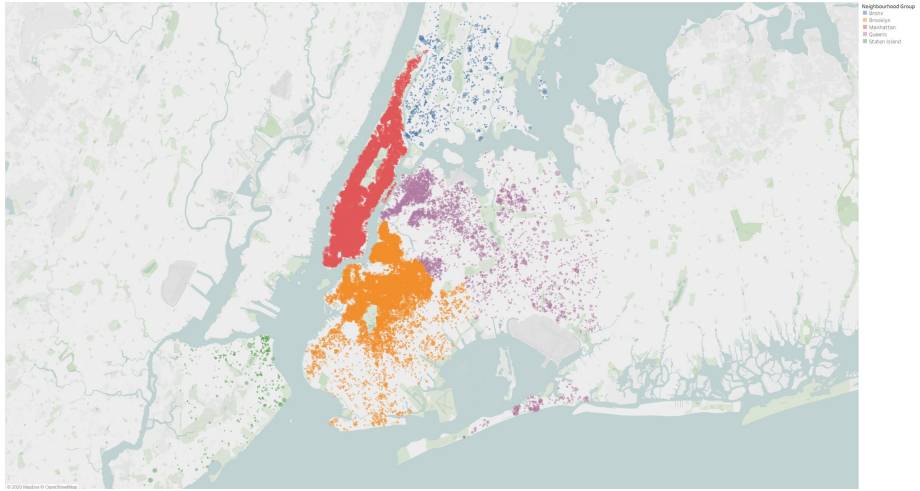
Data Engineering

Data Processing

Our data was already cleaned up and formatted when we downloaded it. So, we did not spend much time of data cleaning, but rather mostly on data manipulation. Using excel and python libraries such as pandas and geopy, we performed calculations to get our RoundedDistance and AveragePrice metrics. However, while doing these calculations, we noticed some outliers that were causing our graphs to be significantly skewed. As such, we elected to remove these rows when rendering certains graphs in order to provide the viewer with a more understandable visual.

Exploratory Data Analysis

When first looking at our data, we mainly used a map we quickly made using Tableau. The Tableau image on our website is actually the same one we used during our initial exploration. We looked for details on price, location, neighborhood, etc. While doing so, we noticed that price was easy to compare to details, but also resulted in interesting statistics. As such, we choose to use price as our common factor between our visualizations.



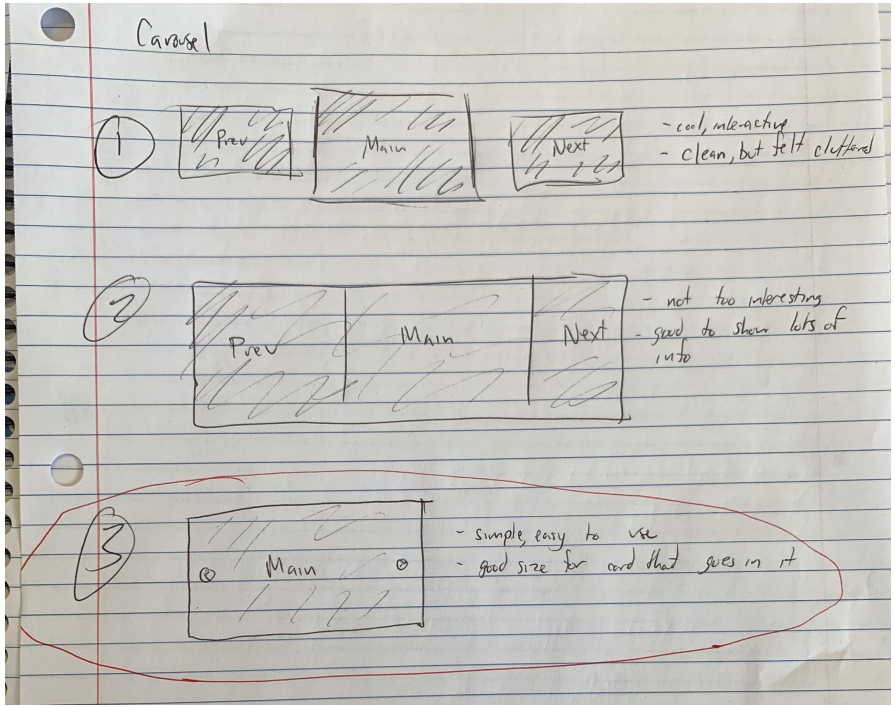
Proposal vs Final Product

There were some differences in the proposal vs final project

- We originally wanted to do a heatmap of prices in locations. We felt that the heatmap would be too cluttered for this data
- In our proposal, we intended to focus on the map as the main, central point of the website
    - After doing some thinking, we decided that it would be cool to explore the average airbnb price as you get further away from a popular landmark
    - We decided to do small multiples instead of the heatmap

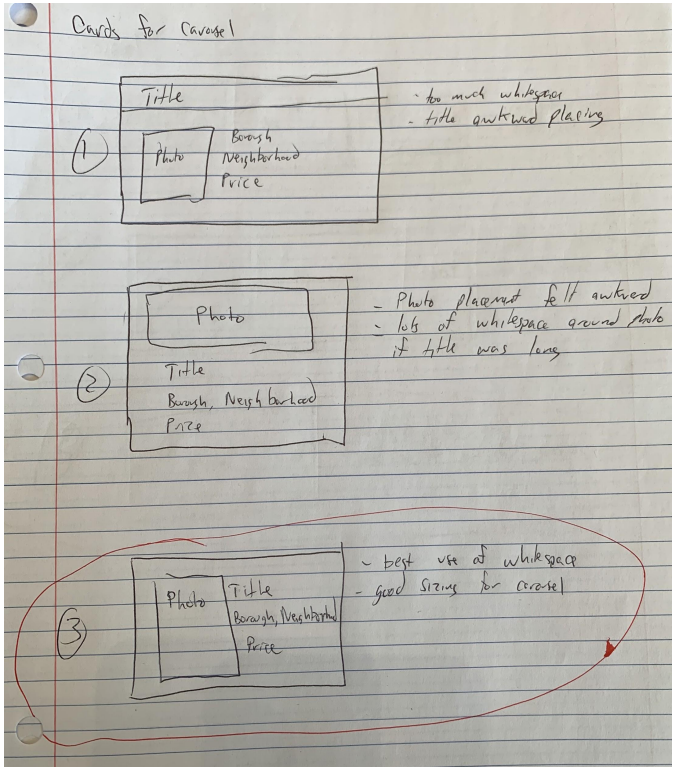
Charts

Carousel - Design Choices



- Initially, I wanted to show multiple cards on the same screen
  - This felt cluttered on implementation
- I settled on showing one, single card at a time
  - This design flowed better with the design of the card itself, and was less cluttered

Carousel Card - Design Choices



- I started by placing the title at the very top in it's own section
  - It looked weird, and there was alot of whitespace
- I also tried a vertical implementation
  - It didn't work with long titles - there was alot of wasted space to the right of the photo
- I settled on a simple design, with the photo on the left and the information on the right

Carousel - Intent and Functionality

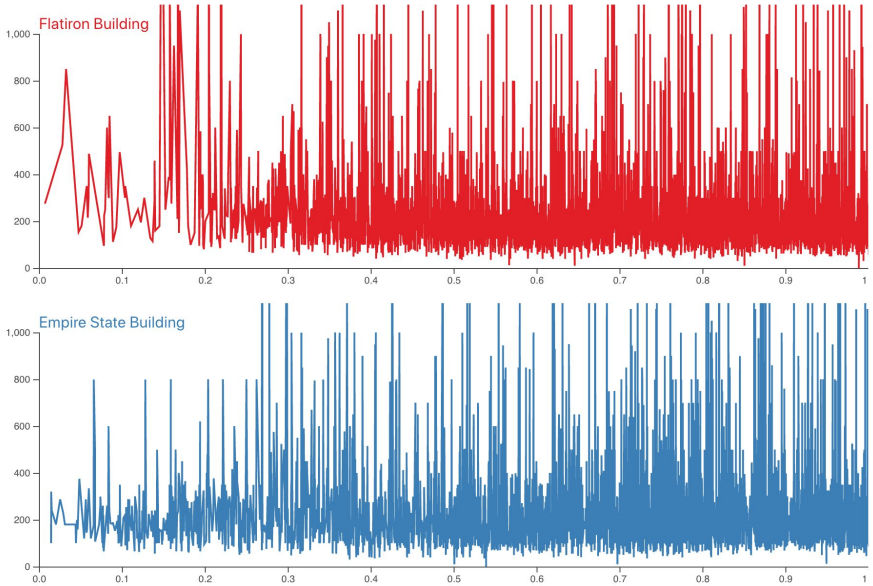
Intent

- The intent of the carousel was to provide a fun, somewhat ridiculous insight into the most expensive airbnb's in the city
- It also shows that some places in our data simply aren't worth it - a one bedroom apartment in chinatown for $9,999 / night is not a good deal

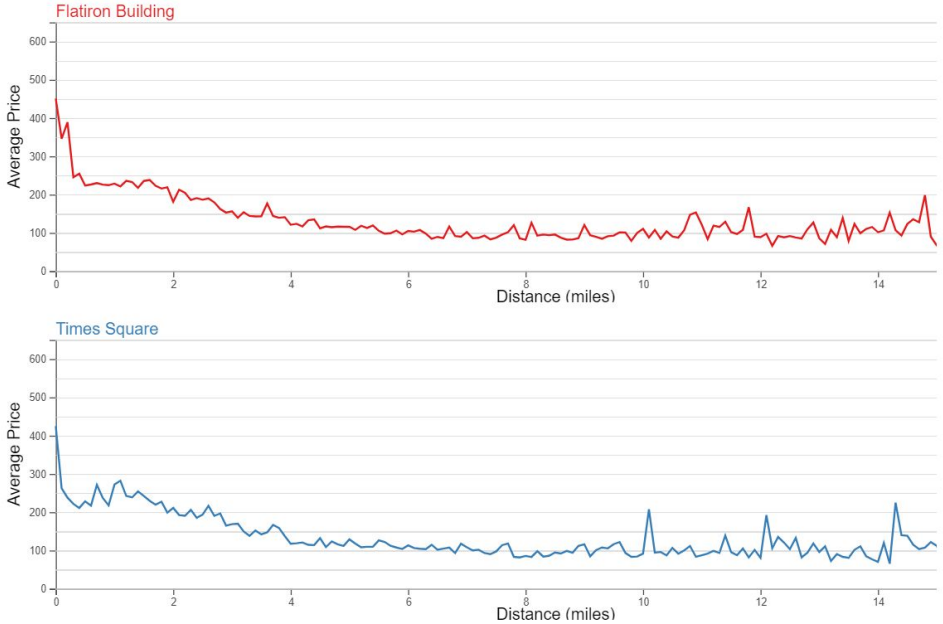
Functionality

- You can click thru the carousel on desktop, or swipe thru it on mobile
- There is a link to the original listing on airbnb

Small Multiples - Design Choices



Old



New

Small Multiples - Design Choices

When initially creating the line charts, we wanted to be able to be able to show every listing. However, as you can see from the “Old” graph on the previous slide, the graph became an extremely dense. No one would be able to understand to get useful information of of this, so we decided to tweak our approach. Our first idea was to create a scatter plot instead. However, that would require us to almost completely redo our code and we were still slightly concerned it could be a little too dense. As such, through some data manipulation and calculations, we were able to get a rounded distance and average of price. This can be seen in out “New” layout on the previous slide. This view is much clearer and allows the user to not be overwhelmed.

Small Multiples - Intent and Functionality

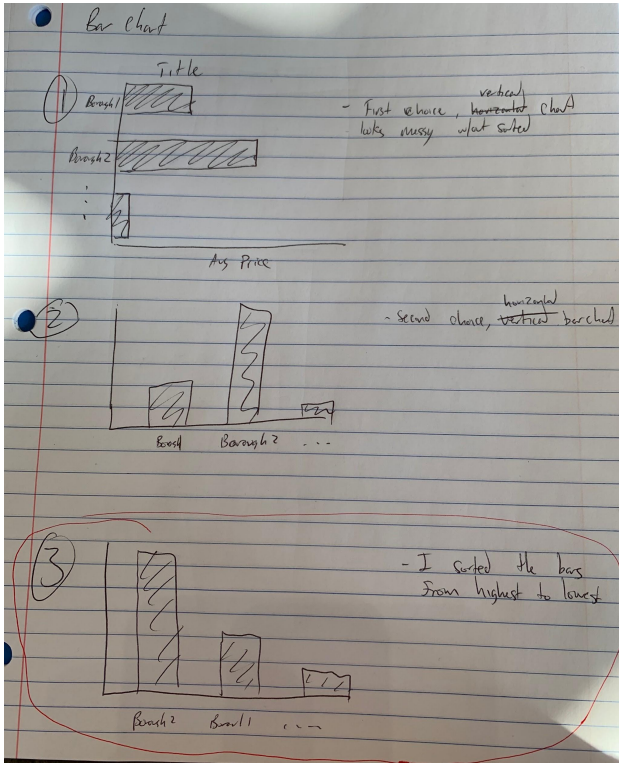
Intent

- The reason behind creating this graph is to gives users a way to see how far from their targeted landmarks they would need to stay in order to save the most money. They would also be able to compare the prices between different locations.

Functionality

- Hovering your mouse of a specific spot on the line will display the distance and average price at that spot.
- Color coded by landmark.

Bar - Design Choices



- Here is the evolution of the bar chart
  - I initially tried a vertical chart, but ended up liking the horizontal version more.
- I also decided to sort the bars from tallest to smallest
  - I liked the look of this more than when they were randomly scattered throughout.
  - It also helps make Manhattan stand out as the most expensive Borough.

Bar - Intent and Functionality

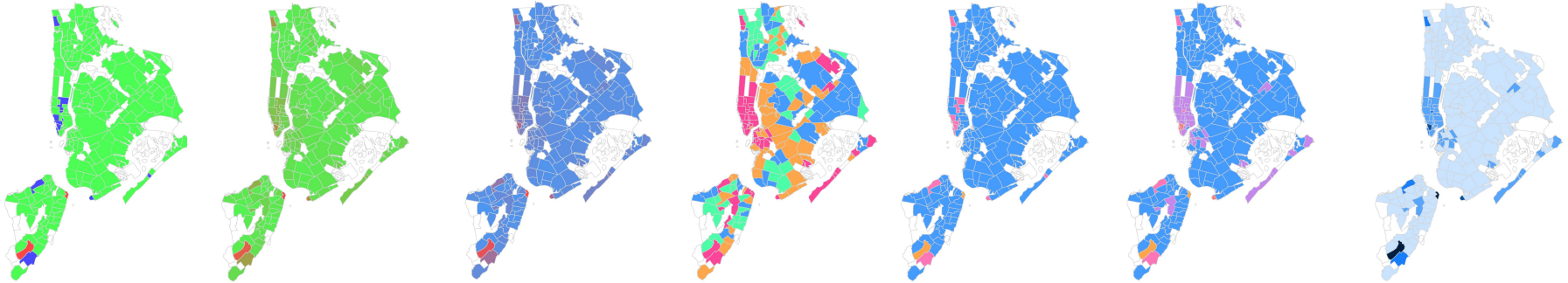
Intent

- The main goal of the bar graph was to give a very general overview of the average price in the five boroughs of NYC

Functionality

- We kept it simple, with simple interactivity on hover.
- It does get the message across that Manhattan is the most expensive borough

Map - Design Choices



- From left to right, the evolution of the map color choices
- In the center one, I experimented with a color scale of many colors... it didn't turn out well lol
- The left 3 looked decent, but the colors were continuous, meaning it would be hard to determine price range groups for the legend
- I settled on the rightmost one, as it went with our color scheme and wasn't too confusing

Map - Intent and Functionality

Intent

- The main goal of the map was to illustrate which neighborhoods were the most expensive in NYC

Functionality

- The map was built to be interactive, so you could hover over different neighborhoods
  - A tooltip would tell you the neighborhood, borough and avg price
- The color scale was built to be simple, easy to read, and consistent with the blue color scheme

Results and Improvements

Results

We were able to answer the questions we asked at the beginning

1. What are some of the most crazy and expensive airbnb listings in NYC?
    a. Stay on a boat, a film location, or even a penthouse with no furniture. NYC has a place for every budget!
2. How do airbnb's average prices changes depending on how close they are to popular landmarks?
    a. Generally, airbnbs get less expensive as you move away from the popular landmarks. One of the best distances we found was about 8 miles away.
3. Which neighborhoods have the most expensive airbnbs?
    a. Stay away from Manhattan if your main goal is to save money. There are some neighborhoods in Manhattan such as Lower East Side or Chinatown that are a little cheaper.

Future Improvements

There were a few improvements that came across our mind that we would like to implement

- Changing the color scale of the map from a dropdown
- Editing the range for the colors on the map to make easier to see differences in lower numbers
- Determining if a scatter plot is better than a line graph for the small multiples graphs
- Plotting some ridiculous airbnb's on an interactive map
    - Linking them to the airbnb website as well
- Making the hover functionality for the line chart work if you hover above or below the line
    - This makes it easier to get the value at a given point
- Pulling carousel information dynamically from airbnb
    - This way, the carousel cards are always up to date

Related Materials

- https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data
- https://www.d3-graph-gallery.com/index.html
- https://nuxtjs.org
- https://www.d3-graph-gallery.com/graph/line_smallmultiple.html