

PROCESS BOOK

NYC AIRBNB

2019

**TEAM
XIANZHI**

TABLE OF CONTENTS

01

Cover Page

02

Table of Content

03

Brainstorming

04

Data Exploration

05

Visualizations

06

Sketches

07

Challenges

08

Contribution

BRAINSTORMING

The first part our project consisted in brainstorming to find a suitable dataset that we all find interesting.

During this phase, we found three interesting datasets. The first was the HIV/AIDS dataset that gave information about the number of cases of people with HIV/Aids in the world and the evolution of the epidemic throughout the years.

The second dataset was about the United Nations' sustainable development goals (SDG). It contained information on all goal, targets and indicators for all countries and displayed the progress of each country in accomplishing those goals.

The third dataset was about Airbnb data for the New York City in 2019. It contains information on all the Airbnb listings, including their exact geographical coordinates, their neighborhood, owner, average review, price, availability.

We ended up choosing the the Airbnb dataset, because we were excited by the potential to do a 3D map visualization to represent the buildings of the city of New York, given that we have the exact latitude and longitude for each listing.

We were also motivated by the fact that creating plots to visualize the listings would greatly help users get insights on what areas and listings correspond to their budgets and needs.

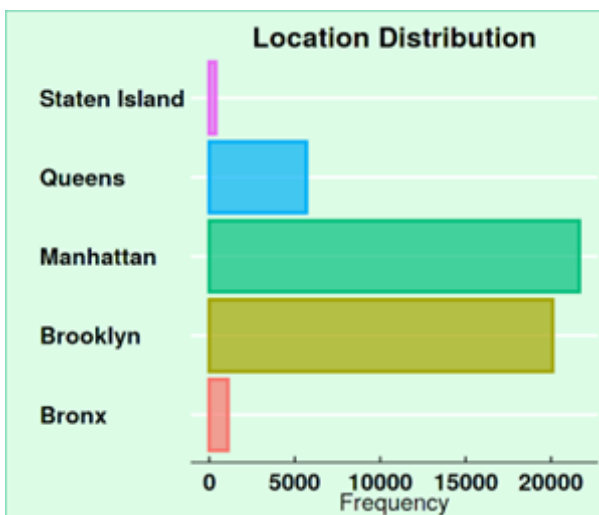
DATA EXPLORATION

Now that we chose our dataset, we started thinking about what kinds of visualizations we want to do. But before, we did some data exploration to get information and insights on the dataset we had.

In our initial analysis, we found that:

- The dataset includes information about 5 different city districts, which are 'Brooklyn', 'Manhattan', 'Queens', 'Staten Island', and 'Bronx'.
- There are a total of 221 different neighborhoods included in the dataset.
- The types of rooms available for rent are 'Private room', 'Entire home/apt', and 'Shared room'.
- For different districts, both average prices and median prices are different. Among all the districts, Manhattan has the highest mean price of 196.8 while the Bronx has the lowest mean price of 87.49

These initial statistics give us a general understanding of the scope and contents of our dataset. We can use this information as a starting point to further explore and analyze the data.



**This information is
a starting point to
further explore and
analyze the data.**

VISUALIZATIONS

Based on our Exploratory Data Analysis, we now have insights to decide what features from our dataset we want to do what kinds of visualizations. Here is a list below!



01 — 2D Neighborhood Map

The purpose of this 2D map is to show how listings are distributed in New York City. Upon clicking on each area, the size of the area would change based on the number of listings, the average price.



02 — Sankey Diagram

The Sankey diagram would help us see the main type of accommodation per area (Entire home or apartment, private room, shared room) as well as the average price in each area.



03 — Dot Plots

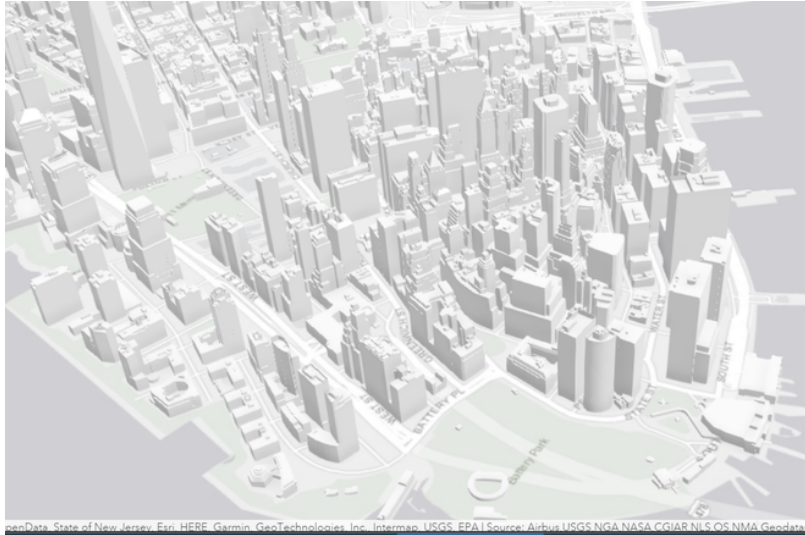
We decided to create two dot plots: one to show the difference in number of reviews between two neighbourhood and to the distribution of reviews during each month. The second was to show the correlation between reviews and price for each area.



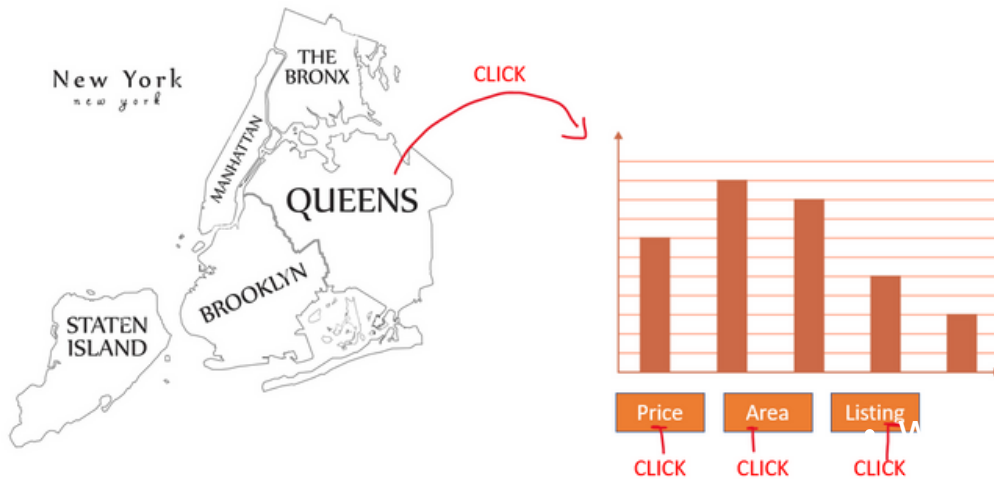
04 — 3D Map of NYC

The most important visualization would consist of a 3D map that was created in ArcGIS, using the longitude and latitude of each listing to include them in the map. Users should be able to hover to see details of the listings. We also wanted to include tools to measure the distance between a place and the other.

SKETCHES



- Initial sketch for the 3D map of NYC
- The arrows show the location of the listings were added
- We also added a search box to search for an area and a distance measure

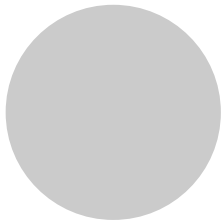


Initial sketch of the 2D Map. At first we wanted to click on each of the five boroughs to display a bar chart according to price, area size or number of listings next to the map.

At the end, we decided to let the map get "deconstructed" when we click on an area, letting all five areas being displayed next to each other with each area corresponding to a bar on the bar chart.

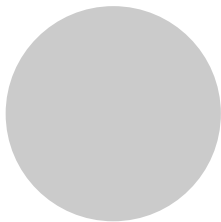
CHALLENGES

In this section, we detail the three main challenges we faced during this project. Challenges related to small bugs and coding difficulties will not be detailed here, since they are minor challenges and they have been fixed



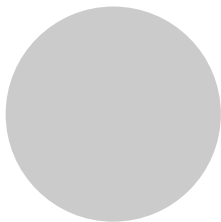
01 — Limited dataset

Our dataset only showed data for the city of New York for the year 2019. This greatly limited the flexibility of the visualizations that we wanted to do (evolution through the years) and gave limited insights to our analysis



02 — 3D Map size

The 3D map was done using ArcGIS, since D3.js has limited feature for 3D visualizations. A main challenge is that since it is in 3D, it actually has a big size and takes time to load



03 — Flexible height for 3D Map

Initially, we wanted to implement the 3D map in a way that the building height would correspond to the price of the listing. This was challenging since we don't have listings in all buildings of the city, which would result in esthetically unappealing visualizations.

CONTRIBUTIONS

In this final section, we detail the contributions of each member in the team. At the beginning, the brainstorming and visualization types were all completed collectively. Then, the work was divided for the implementation as follows:

MARIELLA DAGHFAL

- Website development, layout design, logo design

XINGCHEN LI

- 3D Map, Reviews dotplot implementation

ZOU MENGJIE

- 2D Map, Sankey Diagram and Price dotplot implementation

All of the parts were reviewed by all members. Even if the work was distributed individually, the choices and decisions were made collectively with the agreement of all three members of the team.