NEW YORK CITY AIRBNB

DASHBOARD

IMT 562 FINAL PROJECT REPORT

TEAM 1 : Aditya Challa, Jwala Mukhi Suresh, Yash Patel

LINK TO VISUALIZATION : **NEW YORK AIRBNB DASHBOARD**

Table Of Contents

ABSTRACT

In our project we have used the New York Airbnb dataset from Inside Airbnb to explore the various airbnb listings available in the city. We created an interactive dashboard to help our users to explore and find airbnbs in the five boroughs in NYC that suit their budget. We have included other interesting airbnb features like review scores, availability, accommodates and amenities to help the user narrow down their selection.

INTRODUCTION

About The Data

The dataset includes hosts, geographical availability and necessary metrics to make predictions and draw conclusions about the pricing, listings and ratings of airBnB listings in New York.

Total number of records : **38277**

Features of interest : Neighborhood Groups, Neighborhood Areas, Review Scores, Availability, Price, Room Type, Latitude, Longitude

Link to the source : http://insideairbnb.com/get-the-data.html

(New York City, New York, United States - 04 December,2021)

Motivation

Over the last few years, as Airbnb and competing platforms have grown in popularity, the short-term rental market has attracted considerable attention in cities around the world. Since short-term rentals are more geographically dispersed than hotels, the benefits of tourism (which include the earnings of Airbnb hosts and increased economic activity for local businesses) are spreading to a greater number of largely residential neighborhoods where hotels are often not permitted. (Coles et al., 2017) Since New York is one of the most popular cities for tourism we chose the New York Airbnb dataset for our project.

Initially we planned on working with the Kaggle dataset but we were able to find a much more exhaustive and recent dataset (Upto December-end 2021) from Inside Airbnb. The availability of geospatial data combined with multiple columns for different airbnb features helped us to ask pertinent questions and also to create more robust visualizations. We wanted to use our learnings from the class to create an

interactive visualization that can help the users find their preferred airbnb's that best fits their budget and expectations.

Area of Focus

We followed the **A.S.S.E.R.T** model to design and evaluate our visualizations

- **A**sk a question
  - We created 5 EDA questions to explore the dataset
- **S**earch for Information
  - We found the dataset from Inside Airbnb as a reliable source for our visualization. The site also provided a data dictionary which helped us to understand the features included in our dataset.
- **S**tructure the Data
  - The first step in structuring our data was to link the 5 boroughs to the neighborhood areas so that the user can see their desired information at an area level. Then we normalized the price and availability features to overcome the issue with the wide range.
- **E**nvision the Answer
  - We used a map to answer our initial questions. Then to make our visualization more intuitive we included treemap and box plots to give more details to the user at a granular level. The bubble chart was created to help users find the top 10 Airbnb's for each of the attributes with least effort.
- **R**epresent the Data
  - We made the visualization interactive by adding filters to choose the borough, neighborhood area, and the attribute the users were interested in (Price, Availability, Review score). To improve the effectiveness of our dashboard we included a visual to show the average price, availability or review score for the selected borough.
- **T**ell a Story
  - Our dashboard explores the usage of Airbnb's across the 5 boroughs in New York and focuses on improving the user experience of finding an Airbnb in New York using our interactive dashboard.

DATA QUALITY ISSUES & TRANSFORMATIONS

- The dataset had many missing values. **9504** airbnb listings weren't having a review score.
- There were erroneous values in Price. A few listings had prices as low as 0$. We used the average price instead of using the actual values.
- Price and availability column had a wide range of values. So we normalized them for better visualization.
- Most of the columns were categorical. So we had to plot them against the ordinal to create visualizations.
- Boroughs weren't linked to their corresponding neighborhood areas. We fixed the issue by creating filters to show only relevant areas for the selected borough.

EXPLORATORY DATA ANALYSIS QUESTIONS

1. Which boroughs have the highest and lowest availability of listings ?
2. Which borough has the highest average nightly cost and the lowest average nightly cost amongst listings?
3. Which boroughs have the highest rated AirBnB listings?
4. How many properties are available within a cluster for each attribute (Price, Availability, and Rating)?
5. What is the distribution of properties in each neighborhood for different attributes (Price, Availability, and Rating)?

INITIAL VISUALIZATION

Five Design Sheet

Our team brainstormed the initial ideas of the dashboard through the Five sheet design model. The five sheet design model included steps for our project within which are brainstorming, initial designs and the final design which is often called the realization sheet/final sheet which is as seen in the below figure. The realization sheet included a map of New York city that displays the listings available that are ready to occupy. Our team also intended to do the valuation through a multi linear regression model predicting the price of the listings using multiple factors such as availability, amenities, beds, bathrooms, proximity to Manhattan etc. Subsequently, our design

also had a violin plot which will show the distribution of listings in each of the five neighborhoods of New York city within the dataset.

Furthermore, we also had a radio button selection for the neighborhoods under the name Area on the left pane listing the different neighborhoods of New York city within the dataset. The different attributes of measures such as Pricing, Rating, Valuation and Availability are listed on the left pane with a radio button option as well.
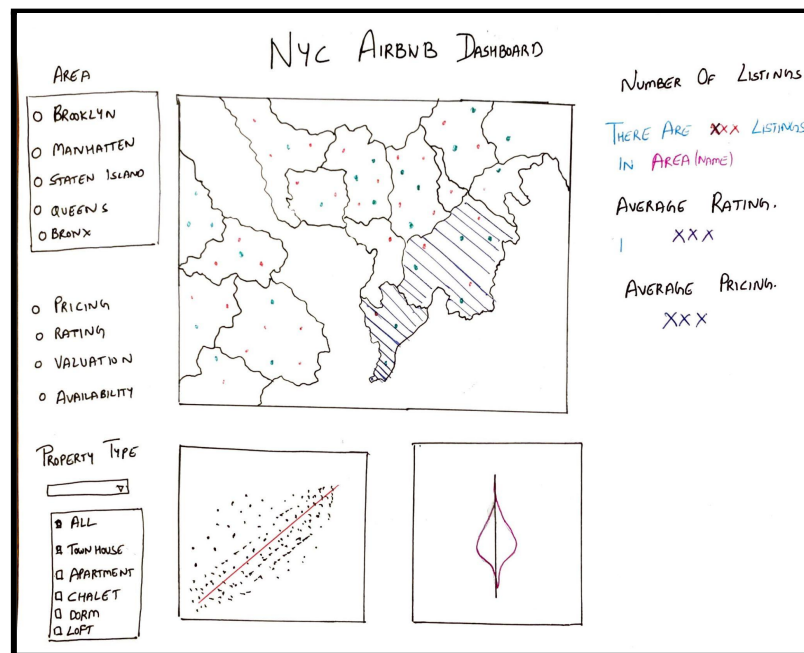

Figure 1 : FDS for our dashboard

USABILITY TESTING

Methods

Our team decided to choose participants who either aspire to travel to New York city very soon (who haven't been there yet even once in their lifetime) and as well as frequent visitors to the city of New York. The intrinsic reason behind choosing participants for our usability test that have travel affinity with New York is to gain insights of the frequent travelers and their experiences trying to find an accommodation in the New York city. We had 5 participants for the usability test out of which 2 were students. Three of the participants visited New York city more than 5 times while the other two haven't been there yet but look forward to visiting the city very soon. We as a team opine that this will help us fine tune our dashboard with the insights from both experienced/seasoned travelers and the new travelers.

Our methods for the usability test include direct testing on the zoom call, where we had a little small conversation about a few current affairs to ease the participants and subsequently asking questions in the test script. Directed testing also helps our team to note the intrinsic expressions of the participants over just answers and opinions. Our team was also readily available to answer any questions from our participants in case they needed help navigating the dashboard.

The test script involves questions pertaining basic demographic questions, followed by the key insights of visualizations and tasks pertaining to the visualizations related to the dashboard. The primary parameters that our team utilized through this usability test are usefulness, completeness, perceptibility, truthfulness, intuitiveness, aesthetics and engagement. See **Appendix** for the Usability Test script.

Participant Details

Beneath are the demographic details of our subjects that our team conducted usability tests for our NYC Airbnb dashboard.

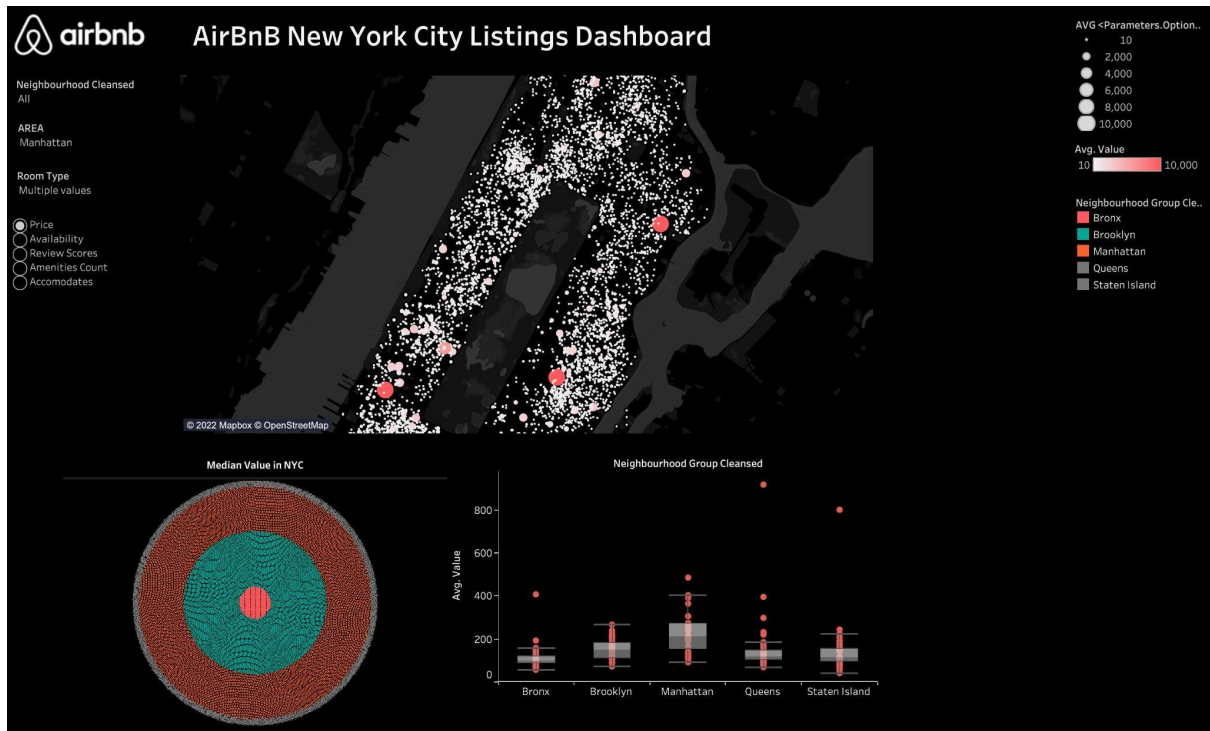| Participant	Age	Gender	Profession	Airbnb User	NYC visitor
Subject 1	27	Female	Software Developer	Yes	Yes
Subject 2	31	Male	IT Auditor	Yes	Yes
Subject 3	35	Male	Software Developer	Yes	Yes
Subject 4	21	Male	Student	No (But aware of what Airbnb is all about)	No
Subject 5	22	Female	Student	No	No

Figure 2 : First version of our dashboard

Observation & Findings

Our team pivoted from the initial five sheet design model even while creating the first version of the dashboard. This is due to the usability test that we conducted based on the visualizations that our team conducted through python code. We created the regression models and especially the violin plot using the seaborn python library to get the feel and the opinion of the users of our dashboard before we fully build it out in tableau.

Feedback from the Users on Python Plots

The figure below shows the violin plot that our team created for the price attribute. The plot shows the listings as a violin representation for different neighborhoods in the city of New York that are within the dataset. However, we didn't receive positive feedback for this plot and the multiple linear regression. The plot also turned out to be cluttered owing to too many listings within a similar price point. Three out of the five users had difficulty interpreting the violin plot. This is purely due to the lack of awareness of the probability distribution and the message that it is trying to convey. They opined that these plots are tricky and difficult to interpret despite being a good choice to represent from a data science standpoint.
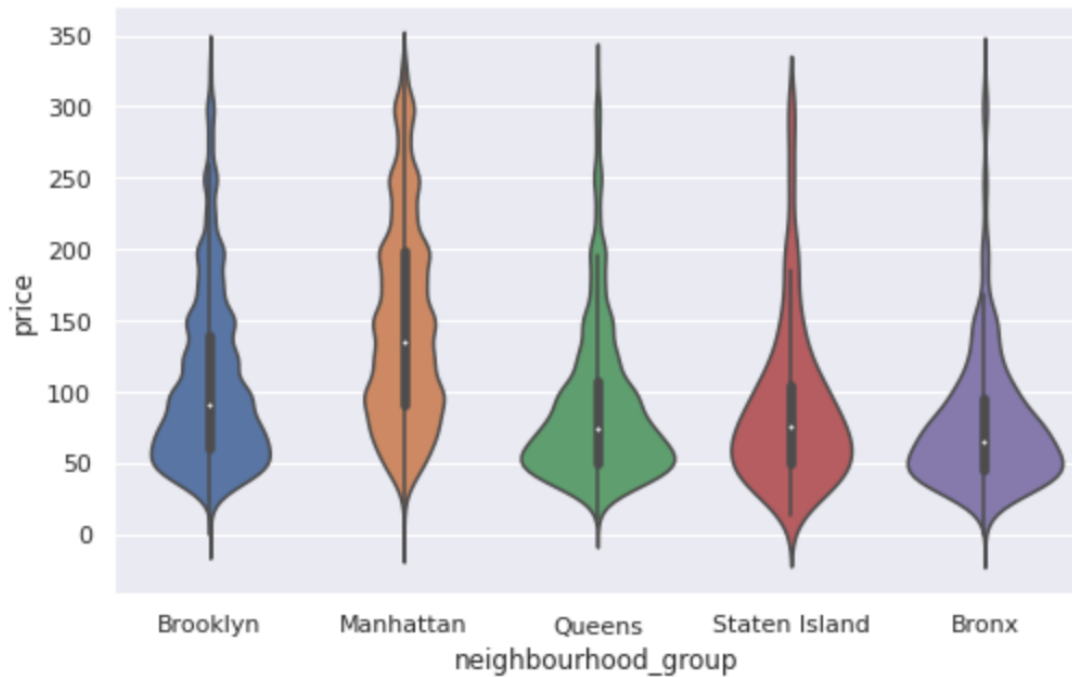
Figure 3 : Price distribution using violin plot

Our team was slightly surprised to hear this feedback and apparently tried to figure out the causation behind these plots being not so intuitive or informative. One of the contributing factors to the confusion created by these plots can be because of the fact that a large percentage of our focus group or subjects do not hail from the technical background. This is an important takeaway for our team as the majority of the user group for our final dashboard might not be technologically literate and might struggle to just hover around the dashboard. As a result, our team tried to incorporate the feedback from our initial usability test and go through the process of five sheet design and brainstorming all over again. While a majority of the visual design and the positions of the visualizations remained constant, we changed the design of the violin plot to the box-cox plot. The box plot is quite intuitive and simple to interpret as it has the upper quartile, lower quartile and the number of data points that are within the range. We also had a bulls eye chart that had the listings of Airbnb in the city of New York. However, from the usability test that our team performed, we got feedback that the bulls eye chart was not that intuitive and informative. The users further expressed that hovering over the chart didn't give them much information and they had a tough time interpreting the concentric circles and the rationale behind those concentric circles.

Final Dashboard: (Presentation Dashboard)

Below is the final version of our Airbnb New york listings dashboard. We have incorporated feedback from the usability tests and made significant and substantial improvements to the dashboard.
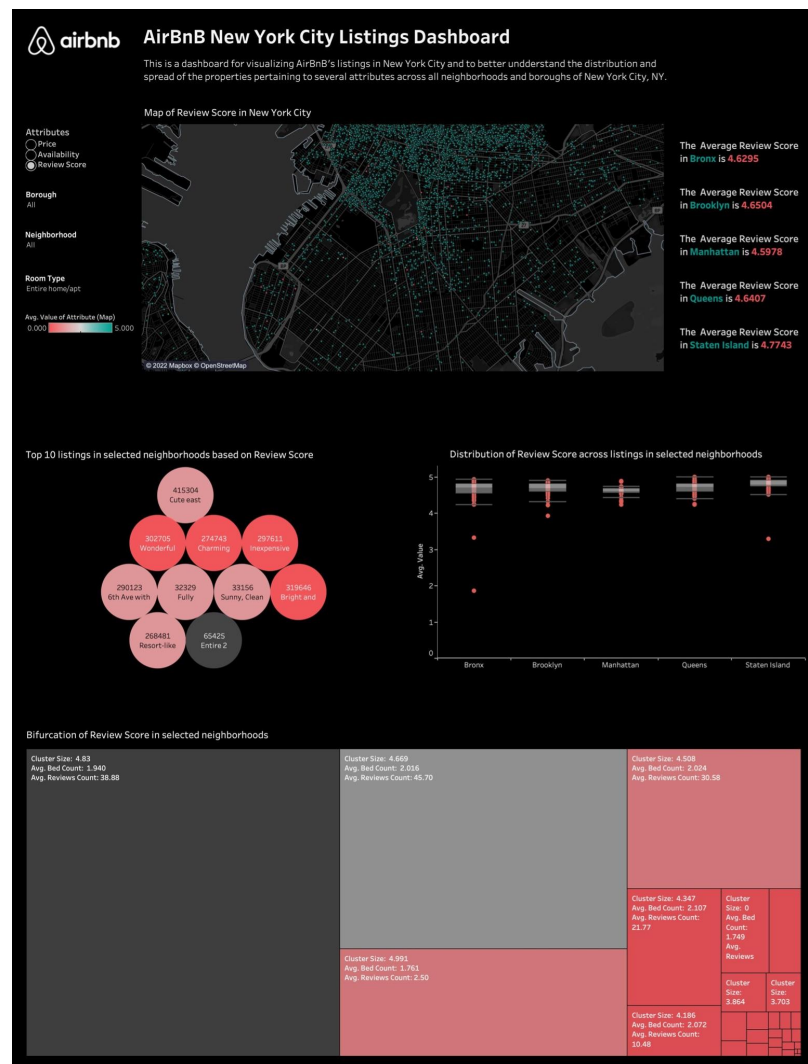

Figure 4 : Final Dashboard

Feedback Incorporated:

The following is the feedback incorporated from the usability test conducted on our subjects based on the first version of the dashboard into our final dashboard:

- Changed the bulls eye chart to the area chart with appropriate legend on the pane to interpret the visualization. The area chart shows the bifurcation of the average of particular attributes selected in the neighborhoods. For example, if the user selects the price attributes on the dashboard, the area chart shows the price alongside the average bed size, review count and cluster size. The cluster size gives the information about the number of listings within that area or region.
- A major visualization change that we made into our final dashboard is that we incorporated to address the concern that was expressed by our subjects. One of the major concerns was that, based on the filters the users expressed concern to identify the top listings so that they can subsequently find them on the map and potentially have a go to book them. Hence, we have created a bubble chart that shows the top 10 listings with when selected appropriate attributes.
- We also addressed the concern of our users to have the text displayed about the average price, availability and rating. Hence on the right pane, we displayed the average rating, price and availability (depending on the attribute the user selects) on the right pane.
- We also segregated the areas under the neighborhood/boroughs. Initially in the dataset even after we select the boroughs, we had difficulty filtering the areas that are within that particular borough. The users pointed out that it was creating confusion. We made an attempt to fix that through first selecting the borough and subsequently listing all the areas under that borough.
- We also changed the size of the display to the size of the points on the map. We decrease the opacity, point the boundaries for the neighborhood so that the users can distinguish the neighborhoods. We couldn't make it any better further as there were too many listings and we tried to avoid the overlapping and clutter as much as possible.
- Our users also mentioned from the initial dashboard to give clarity on what our intention and definition of attribute are. Hence, we had a heading attribute and pushed the legend from the right pane to the left to display the text on the right. This improves the readability of the dashboard and helps readers avoid

the feeling that a lot of information is being thrown at them at once, which one of our users expressed that concern.

- We also chose the Airbnb color palette for our dashboard as recommended from the notes of the usability test. We used a sequential color scheme for our treemap. Our map uses two colors to indicate properties that were high/low in price, availability or review score.

FINAL VISUALIZATION

Insights And Rationale

- **Manhattan has the most number of listings of all New York City Boroughs**

  Manhattan is the most popular borough of New York City, when it comes to tourists as well as the sheer number of things to do. We believe that owing to these reasons, there are more listings in Manhattan than in any other boroughs.

- **Manhattan has the highest average price per night for listings compared to other boroughs**

  The property rates in Manhattan are significantly higher than other boroughs in New York City. This is likely why it has translated into the higher nightly tariffs for AirBnBs.

- **Brooklyn has the least availability of listings**

  Brooklyn has the second highest number of listings among New York City boroughs, however, it is cheaper than Manhattan when factoring in average nightly price for the listings. It is also extremely well connected and the closest borough to Manhattan, therefore, making it a popular choice for AirBnB listings and bookings.

- **Brooklyn has listings with the highest number of reviews**

  The older guests would prefer staying in Manhattan, which would be closer to most attractions. Younger guests would prefer to save money and would not mind adding a slight commute to see attractions, and hence would presumably prefer brooklyn compared to older guests. Borrowing from the statistic that younger patrons are more likely to leave reviews than older patrons, we believe that is why Brooklyn has the listings with the highest number of reviews.

- **The Bronx has the cheapest Airbnb listings.**

  The Bronx has the lowest median income out of the 5 New York City boroughs, as well as the highest incidents of crime per capita of the boroughs. Furthermore, it is farthest from most of the New York City attractions and therefore we infer that it has the cheapest AirBnB listings.

CHALLENGES

The following are the challenges that our team encountered during this project:

- **Usability Test:** One of the major challenges that our team encountered during this project is going through multiple rounds of usability testing. We initially had to create violin and forecasting models on python as they were relatively easy using the library to get the feel of the audience. When we realized they weren't intuitive we had to pivot to a different design of the dashboard and perform the usability test again.
- **Dashboard:** A major challenge for the dashboard that our team encountered was linking the dynamic attributes to the visualizations. For instance, the user selecting price or review attribute showed a very different visualization for the area chart and the box plot, thereby making it difficult to ascertain the appropriate aspect and contrast ratio for the visualizations so that the information provided is meaningful and visually homogeneous. Furthermore, there were also a lot of fields that we deployed in the dashboard that were computed using advanced calculations, and so we needed to create several sanity checks for the same within our dashboard, such as making sure the values in the heatmap and the map were displayed and grouped appropriately.

RECOMMENDATIONS AND FUTURE SCOPE

While we had an exhaustive dataset about the Airbnb listings within the city of New York, the following steps can be incorporated into our future work and scope of the dashboard.

- Add more visualizations pertaining to the review scores of different types and metrics. Maybe a heat map can help.
- We can further incorporate and find exhaustive listings of the other major cities in the United States. We can subsequently create a mega home dashboard where the user gets to select the city from the dropdown. The list contains the names of the cities that Airbnb publishes the data for.

- We can also attach an extension or like a link to the external site which has the prediction of the listings and future forecast as time series computed in python. This will give the information about the likelihood of prices in the next couple of weeks or months more accurately computed perfectly through robust machine learning algorithms.
- Airbnb can populate the datasets on a regular basis to keep dashboards such as our current one to help the airbnb users to get an insight into the pricing and subsequently get an idea when they travel in the future.

CONCLUSION

The dashboard we created can be used by both Airbnb users and hosts. It can also be used by data analysts who want to explore more on the Airbnb rent market. The dashboard has the ability to allow users to explore the data themselves and also provides them with a high level summary based on price, availability or review score.

Usability testing provided us insights on improvements needed and we implemented them in our final visualization.

REFERENCES

Coles, P. A., Egesdal, M., Ellen, I. G., Li, X., & Sundararajan, A. (2017, October 12).

Airbnb Usage Across New York City Neighborhoods: Geographic Patterns and

Regulatory Implications. SSRN.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3048397

Inside Airbnb. Adding data to the debate. (n.d.). Inside Airbnb.

http://insideairbnb.com/get-the-data.html

APPENDIX

Usability Test Script

Subject Demographics

1. What is your name?

2. What do you identify yourself as (in terms of gender)?

3. What is your current Age?

General Questions

1. How well are you aware of AirBnB?

2. How often do you travel? Where do you prefer staying during your travels?

3. Have you ever lived in an AirBnb?

4. What are your priorities when you search for an Airbnb?

5. Do you think your priorities change when you search for an AirBnB in a metropolitan city compared to other places.

6. Have you ever been to New York city? If so, how many times?

7. Did you ever stay in an AirBnB when you visited New York city?

8. Do you recall any search experience of finding an AirBnB particularly in New NewYork? What were the pain points if you recall any?

9. Do you prefer having an interactive dashboard to help you find you an AirBnB?

10. Are you aware of tableau interactive visualizations?

Visualization Specific Questions

1. First thoughts about the dashboard?

2. Were you able to navigate through the dashboard by yourself? How much help did you think you needed to navigate the dashboard?

3. What details or artifacts about the dashboard draw your attention in the first look?

4. What are your thoughts about color coding?

5. Are you able to navigate through the filters on the left pane and correlate the visualizations ?

6. Give your overall comments/ suggestions about this dashboard?

7. Which borough had the highest Airbnb sales in New York city?

8. Which borough has the highest availability within New York city?

9. Can you briefly tell from the dashboard which neighborhood has the expensive listings?

10. What are your first impressions or thoughts about the legend and the pane on the right and left of the visualizations? How interactive and useful are they to you in terms of gathering information from the visualizations and subsequently making insights out of them?

11. What additional attributes do you suggest we add to the visualization to make more sense out of them?

12. Are there any aspects or visualizations within the dashboard that aren't clear or confusing?

13. Any other feedback that you would like to give our team in improving the experience of the dashboard?