# Part One – Draft a Data-Collection Plan

Step 1: Identifying the Situation and Decision:

Airbnb has seen a meteoric growth since its inception in 2008 with the number of rentals listed on its website growing exponentially each year. Airbnb has successfully disrupted the traditional hospitality industry as more and more travelers, not just the ones who are looking for a bang for their buck but also business travelers’ resort to Airbnb as their premier accommodation provider.

New York City has been one of the hottest markets for Airbnb, with over 52,000 listings as of November 2018. This means there are over 40 homes being rented out per square km. in NYC on Airbnb! One can perhaps attribute the success of Airbnb in NYC to the high rates charged by the hotels, which are primarily driven by the exorbitant rental prices in the city.

Airbnb and Airbnb customers are the stakeholders are involved in this data collection plan.

So, I want to perform exploratory data analysis to understand the rental landscape in NYC so that I can make right decision as a business analyst of some company, where to start Hotel in NYC registered with Airbnb as a host to get best revenue or sustainability of the profit for next 5 years after starting the business.

Step 2: Identifying Data for Better Understanding the Situation:

* **Key Performance Indicators:** There five price-determinant categories of Airbnb listings, including host attributes, site and property attributes, amenities and services, rental rules, and online review ratings.
* **Range of Cases to Consider:** I will consider the listing of hotels with listing id for different hosts with host id as cases.
* **Variables:** 
  + name – name of the listing (text)
  + neighborhood – area (location)
  + latitude and longitude – co-ordinate
  + room\_type – listing space type (categorical)
  + price – price in dollars (quantitative)
  + minimum nights – number of nights (quantitative)
  + number\_of\_reviews – Number of reviews (quantitative)
  + last review – latest review (date)

Step 3: Develop a data-gathering plan:

I can collect the data from Airbnb website those are freely available. And all these data are observational data. As of August 2019, this data set contains almost 50 thousand Airbnb listings in NYC.

This data containing different hotels listing from different neighborhood of NYC. I need to resample the data with different neighborhood to ensure that sample is a representative of the population.

I can group the neighborhoods to mitigate potential bias within the sample.

Preview of the Airbnb Sample Data:



Remarks and Comments:

I’ve attached the full excel dataset with submission files.

Instructor Comment or Feedback:

# Part Two – Identify Data Summaries and Visualizations

To inform my decision, I have used neighborhood which is a categorical column and price column which is quantitative. I’ve calculated the quartiles to get summary statistics for different neighborhood groups. Below is the summary statistics for different neighborhood groups:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Summary Stats** | **Bronx** | **Brooklyn** | **Manhattan** | **Queens** | **Staten Island** |
| Min | 0 | 0 | 0 | 10 | 13 |
| Q1 | 45 | 60 | 95 | 50 | 50 |
| Median | 65 | 90 | 150 | 75 | 75 |
| Q3 | 99 | 90 | 220 | 110 | 110 |
| Max | 2500 | 10000 | 10000 | 10000 | 5000 |

From the above summary statistics, we can see that **Brooklyn, Manhattan and Queens** has a spread, big jump for max but quartiles are very less compared to the max values. So, there are outliers those are affecting the statistics. So, we need to be cautious before going for these neighborhoods.

On the other hand, **Staten Island and Bronx** more consistent performers than other neighborhoods.

Also, Staten Island and Manhattan have a very good Q3 quartile.

To reach the final decision, I’ve taken all the neighborhoods, calculated the total price of a specific neighborhood and then plotted bar chart to get the top 10 neighborhoods in the New York City. Below is the chart that I got:

|  |  |
| --- | --- |
| **Neighborhood** | **Price** |
| Williamsburg Total | 563707 |
| Midtown Total | 436801 |
| Upper West Side Total | 415720 |
| Hell's Kitchen Total | 400987 |
| Bedford-Stuyvesant Total | 399917 |
| East Village Total | 344812 |
| Upper East Side Total | 339729 |
| Harlem Total | 316233 |
| Chelsea Total | 277959 |

To take the best decision, we can check which neighborhood comes under which neighborhood group. That will help us to get the best profit with consistency.