# Business Problem

Sample Question 2: Investigate the neighborhoods that have higher or lower Airbnb rental activity. Are there any notable demographic patterns that emerge between such neighborhoods?

# Business Impact

*[It is useful if you can provide an estimated impact of your target/moonshot analysis. If things go exactly as planned, how much does your company/business beneﬁt? If things go much better than planned, how much does your company/business beneﬁt?]*

This will be important for the business as the location analysis will help the business to determine where to place Airbnbs in the future. This will make a very big difference on the amount of rent that can be charged, thereby increasing profitability. If it is in a neighbourhood where the property value goes up then the business value will also increase. It also has a bearing on reducing possible administrative problems such as renters who do not pay rent and building vandalism.

# Data

*[Tell us (at least approximately) how large your dataset is. Highlight both the strengths and weaknesses of your dataset(s).]*

*Strengths*

1. The data set gives reviews per location of business
2. The data set gives the location of each business
3. The dataset gives key features related to the variable of interest, which is the location (providing log and lat information)
4. The data allows us to split the information up and get a more comprehensive pattern of the information

*Weaknesses*

1. *The data set is not complete as there are some missing values*
2. *The data may be too big and we only need a small subset of the information*

# Methods

## **Visualizations**

*[Remember that exploratory data analysis (appropriate visualizations) is the ﬁrst, crucial step in modeling! So you should be describing the EDA that you will be performing on your datasets, not just the "models" you will apply.]*

*The exploratory data analysis that we will be performing on our dataset is first to clean the data, and select our variables. We will remove outliers, null values, check for missing data, identify the shape of the data etc. We can then go on to do a scatter plot , box plot, heat map and correlation matrix to visualize the information well and identify significant correlations.*

## **Models**

*[If you are going to explore a model, please provide a brief description why you think it is useful for your particular project. You must demonstrate you have a high-level understanding of the method.]*

None so far

# Milestones

**Version 1**: Build simple dashboard …

**Version 1**-build a simple dashboard with simple visualizations such as a histogram and box plot

**Version 2**: Build prediction model …

**Version 2**-build more complicated models such as a correlation matrix

**Version 3**: Build XX …

**Version 3**-review visualizations and any negative correlations. Find more interactions if possible.

**Version 4**: Use additional data from … to build YY

**Version 4**-review version 3 and add in anything that may have been left out in our analysis

*[Your Version 1 must be achieved with 100% probability. Notice your Version 4 should be very difﬁcult for you! We want to see you really challenge yourself in what the team could achieve. You should have minimum 3 versions and maximum 5 versions for your project. If you are a larger team with a more complicated project, it may be helpful to have more versions to stay organized.]*

*[You do not have to color-code like this, but your project milestones should ideally contain clear milestones on data, analysis, and visualizations.]*

# Timeline

|  |  |  |
| --- | --- | --- |
| **Date** | **Name** | **Details** |
| **Week 1 (Sep 6 – Sep 12)** | Team Formation | We formed our team and established a Watssap group for communication. |
| **Week 2** | Work on idea formation | We decided to focus on the Airbnb dataset. |
| **Week 3** | Idea should be ﬁnalized | We decide don where our strengths were as team members and discussed our backgrounds so that we could each focus on one part of the assignment that would be best suited. |
| **Week 4** | Datasets sourced | Data sets sourced. |
| **Week 5** | Basic EDA | We decide on the primary and secondary questions we wanted to analyze (time permitting) |
| **Week 6 (Oct 11 – Oct 17)** | FALL READING WEEK | We divided up the work into parts. |
| **Week 7** | Basic EDA/ Writing of midterm report | We began writing the report. |
| **Week 8** | Basic EDA / Deliver midterm report | We reviewed and submitted the report. |
| **Week 9** | In-depth EDA |  |
| **Week 10** | In-depth EDA /  Statistical analysis |  |
| **Week 11** | In-depth EDA /  Statistical analysis |  |
| **Week 12** | Statistical analysis / Machine Learning modeling |  |
| **Week 12** | Statistical analysis / Machine Learning modeling |  |
| **Week 14 (Oct 11 – Oct 17)** | Writing of final report / presentation video rehearsal |  |
| **Week 15 (final exams weeks)** | Final report and presentation video |  |

# Concerns

*[This section should be longer than this and provide more detail, such as what will you do to address the concerns you have.]*

To address some of the concerns we have we will split up the data and work in teams to analyze the relationships between variables.