**IST 652: Scripting for Data Analysis**

**Final Project Proposal**

1. **Team members**

Anupama Mohan Rao

Abhimanyu Pratap Singh Parmar

Sanyukta Shandilya

1. **Topic of investigation:** Airbnb Dataset
2. **Dataset that we are planning to use**

We have decided to use **Seattle Airbnb Open Data** sourced from Kaggle.com for our project. This dataset describes the listing activity of homestays in Seattle, WA and is originally sourced from publicly available information from the Airbnb site.

**Description of the dataset**

This dataset captures Airbnb activity in 102 columns distributed across 3 CSV files, Listings, Reviews and Calendar. Listings contains full description and review score – the major information for analysis. Reviews contains review information including unique id for each reviewer and detailed comments. Calendar contains date and price information including listing id and the price and availability for that day.

1. **Methods of data acquisition and analysis**

Following are the major steps:

Reading & storing the data

Data cleaning

Data querying

Analytics & Modelling

* + **Data Cleaning** to remove all the NULL values and preparing the data for analysis
  + **Data Querying** to understand the data
  + **Exploratory Analysis** to answer business questions

**Libraries:**

* + Numpy
  + Pandas
  + Matplotlib
  + seaborn

1. **Potential Development Tasks**

The purpose of the project is to understand vibe of Seattle neighbourhood listing descriptions, price spikes and general trend in Airbnb listings including:

* Property types and their pricings.
* Effect of time of the year on booking prices throughout the year.
* Most expensive and least expensive neighbourhoods in the city.
* Factors that affect the predicting the price of a listing.