**Madagaskar – Group Project Proposal**

**Barış Sivas**

**Neriman Gürsoy**

**Ozan Barış Baykan**

**Sena Kalemci**

**Tuğba Ünal**

1. **The source and details of the dataset**

Source link: <https://www.kaggle.com/ankit1661/milan-airbnb-dataset-2020>

Milan Airbnb Dataset 2020: This dataset contains 7 tables, but we will use 3 tables and 87 columns which are calender.csv, listings.csv, reviews.csv. These 3 tables consist of approximately 500 MB.

1. **A proposition on what, why and how to work with the data**

Listings.csv is the main table, reiews.csv and calender.csv tables will be used as assistant. We think that Airbnb is one of the most suitable dataset for making data analysis, because of having not only numerical but also categorical variables. We would like to use these variables to make a relation between usage experience and the ratings of venues. For example, demand statistics of type of room(half&full), usage frequency of locations, etc.

1. **A simple outline of the report**

- Selection of database (Milan Airbnb Dataset 2020)

- Create databases

- Preprocessing of the data

- Exploratory Data Analysis

- Analyze

- Summarize or visualize