TEAM 3 – PROJECT 2 PROPOSAL

1. TEAM 3 – Lili, Emily and Meme
2. DATASET IDENTIFIED is Airbnb-Amsterdam, data set contains 7 tables in .csv format <https://www.kaggle.com/erikbruin/airbnb-amsterdam?select=reviews_details.csv>

With Pandas, we are going to extract the data, transform them, and them load back into data frames.

* **Extraction**
  + We will be using Juptyer Notebook to put each CSV into a Pandas Data-Frame.
* **Transform**
  + Copying only the columns needed into a new DataFrame.
  + We will rename columns to fit the tables created in the database.
  + Handle any duplicates.
  + Set index to the previously created primary key.
* **Load**
  + We will be using Mongo to create a connection to database, Mongo will help us convert the raw unstructured data in to a more simplified visualization
  + Checking for a successful connection to the database and confirm that the tables have been created.
  + We Append Data-Frames to tables, assuring to use the index set earlier.

Final report will show :

1. What is the volume of Rent between Seasons?
2. What is the average cost per night?
3. Is there a Trend in Location?
4. What are the average reviews