3. Current Progress

The overall progress of our project is that we have designed our frontend, data processing, and tested one of the algorithms that we would try.

3.1 Data

The data used for this project are divided into two tables: `apartments` table and `users` table. `users` table is obtained through meetup.com API, `apartments` table is found on course website datasets with Airbnb apartment data.

`Users` table stores user information. New user information is appended to the table, and when recommending roommates, algorithm is applied to data in this table in search of optimal candidate.

* no: Primary Key
* firstName: User entered first name
* lastName: User entered last name
* uni: User entered UNI, random generated UNI is assigned to basic data with name initials + 4 random digits
* gender: User gender
* nationality: See countries.txt
* email: User entered email address
* school: See schools.txt
* major: See majors.txt
* smoking: Binary, yes or no
* alcohol: Binary, yes or no
* habit: Binary, early bird or night owl
* roommate: Binary, yes or no

`Apartments` table stores New York Airbnb apartment information. Algorithm is applied to data in this table for finding apartments that satisfies user preference.

* id: Primary Key
* name: Name of apartment
* neighborhood: Apartment neighborhood
* latitude: Latitude
* longitude: Longitude
* distanceLine: Exact distance from Columbia University calculated using longitude and latitude
* distance: Enumerator. <1 mile = near, 1-3 miles = pretty near, 3-7 miles = pretty far, >7 miles = far
* roomType: Enumerator. Private room, entire home/apartment, shared room
* price: Enumerator. <$100 = cheap, $100-$200 = economical, $200-$400 = luxurious, >$400 = expensive
* reviewsPerMonth: Reviews that the apartment gets every month. Used in recommendation algorithm

3.2 Methods

The frontend is deployed with Django, with GBQ interactions embedded in the frontend code. Users can enter their preferences for apartments and roommates using this interactive frontend.

To begin with, there is a base user information database, where existing user information are already. New user would use the frontend website to enter their information and preferences, then this information is appended onto the `users` table in Google Big Query upon user information submission. Meanwhile, recommendation algorithm run on both `users` table for finding optimal roommates, and `apartments` table for recommending most suitable apartments. Finally, the output is displayed on the frontend in the `result` page recommending couple of roommates as well as apartments.

Recommendation algorithms used in this project include KNN, K-means, linear regression. At this point, we have only implemented linear regression model and have not yet compared accuracy between different models. Factors considered for roommate recommendation algorithm include: school, major, gender, alcohol, smoking, habit, nationality . Factors considered for apartment recommendation algorithm include: distance, price, neighborhood, roomType, reviewsPerMonth.

3.3 System

TODO: Put system architecture here.

4. Planned Experiments