**Data Visualization QMSS 5063**

**Final Project Proposal**

**Team Member:**

Zian He [zh2411@columbia.edu](mailto:zh2411@columbia.edu)

Xiaoshu Xu [xx2337@columbia.edu](mailto:xx2337@columbia.edu)

Chuze Zhang [cz2581@columbia.edu](mailto:cz2581@columbia.edu)

Yingtong Zhou [yz3667@columbia.edu](mailto:yz3667@columbia.edu)

**Columbia #1 Rental Site - realrentercolumbia.com**

**Abstract:**

The Big Apple city, one of the most powerful and influential cities, is also the biggest hub and melting pot for people all over the world. However, within millions of people who come to NYC to work, study or sightsee, not a lot of them own their own house/apartment. Over 60% of the apartments in NYC are houses for rent. These rental apartments in NYC, then, offer these people a place to stay while they endeavor to achieve their American dreams. For our study, we would like to focus on rental apartments in NYC. We will closely examine several different aspects of these apartments (more detailed information below), and hopefully, our final project can help future renters to get a clear picture of the overall rental market in NYC and select the one that suits them the most with ease.

We would like to study following aspects related to the rental market:

* **Location, location, location**: How does the rental price vary across different districts? We are probably going to explore the mediating effect of access to public transportation. How does the proximity to a subway station, for example, influence the rental price? At the same time, how does the proximity to a university, like Columbia, affect the rental price? The neighborhood also makes great difference to the rental price, like whether there are restaurants, bars, grocery stores, and gyms nearby? We suspect that changes in infrastructure and the addition of new apartment buildings or offices nearby would have impacts on the rental market.
* **Characteristics of the building**: How does the rental price of different room types (studio, one-bedroom, two-bedroom, etc.) change and how do other attributes like the number of bathrooms influence the rental price? We also consider characteristics like when is the building built, whether it is recently renovated, and whether there is doorman, elevator or in-unit washer/dryer; such amenities would also influence the rental price.
* **Time**: How does average rental price change over time in different districts? We are wondering if certain districts (like the Long Island City) are gaining more popularity in the apartment rental market and whether living close to a university like Columbia is becoming less affordable.
* **Characteristics of renters**: We are also interested in the NYC renters’ profile. What is the average age of renters in different districts? Are certain districts more populated by younger people? Are some districts more racially segregated than others? How about the median or average income of renters in different districts, the racial composition, their average education level, and average commute time to work, etc.? More interestingly, how do these characteristics change over time?
* **Building reviews**: We are also interested in analyzing renters’ ratings and reviews of the buildings. How do ratings vary across districts? And if there is a relationship between rental price and ratings and reviews?
* **Techniques**: We are going to use R packages like ggplot2 and maps, as well as other spatial analysis and text analysis techniques.

**Data Description:**

* Main Dataset: <https://streeteasy.com/blog/methodology-streeteasy-rent-indices/>
  + This dataset was built up by the StreetEasy Research team with the goal of tracking how sale prices and asking rents change over time.
  + It covers changes in monthly rents in Manhattan from 2007 to present, and changes in monthly rents in Brooklyn and Queens from 2010 to present.
  + The rent was estimated at the borough-wide level as well as for geographic submarkets and price tiers within each borough.
* Zillow API: <https://www.zillow.com/howto/api/APIOverview.htm>
  + This API can be used to check the reviews and ratings for real estate agents such that customers can select the best agent for their particular set of needs
  + We also get the business address/name associated with each agency so that we can pinpoint each real estate agency on the map together with their ratings.
* Subway Station:<https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49>
  + We would also like to explore the mediating effect of access to public transportation.
  + This dataset provides detailed geographic information about each subway station in the Manhattan, Brooklyn and Queens area.
* Apartment Ratings: <https://www.apartmentratings.com/ny/new-york/>
  + This website contains lots of ratings and reviews from current or previous residents. It also includes the address, amenities, as well as other information of the buildings.

**Visualizations:**

* Map plot: The main part of our dashboard is a map plot. We want to draw a heat map or a 3-D map plot to give an overview of the rental market in NYC by the district. Including basic characteristics like total rental inventory, average monthly rent, median asking price, etc.
* Key statistics: We would like to investigate how these characteristics change over time (by year, by region). We also want to form some detailed comparisons between districts/boroughs. For each district we would also want to display the rental price change over time, the total rental amount and some other statistics. We would apply various kinds of plot to display our statistics.
* Interactive: We would also make our plot interactive, so that we can analyze according to specific questions. For Columbia students, we can check out some off-campus housing options. Within a specific region such as the upper west side, we can check for some interesting characteristics like rental price, apartment rating, mediating effect of public facilities, renter characteristics, etc. without too much sweating. We can also make other choices by using the interactive plot from the dashboard.