

Of Troubles and Travels

Overview

Within this project, we explored the impact of Columbia students on the Morningside Heights community and also analyzed their top go-to destinations in New York. This data could not only be useful for potential advertisers who could understand the wants and needs of Columbia students, but also car-hailing app companies such as Lyft and Uber who could optimize their driver positions near Columbia.

Data Analysis

We first analyzed the 311 complaints data and focused our results on three districts: Columbia, NYU, and the Upper East Side. We found that the top complaints near Columbia revolved around noise complaints and utilities, which is very expected of a college neighborhood. NYU had the same noise and loud party music complaints, in addition to other similar complaints such as street condition and homeless people. To isolate whether or not Columbia and NYU college students caused such loud party music complaints, we assessed the top complaints near the Upper East Side, which boasts a demographic with very different lifestyles and sleep times. UES had top complaints for street condition and driver parking violations, rather than the loud party music that defined NYC colleges.

Second we analyzed the movement of Columbia students by tracking the drop off locations for taxis that picked people up near Columbia. We recognized that many students would take the subway or an Uber instead, but assumed that the taxi data would still correlate with the movement of the population. Most of the drop offs were concentrated in the Upper West Side, so we focused on drop offs below Columbus Circle 59th Street. Given Columbia student's penchant for finance, we were not surprised to find a clear hub of drop offs near Goldman Sachs and Citibank. Goldman had by far more drop-offs than Citi, but we could not assess the popularity of the other major investment banks since they were concentrated in Midtown. We then split the data into time periods, particularly the 4am-10am morning rush and the 10pm-4am evening out. We found in the morning, drop offs to Goldman, Citi, Hudson Yards, Penn Station, and Park Avenue were very popular, which correlated with our

hypothesis. At night, we found many students going to Chelsea Market, Ktown, and East Village, which hit on a lot of popular bars and clubs. This could be especially useful for nightlife bars and clubs who are looking to advertise to Columbia students by first gauging which destinations were most popular. Also an analysis of rides requested during specific times could be of help to ride-hailing app companies to optimize their driver locations during peak hours. Overall the taxi data confirmed our daily observations of Columbia student's professional and social lives.

Data Cleaning

We use Google's BigQuery platform to access and download data using SQL commands. For some reason, when we were first exploring the 311 complaints data, all the descriptors were not NA, but then several days later when we run the same SQL command again, some descriptors of complaints become NA, which posed some challenges on data cleaning.

The taxi data set is huge. We got about 300000 trips of only one year of yellow cab taxi with pick_up_lat and pick_up_lon in Columbia area. We first tried to use the hexbin feature of basemap to draw the plot, but it took so long and the graph does not overlay on Google map. We then switched to Google heat map, but it does not show properly because there were some '0' data points in the data. Even if we remove those data points, the plot still did not look as we wanted it to. We then restricted the drop_off_lat and drop_off_lon to be in the Manhattan/NYC range, and ran the query again. This time it worked fine and helped us to explore the secret life of Columbia students.

We did not manage to present our result by flask but instead used a website. We just use iframe feature in html to show Google heat map.