Analyze Travel Data to Reduce Turnstile Jumping: -Reynold D’silva

Abstract:

The project aims to resolve the issue of Turnstiles Jumping on NY transit, operated by MTA (my client) by analyzing commuting data and provide insights on the travel activity of individual Subway Stations. I worked with datasets provided by MTA and City of New York. My findings will be further refined by MTA employees, by monitoring the suggested Subway Stations for 7 days via CCTV cameras. Immediate course of action can be to increase Turnstile height or allocating Subway security to the specified locations. MTA has been bankrupt for over two decades and alleviating loss of revenue can be a blessing in disguise for the management. Commuters breaking the law can’t be arrested due to low severity of charge. Turnstile jumping is a multi-faceted issue- lost cost housing houses people that earn less than $50,000 annually (Eligibility criteria for a low-cost residence). These families can’t afford over $500 per month to travel and thereby are compelled to jump turnstiles. City officials should design plans that suit commuters in accordance with their income brackets. As per MTA, they lose $962,500 every day due to Turnstile Jumping. Also, there have been 4093 cases of commuters tripping over turnstiles and hurting themselves, which can be avoided with commute passes designed for low-income residents.

Design:

I’m going to use the MTA dataset to compare the Entries and Exits at all subway stations. The dataset provides data on the entries and exits at most subway station in New York City. Exiting a subway station is free so a station with more exits than entries in a residential area (Tourist locations can possibly distort analysis) can be a valid starting point for this analysis. Also, Unique IDs for turnstiles are provided on the dataset, which can be used to finetune the location pin on the subway station.

Data:

The MTA sample dataset includes more than 2.5 million data points with 9 different categories. Datapoints include station data, Unique IDs of turnstiles and commuting data of over five years. Some of the drawbacks of the data are that data is collected in reverse and turnstiles are reset at different cycles. Certain assumptions have been made to avoid analytical bias.

Algorithms:

A plethora of code in Python has been utilized to refine data and visualize cleaned information.

Tools:

Libraries such as Pandas and NumPy for Data Manipulation

Matplotlib and Tableau for Data Visualization

Communication:

All the above information will be communicated in detail through presentation under the category “Mental Emancipation” on my website www.reynolddsilva.com