EXPLORATORY DATA ANALYSIS: MTA TURNSTILE PROJECT

Presented by: Kalu Uga
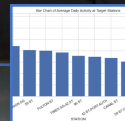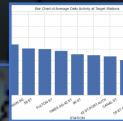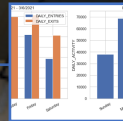
OBJECTIVES

- Problem Statement
- Data Source
- EDA Workflow
- Ridership Pattern
- Anomaly in Ridership Pattern
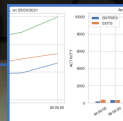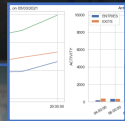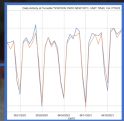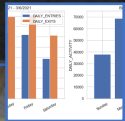- Daily Activity at Busiest Station
- 4 HR Interval Analysis
- Target Stations
- Interactive Map of Target Stations
- Recommendations

PROBLEM STATEMENT

> **Design a model to efficiently gather email addresses of target New York City subway users for WTWY upcoming gala.**

Assumptions:

- Women Tech Women Yes (WTWY) is interested in funding

- Residents in zip codes where median household income is greater than $90,000.00 are more likely to make donations at the gala.

- Students can be targeted better by going directly to colleges and universities.

- A subway ACTIVITY is defined by the sum of ENTRIES and EXITS for the reference frame.

- At subway stations where daily EXITS is within (+/- 2000) of daily ENTRIES, we can assume that the riders are frequent riders.

DATA SOURCE

- MTA Dataset

  http://web.mta.info/developers/data/nyct/turnstile/

- 2019 NYC Median Income.

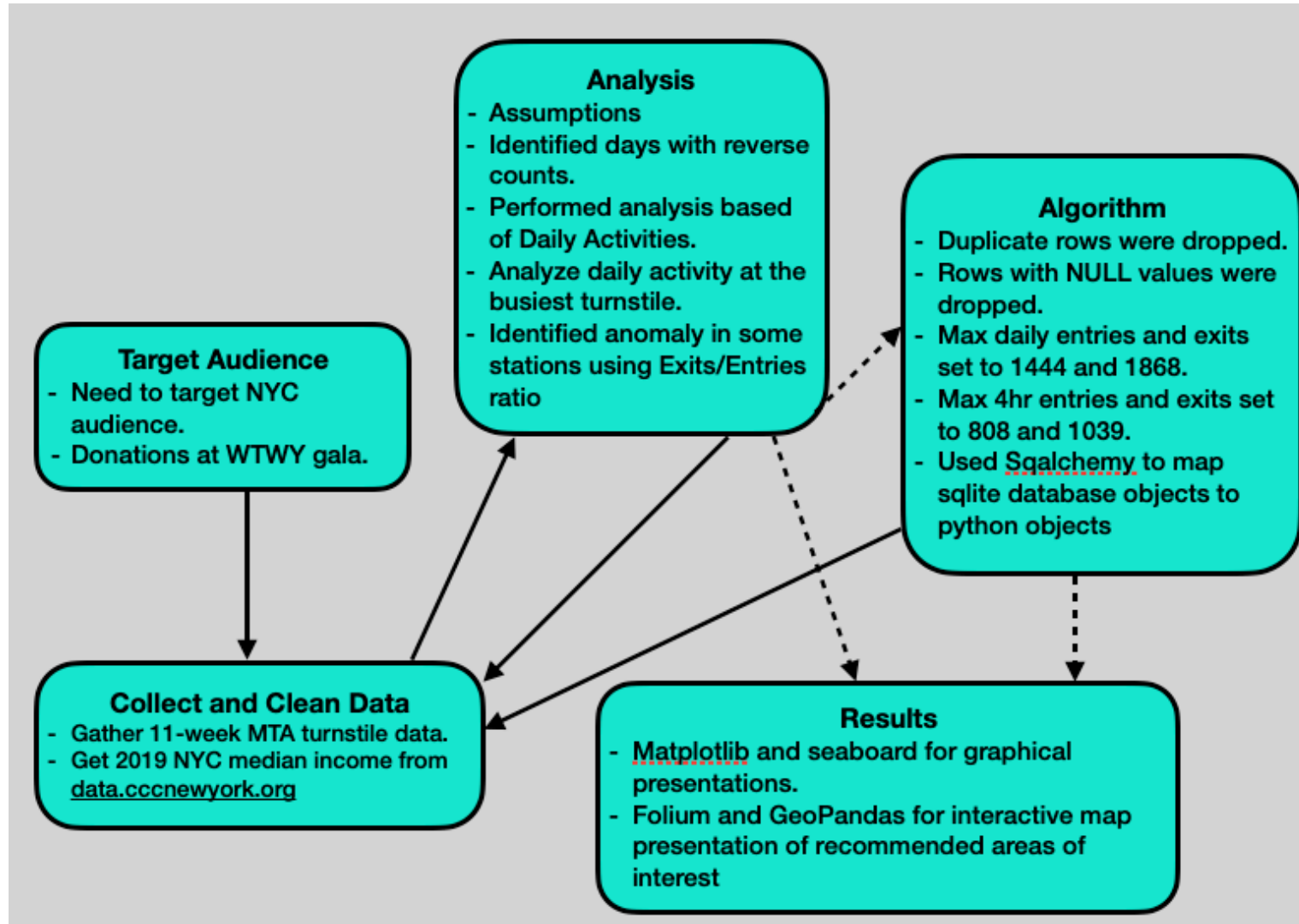 https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a

- NYC GeoJson Cartographic File.

  https://github.com/fedhere/PUI2015_EC/blob/master/mam1612_EC/nyc-zip-code-tabulation-areas-polygons.geojson
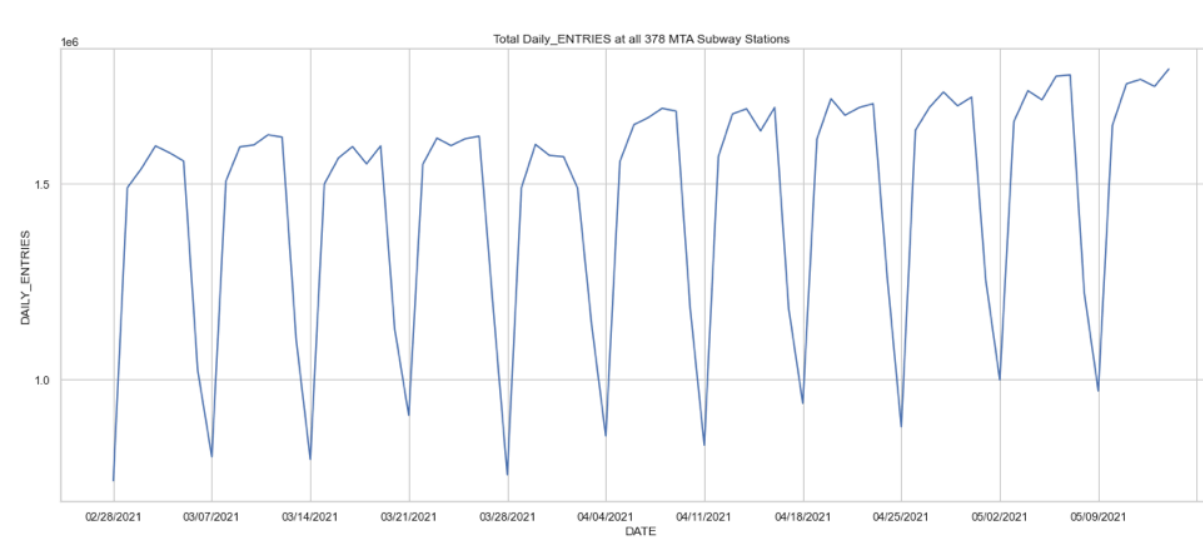
- Interactive Choropleth with Python and Folium

  https://vverde.github.io/blob/interactivechoropleth.html
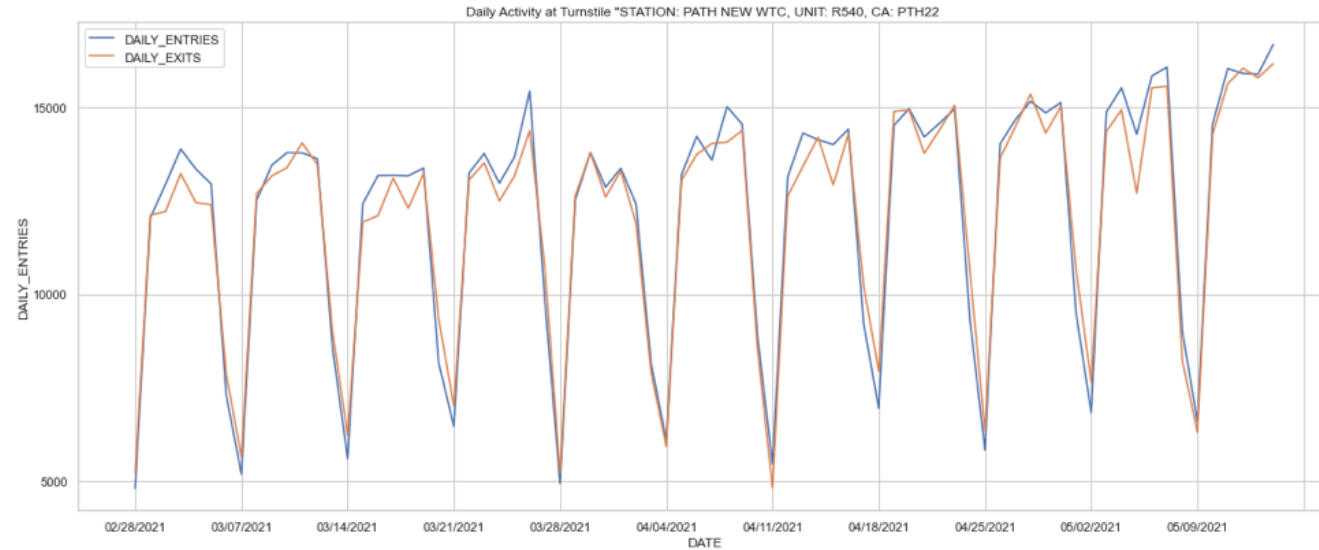
EDA WORKFLOW

RIDERSHIP PATTERN FROM 2/28/2021 TO 5/14/2021

- Increasing trend in subway ridership

- Daily pattern for all 378 stations

- Daily pattern for the busiest turnstile.

Total Daily_ENTRIES at all 378 MTA Subway Stations

Daily Activity at Turnstile "STATION: PATH NEW WTC, UNIT: R540, CA: PTH22

ANOMALY IN RIDERSHIP PATTERN

- Searched for stations where EXITS < 0.25*ENTRIES
- NEWARK BM BW is a PATH train terminal. Looks like payment via METRO CARD is accepted but EXITS are logged mainly by PATH.
- Pattern also fits tourist locations. Tourists get to location by car and leave via the subway.

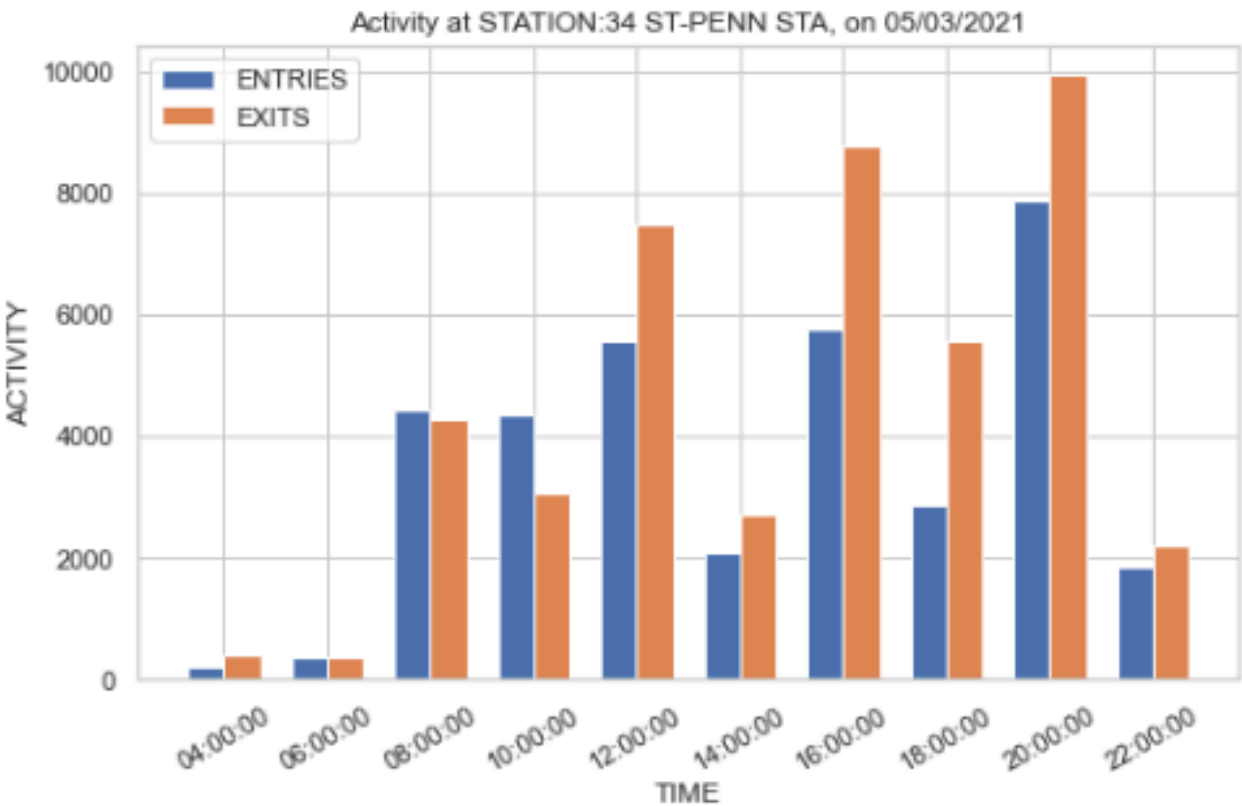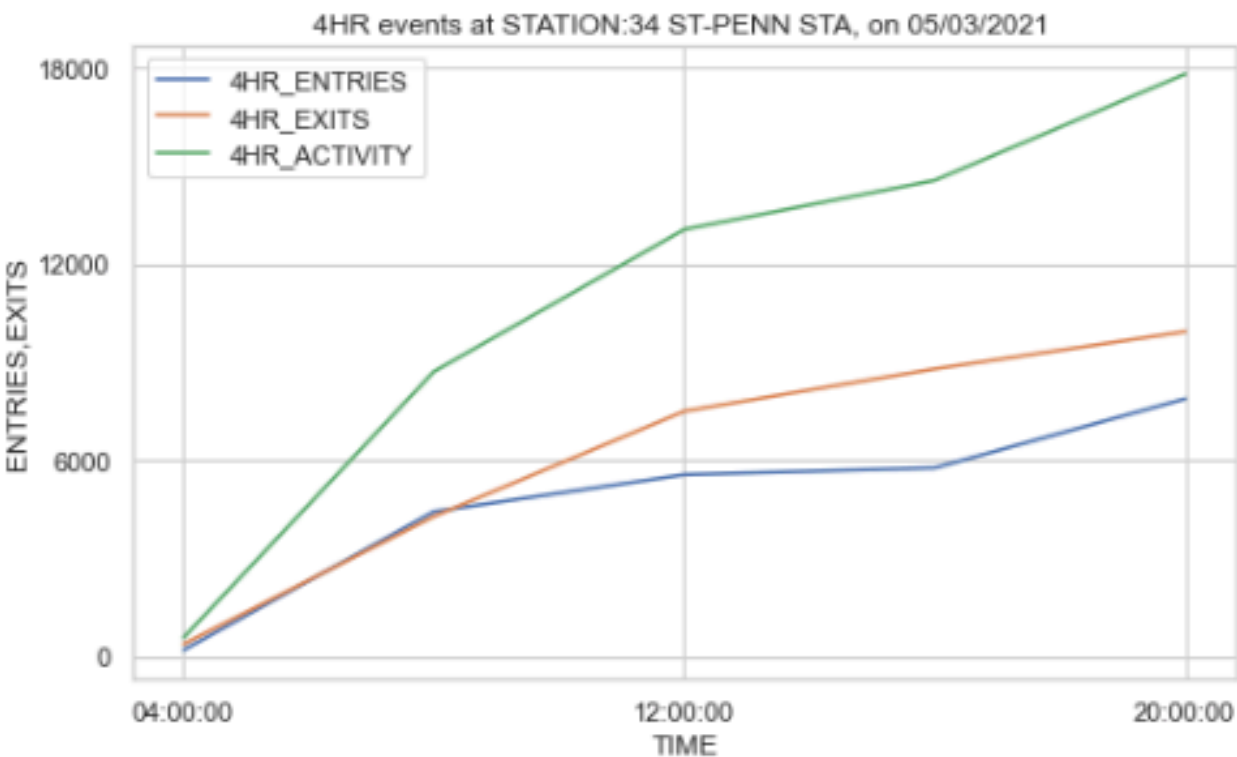Daily Anomaly for "STATION: NEWARK BM BW
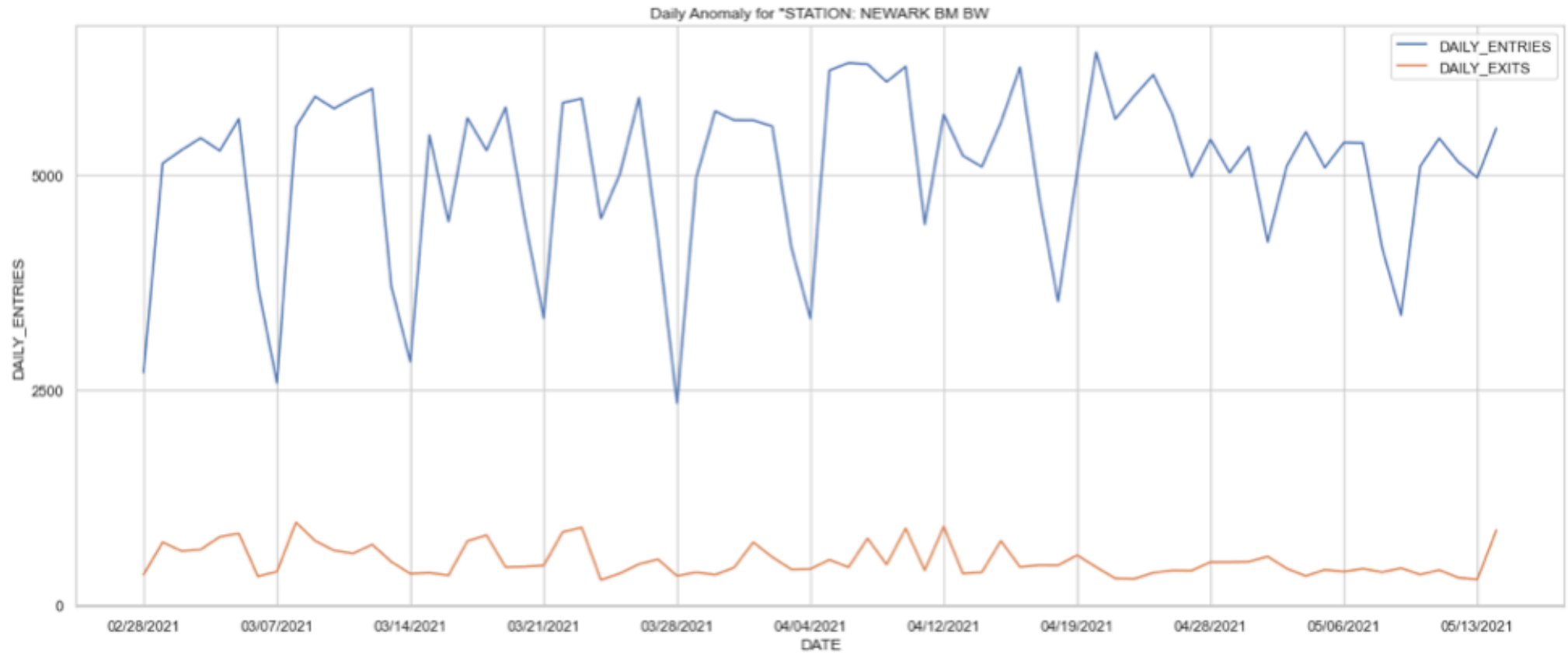
DAILY ACTIVITY AT BUSIEST STATION

- Weekdays are generally busy.
- Weekends are less busy. A major sports event can cause a weekend to see abnormal spike in activity
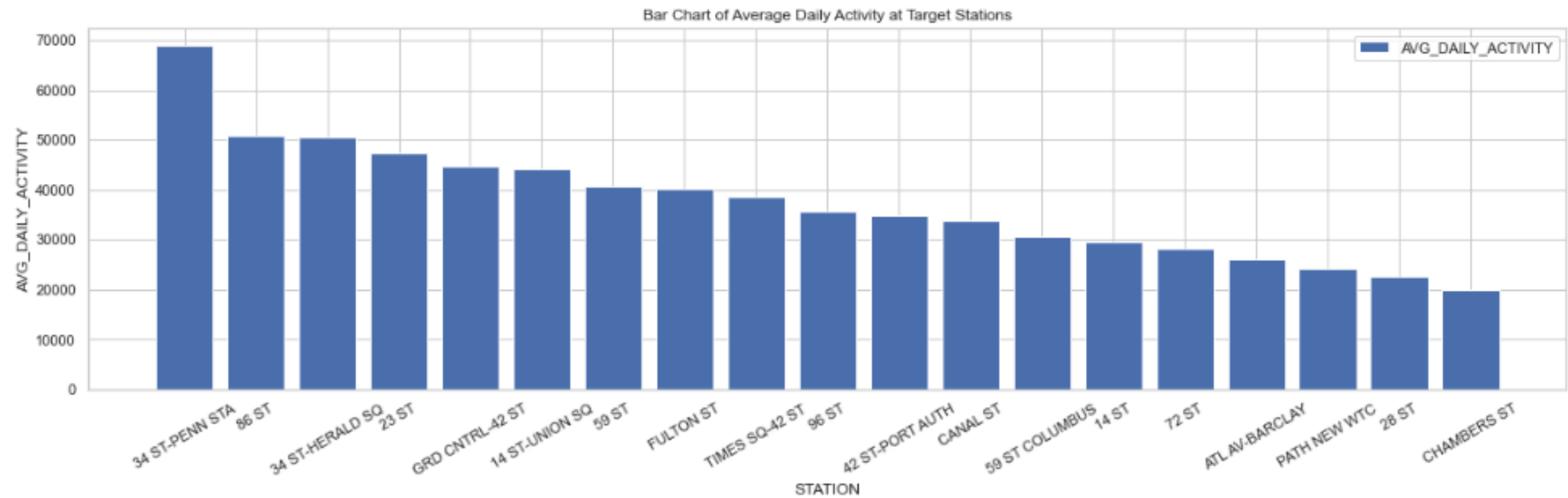

Bar Chart, 34 ST-PENN STA, 2/28/2021 - 3/6/2021
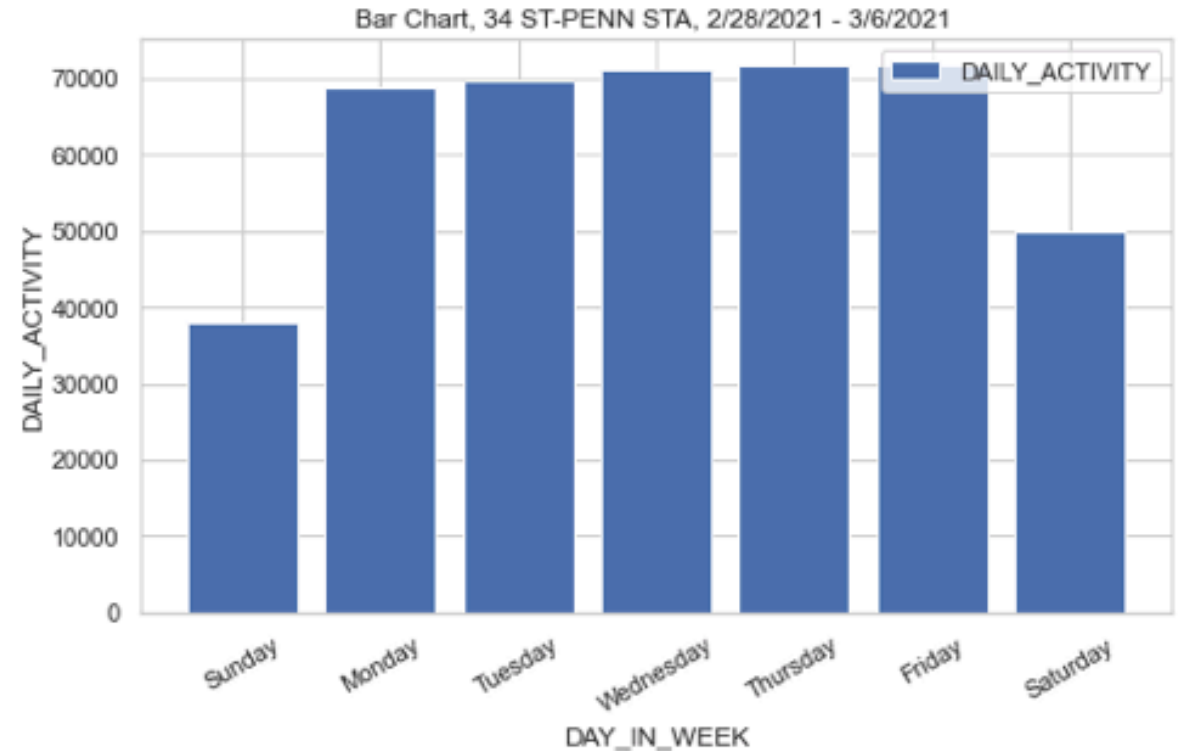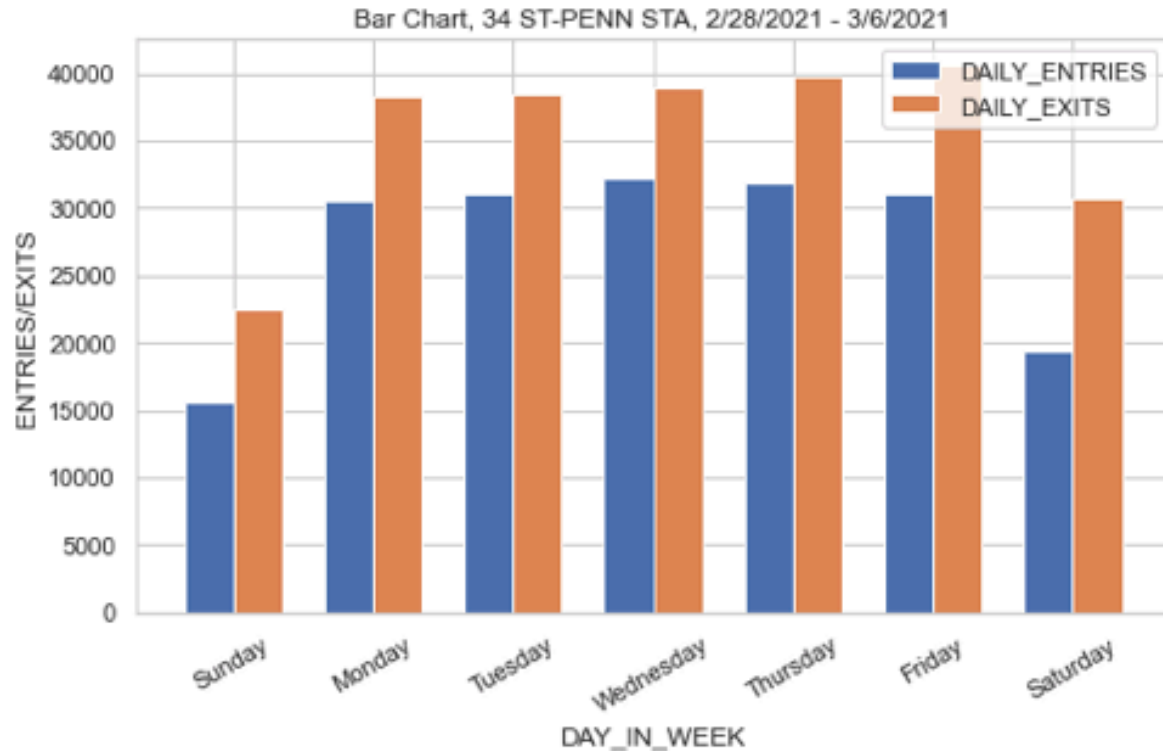
4 HR INTERVAL (MONDAY, MAY 03,2021)
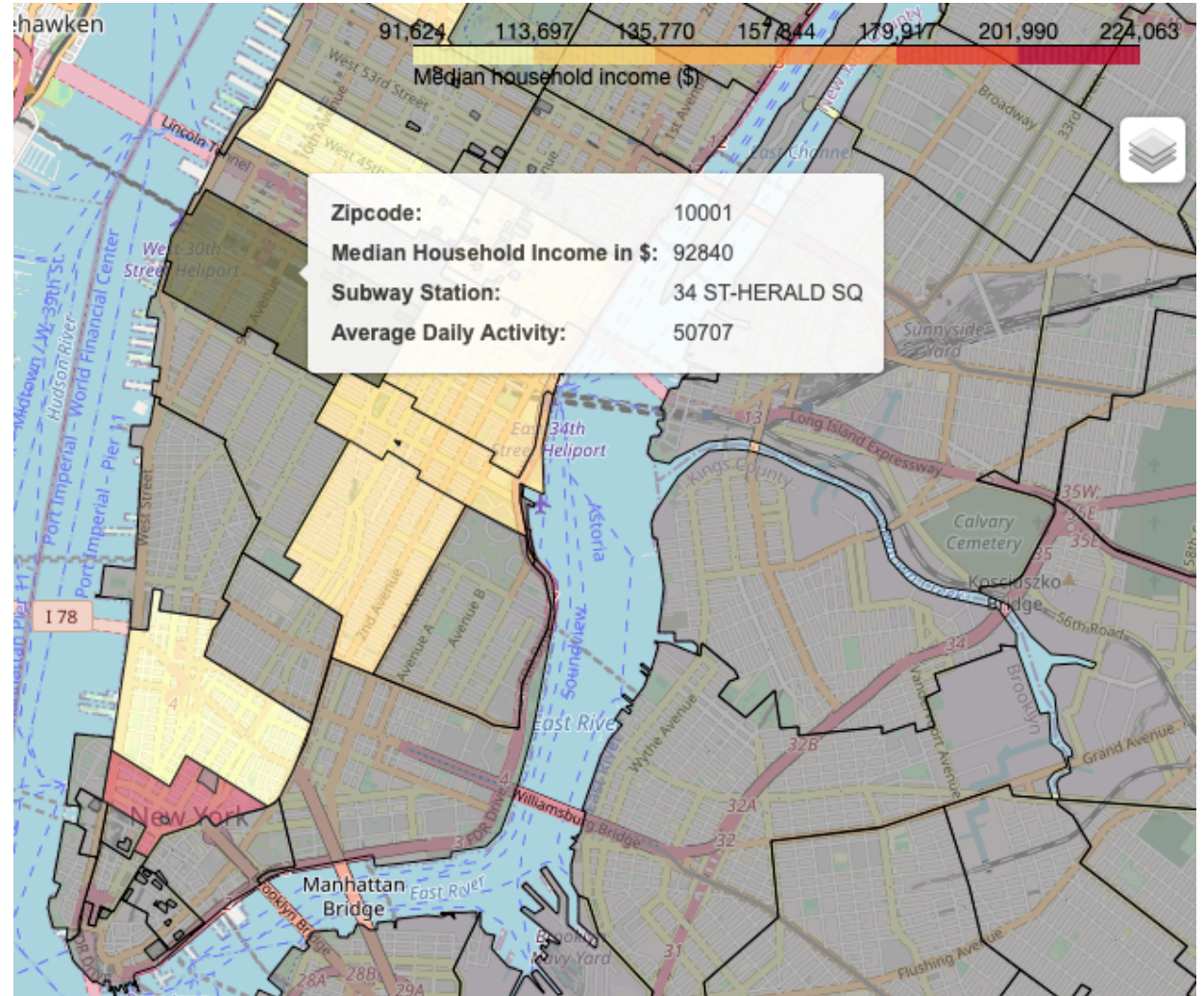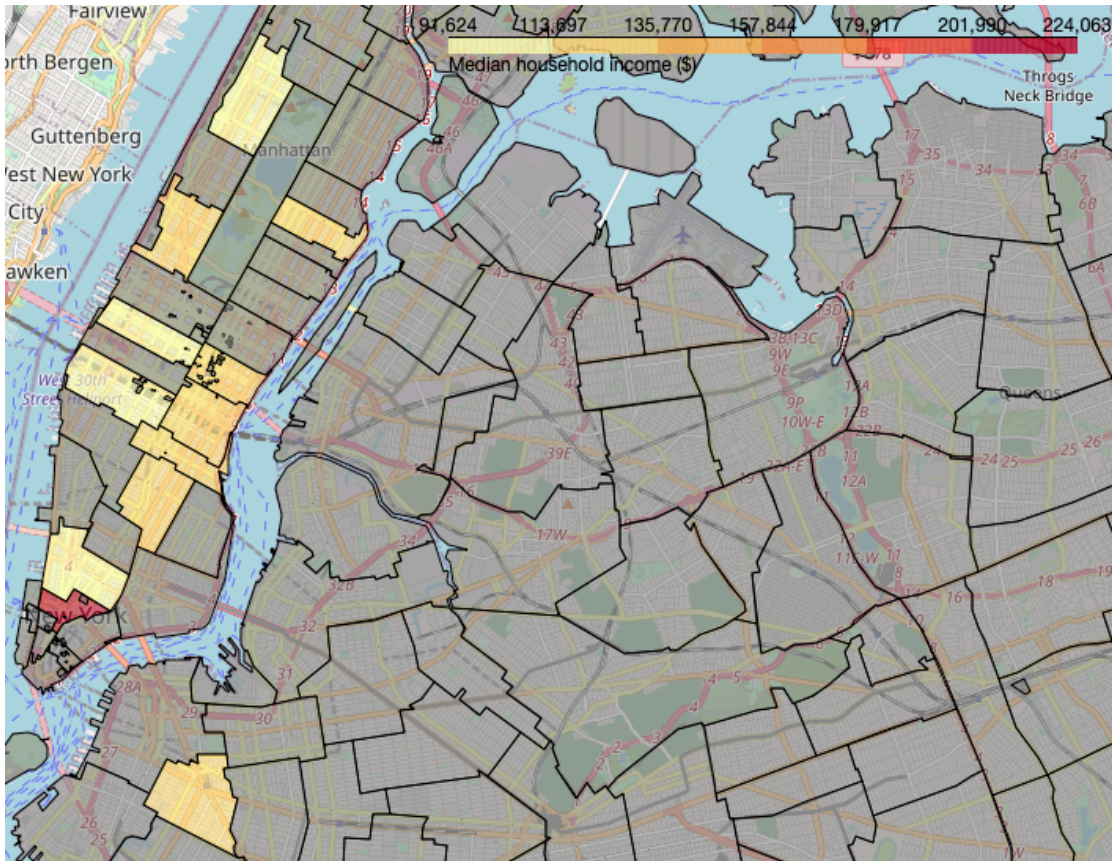
- Active period is between 08:00 a.m. and 08:00 p.m.

TARGET STATIONS

- Average Daily Activity > 19700 (ENTRIES + EXITS)
- Median Household Income > $90,000.00

Bar Chart of Average Daily Activity at Target Stations

INTERACTIVE MAP OF TARGET STATIONS

- Max median income is in lower Manhattan

RECOMMENDATIONS

Based on the current analysis, we recommend the following;

- WTWY agents should focus on the 19 target stations.

- Agents will have a large pool of subway riders between 08:00 am and 08:00 pm during weekdays.

- Target stations in lower Manhattan have a high probability of yielding donors because of the high median income associated with the zip code.