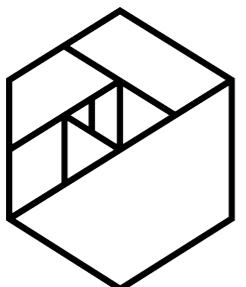


WOMEN IN TECH AND WHERE TO FIND THEM

Exploratory Data Analysis of MTA Turnstiles

Danish James, Flora Xinru Cheng, Luke Newman



METIS

September 27 2019

OBJECTIVES

- Analyze MTA data to:
 - find stations with the highest weekly traffic in May and June 2019
 - Identify stations most likely to attract people who are:
 - women working in or studying technology
 - interested in increasing participation of women in tech

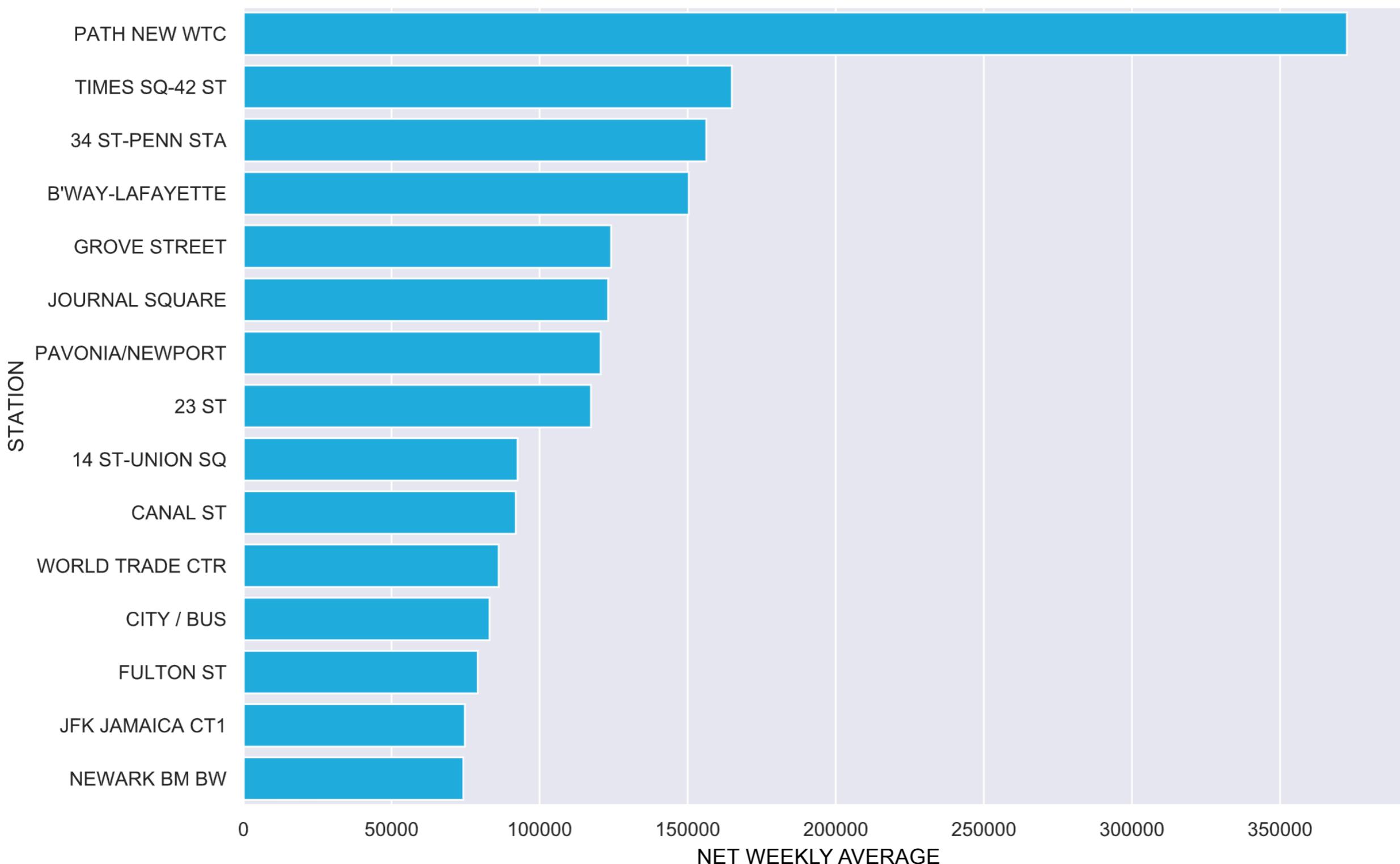


METHODOLOGY

- DATA
 - MTA Turnstiles data (MTA)
 - Additional geographical data:
 - Subway stations (NYC OpenData)
 - Colleges and universities (NYC OpenData)
- TOOLS
 - Python, Pandas, NumPy
 - Matplotlib, Seaborn, GeoPandas
 - GitHub

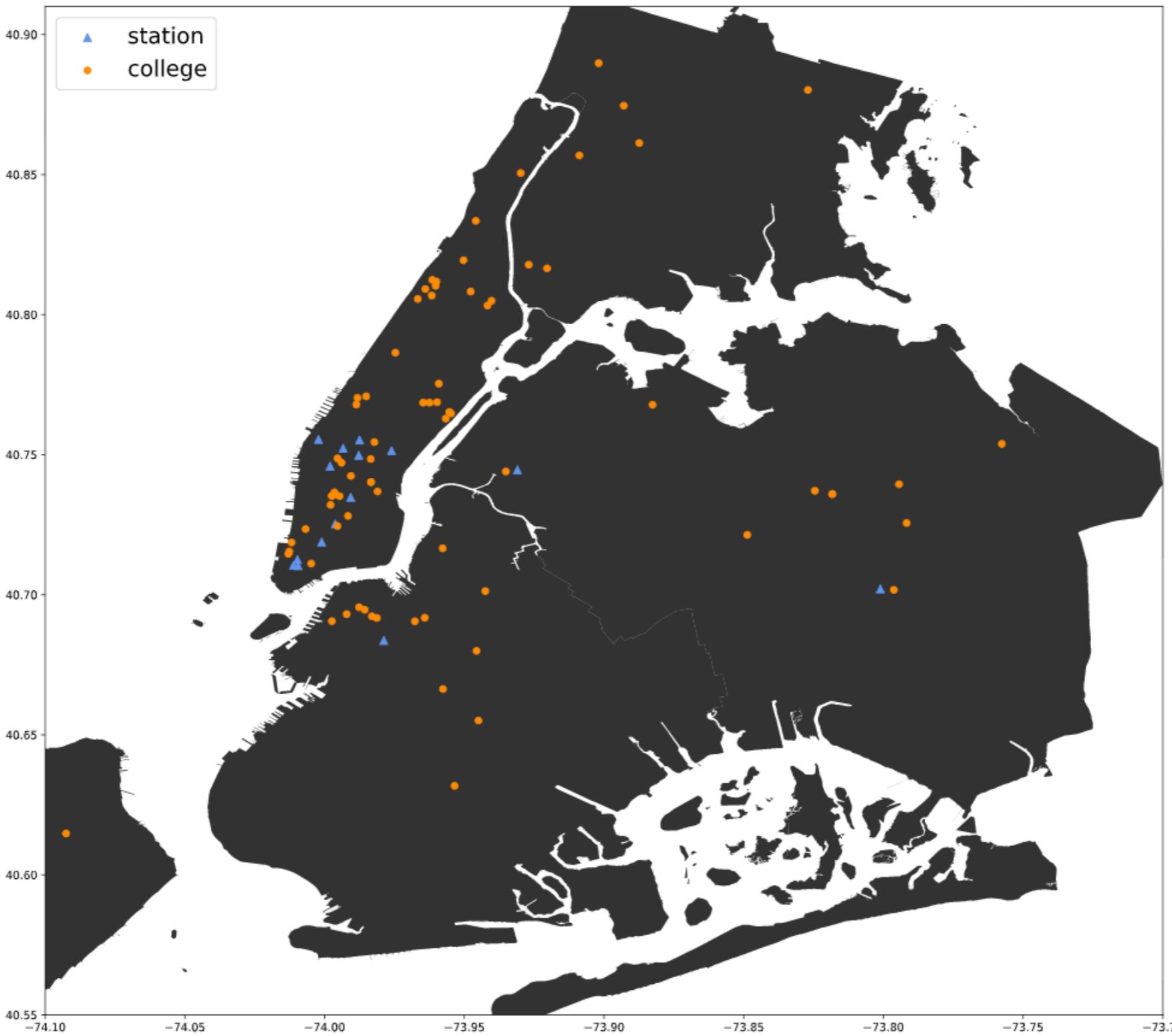
RESULTS

► Plot 1: Top 15 stations with highest weekly ridership values



RESULTS

- Plot 2: Geographical map of top stations and local universities

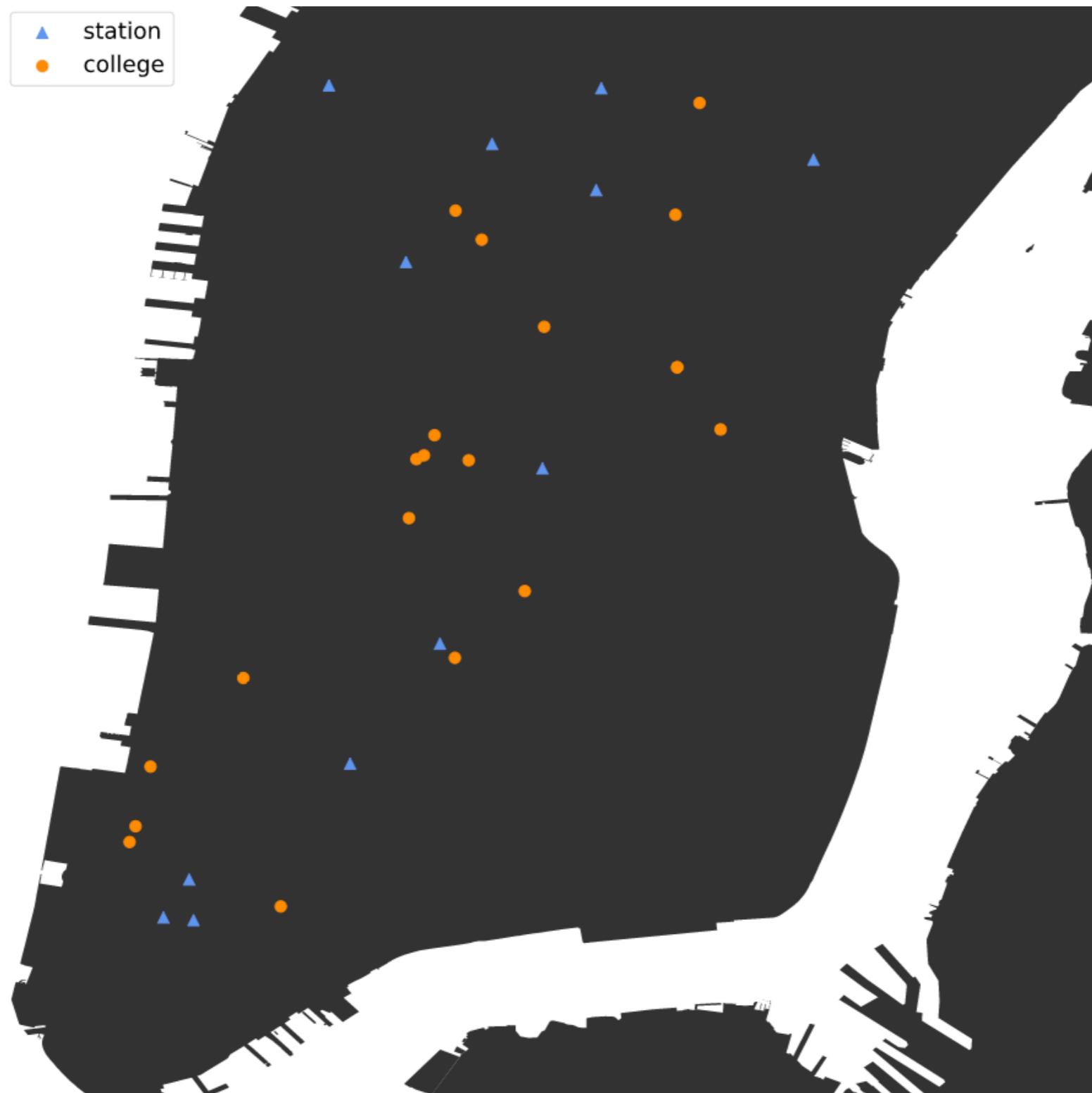


RESULTS

- Plot 2: Geographical map of top stations and local universities

RESULTS

- Plot 2: Geographical map of top stations and local universities



RESULTS

- Plot 2: Geographical map of top stations and local universities



CONCLUSION

- Top stations close to universities
- Put street teams near these stations:
 - 14 ST-UNION SQ
 - 23 ST (8TH AVE Line)
 - FULTON ST



FURTHER WORK

- Plotting location of tech companies in addition to universities
 - Scraping
- Algorithmically compare top stations based on how many tech companies and universities are within a set radius to each



APPENDIX

- Time period: April to June, exclude major holidays and events
- Data cleaning problem: turnstiles resetting
- Reasoning for selecting top recommendations
 - domain knowledge of NYC
 - neighbourhoods
 - outliers - tourism, major transit hubs
- Other factors: Multiple turnstiles per station

