

Using data for optimal train usage

Jing Lee

Metis Bootcamp - Exploratory data analysis

March 23, 2022

Introduction

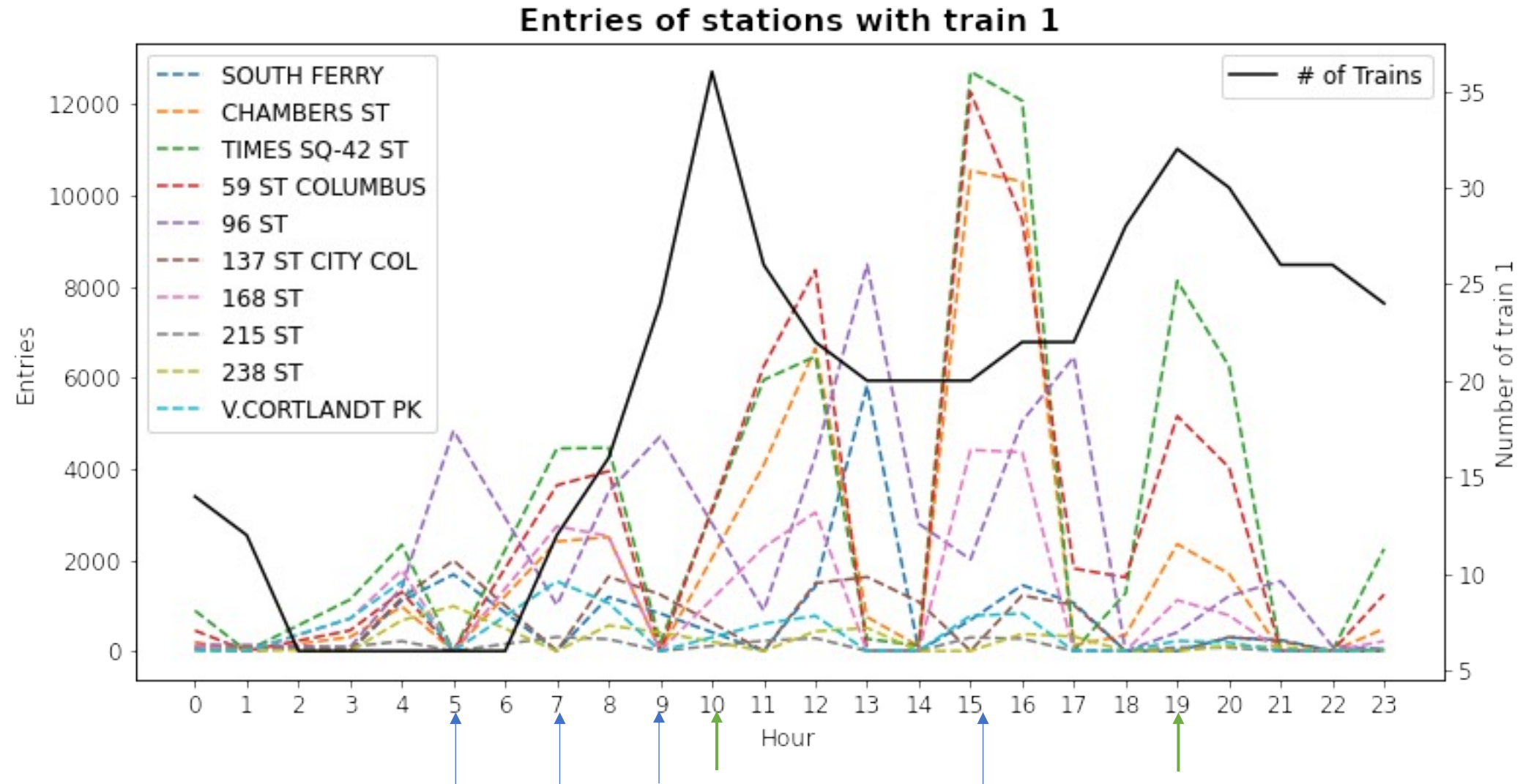
- Motivation:
 - Are trains being optimally used?
- Objectives and goals
 - Provide an efficient train schedule to reduce cost
 - Reduce wait times
 - Have peak number of trains coincide with peak ridership

Methodology

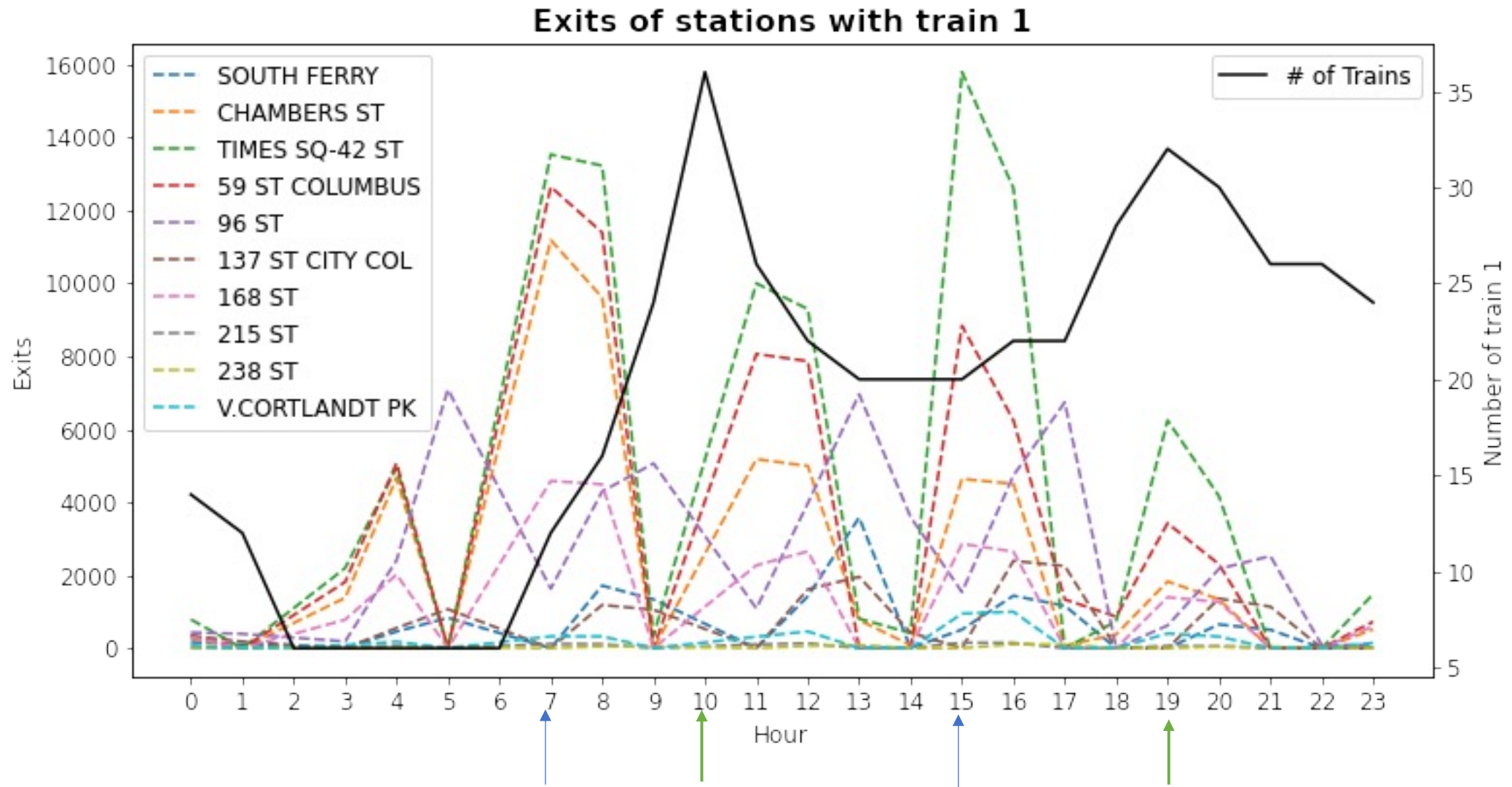
- MTA turnstile data from 09/25/2021 to 12/24/2021
- Analyzing hourly exits and hourly entries on weekday
- Scope: train 1 (Bronx to Manhattan)
- Analyzing number of trains arriving per hour
- Tools: Pandas and Matplotlib

Results

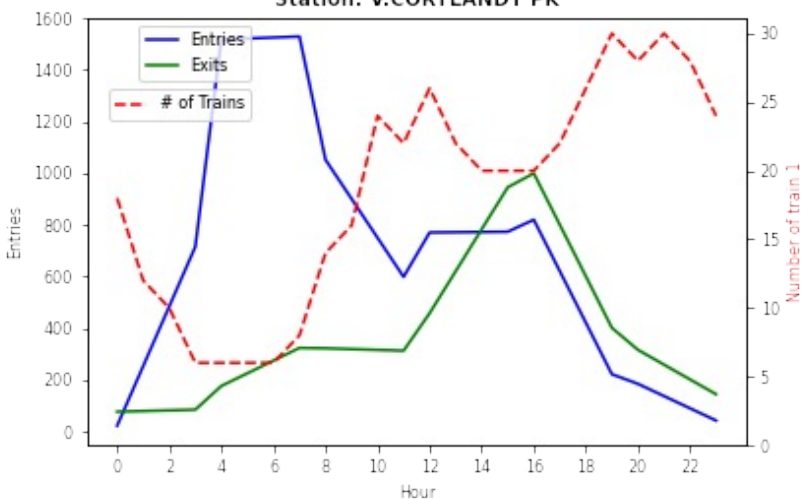
- Peak train arrivals at 10 am and 7 pm
- Peak ridership at 3 pm



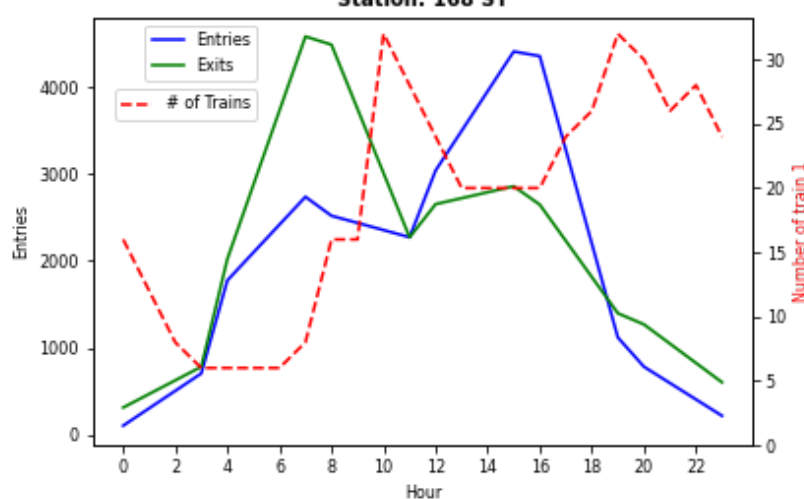
- Peak train arrivals at 10 am and 7 pm
- Peak exits at 7 am and 3 pm.



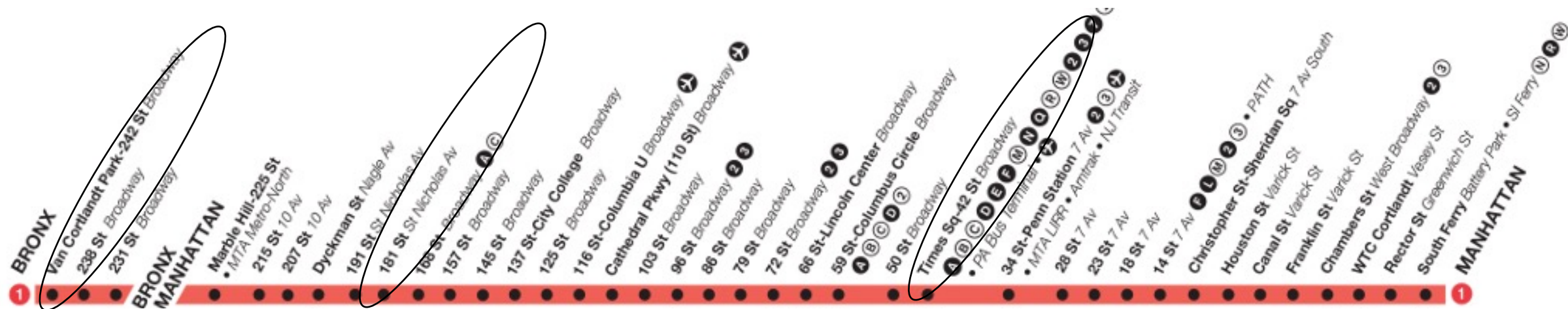
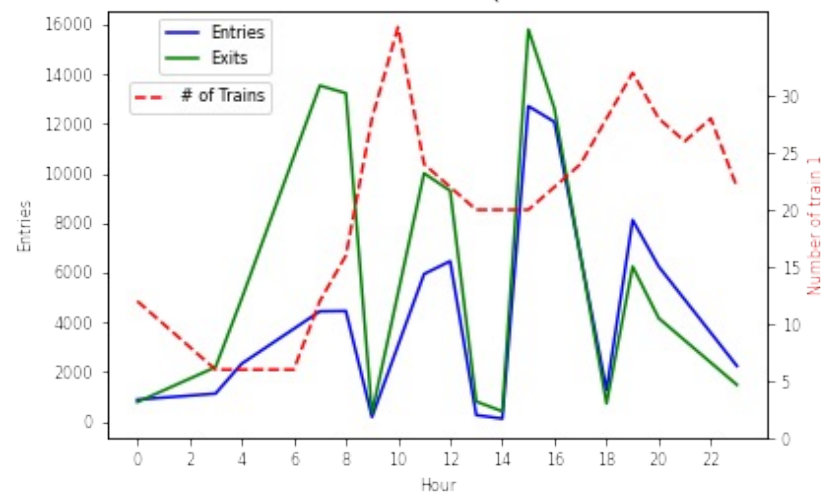
Station: V.CORTLANDT PK



Station: 168 ST



Station: TIMES SQ-42 ST



Conclusions

- Peak train usage is slightly shifted from peak hour of ridership
- Shift trains arrival by 1 hour earlier to overlap with peak ridership

Future work

- Analysis on other trains of the subway
- Take into account that some stations serve multiple lines
- Take into account passenger origin and destinations