Exploratory Data Analysis of the MTA Turnstile Data

Prepared by: Sarah Alabdulwahab

4.3 M. People

Ride the metro every day

5,032 Turnstiles

Across all stations

379 Stations

In NYC

Metropolitan Transportation Authority

Adding turnstile to reduce traffic

Data Collection

- ▷ SQLAlchemy
  - ○ create_engine

- ▷ to_sql

- ▷ read_sql_query

Data

- ▷ Four months of data
- ▷ Main features:
  - ▷ Station
  - ▷ Date
  - ▷ Entries
  - ▷ Exits

Process

Data Cleaning

▷ Drop duplicates

▷ Drop columns

▷ Convert date
  ○ (object → date-time)

Process

Data Cleaning	Feature Engineering

Data Cleaning

▷ Drop duplicates

▷ Drop columns

▷ Convert date
　○ (object → date-time)

Feature Engineering

▷ Unique turnstile ID
　○ C/A, Unit, SCP, & Station

▷ Daily entries and exits

▷ Total daily traffic

Process

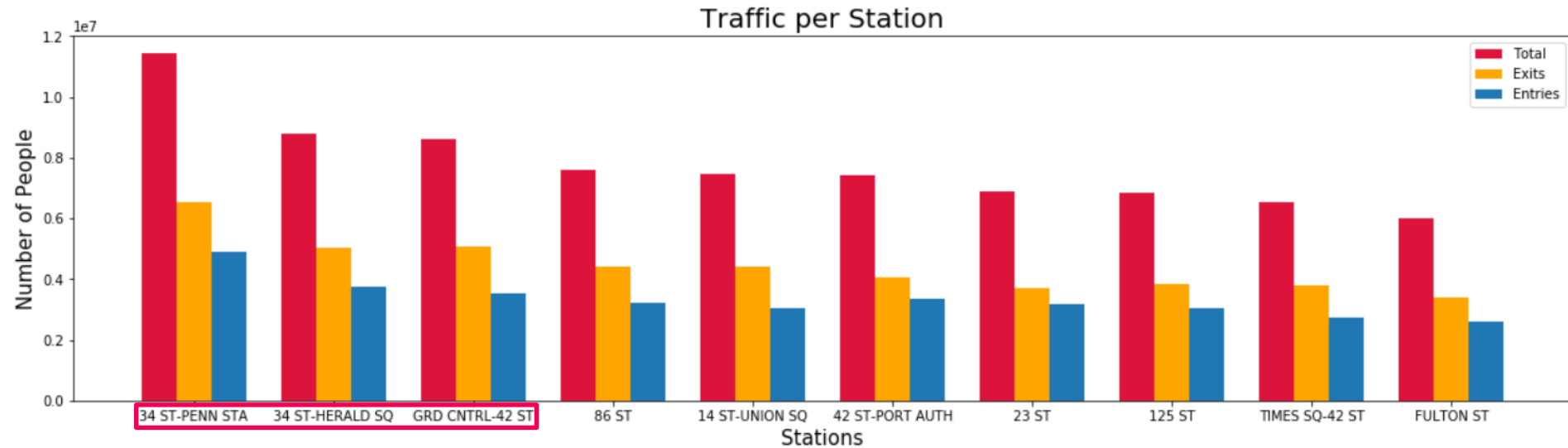| Data Cleaning	Feature Engineering	EDA

- ▷ Drop duplicates

- ▷ Drop columns

- ▷ Convert date
  - ○ (object → date-time)

- ▷ Unique turnstile ID
  - ○ C/A, Unit, SCP, & Station

- ▷ Daily entries and exits
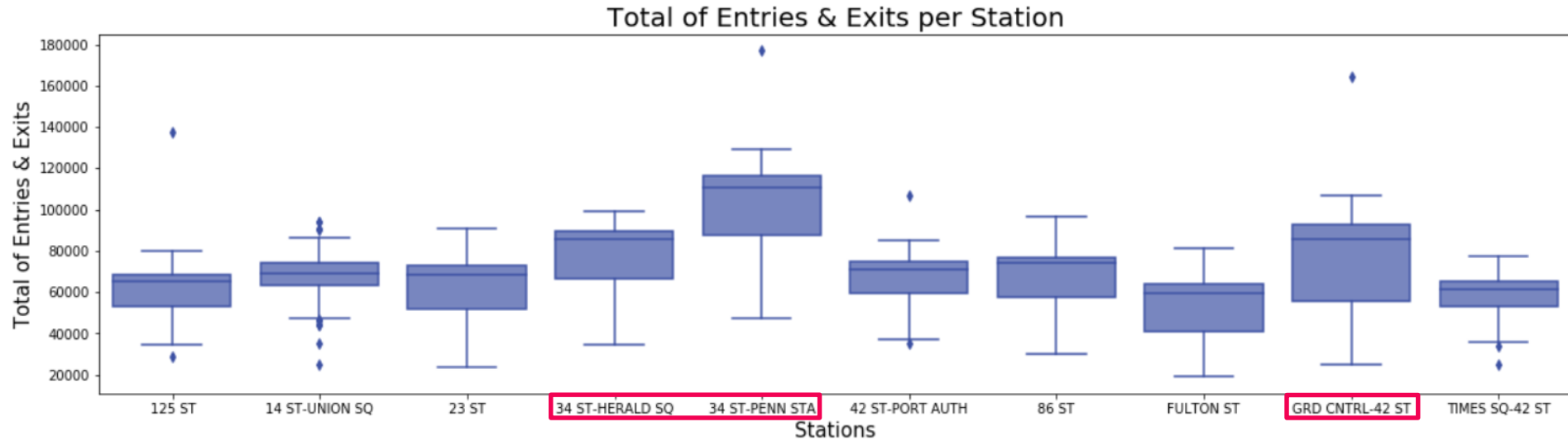
- ▷ Total daily traffic
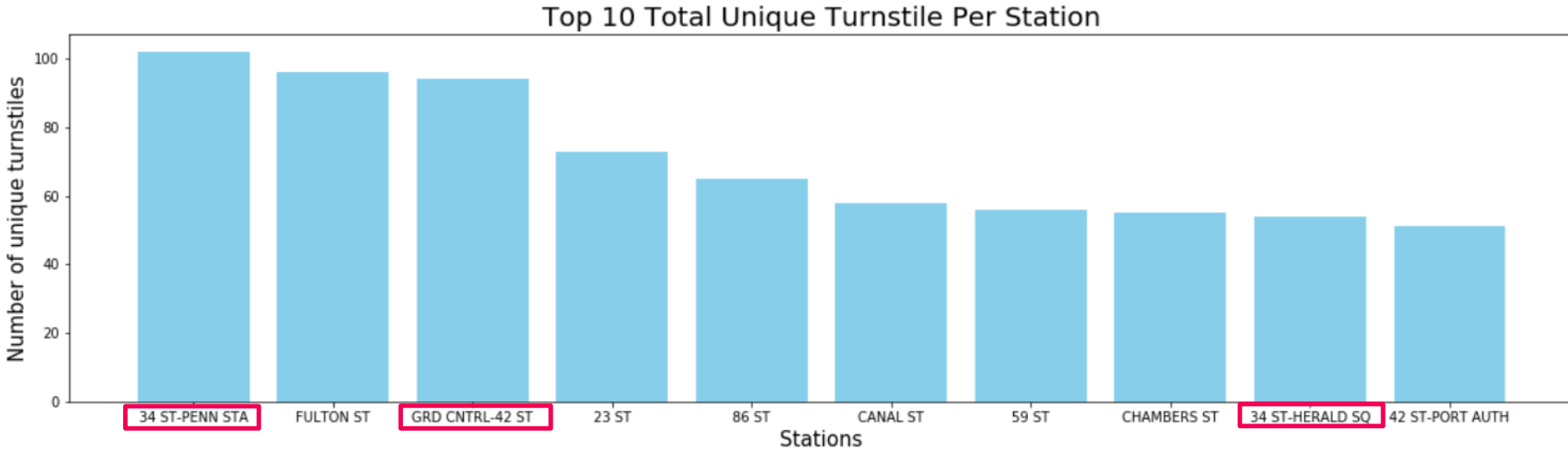
Total VS Exits VS Entries

Daily Traffic

Total of Entries & Exits per Station

Unique Number of Turnstiles

Top 10 Total Unique Turnstile Per Station

Conclusions

▷ Turnstiles should be added in 34 ST-PENN STA, 34 ST-HERALD SQ, and GRD CNTRL-42 ST.

▷ Regarding the number of turnstiles, 34 ST-HERALD SQ should be prioritized.

Thanks!

Any questions?