Exploratory Data Analysis

MTA Turnstile Data & Harassment on NYC Subway

Tara Koplik

Introduction

Client:

Goal:

Provide analysis of both MTA turnstile data & reports of harassment crimes on the subway to inform the scheduling practices for NYPD Transit Officers. Stations with high foot traffic/high reports of Harassment crimes should receive priority.

Methodology

Data

Data Ingestion, Storage, & Extraction
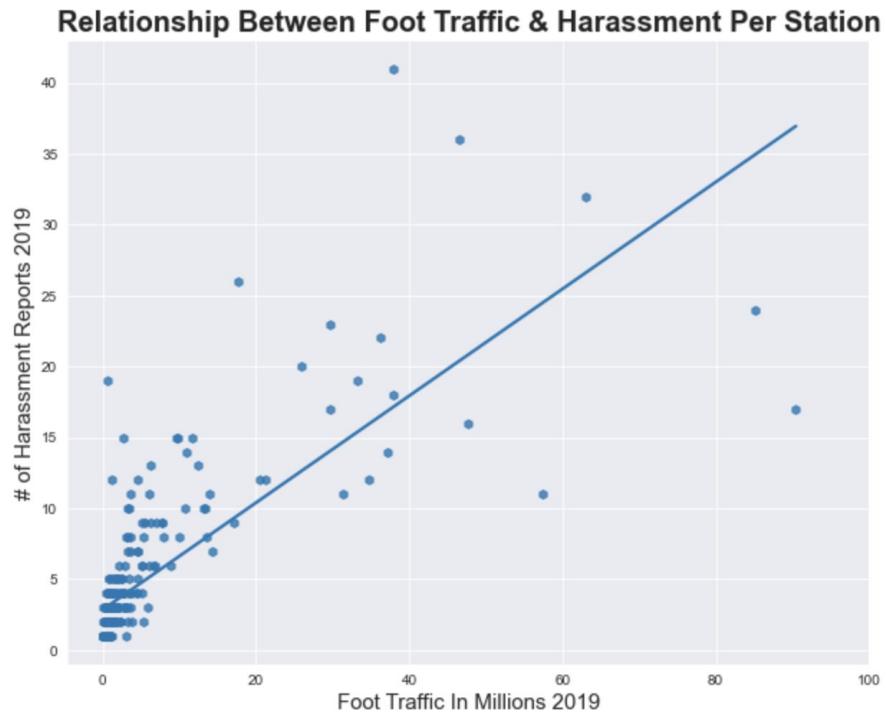
Cleaning

Visualization

HARASSMENT IN THE SECOND DEGREE
Penal Law 240.26

- ❏   strikes, shoves, kicks or otherwise subjects such other person to physical contact, or attempts or threatens to do the same.

- ❏   follows a person in or about a public place or places.

- ❏   engages in a course of conduct or repeatedly commits acts which alarm or seriously annoy such other person and which serve no legitimate purpose.

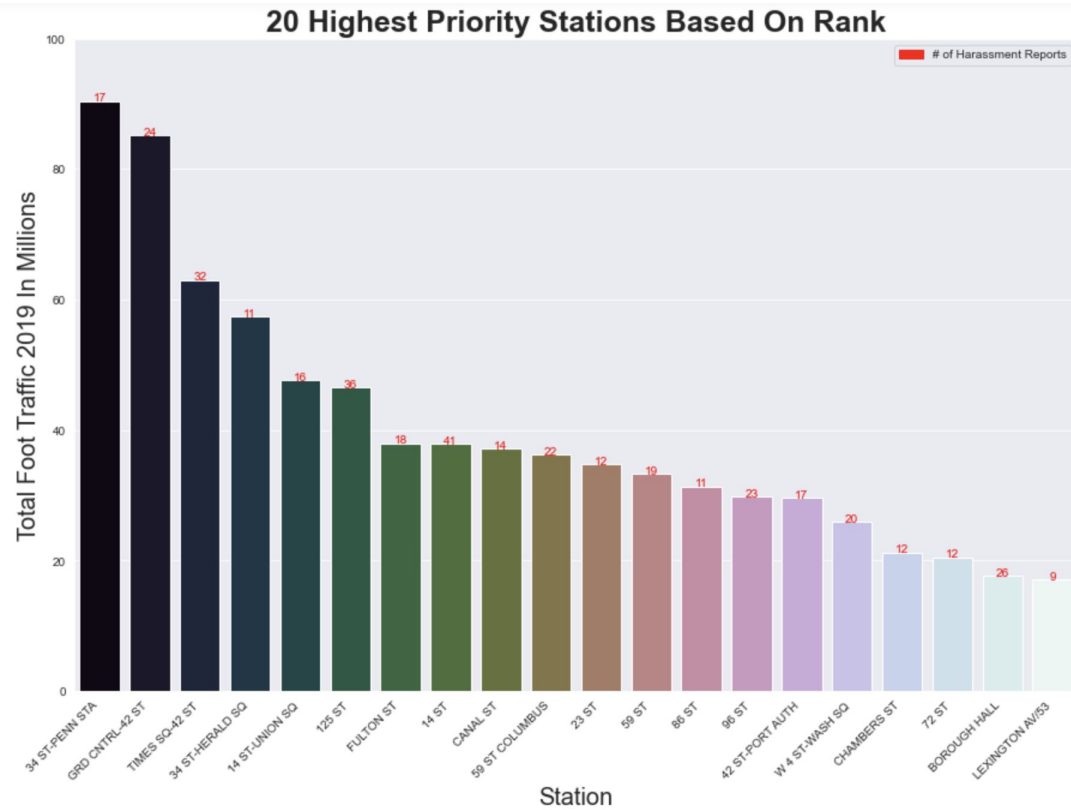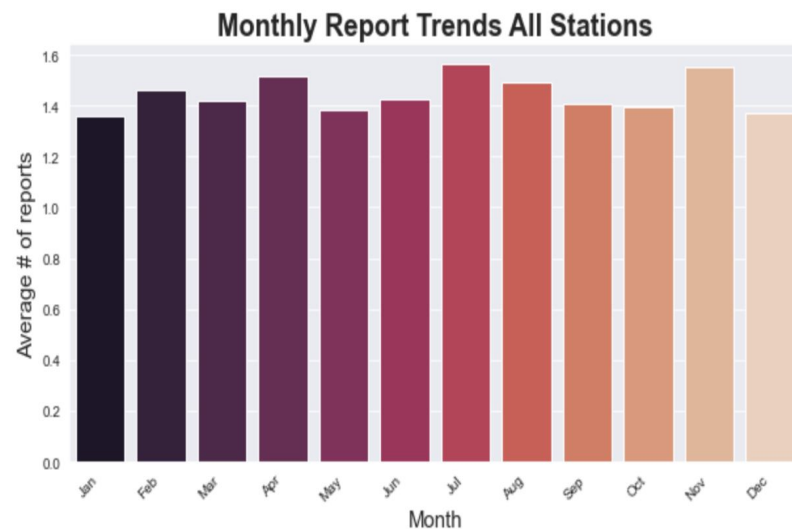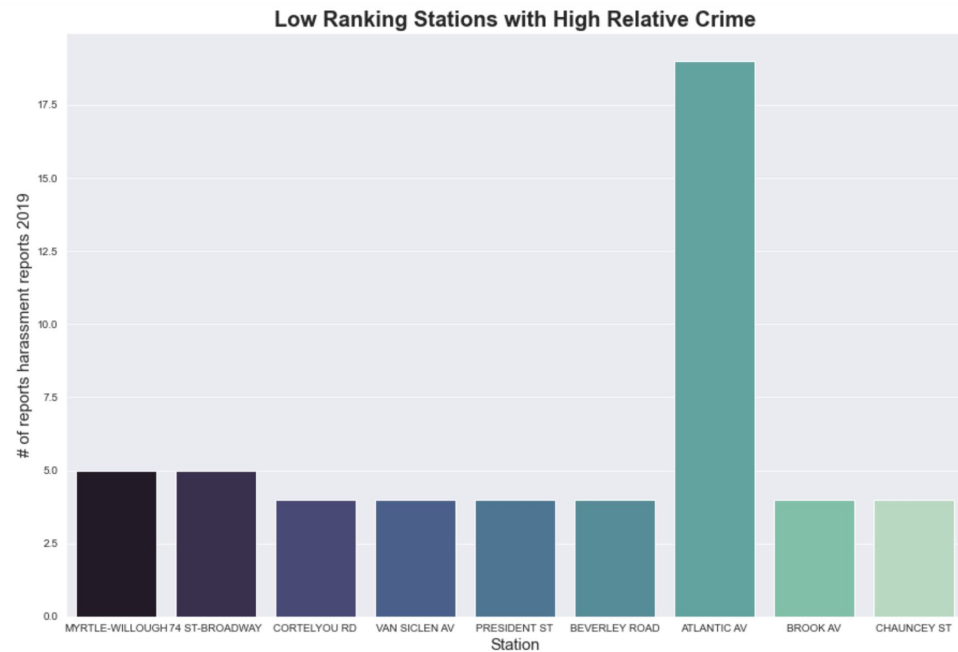- ❏   did so with intent to harass, annoy or alarm

The Data

Relationship Between Foot Traffic & Harassment Per Station

Correlation Coefficient: 0.76

The Data

20 Highest Priority Stations Based On Rank

The Data

Low Ranking Stations with High Relative Crime

Recap & Recommendations

❏ Add additional transit officers to high priority stations & stations with high crime rate relative to the foot traffic.

❏ Monthly report rates are fairly even.

❏ Deeper analysis may be done to see if time series patterns can be narrowed down further.

❏ Remove officers from the schedule in low trafficked/low crime areas.
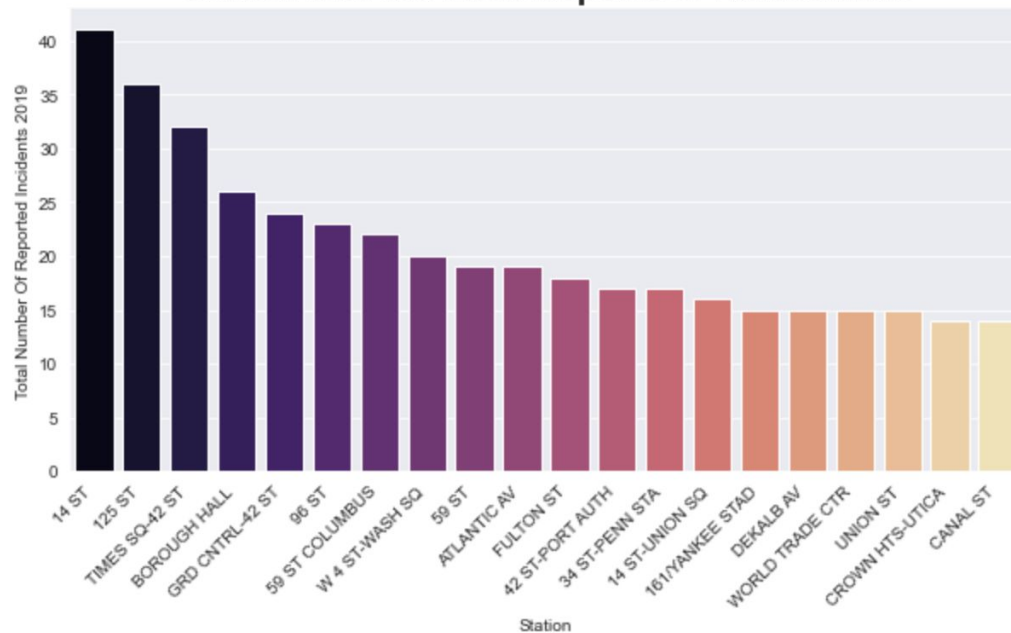
Precinct's near the top priority stations

Future Work

❏ Apply this process to more recent figures

❏ Build a model to predict future trends

❏ Include other prevalent types of crime

Appendix

Station with the Most Reports of Harassment

Appendix

Stations with the Lowest Foot Traffic