The Metropolitan Transportation Authority

Exploratory Data Analysis (EDA)

Outline

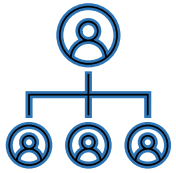
Outline

- Introduction
- Backstory
- Data set
- Algorithm
- Tools
- Analysis result

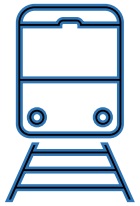
Introduction

Introduction

The growth of population in USA



The government of NY found a solution



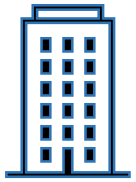
Established the metro since 1965

The metro suffering from some issues

Tring to solve it using Data Science

Backstory

Backstory



Consultant company working in solutions by data science

The company has received an email

The government institution works on improving New York City

Organize the overwhelmed stations specially during rush hours

Reach the required satisfaction levels

Data Set

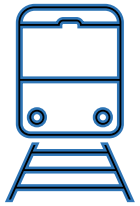
Data Set

The used data from MTA data set
are 3 months of 2021
May, June, July

Data Set

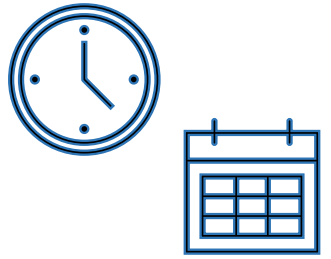
Turnstile

Stations



Line names

Number of
entries and exits



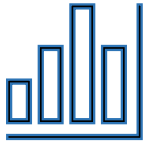
Date and Time

Algorithm

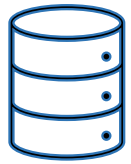
Algorithm

Explore data

Clean data



Plot data



Write queries

Provide a solution

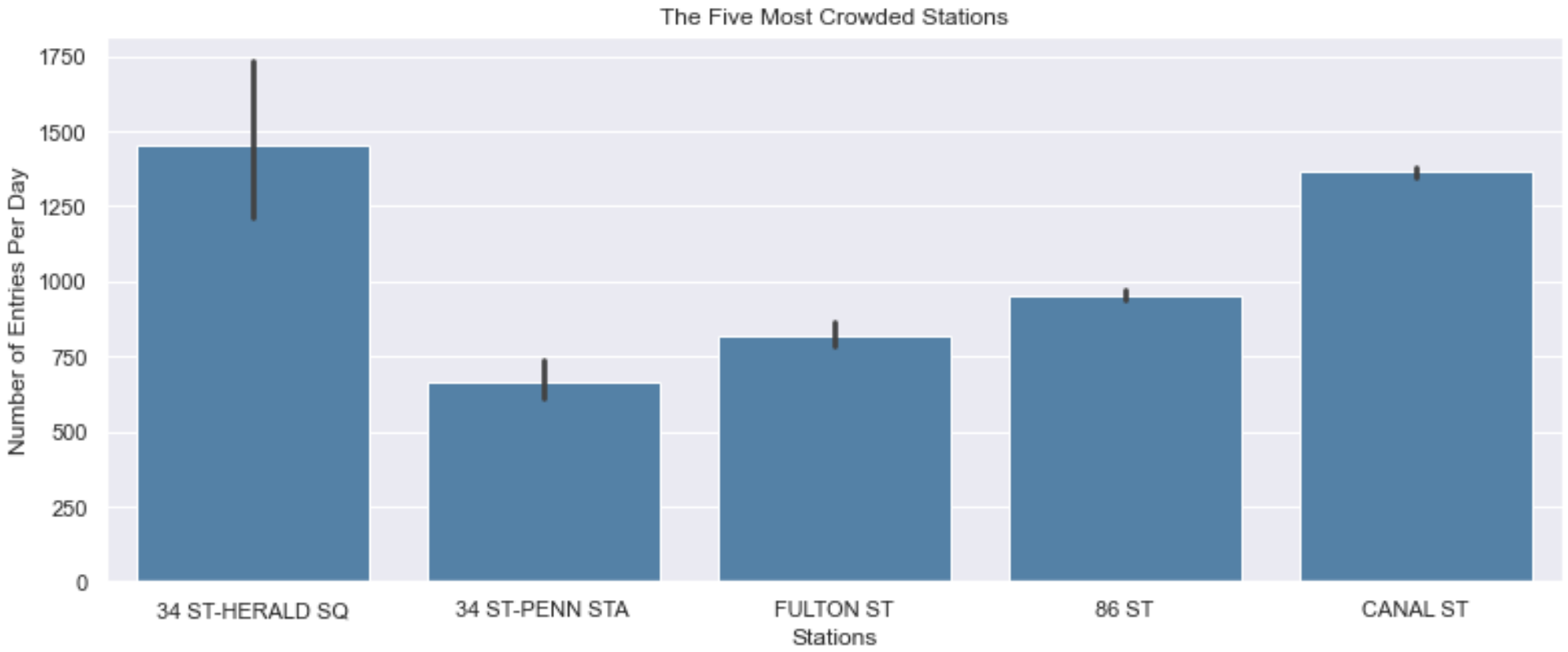
Tools

Tools

- Python NumPy
- Python Pandas
- Python Time
- Python Date
- Python Seaborn
- Python matplotlib
- SQLalchemy
- Python Word Cloud

Analysis Result

Analysis Result



The busiest station is 34 ST PENN STA
The number of entries per day of the top 5 stations

Final Recommendation

Final Recommendation

Short term solution

Expand the busiest stations and increase the turnstiles number

Final Recommendation

Long term solution

Increase the number of stations near to the busiest stations

Thank You

Basma Ghazi Abdullah Alduaiji

2021