T5 Data Science Bootcamp

Exploratory Data Analysis for the Metro Motion on the MTA

By Nouf Alotaibi

METRO MOTION

Metro Motion

Providing repairing and cleaning services for metropolitan transportation in New York City. The company is serving a lot of stations that belong to the Metropolitan Transportation Authority.

Problem Statement

Metro Motion

1.

BACK TO REALITY

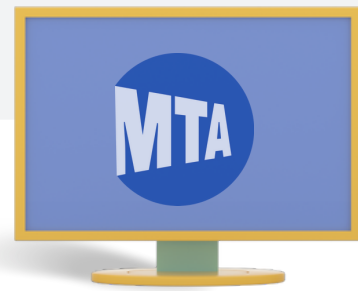Many people are returning to use the transportation service.

2.

SCEDUALE NEEDS TO BE UPDATED

Schedule is not meeting the right times for repairing and cleaning.

3.

RESOURCES AND TIME

Company is wasting a lot of resources and time.

Goals for Metro Motion

WORK ON MTA DATASET

Get insights on the most crowded stations, active days and times.

ESTIMATE THE TIMES

Estimate the times for repairing and cleaning services for stations.

PREPARE A SCHEDULE

Prepare a schedule for the company for when and where to provide the services.
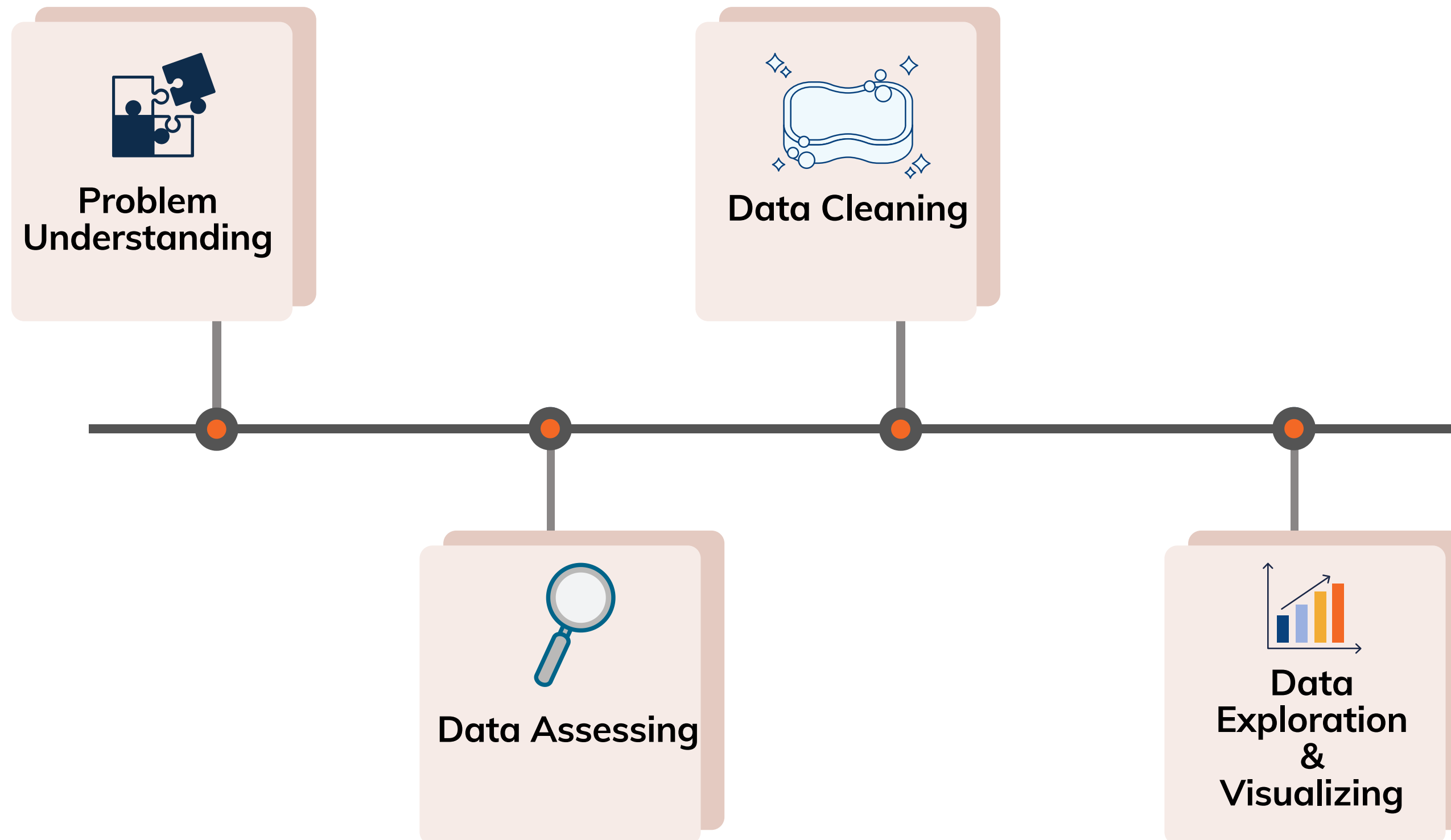
Dataset

- **2019, OCTOBER**
- **2019, NOVEMBER**
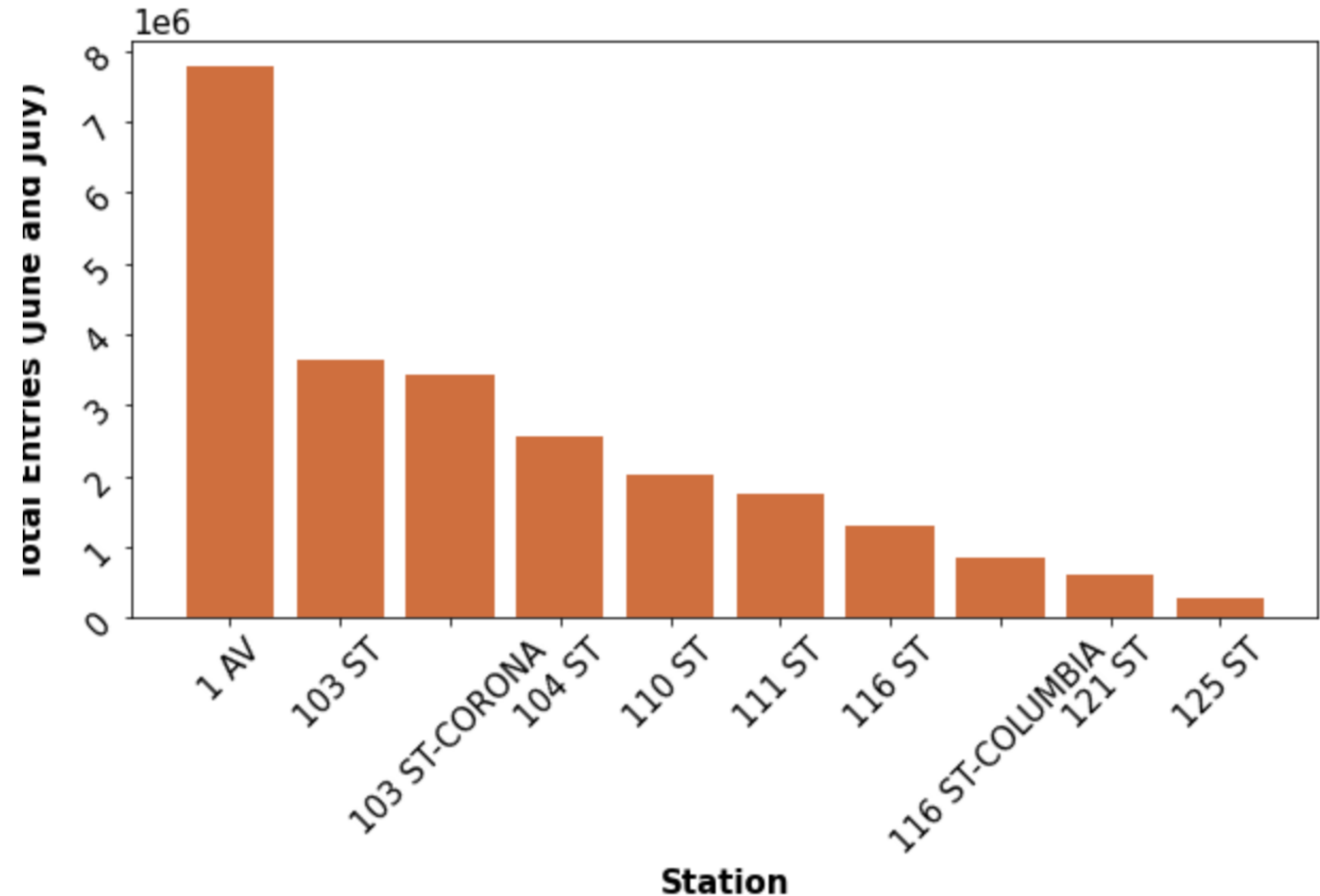
- **2021, JUNE**
- **2021, JULY**

= **3,740,978 RECORDS**

11 ATRRIBUTES

Methodology

Problem Understanding

Data Cleaning

Data Assessing

Data Exploration & Visualizing

Top 10 Active Stations in NYC

As you can see, there is a big difference between **1 AV Station** and other stations.

Top 10 Active MTA Stations in the Year 2021
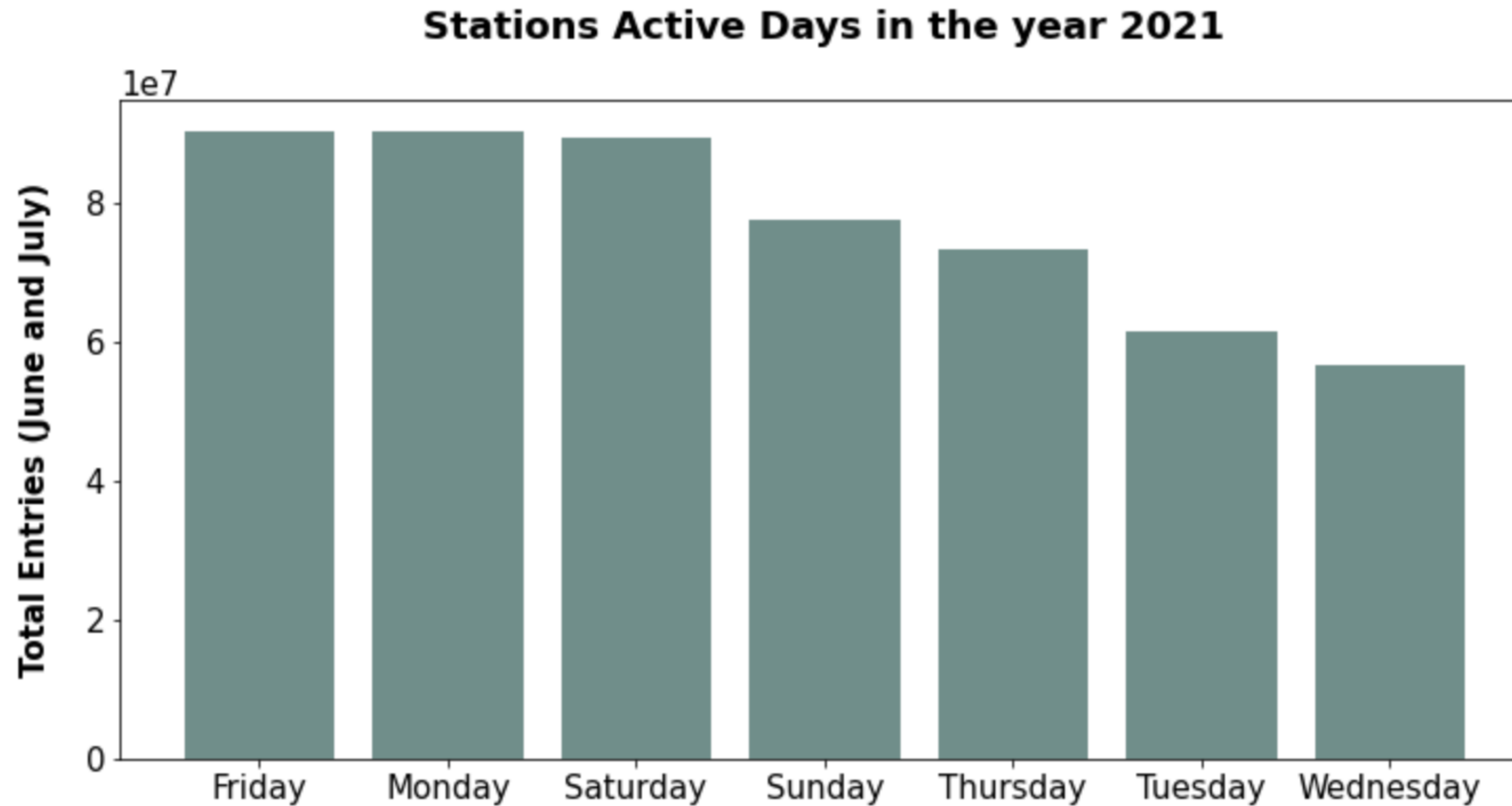
Stations Active Times in NYC

Many turnstiles accrues at Night and Afternoon.

Avoid working in those times in the crowded stations.

Stations Active Times in the Year 2021

Stations Active Days in NYC
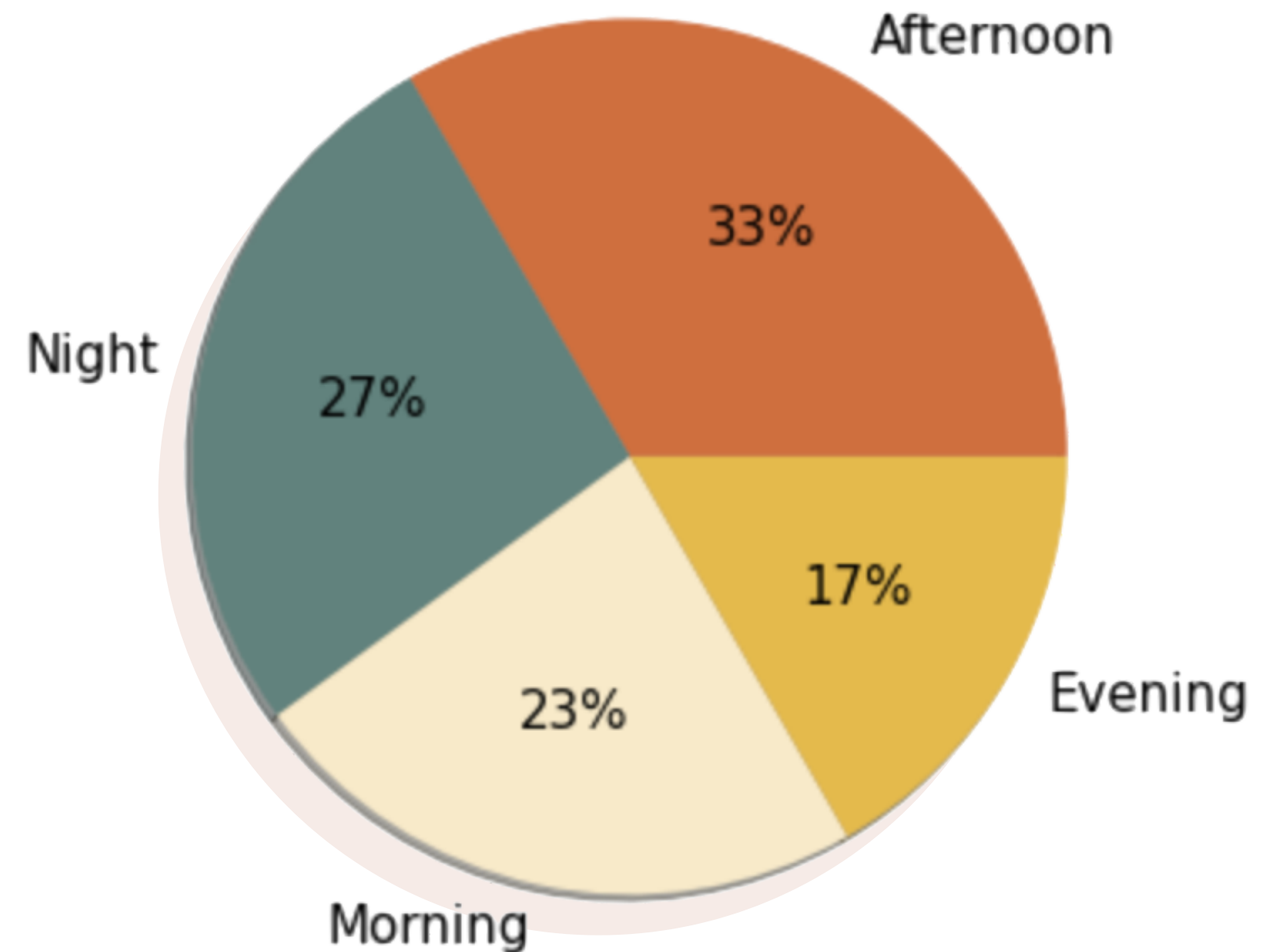
Stations Active Days in the year 2021

Most active days are weekends, and the start of the week.

It gets less by the end of the week.

Conclusion

What's Next for Metro Motion?

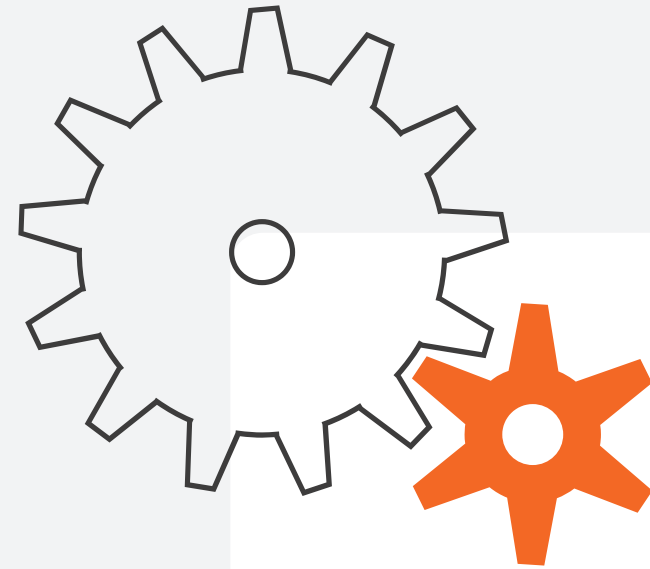Provide the cleaning and repairing services for the **less crowded stations** in the **morning** or at **night**, in the **start of the week or by the end of it.**

Provide the cleaning and repairing for the crowded stations at Evening, **in the middle of week.**

Challanges

- Limited time to explore and work on the dataset.

- Issues with the dataset: Completeness, Validity, Accuracy, Consistency.

Thank you