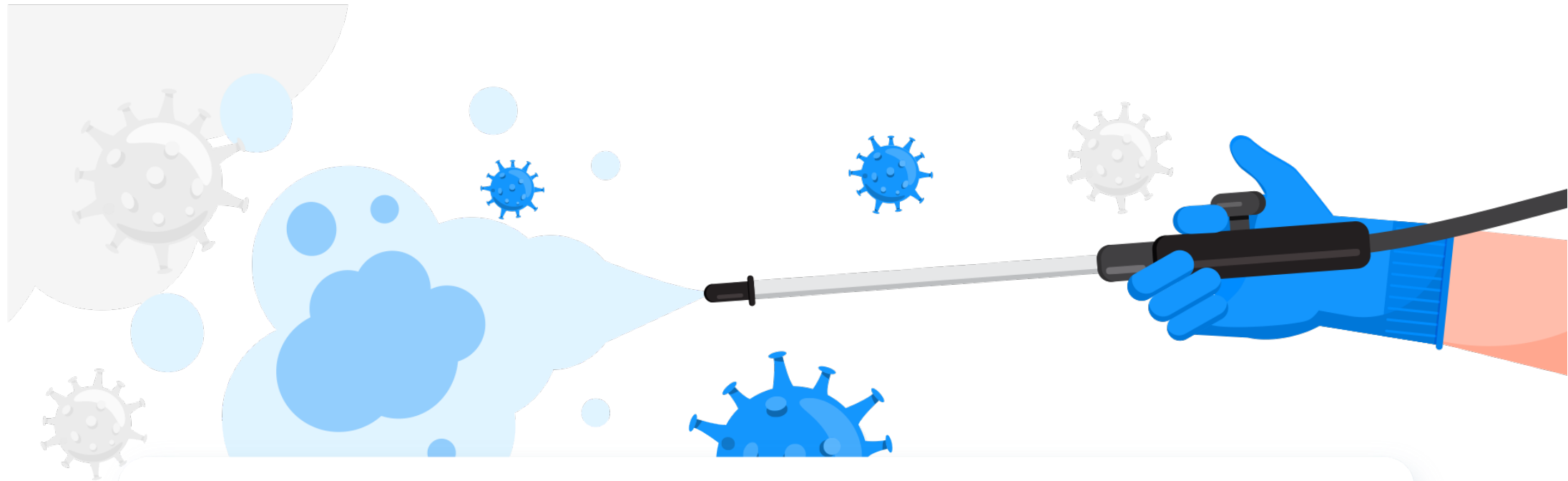


T5 Data Science Bootcamp

Exploratory Data Analysis for the MTA Data

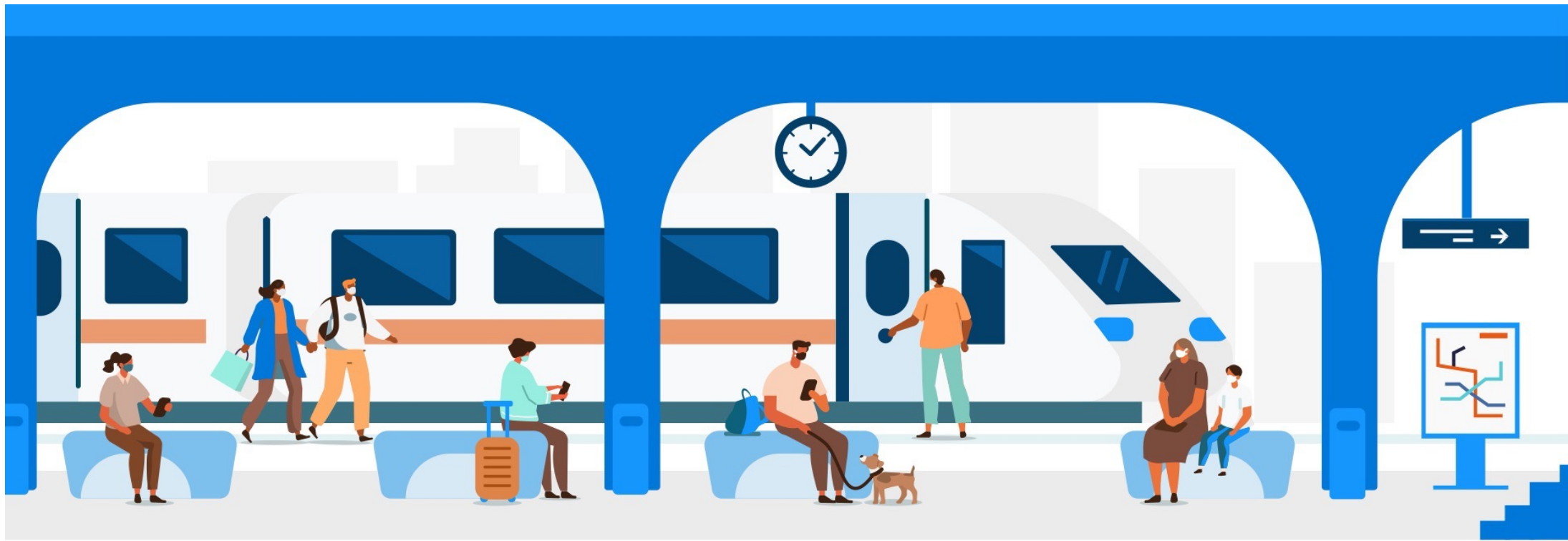
أكاديمية سدايا
SDAIA Academy





Intoduction :

This exploratory analysis supports MTA turnstile's sanitizing company to help them improve their work schedule by estimating the number of sanitizing tours each station needs per day, based on their needs, with the busiest station requiring more than once a day. As I work on this issue, I will answer the questions below.



What are the most active stations?
What are the most active times for each station?
What are the most active days of the week?
What is the best schedule for the company?

Dataset



MTA turnstile data

A series of data files containing numbers of cumulative entries and exits by stations, turnstile, along with their dates and time.



Sample

first three months of 2021 (January-Mar)
2719827 rows 11 columns



Step 1
Renaming columns



Step 2
Removing duplicates
rows



Step 3
Calculating the daily exits and
entries Two Issues:
1- Counter reversed (Solved)
2- Counter reset (Solved)



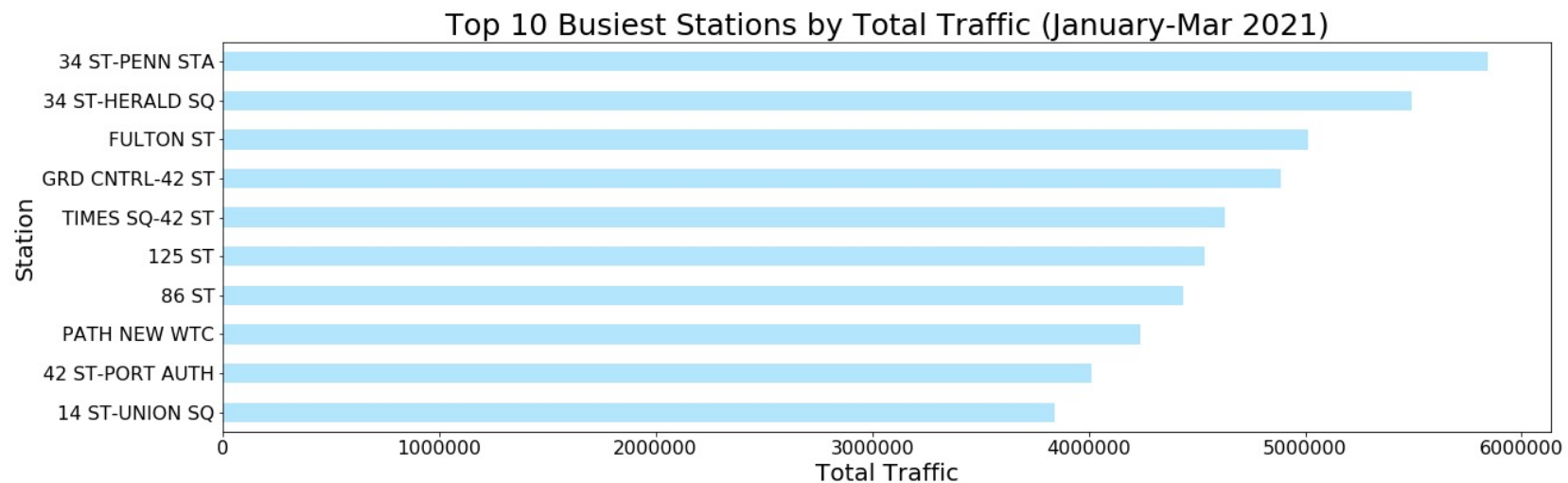
Step 4
Removing outliers
Using Boxplot



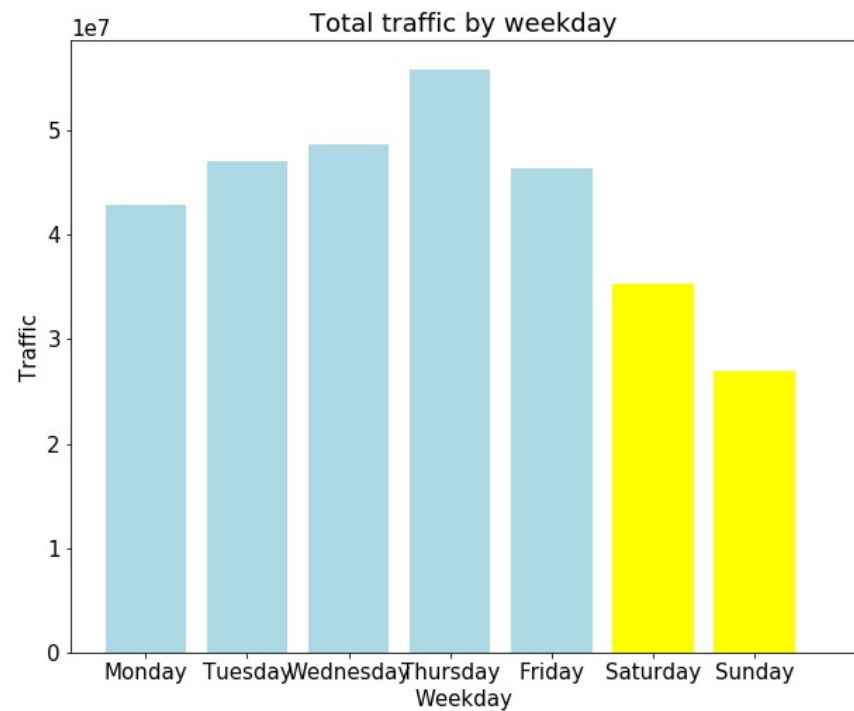
Visualizing data

Now it is time to answer our questions!

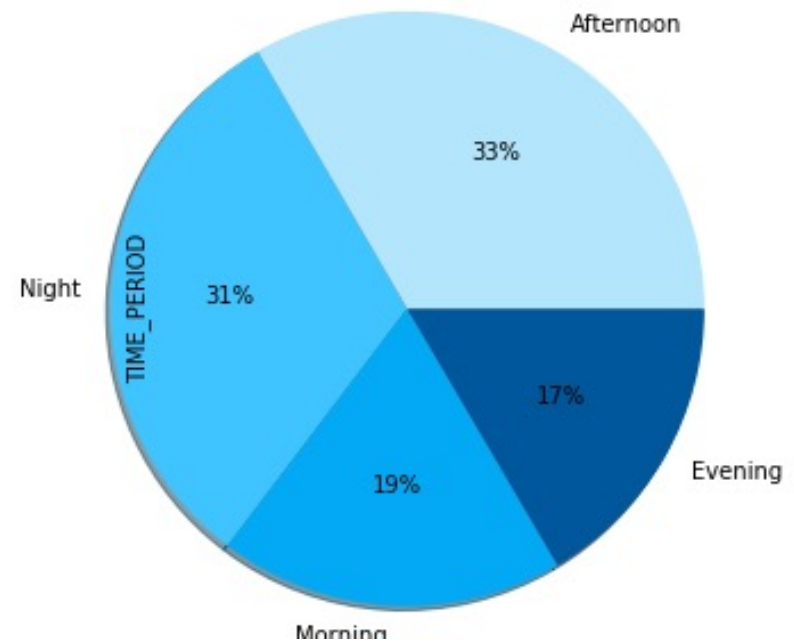
Q1. What are the ten most active stations?



Q2 What are the most active days of the week?



Q3 :What are the most active times for each station?





Conclusion and recommendation

The ten stations
need extra tours.

Focus on weekdays
rather than weekends

Time tours: afternoons
and Night.



Thank you