SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية سدايا
SDAIA Academy

SDAIA Academy T5C04 Bootcamps: Data Science
EDA Module Project Presentation

MTA Turnstile Data: an exploratory analysis

Presented by
Nada Rambu

1 Problem and Solution

Project background and objectives

Women Tech Women Yes (WTWY)

What is missing?

They need to know the most crowded subway stations

How can data help?

By exploring and analyzing the turnstile data of the stations

Assumption

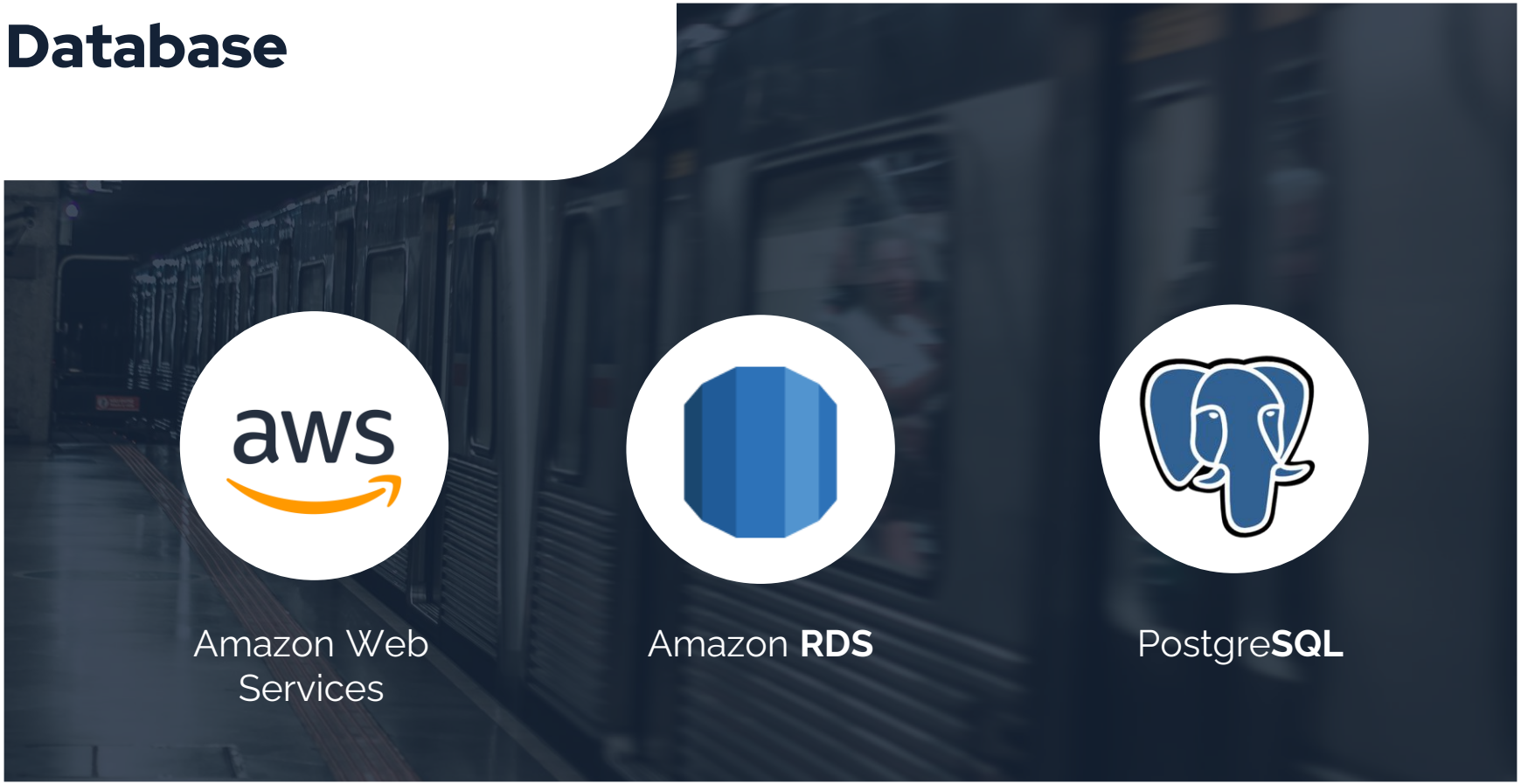
A specific pattern can be found in the data of the same season

2

Data Acquisition

Bringing the data from the its source to be processed

Database



3 Data Cleaning

Preparing the data to be plotted

Data Cleaning

A.

Standardize column names

B.

Merging `Date` and `Time` columns

C.

Converting the merged column to `datetime` data type

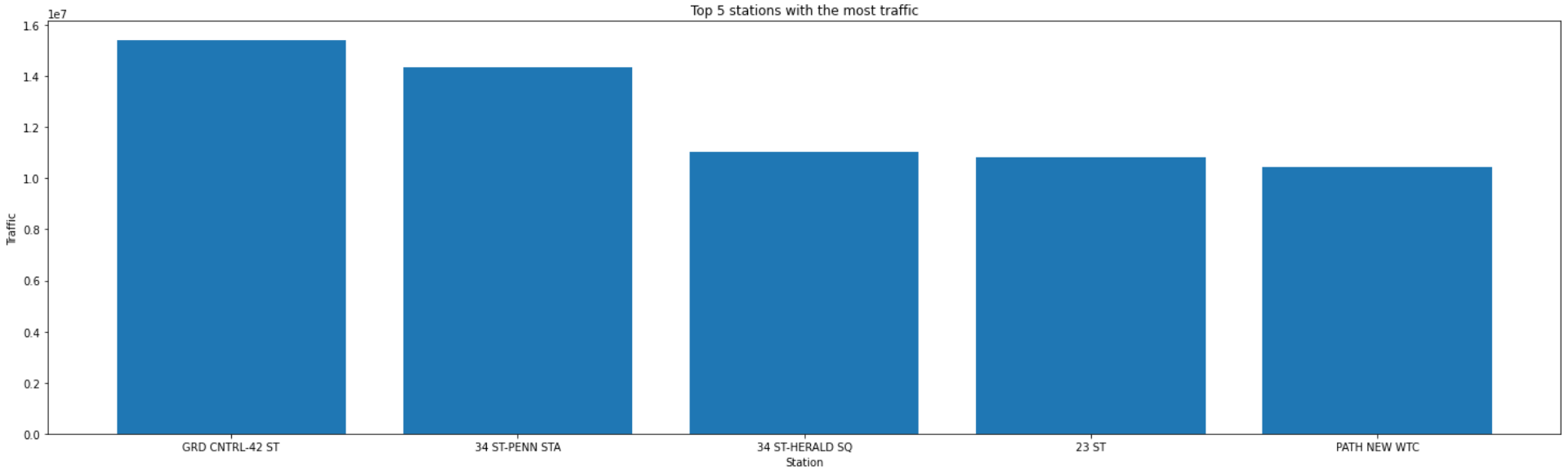
D.

Removed duplicated rows

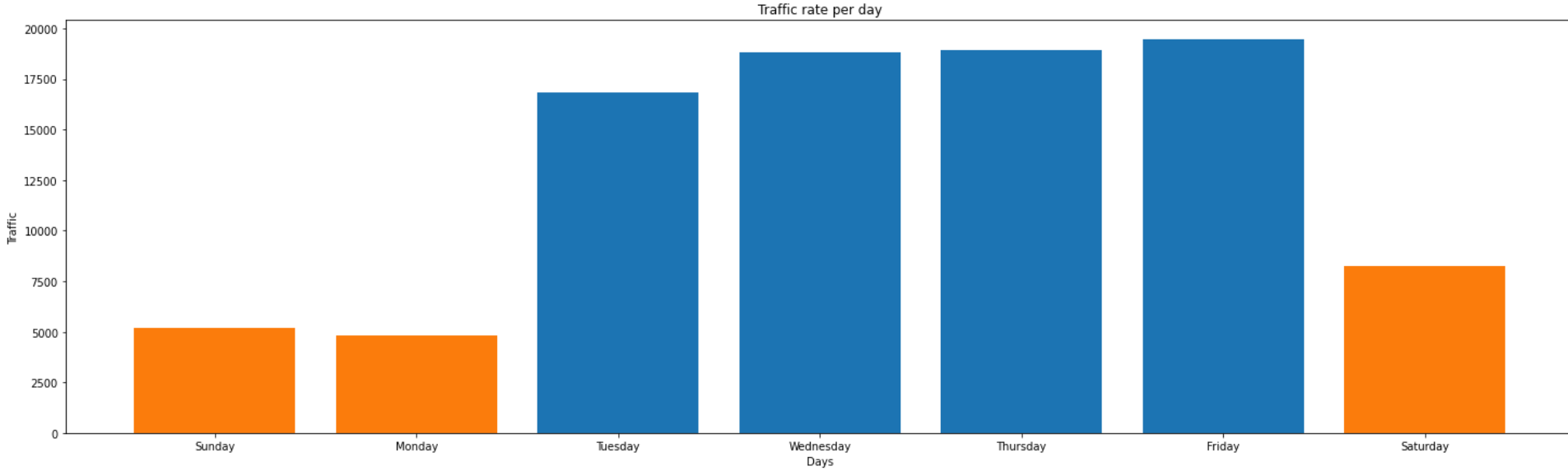
4 Data Visualization

Gaining insights from the data

Graphs


Top 5 stations with the most traffic
1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Traffic
GRD CNTRL-42 ST
34 ST-PENN STA
34 ST-HERALD SQ
23 ST
PATH NEW WTC
Station

Graphs


Traffic rate per day
Traffic
20000
17500
15000
12500
10000
7500
5000
2500
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Days

5 Conclusions

Some recommendations to solve the problem

Recommendations

Top 5 stations with the most traffic:

1. Grand Central-42 St.
2. 34th St.–Penn
3. 34th St.–Herald Square
4. 23rd St.
5. World Trade Center

Top 3 weekdays with the most traffic:

1. Friday
2. Thursday
3. Wednesday

Recommendations

I recommend WTWY to focus on placing their street teams on **Grand Central-42 St., 34th St.–Penn, 34th St.–Herald Square, 23rd St., and World Trade Center** stations on **Friday, Thursday, and Wednesday.**

Thank you!

~~No questions please~~