MTA Analysis For SoftBank Robotics

Presented by:
Amena Mohammed Algroon
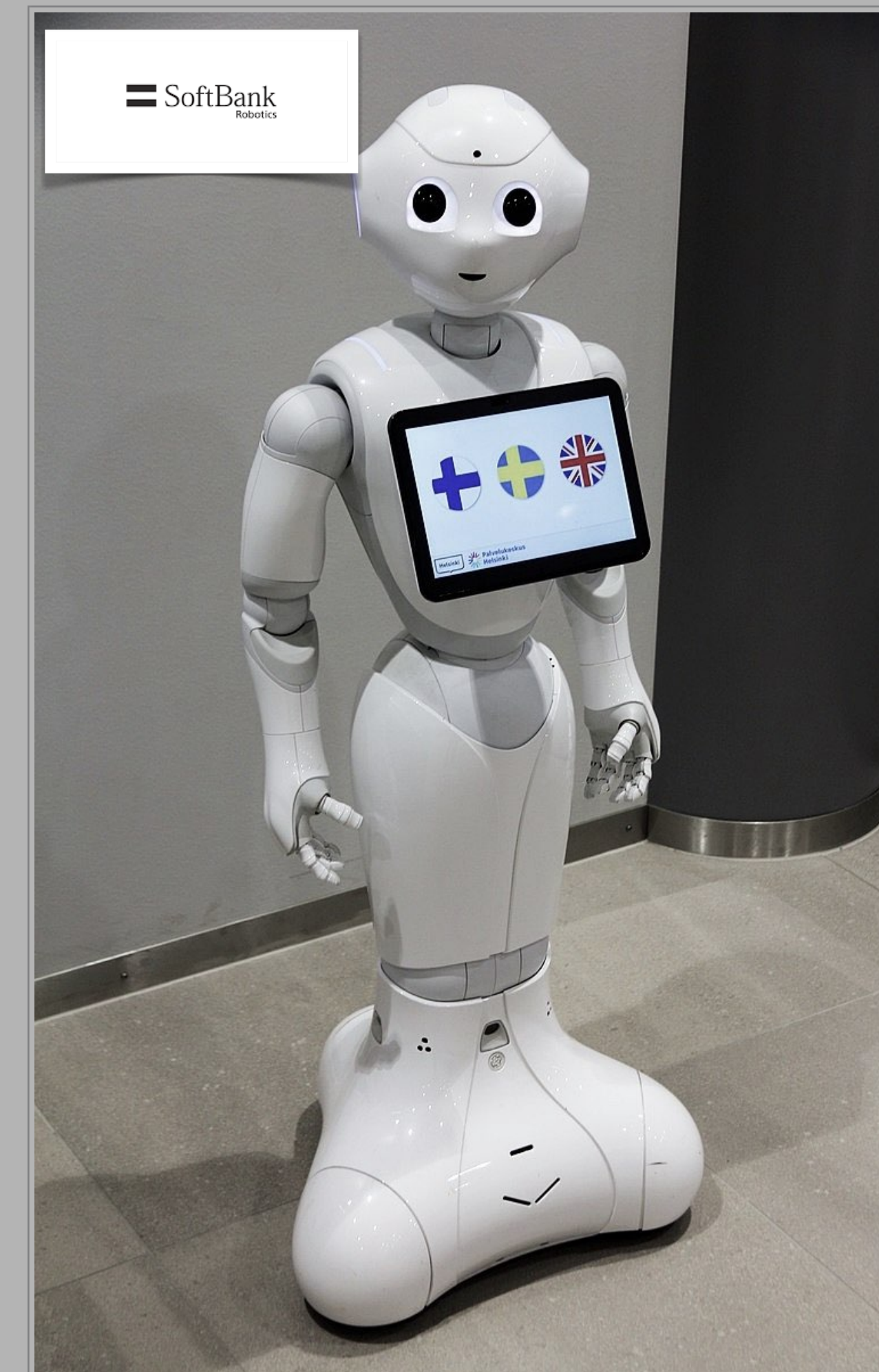
About SoftBank Robotics

Who is SoftBank Robotics?

SoftBank Robotics is constantly exploring and commercializing all robotics solutions that help make people's lives easier, safer, more connected, and more extraordinary.

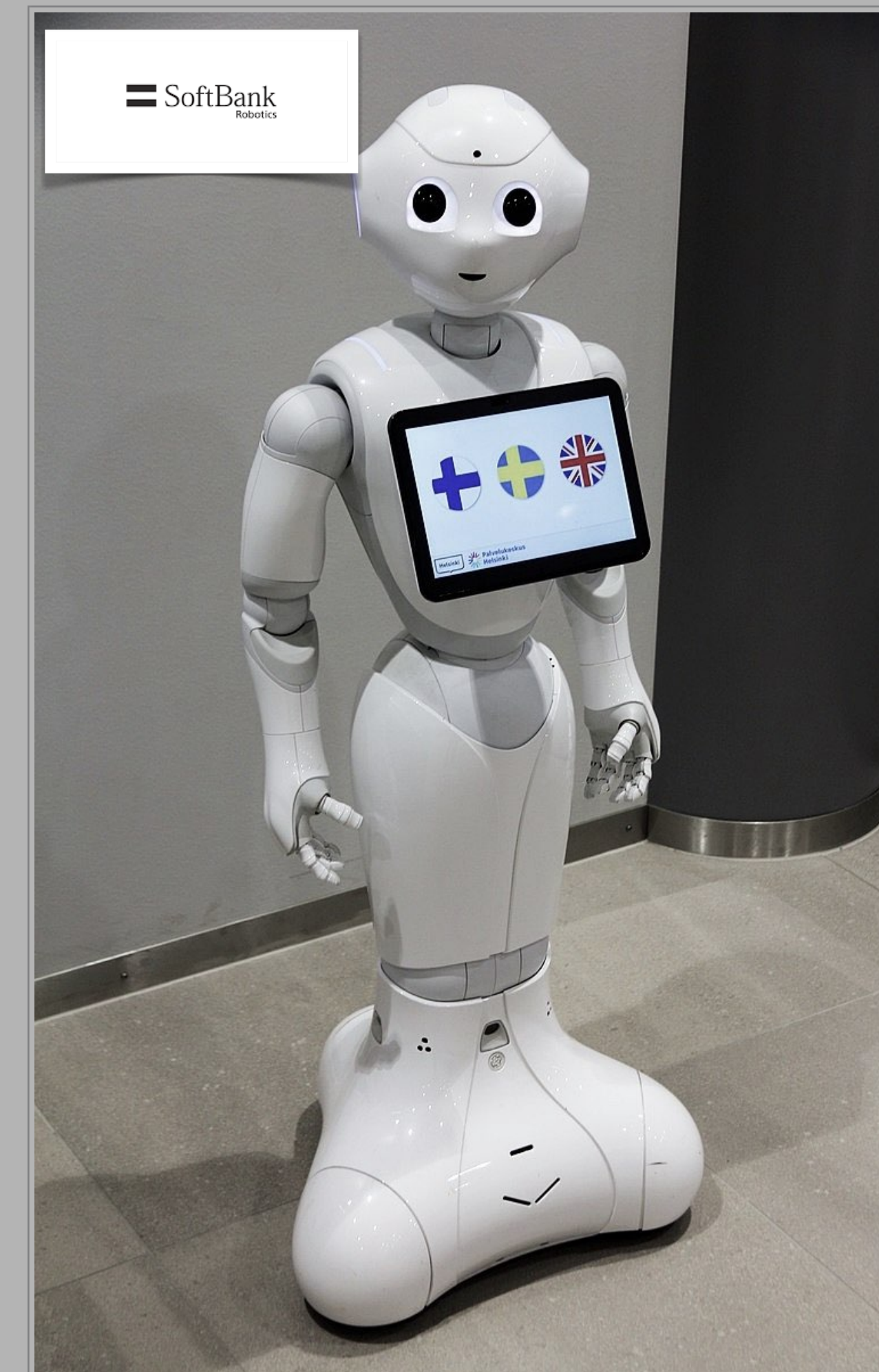# What if you get lost at the busiest station in New York City?

Who is Pepper?

The **solution** is Pepper 'humanoid' robot to guide the people who are lost in New York stations.

Pepper is the world's first social humanoid robot able to recognize faces and basic human emotions. Pepper was optimized for human interaction and is able to engage with people through conversation and his touch screen
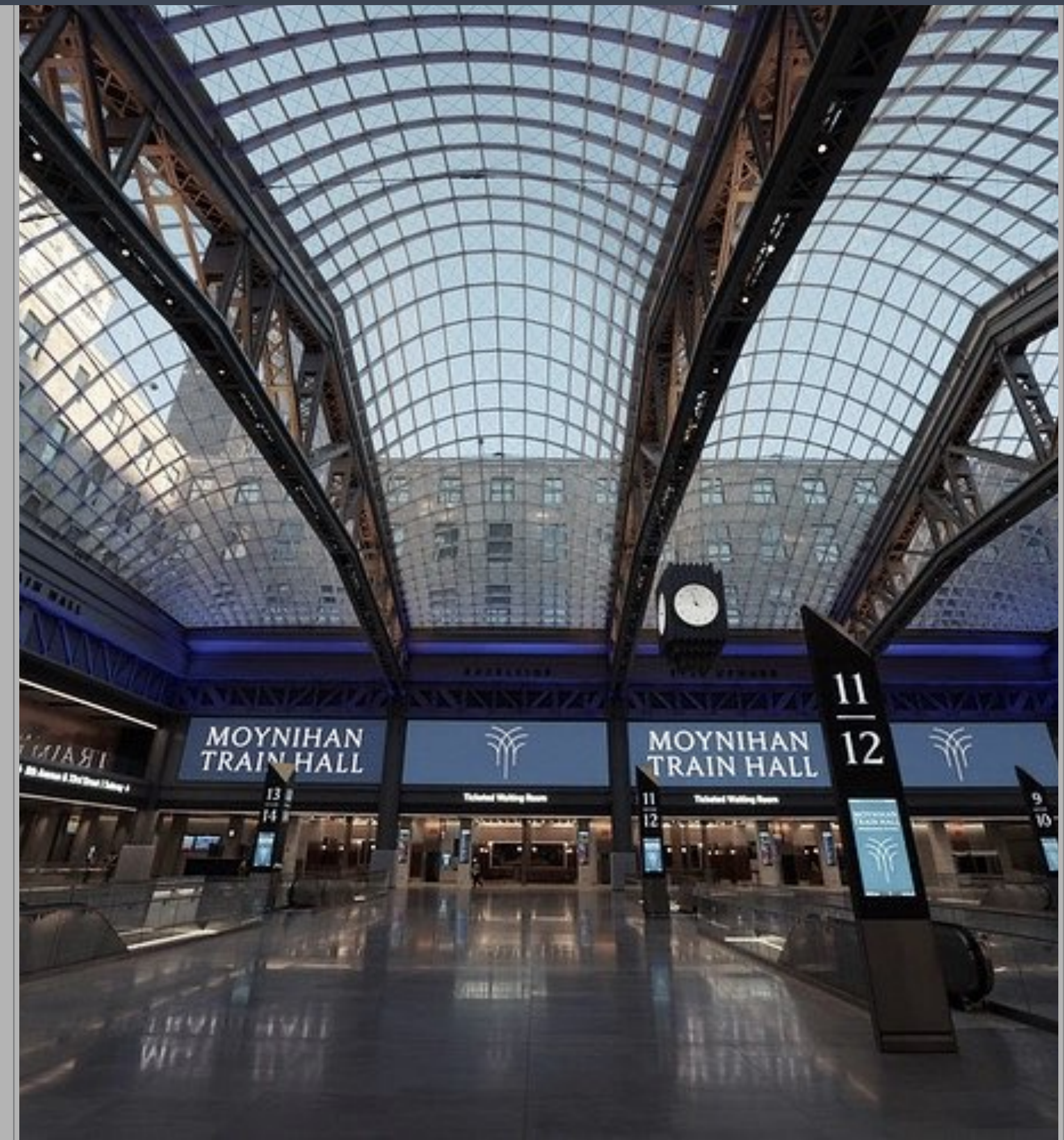
The Metropolitan Transportation Authority (MTA) Dataset

MTA Sample:

Last three months of 2021
(August - October)
-2306607 rows × 11 columns

Analysis

Tools:

SQLite, Python, Jupyter notebook.
NumPy, Pandas, Matplotlib, Seaborn, SQLAlchemy. •

```
1  df=pd.read_sql('SELECT * FROM data group by STATION order by ENTRIES limit 10 ', engine)
2  df
```

	index	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	125465	PTH17	R541	01-00-00	THIRTY THIRD ST	1	PTH	07/31/2021	00:47:54	REGULAR	0	0
1	46804	N012	R035	01-05-00	168 ST	AC1	IND	07/31/2021	00:00:00	REGULAR	46	0
2	159962	R235	R045	01-00-00	GRD CNTRL-42 ST	4567S	IRT	07/31/2021	01:00:00	REGULAR	108	866
3	131596	R106	R305	01-00-00	WTC-CORTLANDT	1	IRT	07/31/2021	01:00:00	REGULAR	144	139
4	45094	N001	R173	01-05-00	INWOOD-207 ST	A	IND	07/31/2021	01:00:00	REGULAR	593	0
5	17482	B023	R211	01-05-00	KINGS HWY	BQ	BMT	07/31/2021	00:00:00	REGULAR	892	0
6	109247	N547	R420	01-04-00	DITMAS AV	F	IND	07/31/2021	01:00:00	REGULAR	922	0
7	91121	N338	R128	01-05-00	SUTPHIN BLVD	F	IND	07/31/2021	01:00:00	REGULAR	1769	0
8	91541	N339	R114	01-05-00	PARSONS BLVD	F	IND	07/31/2021	00:00:00	REGULAR	2026	0
9	128533	PTH22	R540	00-00-00	PATH NEW WTC	1	PTH	07/31/2021	01:08:09	REGULAR	2792	13021

Analysis
Data Cleaning

Check nulls(No nulls)

	C/A	STATION	DATE	TIME	ENTRIES	EXITS
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
209879	False	False	False	False	False	False
209880	False	False	False	False	False	False
209881	False	False	False	False	False	False
209882	False	False	False	False	False	False
209883	False	False	False	False	False	False

2306607 rows × 6 columns

Found duplicate rows

Duplicate Data

```
1
2  #print('\nNumber of rows before discarding duplicates = %d' % (data_df.shape[0]))
```

```
1  #dup = data_df.ENTRIES.duplicated().sum()
2  #print('\nNumber of duplicate rows = %d' % (dup))
```

```
1  dups = data_df.duplicated().sum()
2  print('\nNumber of duplicate rows = %d' % (dups))
3  print('\nNumber of rows before discarding duplicates = %d' % (data_df.shape[0]))
4  data_df2= data_df.drop_duplicates()
5  print('Number of rows after discarding duplicates = %d' % (data_df2.shape[0]))
```
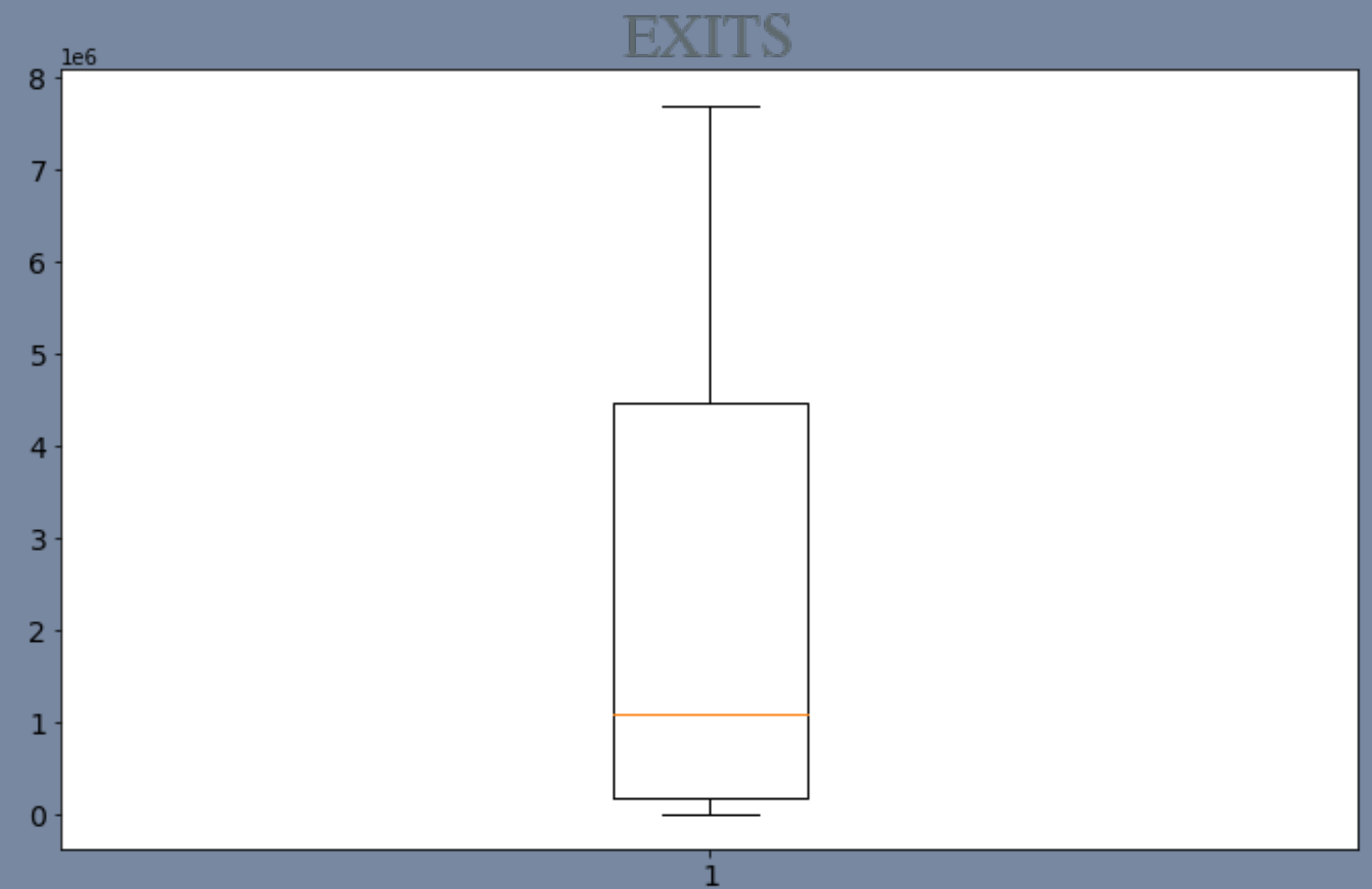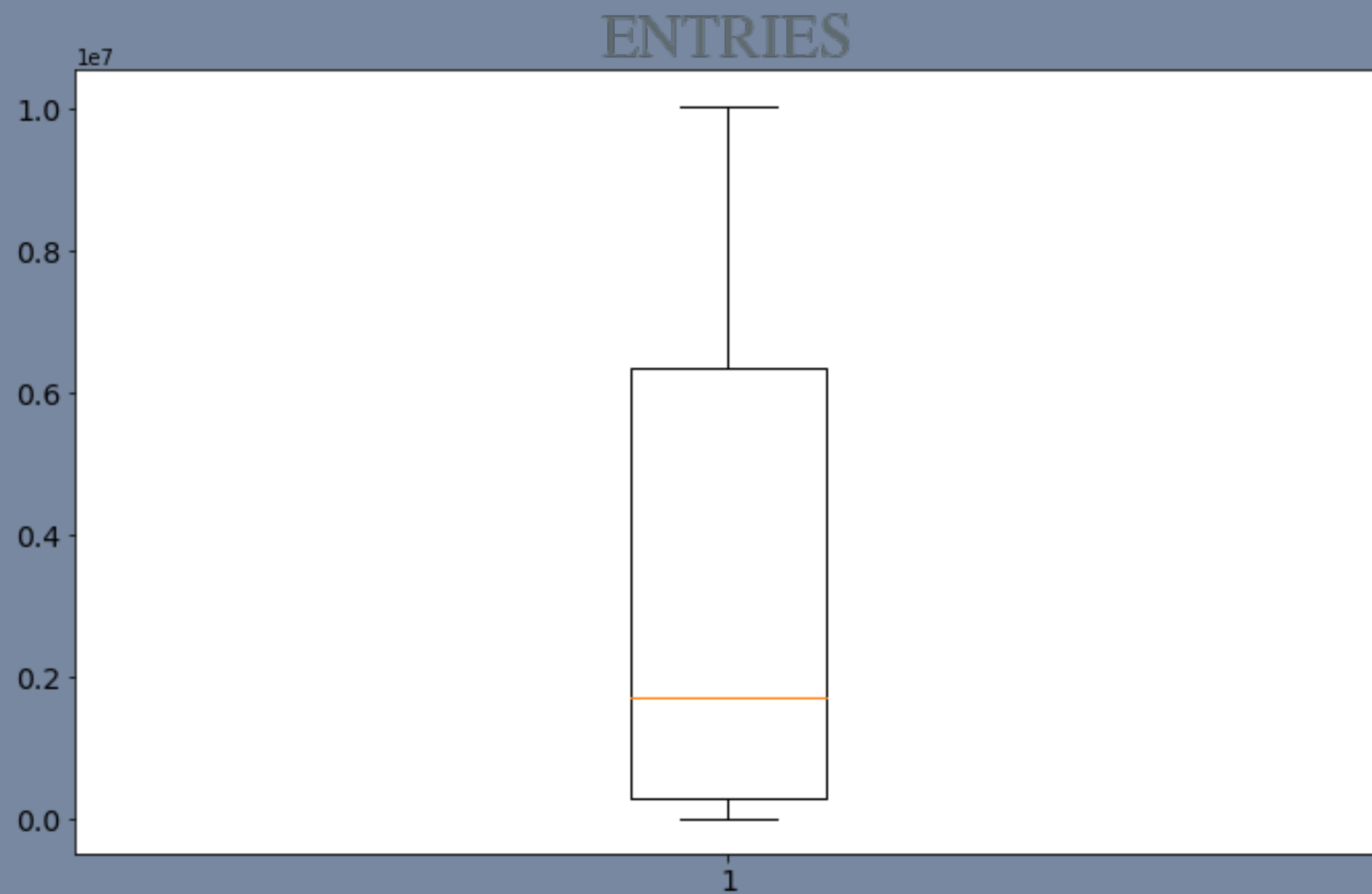
```
Number of duplicate rows = 10

Number of rows before discarding duplicates = 2306607
Number of rows after discarding duplicates = 2306597
```

Analysis
Data Cleaning

Outliers:

Analysis
Data Cleaning

Add new columns:

	C/A	STATION	DATE	TIME	ENTRIES	EXITS	DATE_TIME	MONTH	DAY_WEEK	Entry_Inc	Exits_Inc	TOTAL_PEOPLE
24	A002	59 ST	09/01/2021	00:00:00	7628030	2609899	2021-09-01 00:00:00	9	Wednesday	83.0	21.0	104.0
25	A002	59 ST	09/01/2021	04:00:00	7628037	2609904	2021-09-01 04:00:00	9	Wednesday	7.0	5.0	12.0
26	A002	59 ST	09/01/2021	08:00:00	7628051	2609949	2021-09-01 08:00:00	9	Wednesday	14.0	45.0	59.0
27	A002	59 ST	09/01/2021	12:00:00	7628112	2610083	2021-09-01 12:00:00	9	Wednesday	61.0	134.0	195.0
28	A002	59 ST	09/01/2021	16:00:00	7628256	2610111	2021-09-01 16:00:00	9	Wednesday	144.0	28.0	172.0
...	...	...	...	...	...	...	...	...	...	...	...	...
210200	TRAM2	RIT-ROOSEVELT	09/30/2021	05:00:00	5554	649	2021-09-30 05:00:00	9	Thursday	0.0	0.0	0.0
210201	TRAM2	RIT-ROOSEVELT	09/30/2021	09:00:00	5554	649	2021-09-30 09:00:00	9	Thursday	0.0	0.0	0.0
210202	TRAM2	RIT-ROOSEVELT	09/30/2021	13:00:00	5554	649	2021-09-30 13:00:00	9	Thursday	0.0	0.0	0.0
210203	TRAM2	RIT-ROOSEVELT	09/30/2021	17:00:00	5554	649	2021-09-30 17:00:00	9	Thursday	0.0	0.0	0.0
210204	TRAM2	RIT-ROOSEVELT	09/30/2021	21:00:00	5554	649	2021-09-30 21:00:00	9	Thursday	0.0	0.0	0.0

- Added column DATE_TIME to data is a date column + time column with Change the format %m/%d/%Y %H:%M:%S.

- Also added column month and column day_week to return the month and day.

- Sum the change of entry and exit to get the total people for each timing.

- Then dropped null values from the dataset.

Results

- 34 ST - Penn Station and 34 ST - Herald SQ Station has more visitors than the rest.
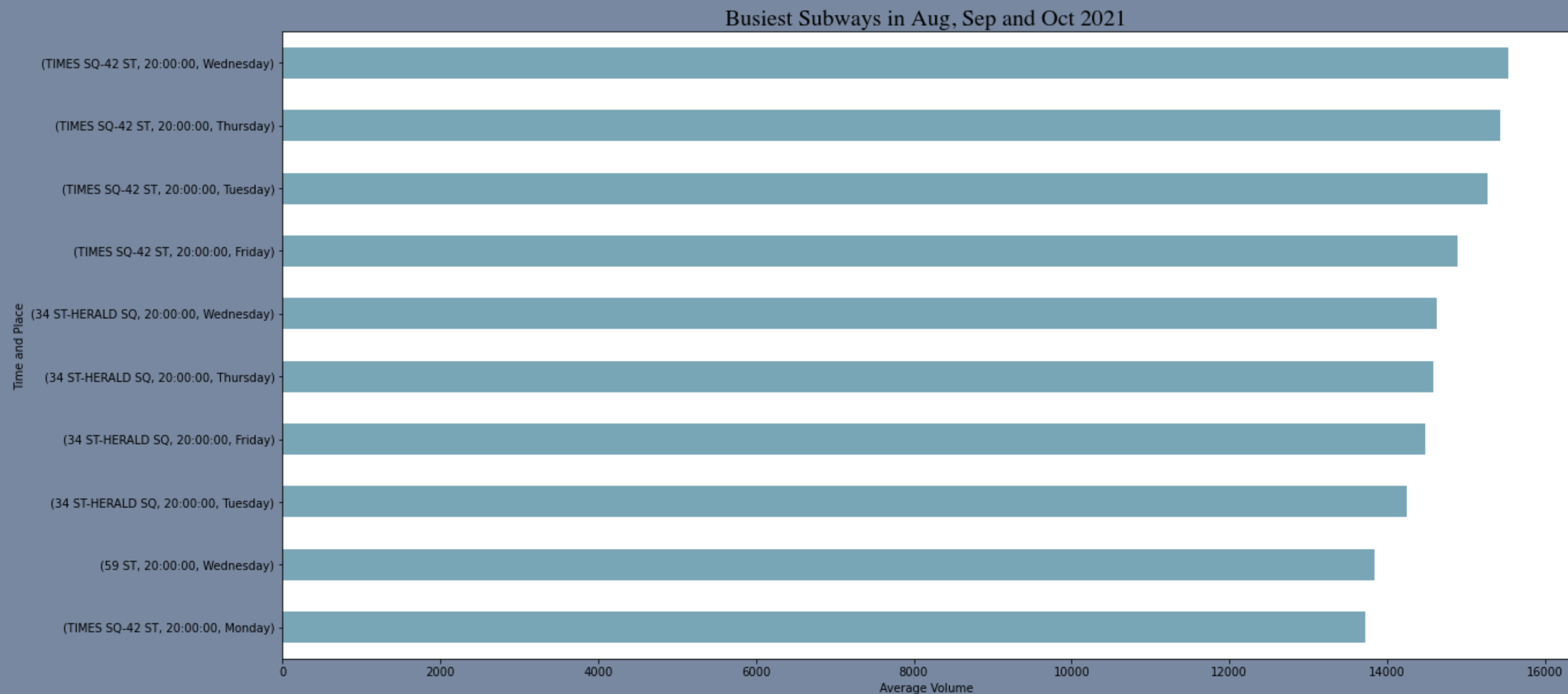

Top 10 Stations by PEOPLE in October 2021

Analysis

- Busiest Subways by the day and time is the TIMES SQ-42 ST at 20:00 on Wednesday has more **visitors** than the rest.

Busiest Subways in Aug, Sep and Oct 2021

In Conclusion

- Given the previous data, MTA officials should focus on increasing the number of robots

 at the 10 busiest MTA stations.

- If time is a limitation, they should focus on weekdays in the late afternoon to late evening between 16:00-20:00.

- Also focus on Wednesday due to the high visitors.