MRT3 - Optimal Scheduling

Group 9

Hauschild, Cruz, Roxas, Rejante, and Pedernal

PROBLEM STATEMENT

In this analysis, our primary focus is on unraveling the intricate patterns of passenger traffic over time across all MRT-3 stations.

Leveraging the timestamp data per station, along with insights into the busiest station and its corresponding high-traffic times, our analysis is poised to deliver valuable recommendations for enhancing the overall efficiency and responsiveness of the MRT-3 system

DATASETS

Our research is based on the dataset DOTC-MRT3 HOURLY RIDERSHIP REPORT from the Freedom of Information website (FOI). It compiles 17,520 observations collected from the whole year of 2022 . These 18 factors listed below are then employed as explanatory variables.

- ID: Unique identification for each datapoint of entry/exit.
- Date: Date of the collected data from 1 January 2022 to 31 December 2022
- Start Time: Start of the ridership duration
- End Time: End of the ridership duration
- Type: Categorical type of ridership (entry/exit)
- Station Name for MRT-3

PREPROCESS

Data set from FOI was presented in a unique way where the columns are all stacked together

DOTC-MRT3 HOURLY RIDERSHIP REPORT

Saturday, August 1, 2020

TIME	North Ave		Quezon Ave	
	Entry	Exit	Entry	Exit
03:00 - 03:59				
04:00 - 04:59	71	0	75	1
05:00 - 05:59	250	0	149	2

"Transformed" panel data

ID	Date	Start Time	End Time	Type	North Ave	Quezon Ave
4456203:00 - 03:59	1/1/2022	3:00:00	03:59:59	Entry		
4456203:00 - 03:59	1/1/2022	3:00:00	03:59:59	Exit		
4456204:00 - 04:59	1/1/2022	4:00:00	04:59:59	Entry	0	0
4456204:00 - 04:59	1/1/2022	4:00:00	04:59:59	Exit	0	0
4456205:00 - 05:59	1/1/2022	5:00:00	05:59:59	Entry	0	0

DEV TOOLS

Data manipulation libraries	Pandas
	Numpy
	Date time from time module
Data visualization libraries	Matplotlib
	Seaborn

DATA PREPROCESSING

Data Extraction

	ID	Date	Start Time	End Time	Type	North Ave	Quezon Ave	GMA Kamuning	Cubao	Santolan	Ortigas	Shaw Blvd	Boni Ave	Guadalupe	Buendia	Ayala Ave	Magallanes	Taft
0	4456203:00 - 03:59	2022-01-01 00:00:00	03:00:00	03:59:59	Entry	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	4456203:00 - 03:59	2022-01-01 00:00:00	03:00:00	03:59:59	Exit	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	4456204:00 - 04:59	2022-01-01 00:00:00	04:00:00	04:59:59	Entry	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4456204:00 - 04:59	2022-01-01 00:00:00	04:00:00	04:59:59	Exit	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4456205:00 - 05:59	2022-01-01 00:00:00	05:00:00	05:59:59	Entry	0	0	0	0	0	0	0	0	0	0	0	0	0

Data contains all-null values for the station. Initial check with the dataframe suggests some data cleaning process to be undertaken.

DATA PREPROCESSING

Unique and Null Values Checking

No significant findings in checking the unique values within the dataframe. However, as expected in the initial scan, only the station columns have null values. It needs more investigation.

	columns	unique values
0	ID	8760
1	Date	365
2	Start Time	24
3	End Time	24
4	Type	2
5	North Ave	2946
6	Quezon Ave	1438
7	GMA Kamuning	2878
8	Cubao	3668
9	Santolan	1087
10	Ortigas	2787
11	Shaw Blvd	3749
12	Boni Ave	2284
13	Guadalupe	2493
14	Buendia	1425
15	Ayala Ave	2968
16	Magallanes	2191
17	Taft	3472

	columns	null values
0	ID	False
1	Date	False
2	Start Time	False
3	End Time	False
4	Type	False
5	North Ave	True
6	Quezon Ave	True
7	GMA Kamuning	True
8	Cubao	True
9	Santolan	True
10	Ortigas	True
11	Shaw Blvd	True
12	Boni Ave	True
13	Guadalupe	True
14	Buendia	True
15	Ayala Ave	True
16	Magallanes	True
17	Taft	True

DATA PREPROCESSING

Unique and Null Values Checking

Upon molding the data, these are the count of null rows in each time duration and type. There is an unusual frequency of null rows. This is due to the fact that these time durations are beyond MRT3's normal operating hours.

Dates were also identifed where MRT3 did not transport passengers for a whole day. These are the same dates of the 2022 Holy Week.

	Start Time	End Time	Type	ID
2	01:00:00	01:59:59	Entry	364
3	01:00:00	01:59:59	Exit	363
6	03:00:00	03:59:59	Entry	363
7	03:00:00	03:59:59	Exit	363
4	02:00:00	02:59:59	Entry	360
5	02:00:00	02:59:59	Exit	360
45	23:00:00	23:59:59	Entry	354
1	00:00:00	00:59:59	Exit	341
46	23:00:00	23:59:59	Exit	312
0	00:00:00	00:59:59	Entry	263
36	18:00:00	18:59:59	Entry	10
44	22:00:00	22:59:59	Exit	8
43	22:00:00	22:59:59	Entry	8

	Date	ID
102	2022-04-13	48
106	2022-04-17	48
105	2022-04-16	48
104	2022-04-15	48
103	2022-04-14	48

DATA PREPROCESSING

Columns and Data Type Checking

No significant findings with the naming of the columns.

```
Index(['ID', 'Date', 'Start Time', 'End Time', 'Type', 'North Ave',
       'Quezon Ave', 'GMA Kamuning', 'Cubao', 'Santolan', 'Ortigas',
       'Shaw Blvd', 'Boni Ave', 'Guadalupe', 'Buendia', 'Ayala Ave',
       'Magallanes', 'Taft'],
      dtype='object', name=0)
```

For the data type, everything seems to be in the proper dtype. However, if you check with the Start Time and End Time columns, they have different classes. The End Time column is of *string* data type while the Start Time column is in its proper data type which is *datetime.time.*

```
The data type of Start Time is <class 'datetime.time'>
The data type of End Time is <class 'str'>
```

```
ID                      object
Date             datetime64[ns]
Start Time              object
End Time                object
Type                    object
North Ave              float64
Quezon Ave             float64
GMA Kamuning           float64
Cubao                  float64
Santolan               float64
Ortigas                float64
Shaw Blvd              float64
Boni Ave               float64
Guadalupe              float64
Buendia                float64
Ayala Ave              float64
Magallanes             float64
Taft                   float64
dtype: object
```

DATA CLEANING

Identified Cleanup Procedures

Based on the identified observations, the following cleanup procedure was applied to the dataset:

- Properly typecast the End Time column
- Drop rows beyond MRT3 operational hours (4:30:00 to 22:30:00)
- Replace null values with zero

Target Market

(Who will benefit from this research?)

TARGET MARKET

The researchers have identified that the optimal scheduling research will benefit two specific target markets, namely:

- MRT Management
- MRT Passengers (In Totals)

EDA FLOW

MONTHLY TRENDS

DAILY TRENDS

HOURLY TRENDS

STATION TRENDS

SYNTHESIS

Adding New Features

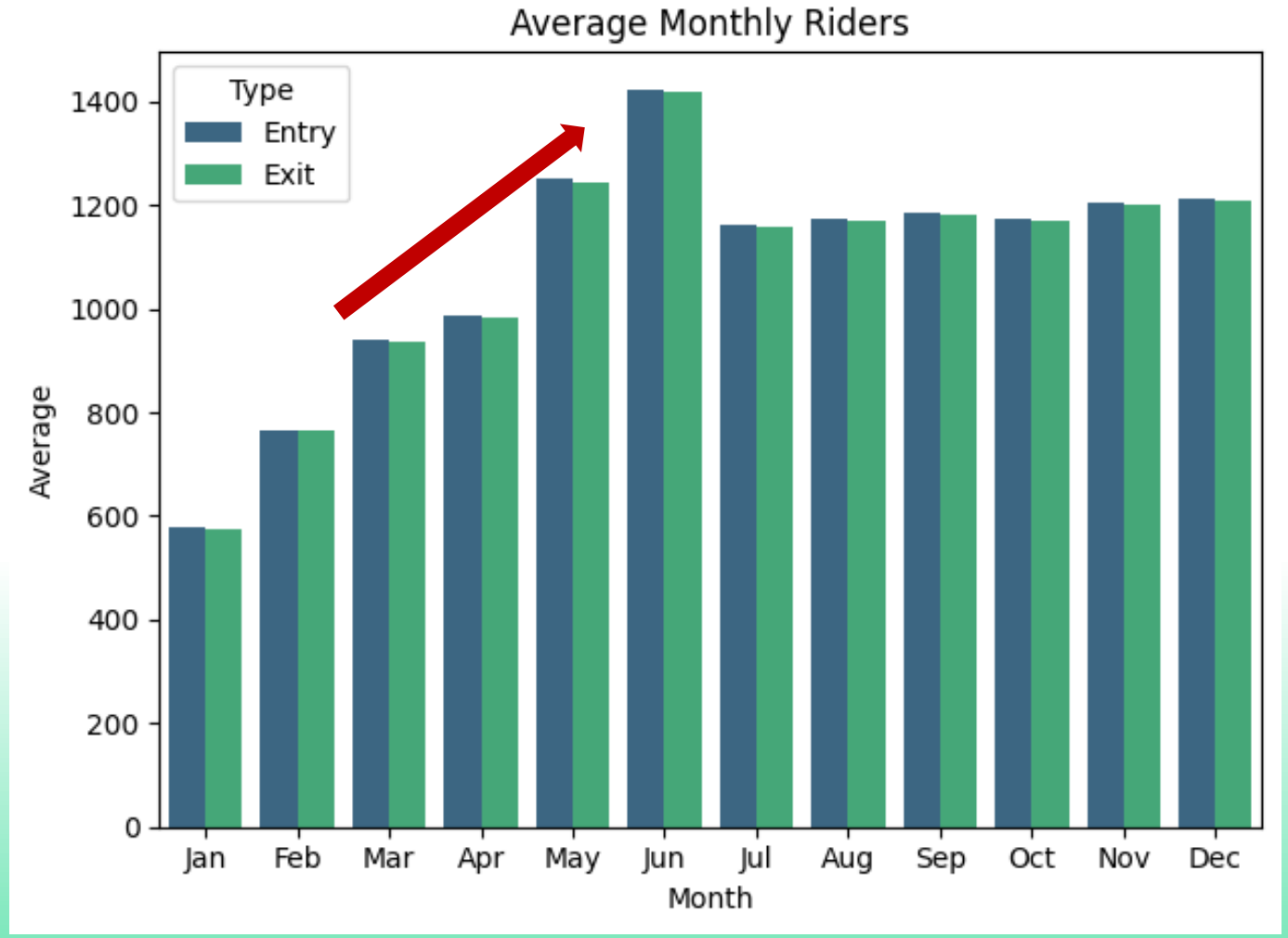
AVERAGE, TOTALS, MONTHS, DAYS OF THE WEEK

	ID	Date	Start Time	End Time	Type	North Ave	Quezon Ave	GMA Kamuning	Cubao	Santolan	...	Boni Ave	Guadalupe	Buendia	Ayala Ave	Magallanes	Taft	Average	Totals	Month	Day of Week
2	4456204:00 - 04:59	2022-01-01	04:00:00	04:59:59	Entry	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Jan	Sat
3	4456204:00 - 04:59	2022-01-01	04:00:00	04:59:59	Exit	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Jan	Sat
4	4456205:00 - 05:59	2022-01-01	05:00:00	05:59:59	Entry	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Jan	Sat
5	4456205:00 - 05:59	2022-01-01	05:00:00	05:59:59	Exit	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Jan	Sat
6	4456206:00 - 06:59	2022-01-01	06:00:00	06:59:59	Entry	227.0	177.0	453.0	482.0	27.0	...	150.0	227.0	58.0	199.0	221.0	525.0	258	3357	Jan	Sat
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
17507	4492620:00 - 20:59	2022-12-31	20:00:00	20:59:59	Entry	0.0	8.0	32.0	101.0	49.0	...	135.0	325.0	44.0	248.0	83.0	439.0	158	2060	Dec	Sat
17508	4492621:00 - 21:59	2022-12-31	21:00:00	21:59:59	Exit	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	1	Dec	Sat
17509	4492621:00 - 21:59	2022-12-31	21:00:00	21:59:59	Entry	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Dec	Sat
17510	4492622:00 - 22:59	2022-12-31	22:00:00	22:59:59	Exit	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Dec	Sat
17511	4492622:00 - 22:59	2022-12-31	22:00:00	22:59:59	Entry	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0	0	Dec	Sat

13870 rows × 22 columns

Monthly Trends

MRT-3 LIBRENG SAKAY PROGRAM INCREASED RIDERSHIP



SERBISYO PUBLIKO

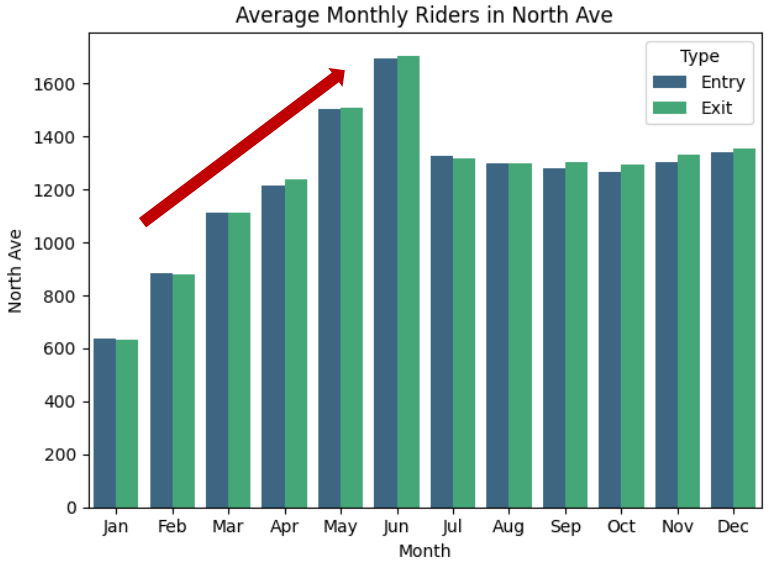
Filtered By: Serbisyopubliko

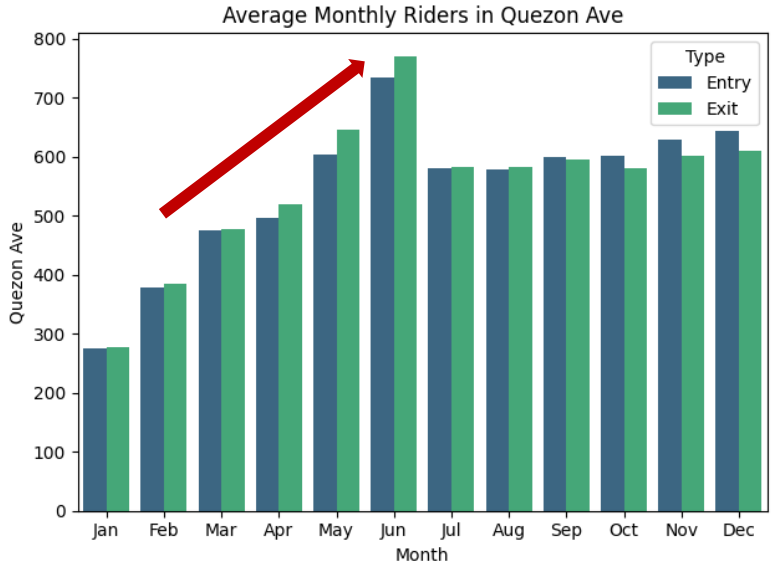
MRT3 extends anew free rides until June 30

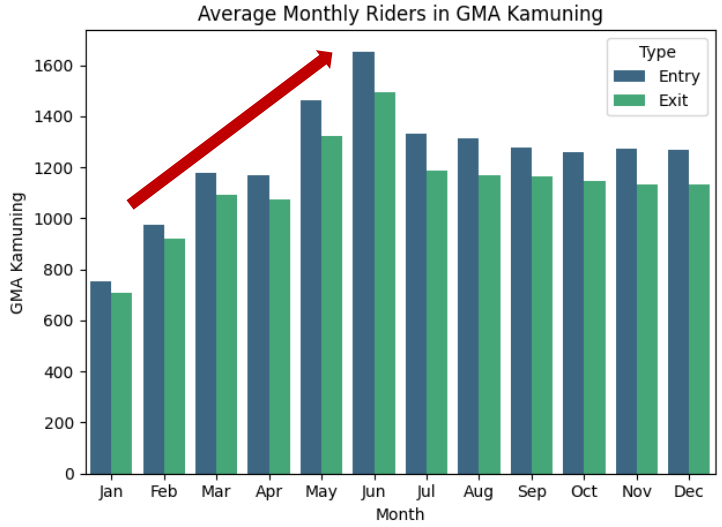
By TED CORDERO, GMA News

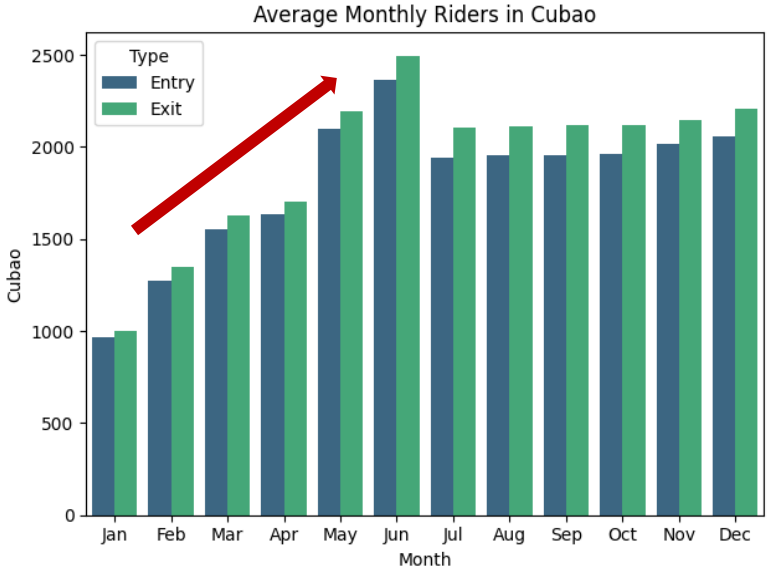
Published May 25, 2022 10:54am
Updated May 25, 2022 11:54am

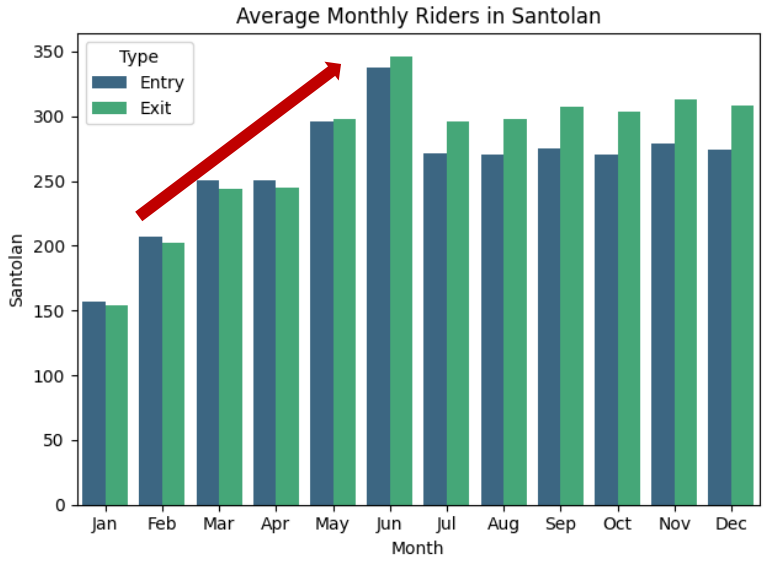
MRT-3 LIBRENG SAKAY PROGRAM INCREASED RIDERSHIP

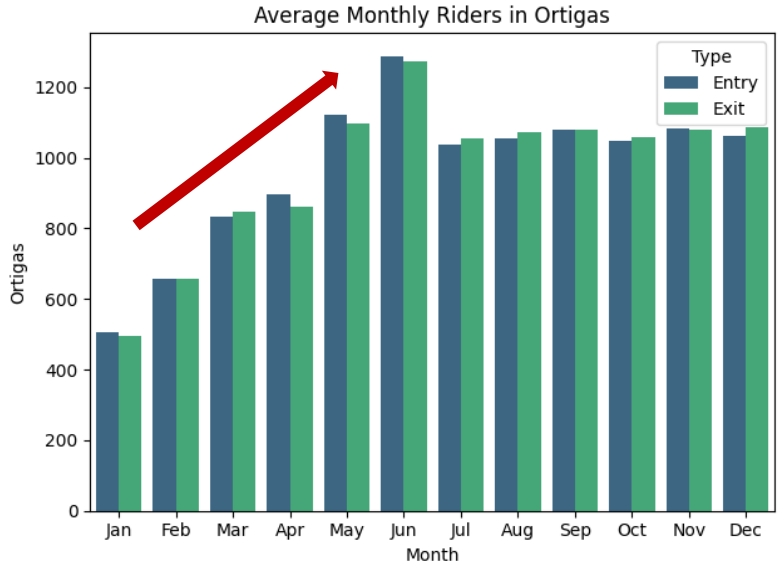




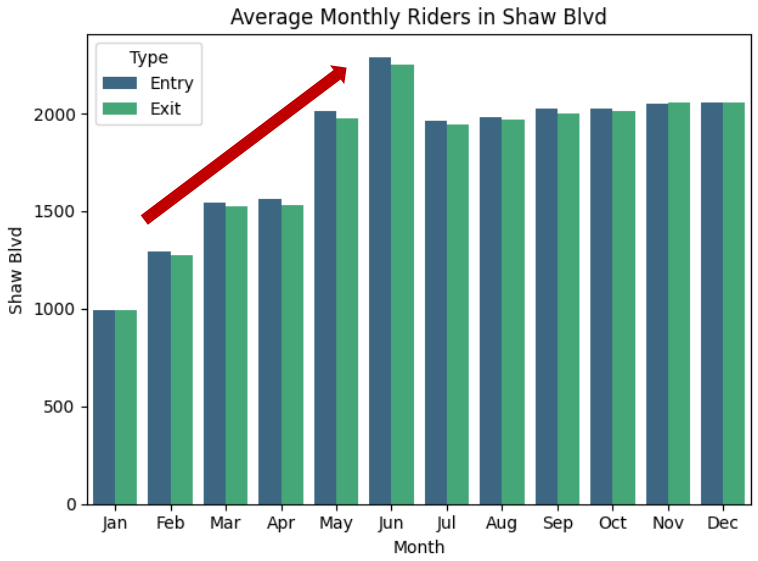


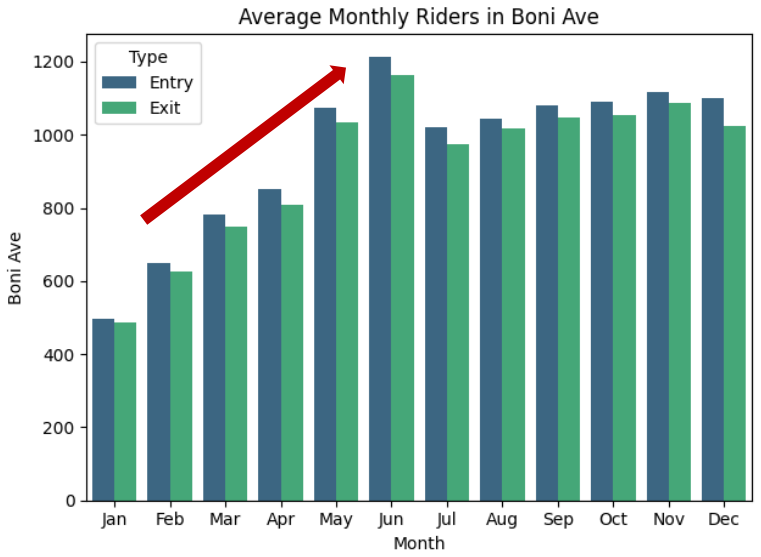


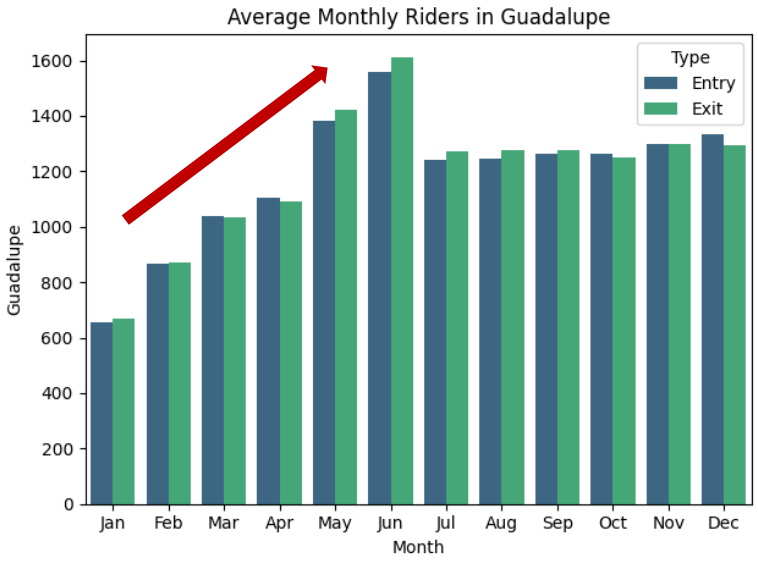


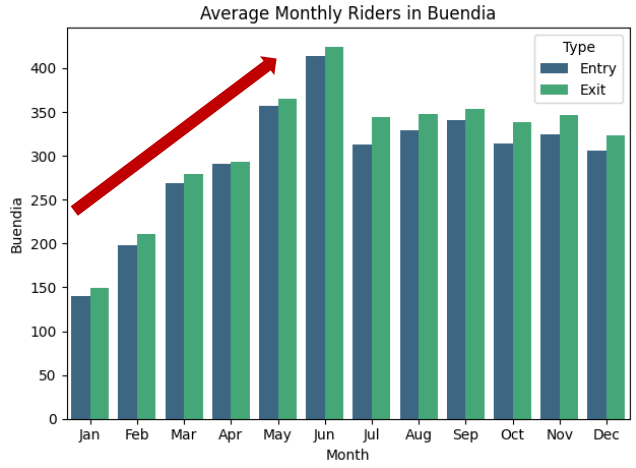


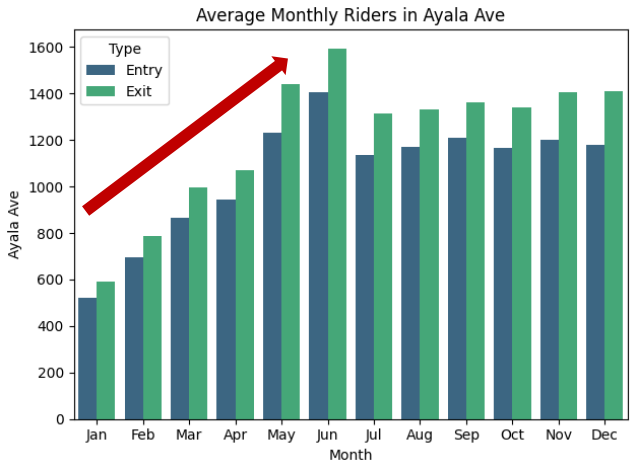
MRT-3 LIBRENG SAKAY PROGRAM INCREASED RIDERSHIP

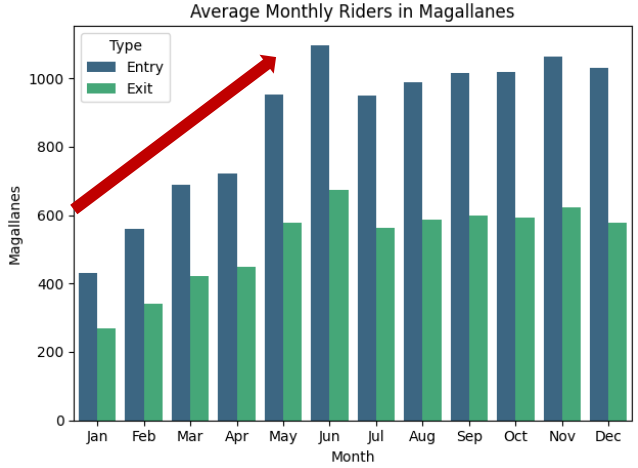


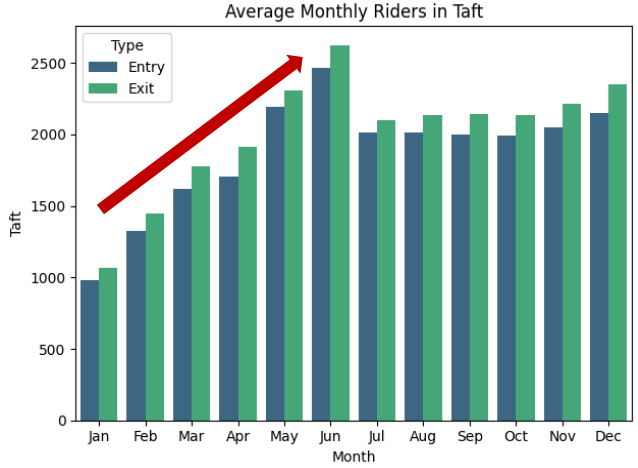




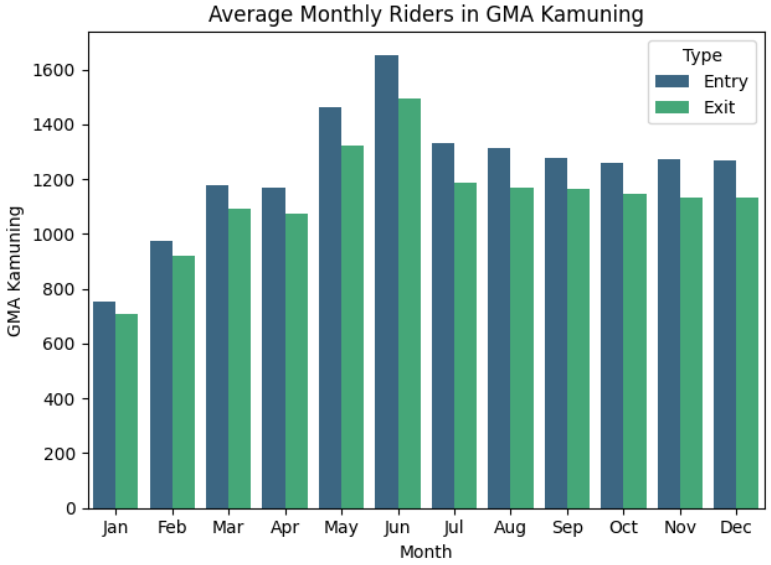


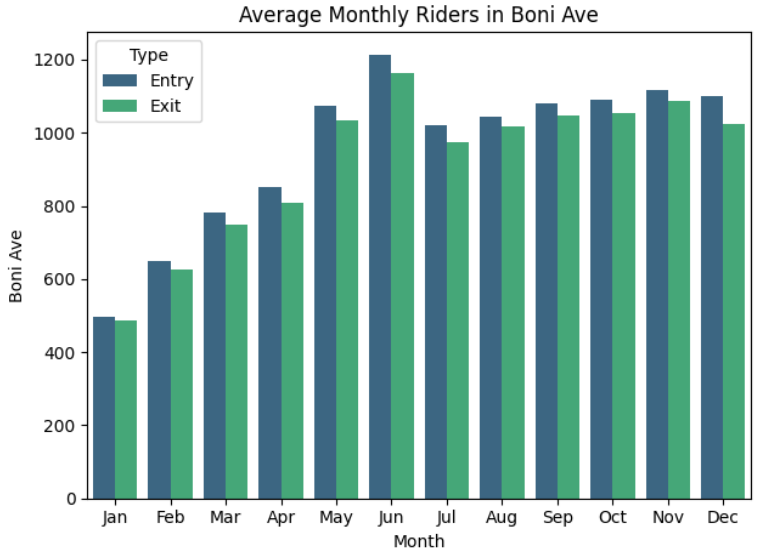


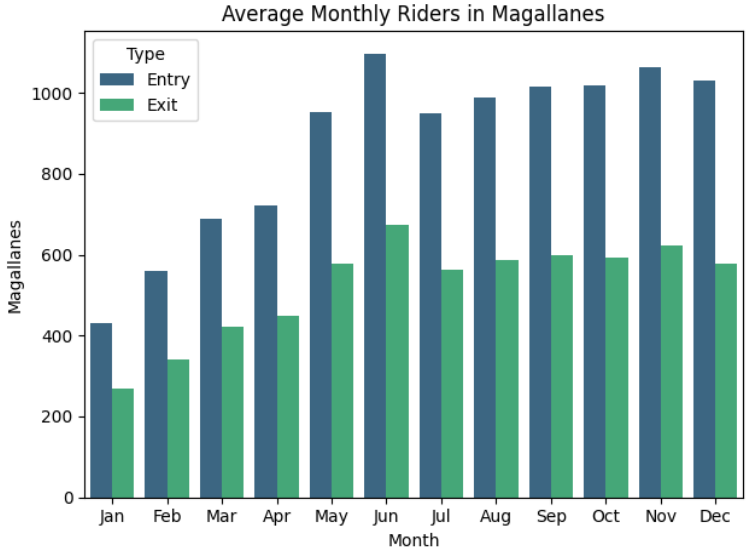


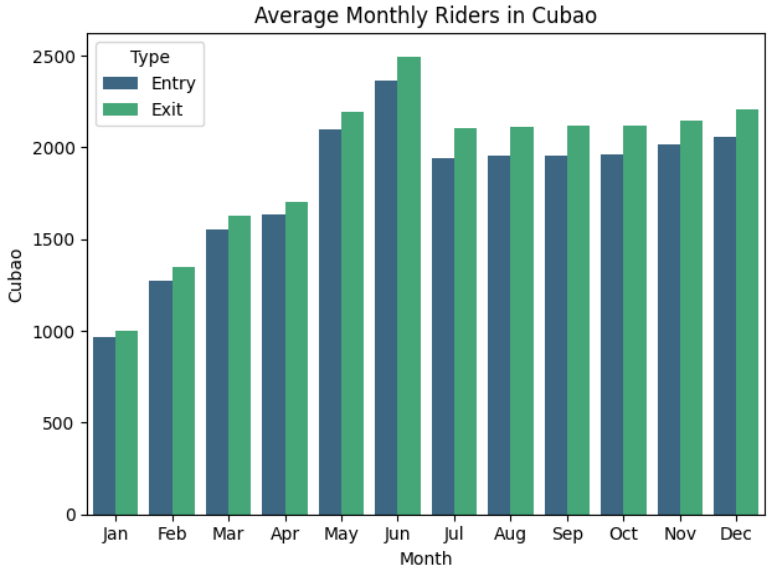


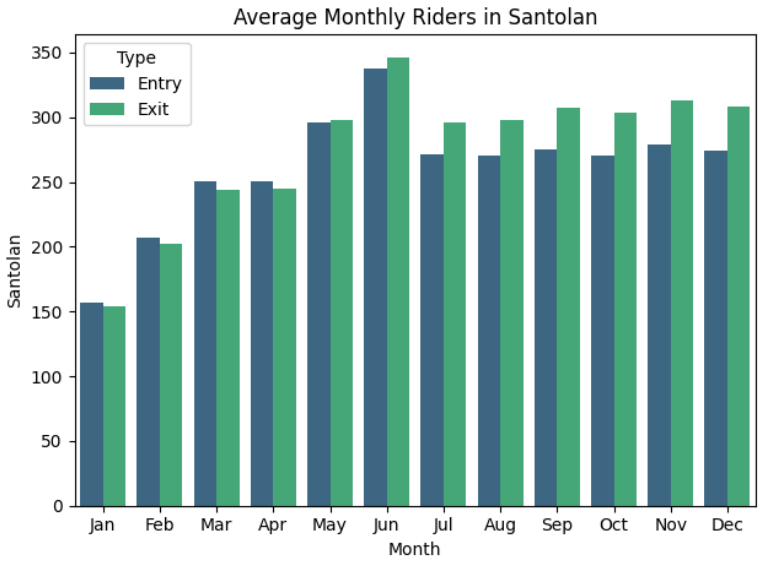
SOME STATIONS ARE MORE CONGESTED THAN OTHERS

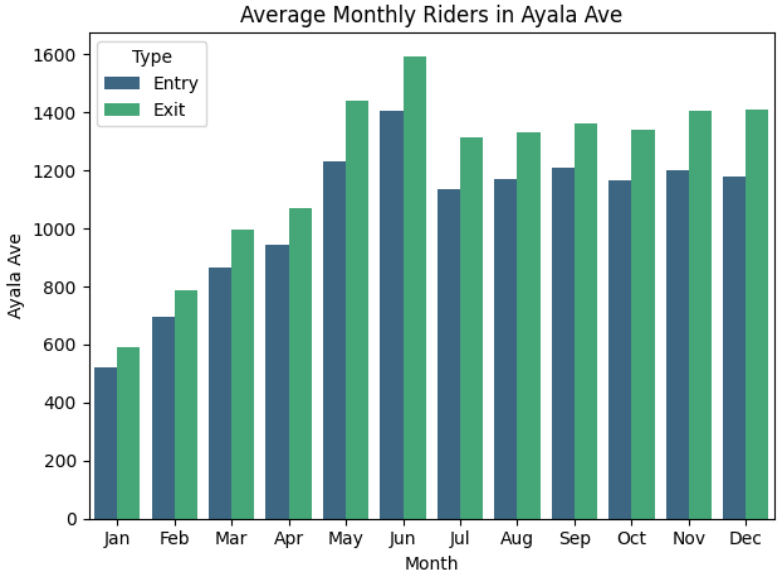






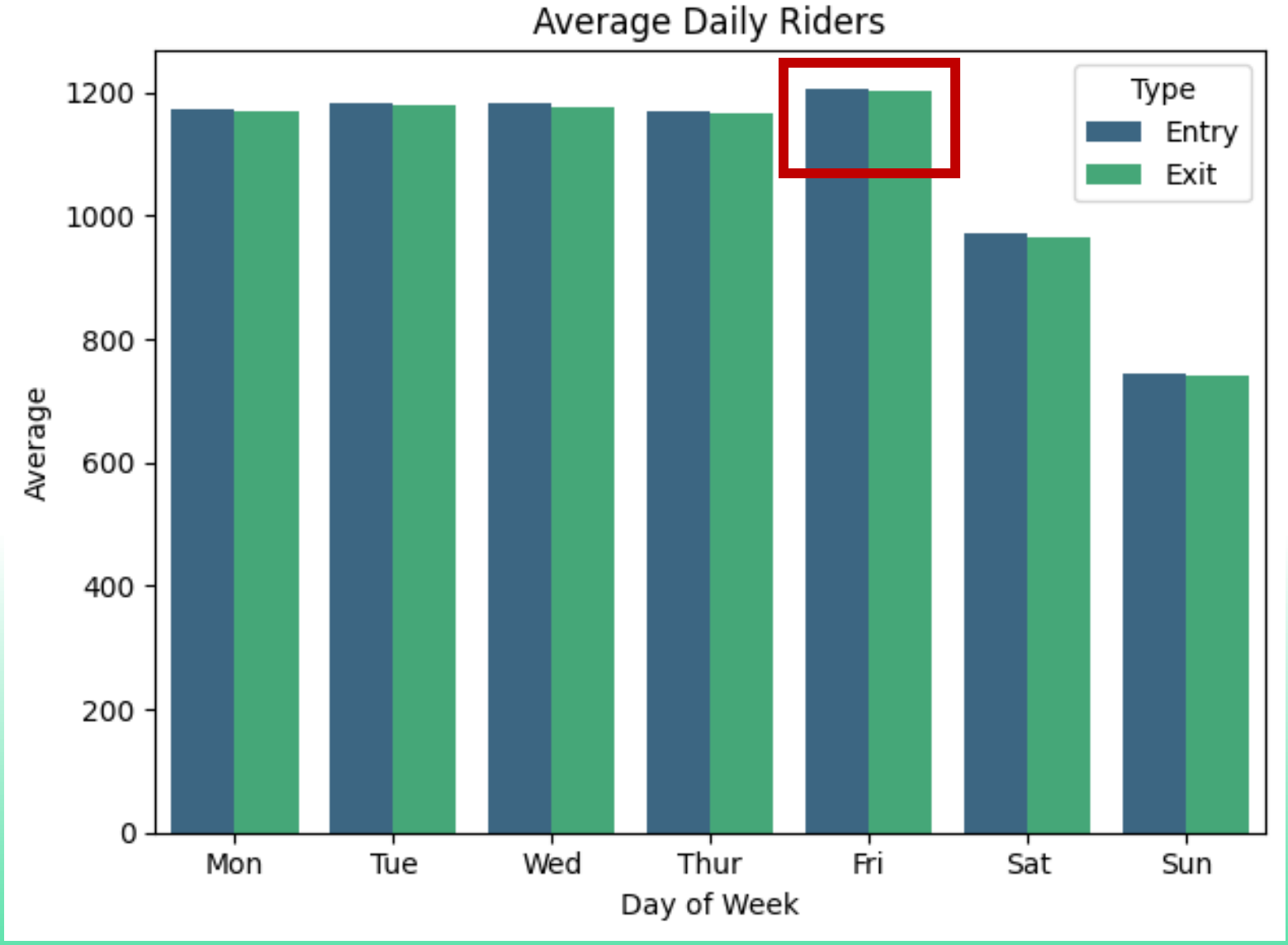




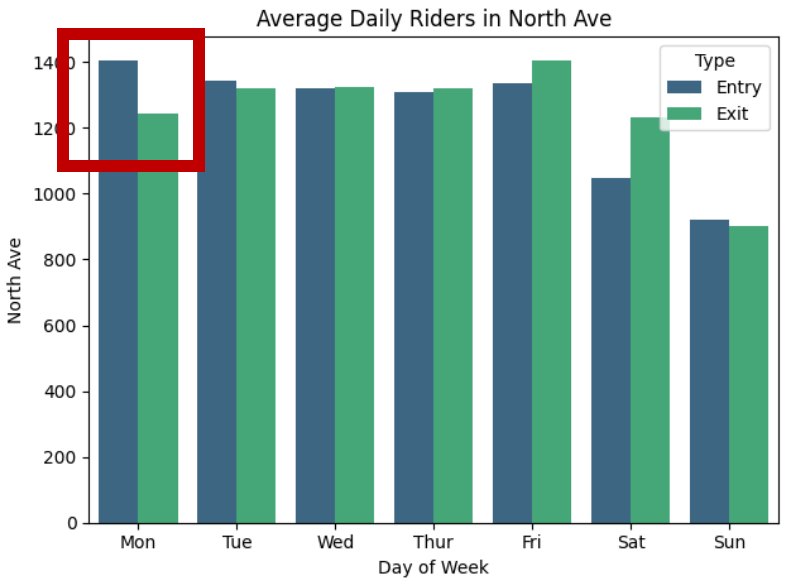


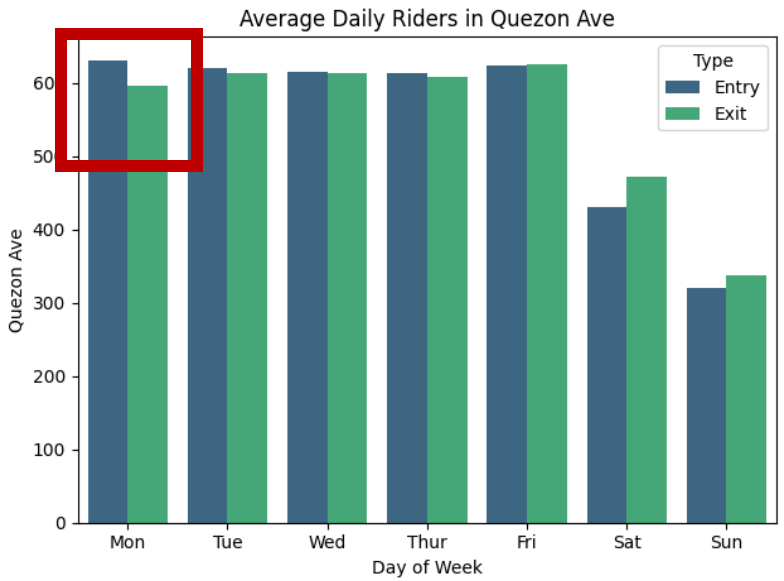
Daily Trends

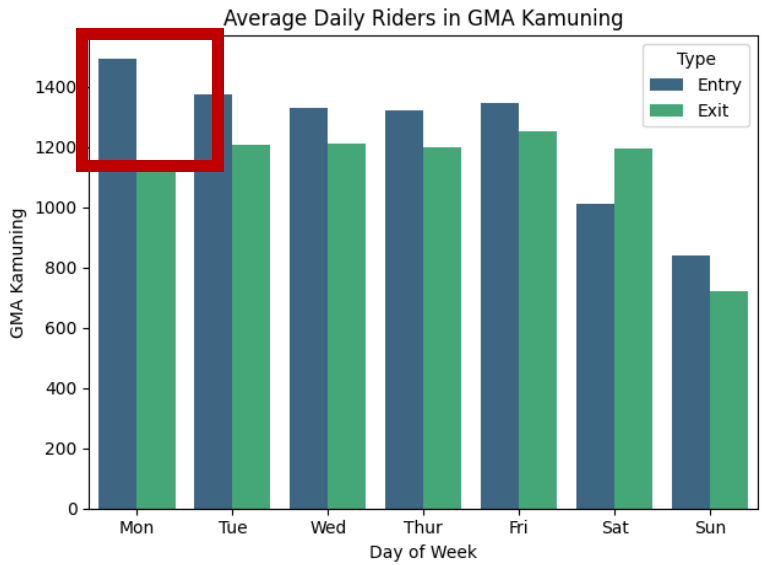
MORE RIDERS ON WEEKDAYS, PEAKING ON FRIDAY

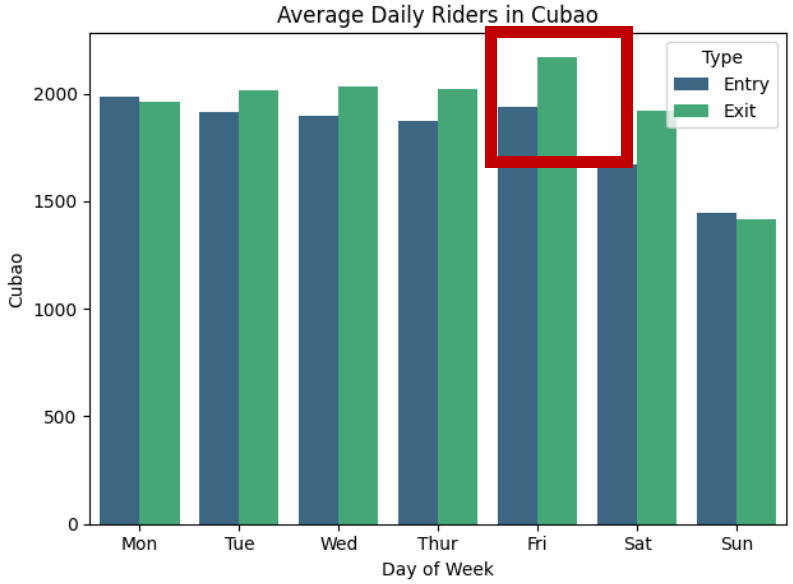


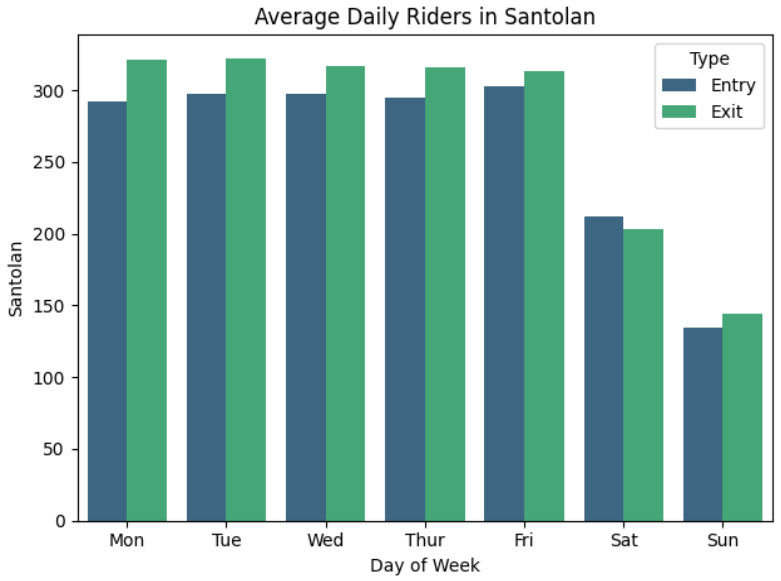
STATIONS NEAR THE NORTH PEAK ON MONDAYS

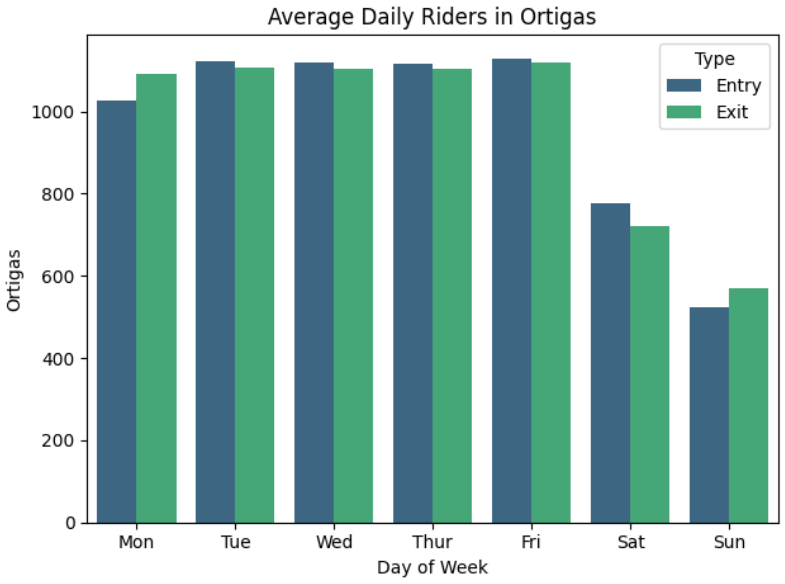




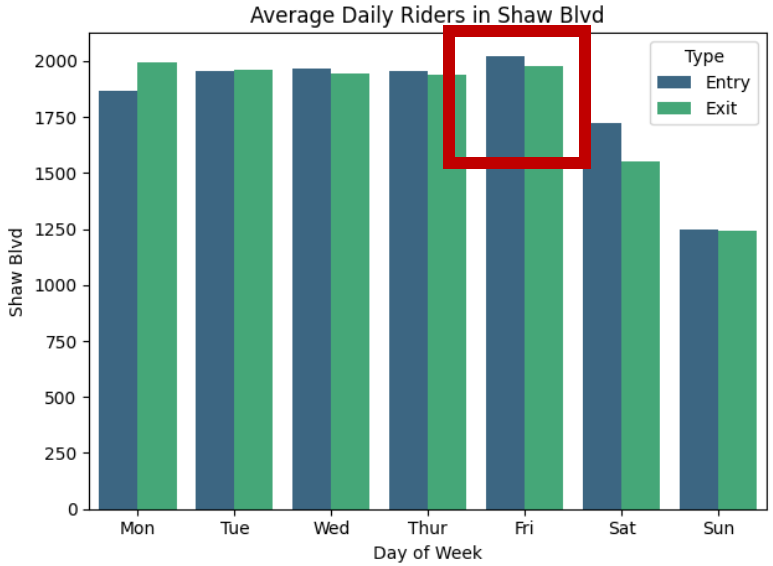


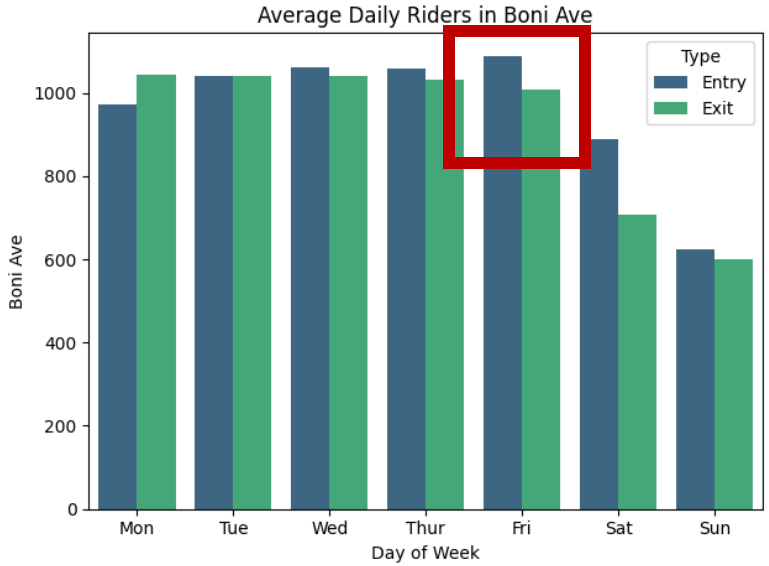


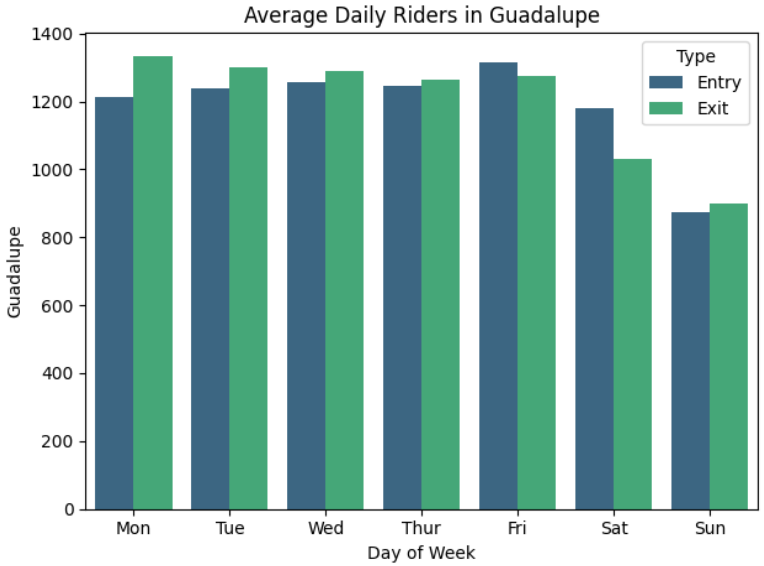


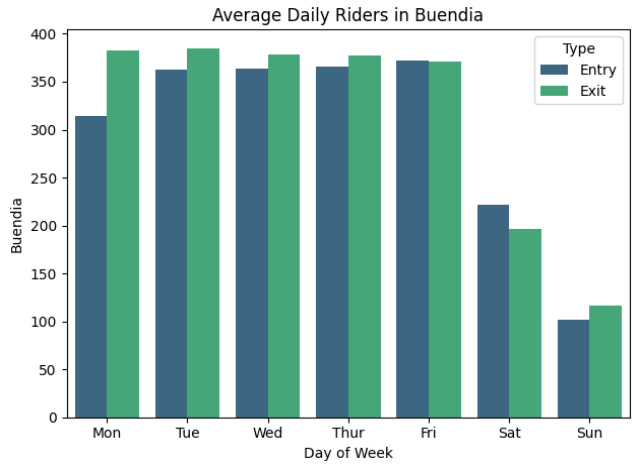


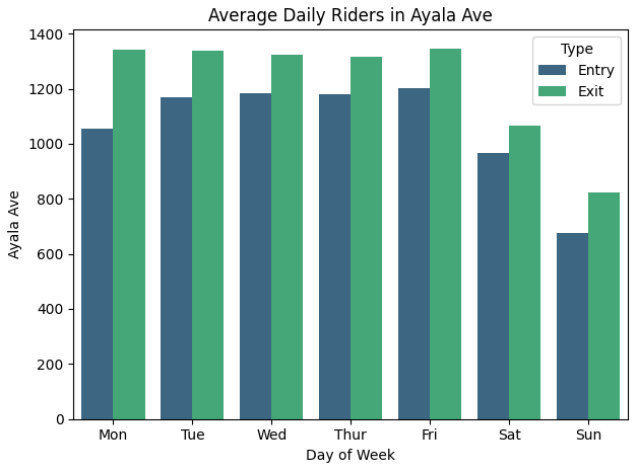
BUT MOST PEAK ON FRIDAYS

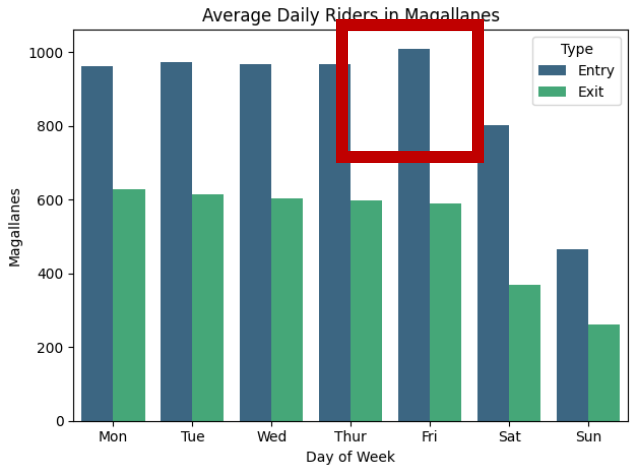


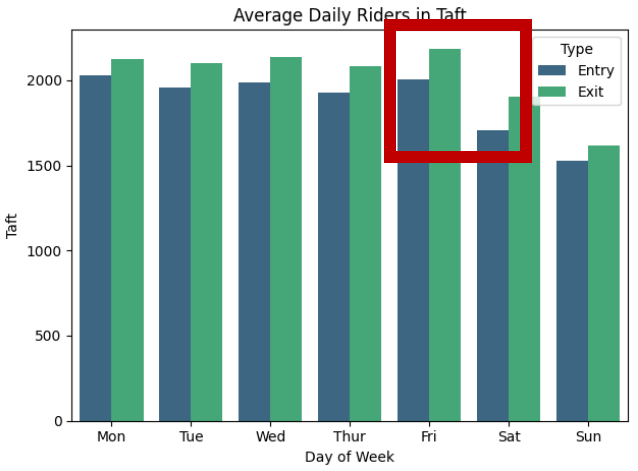




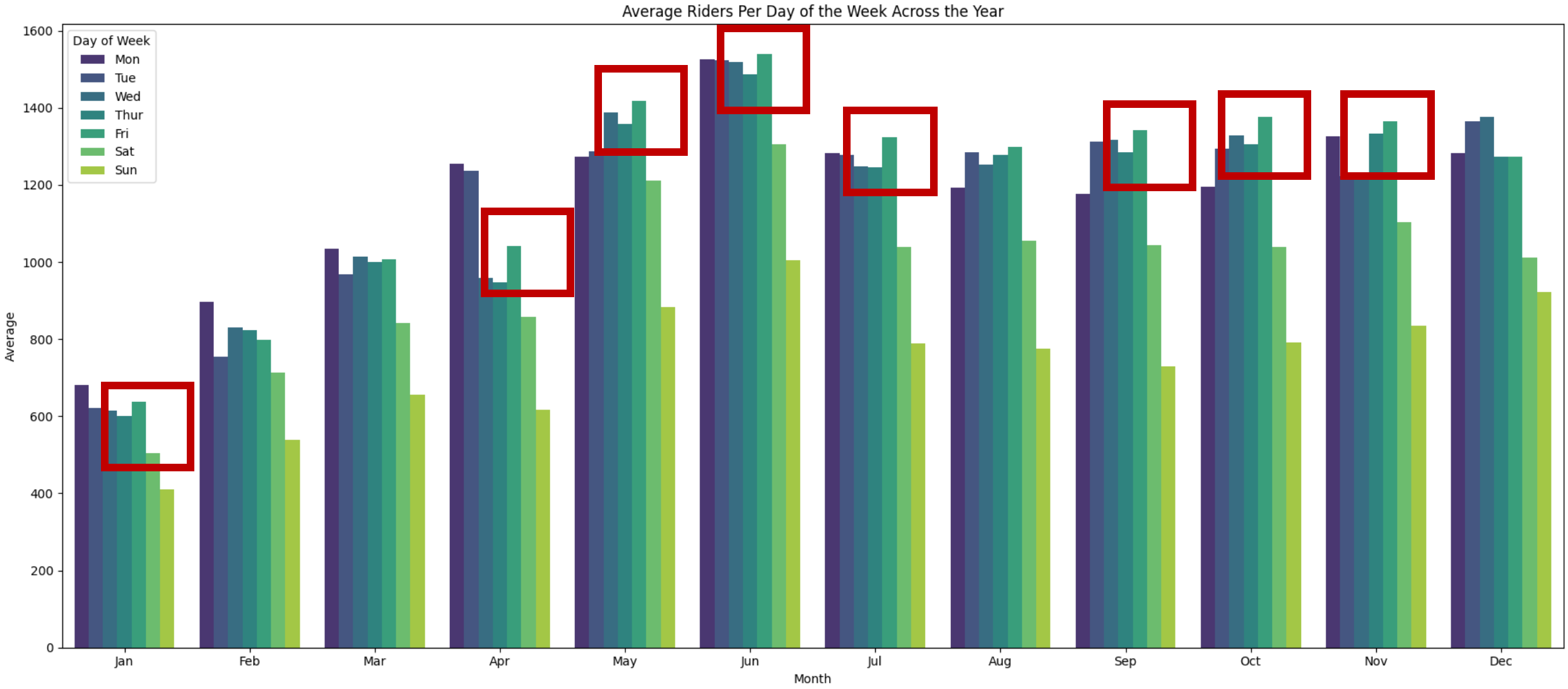






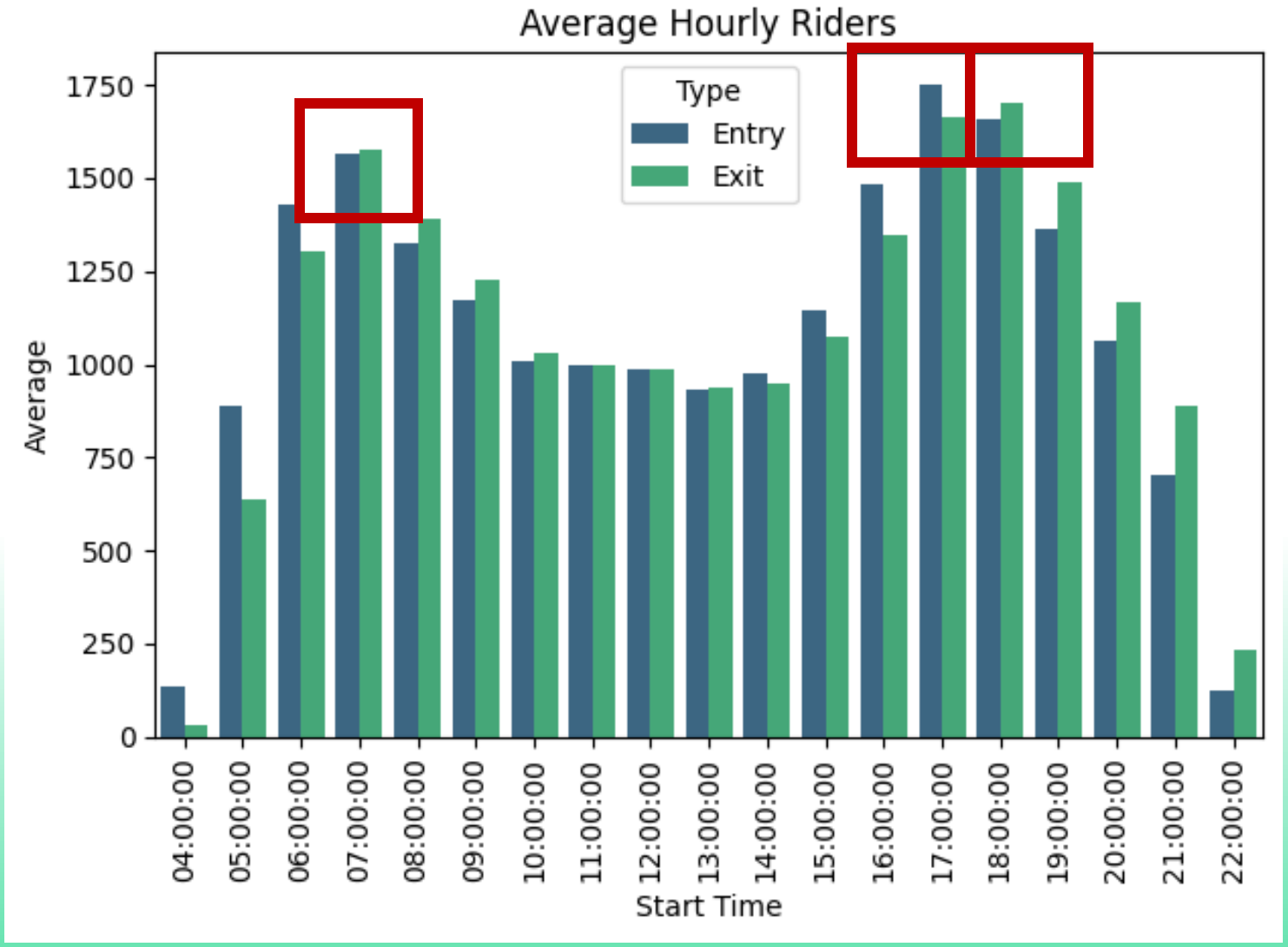


FRIDAY IS THE PEAK DAY THROUGHOUT THE YEAR

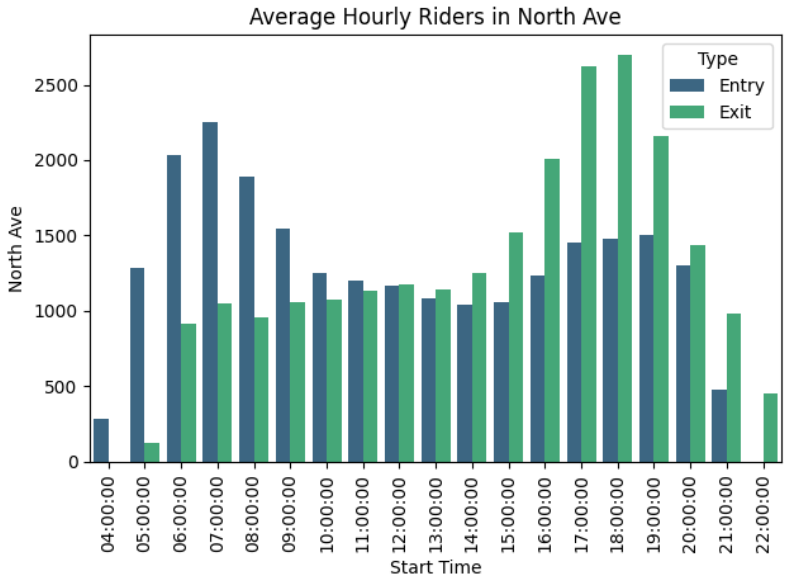


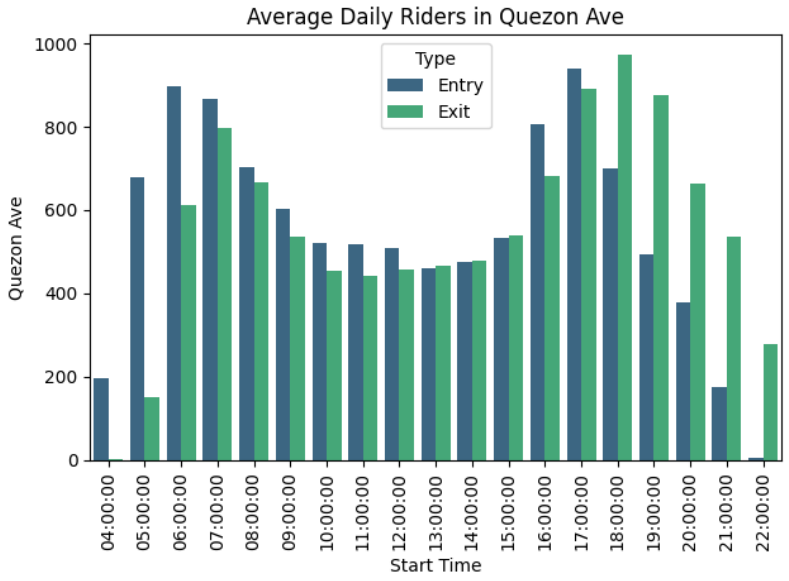
Hourly Trends

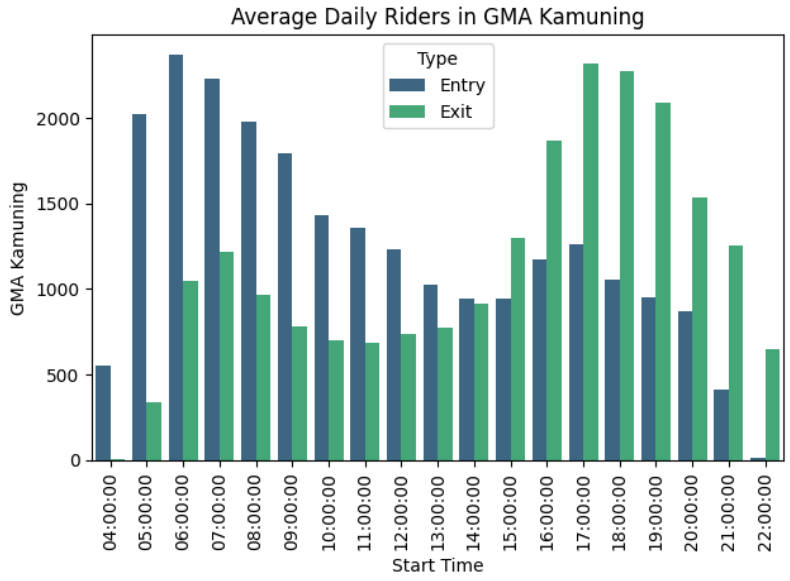
PEAK HOURS ARE 7 AM AND 5-6 PM DUE TO RUSH HOURS

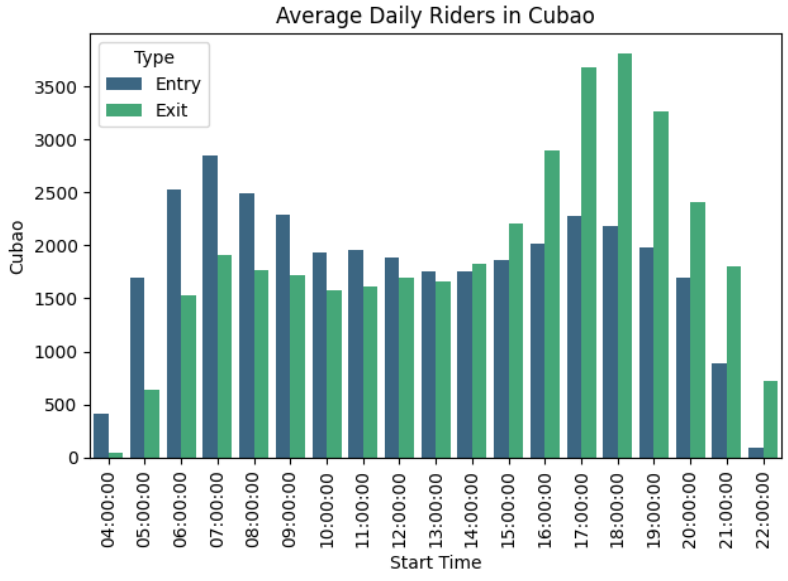


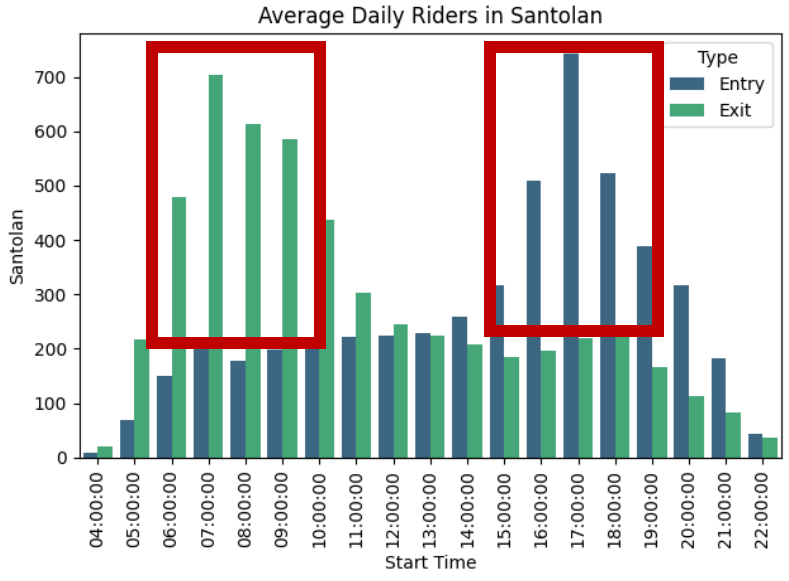
HIGHER EXITS AT PEAK MORNING HOURS AND HIGHER ENTRIES AT PEAK EVENING HOURS NEAR ORTIGAS CENTER AND MAKATI CBD

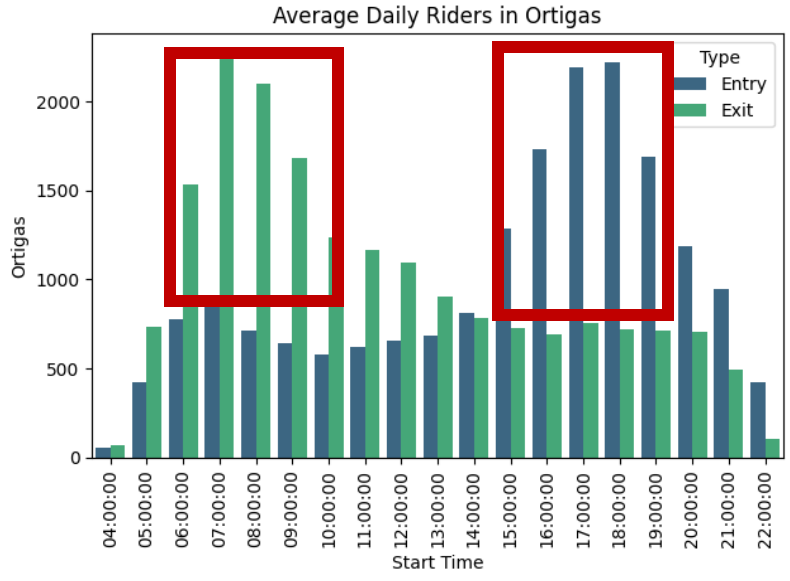




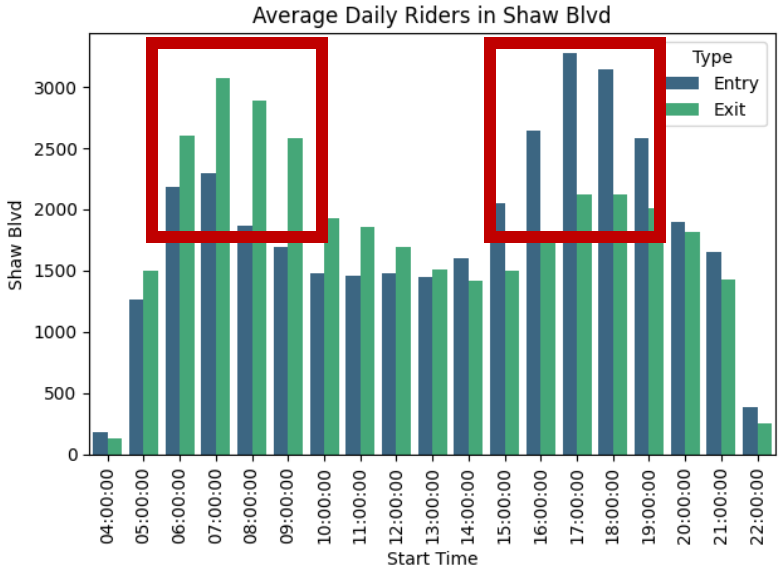


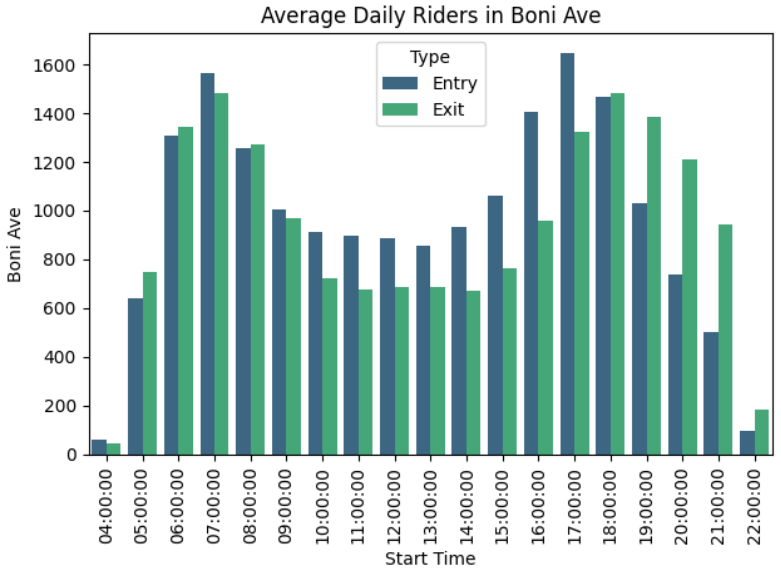


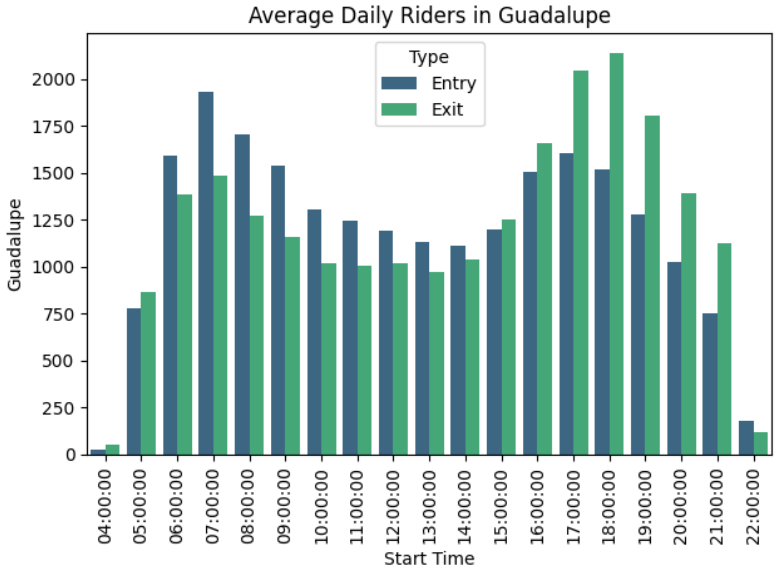


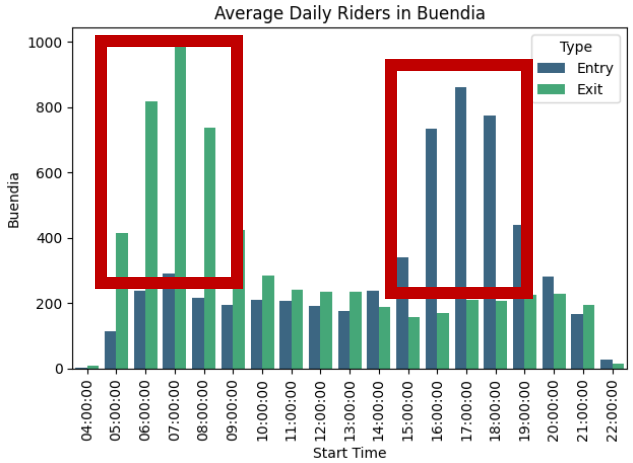


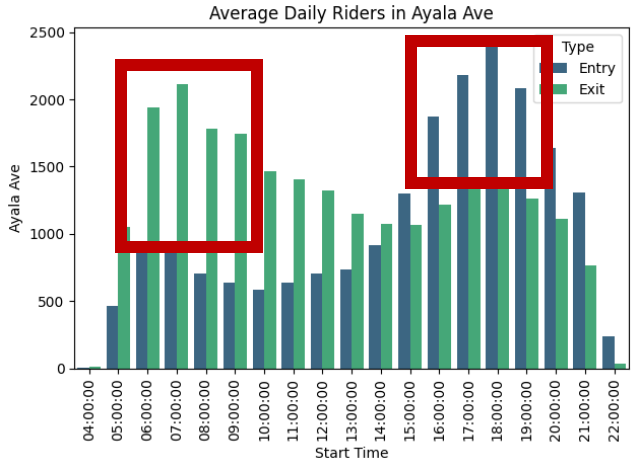
HIGHER EXITS AT PEAK MORNING HOURS AND HIGHER ENTRIES AT PEAK EVENING HOURS NEAR ORTIGAS CENTER AND MAKATI CBD

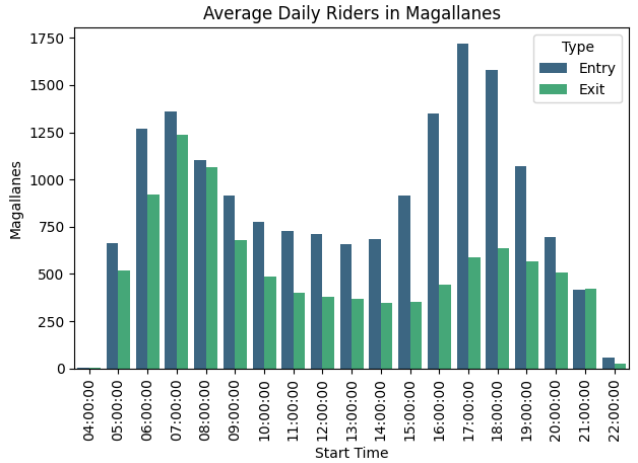


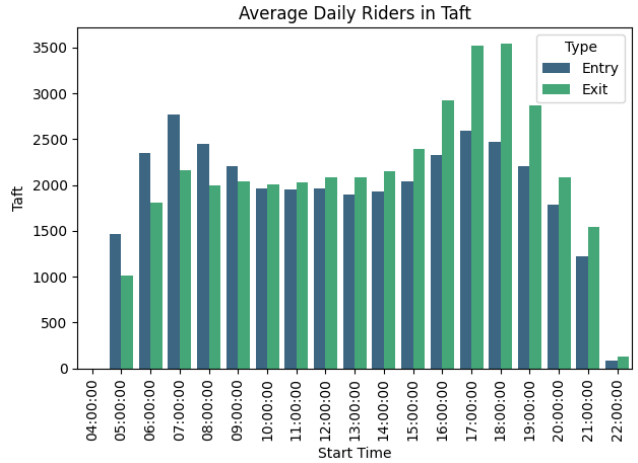




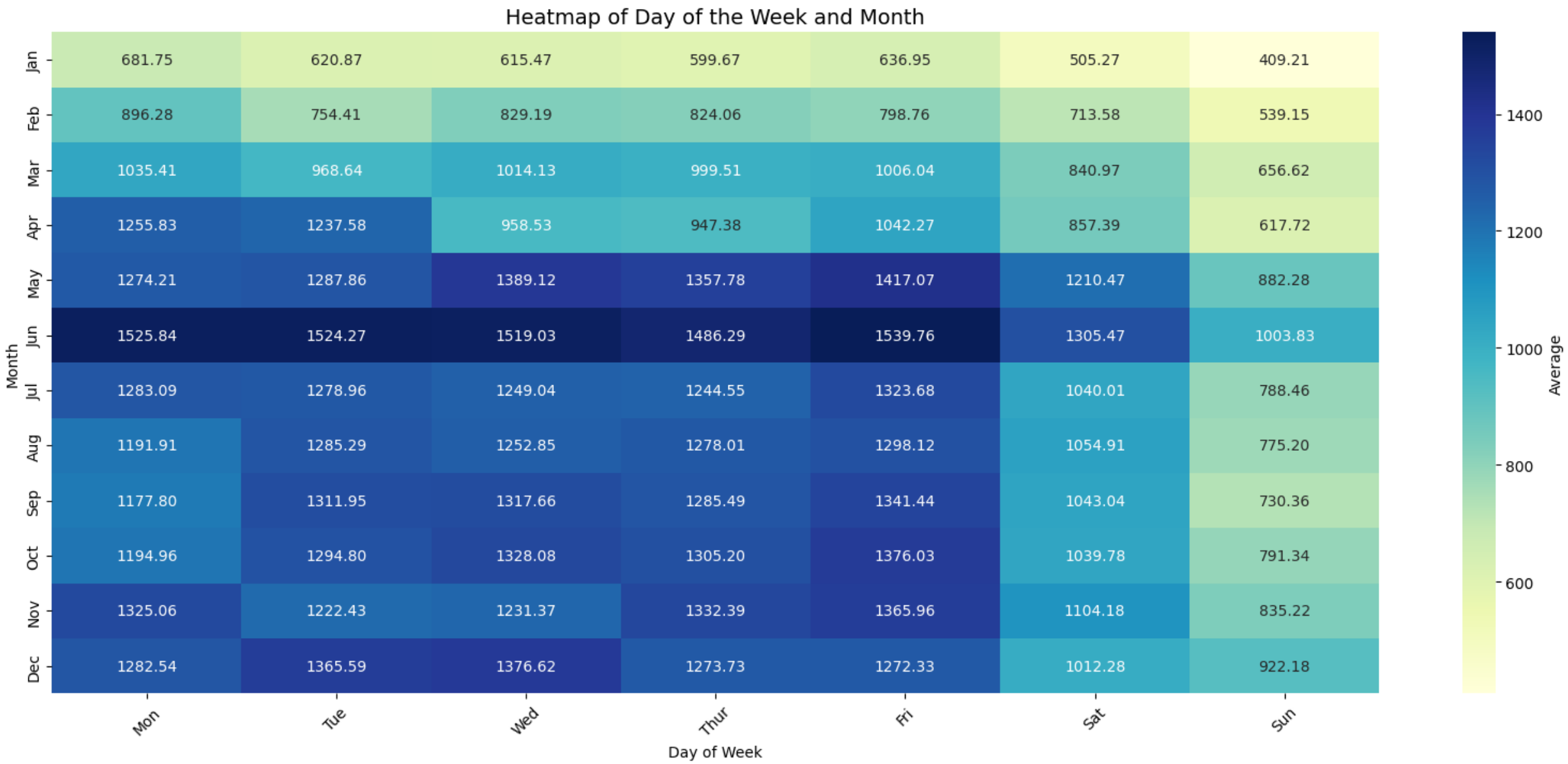


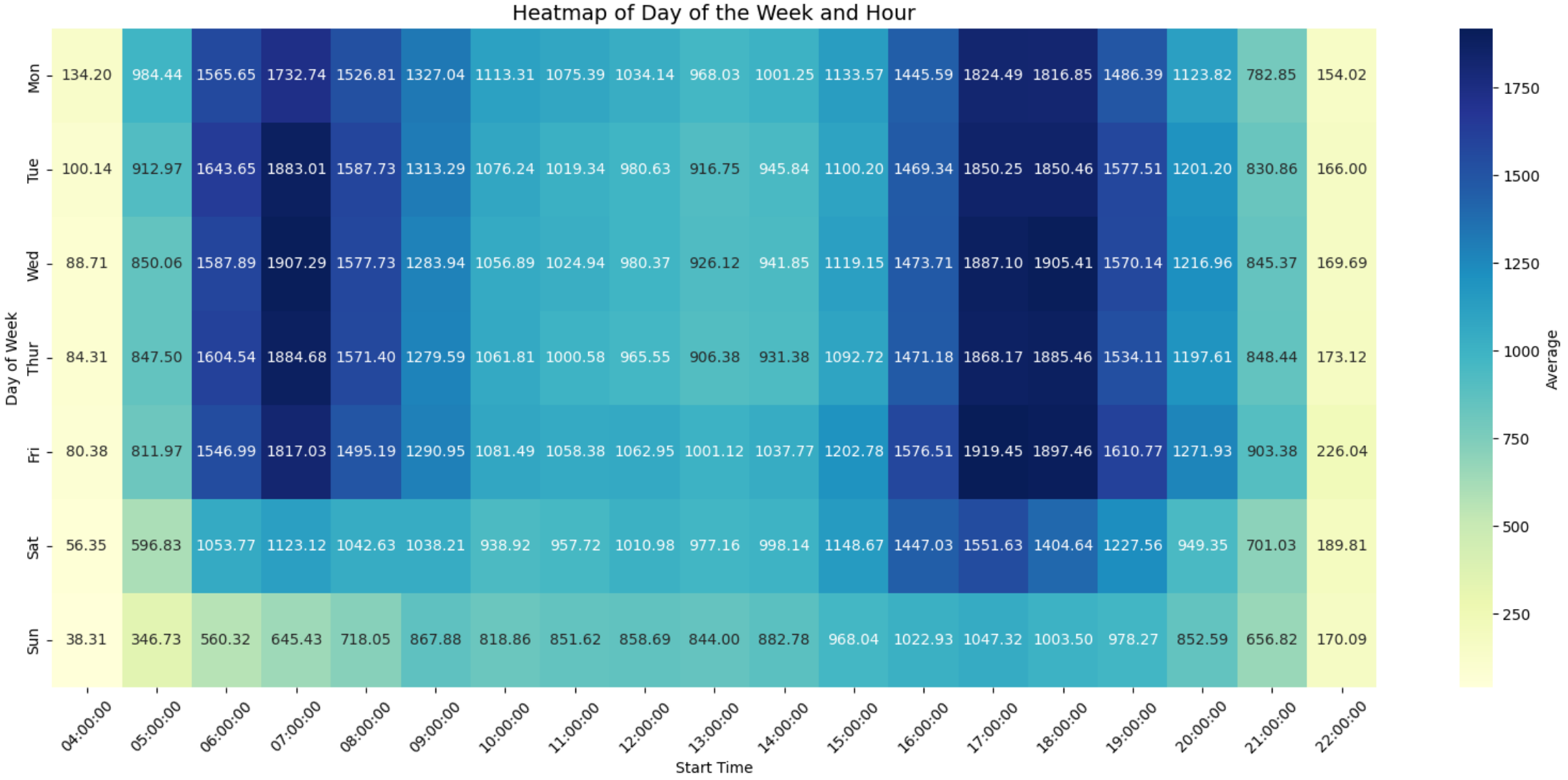


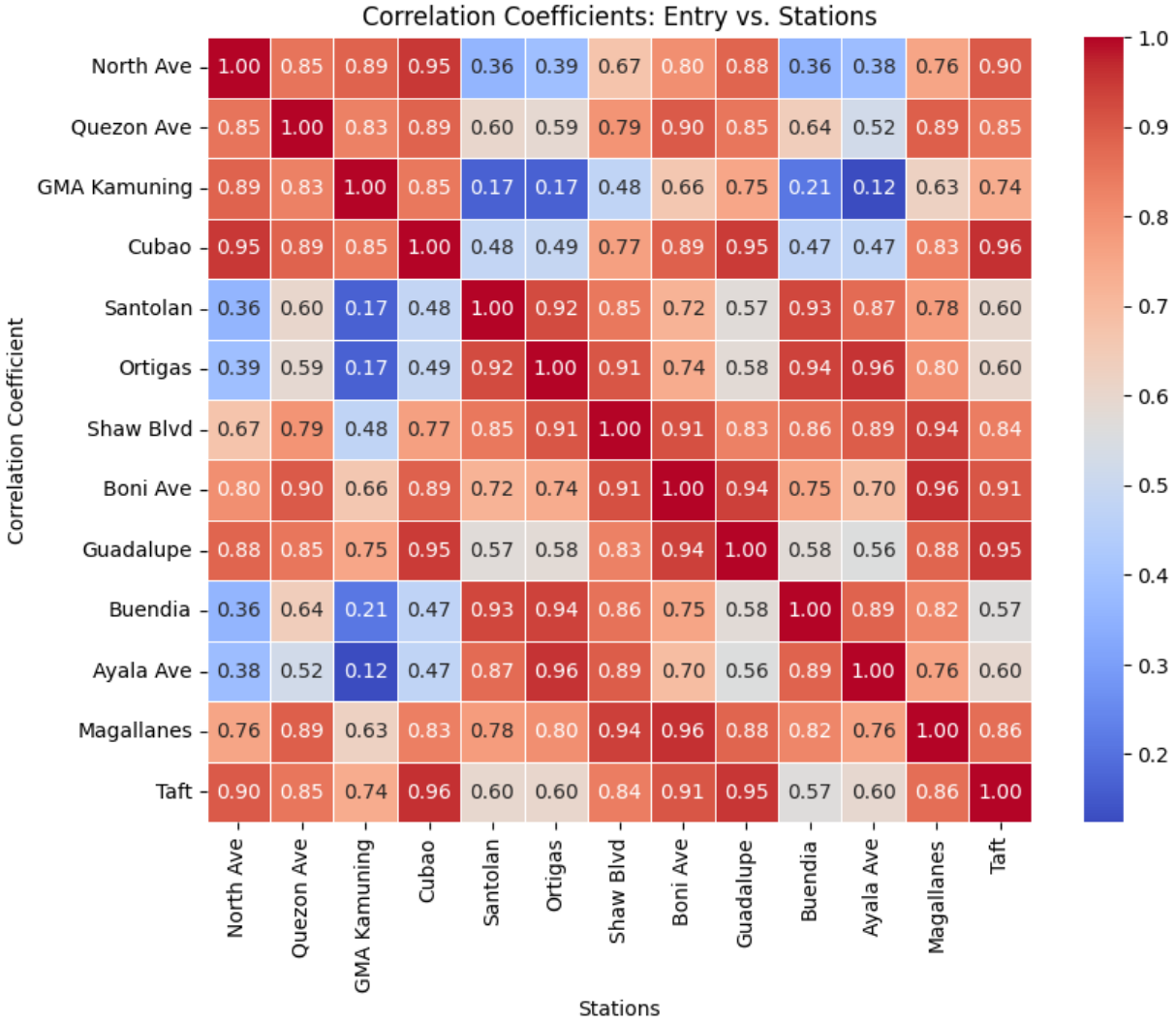


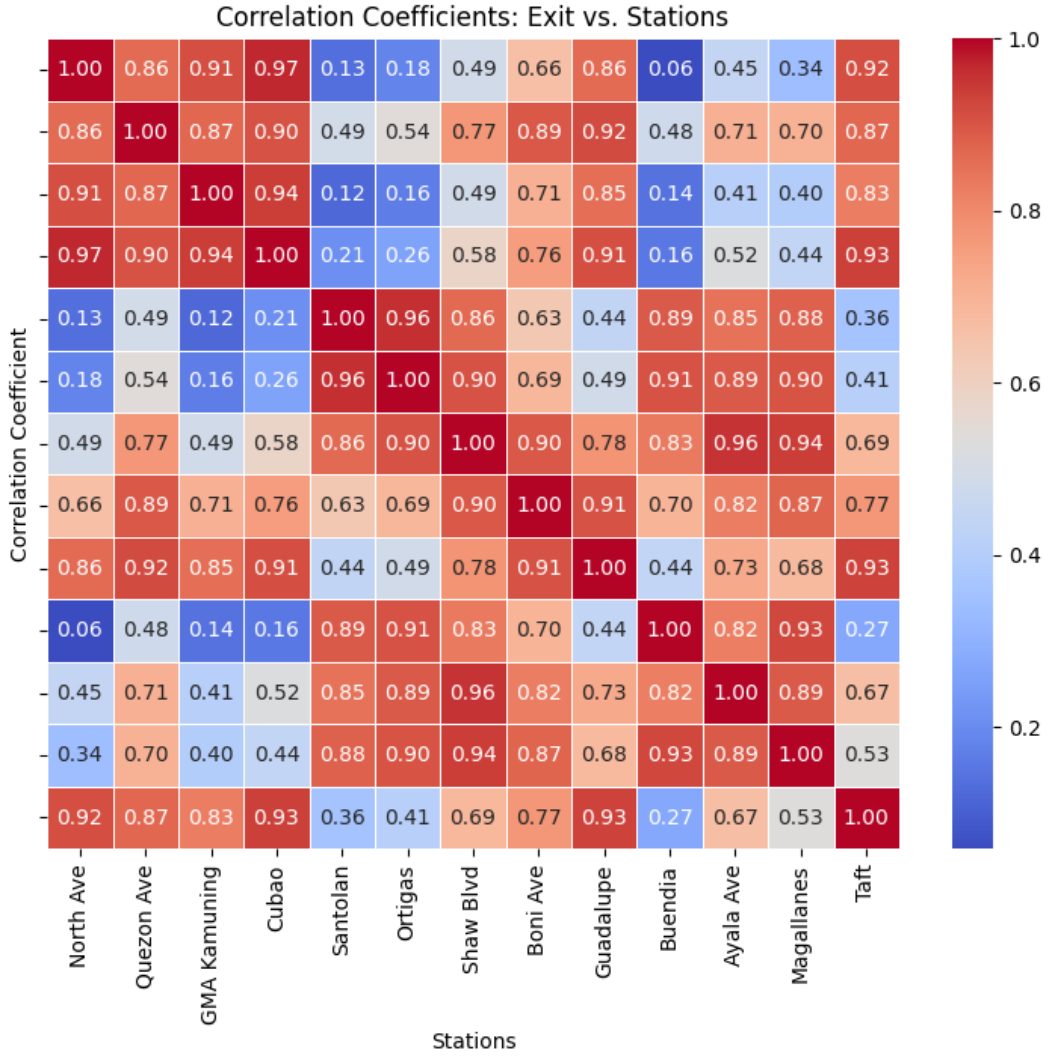


Heatmap Summaries


Heatmap of Day of the Week and Month
Month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
681.75 620.87 615.47 599.67 636.95 505.27 409.21
896.28 754.41 829.19 824.06 798.76 713.58 539.15
1035.41 968.64 1014.13 999.51 1006.04 840.97 656.62
1255.83 1237.58 958.53 947.38 1042.27 857.39 617.72
1274.21 1287.86 1389.12 1357.78 1417.07 1210.47 882.28
1525.84 1524.27 1519.03 1486.29 1539.76 1305.47 1003.83
1283.09 1278.96 1249.04 1244.55 1323.68 1040.01 788.46
1191.91 1285.29 1252.85 1278.01 1298.12 1054.91 775.20
1177.80 1311.95 1317.66 1285.49 1341.44 1043.04 730.36
1194.96 1294.80 1328.08 1305.20 1376.03 1039.78 791.34
1325.06 1222.43 1231.37 1332.39 1365.96 1104.18 835.22
1282.54 1365.59 1376.62 1273.73 1272.33 1012.28 922.18
Mon
Tue
Wed
Thur
Fri
Sat
Sun
Day of Week
Average
1400
1200
1000
800
600


Heatmap of Day of the Week and Hour
Day of Week
Mon 134.20 984.44 1565.65 1732.74 1526.81 1327.04 1113.31 1075.39 1034.14 968.03 1001.25 1133.57 1445.59 1824.49 1816.85 1486.39 1123.82 782.85 154.02
Tue 100.14 912.97 1643.65 1883.01 1587.73 1313.29 1076.24 1019.34 980.63 916.75 945.84 1100.20 1469.34 1850.25 1850.46 1577.51 1201.20 830.86 166.00
Wed 88.71 850.06 1587.89 1907.29 1577.73 1283.94 1056.89 1024.94 980.37 926.12 941.85 1119.15 1473.71 1887.10 1905.41 1570.14 1216.96 845.37 169.69
Thur 84.31 847.50 1604.54 1884.68 1571.40 1279.59 1061.81 1000.58 965.55 906.38 931.38 1092.72 1471.18 1868.17 1885.46 1534.11 1197.61 848.44 173.12
Fri 80.38 811.97 1546.99 1817.03 1495.19 1290.95 1081.49 1058.38 1062.95 1001.12 1037.77 1202.78 1576.51 1919.45 1897.46 1610.77 1271.93 903.38 226.04
Sat 56.35 596.83 1053.77 1123.12 1042.63 1038.21 938.92 957.72 1010.98 977.16 998.14 1148.67 1447.03 1551.63 1404.64 1227.56 949.35 701.03 189.81
Sun 38.31 346.73 560.32 645.43 718.05 867.88 818.86 851.62 858.69 844.00 882.78 968.04 1022.93 1047.32 1003.50 978.27 852.59 656.82 170.09
04:00:00 05:00:00 06:00:00 07:00:00 08:00:00 09:00:00 10:00:00 11:00:00 12:00:00 13:00:00 14:00:00 15:00:00 16:00:00 17:00:00 18:00:00 19:00:00 20:00:00 21:00:00 22:00:00
Start Time
Average
1750
1500
1250
1000
750
500
250


Correlation Coefficients: Entry vs. Stations
Correlation Coefficient
Stations
North Ave 1.00 0.85 0.89 0.95 0.36 0.39 0.67 0.80 0.88 0.36 0.38 0.76 0.90
Quezon Ave 0.85 1.00 0.83 0.89 0.60 0.59 0.79 0.90 0.85 0.64 0.52 0.89 0.85
GMA Kamuning 0.89 0.83 1.00 0.85 0.17 0.17 0.48 0.66 0.75 0.21 0.12 0.63 0.74
Cubao 0.95 0.89 0.85 1.00 0.48 0.49 0.77 0.89 0.95 0.47 0.47 0.83 0.96
Santolan 0.36 0.60 0.17 0.48 1.00 0.92 0.85 0.72 0.57 0.93 0.87 0.78 0.60
Ortigas 0.39 0.59 0.17 0.49 0.92 1.00 0.91 0.74 0.58 0.94 0.96 0.80 0.60
Shaw Blvd 0.67 0.79 0.48 0.77 0.85 0.91 1.00 0.91 0.83 0.86 0.89 0.94 0.84
Boni Ave 0.80 0.90 0.66 0.89 0.72 0.74 0.91 1.00 0.94 0.75 0.70 0.96 0.91
Guadalupe 0.88 0.85 0.75 0.95 0.57 0.58 0.83 0.94 1.00 0.58 0.56 0.88 0.95
Buendia 0.36 0.64 0.21 0.47 0.93 0.94 0.86 0.75 0.58 1.00 0.89 0.82 0.57
Ayala Ave 0.38 0.52 0.12 0.47 0.87 0.96 0.89 0.70 0.56 0.89 1.00 0.76 0.60
Magallanes 0.76 0.89 0.63 0.83 0.78 0.80 0.94 0.96 0.88 0.82 0.76 1.00 0.86
Taft 0.90 0.85 0.74 0.96 0.60 0.60 0.84 0.91 0.95 0.57 0.60 0.86 1.00
North Ave Quezon Ave GMA Kamuning Cubao Santolan Ortigas Shaw Blvd Boni Ave Guadalupe Buendia Ayala Ave Magallanes Taft
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2


Correlation Coefficients: Exit vs. Stations
Correlation Coefficient
Stations
North Ave 1.00 0.86 0.91 0.97 0.13 0.18 0.49 0.66 0.86 0.06 0.45 0.34 0.92
Quezon Ave 0.86 1.00 0.87 0.90 0.49 0.54 0.77 0.89 0.92 0.48 0.71 0.70 0.87
GMA Kamuning 0.91 0.87 1.00 0.94 0.12 0.16 0.49 0.71 0.85 0.14 0.41 0.40 0.83
Cubao 0.97 0.90 0.94 1.00 0.21 0.26 0.58 0.76 0.91 0.16 0.52 0.44 0.93
Santolan 0.13 0.49 0.12 0.21 1.00 0.96 0.86 0.63 0.44 0.89 0.85 0.88 0.36
Ortigas 0.18 0.54 0.16 0.26 0.96 1.00 0.90 0.69 0.49 0.91 0.89 0.90 0.41
Shaw Blvd 0.49 0.77 0.49 0.58 0.86 0.90 1.00 0.90 0.78 0.83 0.96 0.94 0.69
Boni Ave 0.66 0.89 0.71 0.76 0.63 0.69 0.90 1.00 0.91 0.70 0.82 0.87 0.77
Guadalupe 0.86 0.92 0.85 0.91 0.44 0.49 0.78 0.91 1.00 0.44 0.73 0.68 0.93
Buendia 0.06 0.48 0.14 0.16 0.89 0.91 0.83 0.70 0.44 1.00 0.82 0.93 0.27
Ayala Ave 0.45 0.71 0.41 0.52 0.85 0.89 0.96 0.82 0.73 0.82 1.00 0.89 0.67
Magallanes 0.34 0.70 0.40 0.44 0.88 0.90 0.94 0.87 0.68 0.93 0.89 1.00 0.53
Taft 0.92 0.87 0.83 0.93 0.36 0.41 0.69 0.77 0.93 0.27 0.67 0.53 1.00
North Ave Quezon Ave GMA Kamuning Cubao Santolan Ortigas Shaw Blvd Boni Ave Guadalupe Buendia Ayala Ave Magallanes Taft
1.0 0.8 0.6 0.4 0.2

Recommendations

RECOMMENDATIONS

From the EDA conducted here are the insights and recommendations generated by the group for the target market of this research.

1. Station Optimization & Service Planning
2. Dynamic Pricing Scheme
3. Optimal Time to Ride
4. Data Scientists: Modeling

RECOMMENDATIONS

Station Optimization & Service Planning

- Implement strategic measures during peak rush hours, including optimized staffing, increased security, and maintenance scheduling, focusing on high exit stations like Araneta Center-Cubao, Santolan-Annapolis, Ayala, and Taft Avenue.
- Schedule maintenance on Saturdays and in January when rider impact is minimal.

RECOMMENDATIONS

Dynamic Pricing Scheme

- Introduce a dynamic ticket pricing scheme to optimize ridership during low traffic periods, particularly on Fridays and Saturdays, and during off-peak hours.
- Offer reduced ticket prices for early morning and late evening slots, and during midday hours (9:00-16:00) to encourage ridership outside of peak hours.

RECOMMENDATIONS

Optimal Time to Ride

- Riders should aim to travel outside of peak rush hours (6:00-8:00 and 17:00-21:00) to avoid overcrowded trains and delays.
- Traveling during off-peak times will enhance the riding experience by avoiding the rush-hour influx of passengers.

RECOMMENDATIONS

Data Scientists: Modeling

- Utilize the high correlation between end stations to predict boarding and destination patterns in models.
- Acknowledge the low correlation between north and south stations and the centralization of businesses, indicating that riders often commute from the outskirts to the center. Further economic analysis near stations is needed for deeper insights into rider inflow and outflow.

Thank you!