Sushmit Roy
319-610-7375
sushmit86@gmail.com

Problem Statement: Recommendation for improving MTA service using publicly available data set

Data Set: The data set used for this exercise is publicly available from MTA website. Data downloaded is in csv format, consisting of 9 files.

Analysis:

Number of trips and Average trip time:

The plot above shows average number of trips per day and average time in minutes for that particular route on Weekdays / Saturday /Sunday. Summarizing results from above :

- Number of trips on a particular route is more on weekdays compared to weekends.
- There are 27 routes active on weekdays and 23 routes active on weekends.
- Train with second highest average run time (90 minutes) 'D_6_avenue_express' has only 268 round trips per day
- Average time difference between '6X_Lexintgon_Avenue_ Local' and '6X_Lexington_Avenue_Express' is 2 minutes..Average time difference between '7_flushing_local' and '7X_flushing_Express' is 4 minutes.

Based on the above observations, I would recommend the following:
- The number of trips on certain busy routes could be increased on Saturdays.
- 'B_6_avenue_express' route is non operational during weekends. I would recommend to run few trains on this route during weekends
- Increase in number of trips for 'D_6_avenue_express'route.
- Also, decreasing time duration for both 6 and 7 Express lines might help in decreasing commute

time

Hourly Traffic Pattern:
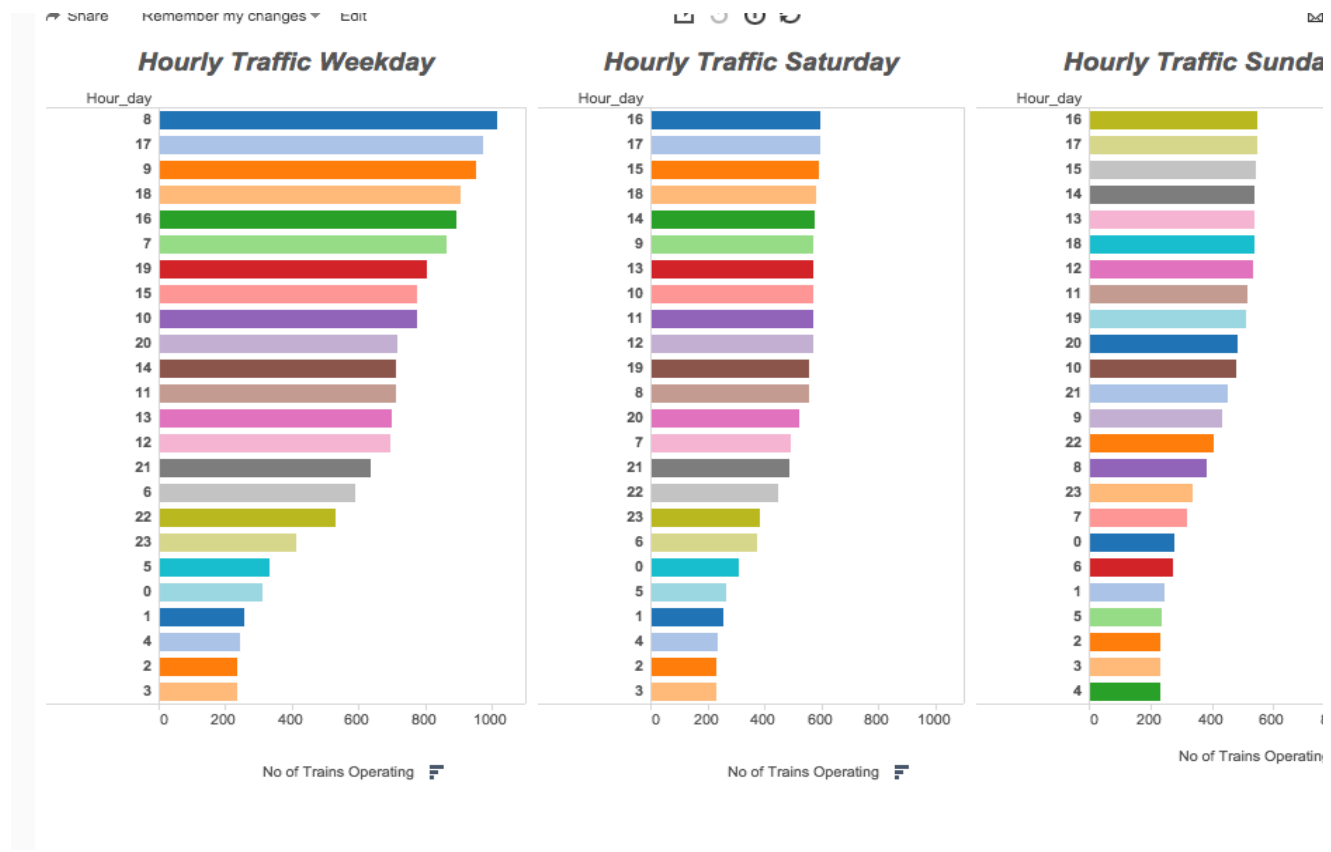
Looking into the hourly pattern of traffic is very helpful in deciding logistics. The above plot shows number of trains running during different hours of the day.

Some of the important points from the above plot:
- Traffic density on Saturday/Sunday is almost half compared to weekdays.
- The hourly traffic pattern between midnight to 5 A.M. in weekends is almost similar to the hourly traffic pattern between midnight to 5 A.M. in weekdays
- The busiest traffic during weekdays is at 8A.M. and 5P.M. . Weekends have more rush during evening.

Based on the above, I would recommend the following:
- Although, it makes sense to have less traffic in weekends but slightly increasing the traffic in Saturdays might help.
- The traffic pattern between midnight to 5A.M. is similar in weekdays and weekends. It might make sense to cut down the traffic on weekdays between midnight to 5A.M.
- A slight increase in traffic between midnight and 5A.M. may help as more people use the Subway service during weekends after midnight

SERVICE CHANGES:

9/7/15 0:00			11/26/15 0:00	
SERVICE STOPPED	SUNDAY SERVICE		SERVICE STOPPED	SUNDAY SERVICE
1	1		1	1
2	2		2	2
3	3		3	3
4	4		4	4
5	5		5	5
6	6		6	6
7	7		7	7
6X	A		6X	A
7X	C		7X	C
A	D		A	D
B	E		B	E
C	F		C	F
D	FS		D	FS
E	G		E	G
F	GS		F	GS
FS	H		FS	H
G	J		G	J
GS	L		GS	L
H	M		H	M
J	N		J	N
L	Q		L	Q
M	R		M	R
N			N	
Q			Q	
R			R	
Z			Z	

From the calendar_dates data we can easily infer the dates of service changes. From the above table, we can conclude that both on 7th September, Monday (labor day) and 26th Nov, Thursday (Thanksgiving day) instead of weekday schedule, MTA will operate on a Sunday schedule. Based on the above data, I would recommend a special holiday schedule be implemented for holidays (instead of running on a Sunday schedule, which has the lowest frequency of trains). Of course, the frequency need not be as high as weekdays but certainly more than that of Sunday service.

There are many more analytical questions that can be answered from this data. Considering the page limit, I would like to limit myself up to this point. The tableau report (see appendix) created also has a dashboard for busiest stations based on traffic.

Appendix:

For analysis, I have used MySQL (5.7) and Tableau (9.1). After downloading the data in csv format, I created an MTA schema in local Mysql database. Data from each text file was then uploaded into corresponding table created in MTA schema. The queries used for the above is attached (MTA_queries.sql) .

For visualization, I used Tableau 9.1 and published it online on Tableau Cloud server. Here are the links to the Dashboard -

Average Time and Number of Daily Runs

Hourly Traffic over 24 Hrs

Station Traffic Weekday

You can access the link using the following credentials

User id : sushmit86@gmail.com
Password : password123

Please let me know if you would like to access the links using your own email. I could give view access to that email.

If you have any questions related to the analysis, please feel free to email or call me.