**Sushmit Roy**

319-610-7375

[sushmit86@gmail.com](mailto:sushmit86@gmail.com)

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

Data Set: The data set used for the 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:**



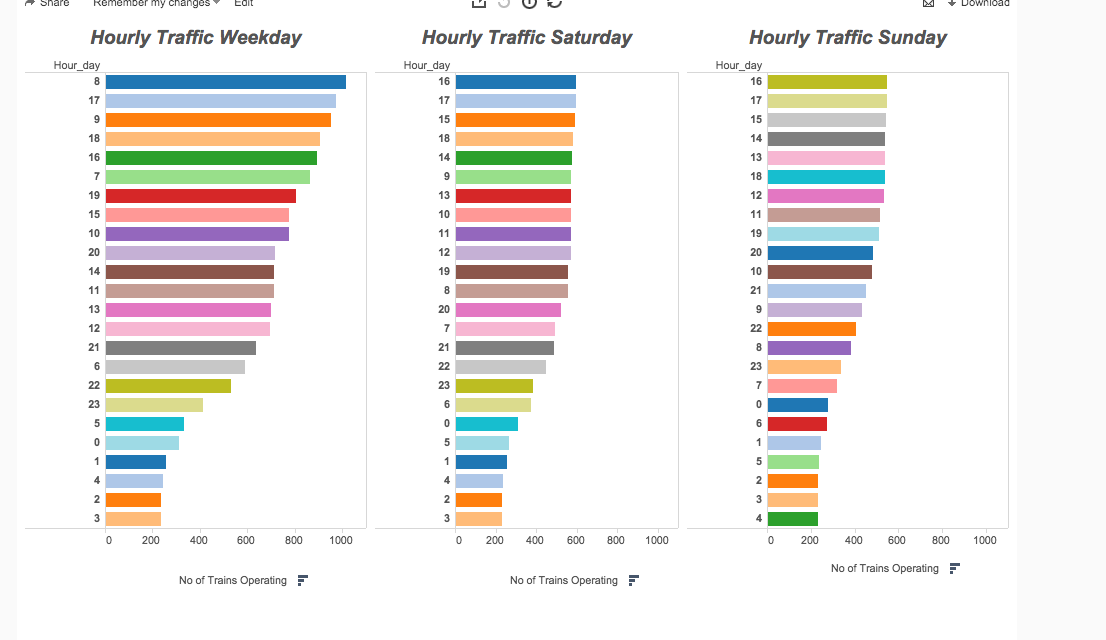
The plot above shows average number of trips per day and average time in minutes for that particular route on Weekdays / Saturday /Sunday. Summarizing results from above

* 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.
* Trains with second highest average run time (90 minutes) ‘D\_6\_avenue\_express’ has only 268 trips per day
* Average time difference between ‘6X\_Lexinton Avenue Local’ and ‘6X\_Lexington\_Avenue\_Express’ is 2 minutes. Similarly, average time difference between ‘7\_flusing\_local’ and ‘7X\_flushing\_Express’ is 4 minutes.

Based on the above observation I would recommend the following:

* The number of trips on certain busy routes can be increased on Saturday.
* It will be worth considering if few trains could run on ‘B\_6\_avenue\_express’ route during weekends.
* Increase in number of trips for ‘D\_6\_avenue\_express’
* Also decreasing time duration both for 6 and 7 Express line might be help decrease commute time.

**Hourly Traffic Pattern:**

****

Looking into hourly pattern of traffic is very helpful in deciding many logistical questions. The plot above shows number of trains running on different hours of the day.

Some of the important points from above plot

* Traffic density on Saturday/Sunday is almost half compared to weekdays.
* The hourly traffic pattern between midnight to 5 AM is almost similar between weekends and weekdays.
* Morning 8AM and 5 PM are the busiest during weekdays and during weekends there is more rush during evening

Based on the above I would recommend the following:

* Although it makes sense to have less traffic on weekends but slightly increasing the traffic during Saturday might help
* The traffic pattern between midnight to 5AM is similar in weekdays and weekends. It might make sense to cut down the traffic on weekdays between midnight to 5AM.
* Similarly increase in traffic between midnight and 5AM 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 there would be service change. From the table above we can conclude that both on 7th September, Monday (labor day) and 26th Nov, Thursday (Thanksgiving day) instead of weekday service MTA will operate Sunday service. Based on the above I would recommend instead of running Sunday service (which has the lowest number of train operating) on Thanksgiving a special holiday schedule can be implemented. Of course the frequency need not be as compared to weekday but certainly more than the Sunday service

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

**Appendix:**

For purpose of this analysis, I have used MySQL (5.6) and Tableau (9.1). After downloading the data in csv format I created a MTA schema in local Mysql database. Data from each text file was uploaded into corresponding table created in MTA schema. The queries used for the above is attached (MTA\_queries.sql) .

For the visualization part I used Tableau9.1 and published it online in Tableau Cloud server. Here are the links to the dashboard

[Average Time and Number of Daily Runs](https://10ay.online.tableau.com/t/sushmit_test/views/Tableau_workbook/AverageTimeandNumberofDailyRuns?:retry=yes&:embed=y&:showShareOptions=true&:display_count=no)

[Hourly Traffic over 24 Hrs](https://10ay.online.tableau.com/t/sushmit_test/views/Tableau_workbook/HourlyTrafficover24Hrs?:retry=yes&:embed=y&:showShareOptions=true&:display_count=no)

[Station Traffic Weekday](https://10ay.online.tableau.com/t/sushmit_test/views/Tableau_workbook/StationTrafficWeekday?:embed=y&:showShareOptions=true&:display_count=no)

You can access the link using the following credentials

User id : [sushmit86a@gmail.com](mailto:sushmit86a@gmail.com)

Password : password123

In case you want to use your own email id please let me know your email address. I can provide view access for the corresponding email address ,although you will have to register to access the dashboard.

In case of any other questions related to the above analysis please feel free to email /call me.