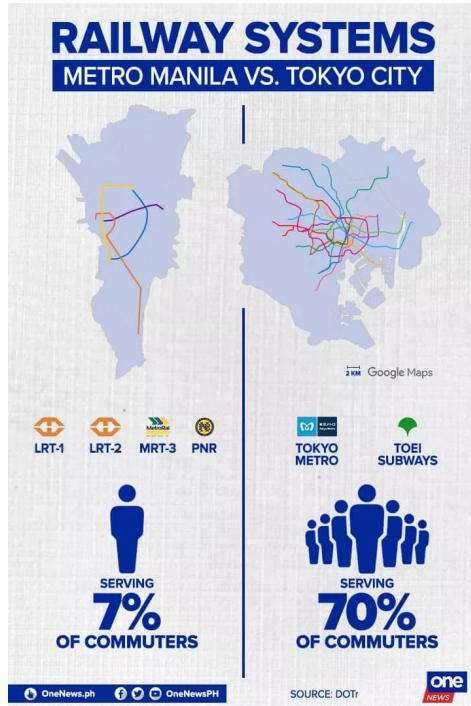


Analysis on MRT3 Ridership 2017

DATASCI FINAL PROJECT

Project **Overview**



One of the most prominent issues that residents face in Metro Manila is the horrendous experience in public transportation.

Although debatably the fastest way to get to places, it is one of the most uncomfortable modes of transportation due to the long queues, overloaded trains and frequent breakdowns.

"It is estimated that millions of man-hours are lost each day as commuters wait for their turn for an hour or more before riding the three rail systems, which are supposed to make travel faster and easier for passengers." (dela Cruz, 2015)

Project **Motivation**

NOT ENOUGH TRAINS

The low supply is not enough to satisfy the current increasing demand.

The best solutions would **require large amounts of budget and resources.**

PROBLEM QUESTION

Given a set of data, how can we determine optimal train schedules and/or suggest improvements to the MRT system?



General **Objective**

The general objective is to **analyze trends and patterns of commuter behaviour and Metro Rail Transit Line 3 traffic activity and ridership** on a monthly, weekly, daily, and hourly basis.

Specific objectives include:



Determine the peak hours and days of traffic activity and ridership in the MRT



Determine the volume and congestion of passengers in specific stations in a certain time



Determine the optimal schedule (time and starting station) of skip trains and the most optimal interval times

Data **Cleaning**

The data wrangling processes are the following:

1. **Manual editing** of data from a csv dataset
2. **Sorting of data:** includes months, stations, etc.
3. **Data type manipulation:** it includes converting the month and day to a datetime format.
4. **Data merging:** combining two dataframes (station values and station points)
5. **Inserting new columns:** includes Summing and Averaging
6. **Geospace Initialization:** cleaning the station points and PH map

Used **Dataset**

Daily number of passengers who have entered and exited per station on average, percentage and total in Metro Rail Transit Line 3 in 2017.

<https://drive.google.com/file/d/1Ly-YzLIS2BjjczC4nduDqGL6CHvRKhNL/view>

Exploratory **Visuals**

Data visualization will mostly deal with time series analysis on station entries. Station exits will **NOT** be included in the analysis. Data visualization will include:

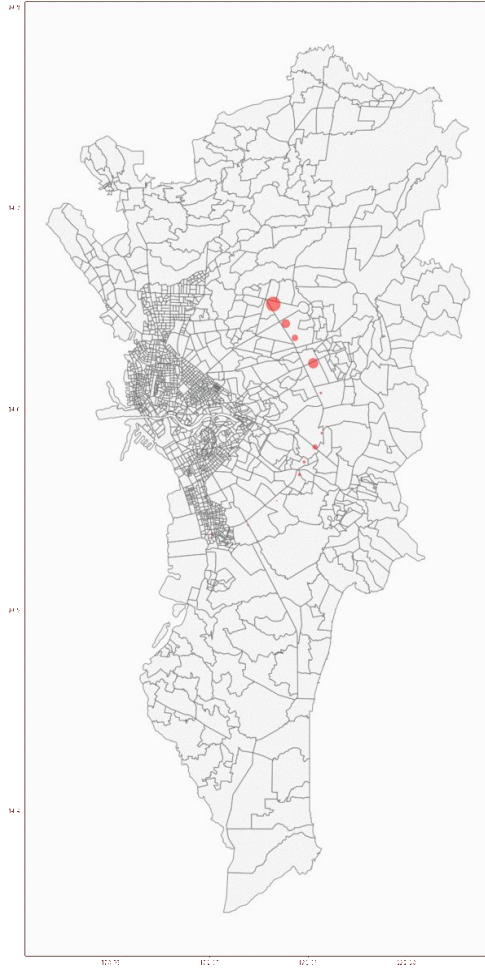


Geospatial

Heatmaps

Time series line graphs

Average number of entries per Station in 2017 at:
4:00 - 4:59 AM

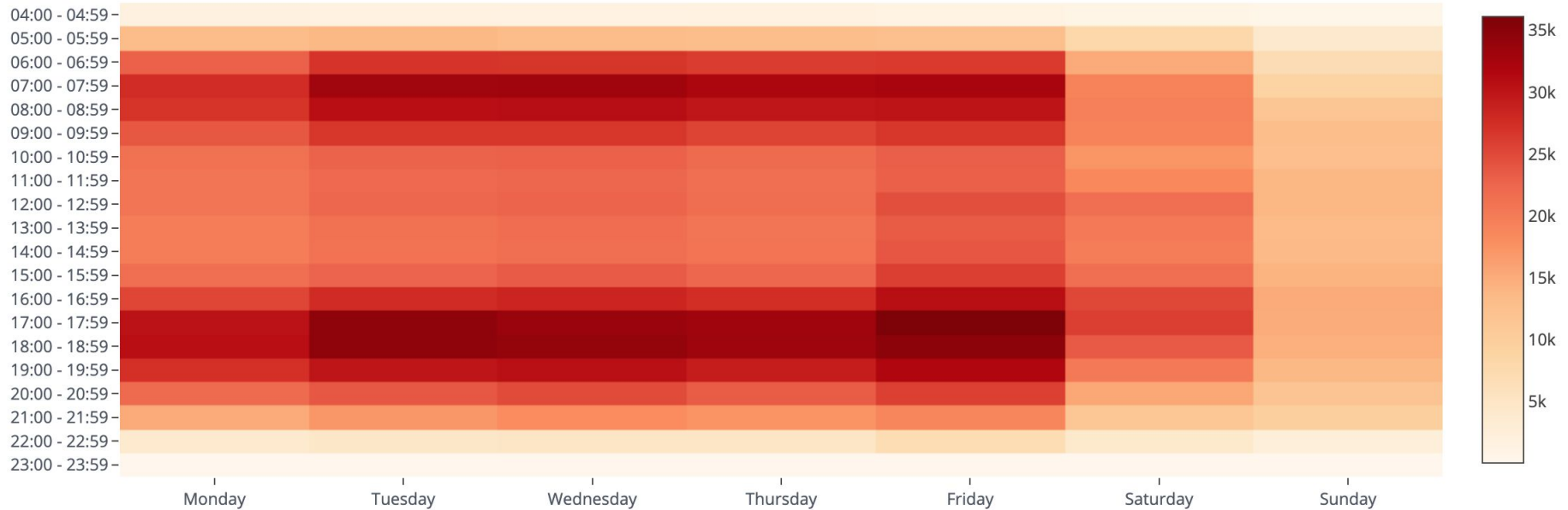


Average commuter traffic activity on different stations in a day in 2017

Observations:

There is always a large traffic activity at the end stations (North ave and Taft), and mid stations.

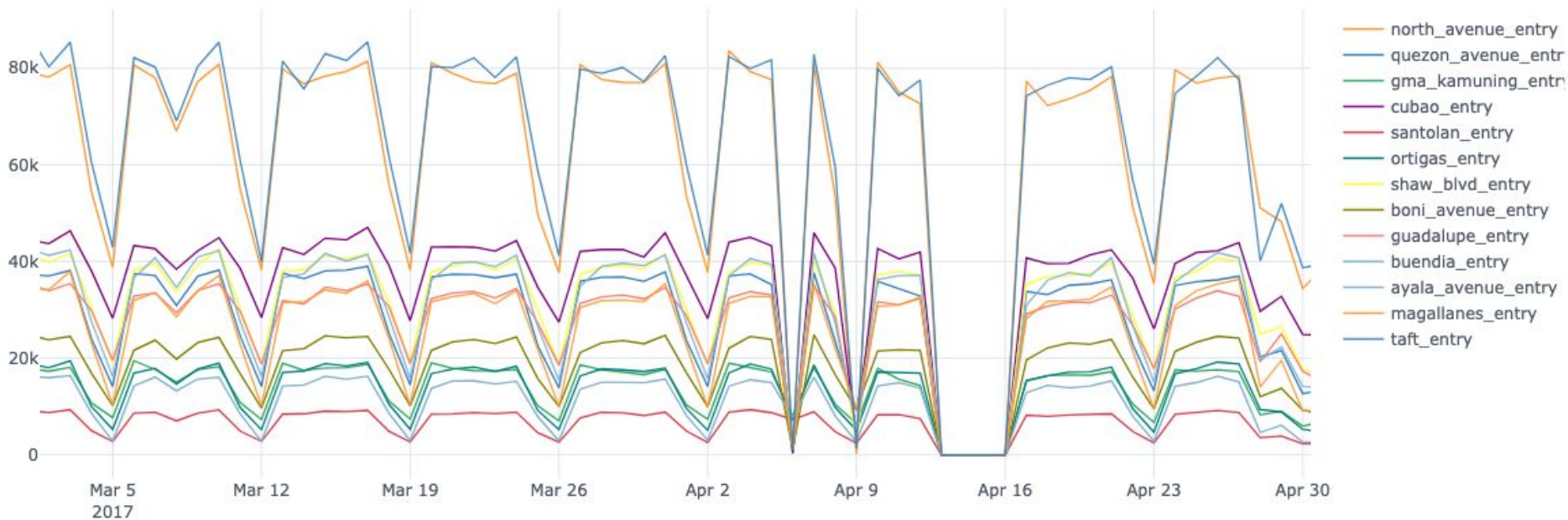
Average commuter traffic at each station at different times in a day



Observations:

1. Most congested hours are start and end of working hours.
2. Peak day is friday, least busiest is sunday

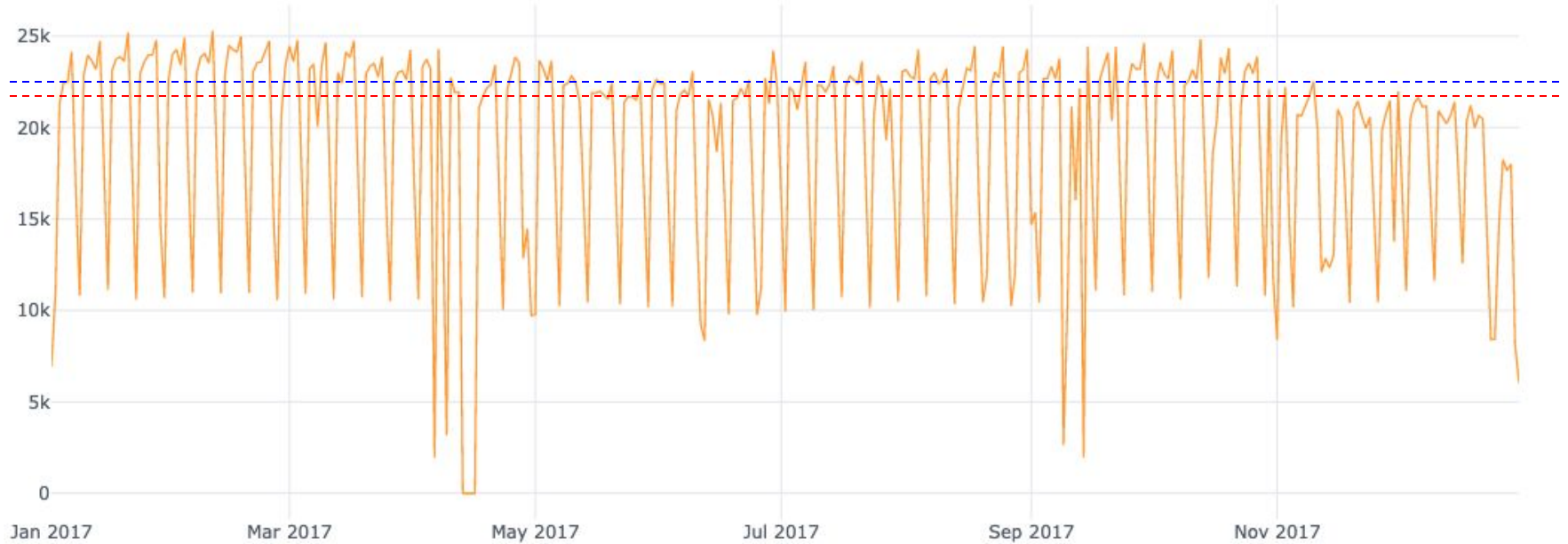
MRT3 2017 Monthly - Daily Ridership (Entries)



Observations:

1. Traffic activity is most congested at the end stations
2. The least commuter traffic activity is found at Santolan station and GMA Kamuning station

Total Averaged MRT3 2017 Monthly - Daily Ridership (Entries)



Observations:

1. There are less commuters during Summer (Apr-Jun) and Ber-months (Nov-Dec) versus the other months

Note: There was no MRT3 activity on Apr 13 - Apr 16, 2017

Test **Hypotheses**



The entry population at the End Stations (Taft and North) are significantly larger than that of the average entries of all stations



The station congestion on Fridays is significantly higher than on the other days of the week



Stations are busiest at the the start and end of the commuters' working hours versus other times in the day.



There are less commuter activity during Summer (Apr-Jun) and Ber-months (Nov-Dec) versus the other months

Statistical **Tests**

Most of the hypothesis underwent Z-Testing used to test a hypothesis on a specific value of the population proportion.

- End station entries versus Overall station entries
- Station congestion at peak hours versus Average station congestion activity within a day
- Station congestion on Fridays versus Average Station congestion of a week

Conclusion and **Recommendations**

Commuter traffic is **most congested at end stations**

7AM-8AM and **5PM-6PM** are the **busiest hours**

Friday is the worst day to ride the MRT3

Skip trains can be reinforced during peak hours to **prioritize the end stations** and opt to skip Santolan and GMA Kamuning stations due to their low commuter traffic activity.



Limitations

- 1 The dataset used is from 2017. The results might not reflect the current trends and may not be accurate for 2019.
- 2 Geospatial Dataset did not have the Ayala station
- 3 The hours provided were in time-range (XX:00 - XX:59) format. A more precise time series analysis could not be possible.



Hamoy Ken