



Grabbing INSIGHTS

Conducting deeper analysis with
the use of Grab-Posisi dataset

Meet Team K.I.M.



Kim

NTU Computer Science & Business



Iffah

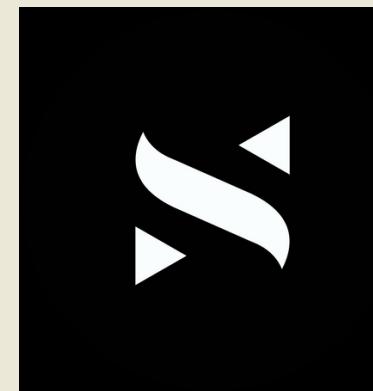
NTU Chemistry & Biological Chemistry



Marcus

NUS Data Science & Analytics

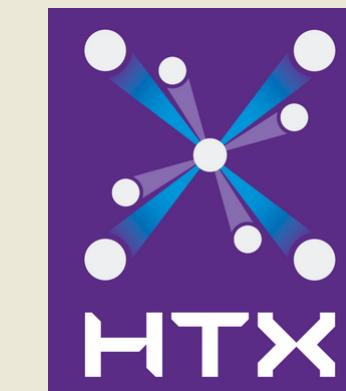
RELEVANT EXPERIENCES:



SUBX



ST Logistics



**Home Team Science and
Technology agency**



Kaggle Competitions

WHAT IS THE ISSUE?

Feedback about **delays** in Grab transport services in both Jakarta and Singapore



decrease in customer satisfaction and trust



increase customer turnover



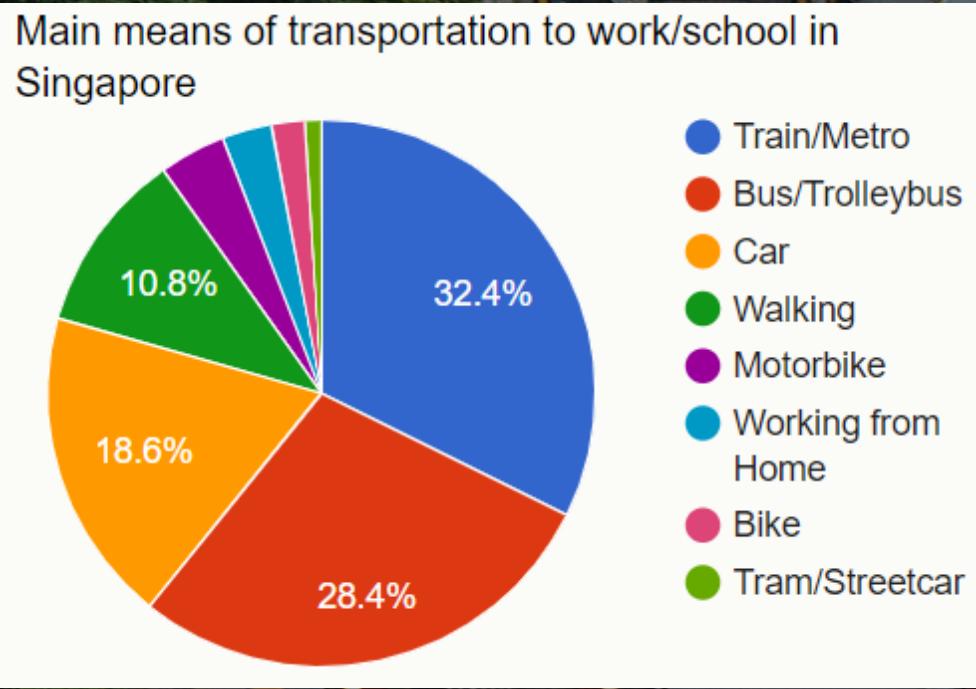
OBJECTIVE

Generate insights and provide recommendations to help Grab alleviate these issues

Road traffic in Singapore and Jakarta

Singapore's Population in 2021

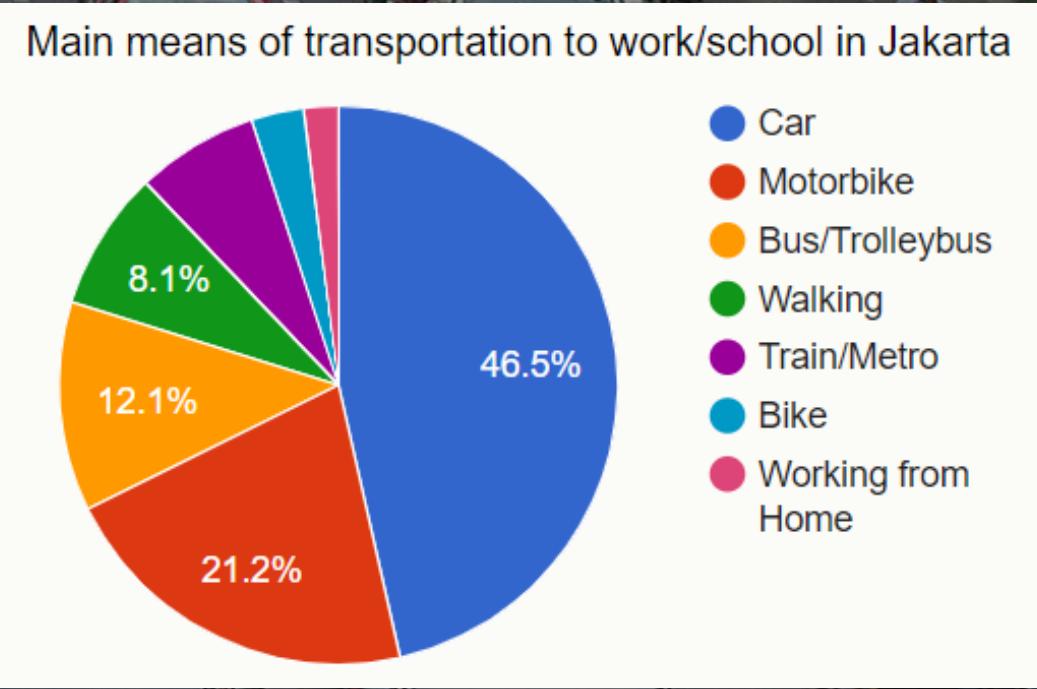
5,886,820



- Implemented numerous private transport charges (COE, ERP) to curb congestions
- Built numerous MRT system to alleviate this issue

Jakarta's Population in 2021

10,135,030



- Poor traffic conditions with frequent unpredictable traffic jams
- Recently built a new MRT system to alleviate this issue

Sources: (all-populations, 2022), (worldpopulationreview, 2022), (numbeo, 2022)

Our initial hypothesis

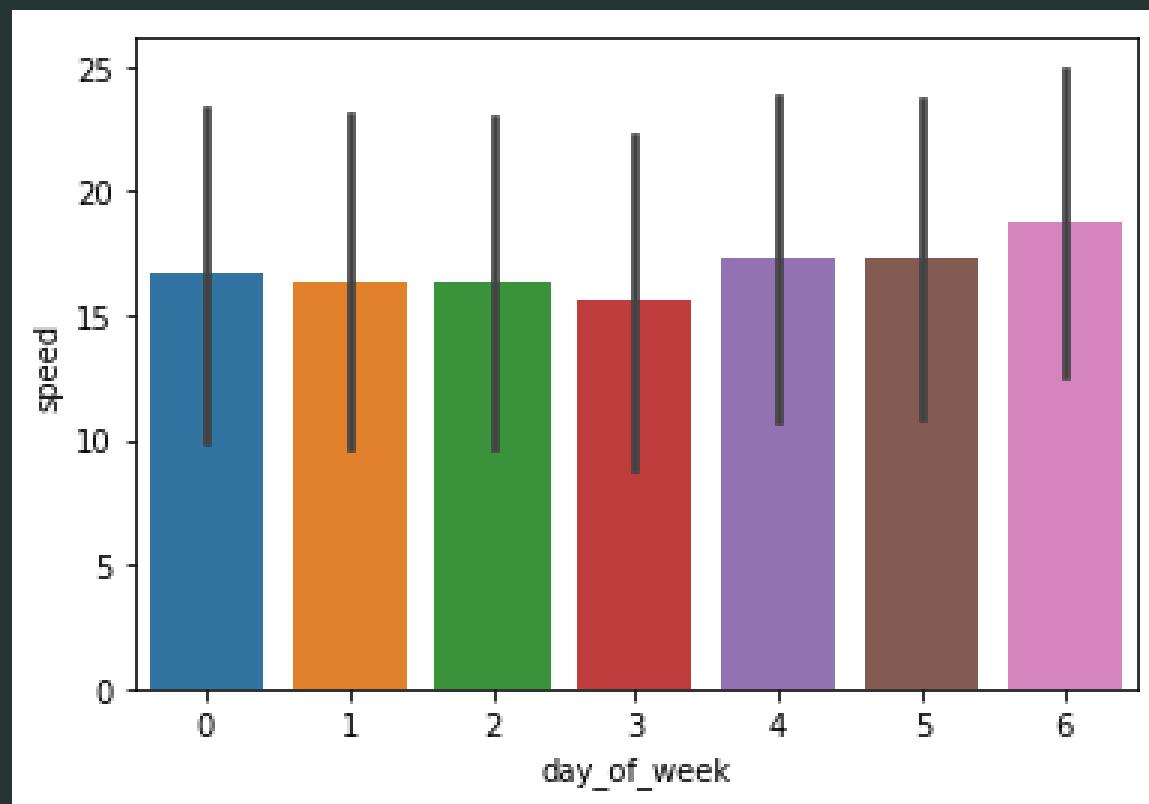
Possible reasons for the delay

- High population density in a particular location which leads to poor network transmission signals
- Unpredictable traffic congestions throughout different time of the day

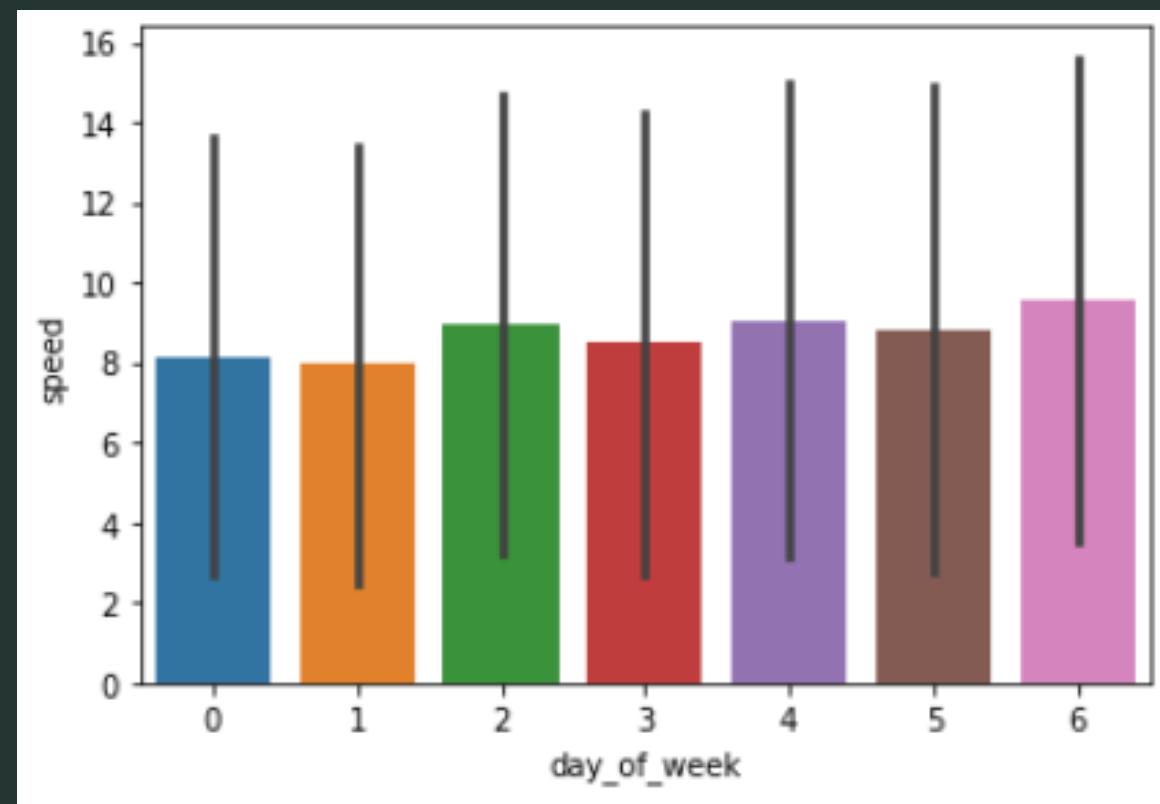
Correlation between speed and day of the week

--- vertical line represents standard deviation (s.d.)

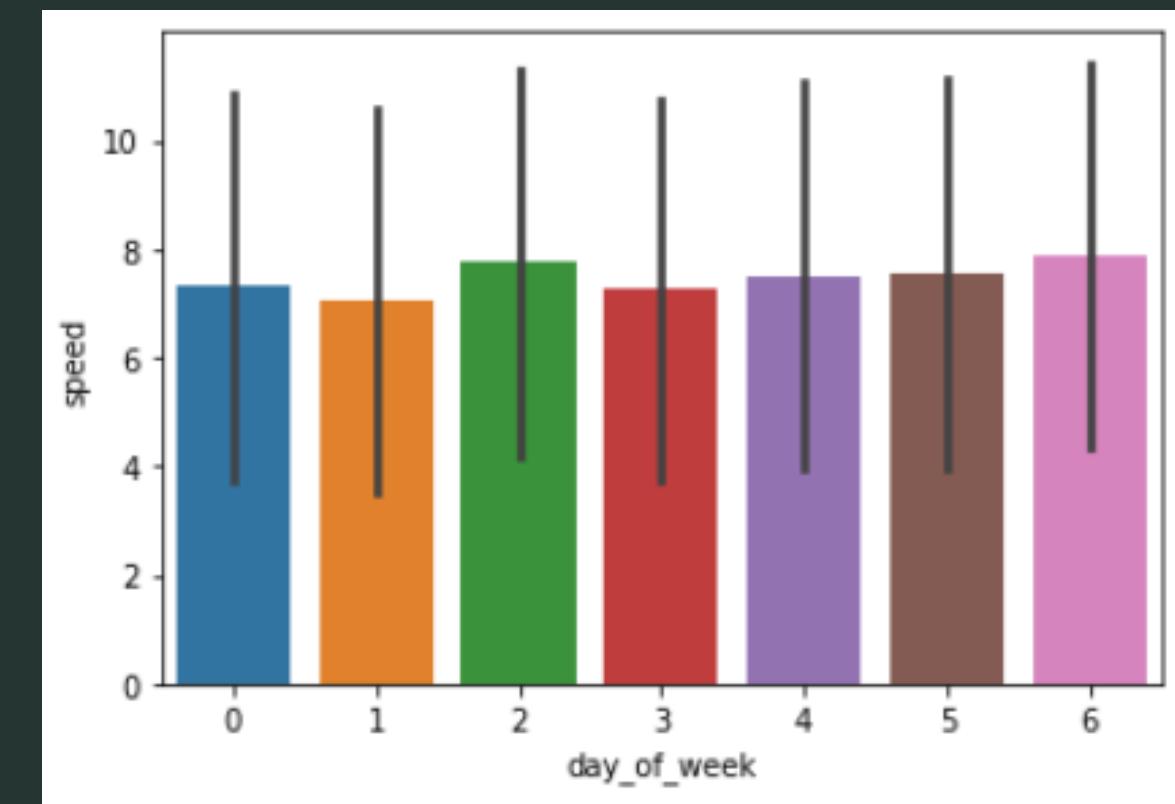
bar graph represents mean speed



Singapore (Car)



Jakarta (Car)



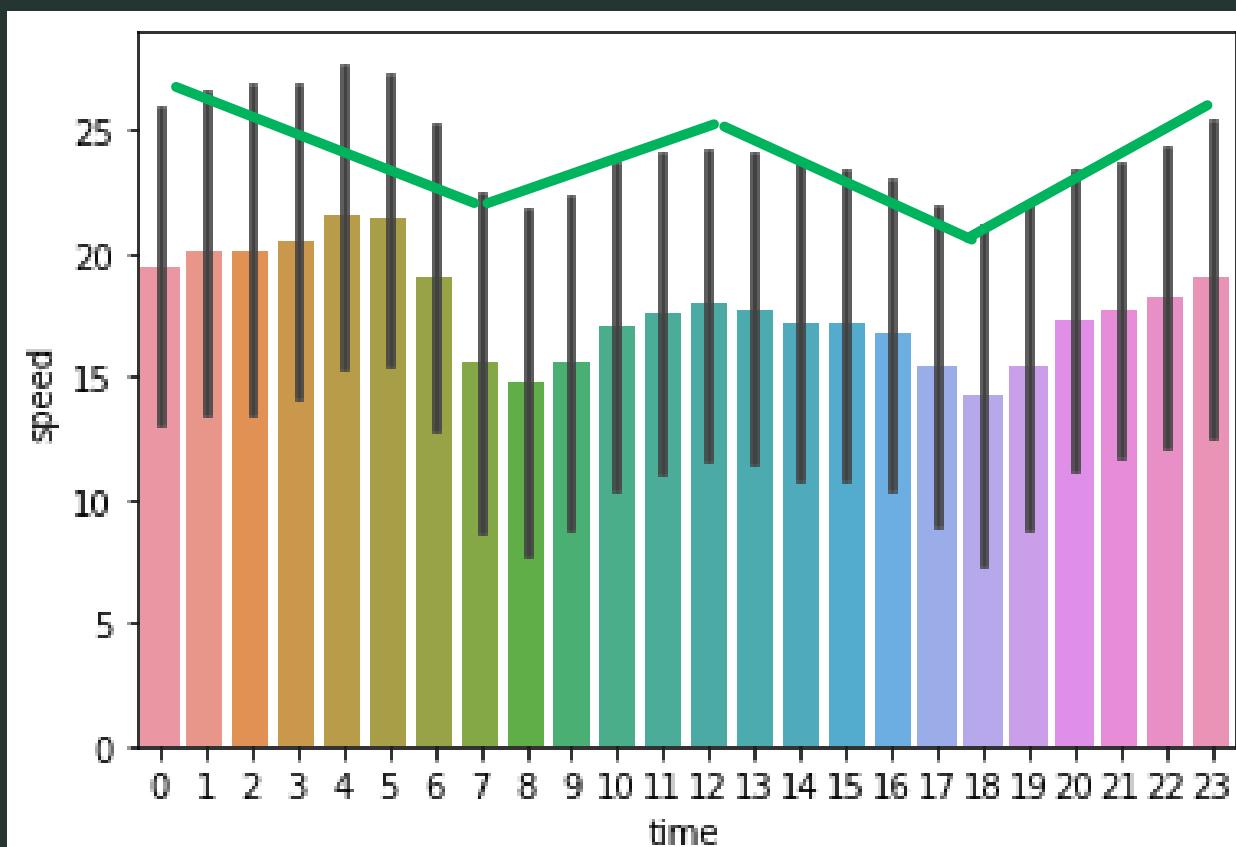
Jakarta (Motorcycle)

- Speed is relatively constant
- Approx. similar s.d. throughout the week

Correlation between speed and time of day

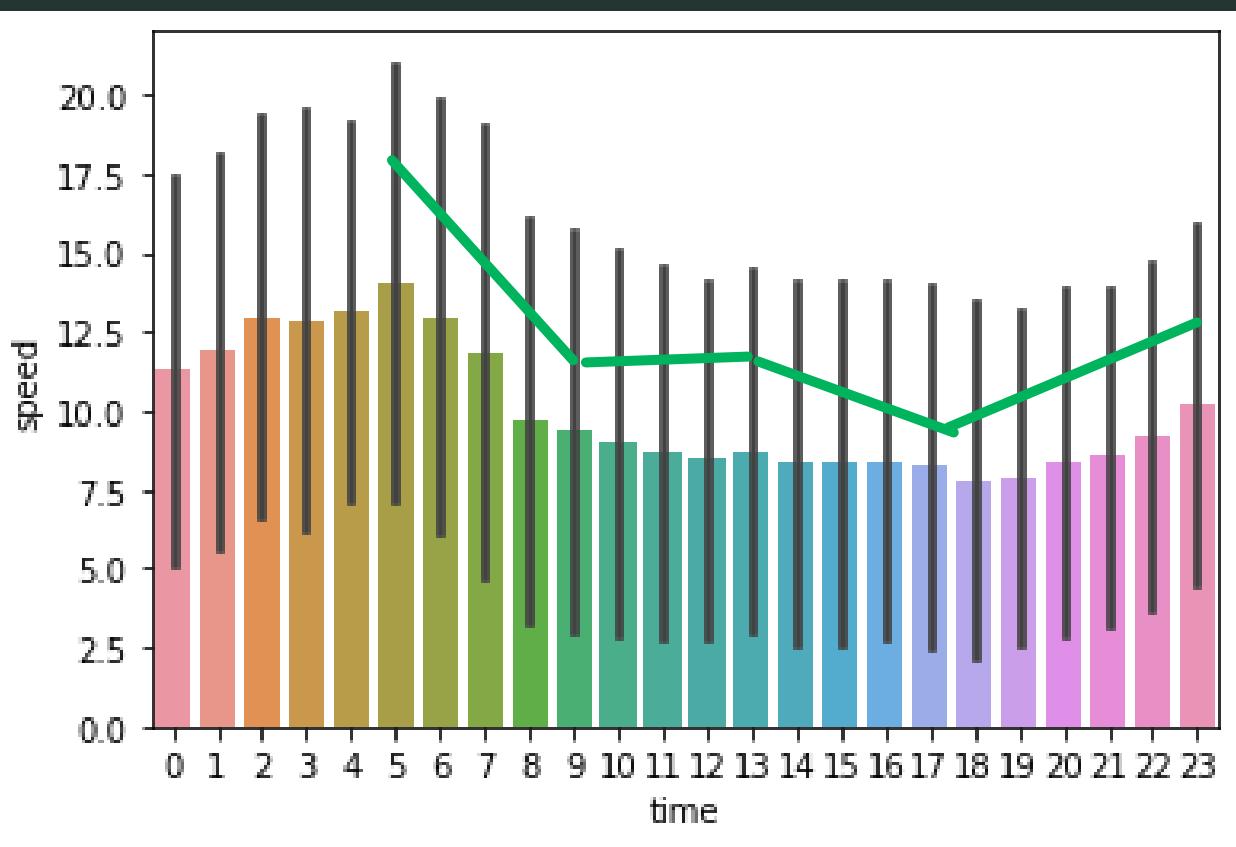
--- vertical line represents standard deviation (s.d.)

bar graph represents mean speed



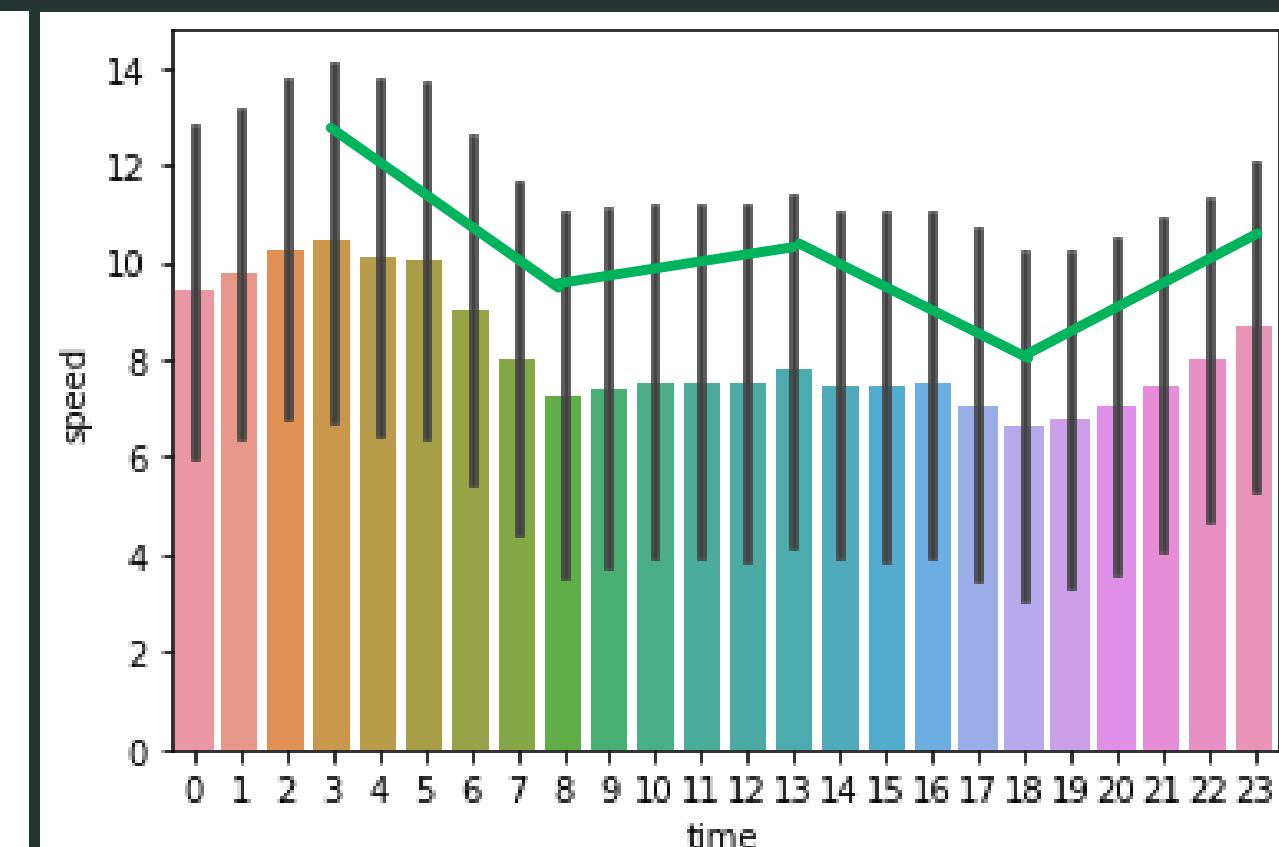
Singapore (Car)

- Slower speed at 8am & 6pm
- Approx. similar s.d. of 12.5



Jakarta (Car)

- Slower speed from 8am-8pm
- Approx similar s.d. of 12.5

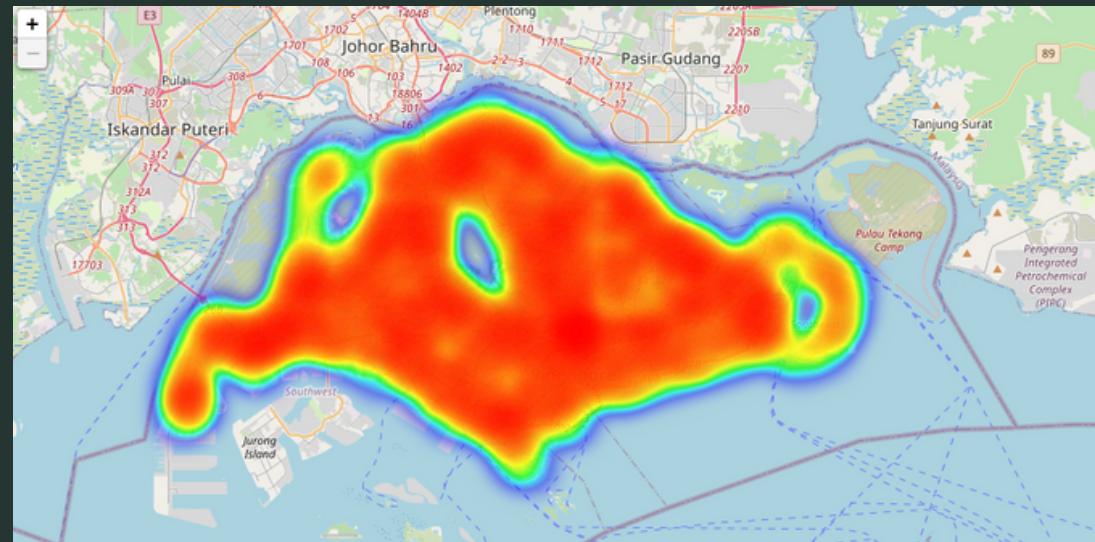


Jakarta (Motorcycle)

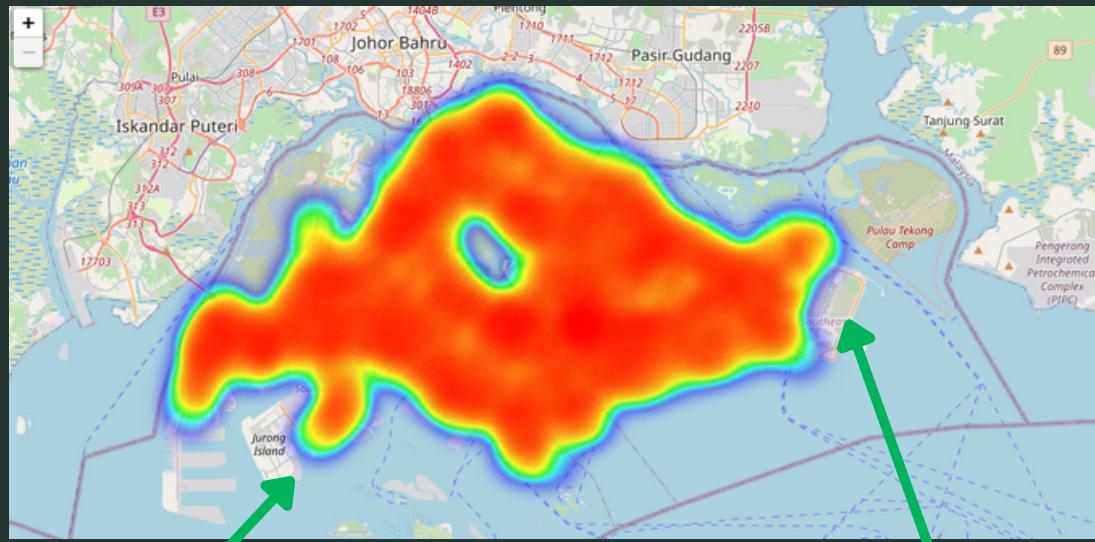
- Slower speed at 8am & 6pm
- Approx similar s.d. of 7

Grab traffic density in Singapore

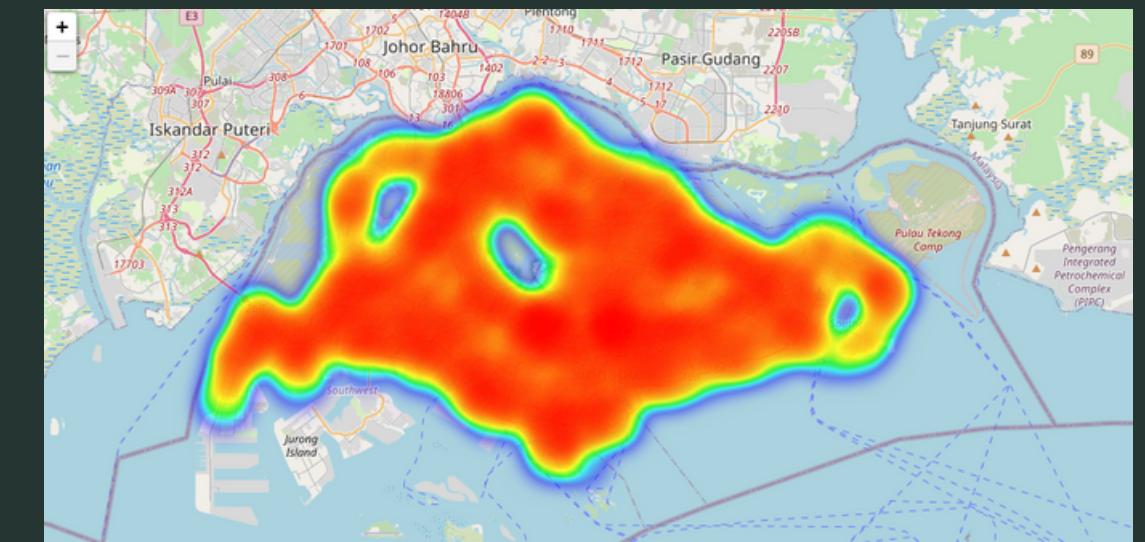
Monday



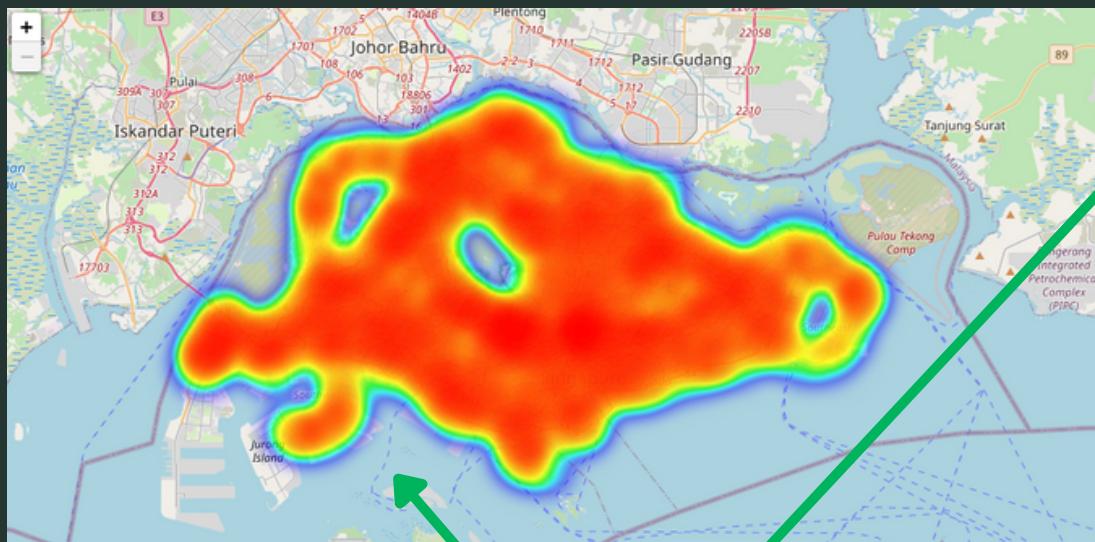
Tuesday



Wednesday



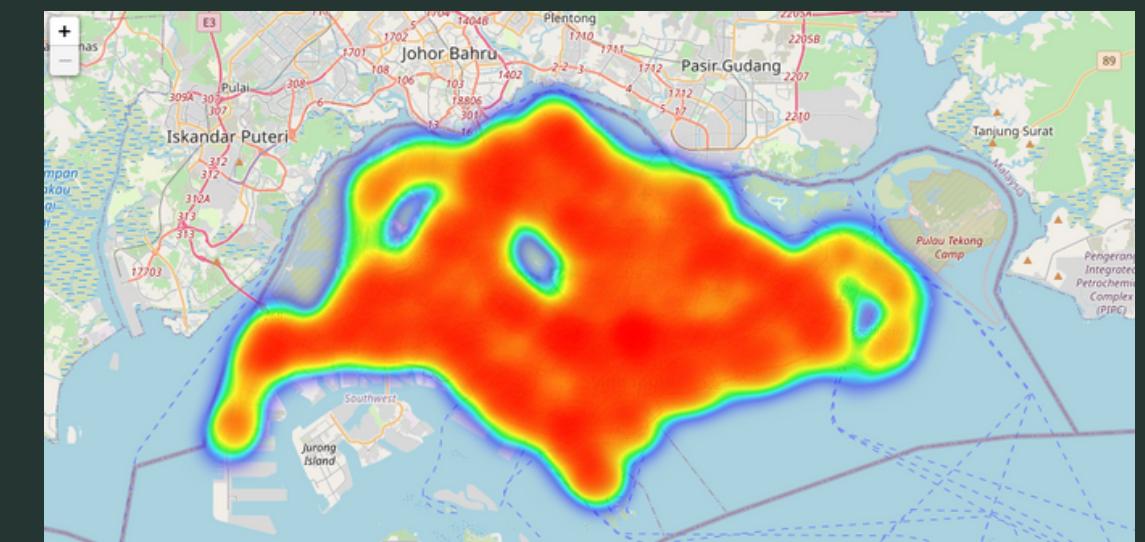
Thursday



Friday



Saturday



higher demand at Jurong Island
(possibly due to allocated shipping days)

lower demand at Changi area
(possibly due to limited departure and arrival flights)



PREDICTIVE MODEL

Gradient Boosting

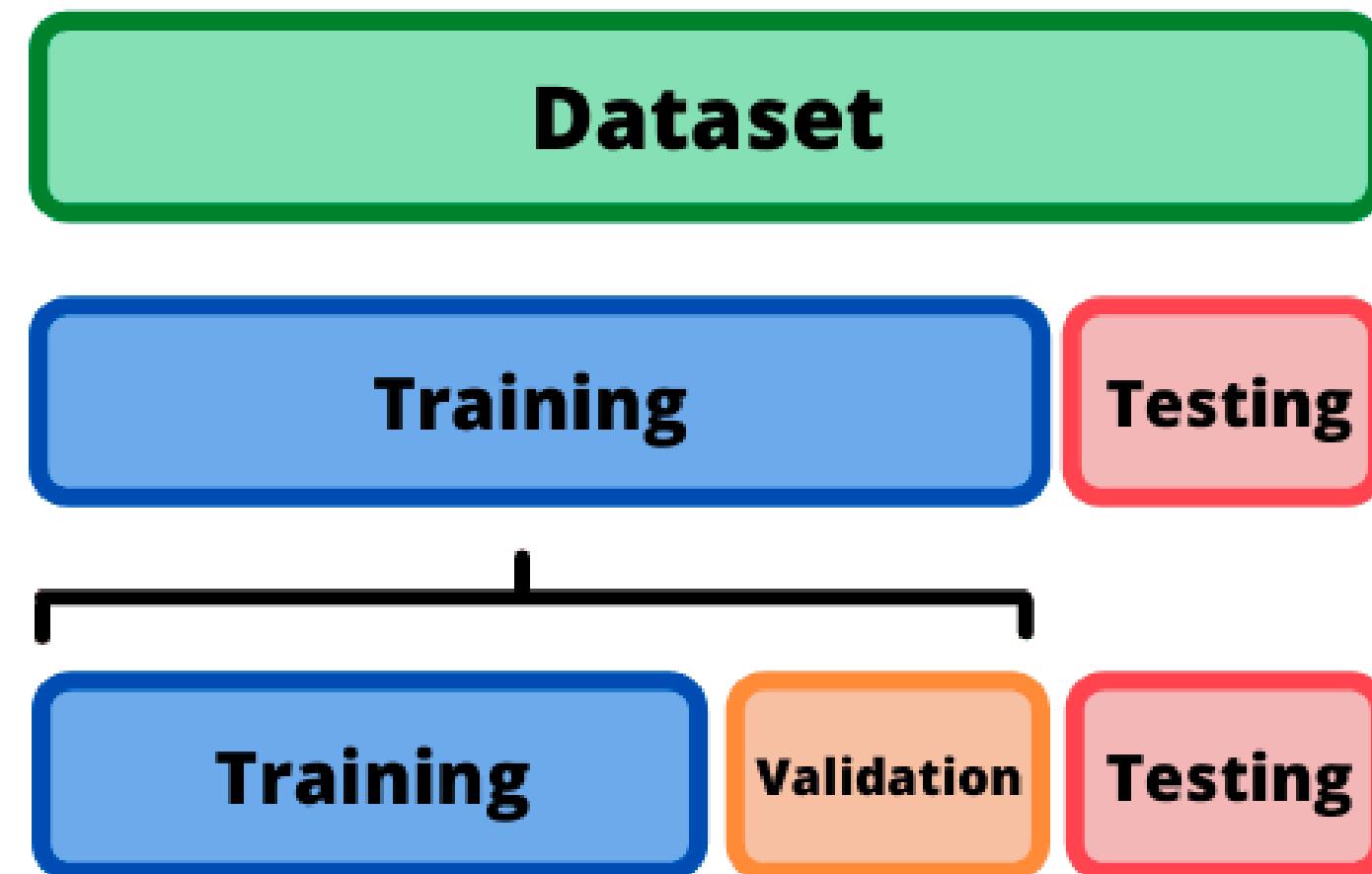


Grab

Dave

Data Preparation

- Need training data, validation data & test data
- 60-20-20 split ratio

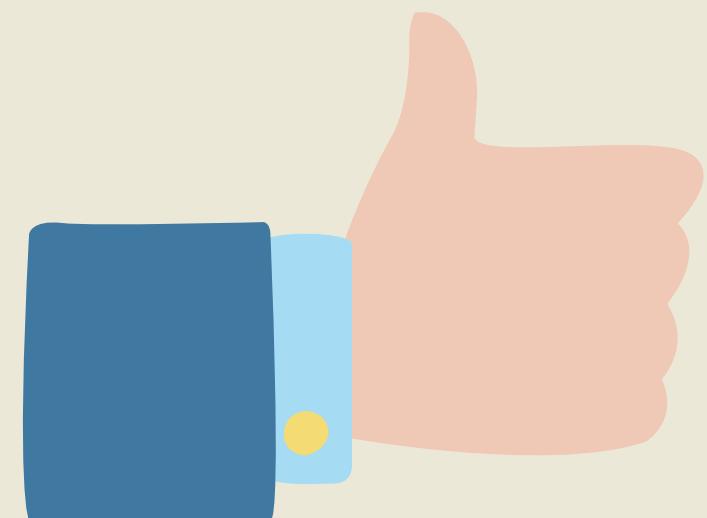


Feature Engineering

- Target Variable: 'speed'
- Drop all negative values for speed
- Features for Singapore:
'rawlat', 'rawlng', 'time', 'day_of_week'
- Features for Jakarta:
'rawlat', 'rawlng', 'time', 'day_of_week', 'driving_mode'
- One hot encoding for 'driving_mode'

Training Model

- Gradient Boosting (ensemble method)
- Mathematically robust
- Insensitive to feature normalization/standardization
- Wide variety of hyperparameter to tune to potentially improve our accuracy



Singapore Model

Initial mean error

- training: 5.36
- validation: 5.36
- naive: 7.17



Test error after hyperparameter tuning (grid-search)

- training: 3.73
- validation: 4.60
- test: 4.83

A reduction from our initial error values and a 32% advantage over the naive model

Jakarta Model

Initial mean error

- training: 4.38
- validation: 4.38
- naive: 5.17



Test error after hyperparameter tuning (grid-search)

- training: 2.71
- validation: 3.88
- test: 3.87

A reduction from our initial error values and a 25% advantage over the naive model

Model Analysis

- Common to both models is that latitude and longitude play a large role in predicting the speed of driver

	Feature	Importance
1	rawlng	0.394675
0	rawlat	0.373960
2	time	0.145504
3	day_of_week	0.085861

Singapore

	Feature	Importance
1	rawlng	0.279888
0	rawlat	0.278713
4	car	0.233970
2	time	0.116278
3	day_of_week	0.091151
5	motorcycle	0.000000

Jakarta

- Time seems to be the next most important factor
- Jakarta model shows that maybe the driving_mode may contribute greater

Model Analysis



Raffles City, Friday evening, 6pm
5.65m/s or 20.35km/h

Model Analysis



Gandaria City, Kebayoran, Friday evening, 6pm

6.10m/s or 21.97km/h for cars

6.12m/s or 22.04km/h for motorcycles

Limitations

1. Target Variable

- lower speed might not necessarily predict congestion
- density could have been an option we could have explored

2. Model not up to expectations

- unsatisfactory level of confidence



— Further Improvements —

1. More representative target variable

- could have averaged out the speed values (smoothing of values)
- density - density of riders in a specific region at specific time intervals

2. Different models

- non-linear SVM would have been a great model too
- more features for our models

3. Feature Variables

- using minutes as a feature could have helped by adding more specificity