TAXI DEMAND FORECAST & ANALYSIS

By The Powerpuff Girls

PREDICTION OF TAXI DEMAND OF A CITY

Summary

1

Summary

- **73** zones
- **Hourly taxi trips** of each zone
- **Weather** data
- **Neighbouring zones** of each zone

Prediction of trips with machine learning techniques!

WORKFLOW

Exploratory Data Analysis

Feature Engineering

Model Training

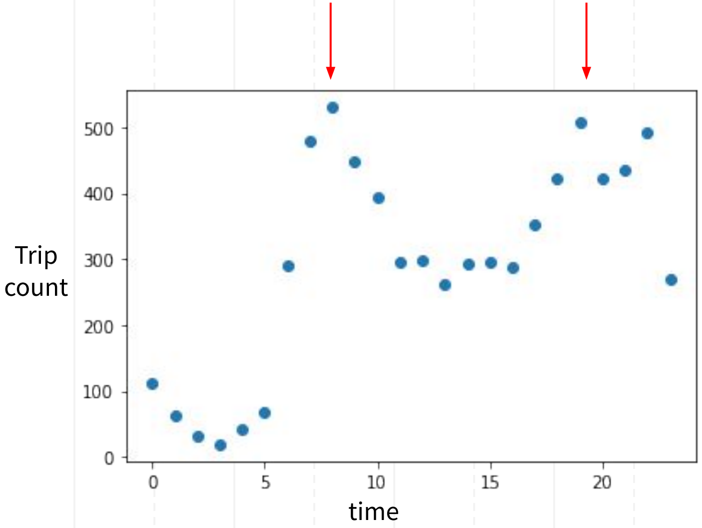
Exploratory Data Analysis And Feature Engineering

What did we learn from the data?

2

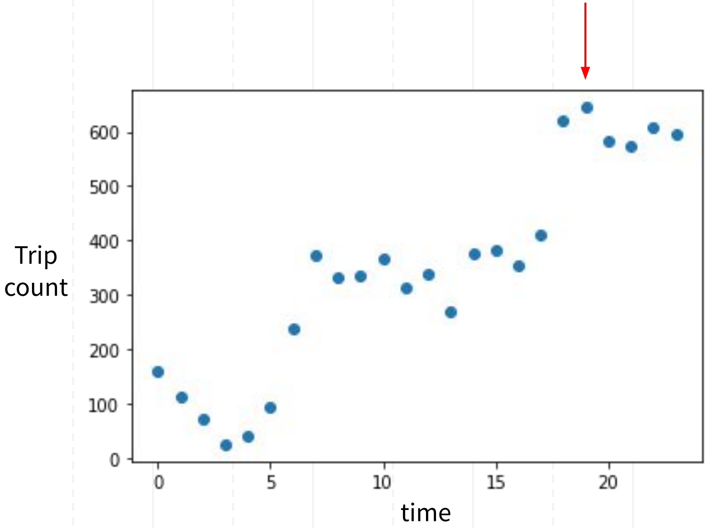
Hourly Trip Count

PEAK TRAFFIC HOUR DETECTION



Weekday

- 06 - 10 am
- 04 -08 pm



Weekend

- 4 -8 pm

FEATURES SELECTED

Weekday / Weekend

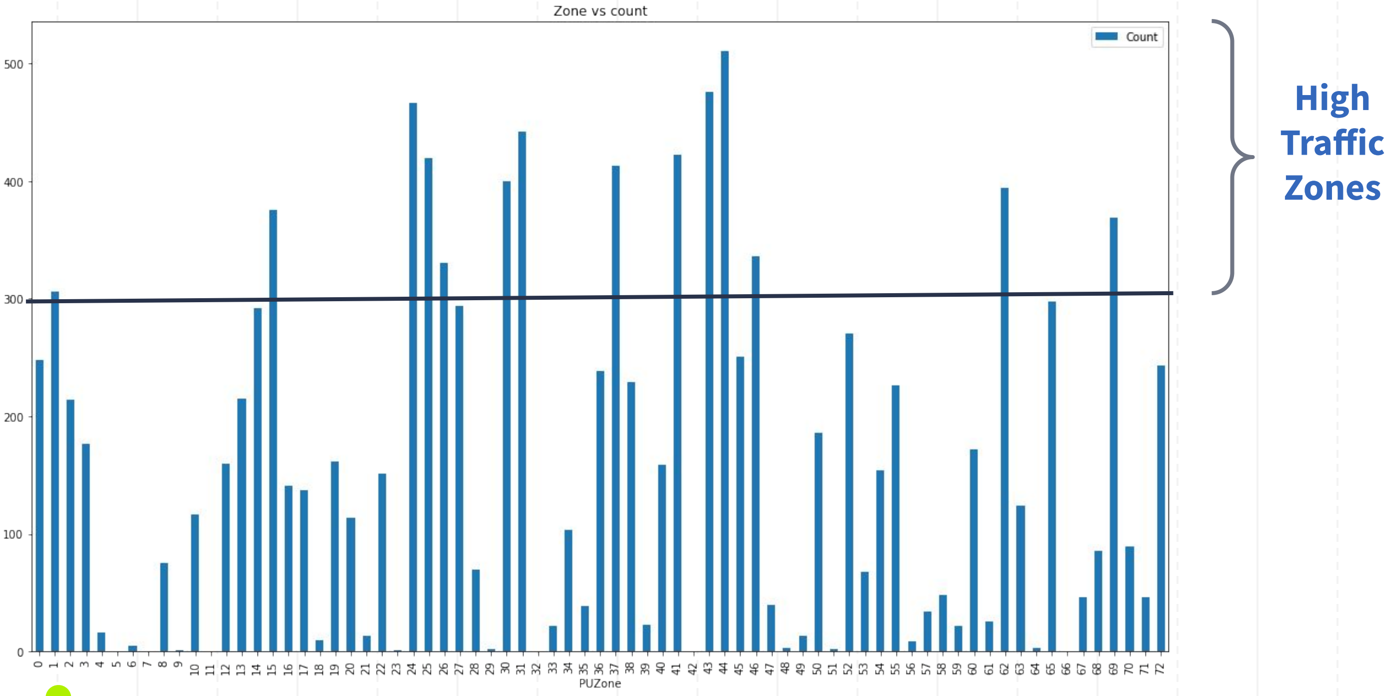
- Average traffic **higher in weekdays** compared to weekends

FEATURES SELECTED

Peak Traffic Hours

- **Weekdays**: 6 - 10 AM and 4 - 8 PM
- **Weekends**: 4 - 8 PM

HIGH TRAFFIC ZONES DETECTION
(Average Trip Count > 300)

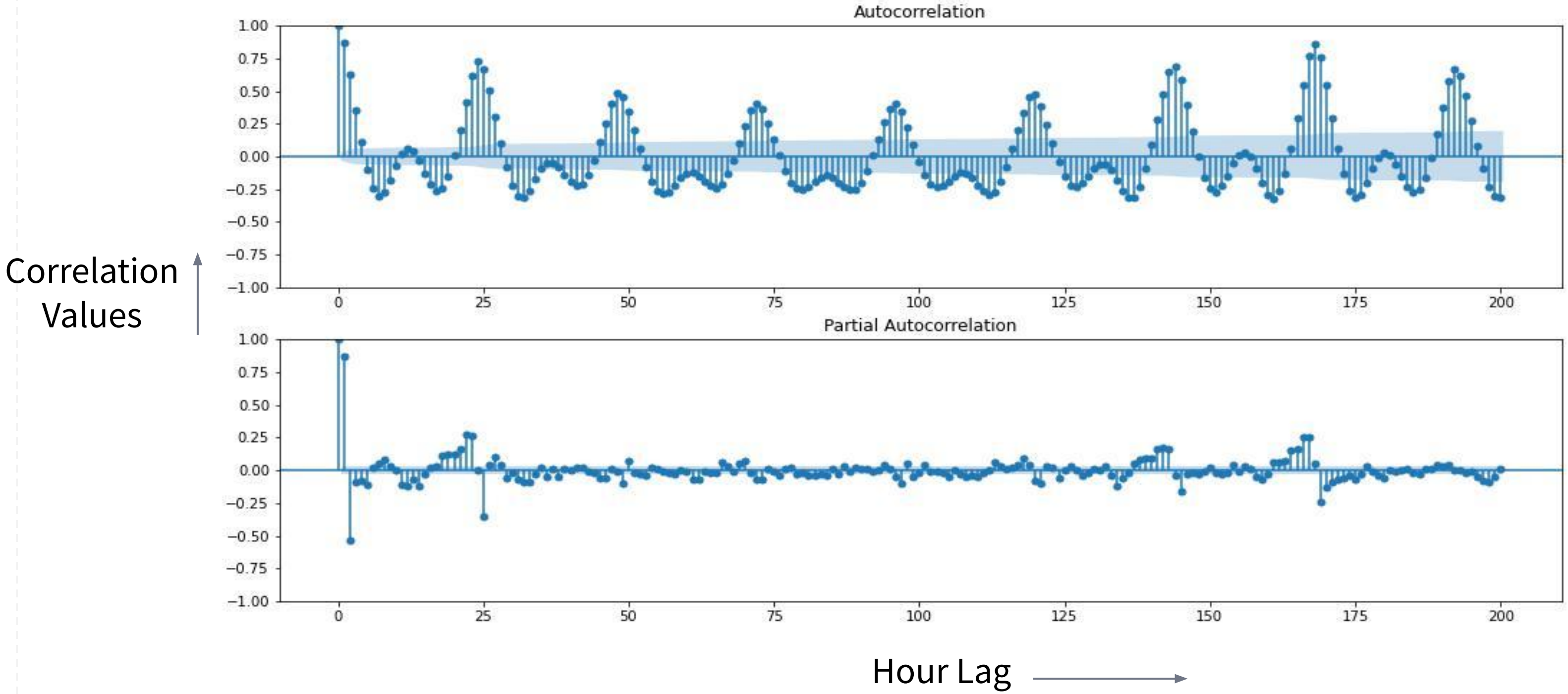


FEATURES SELECTED

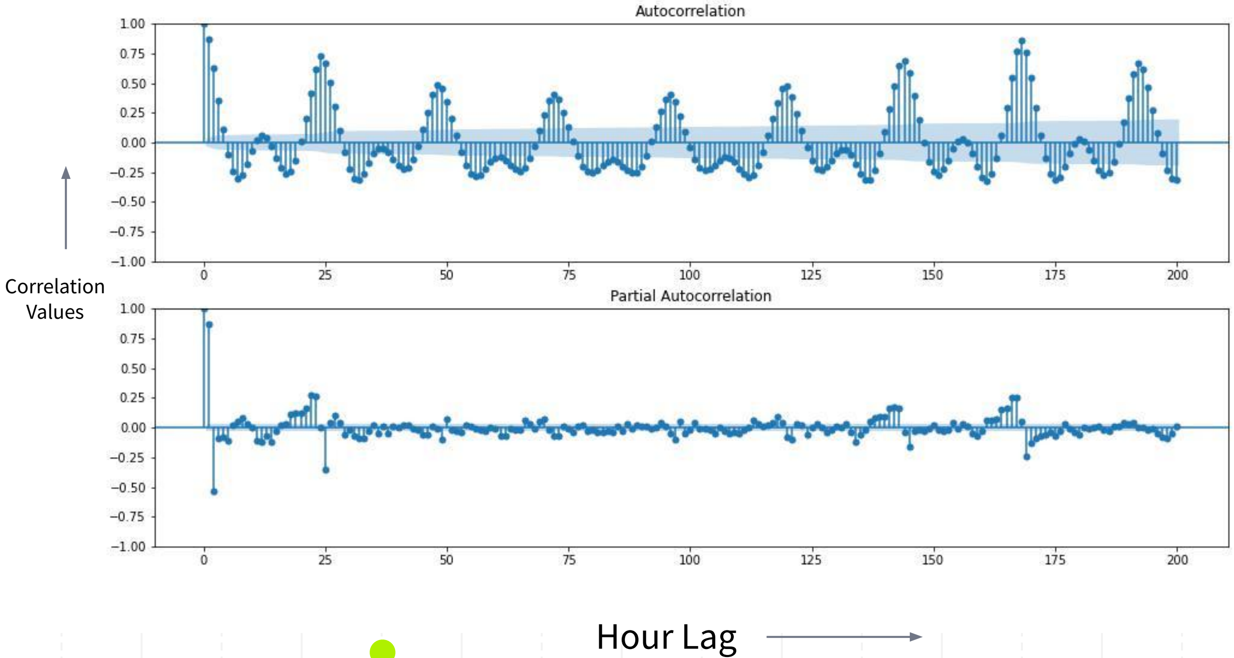
High / Low Traffic Zones

- A new feature created to indicate if an area usually has higher or lower average traffic.

TREND DETECTION OF HOURLY TRIP COUNT WITH ITS PREVIOUS HOURS



TREND DETECTION

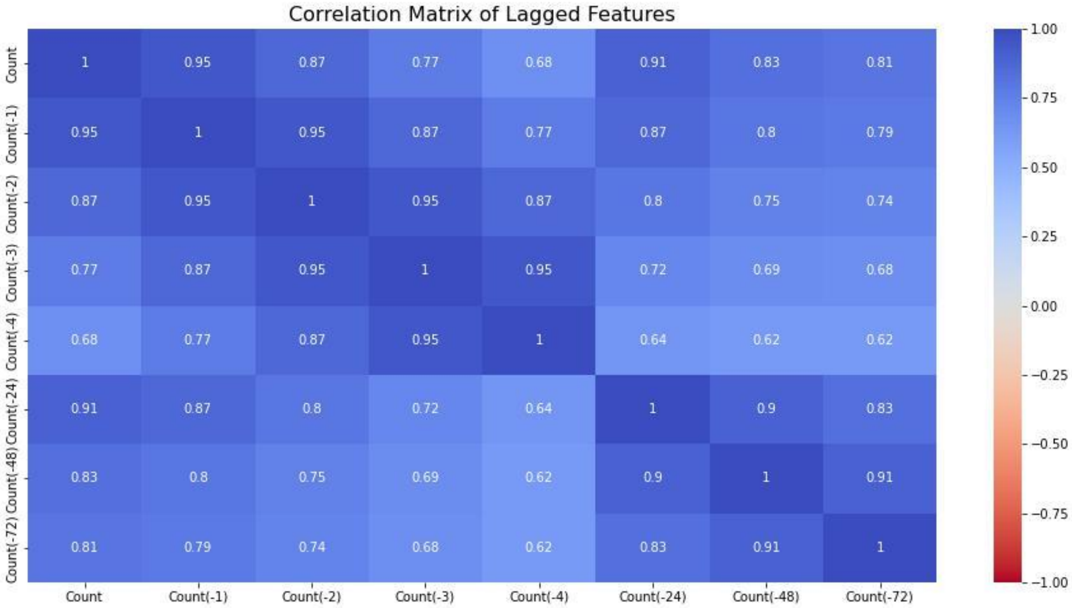


- Periodic tendency of **24 hours**
- Trip count is highly correlated with trips of **previous few hours**

Correlation With Trips of Previous Hours - VERY HIGH

Correlation of trips with trips of

- 1 hours ago
- 2 hours ago
- 3 hours ago
- 4 hours ago
- 24 hours ago
- 48 hours ago
- 72 hours ago



FEATURES SELECTED

Past Traffic Demands

- Today's demand : **Last 24 hours' data**

  Value of trip counts: 1 hour ago, 2 hours ago, ..., 24 hours ago

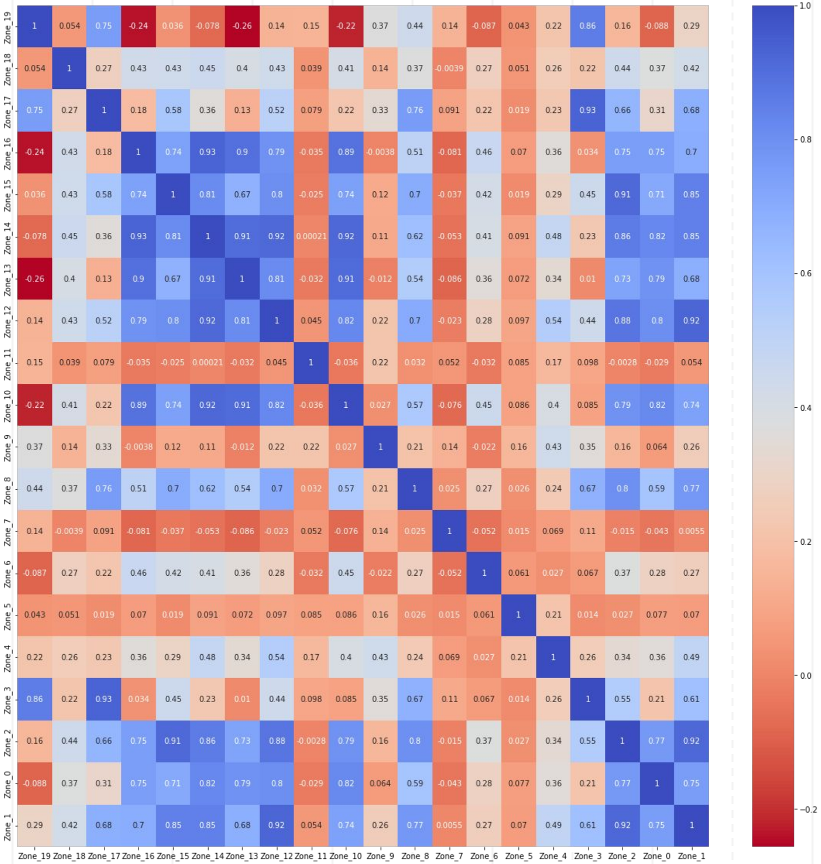
- Hour's demand : **Last 30 days' data for similar hours**

  Value of trip counts: 1 day ago, 2 days ago, ..., 30 days ago

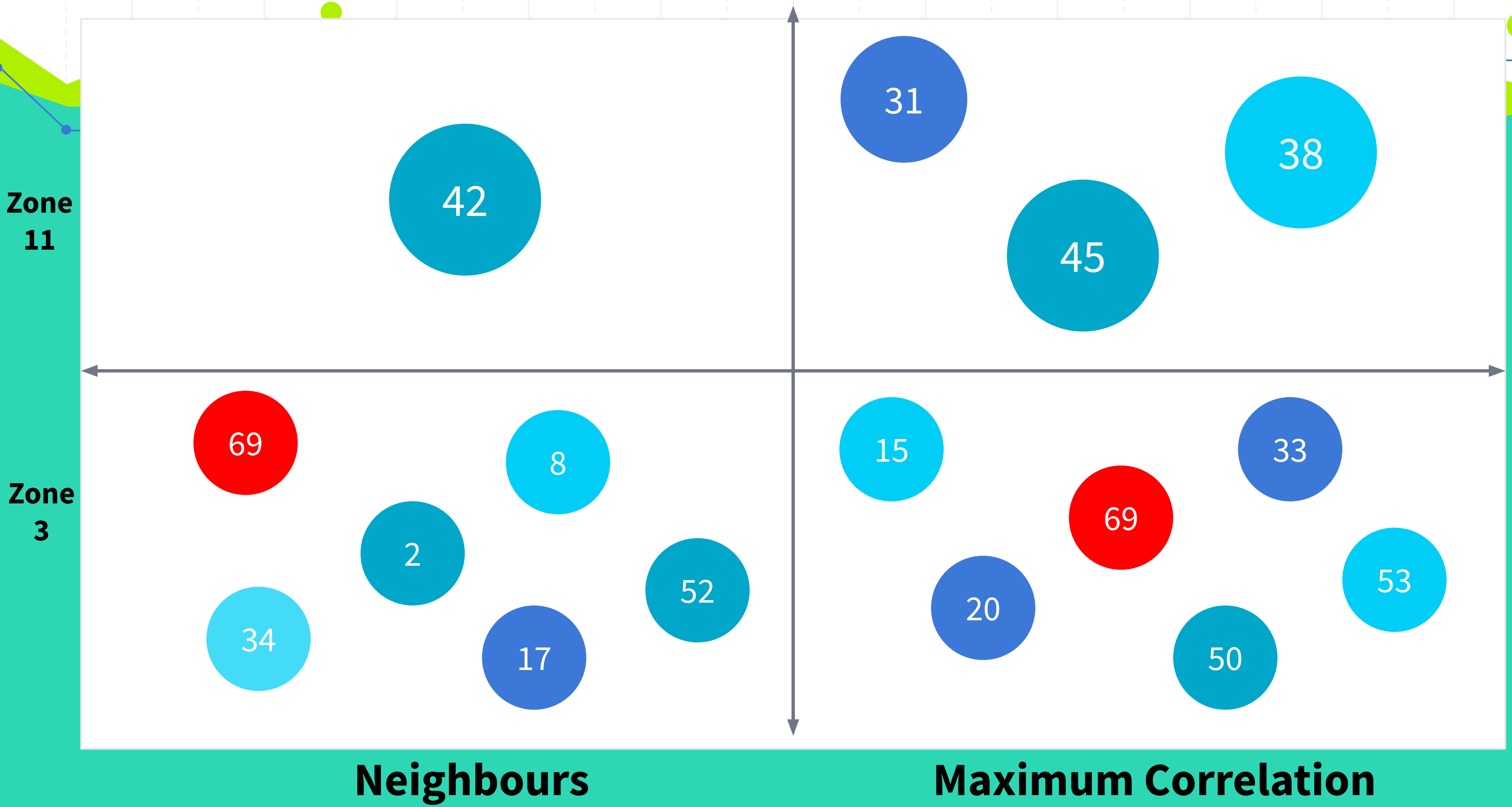
Neighbours of each Zone

73 Zones

CORRELATION BETWEEN ZONES’ TRIP COUNTS

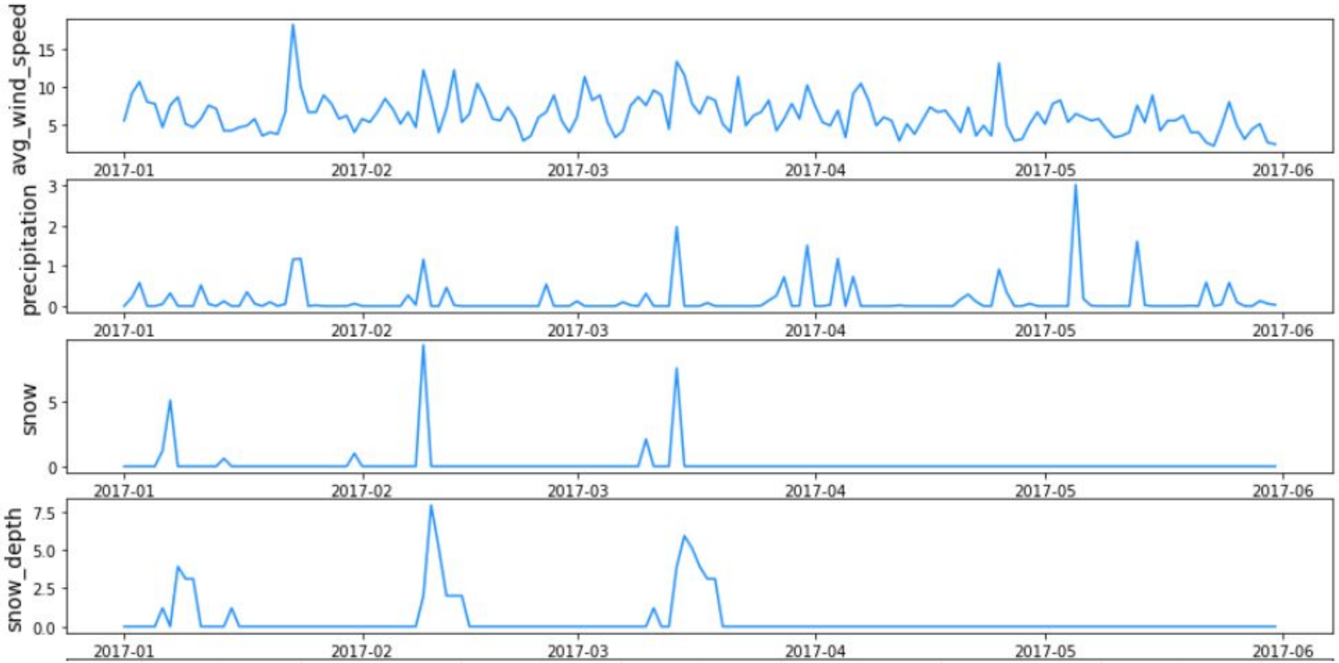


Neighbours VS Maximum Correlated Zones



Weather

WEATHER DATA



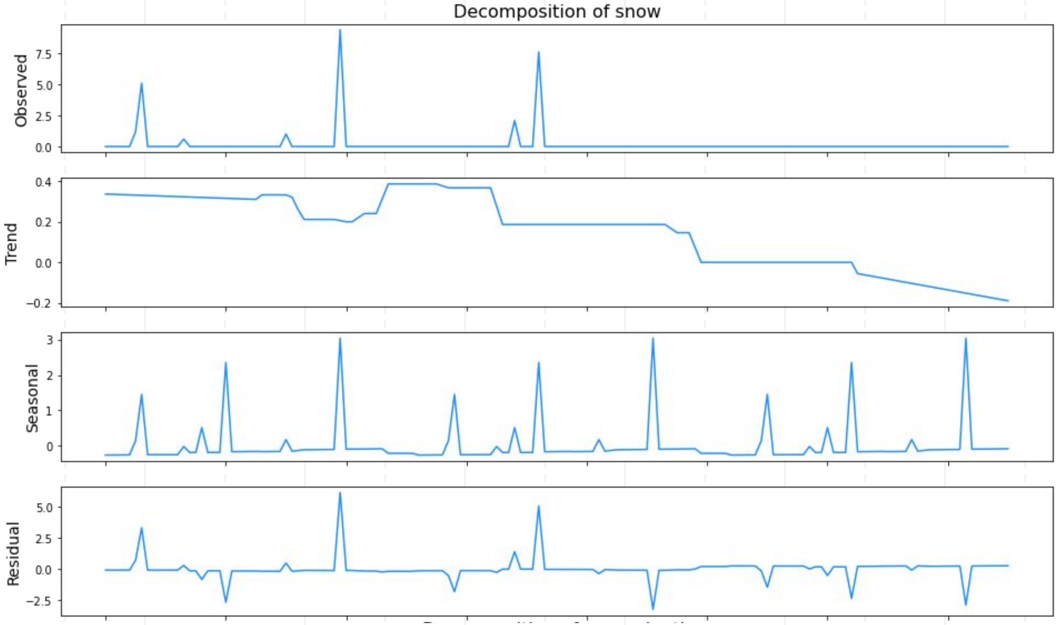
SEASONAL DECOMPOSE
SNOW

OBSERVED

TREND

SEASONAL

RESIDUAL



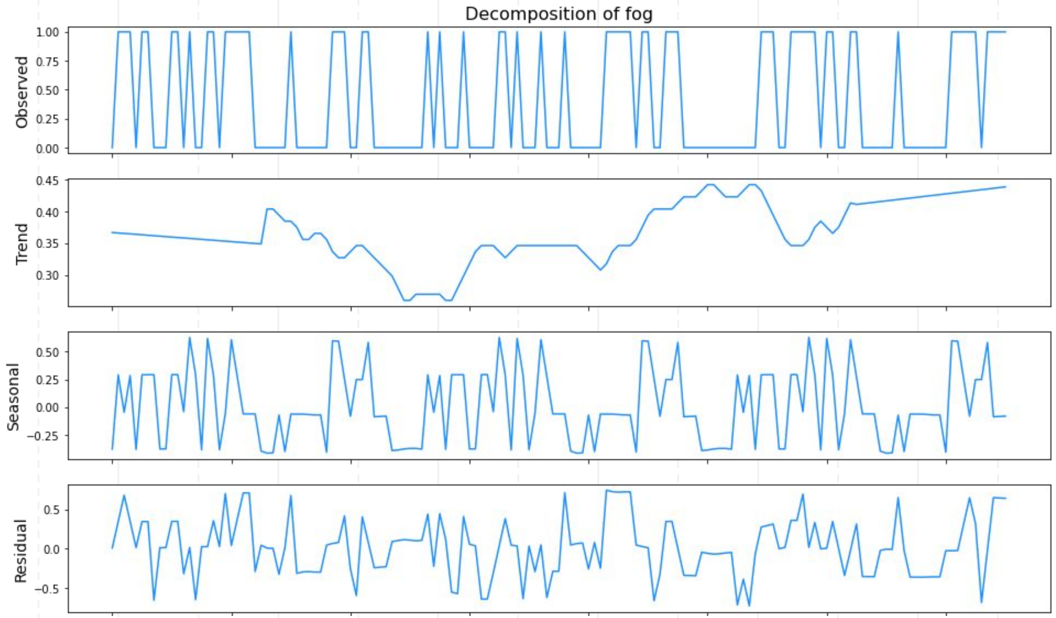
SEASONAL DECOMPOSE
FOG

OBSERVED

TREND

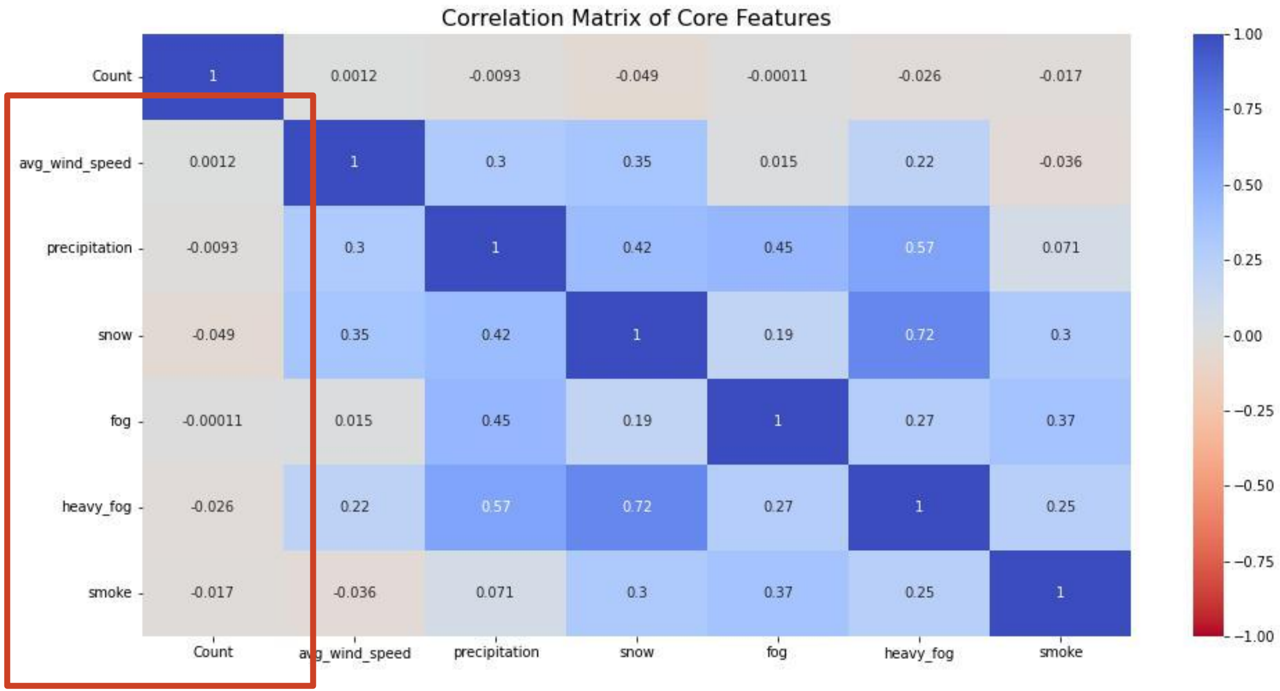
SEASONAL

RESIDUAL



CORRELATION OF TRIP COUNTS WITH WEATHER

Low Correlation



CORRELATION OF TRIP COUNTS WITH WEATHER

Most correlated weather parameters:

- Snow
- Depth of snow
- Fog
- Heavy fog
- Smoke

LIST OF THE FEATURES USED IN MODELS

- Previous 24 hours trip count
- Previous 30 days trip count for same hour
- Weekend / Weekday
- Peak traffic hour or not
- High traffic zone or not
- Weather: Snow, snow depth, fog etc.

ADDITIONAL FEATURE ENGINEERING TECHNIQUES

- Imputation
- Date Extraction
- Grouping
- Creating new features

Training, Testing & Forecasting

3

HOW DID WE CHOOSE OUR MODEL?

Statistical Time Series Prediction Models

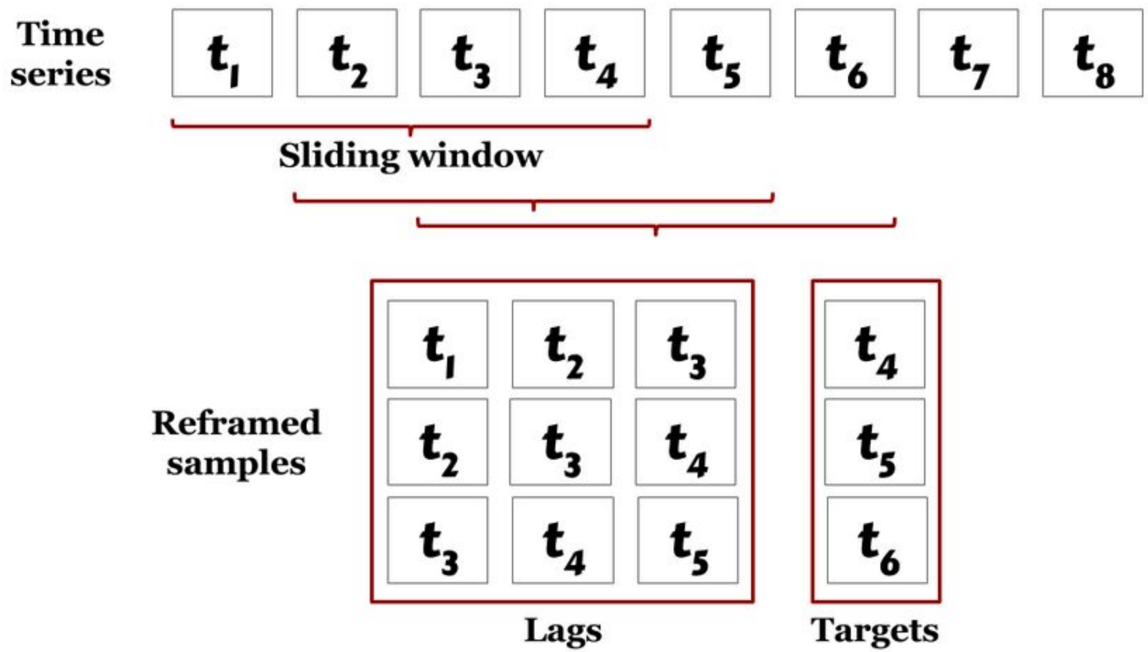
TWO TYPES OF MODELS

Pure Machine Learning Models

- ARIMA, SARIMA, Prophet etc
- Classical statistical approaches
- **Handles continuous data itself.**

- Xgboost, Random Forest etc
- More flexible and can learn complex relations
- **Needs lag variables and seasonal variables.**

How To Convert a Time Series Problem to a Regular Supervised Learning Problem?



Our Models

We ensembled four models using Voting Regressor

Xgboost

- Reigning king of regression problems

MultiLayer Perceptron

- Neural network
- Capable to learn non-linear problems

XGB MLP LGB RF

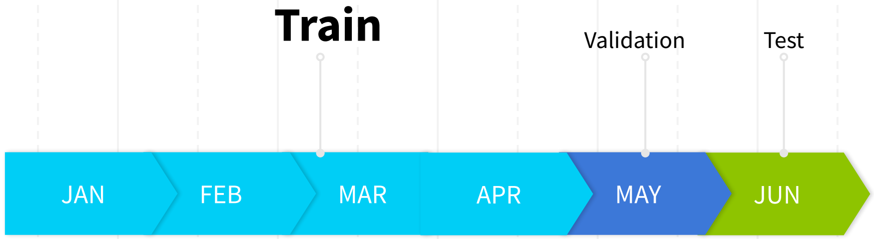
- Faster training speed and accuracy

LightGBM

- Runs efficiently on large data bases

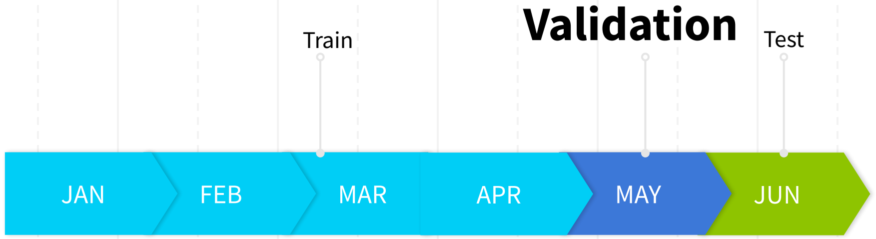
Random Forest

TRAINING

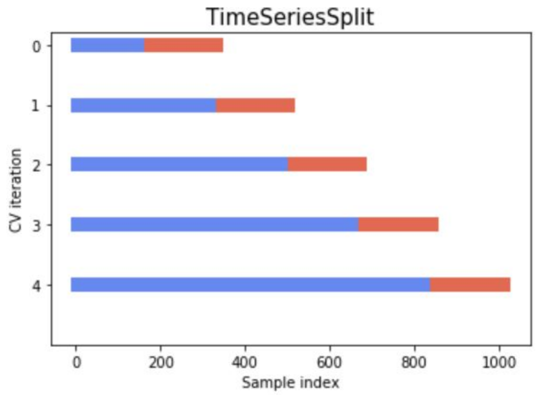


- **Data of 4 months** for training
- An **ensemble model** of Xgboost, Lightgbm, MLP and Random Forest
- Parameter **Tuning**

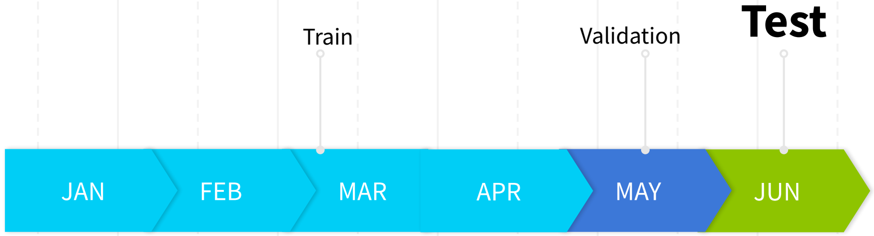
VALIDATION



- **Data of 1 month** for validation
- **Cross validation** using time series split
- **Mean absolute error** as metric



PREDICTION



- We predicted the trip counts for the month of June
- **Mean absolute error** as metric

4 Results & Discussion

Performance Metric - Mean Absolute Error

Predicted value

Actual value

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Number of samples

Mean Absolute Error

14.63

The Given Benchmark was 19.6

WHAT INSIGHTS DID WE GAIN?

PEAK TIME

Weekdays

- **Morning: 6 - 10 am**
  - Offices / schools starting time.
- **Evening: 4 - 8 pm**
  - Everyone returns home.

Weekends

- No rush in the morning.
- **Evening: 4 - 8 pm**
  - Weekend activities.

HIGH TRAFFIC ZONES

Different types of **land use zones**

Commercial : Offices, restaurants, shops etc

Industrial : Factories, Warehouses

Public Use : Hospitals,

Educational institutions, Municipal buildings

NEIGHBOURING ZONES

- Neighbouring zones of **different zone types** have low impact on each other.

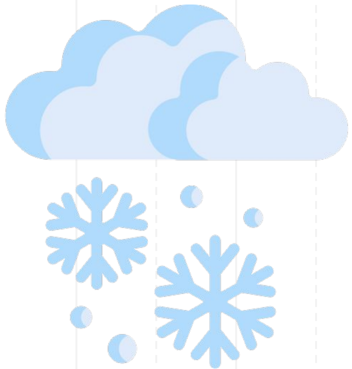
  (Less impact of traffic in a commercial zone on its neighbouring residential area)

- Highly correlated zones might be of the **same kind**.

DEPENDENCY ON TEMPERATURE, PRECIPITATION

- Since temperature or precipitation **do not directly hamper** condition of roads, they **don't affect trip count**

DEPENDENCY ON HEAVY FOG, SNOW



Negative Correlation

- **Snow** can affect condition of roads
- **Fogs, heavy fogs** or smoke can hamper the visibility of drivers
- People do not go out of the house in snow
- These can reduce trip count and taxi demand

Impact of Our Work

5

POTENTIAL IMPACT

- Prediction of taxi demand of an area can help predict the price of a taxi trip, or help with booking a taxi early.

POTENTIAL IMPACT

- Taxi companies can build their business model using this type of prediction system. They can offer deals to customers, come up with a proper pricing model, etc.

POTENTIAL IMPACT

- Collaborating with other stakeholders, such as government agencies for developing better infrastructure for high traffic areas

POTENTIAL IMPACT

- Short-term traffic prediction provides tools for improved road management by allowing the **reduction of delays, incidents and other unexpected events**
- Helps predicting **drivable speeds**

OTHER STUDIES

“

Karimpour, M., Karimpour, A., Kompany, K. and Karimpour, A., 2022. Online Traffic Prediction Using Time Series: A Case study.

- Predict the traffic flow for a **certain intersection**, and **control the signaling** of that intersection
- Using this method for Moallem Blvd. in Mashhad demonstrated that the model is able to predict the traffic flow with 88.74% and 81.96% accuracy for 15 minutes ahead and 1 hour ahead, respectively.

OTHER STUDIES

“

Taylor & Francis. 2022. Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model. [online] Available at: <https://doi.org/10.1080/15472450.2015.1072050> [Accessed 28 April 2022].

- More accurate prediction of traffic based **on both location and time**

TEAM POWERPUFF_GIRLS

Mashiat Mustaq

Mushtari Sadia

J

Ramisa Alam

THANKS!

Any questions?