The Data Incubator

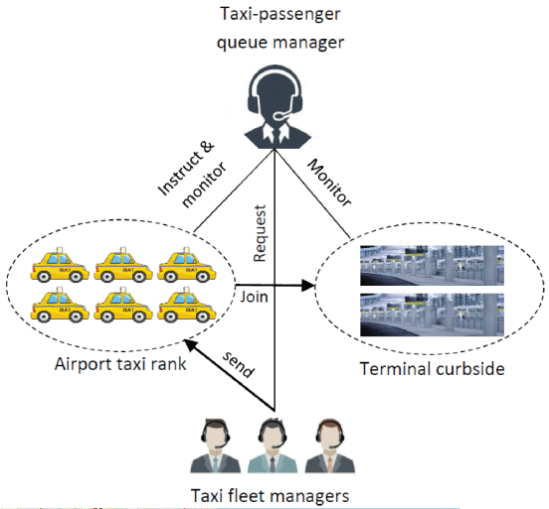
Project Proposal:
Wait Time Prediction for Airport Taxis at O'Hare International Airport (ORD)

Presented by:
Dooman Akbarian

July 2020

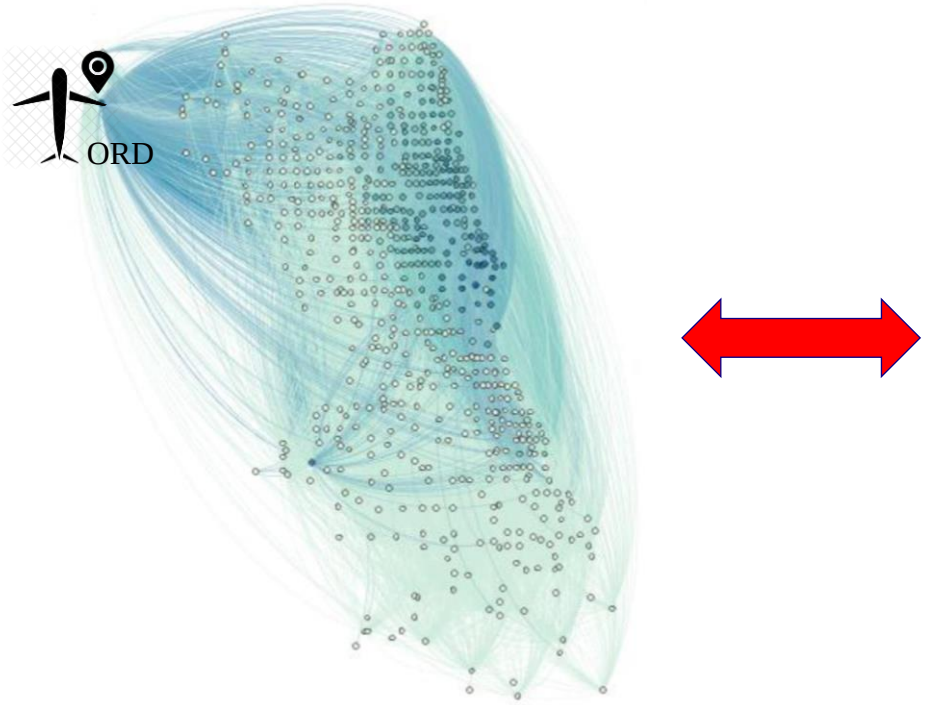
Problem Statement

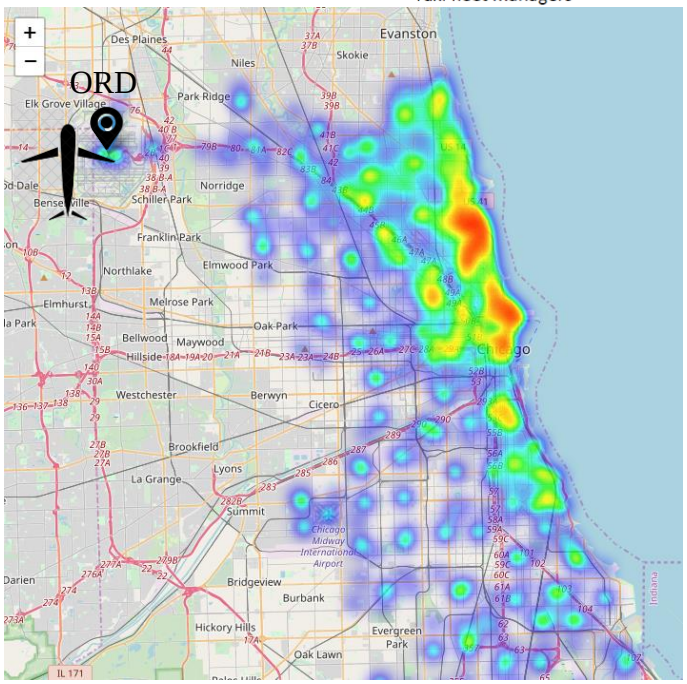
Manual airport taxi demand estimation system



- The human error in manual airport taxi demand estimation causes [1]:
  - Long queues of taxis and traffic congestion.
  - Long queue wait times for the passengers.

- O'Hare International Airport (ORD) is one of the busiest airports in the US, and it is one of the high taxi pickup locations in Chicago. **Therefore, it is essential to develop a ML model that can estimate the taxi queue wait time at ORD.**





1. Rahaman, M.S.,et al., 2018. IEEE, 6,74660-74672.

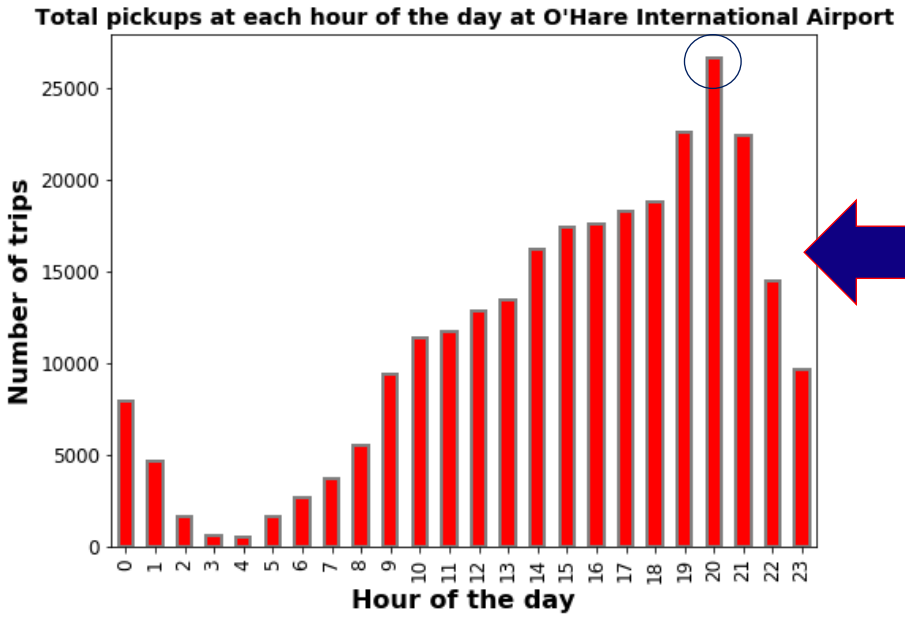
Preliminary Data Analysis

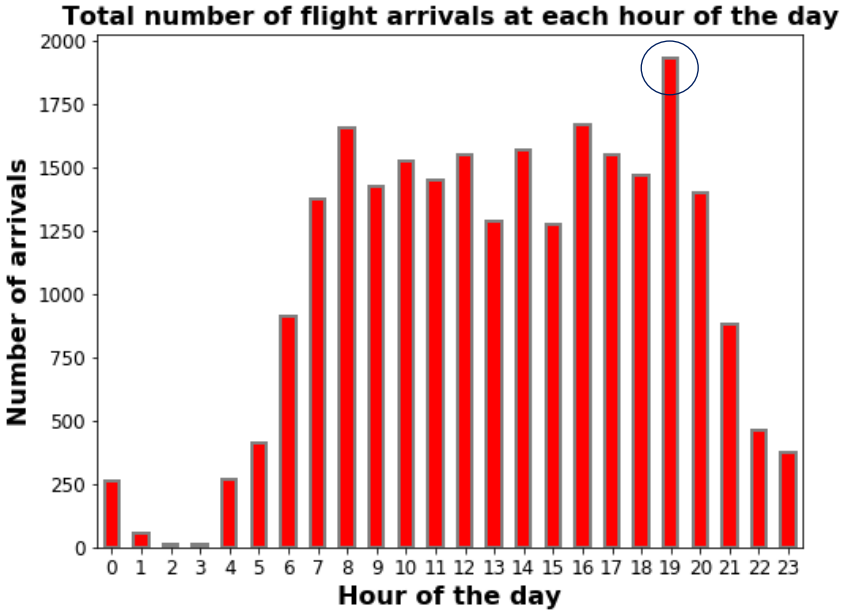
Chicago taxi trip data

- Pickup/ Drop off location
- Pickup/ Drop off time
- Distance of the trip
- Fare and tip amount
- Taxi ID

O'Hare airport flight data

- Number of flights
- Passenger arrivals
- Delay of flights
- Passengers wait time



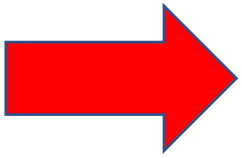


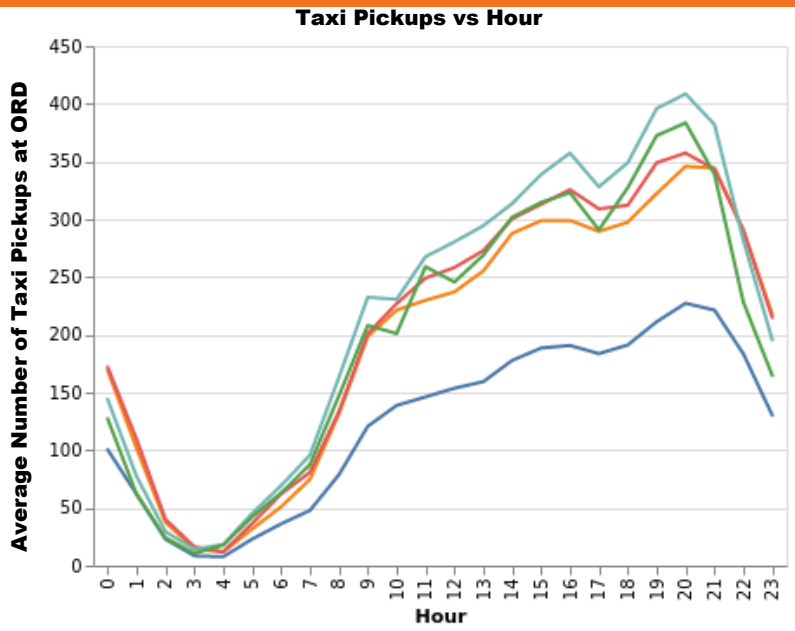
- The maximum number of taxi pickups occurs at 8 p.m. which is related to the maximum number of flight arrivals happens at 7 p.m.

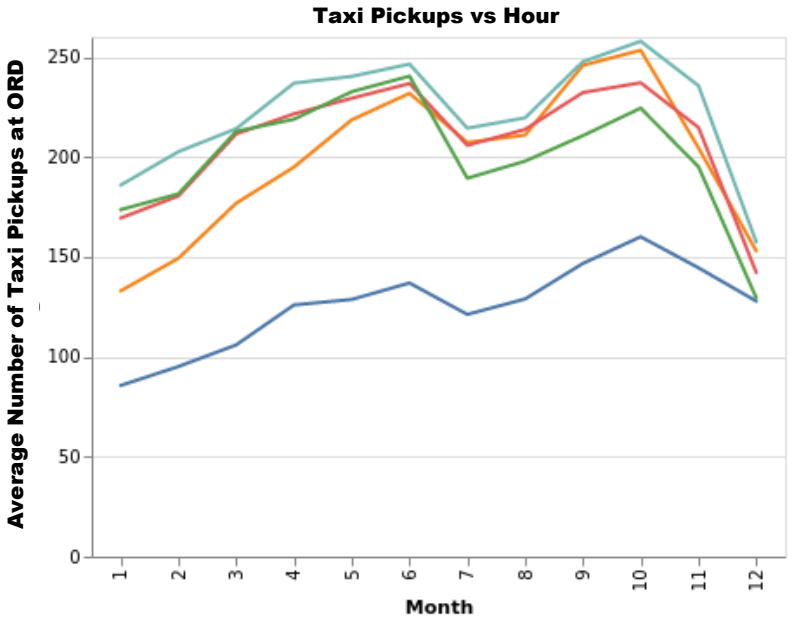
https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p
https://www.transtats.bts.gov/Tables.asp?DB_ID=120

Further Data Analysis after Invitation to Interview

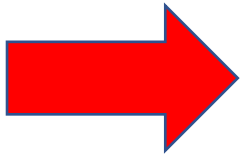
Chicago taxi trip data

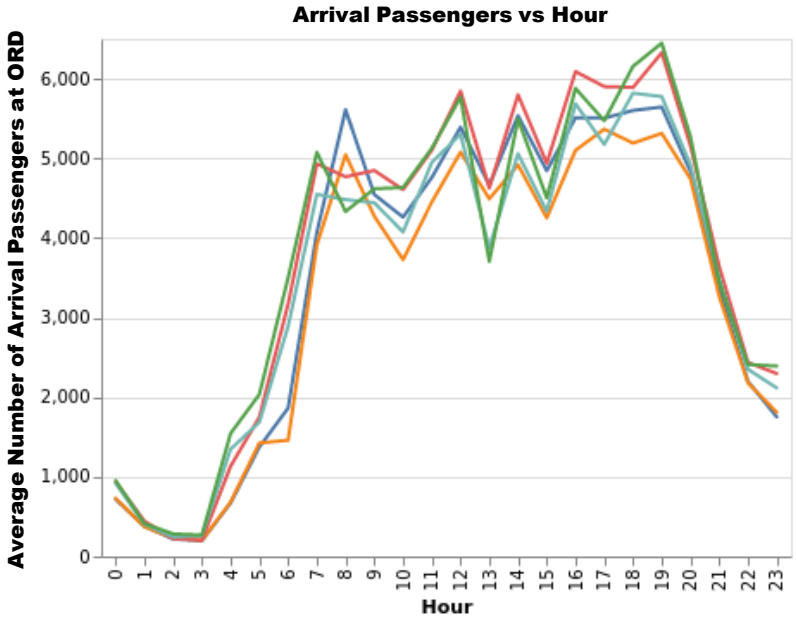


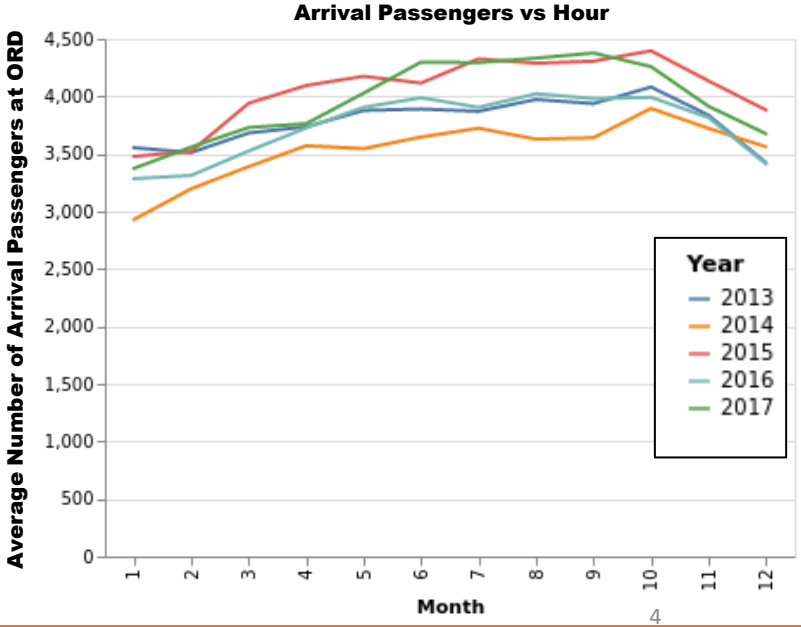




ORD flight data



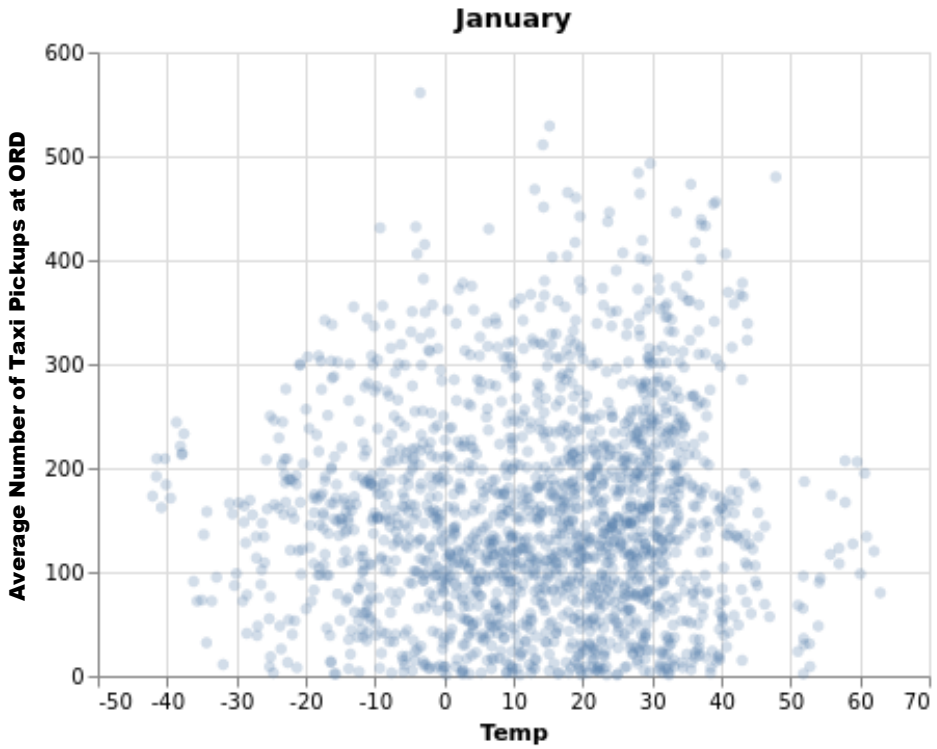


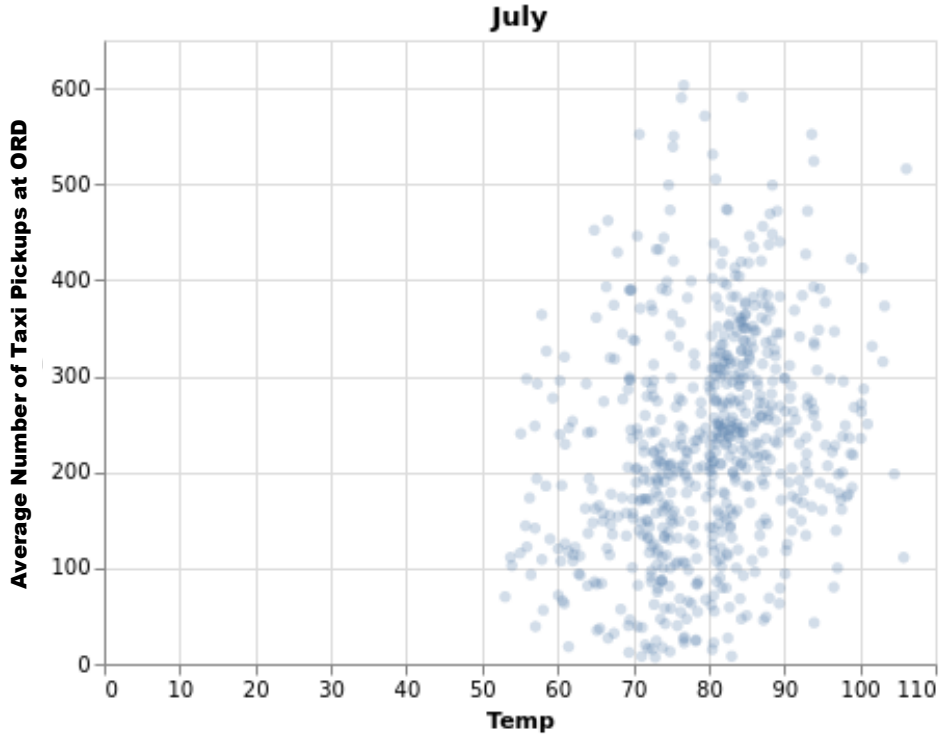


Further Data Analysis after Invitation to Interview

Weather data

Temperature	Precipitation	Clear Cloudy Fog Rain Snow
Wind Speed	Humidity	



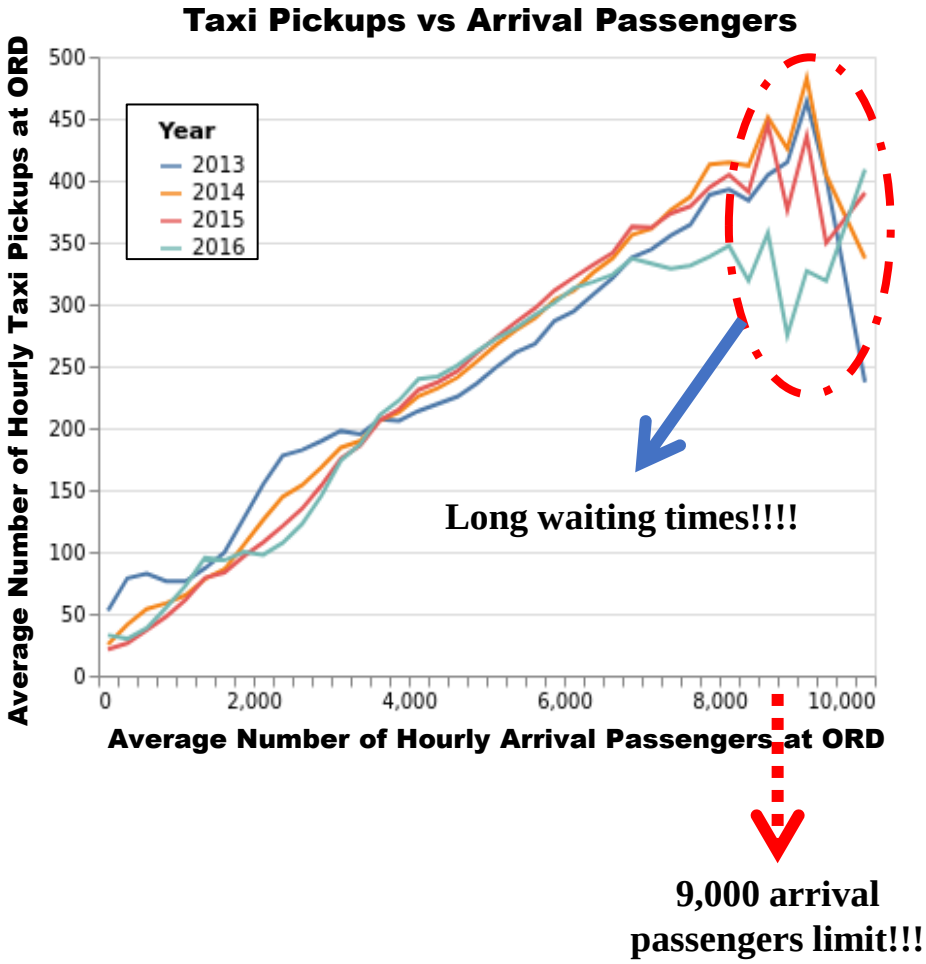


https://mesonet.agron.iastate.edu/ASOS/

Further Data Analysis after Invitation to Interview

➢Using random forest to predict hourly taxi rides (pickups) at O'Hare airport based on features such as: year, month, day of the month, daily hour, temperature, and number of arrival passengers.

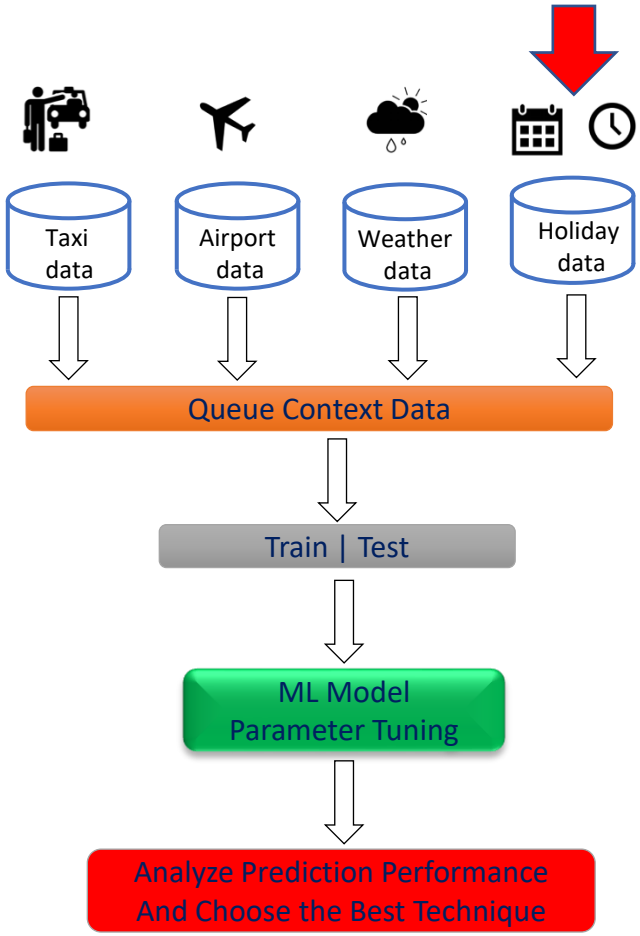
- Number of estimators:100
- Train-test split: 80-20%
- 5 Fold Cross-validation
- RMSE=40
- ***R^2 (Test score): =0.68***



Future Work

- The following ML models will be investigated:
  - Random Forest
  - Linear Regression

- Development of an interactive app for users :



	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Data wrangling/Analysis								
ML model development								
HTML platform								
Testing/Report								

Thank You For Your Attention