

1.A brief on the approach used to solve the problem.

Below are the list of Approach i have taken

1. Read the training hourly car demand data, testing car demand and determine date range in Training which is 2018-08-18 to 2021-02-28 and testing data range from 2021-03-01 to 2022-03-02
2. Extracted Feature from the date column like 'dayofweek', 'working day', 'month', 'year', 'dayofyear', 'weekofyear', 'Quarter', 'peak hour'
2. EDA on Extracted feature with respect to target variable
3. on EDA determine which are best suitable feature
- 4 Once feature extrction done w=move to model building
- 5.Training dataset splited into train and validation set (90% training,10% validation)
- 5 .Various Regression model trained on dataset like linear regresor , RandomForest, SVR,gradient boosting regressor, XGBregressor , on evaluation found XGBregressor is giving the best result with low MSER.
6. XGBRegressor hyper parameter tunning done with RandomisedCV to conclude final model .

2.Which Data-preprocessing / Feature Engineering ideas really worked? How did you discover them?

Feature engineering ideas which I have taken is extracting time series feature using date and hour column .

I have used pandas library to extract date feature from the pandas date time object following are the list of features which I have used

1. Year - rejected
2. Month -worked
3. Week day -worked
4. Working day as extracted feature from week day- worked
5. Day of year -worked
6. Week of year -worked
7. Day of month -worked
8. Season -rejected
- 9 Peak hour -rejected

Once I have extracted all the above mentioned feature I have done univariate analysis of these feature with demand

1. Year - plotted mean demand year wise and found there is slight increase in demand
2. Month wise demand analysis found there is comparatively peak in demand in month of November in each year
3. Demand in working day is less than non working day

4. Since we don't have list of holiday , hence used day of year , and week of year as one the feature to look public holidays like Christmas , new year, independence day , republic day etc.
5. Season - since we haven't have weather information which can highly spike demand of car or degrade demand according to the season , so to kind of season analysis , group month in summer and winter season , but on Eda found that there is no much impact .
6. Peak hour - from the bar graph of demand per hour , predicted peak hour as 9-20 during working day and 10-23 for non working day , but that doesn't work well while integrated the feature with model , hence drop the idea of using it .
 - What does your final model look like? How did you reach it?

3. What does your final model look like? How did you reach it?

Final model is based on XGBRegressor using hour,Month,Week day ,Working day,Day of year,Week of year

Fine tuning of model using Randomised cv and found n_estimators=1000, learning_rate=0.01 giving the best RMSE