APPROACH DOCUMENT TO DEMAND FORECASTING.

Problem relates to forecasting car rentals. From the dataset, its obvious that there are a lot of fluctuations in the demand and its complex to assess as there are multiple factors involved.

While I was exploring with models, I tried Time- series but the attempt to combine the date and hour column caused lot of data type mismatches, so I tried complex Machine learning models and Deep Neural Networks

It took the following steps to finalize the machine learning model to predict with least RMSE score.

1. **Additional Features:**
   I derived several features from the date column such as year, month, week, day, weekday, day of the year.
   Other Boolean features added are: Whether the day is a month end, month beginning, year end or year beginning, quarter end, quarter beginning, or a holiday.
   The parts of the day were segregated to create a new categorical variable, 'Day_part'. The categories under it are: early_morning, morning, late_morning, noon, afternoon, evening, night, midnight.

2. **Exploration of Target variable:**
   The distribution is a bit close to normal and right-skewed. There are several outliers. The average demand on yearly basis has been increasing very steeply. However, in the early years there has been peak demands such as 300 or 350, which is abnormally high.

3. **Handling outliers:**
   The average demand increased on a yearly basis. The peak demands in the past years completely disregarded this average trend so I decided to delete them from the train set.
   The box plot shows more number of outliers, however I deleted rows with demand over 249 to avoid overfitting.

4. **Preprocessing:**
   A preprocessing pipeline has been created, that includes Standard Scaling and creation of polynomial features of degree 2 for the numerical columns, and One Hot encoding for the categorial and Boolean columns.

5. **Training and testing:**
   Some complex models were explored such as Random Forest Regressor(RMSE 33.48), Gradient Boosting Regressor(RMSE 33.9) and Light GBM (33.12). I also created a deep neural network model which gave an RMSE of 36.28. Light GBM has been picked on performance basis. I have made 2 submissions, one with LightGBM model and the other with Random Forest Regressor model.