```
'''
Evaluation metric:
Since our target variable is continous (trip_duration in seconds) we
have used MSE,MAE,R2 for KNN;
MAE for Benchmark; MSE for Linear Regression.

Considering analysis from EDA provided:
1>Some outiliers are removed. trip_duration>6000 are
removed(Considering Box plot for passenger count in EDA analysis)
2>'Short distance' is calculated using features- pickup and dropoff-
(lattitude and longitude) repectively.
3>There are more than 7 lacs rows,hence stratified sampling is done to
reduce the data to 40% to improve execution speed.
  Stratified sampling helps to reduce data without any change in model
performance.


Observations and Conclusions:
1>KNN:  MSE:
        After K>150, the mse error seems to follow a straight pattern.

        R2:
        For K=81 and K=101, the test R2 error value is at optimum and
going down for K>120. As R2 must be high as
        possible and close to 1 we can take K around 80 and 100
        So 80<K100 can be taken.

        MAE:
        Both test and train error are around 230.

2>Benchmark vs KNN:
        KNN's(K=80): MAE: Around 230.
        Bemchmark MAE: Around 350.

        Here the MAE of KNN model seems low and hence KNN model can be
considered a better model.

3>Linear vs KNN
        KNN:MSE:      Test->119682,  Train->115345
        LinearReg:MSE:Test->165526,  Train->164932

        Here the MSE of KNN model seems low and hence KNN model can be
considered a better model.

After evaluating all the 3 models we can consider KNN model for our
problem.
'''
```