2 Main project codes in .ipynb format

1. Code for NFL - nfl.ipynb
2. Code for Bike Prices - bikes.ipynb

nfl.ipynb :-

Steps :

1. Reading and cleaning the data and filling up null values.
2. Finding Correlation matrix and dropping highly correlated columns.
3. Splitting global data into Train and Test set.
4. Training and Applying a Global XGBoost model on the dataset.
5. Finding feature importance of XGBoost model.
6. Clustering the dataset using K-Medoid algorithm.
7. Total clusters = 4 (Checked with R2 score for best cluster size)
8. Dividing cluster 1 into train and test data.
9. Applying XGBoost model to find important features and also checking the prediction curve.
10. Applying linear model on cluster to get the coefficient values and comparing with XGBoost model predictions.
11. Finding Graph for important features using our approach.
12. Using LIME to get the important features output.
13. At the end we have included the code to compare R2 score for each cluster individually and also chaining the k values in cluster and adding few similar code blocks with provide R2 score for K = 3,5,6,7. Since K = 4 is the best among other we showed only K = 4 in code.
14. We have used the clusters to find the global feature importance as well in next few blocks.

bikes.ipynb :-

Steps : All the steps are similar to nfl.ipynb, only difference is that the optimal cluster value for this data set was observed as k =3, so we have divided 3 clusters and did all the calculation based on k value as 3.