Predicting USA Car Commute Time

Instructor: Prof. Daniel Acuna

Team Members: Adarsh Patil, Rahul Rathod, Rashika Singh, Sachin Chaudhury

Problem and Objectives

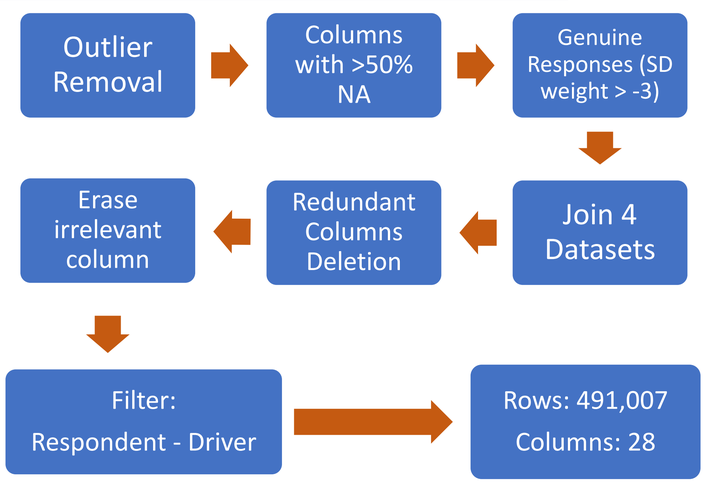
- We aim to predict a car's commute time by an unprecedented approach of analyzing the driver's features.
- There has been no significant research in predicting commute time other than traffic and road condition
- In our data we included driver's demographic, economic, and cultural changes behavior to predict commute time

Data Description

The 2017 Federal Highway Administration's sponsored National Household Travel Survey (NHTS) is the source of the nation's information about travel by US residents in all 50 States and the District of Columbia. The data consists of 4 datasets:

Name	Rows (100k)	Columns	Description
Household	1.3	58	Income, active drivers
Person	2.64	121	Age, sex, race
Vehicle	2.56	54	Age, miles
Trip	9.32	105	Division, time

Data Preprocessing and Cleaning

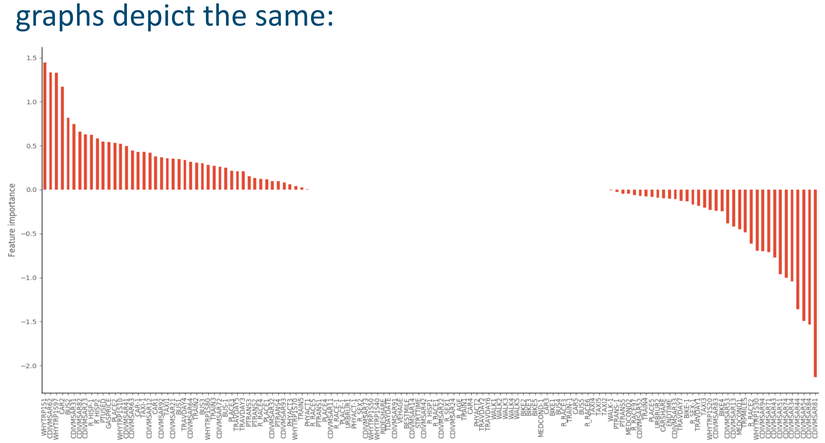


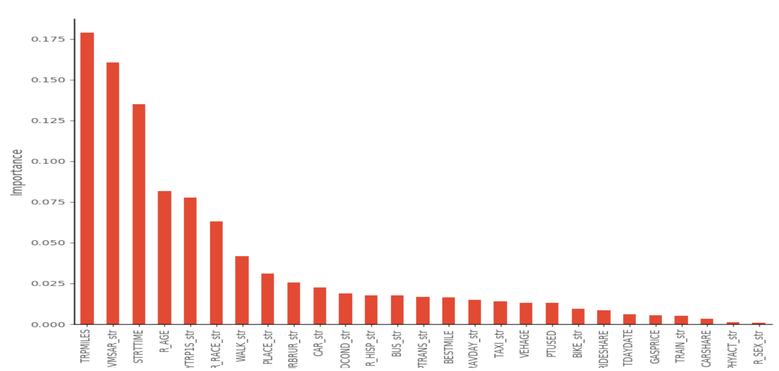
Model Description

MODEL	DESCRIPTION	EVALUATION
Linear Regression	Used for feature importance and prediction	RMSE
Elastic Net Regression	Used for feature importance	Feature importance
Random Forest	Used for feature importance and prediction	RMSE
Gradient Boosting	Used for prediction	RMSE

Feature Importance

In order to select the important features for prediction we used Elastic Net Regression and Random Forest. The following graphs depict the same:





Model Comparison (RMSE)

We choose the 7 important features from Elastic Net Regression and Linear Regression by testing and trying different parameters. Once the parameters were selected, we applied Random Forest and Gradient Boosting to calculate RMSE. In the process to analyze and reduce the RMSE we used combinations for Max Depth and Number of Trees. The following is the result of our analysis:

Random Forest

# of Trees	Max Depth = 7	Max Depth = 8	Max Depth = 9
300	5.97729	5.8353	5.8103
400	5.8813	5.8379	5.8114
500	5.8794	5.8384	5.8121
600	5.8793	5.8375	5.8137

Gradient Boosting

# of Trees	Max Depth = 1	Max Depth = 2	Max Depth = 5
50	6.1522	6.0873	5.9971
100	5.9742	5.8079	5.7765
200	5.8734	5.8036	NE

NE – Not Executable

Conclusion

The top features to predict the commute time were the trip miles, start time, travel day, purpose, state divisions, urban/rural area, age, etc. The algorithm which best predicted the commute time was Gradient Boosting with the least RMSE of 5.737 In this project, we analyzed that despite traffic and geolocation there are different attributes where we could predict commute time.

Data: https://nhts.ornl.gov/