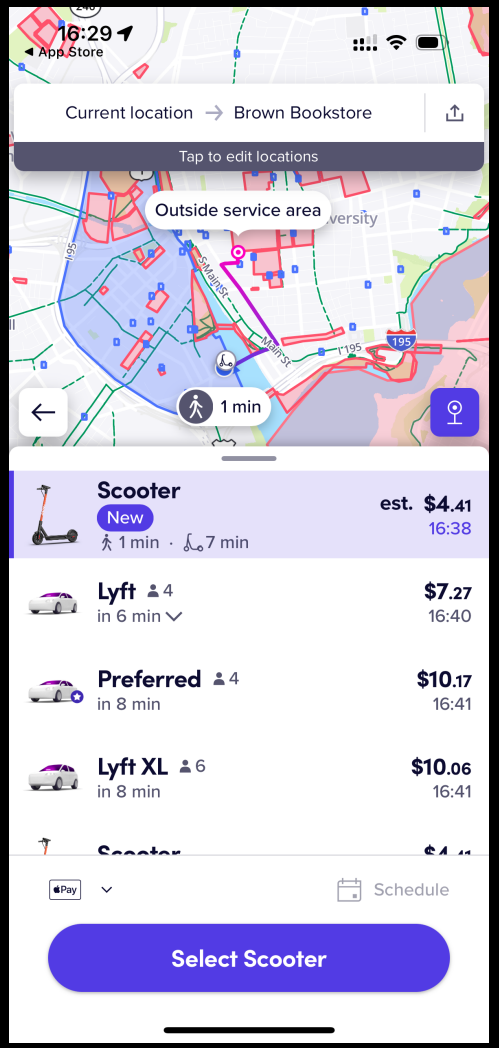
Can we discover the secret behind dynamic cab ride pricing?

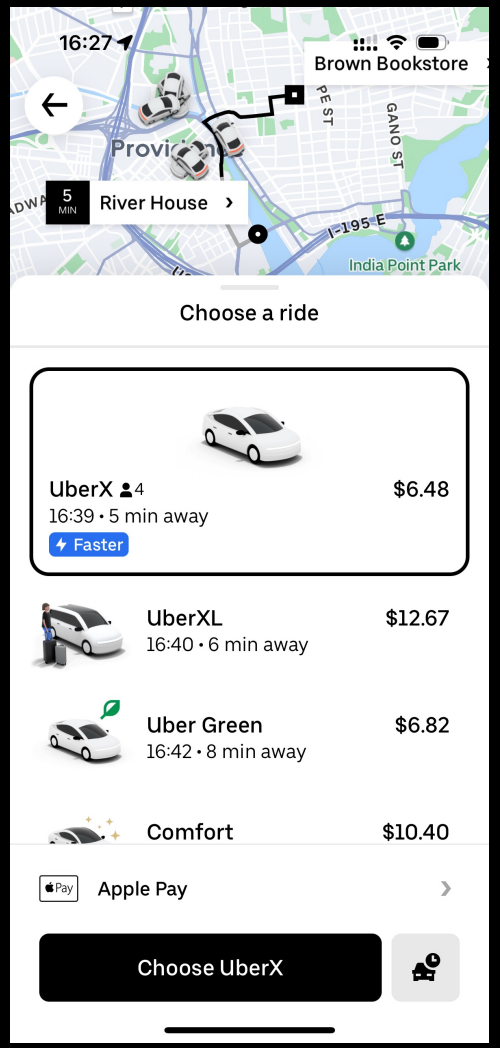
https://stablediffusionweb.com/#demo

FINAL PROJECT PRESENTATION
FOR DATA 1030

Yijia Xue
12/06/2023
Brown University
Github link: https://github.com/Ikea-179/Uber_Lyft_price_prediction#uber_lyft_price_prediction

Recap



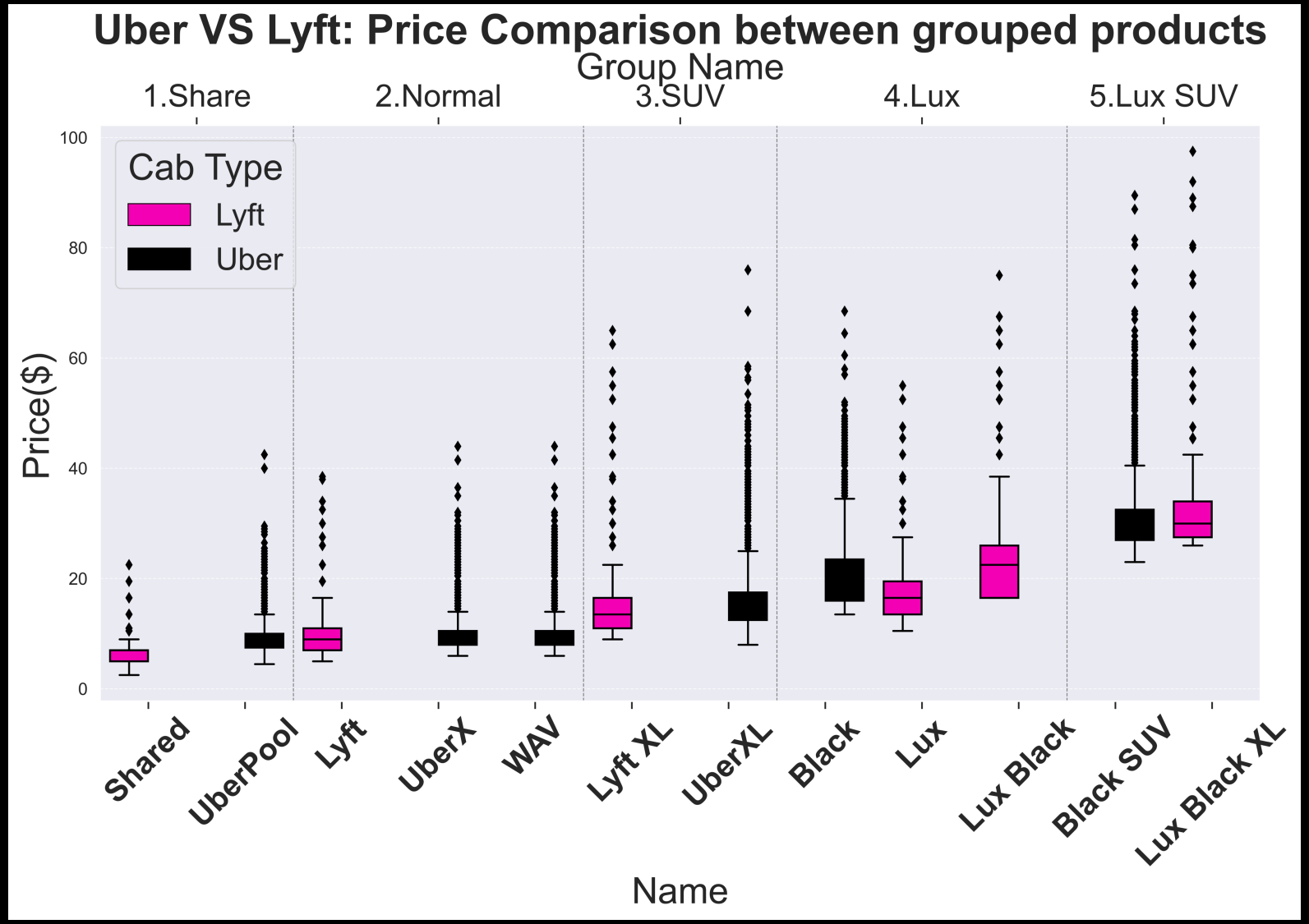


How exactly are these prices calculated?
Is there a pattern to their dynamic pricing rules?

Predict and compare the price of Uber and Lyft rideshares -> Regression Problem

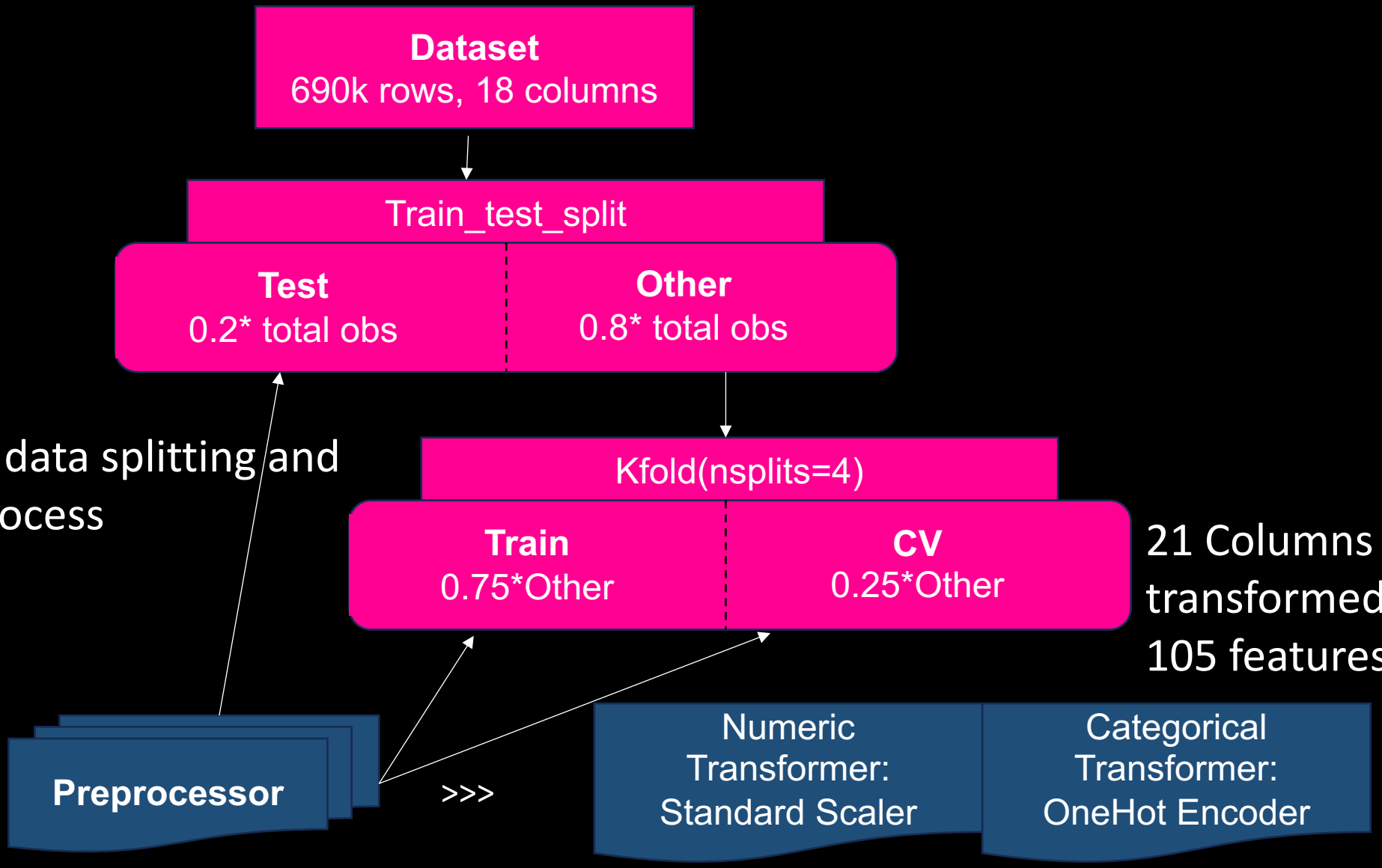
The data is from Kaggle:
[Uber and Lyft Dataset Boston, MA]

Recap



Lyft is
Cheaper than Uber for shared, and SUV cars
Similar price in Normal cars
Giving more options in Luxury cars
More expensive in Luxury SUV car.

Cross Validation



Flowchart of the data splitting and preprocessing process

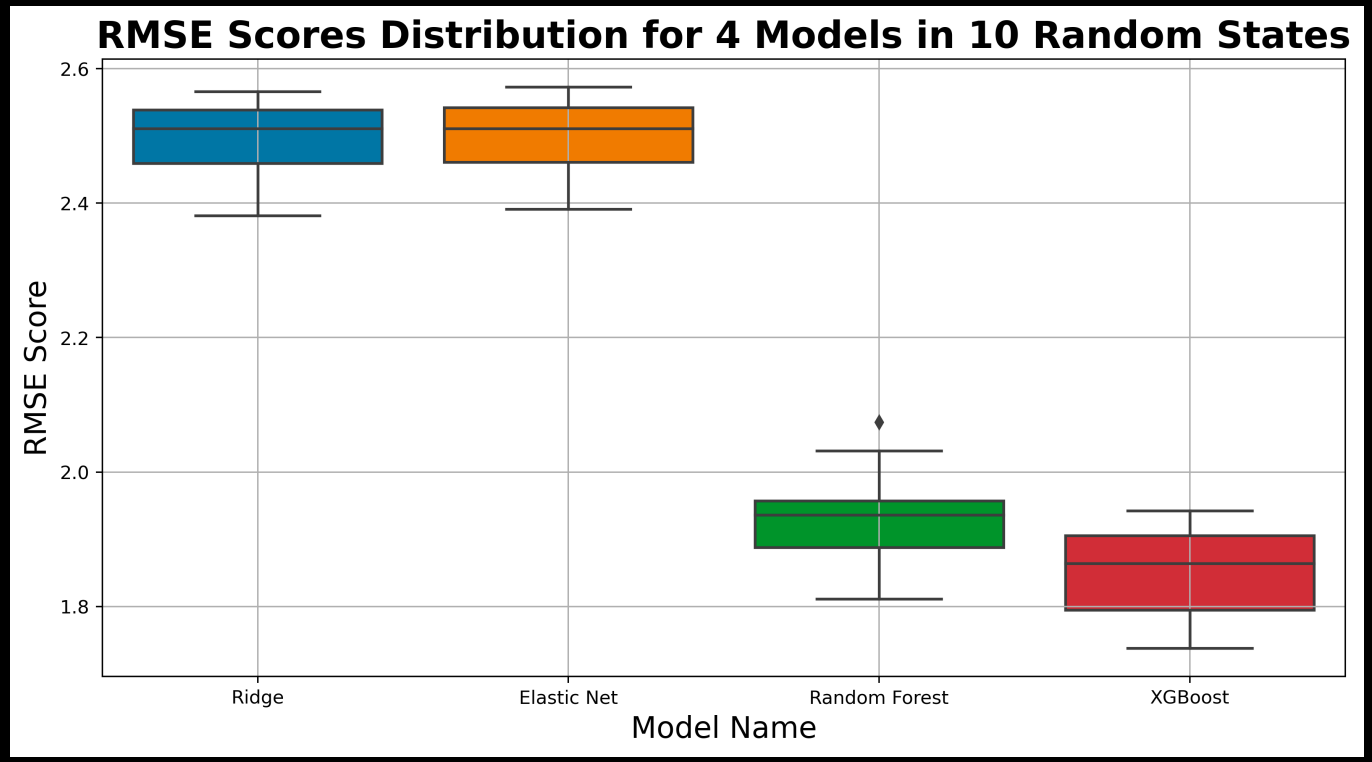
21 Columns transformed into 105 features

Cross Validation

Algorithms	Hyperparameter tuning grids
Ridge Regression	**alpha**: [1e-2, 3e-2, 6e-2, 1.6e-,1 4e-1, 1e0, 2.51e+0,6.31e+0, 1.585e+1, 3.981e+1, 1e+2]
Elastic Net	**alpha**: [1e-2, 3e-2, 6e-2, 1.6e-,1 4e-1, 1e0, 2.51e+0,6.31e+0, 1.585e+1, 3.981e+1, 1e+2] **l1_ratio**: [0.01,0.11,0.21,0.31,0.41,0.5,0.6,0.7, 0.8,0.9, 1]
Random Forest Regression	**n_estimators**: [100,200] **max_depth:** [None,1, 3, 10, 30, 100] **max_features**: [0.25, 0.5, 0.75, 1.0]
XGBoost Regression	**n_estimators**: [100, 200, 300] **max_depth**: [3, 6, 10] **learning_rate**: [0.01, 0.1, 0.2] **colsample_bytree**: [0.3, 0.7]

Hyperparameter tuning grids

Results: Evaluation metrices



Evaluation Result comparison between 4 implemented models

Baseline Score:

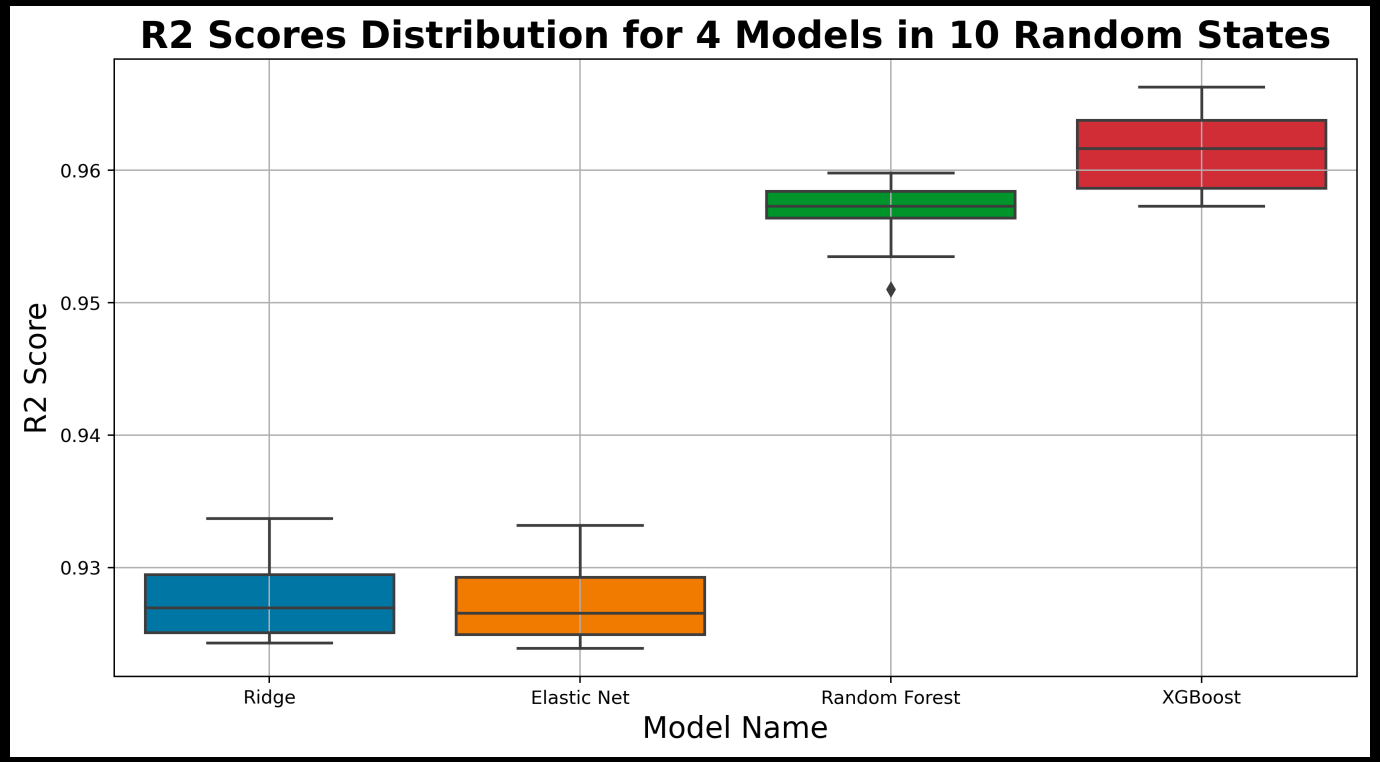
RMSE 9.32

R2 -0.007

Above are measured by letting prediction equal to the mean of y_true for all points in the test set and calculating the metric.

To address uncertainties from data splitting and model variability, training and testing were conducted across 10 random states and getting the same results for each state.

Results: Evaluation metrices



Evaluation Result comparison between 4 implemented models

Baseline Score:

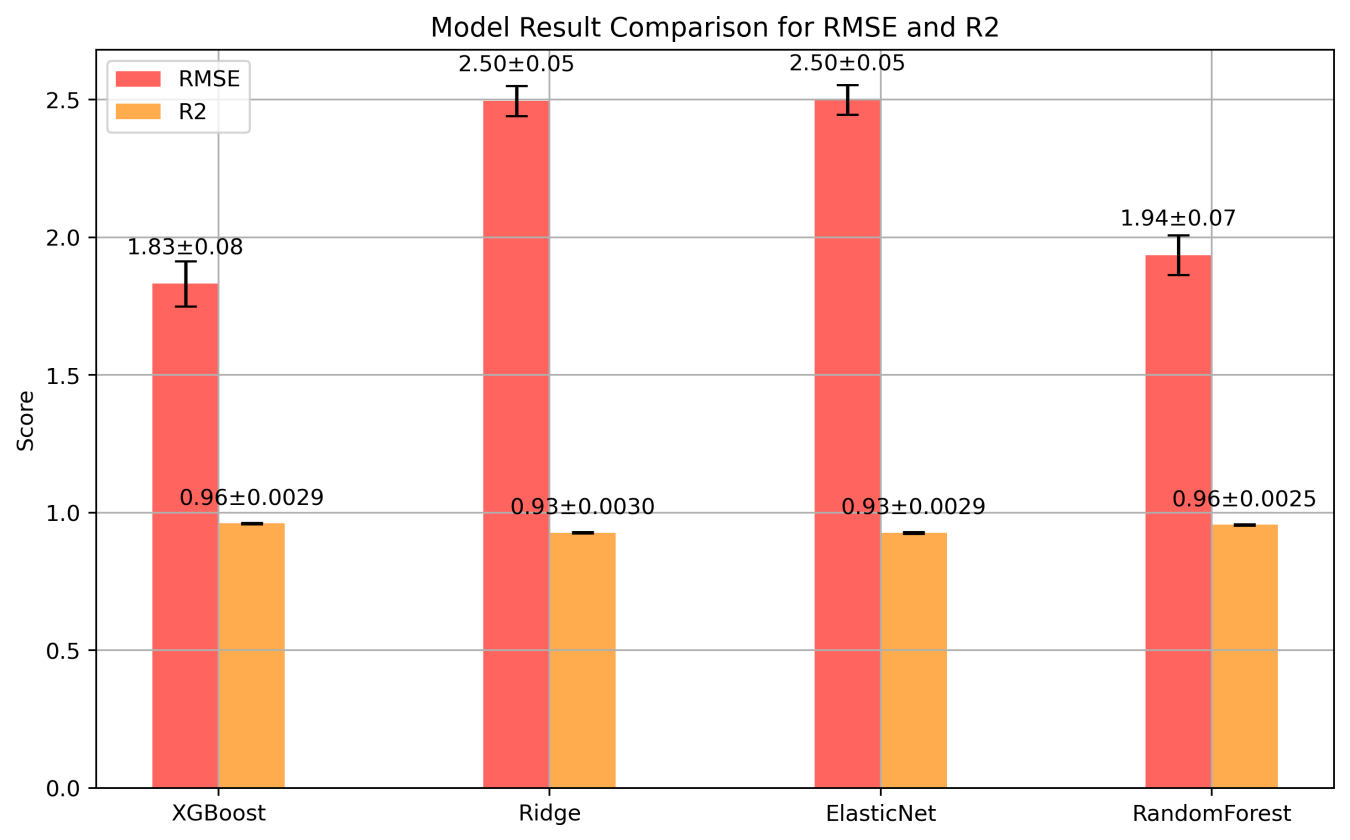
RMSE 9.32

R2 -0.007

Above are measured by letting prediction equal to the mean of y_true for all points in the test set and calculating the metric.

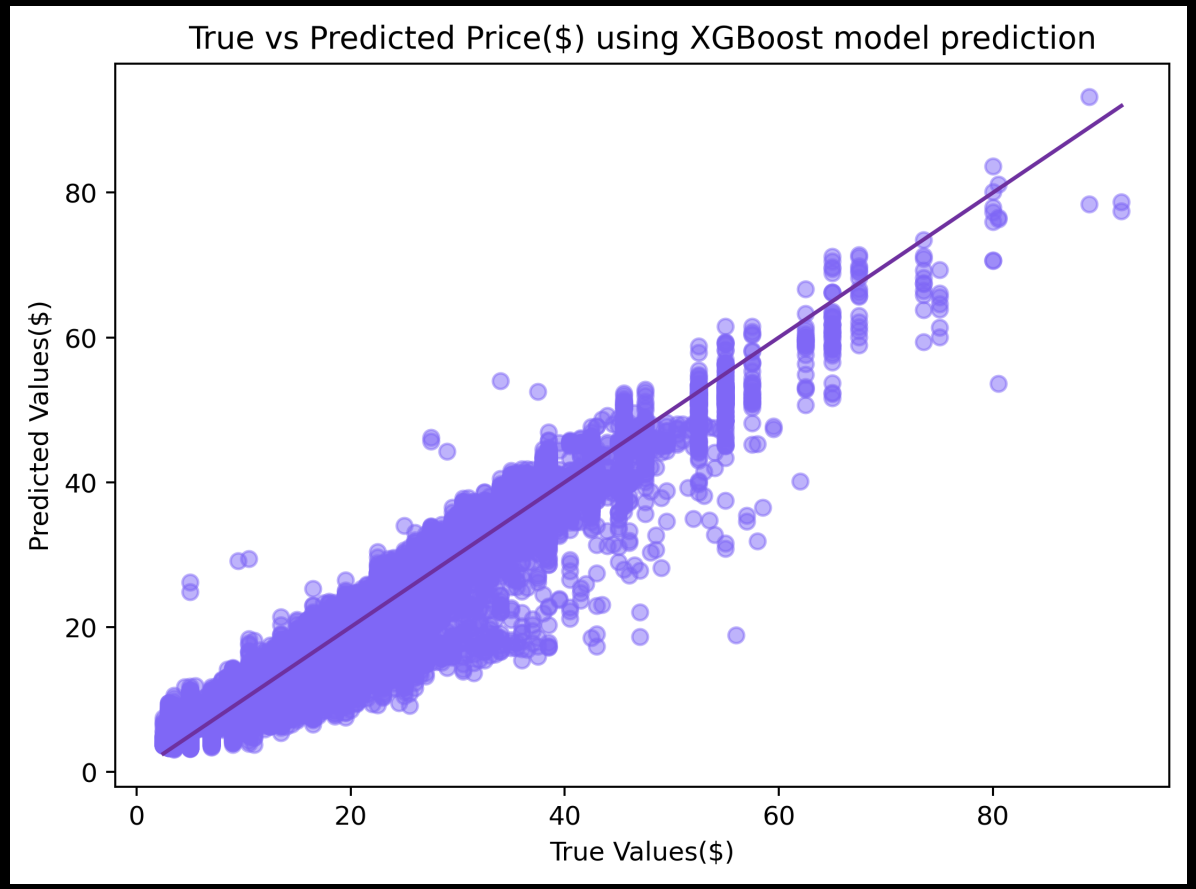
To address uncertainties from data splitting and model variability, training and testing were conducted across 10 random states and getting the same results for each state.

Results



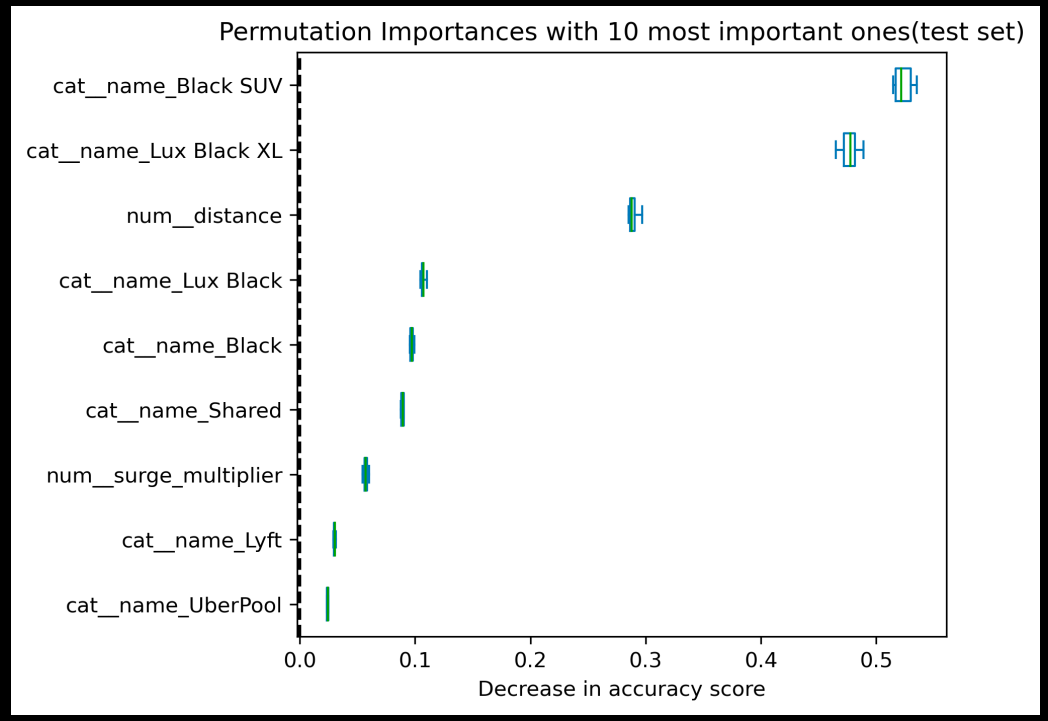
XGBoost appears to be the best-performing one since it has both the highest R2 score and lowest RMSE.

Results

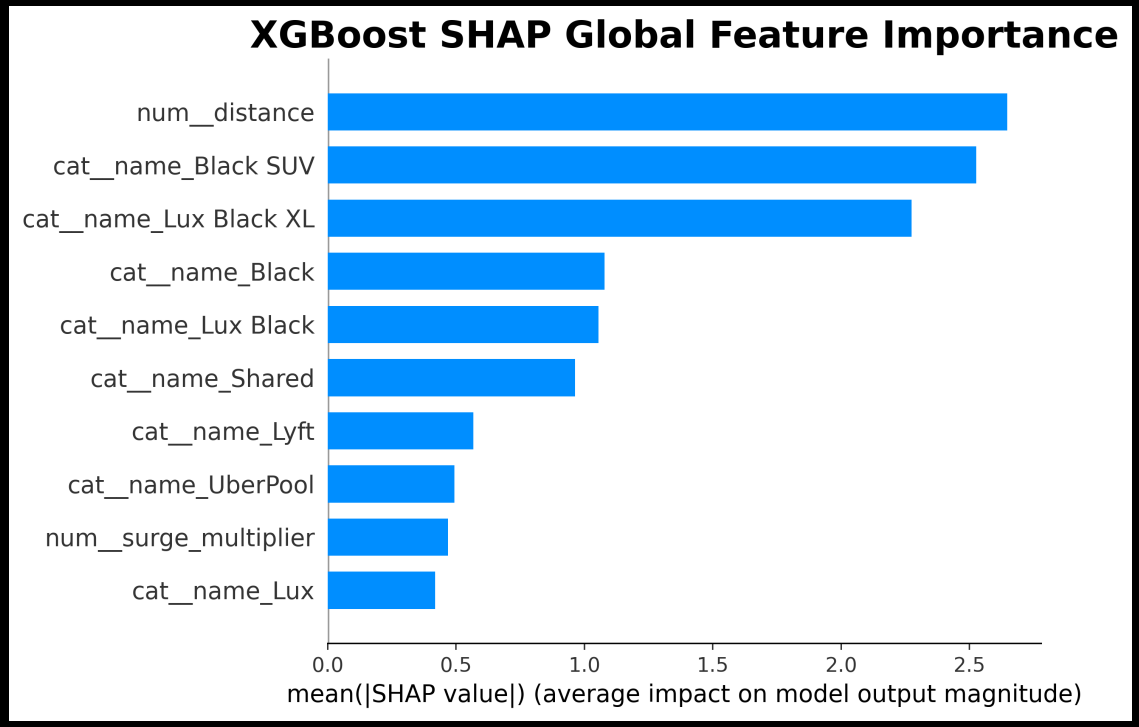


XGBoost appears to be the best-performing one since it has both the highest R2 score and lowest RMSE.

Results: : Global Importance

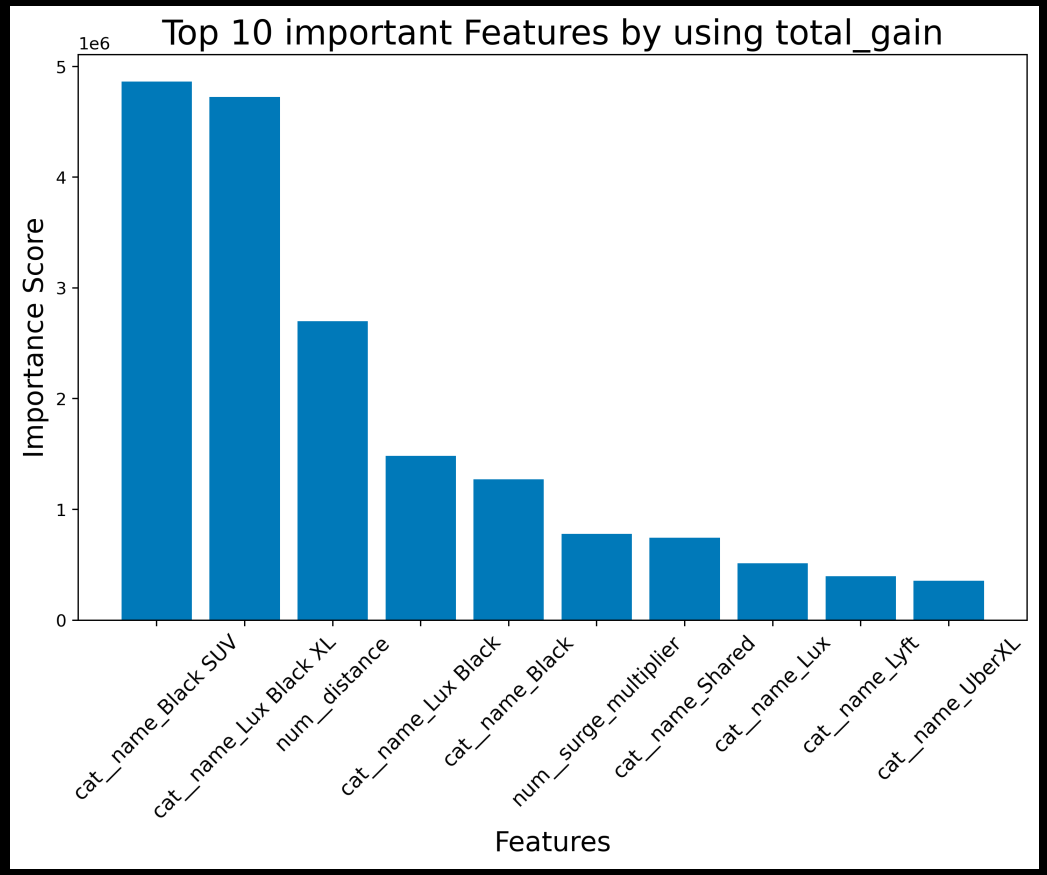


Top 10 most important features as per permutation feature importance.

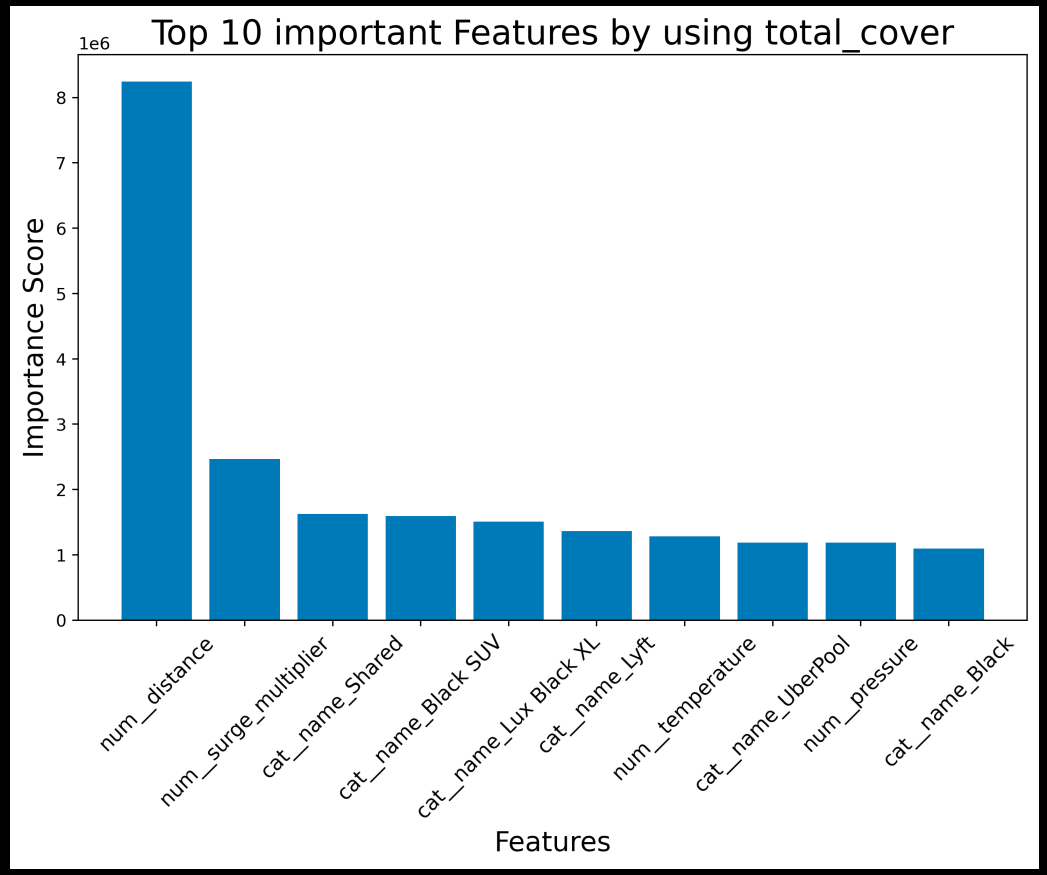


XGBoost SHAP global feature importance

Results : Global Importance



Top 10 most important features as per XGBoost relative importance measure-total_gain

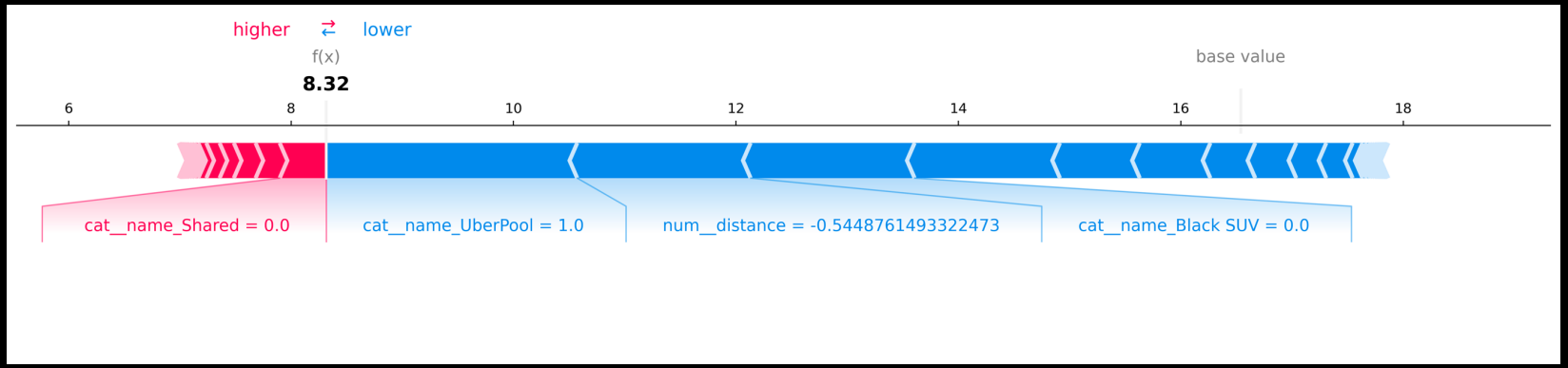


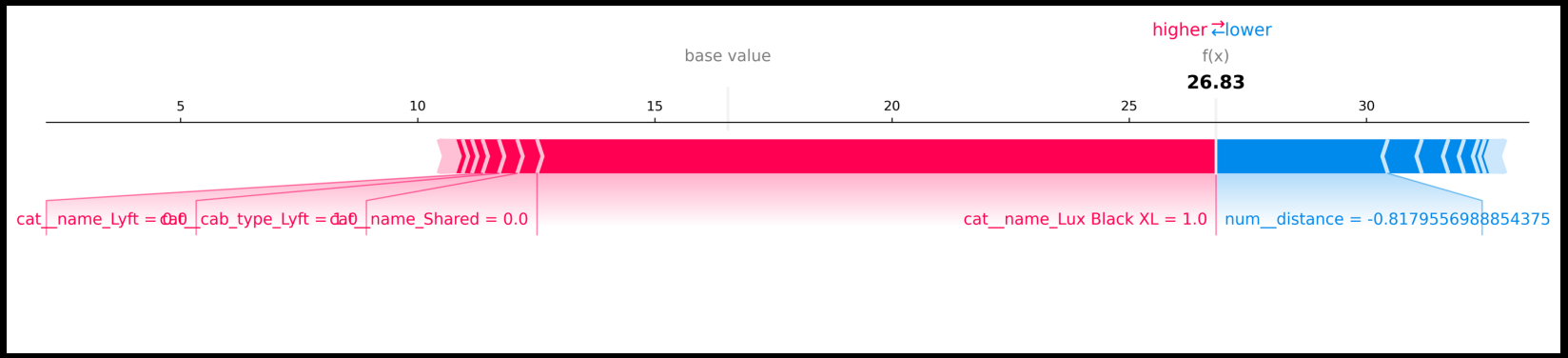
Top 10 most important features as per XGBoost relative importance measure-total_cover.

Results: Global Importance

- num_distance is a critical feature in terms of frequency and data coverage in the model
- The categories related to the type of service (Black SUV, Lux Black XL) are highly impactful in terms of model performance improvement
- Surge_multiplier also plays an important role since it is related to the real-time supply-demand situation
- In consistency with real-world common sense

Results: Shap Local Importance





Shap local value for 2 instances: cat_uberpool plays more importance than the black suv category when the price is low.

Outlook

- One of the primary limitations is the static nature of the dataset, which does not account for real-time variables such as current traffic conditions and driver availability.
- Another promising direction for future research is the deployment of these models in a real-time predictive analytics framework.
- Testing further models such as K-Nearest Neighbors (KNN) regression and models with reduced features

Thank you

https://stablediffusionweb.com/#demo