

SHOWDOWN



**UBER VS. LYFT: WHO WINS THE RIDE
A MACHINE LEARNING SHOWDOWN
GAME?**

When it's pouring and surge pricing is soaring... WHICH APP DO YOU TRUST?

Which factor do you think affects ride-sharing prices the most—distance, time, weather, or app choice?

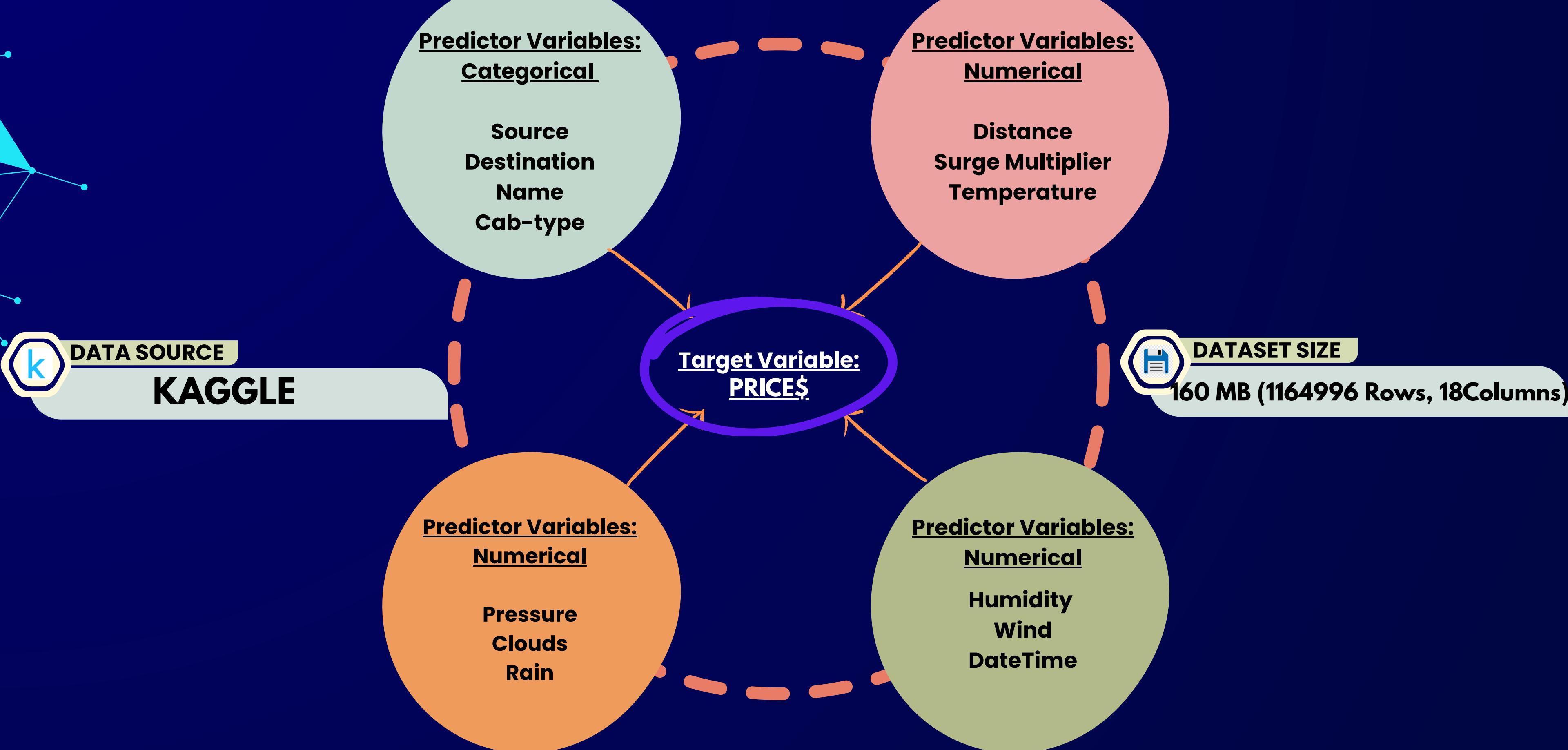
OBJECTIVE

Build a smart model that reveals which ride—Uber or Lyft—gives you the best deal based on weather, location, and rush hour timing.

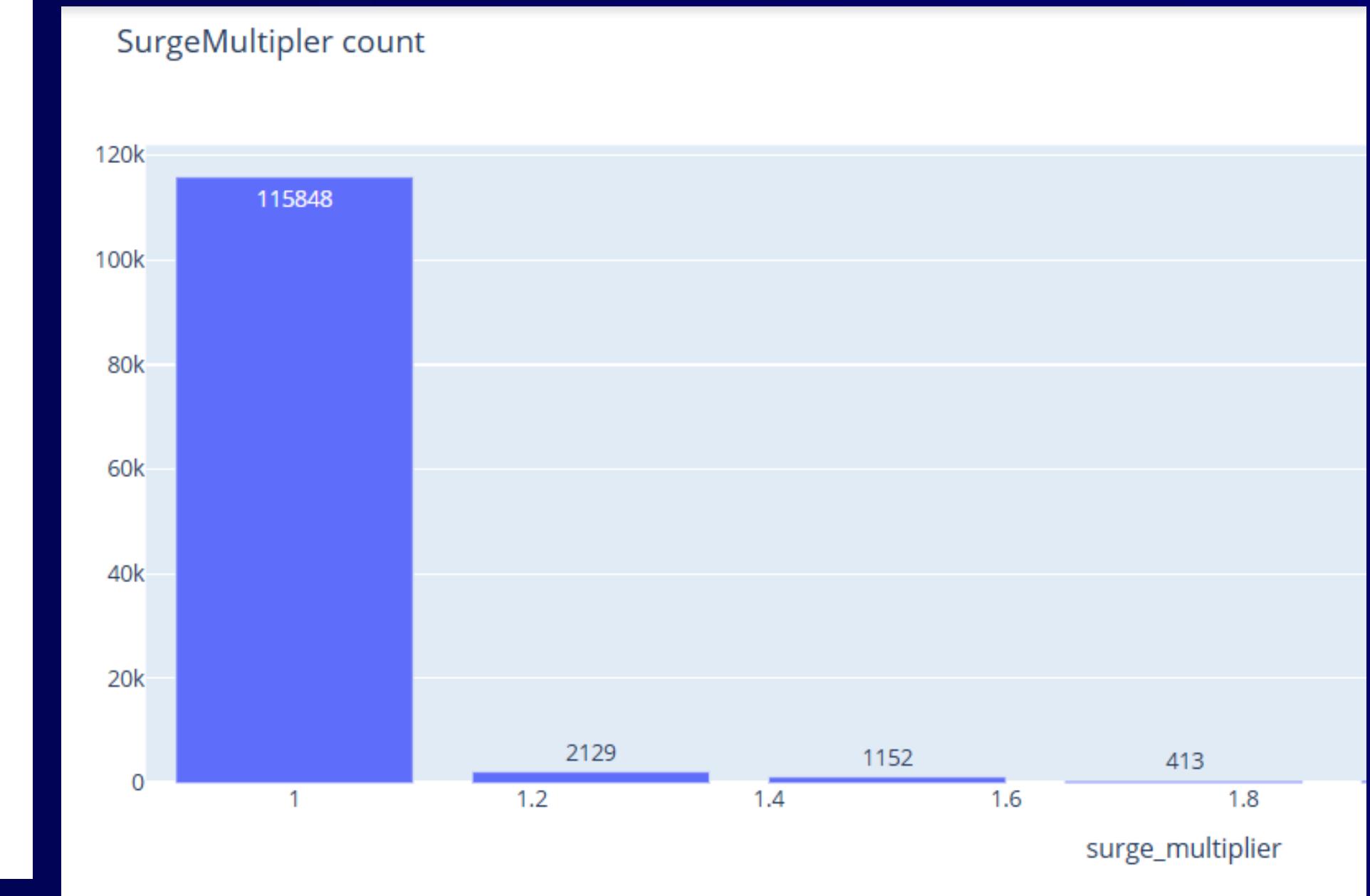
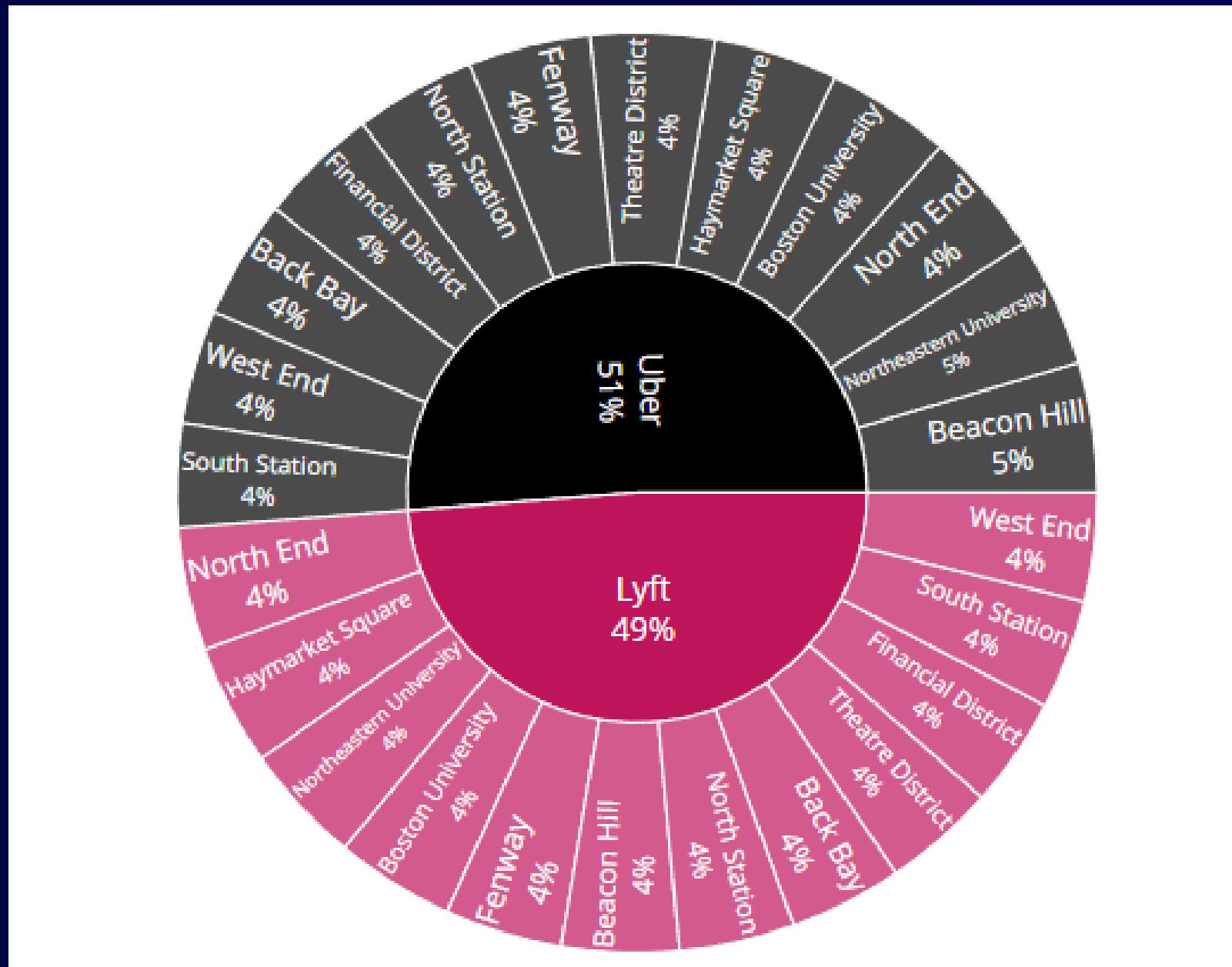
Uncover how distance, location, and time impact ride fares—spotting patterns that drive price changes.

Test multiple ML models to find the most accurate one—then fine-tune it to perfection with hyperparameter tuning.

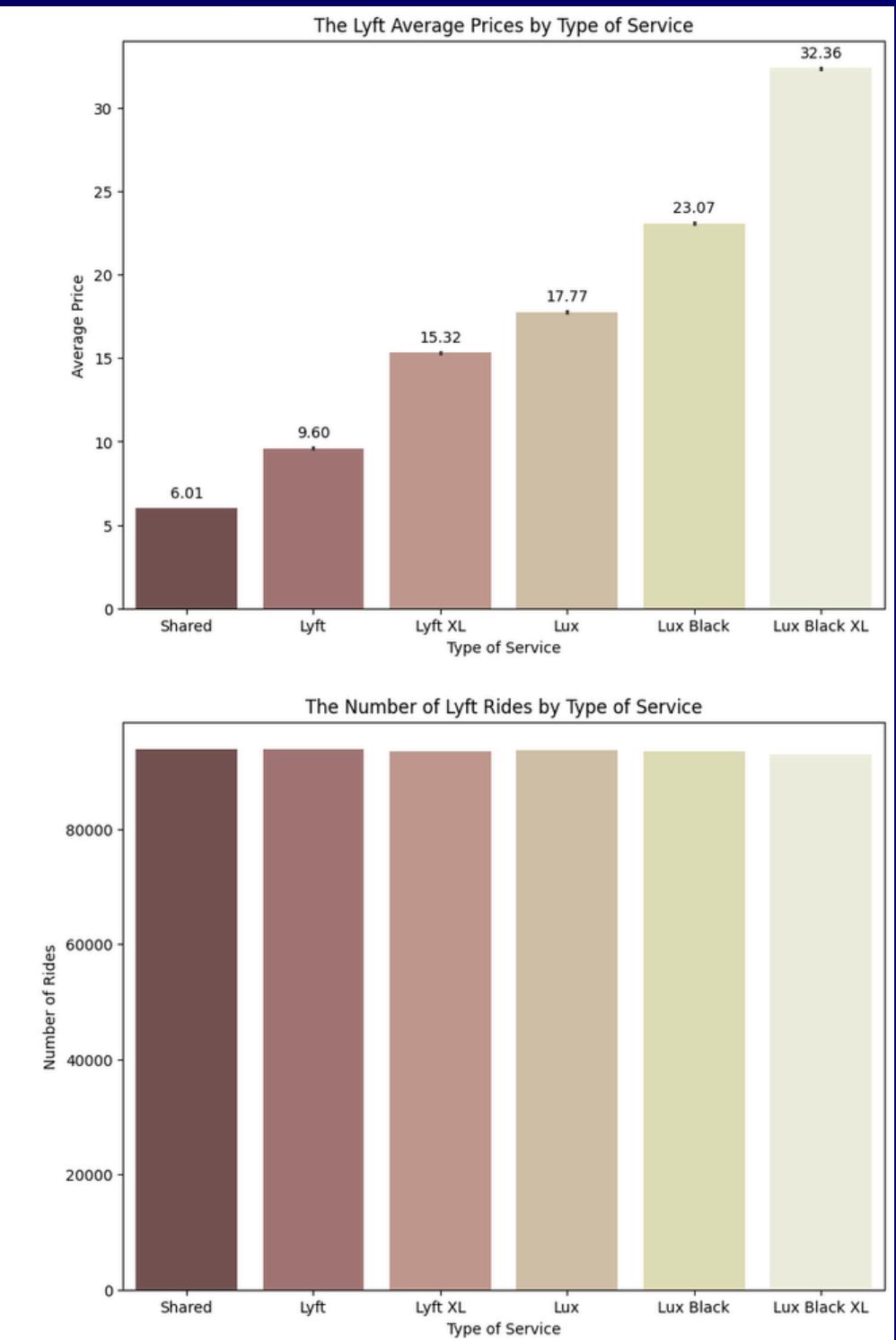
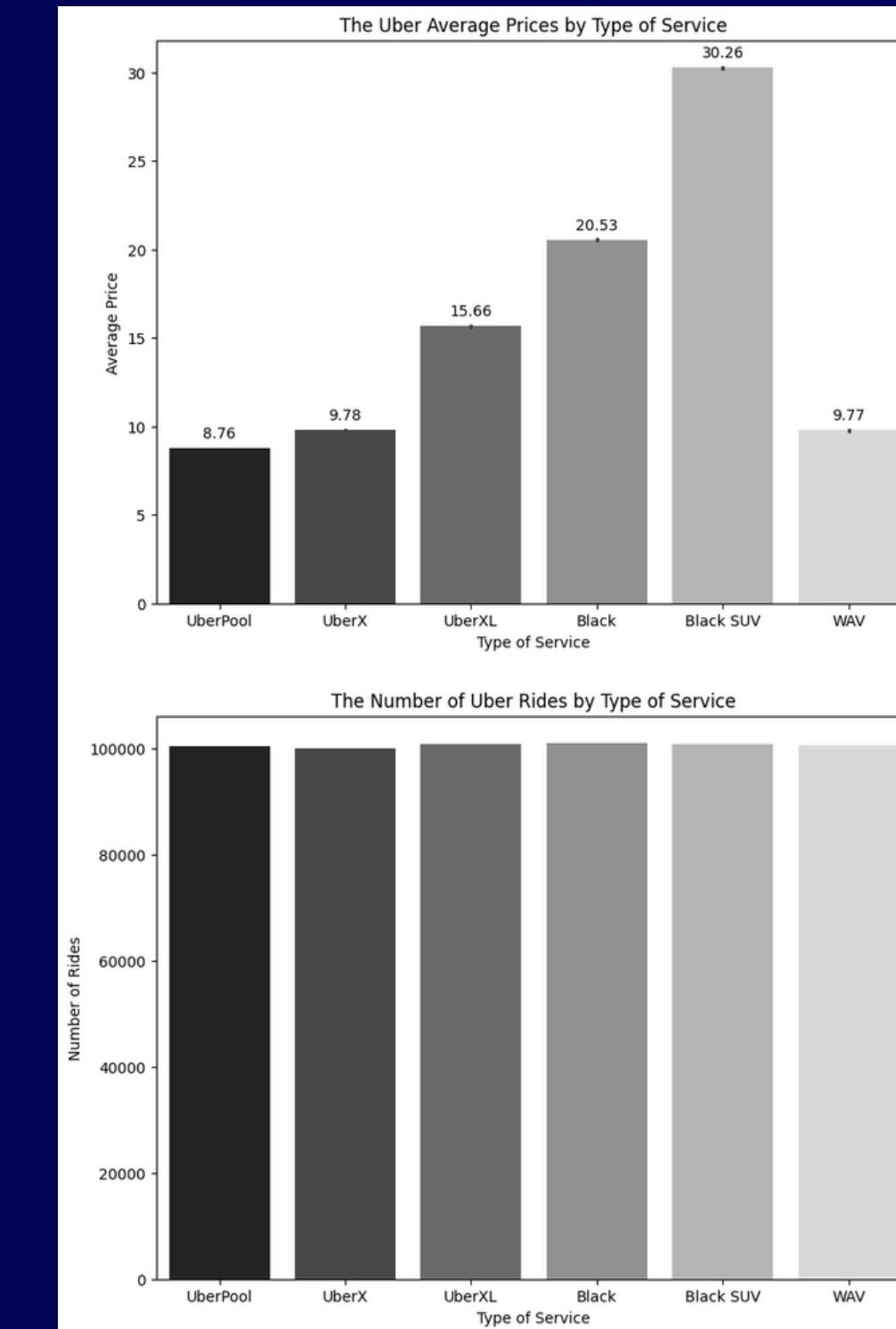
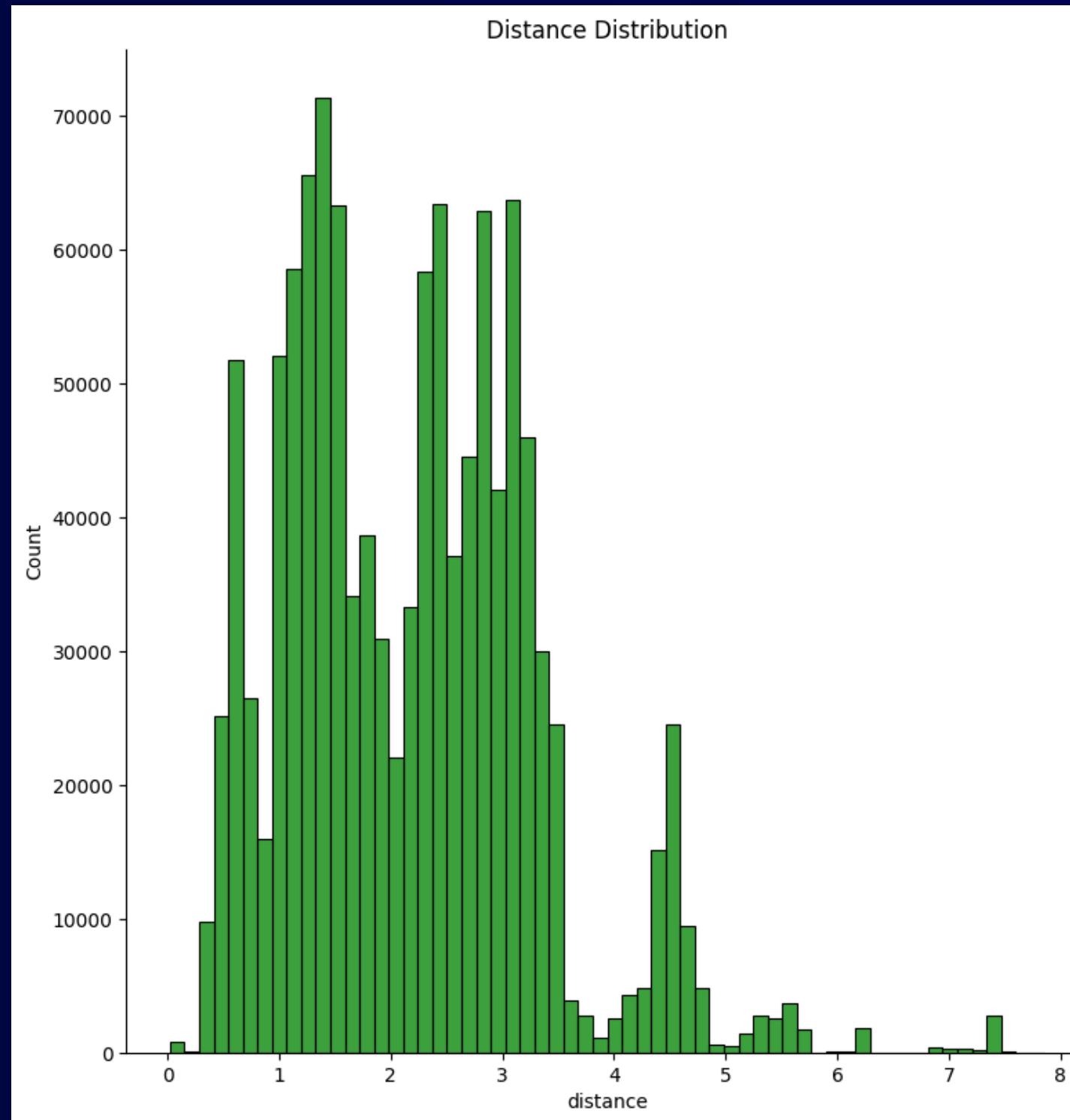
Data Summary



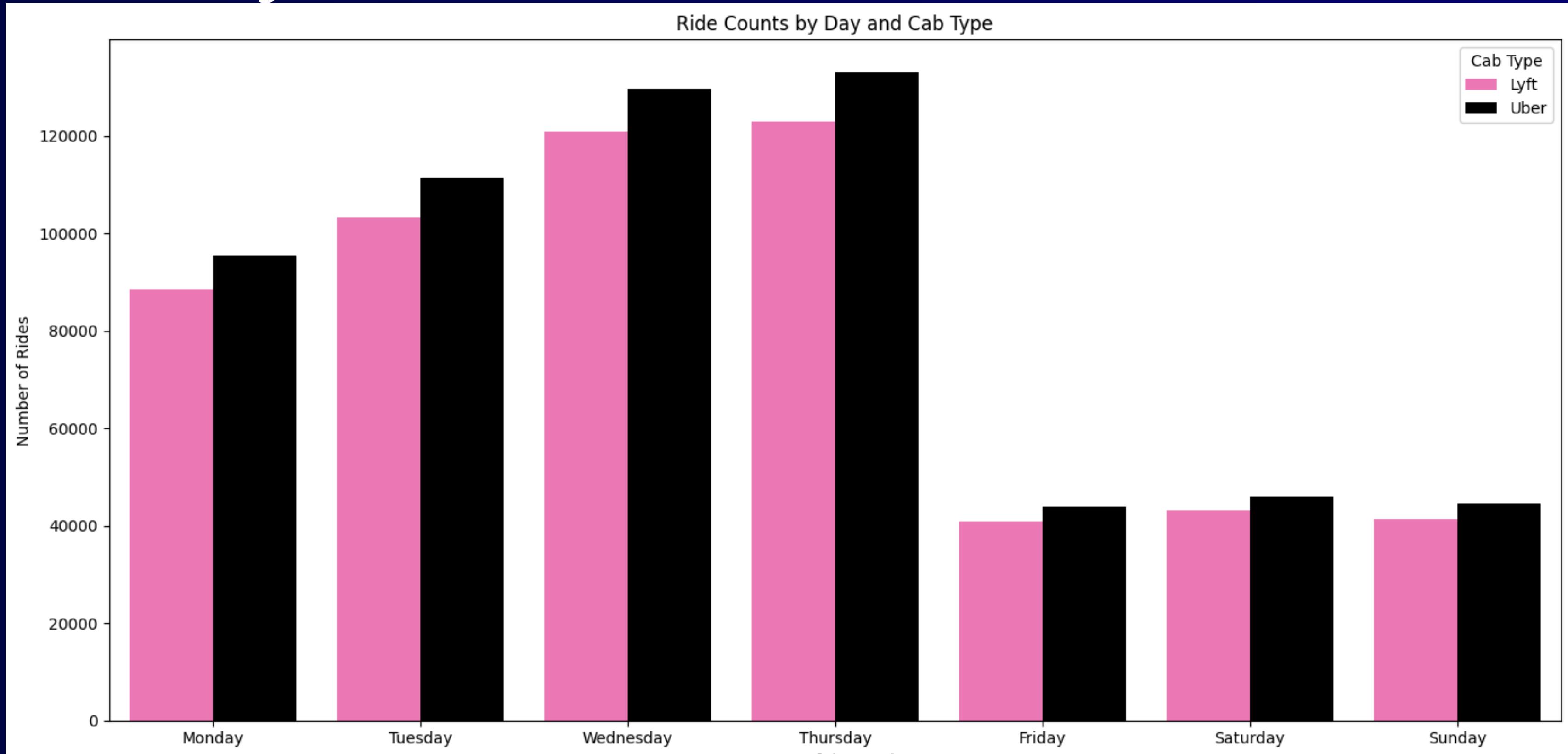
Data Analysis



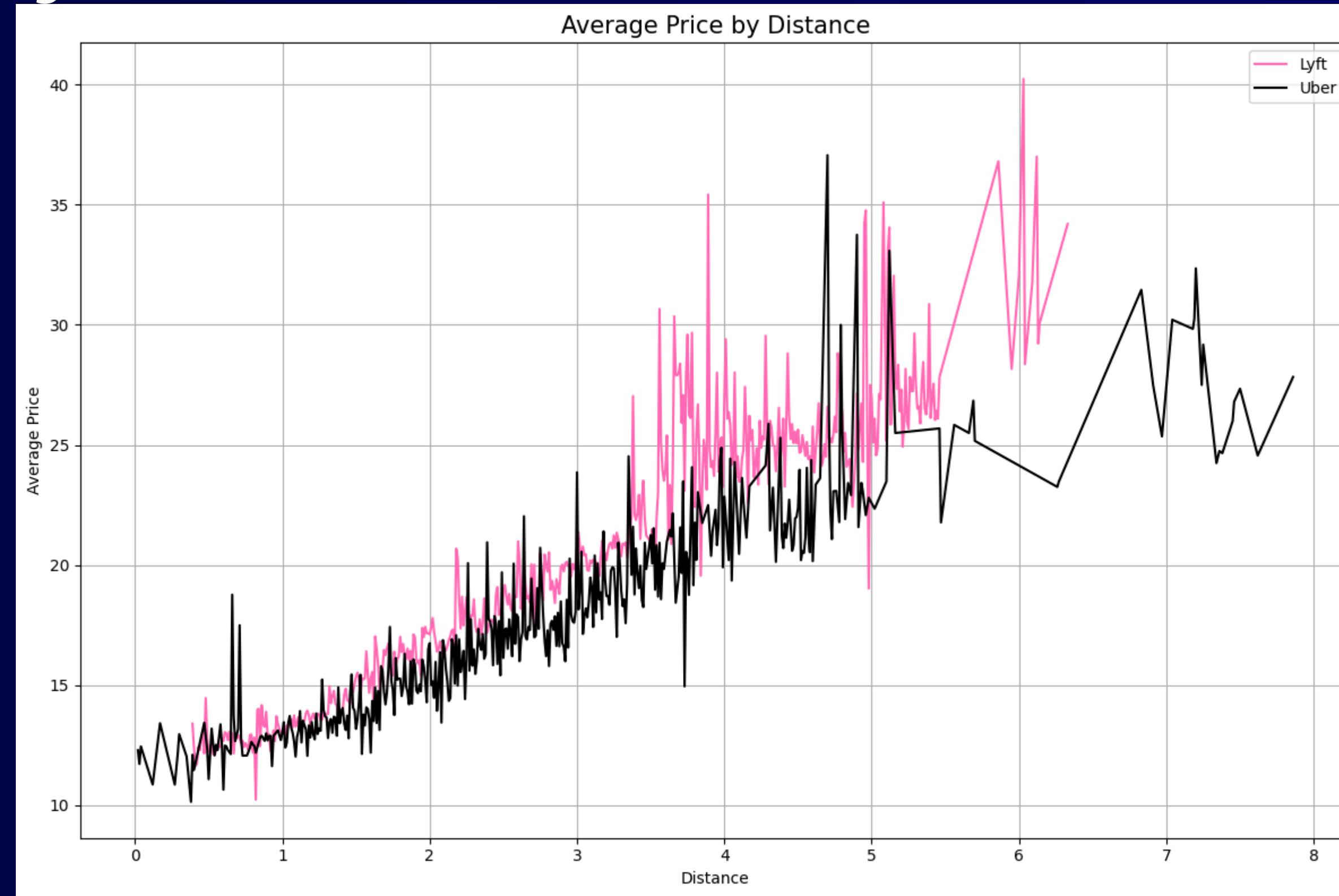
Data Analysis



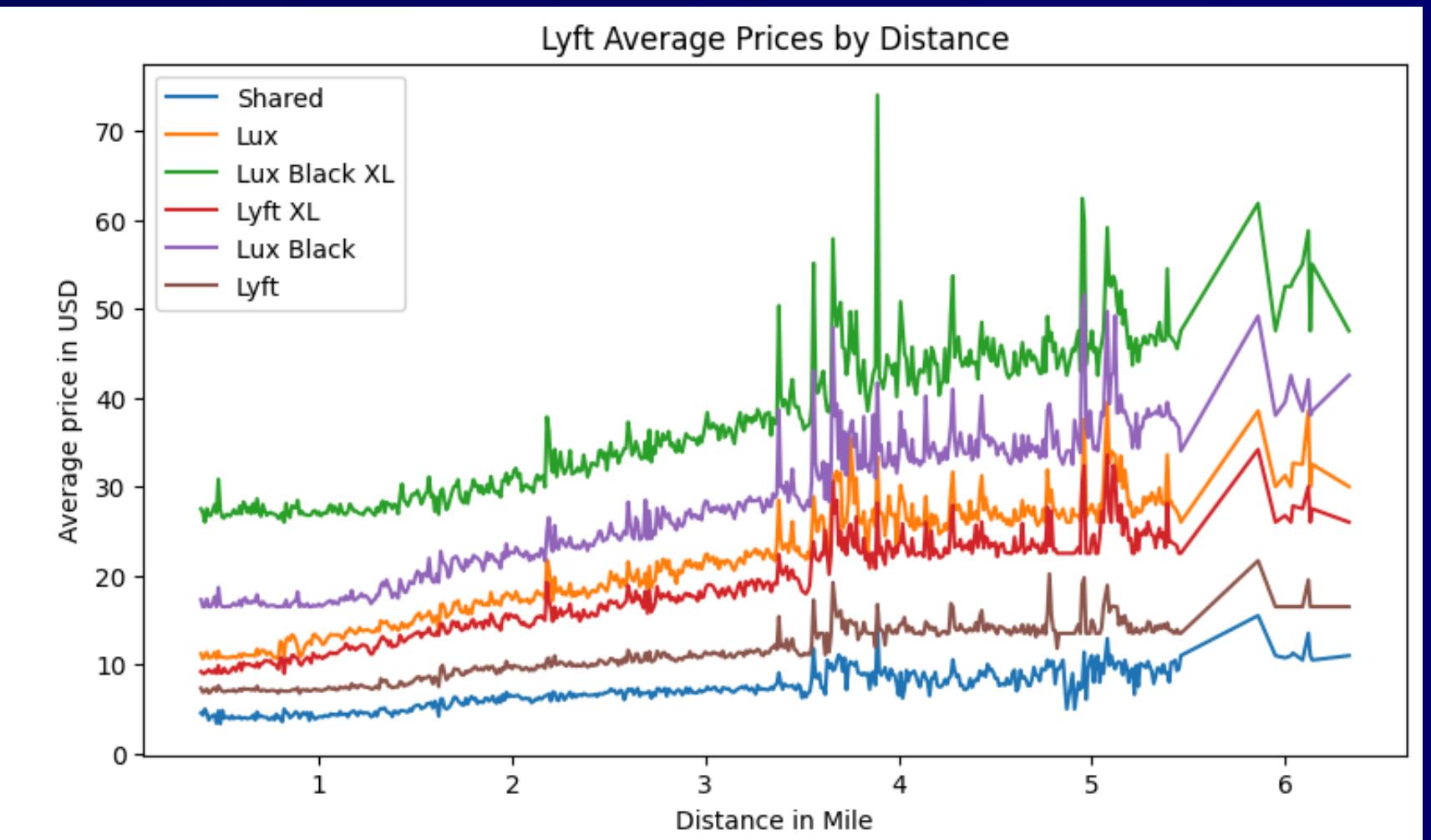
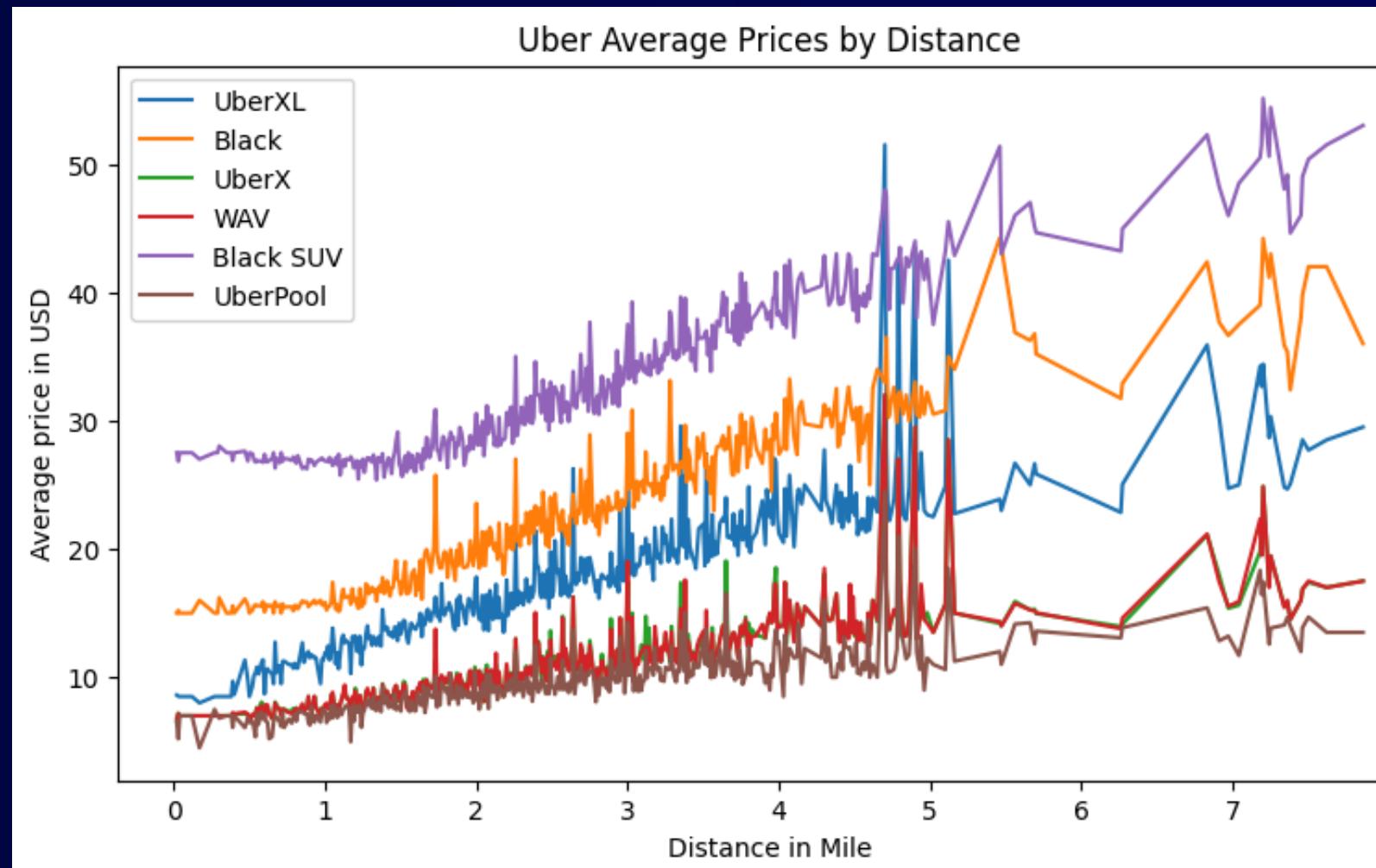
Data Analysis



Data Analysis



Data Analysis



MACHINE LEARNING

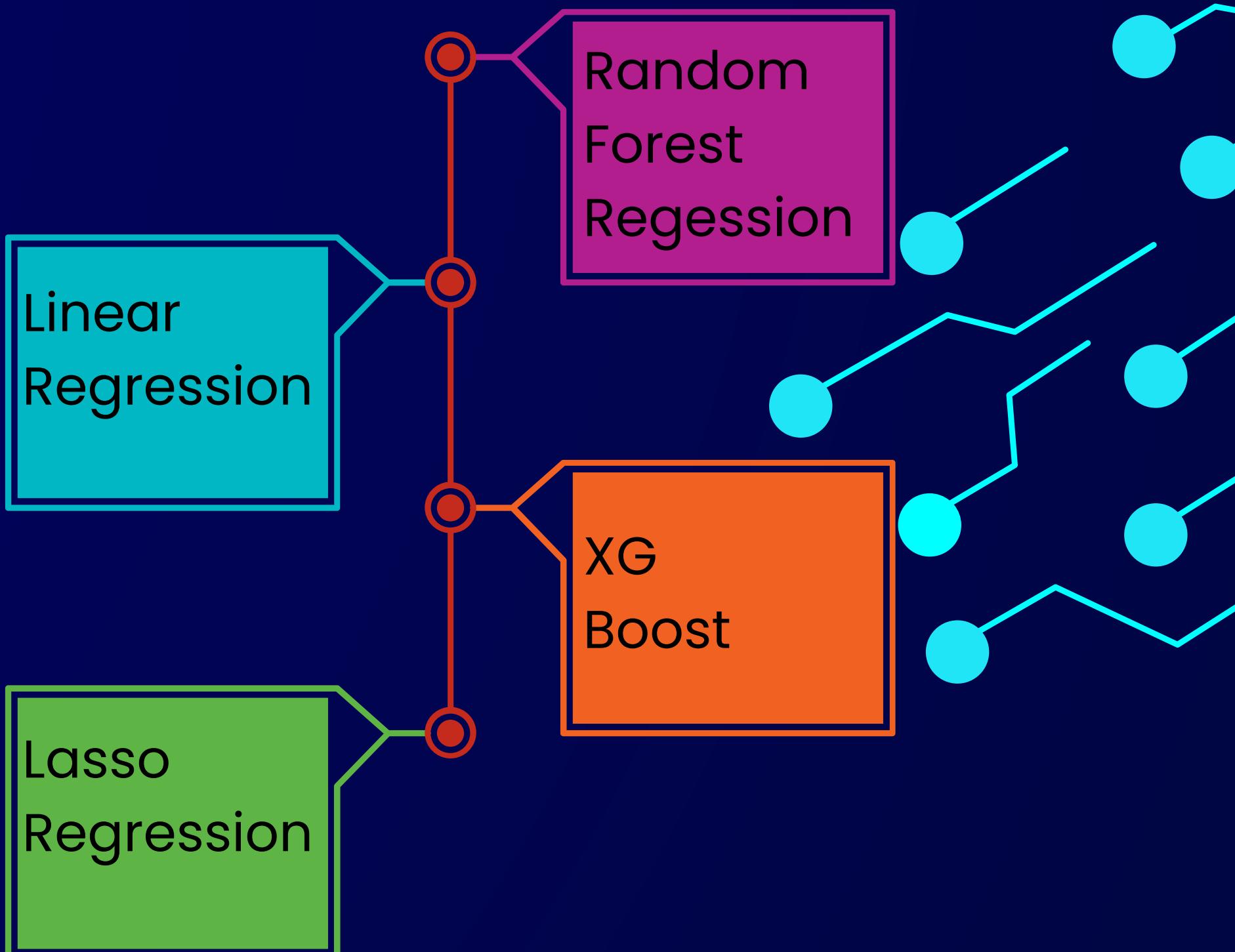
Preprocessing

OneHot Encoding

Standard Scaling

Imputing Null Values

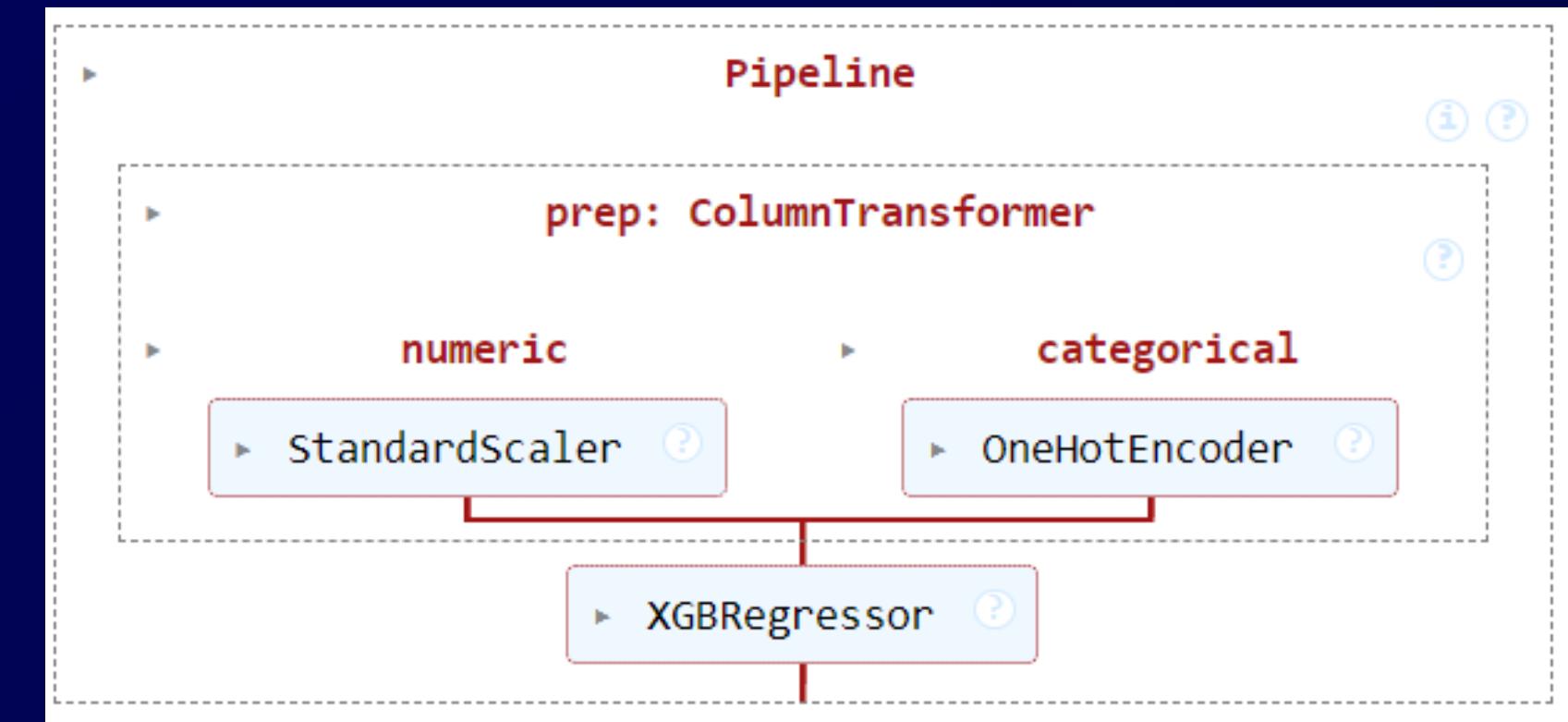
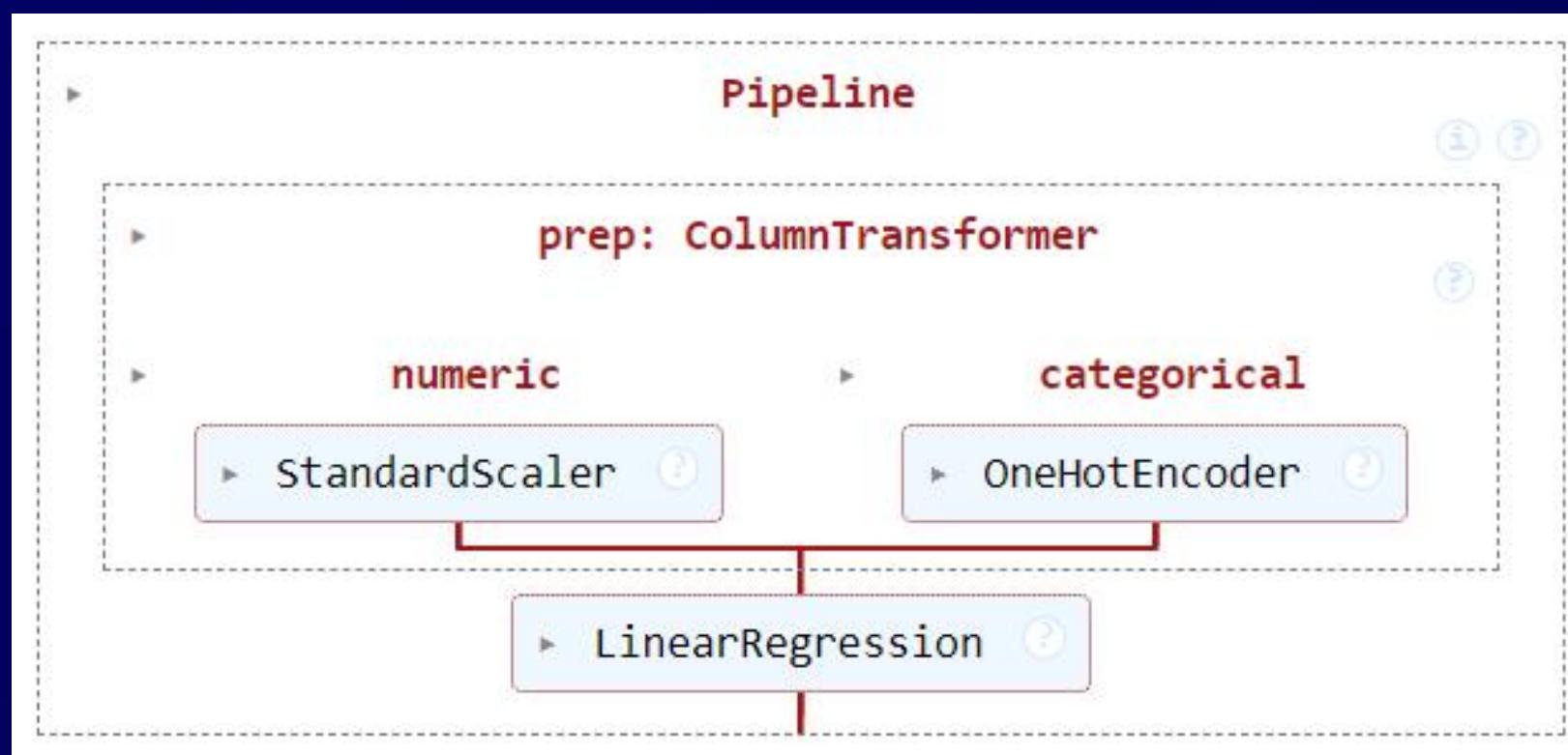
Models Used



Evaluation Metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- R-squared Score

Data Pipelines:



Model Performance:

Linear Regression

	Without Preprocessing	With Preprocessing
MSE	71.863	10.09
MAE	6.949	1.753
R ² Score	0.172	0.883

XG Boost Regression

	Without Preprocessing	With Preprocessing
MSE	71.063	2.804
MAE	6.921	1.109
R ² Score	0.181	0.968



Model Interpretation:

The biggest drivers in cab prices:

- Surge Multiplier, distance_log, cab type (uber), locations (BU, Fenway, Financial District, North Station), cab level (Black SUV, Lux, Black XL)

The biggest negative drivers in cab prices:

- Locations (Beacon Hill, Haymarket Square, South Station, cab type (Lyft), cab level (shared, uber pool, uberX)).

Key Takeaway:

Lyft has overall higher cab fares in the region, with natural factors having little but positive impact on the prices. Also, the prices sequentially drop as we go along the week and the year.

HYPER PARAMETER TUNED MODEL

Random Forest:

- Max Depth = 20
- Min Sample Leaf = 2
- Min Sample Split = 5
- N-Estimators = 100

Accuracy

- MAE = 0.965
- MSE = 2.367
- R² Score = 0.973



XG Boost Regressor

- Learning Rate = 0.2
- Max Depth = 8
- N-Estimators = 200

Accuracy

- MAE = 1.018
- MSE = 2.397
- R² Score = 0.972



Lasso Regression:

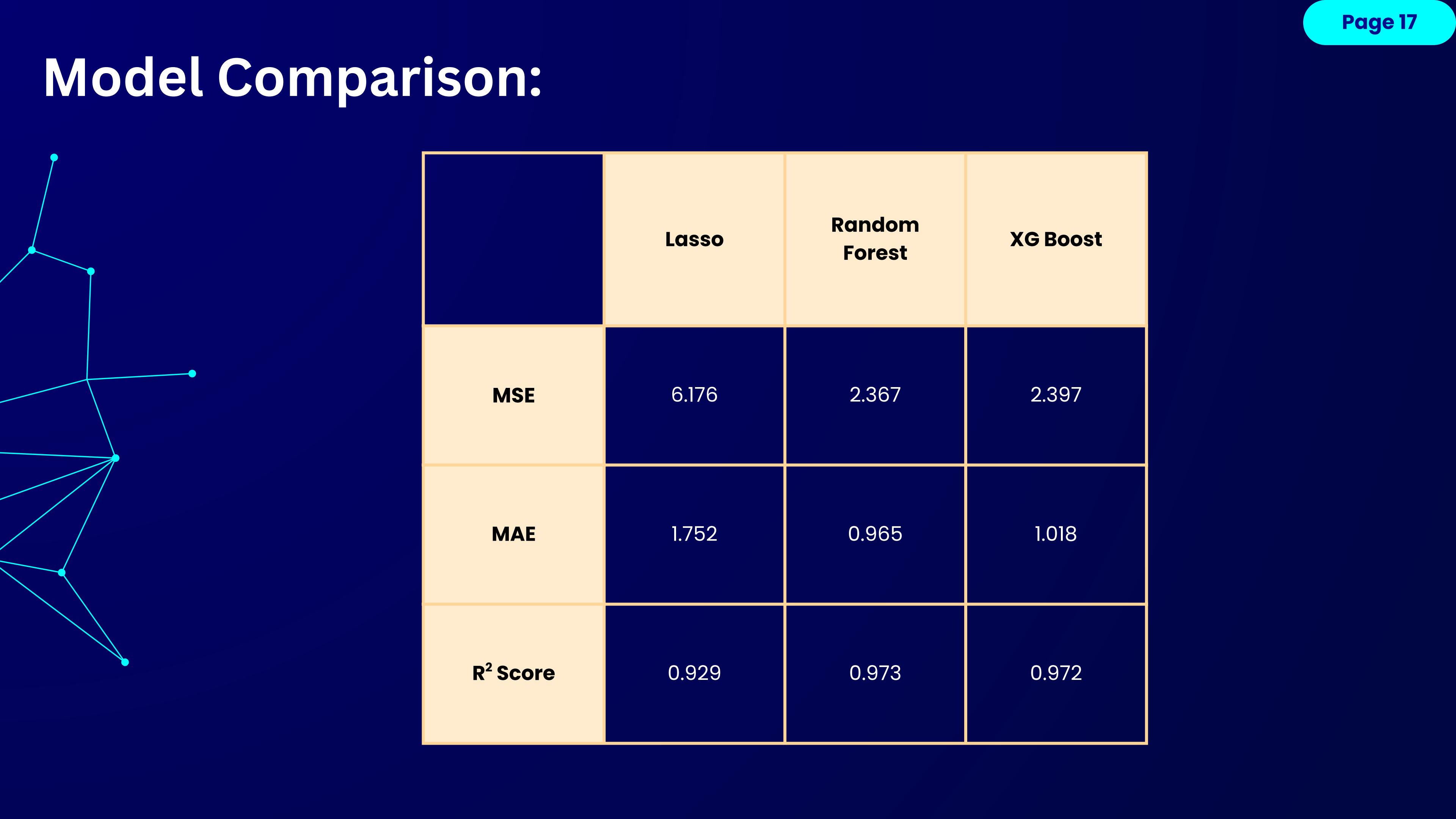
- Best Alpha = 0.001

Accuracy

- MAE = 1.752
- MSE = 6.176
- R² Score = 0.929



Model Comparison:



	Lasso	Random Forest	XG Boost
MSE	6.176	2.367	2.397
MAE	1.752	0.965	1.018
R ² Score	0.929	0.973	0.972

Conclusion:

- Weather conditions doesn't play a significant factor on pricing.
- Only one will get you to Starbucks slightly cheaper.
- Pricing vs distance trend is linear.
- Leave early if you want to save money on cabs.
- Random Forest and XG Boost with hyper parameter tuning gives significantly better accuracy in price prediction.

