**Uber v/s Lyft Prediction**

**Problem Statement:**

Two car-hailing firms, Uber, and Lyft, compete with one another for the same customers and partners. Ride costs on Uber and Lyft are not constant like those on public transportation. They are significantly impacted by the distance, weather conditions, and based on source and destination.

**Target Variable selection :**

The target variable, "classification," was selected as it determines the presence or absence of CKD. This binary classification problem aligns with the medical context and allows us to measure predictive accuracy and clinical significance.

**Which kind of machine learning is suitable for this problem?**

**Multivariable Linear Regression**

* Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. In our case, we have 21 independent variables and one target variable.
* In the below formula, y is the target price we want to predict, and x1,x2... are the independent variables.

**Support Vector Regression**

● Unlike the Linear Regression model which tries to fit a line that passes close through the majority of the points or the line which tries to minimize the error between predicted and real value, SVR tries to fit a hyperplane that covers the majority of the points.

● SVR, has a large fit time complexity, so we have only used 10% of the dataset for training.

**Random Forest Regressor**

* Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression.
* Performance Metrics of the regressor model are mean squared error, Mean Squared Error, R2.
* Generally, Random Forests produce better results, work well on large datasets, since we have a large dataset, we are going with Random Forest Regressor.
* Unlike SVR and Linear Regression, Trees can obtain non-linear relationships between input features and the target variables.
* In our case parameters are n\_estimators = 500, random\_state = 0, n\_jobs = -1.

**Conclusion with insights:**

● Overall Uber is less expensive than Lyft in Boston

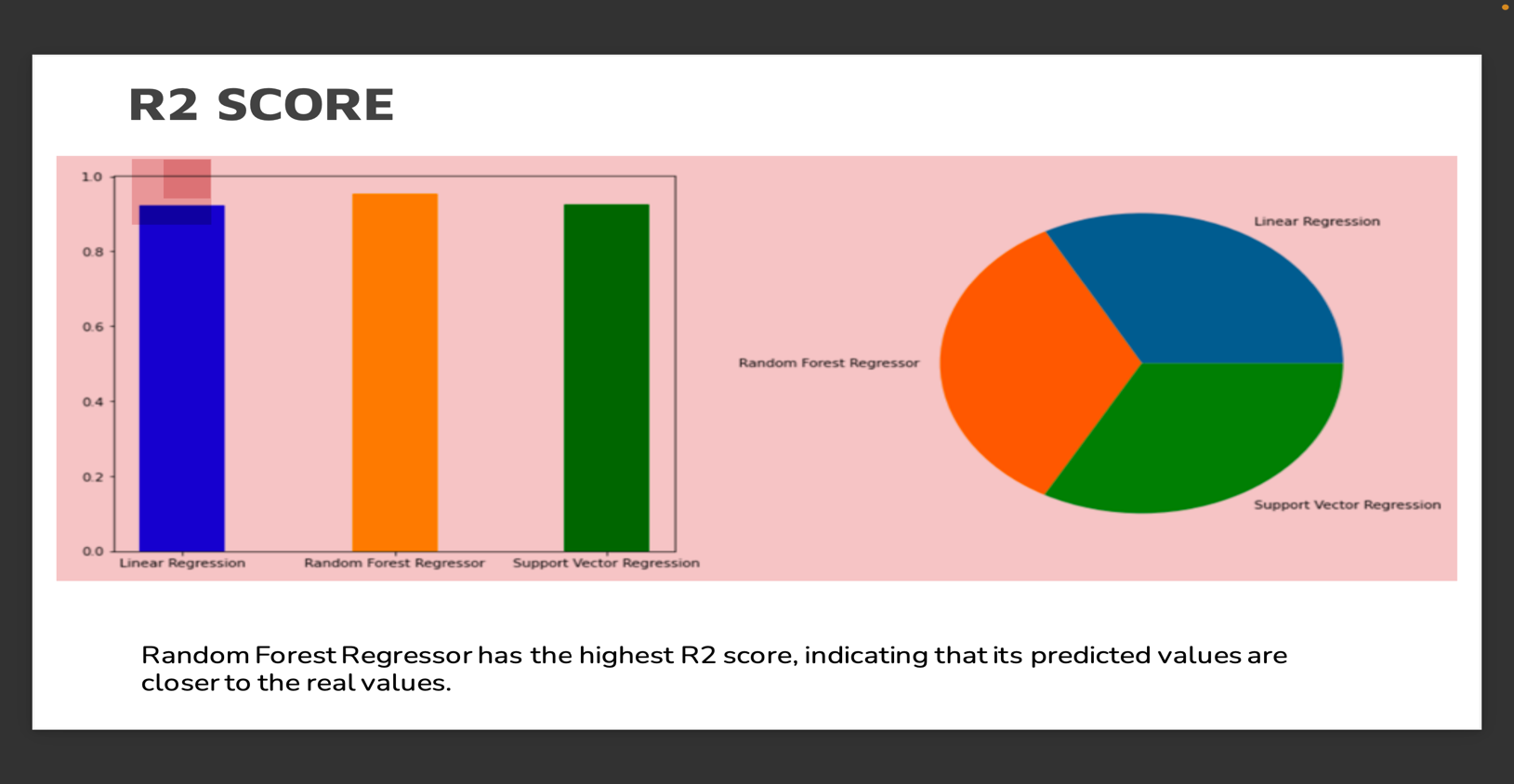
● Uber has more market share in Boston.

● In the economical category Lyft is less expensive compared to Uber. In the Luxury segment is Uber is less expensive.

● Distance and weather conditions significantly impact price.

● Among Support Vector Regression, Linear Regression and Random Forest regressor, Random Forest regressor has the best accuracy score.

● This model can be used as both B2B and B2C. Uber, Lyft or other upcoming competitors can use this model to analyze their competitors pricing. Customers can use our model to find the best prices for their ride.



**Scope for future:**

As we look ahead, there are several avenues to enhance the performance and insights of Chronic Kidney Disease (CKD) prediction.

Advanced Dimensionality Reduction: Using techniques like t-SNE (t-distributed Stochastic Neighbor Embedding) to analyze and visualize complex high-dimensional data, potentially revealing hidden clusters and patterns that can aid diagnosis and prognosis.

Enhanced Sampling Strategies: While undersampling might cause information loss, further research into various sampling strategies, like oversampling methods like SMOTE (Synthetic Minority Over-sampling Technique), can help to better understand how the classifiers react to various data distributions.

**Acknowledging Extra Efforts:**

This project goes beyond expectations by incorporating various dimensions to ensure a comprehensive exploration of Chronic Kidney Disease prediction. By conducting extensive model comparisons, fine-tuning hyperparameters, and focusing on clinical significance through evaluation metrics, the project demonstrates a commitment to delivering a meaningful solution.