

Vivek A

[Click here to view the github](#)

Used Car Cost prediction

Data Science project

Project Objective

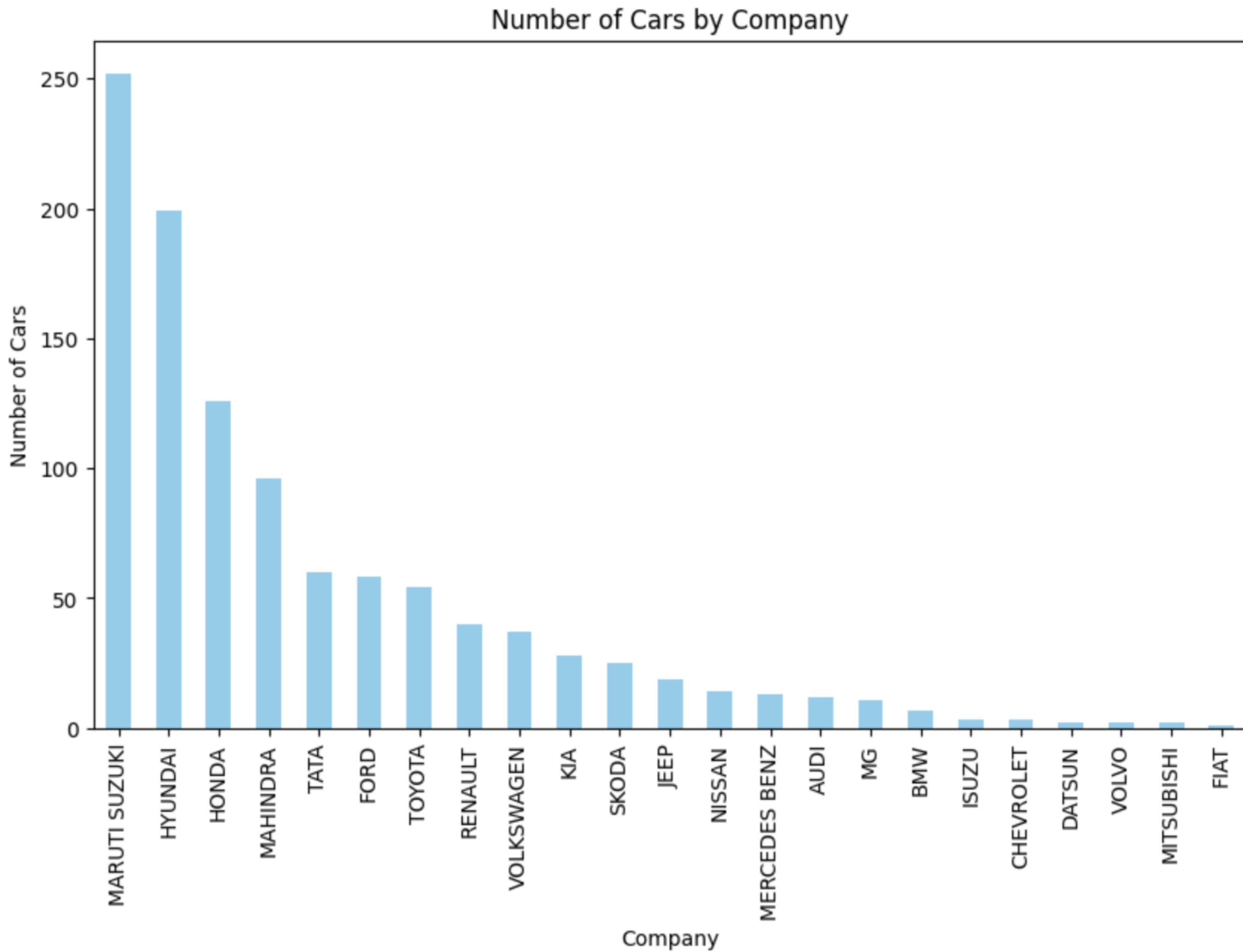
The objective of this project is to develop a machine learning model that accurately predicts the resale price of used cars based on various factors such as brand, model, fuel type, odometer reading, body style, and car age. By analyzing the dataset, we aim to identify key trends in the used car market, such as demand patterns and price determinants, and provide valuable insights for both buyers and sellers. This will help stakeholders make informed decisions regarding the pricing and purchasing of used cars.



Insights from EDA

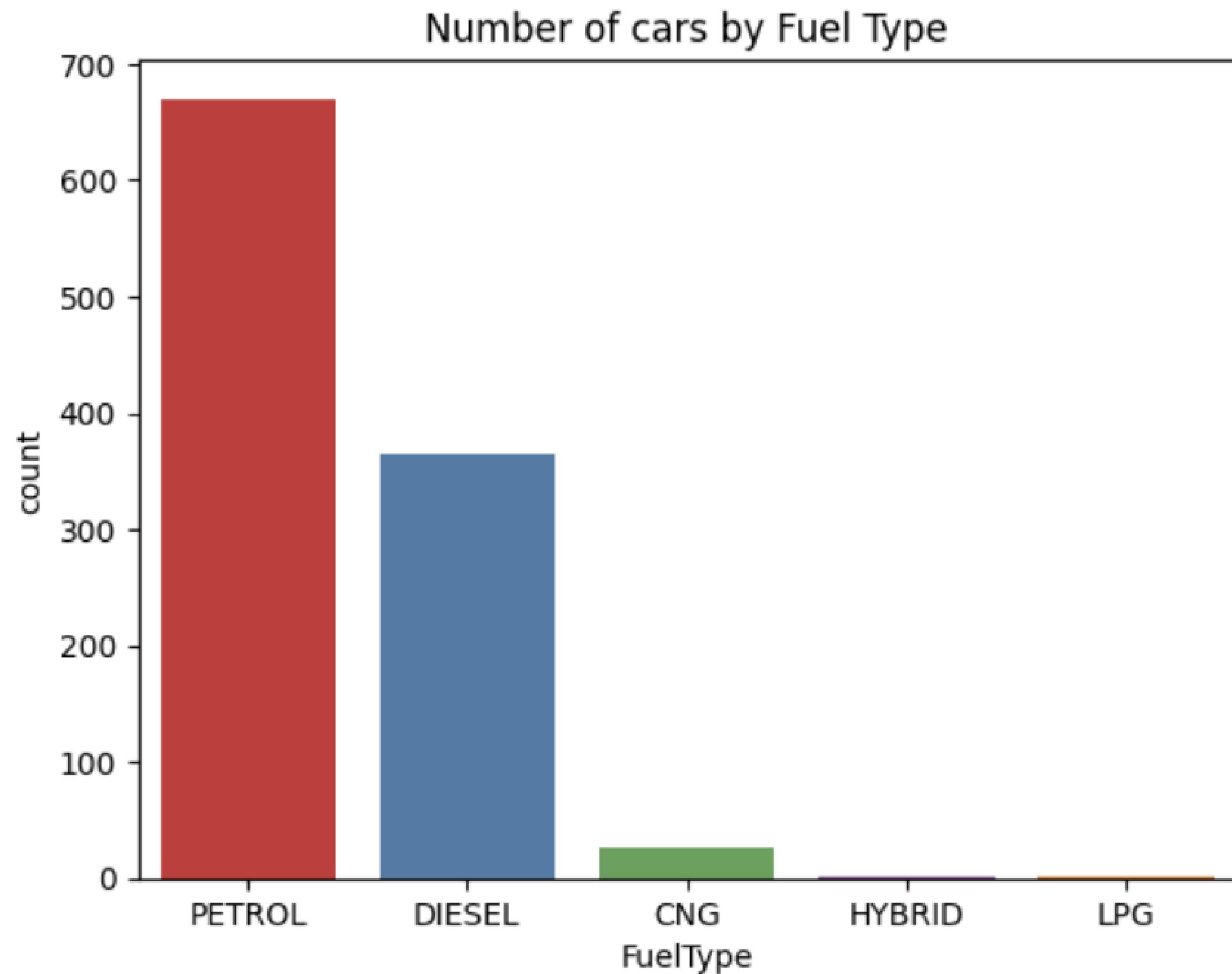
Company Distribution

From the graph, we can observe the distribution of cars in the dataset across various companies. There are a total of 23 companies represented, with Maruti Suzuki, Hyundai, Honda, Mahindra, and Tata emerging as the top five companies with the highest number of used cars for sale. This suggests that these companies' cars are more durable and likely to have a strong resale value, making them popular choices in the used car market.



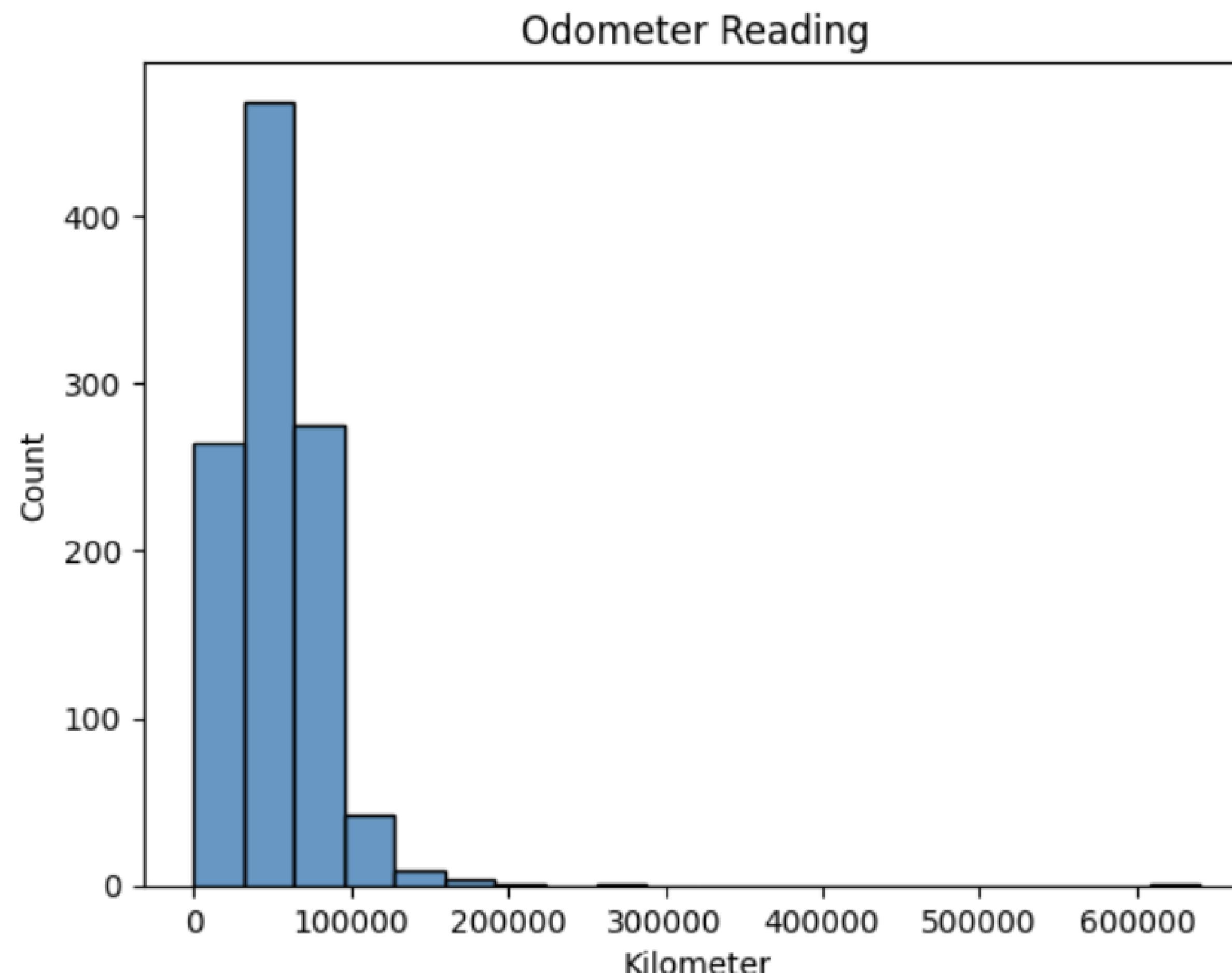
Fuel Type Distribution

Petrol and diesel cars have higher distribution compared to other



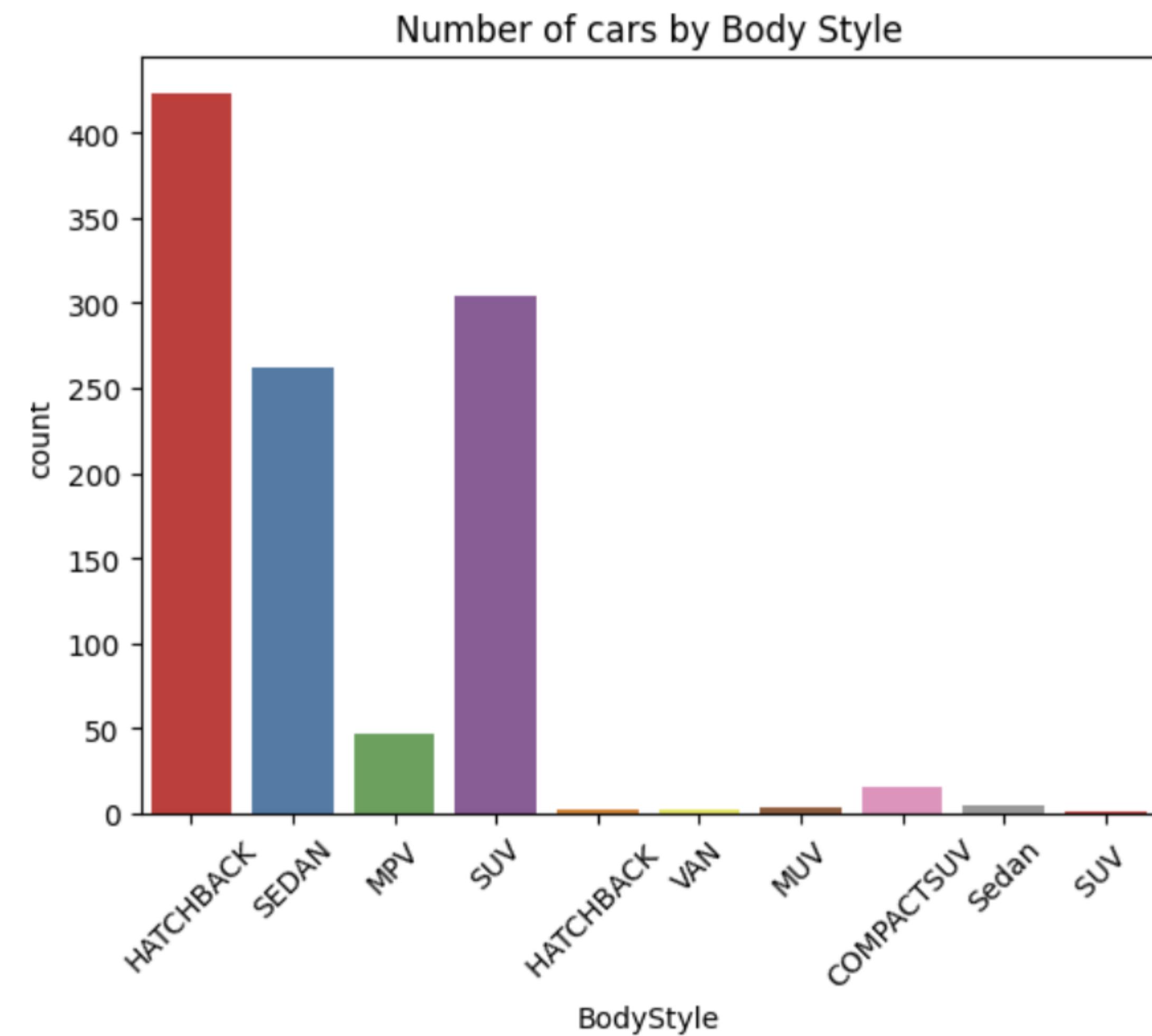
KM Driven among the cars

This graph illustrates the distribution of odometer readings for the cars in the dataset. It shows that most cars have been driven less than 100,000 km, with the majority falling within the 30,000 km to 50,000 km range. Therefore, we can infer that cars with lower odometer readings, particularly under 100,000 km, are in higher demand in the used car market and likely to have better resale value.



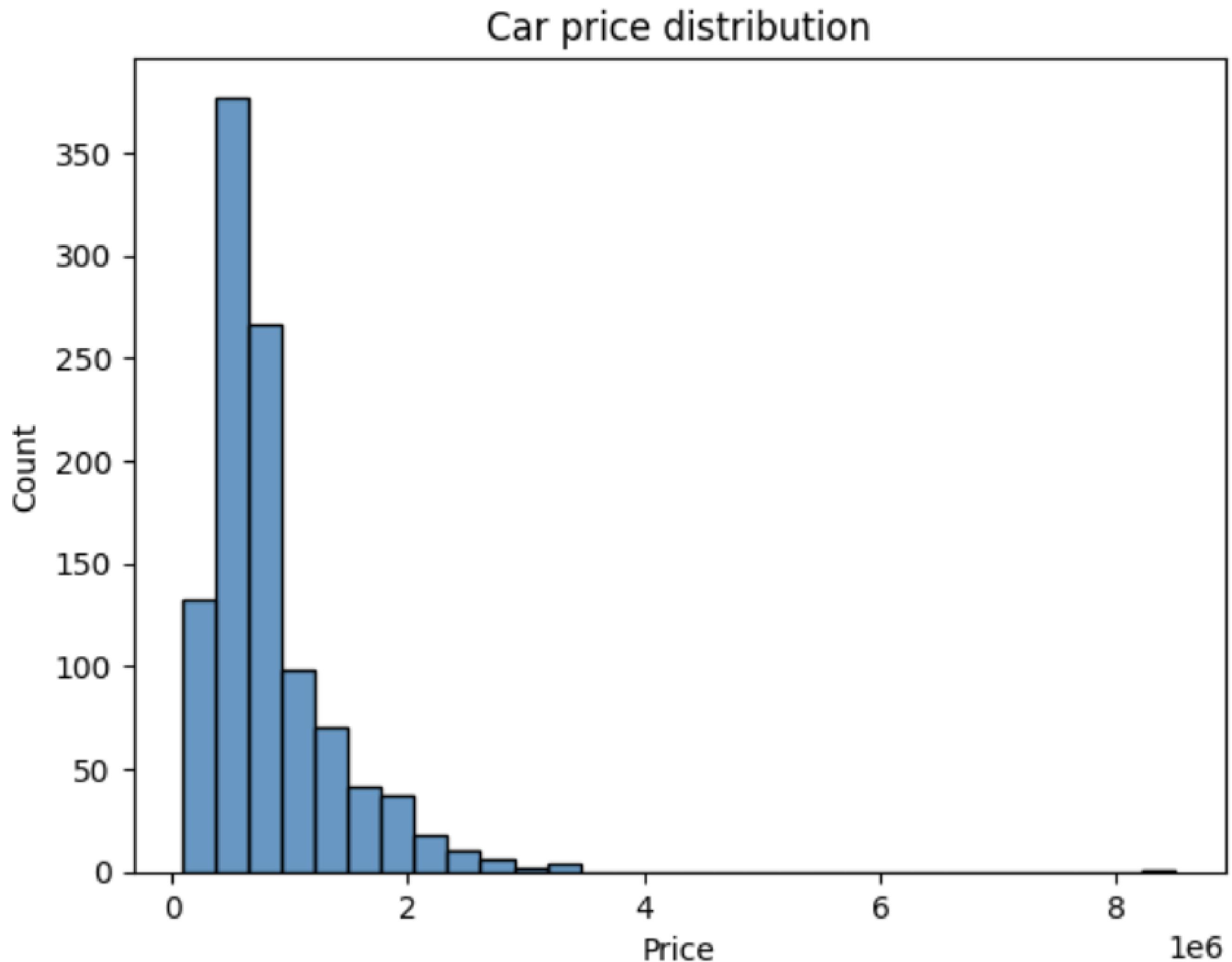
Vehicle Type

According to this graph, most of the cars have HatchBack, SUV and Sedan body style, which tells us about the market demand of these body styles. Therefore, we can assume that cars with HatchBack, SUV and Sedan body style are more in demand in the used car market will have a good resale value



Price distribution

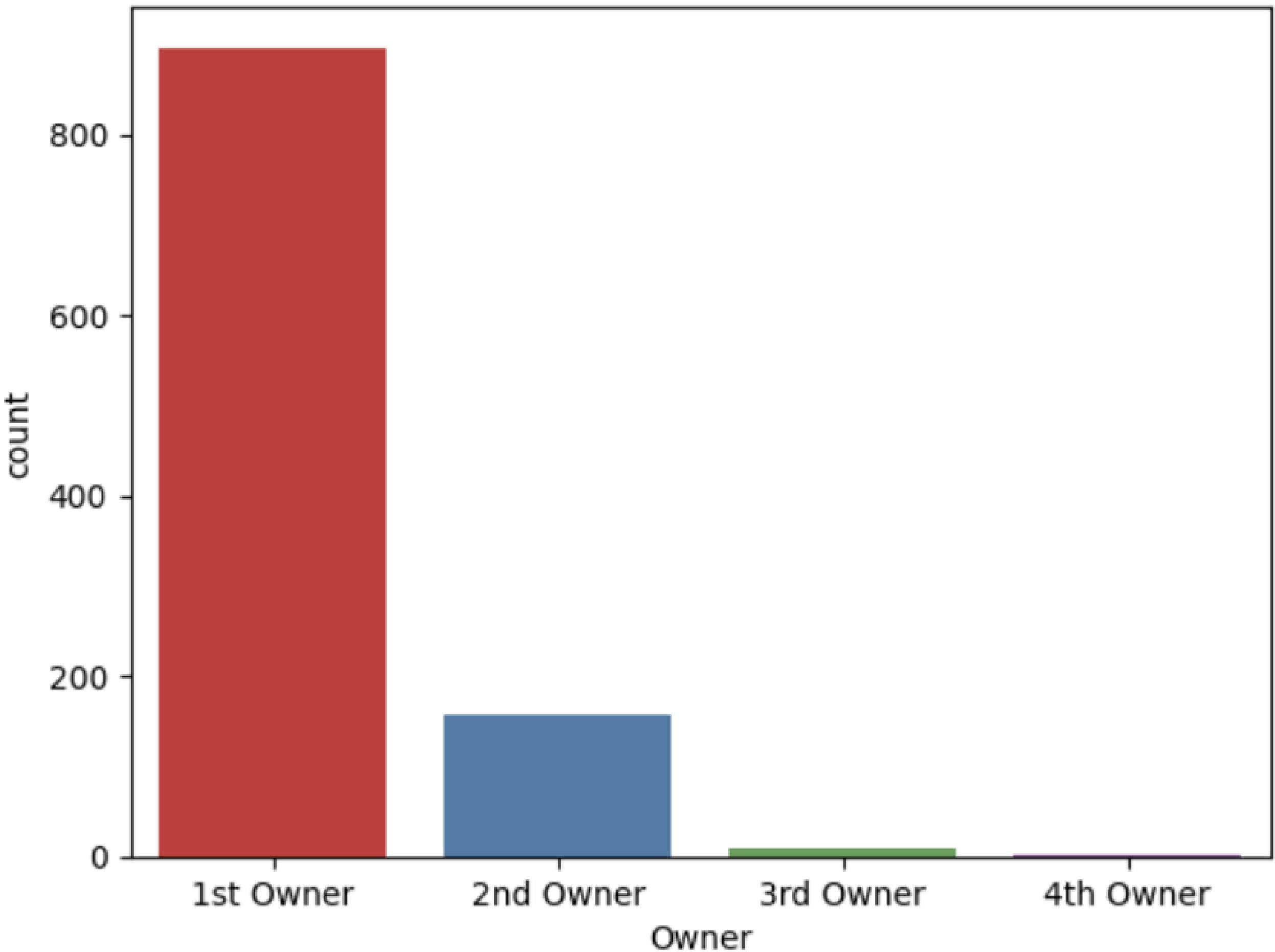
This graph shows the distribution of car prices in the dataset. Most cars are priced between 3 to 9 lakhs, with the highest concentration between 3 to 6 lakhs. This suggests that cars within this price range are more popular in the used car market. Additionally, there are a few cars priced above 20 lakhs, which could either represent luxury vehicles or potential outliers.



Owner Type

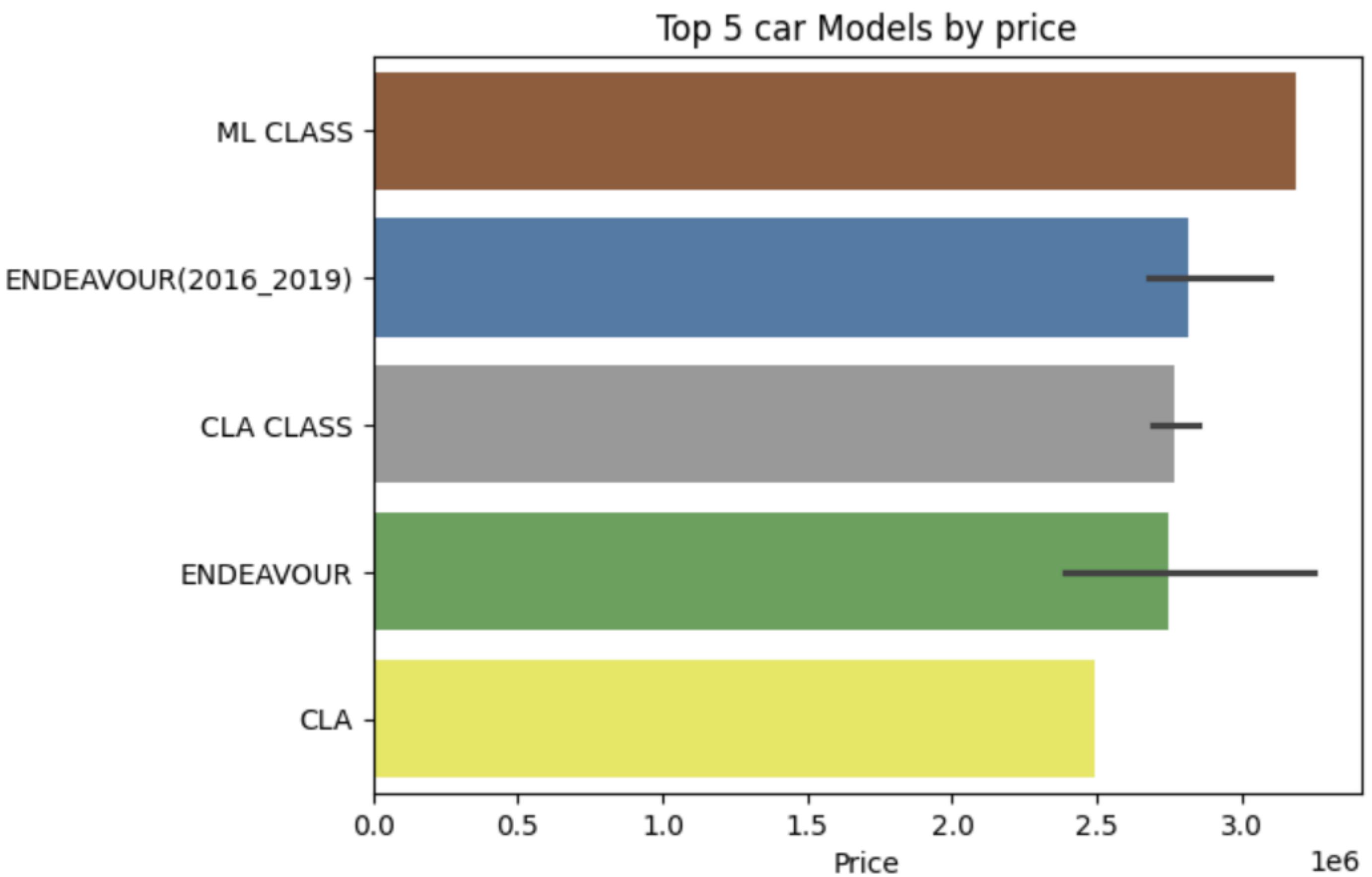
Majority of the distribution is by single owners.

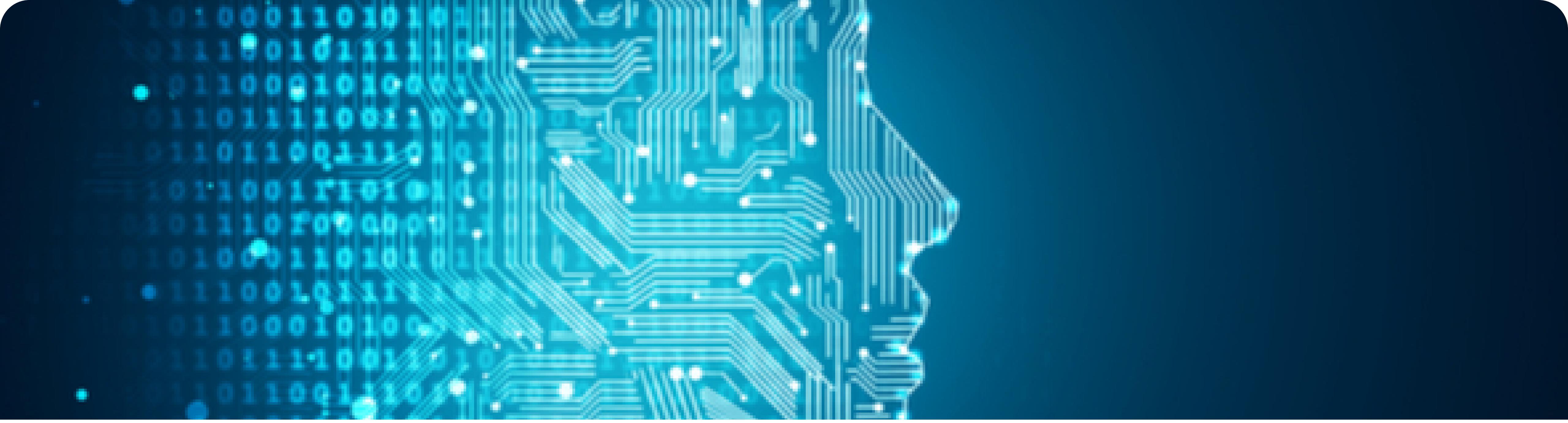
Number of cars by Owner Type



Top 5 Costliest Models

ML Class, Endeavour(2016_2019),
CLA class are the top three models
with highest resale value





Predictive Models

Linear Regression

```
# Model evaluation  
print('Train RMSE:', mean_squared_error(y_train, y_pred_train, squared=False))  
print('Test RMSE:', mean_squared_error(y_test, y_pred_test, squared=False))  
print('R2 Score:', r2_score(y_test, y_pred_test))
```

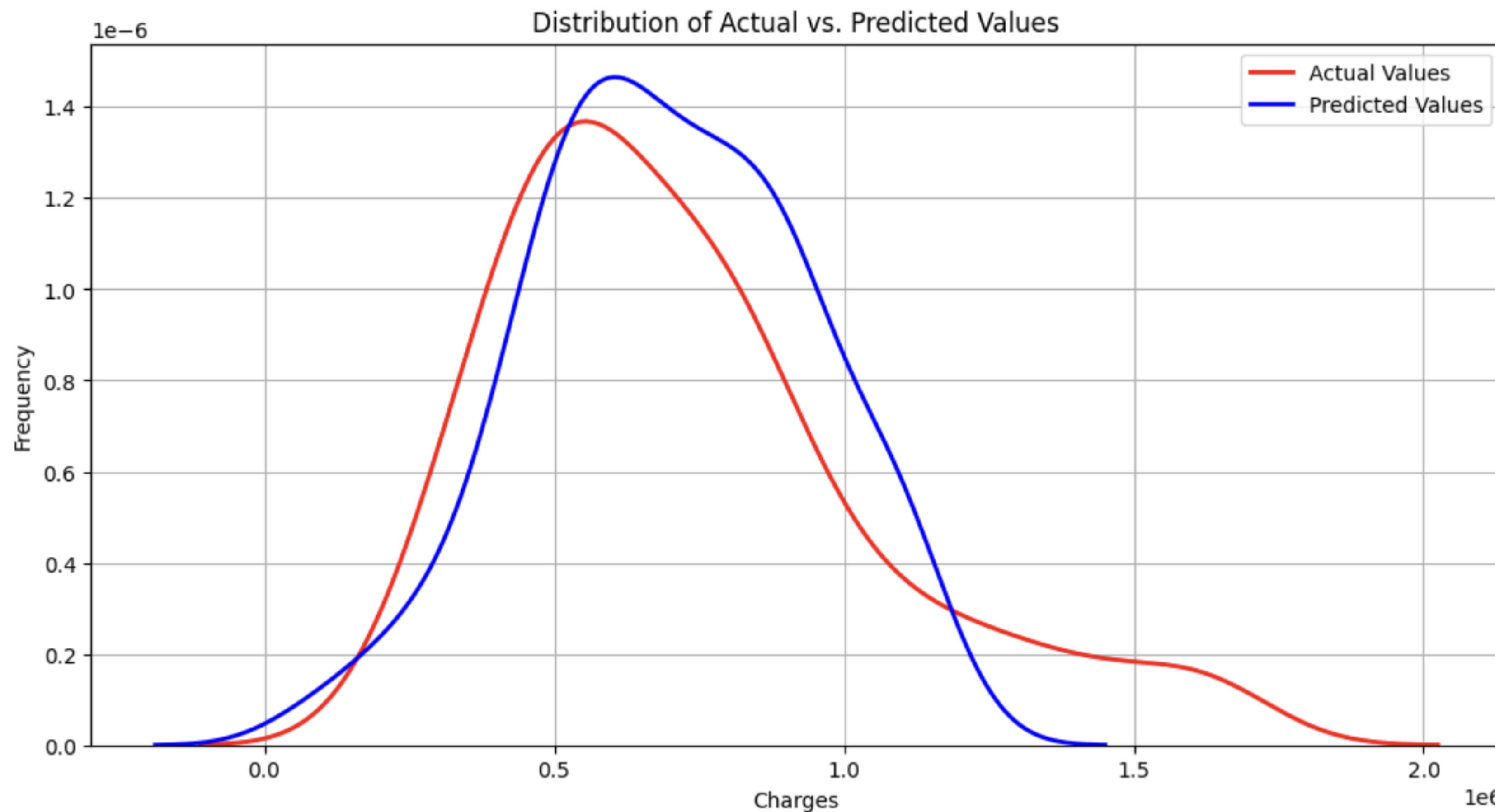
Train RMSE: 228912.33202227353

Test RMSE: 214241.00020621895

R² Score: 0.5739340729922839

- Linear regression model has 57.39 % accuracy

Linear Regression – Actual vs predicted



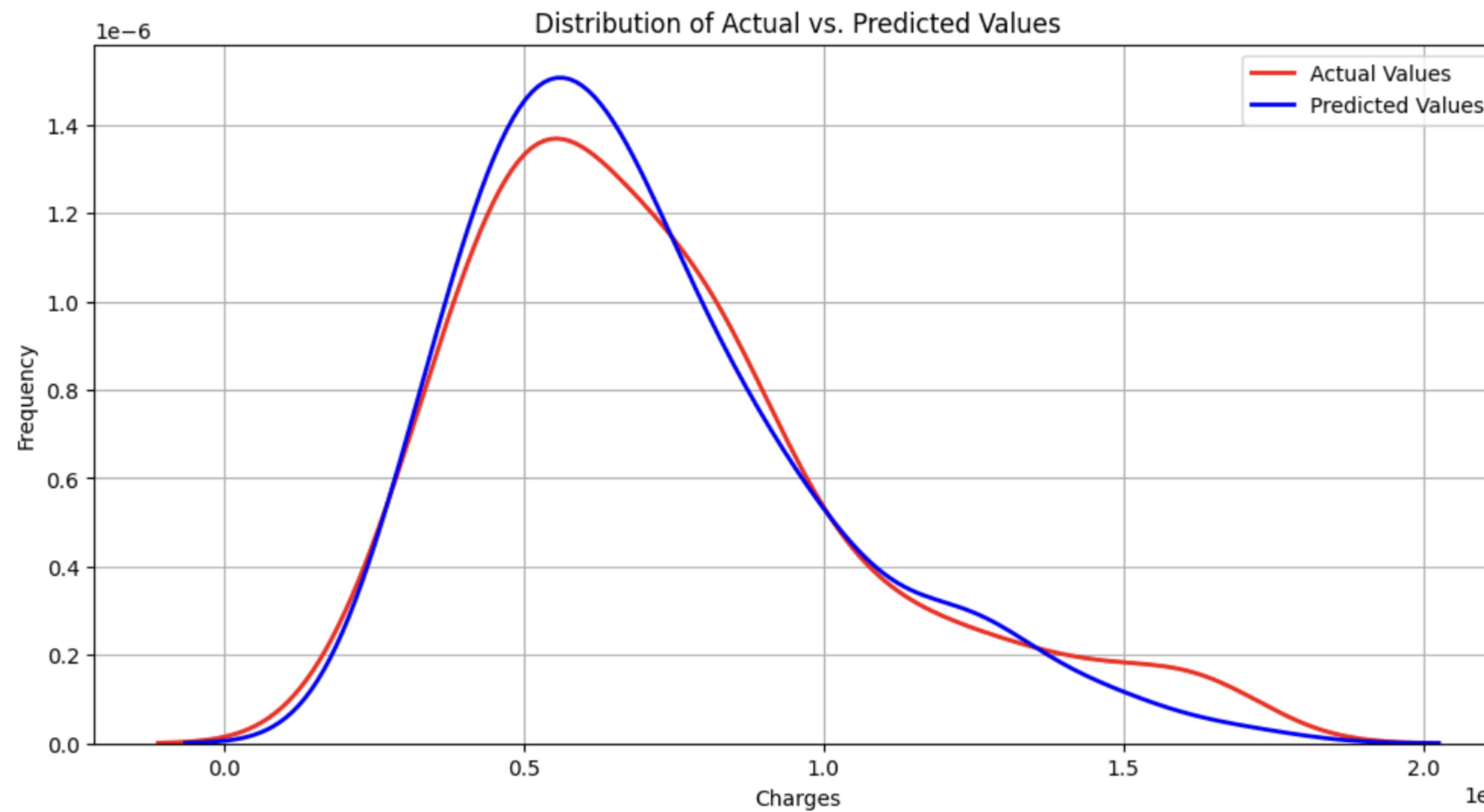
Random Forest

```
#Fitting the model  
rf_model.fit(X_train, y_train)  
  
#Training score  
print(rf_model.score(X_train, y_train))
```

0.8781873430425237

- Random Forest model has 87.82 % accuracy

Random Forest - Actual vs Predicted



Conclusion

From the exploratory data analysis, two key insights emerged about the used car market: demand and price. There is higher demand for low-priced used cars compared to expensive ones, showing that customers are more attracted to budget-friendly options.

In terms of car companies, brands like MG, Mercedes Benz, BMW, Volvo, and KIA have the highest prices, but Maruti Suzuki, Hyundai, Honda, Mahindra, and Tata cars are in greater demand. This indicates that while luxury cars tend to be more expensive, customers prefer to buy new luxury cars rather than used ones.

Most cars run on petrol or diesel, with diesel cars generally priced slightly higher. Additionally, cars in common colors like white, grey, silver, and black are more in demand, but exotic colors such as burgundy, dark blue, and black magic tend to have higher prices.

Conclusion

Regarding odometer readings, the majority of cars have been driven less than 10,000 km, and cars with lower mileage generally command higher prices. In terms of bodystyle, Hatchback, SUV, and Sedan are the most preferred, while MPV, SUV, and Sedan are the most expensive.

Car age also impacts resale value, with cars less than 5 years old being more in demand and priced higher. Additionally, location influences pricing, with Delhi, Maharashtra, and Rajasthan being the top states for higher prices. Car Estate, Star Auto India, and Car Choice are the top dealers with the highest prices.

Lastly, cars with 1st owner type are in higher demand and priced higher. Cars with a warranty provide added assurance to buyers, which increases their price slightly, while cars with a higher quality score tend to have higher prices as well.

Thanks

Thank you everyone who been with me with this journey.
Check the [github repository](#) for more info about the project