Introduction

COMPANY OVERVIEW: GEELY AUTO, A LEADING CHINESE AUTOMOBILE COMPANY, PLANS TO ENTER THE US MARKET

OBJECTIVE: THE PROJECT AIMS TO MODEL THE PRICE OF CARS BASED ON VARIOUS ATTRIBUTES, PROVIDING INSIGHTS TO HELP GEELY AUTO OPTIMIZE ITS CAR DESIGNS AND BUSINESS STRATEGIES FOR THE AMERICAN MARKET

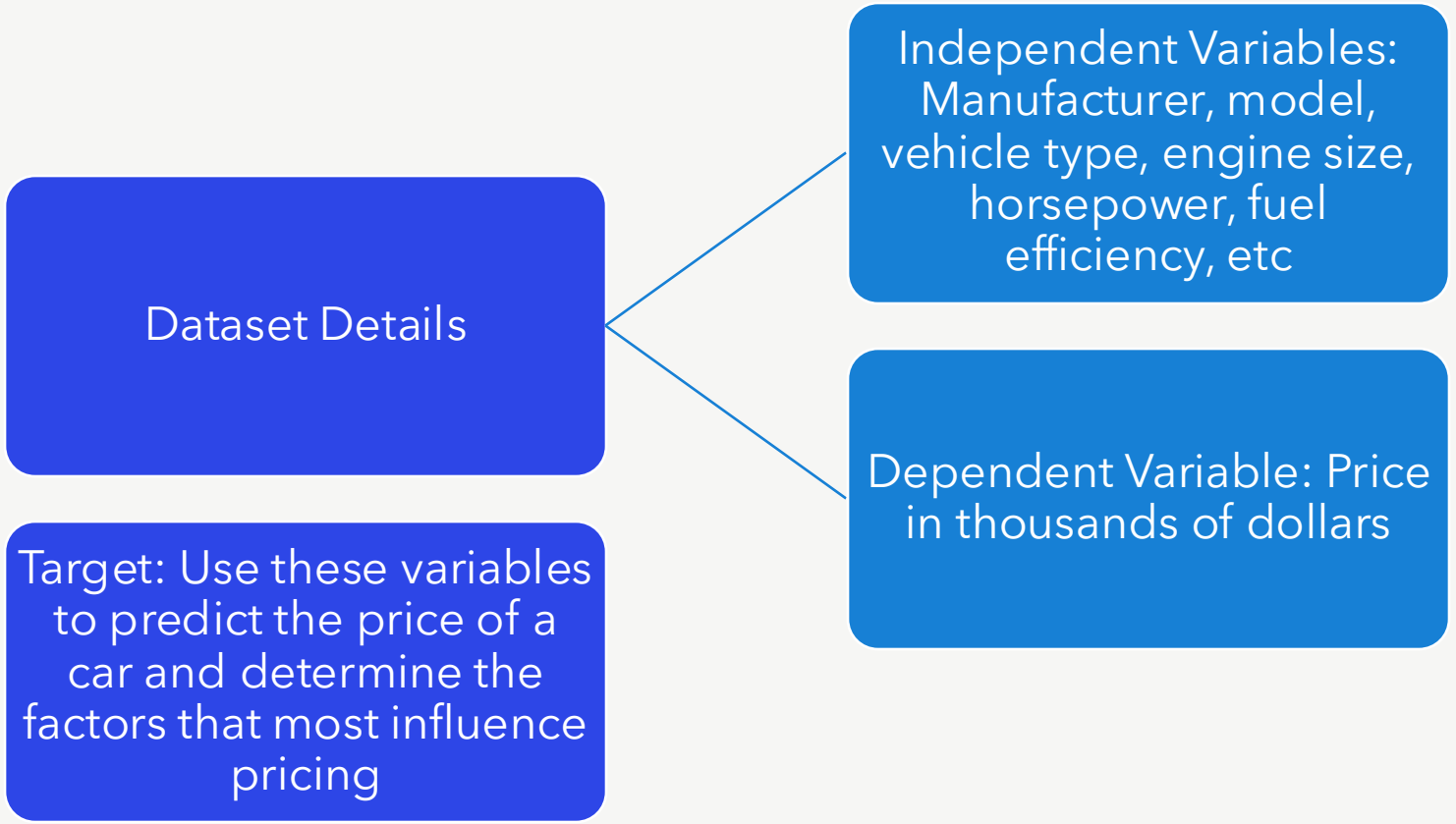
Business Problem

Key Questions

Which variables are significant in predicting car prices?	How can these insights help Geely Auto adjust their business strategy?	What can Geely learn about pricing dynamics in the new market?

Goal: Develop a model to predict car prices based on a variety of features and provide actionable insights for Geely's management

Dataset Overview



Data Exploration

Numerical Feature Distribution

- Analyzed features such as engine size, horsepower, car dimensions, and fuel efficiency
- Dataset shows both positively and negatively skewed features due to cars from different price ranges

Comparison of Mean and Median

- Highlighted that skewness impacts the distribution of data

Key Insights from EDA

Correlation Analysis

- Positive Correlations: Price increases with higher engine size and horsepower
- Negative Correlation: Fuel efficiency decreases as price increases

Linear Regression Insights

- Visualized relationships using scatter plots and linear regression lines to better understand pricing patterns

Manufacturer Insights

Manufacturer Distribution

Dodge and Ford have the highest number of car models in the dataset

Most Expensive Model

Mercedes-Benz CL500 is the priciest model at $85

Top Manufacturer by Average Price

Porsche has the highest average price , followed by Mercedes-Benz

Vehicle Type Distribution

Breakdown

Vehicle Types: The dataset includes 116 passenger vehicles and 40 cars

Common Types

Passenger vehicles are the most common, followed by cars

Price Comparison

The average price of passenger vehicles is slightly higher than cars

Engine Size & Horsepower

Largest Engine

Dodge Viper leads with the largest engine size at 8.0 liters

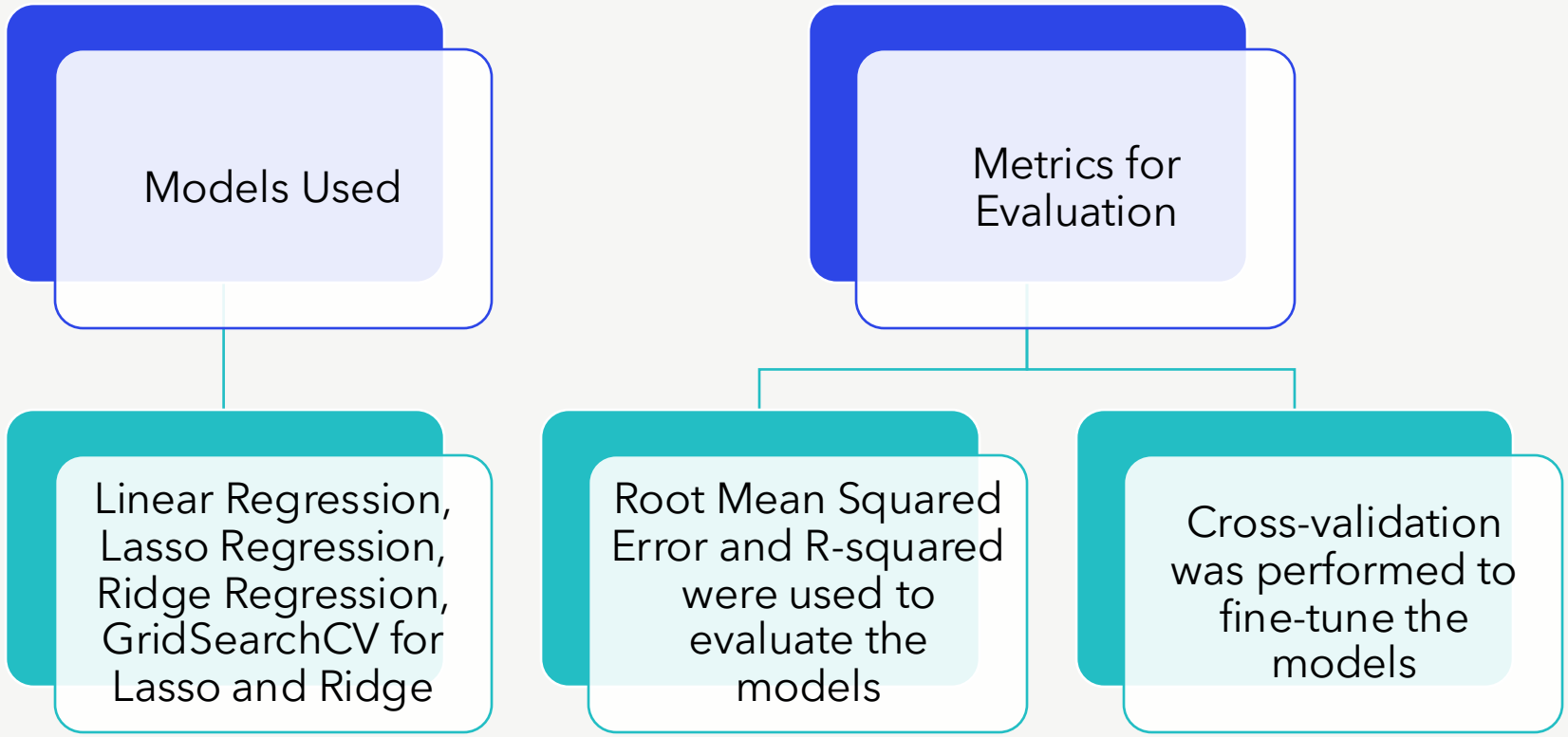
Highest Horsepower

The Dodge Viper also has the highest horsepower

Engine Size vs. Fuel Efficiency

A scatter plot reveals an inverse relationship between engine size and fuel efficiency, where larger engines generally result in lower fuel efficiency

Model Selection



Model Performance

Best Performing Model

GridSearchCV with Lasso Regression had the best performance

Key Results

Test RMSE: 0.0501

Train RMSE: 3.2774

Test R-squared: 0.9434

Train R-squared: 0.9325

Conclusion: The GridSearchCV Lasso model had the lowest error and highest R-squared, making it the most reliable for price prediction

Conclusion

Key Takeaways

- Engine size, horsepower, and fuel efficiency are significant factors influencing car prices
- Fuel efficiency negatively correlates with price, while engine size and horsepower positively correlate
- The Lasso Regression model with GridSearchCV provided the best predictive results