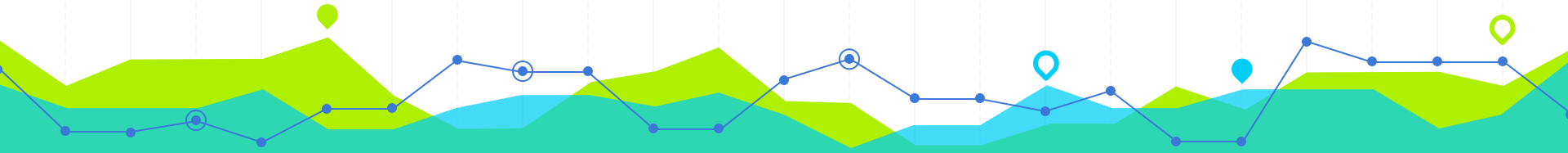


LINEAR REGRESSION ANALYSIS
Car Data

OVERVIEW

QUESTION

Can car value be predicted based off specific attributes of a car?

OBJECTIVE

Collect & analyze car data and build different regression models to make predictions

GOAL

Provide automakers a model to predict car listing price

OUR PROCESS

COLLECT DATA

- Data scraped from car data website
  - Randomly select 25 cars from 45 sitemaps
  - 15 feature, 900 rows
  - https://www.cars-data.com/

BUILD MODELS

- LASSO Regression
  - remove unnecessary features
- Linear Regression
  - baseline after feature removal
- Ridge Regression
  - reduce effects of collinear features

OUR PROCESS

MODEL PERFORMANCE METRICS

- Which model had the highest adjusted R^2
- Which model had the lowest Mean Absolute Error (MAE)
- Which features were most impactful

RESULTS

Model Selection

- **Ridge Regression**
  - highest R^2 adjusted
  - lowest error

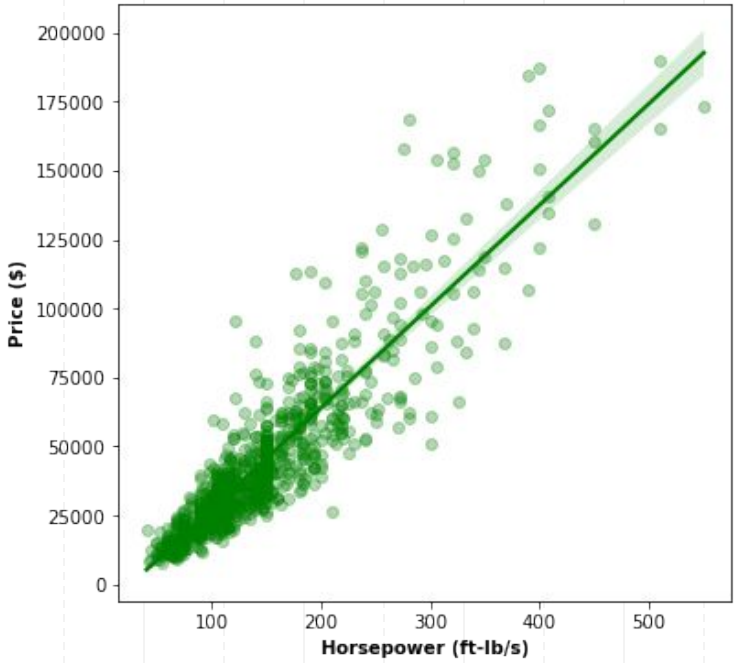
Model Type	# of Features	R^2 Adjusted	MAE
LASSO Regression	95	0.840	$5536
Linear Regression	75	0.859	$5850
Ridge Regression	75	0.873	$5524

RESULTS

Coefficient Analysis

- **Horsepower most impactful**

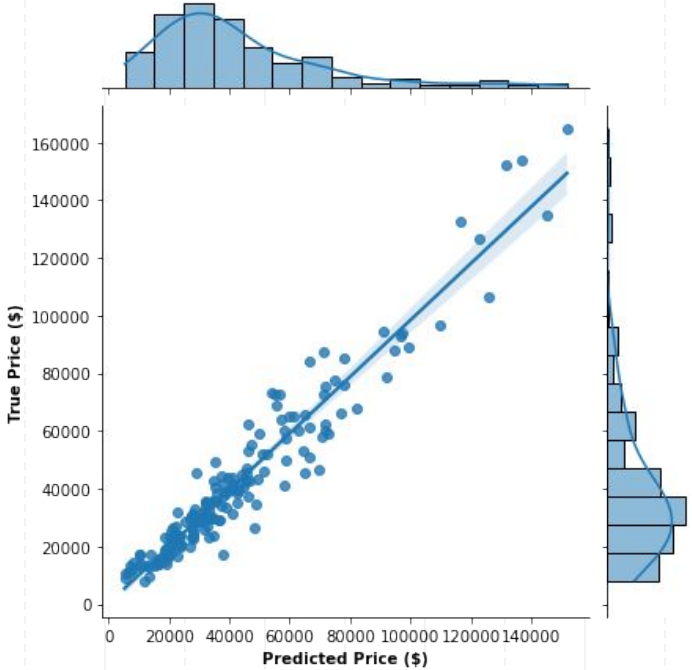
Features	Standardized Coefficients
Horsepower	12532
Torque	8844
Engine Capacity	4306
Acceleration	3926
Year Created	3725



RESULTS

Model Analysis

- **Data right skewed**
  - lower value cars = better prediction



Model Type	# of Features	R^2 Adjusted	MAE
Ridge Regression	75	0.873	$5524

RESULTS

2009 Audi A3 1.9 TDI

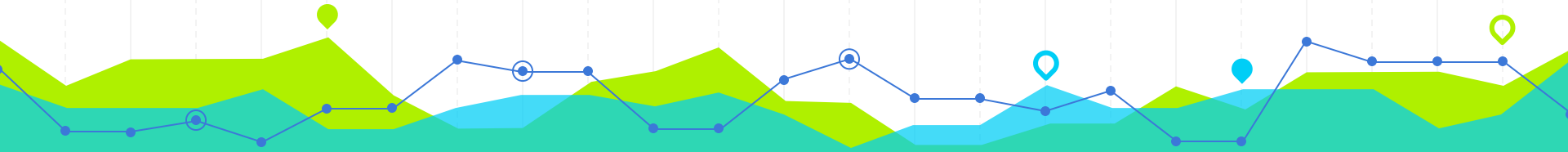
PREDICTED PRICE: $35,085

ACTUAL PRICE: $35,550

2010 BMW ActiveHybrid 7 Series

PREDICTED PRICE: $151,937

ACTUAL PRICE: $164,790



FUTURE WORK

- COLLECT MORE DATA ON HIGHER END CARS
- CATEGORIZE BRANDS INTO BROADER GROUPS
- CREATE A DECISION TREE MODEL

THANKS!

Any questions?