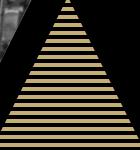


Impact of Car Features on Price & Profitability

SIDDHANT TRIPATHI

15 FEB





Overview



Introduction

- Automotive industry rapidly evolving
- Understanding factors driving demand is crucial

Key Trends

- Growing interest in electric and hybrid
- Varied consumer choices highlight market diversity

Project Scope

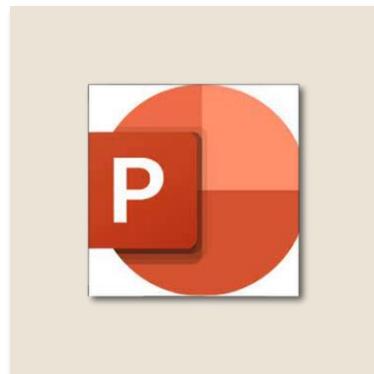
- Analyze impact of features on price
- Optimize for profitability and demand

Tech Stack



MS-Excel 365

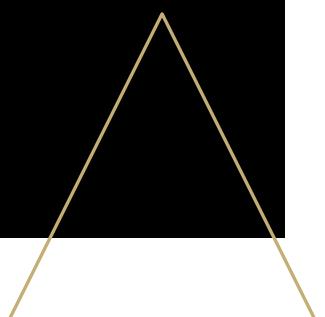
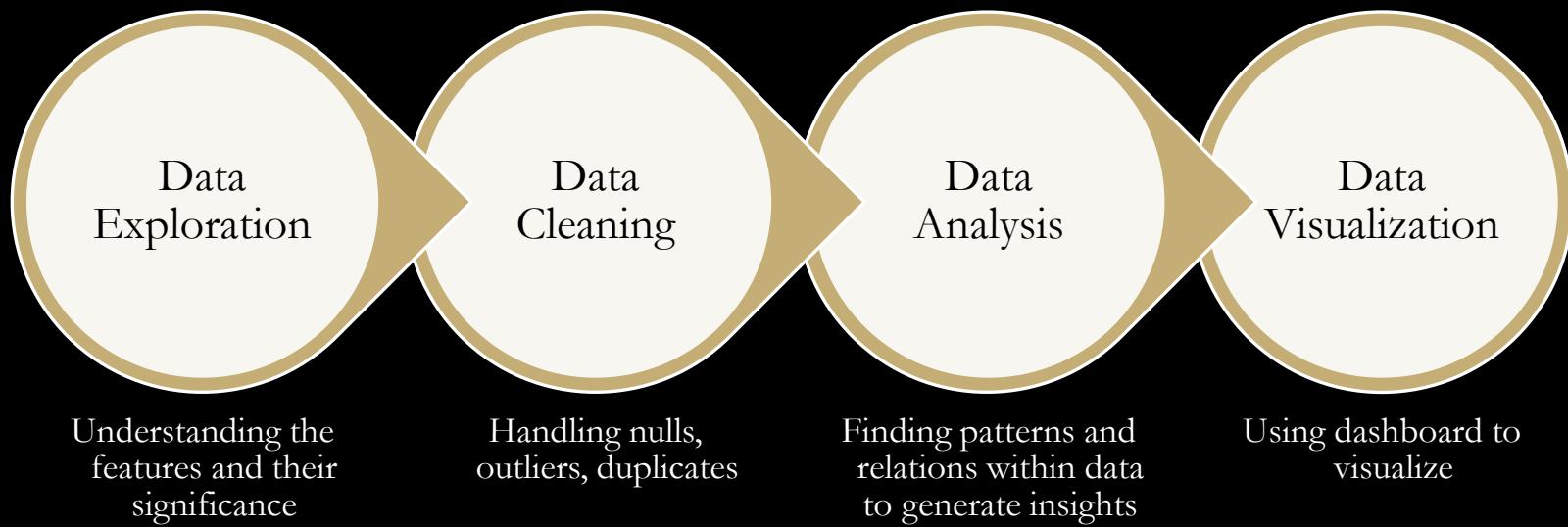
[Analysis File](#)



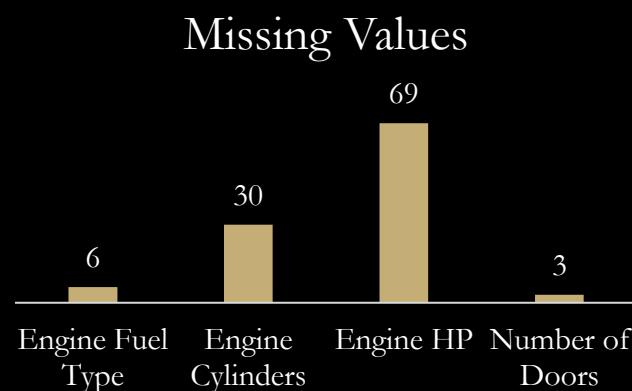
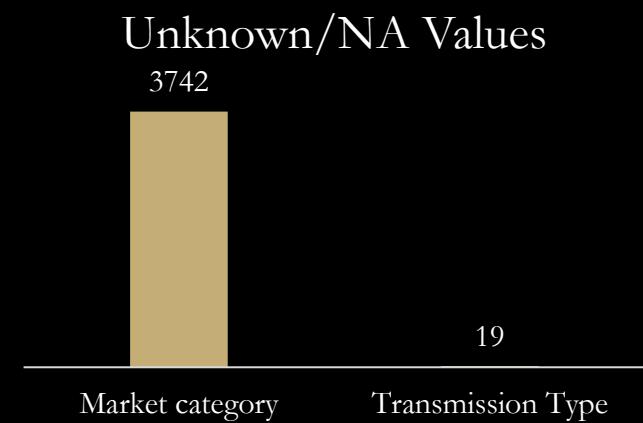
MS-PowerPoint 365



Approach



Data Exploration



Number of Features

16

Number of Records

11,914

Unique Car Brands

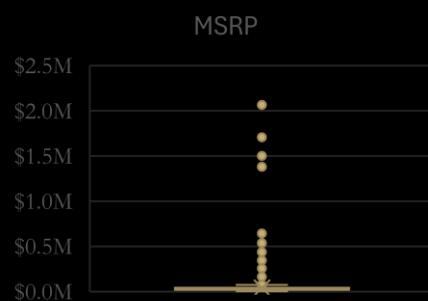
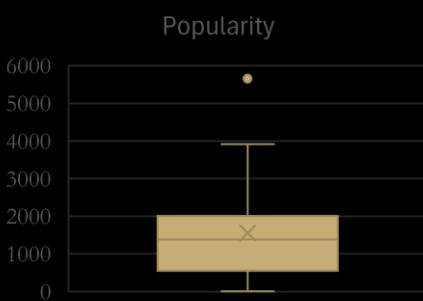
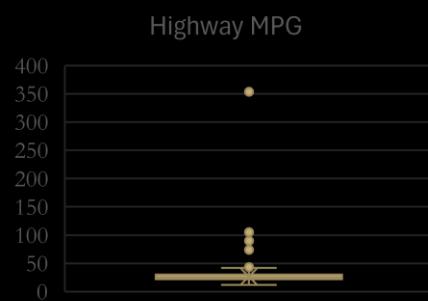
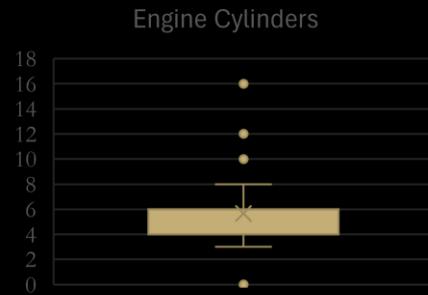
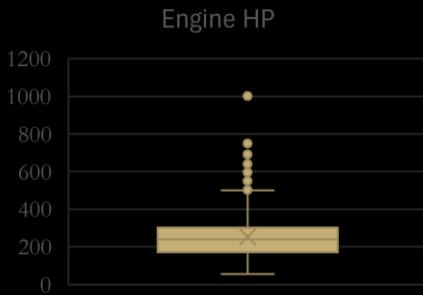
48

Duplicates

715

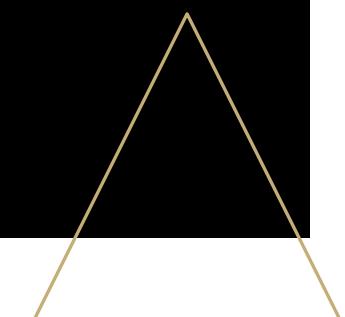


Data Exploration

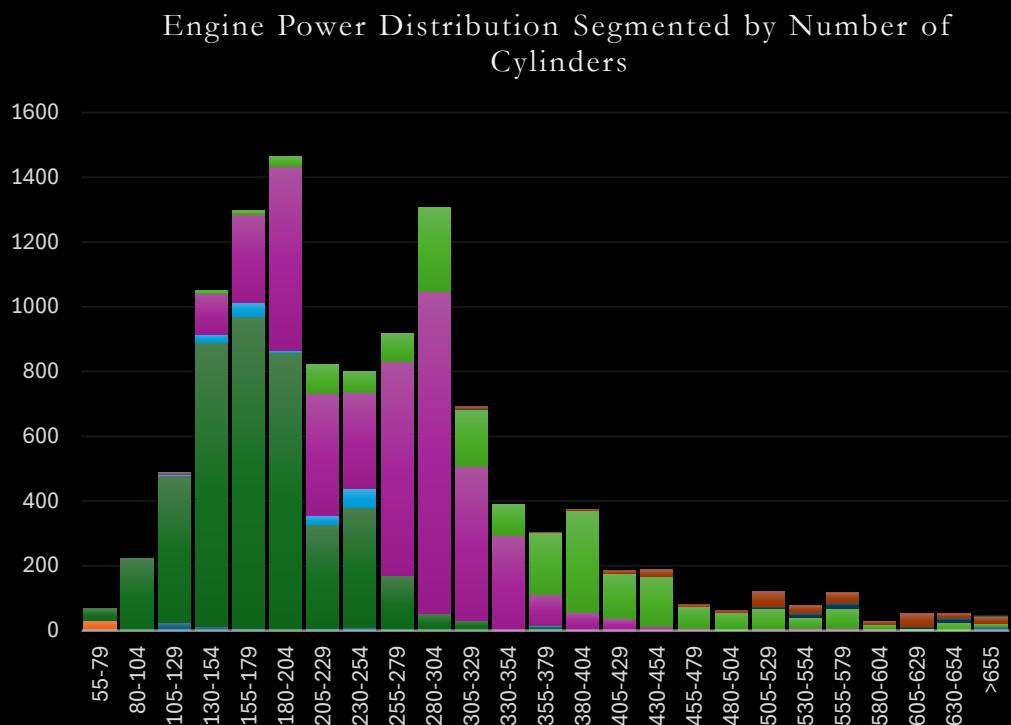


Outliers:

- Engine HP – High-performance cars
- Engine Cylinder – Electric cars & High-performance cars
- City MPG – Hybrid & electric cars
- Highway MPG – 1 Genuine outlier, corrected through online search
- MSRP – High-performance and luxury cars

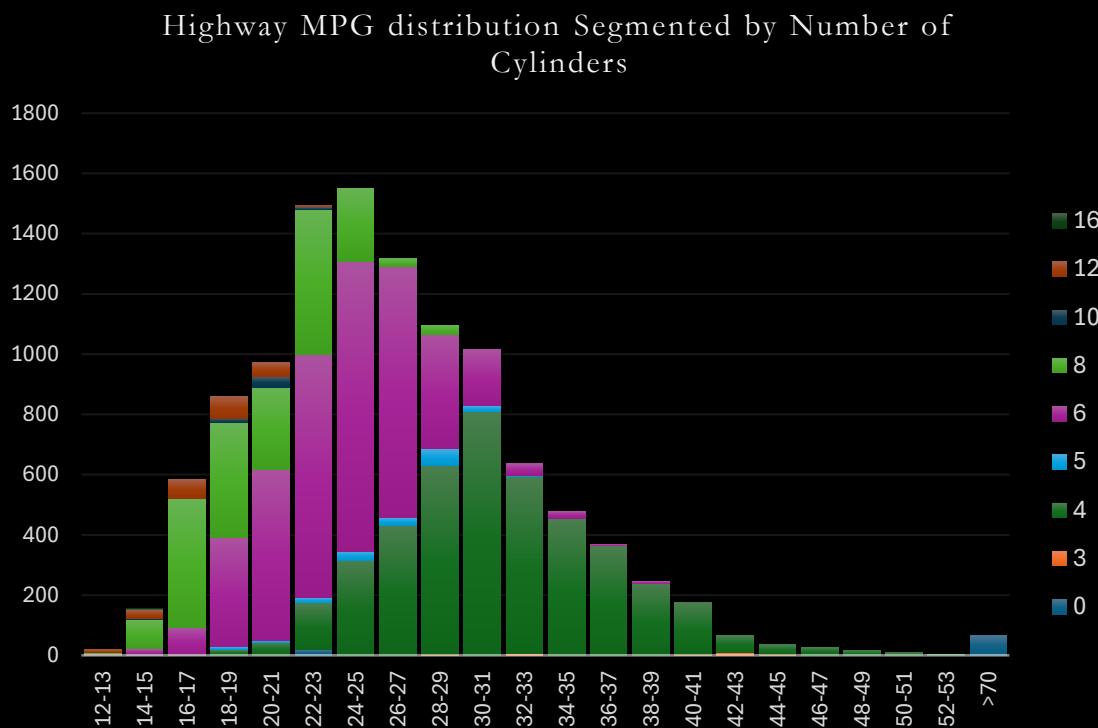


Data Exploration



- Majority cars have 80-380 HP
- Long right tail implies high-performance cars
- Bimodal distribution
- 180-204 HP peak mostly 4 cylinders; 280-304 HP peak mostly 6 cylinders
- Dip between peaks may imply segmentation

Data Exploration



- Majority cars have 14-40 MPG
- Long right tail implies hybrid & electric cars
- Low MPG – 8 cylinders; mid MPG – 6 cylinders; high MPG – 4 cylinders
- >70 MPG only electric cars
- High MPG tail & >70 MPG spike shows rising adoption of green tech

Data Cleaning

- Engine HP missing values – online search
- Engine cylinders missing values – imputed as 0 (electric and rotary engine)
- Unknown transmission corrected by online search
- Number of doors missing values – online search
- Duplicates deleted
- N/A Market Category left as is



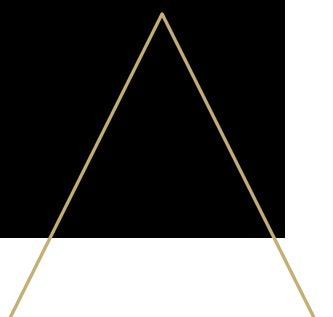
Data Cleaning

Fuel Type Simplification

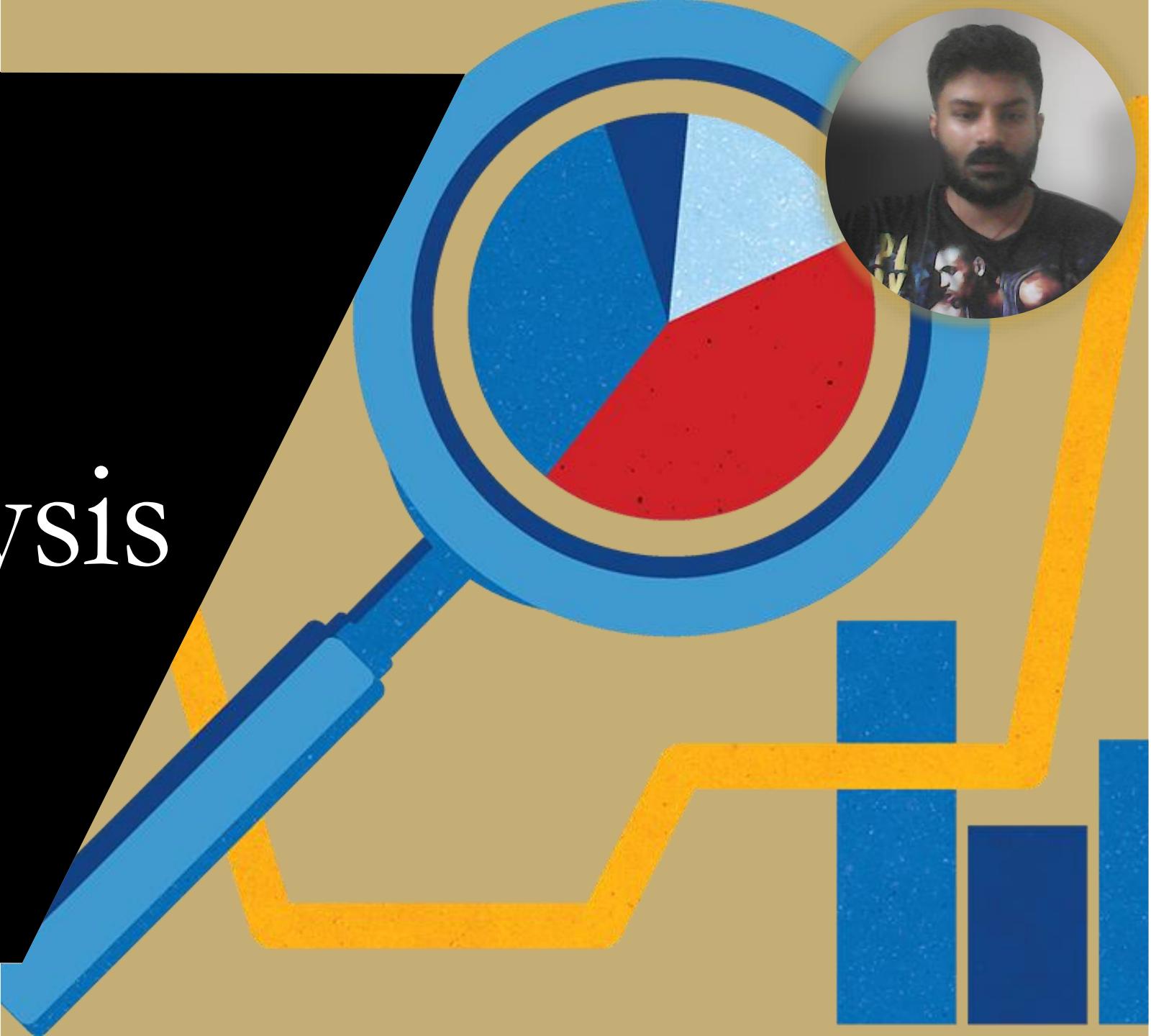


Fuel Type	Simplified Fuel Type
premium unleaded (required), premium unleaded (recommended)	Premium Unleaded
regular unleaded	Regular Unleaded
flex-fuel (unleaded/E85), flex-fuel (premium unleaded recommended/E85), flex-fuel (premium unleaded required/E85), flex-fuel (unleaded/natural gas)	Flex-Fuel
diesel	Diesel
electric	Electric
natural gas	Natural Gas

Reduces complexity & simplifies
encoding for regression analysis

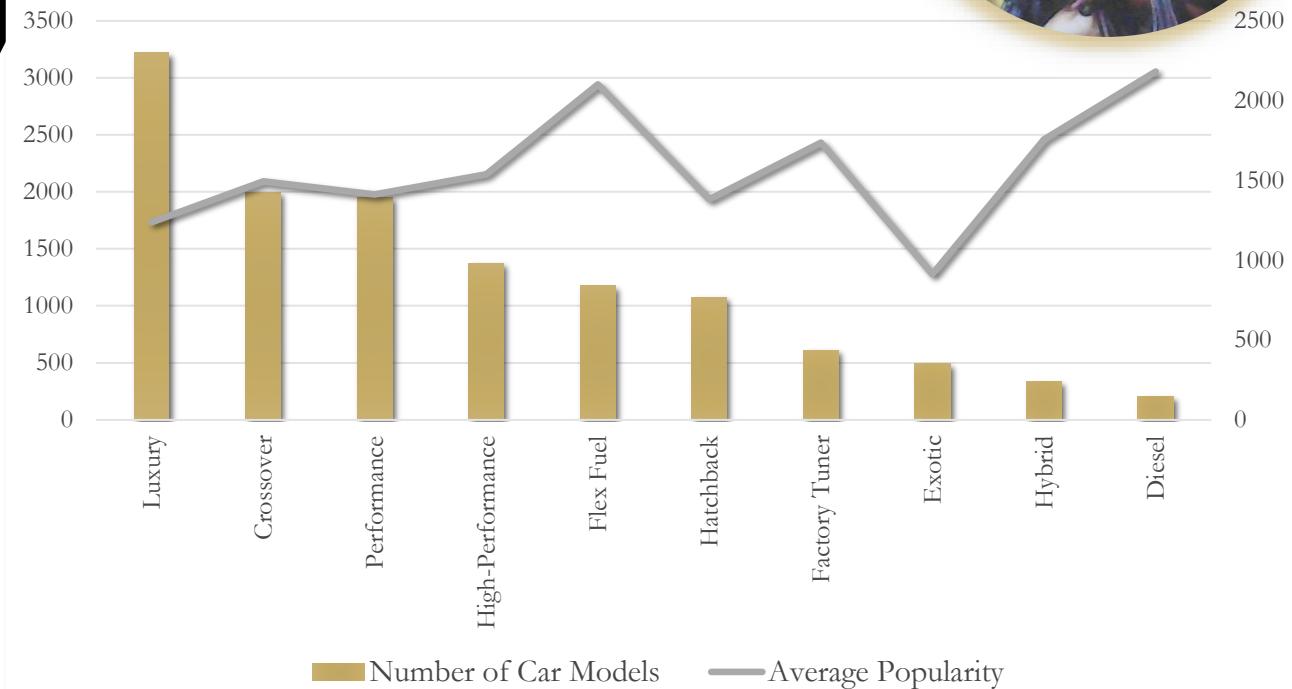


Data Analysis



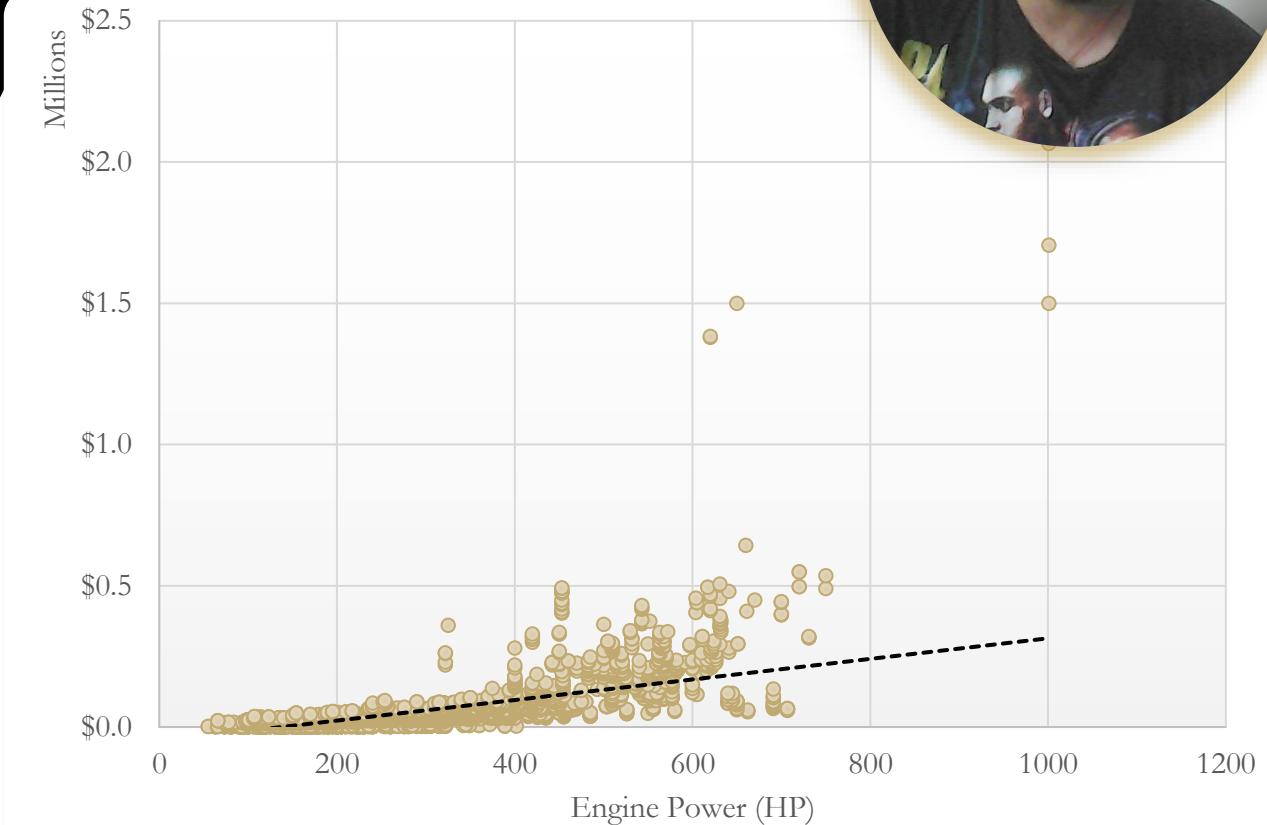
How does the popularity of a car model vary across different market categories?

Market Category and Popularity



Luxury cars have the highest number of models, whereas Flex-fuel and Diesel cars are the most popular

What is the relationship between a car's engine power and price?

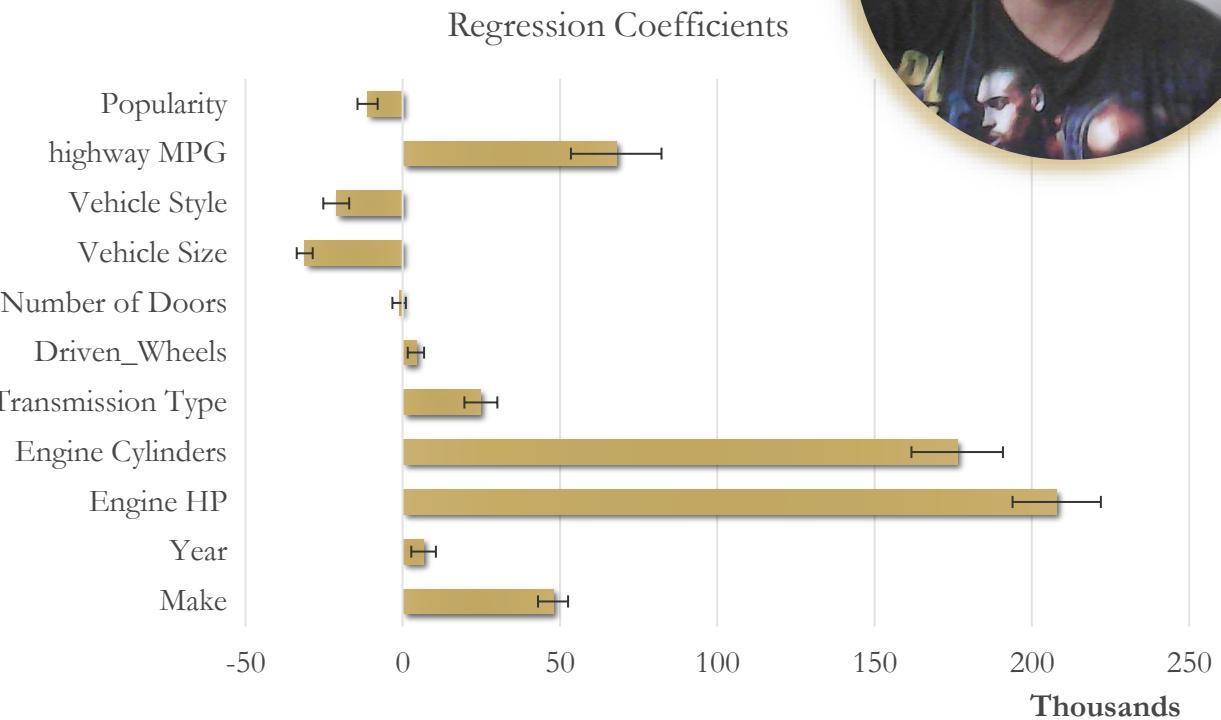


In general, more powerful cars are more expensive – positive correlation

Correlation = 0.656141



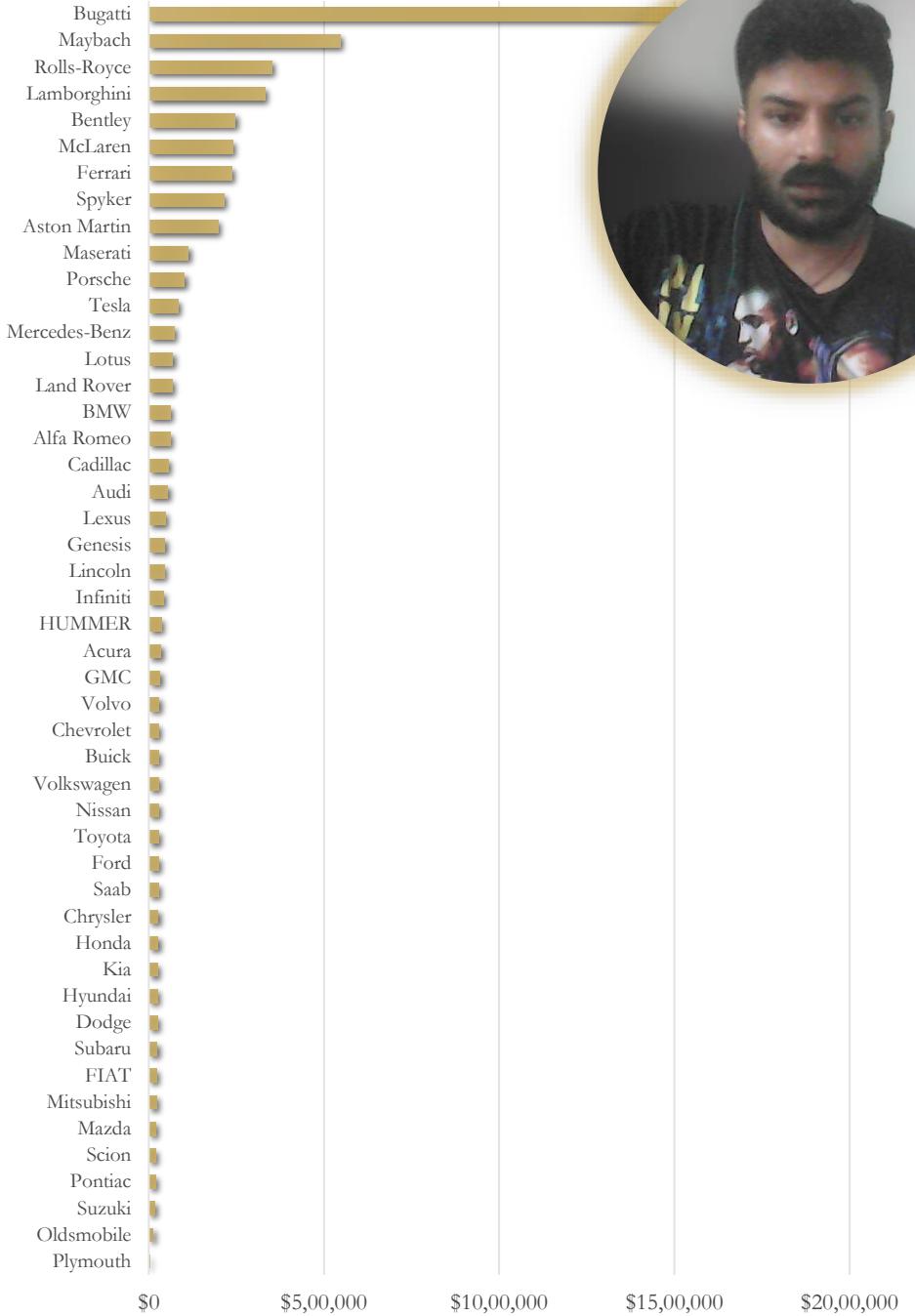
Which car features are most important in determining a car's price?



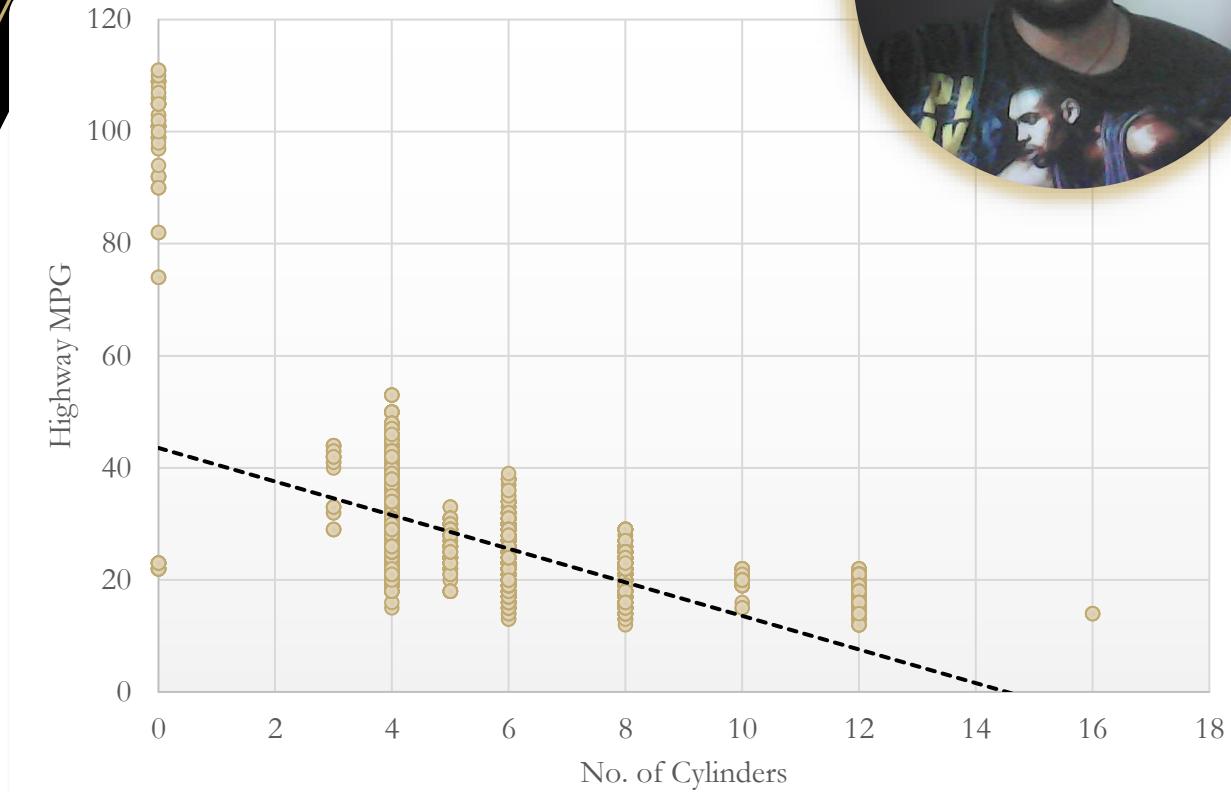
- Engine power and fuel efficiency most important features
- Large Van style cars cheaper
- Branding of car makes it expensive
- More popular cars are priced lower

How does average price of a car vary across different manufacturers?

- Bugatti, Maybach and Rolls-Royce manufacture most expensive cars
- Plymouth, Oldsmobile and Suzuki manufacture the cheapest cars



What is the relationship between fuel efficiency and the number of cylinders in a car's engine?



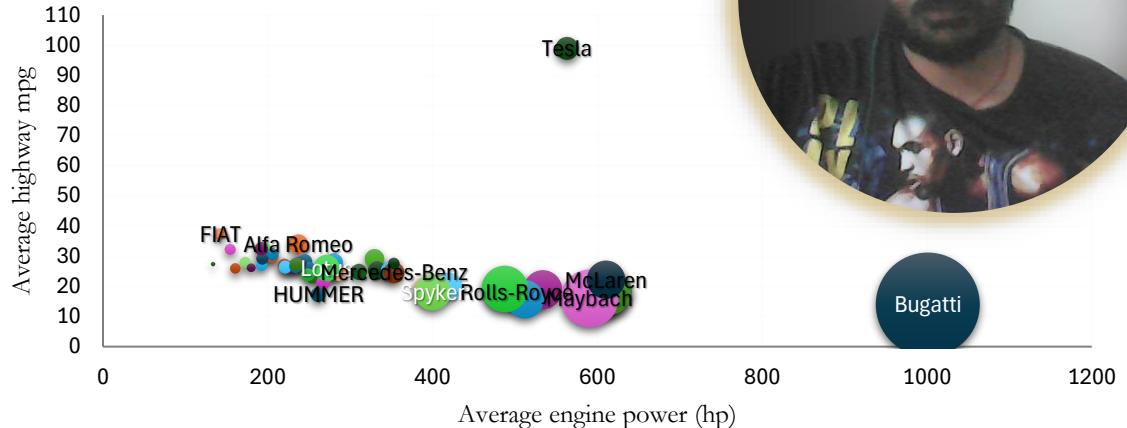
More cylinders decreases fuel efficiency, bigger engines consume more fuel – negative correlation

Correlation = -0.64707

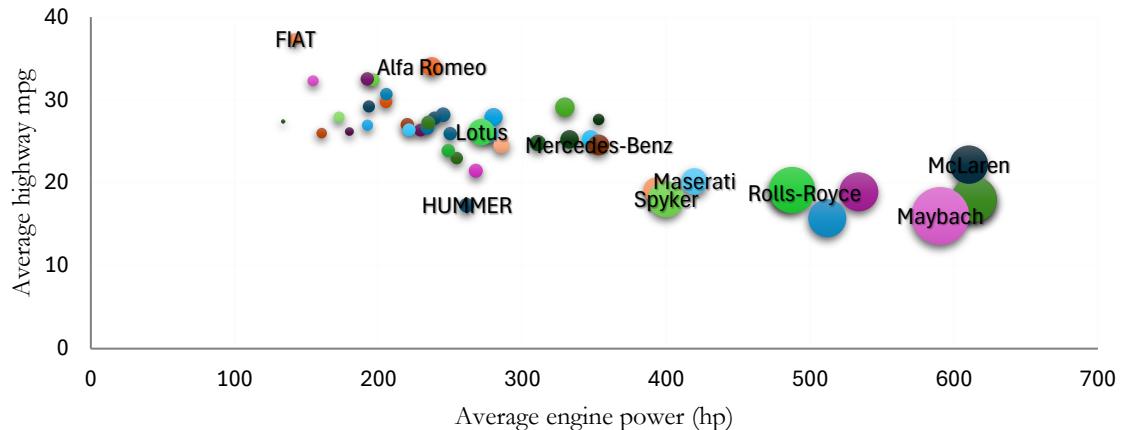


What is the relationship between fuel efficiency, engine power and price of a car across brands?

Engine Power, MPG, and MSRP Relationship Across Brands



Engine Power, MPG, and MSRP Relationship Across Brands



Key Insights



ONE:

Luxury cars have the most models while flex-fuel/diesel are most popular

TWO:

Engine cylinders inversely proportional to MPG, powerful engines increase price

THREE:

Bubble chart segments everyday vs luxury/performance cars

FOUR:

More cylinders increase power but reduce efficiency

FIVE:

Cheaper cars are more popular

SIX:

Recent model year correlates with higher pricing



Summary

- PivotTables and PowerQuery for extensive analysis
- Retaining data integrity despite missing values challenging – imputing data helped
- Dashboards and regression analysis provided pricing insights
- Blended analytical rigor with thoughtful data preparation to derive insights
- Combining statistics with context-aware data cleaning led to an informative modeling exercise