Used Cars EDA Demo

Creative Coders | AI Camp WC2022

December 30th 2022

The Team

Alexandra Ivanova

- Grade 11
- Minimal experience with Python
- Experienced in JAVA

Jayson Kim

- Grade 10
- No experience with Python
- Experienced in C++

Odin Hill

- Grade 9
- No experience with Python
- Minimal experience in JavaScript and C++

Spencer Gilleran

- Grade 11
- No experience with Python
- Accustomed in C# and JAVA

Teddy Schwartz

- Grade 10
- Minimal experience in Python
- Experienced in JavaScript

Haris Hasan

- DSI of this team(Data Scientist Instructor)
- Junior at Purdue University
- Currently studies CS and Statistics

Context

The Goal

- Analyse care resale dataset to find factors influencing the price and how they influence it.

- Power & torque are both dependant on engine speed

- Hypothesis: Power & torque will have a strong correlation between each other

Importance

- Used cars tend to be the only option for many people's transportation.
- For students, our first cars will likely be used cars
- Helps us get the best value possible.

Data Selection and Cleaning

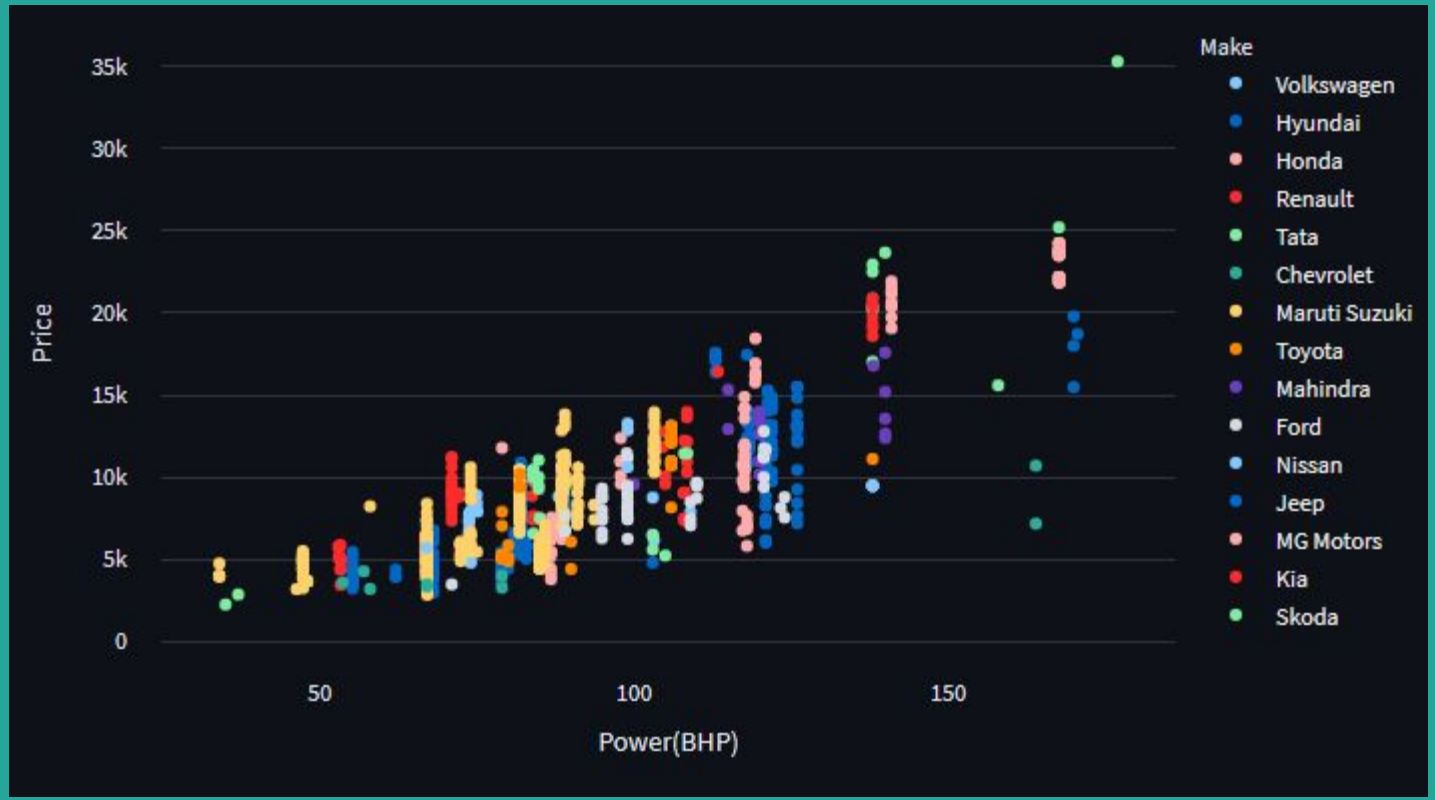
Conversions

What we converted to allow the data to make more sense:

- Price from Indian Rupees to American Dollars
- Mileage from kilometers per liter to miles per gallon
- Fuel tank capacity from liters to gallons
- Number of Owners from ordinal strings (“1st”,”2nd”) to integers(1, 2)

Visualizations

Power(BHP) vs Price

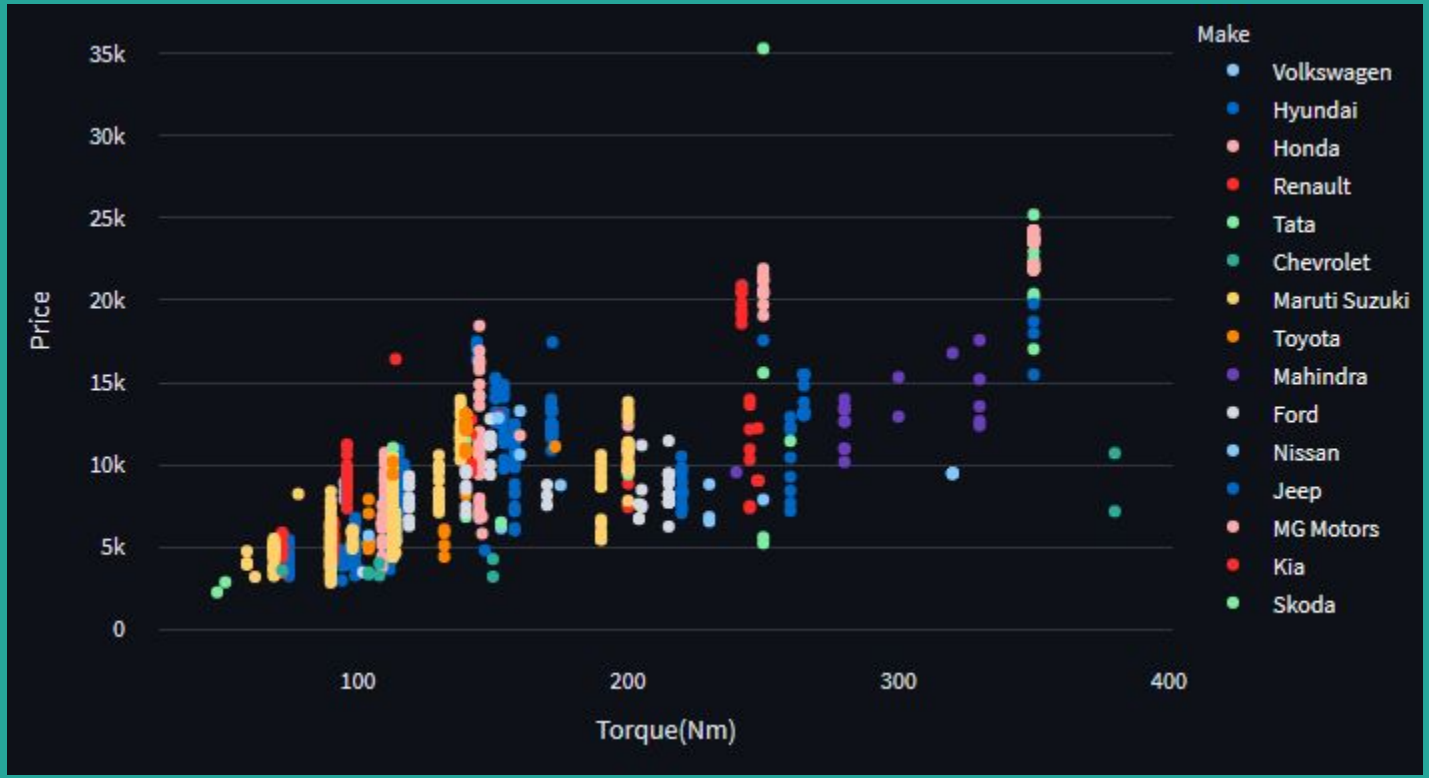


This graph showcases the effect of the power of a car on its price and the positive correlation between the two.

Make Year vs Torque(Nm)

This showcases the fluctuations in torque measured in Nm, over the course of make years.

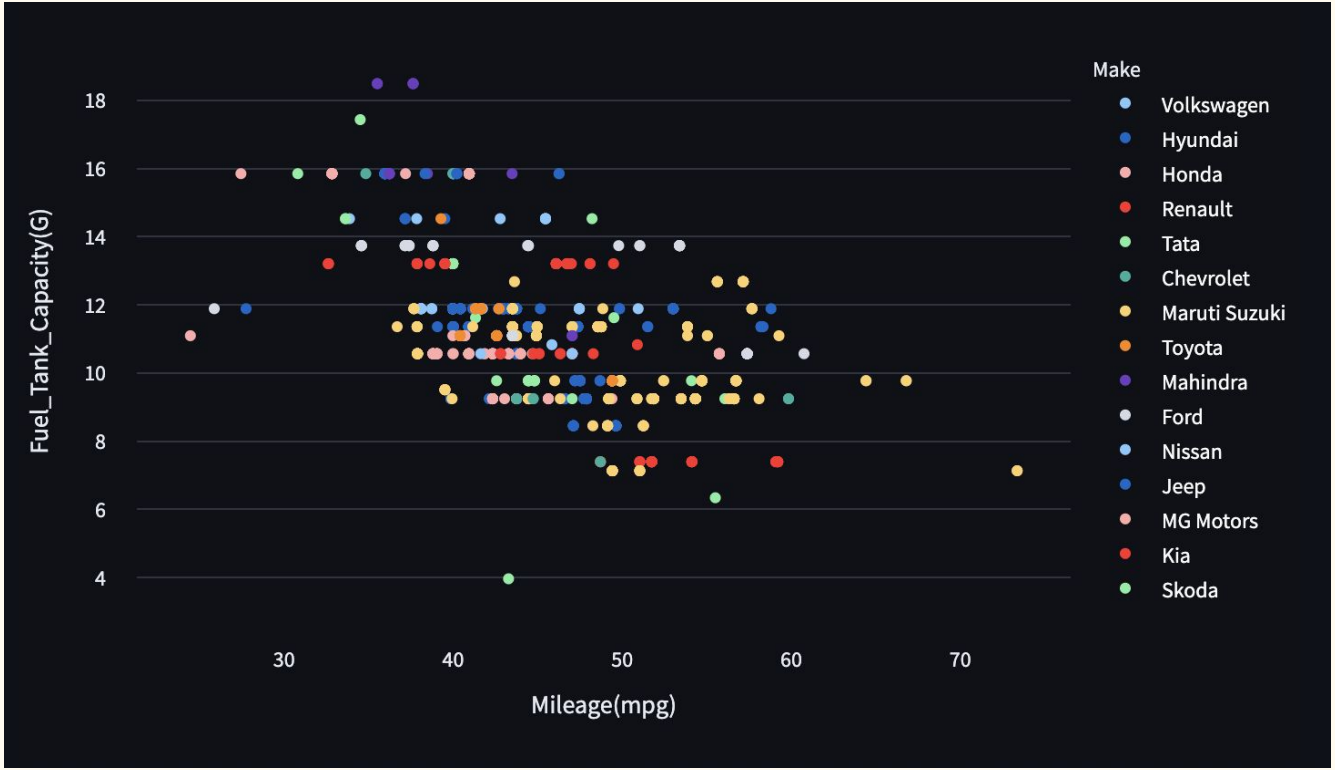
Torque(Nm) vs Price



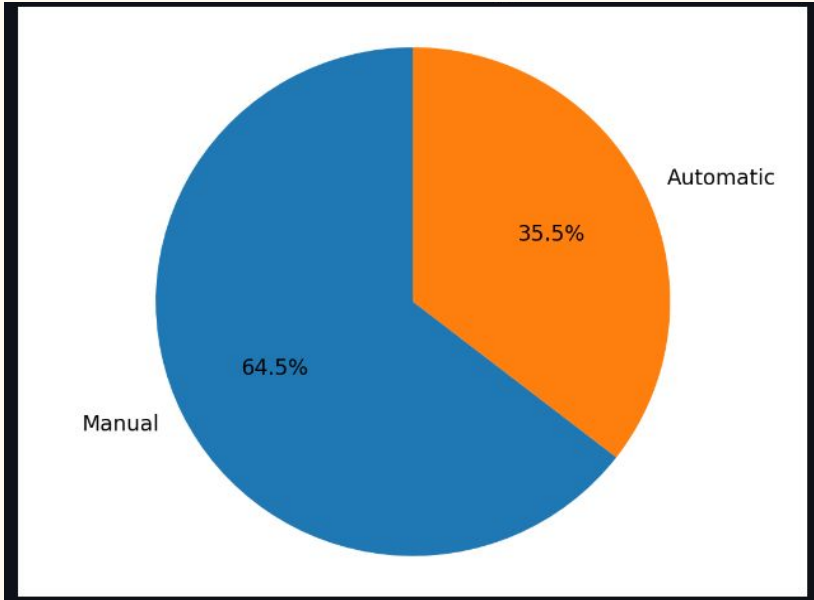
This graph displays the effect of the torque of a car on its price and the positive correlation between the two.

Gas Mileage (mpg) vs Fuel Tank Capacity (gallons)

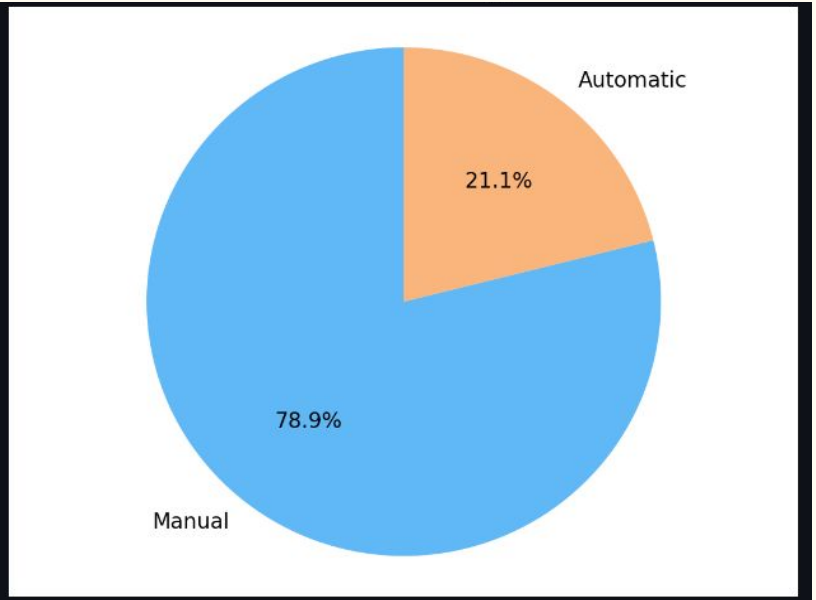
As mileage increases, fuel tank capacity tends to decrease. Most makes tend to have a general fuel tank capacity they stick to.



Top Half of the Car Price Range

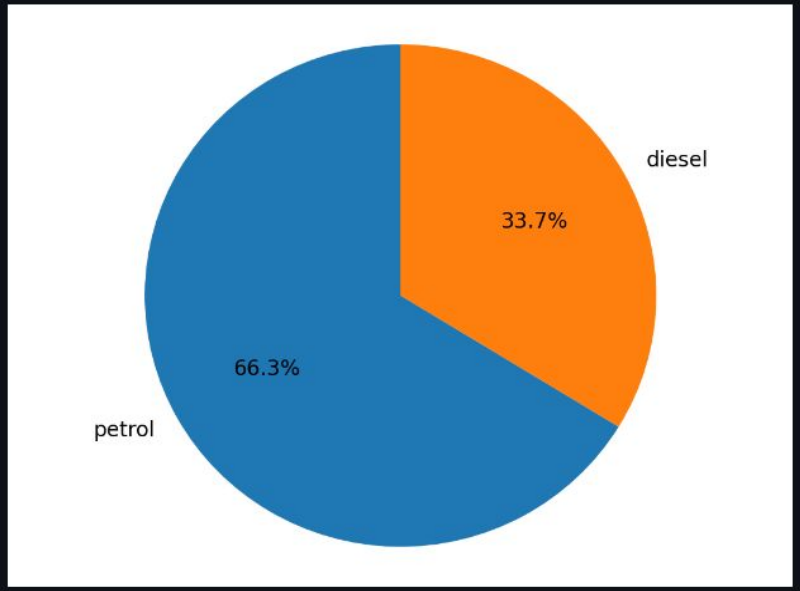


Bottom Half of the Car Price Range

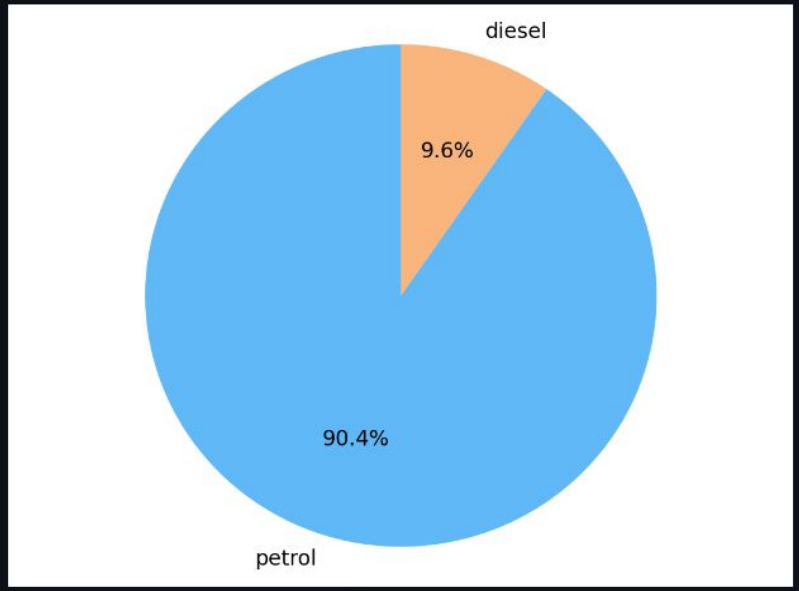


Transmission Type of more expensive and less expensive cars

Top Half of the Car Price Range

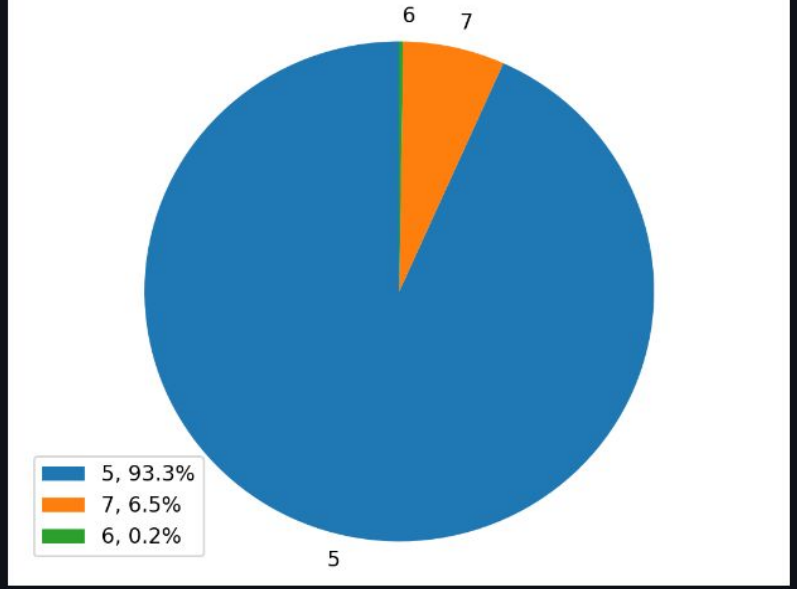


Bottom Half of the Car Price Range

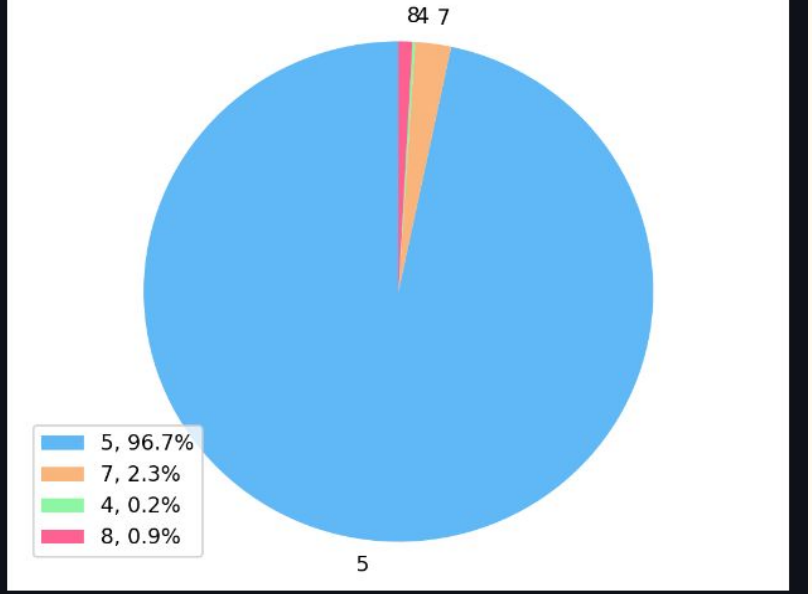


Fuel Type of more and less expensive cars

Top Half of the Car Price Range



Bottom Half of the Car Price Range



Seating Capacity of more and less expensive cars

Insights and Conclusion

Power and Torque Over Time

- Chevrolet shows rapid growth in year range of 2013-2015 and 2022
- Maruti shows almost no improvements in power
- The highest power & torque (250 Nm and 177 BHP) was achieved by the Skoda 2012, 2014, 2019, and 2022
    a. Chevrolet had highest 2013, 2015
    b. Jeep, Hyundai, Honda, Toyota, MG Motors, Tata are highest other years
- 2011-2022 Power nor Torque has significantly increased

Price Insights

- Almost every car valued at 13k or more, has only one previous owner
- Every car valued at over 20k, was made from 2019 to 2022
- Higher power and higher torque, tend to raise the price
- More expensive cars tend to use automatic transmission and diesel rather than less expensive ones

Attributes of Most and Least Expensive Cars Sold

- Most expensive car sold:
  a. Make: Skoda
     i. Power: 250 BHP
     ii. Torque: 177 Nm
     iii. Price: $35,292
- Least expensive car sold:
  a. Make: Tata
     i. Power: 35 BHP
     ii. Torque: 48 Nm
     iii. Price: $2,256

Takeaways

What We Learned:

Alexandra Ivanova

- The basics of Python
- Coding in a group fully online
- What's expected when analyzing data (goal, cleaning, visualizing, insights)
- Various sites for AI Art or Datasets

Jayson Kim

- Learn different modules used for data science
- How to properly analyze data.
- Learning how to work together in coding
- How to clean data and use scatter plots/pie charts to visualize data for analysis

Odin Hill

- How to use modules to clean and analyze data.
- Other Python basics
- How to work with others on code.

Spencer Gilleran

- How to gather good data sets
- How to visualize data using python
- How to work in a group setting with code

Teddy Schwartz

- How to work with modules in Python (Pandas, Numpy, Plotly, Streamlit, Pyplot)
- How to code in groups
- What an EDA is

Demo Time

Click here to go to WebApp