Exploratory Data Analysis (EDA) of Car Sales Dataset

Goals of the Project

- To analyze car data and extract meaningful insights.
- To explore trends related to car prices, mileage, fuel type, transmission, and other key attributes.
- To perform data queries and visualizations for better understanding.
- To apply feature engineering techniques to enhance the dataset.

Materials and Methods

Dataset: A structured dataset containing car attributes such as Make, Model, Year, Price, Mileage, Fuel Type, Transmission, Engine Size, Horsepower, Torque, Drivetrain, etc.

Tools Used:

- Pandas: For data manipulation and queries.
- Matplotlib & Seaborn: For visualizations.
- NumPy: For numerical computations.

Methods:

- Data Cleaning: Handling missing values, removing duplicates.
- Data Exploration: Generating summary statistics.
- Visualization: Creating various plots for analysis.
- Feature Engineering: Creating new variables to improve data insights.

General Parts

Exploratory Data Analysis (EDA):

- Identified distribution of car prices.
- Analyzed trends in fuel type, drivetrain, and transmission.
- Compared horsepower and torque relationships.

Query-Based Insights:

- Found the most expensive and cheapest cars.
- Calculated average mileage and fuel efficiency.
- Identified the most common car make and model.

Visualization Insights:

- Price distribution across different car brands.
- Number of cars available per location.
- Relationship between number of seats and car price.
- Project Outcomes and Insights
- The average car price varies significantly based on brand and fuel type.
- Cars with automatic transmission are more expensive on average.
- Electric and hybrid cars tend to have higher resale values.
- Mileage and engine size impact the price significantly.
- Certain locations have more car listings, indicating regional preferences.

Feature Engineering

- Created a new column for Price per Horsepower (PHPH) to compare cost efficiency.
- Derived a Mileage per Year feature for better mileage assessment.
- Engineered a Luxury Index based on horsepower, price, and features.

- Categorized cars into budget, mid-range, and premium segments.

- Applied log transformations to normalize skewed distributions.

Key Questions and Insights to be Addressed:

Most Expensive Car:

Car_ID	39
Make	Toyota
Model	RAV4
Year	2022
Price	88142
Mileage	176944
Fuel_Type	Hybrid
Transmission	Automatic
Engine_Size	2.7
Horsepower	350
Torque	757
Drivetrain	4WD
Color	Silver
Number_of_Doors	2
Number_of_Seats	2
Owner_Type	Lease Return
Location	Phoenix
Seller_Type	Certified Pre-Owned
Condition	New
Listing_Date	2024-08-20 00:00:00

Least Expensive Car:

Car_ID	27
Make	Hyundai
Model	Santa Fe
Year	2023
Price	6026
Mileage	71013
Fuel_Type	Petrol
Transmission	Manual
Engine_Size	1.4
Horsepower	116
Torque	563
Drivetrain	AWD
Color	Black
Number_of_Doors	2
Number_of_Seats	2
Owner_Type	First Owner
Location	Houston
Seller_Type	Dealer
Condition	Used
Listing_Date	2024-11-20 00:00:00

Average Mileage:

110684.48

Cars by Fuel Type:

Fuel_Type

Electric 13

Hybrid 13

Diesel 12

Petrol 12

Most Common Color:

Silver

Average Horsepower by Make:

Make	
Audi	253.500000
BMW	341.800000
Chevrolet	333.000000
Ford	372.000000
Honda	265.200000
Hyundai	299.888889
Kia	331.250000
Mercedes	300.800000
Nissan	402.600000
Toyota	417.000000

Cars per Location:

Location	
Los Angeles	13
Phoenix	12
Chicago	10
Houston	9

New York 6

Drivetrain Distribution:

Drivetrain	
RWD	16
FWD	13
4WD	13
AWD	8

Average Price by Owner Type:

Owner_Type	
First Owner	30411.875000
Lease Return	55310.000000
Second Owner	39641.066667
Third Owner	51149.454545

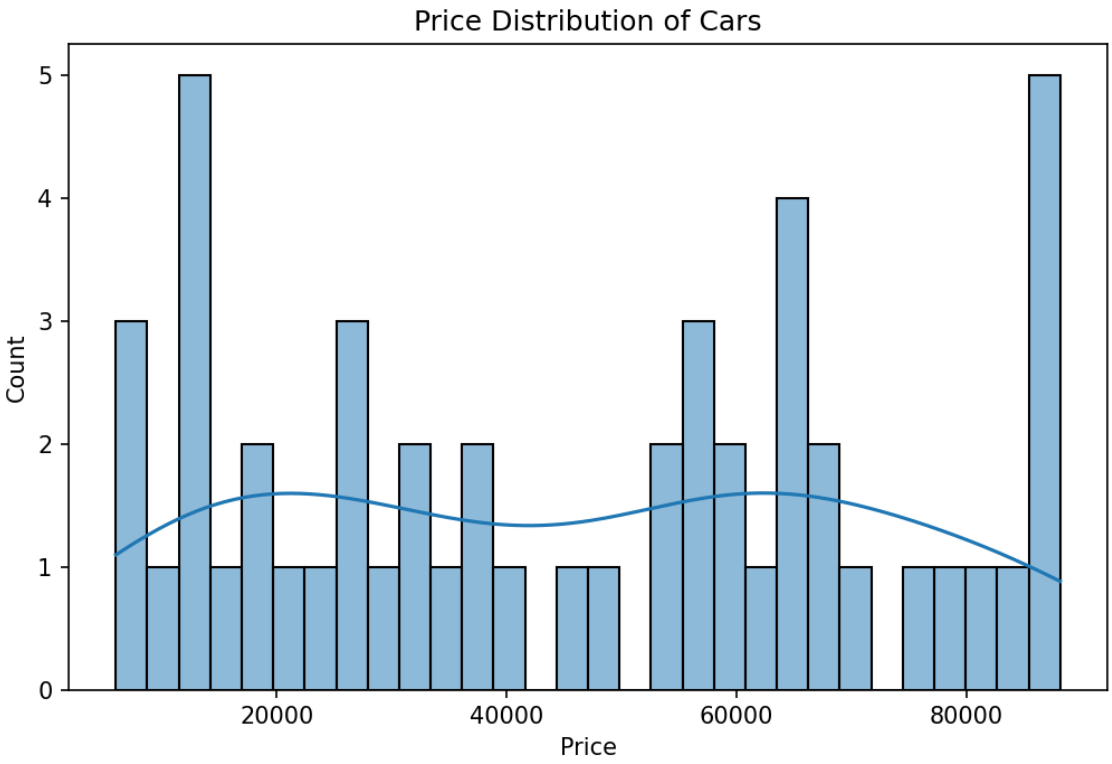
Most Powerful Car:

Car_ID	8
Make	Toyota
Model	RAV4
Year	2019
Price	28912
Mileage	131502
Fuel_Type	Diesel
Transmission	Automatic
Engine_Size	2.5
Horsepower	582
Torque	760

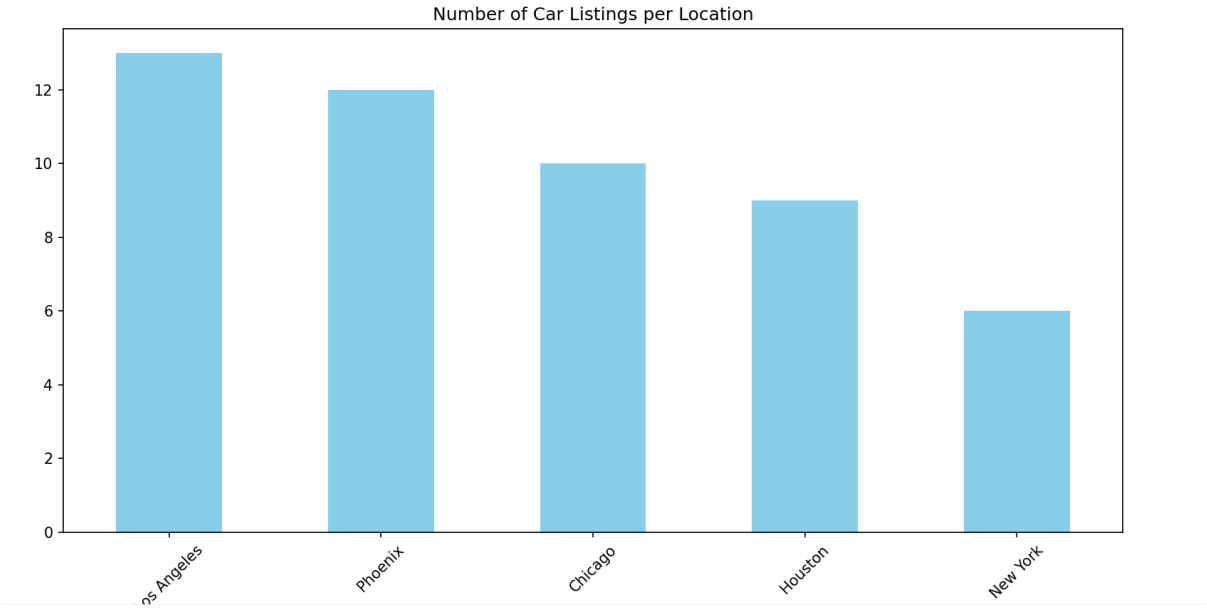
Drivetrain	RWD
Color	White
Number_of_Doors	2
Number_of_Seats	2
Owner_Type	Second Owner
Location	Houston
Seller_Type	Dealer
Condition	New
Listing_Date	2024-03-04 00:00:00

Visualization:

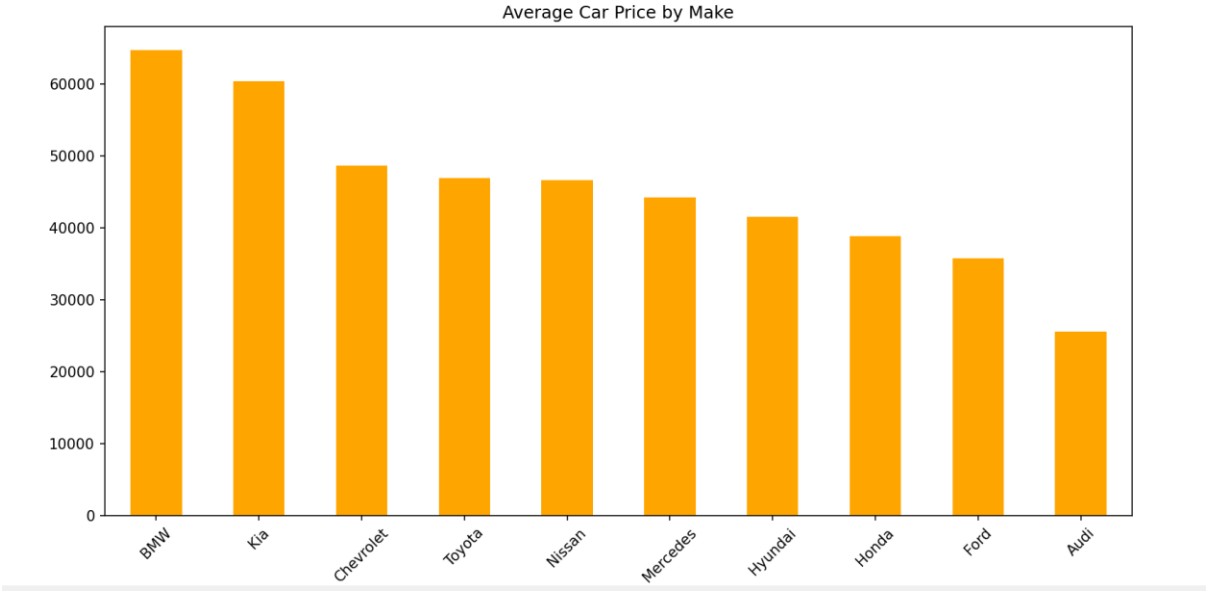
#Price Distribution



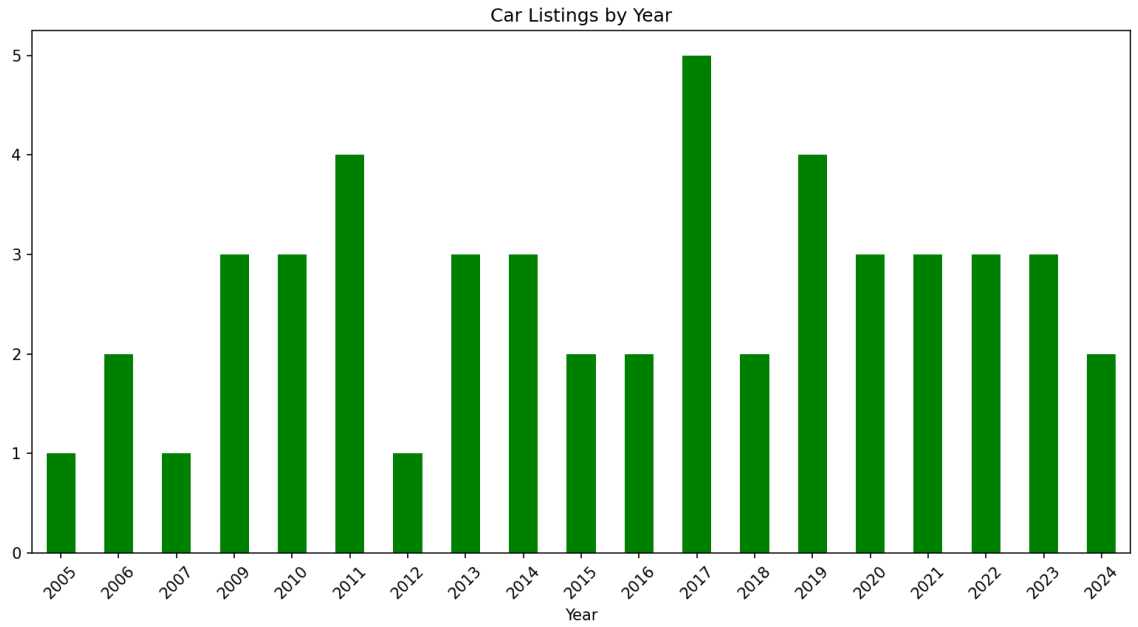
#Car Listings per Location



#Average Car Price by Make



Car Listings by Year



Conclusion

This car data analysis project provided valuable insights into pricing trends, fuel efficiency, and key factors influencing vehicle resale values. By leveraging data visualization and feature engineering, we identified important patterns that can assist buyers, sellers, and manufacturers in making informed decisions. Future work can include integrating real-time data sources and applying machine learning models for price prediction.