Project Milestone-2

Gunadheep Sakthivel

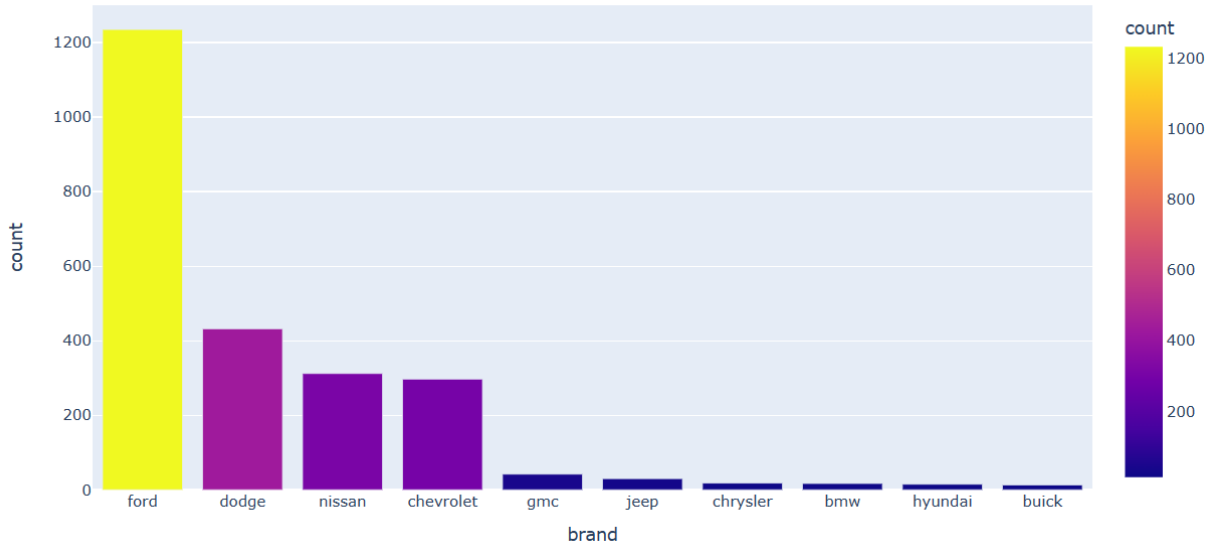
Introduction:

To look at and understand various angles of the US auto market we used this milestone to go deeper into the dataset with data visualization tools. This is why we decided to answer these important questions, for instance: Brand preferences, price distribution, regional sales breakdown, and how vehicle condition is related to pricing, by using tools such as Matplotlib, Seaborn and Plotly.

Interactive Analysis and Visualization

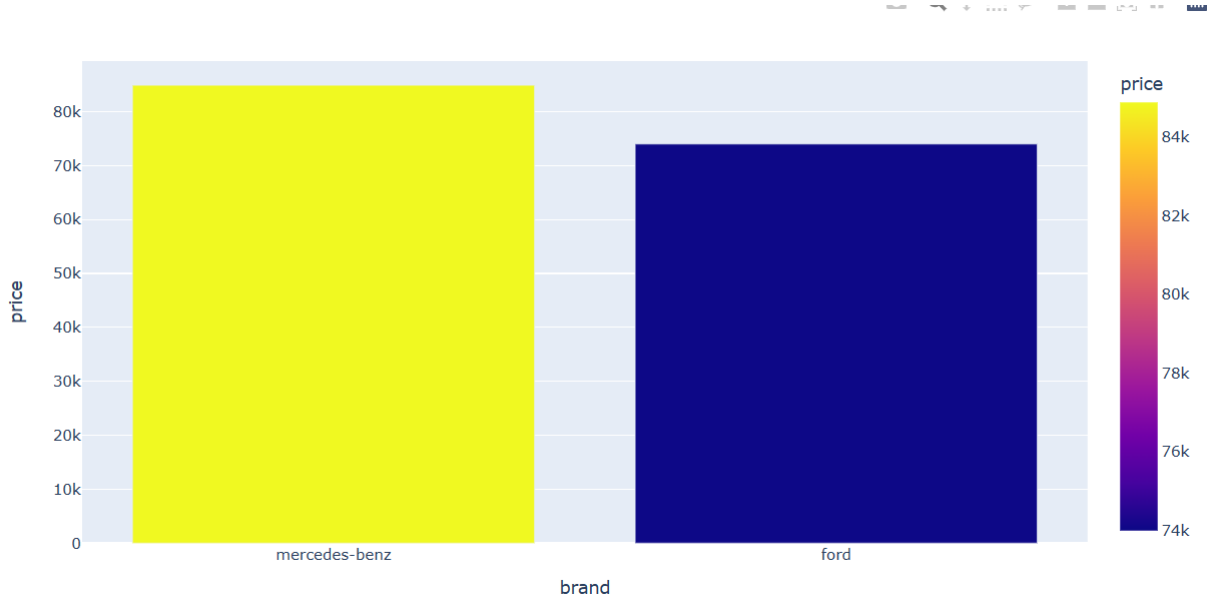
1. Brand Preference

We used a bar chart to display the top 10 brands with the most models after classifying the data by brand. The dataset is dominated by vehicles from major brands such as Ford, Chevrolet, and Dodge, indicating their dominance in the US auto market.



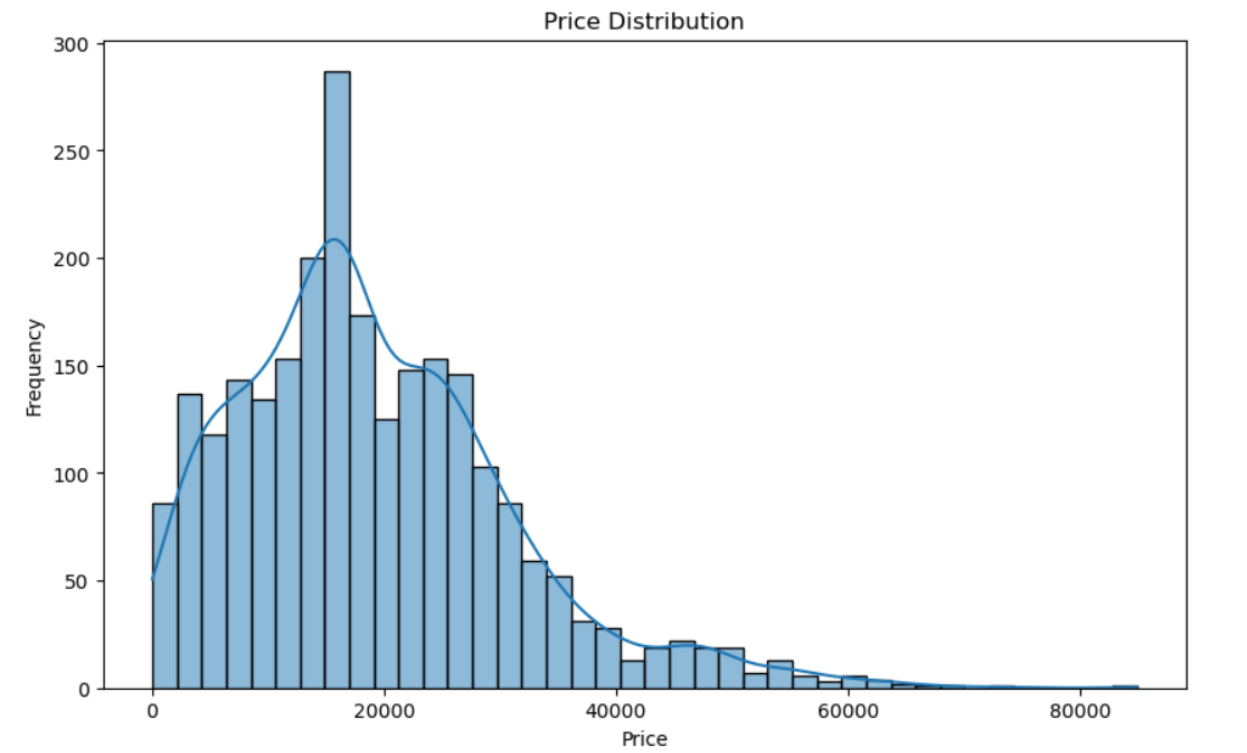
2. Pricey Automobiles

We used a bar chart to display the brands and prices of the top two most costly cars in the dataset. The graphics demonstrated the dominance of the luxury market category, with cars from luxury companies like Ford and Chevrolet costing a premium.



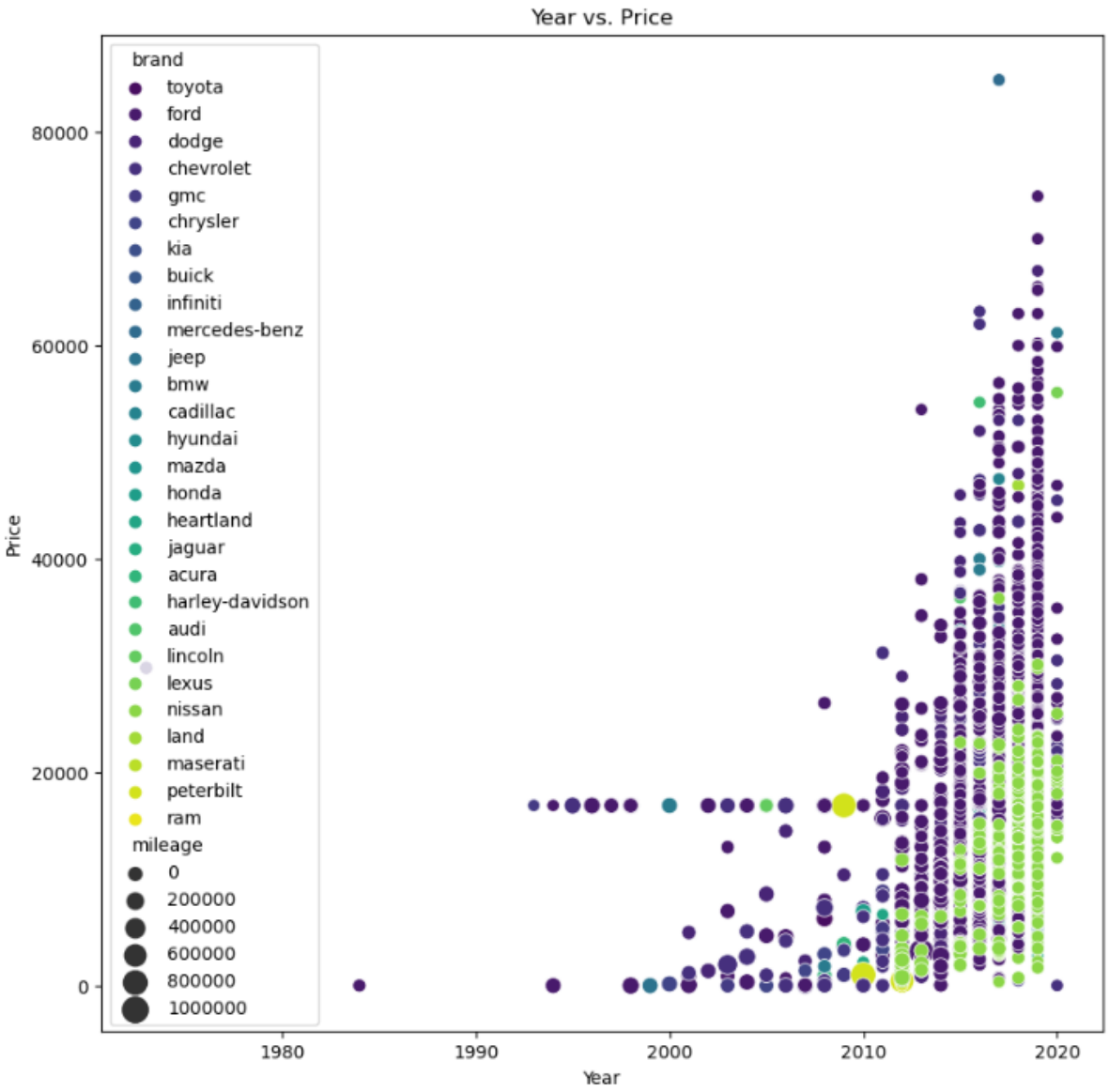
3. Distribution of Prices

The dataset's vehicle price distribution was displayed using a histogram. A small number of outliers in the dataset reflect expensive models, whereas most of the cars in the dataset are priced within a specific range.



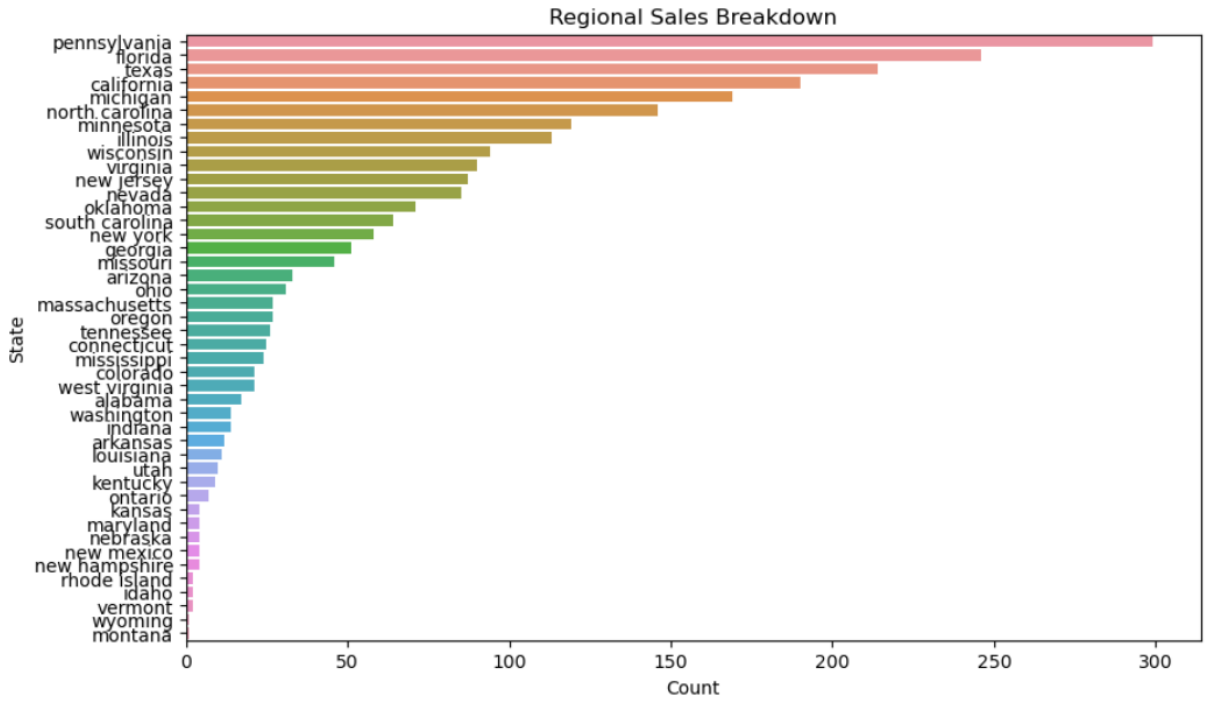
4. Year versus Cost

The association between the year of registration and the cost of vehicles was represented visually using a scatter plot. Prices of modern cars are generally higher, which is indicative of the depreciation factor that affects older models.



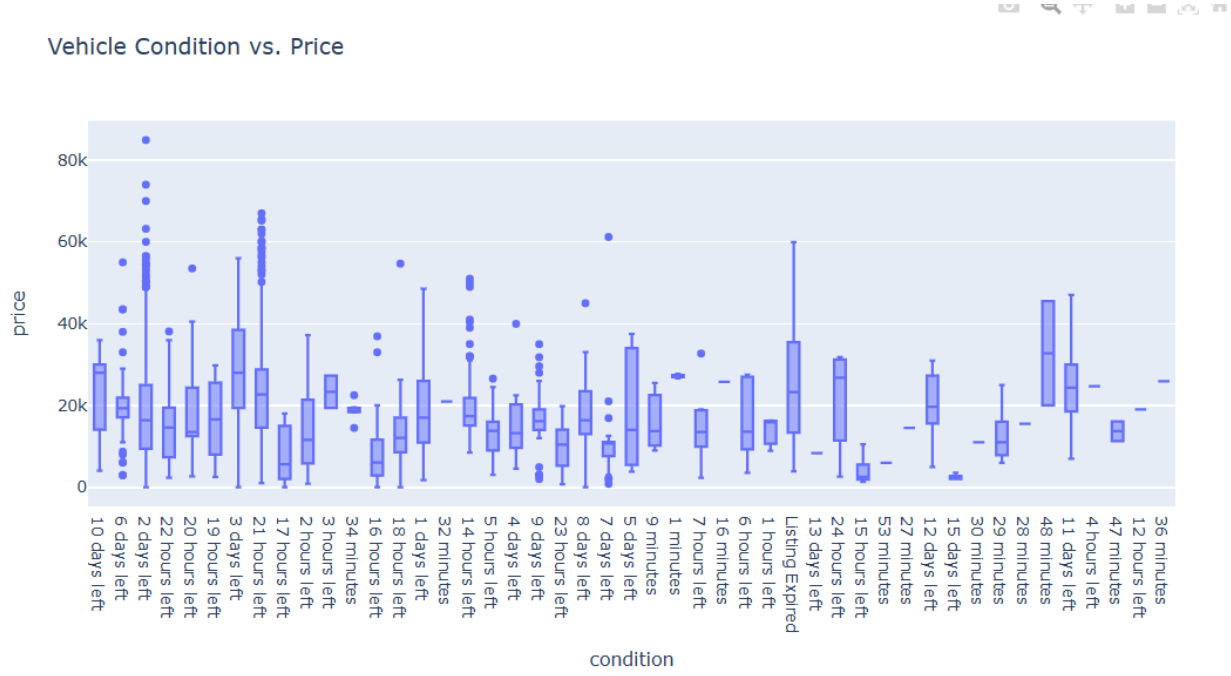
5. Sales by Region Breakdown

The number of cars sold in each state was displayed using a bar chart, which also highlighted regional sales patterns. States with higher car sales than others include Florida, Georgia, and New Jersey, which may indicate preferences among certain regional markets.



6. Car Price vs. Condition

The distribution of pricing according to the car condition (clean title vs. salvage title) was compared using a box plot. Cars with a clean title typically cost more than cars with a salvage title, highlighting how crucial car condition is to pricing.



Self-assessment:
Areas of Success

Interactive Visualizations: Through a joined usage of Plotly and other interactive tools, we managed to produce an exploration dynamic which considerably raised the visual appeal level of our plots and significantly increased the interactivity.
Perceptive Analysis: With the help of our graphics, we established clear correlations in regards to the subject preferences, pricing trends, and regional sales patterns thus answering the question about the goals of our analysis.
Areas for Improvement

Additional Exploration: We may not have established or showcased the most well-versed visualizations that the dataset actually have, but a more detailed investigative review offers a better comprehension of the relationship between the variables and the patterns in the dataset.
Improved Interactivity: It would be very effective if the given visualizations were turned into more interactive elements, which would surely grab viewers' attention and lead them deeper into our dataset.

Conclusion:

With this milestone, we succeeded in producing infographics of benefit to the US auto market, which truly identify multiple angles of this market. Our visualizations shed light on the pricing dynamics, the revving brand preferences, and geographical sales patterns, which are some of the key findings. Going forward we want to improve our visualization techniques and experiment with more complex analytics so our understanding of the market and US automotive dataset will become grounded.