

# Dashboard Drive: Mapping Out the US Car Market with Interactive Visualizations

## Introduction:

In the ever-changing US automotive business, knowing market dynamics, consumer preferences, and regional variances is critical. Our initiative, "Dashboard Drive," aims to explore the complexities of the US car market using interactive visualizations based on a vast dataset obtained from credible platforms. Using innovative data visualization tools, we want to uncover trends, preferences, and correlations in the automotive scene, empowering stakeholders with actionable information.

## Methodology:

### Data Pre-Processing:

Our road to understanding the complexities of the US automobile market begins with thorough data pre-processing, which ensures that the dataset is ready for intelligent analysis. Key phases in our data preprocessing procedure include:

**Data Loading:** We begin by loading the dataset into a Pandas Data Frame using the `read_csv` method. This process helps us to smoothly convert the raw data to a structured representation appropriate for analysis.

```
car_data=pd.read_csv("USA_cars_datasets.csv")
car_data.head()
```

	Unnamed: 0	price	brand	model	year	title_status	mileage	color	vin	lot	state	country	condition
0	0	6300	toyota	cruiser	2008	clean vehicle	274117.0	black	jtezu11f88k007763	159348797	new jersey	usa	10 days left
1	1	2899	ford	se	2011	clean vehicle	190552.0	silver	2fmdk3gc4bbb02217	166951262	tennessee	usa	6 days left
2	2	5350	dodge	mpv	2018	clean vehicle	39590.0	silver	3c4pdogg5jt346413	167655728	georgia	usa	2 days left
3	3	25000	ford	door	2014	clean vehicle	64146.0	blue	1ftfw1et4efc23745	167753855	virginia	usa	22 hours left
4	4	27700	chevrolet	1500	2018	clean vehicle	6654.0	red	3gcporec2jg473991	167763266	florida	usa	22 hours left

**Removing Unnecessary Columns:** During the initial exploration, we detect and eliminate useless columns, such as the 'Unnamed: 0' column, which acts as an index but has no bearing on our study.

```
In [3]: car_data = car_data.drop(columns = ['Unnamed: 0'])
```

```
In [4]: car_data.head()
```

Out[4]:

	price	brand	model	year	title_status	mileage	color	vin	lot	state	country	condition
0	6300	toyota	cruiser	2008	clean vehicle	274117.0	black	jtezu11f88k007763	159348797	new jersey	usa	10 days left
1	2899	ford	se	2011	clean vehicle	190552.0	silver	2fmdk3gc4bbb02217	166951262	tennessee	usa	6 days left
2	5350	dodge	mpv	2018	clean vehicle	39590.0	silver	3c4pdogg5jt346413	167655728	georgia	usa	2 days left
3	25000	ford	door	2014	clean vehicle	64146.0	blue	1ftfw1et4efc23745	167753855	virginia	usa	22 hours left
4	27700	chevrolet	1500	2018	clean vehicle	6654.0	red	3gcporec2jg473991	167763266	florida	usa	22 hours left

**Handling Missing Values:** Identifying and fixing missing values in the dataset is an important part of data pre-processing. We rigorously evaluate each attribute for missing values and use procedures such as imputation or elimination to maintain data integrity.

```
In [5]: #missing data
car_data.isnull().sum().sort_values(ascending=False)
```

```
Out[5]: price          0
brand          0
model          0
year           0
title_status   0
mileage        0
color          0
vin            0
lot            0
state          0
country        0
condition      0
dtype: int64
```

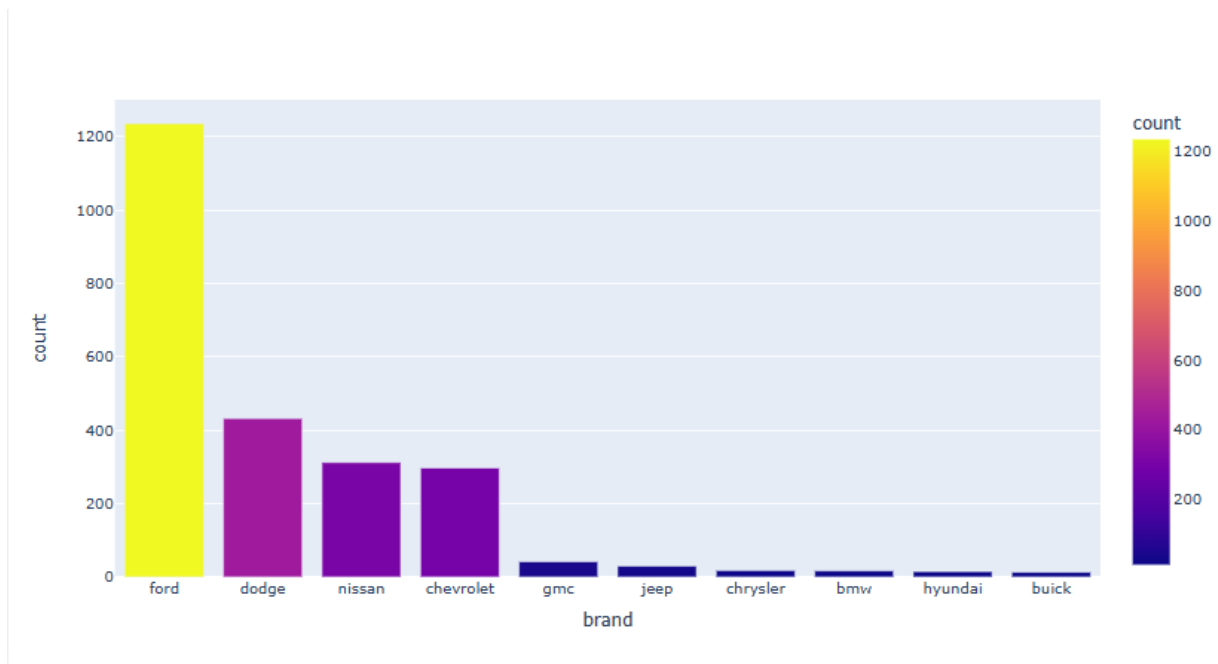
**Handling Zero Values in Price:** When we encounter zero values in the 'price' column, we replace them with the dataset's median price. This approach ensures that our analysis is not influenced by incorrect or missing price information.

```
In [6]: median_price = car_data['price'].median()
car_data['price'] = car_data['price'].astype(int)
car_data['price'].replace(0,median_price ,inplace=True)
```

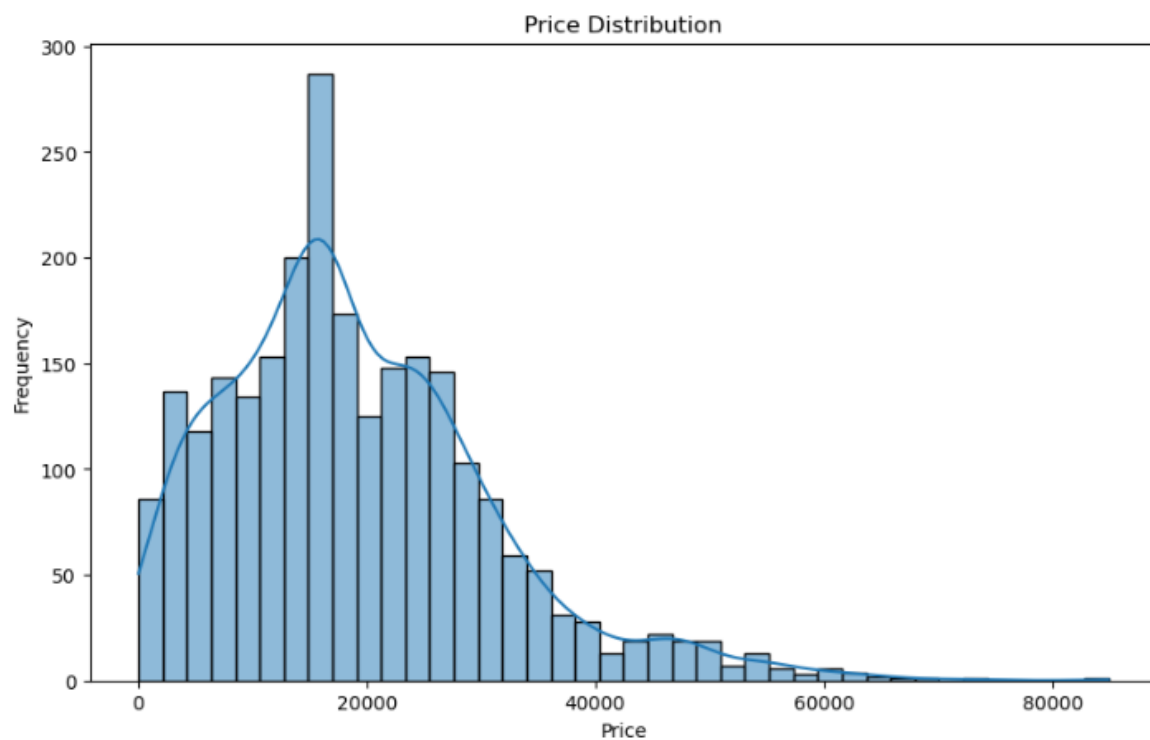
## Visualization:

With our dataset thoroughly pre-processed and sanitized, we move on to the creation of interactive visualizations aimed at uncovering critical facts about the US car market. Our visualization arsenal contains:

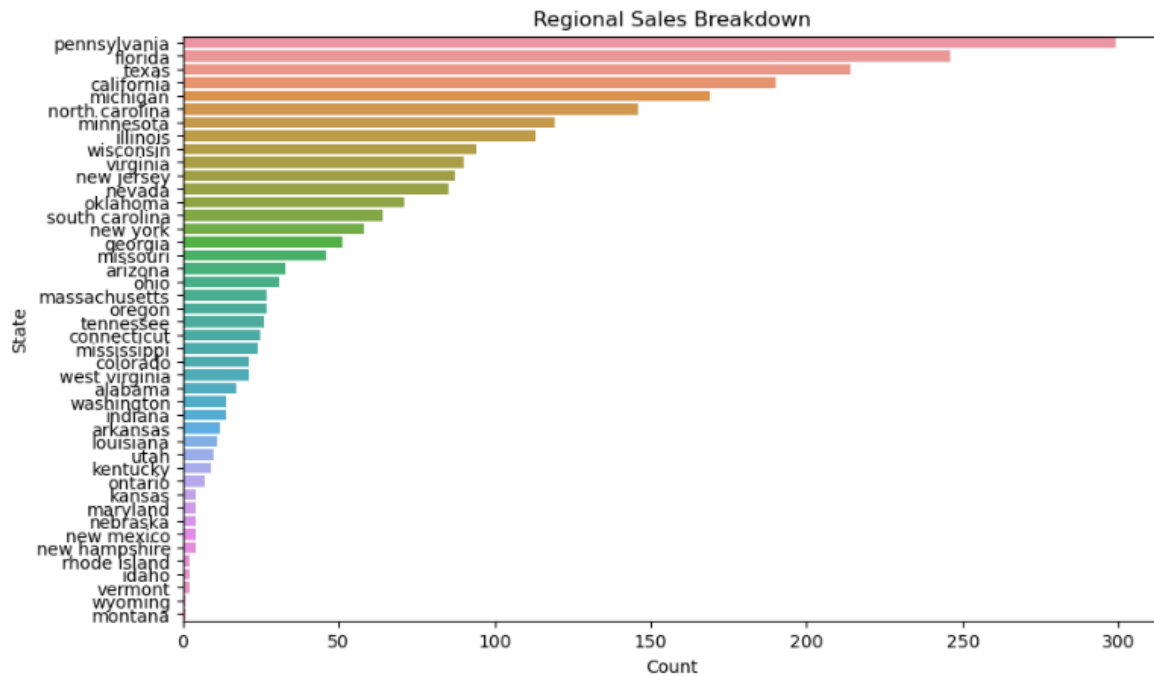
**Brand Preference:** Bar charts show how large brands dominate the market.



**Price Distribution:** Histograms demonstrate the distribution of vehicle prices, highlighting outliers and common price ranges.



**Regional Sales Breakdown:** Bar charts show sales patterns across states, offering information on regional preferences.

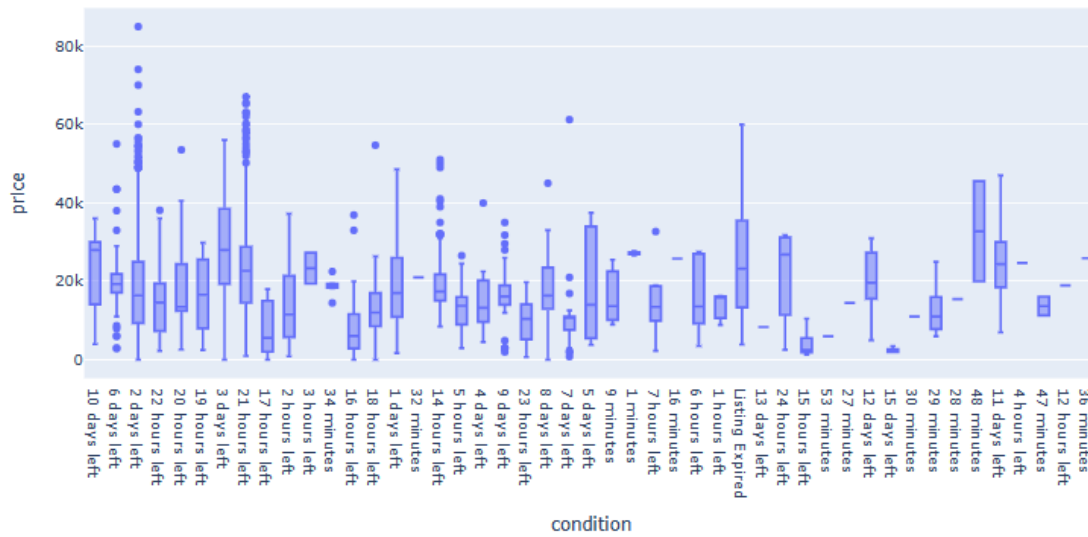


**Price vs Year:** The scatter plot depicting Price vs. Year provides a comprehensive overview of how the prices of vehicles vary based on their year of registration, with additional insights garnered from the brand and mileage size representations. The average price has steadily increased over the past decade.



**Car Price vs. Condition:** Box plots show pricing dependent on car condition, illustrating how condition affects costs. Cars in better condition tend to cost more than cars in worse condition. There is a range of prices for each condition. For example, a car in "excellent" condition could cost anywhere between \$40,000 and \$80,000.

Vehicle Condition vs. Price

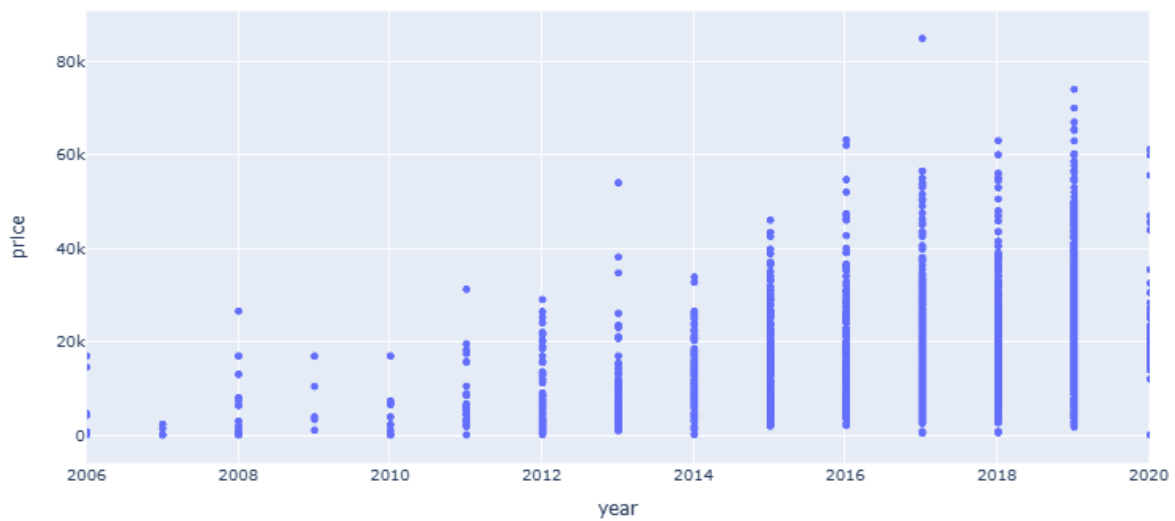


**Interactive Features:** Zooming in on certain data subsets in scatter plots allows users to perform more extensive analysis.

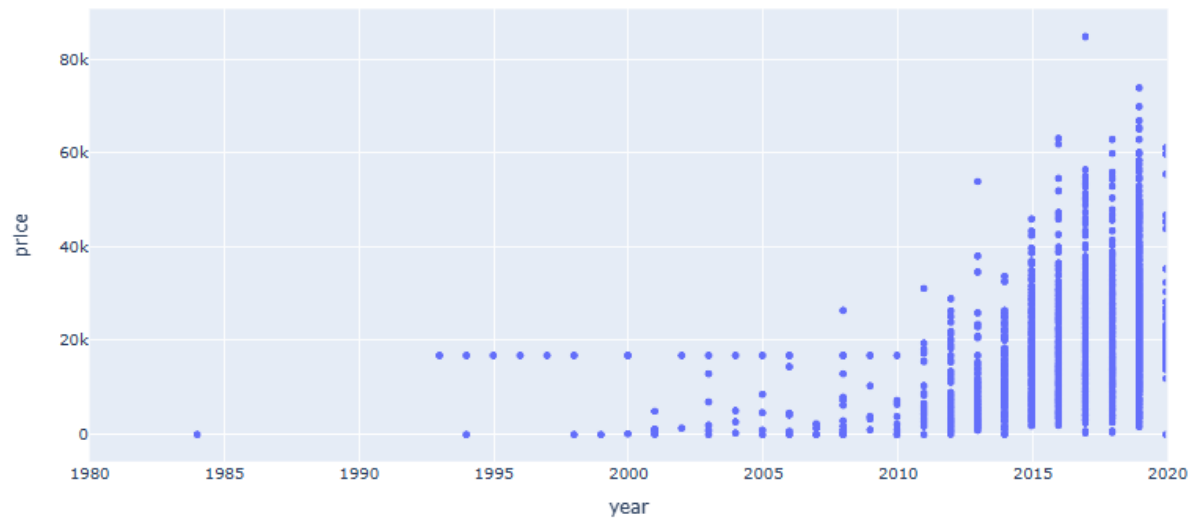
Zoom Level:  2006.00 – 202



Year vs. Price



Year vs. Price



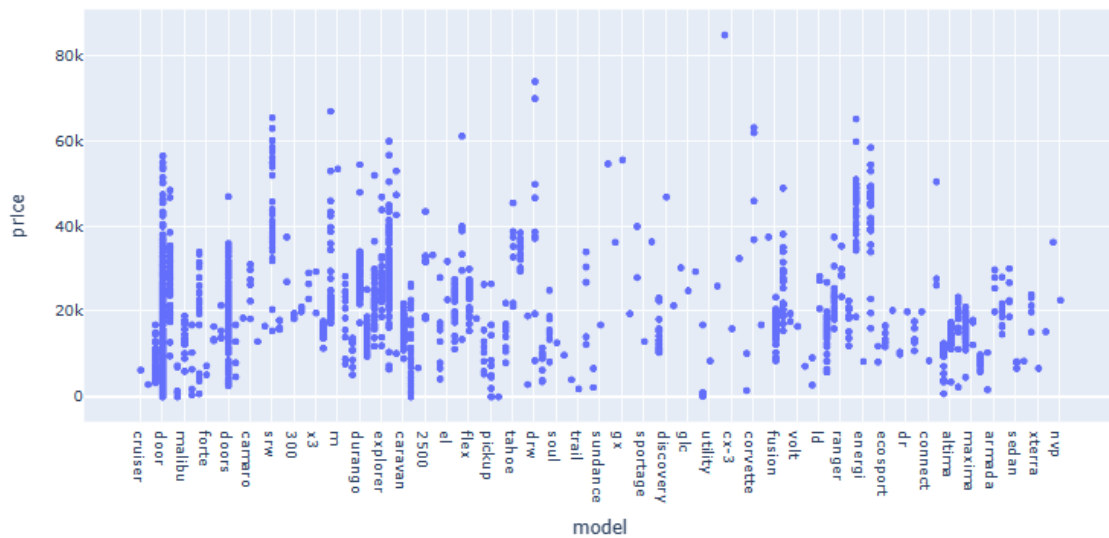
Plot Type:

X Axis:

Y Axis:

Generate Plot

Scatter Plot



## Insights:

**Brand Influence:** Different brands have varied price trends, with some charging larger fees regardless of the year of registration. Luxury companies typically sustain higher price points

throughout multiple model years than mainstream manufacturers. Pricing within the same brand may also vary depending on factors such as model popularity, market demand, and brand reputation.

**Price Distribution:** The bulk of automobiles fall into a given price range, with a few outliers reflecting luxury models that attract higher costs.

**Regional Sales Breakdown:** States such as Florida, Georgia, and New Jersey have larger car sales, demonstrating that preferences differ between areas.

**Car Price vs. Condition:** Cars with a clean title typically attract greater prices than those with salvage titles, highlighting the importance of vehicle condition in pricing.

Our research revealed useful insights into customer behavior, market trends, and geographical variances in the US car market. The dominance of major brands, the impact of vehicle condition on pricing, and regional sales patterns all offers stakeholders useful information for strategic decision-making.

### **Future Work:**

To improve our analysis, future project iterations could include machine learning algorithms for predictive modeling, sentiment analysis of user reviews for brand perception, and real-time data integration for up-to-date insights. Furthermore, improving the interaction elements and broadening the scope to incorporate socioeconomic factors could increase the depth of our investigation.

### **Conclusion:**

"Dashboard Drive" provides a thorough examination of the US automobile business, utilizing interactive graphics to understand complex facts. By identifying trends and patterns, our initiative provides stakeholders with vital insights for effectively navigating the dynamic automotive landscape. Through rigorous research and visualization, we give a road map for informed decision-making and strategic planning in the automobile industry.