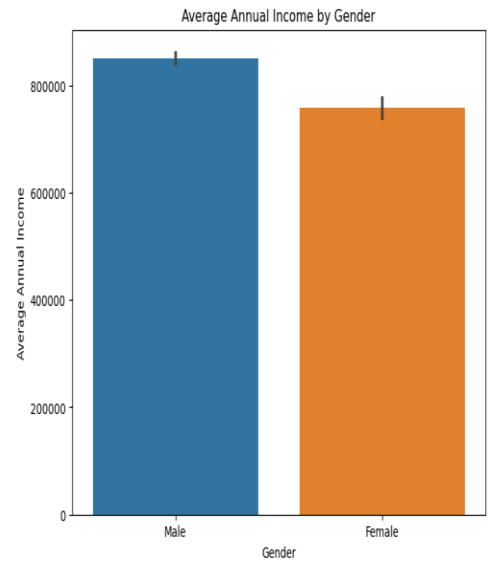
Potential Analysis:

Gender Analysis:



1.The average annual income for **Men** is higher than that of **Women**. Based on the data:

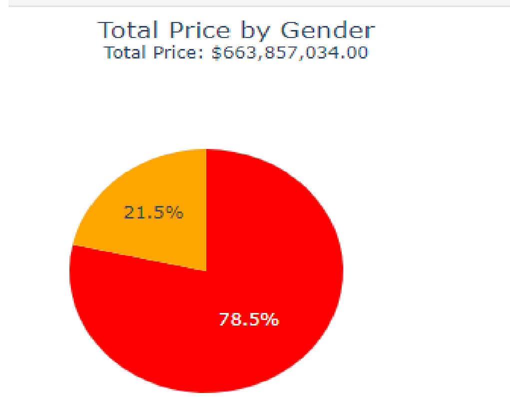
- **Men's average income**: $850,000
- **Women's average income**: $750,000

2.Sales Proportion:

- **78.5%** of total sales come from male customers.
- suggests that **male buyers are**

responsible for the majority of transactions inthisdataset.

- This could be attributed to various factors such as higher purchasing power, preferences for certain car models, or targeted marketing strategies that favor male customers.



Company Analysis:

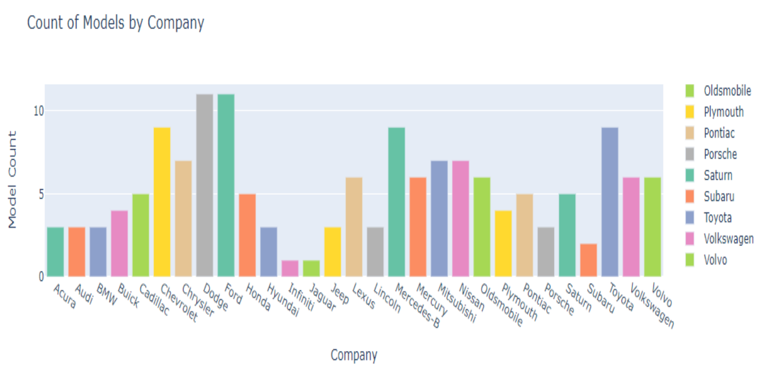
1.I identified the top 10 companies with the highest sales.

2.I also looked at the number of models each company has in the market.

3.The company with the highest sales isn't the one with the most models in the market.

4.The second company, however, has the most models and is the biggest competitor.

5.The other companies are followers or lagging behind in the market.



Models Analysis:

1. I analyzed the top 10 models with the highest sales.

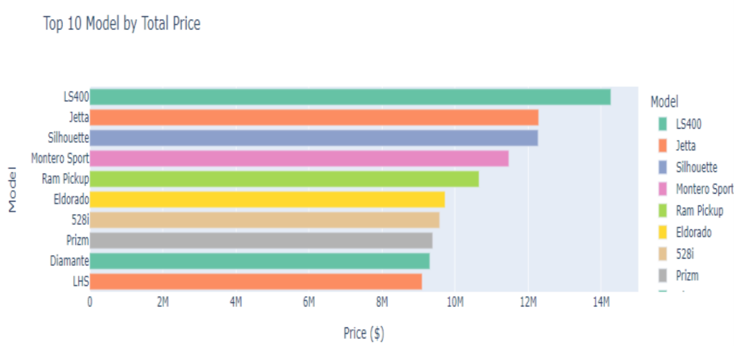
2.This helps identify which model is most popular, so the owning company can focus on stocking it, taking into account turnover to avoid overstocking and leaving space for competitors.

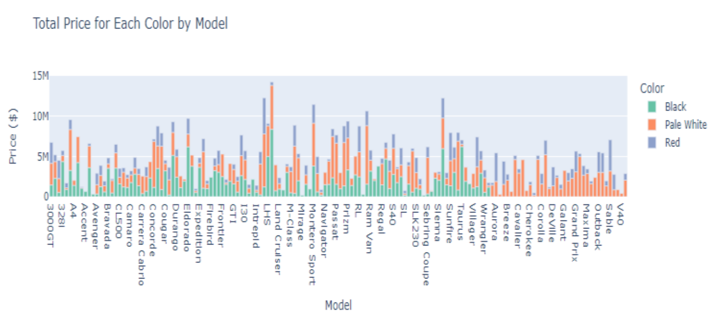
3.Competitors can then add this popular model to their lineup with a unique selling point (USP) to try to outperform the leading company.

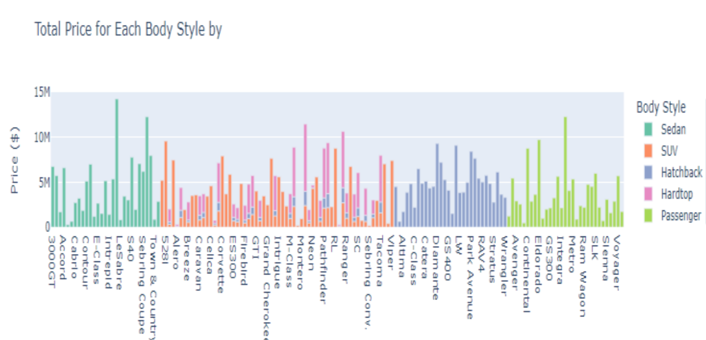
4.Competition is not just about price; factors like service and presentation also play a role.

5.I also found the best-selling body style to be SUVs, though this can vary by time, location, and customer preferences.

6.The most in-demand color was black, which indicates a preference for sophisticated taste and higher purchasing power in that region, as it's also the most expensive color.

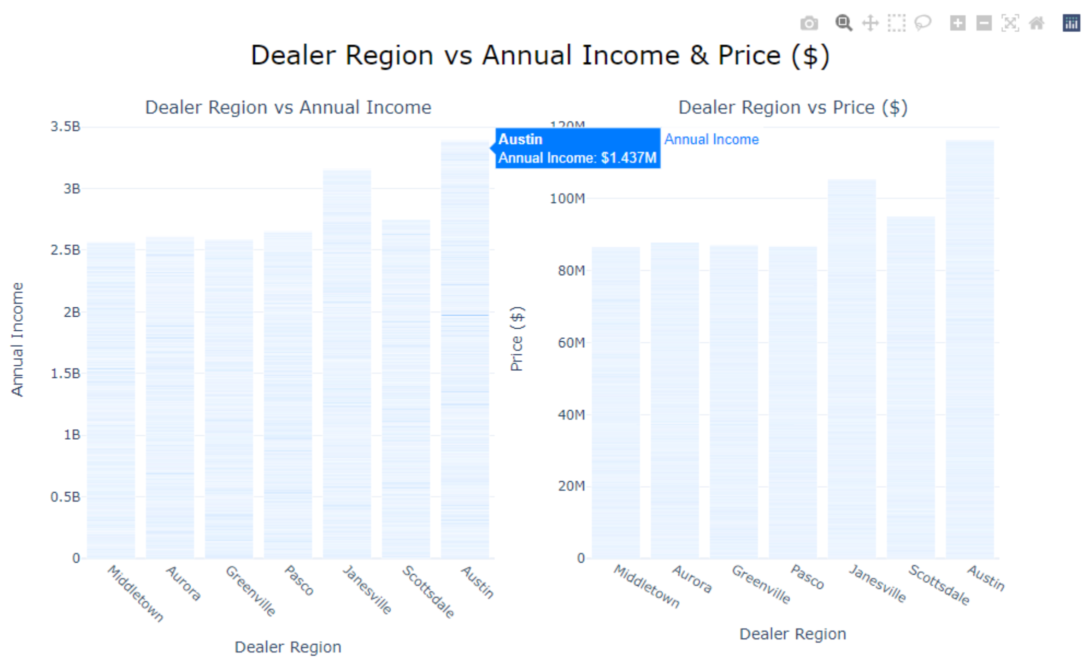






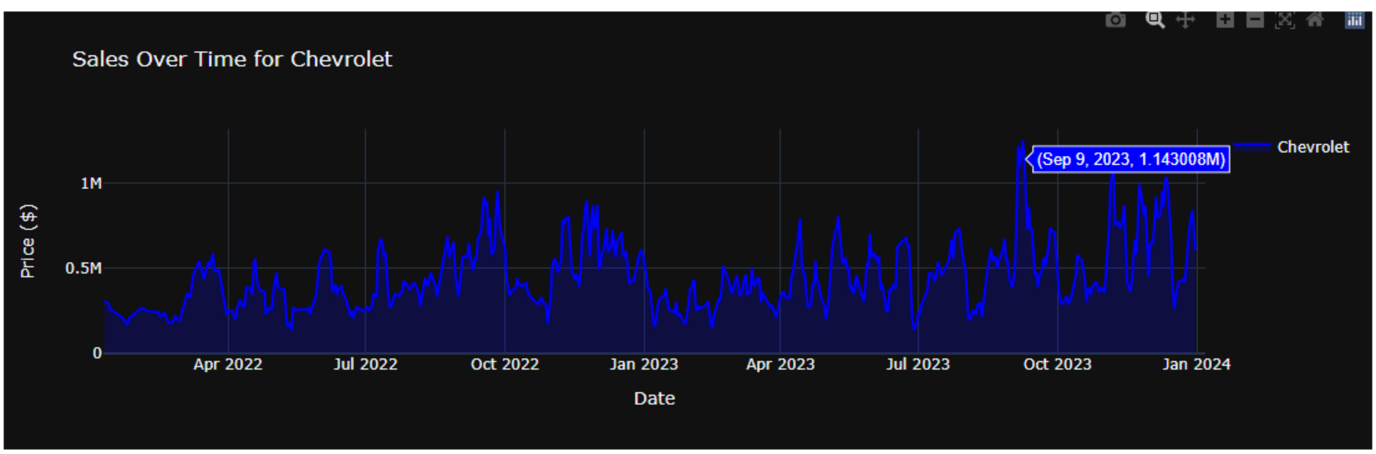
Dealer Region Analysis:

- I analyzed the dealer regions and found that the highest income area also had the highest sales.
- There is a direct relationship between income levels and sales volume.
- Looking at the customer, this region was the best choice for targeting, and the team performed well in this area.
- The marketing team made smart decisions in selecting this region for sales.
- The sales variation across regions wasn't very large, which is a positive sign and reflects successful marketing strategies.



Chevrolet Sales Analysis:

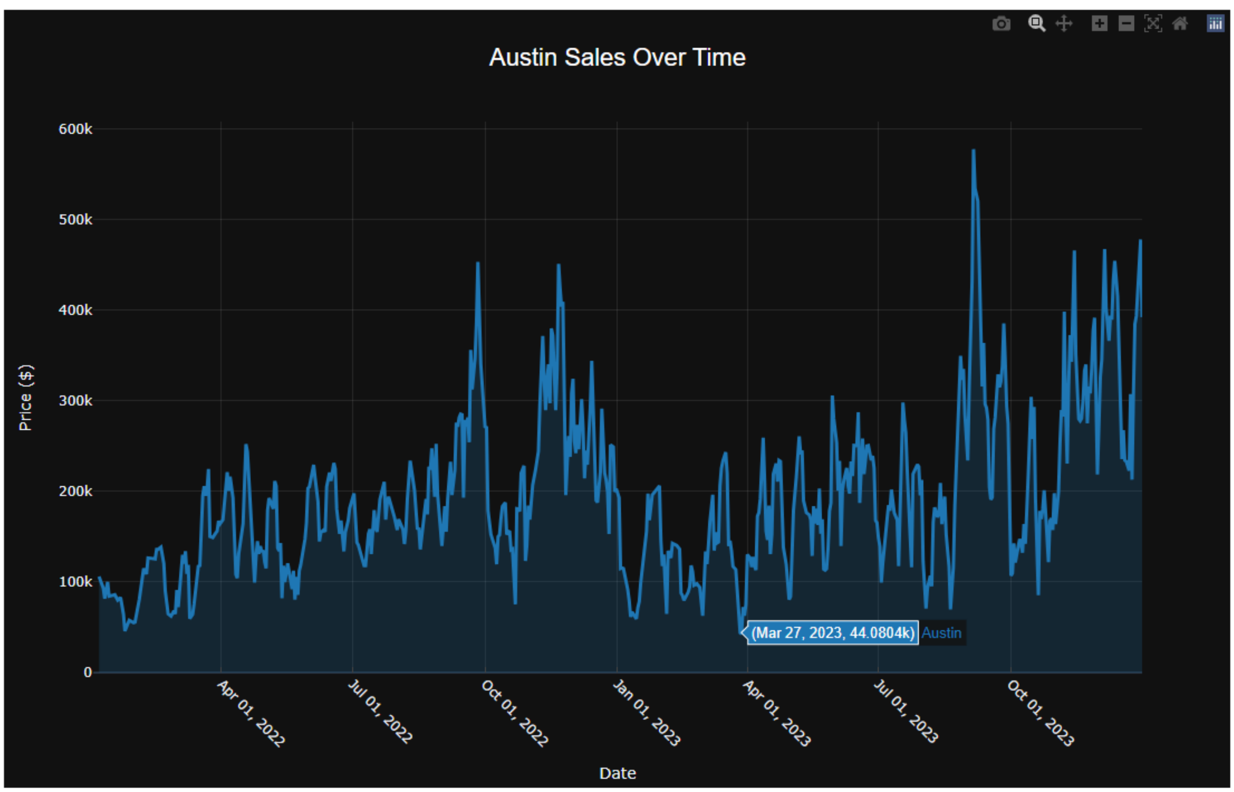
- I analyzed sales for the top company from 2022 to 2024.
- I noticed the company had the lowest sales in June 2023 and the highest in September 2023.
- The company was able to recover quickly from the dip in sales, which is impressive and shows its strength.
- This quick recovery is a key indicator for investors, as it shows resilience.
- The company's ability to bounce back quickly is what helped it maintain its position as number one, highlighting the strength of its team and operations.



Austin Sales Analysis:

- I analyzed the highest income and highest sales region, which is Austin.
- I found that the lowest sales occurred in Marth 2023, while the highest sales were in September, similar to the overall trend.
- The sales trend in this region is stable, as the customer base doesn't fluctuate drastically.
- However, seasonal factors like summer and winter affect customer decisions, with sales being higher in the summer and lower in the winter.

- This should be considered when managing inventory, and it’s important to analyze each region's customer base and how timing influences sales



Feature Engineering and Data Preprocessing for Model Optimization:

- I started by cleaning the ID feature, removing duplicates and converting it to an integer to avoid model issues.
- I deleted unnecessary features.
- I used a correlation matrix, but since the results sometimes aren't accurate, I used Variance Inflation Factor (VIF) to identify and remove highly correlated features.
- I created a new dataset with only the most important features, which best represent the data.
- I applied feature engineering by mapping categorical features to 0 and 1 for easier model processing.
- I checked for outliers before modeling to determine the best scaling technique to use.

Model Development and Optimization:

- I started by separating the features for training.
- First, I tried Linear Regression, but the results were very weak.
- I added features like year, month, and day, which slightly improved the model, but it was still weak.
- Then, I tried XGBoost, but it showed high variance according to cross-validation results.
- I experimented with Ridge regression but found Elastic Net performed better, especially with alpha set to 0.1 and l1_ratio to 1.
- Since the data was nonlinear, I applied Polynomial Features of degree 2, but increasing degrees led to a huge rise in feature count.

- I found Min-Max scaling to be the best scaling method.
- With this setup, I achieved the best result: an R^2 of 0.69 on both training and testing, with minimal overfitting.
- I also tried Random Forest with increased n_estimators, but it resulted in very high variance.
- The best solution was using Polynomial Features with Elastic Net, as XGBoost also showed high variance.