**CAR PRICE PREDICTION: FORD**

**FINAL PROJECT**

**MGIS - 650**

**ROCHESTER INSTITUTE OF TECHNOLOGY**

**PREPARED BY :**

JAINAV MUTHA

DEEP MEHTA

SATHVIK REDDY

CHAITNA MANKALA

**Model For Prediction of Car Price**

* **About Dataset**

The dataset includes information on Ford car models, encompassing key attributes such as production year, price in dollars, transmission type (automatic, manual, semi-auto), mileage (number of miles traveled), fuel type (petrol, diesel, hybrid, electric, other), annual tax, miles per gallon (mpg), and engine size. This comprehensive set of data allows for a detailed analysis of Ford vehicles, providing insights into their specifications, performance, and economic considerations. Whether one is interested in fuel efficiency, tax implications, or the latest models, this dataset offers a well-rounded perspective on Ford cars.

* **Preprocessing of Ford Data**
* In the initial phase of data preprocessing, we conducted a thorough examination of null values within the dataset. Our approach involved either eliminating insignificant missing values, using techniques such as listwise deletion for rows or column deletion for non-critical variables, or employing mean imputation for numerical data where missing values were substituted by the average value.
* Then, we conducted a check for duplicate values within the dataset and subsequently removed any instances identified. This step is crucial to maintain data accuracy and prevent potential biases or distortions in subsequent analyses.
* Then, we removed all negative values from the "Price" attribute.
* Finally, we used Label Encoding for variables like ‘model’, ‘transmission’, and ‘fuelType’ as they had categorical data.
* **Scaling**

After the data cleansing processes, we applied the Min-Max Scaling technique to normalize the dataset. This normalization step is valuable for preventing any feature from disproportionately influencing analyses or machine learning models due to differences in scale.

* **Prediction Model**
* We divided the Ford dataset into test and train respectively.
* Then we implemented multiple prediction models to find the best one

|  |  |
| --- | --- |
| **MODEL** | **RESULT (R2 score)** |
| Linear Regression | 0.72 |
| Random Forest | 0.90 |
| XGBoost | 0.93 |

* From the above table, we can see that the XGBoost algorithm is the most effective, so we chose that.
* The results using XGBoost are, **mean absolute error:** 828.011, **root mean squared error:** 1219.679, **R2 score:** 0.931.

The mean absolute error (MAE) of 828.011 indicates that, on average, the model's predictions deviate by approximately 828 units from the actual values.

The root mean squared error (RMSE) of 1219.679 offers an interpretable measure of the model's prediction error in the original units of the target variable.

The R2 score of 0.931 signifies that the XGBoost model explains about 93.1% of the variance in the target variable, indicating a high degree of explanatory power.

* **Predicting the cost of the same model of cars in different years.**

So, let’s say there is a car, our model says that the price of that car in 2010 is $9393.382.

A black background with white text

Description automatically generated

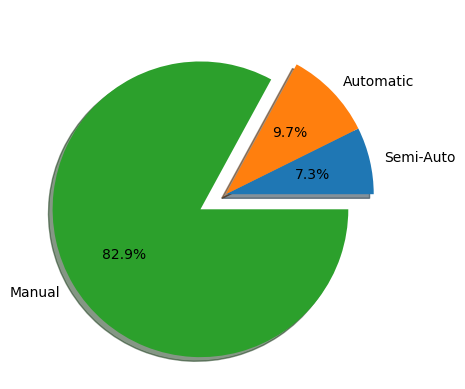
The cost of same model car in 2025 is $ 22868.488.

A black background with white text

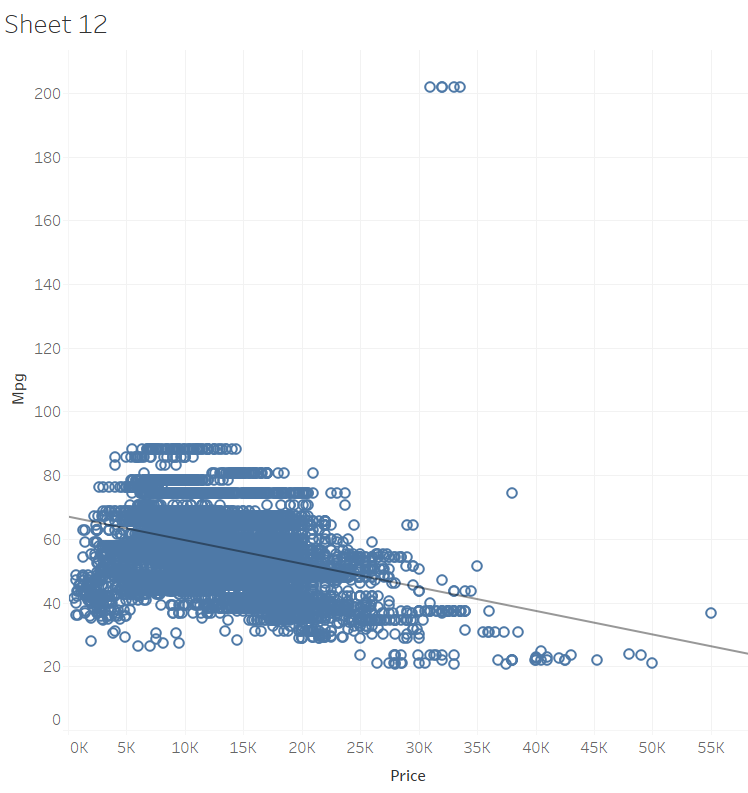
Description automatically generated

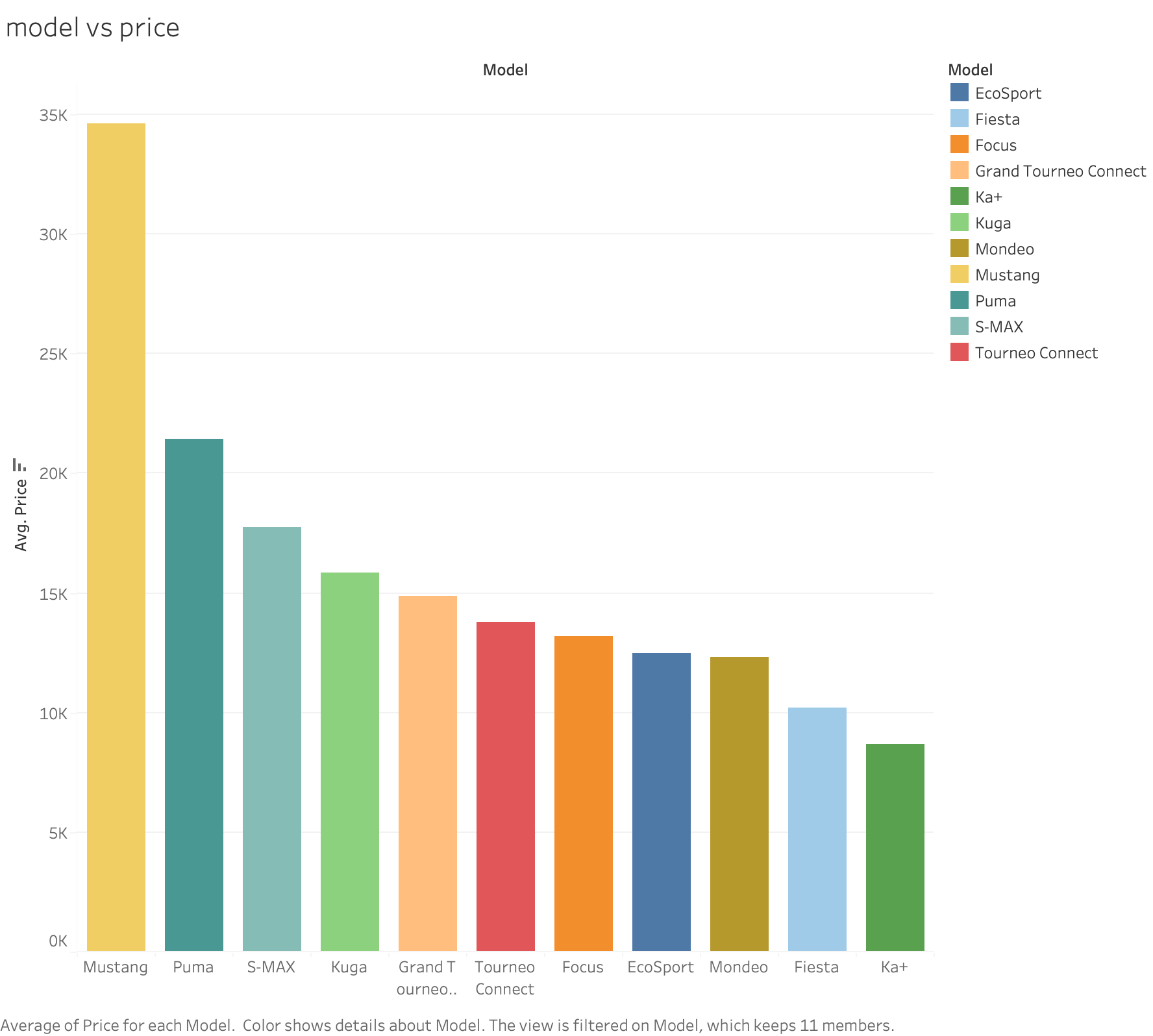
Here we can see how the same car’s price has increased by more than 2 times over 15 years.

**ANALYSIS OF THE FORD DATASET**

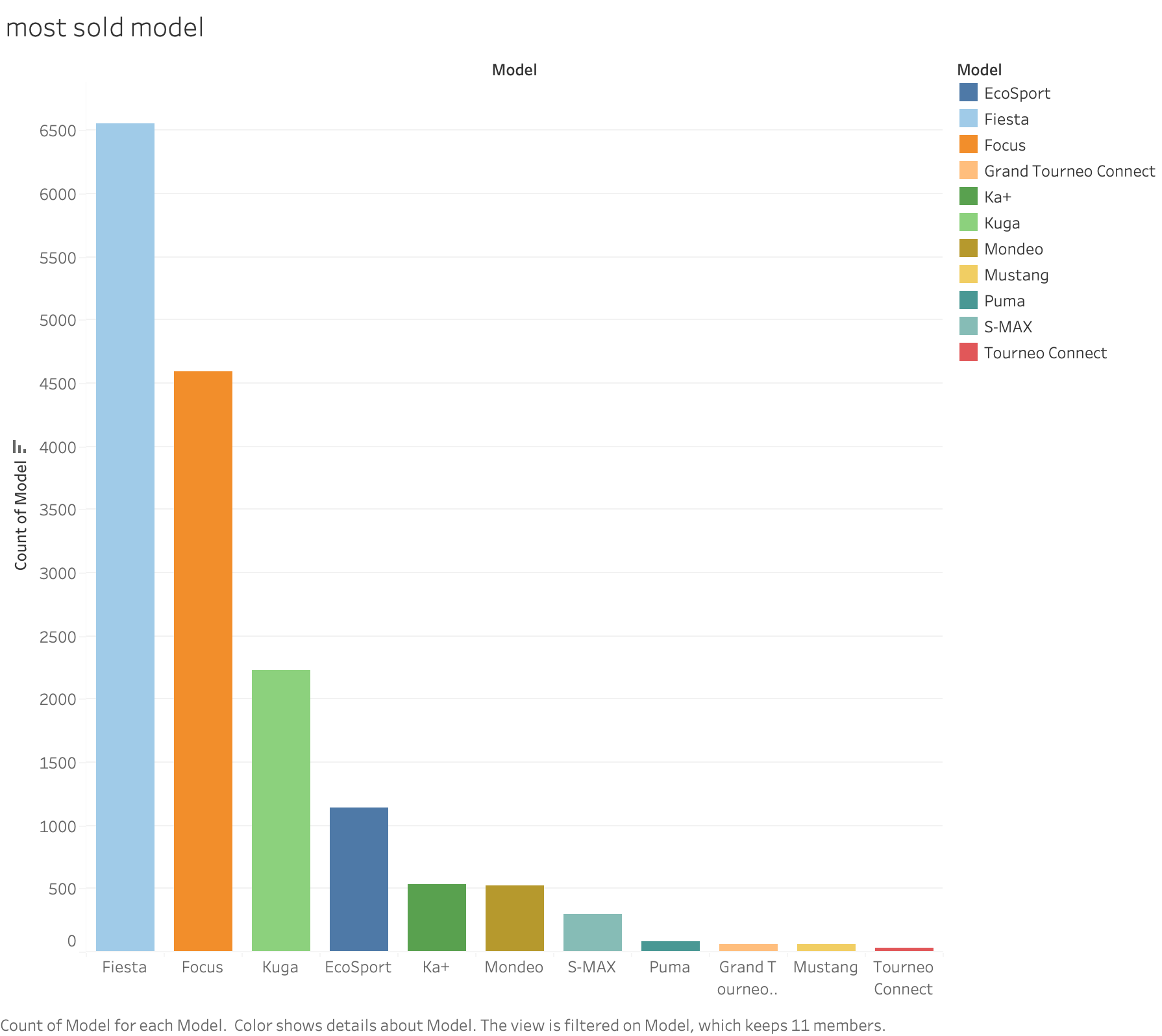


Given the dataset spanning from the 1990s to 2020, the observed prevalence of manual transmissions for Ford vehicles suggests a longstanding consumer preference. To navigate the evolving automotive landscape, Ford could implement a gradual transition strategy, integrating advanced automatic and semi-automatic options, while preserving the brand's legacy associated with manual transmissions. This approach aims to harmonize with historical preferences while strategically aligning with the contemporary market shift towards automatic transmissions.

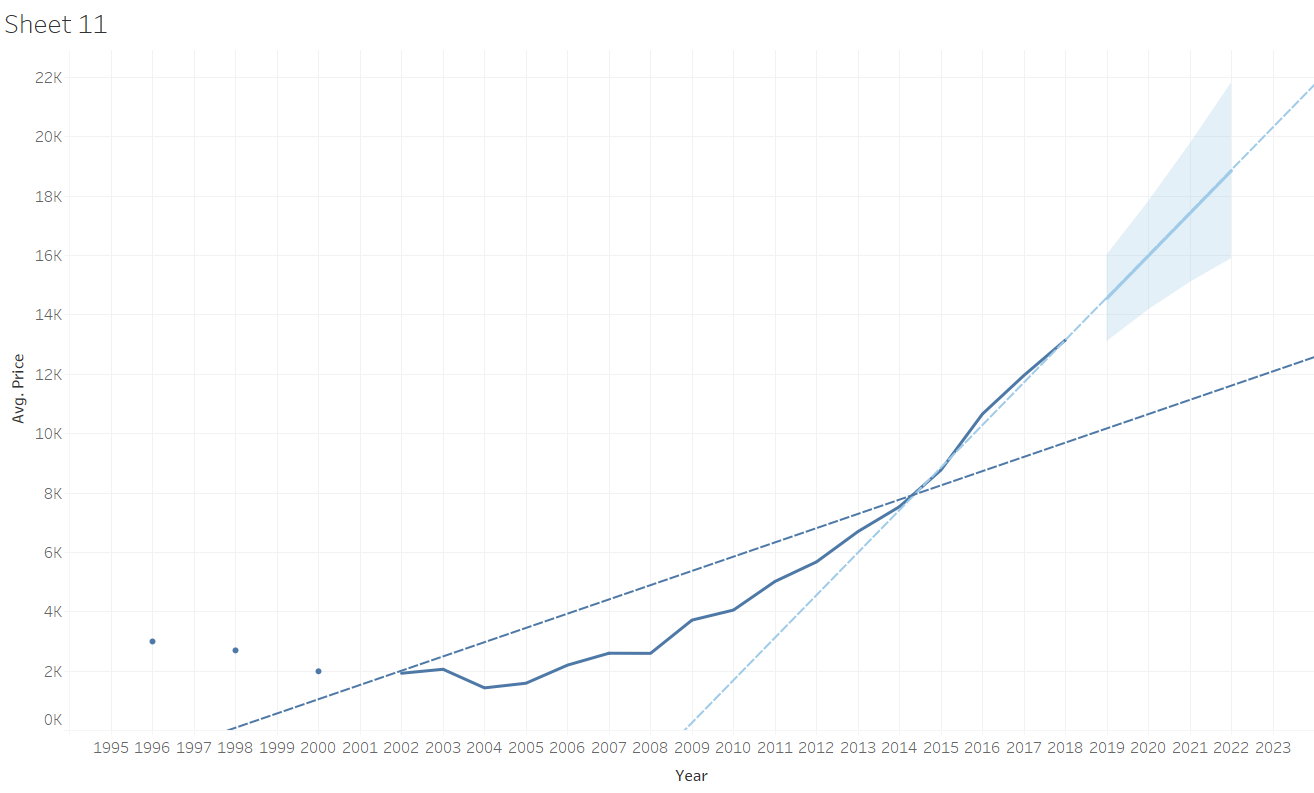
The observed trend of lower mileage correlating with higher prices, especially in sports or luxury cars, may be attributed to several factors. High-performance vehicles often feature larger engines, and their designs prioritize power and precision over fuel efficiency, resulting in lower average mileage. Additionally, owners of sports or luxury cars may use these vehicles sparingly, contributing to the lower mileage figures. The combination of sophisticated engineering, limited use, and the inherent value associated with premium brands collectively influences the pricing dynamics, reinforcing the inverse relationship between mileage and price in this specific automotive segment.



The graph depicts Ford's car model prices, showcasing a significant contrast. The Mustang stands out with an average price of $35k, positioning it as a premium choice. Meanwhile, the Ka+ and Fiesta emerge as the most affordable options at around $10k, appealing to a budget-conscious consumer base. This pricing strategy reflects Ford's effort to cater to diverse preferences and financial considerations within their lineup.

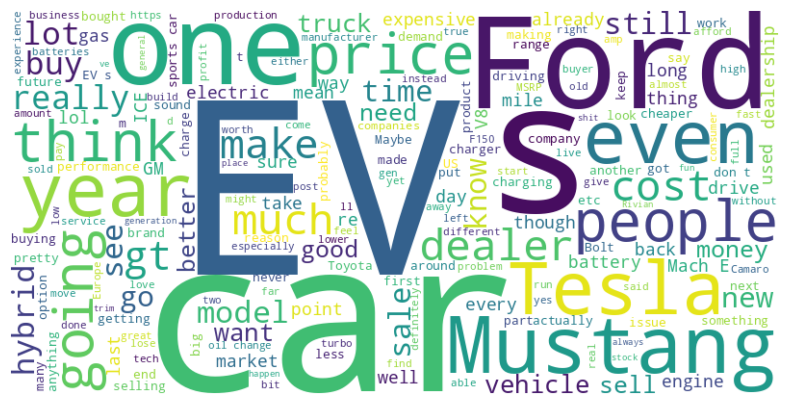
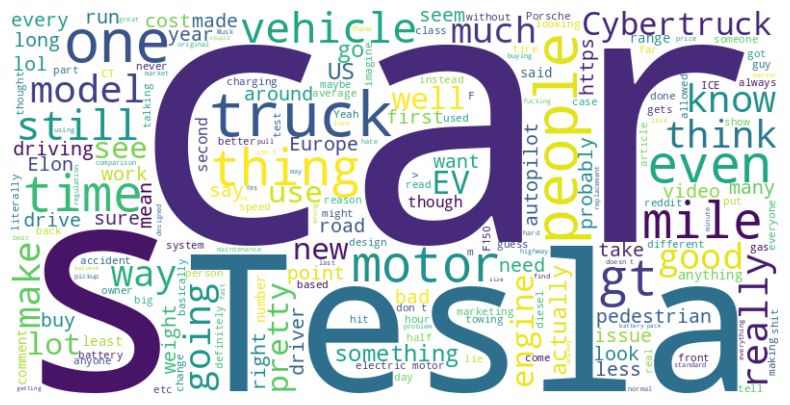


The graph provides a clear indication that the Fiesta holds the distinction of being the best-selling car in Ford's lineup. One plausible explanation for this success could be the comparatively lower average price of the Fiesta. Positioned as an affordable option, it likely appeals to a broader consumer base, contributing to its higher sales figures. Following closely are the Focus and Kuga, occupying a mid-range price category, suggesting a balanced demand in this segment. However, the more premium models such as the Mustang and Puma exhibit notably lower sales, presumably owing to their higher average prices. This observation underscores the impact of pricing on consumer preferences, with affordability playing a significant role in driving sales within Ford's diverse range of car models.



Over the past 15 years, the average price of Ford cars has seen a substantial increase, reflecting broader trends in the automotive industry. In 2005, the average car price for Ford vehicles was a modest $2,000. However, by the year 2020, this figure had surged to $14,000, indicative of evolving market dynamics, technological advancements, and changing consumer expectations. Looking ahead, predictions are suggesting a continued upward trajectory in average car prices for Ford models. Projections indicate that by the near future, around the range of 2023 to 2025, the average price is expected to further climb, reaching a bracket between $16,000 and $22,000. These forecasts consider factors such as inflation, innovations in automotive technology, and shifts in consumer preferences that contribute to the ongoing evolution of car pricing within the Ford brand.

**Analyzing Sentiments: Ford in Comparison with Competitors**



**WordCloud for Ford**   **WordCloud for Tesla**

Some important words and why do they matter?

**EV (Electric Vehicle):** Showcase Ford's commitment to the electric vehicle revolution by emphasizing the technological advancements and eco-friendly features of Ford's electric lineup, promoting a cleaner and more sustainable future.

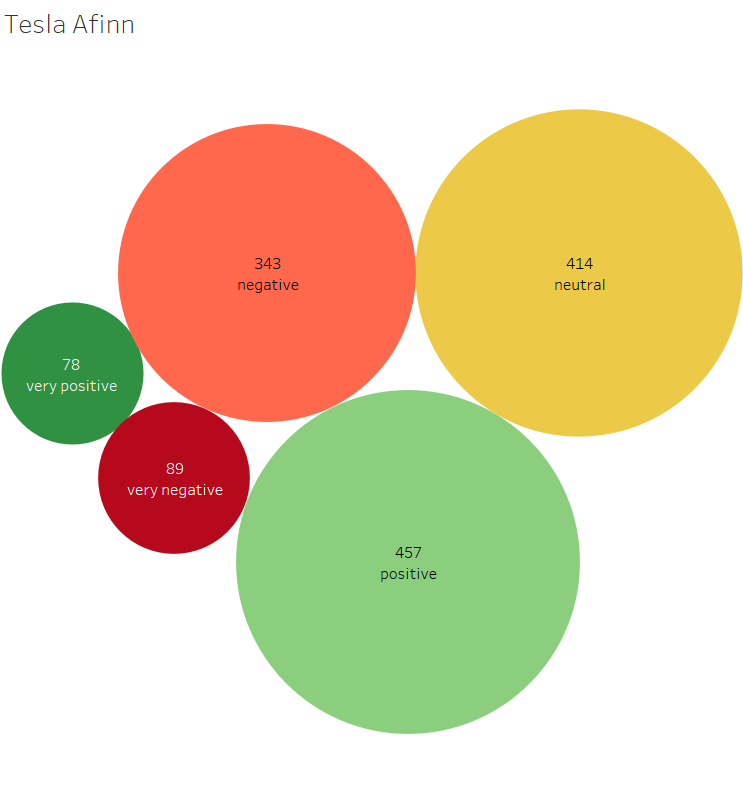
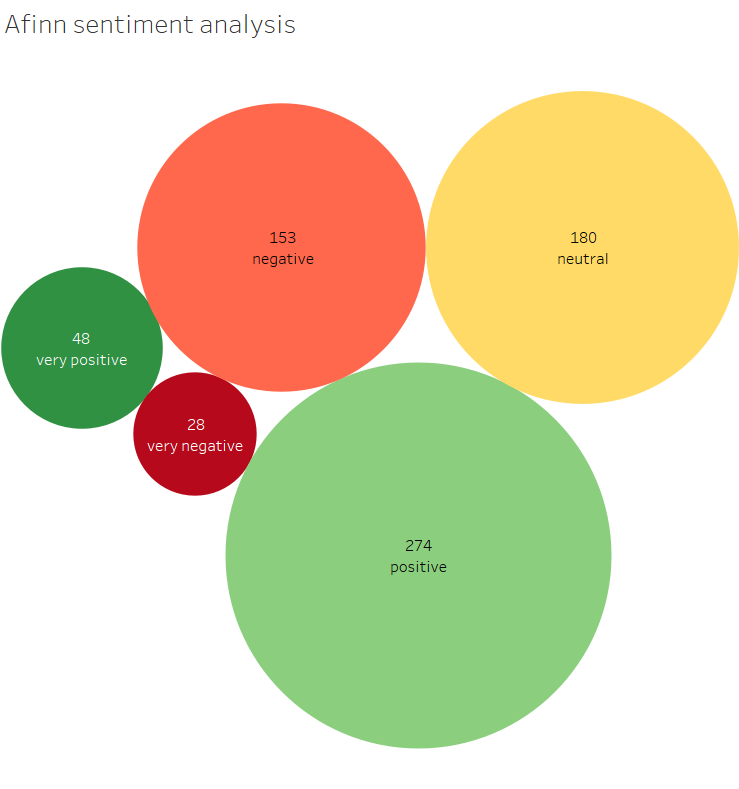
**Mustang**: Leverage the iconic Mustang brand to create excitement and allure around Ford's electric offerings, such as the Mustang Mach-E, emphasizing the fusion of performance and electric power to attract a diverse range of consumers.

**Hybrid:** Highlight Ford's dedication to offering a range of sustainable options with its hybrid lineup, emphasizing the benefits of combining traditional internal combustion engines with electric power to provide efficiency without compromising performance.

**Cost:** Communicate the value proposition of Ford's electric and hybrid vehicles by emphasizing cost-effectiveness, including lower operating costs, potential government incentives, and long-term savings, making sustainable mobility an economically viable choice for consumers.

The prominence of "Tesla" in the Ford word cloud may be indicative of the growing significance of Ford's electric vehicle (EV) endeavors. This heightened visibility suggests an increasing trend of comparison between the two automakers, possibly fueled by Ford's pursuit of EV goals in response to Tesla's leadership in the electric vehicle market. The juxtaposition in the word cloud highlights the evolving dynamics within the automotive industry.

Ford and Tesla both stand at the forefront of the electric vehicle (EV) revolution, each bringing its unique strengths to the market. Ford has a rich history in the automotive industry, known for its durable and reliable vehicles. As executives at Ford, it's crucial to highlight the company's commitment to innovation and sustainability. To enhance Ford's competitiveness against Tesla, it's essential to focus on the integration of cutting-edge technology, ensuring robust charging infrastructure, and emphasizing a seamless user experience. Learning from Tesla's success, Ford can further invest in autonomous driving technology, software updates, and energy efficiency to stay at the forefront of the rapidly evolving EV market. Collaborating with technology partners and prioritizing customer feedback will contribute to Ford's success in providing a compelling alternative to Tesla and solidifying its position in the electric vehicle landscape.



FORD TESLA

Analyzing sentiment percentages for Tesla and Ford reveals distinct patterns. Tesla exhibits a higher positive sentiment percentage, with 33.33% of comments expressing positivity, including 5.68% very positive sentiments. Additionally, Tesla's negative sentiments stand at 31.67%, encompassing 6.49% very negative sentiments and 25.02% negative sentiments. Meanwhile, Ford's sentiment percentages are more evenly distributed. Positive sentiments for Ford constitute 47.14%, with only 7.07% being very positive, while negative sentiments make up 26.50%, comprising 4.09% very negative and 25.61% negative sentiments. These percentages highlight the variance in sentiment distribution between Tesla and Ford, with Tesla leaning more towards positive sentiments despite a larger dataset, while Ford sees a higher proportion of positive sentiments in its smaller dataset.

Several factors could contribute to the observed differences in sentiment distribution between Tesla and Ford. Here are potential reasons and considerations for both positive and negative sentiments:

**Positive Factors:**

* Long-Standing Reputation: Ford, as an established automaker, may benefit from a positive reputation built over decades, contributing to a favorable sentiment base.
* Product Range: Ford's diverse product range, including trucks, SUVs, and electric vehicles, may appeal to a broad consumer base, positively influencing sentiment.

**Negative Factors:**

* Electric Vehicle Perception: Ford's association with traditional combustion engine vehicles might lead to less positive sentiment, especially in comparison to Tesla's focus on electric vehicles.
* Marketing and Branding: Ford may face challenges in marketing and branding, potentially impacting how consumers perceive the brand in the context of emerging trends in the automotive industry.

**Conclusion**

Our project utilized data science techniques to predict Ford car prices, with XGBoost yielding the most accurate results (R2 score of 0.93). Through thorough data preprocessing, we addressed missing values, eliminated duplicates, and normalized the dataset. Insights revealed enduring consumer preferences, pricing strategies, and future price trends. A sentiment analysis compared Ford with Tesla, showcasing distinct patterns. The project contributes valuable insights into Ford's pricing dynamics, aiding in understanding consumer preferences and future market trends.