**Tesla Sales Visualizing the Journey of Electric Excellence**

**Team-1**

**Team members:**

Masrul Akhand

Srujani Mareddy

1. **Data Search:**

We started this endeavor by looking through several different datasets before deciding upon the Tesla dataset. It was found to be both intriguing and relevant, and it had all the features we were looking for. Missing data, null values, data redundancy; it would both be a challenge and a good learning tool. With our choice made, we began by establishing roles and responsibilities.

As our group only has three individuals, we knew we would have to adjust the roles to create a good balance. Masrul started with data engineering, organizing and cleaning the data. Srujani then worked together to analyze the data and create the visualizations. They also built the PowerPoint presentation and Jarred will be the presenter. And for this project document, we have selected to work on it together. Allowing us each to explain the portions we worked on the most effectively.

1. **Data Set and approach:**

Our project concept is to utilize the dataset to find commonalities amongst the pricing of used Tesla. We took the approach that we would be organizing the data and creating visualizations that could put us on a path where our next steps would be to create a new algorithm that could be used for pricing future used Tesla. We found this to be a very logical approach as it has direct real world applications.

The first step was to find an adequate dataset. After looking through many, we found the Tesla dataset on Kaggle, link: <https://www.kaggle.com/datasets/aravindrajpalepu/tesla-used-cars/>

After agreeing upon this dataset, Jarred began working the data.

We utilized Google Colab for the convenience it offers us as a group. We got the dataset downloaded and added into Colab, and began to do preliminary discovery. A quick read of the .csv file, identified the shape and column titles, and then got the column types. We started the cleaning process by identifying how many null values there were. Then we split the variables so the numerical variables would be easier to work.

1. **Introduction and Data Preprocessing:**

Link of the Google Colab file from the start that we used to do project

<https://colab.research.google.com/drive/14i8ntRREdsIdii949qT7AGUIiGiuA_nH?usp=sharing>

We choose to start cleaning by identifying any outliers in price and odometer and eliminating them by using boxplots then identifying the thresholds. Next we looked into the unique values of the different variables and identified a redundancy between location (city and state), state, and zip code. As having the state and the zipcode is more than sufficient for our purposes, location was dropped as the city can be identified by zip code if needed, and we already have a state column. The missing values in driveTrain were dropped as inputting a random selection could give incorrect results for future pricing algorithms. In the year category, every vehicle is from 2018, so any missing values were updated to show this. And then the last action for cleaning was to update the missing values in state. This was done by creating a zip code map and using the zip codes already given to add the correct state. A screenshot of a computer

Description automatically generated

Sample data before preprocessing

1. **Exploratory Data Analysis:**

**Univariate Analysis:**

We performed a univariate analysis to examine the distribution patterns of important variables like 'price' and 'odometer’, ‘model’. Insights into the central tendency and variability of the data were gained using boxplots and histograms. The bar graph on the count of the model follows the Gestalts principles of color harmony and the color palette used is ‘rocket’ from seaborn color palette.

Pie chart: Paint Job is also a feature in the dataset that tells which paint has the highest demand and for which model. Which also tells the individual percentage of each paint job to understand the basic So to understand the basic paint job variation plotted a pie chart which tells the number of datapoints with those paint jobs and the percentage. We used a custom color palette and also the visualization follows the gestalts principles

**Bivariate Analysis:**

A scatterplot was employed to visualize the relationship between 'odometer' and 'price.' This analysis helped identify patterns and potential correlations in the data.

1. **Model Wise Analysis:**

We conducted a model-wise comparison which helps to get insights of how the model and price are related, and the count of each Tesla models across all states.

Also utilizing scatter plots to highlight differences in features such as DAS, odometer readings, and prices across various Tesla models. A box plot to analyze how the Odometer and DAS are `to observe how different types in DAS have the odometer range for every model, which provides insights into central tendency and spread of data.

1. **Geographic Analysis:**

**State Mapping and Visualization**

To enhance the dataset, missing ‘state’ values were filled using a state-zip code mapping. The dataset was then visualized geographically using a choropleth map, showcasing the average price per state for each Tesla model.

The choropleth map is an interactive map which helps to hover over each state and to find the model with highest average price in a state. Helps to understand the sales of and demand for model state wise. Dynamic Insights, animated scatter plot of price and odometer for every model in each state.

1. **Drive Assist System (DAS) Analysis:**

DAS and Price/Odometer Relationship

We explored the relationship between Drive Assist Systems and vehicle characteristics. Scatter plots and boxplots were used to analyze how DAS influences the price and odometer readings of different Tesla models.  
  
**Pair plot:**

Plotted a pair plot over the numerical features in the dataset 'odometer', 'price' against the categorical feature accident history and model. Here it tells how the features are varying for every model as our aim to check how each model has been affected through all the features.

1. **Conclusion:**

This strategy aims to present a comprehensive and insightful exploration of the dataset, highlighting key patterns, relationships, and trends in the used Tesla cars inventory. Key findings include geographic variations in average prices, the impact of Drive Assist Systems on prices and odometer readings, and model-wise comparisons. The visualizations created using Matplotlib, Seaborn enhanced our understanding of the dataset and facilitated clearer communication of trends and patterns.