**Part 1**

**TABLEAU**

**1. Data Preprocessing**

We used the auto-mpg dataset, which contains information on various attributes of cars, including mpg, horsepower, weight, and displacement. Here are the key preprocessing steps:

- **Missing Data:**
  The dataset had missing values in the "horsepower" variable. We addressed this by deleting any saved instances that were missing "horsepower".
- **Data Transformation:**
  The origin column was originally in key format. To clarify the data, we changed the numbers to represent the countries of origin instead.

**2. Create the Dashboard**

This dashboard portrays key factors that explain and find trends in our mpg- auto data.

**Filter**: This allows for the user to adjust the results to only display the origin(s) selected. All the charts are affected by this filter.

**KPI:** We have a very important KPI. The highest MPG is a significant factor to note as our Regression analysis focuses heavily on things that affect the MPG.

**Num Of Cars Per Origin**: A pie chart displaying how many cars there are per origin. We can see that America has a significantly higher distribution of cars than China and Japan.

**Acceleration, Displacement, and MPG per Year**: This line graph displays the trends of these 3 indicators over the course of around 10 years. While acceleration and mpg have improved over time, displacement has taken a slight decline.

**Horsepower Per MPG:** In this chart, the negative correlation between Horsepower and MPG is shown. As MPG in cars increases, the horsepower decreases and vice versa.

**R STUDIO**

**1. Data Preprocessing**

We used the auto-mpg dataset, which contains information on various attributes of cars, including mpg, horsepower, weight, and displacement. Here are the key preprocessing steps:

- **Missing Data:**
  The dataset had missing values in the "horsepower" variable. We addressed this by deleting any saved instances that were missing "horsepower".
- **Data Transformation:**
  The "horsepower" variable was initially a character type. To ensure compatibility with the regression model, we converted this variable into a number.
- **Selection of Variables:**
  Based on exploratory analysis, we selected "weight," "horsepower," and "displacement" as the predictors in the multiple linear regression model. "Weight" and "displacement" were considered key features based on their likely impact on mpg, while "horsepower" was included as it can affect fuel efficiency.

**2. Simple Linear Regression Model**

In the simple linear regression model, we explored the relationship between mpg and a single independent variable (e.g., horsepower). This model helps us understand how mpg varies as horsepower changes, assuming all other factors are constant.

- **Multiple R-squared:**
  0.7745094.
- **Adjusted R-squared:**
  0.7737476.
- **Regression Equation:**

mpg = 40.44456 + -0.006266213 * weight

**3. Multiple Linear Regression Model**

In the multiple linear regression model, we explored the relationship between mpg and multiple independent variables ("weight," "horsepower," and "displacement"). This model allows us to assess the combined effect of multiple factors on mpg.

- **Multiple R-squared:**
  0.7830071

- **Adjusted R-squared:**

  0.7807928

- **Regression Equation:**

  mpg = 39.37395 + -0.004789817 * weight + -0.02057266 * horsepower + -0.0058457 * displacement

**4. Model Performance Evaluation**

For the remaining 98 samples, we used the best-performing model (based on the R-squared and Adjusted R-squared values) to predict the mpg values. We compared the predicted mpg values with the actual mpg values from the dataset.

- **Residual Plot:**
  The residual plot provides insights into the errors made by the model. A good model should have residuals that are randomly scattered around zero. A systematic pattern in the residuals indicates potential issues with the model.
- **Histogram of Residuals:**
  A histogram of the residuals allows us to evaluate whether the errors follow a normal distribution. If the residuals are normally distributed, it indicates that the model assumptions hold true.

**5. Conclusion**

The analysis shows that the relationship between mpg and the predictors (weight, horsepower, and displacement) is strong in the multiple linear regression model. The residual plots and histograms provide useful diagnostics of the model fit, and the prediction of mpg for the remaining 98 samples was relatively accurate. The choice of using horsepower, weight, and displacement as predictors was justified, as they have a significant impact on the mpg of vehicles.

**PART 2**

**PYTHON – Analyzing Call Center Data**

**Distribution of Calls by State**

In this pie chart we see the percentage of calls that were from each state. This is a easy way to visualize the data that will be used in further visualizations. We see that majority of the data is coming from California with Michigan and Maryland coming up after that, and South Carilina as the minority.

**Sentiment Comparison by Channel**

In this bar chart we see the ratings of the customers for each channel of contact. In general the feedback is similar in each of the ratings. We do see some extra negativity in the call center.

**Number of Calls per Date**

In this line graph the number of calls per date is displayed. To properly display the dates, we needed to convert the format and take off the time and extract only the date from the column. This chart shows us which days of the month were more busy vs which days are less busy. An interesting thing to study next would be to see if there is a pattern in terms of the days of the weeks.