University of Europe for Applied Sciences

DATA VISUALIZATION

Assignment 1

Presenting team

Sadaf Zuberi

Sepideh Vasseghi

Muhammed Riswan Puthiyoth

Sanjjyoti Sanjay Dutta

Manali Sanjay Budhale

Under the guidance of

Prof. Sukhmani Kaur Bedi

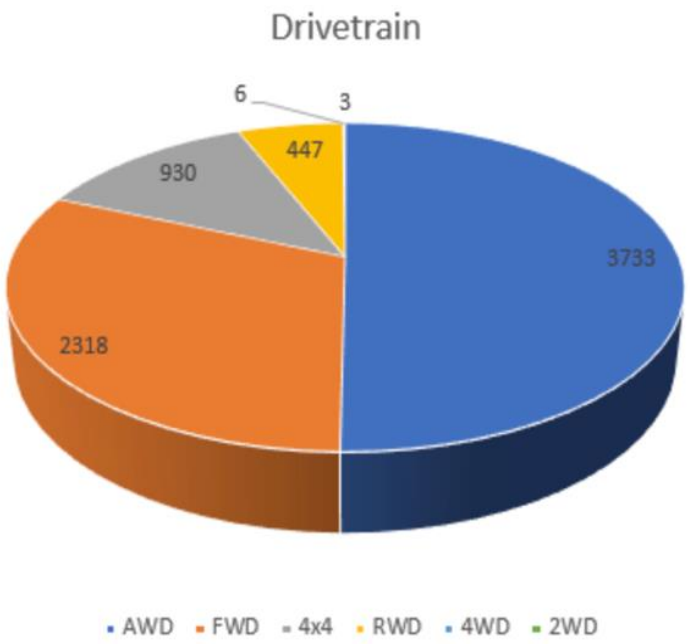
Course

M.Sc. in Data Science

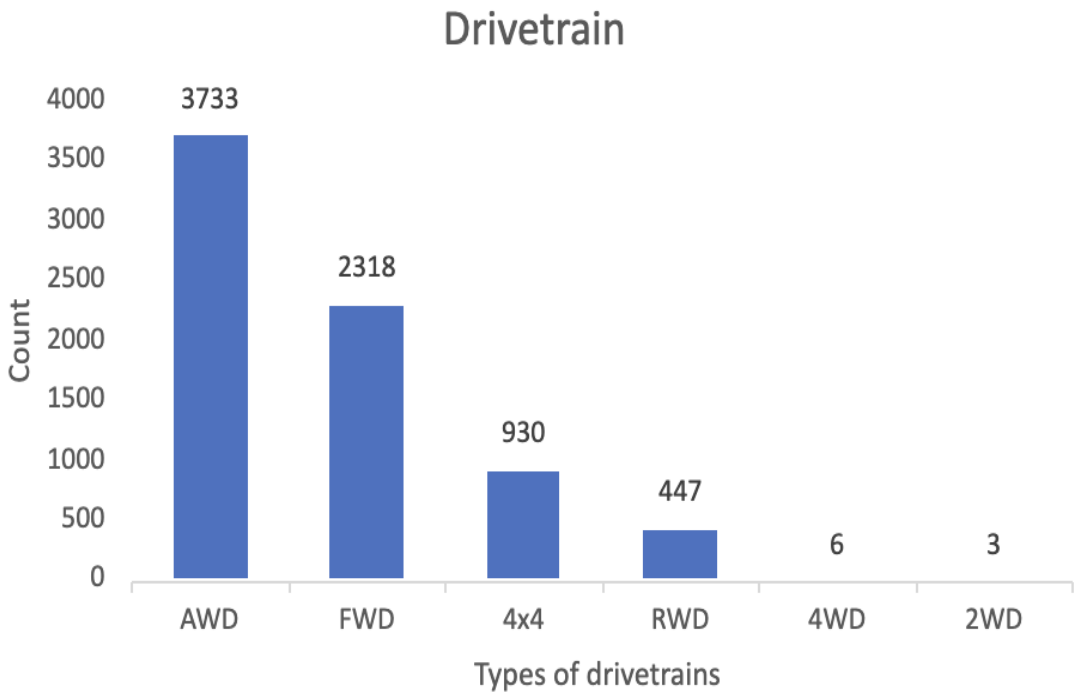
25 April 2023

Data Visualization 1: Distribution of drivetrain axles in used cars

Before:



After:



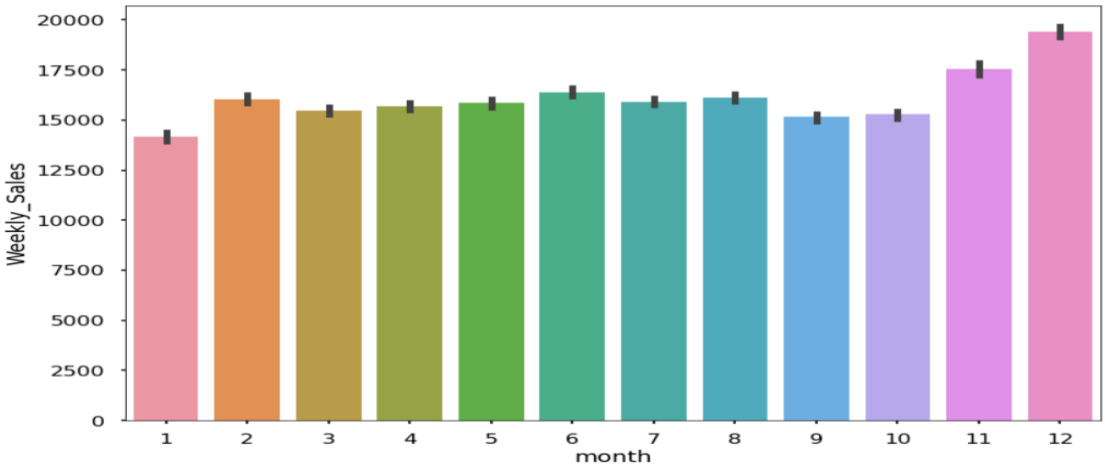
- **Identified Faults & Explanation:**
    1. **No definition**: The chart does not have a clear title, making it difficult for viewers to understand what they are looking at.
    2. **Graph Type**: Wrong graph type has been depicted for showcasing the data, the data in categories are not depicted with true value making it difficult to compare.
    3. **Visibility**: The labels on the pie chart slices are not easily readable, making it tricky for viewers to understand the proportions of the different categories
    4. **Projections**: Unwanted 3D projection of the pie graph making no use in the visualization

- **Fixed Faults & Explanations:**
    1. **Clearly defined title:** Add a clear title to the chart, indicating the purpose and the dataset used.
    2. **Graph Type:** Modified the existing pie chart into a bar graph to make it more visually communicative.
    3. **Increased Visibility:** Labels are presented on a contrasting background to make them more noticeable.
    4. **Deleting 3D projections:** Unwanted 3D projections have been removed.

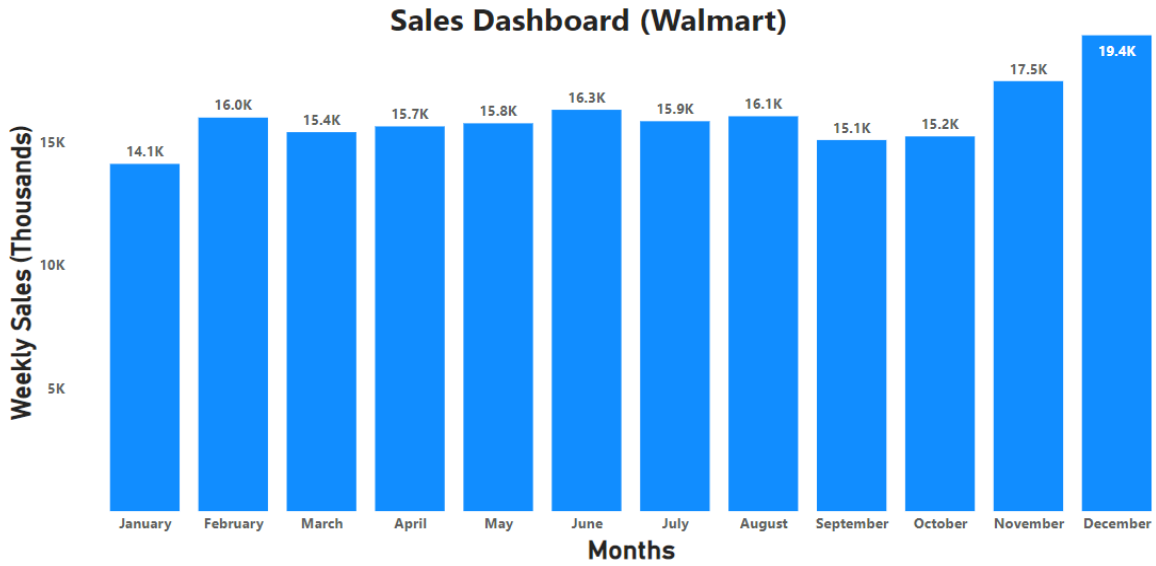
Reference: https://www.kaggle.com/code/omarmaher21/updated-used-vehicles

Data Visualization 2: Weekly sales by month of Walmart stores

Before:



After:



- **Faults Identified & Explained**

1. **Unnecessary colors identified**: In the ‘before’ graph, we can see the multiple colors which do not depict anything. It is only pointless information which will confuse the viewer and cause unnecessary diversions.

2. **The title is missing**: The absence of the heading in this case makes it difficult for a new viewer to comprehend the graph. The title is important because it gives context and is easy for the audience to anticipate by looking at the chart.
3. **More information on the Y-axis**: Too much information was provided on the Y-axis. The excess information on the axis can be reduced, as shown in the 'after' graphic.

- **Faults Fixed** & **Fixes Explained**

1. **Modified the color**: Using only one color in a data visualization can help simplify the visual display and make it easier to interpret for the audience, while also avoiding confusion and distractions.
2. **Heading included**: The inclusion of a clear and informative title to the chart might help the audience understand the message and purpose of the chart.
3. **Modified Y-axis values**: The y-axis had additional data which was modified by adding the values in thousands. It cluttered the extraneous information and gave the graph a clean appearance and precise data.

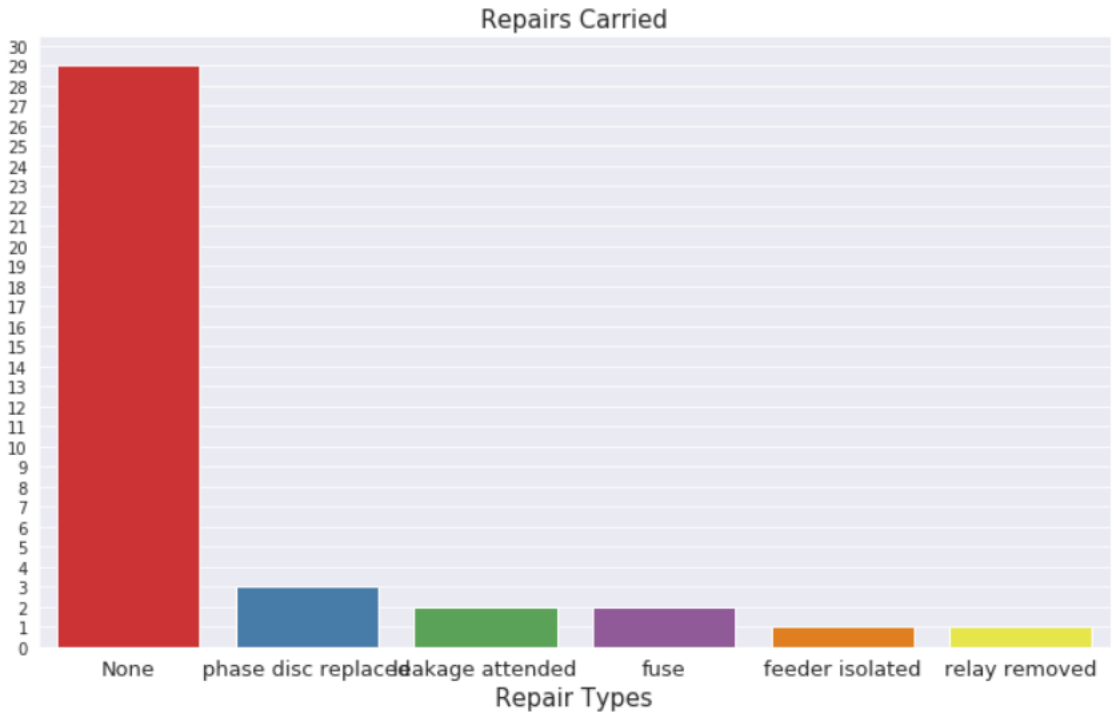
- **Additional work**

1. **Labelling is more appealing and visible**: Labelling is an important part of a data visualization bar chart, making it more appealing and visible can help the audience better understand the data.
2. **The graph is easy to read**: Labelling is essential for making data visualization bar charts more appealing and visible.
3. **A data label is provided for easier comprehension**: Data labels add context to data, allowing the audience to quickly identify and compare values across different categories, leading to a more accurate interpretation of the chart's data.

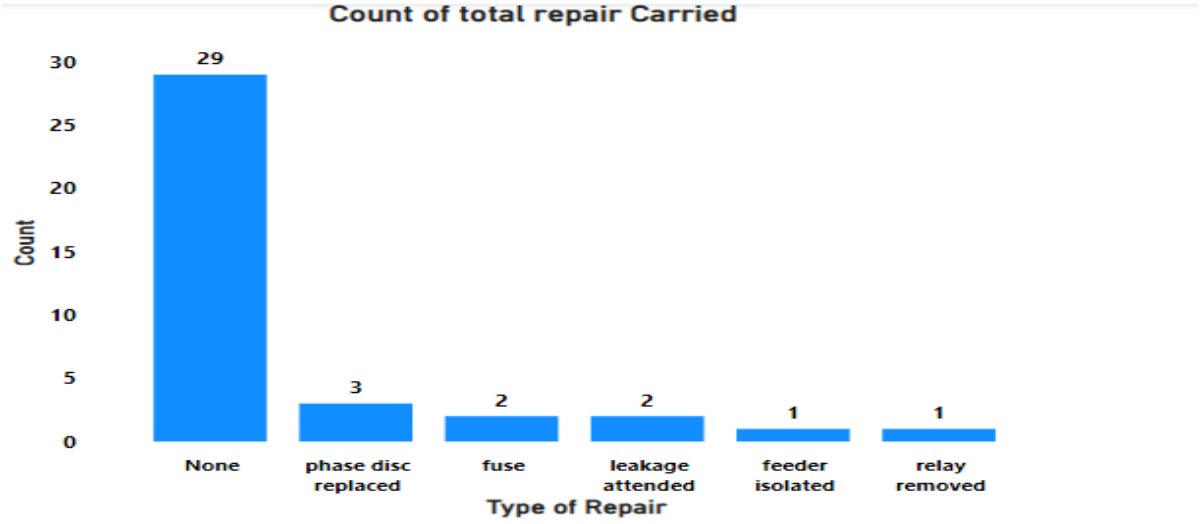
Reference: https://www.kaggle.com/code/aslanahmedov/walmart-sales-forecasting

Data Visualization 3: Type of repairs carried out in electrical analysis

Before:



After:



- **Identified Faults & Explanation**

1. **Interval timings**: The interval counting on Y-axis is crowded and looks unattractive. It makes it difficult for the viewer to read the expected outcome.
2. **Gridlines**: There is no need for gridlines as we can mention the data values above each bar.
3. **Background**: The graph's background makes it less contrasting and difficult to read.
4. **Dynamic coloring**: Different colors have been used on the chart which is an unnecessary development in the chart. Using multiple colors in a bar chart can be confusing and difficult to understand, making it more difficult for viewers.

- **Fixed Faults & Explanation**

1. **Wide interval timings**: The interval timings have been widened so as to keep more space between data labels and make it cleaner and simpler.
2. **Deleted gridlines**: Gridlines have been removed in order for the users to map clearly to their corresponding values.
3. **Background color changed**: The background color here has been changed to white making it more contrasting and pleasing for the reader.
4. **Use of single color**: Using a single color in a data visualization can help simplify the visual display and avoid confusion. Differentiating multiple data sets with different colors may be necessary, but colors that are easily distinguishable and accessible to all viewers are important.

- **Additional changes**

1. **Removal of extra data**: The most important idea is that extra data can be removed from a bar chart to make it more focused on the main information being presented.
2. **Labelling of blank data on X-axis**: Labelling blank or missing data on a bar chart can help the audience better understand the chart and interpret the information.
3. **Text formatting in labels and legends:** They can be made more visually appealing and easier to read by using appealing and consistent formatting such as bolding, color-coding, or font size changes.

Reference: https://www.kaggle.com/code/hashbanger/electrical-failure-analysis