Indian Cars: Exploratory Data Analysis

Around 3.34 Lakh passenger cars were sold in the Indian market in May 2023. The sales increased by over 13% when compared to May last year. The Top 25 Selling Cars constituted over 75% of the cars sold in April 2023.  
  
This dataset consists of 140 columns. Perform Exploratory Data analysis on this dataset. Document the findings and insights using proper graphs to represent the data.

You need to perform Univariate and Bivariate analysis for the given dataset. Below are the steps you can follow for both univariate and bivariate analysis of the dataset.

**Univariate Analysis:**

**1. Understand Dataset:** Get a sense of the dataset structure, including the number of rows and columns.

**2. Data Cleaning:**Handle missing values, if any, by either removing or imputing them based on the nature of the data.

**3. Data Types:**Check the data types of each column. Ensure they are appropriate for the analysis.

**4. Descriptive Statistics:**Calculate basic descriptive statistics such as mean, median, mode, minimum, maximum, and standard deviation for numeric columns.

**5. Histograms:**Create histograms to visualize the distribution of key numeric variables (e.g., sales, price).

**6. Bar Charts:**Plot bar charts to represent the count of categorical variables (e.g., car models, manufacturers).

**7. Box Plots:**Use box plots to identify outliers and understand the distribution of numeric variables.

**8. Pie Charts:**Represent the contribution of different car models or manufacturers using pie charts.

**9. Count Plots:**Visualize the count of observations for categorical variables using count plots.

**Bivariate Analysis:**

**1. Correlation Matrix:**Create a correlation matrix to understand the relationships between numeric variables.

**2. Scatter Plots:**Plot scatter plots to explore the relationship between two numeric variables (e.g., sales vs. price).

**3. Pair Plots:**Use pair plots for a quick overview of relationships between multiple numeric variables.

**4. Bar Charts with Hue:**Enhance bar charts by introducing the 'hue' parameter to represent additional categorical variables.

**5. Heatmaps:**Utilize heatmaps to visually represent the correlation matrix for better insights.

**6. Joint Plots:**Use joint plots to display the distribution of two numeric variables and their relationship.

**7. Box Plots with Hue:**Enhance box plots by introducing the 'hue' parameter to represent additional categorical variables.

**8. Categorical Plots:**Explore relationships between categorical and numeric variables using categorical plots.

**9. Violin Plots:**Visualize the distribution of numeric variables across different categories using violin plots.

**10. Stacked Bar Charts:**Represent the combined contribution of different categories using stacked bar charts.

**11. Insights and Documentation:**Document key findings and insights obtained from both univariate and bivariate analyses. Use graphs and visualizations to support observations.

This approach allows you to explore the dataset thoroughly, gaining insights into individual variables (univariate analysis) and relationships between variables (bivariate analysis). Happy Learning.

Libraries to use for visualization :

1. Seaborn
2. Plotly
3. Matplolib
4. Pandas