**Bright Motor Company Data Analysis**

Data

Bright Motor Company want to analyze the data to get a fair idea about the demand of customers which will help them in enhancing their customer experience. Suppose you are a Data Scientist at the company and the Data Science team has shared some of the key questions that need to be answered. Perform the data analysis to find answers to these questions that will help the company to improve the business.

Data Description

* **Age**: The age of the individual in years.
* **Gender**: The gender of the individual, categorized as male or female.
* **Profession**: The occupation or profession of the individual.
* **Marital\_status**: The marital status of the individual, such as married &, single
* **Education**: The educational qualification of the individual Graduate and Post Graduate
* **No\_of\_Dependents**: The number of dependents (e.g., children, elderly parents) that the individual supports financially.
* **Personal\_loan**: A binary variable indicating whether the individual has taken a personal loan "Yes" or "No"
* **House\_loan**: A binary variable indicating whether the individual has taken a housing loan "Yes" or "No"
* **Partner\_working**: A binary variable indicating whether the individual's partner is employed "Yes" or "No"
* **Salary**: The individual's salary or income.
* **Partner\_salary**: The salary or income of the individual's partner, if applicable.
* **Total\_salary**: The total combined salary of the individual and their partner (if applicable).
* **Price**: The price of a product or service.
* **Make**: The type of automobile

**Basic Steps:**

* + 1. Display the top 5 rows.
    2. Display the last 5 rows
    3. Check the shape of dataset.
    4. Check the datatypes of each feature.
    5. Check the Statistical summary
    6. Check the null values
    7. Check the duplicate values
    8. Check the anomalies or wrong entries.
    9. Check the outliers and their authenticity.
    10. Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.
* **Descriptive Statistics:**
  + What are the mean, median, and standard deviation of the ages of individuals in the dataset?
* **Data Distribution:**
  + What is the distribution of gender in the dataset? Represent it using a pie chart.
* **Correlation Analysis:**
  + Is there a correlation between age and salary? Provide the correlation coefficient and interpret the result.
* **Salary Analysis:**
  + What is the average salary for individuals based on their educational qualifications (Graduate vs. Post Graduate)?
* **Loan Status:**
  + What percentage of individuals have taken a personal loan? How does this compare between males and females?
* **Marital Status and Dependents:**
  + What is the average number of dependents for married individuals versus single individuals?
* **Partner Employment:**
  + How does the employment status of a partner affect the total combined salary?
* **Salary Comparison:**
  + Compare the average salary of individuals whose partners are working versus those whose partners are not working.
* **House Loan Analysis:**
  + What is the proportion of individuals with house loans based on their profession?
* **Salary Distribution:**
  + What is the distribution of salaries for individuals with personal loans versus those without personal loans? Represent it using a box plot.
* **Automobile Make Analysis:**
  + How does the type of automobile relate to the salary of the individuals? Provide insights based on the make of the automobile.
* **Price Analysis:**
  + What is the average price of the product/service in the dataset? How does this price vary based on the individual's total salary?
* **Marital Status and Loans:**
  + Is there a significant difference in the number of personal loans taken by married individuals compared to single individuals?
* **Educational Qualification Impact:**
  + How does educational qualification impact the likelihood of taking a house loan?
* **Dependent Count Analysis:**
  + Analyze the number of dependents based on the profession of the individual. Which profession has the highest average number of dependents?
* **Gender and Salary:**
  + Is there a significant difference in salaries between males and females? Provide statistical evidence.
* **Regression Analysis:**
  + Build a regression model to predict an individual's salary based on age, education, and number of dependents. Discuss the model's accuracy and significance.
* **Loan Status Impact:**
  + How does having a personal loan affect the total combined salary of the individual and their partner?
* **Partner's Salary Contribution:**
  + What is the average partner's salary for individuals with and without house loans?
* **Total Salary Distribution:**
  + Create a histogram showing the distribution of total combined salaries. Identify and discuss any skewness or outliers in the data.