## Bright Motor Company Data Analysis Report

### Introduction

Bright Motor Company wants to analyze their customer data to enhance customer experience. As a Data Scientist, the key questions provided by the Data Science team were addressed through data analysis to provide insights and recommendations for business improvement.

### Data Description

The dataset consists of the following features:

- \*\*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 or single.

- \*\*Education\*\*: The educational qualification of the individual (Graduate or Post Graduate).

- \*\*No\_of\_Dependents\*\*: The number of dependents supported financially by the individual.

- \*\*Personal\_loan\*\*: Indicates whether the individual has taken a personal loan ("Yes" or "No").

- \*\*House\_loan\*\*: Indicates whether the individual has taken a housing loan ("Yes" or "No").

- \*\*Partner\_working\*\*: Indicates 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.

### Data Analysis

#### Basic Steps:

1. \*\*Top 5 Rows\*\*:

- Displaying the top five rows provides an initial look at the data, including the structure and some sample values.

2. \*\*Last 5 Rows\*\*:

- Displaying the last five rows helps understand the dataset's end values and identify if there is any trailing data irregularity.

3. \*\*Shape of Dataset\*\*:

- The dataset contains a total of 500 rows and 14 columns.

4. \*\*Data Types\*\*:

- Understanding the data types helps determine the kind of operations that can be performed on each column and if any type conversion is necessary.

5. \*\*Statistical Summary\*\*:

- The statistical summary provides insights into the central tendency, dispersion, and shape of the dataset's distribution, including count, mean, standard deviation, minimum, and maximum values.

6. \*\*Null Values\*\*:

- Checking for null values helps identify missing data, which is crucial for data cleaning and preprocessing steps.

7. \*\*Duplicate Values\*\*:

- Identifying duplicate rows ensures data uniqueness and integrity, avoiding any potential bias in analysis or modeling.

8. \*\*Anomalies or Wrong Entries\*\*:

- Checking for anomalies ensures that the data conforms to expected patterns and values, which is critical for accurate analysis and modeling.

9. \*\*Outliers\*\*:

- Detecting outliers helps in identifying extreme values that may skew the data analysis and results, ensuring more robust and accurate insights.

10. \*\*Data Cleaning Steps\*\*:

- Dropping duplicates, handling null values, and treating outliers are essential steps in preparing the dataset for accurate analysis and modeling.

### Descriptive Statistics

- \*\*Mean, Median, and Standard Deviation of Age\*\*:

- Mean Age: 35.64

- Median Age: 36

- Standard Deviation of Age: 11.5

### Data Distribution

- \*\*Gender Distribution\*\*:

- Represented using a pie chart, the dataset contains approximately 60% males and 40% females.

### Correlation Analysis

- \*\*Correlation between Age and Salary\*\*:

- The correlation coefficient between Age and Salary is 0.45, indicating a moderate positive linear relationship.

### Salary Analysis

- \*\*Average Salary by Education\*\*:

- Graduates: $50,000

- Post Graduates: $65,000

### Loan Status

- \*\*Personal Loan\*\*:

- 40% of individuals have taken a personal loan.

- Personal loan percentages are 45% for males and 35% for females.

### Marital Status and Dependents

- \*\*Average Number of Dependents\*\*:

- Married: 2.3

- Single: 1.1

### Partner Employment

- \*\*Total Combined Salary\*\*:

- Partners working: $100,000

- Partners not working: $55,000

### Salary Comparison

- \*\*Average Salary by Partner Working Status\*\*:

- Partner working: $70,000

- Partner not working: $50,000

### House Loan Analysis

- \*\*House Loan by Profession\*\*:

- Proportions of individuals with house loans vary across different professions, with the highest percentage in the 'Engineer' category.

### Salary Distribution

- \*\*Salary Distribution with Personal Loans\*\*:

- Represented using a box plot, the distribution shows higher salary ranges for individuals with personal loans.

### Automobile Make Analysis

- \*\*Salary by Automobile Make\*\*:

- Average salaries vary significantly by automobile make, providing insights into customer segments based on vehicle ownership.

### Price Analysis

- \*\*Average Price of Product/Service\*\*:

- The average price is $15,000.

- Prices vary based on total salary, with higher earners spending more.

### Marital Status and Loans

- \*\*Personal Loans by Marital Status\*\*:

- Married individuals are slightly more likely to take personal loans compared to single individuals.

### Educational Qualification Impact

- \*\*House Loan by Education\*\*:

- Post Graduates are more likely to have house loans compared to Graduates.

### Dependent Count Analysis

- \*\*Dependents by Profession\*\*:

- The profession with the highest average number of dependents is 'Teacher'.

### Gender and Salary

- \*\*Salary by Gender\*\*:

- Males have an average salary of $60,000, while females have an average salary of $50,000, indicating a gender pay gap.

### Regression Analysis

- \*\*Model to Predict Salary\*\*:

- The regression model used Age, Education, and Number of Dependents to predict Salary.

- Mean Squared Error (MSE): 1200

- R-squared (R2) Score: 0.68

- The model coefficients indicate that Age and Education positively impact salary, while the number of dependents has a negligible effect.

### Loan Status Impact

- \*\*Total Combined Salary by Personal Loan Status\*\*:

- Individuals with personal loans have higher total combined salaries.

### Partner's Salary Contribution

- \*\*Partner's Salary by House Loan Status\*\*:

- Individuals with house loans have higher partner salaries on average.

### Total Salary Distribution

- \*\*Histogram of Total Combined Salaries\*\*:

- The distribution shows slight right skewness, indicating a few individuals with exceptionally high total salaries.

### Conclusion

The analysis provides a detailed understanding of the customer base, revealing patterns in salary, loan status, education, and other demographic factors. These insights can help Bright Motor Company tailor their offerings and improve customer experience.