Data Analysis & Visualization

Assignment

Description

The Bike buyer's dataset contains the 1000 users from different backgrounds and output variable as whether they buy a bike. Your task is to perform an Exploratory Data Analysis on the dataset. This assignment is part of the continuous assessment and worth of your 30 % module grade.

Dataset

First download the dataset from Moodle's module page. To load the dataset, run the following commands:

```
# read the csv file
bike_buyers.dataset <- read.csv("<full path>/bike_buyers.csv")
```

To get started... see the structure of the data using: `str(bike_buyers.dataset)`

Task

Your task is to perform EDA and calculate the strength of relationships between the variables of the dataset. Consider below as a guideline:

1. Your task is to clean the dataset and prepare it for analysis by e.g. removing/replacing NAs, outliers, and incorrect values. **(20 points)**

2. Begin your analysis with a summary of the variables (use basic statistical methods). Briefly describe your understanding. Prepare plots: pie chart, bar chart, histogram, scatter plot. Each plot should display different variables. Each plot must have a title and meaningful labels. **(20 points)**

3. Focus your analysis on the Purchased.Bike variable: **(20 points)**
   (a) Show the histogram of the Income variable. Describe it briefly. Include summary statistics like mean, median, and variance.
   (b) Group bikers by some Income ranges and summarise those groups separately.
   (c) Explore outliers for Income. You might want to use the boxplot.
   (d) How different attributes are correlated with the Purchased.Bike? Which variables are correlated the most with Purchased.Bike?

4. Create density plots for Income and ggplot comparing Age and Gender. **(10 points)**

5. In your Markdown document, you should use proper headings and commentary for each task. You can get up to **(30 points)** for clarity and quality of the report and the source code. You must show the code blocks and the corresponding output in your knitted document.

Keep in mind the following...

- Acceptable format: Knit your Markdown document in **pdf** output. Use the submission link on Moodle to upload your work. There will be a late submission penalty as per the College policy. I will be very strict on plagiarism. You may want to read Griffith plagiarism policy. I will be awarding ZERO to both the Copyier and Copyee.