**Aster Retail Customers**

This dataset consists of responses from a retail client of Aster Consulting. Pankaj Sinha, Analytics consultant at Aster wants to get some insights from this data to guide marketing strategies of its retail client. This data includes information about customer demographics and purchase behaviour. Data and detailed description of various attributes of data can be found at the end of this document.

Please help Pankaj in performing descriptive analysis of customer dataset by providing suggestions to following questions:

* How should Pankaj classify these attributes of customers into various attribute types?
* Generate frequency chart for Marital Status. There are many categories of Marital status. Can we merge some of the categories to reduce the distinct groups for Marital Status?
* Generate stacked bar chart for Marital Status and Country. Get insights about the percentages of customers with different Marital status across the Countries. Can we say that composition of customers with respect to Marital status changes with respect to country?
* Generate stacked bar chart for Education and Country. Get insights about the percentages of customers with different level of education across the various countries. Can we say that composition of customers based on education changes with respect to country?
* Check if there exists there exists outliers in the data for numerical attributes Income and Age. Also Check if average income of customers exceeds 40,000.
* Check if number of websites, number of store purchases, and number of website purchases meet the assumption of normality. Also find and treat outliers for these variables if any.
* Check if there exists correlation between number of website visits and number of website purchase.
* Check if there exists correlation between number of website visits and number of store purchases.
* Please create new column in data and name it as “Education\_Level”. Populate values for this column in following ways:

If (Education = “Basic”)

Then Education\_Level = “Low”

Otherwise:

Education\_Level= “High”

* Compare if there is difference in terms of website visits between customers with Low and High level of education.
* Compare if there is difference in terms of website purchases between customers with Low and High level of education.
* Can we conclude that meat consumption of customers is higher than Fruits consumption?
* Can we conclude that Fruits consumption of customers is higher than sweets consumption?
* Please visualize average income for different education groups such as Basic, High School, graduation, masters, and PhD. Can we conclude that there is significant difference in the income of customers belonging to different education groups?

**Instructions:**

* For each question, please provide appropriate chart, statistical-test, results, and your conclusion.
* You may use Power BI, Python, or Excel for answering to questions in this assignment.
* Please discuss business implications or relevance for interesting insights which you extract while finding answers to case questions.



Description of Variables:

|  |  |
| --- | --- |
| **Columns** | **Description** |
| ID | Customer ID |
| Age | Age of the customer |
| Education | Customer's level of education |
| Marital\_Status | Customer's marital status |
| Income | Customer's yearly household income |
| MntFruits | amount spent on fruits in the last 2 years |
| MntMeatProducts | amount spent on meat products in the last 2 years |
| MntSweetProducts | amount spent on sweet products in the last 2 years |
| NumWebPurchases | number of purchases made through company's site |
| NumStorePurchases | number of in store purchases |
| NumWebVisitsMonth | number of visits to company's website made in the last month |
| Country | Customer's location |