Bike Store Project

**Introduction**

This project is based on dataset from Company ABC. The company is a store that sells bicycles. The store has a policy in place that ensures data is collected so that insights can be deduced from analyzing collected data from its customers. It contains data about customers that bought or did not buy a bike.

Business data has been collected since the inception of the company. They collect data about sales, whether customers have bought bicycles or not; where the customers live; if the customers are married; how many kids they have; their commute distance to work; and other information per the company’s data policy.

One major benefit of data analysis is the power to derive insights from data. Data from which you would think no decision can be derived from it. But harnessing the power of data analysis can guide you to make informed decisions from the data you analyze. In this project I will analyze the customer data from the bike store and find certain insights that can help the business and provide recommendations.

**Tools**

In this project, I used a spreadsheet application, Microsoft Excel, to analyze and derive my insights. Excel is a powerful tool for analysis. It has loads and lots of analytical functions that help in analyzing data. It has great features that enable visualization. Incredible charts, and Pivot Tables that allow an analyst to group data and analyze them.

I use Microsoft Excel throughout the project.

**Analysis/Insights**

***Data Cleaning***

The data was in an excel sheet formatted file. I went through the data column by column, familiarizing myself with the data before as it is very key for an analyst to be familiar with the data before delving into in-depth analysis. I found blank features during this process. It was up to me to deal with them. I checked for the importance of the blank cells to my whole analysis, and if it would skew or affect my analysis in any way before I made the decision to delete them and replace them with “NULL” values.

That was my first part of the data cleaning process. I checked for duplicate values in the dataset, and this was my second step in cleaning the data. There were a number of duplicate data. There were 151 duplicates present. The salary column for the customers wasn’t standardized. It should be in a “Currency” data type so I fixed that.

I noticed another unseeming problem. The “Marital Status” and “Gender” columns contained single value denotations of their statuses. With M and S for Married and Single respectively, and M and F for Male and Female respectively. It can be confusing so I fixed that by replacing respective vales with the full word.

There wasn’t any column that showed if the customer had children or not. Except for number of children columns. I wanted to find out how many of the customers had children or not so I created a new column where I derived answers for whether they had children based on the number of children they had. I also went ahead to group the ages of the customers into “Old”, “Youth”, and “Middle Age” to see the interest in the various age groups of our customers.

These were the major cleaning steps I took for ensure the data was ready for analysis.

***Insights***

1. How many female, and male customers bought or didn’t buy a bike?
2. How many female, and male customers do we have?
3. How many customers in total?
4. What is the average income of people who bought or didn’t buy a bike?
5. How many customers within the various commute distances bought a bike
6. How many bikes were bought within different age brackets?
7. What is the average age of our customers? Who is the youngest, and oldest customer?
8. How many Single and Married customers bought bikes and didn’t?
9. How likely are Car owners going to buy bikes?
10. What’s the proportion of homeowners who buy bikes?
11. Which occupations bought the most bikes?

Conclusion

Recommendation

**Meta Data**

File format - .xlsx

Total Data - 1027

Total Columns - 13

**Column names**

ID: This column contains unique values used to identify each row

Marital Status: This column contains the marital status of each entry, states whether S or M.

Gender: This column contains the gender description of each entry. M or F.

Income: This value contains the income of the customer. Expressed in currency datatype.

Children: This value contains description of the number of kids each row has.

Education: This value contains description of the educational status of each row.

Occupation: This column contains values with the description of the occupational status.

Home Owner: Contains values about the status of each row, if they own a home.

Cars: This column consists of values that describe the number of cars the customer has

Commute Distance: This column displays about the commute distance of the customer to work.

Region: The region of the customer.

Age: The age of the customer.

Purchased Bike: Whether the customer purchased a bike or not.