**Srilasya Garigipaty**

**Student # 100822953**

**Dashboards and Reporting: The Coffee Cup**

**DATA 1203- Business Analysis an Assessments**

**Date: April 12th, 2021**

**Background of The Coffee Cup and Business Analysis Task**

The Coffee Cup company is a Canadian national Coffee Company that provides a variety of options for coffee enthusiasts and offers a wide variety of drinks and food choices, while providing prompt and courteous service. The Coffee C1up’s mission statement is:

***“To provide the world’s ultimate coffee shop experience with the highest quality products, most-inviting stores, friendliest staff and best value.”***

The Coffee Cup’s goal is to be the neighborhood coffee house, a place where loyalty is earned and where customers go when they desire a freshly brewed cup of hot coffee or a speciality tea selection.

The Coffee Cup **corporates values** is to earn the trust and loyalty from their guests, by showing what they stand for. The following are the values of The Coffee Cup:

**Being Real-** Staff will bring their authentic and unique self to every situation and opportunity to serve one another and guests better than anyone else.

**Be Exceptional-**Staff will challenge themselves and others around them to exceed expectations.

**Be Passionate-** Staff will service for our guests.

The Coffee Cup **core values** is to earn the trust and loyalty from their guests, by showing what they stand for. The following are the **core values** of The Coffee Cup:

**PASSION FOR PEOPLE** - Respecting one another and creating a harmonious Bon environment.

**PASSION FOR SERVICE** - Delivering our uncompromising Integrity, Quality, and Service with a Smile.

**PASSION FOR KNOWLEDGE** - Always staying relevant and ahead of the rest.

**PASSION FOR BUSINESS** - Providing the best in everything Bon.

The **Key Performance Metrics** for The Coffee Cup include:

**Business Performance Metrics**-Profitability, Sales Productivity, Gross Margin, Customer Loyalty, Cost metrics

**Sales Performance Metrics**-Store Performance, Sale Market Analysis, Product Sale Analysis, Product Inventory

**Data Analysis**

**Analysis of the raw data and the Quality of the Data**

The Coffee Cup Data has a total of 4248 rows and 20 columns. The columns include: “Profit”, “Margin”, “Sales”,” Cost of Goods Sold”, “Total Expenses”,” Marketing”,” Inventory”, “Budget Profit”, “Budget Margin”, “Budget Sales”, “Budget COGS”, “Date', 'Market”, “Province”, “City”, “Store#”, “Market Size”, “Product Type”, “Product”, “Type”. There are no missing values in the dataset. It was noted that there was 32 values of 0 in “Profit” column, 66 values of 0 in “Cost of Goods Sold”,66 values of 0 in “Marketing” ,152 values of 0 in “Budget Profit”, 2 values of 0 in “Budget Margin”, 24 values of 0 in “Budget Sales”, 82 values of 0 in “Budget COGS”. Further evaluation is needed to determine why these feature variables in dataset have received a 0. The featured column variables have a wide range of values, and many values are above and below the minimum and maximum bounds. These bounds are calculated as follows:

* **Median**: The median is the mean of the middle two numbers. it is the middle value of the dataset.
* **first quartile (Q1):** The first quartile is the median of the data points to the left of the median. It is also known as 25th.
* **third quartile (Q3):**The third quartile is the median of the data points to the right of the median.
* interquartile range (IQR): Q3 – Q1 or data between 25th percentile to 75th percentile.
* **minimum:** Q1 -1.5 \* IQR (Excluding outliers)
* **maximum:** Q3 + 1.5 \* IQR (Excluding outliers)

In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center. The “Budget Margin”,” Budget Profit”,” Inventory” and “Profit” columns are normally distributed based on histograms plotted on all the features in dataset.

**Quality of the Data**

The Coffee Cup Data is adequate quality. The different dimensions of data quality assessment include completeness, validity, value, accuracy, consistency, volume, and variety of the data, and whether qualitative and quantitative measures be assessed related to the stakeholder’s metrics. The data is not guaranteed to be accurate as there might be mistakes in data recording. All values are in the same formats and data types for each column. There is no duplicate data, but the store #’s are not unique, and the data is consistent throughout. There are no missing values in the data, and it is complete. Since the Coffee Cup is a Canadian National Coffee Chain, it is important that the data includes a variety of cities and provinces. The data has variety and includes various cities in different provinces and their metrics including the type of market in the area. The data also consists of the year 2018 and 2019 data which gives more opportunity to measure trends and get insights. However, it was noted that the data should only include the year 2019 and needs to further be cleaned. The data has value because it is able to provide insights related to the key performance metrics related to business and sales. There is a large volume of data to get insight from, as the data file contains 4248 records with different measures. The data is valid data because it comes from The Coffee Cup company to Data Analyst to help the company facilitate effective decision making. There are quite a few zero values in the data, and analysts need to confirm validity and further information about what caused a zero value in those fields.

**Recommendations to Improve the Data**

Recommendations to improve the data include making additional columns that include a month, year, and day of month. The data can be grouped quarterly, half-yearly and annually. The data includes 2018 as well, and since The Coffee Cup has asked for only the year 2019 data these values need to be cleaned out. Values can be sorted from highest to lowest, or lowest to highest. This will help determine insights into patterns that caused higher than average or lower than average values. The units of the metrics that are unclear can be clarified. For example, marketing only includes number values, but the analysts would be unsure if the number were referring to currency, a time frame, etc. In addition, expenses categories do not have units. It is not clear whether numeric is referring to millions, billions, thousands, etc. Since there is both numerical and categorical data, for data analysis the categorical variables can be label encoded to numerical values in additional columns for analysis in machine learning tools such as R and Python. Based on business metrics and correlation analysis, the data that has little correlation to feature measure company wants to address can be removed to simplify the data for analysis. The data accuracy can be measured by implementing different machine learning algorithms and splitting the data into a modeling set and a testing set. Data can be filtered to include information that is relevant to various parts of business operations. For example, Marketing and Sales, Budget Profits, and Inventory, etc. The data should be appended and analyzed with normalization techniques when there are outdated, incorrect, or duplicated data. If situations where data is outdated or incorrect is found it should be removed, replaced with correct value, or appended with an assumption based on analysis that would most likely have occurred in the given situation.

**List of Questions the Coffee Cup Data Answers for Business and Sales Metrics**

* What was the profit for the years 2018 and 2019 by Province?
* Which province had the highest profit?
* Which province had the lowest profit?
* What city brought in the highest number of sales?
* Which product type had the highest percentage of sales?
* Which market region had the highest expenses?
* Which market region has the highest budget for Cost of Goods Sold?
* Which market region has the lowest budget for Cost of Goods Sold?
* Which market region has the highest cost of goods sold?
* Which market region has the lowest cost of goods sold?

**List of Questions the Coffee Cup Data Does Not Answers and Suggested Data to Answer for Business and Sales Metrics**

* What market capital is the criteria for the market size to be divided into small or major? The data needed to answer this dollar amount of revenue to classify the market size.
* What was the spending of expenses by category for business operations by month and year? The spending expense for various departments in business operations.
* What is the cost of each product for the customer and which product produced the largest sales based on individual cost? The data needed is the price of each product.
* What was the volume of customers within each store number in various cities? The data needed is the volume of customers in each store.
* What was the number of Marketing Campaigns run per quarter per province? The data needed is the number of marketing campaigns in each store.
* The Inventory by Product Type? The data needed is the inventory of each product type for each store location.
* The cost of the Inventory by Product Type? The data needed is the cost of inventory for each store location.
* For areas of low profit, were there any local competitors of coffee shops near the area? The data needed is the number of coffee shops surrounding the store location.
* Hours of the day, and the number of times product type is sold. The data needed is the hourly sales by product type.
* What are the sales for drive-thru, pick up and in-store and what is the number of sales for each product per store? The data needed is the sales for each product in each store for drive thru, pick up, and in store.

Map

Description automatically generated with medium confidence**Interpretation Instructions for Dashboard The Coffee Cup:**

**Top Left Graph Number of Sales by City:** This bar graph display the total sales of all product types by each city in the data set. The bars are displayed in order from highest to lowest with x axis representing sales in thousands, and y- axis the city name. Highest sales city was Kelowna with 77,134 sales, and the lowest sales city was Bathurst with 5,498.

**Top Graph Profit by Provinces:** This geographical map graph displays the profits in various provinces. It is a geographical map of Canada and the shading of blue represents the region in Canada with the highest profit. It can be noted that the larger amounts of profits are earned in the far east regions and far west regions of the country.

**Bottom Left Graph Product Inventory by Market Region:** This is a stacked bar graph that is color coded by market region. Dark green represents Atlantic region, light green represents central region, orange is the north region, yellow is the prairie region, and blue is the western region. The different products are listed on the y-axis and the amount of inventory is listed on the x-axis in thousands.

**Bottom Right Graph Percentage Sales by Product Type:** This donut graph is color coded by product type. Dark blue is coffee, Orange is Espresso, Pink is herbal Tea, and Green is regular tea. Percentages of sales are displayed for each product type.

Chart, scatter chart

Description automatically generated**Interpretation Instructions for Dashboard The Coffee Cup-Dashboard:**

**Left-side Chart Total Inventory Used by Province:** This graph displays the inventory in thousands for each province in Canada. The x-axis displays the province names, the y-axis the inventory in thousands, and the darker the bubble in the chart the larger the inventory is used in that specific province.

**Right-side Chart Total Inventory Used by Market Region:** This graph displays the Budget Cost of Goods Sold in blue bubbles and Cost of Goods Sold in Orange. The market region name is located on the x-axis, and a dual- axis on each side represents the budget cost of goods sold on left side, and cost of goods sold on right side.