## AI-Powered Data Analysis & Automation

This report will be walking through the steps of the data analysis for the data set provided, In this project, Google Sheets, Google AutoML, Power BI, Python and SQL were all used to clean the dataset, analyse data, generate insights and visualizations and provide suggestions and enhance business decision-making.

**Data Cleaning**

The data provided was not analytically robust, as it had missing values and incorrect standardizations. Along with fixing the issues, duplicate values and outliers were also a concern in this dataset. This data set was cleaned using Google sheets and Google AutoML

In the “Date” column of the dataset, there were missing dates. As an average date could not be calculated and putting “0” or “null” in place of the missing values may have lead to the data being misrepresented, the rows containing the missing dates were deleted from the dataset.

A screenshot of a computer program

AI-generated content may be incorrect.FIGURE 1:

The dataset contained missing values in the **“Income”**, **“Credit\_Score”**, and **“Loan\_Amount”** fields. To address this, the dataset was imported into **Google AutoML**, where an **SQL query** was applied to systematically handle the missing data. As detailed in **Figure 1**, missing values in the **“Income”** column were imported using the column’s **average value**, a method that helps maintain the overall distribution of the data without introducing significant bias. Likewise, missing entries in the **“Credit\_Score”** and **“Loan\_Amount”** columns were replaced with their **average values**. This ensures the dataset remains complete and reliable for analysis, providing a reliable foundation for training machine learning models within Google AutoML.

A screenshot of a computer code

AI-generated content may be incorrect.FIGURE 2:

A screenshot of a computer

AI-generated content may be incorrect.FIGURE 3:

The SQL query shown in Figure 2 was applied to pick out outliers within the dataset, looking at the “Income”, “Credit\_Score” and “Loan\_Amount” columns. The query used would return “1” if there were outliers and “0” if there were no outliers found. To be considered an outlier, the value had to have been more than 3 standard deviations away from the average, which is why the query also calculated the averages and standard deviations of each column. The results of this query returned 0 in all rows, meaning that no outliers were detected. This can be seen in Figure 3, which shows a snapshot of the dataset. This query was run for all columns in the dataset, and there were no outliers detected.

A screenshot of a computer

AI-generated content may be incorrect.FIGURE 4:

To ensure that the dataset was reliable for analysis, the query shown in Figure 4 filtered out any duplicates in the dataset that would skew any results. The query returned with no results, meaning that there were no duplicates in the set.

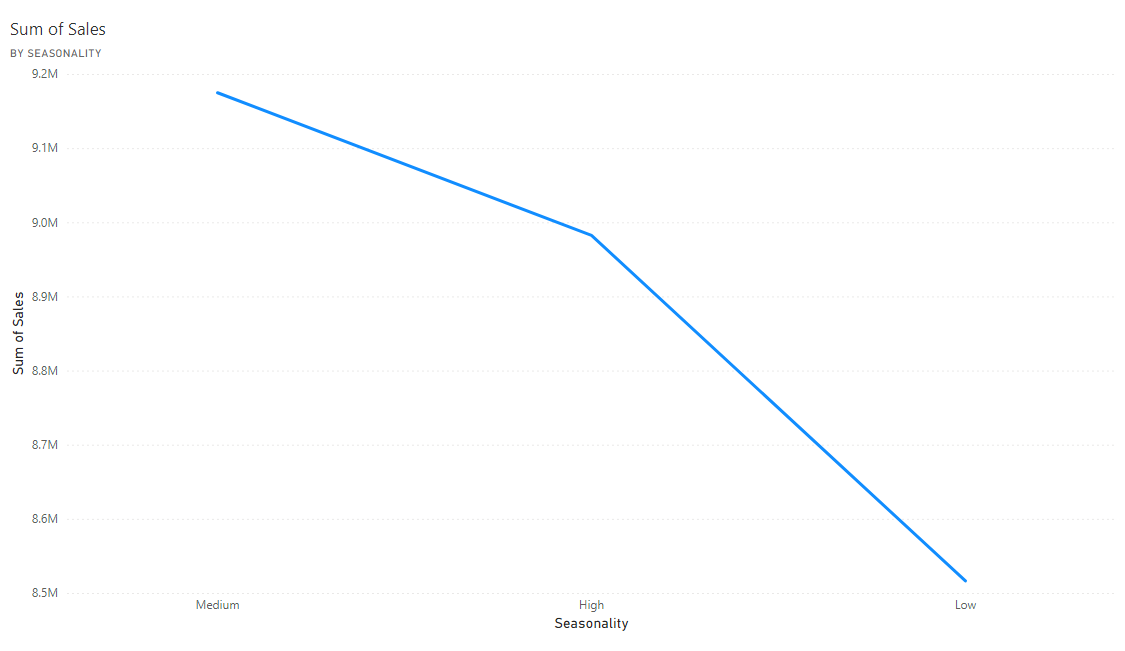
**AI-Powered Data Visualization and Story Telling**

Using Power Bi’s AI features, data visualizations and insights were generated and analyzed. This section looks at trends over time, correlations between variables and any anomalies in the data.

A screen shot of a graph

AI-generated content may be incorrect.Figure 5:

Figure 5 presents a scatter plot showing the relationship between Marketing Spend and

[](https://app.powerbi.com/MobileRedirect.html?action=OpenReport&groupObjectId=3d8f51f7-2266-461c-bddd-2f4cc64eaa0d&reportObjectId=de72a0de-c08e-4519-a0a6-21917f3dc19b&ctid=35d97e90-1583-45fe-b6e9-8b15ae68ef71&reportPage=f4554d140517e0ee3312&pbi_source=copyvisualimage)Sales. The visualization does not show a strong or clear correlation between the two variables, nor does it indicate the presence of any significant outliers. Both axes are scaled in thousands, reflecting the magnitude of the respective measures. Based on the current data, there is no evident linear association suggesting that fluctuations in Marketing Spend directly influence changes in sales performance.

.Figure 6:

|  |
| --- |
|  |
|  |

The graph in figure 6 depicts the sum of sales based on seasonality. The graph’s x-axis represents the seasonality in Low, Medium and High. The y-axis represents the sum of sales in millions. The graph shows a clear positive correlation between seasonality and sales. As seasonality increases from low to high, so do sales. Therefore, sales increase with seasonality. A line graph effectively and clearly illustrates the trend between seasonality and sales. A graph with a line

AI-generated content may be incorrect.Figure 7:

Figure 7 shows a steady increase in purchasing frequency from January to April. There is a significant jump in purchasing frequency from January to April. So, as the year goes on, sales increase. The graph shows a decrease in sales in February; this could be due to seasonal trends in January. Sales then begin to pick up again from March and then they surge in April.

**AI-Driven Predictive and Prescriptive Analytics**

A screenshot of a computer program

AI-generated content may be incorrect.Figure 8:

The python script in Figure 8 is used to train an AI model to predict business trends. The code performs Linear Regression to predict sales based on marketing spend and seasonality. The output of the python script is the Mean Squared Error of the model, which is a measure of how well the model fits the data.

**Summary Report**

The dataset provided had undergone comprehensive pre-processing for anomalies, missing values and duplicates, Missing entries were identified in the “Date”, “Credit\_Score”, “Loan\_Amount” and “Income” columns. The issue with the missing values was solved with mean and average imputation as the missing values were filled in with each column’s average value. No other significant data integrity issues that could distort the set were found.

Insights and Analysis:

Figure 6 illustrates a positive correlation between seasonality and sales. As seasonality increases, so do sales. In figure 7, sales in January take a dip and climb again in March, then there is a surge in April. Seasonality seems to play a big role in the sales of the company, so a suggestion for them would be to focus on seasonal promotions during their low months. To mitigate seasonal fluctuations, the company should implement targeted promotions for low-performing months. Introducing time-limited discounts or marketing campaigns during these months would smooth out declines and improve the overall revenue.