Mini Project Report

On

# “S-MART DATA ANALYSIS”

Submitted in the partial fulfillment for the award of the Degree of Bachelor of Technology

In

Electronics and Communication Engineering

By

# Jilakara Meghana (22030123)

B. Tech, V Semester

Under the guidance of

# Dr.Pabira Kumar Sethy

## Associate Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING SCHOOL OF STUDIES IN ENGINEERING AND TECHNOLOGY

GURU GHASIDAS VISHWAVIDYALAYA, BILASPUR (C.G.)

**SESSION: 2024-25**

# ACKNOWLEDGEMENT

We are highly indebted to **Dr. Pabira Kumar Sethy** for his guidance and constant supervision as well as for providing necessary information regarding the project and also for his support in the project. We owe our special thanks to our Head of Department **Dr. Soma Das** Professor, for encouraging us to acquire courage and knowledge through this project. We would like to express our gratitude towards our parents and faculty members of the Department of Electronics and Communication Engineering, School of Studies in Engineering & Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur CG for their kind co-operation and encouragement which helped us in the completion of this project. We would like to express our special gratitude and thanks to all the staffs of Electronics and Communication Engineering department for giving us such attention and time. Our thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

## Jilakara Meghana (22030123)

## 

## 

# DECLARATION

We the undersigned solemnly declare that this report of the minor project work, entitled “**S-MART DATA ANALYSIS**” is carried out during the course of our study during V semester under the guidance of  **Dr. Pabira Kumar Sethy**, Associate Professor, Department of Electronics and Communication Engineering, School of Studies in Engineering & Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur (C.G.). We further declare that this mini project work is presented for the partial fulfillment of the requirement of degree of Bachelor of Technology in Electronics & Communication Engineering, School of Studies in Engineering & Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur (C.G.).

Date:

Jilakara Meghana

**APPROVAL SHEET**

This mini project report entitled “**S-MART DATA ANALYSIS**” by **Jilakara Meghana**, is approved for the partial fulfillment of the requirement of the degree of Bachelor of Technology in Electronics and Communication Engineering.

## Dr. Pabira Kumar Sethy

(Associate Professor)

Examiners:

Date: .................

Place: Bilaspur

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

# SCHOOL OF STUDIES IN ENGINEERING AND TECHNOLOGY GURU GHASIDAS VISHWAVIDYALAYA, BILASPUR (C.G.)

(A Central University established by the Central University Act 2009 No. 25 of 2009)



**CERTIFICATE**

It is certified that the mini project entitled **“S-MART DATA ANALYSIS”** submitted by **Jilakara Meghana**, in partial fulfillment of the requirements of the award of the degree of Bachelor of Technology in Electronics and Communication Engineering, School of studies in Engineering and Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur, is carried out by them in the Department of Electronics and Communication Engineering during session 2024-25 under supervision and guidance of **Dr. Pabira Kumar Sethy**, Associate Professor, Department of Electronics & Communication Engineering, School of Studies in Engineering & Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur CG.

## Dr. Soma Das

Head of Department Department of Electronics & Communication Engineering School of Studies in Engineering & Technology

Guru Ghasidas Vishwavidyalaya, Bilaspur CG

**ABSTRACT**

This project, "S-MART DATA ANALYSIS," aims to analyze and improve the operations of a local mart by leveraging data collected from customers, feedback from classmates, and insights from staff interviews. The primary goal of the project is to identify trends and patterns that can help optimize product offerings, enhance customer satisfaction, and improve overall business performance. The data was gathered through surveys and interviews, then cleaned using MySQL to ensure consistency and accuracy. Data visualizations were created using Excel and Tableau to identify key trends, such as customer preferences and sales performance. Further analysis was conducted using Python and Pandas to uncover deeper insights into customer behavior and product sales. Interviews with staff provided additional context, helping to refine the analysis and validate findings. Based on the analysis, actionable recommendations were made, including suggestions for improved product placement, inventory management, and targeted marketing strategies. The project demonstrates how data-driven decision-making can significantly enhance the operational efficiency and customer experience at S-MART.

**Contents**

**ABSTRACT**

**Chapter: 1 Introduction** 1

* 1. Project Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
  2. Scope . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
  3. Market . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

**Chapter: 2** **Problem Statement**

* 1. Problem Statement . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

**Chapter: 3** **Project Planning**

* 1. Project Goals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
  2. Timelines . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

**Chapter:4 Methodology**

* 1. Identifying the Problem Statement. . . . . . . . . . . . . . . . . . . . . . . . . . . 6
  2. Data Collection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .6
  3. Data Cleaning . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
  4. Data Analysing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  5. Data Visualization. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9
  6. Presentation of Analysis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

**Chapter:5 Implementation of Analysis**

5.1 Implementation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

* 1. Final Results of Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
  2. Final Insights and Suggestions . . . . . . . . . . . . . . . . . . . . . . . . . . . `19

**Conclusion** 20

**Resources** 21

**APENDIX 1**

**APENDIX 2**

**List of Figures**

Figure 1.1 : Market of Data Analysis

Figure 5.1 : Data Collected from Customer Feedback

Figure 5.2 : Data Cleaned using MySQL Workbench

Figure 5.3 : Products Sold Per Day Table

Figure 5.4 : Products Sold Per Day Graph

Figure 5.5 : Total Payments Table

Figure 5.6 : Total Payments Graph

Figure 5.7 : Sales Per Day Table

Figure 5.8: Sales Per Day Graph

Figure 5.8 : Products Purchased on basis of Gender, Age Table

Figure 5.9 : Products Purchased on basis of Gender, Age Graph

Figure 5.10 : Discount Applied for Type of Visitors Graph

Figure 5.11: Rating Provided Table

Figure 5.12: Rating Provided Graph

Figure 5.13: Products consumed Table

Figure 5.14: Products consumed Graph

S-MART DATA ANALYSIS

1. **INTRODUCTION**

S-MART is a local mart in Bilaspur, that serves as a vital retail hub for the community, providing a wide range of products to its customers. In today’s data-driven world, businesses like S-MART can greatly benefit from leveraging data to improve their operations, enhance customer satisfaction, and make informed decisions. This project, "S-MART Data Analysis," focuses on applying data analysis techniques to optimize the mart’s performance. The objective of the project is to analyze customer feedback, sales data, and insights gathered from staff to uncover trends, identify areas for improvement, and suggest actionable strategies to enhance both the customer experience and business operations.

Data was collected through customer surveys, feedback forms, and staff interviews to understand various aspects of the mart’s offerings, including product preferences, purchasing behaviors, and overall satisfaction. The collected data was then cleaned and processed using MySQL to ensure accuracy and consistency. Visualization tools like Excel and Tableau were used to create insightful charts and dashboards, while Python and Pandas were employed for deeper statistical analysis. Additionally, interviews with the staff provided valuable context and insights, offering a unique perspective on the daily operations and challenges faced at S-MART

**1.1 Project Overview**

The "S-MART Data Analysis" project focuses on utilizing data analytics to improve the operations and customer experience at S-MART, a local retail mart. The goal of this project is to explore customer behavior, feedback, and sales trends in order to identify areas for improvement and provide actionable insights that can guide the mart’s decision-making process.

To achieve this, the project involves a multi-step approach:

**Data Collection**: Data was gathered from various sources, including customer surveys, feedback from classmates, and interviews with staff at S-MART. This provided a comprehensive understanding of customer preferences, product performance, and operational challenges.

**Data Cleaning**: Using MySQL, the raw data was cleaned and pre-processed to ensure it was accurate, consistent, and ready for analysis. This step included handling missing values, removing duplicates, and normalizing data for effective analysis.

**Data Visualization**: Visualizations were created using Excel and Tableau to explore patterns and trends in customer behavior, product sales, and feedback. Dashboards, charts, and graphs helped provide a clear and intuitive view of the data.

**Data Analysis**: Using Python and Pandas, the cleaned data was further analyzed to uncover deeper insights. This included identifying correlations, seasonal trends, and patterns in customer preferences and sales.

1

**Staff Interviews**: Insights were gathered through interviews with the mart’s staff to complement the data analysis. Staff feedback helped validate findings and provided additional context, such as operational challenges and customer service observations.

**Recommendations**: Based on the analysis, actionable recommendations were provided, covering areas such as product placement, inventory management, customer engagement strategies, and targeted marketing campaigns.

The project showcases the power of data-driven decision-making and its potential to enhance business operations. By leveraging both quantitative data and qualitative insights, the project provides valuable suggestions for optimizing S-MART's performance, improving customer satisfaction, and boosting sales.

.

**1.2 Scope**

## As we enter 2025, the future of data analytics in India appears exceptionally promising and pivotal. Data analytics has swiftly become a cornerstone of decision-making across diverse sectors, from healthcare to finance, retail to telecommunications. This transformative power stems from its ability to uncover valuable insights from vast data, enabling businesses to optimise processes, improve customer experiences, and drive innovation.

## Data analytics is a field that examines data to identify patterns and trends to help businesses make decisions. It has a wide scope, and is used in many industries, including:

## Healthcare : Data analytics can help improve patient care, personalize treatment plans, and predict disease outbreaks.

## Finance : Data analytics can help financial institutions detect fraud, assess risk, and manage investments.

## Retail : Data analytics can help retailers understand customer behavior, predict trends, and recommend new products.

## Manufacturing : Data analytics can help manufacturing sectors discover new ways to save costs, solve supply chain issues, and optimize equipment.

## Marketing : Data analytics can help marketers understand consumer preferences, measure campaign success, and target specific audiences.

## Education : Data analytics can help track student performance, personalize learning, and improve outcomes.

## Sports : Data analytics can help teams and coaches improve performance, plan strategies, and monitor player statistics.

## 2

## Market

The data analytics market size was valued at USD 41.05 billion in 2022 and is projected to grow from USD 51.55 billion in 2023 to USD 279.31 billion by 2030, exhibiting a CAGR of 27.3% during 2023-2030. North America dominated the global market with a share of 34.68% in 2022. ccording to a study carried out by an American experience management company, Qualtrics and ServiceNow, around 83% of customers are willing to switch to a brand with a better digital experience, and 70% of customers are more likely to trust brands that provide a great digital experience.

Source: https://www.fortunebusinessinsights.com/data-analytics-market-108882

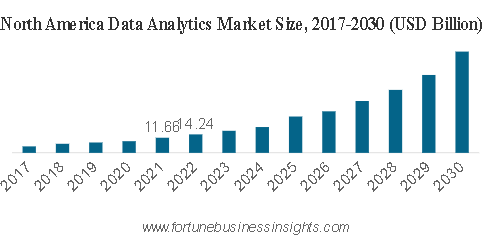


Figure 1.1 : Market of Data Analysis

3

1. **PROBLEM STATEMENT**

The S-MART local mart faces challenges in optimizing its inventory, improving customer satisfaction, and enhancing overall sales strategies due to a lack of comprehensive analysis of customer feedback, purchasing patterns, and sales trends. The absence of actionable insights from the data collected leaves opportunities for operational inefficiencies and missed chances to better cater to customer needs. To address these issues, the mart needs to leverage data-driven analysis to identify key trends, optimize product offerings, and improve customer engagement. This project aims to use data collected from various sources, including customer feedback, staff interviews, and sales data, to provide actionable insights that can enhance the mart's business strategies and decision-making.

4

1. **PROJECT PLANNING**

**3.1 Project Goals**

The main Project Goals include Complete data collection and preparation for analysis, Cleaned data ready for visualization, Initial visualizations in Excel/Tableau complete, Complete Python analysis and data insights, Combined analysis results with staff feedback, Complete draft of report, Final PPT presentation ready, Report submition, presentation

3.2 **Time Lines**

For **S-MART Data Analysis** project, effective project planning and time management will be essential. Below is a structured project plan along with timelines. This plan is considered that with 8 weeks to complete the project.

Breakdown of Activities and Timelines

**Week 1 (Project Initiation & Data Collection)**

* Day 1–2: Define project scope, goals, set up management tools.
* Day 3–7: Collect data via surveys, interviews, and gather available sales data from the mart.

**Week 2 (Data Cleaning)**

* Day 8–13: Import and clean the data using MySQL. Focus on identifying missing values, cleaning duplicates, and normalizing the data.

**Week 3 (Data Visualization)**

* Day 14–21: Set up Excel and Tableau for visualizations, and create initial charts and graphs.
* Day 22–24: Customize Tableau visualizations and ensure interactivity.

**Week 4 (Data Analysis with Python & Pandas)**

* Day 25–30: Perform deeper analysis using Python and Pandas, focusing on statistical analysis, trends, and patterns.

**Week 5 (Integrating Findings & Staff Feedback)**

* Day 31–35: Combine visualizations and data analysis results. Meet with mart staff to review findings and gather feedback.

**Week 6 (Reporting & Documentation)**

* Day 36–45: Draft the report, including the introduction, methodology, data analysis, results, and conclusions. Add the technical appendix and code snippets.

**Week 7 (Presentation Preparation)**

* Day 46–50: Create and finalize the PowerPoint presentation. Rehearse the presentation to ensure a smooth delivery.

**Week 8 (Final Review & Submission)**

* Day 51–53: Review and edit the final report.
* Day 54: Finalize the PPT presentation.
* Day 55: Submit the report and prepare for the presentation
* Day 56: Deliver the final presentation to stakeholders and mart staff

5

1. **METHODOLOGY**

### **Identification of Problem Statement in Data Analysis**

The **problem statement** is a crucial element in any data analysis project, as it defines the scope and purpose of the analysis. Identifying a clear and well-defined problem is the first step in ensuring that the analysis will be focused, effective, and actionable. In the context of data analysis, the problem statement should address a specific issue, question, or challenge that requires data-driven insights to solve or better understand.

### **Data Collection in Data Analysis**

Data can be categorized into two main types: primary and secondary. Primary data is collected directly from the source and can be tailored to specific research needs, while secondary data is already collected by others and is available for reuse. Primary data can further be divided into quantitative and qualitative types. Quantitative data consists of numerical information that can be measured and analyzed statistically, such as sales figures or number of customers. Qualitative data, on the other hand, is non-numerical and typically used to gain deeper insights, like customer feedback or employee opinions. Secondary data can come from various sources, such as public datasets, market reports, or previous research, and can be valuable if it aligns with the objectives of the project.In this project, the project is collected using the Survey of customer feedback, Mart staff interview’s.

Data Cleaning using MySQL

The data cleaning process in MySQL is an essential step to ensure the accuracy, consistency, and usability of the data before it is used for analysis. The goal of data cleaning is to remove or correct any errors, inconsistencies, and missing values that may distort insights and lead to inaccurate conclusions. This process involves several steps, including handling missing data, removing duplicates, correcting data types, handling outliers, and standardizing values. MySQL provides a variety of functions and techniques to perform these tasks efficiently.

The Steps included in the Data Cleaning are as follows :

1. Remove Duplicate Data

Problem: Duplicate records can distort analysis and lead to incorrect conclusions.

Steps:

* + Identify duplicates by checking for repeated rows based on key columns.
  + Use SQL commands like GROUP BY or COUNT to find duplicates.
  + Remove duplicates using DELETE or DISTINCT statements to ensure that only unique records remain.

2. Handle Missing Data

Problem: Missing values can cause errors in analysis, especially when performing statistical operations 6

Steps:

* + Identify missing data by checking for NULL values or blank fields.
  + Decide how to handle missing data: fill with a default value (mean, median, mode), or remove rows/columns with missing values using UPDATE, DELETE, or IS NULL conditions.
  + For important fields, consider collecting additional data if possible.

3. Correct Data Types

Problem: Incorrect data types (e.g., numbers stored as strings) can affect calculations and analysis.

Steps:

* + Identify columns with incorrect data types by reviewing data types in the schema or inspecting sample data.
  + Use functions like CAST() or CONVERT() in MySQL to change data types.
  + For date fields, use STR\_TO\_DATE() to ensure that date values are in the correct format.

4. Standardize Data

Problem: Inconsistent values (e.g., variations in categories like "Yes", "Y", "yes") can complicate analysis.

Steps:

* + Identify inconsistent categories or values across the dataset.
  + Standardize these values using the UPDATE statement to ensure uniformity (e.g., changing "Y" and "yes" to "Yes").
  + Consider applying consistent formats for categories, such as making text case uniform (upper/lower case) or ensuring consistent date formats.

5. Remove Unwanted Characters

Problem: Extra spaces, special characters, or non-printable characters can affect analysis and reporting.

Steps:

* + Use functions like TRIM() to remove leading and trailing spaces.
  + Use REPLACE() to eliminate unwanted characters or symbols.
  + Regularly check for non-printable characters or stray symbols that could interfere with analysis.

7

6. Handle Outliers

Problem: Outliers, or extreme values, can skew statistical analysis and lead to incorrect conclusions.

Steps:

* + Identify outliers by analyzing data distributions or using statistical techniques (e.g., Z-scores, interquartile range).
  + Decide how to handle outliers: either remove them, transform them, or replace them with more reasonable values (such as the mean or median).

Data Analysis using Python, Pandas

Data analysis for data-cleaned datasets is a critical phase where you derive meaningful insights, patterns, and trends from the cleaned data. Once the data has been cleaned—meaning that it has gone through the processes of removing duplicates, handling missing values, correcting errors, standardizing formats, and addressing outliers—you can proceed with a more focused and accurate analysis. Below is an explanation of how data analysis is carried out once the data is clean:

Exploratory Data Analysis (EDA)

* Purpose: To understand the data's structure, detect patterns, identify relationships, and gain insights into the dataset.
* Approach:

Visualizations: Create charts and plots to get a sense of the data distribution, trends, and correlations. Common techniques include histograms, box plots, scatter plots, and line charts.

Descriptive Statistics: Calculate basic summary statistics (mean, median, mode, standard deviation, etc.) to understand the central tendency and spread of the data.

Correlation Analysis: Check for relationships between variables, especially for numerical data, using correlation matrices or heatmaps.

Tools Used:

* matplotlib and seaborn for visualizations.
* pandas for calculating summary statistics and handling data.

8

Data Visualization using Excel, Tableau

Data visualization is a crucial aspect of data analysis, as it helps to communicate insights in an intuitive and easily understandable way. After analyzing your data using tools like Excel or Tableau, the next step is to visualize the results through charts and graphs. Both Excel and Tableau offer a wide range of visualizations that can be tailored to your specific dataset and analytical objectives.

Data Visualization Using Excel

Excel is a widely used tool for data analysis and visualization, especially for smaller datasets. It offers a variety of chart types and customization options to create visually appealing and insightful visualizations.

Prepare Your Data Before visualizing data in Excel, ensure that your data is organized. Rows should represent observations, and columns should represent variables. For effective analysis, clean data should be ready for charting.

Select Your Data Highlight the data range that you want to visualize.

Insert a Chart Go to the Insert tab on the Excel ribbon, and choose from a variety of chart types:

* + Bar Chart: Best for comparing quantities across categories (e.g., sales by region).
  + Line Chart: Ideal for showing trends over time.
  + Pie Chart: Good for showing proportions of a whole (e.g., market share of different products).
  + Scatter Plot: Best for visualizing relationships between two variables (e.g., sales vs. advertising spend).
  + Histogram: Useful for understanding the distribution of numerical data.
  + Heatmaps: Used for visualizing the intensity of data in a matrix format (e.g., customer feedback ratings).

Customize Your Chart

* + Chart Title: Always give a meaningful title to your chart.
  + Axis Titles: Label the X-axis and Y-axis to provide context for the chart.
  + Legend: Ensure the chart has a clear legend if it includes multiple categories or data series.
  + Color: Customize the color scheme to make the chart more visually appealing or easy to interpret.
  + Gridlines: Adjust gridlines for better readability.

Interpret the Data Once the chart is created, analyze the chart to extract insights and identify patterns, trends, or correlations.

9

Data Visualization Using Tableau

Tableau is a powerful and interactive data visualization tool that is designed to handle large datasets and complex visualizations. Unlike Excel, Tableau provides more sophisticated and dynamic visualization capabilities, including dashboards, interactive filters, and mapping features.

Import Your Data Open Tableau and connect it to your data source (Excel, CSV, SQL, etc.). Once the data is imported, Tableau automatically interprets the data types (e.g., numerical, categorical) and organizes them in the "Data" pane.

Drag and Drop to Create Visualizations

* + Rows and Columns: Drag fields from the data pane into the "Rows" and "Columns" shelves to define what should be shown on the X and Y axes.
  + Marks: The "Marks" section allows you to customize the visual appearance of your data (e.g., color, size, shape).
  + Filters: You can apply filters to narrow down the data displayed in the visualization (e.g., filter by region or date range).

Choose the Visualization Type Tableau automatically suggests visualizations based on the data you are using. You can choose from a variety of options:

* + Bar/Column Charts: Used for comparing categories.
  + Line Charts: Ideal for trend analysis over time.
  + Pie Charts: Good for displaying proportions.
  + Scatter Plots: Used for relationships between variables.
  + Heat Maps: For visualizing the intensity of data in a matrix format.
  + Geographical Maps: Useful for visualizing data on a map (e.g., sales by region or country).

Customize Your Visualization

* + Filters: Add filters to allow viewers to interact with the data and focus on specific subsets.
  + Color Encoding: Use color to highlight categories, show intensity, or differentiate data points.
  + Tooltips: Customize tooltips that appear when users hover over a data point, providing additional context.
  + Annotations: Add annotations to your charts to call attention to important insights or findings.

Create Dashboards Once you have multiple visualizations (e.g., bar charts, line graphs, pie charts), you can combine them into a single dashboard. Dashboards allow you to present different aspects of your analysis in one interactive screen. You can also use filters and actions to make the dashboard interactive, so users can explore the data further.

10

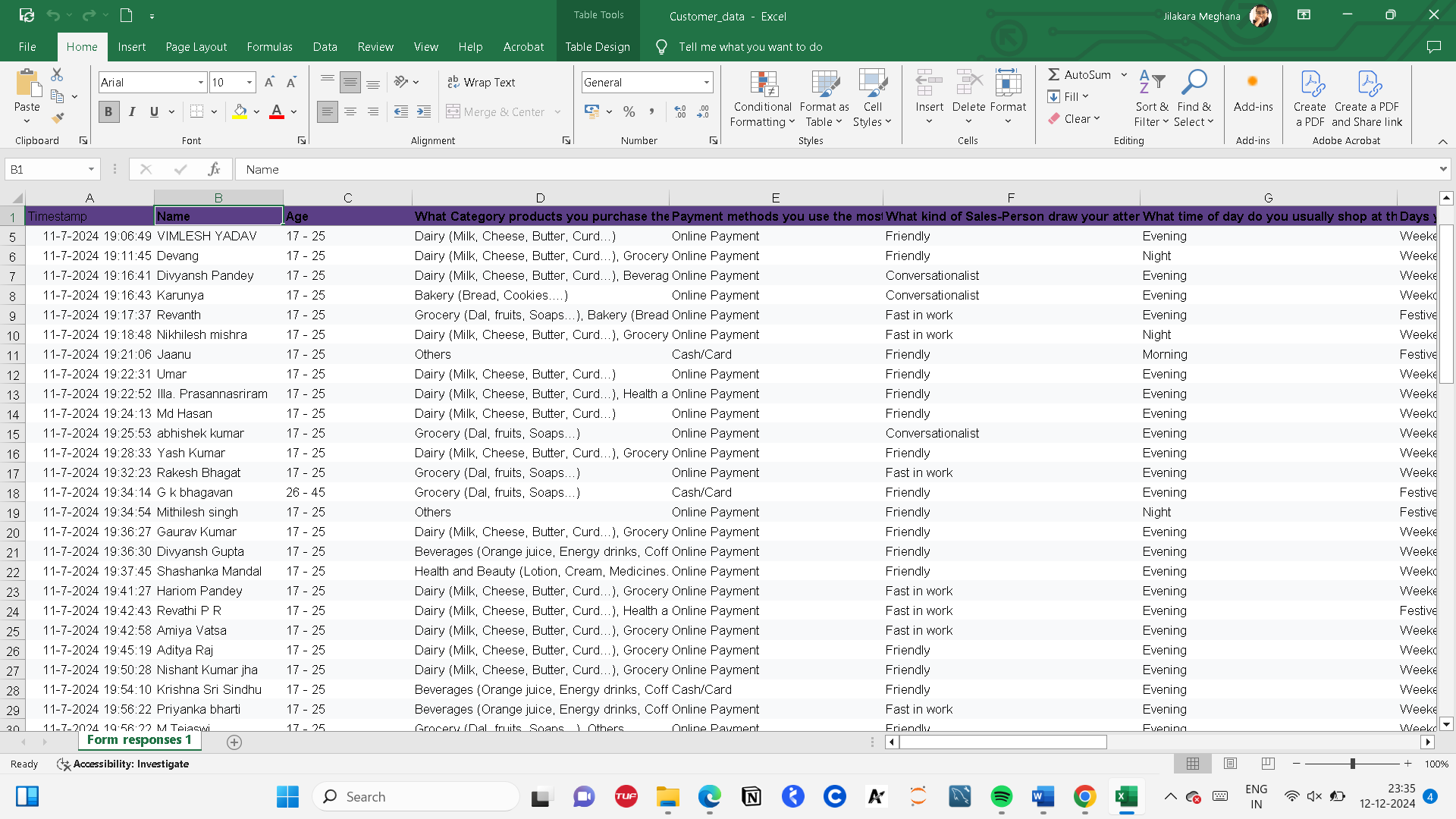
Representation of Data Analyzed

The analysis of the data reveals key trends and insights that guide decision-making. Key findings include correlations between customer preferences and product sales, with certain products showing higher demand in specific regions. Visualization of sales over time highlights seasonal fluctuations, allowing for better inventory planning. Regression analysis suggests that marketing expenditure has a direct impact on sales growth. Customer feedback analysis points to areas of improvement in product quality and customer service. Based on these insights, recommendations for optimizing marketing strategies and inventory management are proposed

11

1. **IMPLEMENTATION OF ANALYSIS**

The project begins with **data collection** through surveys and feedback forms from customers and classmates, which provides valuable insights into their preferences, buying behaviors, and satisfaction levels with the local mart's products and services. The data was gathered in various formats, primarily as CSV files, and included customer demographics, purchase history, and feedback on products and services.

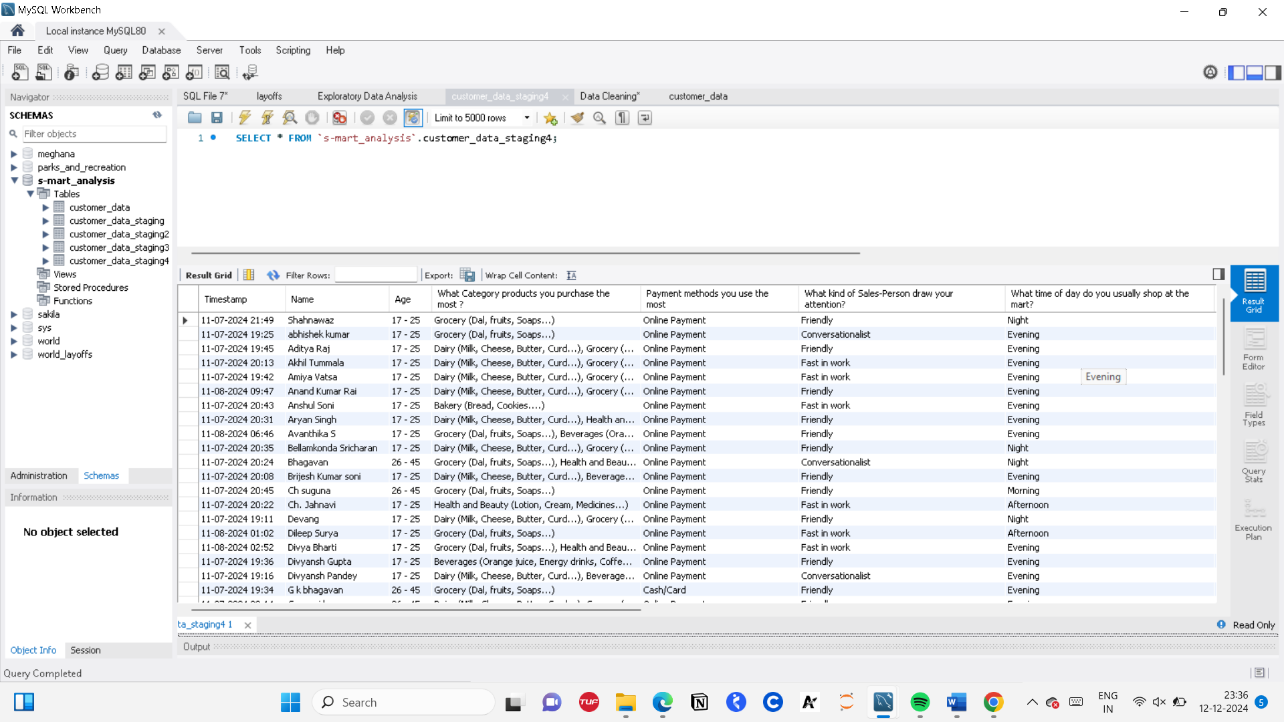


*Figure 5.1 : Data Collected from Customer Feedback*

Once the data was collected, the **data cleaning process** was initiated using **MySQL**. The database was populated with the raw data, and SQL queries were written to address common data issues such as:

* **Removing duplicates**: Identifying and removing any repeated entries that may have resulted from errors in data collection.
* **Handling missing values**: Applying SQL commands to identify missing or NULL values and deciding how to address them, either by filling in missing values with mean, median, or mode or by removing incomplete records.
* **Standardizing data formats**: Ensuring consistency in the data by converting dates to a standard format, correcting any discrepancies in categorical values (e.g., 'Yes' and 'No' standardized to '1' and '0'), and ensuring numeric fields were free from non-numeric characters.

12



*Figure 5.2 : Data Cleaned using MySQL Workbench*

After cleaning the data, **Exploratory Data Analysis (EDA)** was performed using **Pandas** in Python. This step involved the following:

* **Data summarization**: Calculating descriptive statistics such as mean, median, and standard deviation to understand the distribution of the data.
* **Correlation analysis**: Using correlation matrices to identify relationships between variables such as the link between marketing expenditure and sales.
* **Data visualization**: Generating histograms, box plots, and scatter plots using **matplotlib** and **seaborn** to visually inspect data distributions and identify outliers or trends.

With clean and analyzed data, the next step was the **visualization** of the data. **Excel** was used to create bar charts, line graphs, and pie charts for simple and quick visual representation of the results. For more complex and interactive visualizations, **Tableau** was used to design dynamic dashboards. These dashboards allowed for an interactive exploration of the data, enabling filtering by different variables such as region, product type, and time period. The visualizations were created to represent sales trends, customer demographics, and feedback results, making it easier for stakeholders to interpret the findings.

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Labels** |  |  |  |
| **Large (16+ items)** | **Medium (6-15 items)** | **Small (1-5 items)** | **Grand Total** |
| **2** | **29** | **36** | **67** |
| 2 | 13 | 11 | 26 |
|  | 16 | 25 | 41 |
| **2** | **6** | **2** | **10** |
| 1 | 3 |  | 4 |
| 1 | 3 | 2 | 6 |
| **4** | **35** | **38** | **77** |

*Figure 5.3 : Products Sold Per Day Table*

13

*Figure 5.4 : Products Sold Per Day Graph*

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Payment methods** | **Column Labels** |  |  |
| **Row Labels** | **17 - 25** | **26 - 45** | **Grand Total** |
| Female | 26 | 4 | 30 |
| Male | 41 | 6 | 47 |
| **Grand Total** | **67** | **10** | **77** |

*Figure 5.5 : Total Payments Table*

*Figure 5.6 : Total Payments Graph*

14

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column Labels** |  |  |  |  |
| **Afternoon** | **Evening** | **Morning** | **Night** | **Grand Total** |
| **4** | **22** | **3** | **1** | **30** |
| 3 | 20 | 2 | 1 | 26 |
| 1 | 2 | 1 |  | 4 |
| **3** | **35** |  | **9** | **47** |
| 3 | 31 |  | 7 | 41 |
|  | 4 |  | 2 | 6 |
| **7** | **57** | **3** | **10** | **77** |
|  |  |  |  |  |

*Figure 5.7 : Sales Per Day Table*

*Figure 5.8: Sales Per Day Graph*

|  |  |  |  |
| --- | --- | --- | --- |
| **Count of Products** | **Column Labels** |  |  |
| **Row Labels** | **17 - 25** | **26 - 45** | **Grand Total** |
| Female | 26 | 4 | 30 |
| Male | 41 | 6 | 47 |
| **Grand Total** | **67** | **10** | **77** |

*Figure 5.9 : Products Purchased on basis of Gender, Age Graph*

15

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Count of Discount Applied** | **Column Labels** |  |  |  |
| **Row Labels** | **May be** | **No** | **Yes** | **Grand Total** |
| No | 11 | 4 | 5 | 20 |
| Yes | 30 | 2 | 25 | 57 |
| **Grand Total** | **41** | **6** | **30** | **77** |

*Figure 5.10 : Discount Applied for Type of Visitors Table*

*Figure 5.10 : Discount Applied for Type of Visitors Table*

*Figure 5.10 : Discount Applied for Type of Visitors Table*

*Figure 5.11 : Discount Applied for Type of Visitors Graph*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sum of Rating** | **Column Labels** |  |  |  |  |
| **Row Labels** | **2** | **3** | **4** | **5** | **Grand Total** |
| Female | 8 | 30 | 44 | 25 | 107 |
| Male | 8 | 30 | 104 | 35 | 177 |
| **Grand Total** | **16** | **60** | **148** | **60** | **284** |

*Figure 5.12: Rating Provided Table*

*Figure 5.13: Rating Provided Graph*

16

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Count of Products\_2** | **Column Labels** |  |  |  |  |  |  |
| **Row Labels** | **Bakery** | **Beverages** | **Dairy** | **Grocery** | **Health and Beauty** | **Others** | **Grand Total** |
| **Female** | **2** | **6** | **7** | **7** | **2** | **6** | **#** |
| 17 - 25 | 2 | 6 | 6 | 4 | 2 | 6 | # |
| 26 - 45 |  |  | 1 | 3 |  |  | 4 |
| **Male** | **3** | **3** | **22** | **15** | **1** | **2** | **#** |
| 17 - 25 | 3 | 3 | 20 | 12 | 1 | 2 | # |
| 26 - 45 |  |  | 2 | 3 |  |  | 5 |
| **Grand Total** | **5** | **9** | **29** | **22** | **3** | **8** | **#** |
|  |  |  |  |  |  |  |  |

*Figure 5.13: Products consumed Table*

*Figure 5.14: Products consumed Graph*

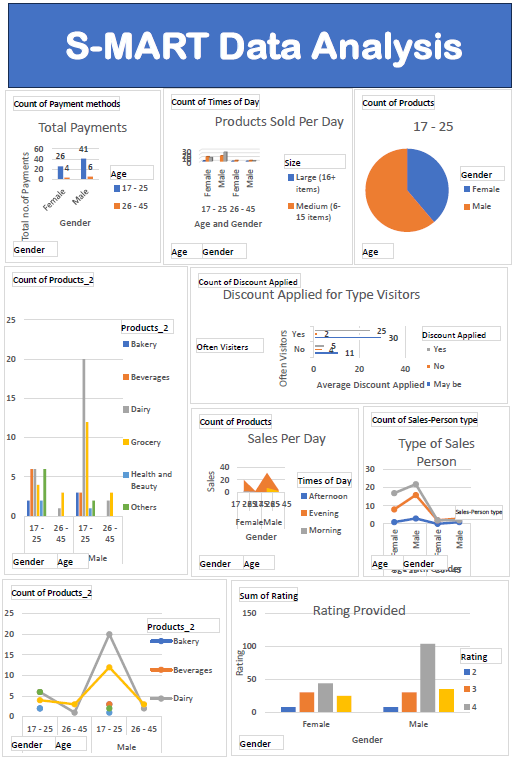
Finally, based on the insights obtained from the analysis, a set of **suggestions** was provided to the mart’s management:

* **Optimizing inventory** based on the sales patterns observed in the analysis.
* **Tailoring marketing strategies** to regions where specific products performed better.
* **Improving customer service** based on feedback and areas identified for improvement.
* **Introducing promotional offers** during peak sales months to increase foot traffic and sales.

The implementation of this analysis, from data collection through visualization and suggestions, provides actionable insights that could help the mart improve operations, increase customer satisfaction, and boost sales.

17

**5.2 Final Analysis of S-MART (Bilaspur, C.G.)**



18

**5.3 FINAL INSIGHTS and SUGGESTIONS**

From this analysis, the results drew gives the insights like :

1. The Grocery Products are being purchased heavily by the customers, followed by Dairy and Bakery , Beverages.
2. The Sales in a Day will be maximum in the Evening.
3. Friendly and Conversationalist Sales-Person are required for better approach towards the Consumers
4. The Marketing of the Mart is Poor which need to be improved by the Poster, Social Media, Adds, etc
5. The Online Payments are happening more than the Cash/Card Payments.
6. Discount is applied to 38% of people out of 100% Customers, in which 72% are regular Customers and 25.9% are Irregular Customers.
7. With respect to the location of mart the performance of Mart can be increased and draw the attention of University Students
8. The Small Size Products (1-5 items) are being sold regularly.
9. The Shopping in the Mart is mostly done in Weekends. So the Mart staff should be active and work at 2x speed in weekends and Festive Days

Customer Suggestions

1. Requires fresh fruits and vegetables daily
2. The Mart should provide more items .
3. If there were the clothes options also,it would be helpful to us
4. yes my suggestion is that they should pay attention on their billing counters because many times whenever i went at billing counters their speed is not that fast i mean sometimes even i have to wait for 10 to 15 minutes that some of them will come and do billing so that's all otherwise overall their services are pretty good
5. Discount Coupon presentation should be more clear so that it becomes more easy understandable to the customers and they can easily get to know that actually how much discount they are getting on that product....
6. Receive the customers in a friendly manner so that they will buy more than they thought

19

**CONCLUSION**

**Achievements**

In conclusion, the **S-MART Data Analysis** project has successfully provided valuable insights into the mart's operations, customer preferences, and sales trends through the application of data analysis techniques and visualization tools. By cleaning and analyzing data from multiple sources, including customer surveys, staff feedback, and sales records, we were able to identify critical trends such as peak sales times, popular products, and customer satisfaction levels.

The key findings reveal that certain products perform better during specific times of the year, and that there is room for improvement in inventory management to avoid stockouts and meet customer demand efficiently. Staff interviews corroborated many of the data-driven insights, highlighting the importance of customer service and product availability in driving satisfaction.

Based on these insights, actionable recommendations were provided, including improving inventory planning, offering targeted promotions, and enhancing the in-store customer experience. These improvements can help S-MART optimize its operations, boost customer satisfaction, and ultimately drive better business outcomes. Future work could focus on further refining inventory forecasting models and exploring customer segmentation for more personalized marketing strategies.

The findings of this project demonstrate the potential of data-driven decision-making in retail operations, and offer a strong foundation for S-MART to grow and thrive in a competitive marketplace.

20

***RESOURCES***

1. MySQL website

[https://www.mysql.com/](%20https://www.mysql.com/%20)

1. Tableau Website

https://www.tableau.com/

1. Jupyter Notebook

https://jupyter.org/

1. "Exploratory Data Analysis with R" by Roger D. Peng https://www.amazon.com/Exploratory-Data-Analysis-R-Roger/dp/0134203597

21

***APENDIX 1***

CREATE TABLE customer\_data\_staging3

LIKE customer\_data;

INSERT customer\_data\_staging3

SELECT \*

FROM customer\_data1;

SELECT\*

FROM customer\_data\_staging3;

WITH duplicate\_cte1 AS

(

SELECT \*,

ROW\_NUMBER() OVER(

PARTITION BY `Name`,Age /\*`What Category products you purchase the most ?`, `Payment methods you use the most`, `What kind of Sales-Person draw your attention?`,

`What time of day do you usually shop at the mart?`,

`Days you enjoy to shop in s-mart ?`, `Is Discount applied on your products ?`,

`What is the typical size of your shopping trip at the mart?`,

`Rate the Marketing(Advertising) of the s-mart ?`,`Would You often visit s-mart ?`,

`What do you like the most about shopping at this mart?`,

`Any other suggestions or feedback for us?`, `Email address`\*/) AS row\_num

FROM customer\_data\_staging

)

SELECT \*

FROM duplicate\_cte1

WHERE row\_num >1;

CREATE TABLE `customer\_data\_staging4` (

`Timestamp` text,

`Name` text,

`Age` text,

`What Category products you purchase the most ?` text,

`Payment methods you use the most` text,

`What kind of Sales-Person draw your attention?` text,

`What time of day do you usually shop at the mart?` text,

`Days you enjoy to shop in s-mart ?` text,

`Is Discount applied on your products ?` text,

`What is the typical size of your shopping trip at the mart?` text,

`Rate the Marketing(Advertising) of the s-mart ?` int DEFAULT NULL,

`Would You often visit s-mart ?` text,

`What do you like the most about shopping at this mart?` text,

`Any other suggestions or feedback for us?` text,

`Email address` text,

`row\_num` INT

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

SELECT\*

FROM customer\_data\_staging4;

INSERT INTO customer\_data\_staging4

SELECT \*,

ROW\_NUMBER() OVER(

PARTITION BY `Name`,Age /\*`What Category products you purchase the most ?`, `Payment methods you use the most`, `What kind of Sales-Person draw your attention?`,

`What time of day do you usually shop at the mart?`,

`Days you enjoy to shop in s-mart ?`, `Is Discount applied on your products ?`,

`What is the typical size of your shopping trip at the mart?`,

`Rate the Marketing(Advertising) of the s-mart ?`,`Would You often visit s-mart ?`,

`What do you like the most about shopping at this mart?`,

`Any other suggestions or feedback for us?`, `Email address`\*/) AS row\_num

FROM customer\_data\_staging;

SELECT\*

FROM customer\_data\_staging4

WHERE row\_num > 1;

DELETE

FROM customer\_data\_staging4

WHERE row\_num > 1;

DELETE

FROM customer\_data\_staging4

WHERE `Name` = '11-08-2024 20:51' ;

INSERT INTO customer\_data\_staging4 (`Timestamp`, `Name`, `Age`, `What Category products you purchase the most ?`, `Payment methods you use the most`, `What kind of Sales-Person draw your attention?`, `What time of day do you usually shop at the mart?`, `Days you enjoy to shop in s-mart ?`, `Is Discount applied on your products ?`, `What is the typical size of your shopping trip at the mart?`, `Rate the Marketing(Advertising) of the s-mart ?`, `Would You often visit s-mart ?`, `What do you like the most about shopping at this mart?`, `Any other suggestions or feedback for us?`, `Email address`, `row\_num`) VALUES("11-07-2024 19:59", "J Prasanna", "17 - 25", "Others", "Cash/Card", "Friendly", "Evening", "Festive days", "No", "Small (1-5 items)", "3", "No", "Chocolates", "Receive the customers in a friendly manner so that they will buy more than they thought:"," ",1) ;

SELECT\*

FROM customer\_data\_staging4;

#Removing Duplicates end here

#Standaradization starts here (Identifying problems and fixing it)

SELECT TRIM(`Name`), TRIM(`What Category products you purchase the most ?`)

FROM customer\_data\_staging4;

UPDATE customer\_data\_staging4

SET `Name` = TRIM(`Name`);

UPDATE customer\_data\_staging4

SET `What Category products you purchase the most ?` = TRIM(`What Category products you purchase the most ?`);

SELECT \*

FROM customer\_data\_staging4

WHERE `Any other suggestions or feedback for us?` LIKE "No%";

UPDATE customer\_data\_staging4

SET `Any other suggestions or feedback for us?` = "No"

WHERE `Any other suggestions or feedback for us?` LIKE "No%";

UPDATE customer\_data\_staging4

SET `Any other suggestions or feedback for us?` = "No"

WHERE `Any other suggestions or feedback for us?` LIKE "NA%";

SELECT \*

FROM customer\_data\_staging4

WHERE `Any other suggestions or feedback for us?` LIKE 'Good%';

UPDATE customer\_data\_staging4

SET `Any other suggestions or feedback for us?` = "No"

WHERE `Any other suggestions or feedback for us?` LIKE '';

SELECT \*

FROM customer\_data\_staging4

WHERE `What do you like the most about shopping at this mart?` LIKE "No%";

UPDATE customer\_data\_staging4

SET `What do you like the most about shopping at this mart?` = "Nothing"

WHERE `What do you like the most about shopping at this mart?` LIKE '';

UPDATE customer\_data\_staging4

SET `Any other suggestions or feedback for us?` = "Good"

WHERE `Any other suggestions or feedback for us?` LIKE 'Good%';

#Standaradization ends here

#No values/blank values starts here

SELECT `Name`,TRIM(`Name`)

FROM customer\_data\_staging4;

UPDATE customer\_data\_staging4

SET `Name`=TRIM(`Name`);

#Mostly done in step 2

#Mo values/blank values end here

#Remove any colums starts here

SELECT\*

FROM customer\_data\_staging4

WHERE `Email address` IS NULL

OR `Email address` = '';

ALTER TABLE customer\_data\_staging4 DROP COLUMN `Email address`;

ALTER TABLE customer\_data\_staging4

DROP COLUMN `row\_num`

# Remove any colums ends here