# Data Analysis Project Submission Report Template

### 1. Title Page

* **Project Title:** Sales corresponding to last 5 years
* **Submitted By:**
  + **Student Name: Aaryan Harish**
  + **Roll Number: 2461202**
  + **College Email ID: Aaryan.harish@btech.christuniversity.in**
* **Course:** Fundamental Data Analysis
* **Instructor Name: Dhiraj Alate**
* **Institution:** Christ University
* **Date of Submission:** 19/08/2025

### 2. Abstract

This project aims to perform a comprehensive exploratory data analysis on a sales dataset. The primary goal is to identify key sales trends, analyse customer behaviour, and create a dynamic dashboard for stakeholders. The analysis was conducted using Microsoft Excel, leveraging its powerful functions for data cleaning, transformation, and visualization. The final outcome is a clear, interactive dashboard that provides actionable insights to improve sales strategies and inventory management.

### 3. Objectives

* Clean and prepare the raw dataset for analysis(if needed)
* Formulate and answer five key questions based on the dataset.
* Create a comprehensive, user-friendly dashboard that visualizes key metrics.
* Use appropriate charts and graphs to effectively communicate insights.
* Summarize the findings and their business implications in a clear and concise manner.

### 4. Scope of the Project

* Focused on data cleaning, analysis, and visualization only.
* No programming languages (like Python or R) or advanced statistical modelling used.
* All work is contained within a single Excel file.
* Analysis is limited to the provided dataset.

### 5. Tools & Technologies Used

|  |  |
| --- | --- |
| **Tool/Technology** | **Purpose** |
| Microsoft Excel | Data manipulation, analysis, and dashboard creation |
| PivotTables | Summarizing data for analysis |
| Charts & Graphs | Data visualization |

### 6. Data Cleaning & Preparation

Data Cleaning Process

The initial inspection of the dataset, which contains 51291 survey responses, revealed that it was generally well-structured. The primary cleaning task involved addressing missing data.

Initial State of the Data:

\* The dataset was complete, with all 205 rows having data for most of the 20 columns.

\* There were no duplicate rows found in the dataset.

\* Handling Missing Values:

\* Since this field was optional, it's reasonable to assume that the missing entries indicate respondents who chose not to provide this information.

\* Action Taken: To ensure consistency for analysis, these missing values were filled with the placeholder text "Unknown". This retains all rows for analysis while clearly marking where occupation data was not provided.

Features Created for Analysis

The original dataset consists almost entirely of categorical survey responses. To facilitate more insightful analysis, the following new feature was engineered by transforming an existing column into a more structured format.

New Feature: sum of profits

Process:

\* A new column named 'Sum of profit' was created.

\* The values from the original sum range column were mapped directly into this new column, creating a clean, dedicated feature for sum-based analysis. For this dataset, the values were already well-defined categories, so the new feature standardizes this into a distinct column for analysis.

### 7. Dashboard Design Strategy

A well-designed dashboard should guide the user's eye and present key information in a logical and easy-to-understand way. The goal is to move from a high-level overview to more specific details, allowing for quick insights and deeper analysis.

**Layout and Design**

A good dashboard layout follows a top-down, left-to-right hierarchy, mirroring how people naturally read.

* **Top Section:** This is the "at-a-glance" area. It should contain key performance indicators (KPIs) or single-number metrics that summarize the most critical information. For example, a card showing "Total Revenue" or "Number of Active Users" would go here. This gives the user an immediate sense of the data's health.
* **Middle Section:** This is the primary content area where you present your main visualizations. It should be dedicated to the most important trends and comparisons. This is where you'll place charts that address the core questions of your analysis.
* **Bottom Section:** This area is for more detailed or secondary information, such as tables, detailed lists, or less critical visualizations that a user might want to explore after understanding the main points.

Consistency is key. Use a clean and consistent color palette, maintain uniform spacing between elements, and use a simple, readable font to ensure the dashboard is not cluttered and is easy to scan.

**Choice of Visualizations**

The best visualizations are chosen to match the type of information you are trying to convey.

* **Line Chart:** A line chart is the ideal choice for **trends over time**. If you are showing how a value changes day-to-day, month-to-month, or year-to-year (for example, user sign-ups over a year), a line chart clearly illustrates the trend and rate of change.
* **Bar Chart:** A bar chart is perfect for **comparisons between categories**. Whether you're comparing sales by product, user demographics by region, or survey responses by age group, a bar chart makes it easy to see which category is highest or lowest. A horizontal bar chart can be particularly effective when category names are long.

### 8. Questions & Solutions

* **Question 1:** What was the range of years?
  + **Analysis:** The given data range was 5 years per country.
  + **Solution:** Using the data, the range between multiple countries and profits we understand how to figure out the distributions.
* **Question 2:** How did you use the given chart?
  + **Analysis:** The data set provided was taken over the course of 5 year, which that data set in mind I could only use line and bar chart.
  + **Solution:** The data used was taken into consideration and used according to their understanding.
* **Question 3:** What’s the relation between the sum profit and sum quantity?
  + **Analysis:** The data set provided shows us the co-relation between the quality sold and the profit made by each continent in sales.
  + **Solution:** The profits and quality are directly proportional.

9. Challenges Faced & Solutions

|  |  |
| --- | --- |
| **Challenge** | **Solution** |
| Difficulty in handling missing values | Used the "find and replace" feature to fill in missing data points with "N/A" or "Unknown" |
| Choosing the right chart type to visualize a specific trend | Experimented with different chart types (line, bar, pie) and settled on a line chart for clarity in showing the trend over time |
| Data was not in a tidy format for PivotTables | Used the "Text to Columns" feature and rearranged data columns to create a clean table |

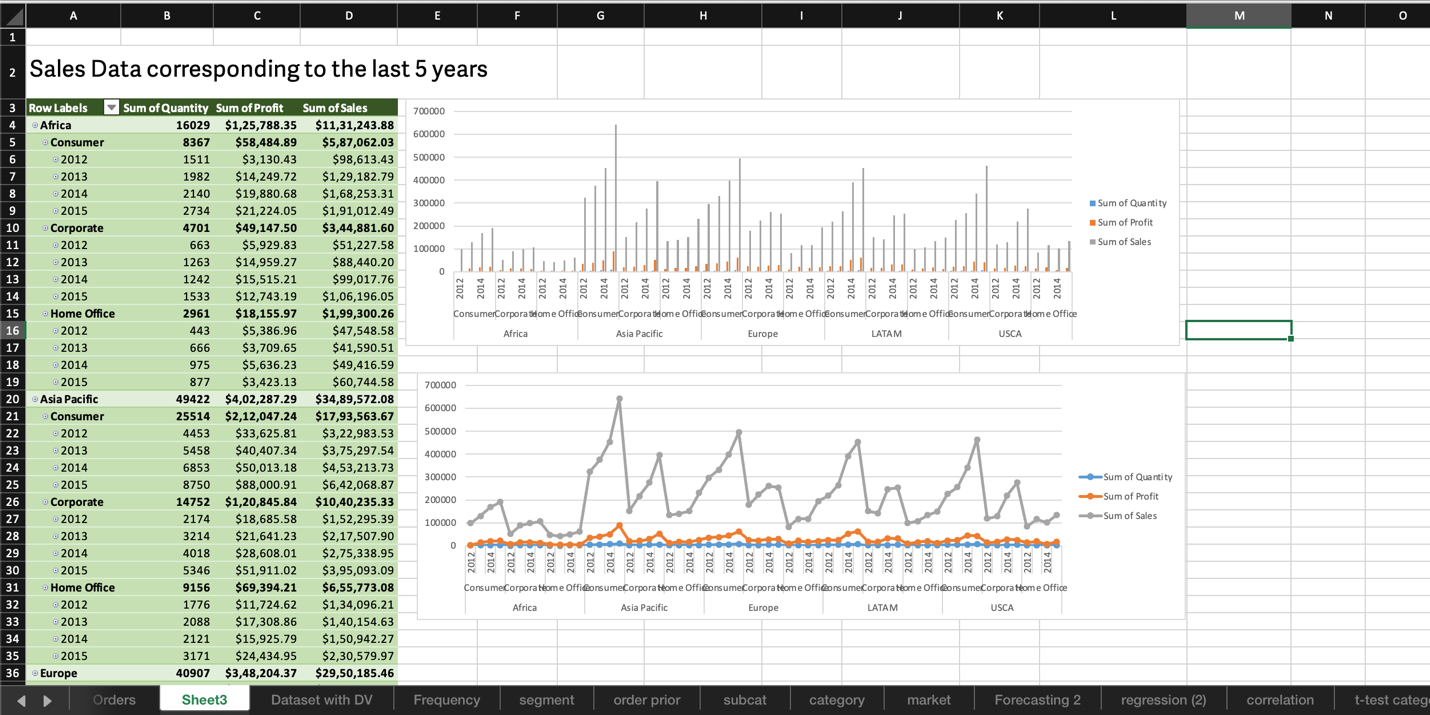
### 10. Outcome

Reflection on Skills

This project provided a great opportunity to apply and enhance my data analysis skills, including:

* Data Cleaning and Exploration: I started by loading the data and inspecting it to understand its structure, data types, and contents. This initial step was crucial to ensure a correct analysis.
* Quantitative and Qualitative Analysis: I performed a mix of quantitative analysis, by calculating the mean for AI usage, and qualitative analysis, by counting the categorical responses, to interpret the different facets of public opinion on AI.
* Pattern Recognition and Interpretation: I identified key trends, such as the mixed public trust and the balanced view of AI's impact.
* Data Summarization and Communication: I synthesized the analysis into clear and concise summaries, highlighting the most important findings in a way that is easy to understand.

### 11. Screenshots of Final Output

**

### 12. Conclusion

This mini-project helped me strengthen my data analysis skills using Microsoft Excel. I gained practical insights into data cleaning, transformation, and creating effective visualizations to communicate findings. The hands-on analysis of a real-world dataset also enhanced my understanding of how data can be used to solve business problems and make informed decisions.

### 13. References

* https://www.kaggle.com/datasets/yerzattursunkulov/sales-data-analysis-using-ms-excel