**Pre-Report for the Madhav Sales (Orders, Details) dataset:**

**Index**

|  |  |  |
| --- | --- | --- |
| **Sl.No** | **Report** | **Page No** |
| 1 | Problem Statement | 1 |
| 2 | Data Requirements | 1 |
| 3 | Data Collection | 1 |
| 4 | Data Validation | 2 |
| 5 | Data Cleaning | 2 |
| 6 | Tools | 2 |
| 7 | Dashboard | 2 |
| 8 | Storytelling | 2 |

The datasets are structured as follows:

1. **Details.csv** contains transactional details:
   * **Order ID**, **Amount**, **Profit**, **Quantity**, **Category**, **Sub-Category**, and **Payment Mode**.
2. **Orders.csv** provides additional order-related information:
   * **Order ID**, **Order Date**, **Customer Name**, **State**, and **City**.
3. **Problem Statement**

The objective is to analyze transactional and order data to derive actionable insights, optimize performance metrics, and improve decision-making processes.

**2. Data Requirements**

To fulfill the objectives, the following data is necessary:

* Order details: Order ID, transaction amount, profit, and quantities sold.
* Product segmentation: Categories and sub-categories of items purchased.
* Customer demographics: Names, states, and cities of customers.
* Transactional metrics: Payment modes and order dates.

**3. Data Collection**

The data has been collected from two CSV files:

* **Details.csv:** Contains granular details about transactions.
* **Orders.csv:** Includes contextual order information like dates and customer details.

**4. Data Validation**

Steps to validate the data:

* Verify that all rows in both datasets have unique **Order ID** values and confirm integrity between datasets by ensuring consistent **Order ID** mapping.
* Check for missing or null values in critical fields.
* Identify and resolve anomalies like mismatched dates or unrealistic amounts.

**5. Data Cleaning**

* Handle missing values in columns (e.g., replacing or imputing where needed).
* Standardize formats for dates and text (e.g., city names).
* Remove duplicates or incorrect entries.
* Resolve inconsistencies between datasets by aligning **Order ID** values.

**6. Tools**

* Data Analysis: Python (Pandas, NumPy).
* Visualization and Storytelling: Tableau, Power BI, or Matplotlib/Seaborn.
* Data Validation and Cleaning: Python libraries and manual inspection

**7. Dashboard**

Key dashboard components:

* **Profitability Analysis:** By product category, sub-category, and location.
* **Sales Trends:** Over time, by payment mode and geography.
* **Customer Insights:** Segment customers based on purchase behavior.

**8. Storytelling**

Transform insights into a narrative:

* Highlight profitable regions, products, and customer groups.
* Identify and address areas of loss or low performance.
* Present actionable recommendations for boosting sales and customer satisfaction.