**I. Problem Analysis and Requirements**

1. **Survey**

**About Company**

Nielsen Holdings plc (NYSE: NLSN) is a global measurement and data analytics company that provides the most complete and trusted view available of consumers and markets worldwide. Nielsen is divided into two business units. Nielsen Global Media, the arbiter of truth for media markets, provides media and advertising industries with unbiased and reliable metrics that create a shared understanding of the industry required for markets to function. Nielsen Global Connect provides consumer packaged goods manufacturers and retailers with accurate, actionable information and insights and a complete picture of the complex and changing marketplace that companies need to innovate and grow. Our approach marries proprietary Nielsen data with other data sources to help clients around the world understand what’s happening now, what’s happening next, and how to best act on this knowledge. An S&P 500 company, Nielsen has operations in over 100 countries, covering more than 90% of the world’s population.

Know more: <https://www.nielsen.com/us/en/>

#### **Dataset Context:**

Nielsen receives transaction-level scanning data (POS Data) from its partner stores on a regular basis. These stores include large-format stores like supermarkets and hypermarkets, as well as smaller traditional grocery stores (Kirana stores), medical stores, etc., using POS machines.

**Data Source**: The data is collected through POS machines, where each transaction contains information about the products scanned at the time of the sale.

[Store Transaction data](https://www.kaggle.com/datasets/iamprateek/store-transaction-data?resource=download)

1. **DataSet Analysis**

Contains store level data by brands and categories for select stores

| File Name | Description |
| --- | --- |
| | **Hackathon\_Ideal\_Data** | | --- |  |  | | --- | | Contains brand-level data for 10 stores over the last 3 months. Serves as the "ideal" reference data. |
| | **Hackathon\_Working\_Data** | | --- |  |  | | --- | | Includes data for selected stores that is either missing or incomplete. |
| | **Hackathon\_Mapping\_File** | | --- |  |  | | --- | | Provides a reference for understanding the column names in the dataset. |
| | **Hackathon\_Validation\_Data** | | --- |  |  | | --- | | Contains data for stores and product groups for which the Total\_VALUE needs to be predicted. |

**Hackathon\_Ideal\_Data.csv**

* **Columns:** 10 (MONTH, STORECODE, QTY, VALUE, GRP, SGRP, SSGRP, CMP, MBRD, BRD).
* **Description:** Contains monthly transaction data with information on store, quantity, value, product group, and brand.
* **Rows:** 14,260.
* **Sample data**

| **MONTH** | **STORECODE** | **QTY** | **VALUE** | **GRP** | **SGRP** | **…** |
| --- | --- | --- | --- | --- | --- | --- |
| M1 | P1 | 25 | 83 | HAIR CONDITIONERS | HAIR CONDITIONERS |  |

**Hackathon\_Mapping\_File.csv**

* **Columns:** 3 (File Name, Column Name, Column Description).
* **Description:** This file contains descriptions of columns in other files, explaining the meaning of each column.
* **Rows:** 24.
* **Sample data:**

| **File Name** | **Column Name** | **Column Description** |
| --- | --- | --- |
| Hackathon\_Ideal\_Data | MONTH | Month ID (M1, M2, M3) |

* **Hackathon\_Validation\_Data.csv**
* **Columns:** 4 (ID, STORECODE, MONTH, GRP).
* **Description:** Contains validation data for transactions with ID and product group information.
* **Rows:** 2,430.
* **Sample data:**

| **ID** | **STORECODE** | **MONTH** | **GRP** |
| --- | --- | --- | --- |
| 1112535 | N1 | M1 | AFTER SHAVE LOTIONS |

**Hackathon\_Working\_Data.csv**

* **Columns:** 14 (MONTH, STORECODE, DAY, BILL\_ID, BILL\_AMT, QTY, VALUE, ...).
* **Description:** Contains detailed daily transaction data, including bill ID, bill amount, and product information.
* **Rows:** 26,985.
* **Sample data:**

| **MONTH** | **STORE CODE** | **DAY** | **BILL\_ID** | **BILL\_AMT** | **QTY** | **VALUE** | **GRP** | **…** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M1 | N1 | 4 | T375 | 225.0 | 1.0 | 225.0 | BUTTER | **…** |

**Sample Submission.csv**

* Columns: 2 (ID, TOTAL VALUE).
* Description: A sample output with the total value of transactions.
* Rows: 9.
* Sample data

| **ID** | **TOTAL VALUE** |
| --- | --- |
| 1112535 | 0 |

The group identifies the analysis topics (Requirements) including FACTs:

| | **Table Name** | | --- |  |  | | --- | | **Field Name** | **Description** |
| --- | --- | --- | --- | --- |
| Fact\_Sales | Sale\_Amount | Total sales (Quantity x Value) |
|  | Quantity | Number of items sold |
|  | Value | Price per unit |
|  | Time\_ID |  |
|  | Location\_ ID |  |
|  | Product\_ID |  |

**3. Object and Requirements**

* **Sales Performance Over Time**Analyzing sales performance by month and store allows us to understand revenue trends and identify peak sales periods. Metrics such as **transaction value (VALUE)** and **quantity sold (QTY)** can highlight growth or decline patterns.
* Required Data: Includes product ID, transaction time, transaction value, and customer-related factors.
* Analysis: Data needs to be cleaned and standardized for easy retrieval and analysis, stored in related tables (e.g., transaction table, product table, customer table).
* **Trends by Product Categories**Categorizing products by groups such as **GRP (main group)**, **SGRP (subgroup)**, and **SSGRP (specific subgroup)** helps identify top-performing product categories. This analysis is essential for recognizing best-selling products, seasonal demand variations, and underperforming items.
* Use visualization tools like Power BI or Tableau to create dashboards that display sales trends, inventory levels, and marketing campaign performance to support decision-making.
* **Store Performance Evaluation**Comparing stores using metrics like total revenue and sales volume highlights top-performing locations and those needing improvement. This insight can guide marketing campaigns or operational adjustments at the store level.

| **Category** | **Objective** | **Detailed Requirements** |
| --- | --- | --- |
| 1. Sales Forecasting | Predict revenue and sales volume to optimize business strategies. | - **By time**: Forecast revenue and sales volume by **month**, **quarter**, or **year**. Detect long-term growth or decline trends.  - **By store**: Predict revenue for each **store** to compare performance and identify potential or underperforming stores.  - **By product group**: Forecast sales volume and revenue for each **product group (GRP, SGRP, SSGRP)**. Identify the products likely to generate the highest future revenue. |
| 2. Product Trend Analysis | Understand product consumption behavior to optimize sales and marketing strategies. | - **Identify best-selling products**: List top products in terms of sales volume across categories (**GRP, SGRP, SSGRP**). Track changes in product rankings over time.  - **Analyze seasonal trends**: Identify peak consumption times (e.g., holidays, summer, Black Friday). Suggest promotional or discount campaigns for these seasons.  - **Underperforming products**: Identify products with low sales or revenue. Assess whether to discontinue or adjust strategies for these items. |
| 3. Inventory Management | Optimize stock levels based on product demand forecasts and sales trends. | - **Replenishment planning**: Use sales trends to identify products that require restocking.  - **Avoid overstocking**: Predict inventory requirements to reduce excess stock and associated holding costs.  - **Track inventory performance**: Monitor turnover rates for different product groups to enhance efficiency. |
| 4. Store Performance Analysis | Evaluate and compare store performance to support decision-making. | - **Performance ranking**: Compare revenue and sales volume across stores. List top-performing and underperforming stores.  - **By region**: Aggregate revenue for stores by geographic area (if available). Assess the performance of each region.  - **Improvement suggestions**: Recommend strategies to increase revenue for underperforming stores (e.g., improving product displays, local promotions). Optimize product or service offerings based on local demand. |
| 5. Invoice Value Analysis | Gain insights into customer purchasing behavior through average bill value analysis. | - **Average bill tracking**: Analyze the average invoice value (**BILL\_AMT**) over time.  - **Customer flow analysis**: Track the number of invoices generated daily, monthly, or seasonally to identify busy periods.  - **Spending trends**: Study changes in customer spending behavior for targeted marketing efforts. |
| 6. Demand Forecasting | Predict product and sales demand to enhance planning and decision-making processes. | - **Future demand analysis**: Use historical data to predict future product demand across categories (**QTY**, **VALUE**).  - **Revenue forecasting**: Project future revenue for better financial planning.  - **Scenario simulation**: Model demand scenarios to prepare for seasonal changes or market fluctuations. |