Power Bi

Pre – Report

By- Santanu Panigrahi

Roll no- 2023JULB01148

Pre-Report for Super Store Sales Datasheet.

Problem statement

To contribute to the success of business by utilizing data analysis technique, specifically focusing on time series analysis to provide sales forecasting.

Data requirement:

- **Historical Sales Data:** Tracks sales details such as dates, products, quantities, revenue, discounts, regions, and channels over 1-2 years for forecasting trends and patterns.
- **Customer Data:** Includes customer demographics, purchase habits, transaction values, and loyalty program insights to understand and segment the customer base.
- **Product Information:** Contains details about products, their categories, lifecycle events (launch/discontinuation), and promotional or seasonal trends.

Data Collection:

- **Internal Sources:** Data from ERP/CRM systems, POS systems, and accounting software provides insights into sales, inventory, and financial performance.
- **External Sources:** Market research reports and public datasets enrich internal data with industry trends, economic factors, or contextual insights like weather.

Data Validation:

1. **Check for Nulls**:
    - Identify and handle missing data using the **Transform** or **Replace Values** options.
    - Example: Replace null sales quantity with 0, if valid.
2. **Data Types**:

- Ensure each column has the correct data type (e.g., dates, numbers, text).

3. **Remove Duplicates**:
    - Use the **Remove Duplicates** feature in Power Query.
4. **Apply Filters**:
    - Filter out invalid data (e.g., negative sales values, future dates).

Data Profiling

- Enable **Column Quality, Distribution, and Profile** in Power Query to:
    - Detect errors, blanks, and duplicates.
    - Visualize data distributions to spot anomalies.

Data cleaning:

- Remove or impute any outliers or anomalous data points.
- Normalize or standardize continuous variables such as cholesterol, blood pressure, and heart rate.
- Encode categorical features (e.g., one-hot encoding for chest pain type).

TOOLS.

- Use Pandas and NumPy for data manipulation, Matplotlib and Seaborn for creating visualizations, and Scikit-learn for building machine learning models.
- Can use Employ Tableau or Power BI to design interactive dashboards and effectively showcase data insights.

Dashboard.

- Understand the audience's needs and technical expertise to tailor the design.
- Use Power Query for transformations and to handle missing or inconsistent data.

- Ensure proper spacing and alignment of visuals for a polished appearance.
- Use efficient data models and aggregations to enhance performance.
- Test the dashboard for functionality and accuracy of data.
- Monitor its usage and relevance, refining the design or adding new features as needed.

Story telling:

Incorporated data analysis technique, specializing in time series analysis, to deliver valuable insights, accurate sales forecasting, and interactive dashboard creation, driving business success.