

Power Bi

Report

By- Santanu Panigrahi

Roll no- 2023JULB01148

S.NO.	CONTENTS
1.	Problem Statement
2.	Data Requirement
3.	Data Collection
4.	Data Validation
5.	Data Cleaning
6.	Tools
7.	Dashboard Analysis
8.	Storytelling
9.	Recommendations
10.	Insights of Each Graph

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.

Predict Sales Trends: Accurately forecast future sales trends to improve inventory management and operational efficiency.

Support Strategic Decisions: Provide data-driven insights to guide strategic decisions, such as marketing campaigns, resource allocation, and product launches.

Identify Seasonal Patterns: Analyze historical data to uncover recurring sales patterns and seasonal demand, ensuring readiness for high-demand periods.

Optimize Revenue: Utilize forecasting to minimize stockouts or overstocking situations, thereby maximizing revenue and customer satisfaction.

Understand Customer Behavior: Integrate customer purchasing habits into the analysis to tailor marketing efforts and improve retention.

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:

- Filter out invalid data (e.g., negative sales values, future dates). 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:

Detect Errors, Blanks, and Duplicates:

- Identify incomplete or incorrect data entries.
- Highlight missing values and allow for efficient imputation or removal.
- Eliminate duplicate records to maintain data integrity.

Visualize Data Distributions to Spot Anomalies:

- Use histograms or box plots to observe data spread and identify outliers.
- Detect inconsistencies, such as unusual spikes or dips in numeric fields.

Assess Data Consistency:

- Ensure uniformity in categorical values (e.g., standardize variations like "USA," "US," and "United States").
- Confirm consistent date formats across the dataset.

Validate Data Ranges:

- Highlight values falling outside expected thresholds (e.g., negative sales values or unrealistically high quantities).

Analyze Relationships Between Columns:

- Check for correlations between columns to validate logical relationships (e.g., higher discounts corresponding to increased sales).

Monitor Data Trends Over Time:

- Spot irregularities in time-series data, such as missing time periods or unexpected trends.

Evaluate Data Completeness:

- Measure how much of the dataset is populated versus missing, enabling informed decisions about additional data collection needs.

Streamline Data Preparation:

- Identify potential preprocessing steps, such as standardization or normalization, directly through profiling insights

Data cleaning:

Remove or Impute Outliers or Anomalous Data Points:

- Apply statistical methods like Z-scores or the IQR (Interquartile Range) to detect and handle outliers.
- Use domain knowledge to decide whether to exclude or replace outliers with median or mean values.

Normalize or Standardize Continuous Variables:

- Rescale continuous data to a common range (e.g., 0-1) for algorithms sensitive to magnitude.
- Apply z-score normalization to ensure all variables have a mean of 0 and a standard deviation of 1 for balanced analysis.

Encode Categorical Features:

- Use one-hot encoding for nominal data to create binary columns for each category.
- Implement label encoding for ordinal features to maintain the inherent order in the data.

Handle Missing Values:

- Use techniques such as mean/mode imputation for numerical data and the most frequent category for categorical data.
- For time-series data, apply forward or backward fill methods to maintain sequence integrity.

Standardize Date and Time Formats:

- Convert dates into a consistent format (e.g., YYYY-MM-DD) to ensure compatibility across tools.
- Extract additional features such as day, month, year, or week number for temporal analysis.

TOOLS.

Data Manipulation and Processing:

- **Pandas:** For handling large datasets with ease, including filtering, grouping, and joining data.
- **NumPy:** Efficient for numerical computations, handling arrays, and performing mathematical operations.^[2]

Data Visualization:

- **Matplotlib:** Ideal for creating static, customizable plots such as line graphs, bar charts, and scatter plots.
- **Seaborn:** Built on Matplotlib, it simplifies creating aesthetically pleasing and informative statistical graphics (e.g., heatmaps, box plots).

Machine Learning and Statistical Analysis:

- **Scikit-learn:** For building, training, and evaluating predictive models (e.g., regression, classification).
- **Stats models:** For detailed statistical tests, time series analysis, and regression modeling.

Interactive Dashboards and Reporting:

- **Power BI:** User-friendly, robust for connecting multiple data sources, and creating dynamic dashboards with drill-down features.
- **Tableau:** Provides advanced visuals, interactive storytelling, and ease of deployment for sharing insights.
- **Google Data Studio:** A free tool for creating simple, shareable dashboards.

Data Storage and Querying:

- ? **SQL:** Essential for querying relational databases, filtering, aggregating, and joining datasets directly from source systems.
- ? **Google Big Query/Amazon Redshift:** For analyzing large-scale datasets using SQL in cloud environments.

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.

Insights:

1. Overall Sales Performance

- **Total Sales:** Displays the total revenue generated within a specific period.
- **Growth Trends:** Highlights year-over-year or month-over-month growth.

2. Regional Analysis

- **Sales by Region:** Identifies top-performing regions and regions needing attention.
- **Contribution Analysis:** Compares sales contributions from different geographical areas.

3. Product Performance

- **Top-Selling Products:** Lists products that drive the highest revenue.
- **Low-Performing Products:** Highlights products with declining or stagnant sales.

4. Customer Insights

- **Customer Segmentation:** Analyzes customer demographics and behaviors.
- **Repeat Customers:** Tracks the percentage of repeat buyers and their contribution to sales.

5. Channel Performance

- **Sales by Channel:** Compares sales across different channels (e.g., online, offline, distributors).
- **Efficiency Metrics:** Evaluates which channels are most effective in reaching sales targets.

6. Time-Based Trends

- **Seasonality:** Displays patterns of sales spikes or dips across seasons.
- **Daily/Monthly Performance:** Helps identify high-performing days or months.

7. Discounts and Promotions

- **Impact of Discounts:** Measures how discounts influence sales and profit margins.
- **Promotional Effectiveness:** Evaluates which campaigns were successful.

8. Inventory Insights

- **Stock Levels:** Tracks inventory turnover and stockout risks.
- **Demand Forecasting:** Predicts future inventory needs based on sales patterns.

9. Key Performance Indicators (KPIs)

- **Conversion Rates:** Tracks how well leads or potential buyers are converted into actual sales.
- **Profit Margins:** Provides insights into the profitability of the business.

10. Actionable Insights

- Highlights areas requiring immediate action, such as:
 - Regions with declining sales.
 - Products that need better marketing support.
 - Over-reliance on discounts affecting profit margins.

Recommendations:

1. Sales Growth Strategies

- **Expand High-Performing Areas:** Increase investments in regions or channels that consistently drive the most revenue.
- **Focus on Top Products:** Allocate more resources to promote best-selling products and optimize their supply chain.
- **Cross-Selling and Upselling:** Introduce bundles or complementary products to increase basket size.

2. Customer-Centric Improvements

- **Personalized Marketing:** Leverage customer segmentation insights to deliver targeted promotions and personalized offers.
- **Loyalty Programs:** Enhance loyalty rewards for repeat customers to improve retention and lifetime value.
- **Feedback Loops:** Collect customer feedback for products with low performance and make necessary improvements.

3. Inventory Optimization

- **Forecast Demand Accurately:** Use historical sales trends to predict demand and avoid stockouts or overstocking.
- **Reduce Carrying Costs:** Eliminate or discount slow-moving inventory to free up warehouse space.
- **Prioritize High-Demand Items:** Focus inventory replenishment efforts on products with consistent or rising demand.

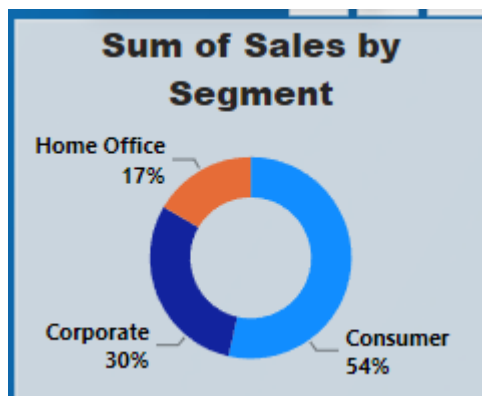
4. Promotional Effectiveness

- **Optimize Discount Levels:** Evaluate the impact of discounts on profitability and avoid excessive markdowns.
- **Seasonal Campaigns:** Align marketing and inventory strategies with identified seasonal peaks.
- **A/B Test Promotions:** Test different offers to find the most effective strategies for boosting sales.

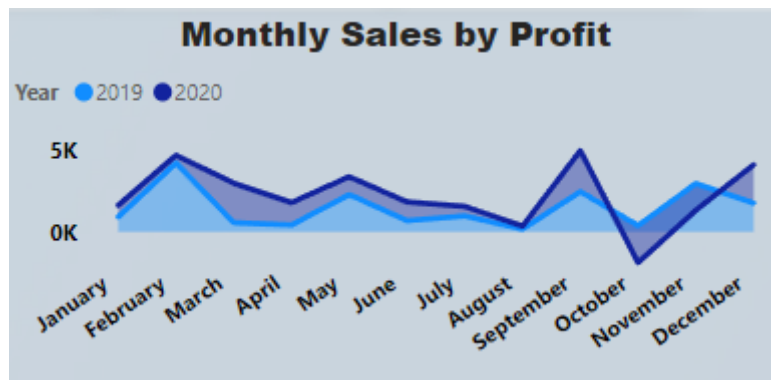
5. Regional Strategy Adjustments

- **Invest in Underperforming Areas:** Analyze reasons for poor performance and tailor strategies for those regions (e.g., localized advertising).
- **Expand in Growing Markets:** Increase presence in regions with upward sales trends.

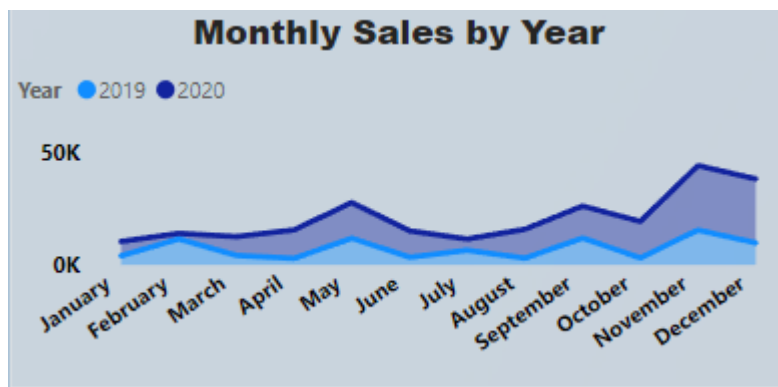
Insights: (of south region)



- **Consumer Segment** is the largest contributor to sales, accounting for more than half of the total. Businesses should continue focusing efforts here to sustain and grow revenue.
- **Corporate Segment** represents a significant portion, but there might be potential to expand further in this area.
- **Home Office Segment** contributes the least, signalling either limited market size or untapped potential.



- **Capitalize on Peak Periods:** Focus marketing and operational efforts in October and December to maximize profits.
- **Address Dips:** Investigate reasons behind lower performance in July 2020 to mitigate similar risks in the future.
- **Sustain Improvements:** Continue leveraging strategies that resulted in 2020's higher profitability in certain months.



- **Leverage Peak Months:** Focus marketing, discounts, and campaigns around November and December to sustain and maximize sales growth.
- **Boost Mid-Year Sales:** Analyse and address factors limiting July and August growth.
- **Replicate Success:** Study the strategies or external factors contributing to sharp increases in October and December 2020 to replicate similar outcomes.