

Supermarket Data Analysis Project Documentation

Table of Contents

1. [Introduction](#)
2. [Project Goals](#)
3. [Data Description](#)
4. [Tools and Technologies Used](#)
5. [Data Analysis Process](#)
6. [Dashboard Features](#)
7. [Results and Insights](#)
8. [Challenges and Solutions](#)
9. [Future Work](#)
10. [Conclusion](#)
11. [Appendix](#)

Introduction

This project involves a comprehensive analysis of supermarket data, conducted using the Bespos software from Sesame Technologies, which is an online billing, inventory, and accounting platform. The analysis aims to derive insights that help in understanding customer behaviour, product demand, and store performance, ultimately assisting store managers in making data-driven decisions to improve profitability.

Project Goals

The main objectives of this project are:

- Analyze and identify the most popular products in individual stores.
- Calculate and compare the average transaction value by store.
- Determine the top stores with the most repeat customers.
- Identify the top 10 popular product categories.
- Analyze the yearly peak sales periods.
- Understand customer spending habits.
- Identify the least bought products in each store.
- Determine the top 5 most profitable products by store.

- Make predictions for product recommendations, stock requirements, and customer traffic for the upcoming month.

Data Description

The analysis utilizes the following tables from the Bespos software:

- **Product Data Table:** Contains information about the products available in the store.
- **Product Batch Data Table:** Tracks batch-specific information, such as expiry dates.
- **Transaction Table:** Includes details of all transactions made in the stores.
- **Transaction Product Data Table:** Provides a breakdown of products involved in each transaction.

Data Cleaning and Preparation

- Unwanted data was removed from the tables, and some columns were renamed for clarity.
- The Transaction Product Data Table was split into multiple tables for easier processing.
- Dummy data were added to columns such as `expiry_date` and `customer_id` to fill in missing information.

Tools and Technologies Used

The following tools and libraries were employed for data analysis and visualization:

- **Python:** Main programming language used.
- **Pandas:** For data manipulation and analysis.
- **NumPy:** For numerical computations.
- **Matplotlib & Seaborn:** For creating static visualizations.
- **Dash (Plotly):** For creating interactive dashboards.
- **Django:** For building a web-based dashboard and frontend interface.
- **HTML/CSS:** For designing and styling the web interface.
- **Jupyter Notebook:** Used for development and testing.

Data Analysis Process

The analysis was conducted in the following stages:

Descriptive Analysis

- **Most Popular Products:** Identified the best-selling products in each store.

- **Average Transaction Value by Store:** Calculated and compared the average spend per transaction across different stores.
- **Repeat Customer Analysis:** Ranked stores based on the number of repeat customers.
- **Product Category Popularity:** Determined the top 10 most popular product categories.
- **Yearly Sales Peaks:** Identified periods of peak sales across the year.
- **Customer Spending Habits:** Analyzed patterns in how customers spend across different stores.
- **Least Bought Products:** Highlighted products with the least sales in each store.
- **Most Profitable Products:** Identified the top 5 products that generate the highest profit in each store.

Predictive Analysis

- **Product Recommendations:** Suggested products for individual customers based on past purchases.
- **Stock Requirements:** Predicted the quantity of products required for the upcoming month.
- **Customer Traffic:** Forecasted the average number of customers expected to visit each store next month.
- **Profit Optimization:** Recommended adjustments to product quantities to maximize profits.

Dashboard Features

Two dashboards were created, one using Dash and the other using Django, each with unique features:

Dash Dashboard

- **Interactive Features:** Dash offers built-in options for zooming, downloading, panning, and auto-scaling.
- **Pages Designed:** Login page, All Store Analysis page, and Store-Specific Analysis page.
- **Graphical Representation:** More attractive and user-appealing graphs created with Dash.

Django Dashboard

- **Comprehensive Web Interface:** Designed using Django, with a Home page, Login page, Registration page, All Store Analysis page, and Store-Specific Analysis page.

- **Store Selection:** Users can select individual stores for specific analysis on the Store-Specific Analysis page.
- **Graph Integration:** Displays all graphs mentioned, including market basket analysis and sales growth trends.

Results and Insights

Key findings include:

- Identification of high-demand products and peak sales periods, which can be used to optimize inventory.
- Insights into customer behaviour, including spending habits and repeat customers, helping to tailor marketing strategies.
- Predictions for future sales and customer traffic, assisting in resource planning.

Challenges and Solutions

Challenges

- **Complex Datasets:** Managing and processing large datasets with millions of rows.
- **Data Gaps:** Missing data in critical columns like expiry_date and customer_id required the addition of dummy data.
- **Frontend Development:** Difficulty in handling the frontend, particularly as this was the first experience with it.

Solutions

- **Data Management:** Split large tables for more manageable processing.
- **Frontend Learning:** Overcame frontend challenges by learning and applying new techniques.

Future Work

The project can be further enhanced by:

- Integrating AI to provide more personalized recommendations.
- Generating automated summaries from graphs for easier interpretation.
- Improving the user interface to be more visually appealing and user-friendly.

Conclusion

This project provides a powerful tool for store managers, enabling them to analyze past performance, predict future needs, and make informed decisions to increase profitability. The dashboards created offer a user-friendly interface to explore and interpret the analysis results.

Appendix

- **Code Snippets:** [Link to code repository](#)
- **Dashboard Screenshots:** [Link to Screenshots](#)