Customer Segmentation and Sales Forecasting Using Data Engineering Techniques

Team members:

1. Mahmoud Adel Saleh
2. Ahmed Mohamed Elsiad
3. Habiba Mohamed
4. Sara Khaled Amin
5. Nada Amin Abdelhay

Supervised by:

Dr. Moshira Ghaleb

1. Introduction

The retail industry faces challenges such as managing inventory, accurately forecasting demand, and targeting the right customer segments. This project provides a comprehensive solution to these issues by using data engineering techniques to clean, process, and analyze retail data for sales forecasting and customer segmentation. The system is designed to assist businesses in making data-driven decisions through an intuitive interface and actionable insights.

Key Features:

- **Customer Segmentation:** Automatically group customers based on their purchasing behavior and characteristics.
- **Sales Forecasting**: The Prophet algorithm predicts future sales trends, allowing businesses to optimize inventory management and make long-term strategic plans.
- **Data Management:** Clean, process, and store data efficiently using Azure Data Warehouse.
- **Interactive Dashboards:** Visualize customer segments, sales performance, and trends through Power BI, providing businesses with a clear understanding of key metrics.
- **Scalability:** Support for handling large datasets, ensuring smooth operations even with growing data volumes.

Target Users:

- **Business Analysts** seeking to derive actionable insights from customer data.
- **Marketing Teams** aiming to personalize marketing efforts based on customer segmentation.
- **Data Scientists** looking for a platform to process, analyze, and model large datasets effectively.

2. System Requirements

For the **Customer Segmentation and Sales Forecasting** software to run effectively, the following hardware and software configurations are recommended:

- **Operating System:**
  - Minimum: Windows 10 or macOS X 10.14
  - Recommended: Windows 11 or macOS X 11.0 and above
- **Processor:**
  - Minimum: Intel Core i3 or AMD Ryzen 3
  - Recommended: Intel Core i5 or higher, or AMD Ryzen 5 and above
- **Memory:**
  - Minimum: 4 GB RAM
  - Recommended: 8 GB RAM or more (for handling large datasets and model training)
- **Storage:**

  Minimum: 500 MB of available disk space

  Recommended: 1 GB or more (for storing datasets and results)

- **Azure Subscription:**
  An active Azure subscription is required to provision resources such as Azure SQL Database and Azure Data Warehouse for data processing and storage.
- **Power BI Account (or Tableau):**
  A Power BI or Tableau account is required for creating interactive dashboards to visualize customer segmentation.
- **Python Environment:**
  Ensure that Python is installed and configured with the following libraries for data analysis and machine learning:

`numpy` (for numerical operations)

`pandas` (for data manipulation)

`scikit-learn` (for machine learning models)

`matplotlib` or `seaborn` (for visualizing results)

- **CSV Datasets:**
  Ensure that the required datasets (e.g., Customers.csv, Sales.csv, Products.csv) are available and accessible either locally or from cloud storage.
- **SQL Management Tools:**
  Tools like Azure Data Studio or SQL Server Management Studio (SSMS) are necessary to query and manage the Azure SQL Database or Data Warehouse.
- **Google Colab or Jupyter Notebook:**
  For data preprocessing and exploratory data analysis, a Google Colab or Jupyter Notebook environment will be used.
- **Additional Requirements:**
  - **Internet Connection:** Required for accessing cloud services (e.g., Azure Data Warehouse) and model deployment.
  - **Browser Compatibility:** Google Chrome, Mozilla Firefox, or Microsoft Edge for accessing the dashboard.

3. Installation Guide

Follow these steps to set up the environment for the **Customer Segmentation** software:

1. **Set up Azure Resources:**
   - Log in to your Azure account.
   - Create an Azure SQL Database for storing cleaned data.
   - Provision other necessary resources, such as Azure Data Factory for ETL processes.
2. **Set up Python Environment:**
   - Install Python (or use Google Colab).
   - Install the required Python libraries by running the following command in the terminal or Colab:
     ```
     pip install numpy pandas scikit-learn matplotlib seaborn
     ```
3. **Download Datasets:**
   - Ensure access to the required datasets (e.g., Customers.csv, Sales.csv, Products.csv).
   - Store the datasets either locally or upload them to a cloud storage solution (e.g., Azure Blob Storage).
4. **Set up Power BI (or Tableau):**
   - Sign in or create an account on Power BI (or Tableau).
   - Download and install the desktop version for local report creation.
5. **Connect to Azure SQL Database:**
   - Using SQL management tools like Azure Data Studio or SSMS, connect to your Azure SQL Database.
   - Load the cleaned data into the database from your Python environment using appropriate SQL queries.
6. **Set up Dashboards:**
   - Open Power BI (or Tableau) and connect it to your Azure SQL Database.

- Start building visualizations to display customer segments and predicted prices.

This guide walks users through setting up the tools and resources for your project. Let me know if you need further clarification or adjustments!

4. Features Overview

The **Customer Segmentation and Price Prediction** software provides several key features to help businesses gain insights from their customer data and make informed decisions.

1. **Customer Segmentation:**
   - Automatically group customers based on various attributes like purchasing behavior, frequency, and demographic details.
   - Provides actionable insights to help businesses target specific customer groups for personalized marketing campaigns.
2. **Data Management and Processing:**
   - Efficiently manage and process large datasets using a structured pipeline, including data cleaning, transformation, and loading into the Azure Data Warehouse.
   - Ensures seamless integration with cloud-based resources for scalable data operations.
3. **Sales Forecasting:**
   - Leverage the Prophet algorithm to predict future sales trends.
   - Helps businesses optimize inventory management and plan for future demand.
4. **Interactive Dashboards:**
   - Use Power BI to visualize customer segments, regional sales performance, and overall trends.
   - Interactive dashboards allow businesses to easily monitor key metrics and trends over time.
5. **Machine Learning Integration:**
   - Integrate machine learning models into the data pipeline using libraries such as scikit-learn.
   - Models include K-Means for segmentation and Linear Regression for price prediction.
6. **Real-time Data Updates:**
   - Sync with Azure Data Warehouse to reflect the latest data in visualizations and predictions.
   - Ensure decisions are based on up-to-date insights.
7. **Scalability and Flexibility:**
   - Designed to handle growing data volumes, supporting scalable solutions through Azure's cloud infrastructure.
   - Easily adaptable to various types of datasets and business needs, allowing for customization and expansion as required.

5. User Interface Guide

Main Features:

- **Dashboard**: View overall sales performance, customer segments, and trends.
- **Customer Segmentation**: Visualize and analyze different customer groups based on behavior.
- **Sales Forecasting**: Access predicted sales trends and insights for future planning.
- **Settings**: Manage user preferences and system configurations.

The user interface is designed for ease of use, ensuring that business users without deep technical knowledge can interact with the data, generate reports, and extract valuable insights for decision-making.

6. Conclusion

This project delivers a full-scale solution for the retail industry, combining data cleaning, machine learning, and data visualization to empower businesses to make informed decisions. By using advanced forecasting and segmentation techniques, businesses can better manage their inventory, target the right customer segments, and anticipate future trends.

The software's ability to handle large datasets, provide real-time updates, and offer actionable insights makes it an invaluable tool for retailers, marketing teams, and inventory managers.