Customer Data Analysis and Improvement

This project aims to enhance customer data management and analysis through a series of well-defined work packages. The goal is to build a comprehensive system for storing, analyzing, and extracting valuable insights from customer data using SQL databases, Python, Azure services, Power BI, and MLOps.

Table of Contents

Project Overview

This project focuses on building a robust system for customer data analysis and improvement. It encompasses database setup, data analysis, and machine learning model deployment. The project is divided into four main work packages, each addressing a specific aspect of data management and analysis.

Dataset Overview

The dataset used in this project is sourced from Kaggle, titled Brazilian E-Commerce Public Dataset by Olist. It consists of extensive e-commerce data from the Olist platform, capturing customer information, order transactions, product reviews, and other activities. The dataset supports insights into customer behavior, sales trends, and more.

Dataset Highlights:

- **Customer Information**: Personal details and segmentation
- **Order History:** Purchase records and order tracking
- **Transactional Data:** Payments and shipping information
- **Product Reviews and Ratings:** Feedback and satisfaction metrics

Work Packages

Work Package 1: Data Management and SQL Database Setup

- Objective: Design and implement a SQL database to store customer-related information, including personal details, transactions, and interactions.
- Tools: Microsoft SQL Server, SQL Server Management Studio
- Deliverables:
    - SQL database schema design
    - Data population scripts
    - SQL queries for extracting and analyzing customer data

Work Package 2: Python Programming and Data Analysis

- **Objective:** Develop Python scripts to interact with the database and conduct comprehensive data analysis.
- **Tools:** Python (Pandas, SQLAlchemy)
- **Deliverables:**
    - Python scripts for data extraction, transformation, and analysis.

Work Package 3: Data Science, Azure Integration, and Power BI

- **Objective:** Conduct data analysis, build predictive models using Python, integrate Azure services, and create interactive reports using Power BI.
- **Tools:** Python (Scikit-learn, Matplotlib), Azure Data Studio, Azure Machine Learning, Power BI
- **Deliverables:**
    - Customer churn prediction model
    - Analysis report with insights and predictions
    - Power BI dashboards for visualizing customer insights

Work Package 4: MLOps, Deployment, and Final Presentation

- **Objective:** Implement MLOps for managing machine learning experiments and deploy the trained models for customer data predictions.
- **Tools:** MLflow, Azure services, Flask/Streamlit (for web application deployment)
- **Deliverables:**
    - Deployed machine learning model or web application for customer predictions.
    - MLflow setup for experiment tracking.

WebPage : https://broccoli-gudtf6dkf2bsdxbm.uaenorth-01.azurewebsites.net/

Technologies Used

- **Database:** Microsoft SQL Server, SQL Server Management Studio.
- **Python:** Pandas, SQLAlchemy, Scikit-learn, Matplotlib.
- **Azure:** Azure SQL Database, Azure Web Service.
- **Power BI:** For creating interactive reports.
- **MLOps:** MLflow for experiment tracking and model management.
- **Web Framework:** Flask/Streamlit (for model deployment).

Deliverables

- **SQL Database:** Schema design and SQL queries for data management and extraction.
- **Python Scripts:** For data extraction, cleaning, and analysis.
- **AI Models:** Trained predictive models for customer analysis.
- **Web Application:** A deployed application using Flask for making predictions with the trained machine learning models.
- **Final Presentation:** Comprehensive project presentation.

Team Members

1. **AbdulRahman Hesham Kamel**
2. **Mohamed Ibrahim Mohamed**
3. **Mohamed Ashraf Safwat**
4. **Ahmed Mohamed Abdelmoneim**