**Bootcamp Project 3: Customer 360 Data Integration Using Azure Synapse and SQL Analytics**

**Project Objective**

The goal of this project was to build a **360-degree customer profile** by integrating various data sources related to customer behavior, purchases, and interactions. The unified view helps the business make **data-driven decisions** around marketing, customer retention, loyalty rewards, and agent performance.

**Architecture Overview**

A multi-layered architecture was implemented to organize and process data efficiently. This followed the **Medallion Architecture Pattern**:

**🔸 Bronze Layer (Raw)**

**A screenshot of a computer

AI-generated content may be incorrect.**

Data was ingested into Azure Data Lake Storage (ADLS)

**🔹 Silver Layer**

**A screenshot of a computer

AI-generated content may be incorrect.**

* **Database Creation:**  
  The script begins by creating a new database named DEMODB using the CREATE DATABASE statement.
* **Switching Database Context:**  
  The context is switched to the newly created DEMODB database with the USE DEMODB; command. This ensures that all subsequent operations are performed within this database.
* **Master Key Creation:**  
  A master key is created for the database using the CREATE MASTER KEY ENCRYPTION BY PASSWORD command. This key is essential for encrypting credentials and other secrets within the database. The password used for encryption is provided as 'Aniketh08@'.
* **Database Scoped Credential for ADLS:**  
  The script creates a database scoped credential named MyADLSCredential. This credential is configured to use a Shared Access Signature (SAS) token for authentication with Azure Data Lake Storage. The IDENTITY is set to 'SHARED ACCESS SIGNATURE', and the SECRET contains the SAS token (represented here as a placeholder).

**A screenshot of a computer program

AI-generated content may be incorrect.**

A screenshot of a computer code

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.Here I have Created an external data source in SQL to connect to Azure Data Lake Storage (ADLS) and set up a CSV file format for reading external tables. The script specifies the ADLS path and credential, then creates a file format with options for comma-separated values, double quotes as string delimiters, and skips the header row.

Used SQL commands for creating external tables that let you access customer and product data stored in CSV files on Azure Data Lake Storage. The tables reference the external data source and file format, so you can query the data directly without importing it into the database.

A screenshot of a computer program

AI-generated content may be incorrect.

A white background with green and blue text

AI-generated content may be incorrect.

**We have Performed few Cleaning techniques such as:**

* Duplicate customers are removed by keeping only the first record for each CustomerID using a common table expression (CTE) with the ROW\_NUMBER() function.
* Product data is cleaned by replacing null values: missing names become 'Unknown', missing categories become 'Miscellaneous', and missing prices are set to 0.
* Loyalty account data is validated by attempting to cast the PointsEarned column to an integer using TRY\_CAST, ensuring only valid numeric values are processed.
* This validates the PointsEarned column in the LoyaltyAccounts table by using TRY\_CAST to safely convert its values to integers. If a value cannot be converted, TRY\_CAST returns NULL instead of causing an error, ensuring only valid numeric data is used for PointsEarned.

**🥇 Gold Layer (Analytics)**

A screenshot of a computer screen

AI-generated content may be incorrect.A screenshot of a computer code

AI-generated content may be incorrect.

* A screenshot of a graph

  AI-generated content may be incorrect.The vw\_AverageOrderValue view calculates key sales metrics for each product and category. It provides the order count, total revenue, and average order value by joining online transactions with the products table and grouping the results by product and category.

A screenshot of a computer code

AI-generated content may be incorrect.

* The vw\_CustomerSegmentation view segments customers based on their total spend, purchase frequency, and loyalty tier. It combines data from online and in-store transactions, as well as loyalty accounts, to compute each customer's total spend (handling missing values with ISNULL), the number of purchases, and their loyalty tier.
* A screenshot of a calendar

  AI-generated content may be incorrect.A screenshot of a computer program

  AI-generated content may be incorrect.The  vw\_PeakPurchaseTimes view, analyzes transaction volumes by day of the week and hour for both online and in-store channels. It combines data from both sources and groups transactions to help identify peak shopping times.
* A screenshot of a white sheet

  AI-generated content may be incorrect.A white screen with colorful text

  AI-generated content may be incorrect.
* The  vw\_AgentPerformance view evaluates customer service agent performance. It calculates the total number of interactions handled by each agent, the number resolved, and the success rate, providing insight into agent effectiveness.

A screenshot of a graph

AI-generated content may be incorrect.

**Agent Performance**

A screenshot of a computer

AI-generated content may be incorrect.

**Average Order Value**

A screenshot of a graph

AI-generated content may be incorrect.

**Customer Segmentation**

A graph with pink squares

AI-generated content may be incorrect.

**Peak Purchase Times**