

Executive Summary

This report follows the completion of an in-depth analysis aimed at segmenting the customer base using clustering techniques. Based on Part One of the assessment, the dataset was enhanced through additional feature engineering, incorporating a mix of transactional, demographic, and hybrid attributes.

Among the clustering algorithms evaluated—K-Means, HDBSCAN, and OPTICS—K-Means was identified as the most effective, dividing customers into three segments. These segments were categorized based on transactional behaviors and demographic characteristics:

- **Cluster 0:** Characterized by average median annual wages, middle-range age, and moderate work experience. This cluster exhibited lower deposit activity but notable loan payment activities.
- **Cluster 1:** Consists of older, more experienced customers with higher wages compared to other clusters.
- **Cluster 2:** The youngest segment, with the least experience, is likely at the early stages of their careers.

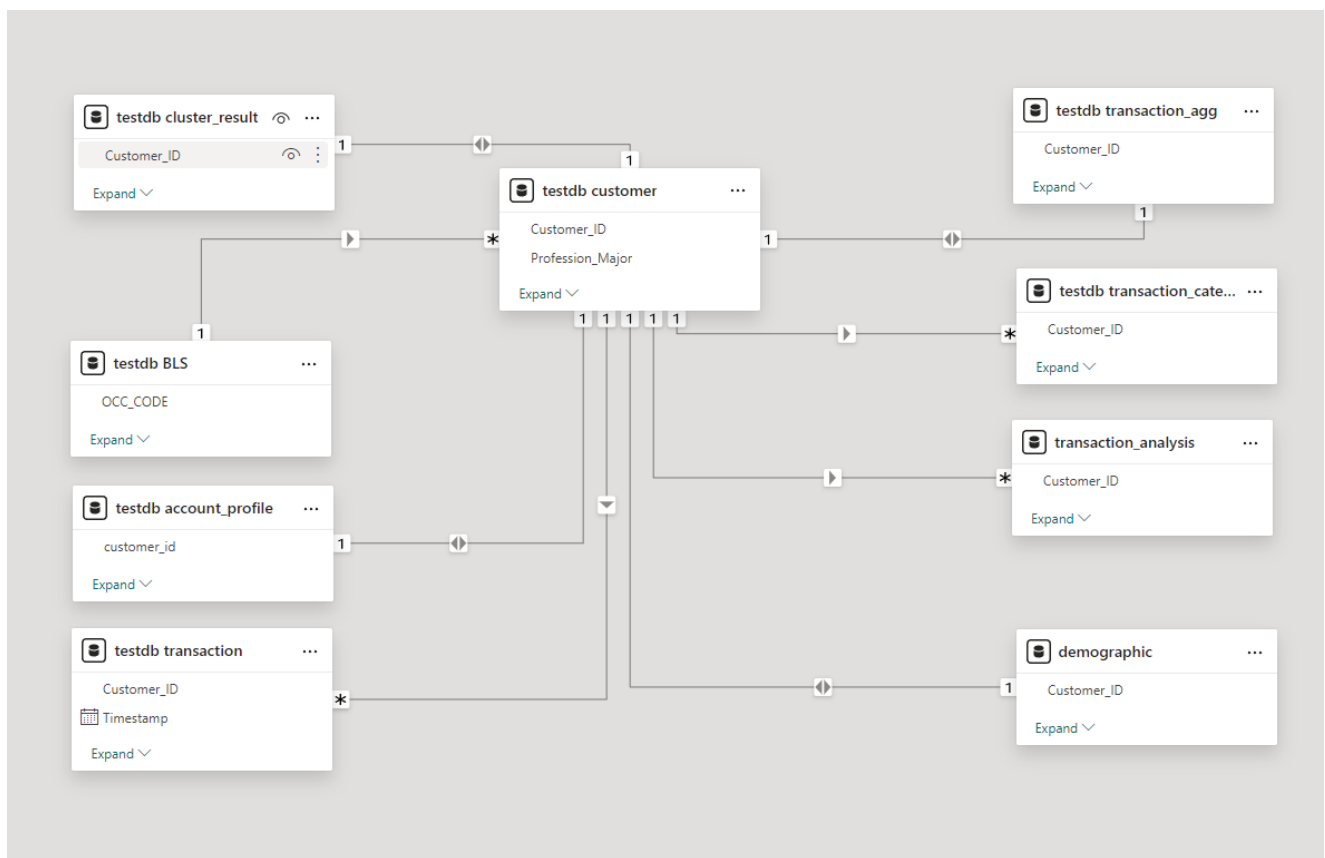
The accompanying dashboard offers a direct visualization of these findings. It includes:

- **Overview Page** that displays the distribution of customer segments and fundamental transactional and demographic data for each group.
- **Transaction Page** and **Demographic Page** provide a deeper dive into the respective behaviors and characteristics of each cluster, offering detailed insights into various key metrics.

This documentation serves to provide a comprehensive understanding of the dashboard's structure, its data sources, and the insights it intends to deliver.

Data Model Description

Below is a diagram of the data tables utilized in the dashboard.



Database Connection

Data is stored in a MySQL database, and a direct connection is established using Power BI's MySQL database connector. This setup allows real-time data fetching from the following tables into Power BI.

Source Table

- **Customer:** Contains detailed demographic and employment information for each customer, including age, family size, gender, annual wage, and profession code. Each record is uniquely identified by a Customer ID.
- **BLS (US Bureau of Labor Statistics):** Provides wage and employment statistics keyed to each profession code. It enriches customer records with detailed occupational data.
- **Transaction:** Logs every transaction per customer, categorized into types like Card, Check, Deposit, Loan Payment, Transfer, and Withdrawal. Each transaction is uniquely identified and linked to a customer ID.
- **Account Profile:** Derived from the Transaction table, it aggregates data by customer ID and transaction type to calculate average transaction amounts and counts.
- **Cluster Result:** Stores outcomes from the clustering model, assigning each customer to a cluster and recording key features used in the model for easy reference and visualization.

Transformed Table

- **Transaction Aggregate:** A transformation of the **Transaction** table, aggregating data by customer ID to compute average transaction amounts and counts, aiding in the analysis of transaction frequencies per customer.
- **Transaction Category Aggregate:** A pivot from the **Account Profile** table designed to optimize the schema for further transformations and merging. Fields include customer ID, transaction type, sum amount, average amount, and count.
- **Transaction Analysis:** Merges the **Transaction Category Aggregate** with unpivoted data from the Cluster Result table. It provides a comprehensive view of each customer's transaction behavior by type, heavily utilized in the transaction behavior analysis section of the dashboard.
- **Demographic:** Combines data from the **BLS**, **Customer**, and **Cluster Result** tables to provide demographic insights such as age, family size, and median annual wages. This table supports demographic analysis on the dashboard.

Dashboard Design

The dashboard is structured into three main pages, each focusing on different aspects of customer data: Customer Segment Analysis Overview, Transaction Behavior Analysis, and Customer Demographic Analysis. These pages are designed to provide a holistic view of customer segments, their transaction behaviors, and demographic profiles.

Customer Segment Analysis Overview:

This page offers a high-level summary of the customer segments, structured into three sections: Summary, Transactional, and Demographical.

- **Summary Section:** Displays the distribution of customers across each cluster alongside key observations derived from transactional and demographic analyses.
- **Transactional Section:** Provides a detailed breakdown of average transaction amounts and frequencies within each cluster, highlighting differences in spending behavior.
- **Demographical Section:** Focuses on demographic variables such as family size and work experience. Includes a scatter plot that illustrates the relationship between age and median annual wage for each cluster, offering insights into economic factors affecting each segment.

Transaction Behavior Analysis:

This page delves deeper into the transactional data, segmented by customer cluster and transaction type:

- **Observations:** Summarizes insights from subsequent charts and plots, providing a narrative on the data presented.
- **Key Metrics by Transaction Type:** Displays behavior metrics for each transaction type within each cluster. Accompanied by detailed observations, this section helps decode patterns in customer spending and transaction activities.
- **Distribution of Transaction Amounts and Counts:** Examines both overall and monthly averages of transaction amounts and counts, offering a comparative perspective on how transaction behaviors vary across clusters.

Customer Demographic Analysis:

Dedicated to exploring the demographic characteristics of each customer cluster, this page is divided into several analytical sections:

- **Observations:** Narrates the primary insights from the demographic data visualizations, guiding the viewer through the findings.
- **Number of Major Occupations:** Shows the prevalent occupations within each cluster, providing a contextual backdrop for the demographic composition of each segment.
- **Demographic Metrics Comparison:** Compares median annual wage, work experience, age, and other relevant metrics across clusters. This comparison highlights the demographic diversities and commonalities that define each customer group.

Published Dashboard

The dashboard is published through Power BI and can be accessible from [[Here](#)]

The interactive report can be accessed from [[Here](#)]