**Valley <> FCC Analytics Engineer – Take Home Assessment**

Scenario: You are working as an Analyst for a financial institution that seeks to understand customer behavior and transaction patterns. The company aims to segment its customers based on transaction behavior and demographics.

Data Sources:

* Transaction DataMart: Contains information about transactions, including transaction amount, timestamp, Customer ID, and transaction type.
* Customer Data: Includes demographic information such as age, gender, profession, work experience and family size.
* Occupational Employment and Wage Statistics: The US Bureau of Labor Statistics (BLS) data released in May 2022.

Tasks

1. Data Ingestion and Cleaning:
   * Ingest the provided transaction data (transaction) and customer data (customer) into a relational database system.
   * Obtain the Annual Median Wages from the US Bureau of Labor Statistics (BLS) and ingest it into the customer table (customer).
   * Ensure that the data is cleaned, normalized, and structured for analysis.
2. Stored Procedure Development:
   * Create Table Account Profile (account\_profile)
   * Write a stored procedure to calculate aggregates of transaction data, including transaction count and average amount per transaction type for each account and add it into a new table (account\_profile)
   * A group of colorful text

     Description automatically generated with medium confidence
   * Write an after-Insert Trigger, executing the stored procedure above and updating the account\_profile table after new data is added to the transaction table.
     + Assume transaction data is updated at the end of every month.
3. Customer Segmentation:
   * Utilize transaction behavior, demographic information, and annual median wage as features for clustering.
   * Analyze the data, perform clustering and present the results in Jupyter notebook.
   * Document the steps, findings and recommend the best solution.

Deliverables:

1. Screenshots of all the tables in the database (Five rows – top and bottom)
2. SQL scripts for stored procedure and trigger
3. Clustering analysis results and code shared in a Jupyter Notebook

Note:

Please specify the code block that was generated by ChatGPT, and the one developed by you. We recommend referring to GPT based tools only when necessary. We are hiring you and not AI.