**Credit Fraud / Credit Score Analysis**

**Objective:**

One of our clients, a credit lender in the USA, has shared some csv files containing historical data as well as customer information with us. Our team is required to work on this data to extract some meaning full insights from the shared data so that the bank can understand further about the credit scores of their customers as well as the credit fraud being committed.

You’ll create an automated dashboard to monitor fraud trends and generate customer credit score reports.

**Metadata:**

The following are the further details regarding the data that has been shared with us.

* Customers: Contains customer demographic and financial information
  + customer\_id: Unique ID for each customer.
  + first\_name, last\_name: names of the customers.
  + date\_of\_birth: date of birth of the customers.
  + email: emails of the customers.
  + phone\_number: phone numbers of the customers.
  + address: street address of the customers.
  + city: city of residence.
  + state: state of residence.
  + postal\_code: zip code.
  + country: country of residence.
  + annual\_income: annual income of the customers.
  + employment\_status: employment status of customers
  + account\_open\_date: date when the account was opened.
  + credit\_score: credit score of the customer.
* Transactions: Detailed credit card transactions information.
  + transaction\_id: unique ID for each transaction.
  + customer\_id: foreign key referencing the customers table.
  + transaction\_date: date and time of the transaction.
  + transaction\_amount: transaction amount.
  + merchant\_name: merchants company names.
  + merchant\_category: categories of business type of merchant.
  + transaction\_city, transaction\_state, transaction\_country: city, state, and country where the transaction occurred.
  + transaction\_status: status of the transaction.
* Accounts: Data about customer accounts.
  + account\_id: unique ID for each account.
  + customer\_id: foreign key referencing the customers table.
  + account\_type: account type.
  + credit\_limit: credit limit for the account.
  + balance: balance on the account according to the credit limit.
  + account\_status: account status.
  + delinquent: Boolean flag indicating whether the account is delinquent (account is behind on payments).
* Credit History: Credit repayment history and behavior of customers.
  + history\_id: unique ID for each credit history record.
  + customer\_id: foreign key referencing the customers table.
  + account\_id: foreign key referencing the accounts table.
  + payment\_date: date when the payment was made.
  + due\_amount: due amount.
  + payment\_amount: payment amount.
  + missed\_payment: Boolean flag indicating whether the payment was missed.
  + days\_late: number of days late, if the payment was missed.
* Fraud Records: Historical records of transactions labeled as fraudulent.
  + fraud\_id: Unique ID for each fraud record.
  + transaction\_id: foreign key referencing the transactions table.
  + fraud\_detected\_date: date when the fraud was detected.
  + fraud\_type: fraud type.
  + investigation\_status: investigation status.
  + fraud\_resolution: resolution for the fraud case.

**Steps:**

* **SQL: Write complex SQL queries to:**
  + Extract customer and transaction data based on defined features (e.g., monthly transaction totals, outlier detection on spending, etc.).
  + Perform joins between customers, transactions, accounts, and credit\_history to gather complete data per customer.
  + Generate new features such as:
  + Average transaction amounts by time periods (daily/weekly).
  + Transaction location consistency (using geographical data).
  + Delinquency rates from credit\_history.
* **Python: Data Preprocessing & Feature Engineering:**
  + Perform data cleaning: Handle missing data, outliers, inconsistent formats.
  + Create customer credit score features based on their credit\_history (e.g., credit utilization, payment history, etc.).
  + Classify customers into credit risk categories (e.g., low-risk, medium-risk, high-risk).
  + Construct fraud-specific features using transactions data (e.g., frequency of transactions, distance between transaction locations, time-of-day activity).
* **Visualization & Reporting (Power BI/Tableau):**
  + Create an interactive dashboard with the following:
    - Fraud Detection:
      * A heatmap of fraudulent transactions by location.
      * A time-series chart showing the trends of fraudulent activity over time.
      * A detailed drill-down report for suspicious customers and transactions.
    - Credit Scoring:
      * Bar charts or histograms visualizing customer credit score distribution.
      * Visualizations showing relationships between credit score and key features (income, account balance, payment history).
      * Scatter plots showing credit utilization vs. delinquency rates.