Data dictionary

Dataset Download link:
https://github.com/akgeoinsys/DataEngg-capstone

1. Credit Card Fraud Detection
Primary Dataset: credit_card_transactions.csv
Size: ~15 MB (50,000 rows)

Columns:

Transaction_ID: Unique ID for each transaction.

Customer_ID: Unique ID for each customer.

Card_Type: Type of credit card (e.g., " Visa ", "MasterCard ").

Transaction_Amount: Amount of the transaction (float, with outliers).

Transaction_Date: Date and time of the transaction (string with inconsistent formats).

Merchant: Name of the merchant (string with leading/trailing spaces).

Location: Location of the transaction (string with missing values).

Fraud_Label: Binary label (0 = legitimate, 1 = fraudulent).

Currency: Currency of the transaction (e.g., "USD", "EUR", missing values).

Transaction_Status: Status of the transaction (e.g., "Success", "Failed", "Pending").

Additional Dataset: customer_details.csv
Size: ~5 MB (10,000 rows)

Columns:

Customer_ID: Unique ID for each customer.

Name: Customer name (string with leading/trailing spaces).

Age: Age of the customer (integer, with missing values).

Gender: Gender of the customer (e.g., "Male", "Female", missing values).

Address: Customer address (string with missing values).

Phone_Number: Phone number (string with inconsistent formats).

Download Links:
credit_card_transactions.csv
Customer_details.csv

2. Loan Default Prediction

Primary Dataset: loan_data.csv
Size: ~10 MB (30,000 rows)

Columns:

Loan_ID: Unique ID for each loan.

Customer_ID: Unique ID for each customer.

Loan_Amount: Amount of the loan (float, with outliers).

Interest_Rate: Interest rate on the loan (float, with missing values).

Term: Loan term in months (integer, with missing values).

Employment_Status: Employment status of the customer (e.g., " Employed ", "Self-Employed", missing values).

Annual_Income: Annual income of the customer (float, with outliers).

Default_Status: Binary label (0 = not defaulted, 1 = defaulted).

Loan_Type: Type of loan (e.g., "Personal", "Business", missing values).

Application_Date: Date of loan application (string with inconsistent formats).

Additional Dataset: customer_employment.csv
Size: ~3 MB (10,000 rows)

Columns:

Customer_ID: Unique ID for each customer.

Job_Title: Job title (string with leading/trailing spaces).

Company_Name: Company name (string with missing values).

Years_of_Experience: Years of experience (integer, with missing values).

Download Links:
loan_data.csv
Customer_employment.csv

3. Insurance Claim Analysis

Primary Dataset: insurance_claims.csv
Size: ~12 MB (40,000 rows)

Columns:

Claim_ID: Unique ID for each claim.

Policy_ID: Unique ID for each policy.

Customer_ID: Unique ID for each customer.

Claim_Amount: Amount claimed (float, with outliers).

Claim_Date: Date of the claim (string with inconsistent formats).

Incident_Type: Type of incident (e.g., " Accident ", "Theft", missing values).

Location: Location of the incident (string with missing values).

Fraud_Flag: Binary label (0 = legitimate, 1 = fraudulent).

Claim_Status: Status of the claim (e.g., "Approved", "Rejected", "Pending").

Policy_Type: Type of policy (e.g., "Auto", "Home", missing values).

Additional Dataset: policy_details.csv
Size: ~4 MB (15,000 rows)

Columns:

Policy_ID: Unique ID for each policy.

Policy_Start_Date: Start date of the policy (string with inconsistent formats).

Policy_End_Date: End date of the policy (string with missing values).

Premium_Amount: Premium amount (float, with outliers).

Download Links:
insurance_claims.csv
Policy_details.csv

4. Customer Segmentation for Banking

Primary Dataset: customer_data.csv
Size: ~8 MB (25,000 rows)

Columns:

Customer_ID: Unique ID for each customer.

Age: Age of the customer (integer, with missing values).

Gender: Gender of the customer (e.g., " Male ", "Female", missing values).

Income: Annual income of the customer (float, with outliers).

Transaction_Frequency: Number of transactions per month (integer, with missing values).

Average_Transaction_Amount: Average amount per transaction (float, with outliers).

Segment: Customer segment (e.g., "Premium", "Standard", missing values).

Account_Type: Type of account (e.g., "Savings", "Current", missing values).

Last_Login_Date: Last login date (string with inconsistent formats).

Additional Dataset: transaction_details.csv
Size: ~6 MB (50,000 rows)

Columns:

Transaction_ID: Unique ID for each transaction.

Customer_ID: Unique ID for each customer.

Transaction_Date: Date of the transaction (string with inconsistent formats).

Transaction_Type: Type of transaction (e.g., "Deposit", "Withdrawal", missing values).

Amount: Transaction amount (float, with outliers).

Download Links:
customer_data.csv
Transaction_details.csv

5. Stock Market Data Analysis

Primary Dataset: stock_data.csv
Size: ~20 MB (50,000 rows)

Columns:

Stock_ID: Unique ID for each stock.

Stock_Name: Name of the stock (string with leading/trailing spaces).

Date: Date of the stock data (string with inconsistent formats).

Open_Price: Opening price of the stock (float, with outliers).

Close_Price: Closing price of the stock (float, with missing values).

High_Price: Highest price of the stock during the day (float, with outliers).

Low_Price: Lowest price of the stock during the day (float, with missing values).

Volume: Trading volume of the stock (integer, with missing values).

Market_Cap: Market capitalization (float, with outliers).

Additional Dataset: company_details.csv
Size: ~5 MB (10,000 rows)

Columns:

Stock_ID: Unique ID for each stock.

Company_Name: Name of the company (string with leading/trailing spaces).

Sector: Sector of the company (e.g., "Technology", "Healthcare", missing values).

CEO: CEO of the company (string with missing values).

Download Links:
stock_data.csv
Company_details.csv

Download link :

https://github.com/akgeoinsys/DataEngg-capstone