Problem statement:

To conduct a thorough exploratory data analysis (EDA) and deep analysis of a comprehensive dataset containing basic customer details and extensive credit-related information. The aim is to create new, informative features, calculate a hypothetical credit score, and uncover meaningful patterns, anomalies, and insights within the data.

Data Dictionary:

Column Name	Description
ID	Represents a unique identification of an entry
Customer_ID	Represents a unique identification of a person
Month	Represents the month of the year
Name	Represents the name of a person
Age	Represents the age of the person
SSN	Represents the social security number of a person
Occupation	Represents the occupation of the person
Annual_Income	Represents the annual income of the person
Monthly_Inhand_Salary	Represents the monthly base salary of a person
Num_Bank_Accounts	Represents the number of bank accounts a person holds
Num_Credit_Card	Represents the number of other credit cards held by a person
Interest_Rate	Represents the interest rate on credit card
Num_of_Loan	Represents the number of loans taken from the bank
Type_of_Loan	Represents the types of loan taken by a person
Delay_from_due_date	Represents the average number of days delayed from the payment date
Num_of_Delayed_Payment	Represents the average number of payments delayed by a person
Changed_Credit_Limit	Represents the percentage change in credit card limit
Num_Credit_Inquiries	Represents the number of credit card inquiries
Credit_Mix	Represents the classification of the mix of credits
Outstanding_Debt	Represents the remaining debt to be paid (in USD)
Credit_Utilization_Ratio	Represents the utilization ratio of credit card
Credit_History_Age	Represents the age of credit history of the person

Payment_of_Min_Amount	Represents whether only the minimum amount was paid by the person
Total_EMI_per_month	Represents the monthly EMI payments (in USD)
Amount_invested_monthly	Represents the monthly amount invested by the customer (in USD)
Payment_Behaviour	Represents the payment behavior of the customer (in USD)
Monthly_Balance	Represents the monthly balance amount of the customer (in USD)

Insights:

- Age Distribution: Most customers fall within the middle age range (between 30 and 50 years). There are fewer customers in the younger and older age groups.
- Annual Income Distribution: The majority of customers have annual incomes between 25000 and 75000. Fewer customers fall into higher income brackets.
- Occupation Diversity: Customers come from various professions, including writers, programmers, executives, doctors, lawyers, teachers, and more. This diversity suggests a broad occupational background among the customer base.
- Credit Mix: The credit mix is categorized into 'Bad,' 'Standard,' and 'Good.' The 'Standard' category has the highest count, indicating a balanced mix of credit profiles.
- distribution of customers by credit score:
- Good (670-740): Largest group, 44.2%.
- Fair (580-670): Second largest, 25.0%.
- Very Good (740-800): 17.9%.
- Poor (<580): 11.2%.
- Exceptional (>=800): Smallest group, 1.7%.
- Most customers have "Good" scores, while a significant portion falls into "Fair" or "Poor," indicating potential risk.

Positive Relationships:

- Age: Older individuals tend to have higher credit scores.
- Credit History Age: Longer credit history is associated with higher credit scores.

Negative Relationships:

- Num Bank Accounts: Having more bank accounts is associated with lower credit scores.
- Delay_from_due_date: Longer delays from due dates lead to lower credit scores.
- Num_of_Delayed_Payment: More delayed payments are linked to lower credit scores.
- Outstanding_Debt: Higher outstanding debt correlates with lower credit scores.

- Num_Credit_Inquiries: More credit inquiries result in lower credit scores.
- Num_of_Loan: Having more loans is associated with lower credit scores.
- Neutral Relationship: Credit_Utilization_Ratio: No statistically significant relationship with credit scores.