Column-by-Column Breakdown & Inter-Relationships

For the Synthetic Data

1. Customer_id

- **Description:** A unique identifier for each customer (e.g., SCB887716372). This allows the institution to link multiple loan accounts to a single individual.

2. Loan_Account_id

- **Description:** A unique identifier for each specific loan account (e.g., 86893497). A single customer can have multiple Loan_Account_ids for different products.

3. Product_Type

- **Description:** The category of the loan. The types are:
    - Personal loan
    - Auto loan
    - Credit card
    - Education loan
    - Business loan
- **Relationship:** This column is a primary driver for several others:
    - Interest_Rate**:** Each product type has a predefined risk profile and thus a different interest rate (see below).
        - Product Type Interest Rate

          Education Loan 8.0%

          Personal Loan 8.5%

          Auto Loan 9.0%

          Business Loan 9.5%

          Credit Card 10.0%

- Loan_Amount_SGD**:** Each product has different minimum and maximum loan amount limits.
    - Product Type Loan Amount Range (SGD)

Credit Card	5,000 - 50,000
Personal Loan	5,000 - 100,000
Education Loan	5,000 - 100,000
Auto Loan	5,000 - 200,000
Business Loan	5,000 - 500,000

4. Loan_Amount_SGD

- **Description:** The original principal amount of the loan disbursed to the customer, in Singapore Dollars (SGD).
- **Relationship:** It is the key input for calculating the Equated Monthly Installment (Current_EMI_SGD).

5. Outstanding_Balance_SGD

- **Description:** The remaining amount the customer owes on the loan at the time of the data snapshot. This includes the remaining principal plus any accrued interest and penalties.
- **Formula & Relationship:** This is a calculated value, not a simple subtraction. It is derived using an **amortization formula** and is influenced by:
    - Loan_Amount_SGD
    - Interest_Rate
    - Tenure (total number of installments)
    - Number_of_Past_Payments (installments already paid)
    - Day_Past_Due (delinquency can incur penalty interest, increasing the balance)
      The formula ensures the balance decreases over time but can be adjusted upwards for penalties.

6. Day_Past_Due

- **Description:** The number of days the customer's payment is overdue from the Installment_Due_Date. A value of 0 means the account is current.
- **Formula/Logic:** Day_Past_Due = max(0, (Installment_Due_Date - Last_Payment_Date).days - 30). A 30-day grace period is often assumed before an account is formally considered "past due".
- **Relationship:** This is a **critical risk indicator**. It directly influences:
    - Current_EMI_SGD**:** The EMI might be recalculated to include penalties.
    - Outstanding_Balance_SGD**:** Penalty interest accrues on overdue amounts.
    - Credit_Score**:** Higher days past due severely negatively impact the credit score.
    - Repayment_Irregularity_Flags**:** A high value will set this to True.
    - Contact_History_***:** Higher days past due trigger more collection attempts.

7. Tenure

- **Description:** The total duration of the loan, expressed in **months** (e.g., 60 = 5 years).
- In this dataset, the possible values are:
    - 12 months (1 year)
    - 24 months (2 years)
    - 36 months (3 years)
    - 48 months (4 years)
    - 60 months (5 years)

8. Interest_Rate

- **Description:** The annual interest rate charged on the loan (e.g., 8.0%).
- **Relationship:** Determined by the Product_Type. It is a key input for the EMI calculation.

9. Current_EMI_SGD

- **Description:** The Equated Monthly Installment amount the customer is currently required to pay. This amount may change from the original EMI if the account becomes delinquent and penalties are added.
- **Formula:** The standard EMI is calculated using the formula:
  EMI = [P * r * (1+r)^n] / [(1+r)^n - 1]
  - P = Loan_Amount_SGD (Principal)
  - r = Interest_Rate / 12 x 100 (Monthly interest rate)
  - n = Tenure (Number of monthly installments)
- **Relationship:** Can be **adjusted upwards** by a multiplier (e.g., 1.02 for 2% penalty) based on the Day_Past_Due value.

10. Installment_Due_Date

- **Description:** The next upcoming date on which the customer's EMI payment is due.

11. Last_Payment_Date

- **Description:** The date on which the customer last made a payment.
- **Relationship:** Used to calculate Day_Past_Due by comparing it to the previous Installment_Due_Date.

12. Partial_Payment_Indicator

- **Description:** A boolean (True/False) flag indicating if the customer has a history of making partial payments (paying less than the full EMI amount).
- **Relationship:** A True value is a negative factor in the Credit_Score calculation.

13. Number_of_Past_Payments

- **Description:** The count of EMI installments the customer has successfully paid so far.

14. Payment_Frequency

- **Description:** Categorizes the customer's payment habit as Regular or Irregular.
- **Relationship:** Derived from an on_time_percentage metric. If the customer pays on time more than 80% of the time, they are Regular; otherwise, they are Irregular.

15. Amount_Paid_Each_Month_SGD

- **Description:** The typical amount the customer pays each month. It is the actual EMI amount without any Penalities.

16. Bounce_History

- **Description:** A count of how many times the customer's payment (e.g., cheque, direct debit) has bounced due to insufficient funds.
- **Relationship:** A higher count negatively impacts the Credit_Score.

17. Settlement_History

- **Description:** Indicates the status of any settlement negotiations. Values include:
    - Settled: The loan was fully paid off, potentially at a discounted amount.
    - Partial Settlement: A settlement offer was made and partially paid.
    - Not Settled: No settlement activity.
    - Under Negotiation: actively discussing a settlement.
- **Relationship:** A marker of severe distress, often correlated with high Day_Past_Due and a low Credit_Score.

18. Repayment_Irregularity_Flags

- **Description:** A boolean (True/False) flag that is triggered if the customer's payment behavior is deemed irregular (e.g., on_time_percentage < 0.7).

19-22. Contact_History_Call_Attempts**,** SMS**,** WhatsApp**,** EmailLogs

- **Description:** These four columns count the number of outbound contact attempts made by the bank's collections team via each channel.
- **Relationship:** The counts are highest for accounts with high Day_Past_Due. The sum of these columns gives the total outreach effort for an account.

23. Channel_used

- **Description:** The primary channel used for the most recent contact attempt (e.g., IVR, SMS, Call).

24. Response_Outcome

- **Description:** The result of the last contact attempt. Values include Connected, Promised to pay, Ignored, Disconnected, Paid fully, Partial paid.

25. Average_Handling_Time

- **Description:** The average time (in Minutes) spent by agents on calls with customers. Shorter times can indicate simple queries or refusal to engage; longer times can indicate complex negotiations.

26. Name

- **Description:** Customer's name, generated based on Singapore's multi-ethnic demographic distribution (Chinese, Malay, Indian, Western).

27. Age

- **Relationship:** Can influence Occupation, Income_Band_SGD, and Smartphone_Penetration (e.g., older customers may have lower penetration). The Age is ranging from 25-70.

28. Gender

- Male, Female, others

29. Occupation

- **Relationship:** Influences Income_Band_SGD and Employeement_Type.

30. Income_Band_SGD

- **Description:** The customer's annual income range. A key variable for creditworthiness assessment.
- **Relationship:** Higher income bands are generally associated with higher Loan_Amount_SGD and better Credit_Score.

31. Employeement_Type

- **Description:** The nature of the customer's employment (e.g., Full time, Part time, Freelance, Unemployed).
- **Relationship:** Full time is typically seen as more stable and less risky than Freelance or Unemployed.

32. Address

- **Description:** A synthetically generated Singaporean address, including district and postal code.

33. Urban_Rural_Tag

- **Description:** In the context of Singapore, this is always Urban.

34. Language_Preference

- **Description:** The customer's preferred language for communication (English or Regional languages).

35. Mobile_Number_Active_Status

- **Description:** Whether the customer's mobile number is active. Crucial for collections outreach.

36. Email_Activity

- **Description:** Indicates if the customer is active on email.

37. App_Login_Frequency

- **Description:** The average number of times the customer logs into the bank's mobile app per month. A proxy for digital engagement.

38. UPI_Transactions

- **Description:** In a Singapore context, this likely refers to the number of digital/instant payment transactions (e.g., via PayNow) per month. Indicates financial activity.

39. Online_Banking_Activity

- **Description:** The average number of online banking logins or transactions per month.

40. Smartphone_Penetration

- **Description:** Likelihood of the customer using a smartphone (High, Medium, Low).
- **Relationship:** Calculated based on Age and Occupation. Influences the effectiveness of channels like WhatsApp and App notification.

41. Preferred_Channel

- **Description:** The channel the customer is most responsive on (e.g., Email, WhatsApp, Call).

42. Call_SMS_Activity_Patterns

- **Description:** The customer's general responsiveness to calls and SMS (High, Medium, Low).

43. WhatsApp_OTT_usage_Indicator

- **Description:** A boolean indicating if the customer uses Over-The-Top (OTT) messaging apps like WhatsApp.

44. Credit_Score

- **Description:** A numerical representation (typically 300-850) of the customer's creditworthiness.
- **Formula/Relationship:** This is a **composite score** derived from multiple factors:
    - **Base Score:** ~650
    - **+/– Payment History:** Based on on_time_percentage.
    - **– Bounce History:** Penalty for each bounce.
    - **– Delinquency Flag:** Penalty if Delinquency_on_other_Loans is True.
    - **– Partial Payments:** Penalty if Partial_Payment_Indicator is True.
      The score is capped between 300 and 850.

45. Recent_Inquiries

- **Description:** The number of recent credit applications made by the customer at other institutions. Many inquiries in a short time can signal financial stress.

46. Loan_Exposure_Across_Banks

- **Description:** The number of other loans the customer holds across different banks. High exposure increases risk.

47. Delinquency_on_other_Loans

- **Description:** A boolean indicating if the customer has been delinquent on loans with other banks. A major negative factor for Credit_Score.

48. Recent_Score_Change

- **Description:** The change in the customer's Credit_Score over a recent period (e.g., last 6 months). Can be positive or negative.

49-52. Macroeconomic
Indicators: Unemployeement_rate_region**,** Inflation_Rate**,** Interest_Rate_Trend**,** Economic_Stress_Index

- **Description:** These are external, macroeconomic variables applied to all customers, representing the broader Singapore economic environment.
- **Relationship:** They are used for **portfolio-level risk analysis**. For example, a high Unemployeement_rate_region or Inflation_Rate could predict a future increase in defaults across the entire loan book. They are not customer-specific.

53. Do_Not_Call_Registry_Data

- **Description:** A boolean indicating if the customer is registered on Singapore's Do Not Call (DNC) registry. This legally restricts telemarketing calls and SMSes, impacting collections strategy.

54. Regional_Time_Restrictions

- **Description:** The preferred time of day to contact the customer (Morning, Afternoon, Evening, Night).

55. Communication_Complaince_Limits

- **Description:** Any other compliance restrictions on communication, such as not calling on Weekends or Holidays.