**Tier 1: High Potential Impact ( Must-Consider )**

1. **Credit\_Utilization\_Ratio**: Directly affects credit scores, as high utilization negatively impacts scores.
2. **Credit\_History\_Age**: Longer credit history generally improves credit scores.
3. **Payment\_Behaviour**: Timely payments significantly boost credit scores.
4. **Outstanding\_Debt**: High outstanding debt can negatively affect credit scores.
5. **Credit\_Score** (Target Variable): Ensure this is your outcome variable.

**Tier 2: Medium to High Potential Impact ( Strongly Consider )**

1. **Num\_of\_Delayed\_Payment**: Direct indicator of payment reliability.
2. **Delay\_from\_due\_date**: Similar to above, indicates payment discipline.
3. **Monthly\_Inhand\_Salary** & **Annual\_Income**: Income level can influence creditworthiness.
4. **Type\_of\_Loan** & **Num\_of\_Loan**: Diversified loan types and numbers can impact credit scores.
5. **Credit\_Mix**: A diverse credit mix is often viewed positively.

**Tier 3: Potential Impact ( Consider with Caution/Transformation )**

1. **Interest\_Rate**: Might influence credit score indirectly through debt burden.
2. **Changed\_Credit\_Limit**: Could indicate credit health if limits are increased responsibly.
3. **Num\_Credit\_Inquiries**: Too many inquiries can slightly lower credit scores.
4. **Total\_EMI\_per\_month**: High EMI could strain finances, potentially affecting scores.
5. **Amount\_invested\_monthly**: Positive savings/investment habits might support creditworthiness.

**Tier 4: Low Potential Direct Impact ( Possibly for Additional Context or Feature Engineering )**

1. **ID**, **Customer\_ID**, **Month**, **Name**, **SSN**: Primarily identifiers or irrelevant to credit scoring.
2. **Age**: While age can influence financial stability, its direct impact on credit scores is less clear.
3. **Occupation**: Could potentially be used in feature engineering (e.g., job stability), but direct impact is low.
4. **Num\_Bank\_Accounts** & **Num\_Credit\_Card**: While indicative of financial activity, the direct impact on credit scores is less significant without context (e.g., usage patterns).
5. **Monthly\_Balance**: Less directly influential than credit utilization ratio but could support feature engineering.
6. **Payment\_of\_Min\_Amount**: Potentially useful in assessing payment behavior but might be redundant with Payment\_Behaviour.

**Feature Selection Strategy:**

* **Initial Model:** Start with all **Tier 1** and **Tier 2** features.
* **Iterative Refinement:**
  + Add **Tier 3** features one by one, assessing model performance improvement.
  + Consider **Tier 4** features for engineering new, more informative features (e.g., combining Age and Credit\_History\_Age for a "Credit Maturity" score).

**df\_new['Age'] = pd.to\_numeric(df\_new['Age'], errors='coerce')**

**What it does:**

1. **pd.to\_numeric()**: This is a pandas function used to convert a scalar, array, or Series (a one-dimensional labeled array of values) to a numeric data type (either int64 or float64, depending on the presence of decimal points).
2. **df\_new['Age']**: This specifies the column ('Age') within the DataFrame (df\_new) that you want to convert to a numeric type.
3. **errors='coerce'**: This parameter dictates how to handle values that cannot be converted to a numeric type. With 'coerce', any value that cannot be converted will be replaced with NaN (Not a Number), which is pandas' way of representing missing or null values in numeric columns.