**Project Name: Credit Risk Analysis**

**Project Objective:**

The goal of this project was to identify patterns that indicate whether a client might face difficulty in paying back their installments. These patterns can help guide actions like denying loans, reducing loan amounts, or charging higher interest rates for high-risk applicants.

**Approach/ Steps Followed:**

1. **Data Upload**
2. **Remove Columns with Excessive Missing Values**: Columns with more than 50% missing data were removed—41 columns from current applications and 4 from previous applications.
3. **Data Understanding**: Studied the column descriptions to develop a thorough understanding of the data.
4. **Individual Column Analysis and Cleaning**:
    - **Parameters Checked**: Null values, unexpected entries, potential inferences, and actions to take for each column.
    - **Outcome**:
        - Null values for numerical columns were replaced with 0 or the mean, depending on the context.
        - Outliers in numerical columns were adjusted using the IQR method.
5. **Merging Dataframes**: Cleaned dataframes were merged after initial cleaning.
6. **Identifying Data Types**: Differentiated between numeric and categorical data.
7. **Cleaning Categorical Columns**: Focused on columns with 10 or fewer unique values. Replaced nulls with the mode or dropped columns with high null values (~49%).
8. **Segmenting Data Based on Target Group**: Created two separate dataframes (Target = 0 or 1) for in-depth analysis.
9. **Running Univariate, Bivariate, and Multivariate Analysis**:
    - **Tools Used**: Pie charts, heatmaps, scatter plots, and bar plots to explore correlations.
10. **Conclusion**: Summarized the findings and identified significant correlations.

**Key Findings from the Analysis**

**Pie Chart Observations:**

1. **Loan Type**: Revolving loans have fewer repayment issues compared to cash loans.
2. **Gender**: Females are generally better at loan repayments than males.
3. **Car Ownership**: Car owners have a slightly better repayment rate.
4. **Family Accompaniment**: Applicants accompanied by family during loan application have better repayment records.

5. **Income Type**: State servants have the best repayment history, followed by working-class individuals.
6. **Education**: Applicants with higher education levels are the most reliable in repaying loans.
7. **Family Status**: Married individuals are more reliable in repayments compared to singles.

**Heatmap Observations:**

Identified strong correlations between:

- **AMT_GOODS_PRICE**, **AMT_CREDIT**, and **AMT_ANNUITY**
- **CNT_CHILDREN** and **CNT_FAM_MEMBERS**
- **DAYS_LAST_DUE** and **DAYS_TERMINATION**

**Scatter and Count Plot Observations:**

- No specific combinations showed distinct behavior differences between those who struggle with payments and those who do not.
- Some categories like **Cash Loans** and **House/Apartment owners** displayed unique patterns.

**Conclusion:**

Based on the analysis, the following parameters have the most significant impact on loan repayment:

- **Gender**
- **Education Level**
- **Family Status**

These factors are critical for assessing a client's risk profile and guiding loan-related decisions.

**Project Outcomes:**

- For the use case, the project provided crucial insights to aid decision-making in the loan approval process.
- Personally, this project has enhanced my ability to manage large datasets and handle a wide range of parameters, even without direct guidance from clients.