Project Report: Financial Loan Portfolio Analysis

1. Executive Summary

This report details the end-to-end analysis of a comprehensive financial loan dataset. The primary objective of this project was to assess the health of the loan portfolio, identify key business trends, and understand the characteristics of borrowers. By leveraging Python and its data science libraries, we transformed raw data into actionable insights. The analysis focused on calculating critical Key Performance Indicators (KPIs), segmenting the portfolio into 'Good' and 'Bad' loans, and creating a suite of visualizations to uncover underlying patterns in lending activity. The findings from this analysis provide a clear view of the portfolio's performance and highlight strategic areas for business growth and risk management.

2. Data Preparation and Cleaning

The initial phase of the project involved robust data preparation to ensure the accuracy and reliability of the analysis. The raw dataset contained missing values and inconsistent data types that needed to be addressed.

- **Data Loading:** The dataset was loaded into a Pandas DataFrame.
- **Handling Missing Values:** A simple imputation strategy was employed. Missing numerical values (e.g., dti, int_rate) were filled with the median of their respective columns, while missing categorical values were filled with the mode. This ensured that no data was lost while minimizing the impact of missing entries.
- **Data Type Conversion:** Columns containing dates (issue_date, last_payment_date, etc.) were converted from text format to a proper datetime format. This was crucial for performing time-series analysis.
- **Feature Engineering:** To facilitate a more meaningful analysis, a new categorical feature, LOAN_CONDITION, was created. Loans with statuses of 'Fully Paid' or 'Current' were classified as 'Good Loans,' while all others (e.g., 'Charged Off') were classified as 'Bad Loans.' This new feature became the cornerstone of the portfolio health assessment.

3. Key Performance Indicator (KPI) Analysis

A set of core KPIs was calculated to provide a high-level snapshot of the portfolio's performance.

- **Total Loan Applications:** The portfolio consists of **38,577** loan

applications.

- **Total Funded Amount:** A total of **$425,153,350** has been disbursed to borrowers.
- **Total Amount Received:** The total amount paid back by borrowers to date is **$458,527,333.**
- **Average Interest Rate:** The average interest rate across all loans is **12.05%.**
- **Average Debt-to-Income (DTI) Ratio:** The average borrower has a DTI of **13.33%,** providing insight into the overall financial health of the customer base.

Month-To-Date (MTD) metrics were also calculated for the most recent month in the dataset to monitor current performance against historical averages.

4. Portfolio Health: Good vs. Bad Loan Segmentation

The analysis of the LOAN_CONDITION segment revealed critical insights into the portfolio's risk profile.

- **Good Loans:**
  - Constitute **85.95%** of all applications.
  - Account for **$367.7 million** in funded amounts.
  - Have generated **$430.7 million** in received payments, indicating a healthy return.
- **Bad Loans:**
  - Make up **14.05%** of all applications.
  - Account for **$57.4 million** in funded amounts.
  - Have only returned **$27.8 million** in payments, representing a significant loss.

This segmentation clearly quantifies the risk and performance within the portfolio and serves as a vital metric for credit risk assessment.

5. Visual Analysis and Key Insights

A series of visualizations were created to explore the data from multiple dimensions:

- **Monthly Lending Trends:** A line chart of loan applications over time revealed clear seasonality and growth trends in lending activity, allowing for better financial forecasting.
- **Regional Analysis:** A bar chart of loans by state identified California, New York, and Florida as the top three states with the highest volume of loan applications. This insight is crucial

for tailoring regional marketing strategies.

- **Loan Purpose Breakdown:** The majority of loans were taken for **debt consolidation**, followed by **credit card refinancing**. This highlights the primary financial needs of the customer base.
- **Loan Term Distribution:** The analysis showed a significant preference for **36-month loan terms** over 60-month terms, indicating that most borrowers prefer shorter repayment periods.
- **Borrower Employment Length:** A bar chart showed that individuals with **10+ years of employment** represent the largest group of borrowers, suggesting that employment stability is a strong factor in loan applications.
- **Home Ownership Status:** An interactive treemap demonstrated that borrowers who **rent** or have a **mortgage** make up the vast majority of the loan portfolio, far surpassing those who own their homes outright.

6. Conclusion and Recommendations

This data analysis project provided a multi-faceted view of the financial loan portfolio. The key takeaway is that while the majority of the portfolio is healthy, the 14% of 'Bad Loans' represent a substantial financial loss that requires strategic intervention.

Based on the insights gathered, the following recommendations are proposed:

- **Refine Risk Models:** Use insights from the 'Bad Loan' segment (e.g., purpose, grade, home ownership) to strengthen credit risk assessment models and potentially adjust lending criteria.
- **Targeted Marketing:** Focus marketing and sales efforts on high-performing regions and borrower profiles (e.g., individuals with stable employment seeking debt consolidation).
- **Product Optimization:** Given the popularity of 36-month terms, consider developing more products or offering more competitive rates around this term length.

7. Tools and Technologies Used

- **Language:** Python
- **Libraries:**
  - **Pandas:** For data manipulation and cleaning.
  - **NumPy:** For numerical operations.
  - **Matplotlib & Seaborn:** For creating static charts and graphs.
  - **Plotly:** For creating interactive visualizations.