**Dynamic Risk Profiling and Performance Analysis in Consumer Lending - Excel Case Study**

### **Background:**

The consumer lending landscape has undergone a significant transformation in recent years, driven by evolving economic conditions, technological advancements, and changing consumer behaviors. In this dynamic environment, lenders face the challenge of managing a diverse loan portfolio while balancing profitability with risk. The ability to effectively analyze and interpret loan data is crucial for making informed decisions and maintaining a competitive edge.

The loan dataset in this case study offers a rich source of information, encompassing various aspects of the lending process, including borrower demographics, loan characteristics, financial metrics, and repayment histories. It provides an opportunity to delve deep into the patterns and trends that govern consumer lending, offering insights into risk management, borrower behavior, and financial performance.

### **Objective:**

The objective of this case study is to apply advanced data analysis techniques using Excel to deeply understand and effectively manage a loan portfolio. This involves assessing loan risk profiles, evaluating financial performance metrics, gaining insights into borrower behaviors, and ultimately, developing a comprehensive, interactive Excel dashboard. These efforts aim to provide strategic recommendations for optimizing loan portfolio management, mitigating risks, and enhancing decision-making processes for financial institutions and stakeholders in the consumer lending sector.

Data Source: ***loan\_final313\_.csv***

The loan dataset provides a detailed snapshot of consumer lending, encompassing a wide array of information about individual loans.

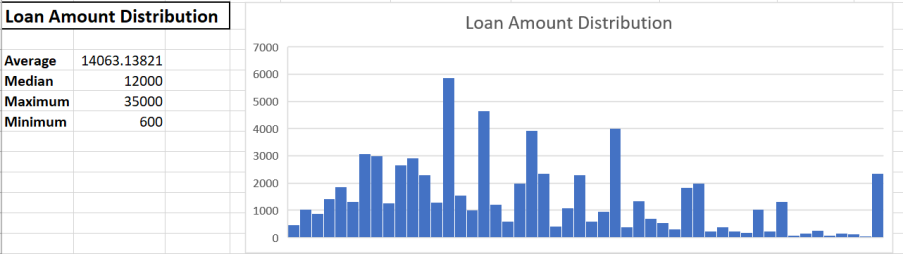
1. **id**: A unique identifier for each loan.
2. **year**: The year when the loan was issued.
3. **issue\_d**: The specific date when the loan was issued.
4. **final\_d**: A date-related field, possibly indicating the final date of the loan term or the date of final payment.
5. **emp\_length\_int**: Employment length of the borrower, in years, as an integer.
6. **home\_ownership**: The housing status of the borrower, indicating whether they own a home, rent, or have another arrangement.
7. **home\_ownership\_cat**: Categorical encoding of the home ownership status.
8. **income\_category**: A categorical representation of the borrower's income level.
9. **annual\_inc**: The annual income of the borrower.
10. **income\_cat**: A categorical encoding of the borrower's income.
11. **loan\_amount**: The amount of money borrowed.
12. **term**: The term of the loan, typically in months or years.
13. **term\_cat**: Categorical encoding of the loan term.
14. **application\_type**: Indicates whether the loan application was individual or joint.
15. **application\_type\_cat**: Categorical encoding of the application type.
16. **purpose**: The purpose of the loan (e.g., debt consolidation, home improvement).
17. **purpose\_cat**: Categorical encoding of the loan purpose.
18. **interest\_payments**: A descriptor of the interest payments (e.g., high, low).
19. **interest\_payment\_cat**: Categorical encoding of the interest payments.
20. **loan\_condition**: Status of the loan (Good Loan and Bad Loan).
21. **loan\_condition\_cat**: Categorical encoding of the loan condition.
22. **interest\_rate**: The interest rate on the loan.
23. **grade**: The loan grade, which is typically an assessment of the loan's risk.
24. **grade\_cat**: Categorical encoding of the loan grade.
25. **dti**: The debt-to-income ratio of the borrower.
26. **total\_pymnt**: The total payments made on the loan to date.
27. **total\_rec\_prncp**: The total principal received to date.
28. **recoveries**: The amount recovered on the loan after default.
29. **installment**: The monthly payment owed by the borrower.
30. **region**: The geographical region of the borrower.

## **Part 1: Excel Data Analysis: Manipulation, Formulas, and Functions**

**1. Missing Data Analysis: Identify any missing data in the loan dataset. What patterns, if any, can be observed in the missing data?**

There are no missing values present in the data.

**2. Loan Amount Distribution: Analyze the distribution of loan amounts. What is the average, median, and range of loan amounts?**

****

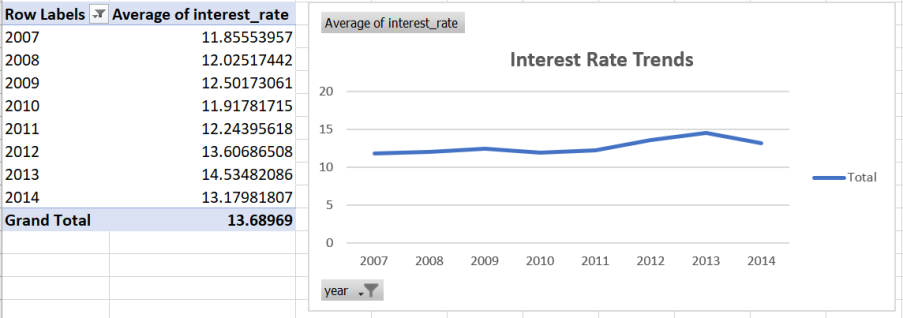
**Interpretation:**

· The loan portfolio is mostly composed of small to mid-size loans.

· The wide range (₹600 to ₹35,000) suggests that the lender serves a diverse borrower base.

· The histogram helps identify popular loan sizes and spot any outliers or gaps.

**3. Interest Rate Trends: Examine how interest rates have changed over the years. Create a line chart to visualize this trend.**

****

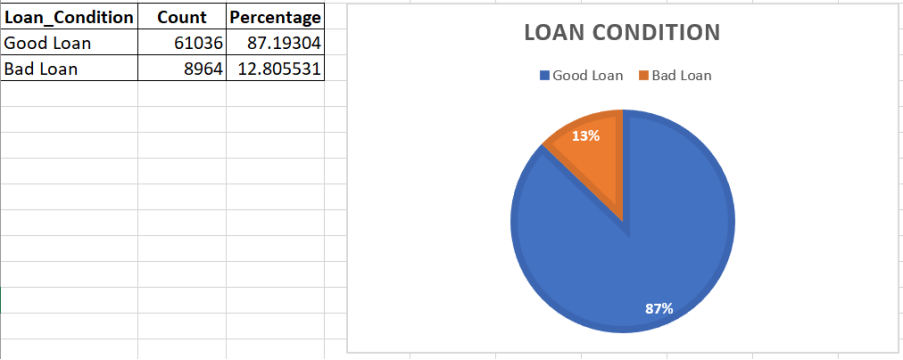
**Interpretations :**

- There is a gradual increase in average interest rates from 2010 to 2013.

- The spike in 2013 might indicate market tightening or increased lending risk.

- The trend can be useful for analyzing how market conditions or policy changes influenced lending rates over time.

**4. Loan Status Analysis: Categorize loans based on their condition (e.g., Good Loan, Bad Loan). What percentage of loans falls into each category?**

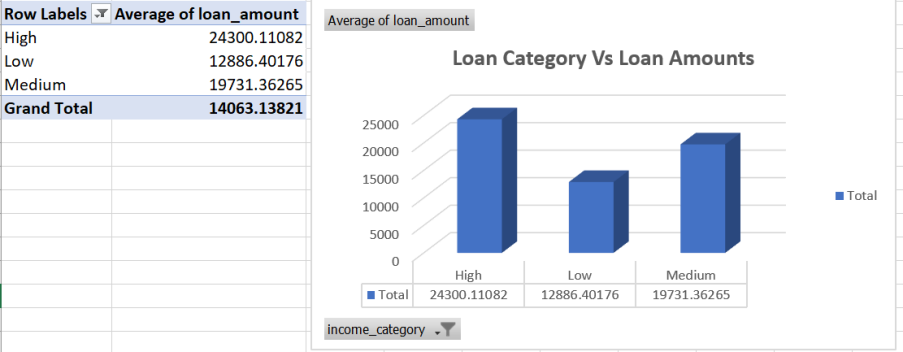
****

**Interpretation:**

· The lender has a high-quality loan portfolio, with the vast majority (87%) of loans performing well.

· Only 13% are non-performing or “bad” loans, which could mean defaults, late payments, or collections.

**5. Income Category and Loan Amount Correlation: Investigate if there is a correlation between income categories and loan amounts.**

****

**Interpretations :**

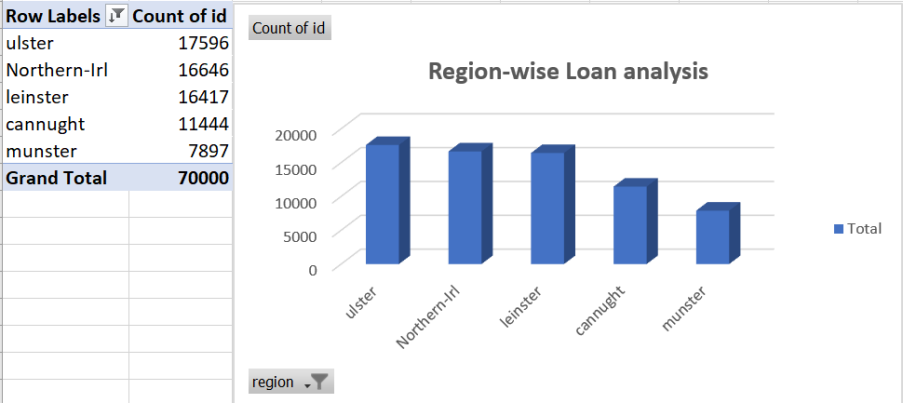
· There’s a direct correlation between income level and average loan amount.

· Lenders tend to approve larger loans for high-income borrowers—likely due to their higher repayment capacity and lower risk.

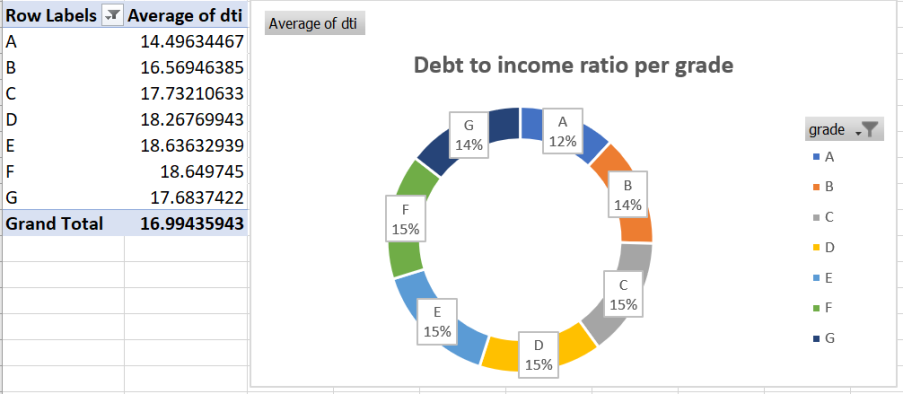
· Low-income borrowers may be more credit-constrained, or lenders may limit loan size due to higher perceived risk.

**6. Region-wise Loan Analysis: Which region has the highest number of loans? Visualize the distribution of loans across different regions.**

Ulster (17,596) is the top-performing region in terms of loan volume.

****

**7. Debt-to-Income Ratio Insights: Calculate and analyze the average debt-to-income ratio for each loan grade.**

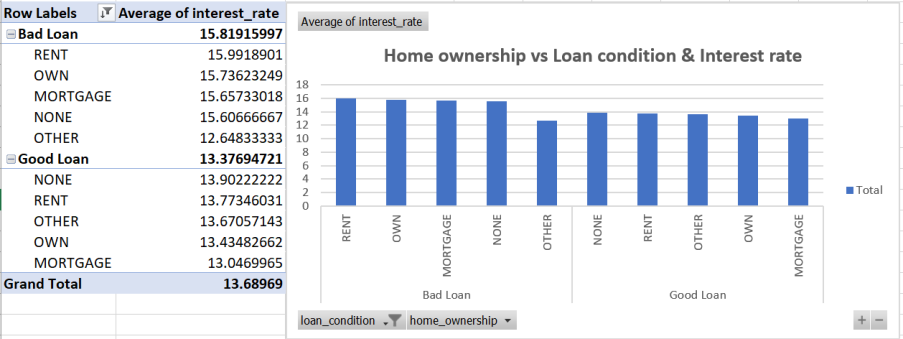
****

**Interpretation:**

· Higher grades (A, B) have lower DTI, indicating less financial burden and lower credit risk.

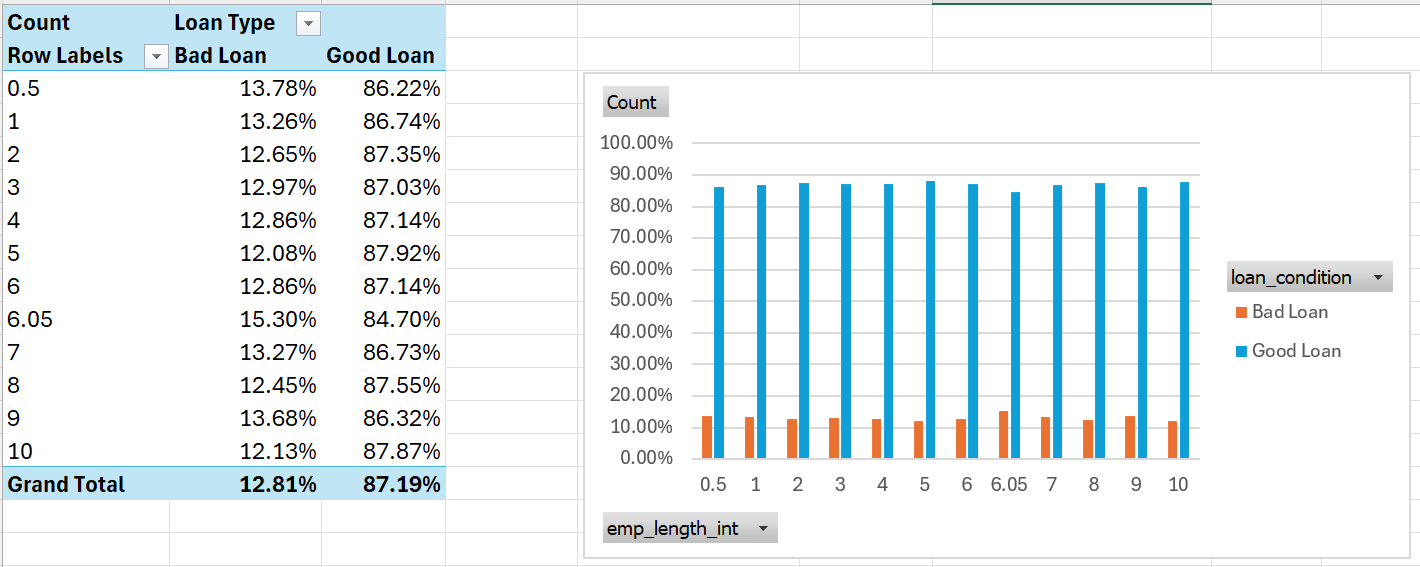
· Lower grades (D to G) have higher DTI, suggesting borrowers in these categories are more leveraged and riskier.

**8. Home Ownership Impact: Examine how home ownership status (rent, own, mortgage) affects loan conditions and interest rates.**

****

Mortgage and Own statuses are typically linked to lower risk and receive better rates.

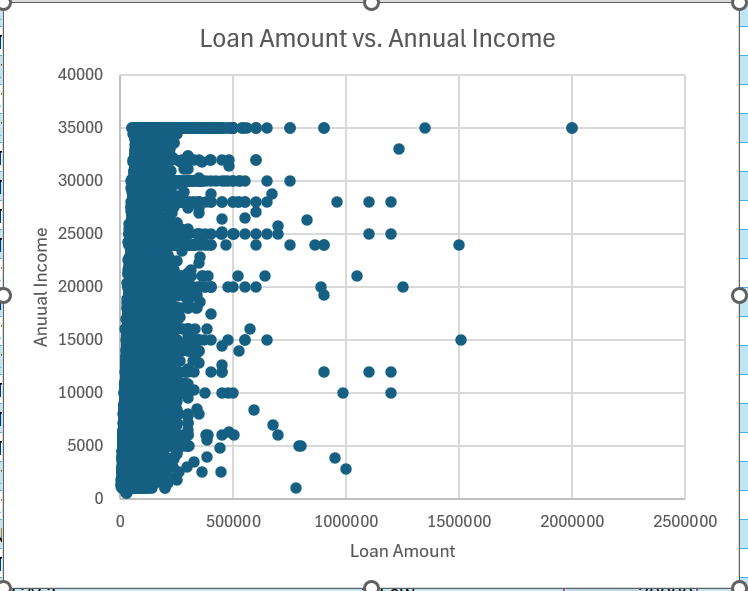
**9. Employment Length and Loan Condition: Is there a relationship between the length of employment and the condition of loans?**

****

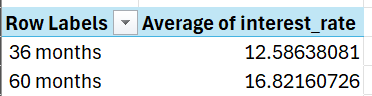
The data reveals that default rates are higher among borrowers with shorter employment history (e.g., 0–2 years). As employment length increases, default rates generally decline, suggesting greater financial stability and repayment ability among long-term employees.

**10. Loan Amount vs. Annual Income: Create a scatter plot to analyze the relationship between loan amounts and borrower's annual income.**

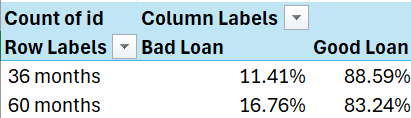
The scatter plot shows a weak to moderate positive correlation between annual income and loan amount. While higher-income borrowers do tend to take larger loans, there's a significant spread indicating that loan size isn't strictly tied to income. This could imply that creditworthiness, purpose, and other factors influence loan amounts.



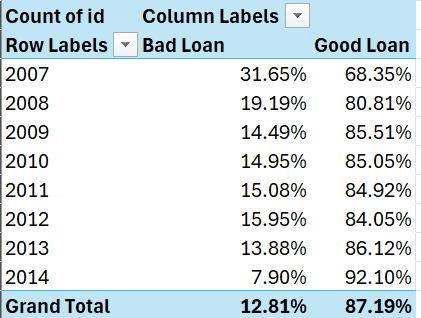
**11.Loan Term Analysis: Compare the terms of loans (e.g., 36 months vs. 60 months). What are the differences in interest rates and default rates?(*Steps: 1. Identify Loan Term Lengths, 2. Calculate Average Interest Rates for Each Term, 3. Categorize Loans by Term and Condition, 4. Calculate Default Rates for Each Term)***

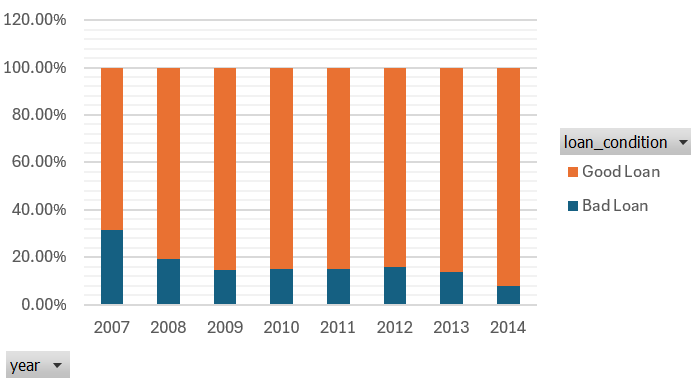
**Average Interest rate:**

**Default rates :**



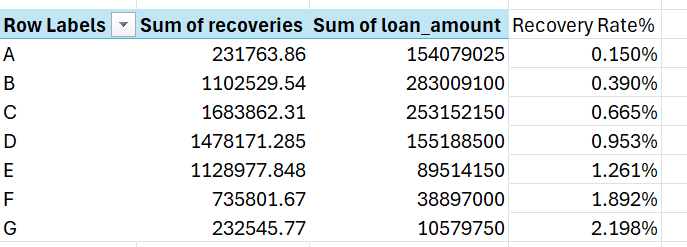
* Loans with longer terms (60 months) have a higher average interest rate and higher default rate compared to shorter-term loans (36 months).
* This suggests that longer-term loans are riskier and possibly issued to higher-risk borrowers, hence the higher interest rates.

**12.Default Rate Calculation: Calculate the default rate (percentage of loans that are in default) for each year. (*Default Rate = Divide the number of loans categorized as "Bad Loan" by the total number of loans for that term.*)**



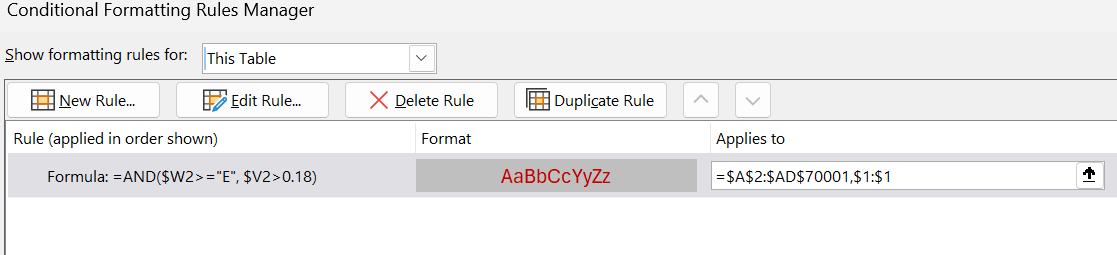
* The default rate has significantly decreased over time, from a high of 31.65% in 2007 to just 7.90% in 2014.
* This trend indicates improved loan underwriting practices, economic conditions, or borrower quality over the years.
* The most notable drop occurred between 2007 and 2009, aligning with the post-crisis tightening of lending standards.

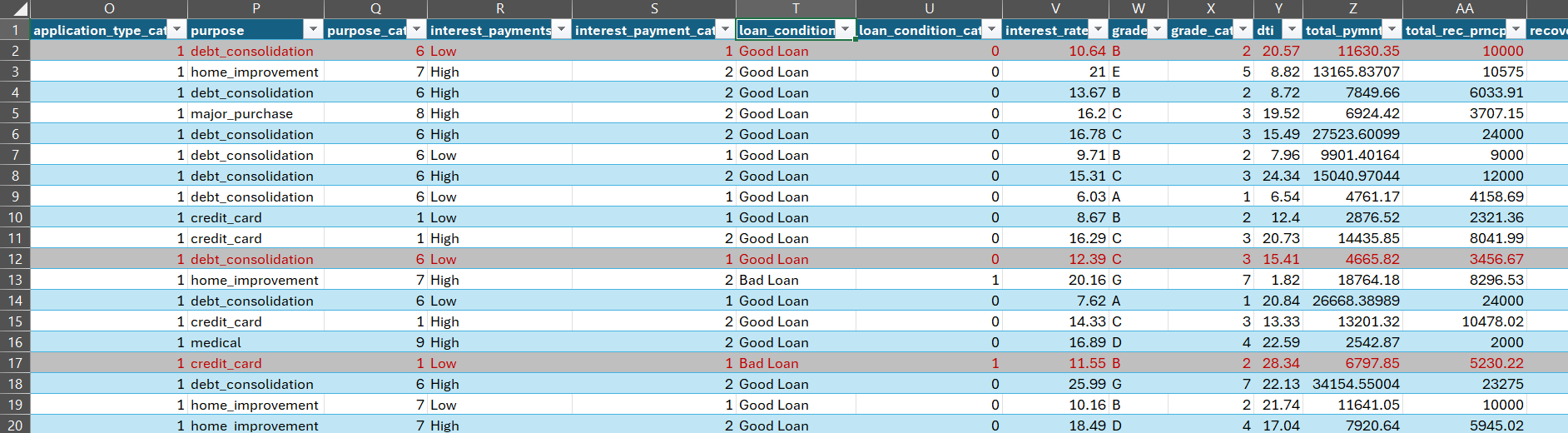
**13. Recovery Rate Analysis: Analyze the recovery rate (percentage of defaulted amount recovered) across different loan grades.(Can be done with recoveries and loan\_condition (for defaulted loans).)**



1. Recovery rate increases with riskier loan grades:
   * Grade A has the lowest recovery rate (0.150%).
   * Grade G has the highest recovery rate (2.198%).
   * This might indicate that lenders recover more from riskier loans — likely due to higher interest rates or better recovery strategies being applied.
2. Lower-grade loans (F & G) show higher returns on recovery:  
   * Although they are riskier and more prone to default, loan amounts are smaller, and recoveries tend to be more efficient.
3. Middle grades (C, D, E) show steady recovery increase:  
   * These represent a moderate risk-return tradeoff and could be the sweet spot for balancing risk and recovery.

**14.Conditional Formatting for High-Risk Loans: Use conditional formatting to highlight loans with high interest rates and low grades**

****

****

To visually identify and monitor potentially high-risk loans, conditional formatting was applied to the dataset using two key criteria:

* Loan Grade: Grades E, F, and G were considered low-quality and high-risk.
* Interest Rate: Interest rates greater than 18% were flagged as risk indicators.

here:

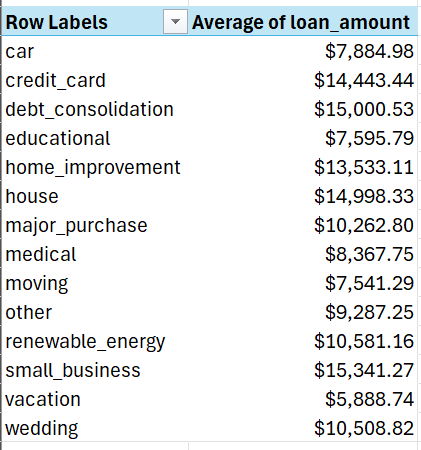
* $W2 refers to Loan Grade.
* $V2 refers to Interest Rate.

This rule highlights rows meeting both conditions with a Grey fill color with red letters, making high-risk loans instantly noticeable.

**15.Loan Purpose Analysis: Categorize loans by their stated purpose. Which purpose has the highest average loan amount?**

The top 3 loan purposes with the highest average loan amount are:

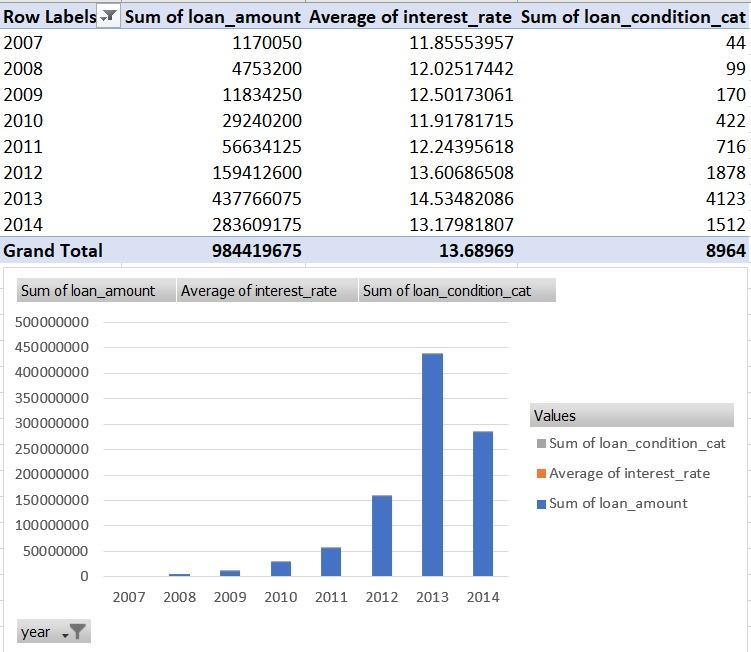
1. Small Business – $15,341.27
2. Debt Consolidation – $15,000.53
3. House – $14,998.33

The lowest average loan amounts were for

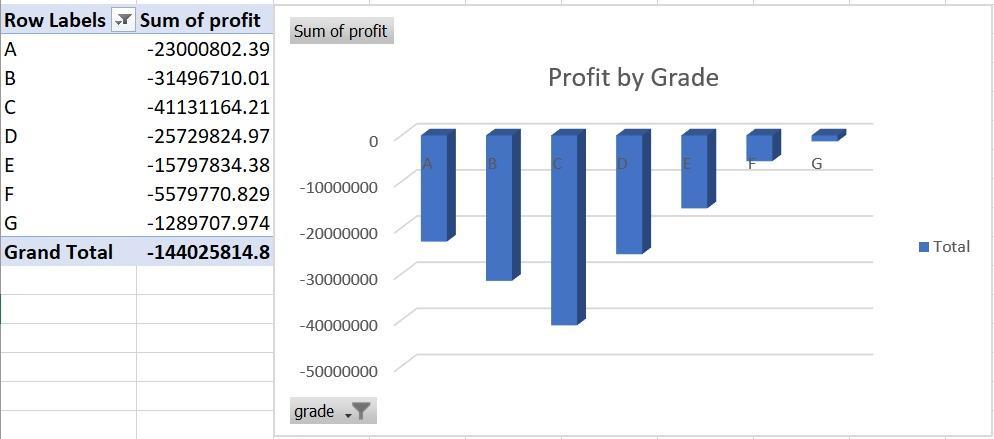
* Vacation – $5,888.74
* Moving – $7,541.29
* Educational – $7,595.79

**16 . Aggregate Loan Analysis by Year: Summarize total loan amounts, average interest rates, and default rates for each year.( *Steps: 1. Summarize Total Loan Amounts by Year, 2. Calculate Average Interest Rates by Year, 3. Determine Default Rates by Year*)**

* **Loan Volume Growth:**  
   The total loan amount issued grew significantly between 2010 and 2013, peaking in 2013 at over ₹437 million. This indicates a rapid expansion in lending activity during this period.
* **Interest Rate Trends:**  
   Average interest rates gradually increased over the years, peaking in 2013 at 14.53%, and slightly dropped to 13.18% in 2014.

****

**17.Grade-wise Profitability Analysis: Calculate the profitability of loans (total payments received minus the loan amount) for each grade.**

****

The analysis shows that the business is experiencing significant financial losses. Grade F is the largest contributor to these losses, followed by Grade C. Grade G, while still negative, is performing relatively better than the others.

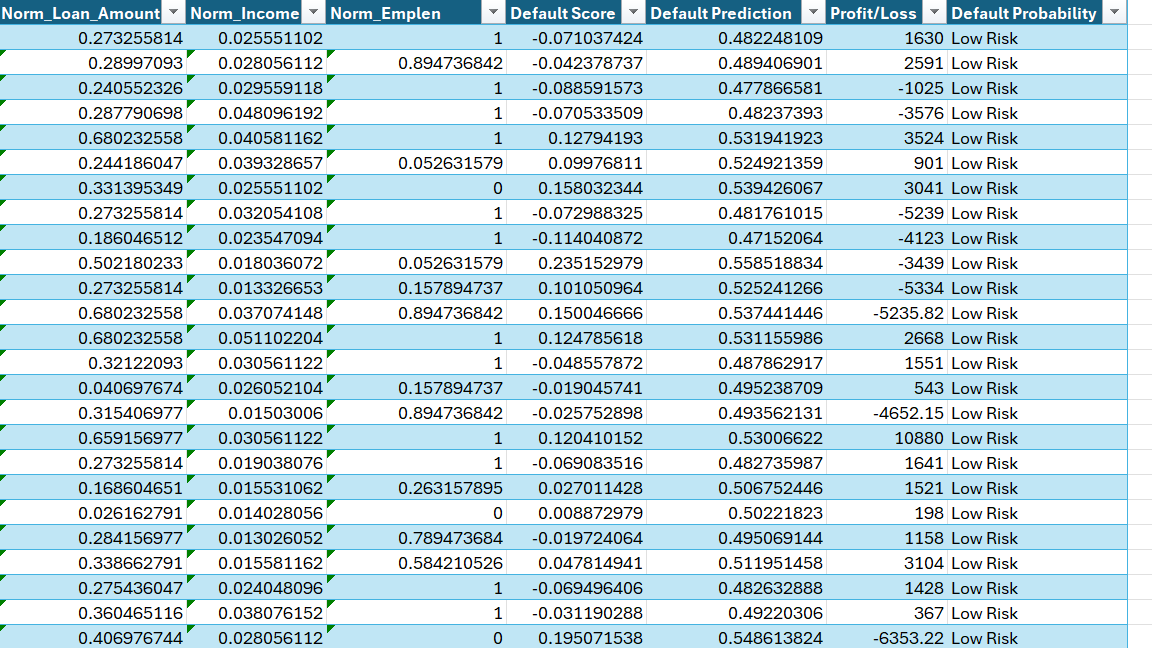
**18. Predictive Analysis for Loan Defaults: Utilize Excel's advanced functions to predict the likelihood of default based on factors like loan amount, income, and employment length.**

To estimate the likelihood of loan default, we built a predictive model in Excel using normalized values of loan amount, annual income, and employment length. A custom weighted formula was applied to generate a Default Score, which was then converted to a probability using the sigmoid function. Based on a threshold (e.g., > 0.56), loans were classified as either High Risk or Low Risk.

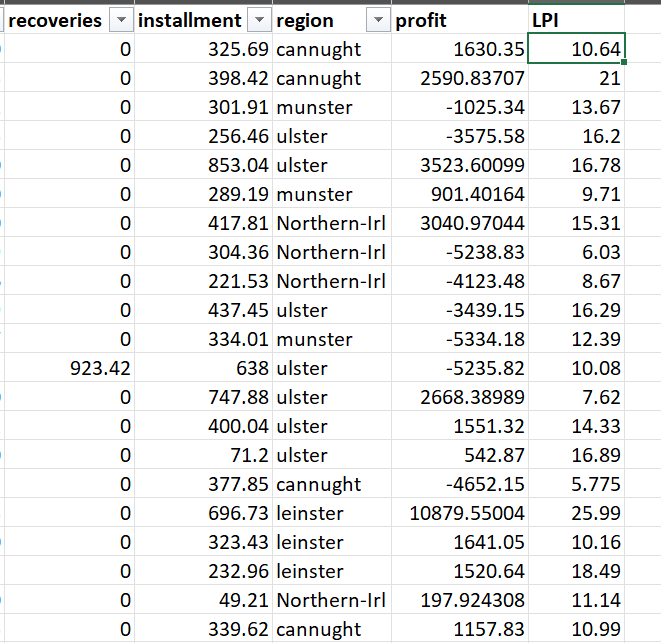
|  |
| --- |

Findings:

* Majority of loans were classified as Low Risk.
* High Risk loans typically had:
  + Higher loan amounts
  + Lower income
  + Shorter employment history

****

**19. Loan Profitability Index Calculation: Develop a "Loan Profitability Index" (LPI) for each loan, a composite metric that factors in the interest rate, loan amount, loan term, and the loan condition (whether it's in good standing, late, or defaulted). (This can be calculated with available data like interest\_rate, loan\_amount, loan\_term, and loan\_condition.)**

****

LPI Range Observed:  
 Values range from as low as 5.78 to as high as 26.06.

Top Performing Loans (High LPI):

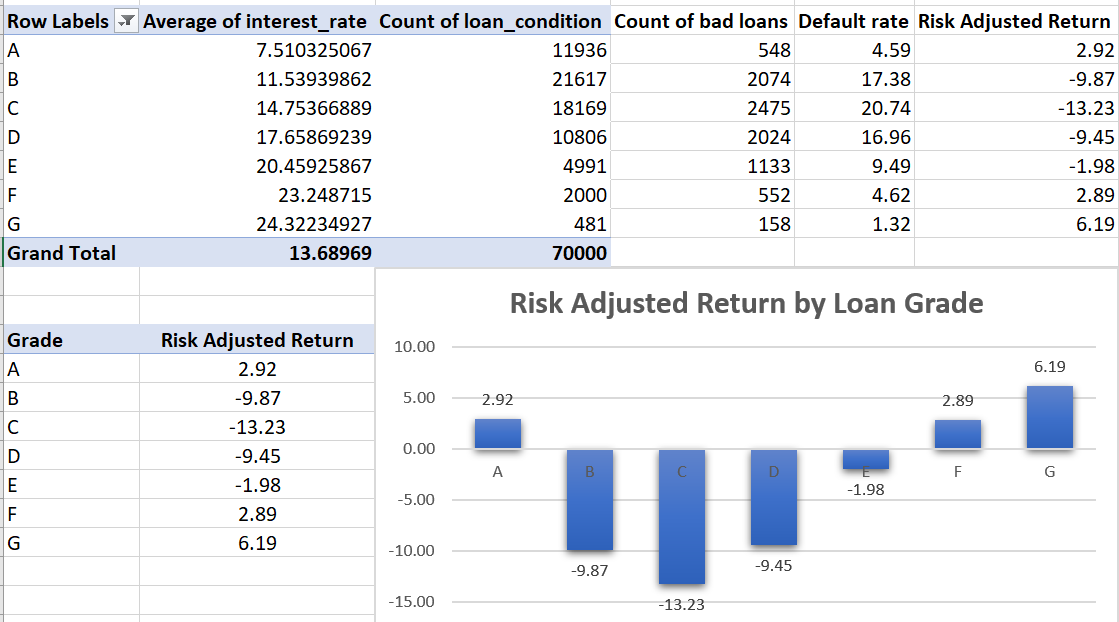
LPI > 20 indicates very high profitability.

* Example: Leinster region has an LPI of 25.99 (indicating strong return vs. risk).

Low Performing Loans (Low LPI):

* LPI < 10 shows reduced profitability, often due to defaults or late payments.
* Example: Northern-Irl loan with LPI = 5.78, likely due to high loss or default

**20. Risk-Adjusted Return on Loans: Calculate the risk-adjusted return on loans for each grade category. This metric should reflect not only the return (in terms of interest received) but also the risk (chance of default).**

****

The chart shows that **grades G, A, and F** offer the highest risk-adjusted returns, making them the **most profitable.** In contrast, **grades C, B, and D** yield **negative returns, indicating poor lending options.**

**21. Calculating Weighted Average Interest Rate for Each Income Category: Calculate the weighted average interest rate for each income category, taking into account the loan amount as the weight. (Method: Use a combination of SUMPRODUCT and SUMIF**

**functions.)**



*SUMPRODUCT((loan\_final313\_[income\_category]=A2)\*(loan\_final313\_[interest\_rate])\*(loan\_final313\_[loan\_amount])) /*

*SUMIF(loan\_final313\_[income\_category], A2, loan\_final313\_[loan\_amount]*

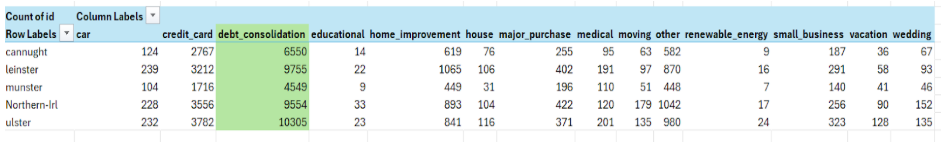
Higher Interest Rates for Lower Incomes:

* Borrowers in the Low income category are charged the highest weighted average interest rate.
* Indicates greater perceived risk or lower creditworthiness.

Decreasing Trend:

* As income increases, the average interest rate decreases, showing a clear inverse relationship between income and interest charged.

**22. Determining the Most Common Loan Purpose for Defaulted Loans in Each Region: Find out the most frequent loan purpose for loans that have defaulted in each region. (Method: Utilize a combination of IF, MODE, and MATCH functions.) (With purpose, loan\_condition, and region.)**

****

* Across all regions, “debt\_consolidation” is the most frequent loan purpose among defaulted loans.
* This strongly suggests that borrowers taking loans for debt consolidation are more likely to default, indicating a higher credit risk profile.

### **Part 2: Building an Excel Dashboard**

**Create a comprehensive dashboard in Excel that provides real-time insights into the performance and risk profile of the loan portfolio. The dashboard should enable stakeholders to quickly assess key metrics, identify trends, and make informed decisions regarding loan management strategies.**

**Excel Pivot Tables and Dashboard Link:** [**loan\_final313\_.xlsx**](https://docs.google.com/spreadsheets/d/15sDan-LI8jnZ5LC6xYzN82_5CeZGYvu9/edit?usp=sharing&ouid=112633500626026739476&rtpof=true&sd=true)