Financial Risk Assessment Data Analysis

Welcome to my exploration of financial risk assessment data, where I experiment with real-world datasets to develop my analytical skills for my portfolio. In this project, I focus on data cleaning, preparation, and visualisation to uncover insights into loan risks and borrower profiles. By analysing key metrics such as credit scores, loan-to-value ratios, and debt-to-income ratios, I aim to identify trends and patterns that influence risk ratings and loan decisions. Through this process, I refine my ability to interpret financial data and present meaningful insights using data-driven storytelling.

Data Cleaning and Preparation

1 Excel Data Cleaning

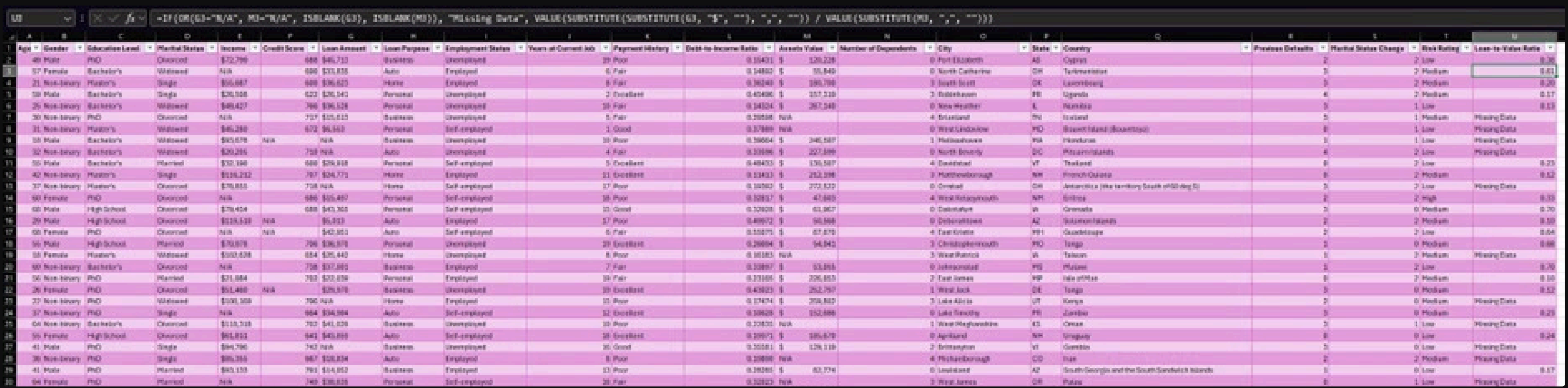
Imported CSV data into Excel for initial cleaning. Fixed data types, filled null values as "N/A", checked for spelling errors and duplicate columns. Added a new column for loan-to-value ratio to enhance analysis.

2 Loan-to-Value Ratio Calculation

Implemented a formula to calculate LTV ratio, handling blank cells and currency formatting. LTV ratio provides a clear measure of risk by showing how much of the asset's value is financed by the loan.

3 Power BI Data Loading

I then loaded the data into Power Query in Power Bi and addressed issues during Power BI import, including removing rows with missing names and adjusting data types. Initially replaced N/A values with 0, but later filtered them out completely to prevent data skewing.



Understanding Loan-to-Value Ratio

Definition of LTV Ratio

Loan-to-Value Ratio (LTV) is a measure of risk in the context of a specific loan, typically in mortgage or asset-backed lending situations. It is calculated based on the loan amount relative to the value of a single asset (e.g., a house, car, or property).

Importance in Risk Assessment

LTV ratio provides a clear measure of risk by showing how much of the asset's value is financed by the loan. It's a key metric for visualising risk levels and can be effectively plotted using scatter plots or histograms.

U2

```
=IF(OR(G2="N/A", M2="N/A", ISBLANK(G2), ISBLANK(M2)), "Missing Data", VALUE(SUBSTITUTE(SUBSTITUTE(G2, "$", ""), ",", "")) / VALUE(SUBSTITUTE(M2, ",", "")))
```

rital Status	Income	Credit Score	Loan Amount	Loan Purpose	Employment Status	Years at Current Job	Payment History	Debt-to-Income Ratio	Assets Value	Number of Dependents	City	St
orced	$72,799	688	$45,713	Business	Unemployed	19	Poor	0.15431	$ 120,228	0	Port Elizabeth	AS
lowed	N/A	690	$33,835	Auto	Employed	6	Fair	0.14892	$ 55,849	0	North Catherine	O
gle	$55,687	600	$36,623	Home	Employed	8	Fair	0.36240	$ 180,700	3	South Scott	O
gle	$26,508	622	$26,541	Personal	Unemployed	2	Excellent	0.45496	$ 157,319	3	Robinhaven	PI
lowed	$49,427	766	$36,528	Personal	Unemployed	10	Fair	0.14324	$ 287,140	0	New Heather	IL
orced	N/A	717	$15,613	Business	Unemployed	5	Fair	0.29598	N/A	4	Brianland	TN
lowed	$45,280	672	$6,553	Personal	Self-employed	1	Good	0.37889	N/A	0	West Lindaview	M

This is the formula i Used to calculate the LTV ratio. Although it is just (Loan Amount / Assets Value), excel could not do this with all the blanks so i wrote this script to replace the section when their is blanks to "Missing Data" and i also removed the ($) and the (,) signs in my script as it was creating an error. I then later found out in Power Bi it was better for the graphs to filter them the missing data out. So I corrected that.

Data Cleaning Challenges

1

Initial Data Import

Encountered issues with four rows containing hours but no names. These rows were removed to maintain data integrity and consistency.

2

Data Type Adjustments

Changed all data types, addressing mixed formats. Initially replaced N/A values with 0 to handle missing data, but this approach was later reconsidered.

3

Final Data Preparation

Realised that replacing N/A with 0 could skew the data incorrectly when plotting graphs. Decided to filter out these values instead to ensure accurate analysis.

Cleaned Data Overview

Excel Cleaning

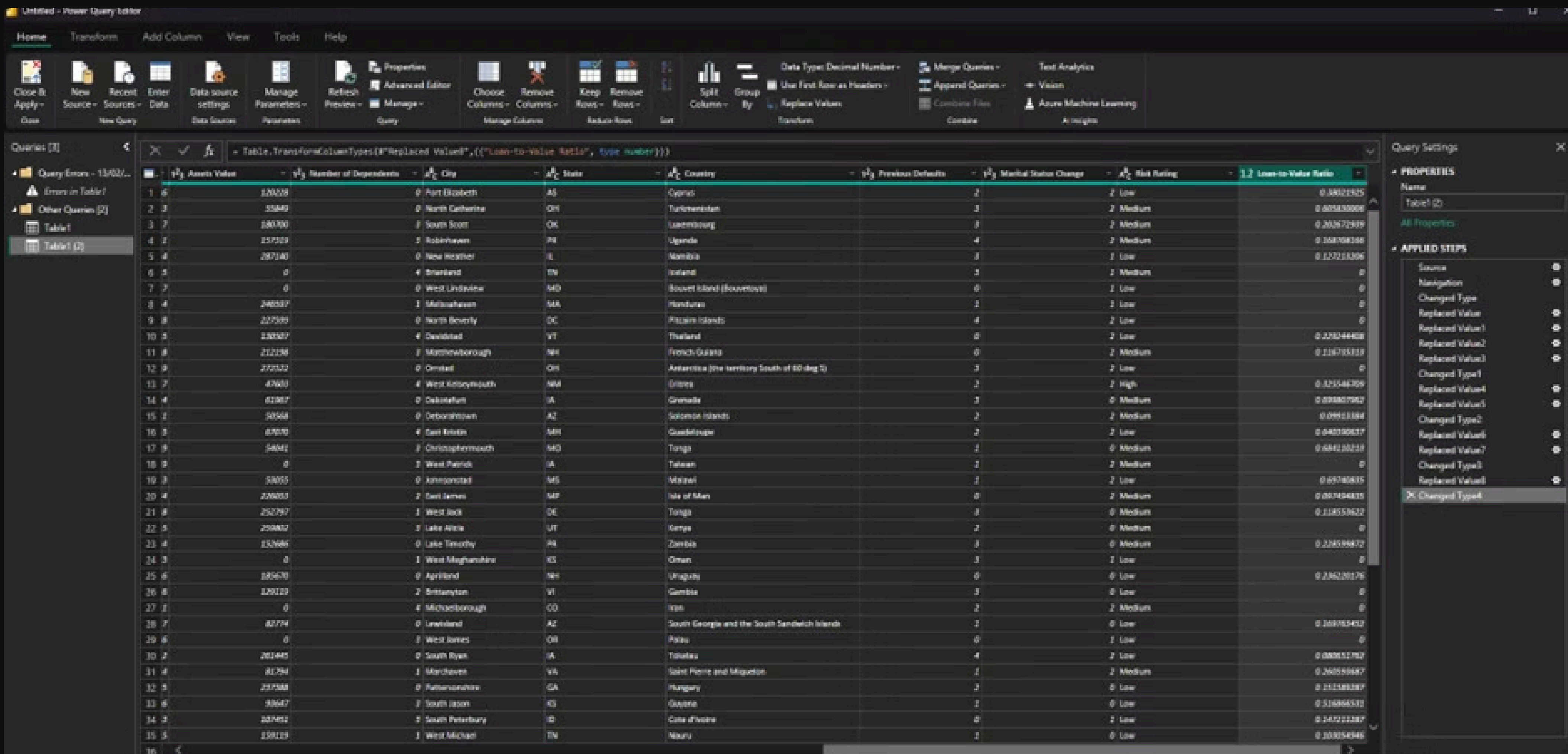
Initial data cleaning performed in Excel, addressing basic formatting issues and adding calculated columns like LTV ratio.

Power Query Refinement

Further data refinement in Power Query, ensuring all data types are set correctly and consistent across the dataset.

Final Dataset

The resulting dataset is clean, consistent, and ready for in-depth analysis and visualization in Power BI.



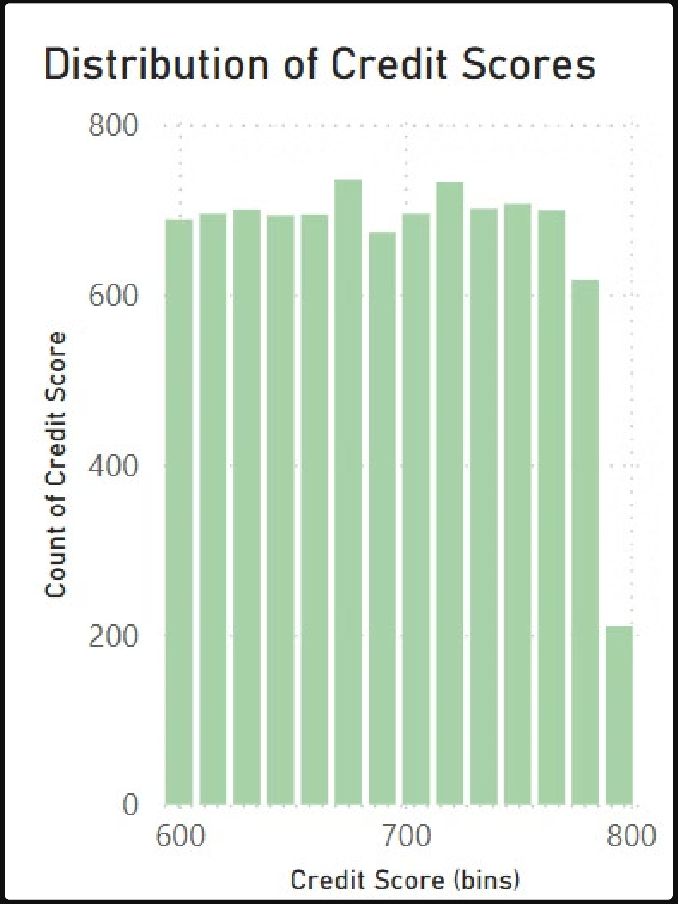
Power BI Dashboard

I then closed and applied all the data cleaning and began creating a dashboard with the Data. These are the types of graphs I believe would be most effective for visualising data from a financial risk assessment dataset, so I included them in the dashboard.

Credit Score and Loan Amount Analysis

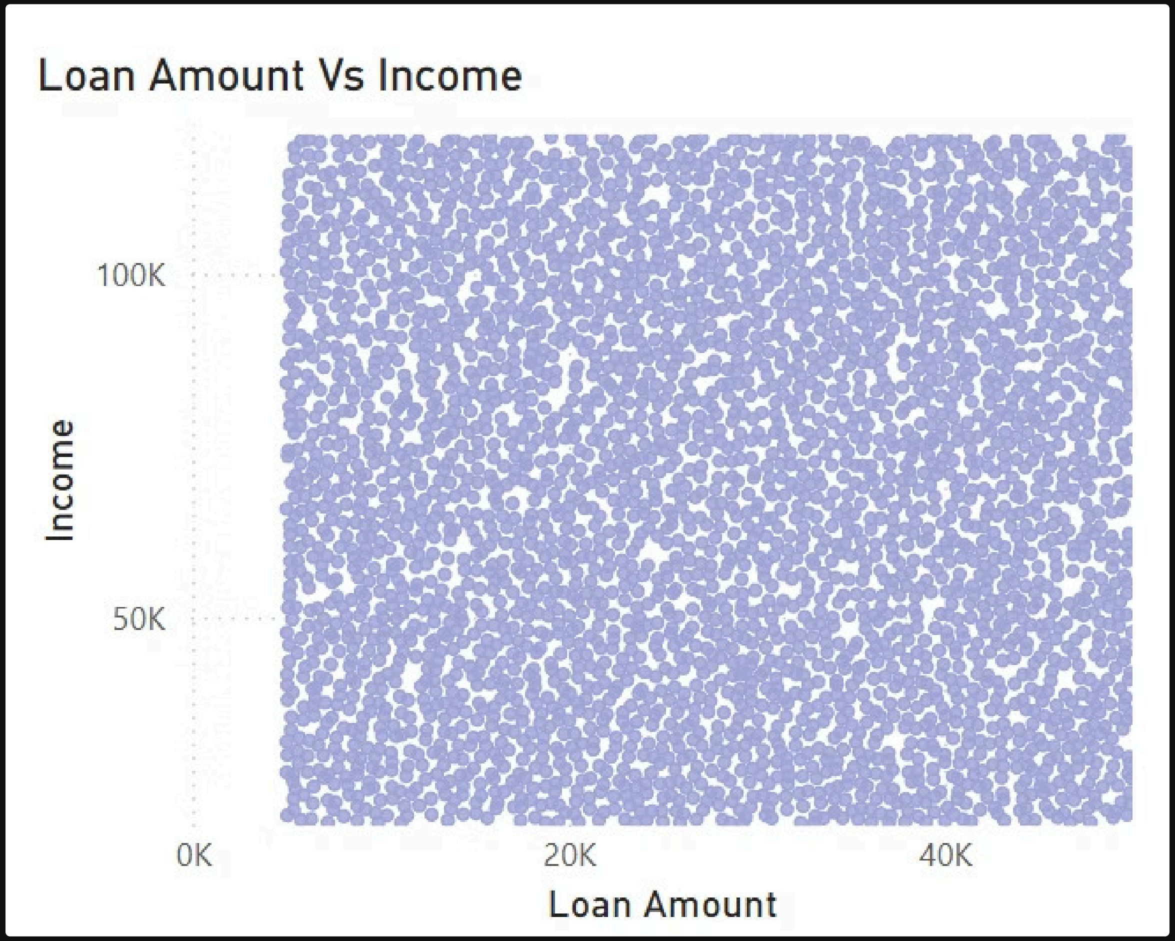
Distribution of Credit Scores

The histogram shows the spread of credit scores in the dataset, highlighting how many people fall within certain ranges. This visualisation helps in identifying the overall creditworthiness of the applicants.



Loan Amount vs. Income

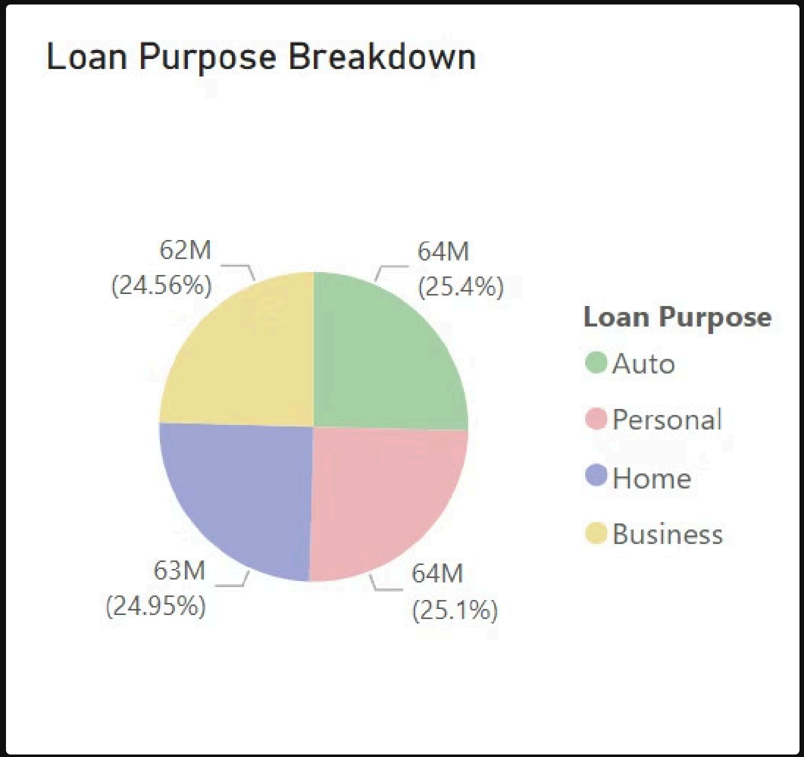
This scatter plot illustrates the relationship between loan amount and income. It helps identify if there's a correlation between higher loan amounts and income levels, while also highlighting potential outliers in the dataset. As you see here there is no correlation in this particular data set.



Loan Purpose and Risk Analysis

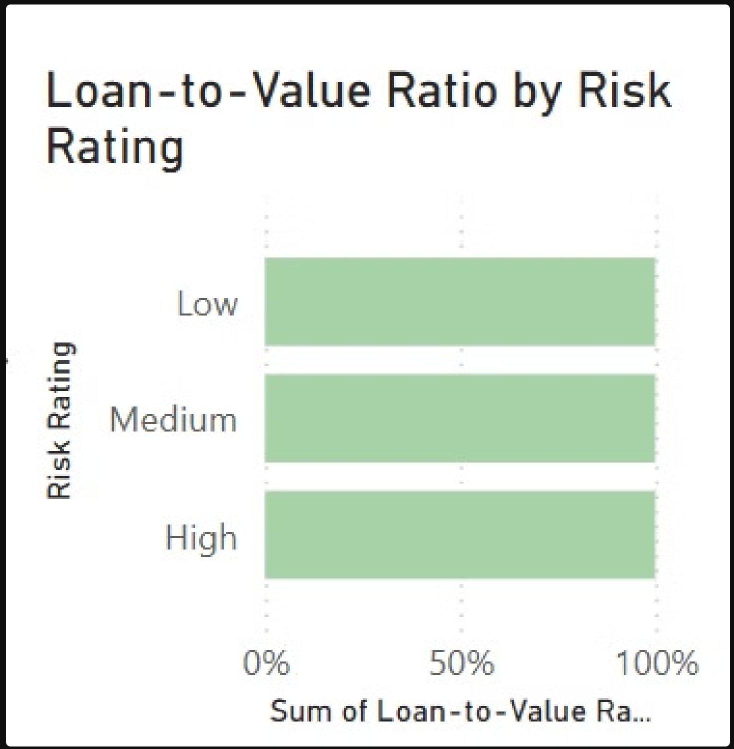
Loan Purpose Breakdown

This pie chart shows the distribution of loan purposes, such as home purchase, education, and debt consolidation. It provides insights into the primary reasons for loan applications and potential risk factors associated with different purposes.



Loan-to-Value Ratio by Risk Rating

This stacked bar chart displays the Loan-to-Value ratio across different risk ratings. It helps visualize any correlation between higher LTV ratios and higher risk assessments, providing valuable insights for risk management strategies.

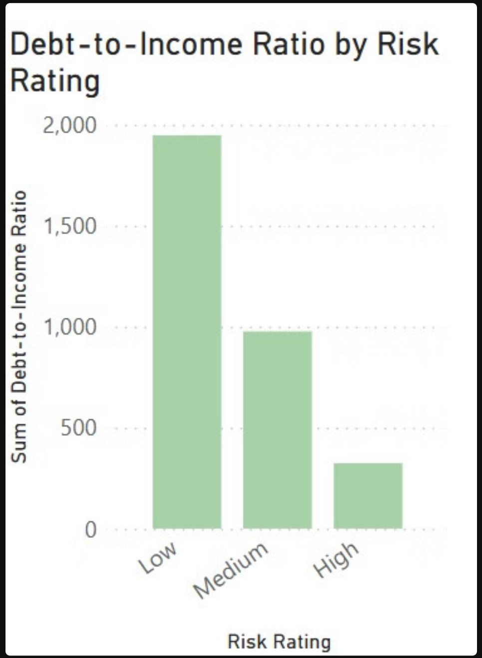


Made with Gamma

Debt-to-Income and Default Analysis

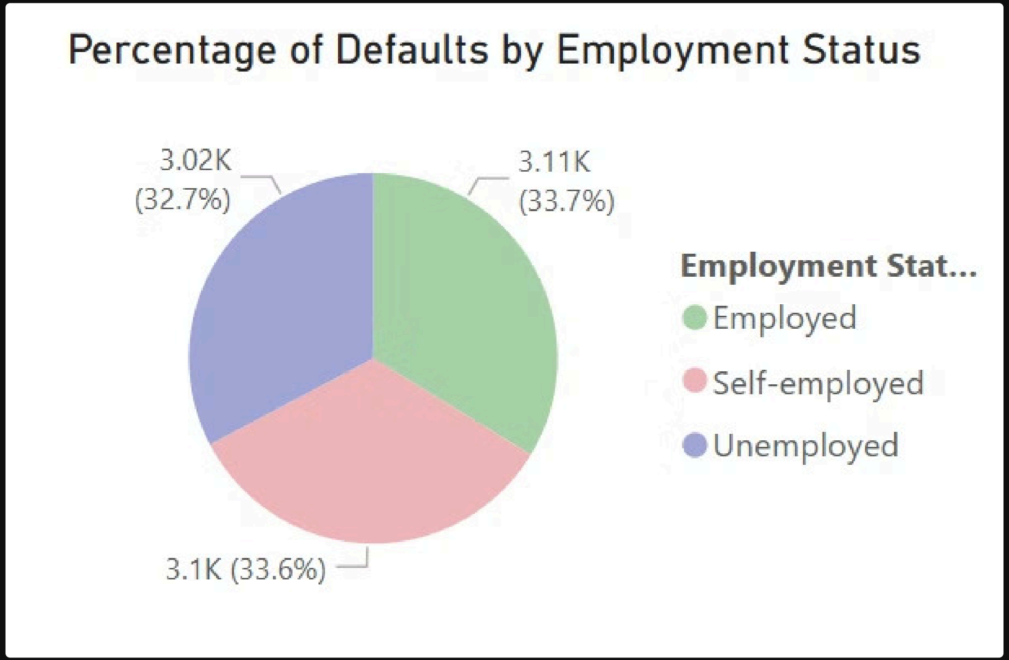
Debt-to-Income Ratio by Risk Rating

This clustered bar chart illustrates the debt-to-income ratio across different risk ratings. It helps assess whether higher debt-to-income ratios correspond with higher-risk ratings, providing crucial information for loan approval processes.



Percentage of Defaults by Employment Status

This chart shows the percentage of people with previous defaults grouped by their employment status. It reveals trends in risk related to employment status, such as whether unemployed individuals have a higher chance of defaults.



Next Steps

The next steps with the findings from these graphs would be to conduct a deeper analysis to identify patterns, correlations, and potential risk factors. For example, if the Loan-to-Value and Debt-to-Income ratios are consistently high for applicants with lower credit scores, stricter lending criteria may be necessary to mitigate risk. Similarly, if certain employment statuses or loan purposes show a higher percentage of defaults, the institution could refine its approval process or adjust interest rates accordingly. These insights can also guide policy adjustments, improve risk assessment models, and inform future decision-making. Additionally, further data segmentation and predictive modeling could be used to enhance risk evaluation and optimise loan approval strategies.