

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.

Financial Risk Assessment Data

0.23

Average of Loan-to-Value Ratio

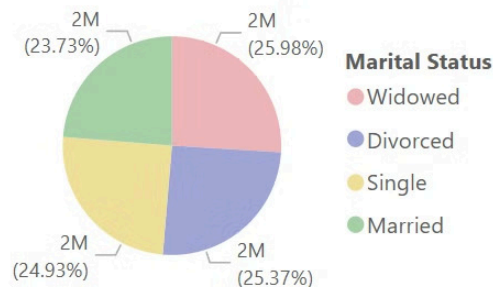
698.80

Average of Credit Score

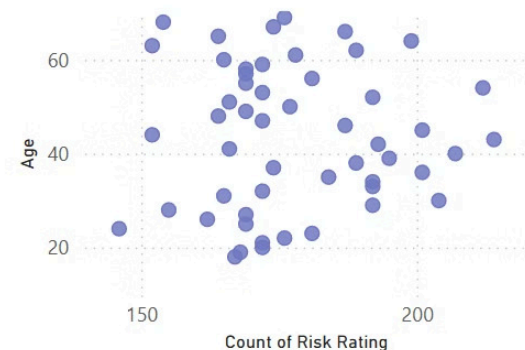
254M

Sum of Loan Amount

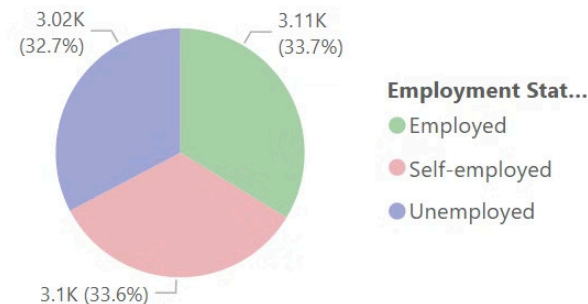
Credit Score by Marital Status



Risk Rating by Age



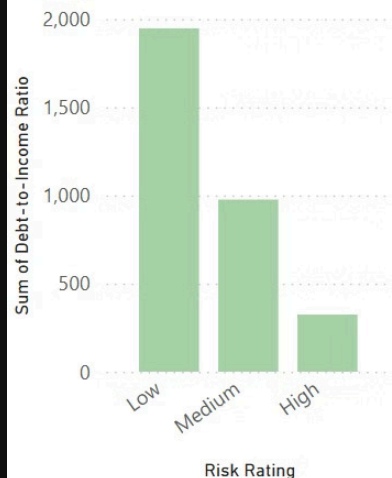
Percentage of Defaults by Employment Status



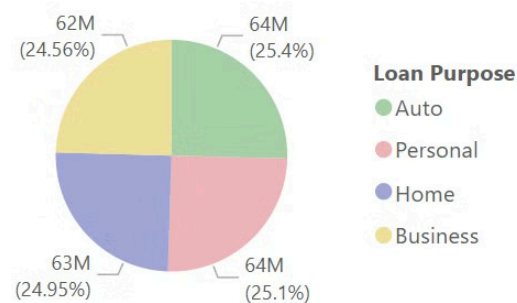
Loan-to-Value Ratio by Risk Rating



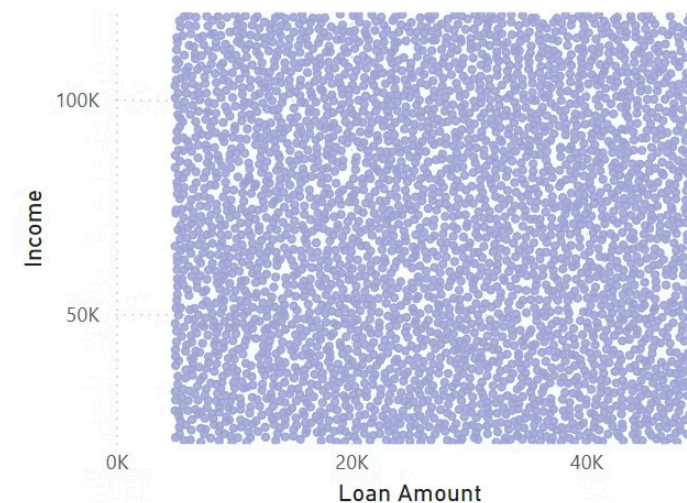
Debt-to-Income Ratio by Risk Rating



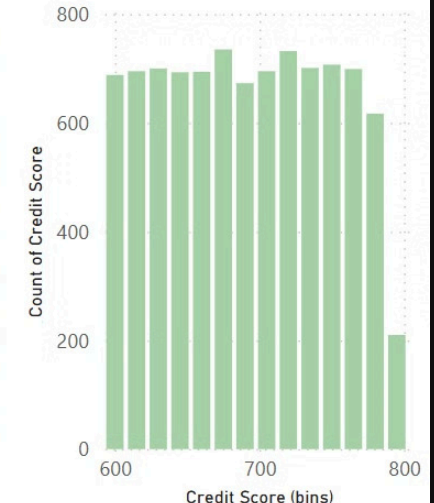
Loan Purpose Breakdown



Loan Amount Vs Income



Distribution of Credit Scores



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.

=IF(OR(G3="N/A", M3="N/A", ISBLANK(G3), ISBLANK(M3)), "Missing Data", VALUE(SUBSTITUTE(SUBSTITUTE(G3, "\$", ""), ", ", "")) / VALUE(SUBSTITUTE(M3, ", ", "")))																					
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
Age	Gender	Education Level	Marital Status	Income	Credit Score	Loan Amount	Loan Purpose	Employment Status	Years of Current Job	Payment History	Debt-to-Income Ratio	Assets Value	Number of Dependents	City	State	Country	Previous Defaults	Marital Status Change	Risk Rating	Loan-to-Value Ratio	
40	Male	PhD	Divorced	\$72,796	688	\$46,713	Business	Unemployed	29	Poor	0.19431	\$	120,228	0	Port Elizabeth	AE	Cyprus	2	2	Low	0.36
57	Female	Bachelor's	Widowed	N/A	680	\$71,836	Auto	Employed	0	Fair	0.14802	\$	10,849	0	North Catherine	OM	Turkmenistan	2	2	Medium	0.61
21	Non-binary	Master's	Single	\$66,667	680	\$36,623	Home	Employed	0	Fair	0.36240	\$	186,796	3	South Scott	OK	Luxembourg	3	2	Medium	0.28
39	Male	Bachelor's	Single	\$26,508	623	\$26,543	Personal	Unemployed	2	Excellent	0.45486	\$	157,310	2	Bucktown	FB	Uganda	4	2	Medium	0.17
26	Non-binary	Bachelor's	Widowed	\$49,427	796	\$36,528	Personal	Unemployed	10	Fair	0.14524	\$	267,248	0	New Heather	AL	Norfolk	3	1	Low	0.13
30	Non-binary	PhD	Divorced	N/A	727	\$25,613	Business	Unemployed	0	Fair	0.20286	N/A		4	Stratford	TM	Iceland	5	1	Medium	Missing Data
31	Non-binary	Master's	Widowed	\$46,288	672	\$6,163	Personal	Self-employed	1	Good	0.37389	N/A		0	West Lindaville	PD	British Indian Ocean Territory	0	1	Low	Missing Data
38	Male	Bachelor's	Widowed	\$65,678	N/A	N/A	Business	Unemployed	10	Poor	0.39664	\$	246,187	1	Midwesthaven	VA	Marshall	1	1	Low	Missing Data
32	Non-binary	Bachelor's	Widowed	\$20,256	718	N/A	Auto	Unemployed	4	Fair	0.33186	\$	227,189	0	North Beverly	DC	Pitcairn Islands	4	2	Low	Missing Data
55	Male	Bachelor's	Married	\$70,198	680	\$26,918	Personal	Self-employed	0	Excellent	0.48433	\$	130,187	4	Eastbrook	VT	Thailand	0	2	Low	0.23
42	Non-binary	Master's	Single	\$116,232	787	\$24,771	Home	Employed	11	Excellent	0.12433	\$	212,396	3	Matthewborough	NH	French Guiana	0	2	Medium	0.12
37	Non-binary	Master's	Divorced	\$76,810	718	N/A	Home	Self-employed	17	Poor	0.18782	\$	272,120	0	Orwell	OM	Antarctica (the territory South of 60 deg S)	3	2	Low	Missing Data
60	Female	PhD	Divorced	N/A	686	\$15,487	Personal	Self-employed	18	Poor	0.10817	\$	47,680	4	West Fitzpatrick	NH	Switzerland	2	2	High	0.10
60	Male	High School	Divorced	\$79,404	688	\$43,365	Personal	Self-employed	10	Good	0.12028	\$	61,967	0	Dorchester	VA	Grenada	3	0	Medium	0.70
29	Male	High School	Divorced	\$116,118	N/A	\$6,163	Auto	Employed	17	Poor	0.48972	\$	96,168	0	Dorchester	AZ	Saint Helena	2	2	Medium	0.10
60	Female	PhD	Divorced	N/A	N/A	\$42,163	Auto	Self-employed	0	Fair	0.15275	\$	87,670	4	East Kismet	PR	Guadeloupe	2	2	Low	0.64
55	Male	High School	Married	\$79,878	796	\$36,178	Personal	Unemployed	20	Excellent	0.26894	\$	94,843	3	Christchurch	PD	Togo	1	0	Medium	0.88
50	Female	Master's	Widowed	\$103,628	684	\$25,443	Home	Unemployed	8	Poor	0.16183	N/A		3	West Patrick	VA	Taiwan	1	2	Medium	Missing Data
60	Non-binary	Bachelor's	Divorced	N/A	718	\$17,183	Business	Employed	7	Fair	0.13897	\$	93,186	0	Johnston	PD	Myanmar	1	2	Low	0.70
50	Non-binary	PhD	Married	\$71,884	780	\$22,030	Personal	Employed	10	Fair	0.17165	\$	226,153	2	East James	WP	Isle of Man	0	2	Medium	0.10
26	Female	PhD	Divorced	\$61,488	N/A	\$26,178	Business	Unemployed	10	Excellent	0.43823	\$	252,797	1	West Jack	DE	Togo	3	0	Medium	0.12
27	Non-binary	PhD	Widowed	\$103,108	796	N/A	Home	Employed	10	Poor	0.17474	\$	258,183	3	Lake Alexia	UT	Korea	2	0	Medium	Missing Data
34	Male	PhD	Single	N/A	684	\$34,184	Auto	Self-employed	12	Excellent	0.10828	\$	152,686	0	Lake Trinity	PR	Zambia	3	0	Medium	0.20
64	Non-binary	Bachelor's	Divorced	\$110,718	712	\$41,628	Business	Unemployed	10	Poor	0.22830	N/A		1	West Highdown	KS	Oman	3	1	Low	Missing Data
56	Female	High School	Divorced	\$61,811	641	\$43,189	Auto	Self-employed	10	Excellent	0.10971	\$	186,670	0	Apollonia	NH	Uruguay	0	0	Low	0.24
41	Male	PhD	Single	\$84,786	742	N/A	Business	Unemployed	10	Good	0.15511	\$	126,119	2	Brimingham	VI	Gambia	3	0	Low	Missing Data
30	Non-binary	PhD	Single	\$86,386	867	\$28,134	Auto	Employed	8	Poor	0.10880	N/A		4	Richardborough	CD	Iran	2	2	Medium	Missing Data
41	Male	PhD	Married	\$83,133	761	\$14,163	Business	Employed	13	Poor	0.18181	\$	82,774	0	Lowland	AZ	South Georgia and the South Sandwich Islands	1	0	Low	0.17
64	Female	PhD	Married	N/A	740	\$38,638	Personal	Self-employed	10	Fair	0.10723	N/A		0	West James	CB	Poland	0	1	Low	Missing Data

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.

=IF(OR(G2="N/A", M2="N/A", ISBLANK(G2), ISBLANK(M2)), "Missing Data", VALUE(SUBSTITUTE(SUBSTITUTE(G2, "\$", ""), ", ", "")) / VALUE(SUBSTITUTE(M2, ", ", "")))												
D	E	F	G	H	I	J	K	L	M	N	O	P
Martial 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	State
Forced	\$72,799	688	\$45,713	Business	Unemployed	19	Poor	0.15431	\$ 120,228	0	Port Elizabeth	AS
Widowed	N/A	690	\$33,835	Auto	Employed	6	Fair	0.14892	\$ 55,849	0	North Catherine	OR
Single	\$55,687	600	\$36,623	Home	Employed	8	Fair	0.36240	\$ 180,700	3	South Scott	OR
Single	\$26,508	622	\$26,541	Personal	Unemployed	2	Excellent	0.45496	\$ 157,319	3	Robinhaven	PR
Widowed	\$49,427	766	\$36,528	Personal	Unemployed	10	Fair	0.14324	\$ 287,140	0	New Heather	IL
Forced	N/A	717	\$15,613	Business	Unemployed	5	Fair	0.29598	N/A	4	Brianland	TH
Widowed	\$45,280	672	\$6,553	Personal	Self-employed	1	Good	0.37889	N/A	0	West Lindaview	

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

Initial Data Import

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

1

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.

Table: TransformColumnTypes(If Replaced Value, ({Loan-to-Value Ratio, type number}))

	Assets Value	Number of Dependents	City	State	Country	Previous Defaults	Marital Status Change	Risk Rating	Loan-to-Value Ratio
1	6	120228	Port Elizabeth	AS	Cyprus	2	2	Low	0.38021825
2	3	55849	North Catherine	OH	Turkmenistan	3	2	Medium	0.605833006
3	7	180700	South Scott	OK	Laos	3	2	Medium	0.202672919
4	2	157323	Robbinston	PR	Uganda	4	2	Medium	0.168708388
5	4	287140	New Heather	IL	Namibia	3	2	Low	0.127213396
6	5	0	Briarland	TN	Iceland	3	2	Medium	0
7	7	0	West Underline	MD	Bouvet Island (Bouvetøya)	0	2	Low	0
8	4	246537	Melrosehaven	MA	Honduras	2	2	Low	0
9	8	227539	North Beverly	OC	Pitcairn Islands	4	2	Low	0
10	5	130507	Devilsdell	VT	Thailand	0	2	Low	0.228244408
11	8	212238	Matthewborough	NH	French Guiana	0	2	Medium	0.116795313
12	8	272522	Orrsted	OH	Antarctica (the territory South of 60 deg S)	3	2	Low	0
13	7	47603	West Kinsmouth	NM	Ghana	2	2	High	0.321546709
14	4	61367	Oakdale	IA	Grenada	3	0	Medium	0.898807982
15	2	30568	Deborahstown	AZ	Solomon Islands	2	2	Medium	0.09911384
16	5	67070	East Evelyn	MH	Guadeloupe	2	2	Low	0.640390677
17	9	54041	Christophersmouth	MO	Tonga	2	0	Medium	0.684138213
18	9	0	West Patrick	IA	Taiwan	2	2	Medium	0
19	3	50055	Johnsonstad	MS	Malawi	3	2	Low	0.69740815
20	4	226053	East James	MP	Isle of Man	0	2	Medium	0.067454833
21	8	252797	West Jack	OE	Tonga	3	0	Medium	0.118553622
22	5	259802	Lake Alice	UT	Kenya	2	0	Medium	0
23	4	152686	Lake Timothy	PR	Zambia	3	0	Medium	0.228599872
24	3	0	West Meghamshire	KS	Oman	3	2	Low	0
25	6	185670	Apriland	NH	Uruguay	0	0	Low	0.236230176
26	8	129123	Brittanyton	VT	Gambia	3	0	Low	0
27	1	0	McDonoughborough	CO	Iran	2	2	Medium	0
28	7	82774	Lowland	AZ	South Georgia and the South Sandwich Islands	2	0	Low	0.189763452
29	6	0	West James	OR	Palau	0	2	Low	0
30	2	261445	South Ryan	IA	Tuvalu	4	2	Low	0.080611762
31	4	81794	Marchaven	VA	Saint Pierre and Miquelon	2	2	Medium	0.240593687
32	5	237588	Pattersonshire	GA	Hungary	2	0	Low	0.151381387
33	6	93647	South Jason	KS	Guinea	2	0	Low	0.516866511
34	3	107452	South Peterbury	ID	Cote d'Ivoire	0	2	Low	0.247211287
35	5	159123	West Michael	TN	Nauru	2	0	Low	0.103054946
36									

Power BI Dashboard

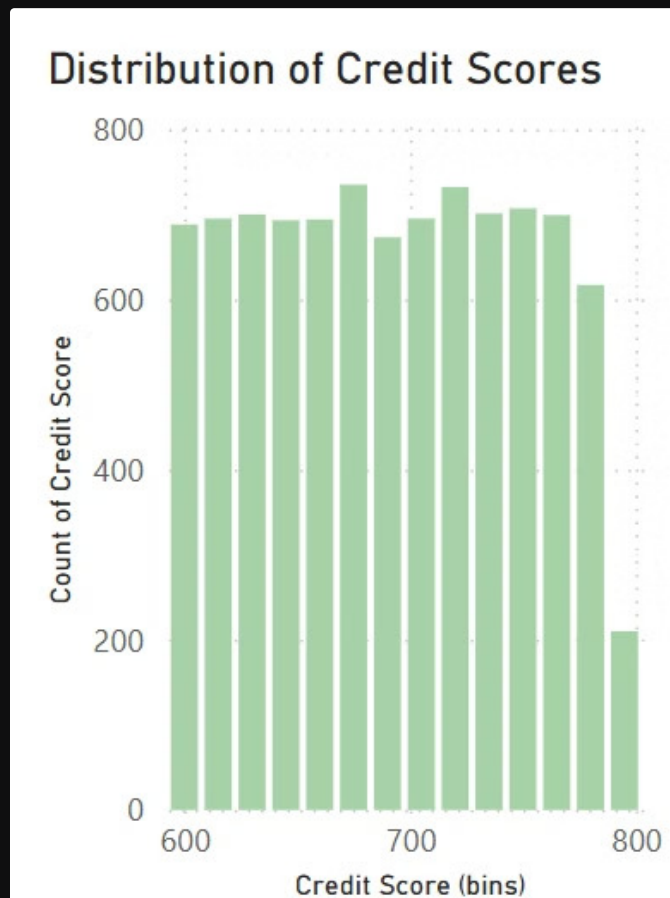
I then closed and applied all the data cleaning and began creating a dashboard with the Data. These are the types of graphs I believe would be most effective for visualising data from a financial risk assessment dataset, so I included them in the dashboard.



Credit Score and Loan Amount Analysis

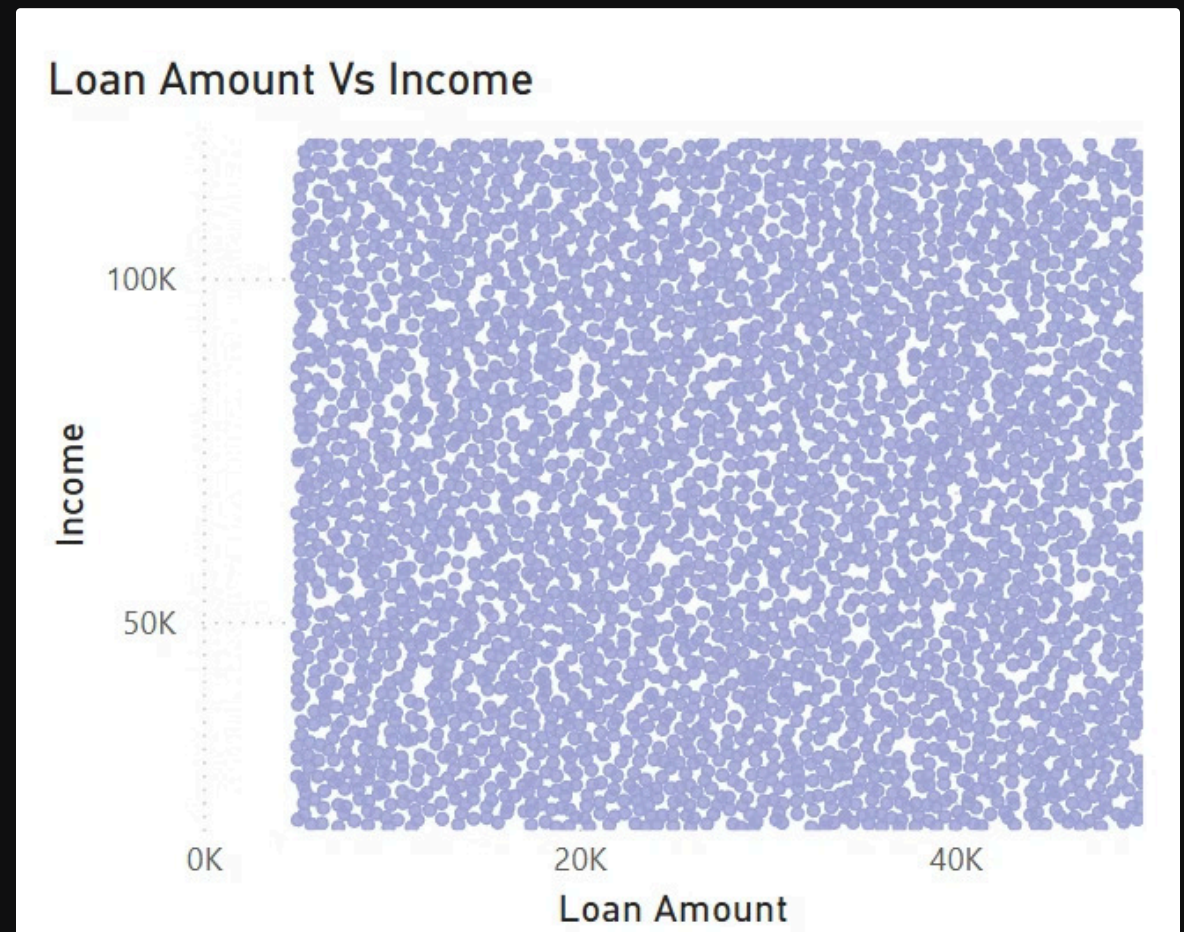
Distribution of Credit Scores

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

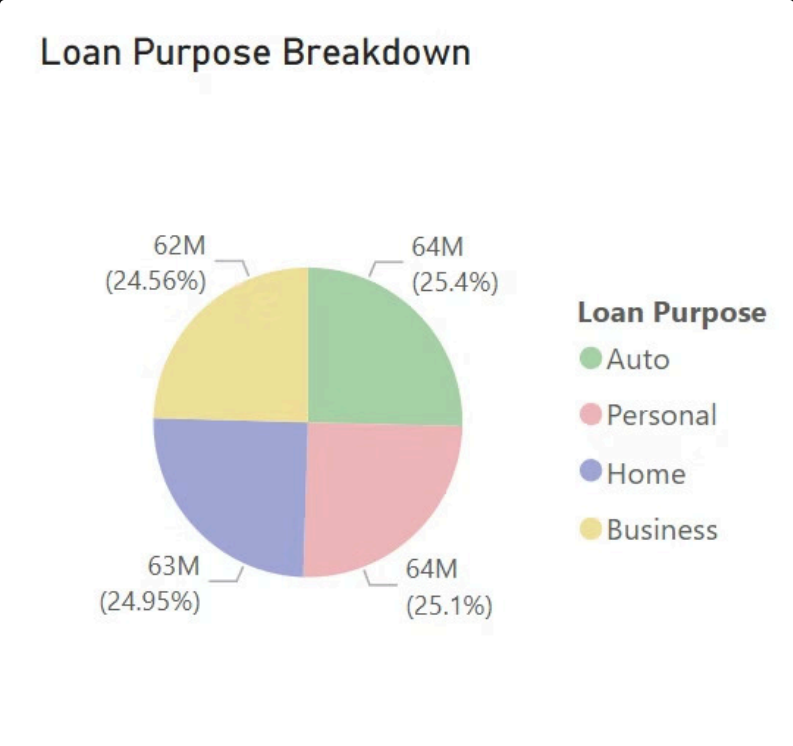
This scatter plot illustrates the relationship between loan amount and income. It helps identify if there's a correlation between higher loan amounts and income levels, while also highlighting potential outliers in the dataset. As you see here there is no correlation in this particular data set.



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

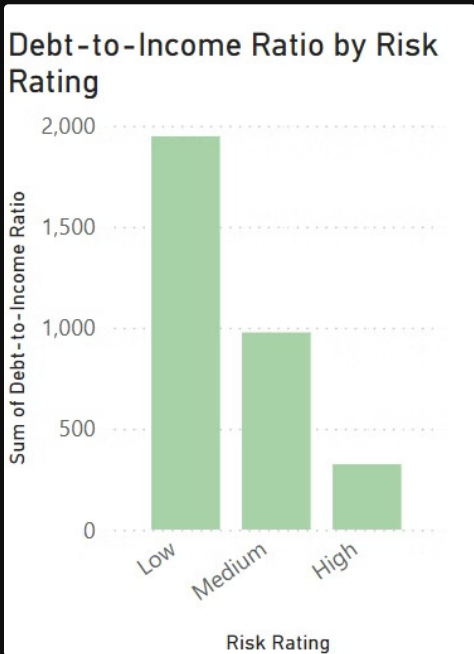
This stacked bar chart displays the Loan-to-Value ratio across different risk ratings. It helps visualize any correlation between higher LTV ratios and higher risk assessments, providing valuable insights for risk management strategies.



Debt-to-Income and Default Analysis

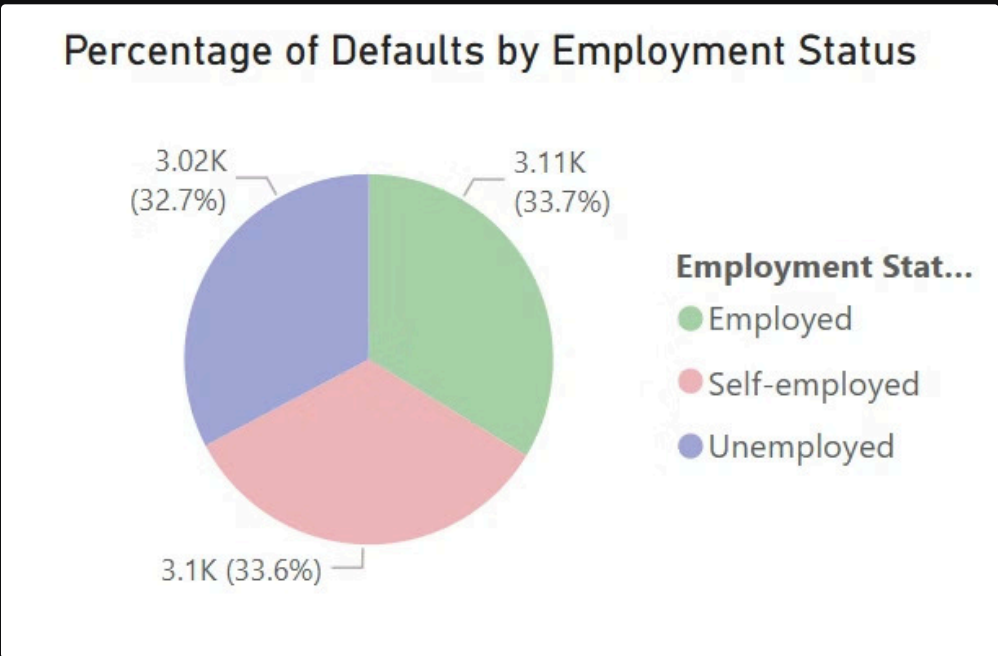
Debt-to-Income Ratio by Risk Rating

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.