Introduction

This report examines a dataset containing loan application records and the approval rate associated and reasons for whether the loan was approved or rejected. The data has detailed borrower demographics, such as education and sex, financial and credit history, loan details and outcomes.

By analysing these attributes, we aim to uncover any correlations between the data and see if there are predictors that can lead to increased loan approval. The study highlights the interplay of borrower characteristics, economic indicators and creditworthiness in shaping lending decisions. The research questions we asked are outlined below and we used these to create graphs and insights into the data.

Research Questions

1. who gets more loan approvals men or women?

2. how loan approval dependent on age, education, income

3. Credit Score Influence: Higher credit scores lead to higher loan approval rates.

4. Previous Defaults Impact: Applicants with previous loan defaults are less likely to get approved.

5. Home Ownership Effect: Homeowners (OWN or MORTGAGE) have higher approval rates than renters.

6. Credit History Length: A longer credit history increases the likelihood of loan approval.

7. Income vs. Interest Rate: Higher-income individuals receive lower interest rates on loans.

8. Loan Intent Relation: The purpose of the loan (e.g., EDUCATION, MEDICAL) affects approval chances.

9. Education Level: Higher education levels correlate with higher loan amounts approved.

10. Employment Experience: More years of employment experience increase approval rates.

11. Loan Amount Requested: Larger loan amounts decrease the likelihood of approval.

12. Loan Percent Income: A higher loan-to-income ratio reduces the chance of loan approval.

Data Collection and Cleaning

We looked for datasets on Kaggle, we looked at various datasets and tried to find datasets that had a large enough amount of data so that we could work with the data to analyse the data and create the relevant graphs to display what we found.   
We decided on the loan dataset as it was large enough and had a lot of different loan characteristics to explore, making it an ideal choice for analysis.

The data provides detailed demographic data, including age, gender and education levels of borrowers. It also has some financial and employment details, like income, employment experience and residential ownership status. The data also has loan characteristic data like the loan amount, the purpose of the loan, interest rates and the loan outcomes. The dataset has the credit and financial history of the borrower which include, length of credit history in years, credit scores and whether the borrower has had previous loans and if they have defaulted on them, meaning they haven’t paid them back.

How to increase your loan approval rate?

## Analysis of the Education Chart

Key Observations:

High School: A significantly higher number of applications were denied compared to approvals.

Bachelor's and Associate's Degrees: Similar trends are observed, with denials outnumbering approvals, but the absolute number of applications is higher than for High School graduates.

Master's Degree: There are still more denials than approvals, though the approval numbers seem relatively higher compared to lower education levels.

Doctorate: The number of applications is much smaller, and denials dominate this group.

Higher education levels, such as Master's and Doctorate, seem to improve the likelihood of approval compared to lower levels, but even here, denials dominate.

The trend suggests that while education level influences approval rates, it is not the sole factor. Other variables such as income, credit score, or employment history likely play a significant role. There is also not a big disparity between the loan and approval rates with respect to education level, higher education levels do tend to get more loans approved but not at a much higher rate, which is what we thought would be the case. The two education levels with the closest application levels are ‘High School’ and ‘Associate’ which have very similar approval rates further indicating that other factors are considered when loan applications are submitted.

## Analysis of the Education Proportionate Chart

Minimal Variation Across Education Levels:

Across all education levels, the proportion of denials consistently hovers around 77-78%, indicating a general trend of low loan approval rates.

The proportion of approvals (green bars) and denials (red bars) is relatively stable, regardless of the education level:

Associate's: 22.0% approvals, 78.0% denials.

Bachelor's: 22.5% approvals, 77.5% denials.

Doctorate: 22.9% approvals, 77.1% denials.

High School: 22.3% approvals, 77.7% denials.

Master's: 21.8% approvals, 78.2% denials.

Highest Approval Rate:

Applicants with a Doctorate have the highest approval rate at 22.9%. They also have the lowest application rate, this is likely due to less people having Doctorates which could mean the data is misrepresented.

Lowest Approval Rate:

Applicants with a master’s degree have the lowest approval rate at 21.8%, which is slightly lower than others.

Education Level Alone May Not Be a Decisive Factor:

The similar approval and denial percentages across all education levels suggest that education level alone is not a strong determinant for loan approval.

Other factors, such as income, credit score, employment history, and loan amount, might have a more significant impact.

High Denial Rates Across All Levels:

The consistently high denial rates could indicate strict loan approval criteria or a general lack of eligibility among applicants, regardless of education level.

Doctorate Applicants:

Although they have the highest approval percentage, the difference is marginal and not statistically significant in this dataset.

## Analysis of the Income Group charts

Key Observations:

0-20k Income Group:

Significantly more approved applications than denied ones.

Represents the highest approval rate compared to other groups.

20k-40k Income Group:

Denials dominate this group, with a smaller but notable number of approvals.

40k-60k Income Group:

Denials continue to outnumber approvals by a significant margin.

60k-80k Income Group:

Very few approvals and a sharp increase in the number of denials compared to the previous group.

80k-100k Income Group:

The trend of denials dominating approvals is most pronounced in this group, showing the largest discrepancy.

Insights:

Applicants in the 0-20k income group have the highest absolute number of approvals, likely due to fewer risks associated with their loans or potentially smaller loan requests.

Approval numbers consistently drop as income increases, which might seem counterintuitive and may warrant deeper exploration of other factors influencing these trends (e.g., credit score, loan amount).

Key Observations:

0-20k Income Group:

83.4% approved and only 16.6% denied—the highest approval rate across all income groups.

20k-40k Income Group:

Approval percentage drops sharply to 41.5%, with 58.5% denied.

40k-60k Income Group:

Approval percentage continues to decline to 25.8%, with denials making up 74.2% of applications.

60k-80k Income Group:

Approval rate further drops to 18.3%, with 81.7% denied.

80k-100k Income Group:

The lowest approval rate was at 15.2%, with the highest denial rate at 84.8%.

Insights:

Approval rates decrease steadily as income increases, while denial rates increase. This could indicate that higher-income applicants are applying for larger loans or riskier terms that are more likely to be denied.

The 0-20k income group stands out with an extremely high approval rate, which could be due to smaller loan requests or other factors that make them less risky for lenders.

Overall Analysis

Approval Trends:

Contrary to typical expectations, higher-income groups are not favored for loan approvals. This suggests that factors other than income, such as loan amount requested, credit history, or previous defaults, might play a more significant role in approval decisions.

Significance of the 0-20k Group:

This group has the highest approval rate, possibly because their loan applications are more modest or align better with lender expectations.

Potential Bias Against Higher-Income Groups:

The steep decline in approval rates for higher-income groups may indicate that lenders perceive these applicants as higher-risk, potentially due to larger loan amounts or financial liabilities not reflected in the income figures.

## Analysis of the Home Ownership chart

The chart presents the loan approval and denial rates categorized by the type of home ownership. It is a stacked bar chart where the red segments represent denied loans and the green segments represent approved loans. Here is a detailed analysis:

1. General Trends

Across all home ownership categories, the proportion of denied loans (red) is consistently higher than approved loans (green).

The denial rates vary significantly depending on the type of home ownership.

2. Breakdown by Home Ownership Categories

Mortgage:

Denial Rate: 88.4%

Approval Rate: 11.6%

This group experiences one of the highest denial rates, indicating applicants with mortgages are less likely to have their loans approved.

Other:

Denial Rate: 66.7%

Approval Rate: 33.3%

This category shows a relatively better approval rate compared to others, with one-third of the applicants having their loans approved.

Own:

Denial Rate: 92.5%

Approval Rate: 7.5%

The group with the highest denial rate, indicating owning a home outright may not be a significant advantage for loan approval in this dataset.

Rent:

Denial Rate: 67.6%

Approval Rate: 32.4%

Similar to the "Other" category, renters have a relatively higher approval rate compared to mortgage holders and homeowners.

3. Key Observations

Applicants categorized under "Own" face a surprisingly high denial rate (92.5%), which contradicts the general assumption that owning a home outright may lead to better loan approval chances. This may indicate additional underlying factors influencing the decisions.

"Other" and "Rent" categories have the most favorable approval rates, suggesting less stringent criteria for these groups or differences in applicant profiles.

Applicants with a mortgage face a lower denial rate compared to those who own their homes outright, which could imply that having a mortgage is less of a deterrent than expected.

## Analysis of the Loan Approval Status Distribution by Loan Intent chart

1. General Trends

Across all loan intents, denial rates (red) are higher than approval rates (green), suggesting that approval is generally challenging irrespective of the intent.

The approval rates vary significantly depending on the purpose of the loan.

2. Breakdown by Loan Intent

Debt Consolidation:

Approval Rate: 30.3%

Denial Rate: 69.7%

Among all intents, debt consolidation has the highest approval rate. This indicates that lenders may be more inclined to approve loans intended for consolidating existing debts, possibly viewing it as a step toward better financial management.

Education:

Approval Rate: 17.0%

Denial Rate: 83.0%

Education loans face high denial rates, reflecting the potential uncertainty associated with repayment due to the long-term nature of education investments and possibly low current income levels of applicants.

Home Improvement:

Approval Rate: 26.3%

Denial Rate: 73.7%

Home improvement loans perform relatively well compared to other categories, likely due to the association of such loans with increasing property value, which may be viewed positively by lenders.

Medical:

Approval Rate: 27.8%

Denial Rate: 72.2%

Medical loans also exhibit a moderate approval rate. This may be because such loans are often for necessary expenses, and lenders might see a moral obligation to support applicants in these situations.

Personal:

Approval Rate: 20.1%

Denial Rate: 79.9%

Personal loans show a lower approval rate, as these loans are less tied to specific investments or assets, which could increase the perceived risk for lenders.

Venture:

Approval Rate: 14.4%

Denial Rate: 85.6%

Loans intended for business ventures have the lowest approval rate, reflecting the high-risk nature of startup and entrepreneurial financing.

3. Key Observations

Debt Consolidation Stands Out:

It has the highest approval rate, likely because lenders view these loans as part of a strategy to reduce financial liabilities, thus lowering risk.

High Denial Rates for Education and Ventures:

Education loans are likely denied due to applicants' lower income levels and uncertain future earning potential.

Venture loans are perceived as the riskiest due to the inherently uncertain nature of startups and new business initiatives.

Moderate Performance of Home Improvement and Medical Loans:

These categories have slightly better approval rates, possibly because they are tied to tangible outcomes (e.g., increased property value, critical medical care).

4. Implications

Loan approval decisions are highly sensitive to the intent of the loan, with lenders favoring categories that either reduce financial risk (e.g., debt consolidation) or have tangible outcomes (e.g., home improvement).

Higher denial rates for personal and venture loans suggest that these intents may benefit from additional collateral, stronger income proofs, or co-signers to improve approval chances.

## Analysis of the Chart: Loan Approval Status by Credit History Length

This chart illustrates the percentage of loan approvals (green) and denials (red) based on the length of an applicant's credit history. Each bar represents a specific range of credit history duration, and the percentages are stacked within each bar.

1. General Trends

Loan denials (red) dominate across all credit history length groups, with denial rates consistently above 75%.

Loan approval rates (green) slightly decline as the length of credit history increases, but the changes are marginal.

2. Breakdown by Credit History Length

2–4 Years:

Approval Rate: 23.2%

Denial Rate: 76.8%

Applicants with shorter credit histories have the highest approval rate among the groups, indicating that lenders may consider shorter credit histories more favorably than expected.

4–6 Years:

Approval Rate: 22.4%

Denial Rate: 77.6%

The approval rate drops slightly compared to the 2–4 year group, suggesting that moderate-length credit histories don’t offer much advantage.

6–8 Years:

Approval Rate: 21.7%

Denial Rate: 78.3%

As credit history length increases, the approval rate continues to decline, hinting at lenders’ preferences for other financial indicators over longer credit histories.

8–10 Years:

Approval Rate: 21.2%

Denial Rate: 78.8%

Denial rates increase marginally, and approval rates drop further, continuing the downward trend with longer credit history lengths.

>10 Years:

Approval Rate: 21.1%

Denial Rate: 78.9%

Applicants with the longest credit histories have the lowest approval rate, suggesting diminishing returns for having an extended credit history.

3. Key Observations

No Significant Advantage for Longer Credit Histories:

Contrary to expectations, longer credit histories (beyond 4 years) do not substantially improve approval rates. Approval rates are higher for shorter credit histories (2–4 years).

High Denial Rates Across All Groups:

The consistent denial rates above 75% for all credit history lengths indicate that other factors, such as credit score, income, or debt-to-income ratio, might play a more decisive role in loan approvals.

Declining Approval Rates:

Approval rates drop progressively as credit history length increases, which may reflect underlying trends such as over-reliance on credit or other financial liabilities associated with longer histories.

4. Implications

Length of Credit History Alone Is Not Sufficient:

The data suggests that simply having a longer credit history is not enough to secure a loan. Lenders likely weigh other factors more heavily, such as the quality and consistency of the credit history (e.g., on-time payments, debt levels).

Shorter Credit Histories Are Not a Barrier:

Applicants with shorter credit histories (2–4 years) have slightly better approval rates, implying that lenders are not overly biased against newer borrowers.

## Analysis of the Loan Approval Status by Loan Percent Income chart

1. General Trends

There is a clear relationship between the loan percent income and the likelihood of loan approval.

As the percentage of income that the loan represents increases, the loan denial rate generally rises, especially in lower-income brackets.

Approval rates are significantly higher in the middle ranges of loan percent income (20%-60%).

2. Breakdown by Loan Percent Income Groups

0–10%:

Approval Rate: 11.8%

Denial Rate: 88.2%

Surprisingly, applicants requesting loans that constitute only 0–10% of their income face the highest denial rates. This could be due to other factors such as poor creditworthiness or other unfavorable financial indicators.

10–20%:

Approval Rate: 16.7%

Denial Rate: 83.3%

While marginally better than the 0–10% group, this category still has a high denial rate, which may again point to external factors beyond the loan-to-income ratio.

20–30%:

Approval Rate: 39.3%

Denial Rate: 60.7%

A notable improvement in approval rates begins here, indicating that this range might represent a more balanced loan-to-income ratio that aligns better with approval criteria.

30–40%:

Approval Rate: 72.1%

Denial Rate: 27.9%

This group shows one of the highest approval rates, suggesting that loans in this range are perceived as affordable by lenders.

40–50%:

Approval Rate: 75.6%

Denial Rate: 24.4%

Similar to the 30–40% group, this range enjoys a high approval rate, reinforcing the notion that loans within 30–50% of income are generally viewed as manageable.

50–60%:

Approval Rate: 71.9%

Denial Rate: 28.1%

Approval rates remain strong but dip slightly compared to the 40–50% group, likely reflecting an increased risk associated with higher loan-to-income ratios.

60–70%:

Approval Rate: 66.7%

Denial Rate: 33.3%

Approval rates begin to decline as loan amounts become a larger proportion of income.

3. Key Observations

Counterintuitive Denial Rates at Lower Loan Percentages:

Applicants in the 0–10% and 10–20% income groups face surprisingly high denial rates. This may be influenced by:

Poor credit histories.

Low income despite the low loan request amount.

Other unfavorable financial metrics not reflected in the loan percent income alone.

Optimal Approval Ranges:

Loan amounts that are 20–50% of income appear to strike a balance between affordability and lender risk, yielding the highest approval rates.

Declining Approval Rates Beyond 50%:

As the loan percent income exceeds 50%, approval rates begin to drop, reflecting lender concerns about the borrower’s ability to manage repayment alongside existing financial obligations.

4. Implications

The relationship between loan percent income and approval status is not linear, suggesting that other factors (e.g., credit score, employment stability, existing debt) heavily influence approval decisions.

Lower approval rates at both ends of the spectrum (0–10% and >60%) highlight that both excessively conservative and ambitious loan requests are viewed less favorably.

## Analysis of Feature Importances in Predicting Loan Approval

This visualization and table depict the relative importance of various features in determining loan approval outcomes. The features are ranked based on their importance percentage, with higher values indicating greater influence on the predictive model. Here’s a detailed analysis:

1. Key Observations

Dominant Importance of Previous Loan Defaults:

The feature Previous\_Loan\_Defaults stands out as the most critical determinant, accounting for 77.28% of the importance.

This indicates that the history of loan defaults is overwhelmingly influential in predicting loan approval. Applicants with prior defaults likely face significantly higher rejection rates.

Person Home Ownership:

Person\_Home\_Ownership ranks second, contributing 17.47% to the model's predictive power.

This suggests that the type of home ownership (e.g., rent, mortgage, own) plays a substantial role in influencing loan decisions, possibly due to its correlation with financial stability and collateral availability.

Loan Intent:

Loan\_Intent ranks third with 5.30% importance.

The purpose of the loan (e.g., education, medical, debt consolidation) moderately impacts approval outcomes, reflecting lender preferences for specific loan intents that align with lower risk profiles.

Negligible Influence of Other Features:

All remaining features, including Credit\_Score, Loan\_Amount, and Person\_Income, contribute less than 1% each to the model's predictive power.

Surprisingly, Credit\_Score, often considered a critical metric for loan decisions, has almost no impact (0.00045%). This suggests that the dataset or lending policies prioritize other factors over credit scores.

2. Implications

Overwhelming Reliance on Defaults:

Lenders heavily rely on the history of loan defaults, which might overshadow other potentially significant indicators like income or credit score.

This reliance could lead to challenges for applicants with a past default but an otherwise strong financial profile.

Role of Home Ownership:

The prominence of home ownership underscores its importance as a proxy for financial stability or collateral. Applicants with favorable ownership statuses (e.g., owning a home) are likely viewed as less risky.

Underuse of Traditional Metrics:

Features like Credit\_Score and Person\_Income, which are typically considered cornerstones of credit risk assessment, appear underweighted. This may indicate an unconventional or limited dataset or an alternative decision-making approach by lenders.