# Exploratory Correlation Analysis of Loan Application Attributes

1. Introduction:

The objective of this analysis is to gain insights into the loan application data. By examining correlations, feature importance, and key variable distributions, we aim to understand the factors influencing loan approval decisions.

2. Variable Correlation:

Correlation Heatmap of Selected Features
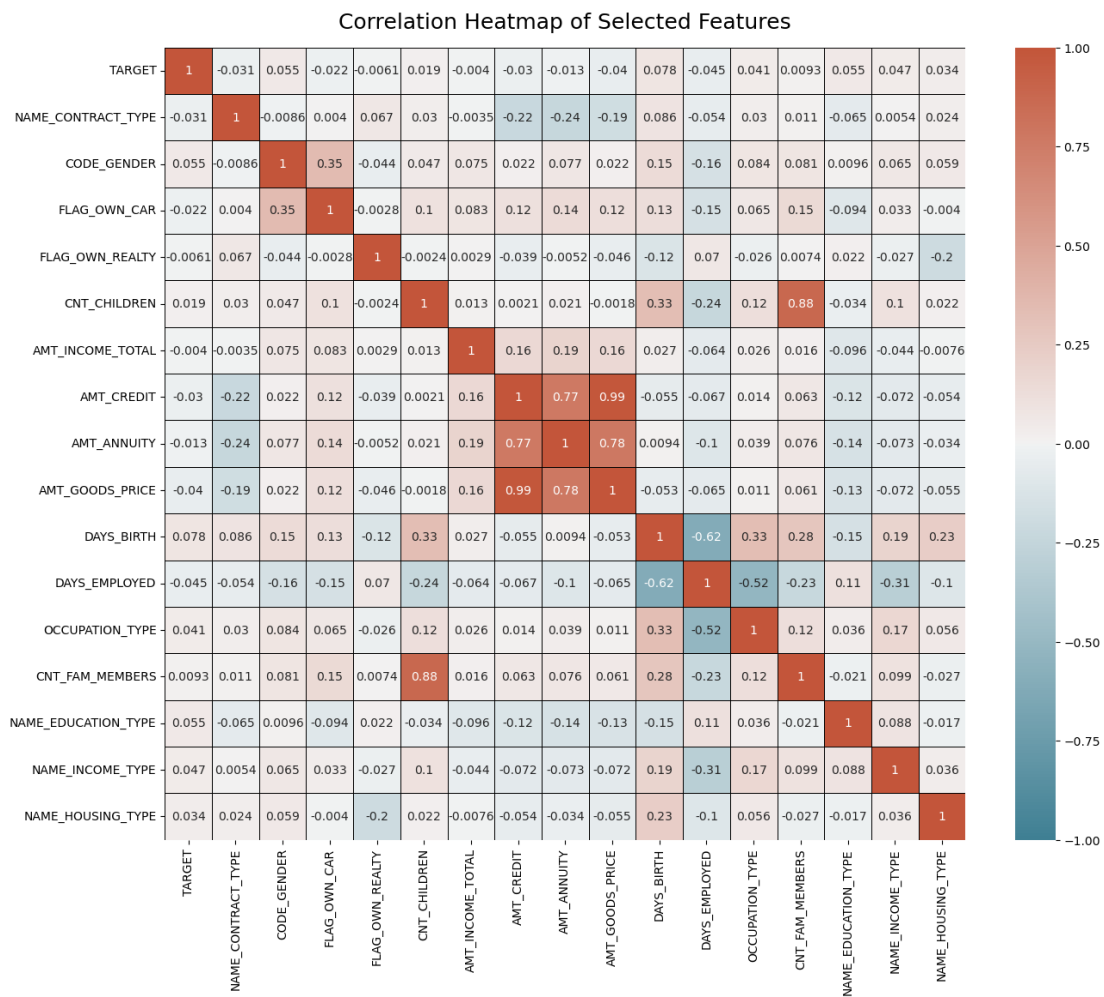
Legend:

Heatmap Colors: The heatmap uses a color spectrum that transitions from bright aqua (indicating negative correlation) to white (indicating little to no correlation) to vibrant magenta (indicating positive correlation).

Annotations: Each cell contains a numerical value representing the correlation coefficient between the respective row and column variables. A value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and values close to 0 indicate weak correlations.

Borders: The black lines separating cells help in visually distinguishing between different correlation values..

Findings:

1.Prominent Positive Correlations:

AMT_CREDIT and **AMT_ANNUITY** showcase a strong positive correlation, indicating that higher loan amounts typically come with higher annuity payments.

AMT_CREDIT and **AMT_GOODS_PRICE** are positively correlated, suggesting that as the price of goods for which the loan is applied increases, the credit amount also tends to increase.

2.Notable Negative Correlations:

DAYS_BIRTH (Age) and **DAYS_EMPLOYED** (Employment Duration) are negatively correlated, implying that as applicants age, their current employment duration tends to decrease, possibly due to factors like career transitions or retirement.

3. Feature Importance:

In descending order of importance for predicting credit default, the variables are:
DAYS_BIRTH, DAYS_EMPLOYED, AMT_ANNUITY, AMT_CREDIT, AMT_INCOME_TOTAL, AMT_GOODS_PRICE, OCCUPATION_TYPE.

4. Grouping and Analysis: Detailed Insight
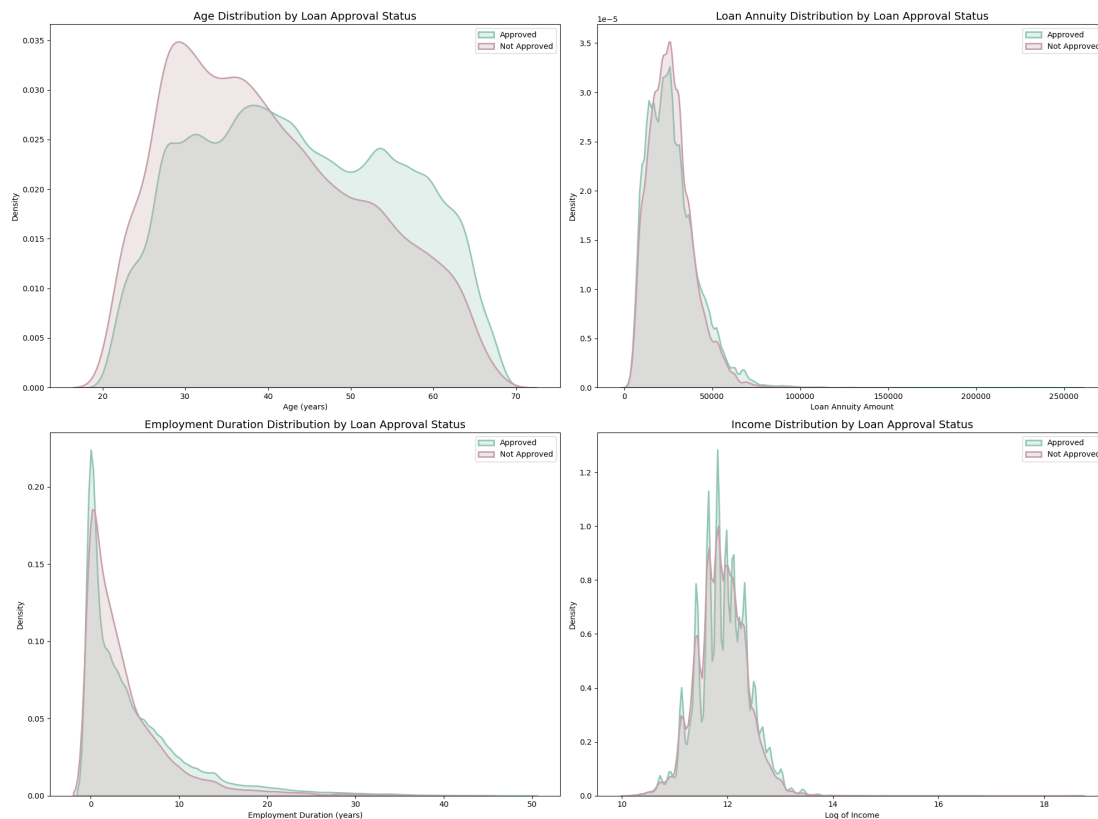
Legend:

Colors: Green for approved loans, Pink for non-approved loans.

Subplot 1: Age Distribution by Loan Approval Status.
Prime Age for Approval: Most approved loans are clustered among applicants aged 30 to 60, with a peak in approvals around the early 40s.
Challenges for Younger Applicants: A noticeable spike in disapprovals is seen among applicants in their late 20s to early 30s, possibly reflecting concerns over financial stability or lack of established credit history.
Complex Age Dynamics: The overlap in age distribution between 30 to 50 years suggests a nuanced approach to loan approvals, where age is a factor but not the sole determinant.

Subplot 2: Loan Annuity Distribution by Loan Approval Status.
Preference for Higher Annuities: A clear trend of higher loan approval rates is observed for applicants with larger annuity amounts, hinting at a preference for financially robust applicants.
Mixed Approvals in Mid-Range: The overlapping densities in mid-range annuities indicate a more complex decision-making process, where other factors besides annuity size come into play.

Subplot 3: Employment Duration Distribution by Loan Approval Status.
Stable Employment Favored: Longer durations of employment correlate with higher approval rates, emphasizing the value placed on employment stability.
Shorter Employment, Increased Disapprovals: Applicants with short or zero employment durations (indicating current unemployment) are more frequently disapproved, underlining the importance of consistent employment history in loan assessments.

Subplot 4: Income Distribution by Loan Approval Status.
High Income, High Approval Rate: There's an evident trend of higher approval rates for applicants with higher incomes, reinforcing the idea that income level is a critical factor in loan approvals.
Variability at Lower Incomes: Among applicants with lower incomes, the approval rates vary widely, suggesting that for these individuals, factors other than income are pivotal in the loan decision process.

5. Conclusion:
Through this analysis, we've unearthed significant insights into the dynamics of loan approvals. Age, employment duration, and credit metrics play pivotal roles in the decision-making process. However, it's evident that loan decisions are multifaceted, relying on a combination of various applicant attributes. Further analyses can delve deeper into individual feature interactions, providing even more nuanced insights.

Data gathered from:
https://www.kaggle.com/datasets/mishra5001/credit-card?select=previous_application.csv
GitHub Link:
https://github.com/Sophiaxiaoyu/Correlation-Analysis-of-Loan-Application-Attributes