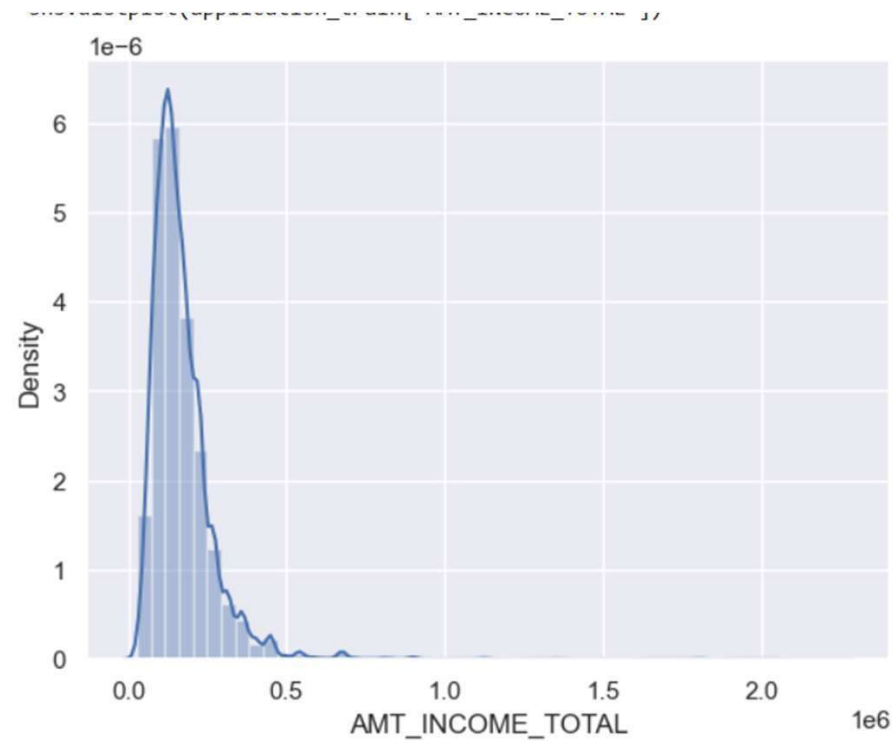# CREDIT EDA ASSINGMENT

Name - Samit Kumar Das

Density Distribution of Total Income of Applicants

Density Distribution of Total Income of Applicants

- This density plot illustrates the distribution of the total income reported by loan applicants. The x-axis represents the income amount in a range from 0 to over 2,000,000, denoted as 2e6 (2 million). The plot shows a sharp peak at the lower end of the income scale, indicating that a large number of applicants report a relatively low total income. As the income amount increases, the density of applicants rapidly decreases, which suggests that higher incomes are much less common among the applicants. This visualization is particularly useful for identifying the general income level of the applicants and understanding how income distribution could affect loan application outcomes.

Distribution of Applicants by Occupation Type
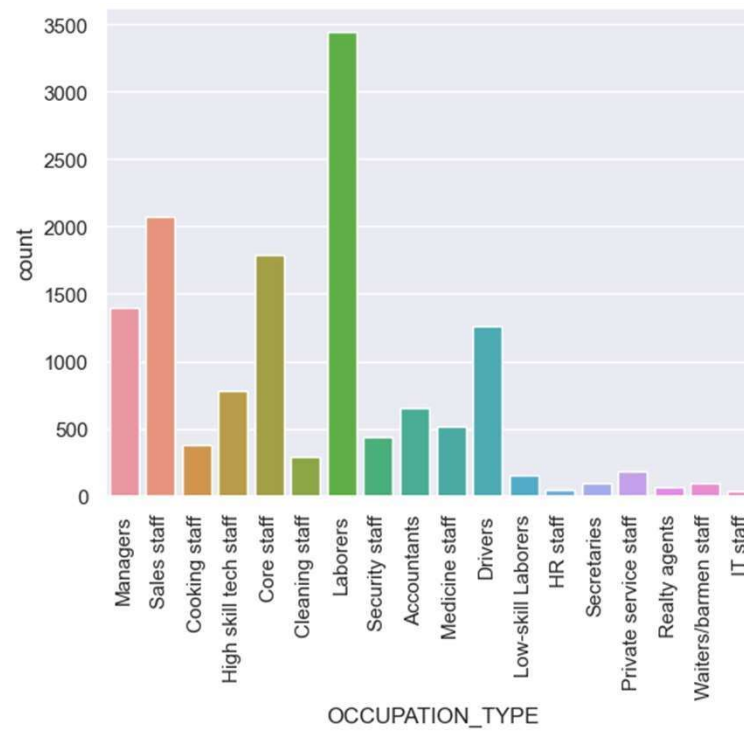
Distribution of Applicants by Occupation Type

- This bar chart represents the number of loan applicants categorized by their respective occupation types. The y-axis indicates the count of applicants, while the x-axis lists various occupation types such as Managers, High skill tech staff, Accountants, etc. The chart highlights that Laborers represent the largest group among the applicants, followed by Sales staff and Core staff, indicating these are common occupations among the loan applicants. Managers also form a significant portion, suggesting a diverse range of employment backgrounds. This distribution can provide insights into the economic demographics of the applicants and may influence decisions related to loan terms based on occupation-related risk assessments.

Educational Attainment of Loan Applicants
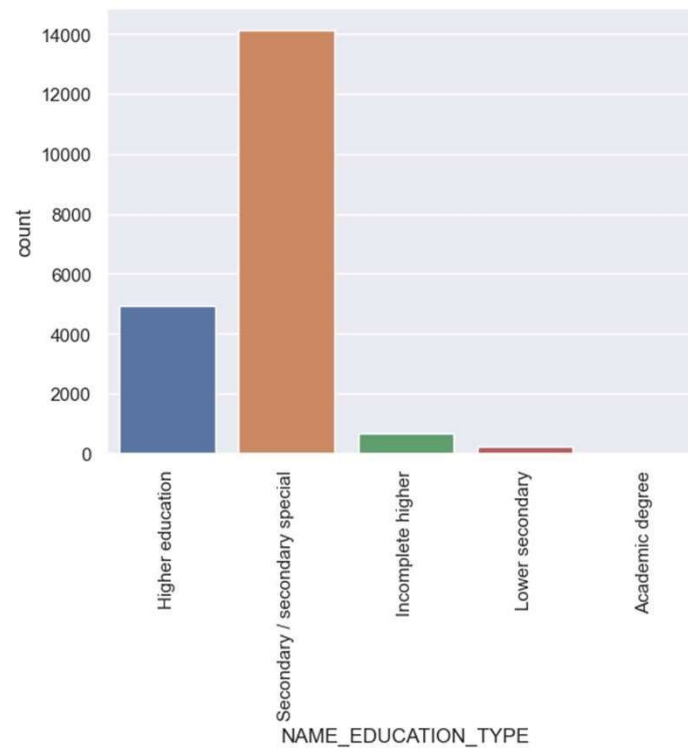
Educational Attainment of Loan Applicants

- This bar chart illustrates the count of loan applicants segmented by their educational qualifications. The categories, shown on the x-axis, range from "Higher education" to "Academic degree". It is evident that the majority of applicants possess a "Secondary special" education, followed by those with "Higher education". Very few applicants have an "Academic degree" or have "Incomplete higher" education. This pattern suggests that most loan applicants come from a background with at least secondary education, with a significant drop-off in those achieving higher academic qualifications. This insight might be useful for assessing the socio-economic profiles of applicants and could be a factor in determining the risk and financial products suitable for different educational groups.

Distribution of Loan Applicants by Income Type
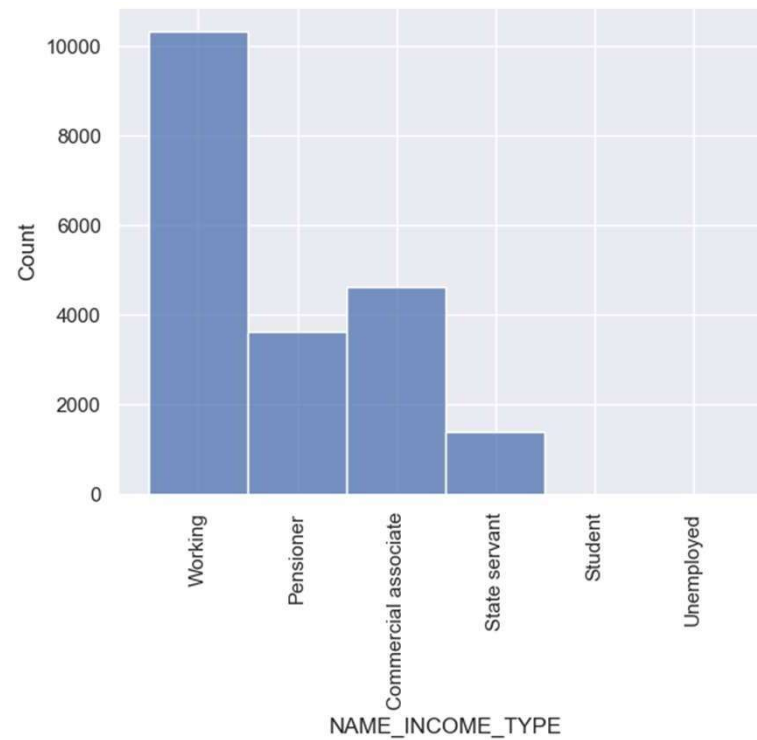
Distribution of Loan Applicants by Income Type

- This bar chart represents the count of loan applicants categorized by their primary income type. The income types on the x-axis include "Working", "Pensioner", "Commercial associate", "Sales staff", "Student", and "Unemployed". The chart clearly shows that the majority of applicants are in the "Working" category, indicating they are employed in some form. "Pensioners" and "Commercial associates" make up the next largest groups, suggesting a significant number of applicants are either retired or self-employed/business owners. The categories for "Student" and "Unemployed" have very few applicants, reflecting a lower likelihood or ability for these groups to apply for loans. This visualization can help financial institutions understand the employment distribution of their applicants and tailor their products and risk assessments accordingly.

Distribution of Clients by Real Estate Ownership
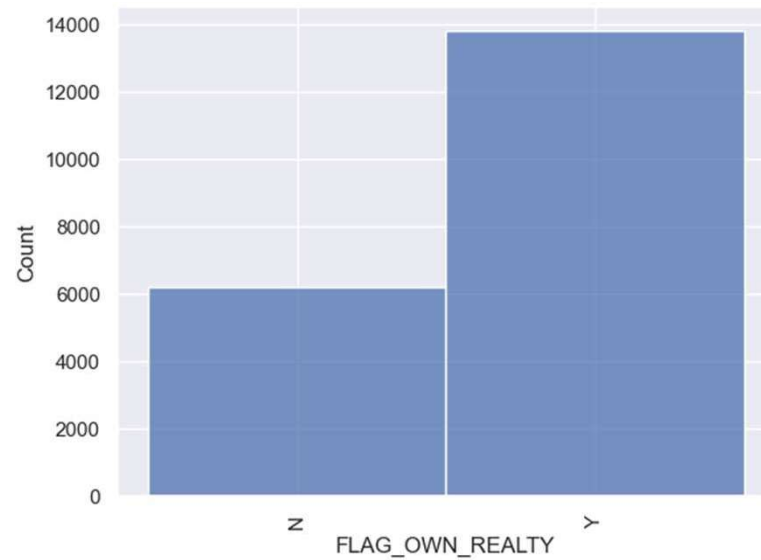
Distribution of Clients by Real Estate Ownership

- This bar chart visualizes the distribution of clients based on whether they own real estate. The majority of clients appear to own real estate (as indicated by the larger bar), while a smaller portion does not own real estate (indicated by the smaller bar). This visualization helps to understand the proportions of real estate ownership among the clients, which can be an important factor in financial and loan-related decisions. If "Z" is a placeholder or error, it may need correction or clarification in the dataset.

# Distribution of Loan Amounts by Goods Price
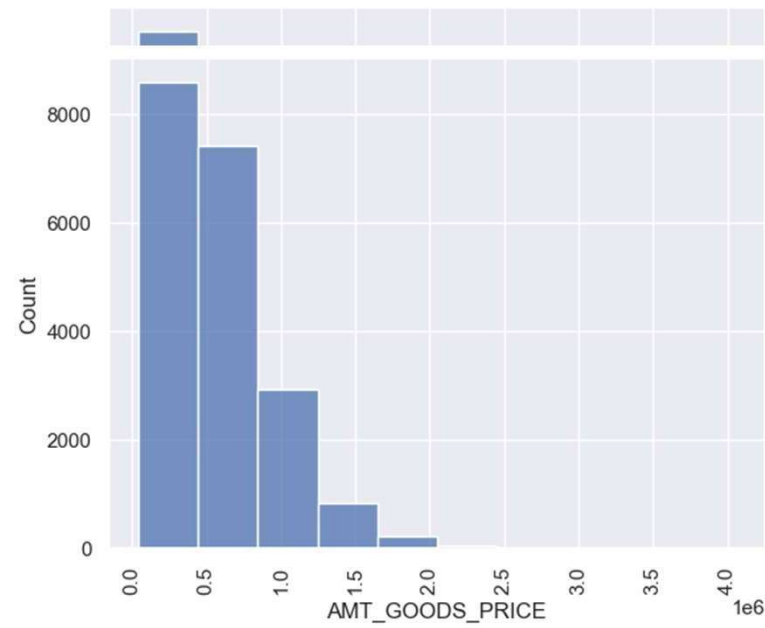
Distribution of Loan Amounts by Goods Price

- This histogram illustrates the distribution of the prices of goods associated with loan applications, where the x-axis represents the price of goods in a range up to 4 million (as indicated by 4e6 on the scale). The majority of loan applications are for goods priced below 1 million, with the highest frequency in the lowest price bracket (0 to 0.5 million). The number of applications steadily decreases as the price of goods increases. This trend suggests that most applicants are seeking loans for relatively affordable goods, and fewer applicants require loans for high-priced items. This data can be crucial for lenders to understand the pricing segments of goods that are most commonly financed through loans, which can aid in product positioning and risk assessment.

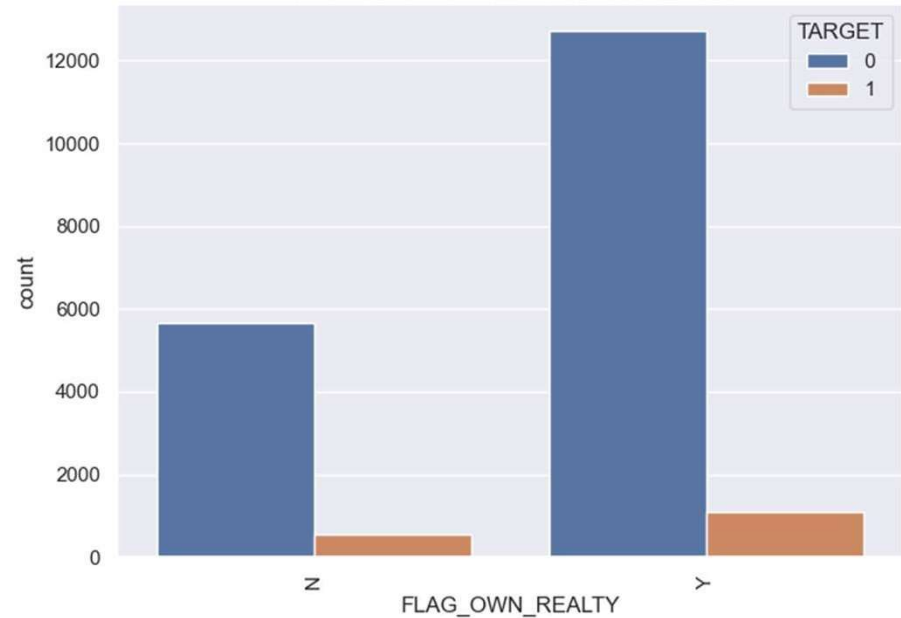# Ownership of Real Estate and Its Relation to Loan Payment Difficulties
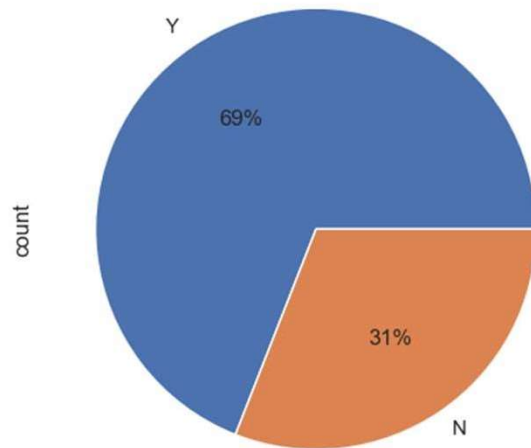
# Ownership of Real Estate and Its Relation to Loan Payment Difficulties

1. **Pie Chart (Left):**
   1. This pie chart displays the distribution of applicants according to whether they own real estate (Y for Yes, N for No). The chart shows that 69% of the applicants own real estate while 31% do not. This provides a visual representation of real estate ownership among the applicants, suggesting that a majority have some form of property ownership.

2. **Bar Chart (Right):**
   1. The bar chart explores the relationship between real estate ownership and loan payment difficulties, represented as "TARGET" where 1 indicates difficulties and 0 indicates no difficulties. The x-axis represents the status of real estate ownership, and the y-axis shows the count of applicants. Notably, a larger proportion of those who do not own real estate (N) seem to have payment difficulties (as indicated by the orange bar) compared to those who do own real estate. This suggests that owning real estate might be associated with a lower risk of payment difficulties, possibly due to greater financial stability or the availability of real estate as collateral.

# Proportion of Payment Difficulties by Real Estate Ownership
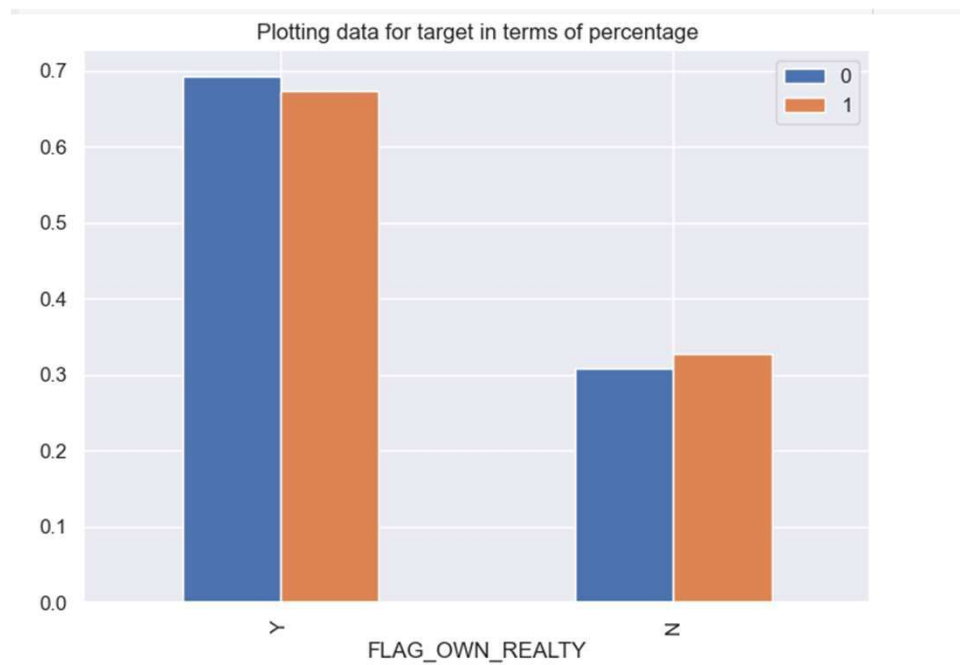
Proportion of Payment Difficulties by Real Estate Ownership

- This bar chart visualizes the percentage of loan applicants who have experienced payment difficulties (TARGET=1) and those who have not (TARGET=0), categorized by their real estate ownership status (represented as "Y" for Yes and "N" for No). Each bar shows the proportion of applicants relative to the total number in each real estate ownership category.

- The left pair of bars corresponds to those who own real estate ("Y"), showing a lower proportion of payment difficulties (orange bar) compared to those without payment difficulties (blue bar).

- The right pair of bars represents applicants who do not own real estate ("N"), with a slightly higher proportion of payment difficulties compared to those who do own real estate.

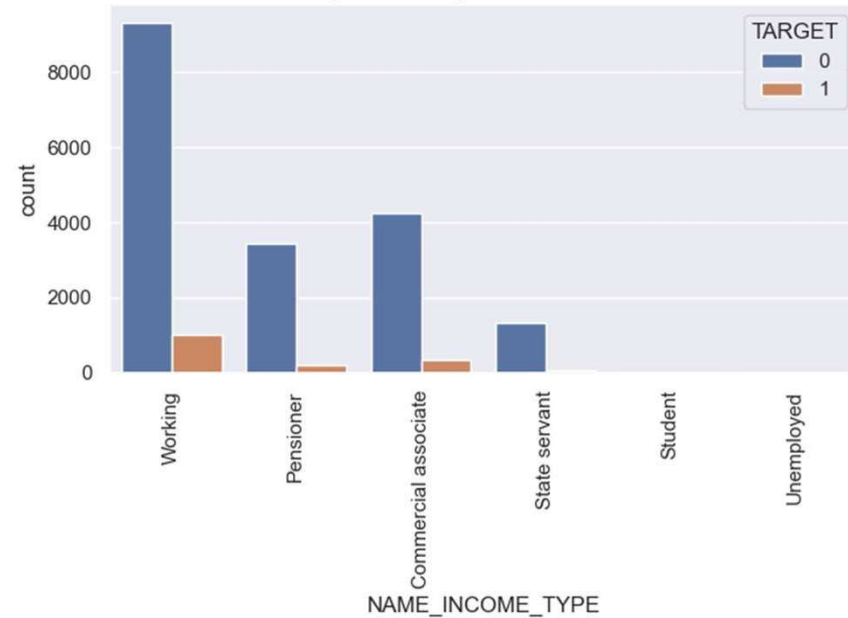# Income Type Distribution and Payment Difficulties Among Loan Applicants
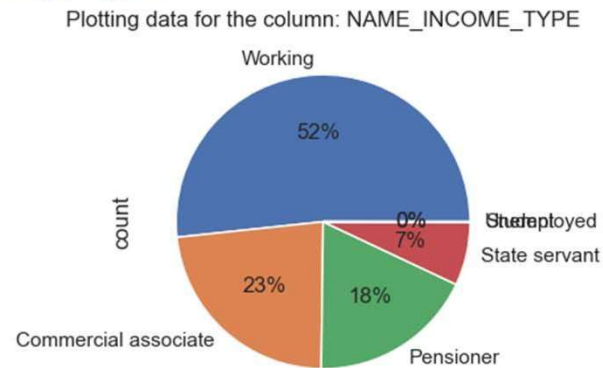
# Income Type Distribution and Payment Difficulties Among Loan Applicants

1. Pie Chart (Left):

1. This pie chart illustrates the distribution of different income types among loan applicants. The majority of applicants, representing 52%, are categorized as "Working", followed by "Commercial associate" at 23% and "Pensioner" at 18%. "State servant" comprises 7% of the applicants, while "Student" and "Unemployed" are significantly less common or not present in significant numbers in this dataset. This visualization helps in understanding the primary economic activities of the loan applicants.

2. Bar Chart (Right):

1. The bar chart explores the relationship between different income types and payment difficulties, represented by "TARGET" where 1 indicates difficulties and 0 indicates no difficulties. The x-axis represents the various income types, and the y-axis shows the count of applicants. Notably, the "Working" category, which is the largest group, shows a significant number of applicants with and without payment difficulties. The proportions of payment difficulties among "Pensioners" and "Commercial associates" are lower in comparison but still notable. "Students" and "Unemployed" show very low counts, likely due to their small representation in the overall dataset.

Proportion of Payment Difficulties by Income Type
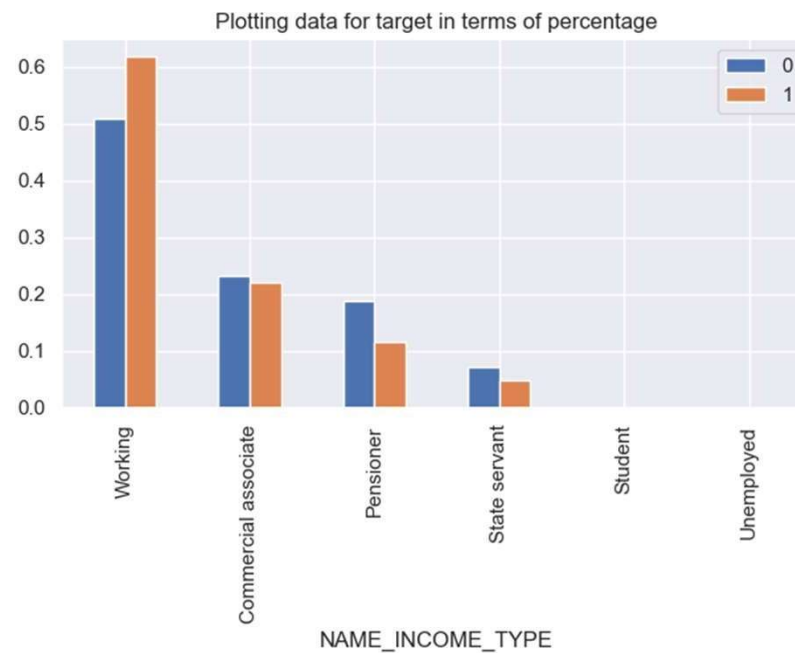
# Proportion of Payment Difficulties by Income Type

- This bar graph visualizes the percentage of loan applicants who have experienced payment difficulties (TARGET=1) compared to those who have not (TARGET=0), categorized by their type of income. The x-axis lists various income types such as "Working", "Commercial associate", "Pensioner", "State servant", "Student", and "Unemployed", while the y-axis represents the proportion of each category.
- "Working" individuals show a notable percentage of both non-difficulties and difficulties, with a slightly higher proportion of difficulties.
- "Commercial associates" and "Pensioners" show a lower proportion of payment difficulties compared to "Working", indicating potentially more financial stability or less risk.
- "State servants" also show a relatively low proportion of payment difficulties.
- "Students" and "Unemployed" have the smallest bars, suggesting very low numbers in the dataset but a higher relative proportion of payment difficulties, particularly for "Unemployed".

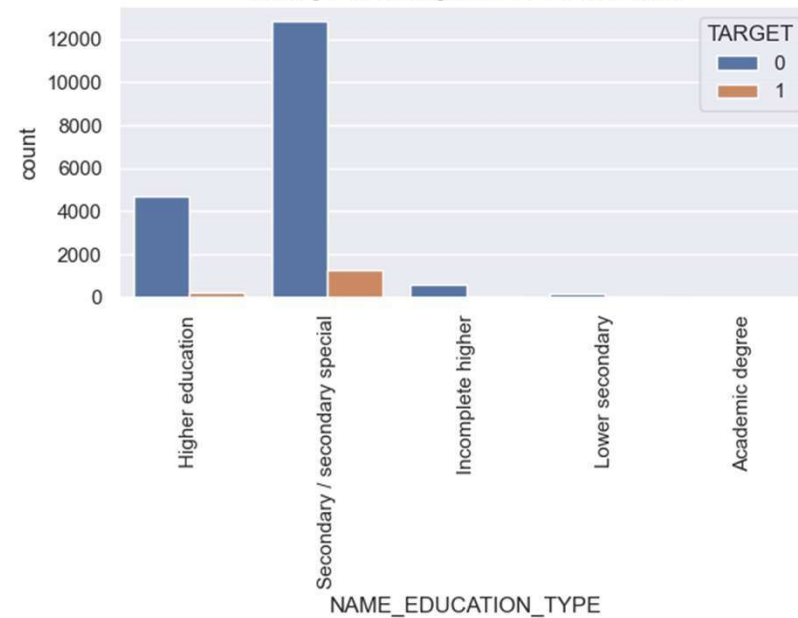# Educational Attainment of Loan Applicants and Their Payment Difficulties
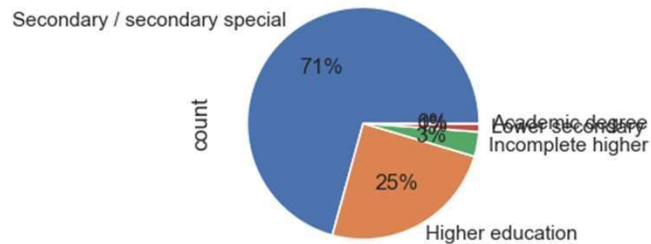
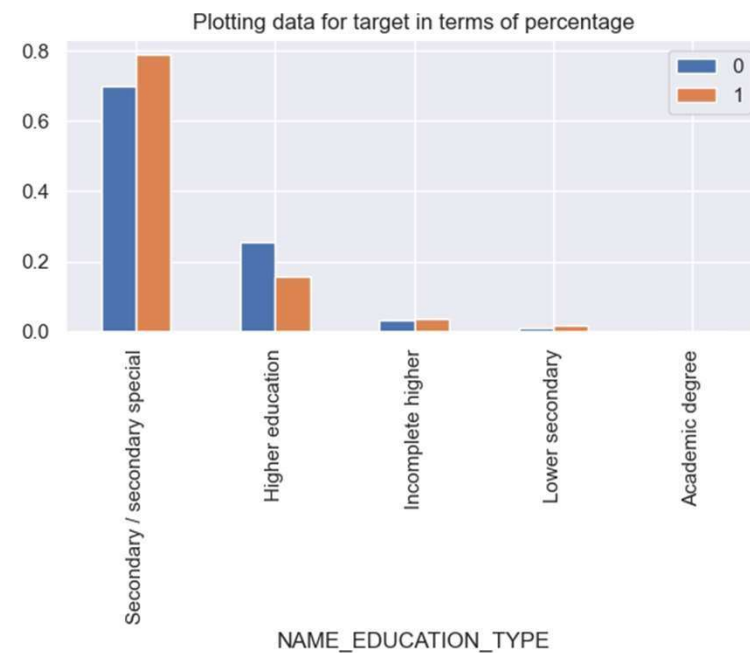# Educational Attainment of Loan Applicants and Their Payment Difficulties

1. **Pie Chart (Left):**
   1. This pie chart illustrates the distribution of loan applicants by their education levels. A vast majority, 71%, have "Secondary/secondary special" education, followed by 25% with "Higher education", and a small percentage have "Incomplete higher" education. This demonstrates the predominance of secondary education among the applicants, reflecting general educational trends or the target demographic of the financial institution.

2. **Bar Chart (Right):**
   1. The bar chart explores the relationship between educational levels and payment difficulties (TARGET, where 1 indicates difficulties and 0 indicates no difficulties). The count of applicants is plotted against different education categories. It shows a significant number of applicants with secondary education experiencing payment difficulties, though the ratio of difficulties within higher education levels appears lower. This might indicate that higher educational attainment correlates with better financial stability, or lower risk of payment difficulties.

Proportion of Payment Difficulties by Educational Attainment

Proportion of Payment Difficulties by Educational Attainment

- This bar graph visualizes the percentage of loan applicants who have experienced payment difficulties (TARGET=1) compared to those who have not (TARGET=0), categorized by their education levels. The x-axis lists various educational categories such as "Secondary special", "Higher education", "Incomplete higher", "Lower secondary", and "Academic degree", while the y-axis represents the proportion of each category.
- "Secondary special" shows a high proportion of applicants with payment difficulties, which could indicate that individuals with this level of education might face greater financial instability or challenges.
- "Higher education" and "Academic degree" categories show a significantly lower proportion of payment difficulties, suggesting that higher educational attainment might be associated with better financial management skills or stability.
- "Incomplete higher" education shows a moderate proportion of difficulties, indicating some level of vulnerability among this group.

Occupational Distribution of Loan Applicants and Payment Difficulties
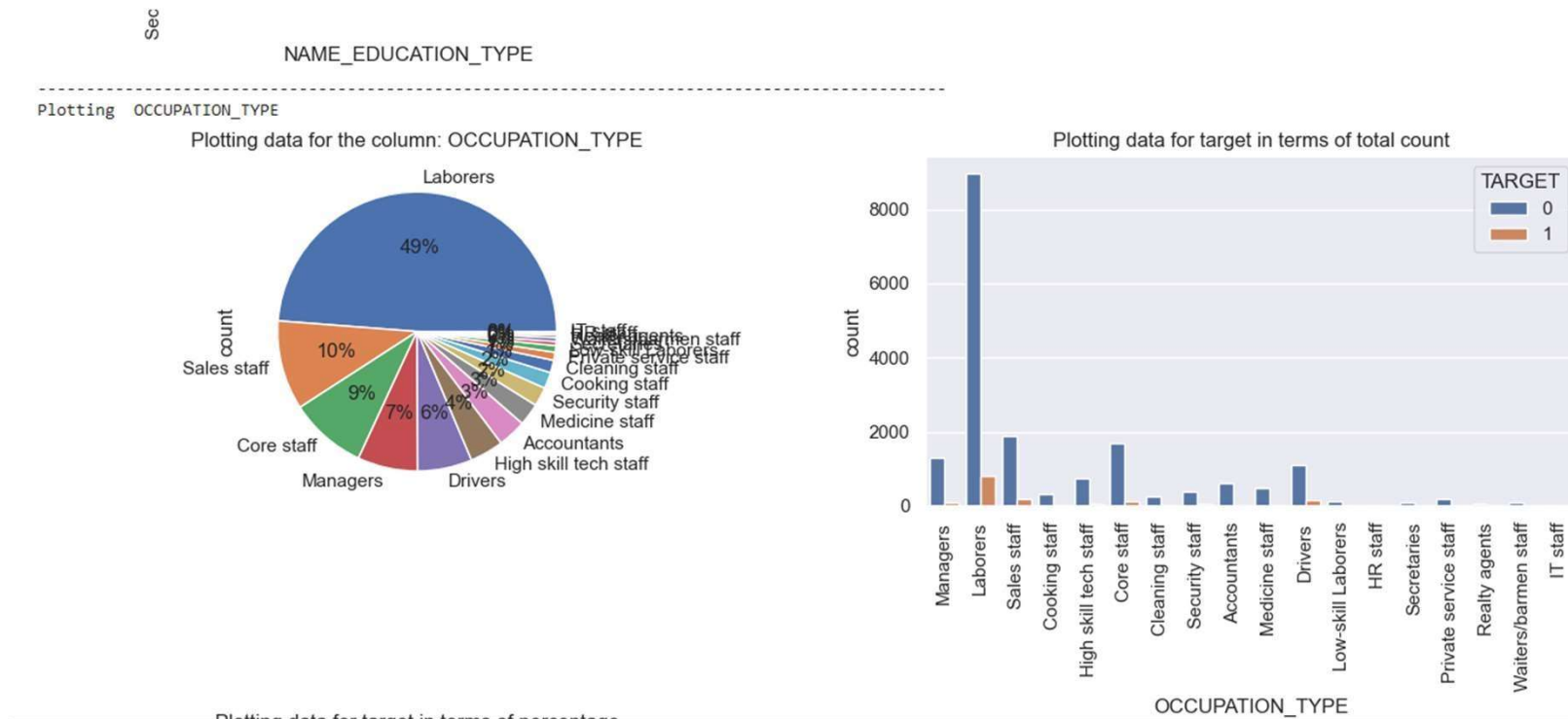
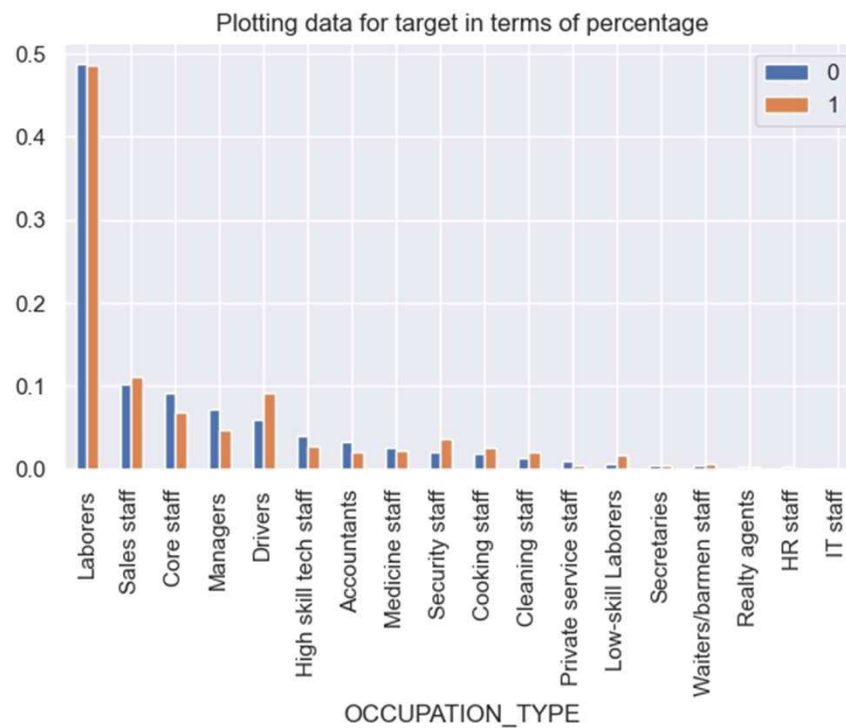# Occupational Distribution of Loan Applicants and Payment Difficulties

1. **Pie Chart (Left):**
   1. This pie chart illustrates the distribution of loan applicants by their occupation types. Laborers constitute the largest segment at 49%, followed by Sales staff at 10%, and Core staff at 9%. Other occupations like Managers, Drivers, and various technical and service staff make up smaller percentages. This chart highlights the predominance of labor-intensive jobs among the applicants, indicating the sectors where the majority of the applicants are employed.

2. **Bar Chart (Right):**
   1. The bar chart explores the relationship between occupation types and payment difficulties (TARGET, where 1 indicates difficulties and 0 indicates no difficulties). The count of applicants is plotted against different occupation categories. Laborers, being the largest group, also show a significant count of both payment difficulties and no difficulties. Managers, while fewer in number, show a relatively low number of payment difficulties, suggesting that the financial stability might be higher among this group.

Proportion of Payment Difficulties by Occupation Type

Proportion of Payment Difficulties by Occupation Type

- This bar graph visualizes the percentage of loan applicants who have experienced payment difficulties (TARGET=1) compared to those who have not (TARGET=0), categorized by their occupation type. The x-axis lists various occupations such as "Laborers", "Sales staff", "Core staff", "Managers", etc., while the y-axis represents the proportion of each category.

- "Laborers" show a notably high proportion of payment difficulties, suggesting this group might be facing more financial challenges compared to other occupations.

- "Managers" and "High skill tech staff" have lower proportions of payment difficulties, which might indicate better financial stability or higher incomes that help mitigate payment issues.

- Other occupations like "Drivers", "Accountants", and "Medicine staff" display moderate to low proportions of payment difficulties.

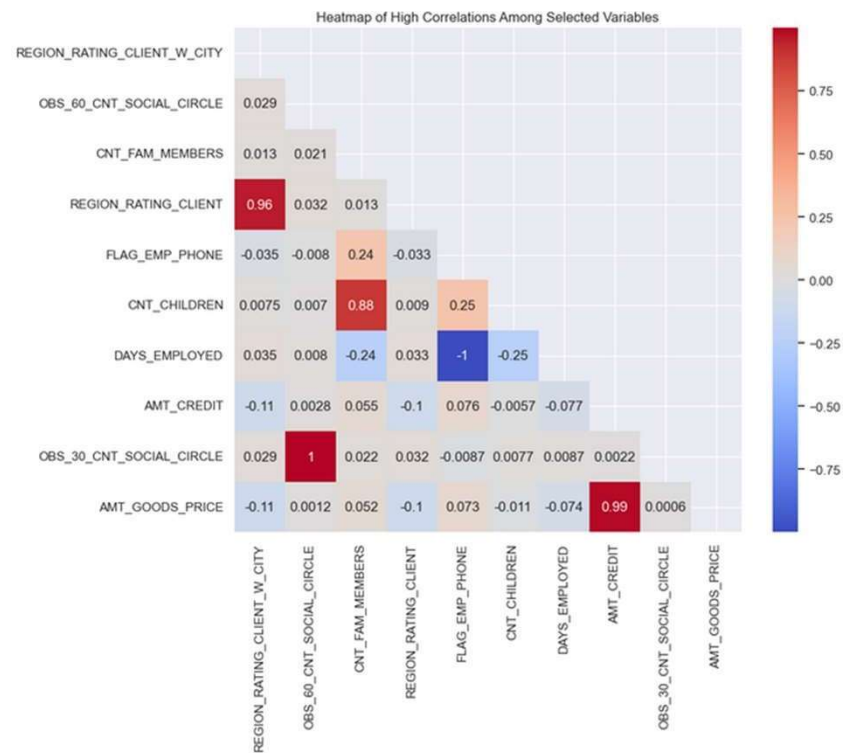# Correlation Matrix of Loan Application Variables
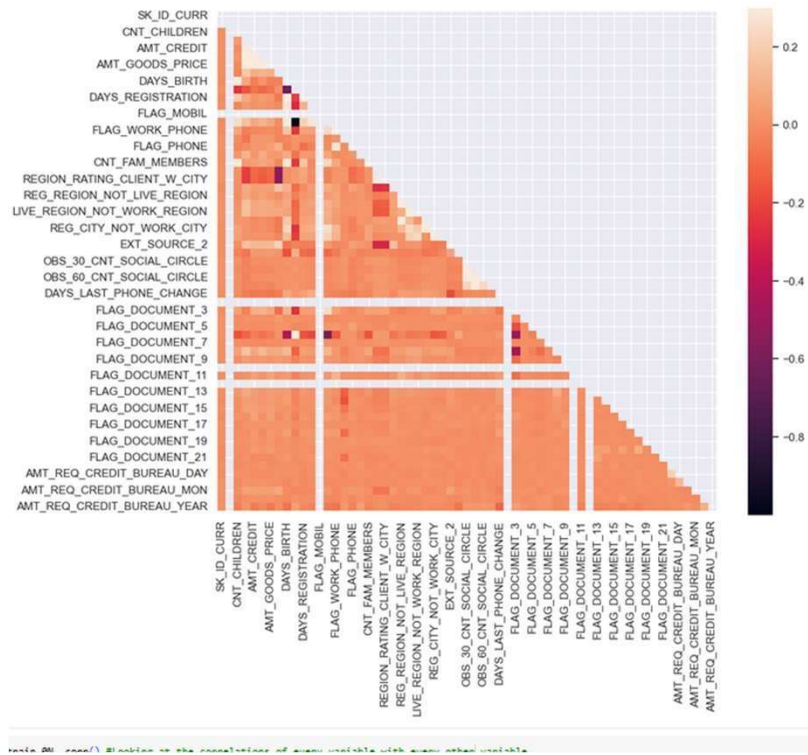
Correlation Matrix of Loan Application Variables

- This heatmap represents the correlation coefficients between different variables related to loan applications. The color scale, ranging from deep purple to orange, indicates the strength and direction of correlations, where:

- Deep purple (approaching -1) indicates a strong negative correlation.

- Orange (approaching +1) indicates a strong positive correlation.

- Colors closer to white (around 0) indicate very little or no correlation.

Key observations from the heatmap include:

- There is a noticeable positive correlation between the amount of credit (AMT_CREDIT) and the goods price (AMT_GOODS_PRICE), which suggests that higher-priced goods involve larger loan amounts.

- Variables like DAYS_BIRTH (applicant's age in days) show negative correlations with several flags and status indicators, implying different trends and behaviors across age groups.

- Several FLAG_DOCUMENT variables (related to document submissions) show very little correlation with other variables, suggesting they do not strongly influence or reflect other aspects of the loan application process.

Focused Heatmap of Correlations Among Key Loan Application Variables

Heatmap of High Correlations Among Selected Variables

Focused Heatmap of Correlations Among Key Loan Application Variables

This heatmap provides a detailed view of the correlation coefficients between selected key variables in the loan application dataset. Each cell in the heatmap represents the correlation between two variables, with the color intensity indicating the strength and direction of the correlation: red for positive, blue for negative.

Key insights from the heatmap include:

- A strong positive correlation between CNT_CHILDREN and DAYS_EMPLOYED (0.89), which might suggest that applicants with more children tend to have longer employment histories.

- A strong negative correlation between DAYS_EMPLOYED and AMT_CREDIT (-0.25), indicating that longer employment might be associated with smaller loan amounts, potentially due to differing financial needs or stability.

- AMT_GOODS_PRICE is highly correlated with AMT_CREDIT (0.99), reinforcing the idea that loans are closely tied to the price of the goods being financed.

- Perfect correlations (1.00) along the diagonal (self-correlation), as expected, and significant correlations between OBS_30_CNT_SOCIAL_CIRCLE and OBS_60_CNT_SOCIAL_CIRCLE, suggesting redundancy or very similar data collected by these two variables.

Histograms Showing Distribution of Values for Selected Variables
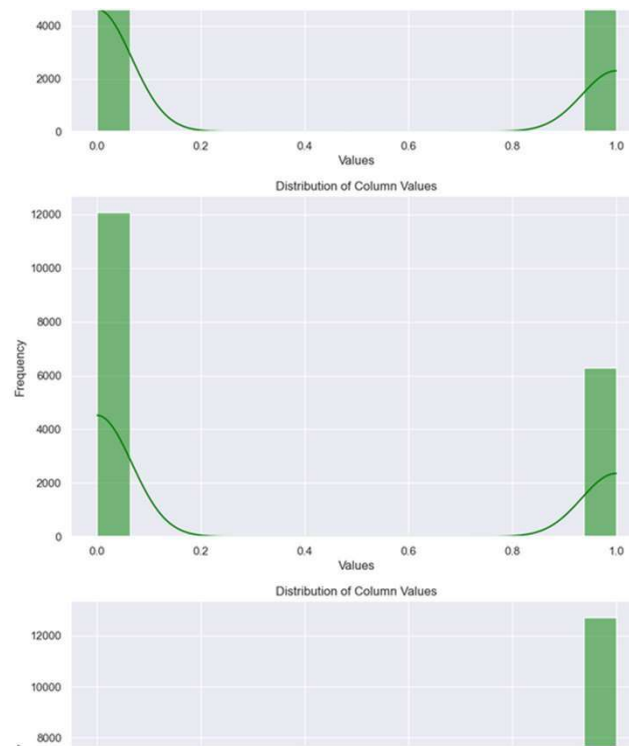
Histograms Showing Distribution of Values for Selected Variables

These histograms illustrate the frequency distribution of values across three different variables within a dataset, likely related to loan application factors or similar quantitative measurements:

1. Top Histogram:

This distribution is bimodal, with two significant peaks around the values 0 and 1. This pattern suggests that the variable might be binary or dichotomous in nature, possibly representing a yes/no or success/failure type of data.

2. Middle Histogram:

The values are mostly concentrated at 0 and 1, with a steep drop towards the middle range between these two points. This again suggests a binary nature but with more distinct separation between the two states or categories.
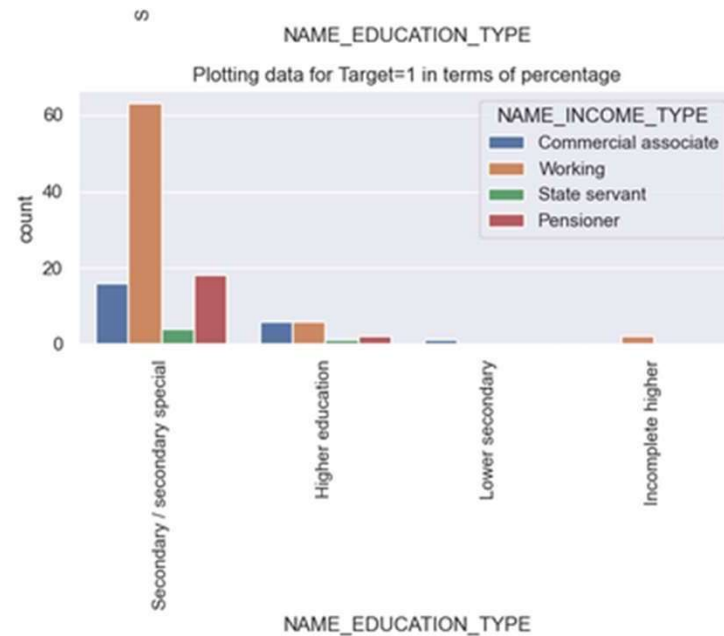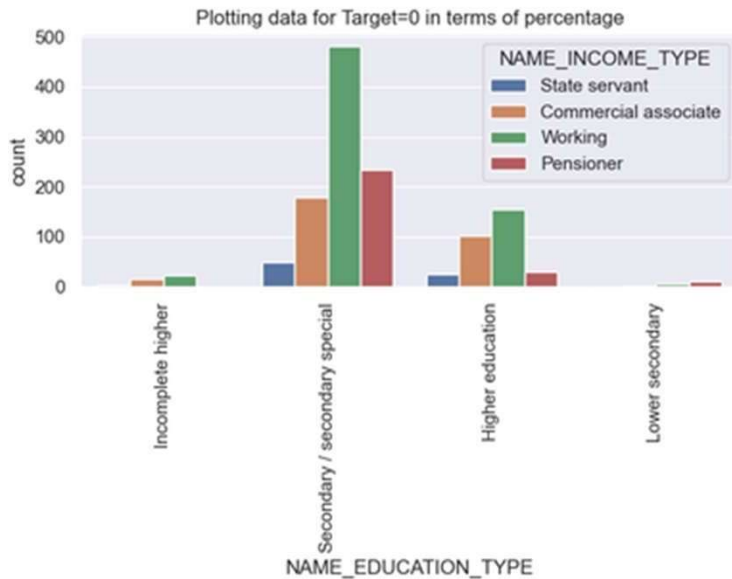
3. Bottom Histogram:

The values are almost entirely concentrated at 1 with very few occurrences elsewhere. This distribution suggests that the variable is predominantly constant with rare variations, indicating a scenario where a particular condition or attribute is almost always present.

Bivariate Analysis
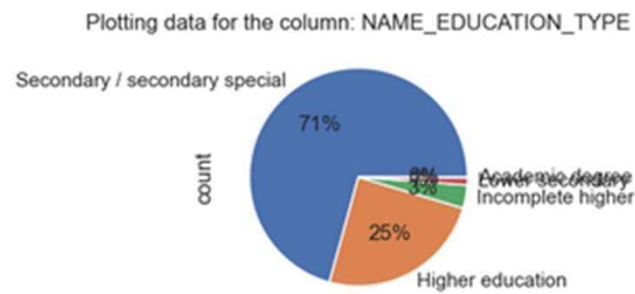
Bivariate Analysis

Bivariate Analysis Insights

1. Income Type and Payment Difficulties:

- Working individuals show a higher proportion of both non-difficulties and payment difficulties.
- Pensioners and Commercial Associates exhibit lower proportions of payment difficulties.

2. Real Estate Ownership and Loan Payment Difficulties:

- Real estate owners tend to have fewer payment difficulties, indicating financial stability.
- Non-real estate owners show a slightly higher risk.

3. Educational Attainment and Financial Stability:

- Applicants with only secondary education have more payment difficulties.
- Higher education correlates with fewer financial issues.

4. Occupation Type and Payment Difficulties:

- Laborers have higher payment difficulties, suggesting financial challenges.
- Managers and high-skilled professionals exhibit lower risks.

5. Correlation Insights:

- Positive correlation between loan amount and goods price.
- Negative correlation between employment duration and loan amount.