Comprehensive Loan Approval Analysis Report

A Deep Dive into Financial Inclusion

Under the guidance of:

Dr. Varun Aggarwala

Group Members:

Bhawana Thawarani

Neha Roy Choudhury

Piyush Borse

Prajwal Wagh

Prateek Majumder

Yuvraj Singh Srinet

The Problem

Banks meticulously evaluate many factors when reviewing loan applications to determine eligibility. These factors encompass many financial and personal considerations, including credit history, income stability, employment status, debt-to-income ratio, and the loan's purpose. By scrutinizing these elements, banks aim to assess the applicant's financial health, repayment capacity, and overall creditworthiness. Our analysis delves into these factors to identify applicants with a higher likelihood of securing loan approval. Understanding the intricate interplay of these variables enables us to provide valuable insights and guidance to individuals seeking financial assistance.

Why is it important to address?

The importance of understanding the factors influencing loan approval lies in its profound impact on individuals' financial well-being and broader economic stability. For individuals, securing a loan can facilitate significant life goals such as purchasing a home, starting a business, or funding education. Conversely, being denied a loan can hinder these aspirations and lead to financial strain.

Moreover, from a macroeconomic perspective, access to credit fuels economic growth by stimulating consumption and investment. When banks make informed decisions about loan approvals, they allocate resources efficiently, mitigating risks of defaults and financial instability. Conversely, indiscriminate lending can lead to credit bubbles and financial crises.

By comprehensively analyzing the factors affecting loan approval, individuals can better position themselves to meet lenders' criteria and improve their chances of securing financing. Additionally, banks can enhance their risk management practices, fostering a healthier lending ecosystem that supports both individual financial aspirations and broader economic growth.

The insights obtained from examining loan applicant data can help financial organizations make strategic decisions. Understanding consumer preferences, risk profiles, and market trends allows institutions to create more tailored marketing campaigns, product offerings, and pricing structures.

How are we addressing this Problem

Utilizing a dataset of loan applicants, we aim to glean profound insights into the determinants of loan approval by meticulously scrutinizing a myriad of applicant attributes.

Key techniques employed in this analysis include:

Descriptive Statistics: Employed to summarize and describe the essential features of the dataset, providing a snapshot of the loan applicant population's characteristics.

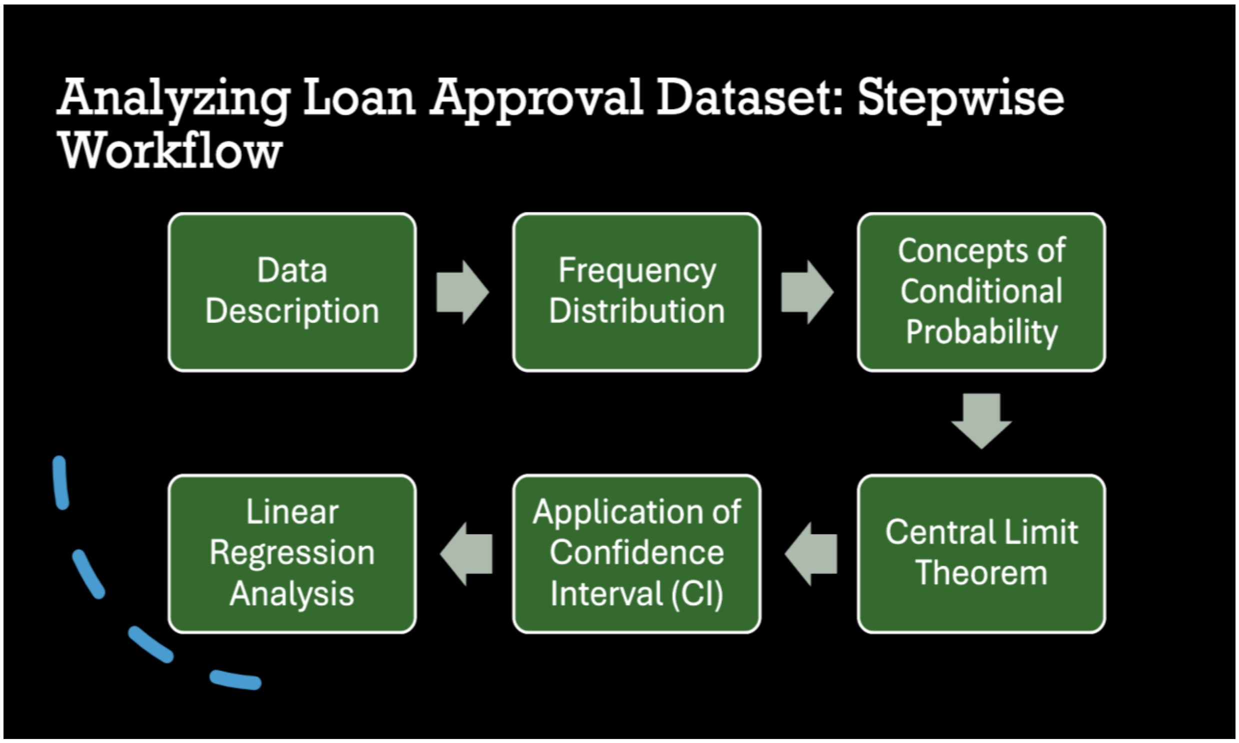
Data Distributions: Investigated to understand the spread and variability of different variables within the dataset, shedding light on the underlying patterns and trends.

Statistical Analysis: Conducted to discern relationships and associations between various applicant characteristics and loan acceptance outcomes, facilitating a deeper understanding of the factors influencing loan approval.

Conditional Probability: Utilized to assess the likelihood of loan approval given specific applicant traits or conditions, offering valuable insights into the conditional dependencies within the dataset.

Regression Analysis: Applied to model the relationship between independent variables (such as applicant attributes) and the dependent variable (loan acceptance), enabling the prediction and interpretation of loan approval probabilities based on different applicant profiles.

By leveraging these advanced analytical techniques, we endeavor to unravel the intricate dynamics of loan approval processes, empowering lenders with actionable insights to optimize decision-making and enhance the efficiency of their lending operations.



Data Description:

```
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Gender              499 non-null    object
 1   Married             499 non-null    object
 2   Dependents          499 non-null    object
 3   Education           499 non-null    object
 4   Self_Employed       499 non-null    object
 5   Applicant_Income    499 non-null    int64
 6   Coapplicant_Income  499 non-null    float64
 7   Loan_Amount         499 non-null    int64
 8   Term                499 non-null    float64
 9   Credit_History      499 non-null    float64
 10  Area                499 non-null    object
 11  Status              499 non-null    object
dtypes: float64(3), int64(2), object(7)
```

Sample Data :

	Gender	Married	Dependents	Education	Self_Employed	Applicant_Income	Coapplicant_Income	Loan_Amount	Term	Credit_History	Area	Status
0	Male	No	0	Graduate	No	584900	0.0	15000000	360.0	1.0	Urban	Y
1	Male	Yes	1	Graduate	No	458300	150800.0	12800000	360.0	1.0	Rural	N
2	Male	Yes	0	Graduate	Yes	300000	0.0	6600000	360.0	1.0	Urban	Y
3	Male	Yes	0	Not Graduate	No	258300	235800.0	12000000	360.0	1.0	Urban	Y
4	Male	No	0	Graduate	No	600000	0.0	14100000	360.0	1.0	Urban	Y

Assumptions

The data collection process involved employing a method of random sampling, which was deliberately chosen to guarantee that the subset of individuals studied accurately mirrors the entire population of loan applicants. This approach aimed to minimize bias and ensure that the insights drawn from the sampled data could be generalized to the broader applicant pool with a high degree of confidence.

While our analysis assumes that each data point is independent and unrelated to the others, it is essential to acknowledge that real-world data frequently exhibits nuances and interrelationships among variables. Despite this assumption of independence, variations and correlations may indeed exist within the dataset, reflecting the complex nature of the factors influencing loan application outcomes. Thus, while independence is a simplifying assumption often made in statistical analyses, recognizing and accounting for the potential presence of such variations is crucial for accurately interpreting and extrapolating findings from the data.

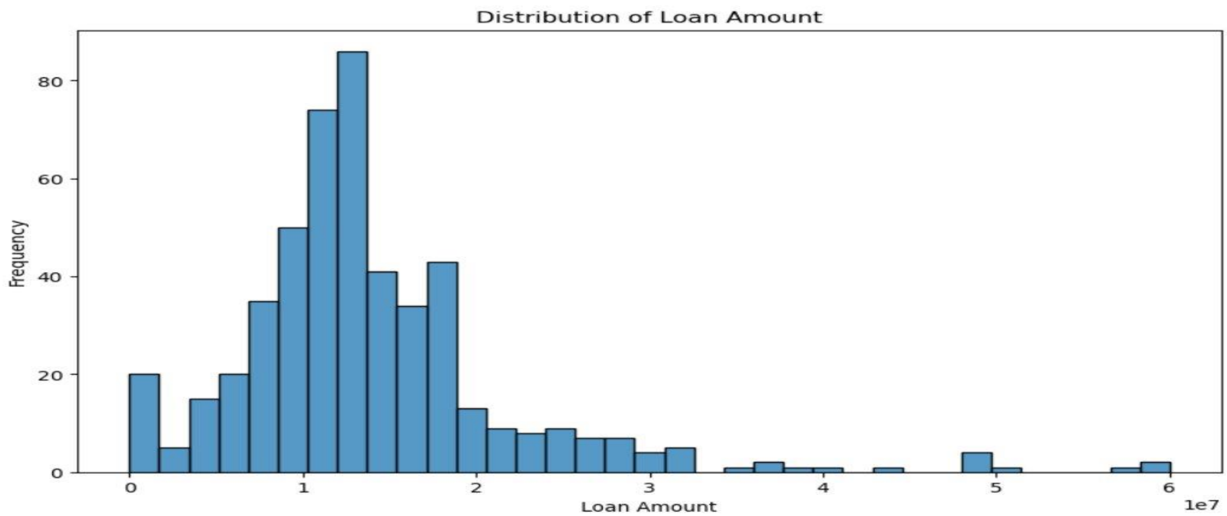
Limitations

- Data quality (accuracy, completeness, consistency) can vary, impacting analysis validity.
- Missing data, outliers, and errors in entry may distort results.
- The observations might not be applicable on a large population sample due to the small dataset size.
- Potential selection bias due to over or underrepresentation of certain applicant groups.
- Credit score data available only in binary form, lacking granularity (for e.g., ranging between 700-800).

Data Visualization and Analysis

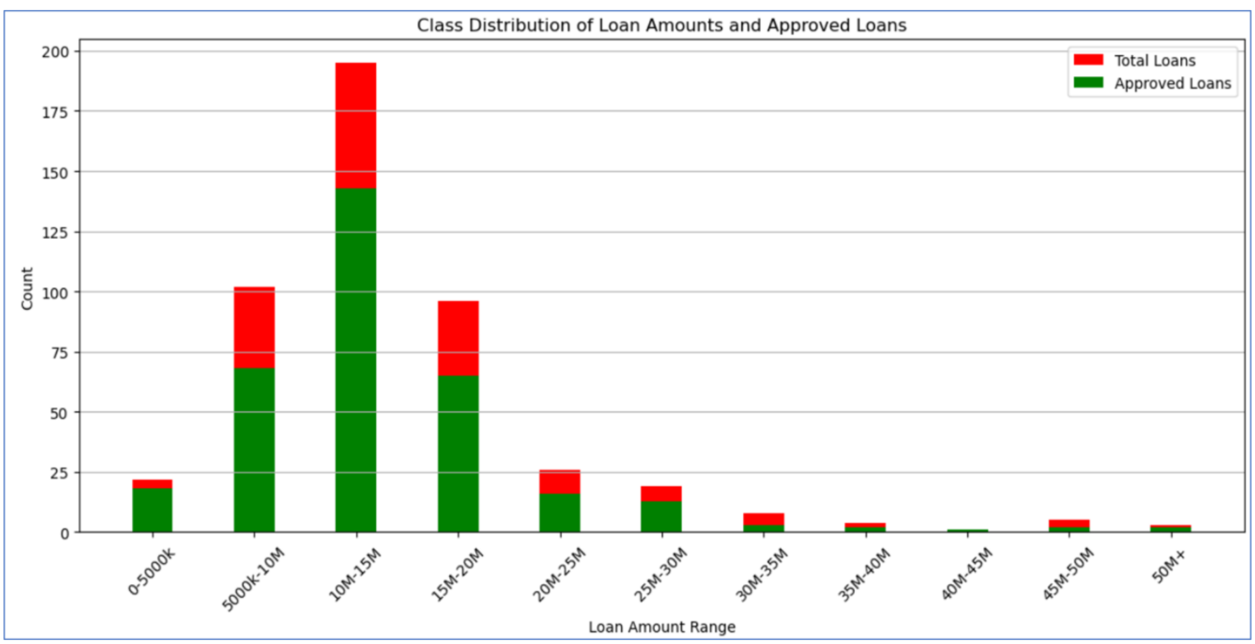
1.Frequency Distribution of Loan Amount :

The frequency distribution of loan amounts provides a comprehensive overview of the distribution patterns and prevalence of different loan values within a given dataset. By organizing the loan amounts into distinct categories or intervals and tallying the frequency of occurrences within each category, this analysis offers valuable insights into the typical loan amounts sought by applicants. This statistical technique enables us to identify common loan denominations, assess the spread and concentration of loan values, and uncover any notable trends or outliers. Understanding the frequency distribution of loan amounts is essential for financial institutions to tailor their lending strategies, assess risk exposure, and make informed decisions regarding loan approvals and product offerings.

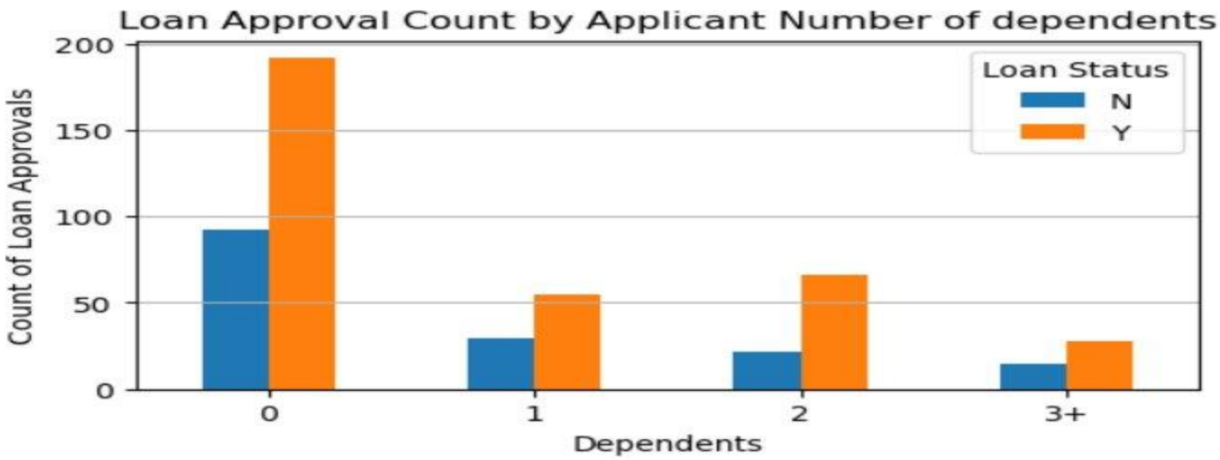


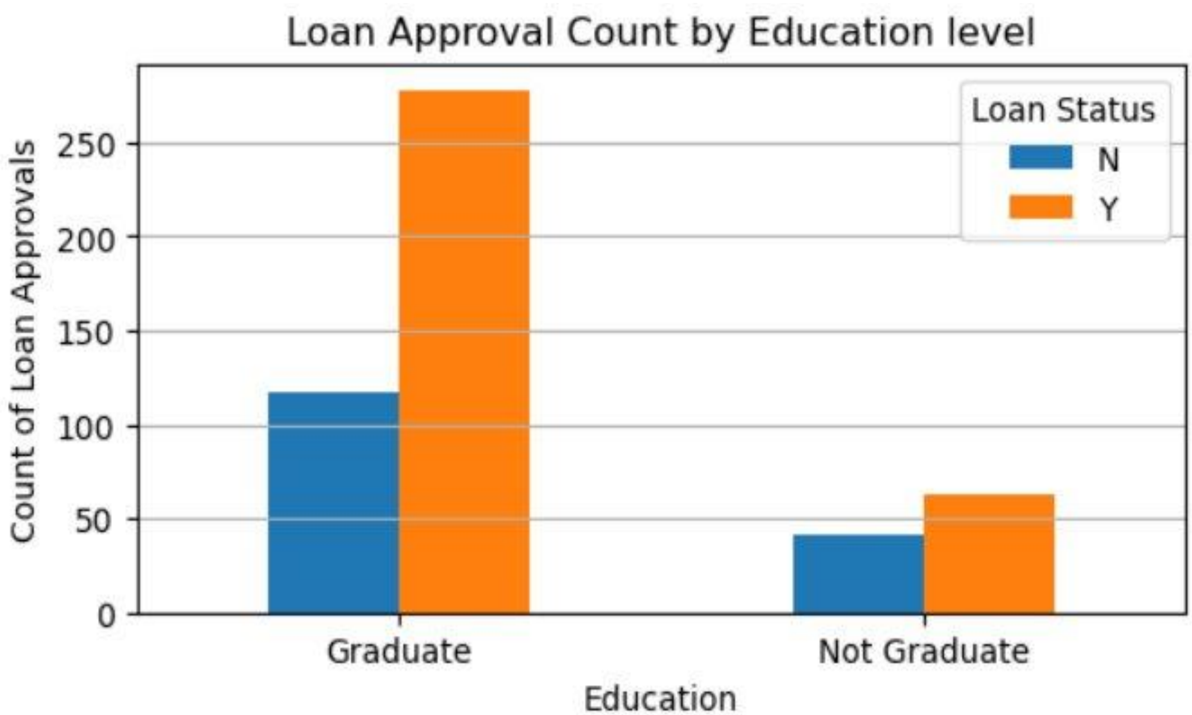
2.Loan Approval counts based on Loan Amounts:

The loan approval counts based on loan amounts provide a breakdown of approved loans across different value ranges. By categorizing loan amounts and recording approval counts within each category, this analysis reveals patterns in borrower preferences and lender risk tolerance. It helps financial institutions refine lending strategies and optimize decision-making processes to meet borrower needs while managing risks effectively.

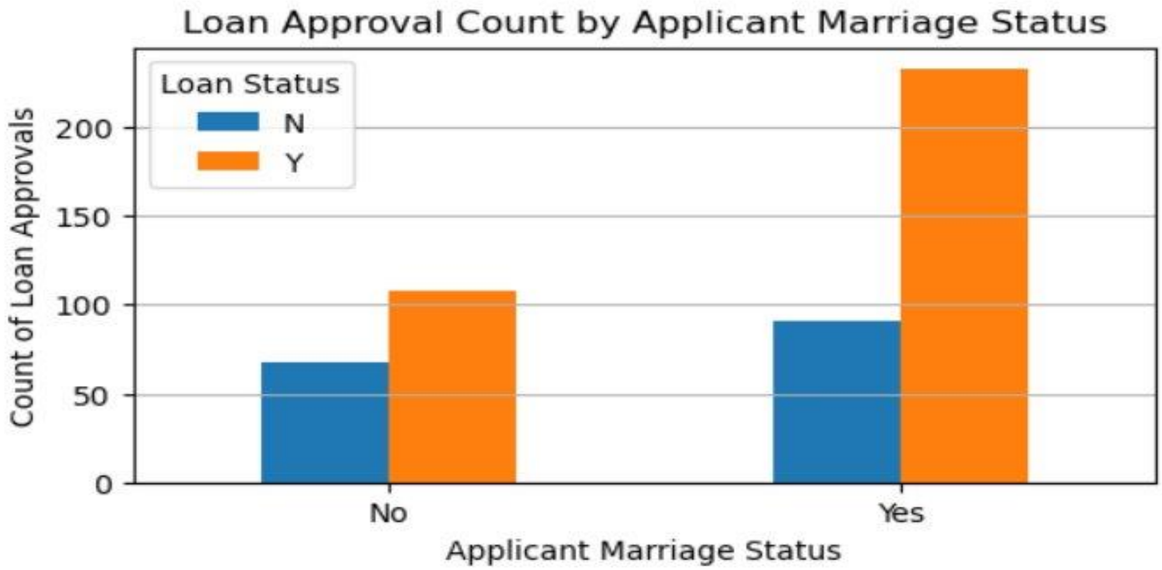


3.Loan Approval Count By Applicant Number of Dependents

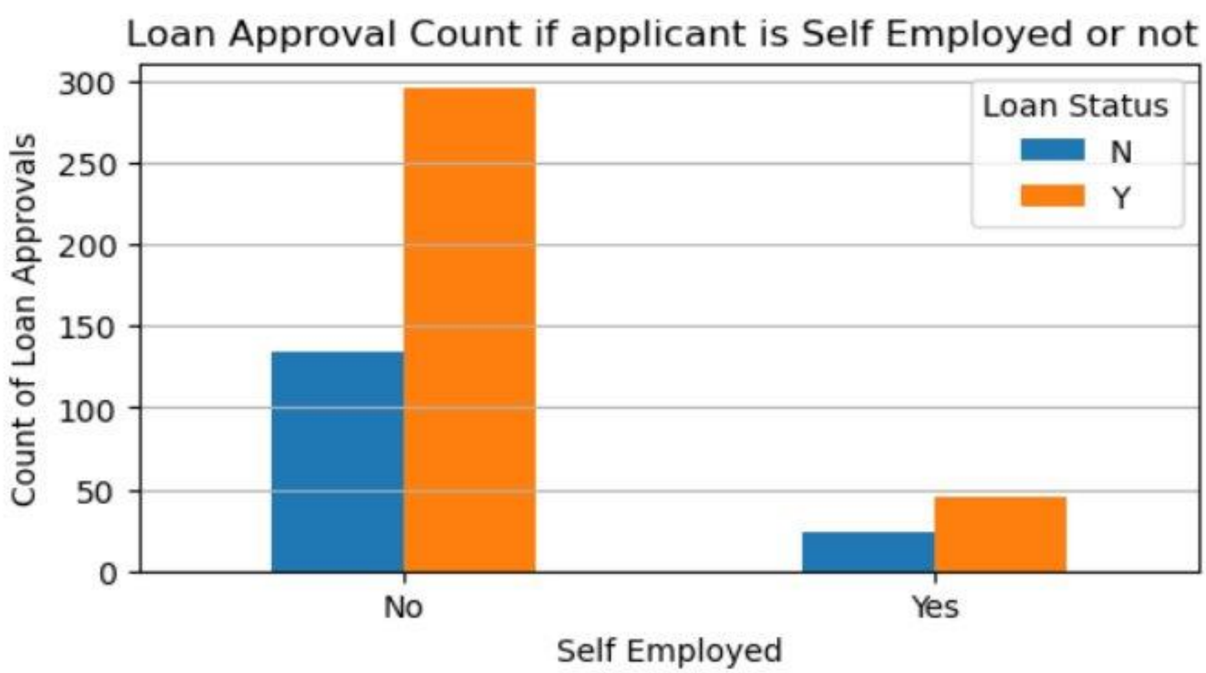




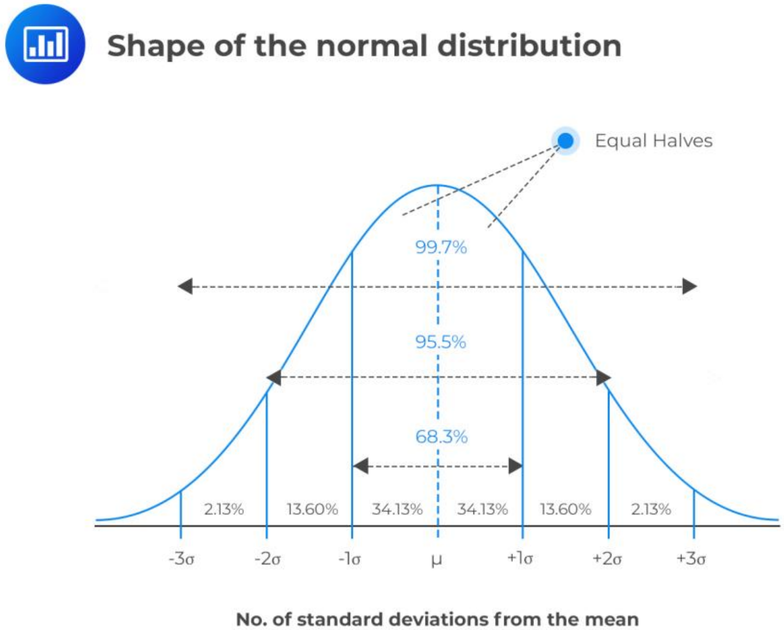
5.Loan Approval Count by Applicant Marriage Status



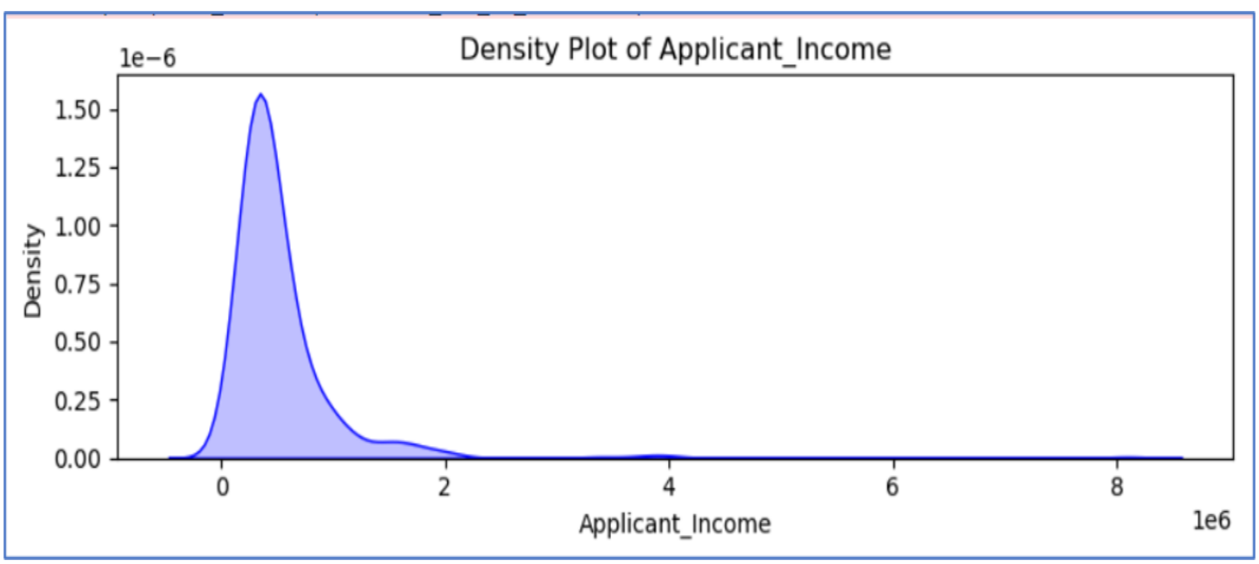
6.Loan Approval Count if applicant is self-employed or not

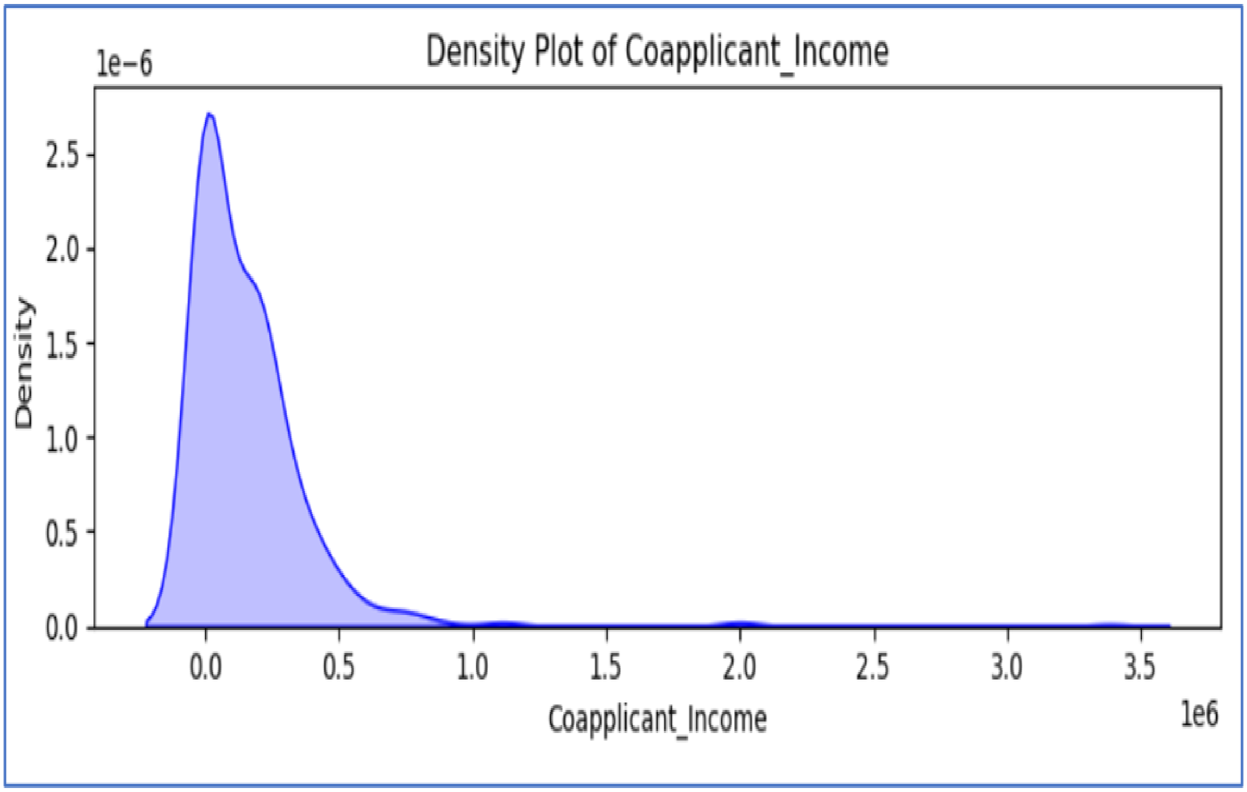


Normal Distribution Analysis :

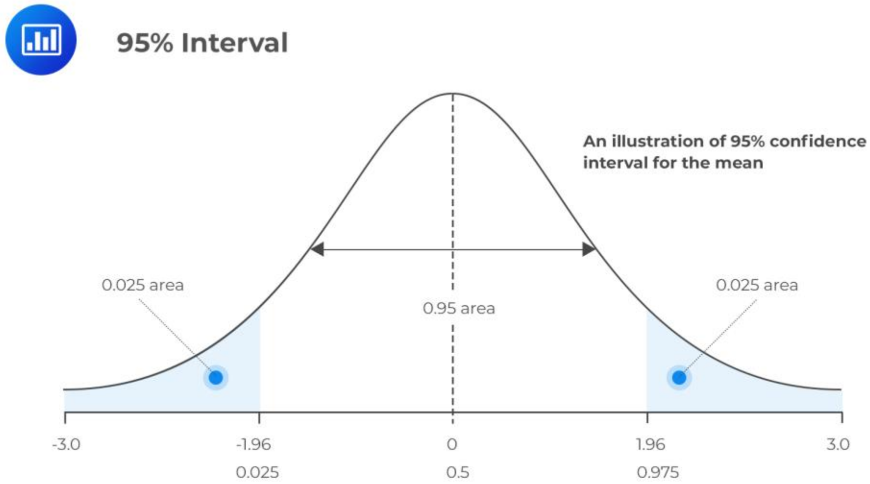


- The normal distribution is a continuous probability distribution with a bell-shaped curve. It is symmetric around the mean, which indicates that data closer to the mean occur more frequently than data farther away from the mean.
- We have used normal distribution to visualize and identify the relevant patterns in the dataset based on mean and the standard deviation obtained.



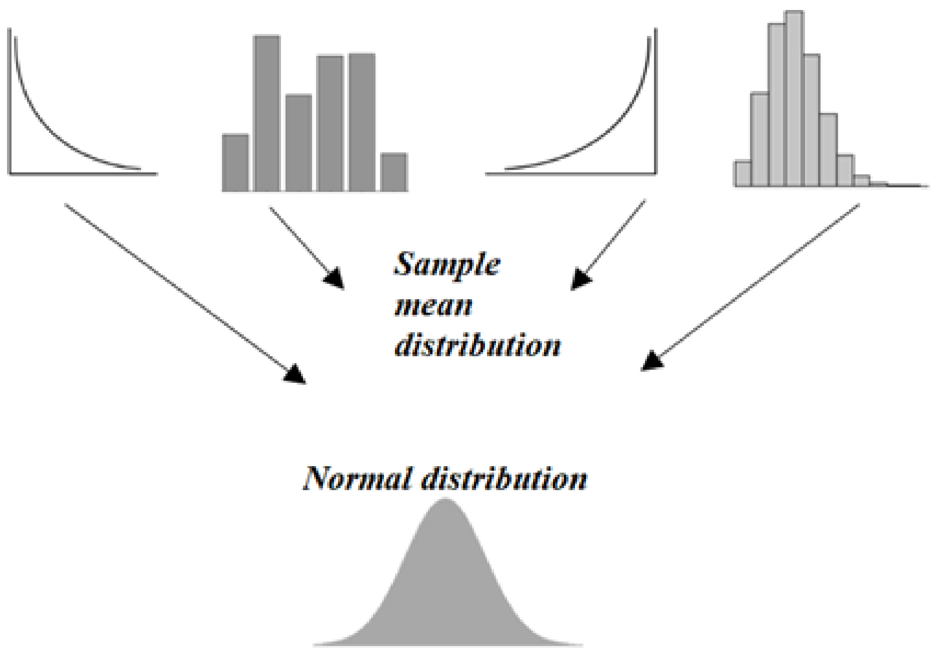


Confidence Intervals:

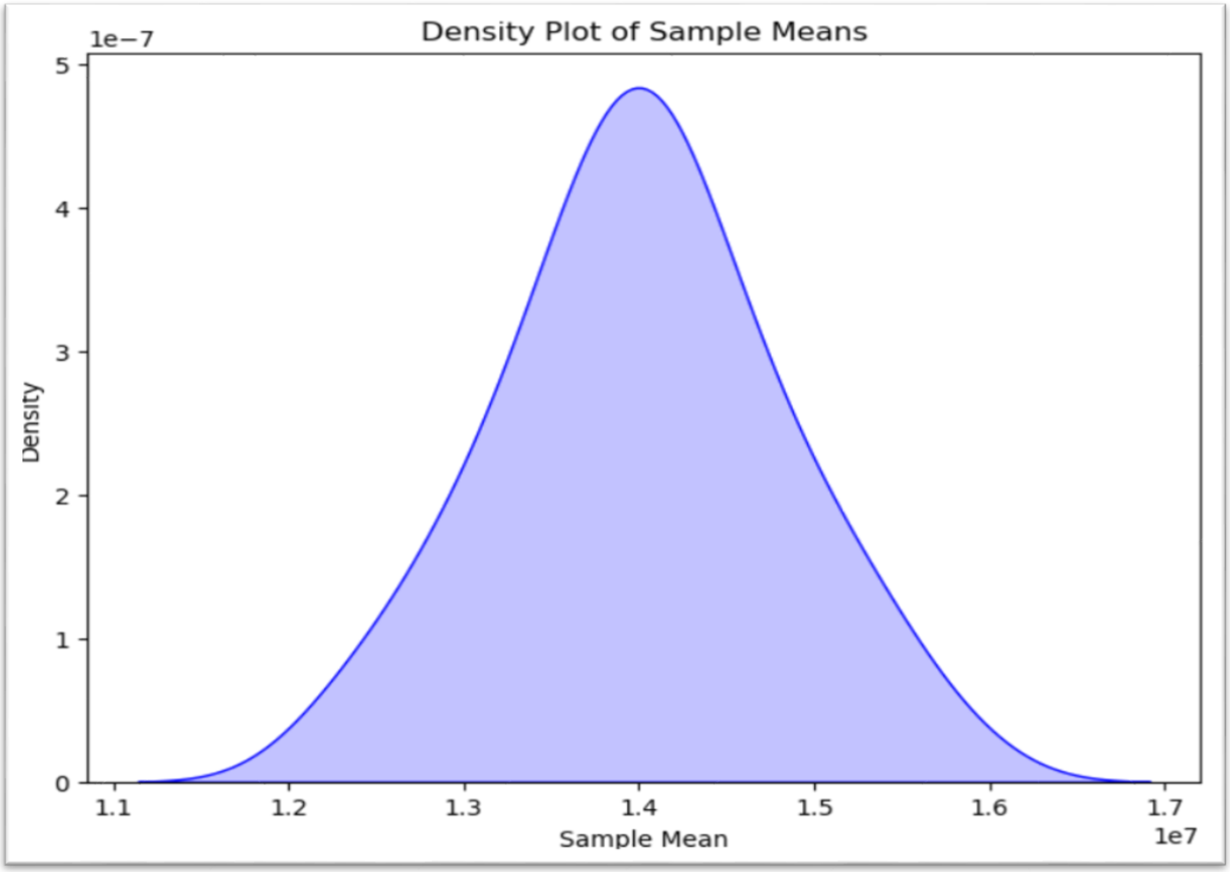


- A confidence interval is a statistical range, based on sample data, that is likely to contain the true value of a population parameter with a specified level of confidence.
- We calculated a confidence interval for the mean of loan amounts by using the sample mean, sample standard deviation, sample size, and desired confidence level to estimate where the true population mean lies with a specified level of confidence.
- The 95% confidence interval for the loan amount is estimated to be between 11,843,024.39 and 16,061,985.63.

Central Limit Theorem



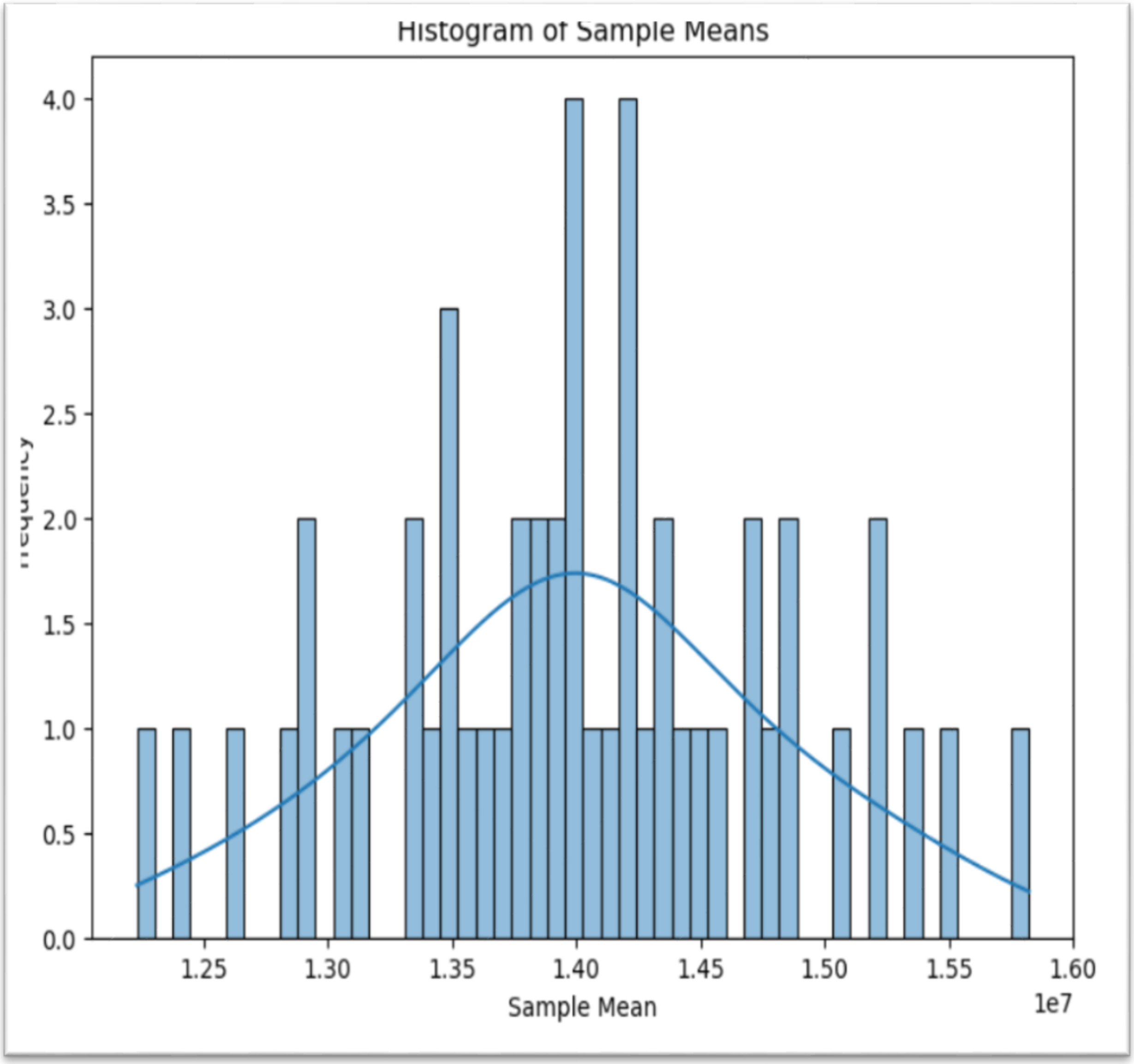
- We used the Central Limit Theorem to demonstrate that the sum or average of independent and identically distributed random variables approaches a normal distribution, regardless of the original distribution, when a large enough sample size is employed.
- In probability theory, the central limit theorem (CLT) states that the distribution of a sample variable approximates a normal distribution (i.e., a "bell curve") as the sample size becomes larger, regardless of the population's actual distribution shape.



Sample Means:

[12233000.0,12431000.0,12611000.0,12855000.0,
12924000.0,12926000.0,13086000.0,13113000.0,
13316000.0,13325000.0]

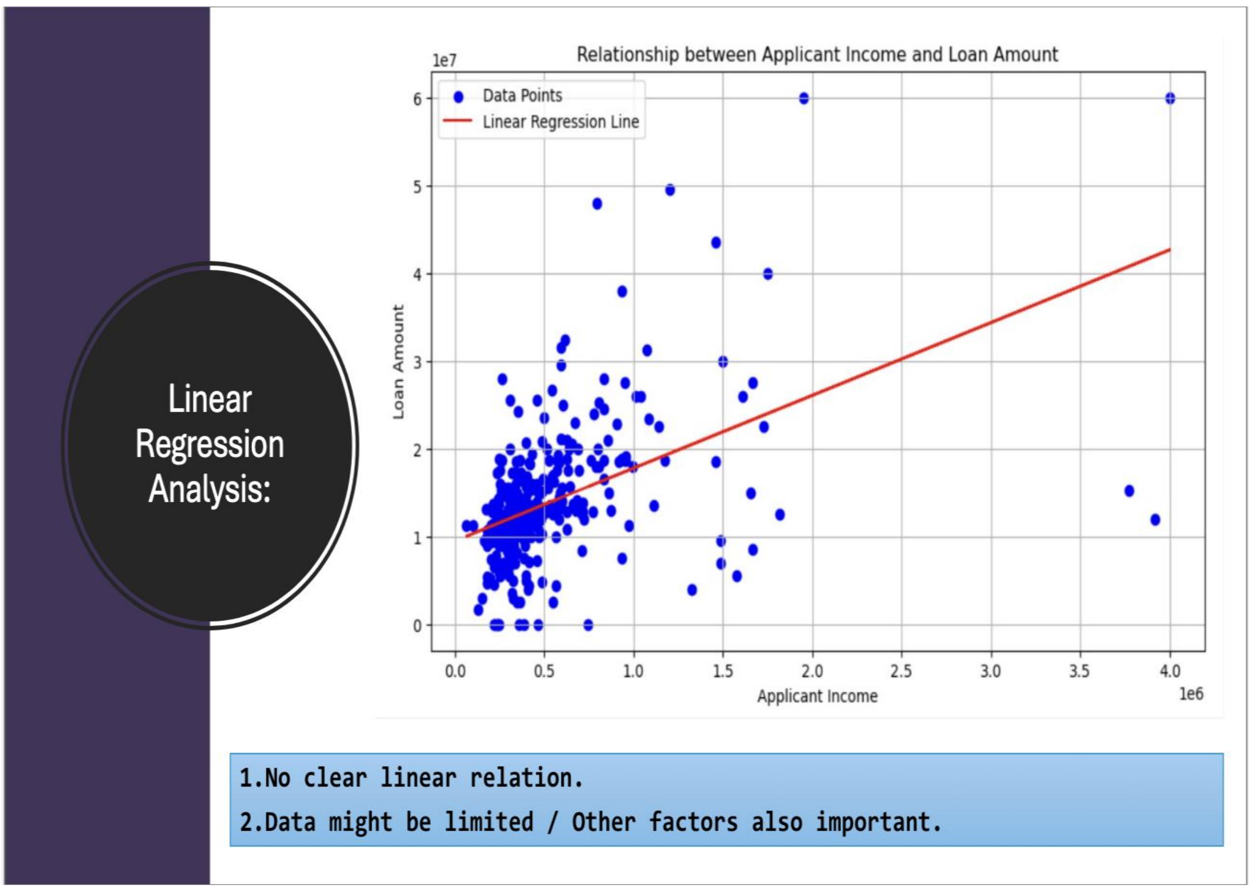
Linear Regression :




```
Mean of Sample Means: 14004200.0
Standard Deviation of Sample Means: 789163.2784158168
```

Linear Regression :

Linear regression between applicant income and loan amount involves exploring the relationship between these two variables using a linear model. This statistical technique aims to understand how changes in applicant income affect the amount of loan approved. By fitting a straight line to the data points representing applicant income and corresponding loan amounts, linear regression provides insights into the direction and strength of this relationship. Analyzing this relationship is essential for financial institutions to assess the affordability of loans for applicants and tailor lending practices accordingly.



Marginal Probability :

The marginal probability of a loan being approved or rejected refers to the probability of each outcome occurring independently of other variables. It provides a straightforward measure of the likelihood of loan approval or rejection without considering any additional factors. Calculating marginal probabilities is essential for understanding the overall distribution of loan approval outcomes and can serve as a baseline for more complex analyses involving multiple variables. This analysis helps financial institutions gauge the general likelihood of loan approval or rejection, facilitating informed decision-making in lending practices

```
# Marginal Probability of Loan Rejection
prob_loan_rej = marginal_prob(loan_data, 'Status', 'N')
print("Marginal Probability of Loan Rejection:", prob_loan_rej)
```

```
Marginal Probability of Loan Rejection: 0.3166
```

```
# Marginal Probability of Loan Approval
prob_loan_approval = marginal_prob(loan_data, 'Status', 'Y')
print("Marginal Probability of Loan Approval:", prob_loan_approval)
```

```
Marginal Probability of Loan Approval: 0.6834
```

Marginal Probability of Loan getting rejected: 31.7 %

Marginal Probability of Loan getting accepted: 68.3 %

Conditional Probability

- Conditional probability is the likelihood of an event occurring given that another event has already occurred.
- We calculated the conditional probability by dividing the count of cases where the specified conditions are met, and the target event (loan approval) occurs by the total count of cases where the conditions are met.
- It provided insights into the likelihood of loan approval given certain applicant characteristics or conditions.

Conditional Probability: Results Obtained

1. Conditional probability of loan approval given the gender of the applicant

P(Status = Y | Gender = Male) = (Number of Male applicants with Status = Y) / (Total number of Male applicants)

P(Status = Y | Gender = Female) = (Number of Female applicants with Status = Y) / (Total number of Female applicants)

```
Conditional Probability of Loan Approval Given Gender (Male): 0.6959
Conditional Probability of Loan Approval Given Gender (Female): 0.6250
```

2. Conditional probability of loan approval given the education level of the applicant

P(Status = Y | Education = Graduate) = (Number of Graduate applicants with Status = Y) / (Total number of Graduate applicants)

P(Status = Y | Education =Not Graduate) = (Number of Not Graduate applicants with Status = Y) / (Total number of Not Graduate applicants)

```
Conditional Probability of Loan Approval Given Education (Graduate): 0.7038
Conditional Probability of Loan Approval Given Education (Not Graduate): 0.6058
```

More observations:

```
Conditional Probability of Loan Approval Given Education (Graduate): 0.7038
Conditional Probability of Loan Approval Given Education (Not Graduate): 0.6058
```

```
Conditional Probability of Loan Approval Given Credit History (Good): 0.7859
Conditional Probability of Loan Approval Given Credit History (Bad): 0.0946
```

```
Conditional Probability of Loan Approval Given Area (Urban): 0.6352
Conditional Probability of Loan Approval Given Area (Rural): 0.6069
```

```
Conditional Probability of Loan Approval Given if Candidate is selfemployed :yes: 0.1132
Conditional Probability of Loan Approval Given if Candidate is selfemployed :No: 0.8868
```

Probability of Loan Approval Given Multiple Conditions:

```
{'Gender': 'Female', 'Education': 'Graduate', 'Credit_History': 1.0}
Probability of Loan Approval Given Multiple Conditions: 0.7377
```

```
{'Gender': 'Female', 'Education': 'Graduate', 'Credit_History': 1.0, 'Married':"Yes"}
Probability of Loan Approval Given Multiple Conditions: 0.7895
```



```
[illegible]ducation': 'Not Graduate', 'Married':"No",}
[illegible]pproval Given Multiple Conditions: 0.6364
```

```
{'Gender': 'Female', 'Education': 'Not Graduate', 'Married':"Yes",}
Probability of Loan Approval Given Multiple Conditions: 0.6667
```

```
{'Gender': 'Male', 'Education': 'Not Graduate', 'Married':"No",}
Probability of Loan Approval Given Multiple Conditions: 0.6000
```

Probability of Loan Approval Given Multiple Conditions:

```
{'Gender': 'Male', 'Education': 'Not Graduate',"Area":"Rural"}
Probability of Loan Approval Given Multiple Conditions: 0.5161

{'Gender': 'Male', 'Education': 'Graduate',"Area":"Rural"}
Probability of Loan Approval Given Multiple Conditions: 0.6489

{'Gender': 'Female', 'Education': 'Graduate',"Area":"Rural"}
Probability of Loan Approval Given Multiple Conditions: 0.5333

{'Gender': 'Male', 'Education': 'Graduate',"Area":"Urban"}
Probability of Loan Approval Given Multiple Conditions: 0.7048

{'Gender': 'Female', 'Education': 'Graduate',"Area":"Urban"}
Probability of Loan Approval Given Multiple Conditions: 0.5238
```

Observations:

- Male Gender (Loan Approval): 69.59% probability of approval. Female Gender (Loan Approval): 62.50% probability of approval. This indicates moderate favoritism towards male applicants in loan approval.
- Graduate Education (Loan Approval): 70.38% probability of approval. Non-Graduate Education (Loan Approval): 60.58% probability of approval. Graduates have a higher likelihood of loan approval, emphasizing the importance of education in access to financial resources.
- Good Credit History (Loan Approval): 78.59% probability of approval. Bad Credit History (Loan Approval): 9.46% probability of approval. A good credit history significantly increases the chances of loan approval, underscoring the role of creditworthiness in financial inclusion.
- Urban Area (Loan Approval): 63.52% probability of approval. Rural Area (Loan Approval): 60.69% probability of approval. Both urban and rural residents have moderate chances of loan approval, indicating an inclusive lending approach across different geographic areas.

- Self-Employment (Loan Approval): 11.32% probability of approval for self-employed applicants, compared to 88.68% for non-self-employed applicants. This highlights significant barriers to loan approval for self-employed individuals, suggesting potential areas for improving financial inclusion initiatives.

Interpretations:

- These findings underscore the importance of addressing disparities in access to financial services based on gender, education, credit history, and employment status.
- The disparity in loan approval probabilities between male and female applicants with similar educational backgrounds and residing in urban areas indicates potential gender-based bias in lending practices.
- Despite having comparable qualifications and residing in the same geographic location, female applicants face lower odds of loan approval, highlighting the existence of systemic inequalities in the lending process.

Scope of improvement in the Financial System:

- Addressing Geographical Bias: Ensure fair access to financial resources in rural and urban areas through equitable credit distribution policies and infrastructure investments.
- Mitigating Gender-Based Bias: Enforce anti-discrimination laws and offer tailored financial literacy programs to empower women in financial decision-making.
- Promoting Inclusive Practices: Foster partnerships and transparency in lending processes to promote financial inclusion for marginalized communities.
- Advancing Gender Equality: Support women's entrepreneurship and leadership in finance through funding and diversity initiatives.
- Empowering Marginalized Communities: Facilitate access to alternative financing and advocate for financial rights through community-led initiatives