

Loan Dataset Report

Introduction:

Dataset Overview:

Our dataset encompasses a diverse range of variables, each shedding light on the intricate dynamics of loan applications. From fundamental applicant details such as Gender, Marital Status, and Education to more nuanced factors like Employment Status, Loan Amount, and Residential Type, every aspect has been meticulously recorded.

Key Attributes:

1. Gender: A demographic identifier providing insights into the gender distribution among loan applicants.
2. Marital Status (Married, Not Married): Categorization based on marital status aiding in demographic segmentation.
3. Education (Graduate, Non-graduate): Classification based on educational background for further analysis.
4. Employment Status (Employed, Unemployed): Distinction between employed and unemployed applicants, crucial for risk assessment.
5. Loan Amount: The principal amount applied for, providing a measure of financial need and capacity.
6. Residential Type (Urban, Semi-urban, Rural): Geographic classification enabling analysis across different residential areas.

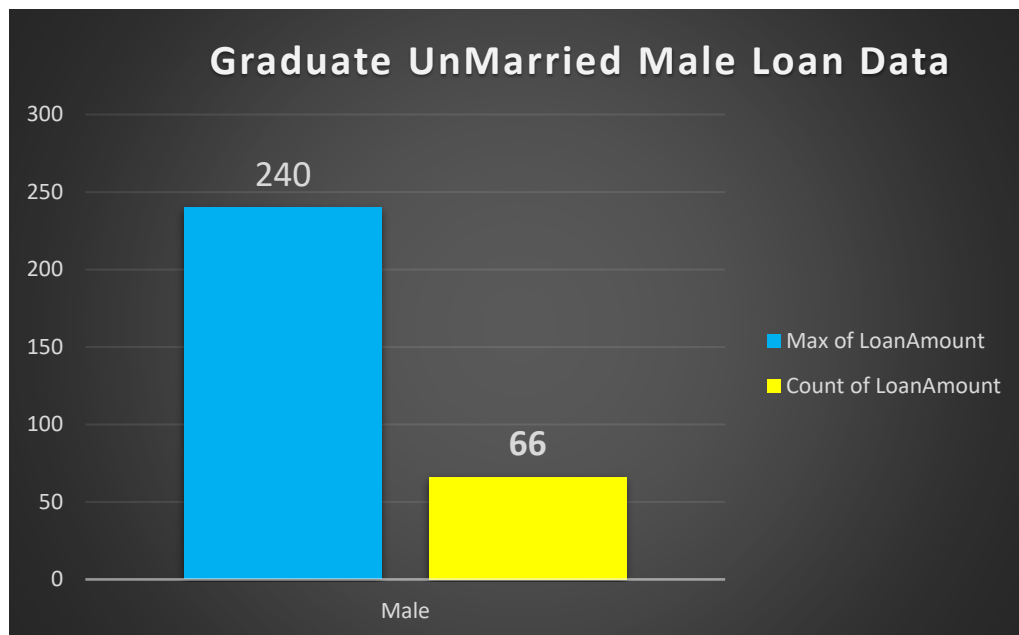
Questionnaire:

- Q1. How many male graduates who are not married applied for Loan? What was the highest amount?
- Q2. How many female graduates who are not married applied for Loan? What was the highest amount?
- Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?
- Q4. How many female graduates who are married applied for Loan? What was the highest amount?
- Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rural on the basis of amount.

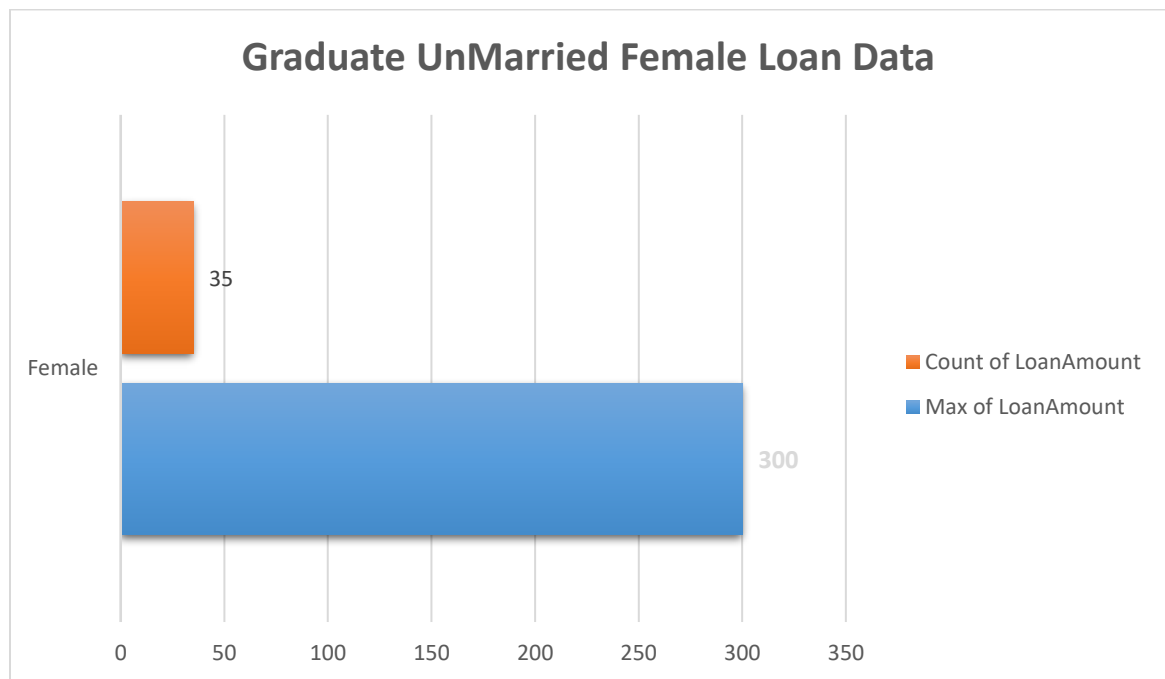
Analytics:

Q1. How many male graduates who are not married applied for Loan? What was the highest amount?

Gender	Married	Education
Female	No	Graduate
Male	Yes	Not Graduate
(blank)	(blank)	(blank)

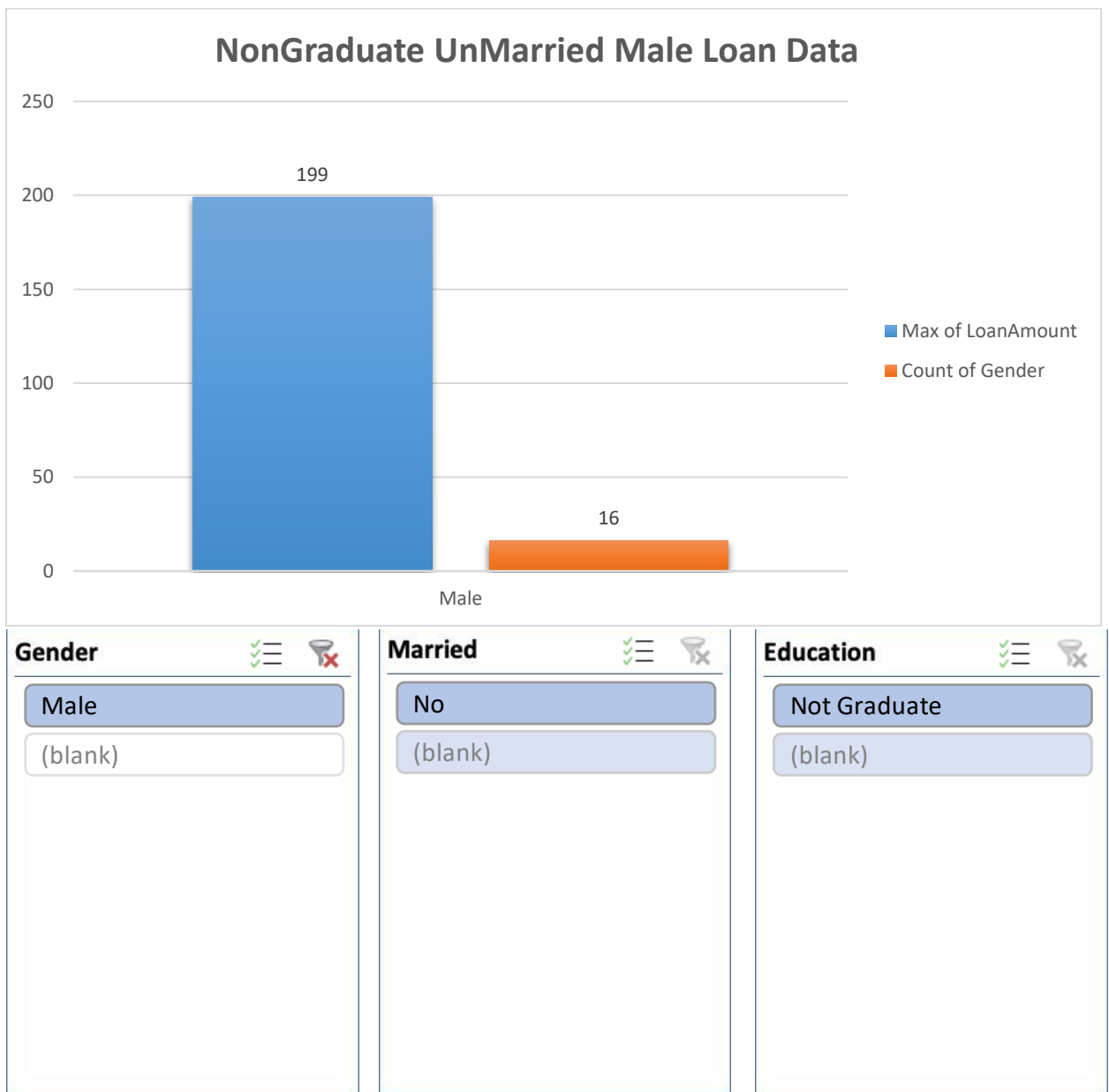


Q2. How many female graduates who are not married applied for Loan? What was the highest amount?

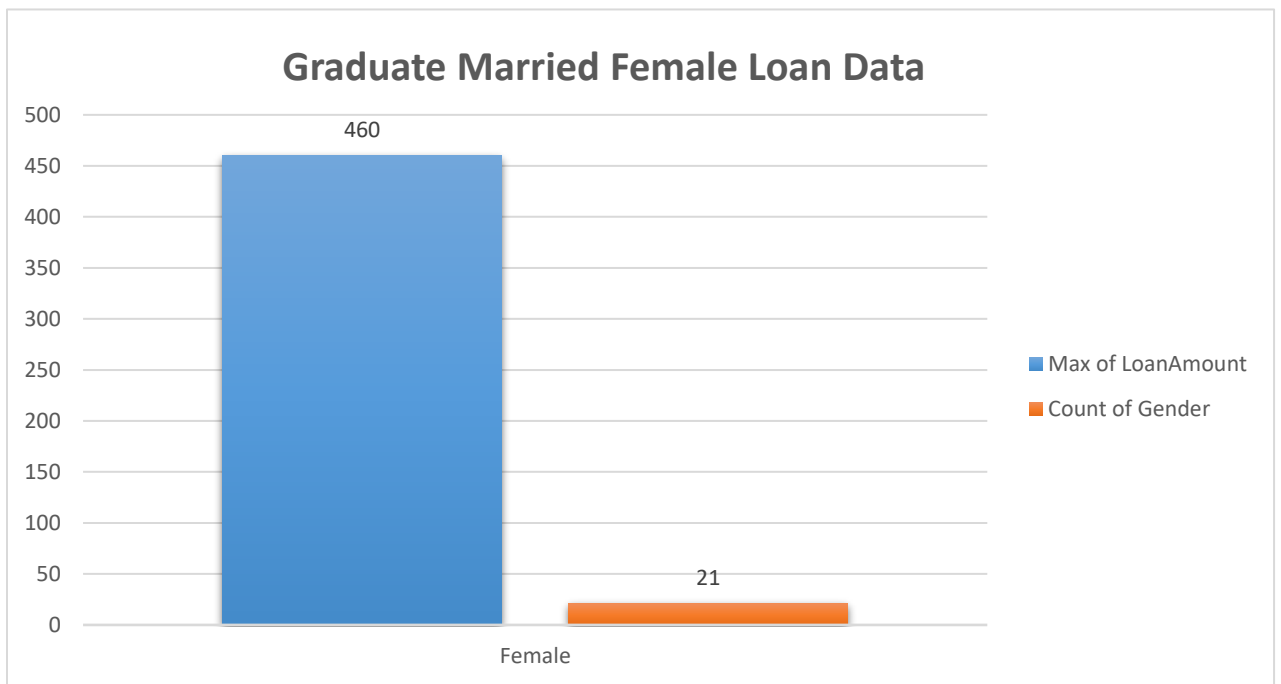


Gender	Married	Education
<div>Female</div>	<div>No</div>	<div>Graduate</div>
<div>Male</div>	<div>Yes</div>	<div>Not Graduate</div>
<div>(blank)</div>	<div>(blank)</div>	<div>(blank)</div>

Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?



Q4. How many female graduates who are married applied for Loan? What was the highest amount?



Gender ✓≡ ⌵

Female

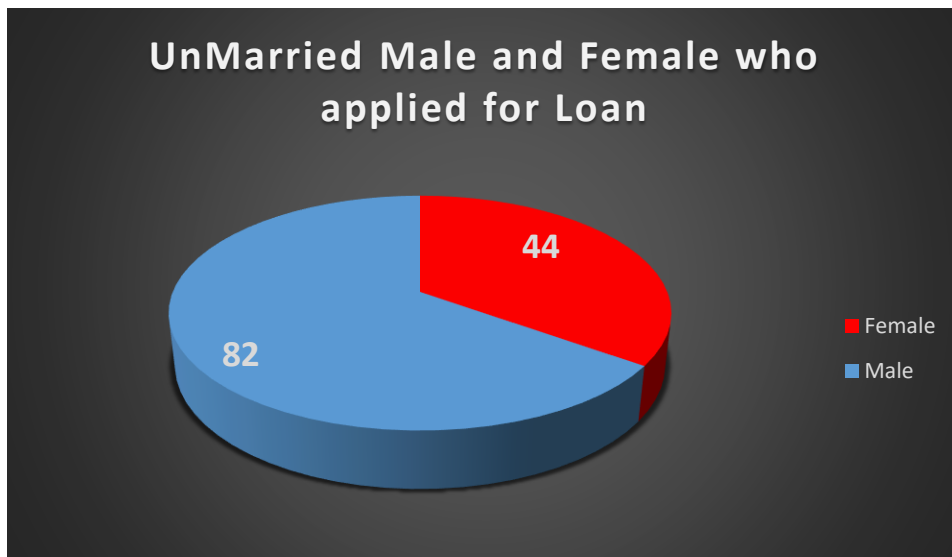
Married ✓≡ ⌵




Yes

Education ✓≡ ⌵

Graduate




Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rural on the basis of amount.



Gender   

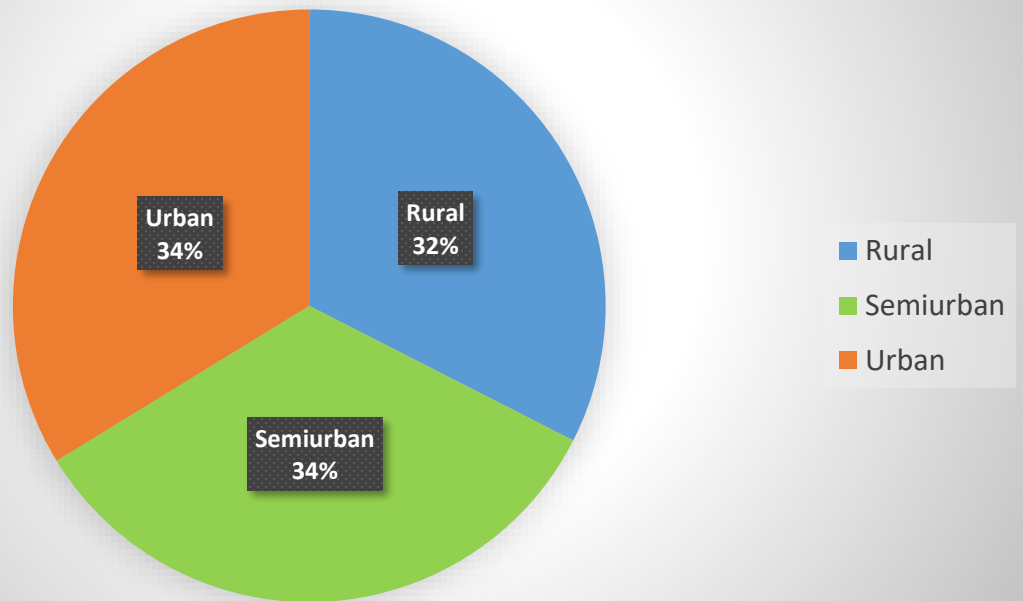
Female

Male

Married   

No

Comparision of Loan Amount on basis of Property Area



Property_Area ✓☰⚙

Rural

Semiurban

Urban

(blank)

LoanAmount ✓☰⚙

28

40

46

50

55

59

60

64

65

Conclusion:

Our analysis, using varied visualization techniques, revealed valuable insights, enhancing comprehension and decision-making. Visualizing data clarified complex findings, facilitating actionable strategies. This highlights the pivotal role of data visualization in extracting meaningful insights and informing decisions effectively.

Regression:

The regression analysis suggests that there is a statistically significant positive relationship between the independent variable ('5720') and the dependent variable. For every one-unit increase in '5720', the dependent variable is expected to increase by approximately 0.0059 units. However, it's important to note that the model only accounts for about 21.1% of the total variance in the dependent variable.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.45908096
R Square	0.21075532
Adjusted R Square	0.20858707
Standard Error	56.0766111
Observations	366

ANOVA

	df	SS	MS	F	Significance F
Regression	1	305655.205	305655.205	97.2004502	1.7676E-20
Residual	364	1144629.42	3144.58631		
Total	365	1450284.62			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%
Intercept	106.07753	4.10024098	25.8710478	1.7585E-84	98.014396	114.140665	98.01439
5720	0.0058851	0.00059692	9.85902887	1.7676E-20	0.00471125	0.00705895	0.004711

Correlation:

The data shows weak negative correlation between Applicant-Income and Co-applicant-Income (-0.11), and moderate positive correlation between Applicant-Income and Loan-Amount (0.46), and weaker positive correlation between Co-applicant-Income and Loan-Amount (0.14).

	<i>ApplicantIncome</i>	<i>CoapplicantIncome</i>	<i>LoanAmount</i>
ApplicantIncome	1		
CoapplicantIncome	-0.110334799	1	
LoanAmount	0.458768926	0.144787815	1

Anova (Single Factor) :

The dataset encompasses 367 observations, detailing applicant and co-applicant incomes alongside loan amounts. On average, applicants possess a higher income, averaging around \$4805.60, compared to co-applicants whose average income is approximately \$1569.58. Loan amounts vary widely, averaging \$134.28. ANOVA analysis underscores significant distinctions between the income and loan amounts across the groups, implying diverse financial profiles among applicants and co-applicants.

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
		176365	4805.59945	24114831.0
ApplicantIncome	367	5	5	9
CoapplicantIncome	367	576035	1569.57765	5448639.49
			7	1
			134.277929	3964.14112
LoanAmount	367	49280	2	4

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	4202537452	2	2101268726	213.200984	5.87569E-79	3.00392057
Within Groups	1082168110	1098	9855811.57			
Total	1502421856	1100				

Anova two factor without Replication:

The ANOVA results indicate significant variation both within rows ($p = 0.441$) and between columns ($p < 0.001$). This suggests that there are meaningful differences among the row categories and column categories in the dataset, warranting further investigation into the factors influencing these variations.

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Rows	1004340909	365	2751618.93	1.015674698	0.440986529	1.1881716
Columns	379216841.8	1	379216841.8	139.9761235	1.47092E-27	3.867061668
Error	988841123.7	365	2709153.763			
Total	2372398875	731				

Descriptive Statistics:

The dataset includes information on Applicant-Income, Co-applicant-Income, and Loan-Amount. The largest Applicant-Income recorded is \$72,529, while the smallest is \$0. For Co-applicant-Income, the largest value is \$24,000, and the smallest is \$0. Additionally, the Loan-Amount ranges from a maximum of \$550 to a minimum of \$0. Confidence levels for these variables at a 95.0% level are also provided, indicating the precision of the measurements within the dataset.

Largest(1)	72529	Largest(1)	24000	Largest(1)	550
Smallest(1)	0	Smallest(1)	0	Smallest(1)	0
Confidence	504.0756	Confidence	239.6059	Confidence	6.462910
Level(95.0%)	067	Level(95.0%)	543	Level(95.0%)	219