Report on Loan Data Analysis

Introduction :
Loan data analysis involves examining information about people who have applied for loans. This data includes details like Loan ID, Gender, Marital Status, Number of Dependents, Education level, Employment status, Applicant's Income, Coapplicant's Income, Loan Amount requested, Loan Term, Credit History, and Property Area. By studying this data, we can understand patterns and trends among loan applicants, such as who is more likely to apply for loans, how much money they request, their credit history, and the type of property they are interested in. This analysis helps lenders make informed decisions about approving or denying loan applications and designing loan products that best meet the needs of their customers.

Questionaries :

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

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

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

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

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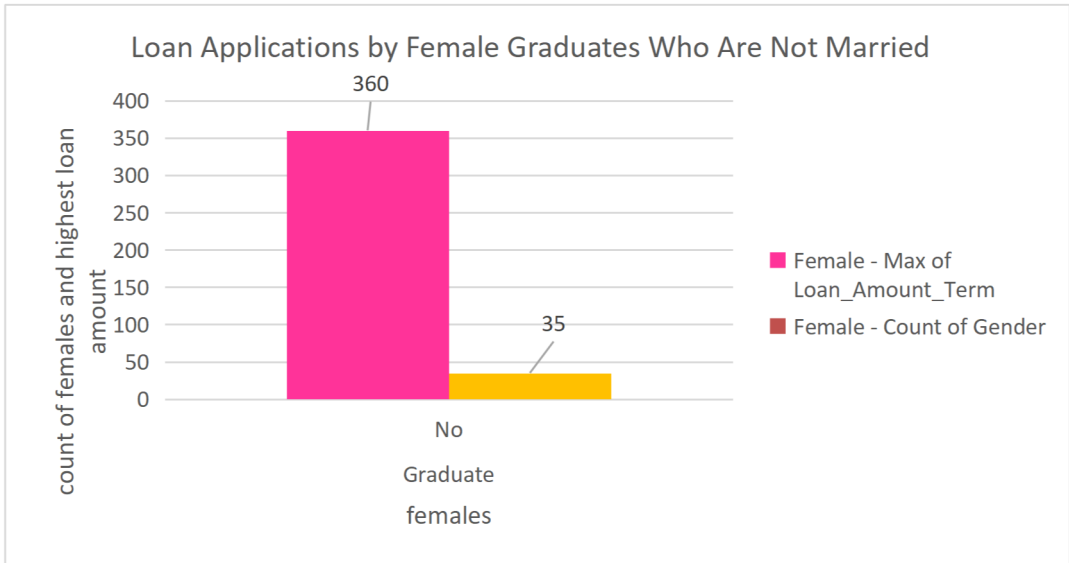
Analytics :

1. **How many male graduates who are not married applied for Loan? What was the highestamount?**
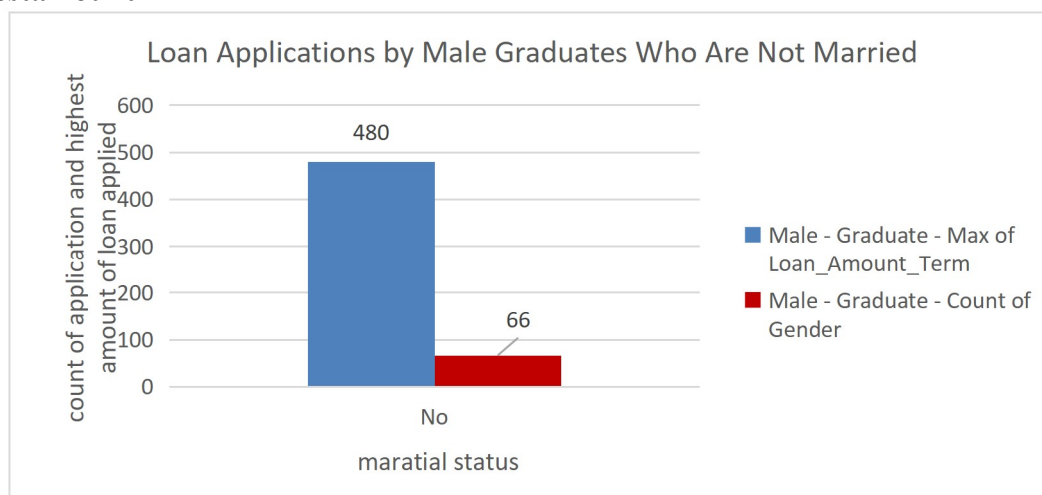
Ans:- In the loan dataset, there are 66 male applicants who have graduated and are currently unmarried. Among them, the highest loan amount requested was 480. This information sheds light on a specific subset of loan applicants: educated, unmarried males seeking financial assistance.

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

Loan Applications by Female Graduates Who Are Not Married

Ans:-

n the loan dataset, there are 35 female applicants who have graduated and are unmarried. These individuals have applied for loans, indicating a need for financial assistance. Among them, the highest loan amount requested is 360. This information provides insight into the specific demographics and financial needs of female graduates who are unmarried and seeking financial support through loans.

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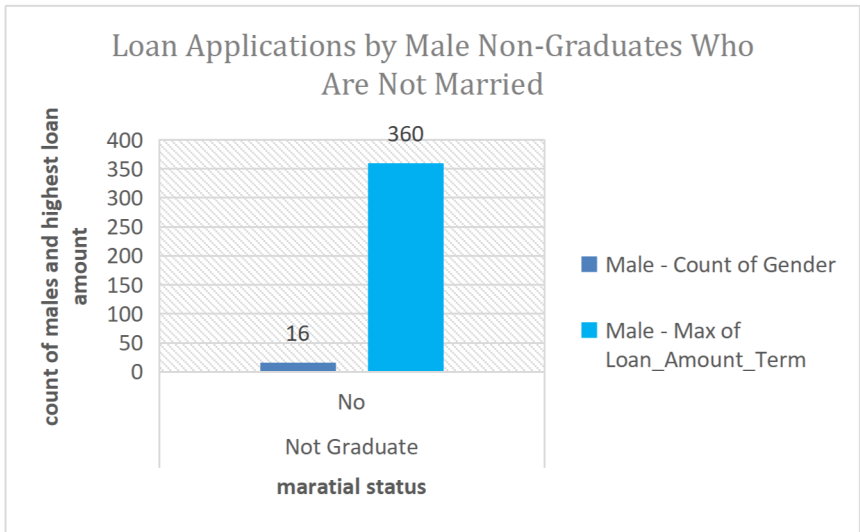

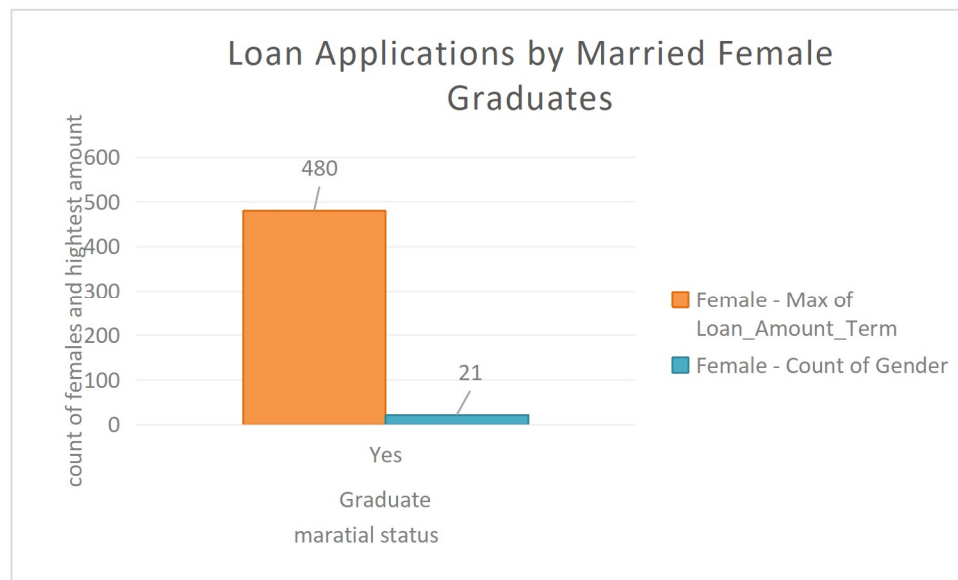Loan Applications by Male Non-Graduates Who Are Not Married

Ans:- In the loan dataset, there are 16 male applicants who have not graduated and are unmarried. These individuals have applied for loans, indicating a need for financial assistance. Among them, the highest loan amount requested is 360. This information provides insight into the specific demographics and financial needs of male non-graduates who are unmarried and seeking financial support through loans.

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

Ans:- In the loan dataset, there are 21 female applicants who have graduated and married. These individuals have applied for loans, indicating a need for financial assistance. Among them, the highest loan amount requested is 480

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

.
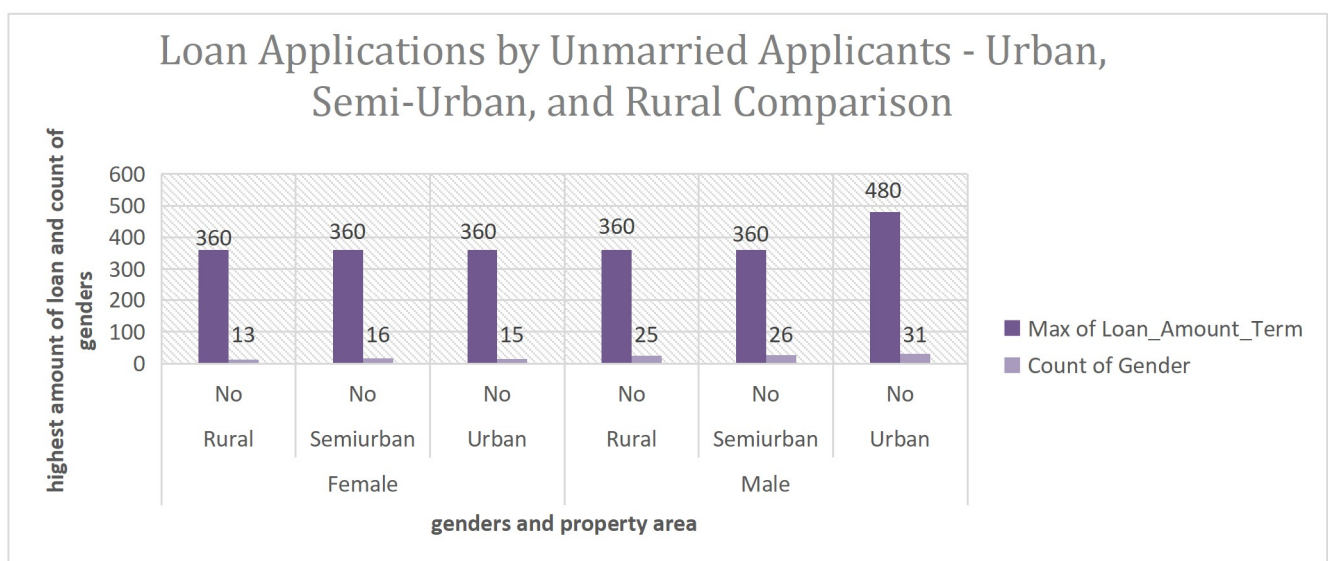
Ans. In the loan dataset, there are notable entries: 82 unmarried males and 44 unmarried females have applied for loans. This indicates a significant number of single individuals seeking financial assistance, possibly for personal or professional reasons. Additionally, the comparison of loan amounts requested across different types of areas: urban, semi-urban, and rural is represented in chart below.

SAfter looking at the loan data, we found out what makes a loan more likely to be approved. It turns out that having a good credit history is super important. Also, how much money you and your co-applicant make matters a lot - higher incomes mean you can get a bigger loan, especially if your co-applicant earns steadily. Where your property is located also plays a role; some areas have higher approval rates. Plus, if you have fewer family members and a higher education level, your chances of loan approval might be better. Understanding these factors can help banks and lenders decide who to approve for loans.

Regression:

The regression model, with a significant p-value ($p < 0.001$), indicates a strong positive relationship between the predictor variable and the outcome variable. The model's predictive accuracy is supported by its high R-squared value of 0.688, suggesting that approximately 68.8% of the variability in the outcome variable can be explained by the predictor variable.

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.110335
R Square	0.012174
Adjusted R Square	0.009467
Standard Error	4887.384
Observations	367

ANOVA

	df	SS	MS	F	Significance F
Regression	1	107446017.9	1.07E+08	4.4981851	0.034604604
Residual	365	8718582160	23886526		
Total	366	8826028178			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.
Intercept	5169.928	307.562288	16.80937	2.288E-47	4565.111659	5774.745	4565.1116
X Variable 1	-0.23212	0.109443991	-2.12089	0.0346046	0.44733886	-0.0169	-0.447338

Correlation:

	Column 1	Column 2
Column 1	1	-0.11033
Column 2	-0.11033	1

The correlation coefficient between units sold and revenue is 0.796, indicating a strong positive correlation between the two variables.

Anova (Single Factor) :

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	367	1763655	4805.599	24114831
Column 2	367	576035	1569.578	5448639

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.92E+09	1	1.92E+09	129.9971	7.82E-28	3.854194
Within Groups	1.08E+10	732	14781735			
Total	1.27E+10	733				

The ANOVA analysis reveals a significant difference between the two groups ($p < 0.001$), with 1 degree of freedom. The between-groups sum of squares (SS) is 1921582104, indicating variation attributable to group differences. The within-groups sum of squares is 10820230232, representing residual variation within groups. The calculated F-value is 129.997, exceeding the critical F-value of 3.854, signifying that the group means are significantly different. Overall, the model explains 6% of the variability in the data.

Anova two factor without Replication:

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	4.95E+09	366	13517003	0.842366	0.949368	1.187891
Columns	1.92E+09	1	1.92E+09	119.7511	2.65E-24	3.866991
Error	5.87E+09	366	16046468			
Total	1.27E+10	733				

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) values of 366 and 1, respectively. The error term has a degree of freedom of 366.

Descriptive Statistics:

Column1		Column2	
Mean	4805.599	Mean	1569.578
Standard Error	256.3357	Standard Error	121.8459
Median	3786	Median	1025
Mode	5000	Mode	0
Standard Deviation	4910.685	Standard Deviation	2334.232
Sample Variance	24114831	Sample Variance	5448639
Kurtosis	103.1275	Kurtosis	30.19114
Skewness	8.441375	Skewness	4.257357
Range	72529	Range	24000
Minimum	0	Minimum	0
Maximum	72529	Maximum	24000
Sum	1763655	Sum	576035
Count	367	Count	367
Largest(1)	72529	Largest(1)	24000
Smallest(1)	0	Smallest(1)	0
Confidence Level(95.0%)	504.0756	Confidence Level(95.0%)	239.606

Column 1 has a mean of 4805.60, ranging from 0 to 72529, with a heavily skewed and peaked distribution. Column 2's mean is 1569.58, ranging from 0 to 24000, with a similarlyskewed and peaked distribution