**Title:** Loan Prediction using Machine Learning

. **Introduction:** .

**In this project, we will predict whether a customer will get the loan from bank or not.**

**. DESCRIPTION** .

Dream Housing Finance company deals in all home loans. They have a presence across all urban, semi-urban, and rural areas. Customer-first applies for a home loan after that company validates the customer eligibility for a loan.

The company wants to automate the loan eligibility process (real-time) based on customer detail provided while filling the online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History, and others. To automate this process, they have given a problem to identify the customer's segments, those are eligible for loan amount so that they can specifically target these customers. Here they have provided a partial data set.

**COLUMN DESCRIPTION:**

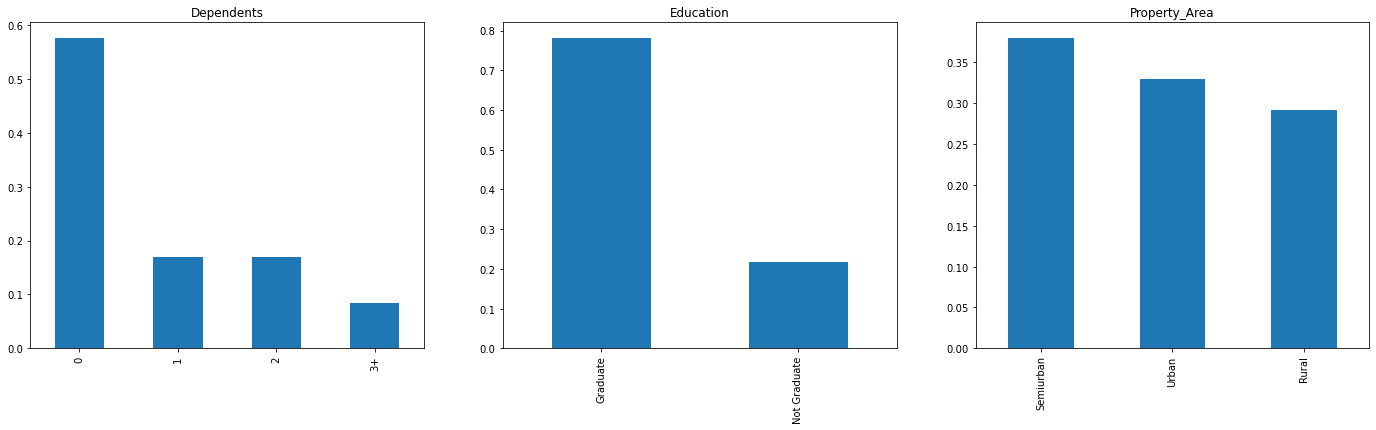
| Key Name | Description |
| --- | --- |
| Loan\_ID | Unique Loan ID |
| Gender | Male/ Female |
| Married | Applicant married (Y/N) |
| Dependents | Number of dependents |
| Education | Applicant Education (Graduate/ Under Graduate) |
| Self\_Employed | Self-employed (Y/N) |
| ApplicantIncome | Applicant income |
| CoapplicantIncome | Coapplicant income |
| LoanAmount | Loan amount in thousands |
| Loan\_Amount\_Term | Term of a loan in months |
| Credit\_History | credit history meets guidelines |
| Property\_Area | Urban/ Semi-Urban/ Rural |
| Loan\_Status | Loan approved (Y/N) |

. **APPROACH** .

My approach towards solution for this problem is: visualization of data,preprocessing the data elements,training-evaluating the model, and f1 score for all the models.

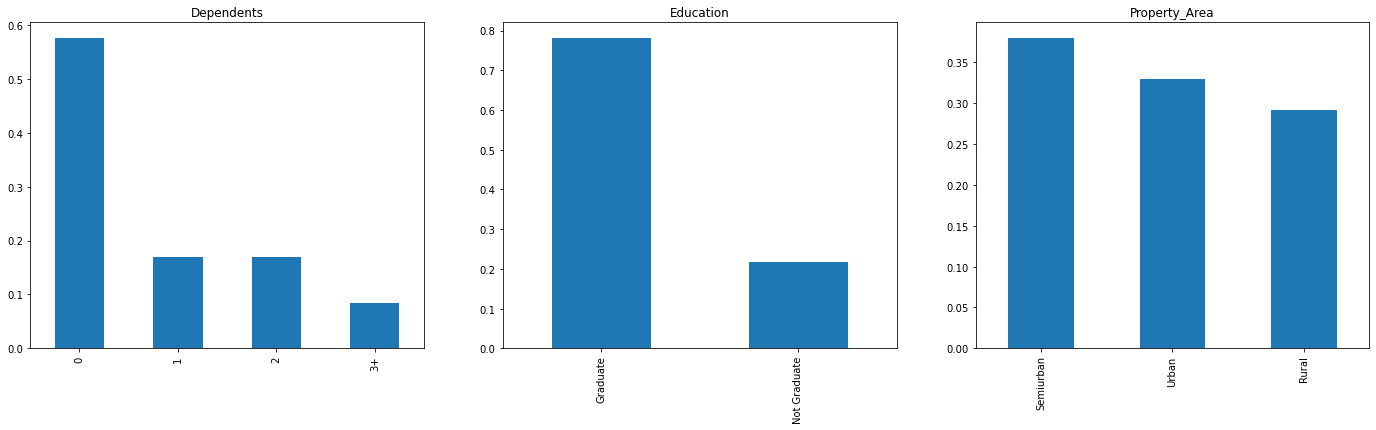
**. VISUALTIZATION .**

1. **Graduates vs Non Graduates**

****

* Graduated people are more likely to get the loan but there are some non graduated people too.

**2)Property Area**

****

* Customers are somewhat equal in urban and rural.

**. ALGORITHMS .**

I have used different algorithms in these model,they are mentioned below:

1. DecisionTreeClassifier
2. Support Vector Classifier
3. Logistic Regression
4. Linear Discriminant Analysis
5. Random Forest
6. K- Neirest Neighbour
7. Naive Bayes

**F1 score:**

0.90(Highest)

**Accuracy:**

0.84

**. CONCLUSION .**

After applying all these Visualizations and algorithms,we can conclude that

Logistic regression gives the best accuracy on this Dataset.