# **DMSL ASSIGNMENT – LOAN SANCTION STATUS DATASET INTERPRETATION OF THE LOGISTIC REGRESSION OUTPUT**

The **objective** of this analysis was to predict whether loan will be sanctioned or not. Logistic regression was used to classify outcomes into two categories: **loan sanctioned (Yes) or not sanctioned (No)**.

**Original Dataset output:**   
The Logistic Regression model was improved by removing factors with high p-values (above 0.05), indicating low significance. Variables like **Self\_Employed, ApplicantIncome, Loan\_Amount\_Term, and Education etc** were eliminated step by step due to their less significance.

# **A screenshot of a spreadsheet AI-generated content may be incorrect.**

In the end, only **Married (p = 0.023)** and **Credit\_History (p = 0.0)** remained, as they had p-value < 0.05 indicating high significance.  **A table with numbers and a number on it

AI-generated content may be incorrect.**

General equation for Sigmoid curve =**P(Y=1)= (ez)/(1+ez) [ S- CURVE ]**

A screenshot of a calculator

AI-generated content may be incorrect.i.e. the probability of loan getting sanction

where z = beta0 + beta1x1 + beta2x2……+ betanxn  =**[2.811+(0.509\*Married)+(3.860\*Credit\_History)**

* The **Wald test** checks if each variable in the model is significant.
* **Chi-square = 180.08:** High value indicates a good model fit.
* **LL= -**262.239: It checks the goodness of fit.

**NOTE:** In a good regression model, the actual and predicted values should be same**.** However, in reality, they are not always exactly equal.

**The classification or confusion matrix:**   
It compares **predicted values** with **actual values** to see that how well the model classifies the data.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a calculator

AI-generated content may be incorrect.



**The cut off = 0.5 (By default)**   
It is changing parameter just like level of confidence.

**TRUE CLASSIFICATION:** Following are cases where the model made the right prediction:

* **True Positives** (**378)**: It means that the model has **correctly made prediction** that 378 loan cases are eligible for loan sanctioning based on marital status and credit history
* **True Negatives** (**82)** : It means that the model **correctly predicted** rejection of 82 cases based on marital status and credit history, them being risky.

**MISCLASSIFICATION:** Following are cases where the model made the wrong prediction:

* **False Positives (97):** The model **wrongly predicted** that 97 loan cases were eligible for loan sanctioning, but in reality, they were risky and should have been rejected.
* **False Negatives (7):** The model **wrongly predicted** that 7 loan cases were not eligible for loan sanctioning, but in reality, they were actually safe and should have been approved.

**(AUC) Area under Curve: 0.89427556**

1. It states that the classification model has overall accuracy of 89.42%.
2. AUC represents that if there are 100 data points, then 89.42 data points out of them lie below the curve that is the sigmoid curve (which takes the shape of yes curve).
3. As, AUC is **closer to 1** so it has higher reliabilityin differentiating between sanctioned and not sanctioned loans.

**CONCLUSION**

The logistic regression model is good at predicting loan eligibility with **81.56% accuracy** and an **AUC of 0.894**. It correctly approved most eligible loan cases but had also wrongly approved risky loan cases (97 False Positives) **leading to financial losses**. It misses some eligible loan cases (7 False Negatives) which may result in **customer loss** as they might go to competitors.