**Project Title**: Predicting Customer Churn in a Telecommunications Company

**Data Description:** 0 Age 1000 non-null int32

1 Gender 1000 non-null object

2 Location 1000 non-null object

3 Num\_Calls 1000 non-null int32

4 Call\_Duration 1000 non-null float64

5 Data\_Usage 1000 non-null float64

6 Monthly\_Charges 1000 non-null float64

7 Total\_Charges 1000 non-null float64

8 Tenure 1000 non-null int32

9 Churn\_Status 1000 non-null int32

**Find the missing value**: There is no outliers in this data.

**Check the Ouliers:** In the column ‘Total\_Charges’ there is an outliers.

**Remove The Outliers:** Remove The outliers Using IQR method.  
**LabelEncodeing:** Then label Encoding two categorical column ‘Gender’ and’ location’

**Correlation:** ‘Monthly \_charge ‘and ‘Total \_charge ‘has correlation.

**Create a new feature for average monthly usage:** Avg\_Monthly\_Usage, High\_Data\_Usage, Call\_Intensity These are new features.

**Model Fit:** Then fit the Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and check highest accuracy 66.16 for logistic regression

**After cross validation:** The accuracy is 69.29 after cross-validation.

**Hyparparameter tuning:**67.5

**Most Important Feature:** Call intensity has most important fetaures