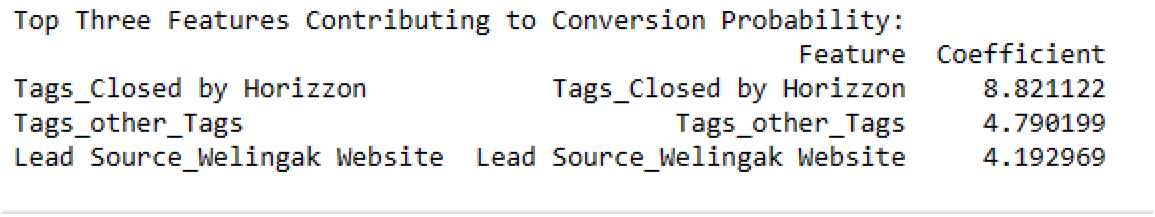
SUMMARY REPORT

- First step is to Import necessary libraries such as numpy, pandas, matplotlib, seaborn, sikit learn, statsmodel and loading the dataset.
- Next is to identify the null values and impute/remove the null values according to the effect of the Variable on the dataset.
- If there are more than 30-40% null values then we will simply drop the rows. If the null values are less then we will impute the values according to the type of the variable (i.e. if categorical then impute with the mode, if numerical impute with the mean.
- Some variables having more than 30-40% null values are kept because they predict the conversion rate of lead significantly. In this the rows containing the null values are dropped off.
- Identify the outliers using boxplot.
- Identifying the data imbalance using countplot.
- EDA (Exploratory Data Analysis)
- Dummy variable creation.
- Train Test Split.
- Model building.
- Feature Selection- It is one firstly by RFE (Recursive Factor Elimination) and after that by manual selection on the basis of p-value and VIF (Variance Influence Factor).
- Model Evaluation- It is done by confusion matrix, ROC –curve, accuracy, sensitivity, specificity, recall and precision.

OBSERVATIONS

- The logistic regression model predicts the probability of the Target variable having a certain value, rather than predicting the value of the target variable directly. Then a cutoff of probability is used to obtain the predicted value of the target variable.
- Here the logistic regression model is used to predict the probability of conversion of a customer.

- Final Logistic Regression model is built with 14 features.
- Top three features contributing in lead conversion are-



```
Top Three Features Contributing to Conversion Probability:
                                                 Feature  Coefficient
Tags_Closed by Horizzon             Tags_Closed by Horizzon     8.821122
Tags_other_Tags                             Tags_other_Tags     4.790199
Lead Source_Welingak Website  Lead Source_Welingak Website     4.192969
```