

## SUMMARY

- Firstly understood the business problem statement which primarily focused on the lead conversion
- Then, tried to analyse data in various data reading steps. data types of columns  
Checked the shape of data, etc
- Then, tried to analyse & understand the trends with by using various plots of seaborn etc
- Used logistic regression model with dummy var. creation and used it for optimal results
- Looked for correlations and collinearity & splitted the data sets and scaled them
- Used sklearn and statsmodels for building LR model, with the help of RFE and manual feature selection
- Checked optimal cutoff using ROC Curve method and model evaluation metrics.

### **Imp features:**

- Lead Origin\_Lead Add Form
- TotalVisits
- Lead Source\_Welingak Website, Google Website, Direct traffic
- What is your current occupation\_Not available
- Total Time Spent on Website
- Specialization\_Not available
- Country\_Not available
- What is your current occupation\_Working Professional
- Lead Origin\_Landing Page Submission
- Specialization\_Hospitality Management

### **Methodology:**

- Understanding the data
- Data Preparation
- EDA
- Creation of Dummy Var.
- Train Test Split
- Feature Scaling
- Checked Correlations
- Model Building (Feature Selection Using RFE, improvising the model further inspecting VIF and p- values)
- Built final model
- Model evaluation with different metrics.

**Accuracy= 80%, Precision=71%, Recall=73%**