## Summary

**Problem Description**

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

**Analysis:**

To analyze the dataset we tried to get the look and feel of the data, we observed following things:

- Number of rows and columns
- Data types of each columns
- Checking first few rows how data looks
- Checking how the data is spread.
- Checking for duplicates, if any.

For data cleaning we checked for discrepancies in the dataset.

- Checking for any column names correction
- Checking for null values and imputing them with appropriate methods

We used mode imputation for categorical columns.
We used mean imputation for numerical columns if there is no skewness in data.
We used median imputation for numerical columns if there is skewness in the data.

**Approach**

From above problem description we conclude that the above problem is the classification problem, hence we choose logistic Regression to calculate the Lead rate.
Below are the steps followed to solve this problem:

1. Data Reading and Understanding
2. Data Cleaning
3. Data visualization and Outlier Treatment
4. Feature Scaling
5. Model Building
6. Model Evaluation on Train set
7. Prediction on Train Set
8. Conclusion