**Customer Segmentation Problem**

**PROBLEM STATEMENT:**

* A digital arm of a bank faces challenges with lead conversions. The primary objective of this division is to increase customer acquisition through digital channels. The division was set up a few years back and the primary focus of the division over these years has been to increase the number of leads getting into the conversion funnel.
* They source leads through various channels like search, display, email campaigns and via affiliate partners. As expected, they see differential conversion depending on the sources and the quality of these leads.
* They now want to identify the leads' segments having a higher conversion ratio (lead to buying a product) so that they can specifically target these potential customers through additional channels and re-marketing. They have provided a partial data set for salaried customers from the last 3 months. They also capture basic details about customers.

**OBJECTIVE:**

* Need to identify the customers with a high probability of conversion

**DATA DICTIONARY:**

**Input variables:**

|  |  |
| --- | --- |
| ID | Unique ID (cant be used for predictions) |
| Gender | Sex of the applicant |
| DOB | Date of Birth of the applicant |
| Lead Creation Date | Date on which Lead was created |
| City Code | Anonymised Code for the City |
| City Category | Anonymised City Feature |
| Employer Code | Anonymised Code for the Employer |
| Employer\_Category1 | Anonymised Employer Feature |
| Employer\_Category2 | Anonymised Employer Feature |
| Monthly Income | Monthly Income in Dollars |
| Customer Existing Primary Bank Code | Anonymised Customer Bank Code |
| Primary Bank Type | Anonymised Bank Feature |
| Contacted | Contact Verified (Y/N) |
| Source | Categorical Variable representing source of lead |
| Source Category | Type of Source |
| Existing EMI | EMI of Existing Loans in Dollars |
| Loan Amount | Loan Amount Requested |
| Loan Period | Loan Period (Years) |
| Interest Rate | Interest Rate of Submitted Loan Amount |
| EMI | EMI of Requested Loan Amount in dollars |
| Var1 | Categorical variable with multiple levels |
| Approved | (Target) Whether a loan is Approved or not (0/1) |

**Expected Deliverables:**

* Create a power point presentation explaining the approach, process and key findings from the case study
* Submit the R/Python code along with presentation

**DEPLOYED MODELS:**

* Logistic regression and Decision Tree
* Plotted roc\_auc score and curve