Assignment Submission

Lead Scoring Case Study

Submission by:

Ravi Kumar Desai

Aditi Rani

DS Batch 69

Problem Statement & Analysis Approach

Problem Statement: To help 'X Education' company increase lead conversion ratio for its online courses offered to industry professionals

Analysis Approach: Using logistic regression to analyze problem statement and come back with meaningful insights. Below steps were taken with respect to logistic regression approach

Data Cleaning
1. Checked missing values and removed columns with missing values>30%
2. Filtered out rows with high number of missing values (except 'Country')
3. Dropped irrelevant columns or columns which do not give much insight

Data Preparation
1. Created dummy variables for categorical columns with multiple values

Built Logistic Regression Model:
1. Divided the data into test and train set
2. Used minmax scaler to scale various features to same scale
3. Removed columns with high p-value i.e. statistically insignificant/low association with conversion rate
4. Refined the model to increase accuracy

Testing:
Tested the model on test data

Key Observations

Heat Map: Darker color indicates high co-relation while lighter one shows lower co-relation.
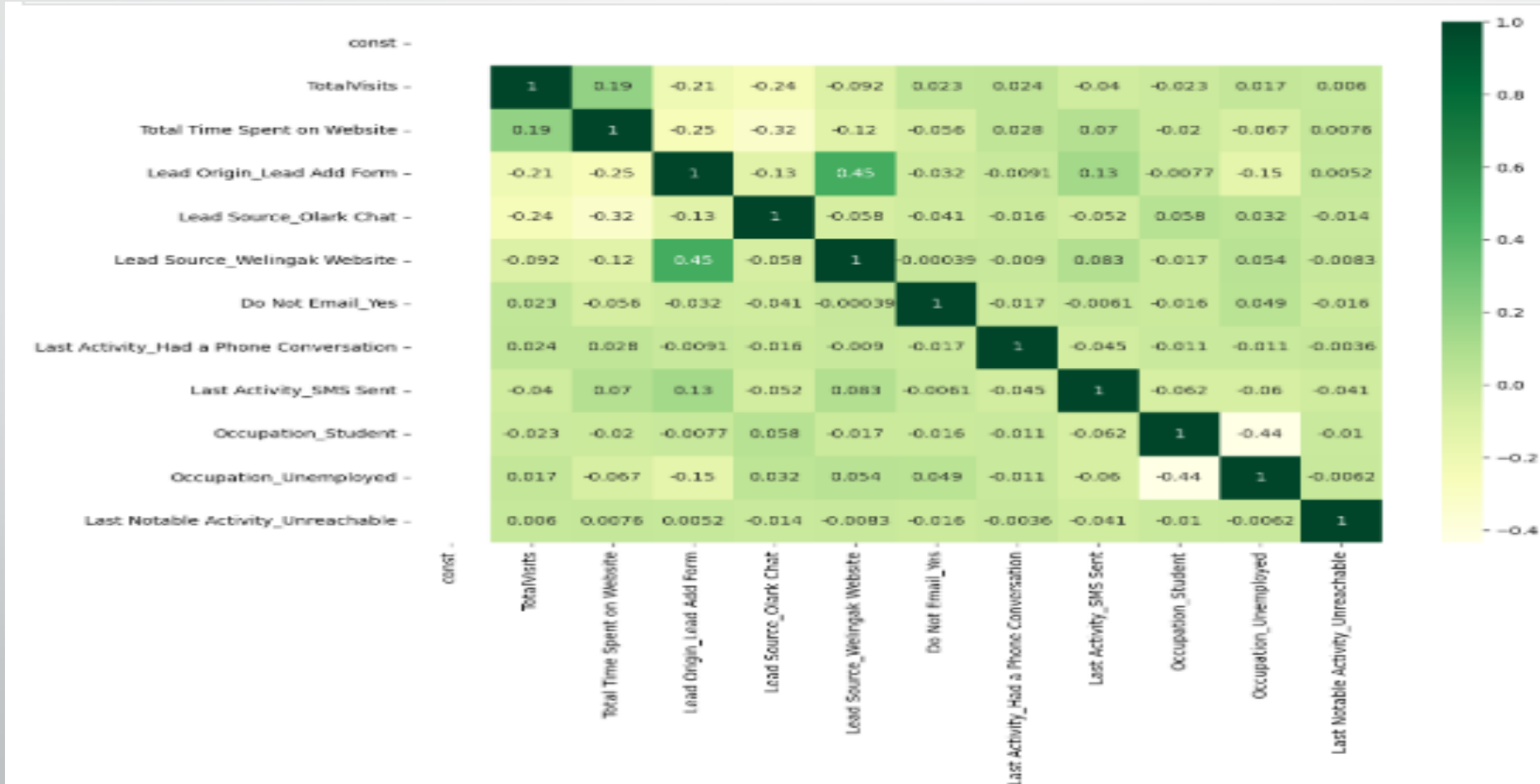
Key Observations Contd..

Attributes/Field with high positive coefficient is positively related with conversion rate and the same with high negative values are negatively co-related

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.1243	0.181	0.687	0.492	-0.231	0.479
TotalVisits	11.1552	2.528	4.412	0.000	6.199	16.111
Total Time Spent on Website	4.4982	0.175	25.741	0.000	4.156	4.841
Lead Origin_Lead Add Form	4.1157	0.233	17.659	0.000	3.659	4.572
Lead Source_Olark Chat	1.4903	0.114	13.052	0.000	1.266	1.714
Lead Source_Welingak Website	2.3323	1.031	2.263	0.024	0.312	4.352
Do Not Email_Yes	-1.4056	0.179	-7.867	0.000	-1.756	-1.055
Last Activity_Had a Phone Conversation	2.8622	0.794	3.606	0.000	1.306	4.418
Last Activity_SMS Sent	1.2543	0.077	16.264	0.000	1.103	1.406
Occupation_Student	-2.3066	0.258	-8.925	0.000	-2.813	-1.800
Occupation_Unemployed	-2.5201	0.171	-14.694	0.000	-2.856	-2.184
Last Notable Activity_Unreachable	2.9375	0.797	3.687	0.000	1.376	4.499

Explanation of Results in Business terms

Below are key highlights after building model and interpreting the result

Looking at attributes with high positive coefficient :

1. People visiting website more and spending more time on website are good potential leads
2. Candidates filling forms are good potential lead
3. Leads coming via 'Wellingak Website', followed by 'Olark Chat' should be explored
4. Follow-up with potential lead with whom conversation was done previously via phone followed by via text.

Attributes with negative co-efficient:

1. Put 'Students' and 'Unemployed' people off the target list they are negatively co-related, which means their conversion rate is less and effort can be utilized in a fruitful direction

Thank You!

Submission by:

Ravi Kumar Desai

Aditi Rani

DS Batch 69