Lead - Score Case Study Presentation - DSC43

Kalyani Gupta & Thippesh

Business Problem

- ➔ X Education Company offers online courses to Business professionals
- ➔ X Education gets leads from various sources and groups of people
- ➔ The conversion rate of people interested to actually enrolling in their program is very low.
- ➔ The firm wants to identify the Hot leads, ie. leads with higher possibility of enrolling with their program
- ➔ This would help sales team to target hot leads and add more to the business.
- ➔ Our task is to identify hot leads and help sale grow up by focussing communication with hot leads.

Business Expectation

- ➔ X Education want to find Hot leads for their platform/courses
- ➔ Find sources and analyse attributes that derive and bring more Hot leads
- ➔ Build a model to identify such hot leads

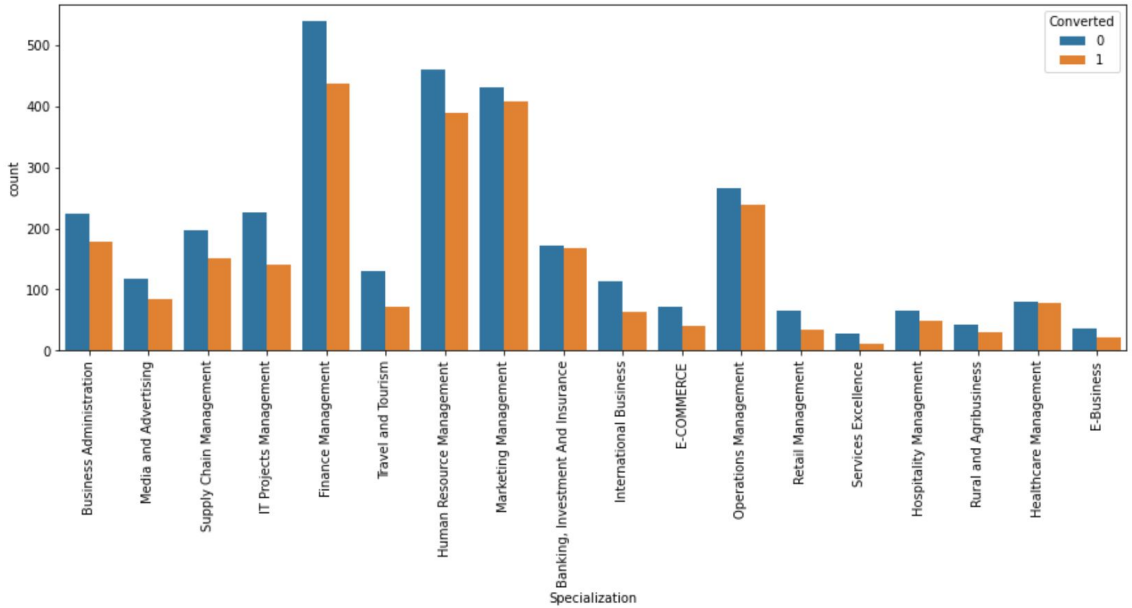
Step Followed

1. Data Analysis
2. Data Clean and Preparation
    a. Replacing 'Select' default value as null
    b. Removing columns with more than 45% null value
    c. Handle remaining null values
    d. Drop Columns that do not add to insight of the model, single value dominant fields
    e. Handle outliers in data
3. EDA
    a. Categorical fields analysis with conversion rate
    b. Numerical Field range with conversion rate
4. Creating dummies for categorical fields
5. Test and Train split data
6. Running Logistic Regression model
7. Testing the final model on test data set
8. Validation of model
9. Conclusion and Recommendations

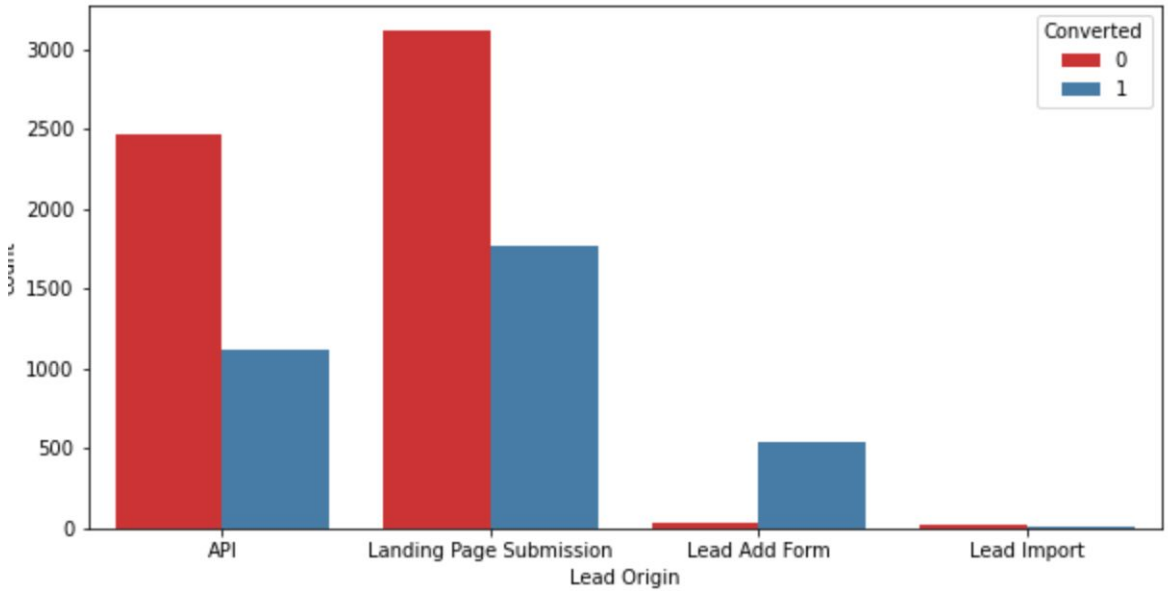
Data Cleaning and Manipulation

1. Replacing 'Select' ie. not specified to null
2. Removing >45% null columns like 'How did you hear about X Education' and 'Lead Profile'
3. Grouping small sets to a single set eg. Country into India, Outside India
4. Remove data set keys like Lead Number and Prospect id which are not use for analysis
5. Remove Columns which are single value dominant and no insight provided to the model for example: 'Search', 'Magazine', 'Newspaper Article', 'X Education Forums', 'Newspaper', 'Digital Advertisement', 'Through Recommendations',
6. Lastly Removing rows with null value considering the set is small and doesn't impact our analysis

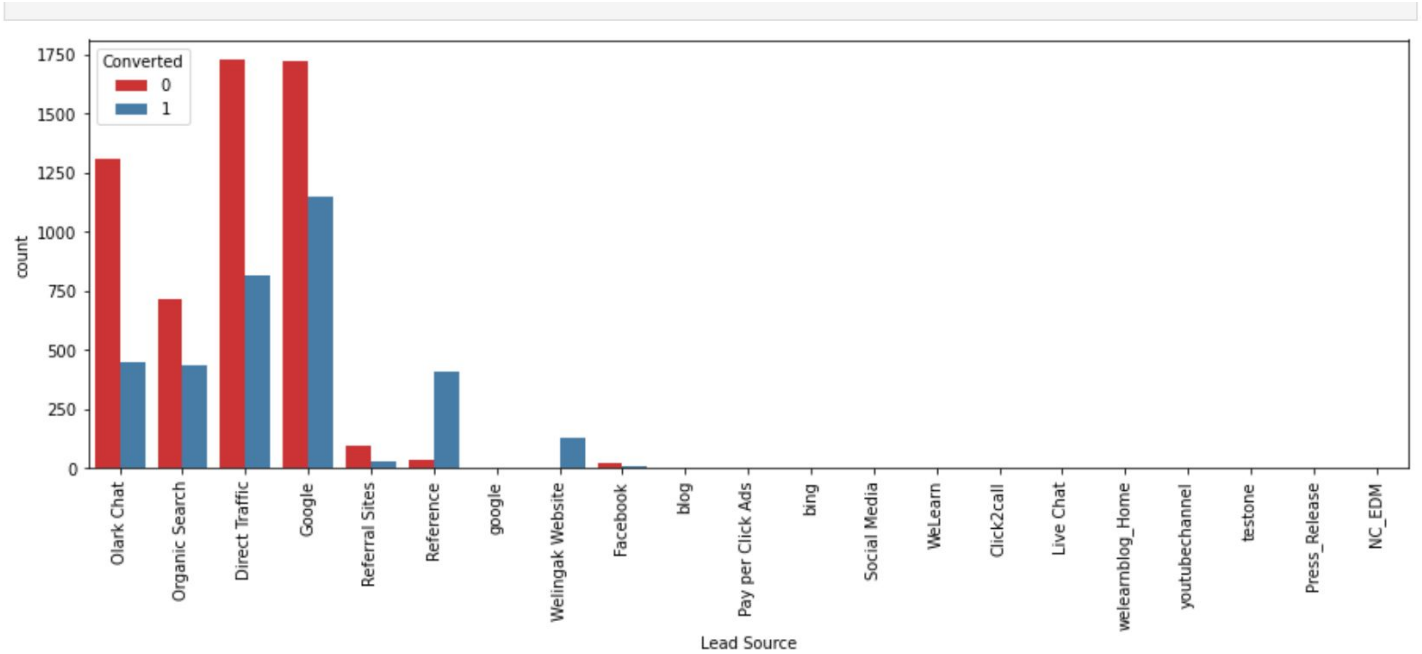
Graphs conversion trend



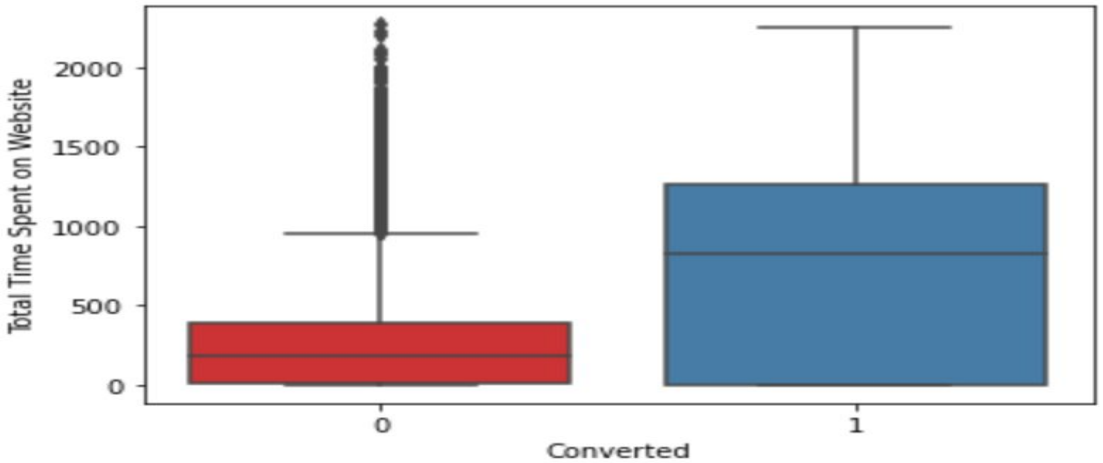
Finance, HR and Marketing Management are more likely to take the course, higher conversion percentage



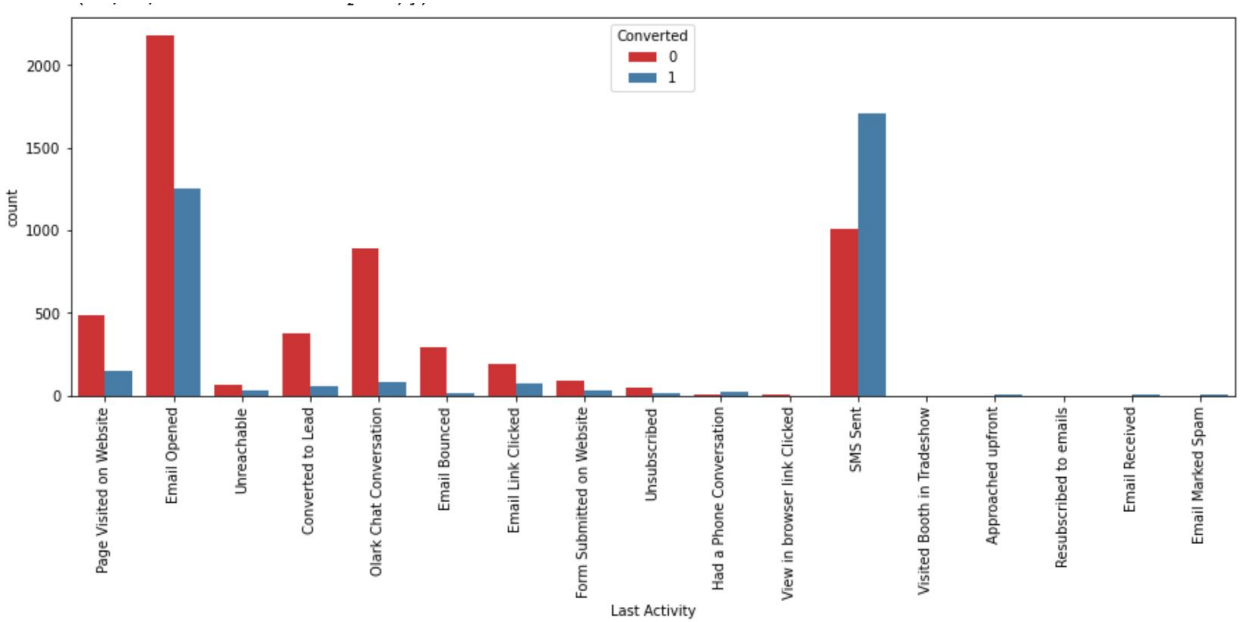
To improve overall lead conversion rate, focus more on improving lead converion of API and Landing Page Submission origin.



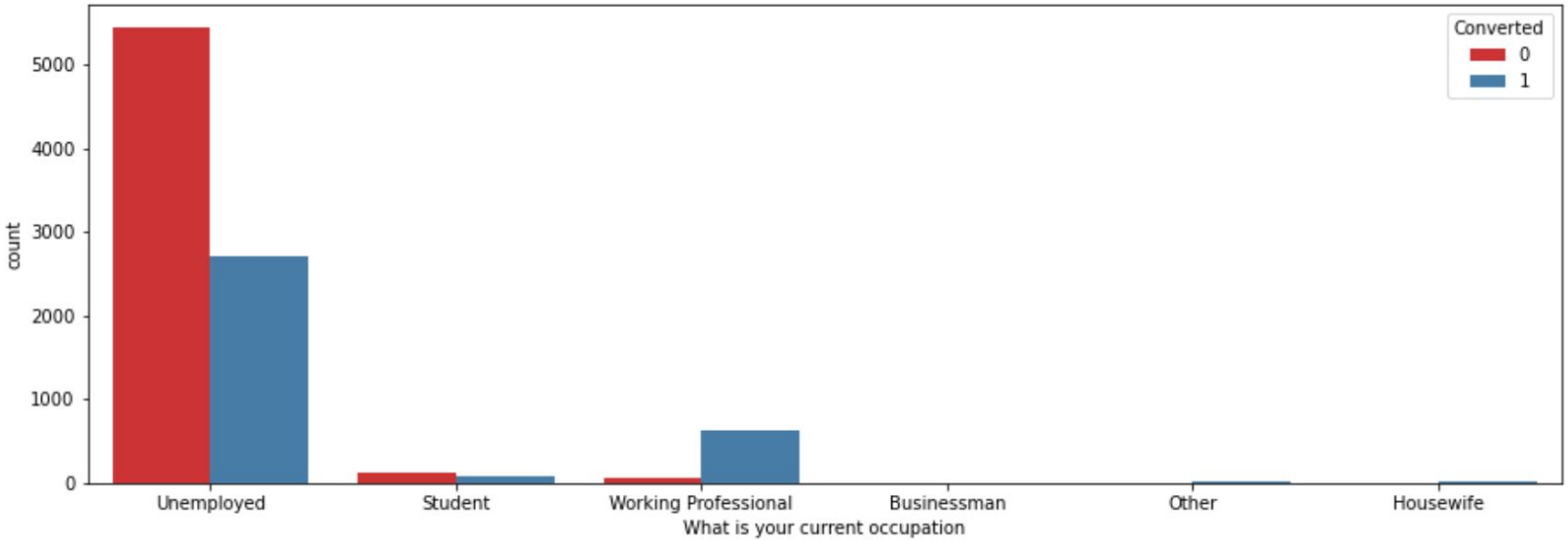
To improve overall lead conversion rate, focus more on improving lead conversion of Reference lead source. Highest probability of enrolling in course



To improve overall lead conversion rate, focus more on users who are spending more time of the website



Higher conversion rate is noticed via SMS connecting method. and also Email opened are have shown significant conversion rate. Where as connecting via Qlark Chat Conversation has least like to convert, and should not be our target

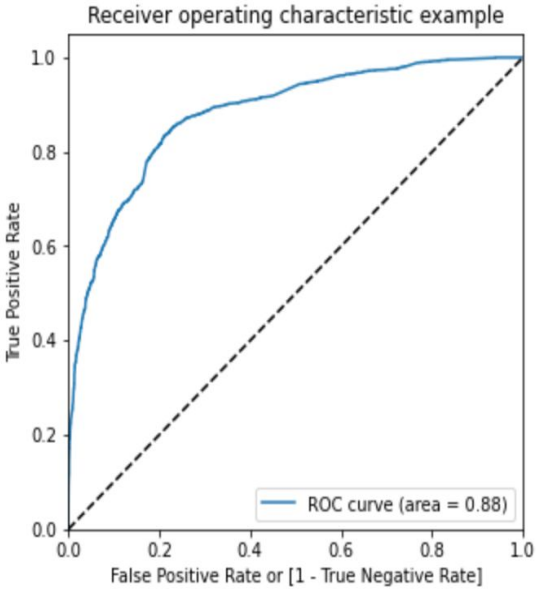


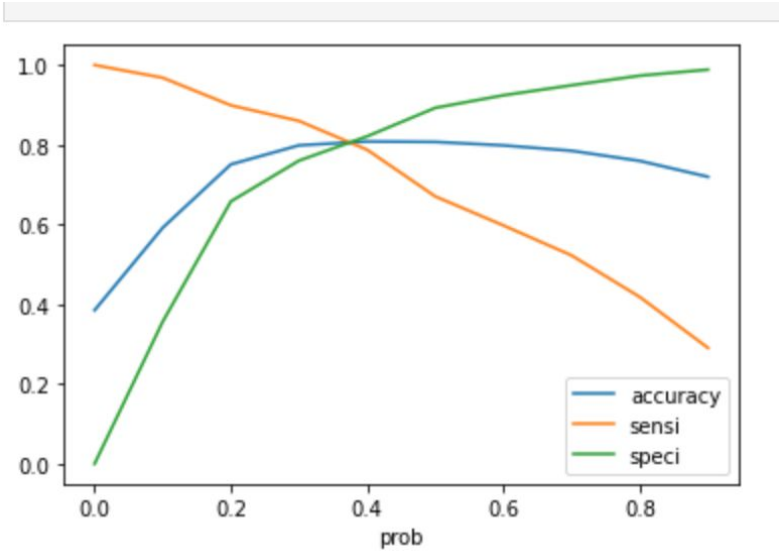
Unemployed and Working professionals have good conversions, in which most of working professionals get converted more likely

Model Built

1. Logistic Regression
2. Running RFE on 20 attributes/columns
3. Removing fields with high p value i.e >0.05 and high VIF i.e >5 to get to final model fields
4. Overall accuracy came out to be 79-80%

ROC curve





Cut off found is 0.38 from above

Conclusion

Top features X Education sales team should focus on deciding to following potential buyers of the online courses are :

1. Contact Lead Origin is from Lead Add Form
2. Contact Occupation is Working Professional
3. Contact Lead Source is Welingak Website
4. Contact People who spend more time on website
5. Avoid Users who opted for Do not email
6. Avoid people from Olark Chat