

# Lead Score Case Study

---

JAINET

SOURAV

NILESH



# Problem Statement

---

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

# Business Objective

---

Knowing the most promising leads

Build a model that identifies the hot leads

Deployment of the model for the future use

# Solution Methodology

---

1. Data Cleaning and manipulation
- 2.EDA
- 3.Feature scaling and dummy variable creation
- 4.Classification technique
- 5.Validation of the model
- 6.Model Presentation
- 7.Conclusions

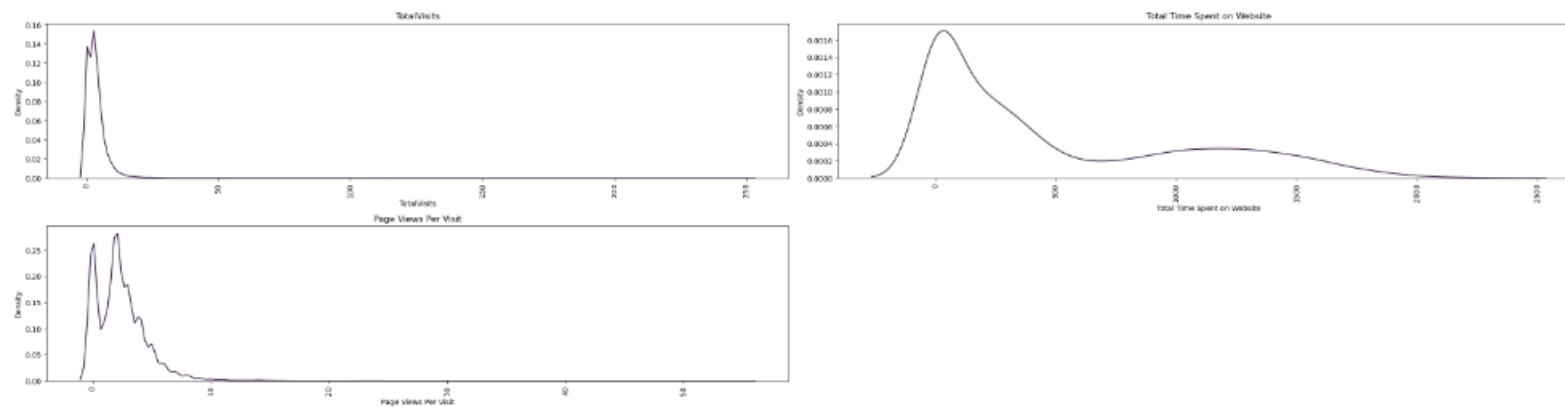
# Data Manipulation

---

1. Total number of rows = 37 , Total number of columns = 9240,
2. Not so important columns have been dropped
3. Removing columns which are not required for the analysis
4. Features which do not have enough variance have been dropped.
5. Dropping columns having more than 35% missing values

# EDA

---



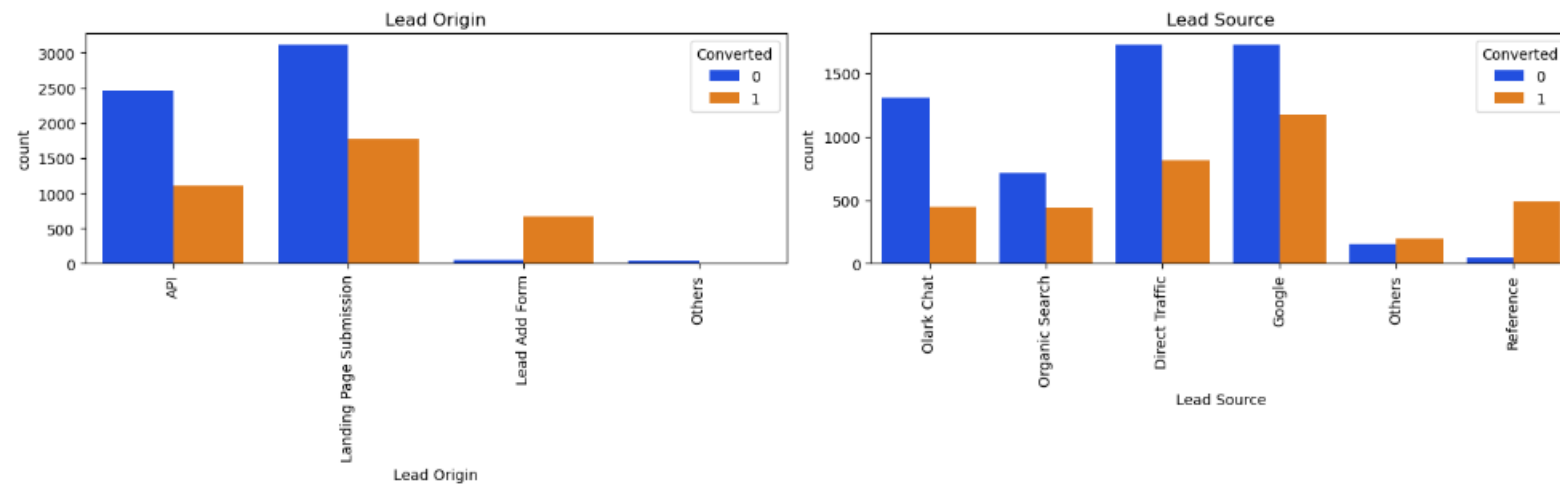
## Observation :

### Uni-variate Analysis - Numerical values:

- The max probability for TotalVisits is found to be around 15-20. It increases initially but decreases further.
- The max probability for PageViewsPerVisit is found to be around to be 3-5
- The probability of time spent is found to be high for time between 0-300 seconds and decreases further.

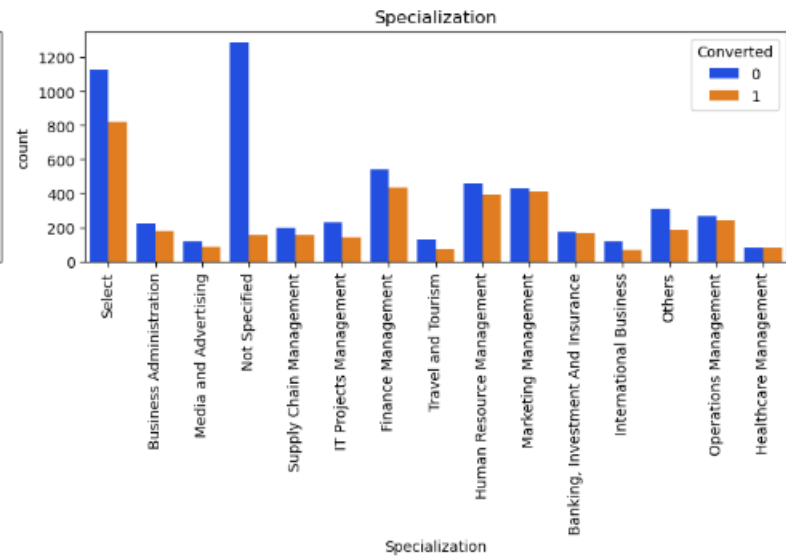
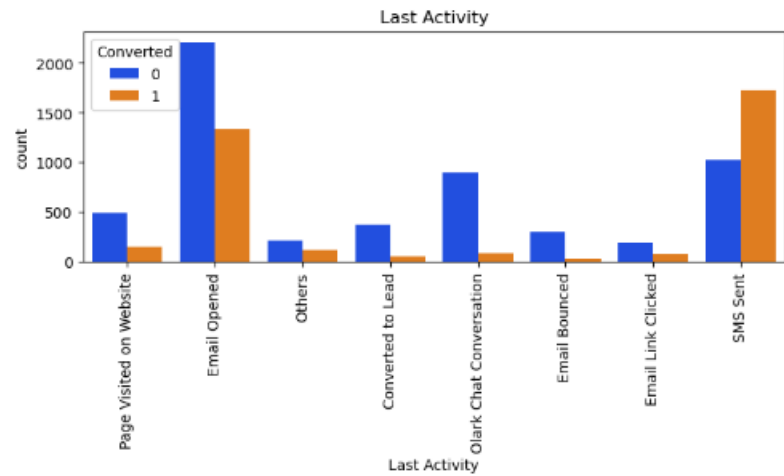
# EDA

---



# EDA

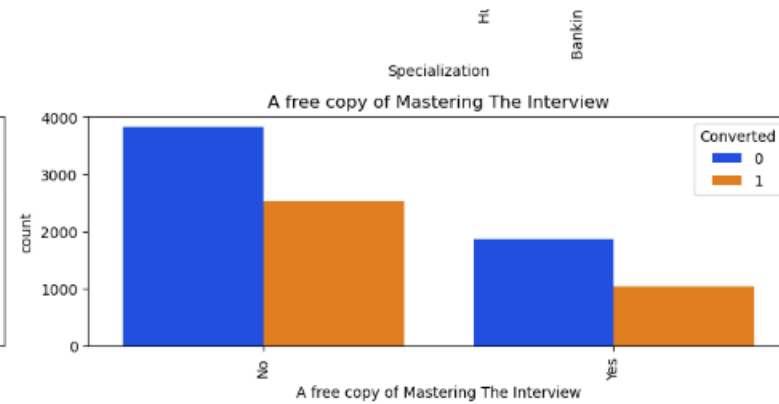
---





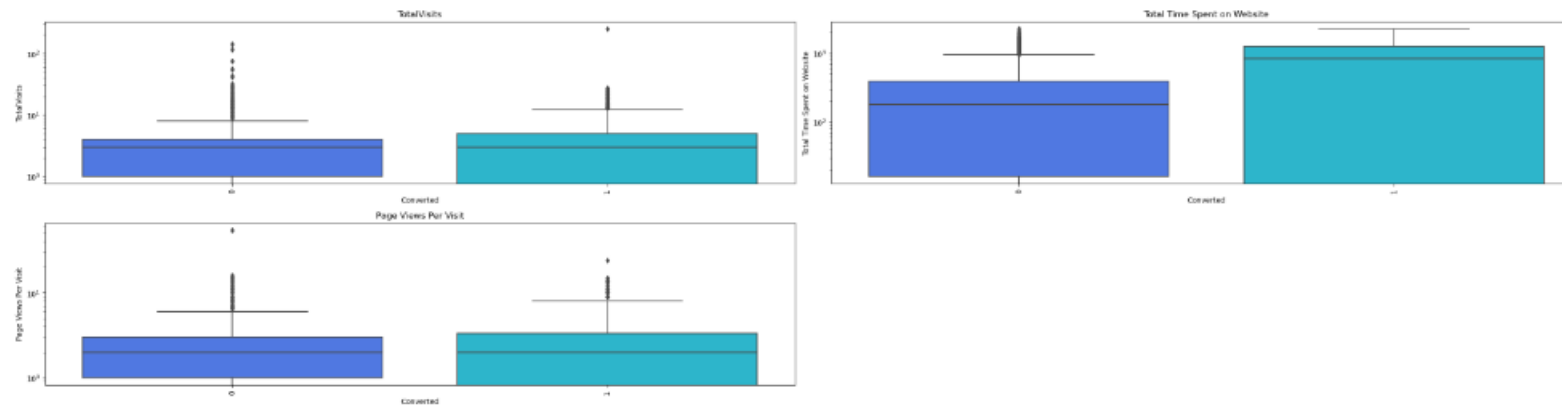
# EDA

---



# EDA

---



# DATA Conversion

---

- ❖ Numerical Variables are normalised
- ❖ Dummy variables are created for object type variables
- ❖ Total rows for analysis 8792
- ❖ Total column for analysis 43

# MODEL Building

---

- ❑ Splitting the dataset into training and test dataset
- ❑ The first basic steps for regression is performing a train test split
- ❑ Use RFE for feature for Feature selection
- ❑ Running RFE with 15 variables as output
- ❑ Building Model by removing the variables whose p-values are greater than 0.5 and vif greater than 5
- ❑ Prediction on test data set
- ❑ Then measure the overall accuracy

# Conclusion

---

- **Conversion Rate for hot leads is increases from 73% to 96%. This means they have a 96% probability of getting converted to a lead.**
- **Focusing on Hot Leads will increase the chances of obtaining more value to the business as the number of people we contact are less but the conversion rate is high.**

**From our model, we can conclude following points :**

- The customer/leads who fills the form are the potential leads.
- We must majorly focus on working professionals.
- We must majorly focus on leads whose last activity is SMS sent or Email opened.
- It's always good to focus on customers, who have spent significant time on our website.
- It's better to focus least on customers to whom the sent mail is bounced back.
- If the lead source is referral, he/she may not be the potential lead.
- If the lead didn't fill specialization, he/she may not know what to study and are not right people to target. So, it's better to focus less on such cases.

**We know that the relationship between  $\ln(\text{odds})$  of 'y' and feature variable "X" is much more intuitive and easier to understand.**

# Recommendations

---

- **It's good to collect data often and run the model and get updated with the potential leads. There is a belief that the best time to call your potential leads is within few hours after the lead shows interest in the courses.**
- **Along with phone calls, it's good to mail the leads also to keep them reminding as email is as powerful as cold calling.**
- **Reducing the number of call attempts to 2-4 and increasing the frequency of usage of other media like advertisements in Google, or via emails to keep in touch with the lead will save a lot of time.**
- **Focusing on Hot Leads will increase the chances of obtaining more value to the business as the number of people we contact are less but the conversion rate is high.**

# Acknowledgement

---

Special thanks to Upgrad for the assignment and the tutors and specialists who helped us in achieving to complete this assignment