# Lead Scoring Case Study Data Analysis

## Context/Abstract/Background:

An education company sells online courses to the professionals. The company wants to increase the conversion rate of the leads to customers from 30% to 80% by communicating and providing education on the products.

## Business Scenario:

* Why this data might have collected?
* As an employee of X education company, wants to increase the conversion rate of the leads to customers
* In order to increase profit, they can work on the factors that are reducing the conversion rate.
* As a professional, factors like cost of the course, content of the course, job assistance, reviews by other customers etc. matter to convert them into customer.
* Research and reading on what this data can be about.
  + What factors can contribute to conversion rate.

Cost of the course

Content of the course

Job assistance

reviews by other customers and many more

* Possible use-cases that be addressed using that data.
  + Where your analysis can provide value?
    - Revenue Increase
      * If conversion rate is increased, then it would lead to increase in the revenue.
    - Reduce Revenue losses
      * If the factors that are affecting the conversion rate is addressed, then the leads wouldn’t have to opt for other education companies.

## Business Problem:

* If you have a client or it’s a take away assignment, understand business problem.
* Hypothetically, put yourself as a stakeholder and think of problem that you’re facing and

See if it can be solved by this data. Or you can gain some insight why this problem might be happening?

To understand the possible factors involved in the conversion rate. Getting some insights on what are the main features and if there are any patterns in such occurrences.

Finally, provide some data-driven recommendations to maximize the conversion rate.

Prepare a predictive model to predict conversion rate.

## Data:

There is a file containing the data information:

Leads data:

It includes the details of the professionals those who became leads. Factors like lead score, lead origin, current occupation, specialization etc are recorded.

### Data Dictionary:

Leads data:

1. Prospect ID:
2. This represents a unique ID with which the customer is identified.
3. This is an object
4. This has 730 unique values in the data.
5. Lead Number:
6. This represents a lead number assigned to each lead procured.
7. This is an integer value.
8. Lead Origin:
9. This represents the origin identifier with which the customer was identified to be a lead.
10. This is an object value.
11. This has 4 unique values in the data.
12. Lead Source:
13. This represents the source of the lead.
14. This is an object value.
15. This has 2 unique values in the data.
16. Do Not Email:
17. This represents an indicator variable selected by the customer for email courses.
18. This is an object value.
19. This has 2 unique values in the data.
20. Do Not Call:
21. This represents an indicator variable for calling the customer.
22. This is an object value.
23. This has 2 unique values in the data.
24. Converted:
25. This represents the target variable.
26. This is an integer value.
27. This has 2 unique values in the data.
28. TotalVisits:
29. This represents the total number of visits made by the customer on the website.
30. This is an integer value.
31. Total Time Spent on Website:
32. This represents the total time spent by the customer on the website.
33. This is an integer value.
34. Page Views Per Visit:
35. This represents an average number of pages on the website viewed during the visits.
36. This is a float value.
37. Last Activity:
38. This represents the last activity performed by the customer.
39. This is an object value.
40. Country:
41. This represents the country of the customer.
42. This is an object value.
43. Specialization:
44. This represents the industry domain in which the customer worked before.
45. This is an object value.
46. How did you hear about X Education:
47. This represents the source from which the customer heard about X Education.
48. This is an object value.
49. What is your current occupation:
50. This represents whether the customer is a student, unemployed or employed.
51. This is an object value.
52. What matters most to you in choosing this course:
53. This represents an option selected by the customer indicating what is their main motto behind doing this course.
54. This is an object value.
55. Search, Magazine, Newspaper Article, X Education Forums, Newspaper, Digital Advertisement:
56. These represent whether the customer had seen the ad in any of the listed items.
57. These are object values.
58. They have 2 unique values in the data.
59. Through Recommendations:
60. This indicates whether the customer came in through recommendations.
61. This is an object value.
62. This has 2 unique values in the data.
63. Receive More Updates About Our Courses:
64. This indicates whether the customer chose to receive more updates about the courses.
65. This is an object value.
66. This has 2 unique values in the data.
67. Tags:
68. This represents tags assigned to customers indicating the current status of the lead.
69. This is an object value.
70. Lead Quality:
71. This indicates the quality of lead based on the data and intuition the the employee who has been assigned to the lead.
72. This is an object value.
73. Update me on Supply Chain Content:
74. This indicates whether the customer wants updates on the Supply Chain Content.
75. This is an object value.
76. This has 2 unique values in the data.
77. Get updates on DM Content:
78. This indicates whether the customer wants updates on the DM Content.
79. This is an object value.
80. This has 2 unique values in the data.
81. Lead Profile:
82. This represents a lead level assigned to each customer based on their profile.
83. This is an object value.
84. City:
85. This represents the city of the customer.
86. This is an object value.
87. Asymmetric Activity Index, Asymmetric Profile Index:
88. These are the indices assigned to each customer based on their activity and their profile.
89. This is an object value.
90. Asymmetric Activity Score, Asymmetric Profile Score:
91. These are the scores assigned to each customer based on their activity and their profile.
92. This is an int value.
93. I agree to pay the amount through cheque:
94. This indicates whether the customer has agreed to pay the amount through cheque or not.
95. This is an object value.
96. This has 2 unique values in the data.
97. A free copy of Mastering the Interview:
98. This indicates whether the customer wants a free copy of 'Mastering the Interview' or not.
99. This is an object value.
100. This has 2 unique values in the data.
101. Last Notable Activity:
102. This represents the last notable activity performed by the student.
103. This is an object value.

## Analysis:

### Descriptive Analytics:

To understand what has happened in Data or to provide a summary on the data.

### Exploratory Analytics:

To understand why something happened in Data or to understand relation between events and possible cause of it.

You try to answer question what has happened and factors between them?

Question: Use framework of design thinking to collect all possible questions that you might have.

List down around 10 (sufficient number of questions) different question which adds different perspective.

Visualization

* Why the data was collected?
* Who is your stakeholder or to whom you want to present the data?
* What is the size of your data?
* What kind of data they are numerical or categorical data, date-time element?
* You’ve to judge that your plot gives good clarity and representation to your data?

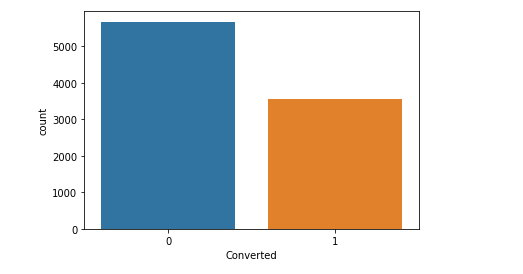
See the 5 points for visualisation and different types of charts that you should use to present your analytics.

Analysis points:

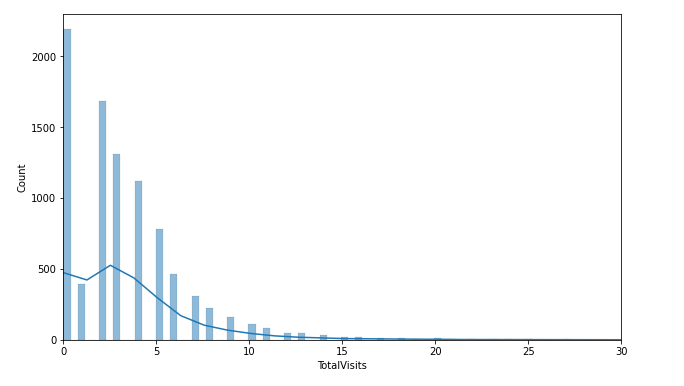
Univariate Analysis

* Solving analytical question and representing them with appropriate charts and plots.

1. What is the conversion rate?



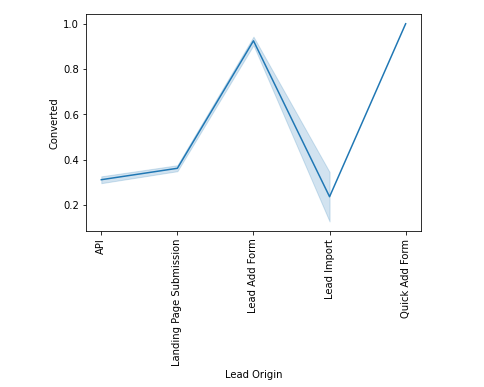
1. What is the maximum and minimum total visits?



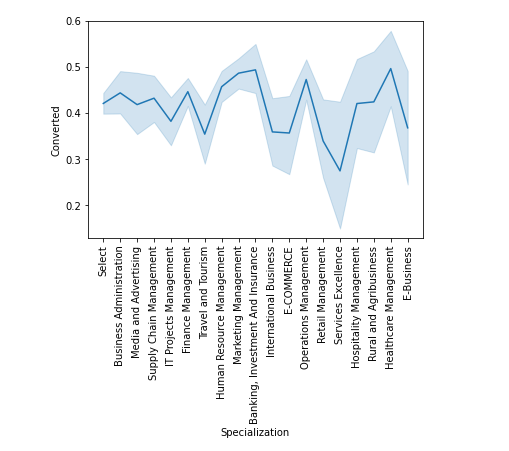
Bi-variate Analysis

* Solving analytical question and representing them with appropriate charts and plots.

1. What is the conversion rate based on lead origin?



1. What is the conversion rate based on specialization?



### Predictive Analytics:

To Predict what will happen in future by understanding occurrences of the past. We create a model based on our data to predict future events.

### Prescriptive Analytics:

To prescribe or to advise what course of action should be taken based on previous Exploratory and Predictive Analytics. Here try to give recommendation based on our insights from data and help client to make data-driven decision.

List down around 10 (sufficient number of questions) different questions which add different perspective.

1. How many bikes are shared on weekdays?
2. How many bikes are shared on weekends?
3. How many bikes are shared on holidays?
4. Which month bikes are shared maximum and minimum?
5. Which season bikes are shared maximum and minimum?
6. How many bikes are shared on a year?
7. Which date the casual users count is maximum and minimum?
8. Which date the registered users count is maximum and minimum?
9. Which day of the week bikes are shared maximum?
10. Which weather demands for more bikes?