Lead Score Case Study

Group Members

Yasaswi Racha
Varun Duggal
Rajat Gupta

Problem Statement

- X Education sells online courses to industry professionals.
- X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.
- To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. Business Objective:
- X education wants to know most promising lead

Problem Statement (contd)

Business Objective:

- X education wants to know most promising leads.

- For that they want to build a Model which identifies the hot leads. Deployment of the model for the future use

Solution Methodology

- Data cleaning and data manipulation.

1. Check and handle duplicate data.
2. Check and handle NA values and missing values.
3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
4. Imputation of the values, if necessary.
5. Check and handle outliers in data.

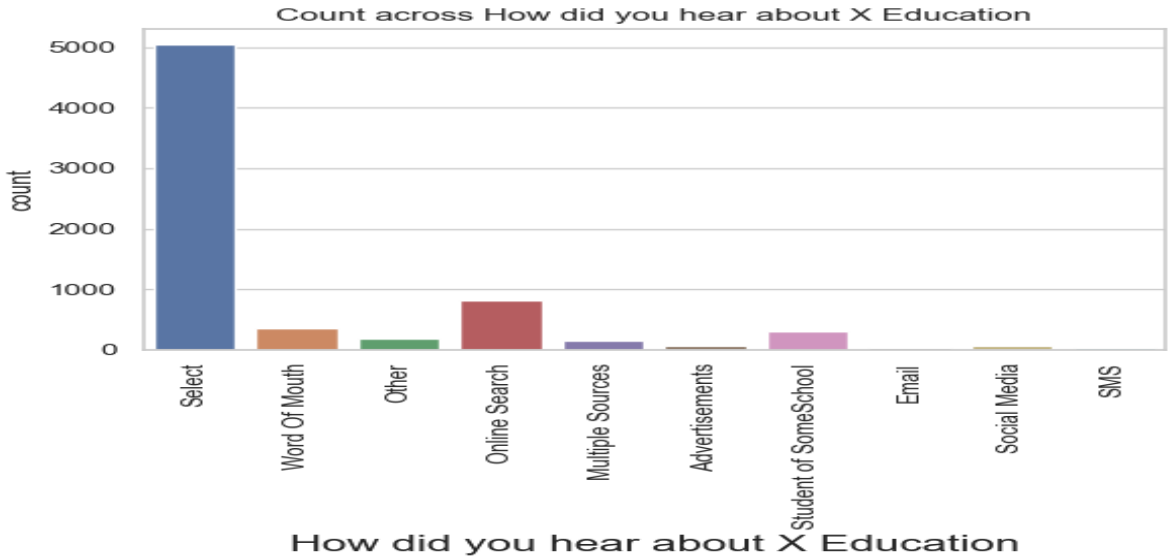
- EDA

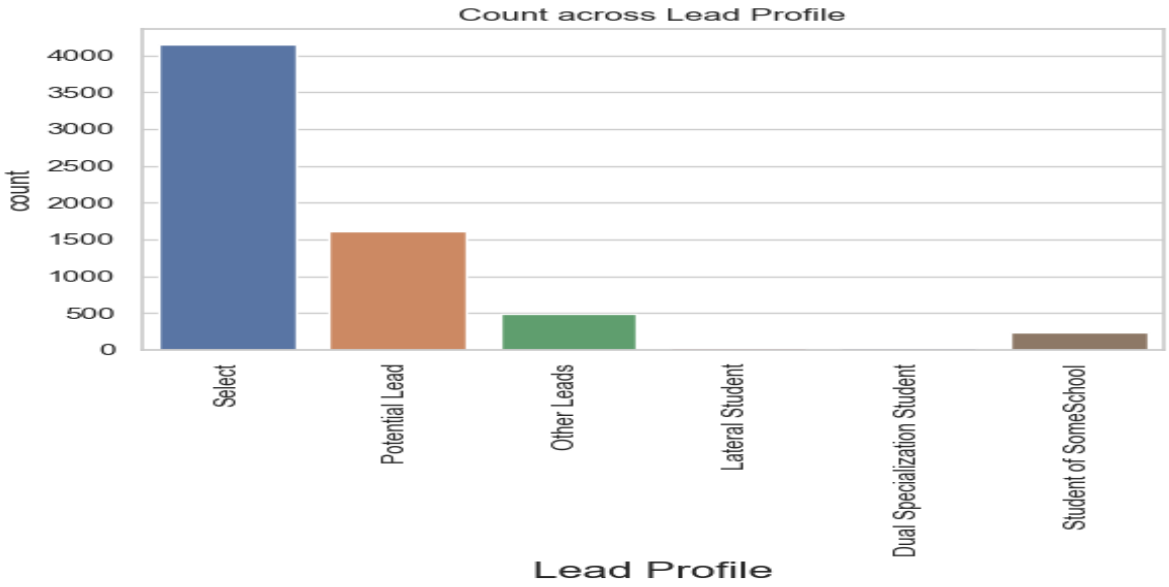
1. Univariate data analysis: value count, distribution of variable etc.
2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.
   - Feature Scaling & Dummy Variables and encoding of the data.
   - Classification technique: logistic regression used for the model making and prediction.
   - Validation of the model. Model presentation.
   - Conclusions and recommendations.

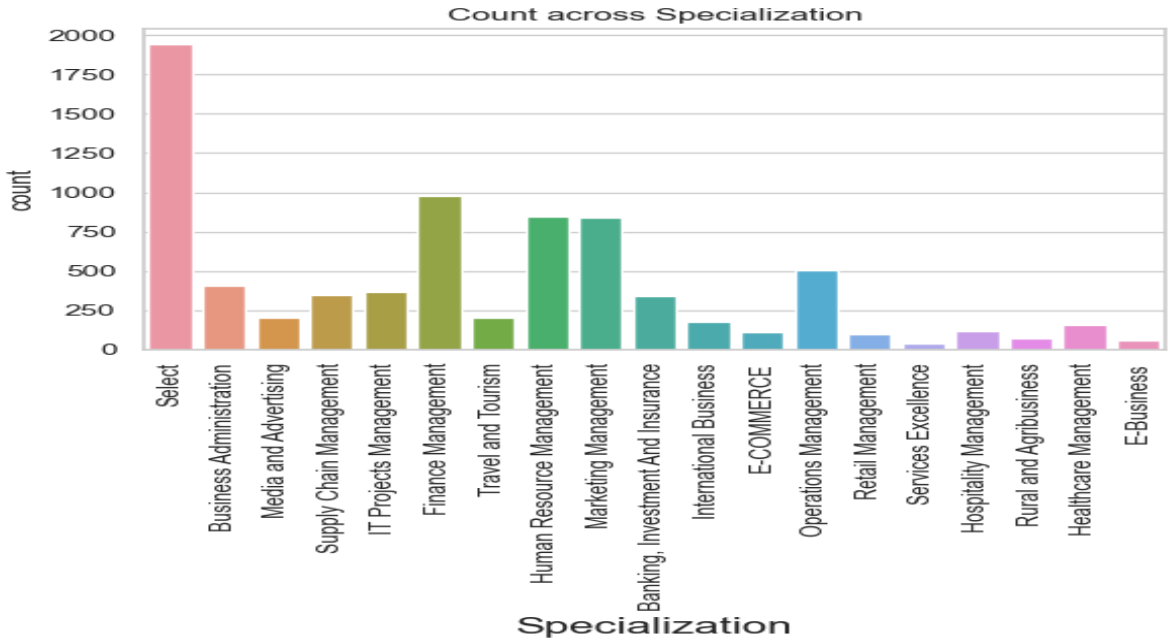
Data Manipulation

- Data Manipulation
  - Total Number of Rows =37
  - Total Number of Columns =9240.
  - Single value features like “Magazine”, “Lead Profile', 'How did you hear about X Education, City, Country, etc. have been dropped.
  - Removing the “Prospect ID” and “Lead Number” which is not necessary for the analysis. After checking for the value counts for some of the object type variables, we find some of the features which has no enough variance, which we have dropped, the features are: Do Not Call', 'Search', 'Magazine', 'Newspaper Article', 'X Education Forums', 'Newspaper', 'Digital Advertisement', 'Through Recommendations', 'Receive More Updates About Our Courses’.

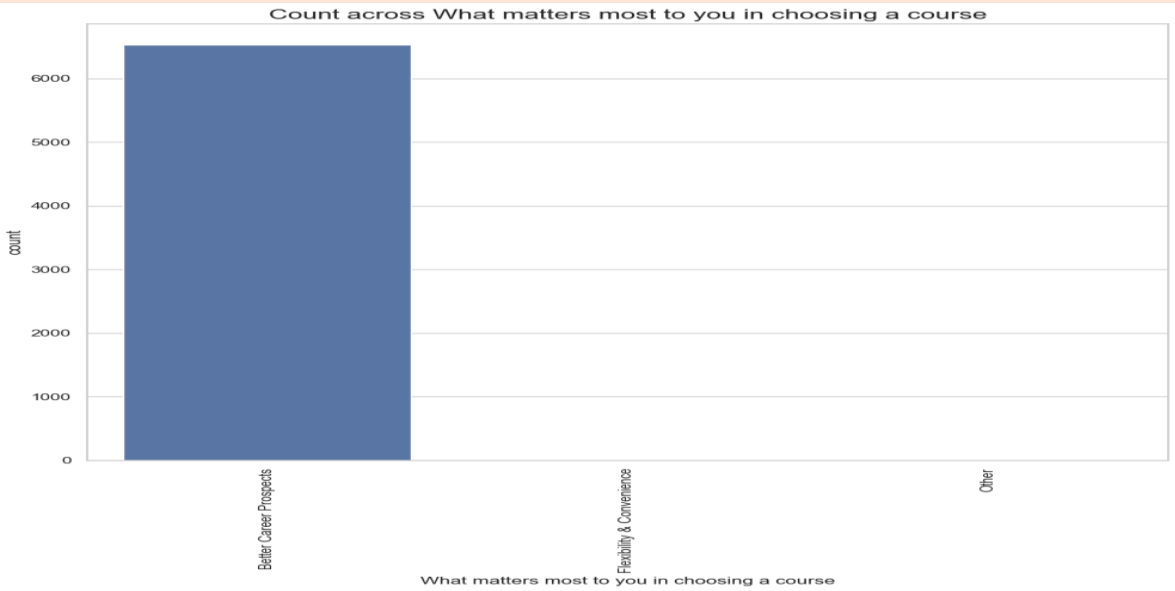
EDA Graphs

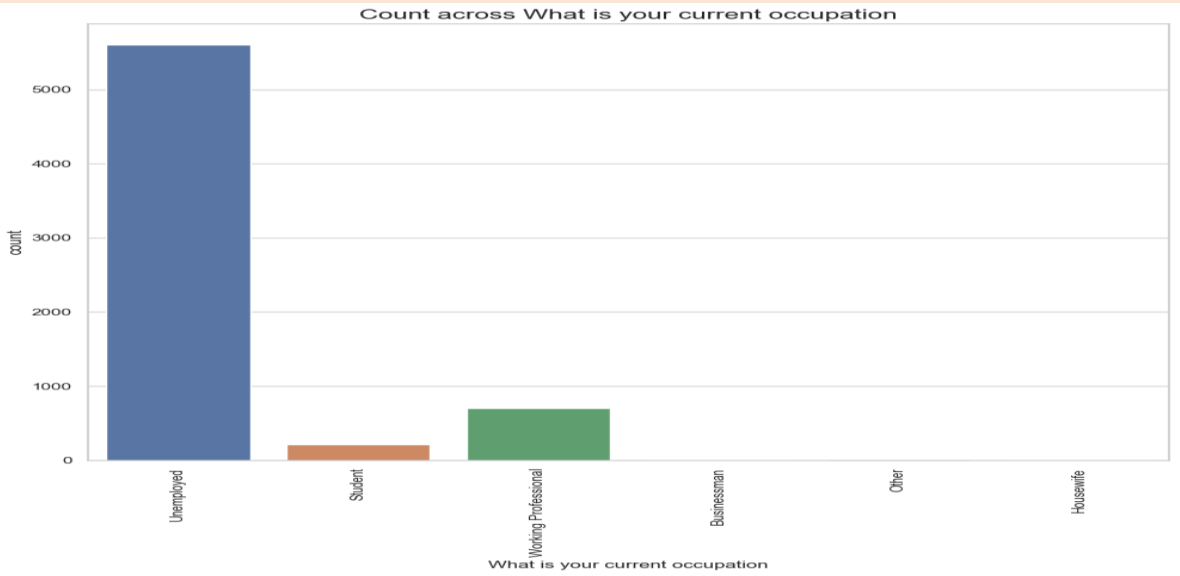

Count across How did you hear about X Education
count
5000
4000
3000
2000
1000
0
Select
Word Of Mouth
Other
Online Search
Multiple Sources
Advertisements
Student of SomeSchool
Email
Social Media
SMS
How did you hear about X Education

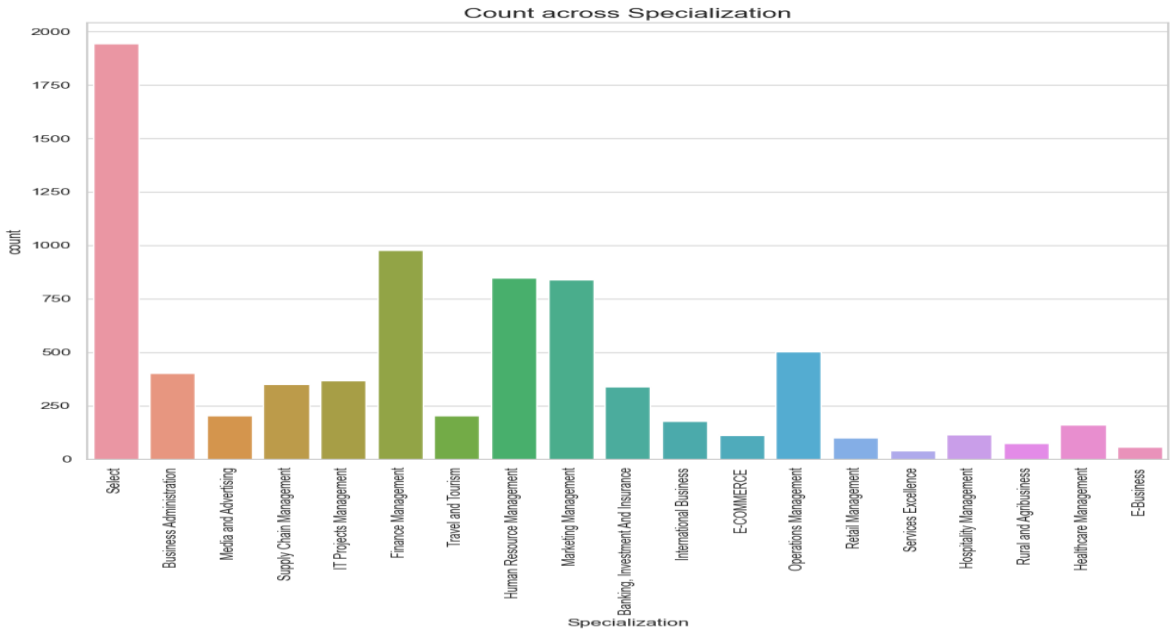

Count across Lead Profile
count
4000
3500
3000
2500
2000
1500
1000
500
0
Select
Potential Lead
Other Leads
Lateral Student
Dual Specialization Student
Student of SomeSchool
Lead Profile


Count across Specialization
count
2000
1750
1500
1250
1000
750
500
250
0
Select
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

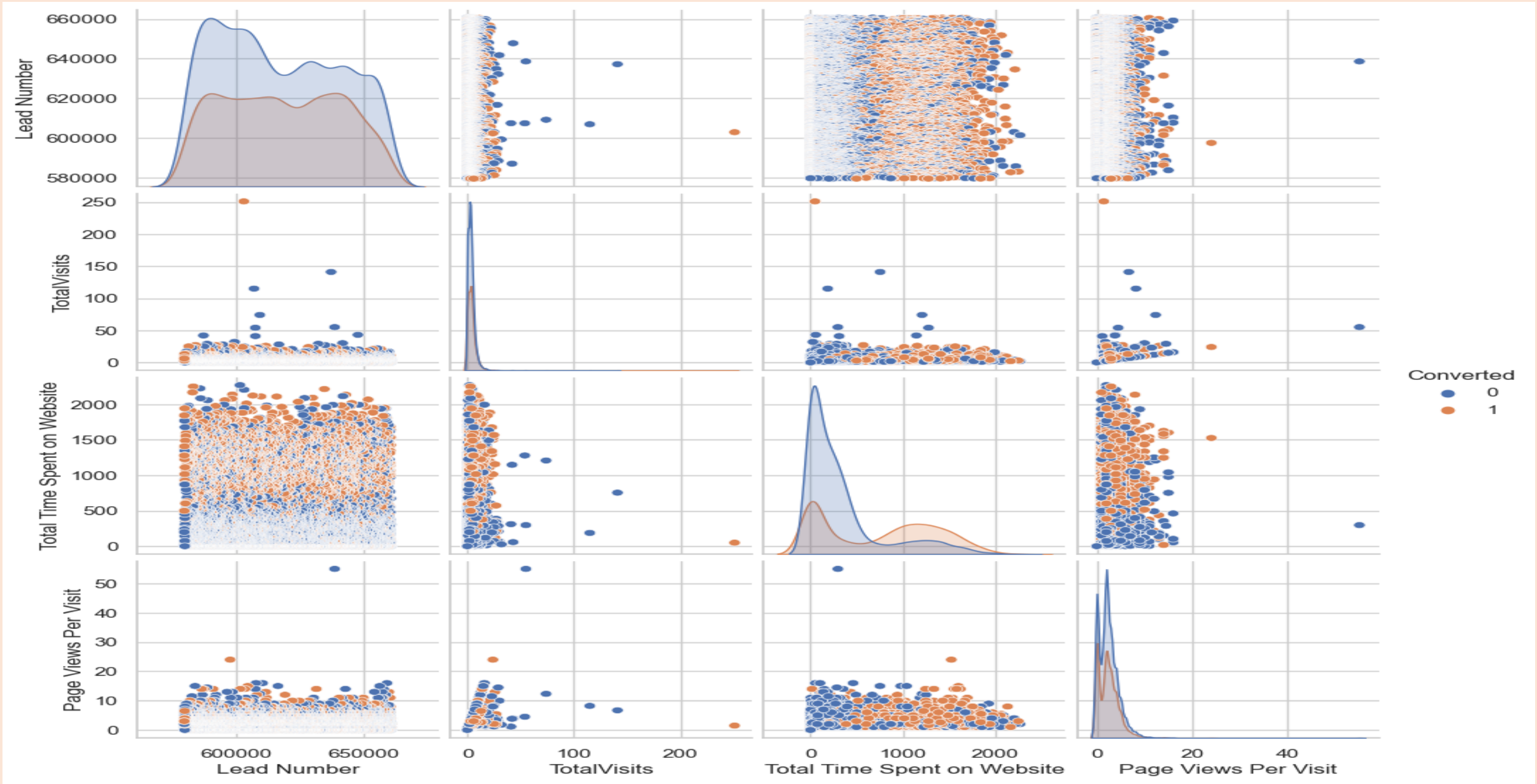
EDA Graphs


Count across What matters most to you in choosing a course
count
0
1000
2000
3000
4000
5000
6000
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course

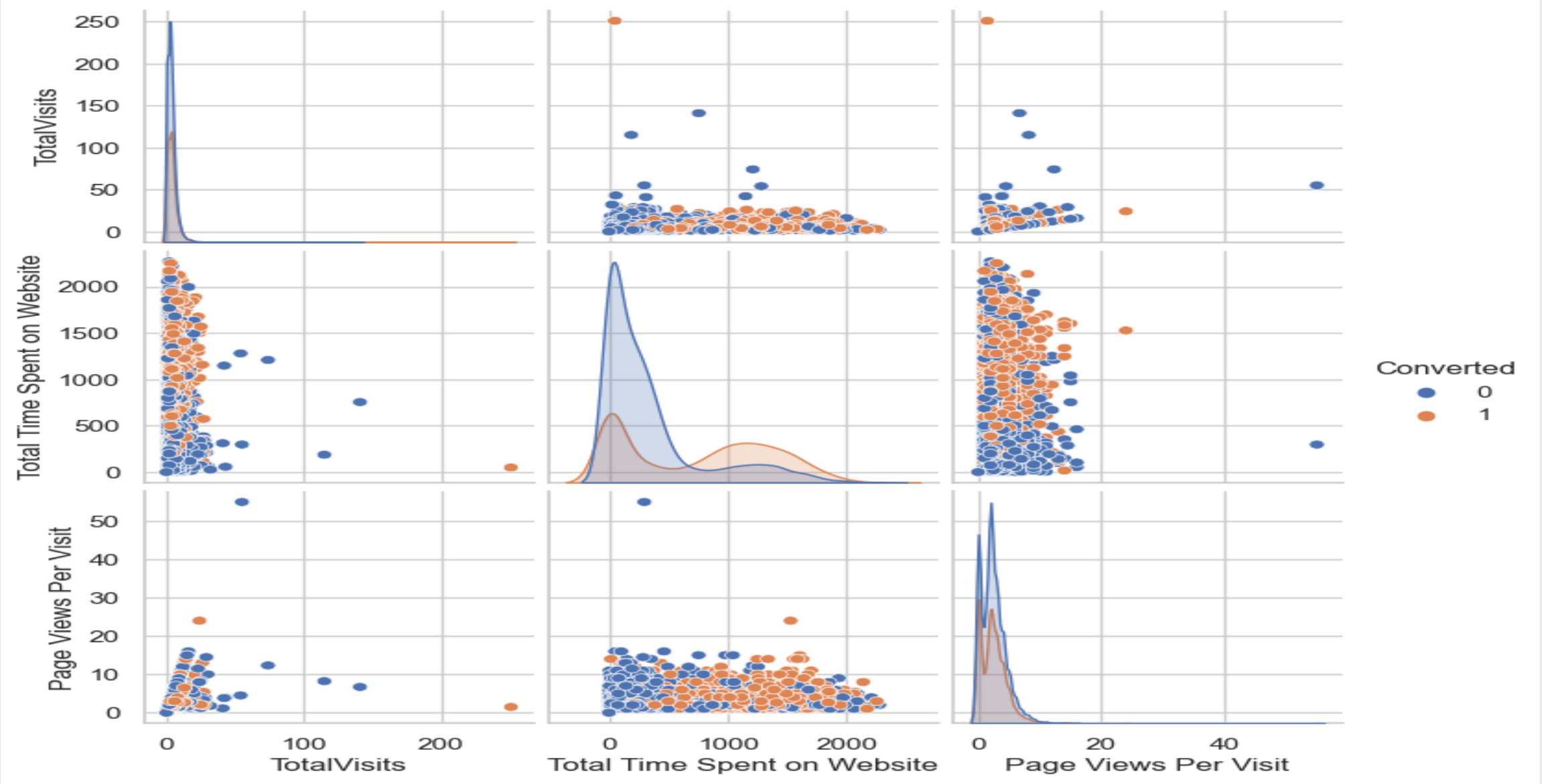

Count across What is your current occupation
count
0
1000
2000
3000
4000
5000
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation


Count across Specialization
count
0
250
500
750
1000
1250
1500
1750
2000
Select
Business Administration
Media and Advertising
Supply Chain Management
IT Projects Management
Finance Management
Travel and Tourism
Human Resource Management
Marketing Management
Banking, Investment And Insurance
International Business
E-COMMERCE
Operations Management
Retail Management
Services Excellence
Hospitality Management
Rural and Agribusiness
Healthcare Management
E-Business
Specialization

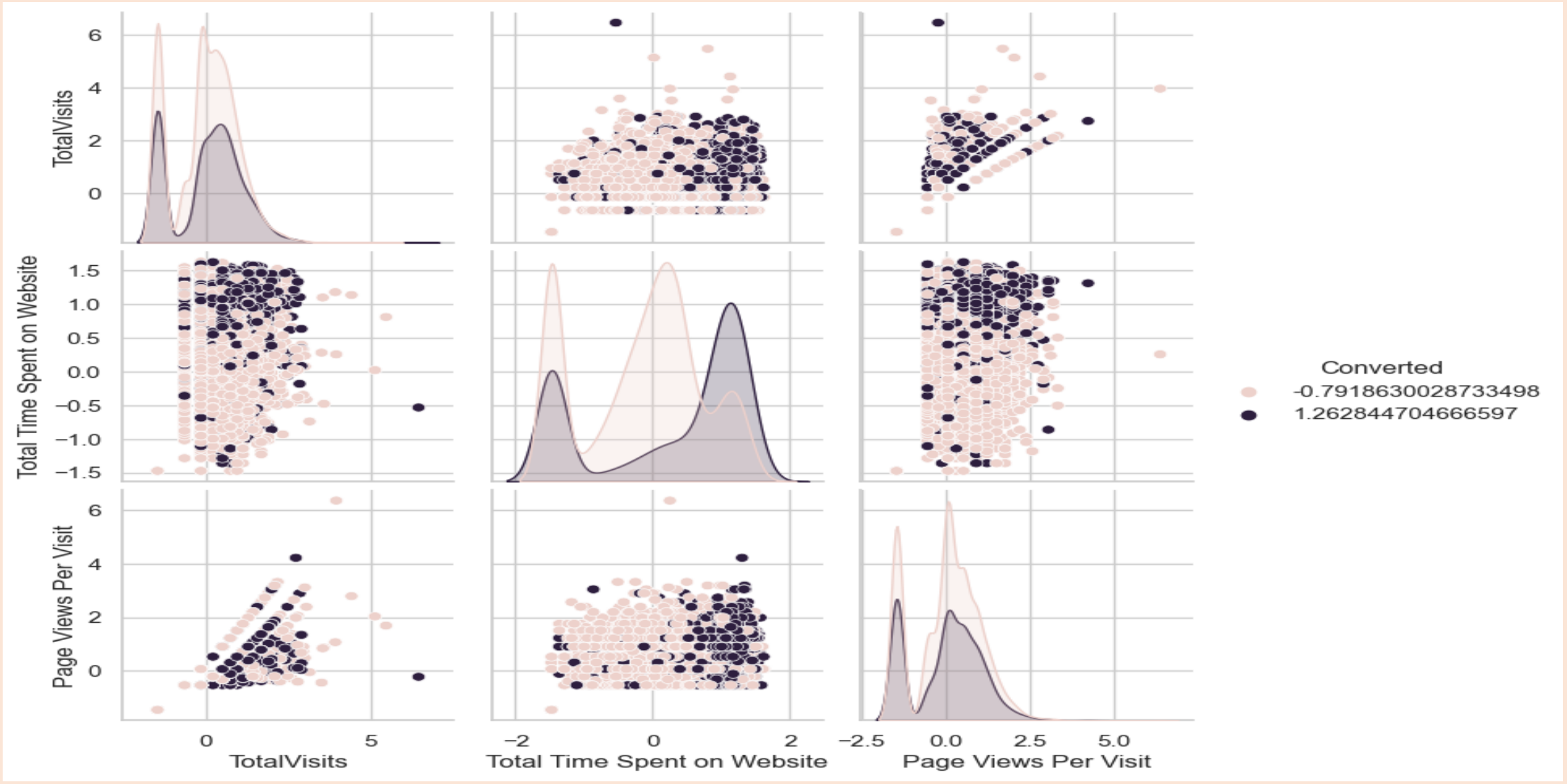
EDA Graphs


Lead Number
660000
640000
620000
600000
580000
TotalVisits
250
200
150
100
50
0
Total Time Spent on Website
2000
1500
1000
500
0
Page Views Per Visit
50
40
30
20
10
0
600000
650000
Lead Number
0
100
200
TotalVisits
0
1000
2000
Total Time Spent on Website
0
20
40
Page Views Per Visit
Converted
0
1

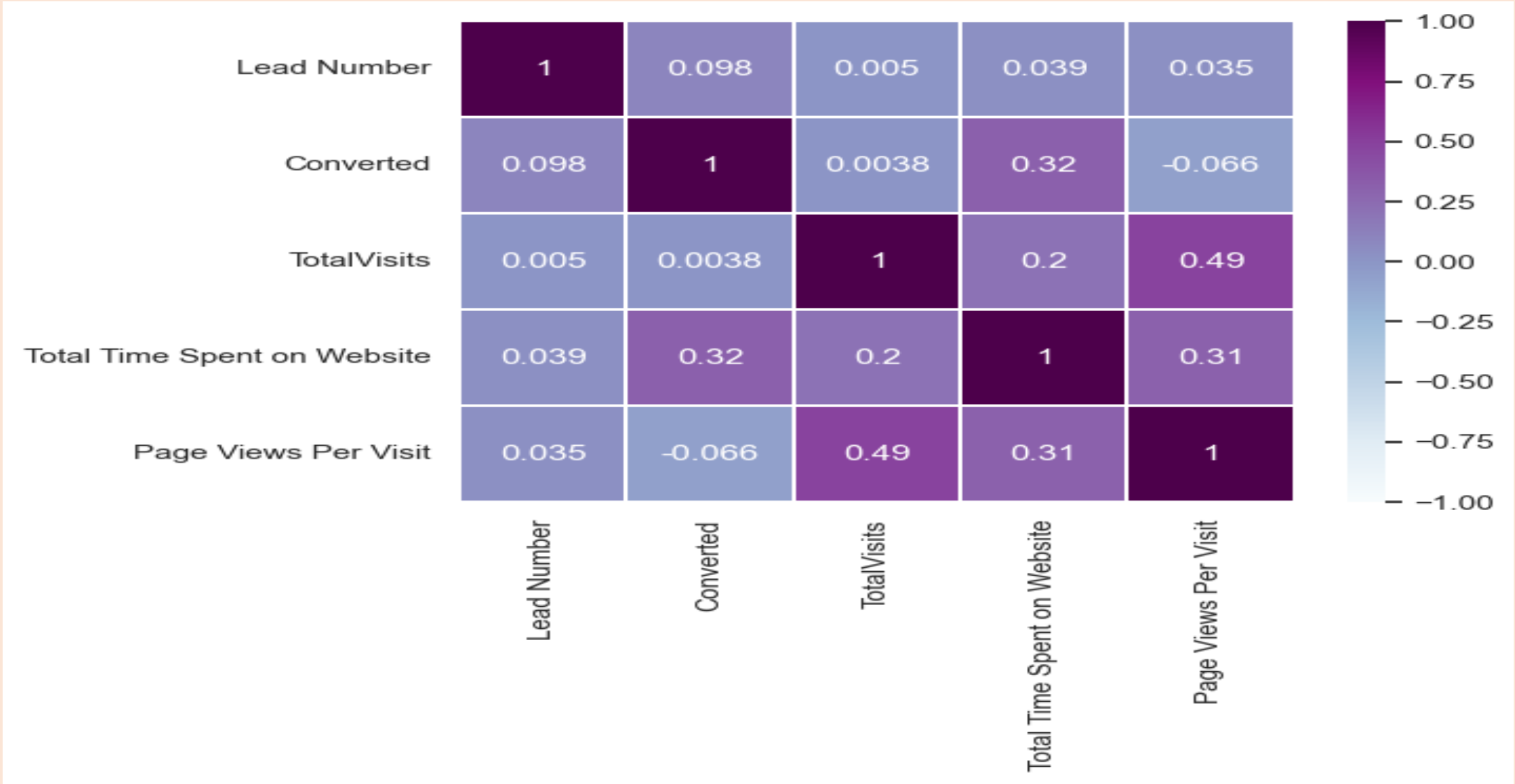
EDA Graphs


TotalVisits
Total Time Spent on Website
Page Views Per Visit
0
100
200
0
1000
2000
0
20
40
Converted
0
1

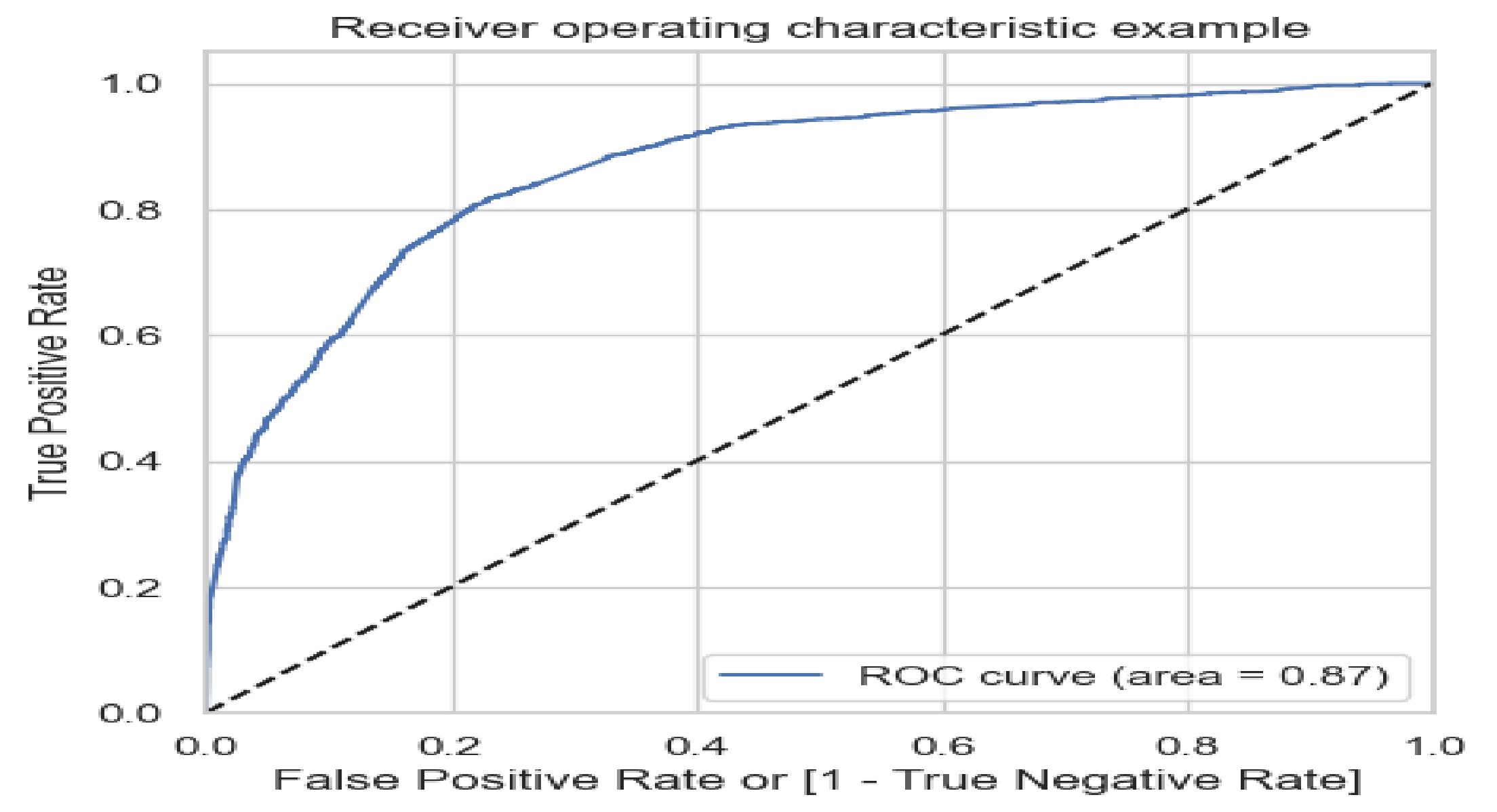
EDA Graphs


TotalVisits
Total Time Spent on Website
Page Views Per Visit
Converted
-0.7918630028733498
1.262844704666597

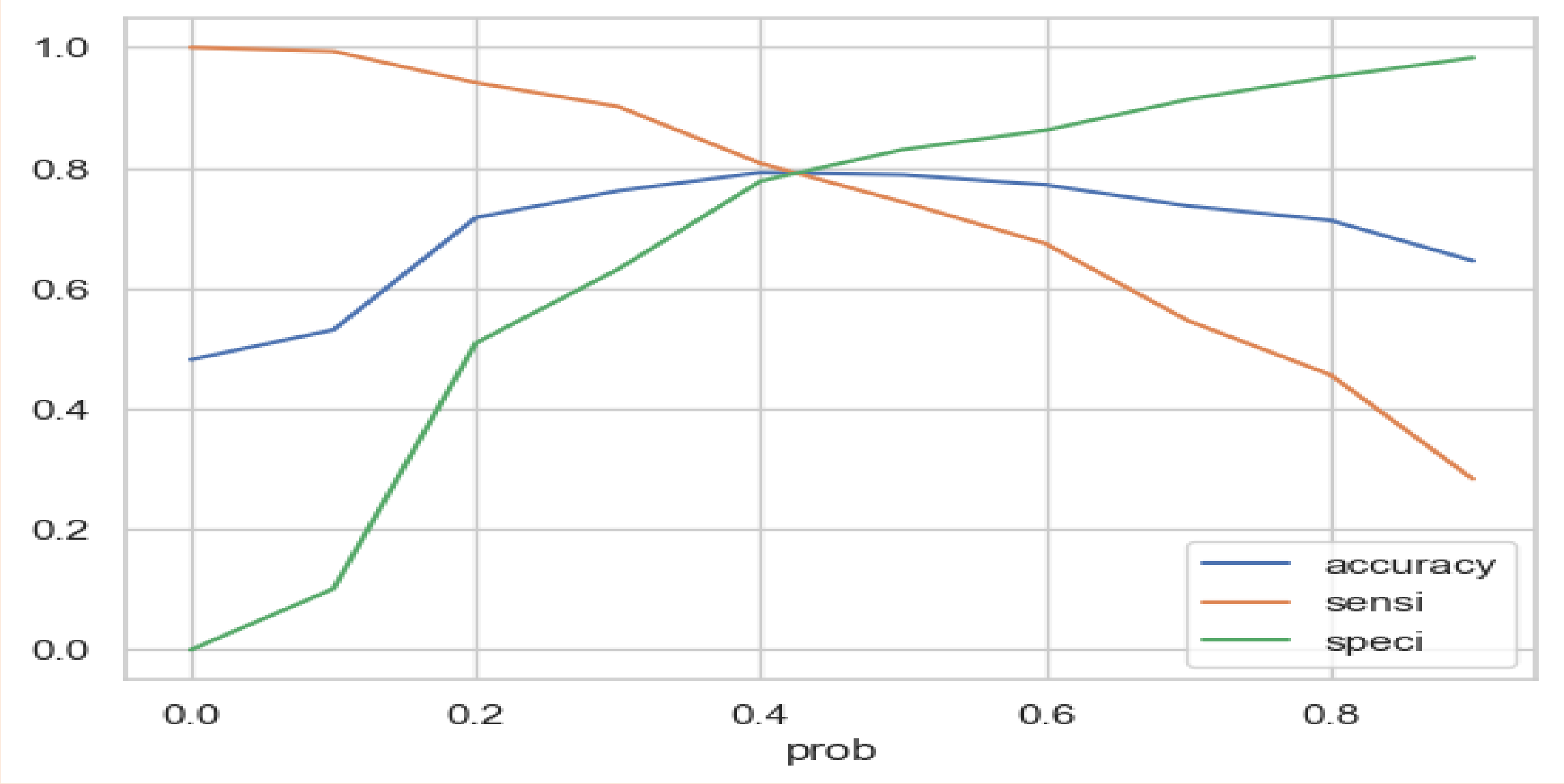
Correlation



ROC Curve Graphs



Accuracy sensitivity and specifivity Graphs



Data Conversion

- Numerical Variables are Normalised
- Dummy Variables are created for object type variables
- Total Rows for Analysis: 6373
- Total Columns for Analysis: 75

Model Building

- Splitting the Data into Training and Testing Sets
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection
- Running RFE with 15 variables as output
- Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5
- Predictions on test data set

Conclusion

- The most numbers of leads are from India and in terms of city highest number are from Mumbai.
- First, sort out the best prospects from the leads you have generated. 'Total Visits' , 'Total Time Spent on Website' , 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted.
- The leads are joined course for Better Career Prospects, most of having Specialization from Finance Management. Leads from HR, Finance & marketing management specializations are high probability to convert.
- Talking to last notable Activity, making improvement in customer engagement through email & calls will help to convert leads. As the leads which are opening email have high probability to convert.