Problem Statement : We have been provided with a leads dataset from the past with around 9000 data points. This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not.

The target variable, in this case, is the column ‘Converted’ which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn’t converted.

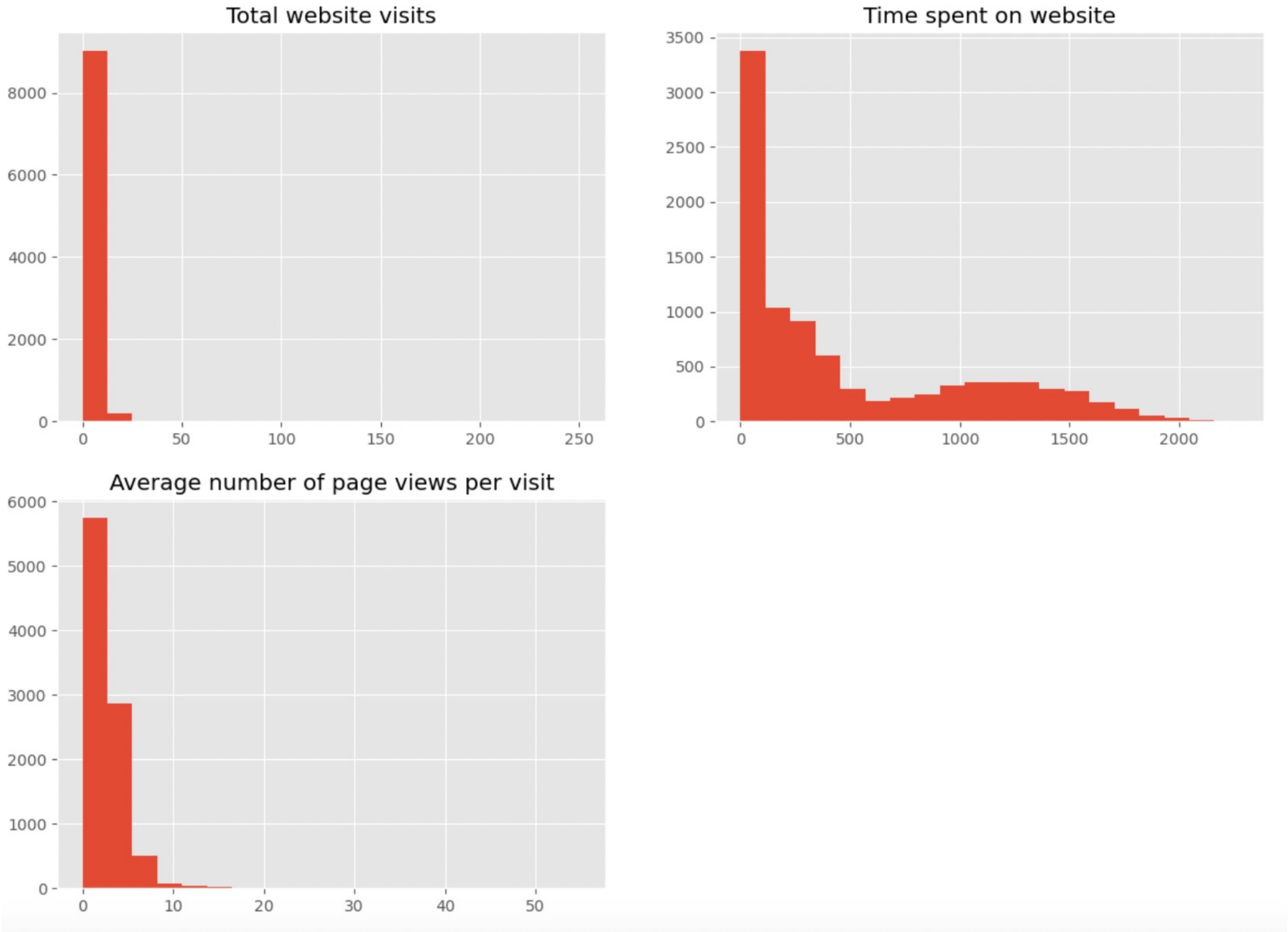
Another thing that we need to check out for are the levels present in the categorical variables. Many of the categorical variables have a level called 'Select' which needs to be handled because it is as good as a null value

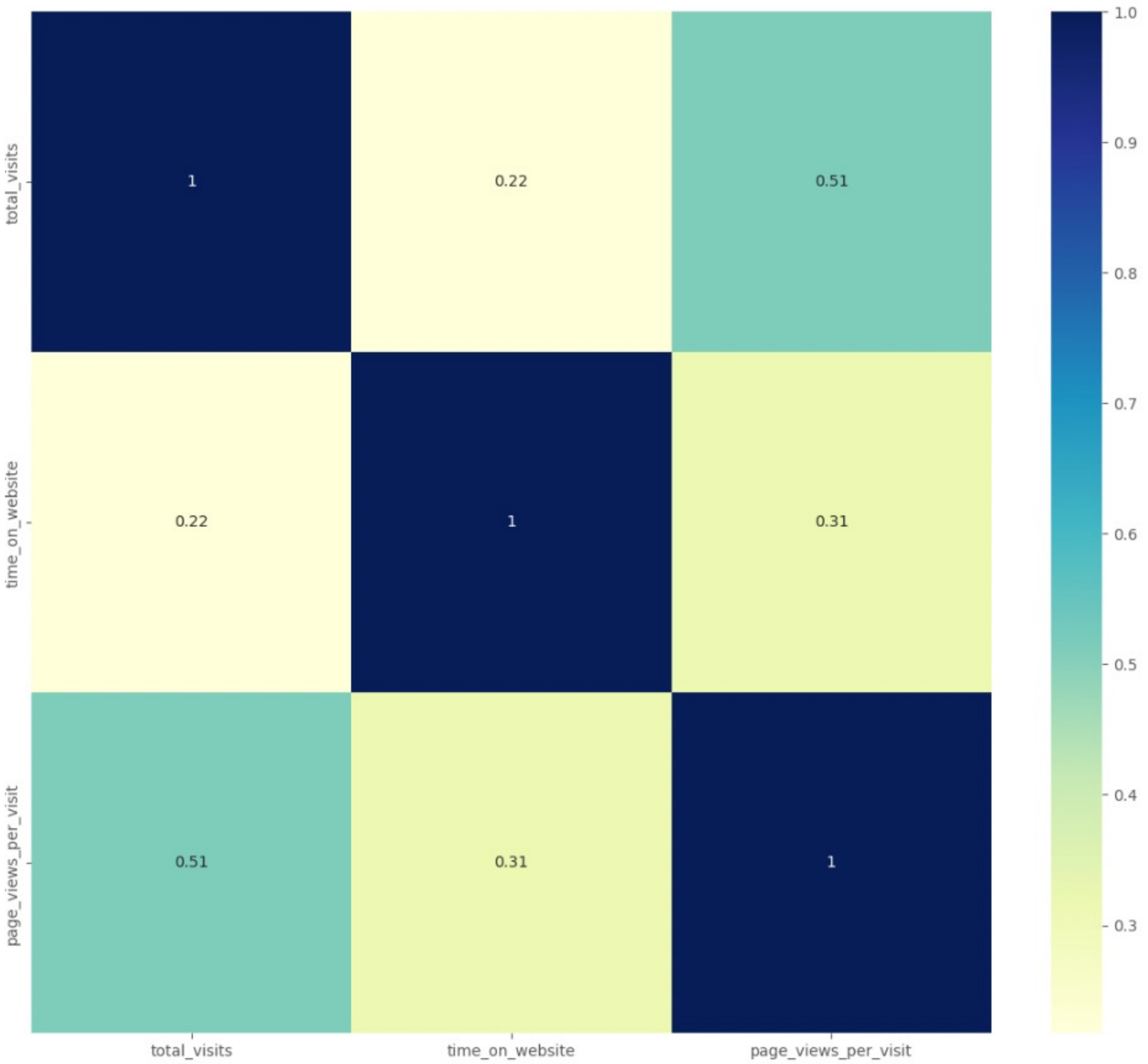
Approach

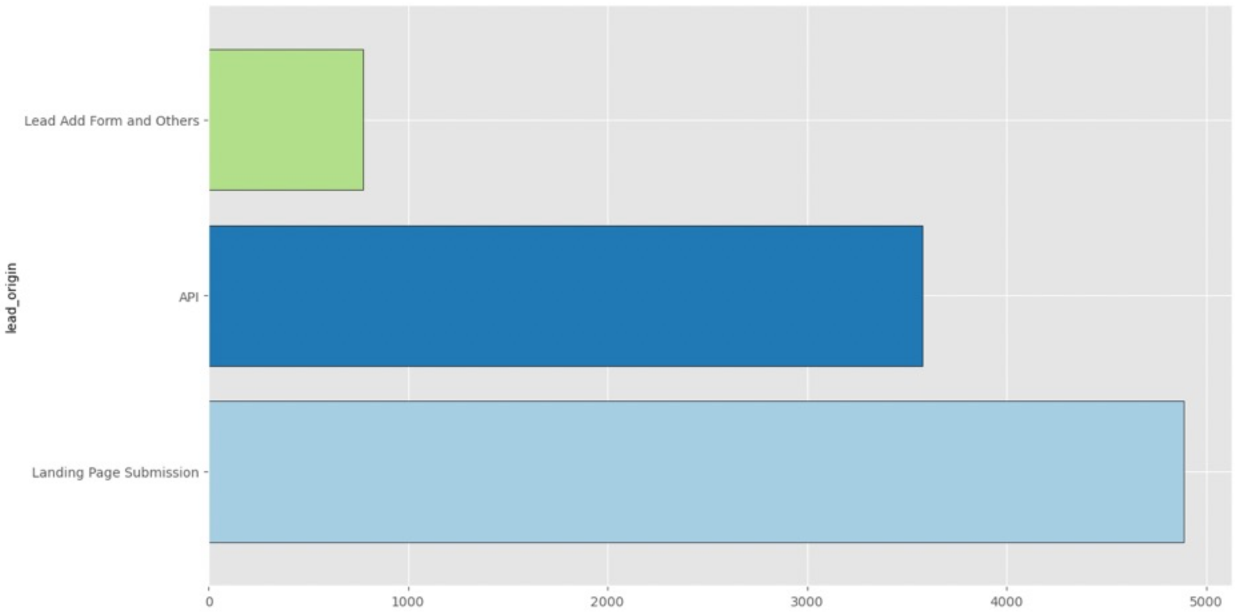
1 Reading and Understanding Data

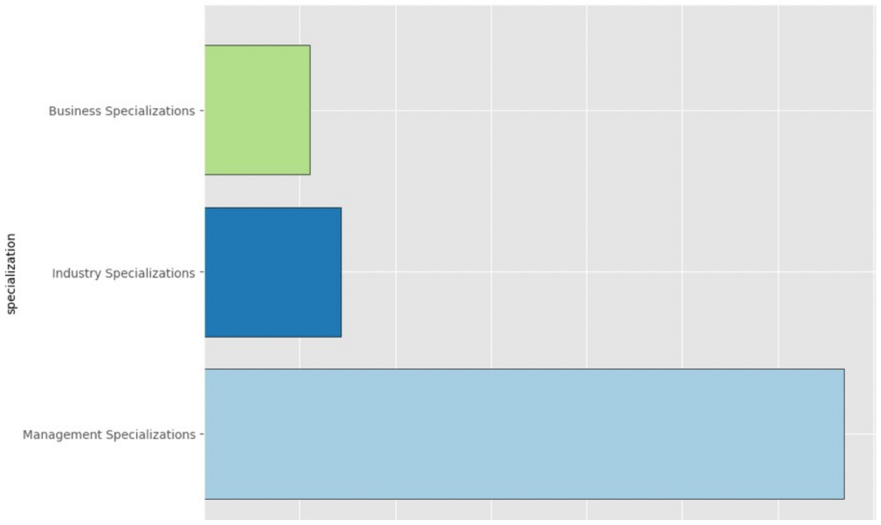
2 Data Cleaning

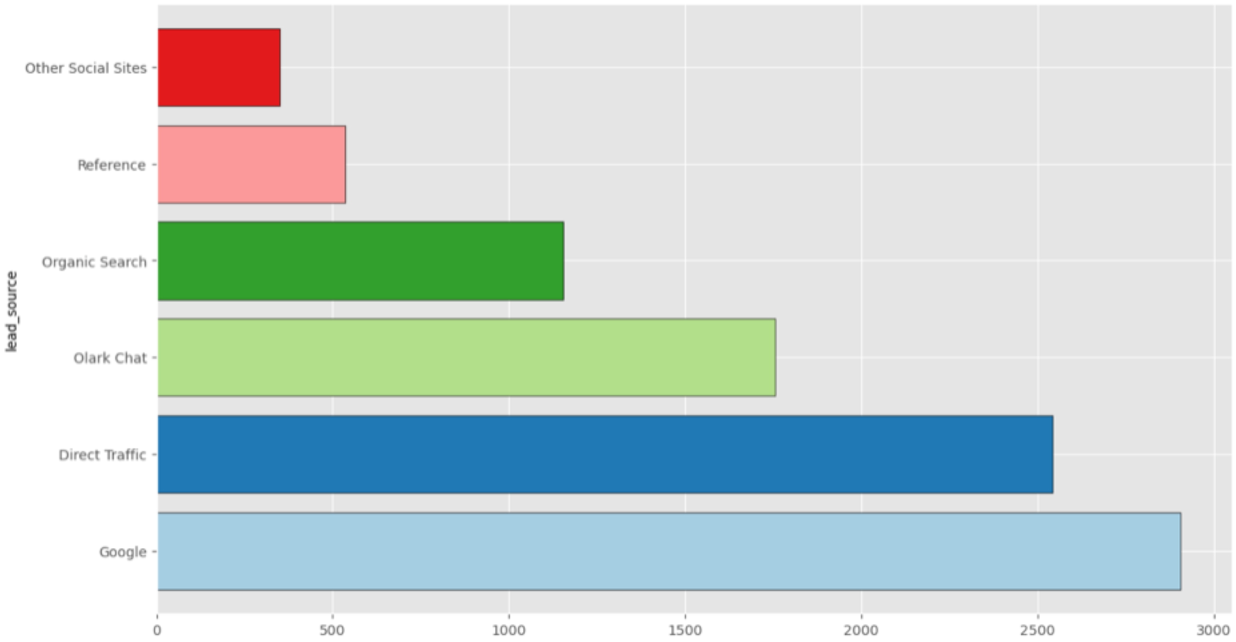
- 2.1 Rename column names
- 2.2 Drop prospect_id column
- 2.3 Replace "Select" category with null values
- 2.4 Handle null values and sales generated columns
    - 2.4.1 Drop columns that have null values > 40% or Sales generated columns
    - 2.4.2 country column
    - 2.4.3 course_selection_reason column
    - 2.4.4 occupation column
    - 2.4.5 specialization column
    - 2.4.6 city column
- 2.5 Handle categorical columns with low number of missing values and low representation of categories
    - 2.5.1 lead_origin column
    - 2.5.2 lead_source column
- 2.6 Handle Binary columns
- 2.7 Handle Numerical columns
    - 2.7.1 lead_number column: change datatype
    - 2.7.2 total_visits column
    - 2.7.3 page_views_per_visit column

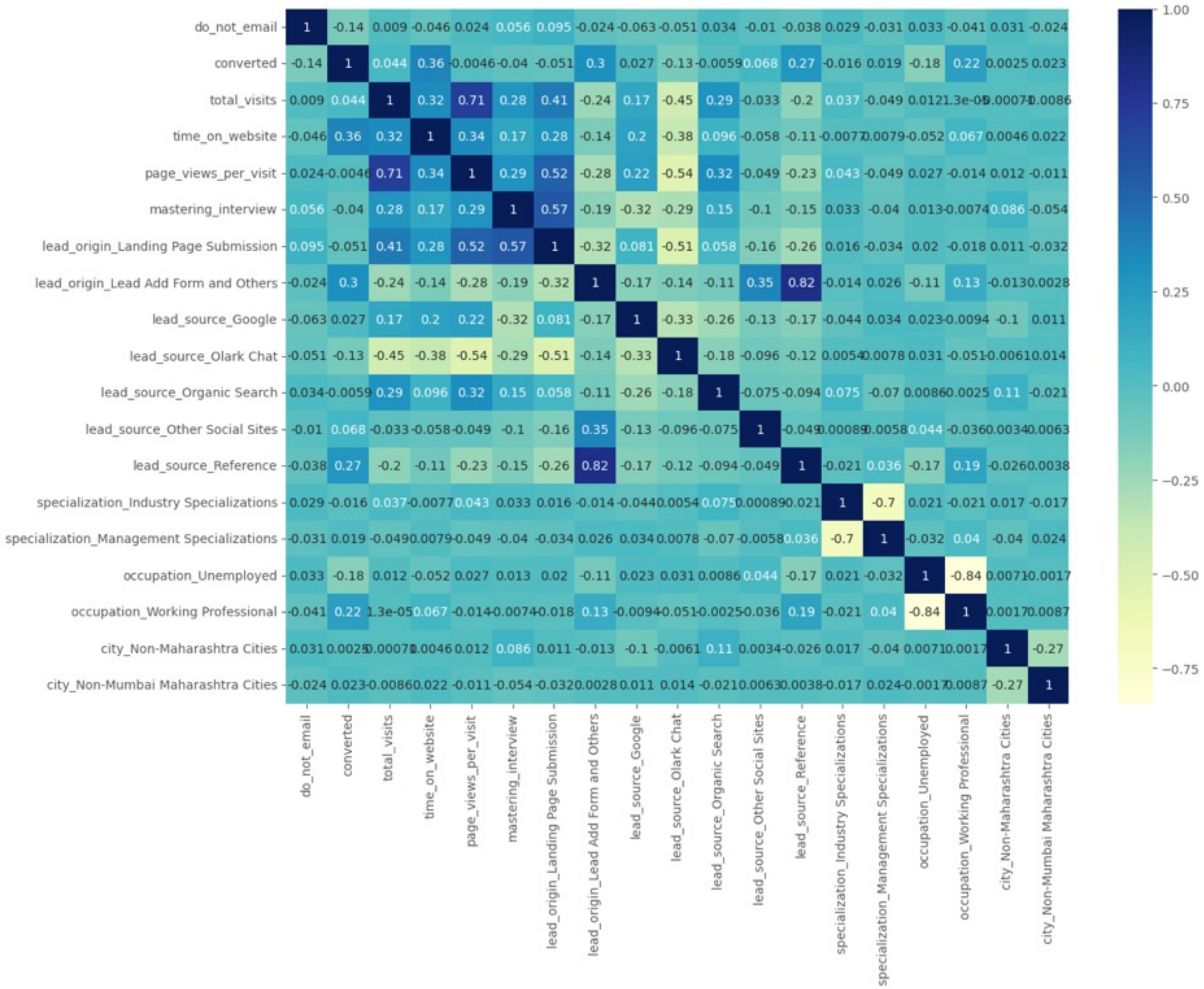

Total website visits
0
2000
4000
6000
8000
0
50
100
150
200
250
Time spent on website
0
500
1000
1500
2000
2500
3000
3500
0
500
1000
1500
2000
Average number of page views per visit
0
1000
2000
3000
4000
5000
6000
0
10
20
30
40
50

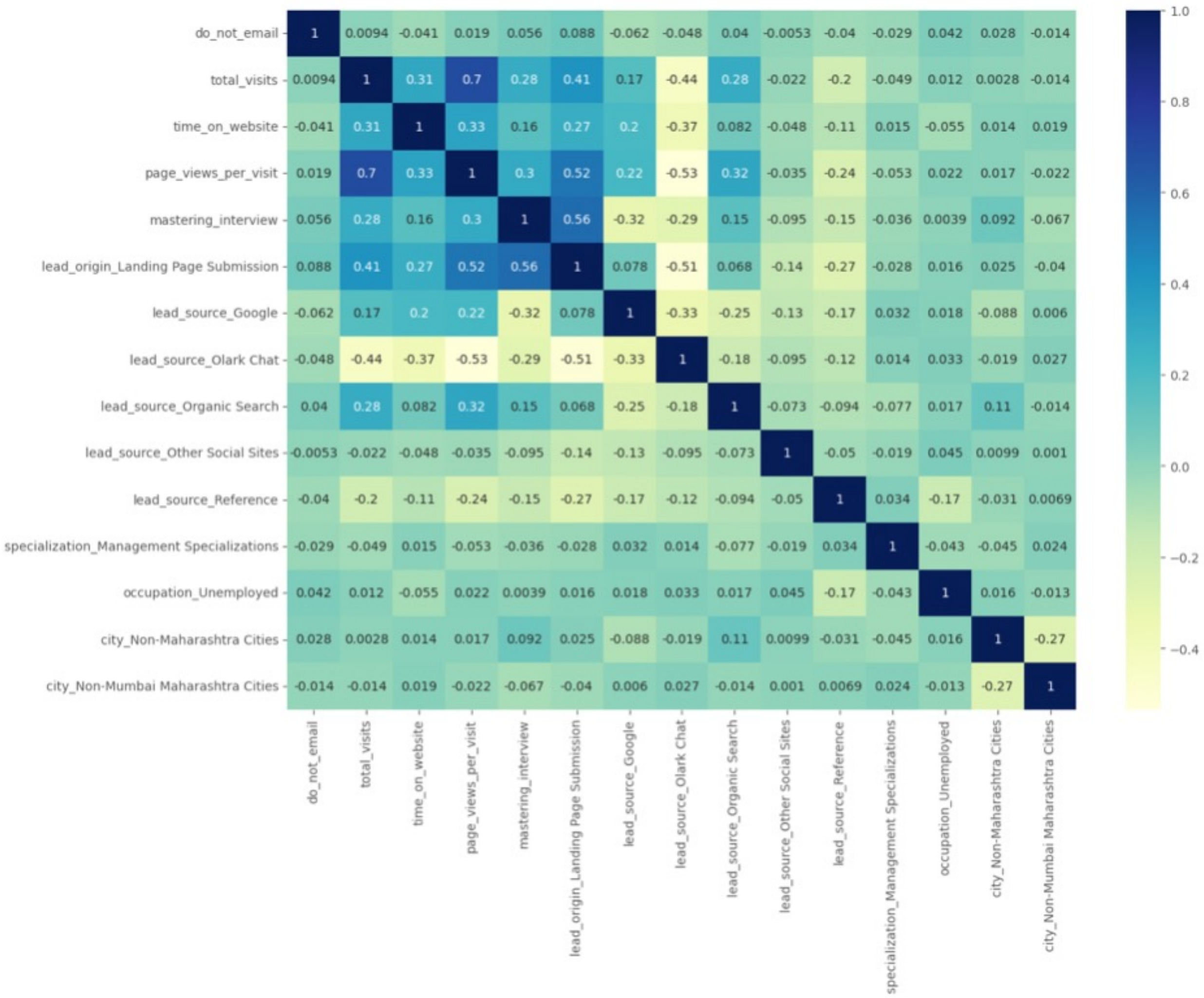

total_visits
time_on_website
page_views_per_visit
1
0.22
0.51
0.22
1
0.31
0.51
0.31
1
total_visits
time_on_website
page_views_per_visit
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3


lead_origin
Lead Add Form and Others
API
Landing Page Submission
0
1000
2000
3000
4000
5000


specialization
Business Specializations
Industry Specializations
Management Specializations


lead_source
Other Social Sites
Reference
Organic Search
Olark Chat
Direct Traffic
Google
0
500
1000
1500
2000
2500
3000


do_not_email 1 -0.14 0.009 -0.046 0.024 0.056 0.095 -0.024 -0.063 -0.051 0.034 -0.01 -0.038 0.029 -0.031 0.033 -0.041 0.031 -0.024
converted -0.14 1 0.044 0.36 -0.0046 -0.04 -0.051 0.3 0.027 -0.13 -0.0059 0.068 0.27 -0.016 0.019 -0.18 0.22 0.0025 0.023
total_visits 0.009 0.044 1 0.32 0.71 0.28 0.41 -0.24 0.17 -0.45 0.29 -0.033 -0.2 0.037 -0.049 0.012 1.3e-05 0.00071 -0.0086
time_on_website -0.046 0.36 0.32 1 0.34 0.17 0.28 -0.14 0.2 -0.38 0.096 -0.058 -0.11 -0.0077 0.0079 -0.052 0.067 0.0046 0.022
page_views_per_visit 0.024 -0.0046 0.71 0.34 1 0.29 0.52 -0.28 0.22 -0.54 0.32 -0.049 -0.23 0.043 -0.049 0.027 -0.014 0.012 -0.011
mastering_interview 0.056 -0.04 0.28 0.17 0.29 1 0.57 -0.19 -0.32 -0.29 0.15 -0.1 -0.15 0.033 -0.04 0.013 -0.0074 0.086 -0.054
lead_origin_Landing Page Submission 0.095 -0.051 0.41 0.28 0.52 0.57 1 -0.32 0.081 -0.51 0.058 -0.16 -0.26 0.016 -0.034 0.02 -0.018 0.011 -0.032
lead_origin_Lead Add Form and Others -0.024 0.3 -0.24 -0.14 -0.28 -0.19 -0.32 1 -0.17 -0.14 -0.11 0.35 0.82 -0.014 0.026 -0.11 0.13 -0.013 0.0028
lead_source_Google -0.063 0.027 0.17 0.2 0.22 -0.32 0.081 -0.17 1 -0.33 -0.26 -0.13 -0.17 -0.044 0.034 0.023 -0.0094 -0.1 0.011
lead_source_Olark Chat -0.051 -0.13 -0.45 -0.38 -0.54 -0.29 -0.51 -0.14 -0.33 1 -0.18 -0.096 -0.12 0.0054 0.0078 0.031 -0.051 -0.0061 0.014
lead_source_Organic Search 0.034 -0.0059 0.29 0.096 0.32 0.15 0.058 -0.11 -0.26 -0.18 1 -0.075 -0.094 0.075 -0.07 0.0086 -0.0025 0.11 -0.021
lead_source_Other Social Sites -0.01 0.068 -0.033 -0.058 -0.049 -0.1 -0.16 0.35 -0.13 -0.096 -0.075 1 -0.049 0.00089 -0.0058 0.044 -0.036 0.0034 0.0063
lead_source_Reference -0.038 0.27 -0.2 -0.11 -0.23 -0.15 -0.26 0.82 -0.17 -0.12 -0.094 -0.049 1 -0.021 0.036 -0.17 0.19 -0.026 0.0038
specialization_Industry Specializations 0.029 -0.016 0.037 -0.0077 0.043 0.033 0.016 -0.014 -0.044 0.0054 0.075 0.00089 -0.021 1 -0.7 0.021 -0.021 0.017 -0.017
specialization_Management Specializations -0.031 0.019 -0.049 0.0079 -0.049 -0.04 -0.034 0.026 0.034 0.0078 -0.07 -0.0058 0.036 -0.7 1 -0.032 0.04 -0.04 0.024
occupation_Unemployed 0.033 -0.18 0.012 -0.052 0.027 0.013 0.02 -0.11 0.023 0.031 0.0086 0.044 -0.17 0.021 -0.032 1 -0.84 0.0071 -0.0017
occupation_Working Professional -0.041 0.22 1.3e-05 0.067 -0.014 -0.0074 -0.018 0.13 -0.0094 -0.051 -0.0025 -0.036 0.19 -0.021 0.04 -0.84 1 0.0017 0.0087
city_Non-Maharashtra Cities 0.031 0.0025 0.00071 0.0046 0.012 0.086 0.011 -0.013 -0.1 -0.0061 0.11 0.0034 -0.026 0.017 -0.04 0.0071 0.0017 1 -0.27
city_Non-Mumbai Maharashtra Cities -0.024 0.023 -0.0086 0.022 -0.011 -0.054 -0.032 0.0028 0.011 0.014 -0.021 0.0063 0.0038 -0.017 0.024 -0.0017 0.0087 -0.27 1
do_not_email
converted
total_visits
time_on_website
page_views_per_visit
mastering_interview
lead_origin_Landing Page Submission
lead_origin_Lead Add Form and Others
lead_source_Google
lead_source_Olark Chat
lead_source_Organic Search
lead_source_Other Social Sites
lead_source_Reference
specialization_Industry Specializations
specialization_Management Specializations
occupation_Unemployed
occupation_Working Professional
city_Non-Maharashtra Cities
city_Non-Mumbai Maharashtra Cities
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75



	do_not_email	total_visits	time_on_website	page_views_per_visit	mastering_interview	lead_origin_Landing Page Submission	lead_source_Google	lead_source_Olark Chat	lead_source_Organic Search	lead_source_Other Social Sites	lead_source_Reference	specialization_Management Specializations	occupation_Unemployed	city_Non-Maharashtra Cities	city_Non-Mumbai Maharashtra Cities
do_not_email	1	0.0094	-0.041	0.019	0.056	0.088	-0.062	-0.048	0.04	-0.0053	-0.04	-0.029	0.042	0.028	-0.014
total_visits	0.0094	1	0.31	0.7	0.28	0.41	0.17	-0.44	0.28	-0.022	-0.2	-0.049	0.012	0.0028	-0.014
time_on_website	-0.041	0.31	1	0.33	0.16	0.27	0.2	-0.37	0.082	-0.048	-0.11	0.015	-0.055	0.014	0.019
page_views_per_visit	0.019	0.7	0.33	1	0.3	0.52	0.22	-0.53	0.32	-0.035	-0.24	-0.053	0.022	0.017	-0.022
mastering_interview	0.056	0.28	0.16	0.3	1	0.56	-0.32	-0.29	0.15	-0.095	-0.15	-0.036	0.0039	0.092	-0.067
lead_origin_Landing Page Submission	0.088	0.41	0.27	0.52	0.56	1	0.078	-0.51	0.068	-0.14	-0.27	-0.028	0.016	0.025	-0.04
lead_source_Google	-0.062	0.17	0.2	0.22	-0.32	0.078	1	-0.33	-0.25	-0.13	-0.17	0.032	0.018	-0.088	0.006
lead_source_Olark Chat	-0.048	-0.44	-0.37	-0.53	-0.29	-0.51	-0.33	1	-0.18	-0.095	-0.12	0.014	0.033	-0.019	0.027
lead_source_Organic Search	0.04	0.28	0.082	0.32	0.15	0.068	-0.25	-0.18	1	-0.073	-0.094	-0.077	0.017	0.11	-0.014
lead_source_Other Social Sites	-0.0053	-0.022	-0.048	-0.035	-0.095	-0.14	-0.13	-0.095	-0.073	1	-0.05	-0.019	0.045	0.0099	0.001
lead_source_Reference	-0.04	-0.2	-0.11	-0.24	-0.15	-0.27	-0.17	-0.12	-0.094	-0.05	1	0.034	-0.17	-0.031	0.0069
specialization_Management Specializations	-0.029	-0.049	0.015	-0.053	-0.036	-0.028	0.032	0.014	-0.077	-0.019	0.034	1	-0.043	-0.045	0.024
occupation_Unemployed	0.042	0.012	-0.055	0.022	0.0039	0.016	0.018	0.033	0.017	0.045	-0.17	-0.043	1	0.016	-0.013
city_Non-Maharashtra Cities	0.028	0.0028	0.014	0.017	0.092	0.025	-0.088	-0.019	0.11	0.0099	-0.031	-0.045	0.016	1	-0.27
city_Non-Mumbai Maharashtra Cities	-0.014	-0.014	0.019	-0.022	-0.067	-0.04	0.006	0.027	-0.014	0.001	0.0069	0.024	-0.013	-0.27	1

Conclusion-

The Sensitivity and Specificity, Accuracy, Precision and Recall score we got from test set are almost accurate.
We have high recall score than precision score which is a sign of good model.
In business terms, this model has an ability to adjust with the company’s requirements in coming future.
This concludes that the model is in stable state.
Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :

- Lead Origin Lead Add Form
- Total Time Spent on Website
- What is your current occupation Working Professional