

LEAD GENERATION

PRESNTATION

THE PERSNTATION OF THELEAD GENERATION
REQUIRED THE CUSTOMER CONVERSION TO
THEIR NEED

OUR DATASET

	Prospect ID	Lead Number	Lead Origin	Lead Source	Do Not Email	Do Not Call	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	...	Get updates on DM Content	Lead Profile
0	7927b2df-8bba-4d29-b9a2-b6e0beafe620	660737	API	Olark Chat	No	No	0	0.0	0	0.0	...	No	Select
1	2a272436-5132-4136-86fa-dcc88c88f482	660728	API	Organic Search	No	No	0	5.0	674	2.5	...	No	Select
2	8cc8c611-a219-4f35-ad23-fdfd2656bd8a	660727	Landing Page Submission	Direct Traffic	No	No	1	2.0	1532	2.0	...	No	Potential Lead
3	0cc2df48-7cf4-4e39-9de9-19797f9b38cc	660719	Landing Page Submission	Direct Traffic	No	No	0	1.0	305	1.0	...	No	Select
4	3256f628-e534-4826-9d63-4a8b88782852	660681	Landing Page Submission	Google	No	No	1	2.0	1428	1.0	...	No	Select

THIS DATASET REQUIRE ONLY 90% OF ANALYSIS

THIS IS THE EXAMPLE OF DATASET GIVEN FOR GIVEN LOGISTIC REGRESSION

CONTENT:-

STEPS FOR APPLYING LOGISTIC REGRESSION:-

- GETTING IMPORTANT INFORMATION ABOUT THE DATA
- CLEANING THE DATASET
- PLOTING FOR IMPORTANT COLUMN TO UNDERSTAND BETTER
- CREATING THE DUMMY VARIABLES
- APPLYING UNIVARIATE LOGISTIC REGRESSION
- APPLYING STATSTIC LOGISTIC REGRESSION MODEL
- CALCULATING THE REF(IF NEEDED)
- CALCULATING THE VIF

INFO DESCRIBE

```
1 ld.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9240 entries, 0 to 9239
Data columns (total 37 columns):
 # Column           Non-Null Count Dtype
 --- -----
 0 Prospect ID      9240 non-null  object
 1 Lead Number      9240 non-null  int64
 2 Lead Origin      9240 non-null  object
 3 Lead Source      9204 non-null  object
 4 Do Not Email     9240 non-null  object
 5 Do Not Call      9240 non-null  object
 6 Converted        9240 non-null  int64
 7 TotalVisits      9103 non-null  float64
 8 Total Time Spent on Website 9240 non-null  int64
 9 Page Views Per Visit   9103 non-null  float64
 10 Last Activity     9137 non-null  object
 11 Country          6779 non-null  object
 12 Specialization    7802 non-null  object
 13 How did you hear about X Education 7033 non-null  object
 14 What is your current occupation 6550 non-null  object
 15 What matters most to you in choosing a course 6531 non-null  object
 16 Search            9240 non-null  object
 17 Magazine          9240 non-null  object
 18 Newspaper Article 9240 non-null  object
 19 X Education Forums 9240 non-null  object
 20 Newspaper         9240 non-null  object
```

AS WE CAN SEE THE INFORMATION PROVIDE

- By using info we Get maximum insights from a data set
- Analysts try to find different patterns, relations, and anomalies in the data using some statistical graphs and other visualization techniques.Following things are part of EDA :

- Get maximum insights from a data set
- Uncover underlying structure
- Extract important variables from the dataset
- Detect outliers and anomalies(if any)
- Test underlying assumptions
- Determine the optimal factor settings

From the info(),We can observe that those columns that have symbols are in object form as well as some columns should be of an integer type but are of an object type.

We will use describe() method, which shows basic statistical characteristics of each numerical feature (int64 and float64 types): number of non-missing values, mean, standard deviation, range, median, 25%, 50%, 75% quartiles.

HANDLING MISSING VALUES

we have lot off missing values here
if the variable have more then 40%
missing value remove the data

Also here are lots of the
variables that does't make
sense should be removed

Here a lot of data is missing
we need to clean it

FROM THE HEAD

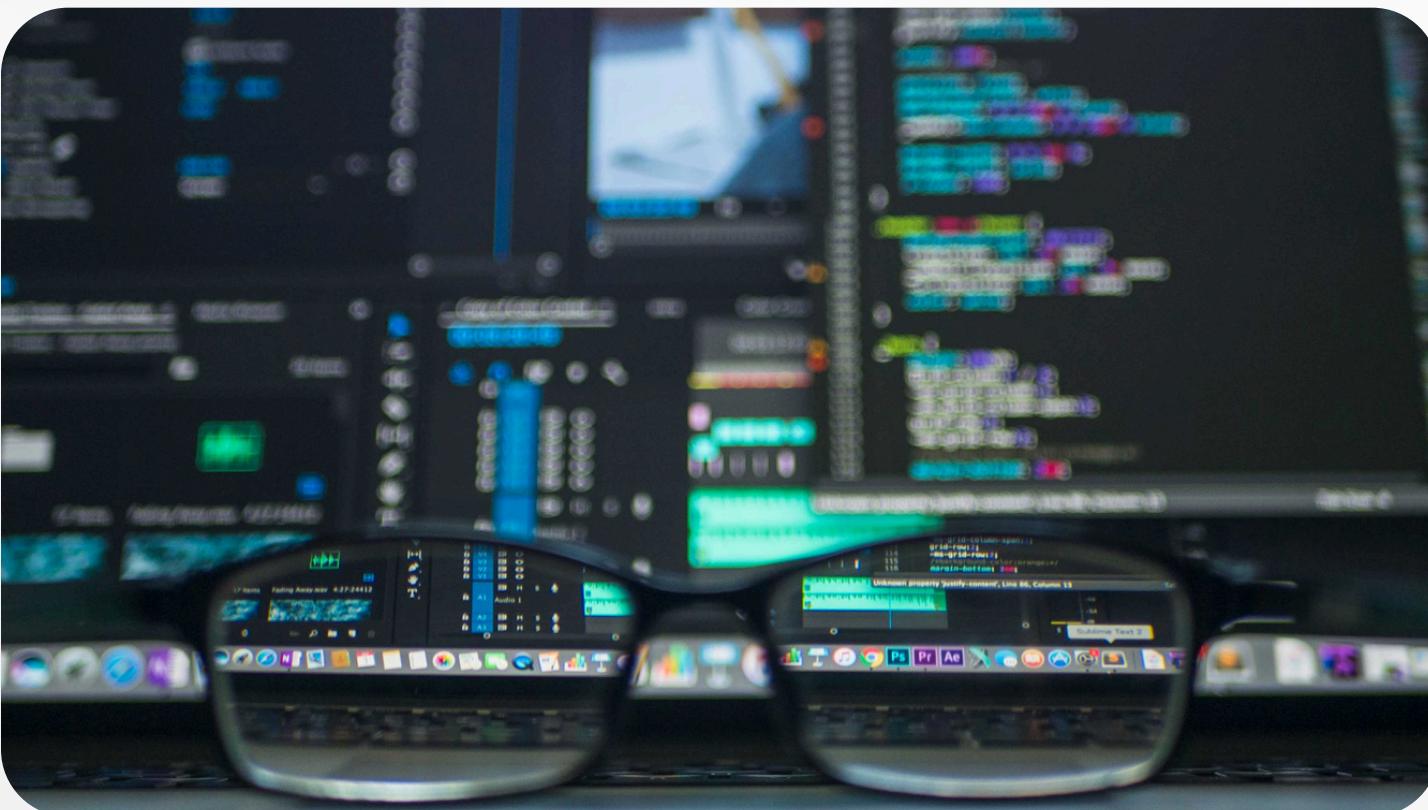
1	ld.isnull().sum().head(15)
Prospect ID	0
Lead Number	0
Lead Origin	0
Lead Source	36
Do Not Email	0
Do Not Call	0
Converted	0
TotalVisits	137
Total Time Spent on Website	0
Page Views Per Visit	137
Last Activity	103
Country	2461
Specialization	3380
How did you hear about X Education	7250
What is your current occupation	2690
	dtype: int64

FROM THE TAIL

1	ld.isnull().sum().tail(15)
Through Recommendations	0
Receive More Updates About Our Courses	0
Tags	3353
Lead Quality	4767
Update me on Supply Chain Content	0
Get updates on DM Content	0
Lead Profile	6855
City	3669
Asymmetrique Activity Index	4218
Asymmetrique Profile Index	4218
Asymmetrique Activity Score	4218
Asymmetrique Profile Score	4218
I agree to pay the amount through cheque	0
A free copy of Mastering The Interview	0
Last Notable Activity	0
	dtype: int64

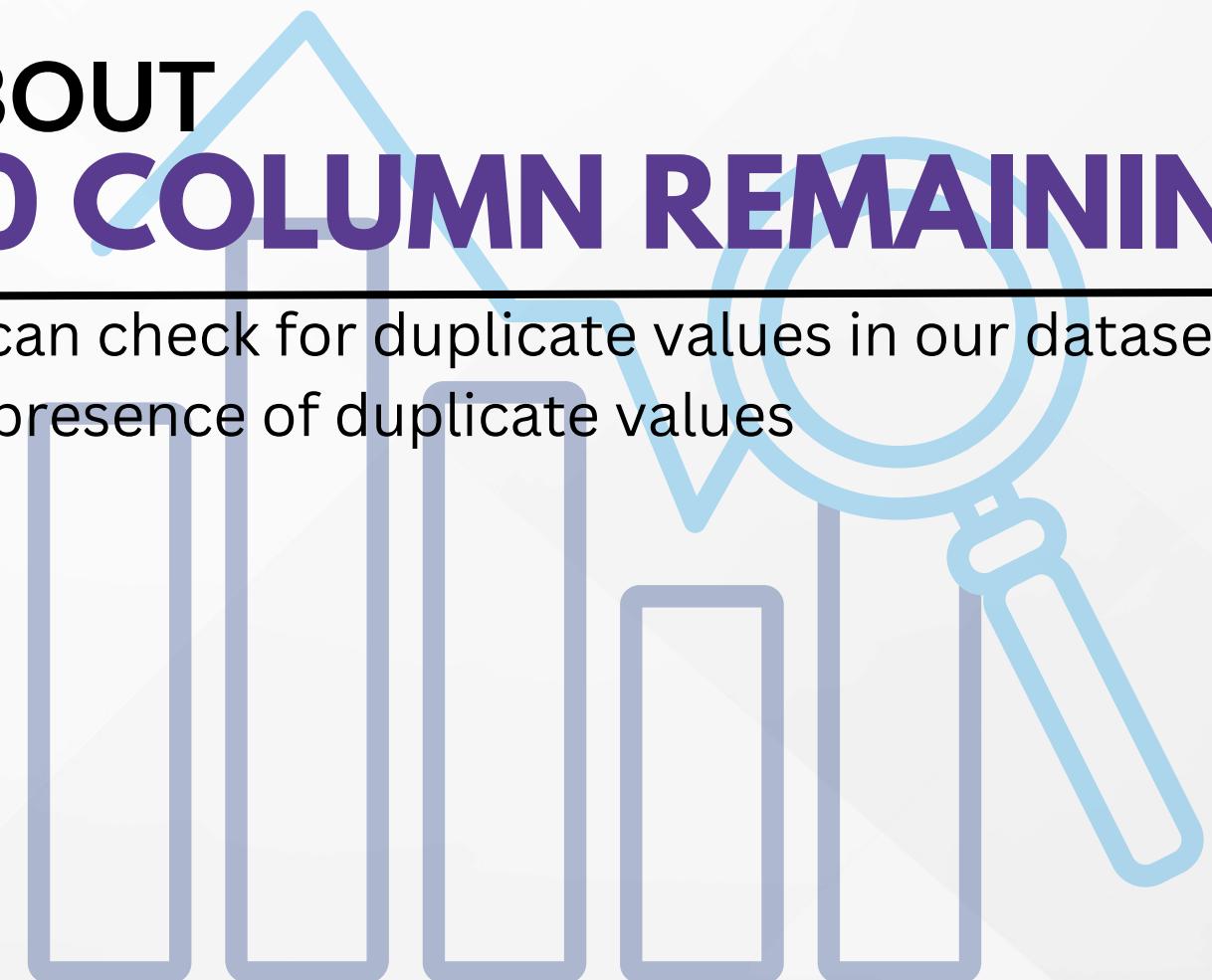
DATA CLEANING

- we use to separate the categorical column from numerical column
- also dropping those columns which have no use
- so after cleaning and dropping the column the resultant dataset required to analysis



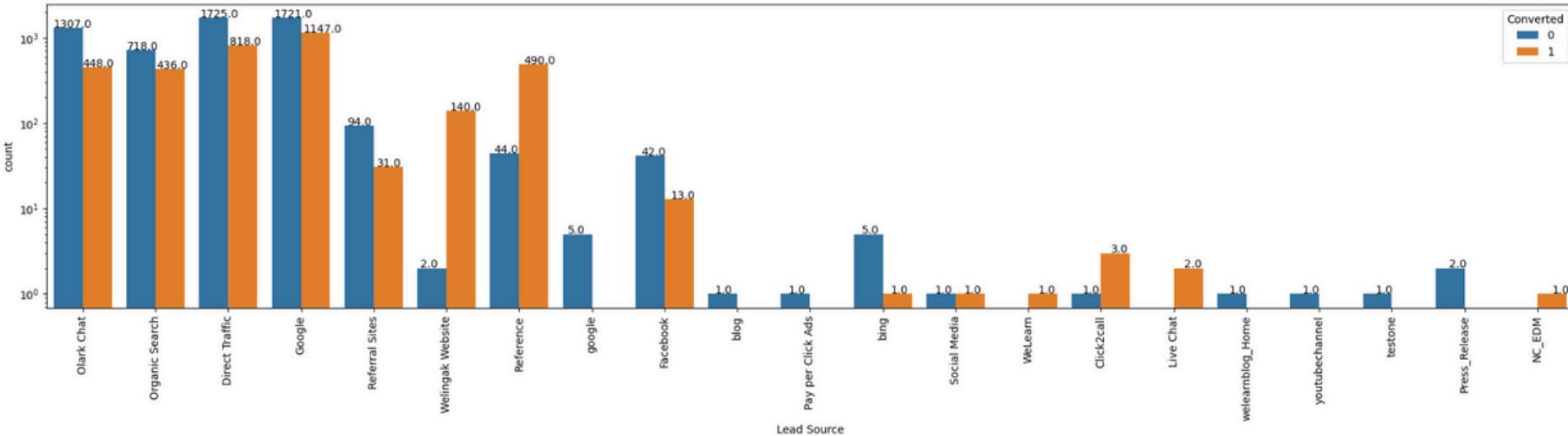
**ABOUT
80 COLUMN REMAINING**

We can check for duplicate values in our dataset as the presence of duplicate values

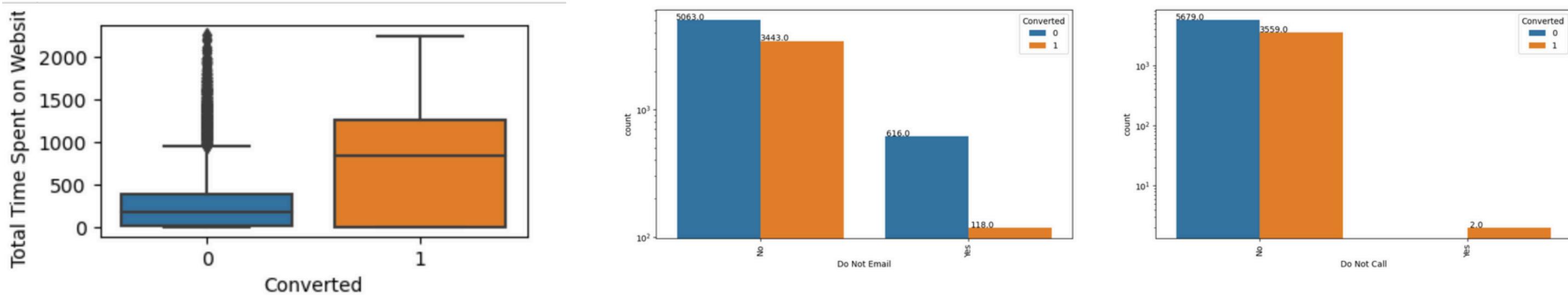


LEAD GENERATION

EDA



Google and Direct traffic generates maximum number of leads. Conversion Rate of reference leads and leads through welingak

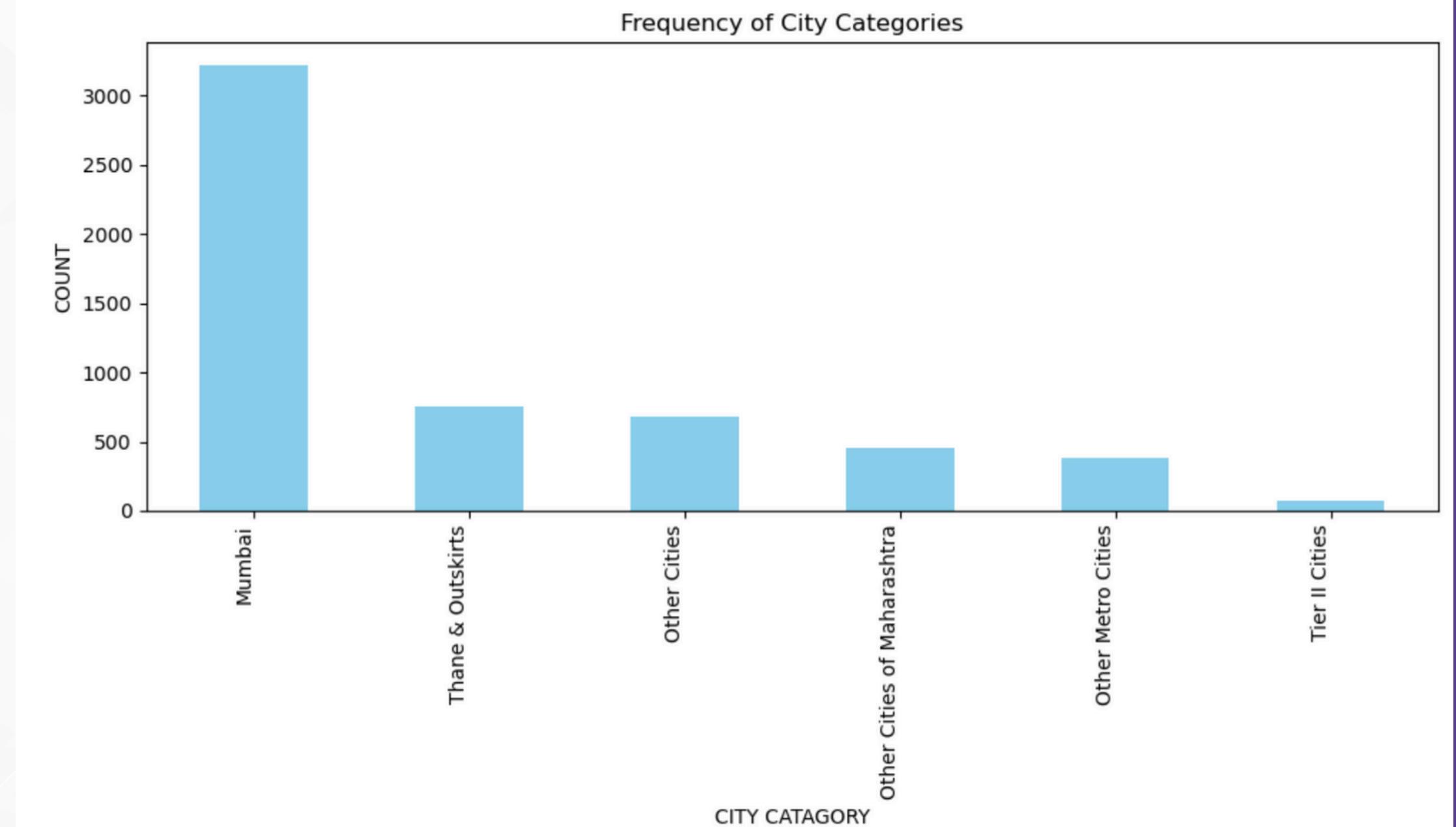


The eda here shows how the variables affecting the lead generation

FREQUENCY OF CITY PLOTTING THE GRAPH

Here city shows places interested in conversion of the lead generation

Here the mumbai city shows the most interest but also city variable has a lot of null values so we need to impute it

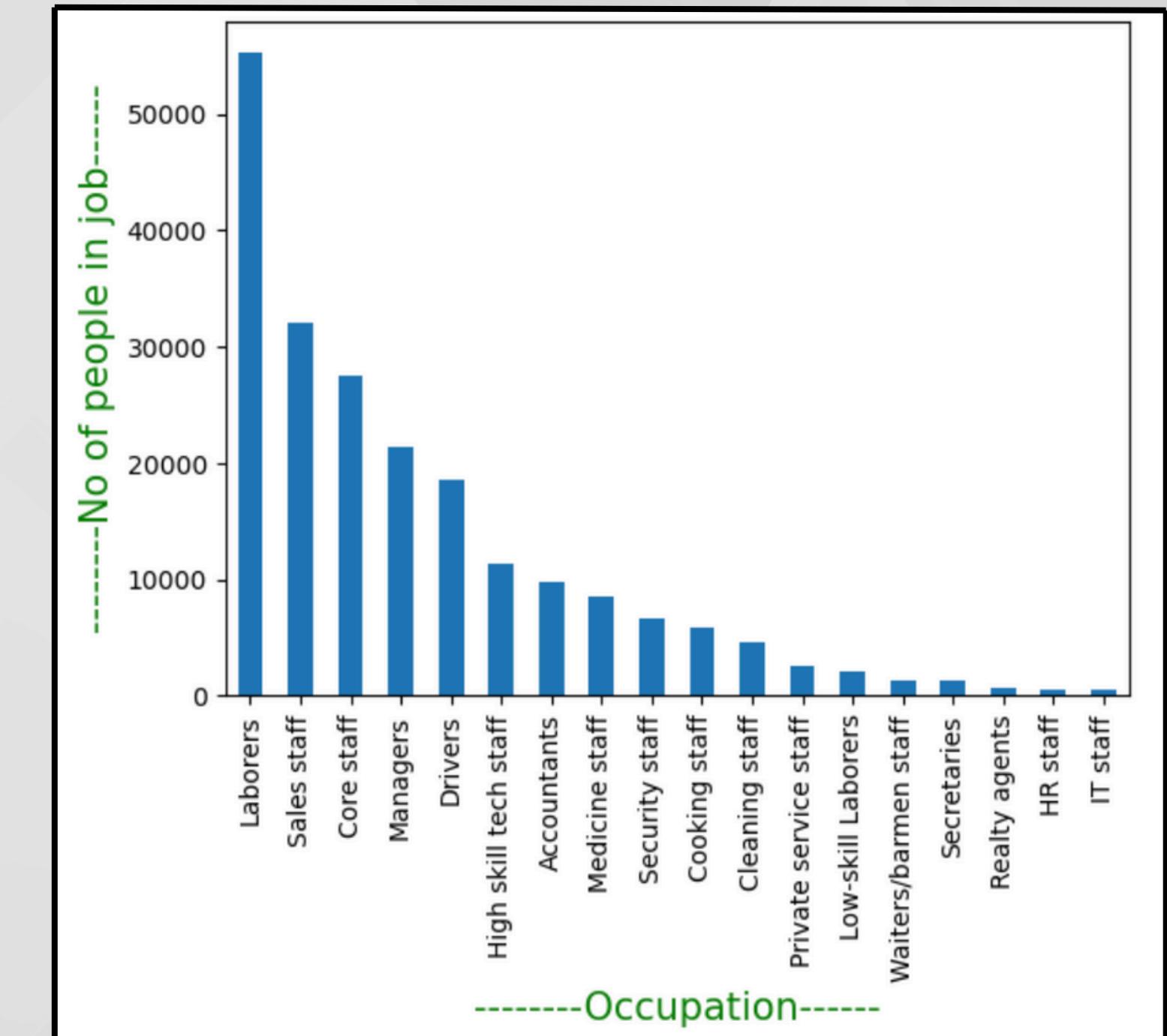


DATA PLOTING UNIVARIATE ANALYSES

Here we plotted the bar chart for the occupation had by no of peoples

This allows to calculate how many people according to their occupation eligible for the loan

It will sort out the problem for the non eligible candidates like Laborers, Drivers, Security staff etc.



CLIENT INCOME & CREDIT BIVARIATE ANALYSIS

This plot tells the client income and according to it what credit they section the loan

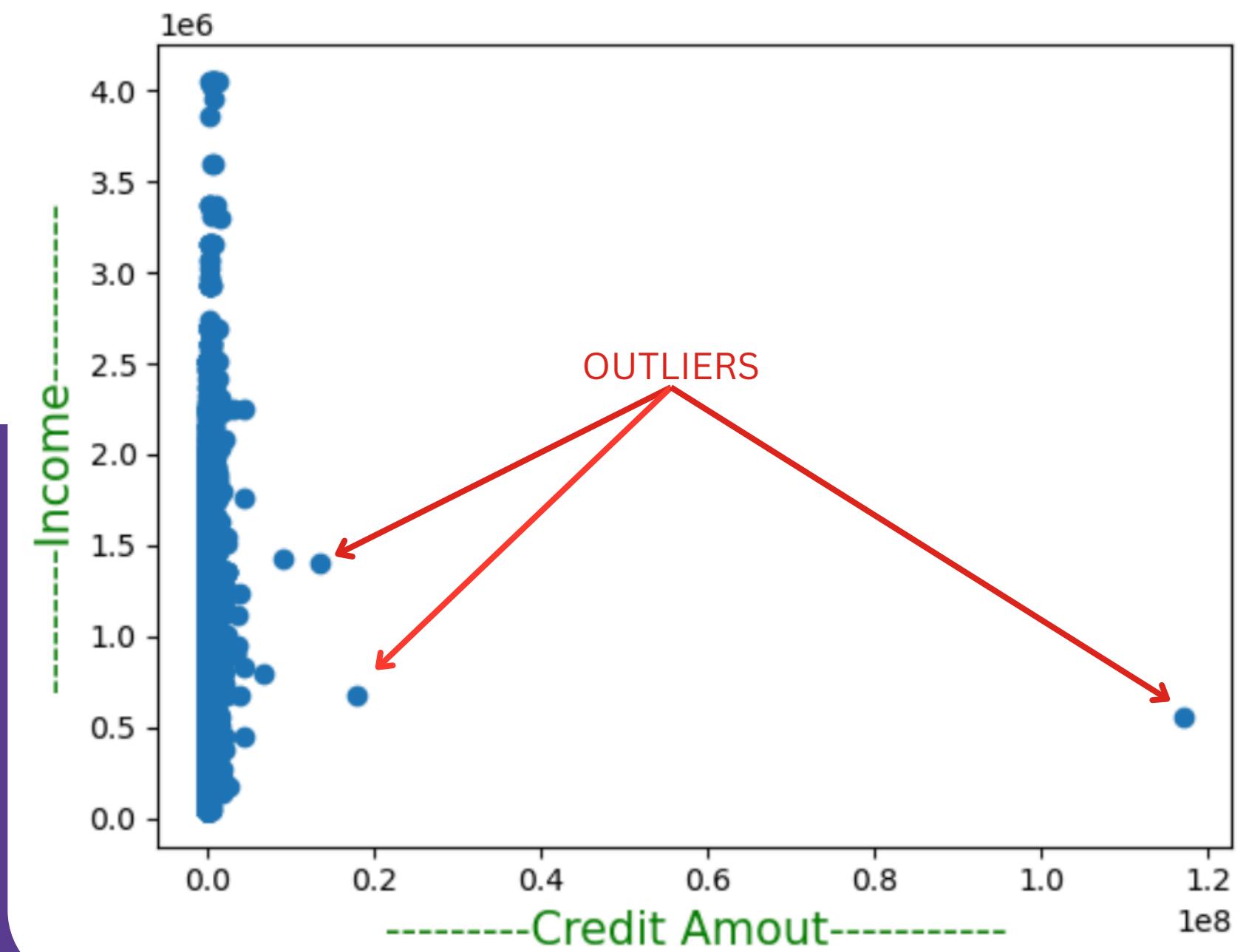
We plotted the scatter graph to see the outlier on that client who credit too high but those are very less



Amount of income allow them to apply for the credit on the basis of which they will repay the loan

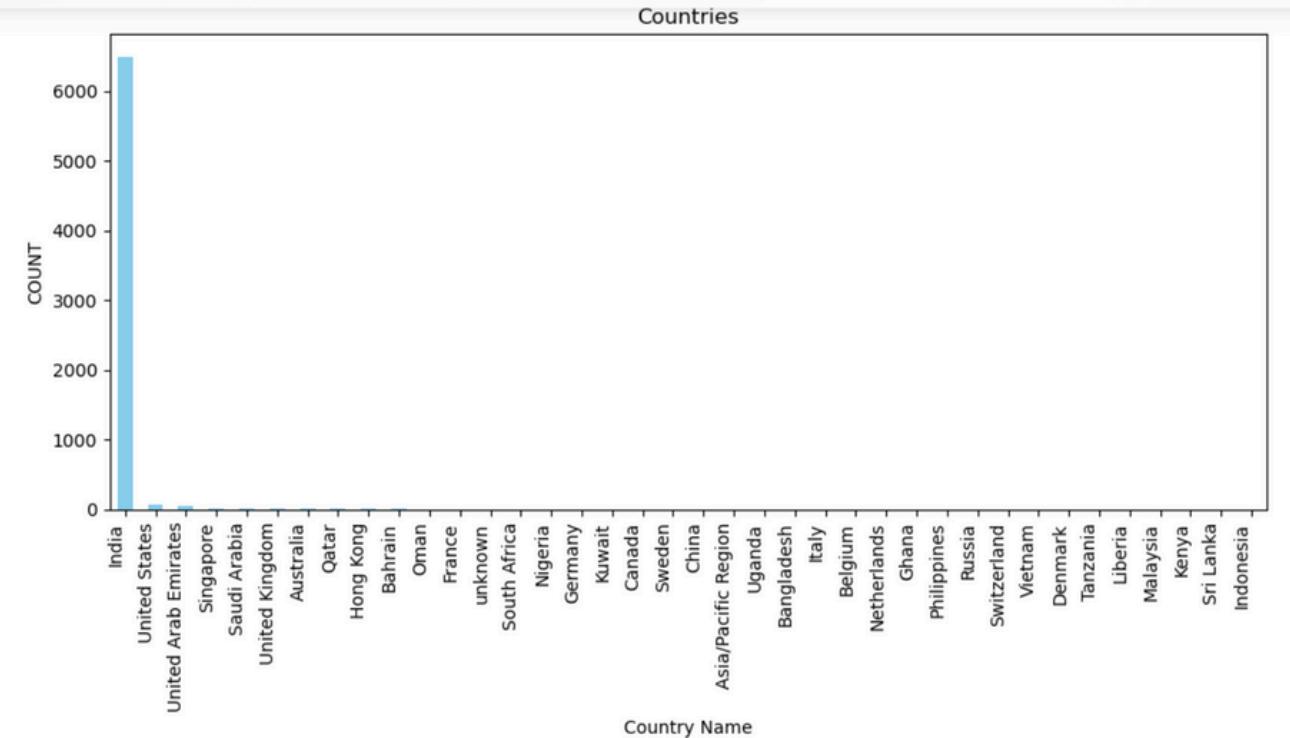
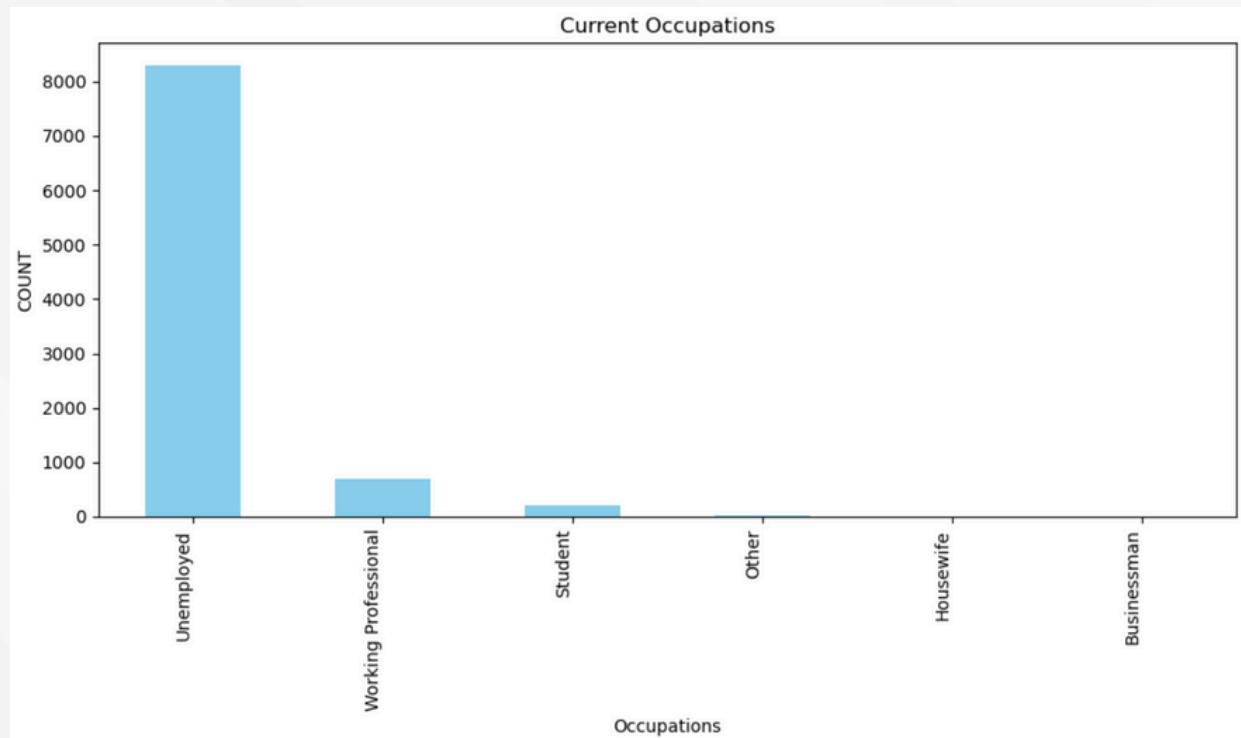


Credit amount on loan will tells on how much they have to repay it



IMPUTING THE VARIABLES ON

'WHAT IS YOUR CURRENT OCCUPATION'
SPECIALIZATION
COUNTRY



here imputation is necessary to stable
the datafram

Mostly indians are involve in the
lead generation

Mostly unemployed people
participated in conversion

Buisnessman showed less
interest on lead generation

DATA MANIPULATION

CONVENTION AND DUMMY CREATION



For the conversion we use to convert all numerical values into binary it is necessary in logistic regression



It will allow binary outcomes and the data fitting becomes possible



Datafram had a lot of variables with many levels so it required dummy variable for sorting the datafram



After that all object variables should be deleted

MULTIVARIAT HEATMAP ANALYSIS

1	0.00048	0.0013	0.25	0.83	0.015	0.39	0.17	0.19	0.57	0.24	-0.2	-0.24	0.014	0.021	-0.051	0.34	0.34	0.36	0.36	0.33	0.28	0.28
0.00048	1	0.018	0.24	0.11	0.025	0.0033	-0.02	-0.0073	-0.0063	-0.013	0.014	-0.01	-0.0038	0.29	0.17	0.28	-0.1	-0.058	-0.077	-0.1	-0.023	-0.04
-0.0013	0.018	1	0.35	0.15	0.17	-0.013	-0.11	-0.046	-0.16	-0.024	0.22	-0.2	-0.014	0.32	-0.32	0.17	-0.0037	0.12	0.099	0.041	0.21	0.21
0.25	0.24	0.35	1	0.65	0.23	0.1	-0.11	-0.0052	-0.066	0.028	0.27	-0.37	-0.011	0.79	-0.77	0.5	0.3	0.44	0.48	0.34	0.63	0.62
0.83	0.11	0.15	0.65	1	0.11	0.33	0.073	0.16	0.37	0.18	-0.054	-0.3	0.0077	0.36	-0.38	0.48	0.42	0.51	0.51	0.43	0.51	0.51
0.015	0.025	0.17	0.23	0.11	1	0.024	-0.2	-0.051	-0.12	-0.015	0.18	-0.17	-0.0049	0.26	-0.18	0.16	-0.037	0.052	0.052	-0.064	0.11	0.12
0.39	0.0033	-0.013	0.1	0.33	0.024	1	0.44	0.084	0.25	0.086	-0.073	-0.11	-0.0062	0.006	-0.0063	0.14	0.14	0.14	0.17	0.12	0.12	0.12
0.17	-0.02	-0.11	-0.11	0.073	-0.2	0.44	1	0.074	0.2	0.057	-0.15	0.053	0.0028	-0.16	0.13	-0.017	0.091	0.03	0.02	0.075	-0.019	-0.037
0.19	-0.0073	-0.046	-0.0052	0.16	-0.051	0.084	0.074	1	-0.29	0.069	-0.099	-0.017	0.0031	-0.084	0.04	-0.012	0.09	0.088	0.092	0.092	0.051	0.052
0.57	-0.0063	-0.16	-0.066	0.37	-0.12	0.25	0.2	-0.29	1	0.18	-0.28	-0.008	0.013	-0.21	0.2	0.12	0.21	0.11	0.17	0.24	0.073	0.077
0.24	-0.013	-0.024	0.028	0.18	-0.015	0.086	0.057	0.069	0.18	1	-0.37	-0.29	-0.0014	-0.047	0.0075	0.065	0.12	0.088	0.12	0.1	0.044	0.047
-0.2	0.014	0.22	0.27	-0.054	0.18	-0.073	-0.15	-0.099	-0.28	-0.37	1	-0.39	0.0031	0.34	-0.28	0.087	-0.11	0.00042	-0.012	-0.11	0.14	0.14
-0.24	-0.01	-0.2	-0.37	-0.3	-0.17	-0.11	0.053	-0.017	-0.008	-0.29	-0.39	1	0.0095	-0.31	0.31	-0.23	-0.086	-0.17	-0.19	-0.069	-0.25	-0.25
0.014	-0.0038	-0.014	-0.011	0.0077	-0.0049	-0.0062	0.0028	0.0031	0.013	-0.0014	0.0031	0.0095	1	-0.0093	0.011	-0.00089	-0.026	-0.0075	0.0067	-0.0021	9.8e-05	-0.0021
0.021	0.29	0.32	0.79	0.36	0.26	0.006	-0.16	-0.084	-0.21	-0.047	0.34	-0.31	-0.0093	1	-0.47	0.37	-0.026	0.16	0.18	-0.03	0.33	0.32
-0.051	0.17	-0.32	-0.77	-0.38	-0.18	-0.0063	0.13	0.04	0.2	0.0075	-0.28	0.31	0.011	-0.47	1	-0.22	-0.33	-0.39	-0.38	-0.34	-0.41	-0.42
0.34	0.28	0.17	0.5	0.48	0.16	0.14	-0.017	-0.012	0.12	0.065	0.087	-0.23	-0.00089	0.37	-0.22	1	0.11	0.2	0.2	0.11	0.26	0.27
0.34	-0.1	-0.0037	0.3	0.42	-0.037	0.14	0.091	0.09	0.21	0.12	-0.11	-0.086	-0.026	-0.026	-0.33	0.11	1	0.29	0.28	0.37	0.17	0.18
0.36	-0.058	0.12	0.44	0.51	0.052	0.14	0.03	0.088	0.11	0.088	0.00042	-0.17	-0.0075	0.16	-0.39	0.2	0.29	1	0.3	0.3	0.28	0.27
0.36	-0.077	0.099	0.48	0.51	0.052	0.17	0.02	0.092	0.17	0.12	-0.012	-0.19	0.0067	0.18	-0.38	0.2	0.28	0.3	1	0.33	0.38	0.39
0.33	-0.1	0.041	0.34	0.43	-0.064	0.12	0.075	0.092	0.24	0.1	-0.11	-0.069	-0.0021	-0.03	-0.34	0.11	0.37	0.3	0.33	1	0.27	0.28
0.28	-0.023	0.21	0.63	0.51	0.11	0.12	-0.019	0.051	0.073	0.044	0.14	-0.25	9.8e-05	0.33	-0.41	0.26	0.17	0.28	0.38	0.27	1	0.53
0.28	-0.04	0.21	0.62	0.51	0.12	0.12	-0.037	0.052	0.077	0.047	0.14	-0.25	-0.0021	0.32	-0.42	0.27	0.18	0.27	0.39	0.28	0.53	1

- The analysis of the heatmap clearly shows that there are less correlation
- The map indicates that diagonally most variables equated have the darker side
- The map indicates that more interaction od conversion given to the E-mail and sms
- It indicates that the more converted people comes from india by unemployed people
- Some remaining variables do not have anything to do with conversion

MODEL BUILDING SUMMARY

Generalized Linear Model Regression Results

Dep. Variable:	Converted	No. Observations:	5097
Model:	GLM	Df Residuals:	5044
Model Family:	Binomial	Df Model:	52
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	nan
Date:	Sat, 21 Sep 2024	Deviance:	1.3106e+05
Time:	18:25:25	Pearson chi2:	6.41e+18
No. Iterations:	100	Pseudo R-squ. (CS):	nan
Covariance Type:	nonrobust		

- Here the summary of the model says observation 5095 after cleaning and manipulating the data
- The scale is at exactly 1 which fits the model

RFE & VIF MODEL

Summary static model

VIF

		coef	std err	z	P> z	[0.025	0.975]
	const	1.935e+14	9.56e+06	2.02e+07	0.000	1.94e+14	1.94e+14
	Do Not Email	-1.42e+14	3.85e+06	-3.69e+07	0.000	-1.42e+14	-1.42e+14
	Do Not Call	3.513e+15	4.77e+07	7.37e+07	0.000	3.51e+15	3.51e+15
	TotalVisits	9.362e+13	1.11e+06	8.4e+07	0.000	9.36e+13	9.36e+13
	Total Time Spent on Website	7.618e+14	1.08e+06	7.04e+08	0.000	7.62e+14	7.62e+14
	Page Views Per Visit	-9.522e+13	1.31e+06	-7.28e+07	0.000	-9.52e+13	-9.52e+13
	Search	-2.387e+12	2.71e+07	-8.82e+04	0.000	-2.39e+12	-2.39e+12
	Newspaper Article	4.358e+15	6.73e+07	6.47e+07	0.000	4.36e+15	4.36e+15
	X Education Forums	-19.9620	3.92e-07	-5.1e+07	0.000	-19.962	-19.962
	Newspaper	-5.794e+15	6.73e+07	-8.61e+07	0.000	-5.79e+15	-5.79e+15
	Digital Advertisement	-1.283e+14	4.76e+07	-2.69e+06	0.000	-1.28e+14	-1.28e+14
	Through Recommendations	-2.61e+14	3.2e+07	-8.16e+06	0.000	-2.61e+14	-2.61e+14

Most of the P-values in summary is 0 so their is no need for RFE

	Features	VIF
13	Lead Origin_Lead Add Form	58.80
22	Lead Source_Reference	44.89
23	Lead Source_Welingak Website	14.76
18	Lead Source_Direct Traffic	7.81
12	Lead Origin_Landing Page Submission	7.50
19	Lead Source_Google	6.25
45	Specialization_Others	4.96
21	Lead Source_Organic Search	3.50
20	Lead Source_Olark Chat	3.41
11	A free copy of Mastering The Interview	3.20
37	Specialization_Finance Management	2.15
51	Last Notable Activity_Email Opened	2.08
53	Last Notable Activity_SMS Sent	1.98
39	Specialization_Human Resource Management	1.97
42	Specialization_Marketing Management	1.93
4	Page Views Per Visit	1.92
44	Specialization_Operations Management	1.59
36	Specialization_Business Administration	1.48
2	TotalVisits	1.41
46	Specialization_Supply Chain Management	1.40
40	Specialization_IT Projects Management	1.40
3	Total Time Spent on Website	1.33
47	Specialization_Travel and Tourism	1.25
43	Specialization_Media and Advertising	1.24
41	Specialization_International Business	1.21
38	Specialization_Healthcare Management	1.20
52	Last Notable Activity_Page Visited on Website	1.17

CONCLUSION

PARTICIPATION



The participation of the diffrent background people from deffrent places shows lots of interaction in a lead conversion

GROWTH



If the proper target customer engage for the lead conversion then the growth of lead generation is possible

LEADS CONVERSION



The lead conversion depend only the basic steps to approach the customers like sms email phone call not much with their personal information

THANK YOU