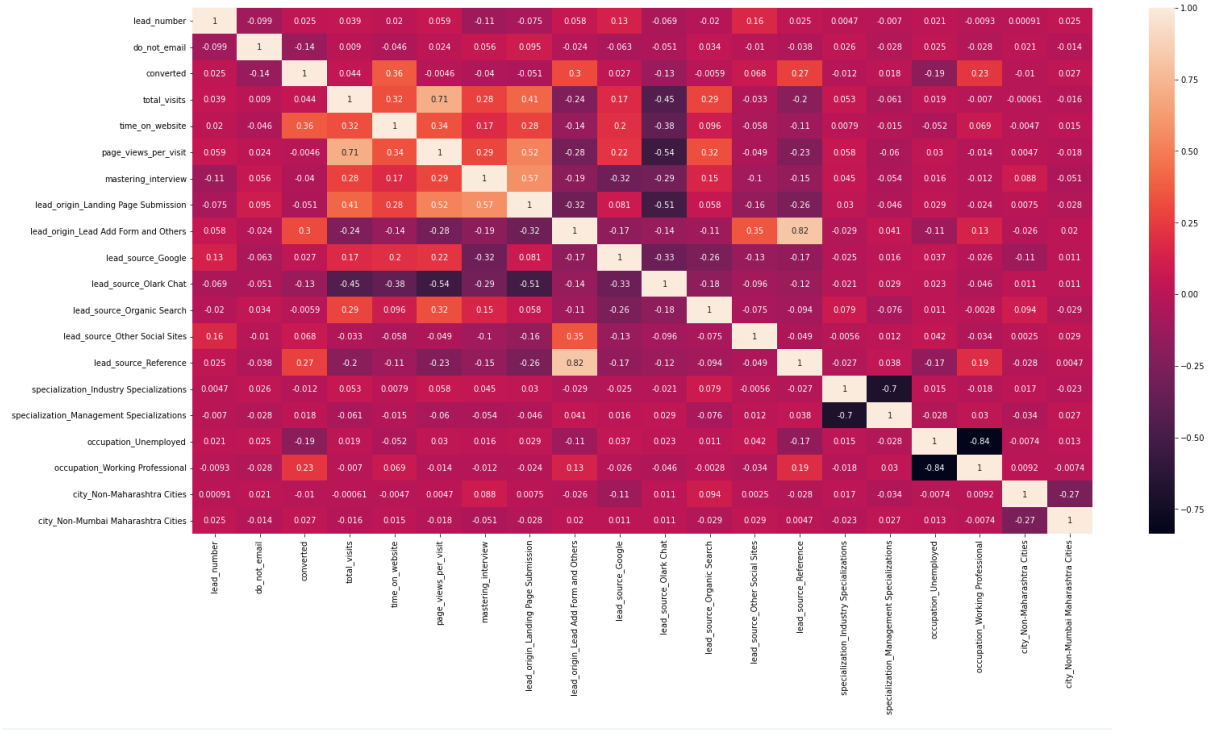
Team Members-

1. Mohd Afraaz Firoz Khan Roll no: 220940325042
2. Shubham Bharat Chaudhari Roll no:220940325069
3. Mandar Manish Ghaisas Roll no:220940325041
4. Aishwarya Devdas Bhalbhar Roll no:220940325005
5. Rahul Vinayrao Joshi Roll no:220940325053

Assignment Subjective Question Solution



1)Which are the top three variables in your model which contribute most towards the probability of a lead getting converted?

From Correlation matrix, it is defined that

1)Time_on_website originally called as "total_time_spent_on_website"

-> The Correlation Matrix defines that the Time_on_website has 36% Positive correlation with converted, which is the target column.

2)Lead_origin_Landing Page Submission originally called lead_origin

-> From the Correlation matrix we can check that, Lead_origin_Landing Page Submission 3% Positive of correlation with the target column

3)lead_source_Reference originally called as lead_source

-> Correlation Matrix shows the correlation of lead_source of 27% Positive with converted[Target column]

4)occupation_Working Professional originally known as occupation

-> occupation_Working Professional has 23% Positive correlation with converted [Target column]

5)occupation_unemployed

-> occupation_unemployed has a negative 19% correlation with converted which is the target column.

2. What are the top 3 categorical/dummy variables in the model which should be focused the most on in order to increase the probability of lead conversion?

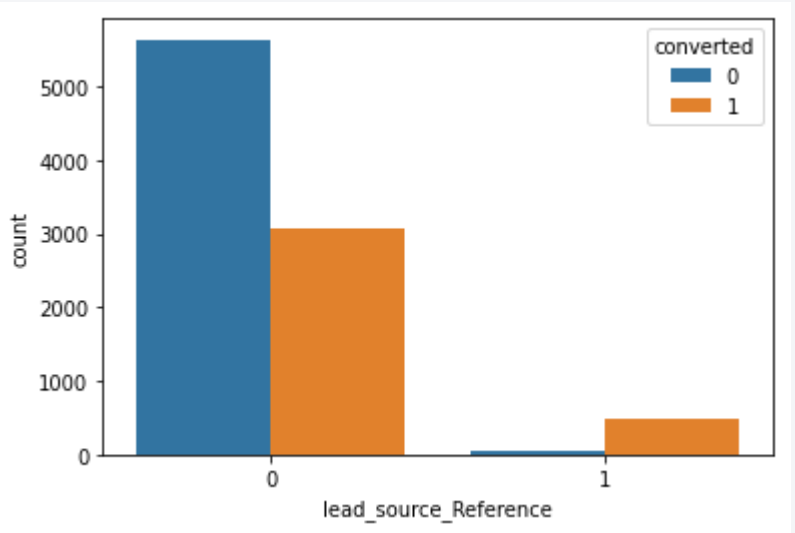
Ans: From the Correlation matrix, it is understood that,

1)Lead Source_Reference

2)lead_origin_Lead Add Form and Others Originally called as Lead Source_Social Media

3)Lead Source_Olark Chat

3.X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (i.e. the customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as many of such people as possible. Suggest a good strategy they should employ at this stage.

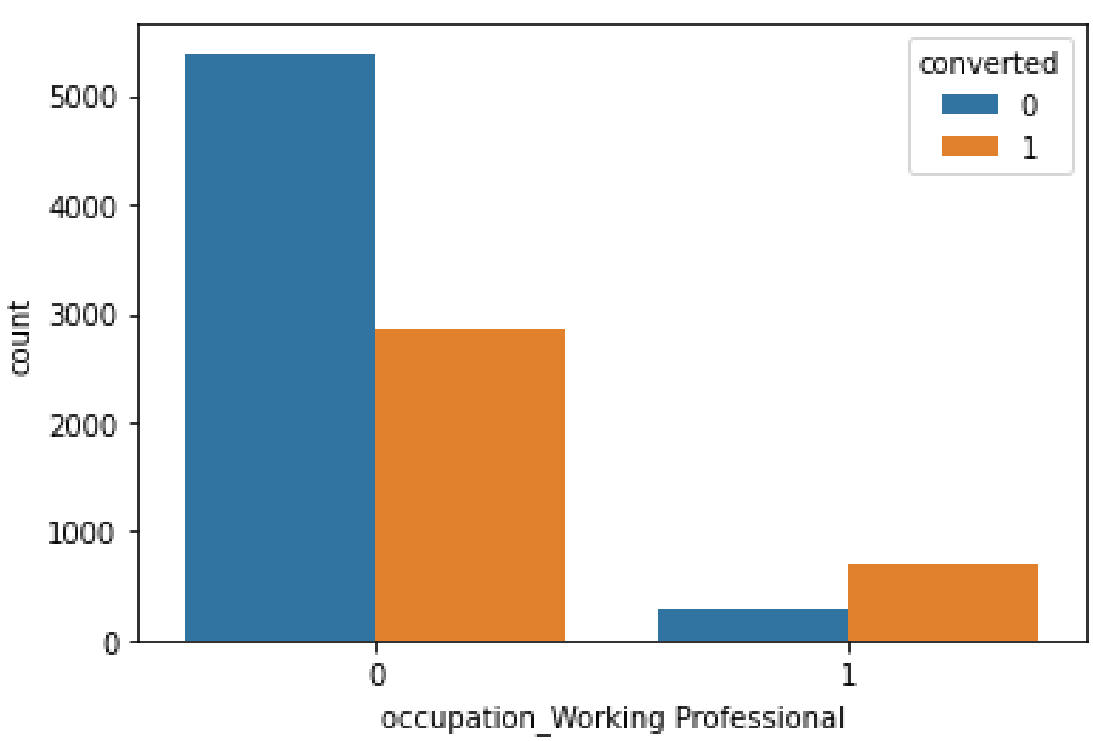


```
pd.crosstab(df.lead_source_Reference, df.converted)
```

converted	0	1
lead_source_Reference		
0	5635	3071
1	44	490

1)The Above countplot shows that, the lead source coming from reference has more probability of conversion, thus company can focus on people whose lead source is reference

2)The correlation matrix shows that people spending more time on website [column-total_time_spent_on_website] have higher rate of conversion, thus company should focus on people spending more time on website



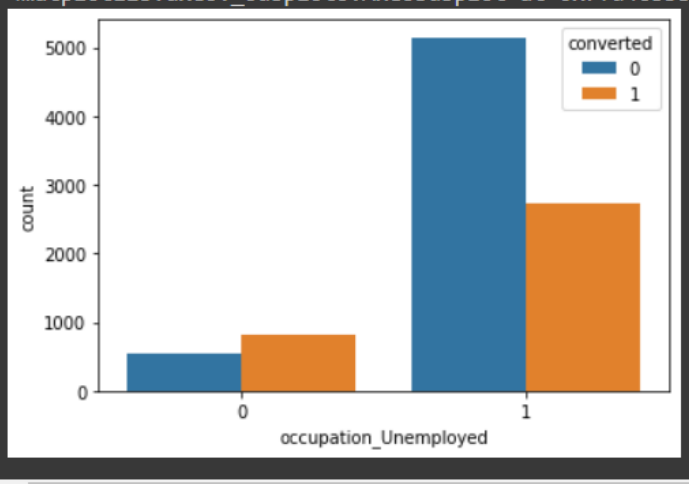
from above plot we can say that,

3)People having a working profession as Profession have higher probability of getting converted, The company can focus on people having working profession as Professional.

4. Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimise the rate of useless phone calls. Suggest a strategy they should employ at this stage.

```
sns.countplot(df['occupation_Unemployed'], hue=df['converted'])
```

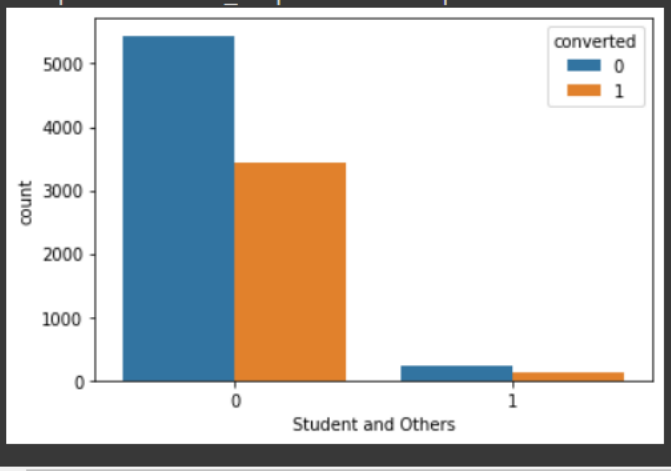
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pas
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4cbb8bdc0>
```



At a time when company wants to minimise the rate of useless phone calls, the team may not focus on people having occupation as unemployed. Because from the above plot we can understand that, the occupation_umemployed people are less converted

```
sns.countplot(occupation_analysis['Student and Others'], hue=df['converted'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass t
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7fd4cba36e20>
```



The people being Students and others have less probability of converting, lokks like they are already studying and not want to enroll in courses specially designed for the working professionals, so at time of minimizing the rate of useless phone calls, such people are ignored