

# Lead Scoring Case Study

# Problem Statement

- X Education sells online courses to industry professionals
- On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- X Education gets a lot of leads, its lead conversion rate is very poor. Only 30 %
- 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
- The objective is to identify the most promising leads

# Analysis Approach

1. Data cleaning and data manipulation
2. Exploratory Data Analysis
3. Feature Scaling , dummy variable creation
4. Encoding of the data
5. Creation of logistic regression model
6. Validation of the model
7. Conclusions

# Data cleaning and data manipulation

- Total rows: 9240, columns: 37
- Replaced Select value with Nan
- Dropped columns with only all unique values
- Dropped columns with more than 45% null values
- Replaced missing values with appropriate values
- Managed outliers
- Finally we have 14 columns

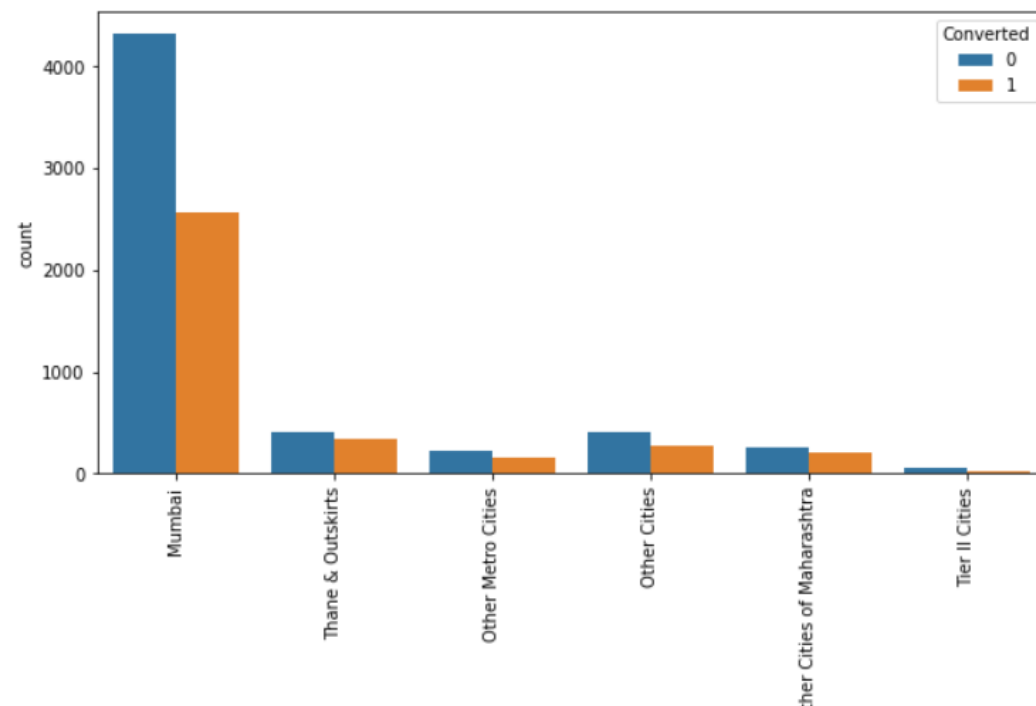
# EDA

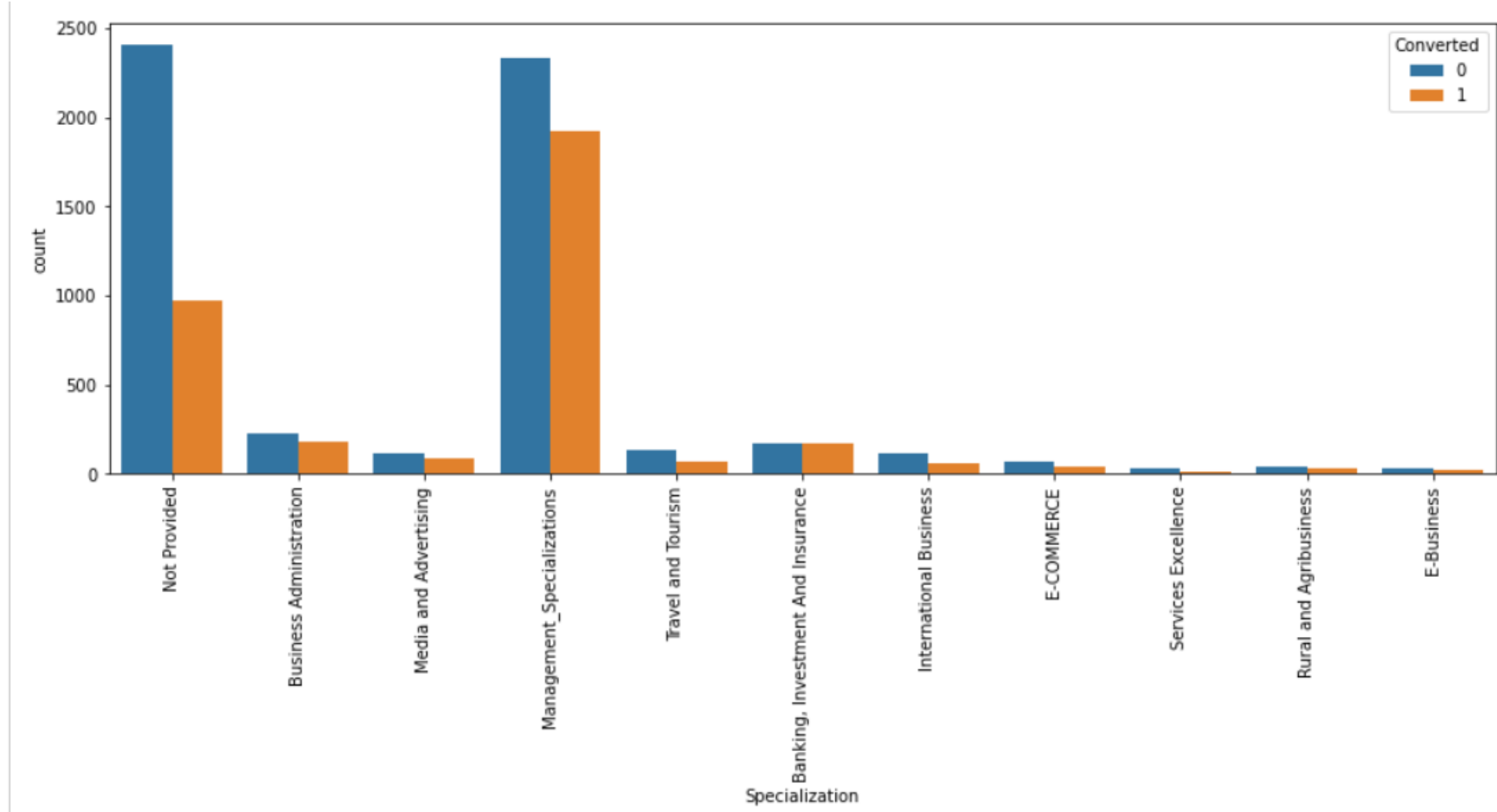
In [25]: *#plotting spread of City columnn after replacing NaN values*

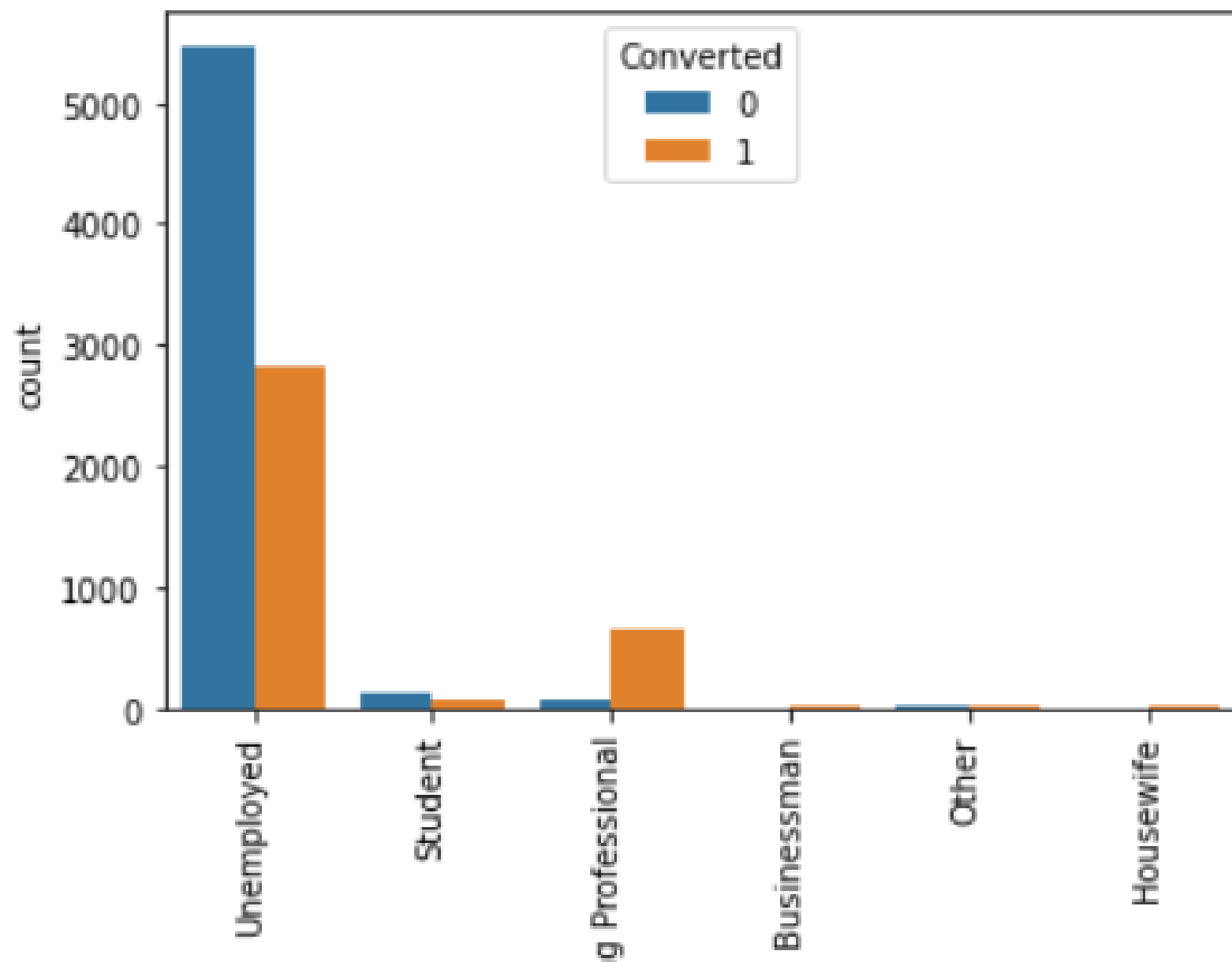
```
plt.figure(figsize=(10,5))
s1=sns.countplot(leads_new.City, hue=leads_new.Converted)
s1.set_xticklabels(s1.get_xticklabels(),rotation=90)
plt.show()
```

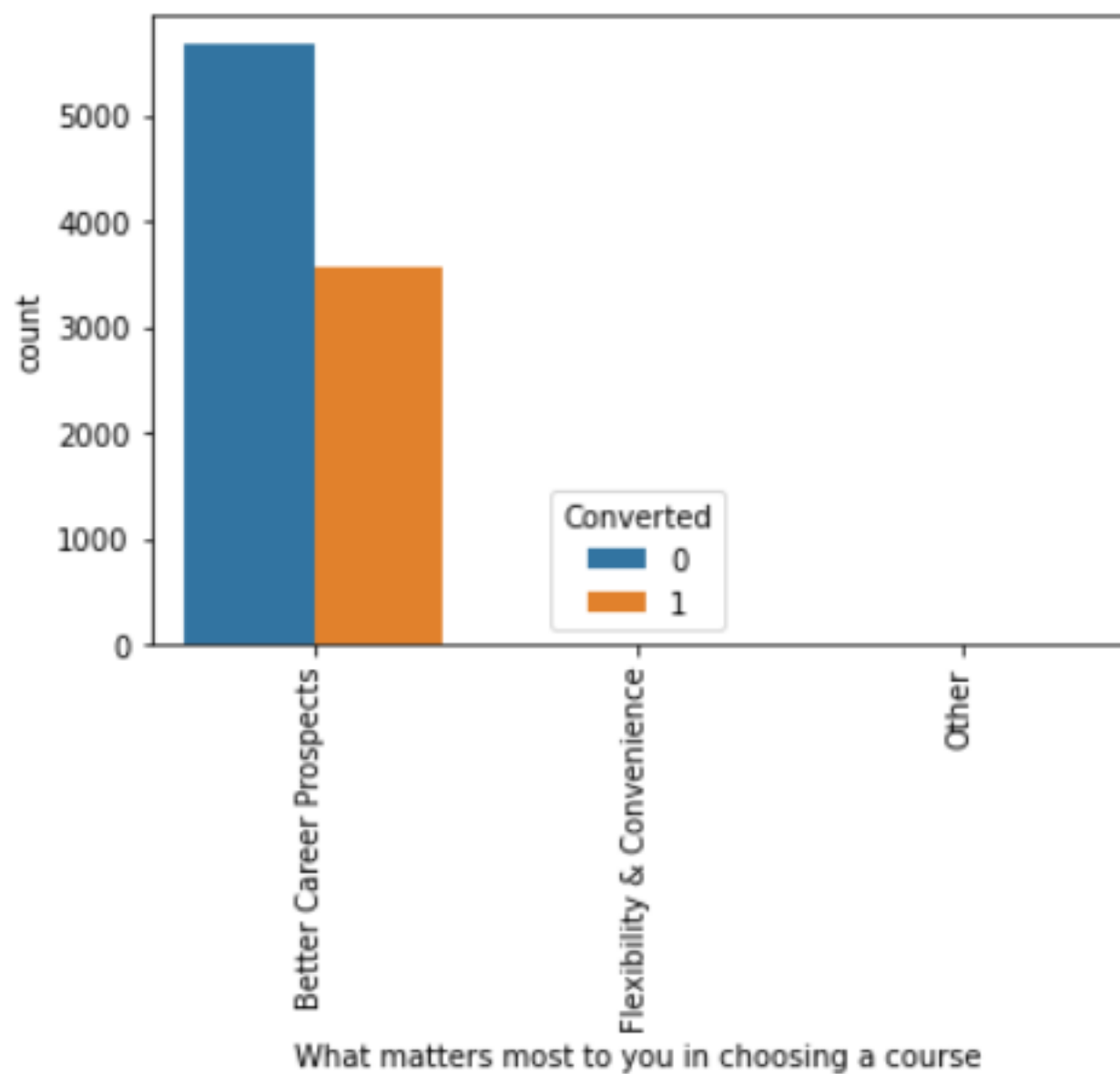
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following argument 'x'. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

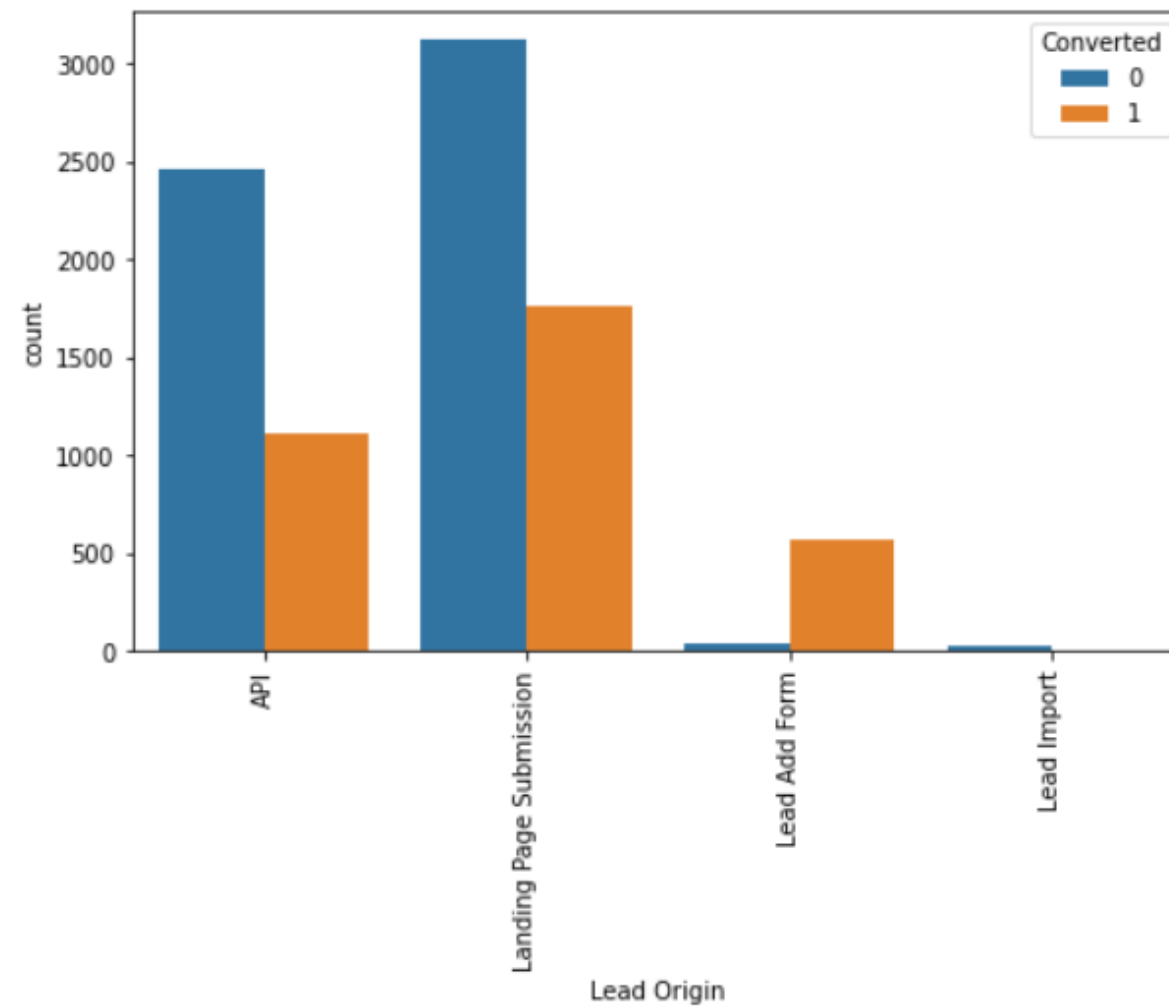








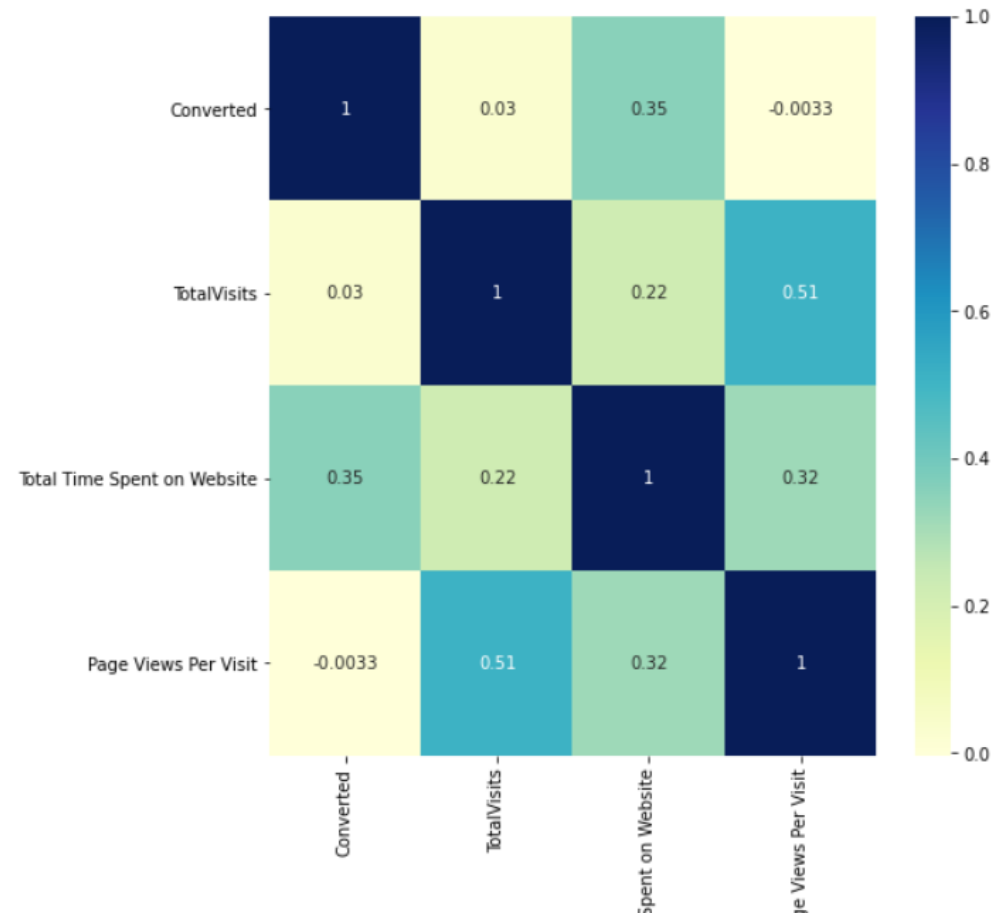




# Correlation

```
In [87]: # correlations of numeric values
```

```
plt.figure(figsize=(8,8))  
sns.heatmap(leads_new.corr(), cmap="YlGnBu", annot=True)  
plt.show()
```



# Final Result

*: # Observation: after running the model on the Test Data these are the numbers we got:*

*#Accuracy : 92.78%*

*#Sensitivity : 91.98%*

*#Specificity : 93.26%*

*: # based on the above numbers we can say that the model is able to predict the Conversion Rate very well*