

Name:P.Asha Belcilda

Rollno:225229104

Lab-7 Data Visualization in SeABorn

```
In [1]: #import necessary packages
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

1. Visualizing Statistical Relationships

Import train_upvote_mini.csv file

```
In [5]: df=pd.read_csv("/train_upvote_mini.csv")
df.head()
```

Out[5]:

	ID	Tag	Reputation	Answers	Username	Views	Upvotes
0	52664	a	3942.0	2.0	155623	7855.0	42.0
1	327662	a	26046.0	12.0	21781	55801.0	1175.0
2	468453	c	1358.0	4.0	56177	8067.0	60.0
3	96996	a	264.0	3.0	168793	27064.0	9.0
4	131465	c	4271.0	4.0	112223	13986.0	83.0

What is its size?

```
In [6]: df.shape
```

Out[6]: (15440, 7)

Show the types of each feature

```
In [7]: df.dtypes
```

```
Out[7]: ID           int64
Tag          object
Reputation   float64
Answers      float64
Username     int64
Views        float64
Upvotes      float64
dtype: object
```

How many unique "tag" available?

```
In [8]: df['Tag'].unique()
```

```
Out[8]: array(['a', 'c', 'r', 'j', 'p', 's', 'h', 'o', 'i', 'x'], dtype=object)
```

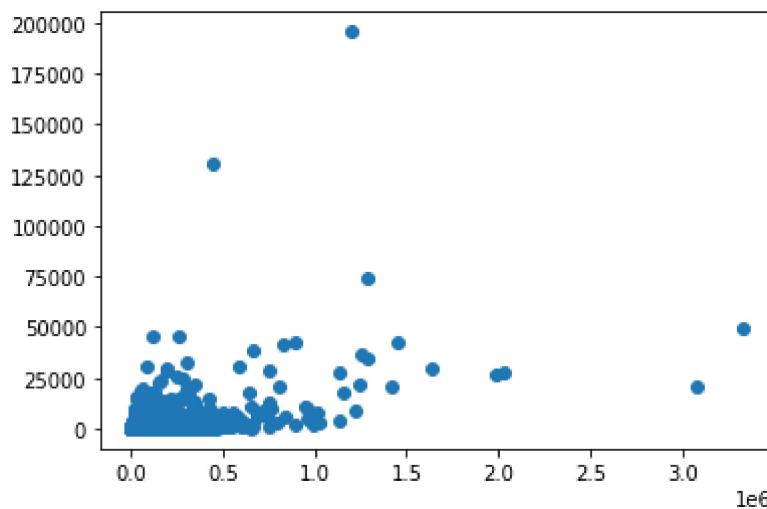
Visualize with Scatterplot

Does no. of views correlate no of upvotes?.

Show scatterplot (inherited from matplotlib) and relplot between "views" and "upvotes"

```
In [9]: plt.scatter(x=df['Views'],y=df['Upvotes'])
```

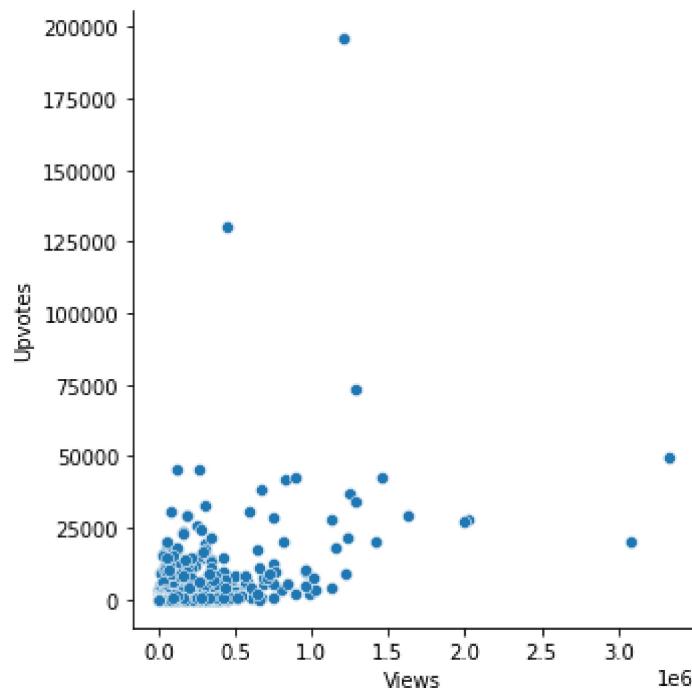
```
Out[9]: <matplotlib.collections.PathCollection at 0x7f0731677190>
```



Plot relplot between "Views" and "Upvotes"

```
In [10]: sns.relplot(data=df, x="Views", y="Upvotes")
```

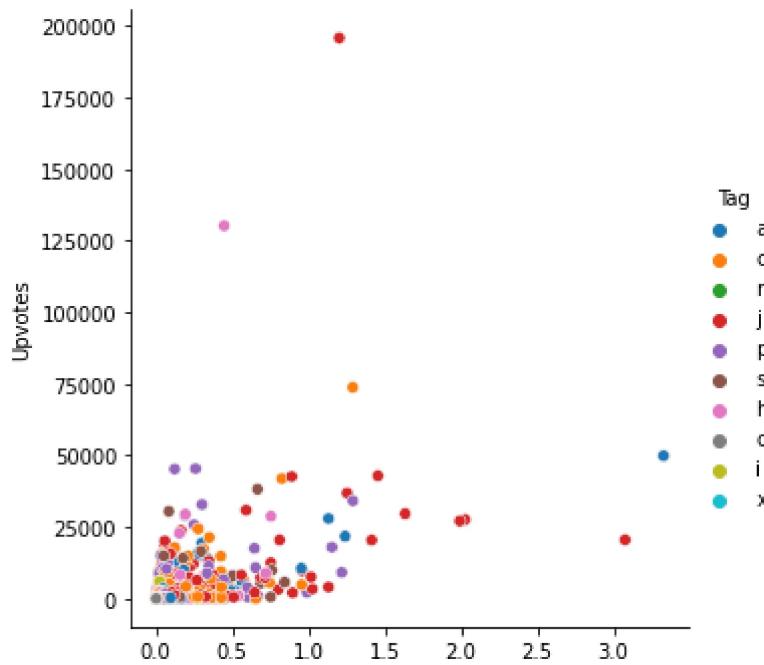
```
Out[10]: <seaborn.axisgrid.FacetGrid at 0x7f072ee57280>
```



Next, we want to see the tag associated with data.

Plot relplot between "Views" and "Upvotes" with hue as "Tag"

```
In [11]: sns.relplot(data=df , x="Views" , y="Upvotes" , hue='Tag')
```

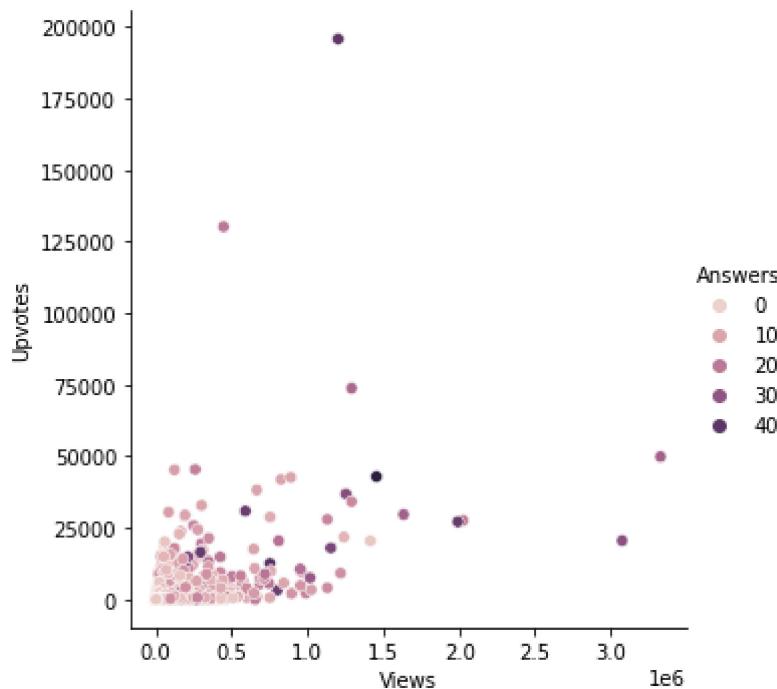


Hue Plot

Plot relplot between "Views" and "Upvotes" with hue as "Answers"

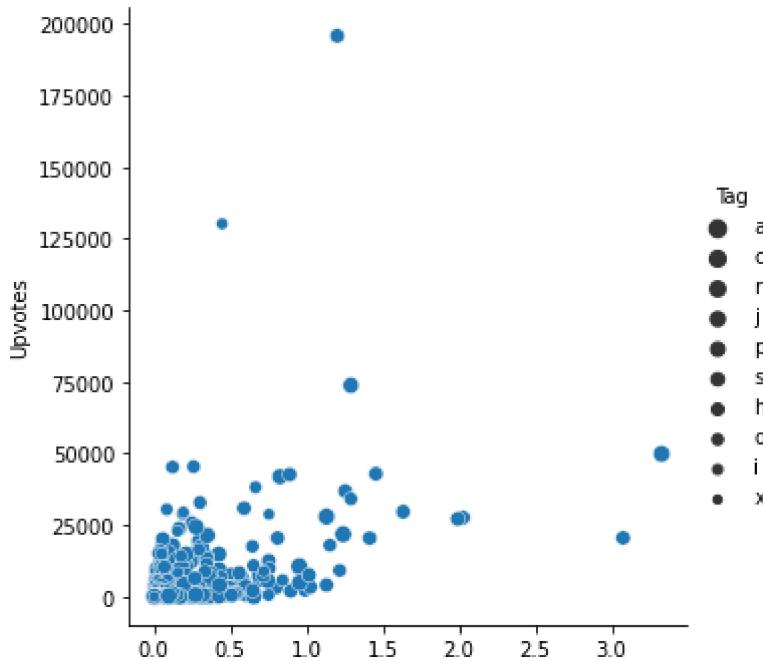
```
In [12]: sns.relplot(data=df , x="Views" , y="Upvotes" , hue='Answers')
```

```
Out[12]: <seaborn.axisgrid.FacetGrid at 0x7f072ed5bca0>
```



Plot relplot between "Views" and "Upvotes" with size as "Tag"

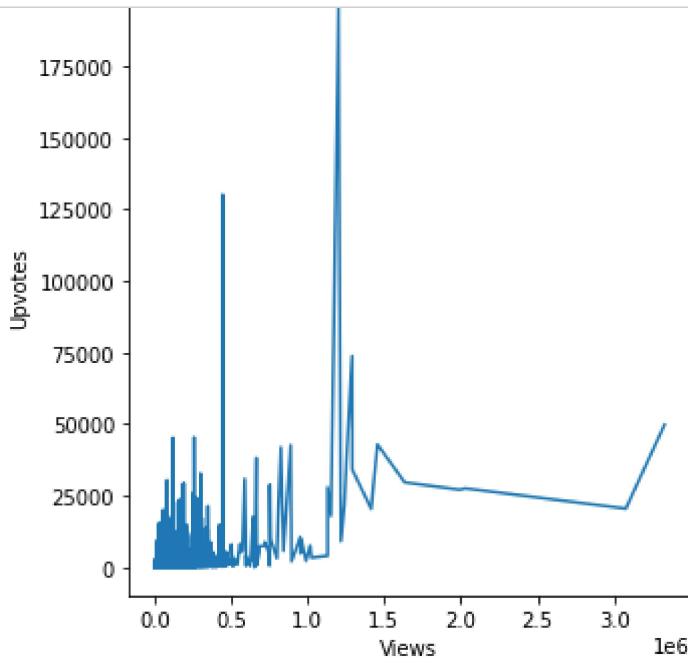
In [13]: `sns.relplot(data=df , x="Views" , y="Upvotes" , size='Tag')`



Does no of times question answered impact the no. of upvotes?

Plot line chart using relplot between "Answers" and "Upvotes"

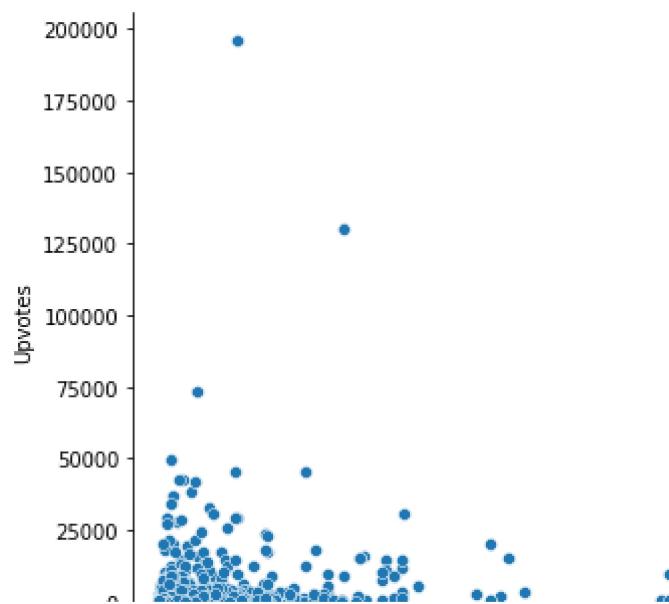
In [14]: `sns.relplot(data=df , x="Views" , y="Upvotes" , kind='line')`



Does Reputation score of question author impact no of upvotes?. Draw replot.

In [15]: `sns.relplot(data=df , x="Reputation" , y="Upvotes")`

Out[15]: <seaborn.axisgrid.FacetGrid at 0x7f072e93d940>



2. Visualizing Categorical Data

Various Categorical Plots in Seaborn

In [17]: `df1=pd.read_csv("/train_hr_mini.csv")
df1.head()`

Out[17]:

	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age
0	65438	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35
1	65141	Operations	region_22	Bachelor's	m	other	1	35
2	7513	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	35
3	2542	Sales & Marketing	region_23	Bachelor's	m	other	2	35
4	48945	Technology	region_26	Bachelor's	m	other	1	42

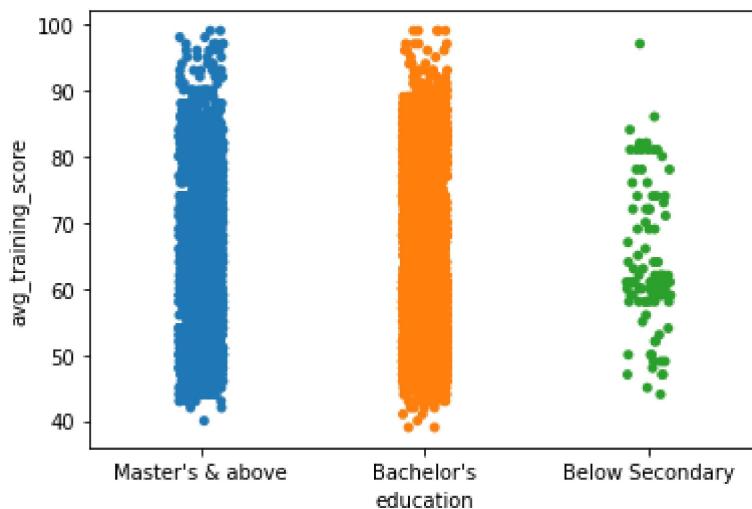
In [18]: `#shape
df1.shape`

Out[18]: (6397, 14)

Show Jitter plot between education and avg_training_score

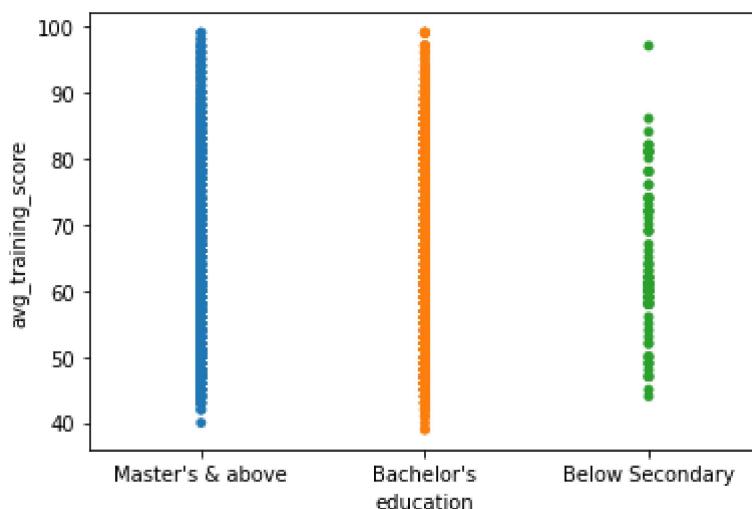
```
In [19]: sns.stripplot(data=df1,x="education",y="avg_training_score",jitter=True)
```

```
Out[19]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



```
In [20]: sns.stripplot(data=df1,x="education",y="avg_training_score",jitter=False)
```

```
Out[20]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



Swarm Plot

Plot Swarm plot between education category and avg_training_score

```
In [21]: sns.swarmplot(data=df1,x="education",y="avg_training_score")
```

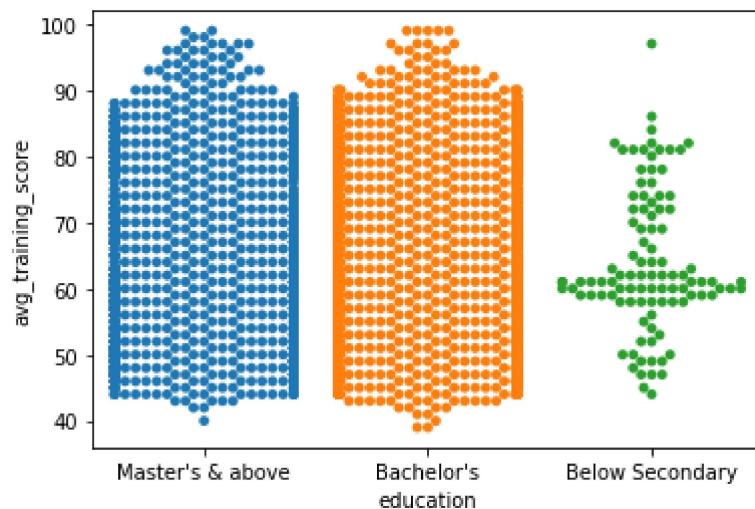
```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning:  
g: 74.2% of the points cannot be placed; you may want to decrease the size of t  
he markers or use stripplot.
```

```
    warnings.warn(msg, UserWarning)
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning:  
g: 88.1% of the points cannot be placed; you may want to decrease the size of t  
he markers or use stripplot.
```

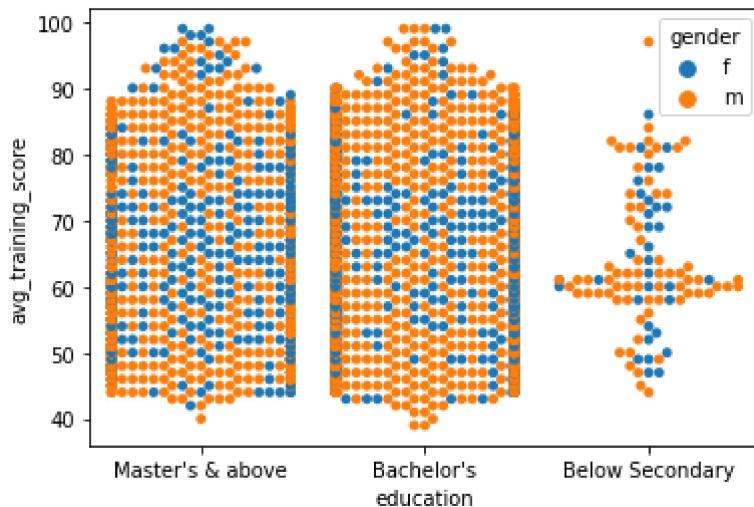
```
    warnings.warn(msg, UserWarning)
```

```
Out[21]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



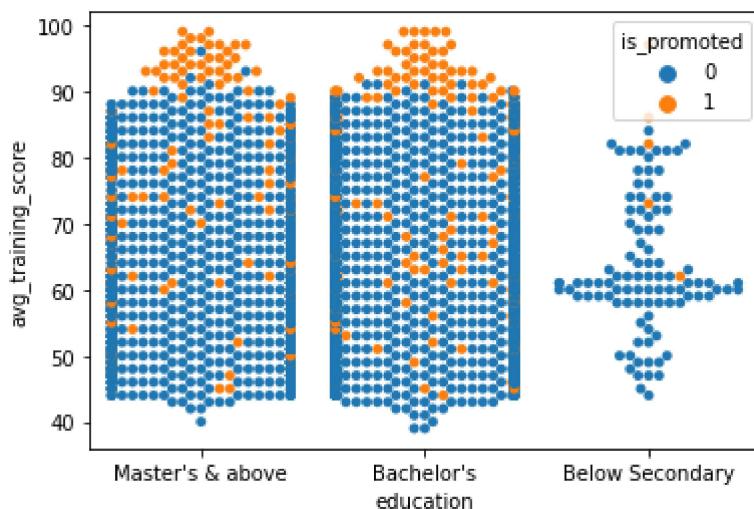
```
In [22]: import warnings  
warnings.filterwarnings('ignore')  
sns.swarmplot(data=df1,x="education",y="avg_training_score",hue="gender")
```

Out[22]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>



```
In [23]: sns.swarmplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

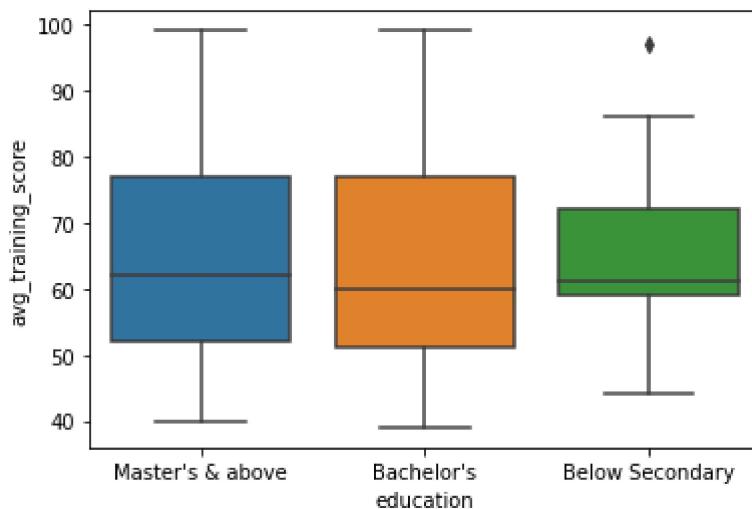
Out[23]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>



Box Plot

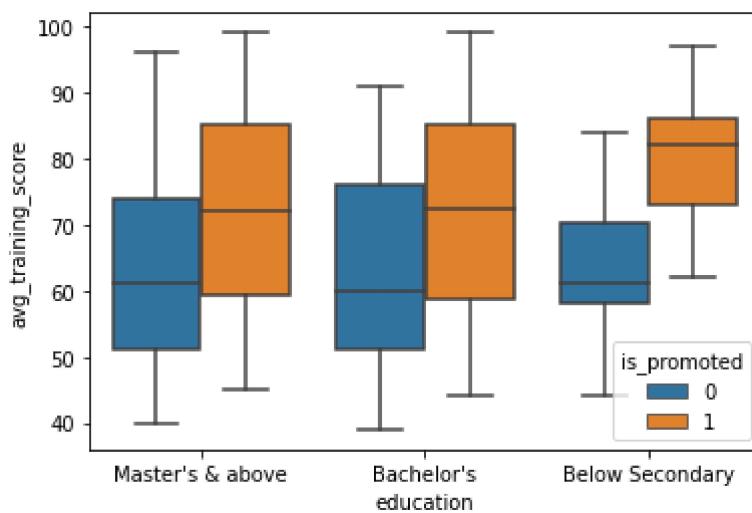
```
In [24]: sns.boxplot(data=df1,x="education",y="avg_training_score")
```

```
Out[24]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



```
In [25]: sns.boxplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

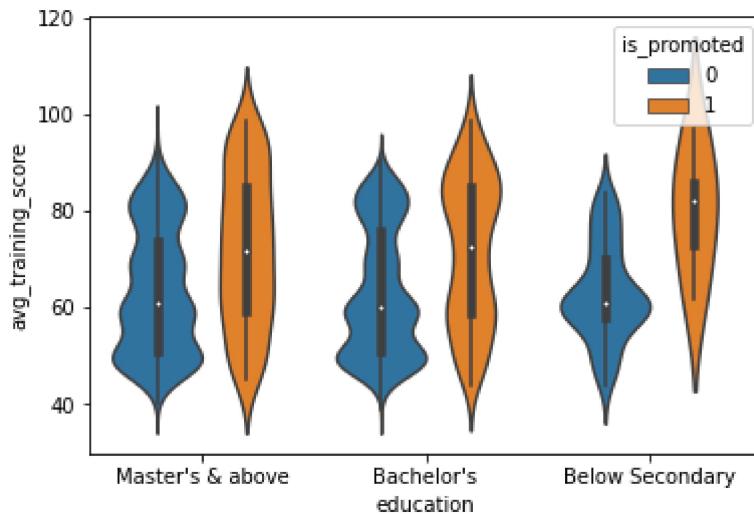
```
Out[25]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



Violin Plot

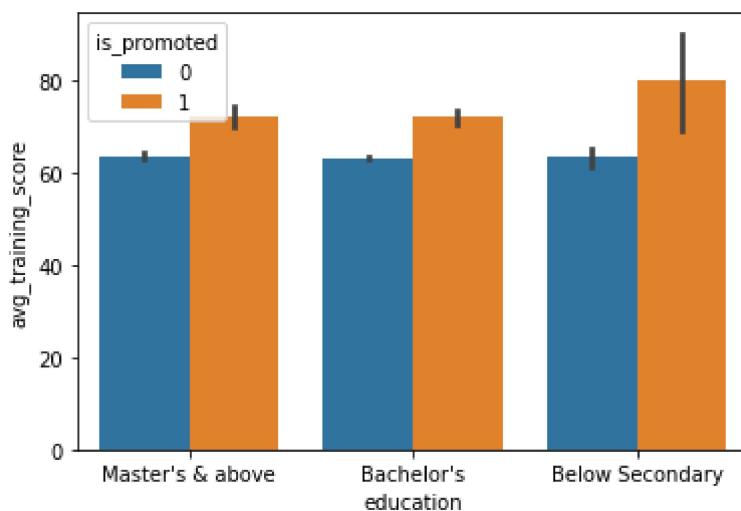
```
In [27]: sns.violinplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

```
Out[27]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



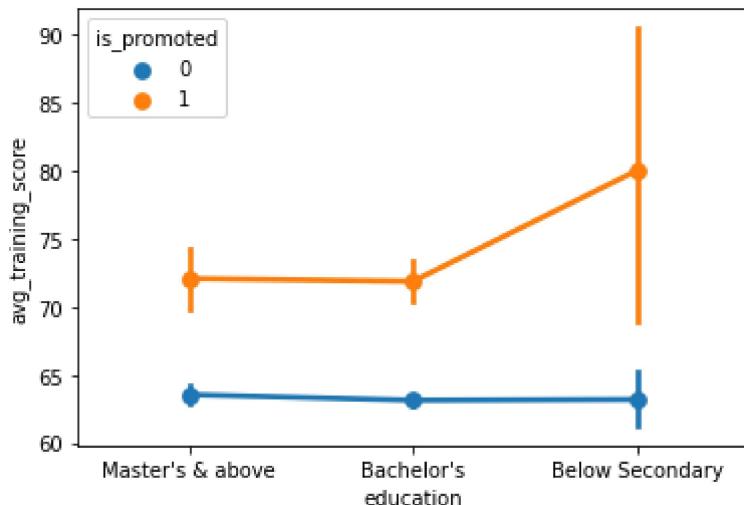
```
In [28]: sns.barplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

```
Out[28]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



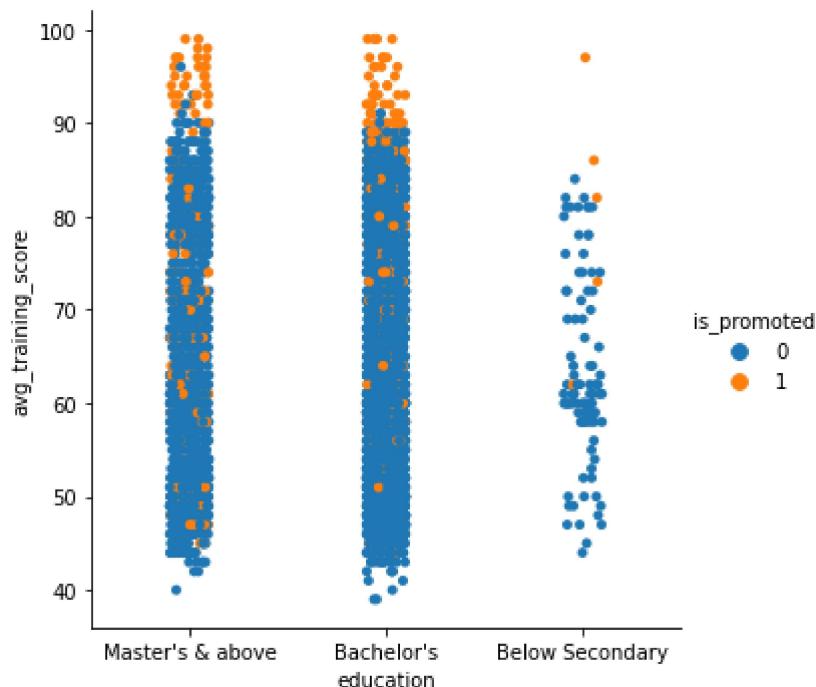
```
In [29]: sns.pointplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

```
Out[29]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



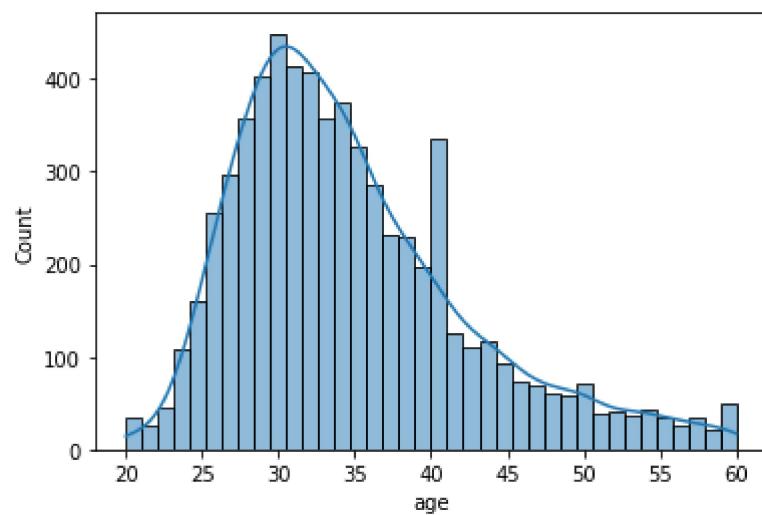
```
In [30]: sns.catplot(data=df1,x="education",y="avg_training_score",hue="is_promoted")
```

```
Out[30]: <seaborn.axisgrid.FacetGrid at 0x7f072c5e94f0>
```



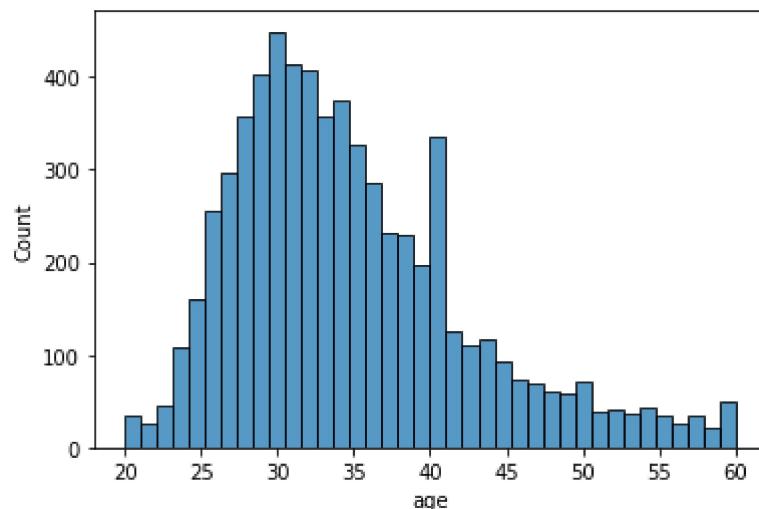
```
In [31]: sns.histplot(x='age',data=df1,kde=True)
```

```
Out[31]: <AxesSubplot:xlabel='age', ylabel='Count'>
```



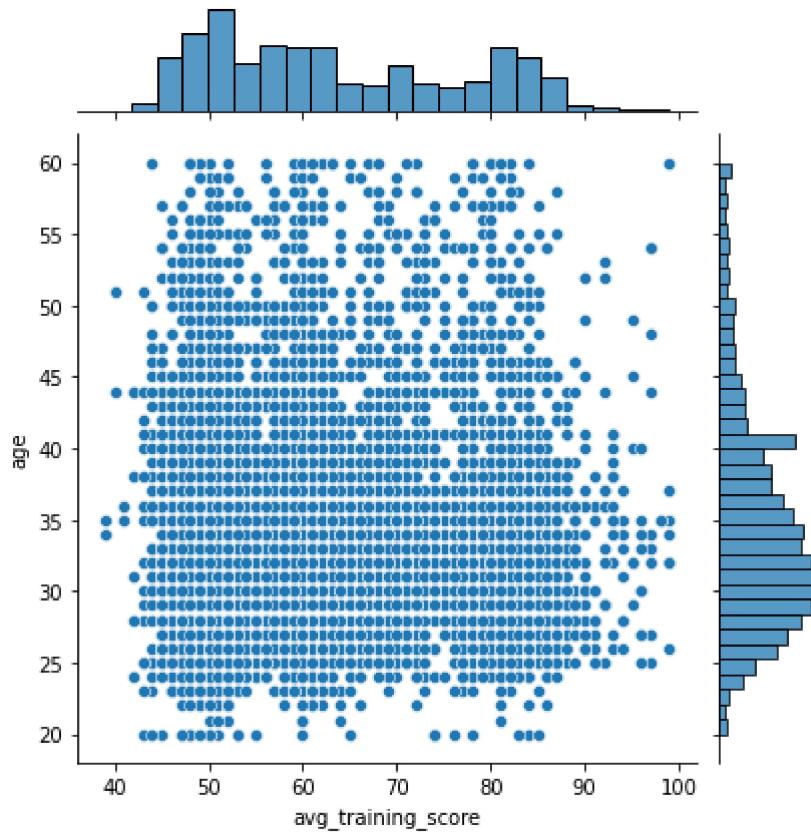
```
In [32]: sns.histplot(x='age', data=df1)
```

```
Out[32]: <AxesSubplot:xlabel='age', ylabel='Count'>
```



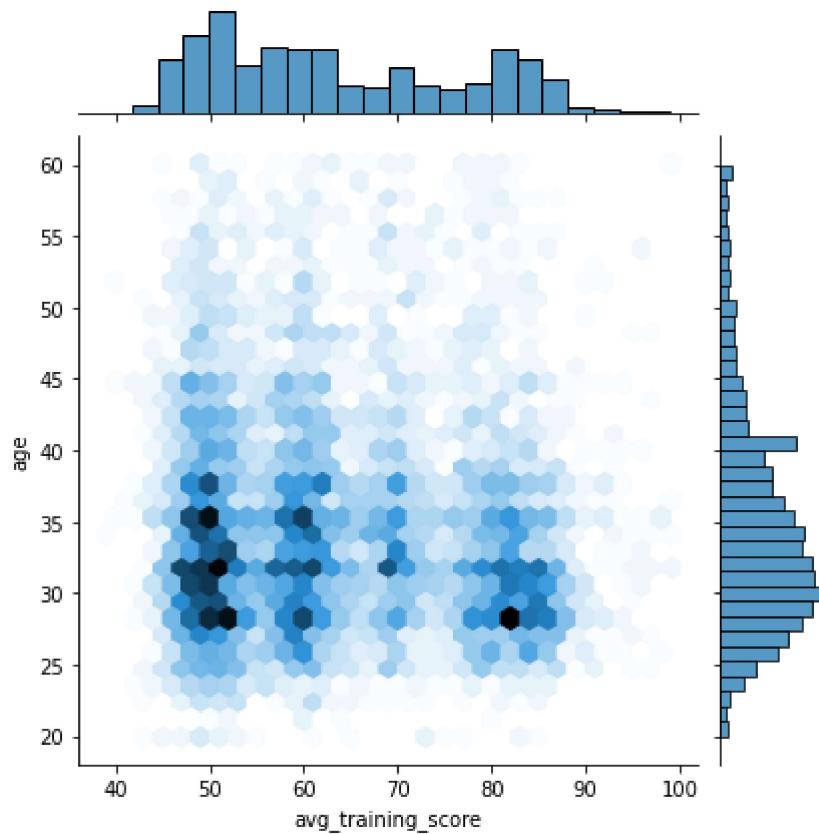
```
In [34]: sns.jointplot(x="avg_training_score",y="age",data=df1)
```

```
Out[34]: <seaborn.axisgrid.JointGrid at 0x7f072a259af0>
```



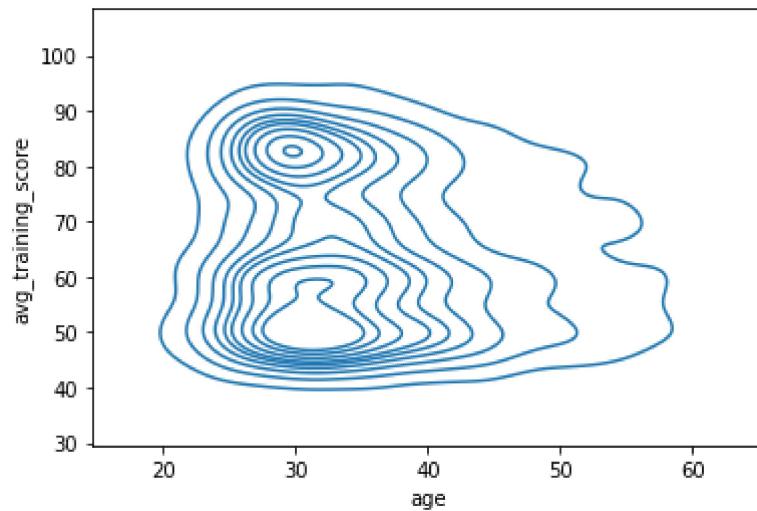
```
In [35]: sns.jointplot(x="avg_training_score",y="age",kind='hex',data=df1)
```

```
Out[35]: <seaborn.axisgrid.JointGrid at 0x7f072a21fc10>
```



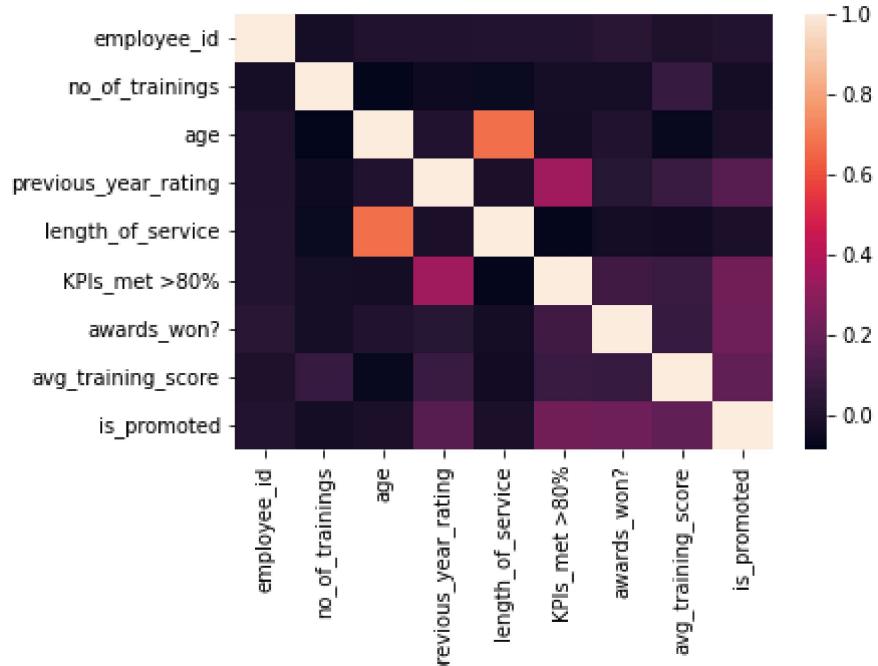
```
In [36]: sns.kdeplot(x="age",y="avg_training_score",data=df1)
```

```
Out[36]: <AxesSubplot:xlabel='age', ylabel='avg_training_score'>
```



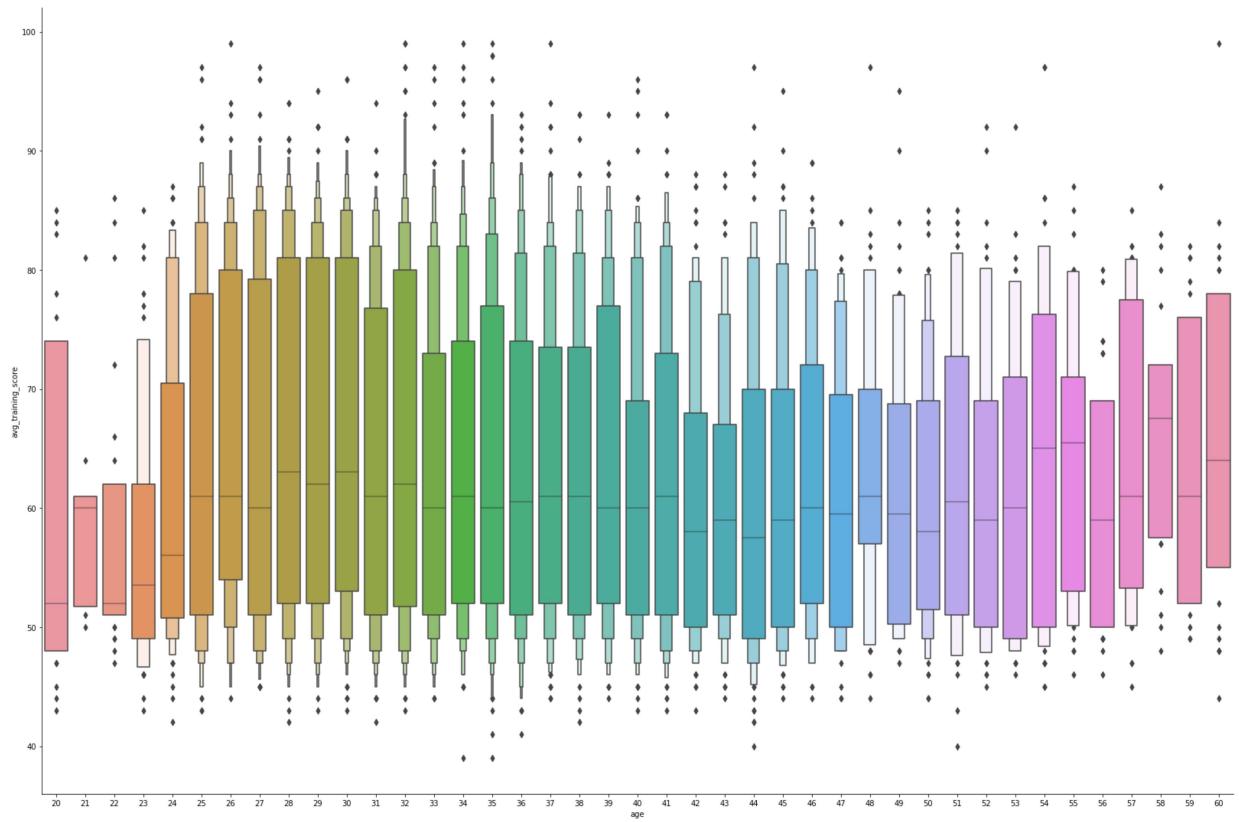
In [38]: `sns.heatmap(df1.corr())`

Out[38]: <AxesSubplot:>



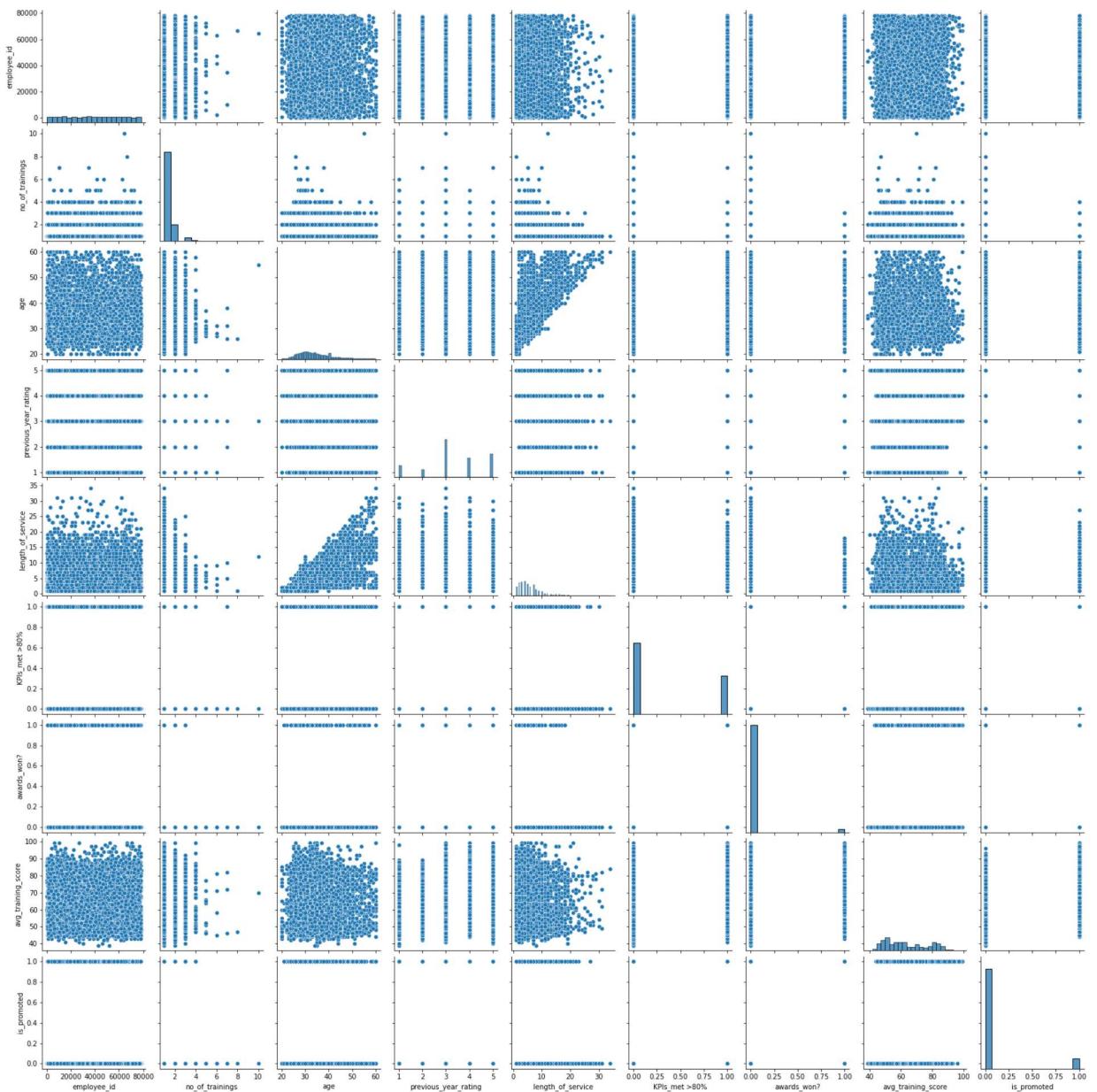
In [41]: `sns.catplot(x='age',y='avg_training_score',data=df1,kind='boxen',height=15, aspect=1)`

Out[41]: <seaborn.axisgrid.FacetGrid at 0x7f0729c87b50>



In [42]: `sns.pairplot(df1)`

Out[42]: <seaborn.axisgrid.PairGrid at 0x7f072a2b1a00>



In []: