

## LAB 7 : DATA VISUALIZATION IN SEABORN

DINESH KUMAR K

225229108

In [2]:

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
```

### 1. Visualizing Statistical Relationships

In [3]:

```
1 data = pd.read_csv('train_upvote_mini.csv')
2 data.head()
```

Out[3]:

	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

In [4]:

```
1 data.shape
```

Out[4]: (15440, 7)

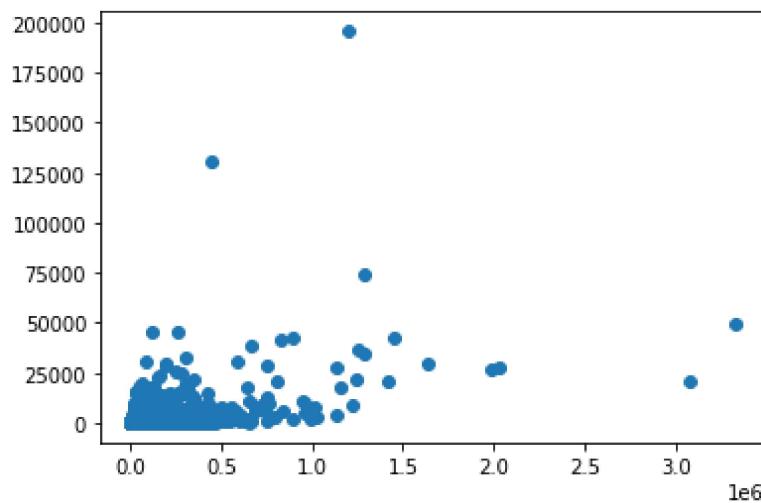
In [5]:

```
1 data['Tag'].unique()
```

Out[5]: array(['a', 'c', 'r', 'j', 'p', 's', 'h', 'o', 'i', 'x'], dtype=object)

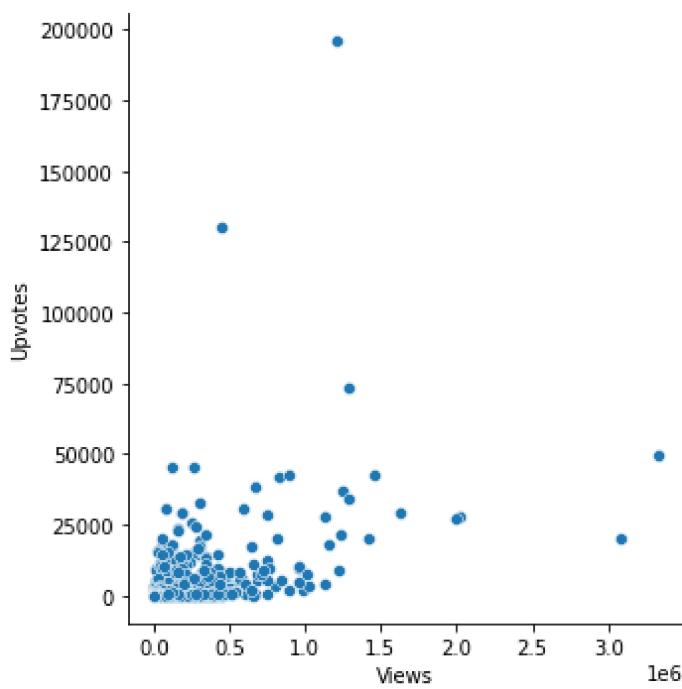
```
In [6]: 1 plt.scatter(x=data['Views'],y=data['Upvotes'])
```

```
Out[6]: <matplotlib.collections.PathCollection at 0x7f152220e700>
```



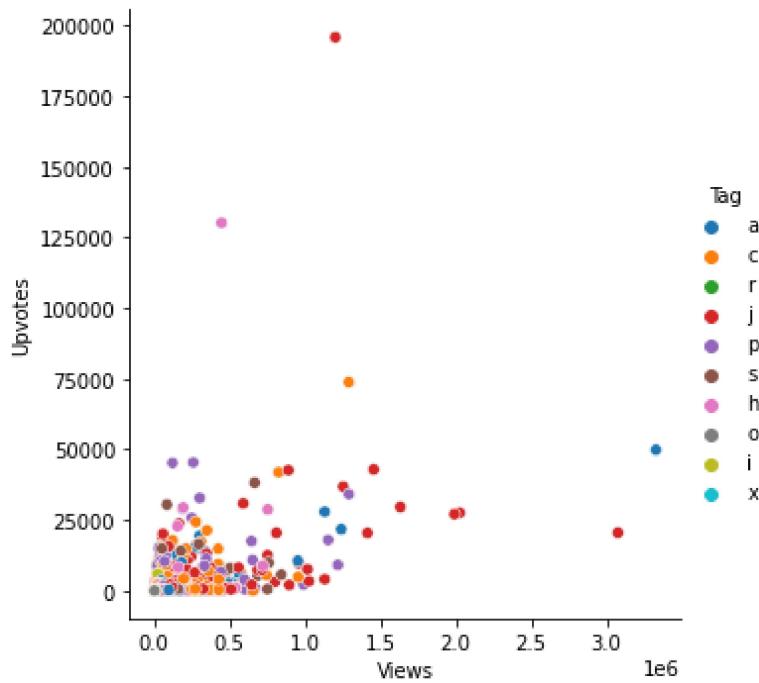
```
In [7]: 1 sns.relplot(x=data['Views'],y=data['Upvotes'])
```

```
Out[7]: <seaborn.axisgrid.FacetGrid at 0x7f1522356910>
```



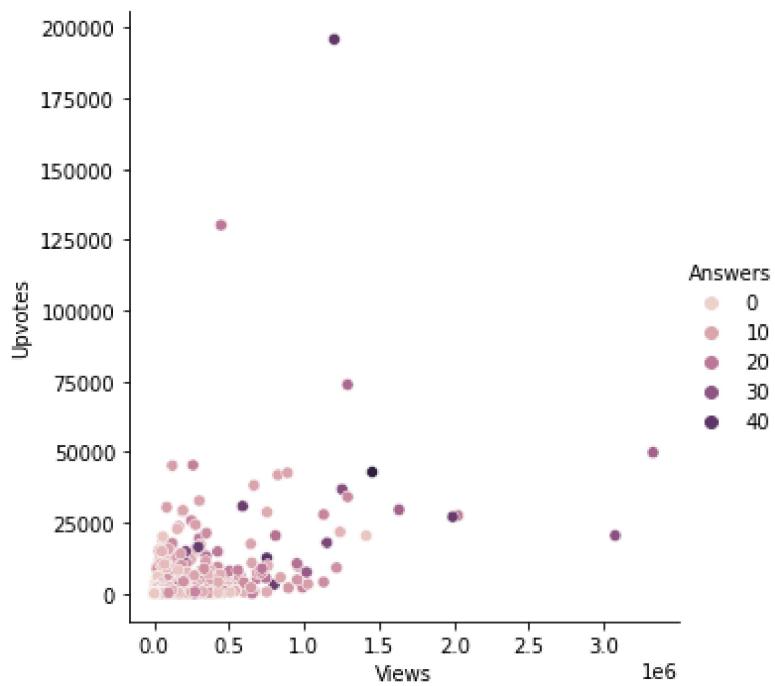
```
In [13]: 1 sns.relplot(data=data,x='Views',y='Upvotes',hue='Tag')
```

Out[13]: <seaborn.axisgrid.FacetGrid at 0x7f1550a86610>



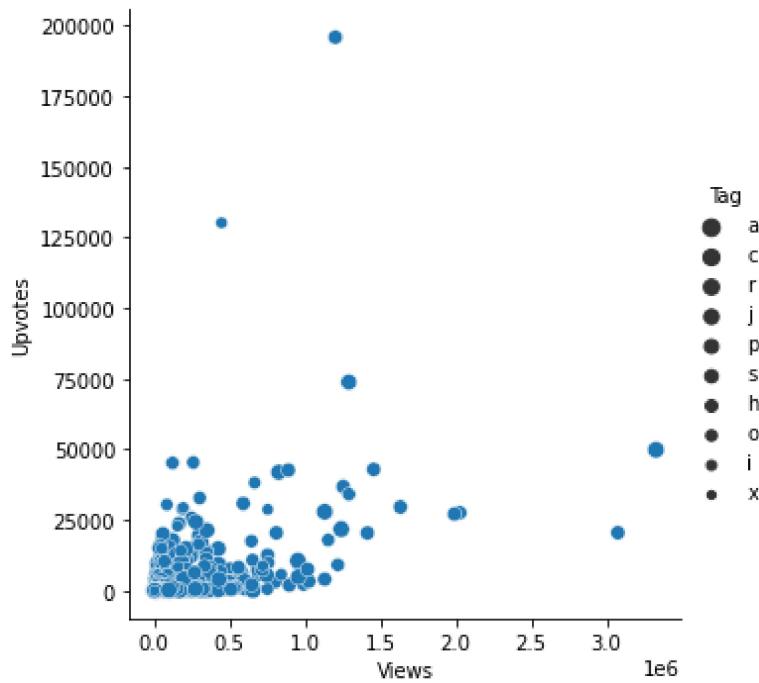
```
In [14]: 1 sns.relplot(data=data,x='Views',y='Upvotes',hue='Answers')
```

Out[14]: <seaborn.axisgrid.FacetGrid at 0x7f151f867880>



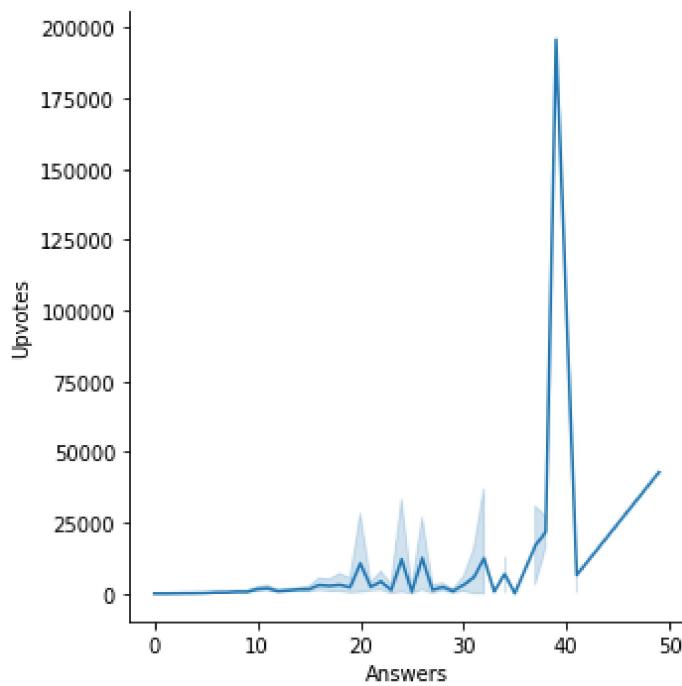
```
In [15]: 1 sns.relplot(data=data,x='Views',y='Upvotes',size='Tag')
```

Out[15]: <seaborn.axisgrid.FacetGrid at 0x7f151f888ac0>



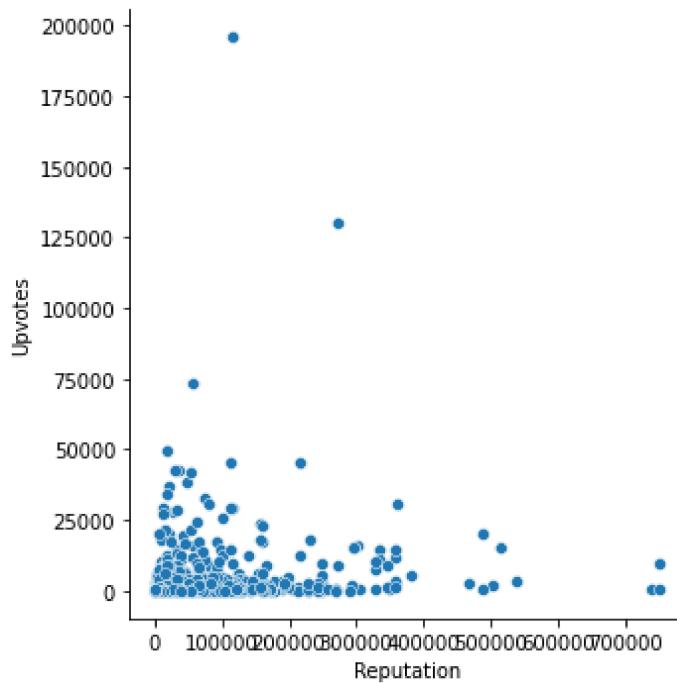
```
In [17]: 1 sns.relplot(data=data,x='Answers',y='Upvotes',kind='line')
```

Out[17]: <seaborn.axisgrid.FacetGrid at 0x7f151e9a5fa0>



```
In [19]: 1 sns.relplot(data=data, x="Reputation", y="Upvotes")
```

```
Out[19]: <seaborn.axisgrid.FacetGrid at 0x7f151ea2b430>
```



```
In [20]: 1 df1=pd.read_csv('train_hr_mini.csv')
2 df1.head()
```

```
Out[20]:
```

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

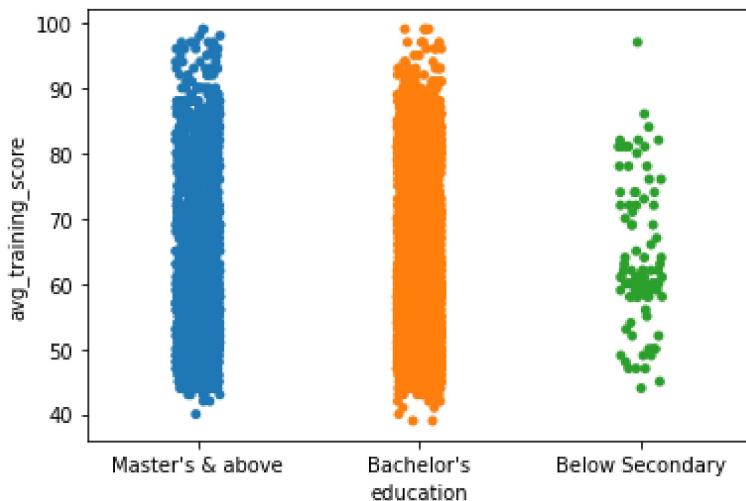
◀ ▶

```
In [22]: 1 df1.shape
```

```
Out[22]: (6397, 14)
```

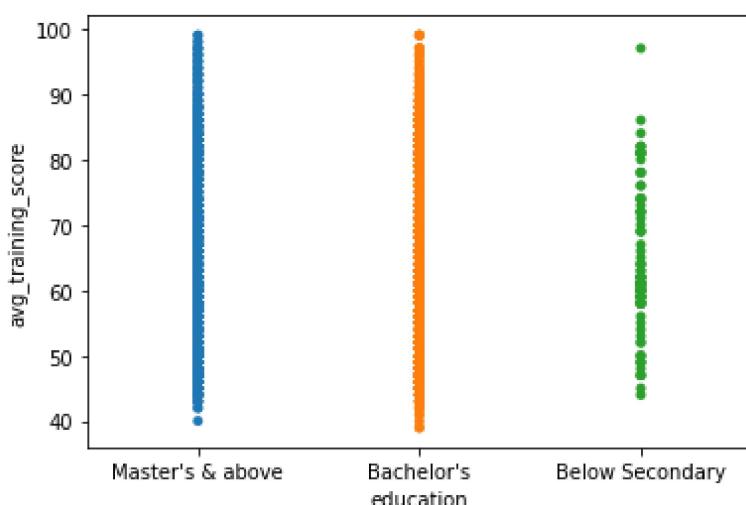
```
In [23]: 1 sns.stripplot(data=df1, x="education", y="avg_training_score", jitter=True)
```

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



```
In [24]: 1 sns.stripplot(data=df1, x="education", y="avg_training_score", jitter=False)
```

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



```
In [25]: 1 sns.swarmplot(data=df1, x="education", y="avg_training_score")
```

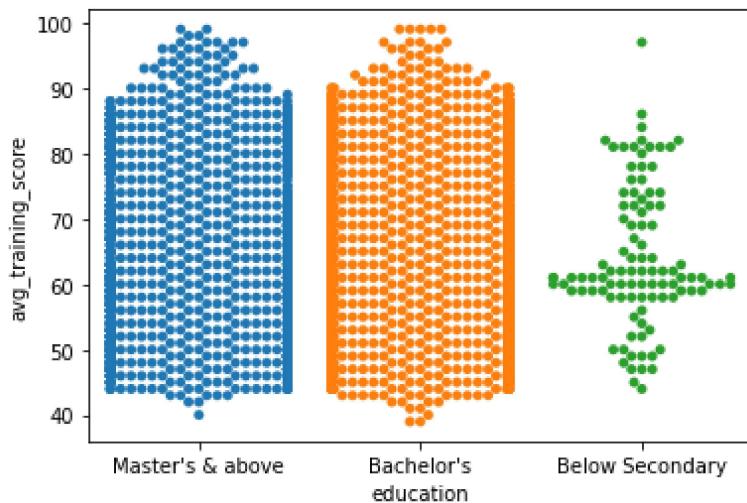
```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning:  
g: 74.2% of the points cannot be placed; you may want to decrease the size of  
the markers or use stripplot.
```

```
    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  
the markers or use stripplot.
```

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

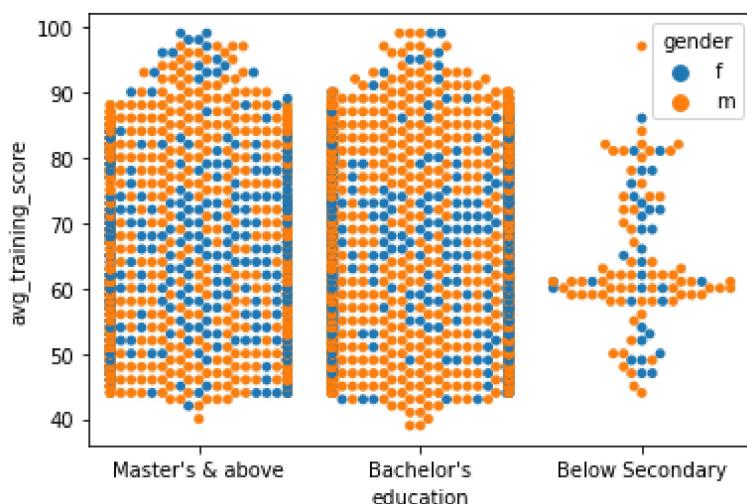
```
Out[25]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



```
In [26]: 1 sns.swarmplot(data=df1, x="education", y="avg_training_score", hue ='gender')

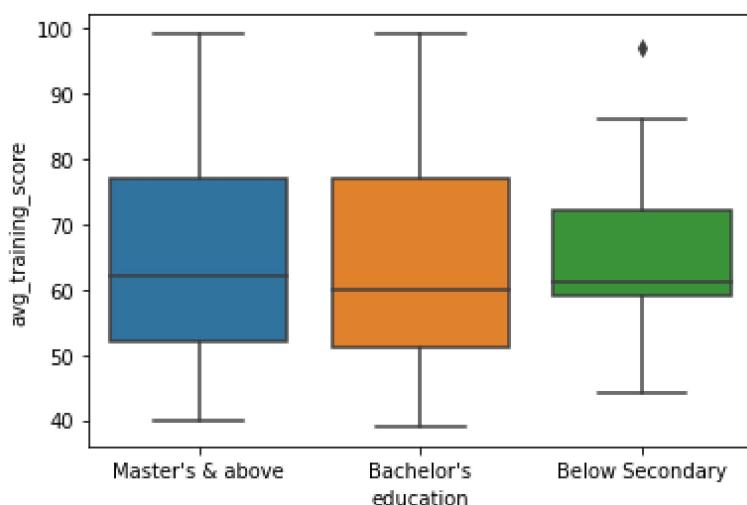
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning:
g: 74.2% of the points cannot be placed; you may want to decrease the size of
the markers or use stripplot.
    warnings.warn(msg, UserWarning)
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning:
g: 88.1% of the points cannot be placed; you may want to decrease the size of
the markers or use stripplot.
    warnings.warn(msg, UserWarning)
```

Out[26]: <AxesSubplot:xlabel='education', ylabel='avg\_training\_score'>



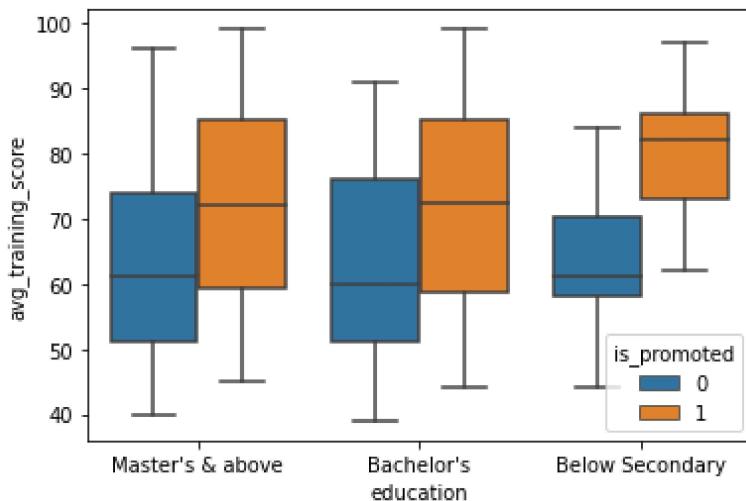
```
In [27]: 1 sns.boxplot(data=df1, x="education", y="avg_training_score")
```

Out[27]: <AxesSubplot:xlabel='education', ylabel='avg\_training\_score'>



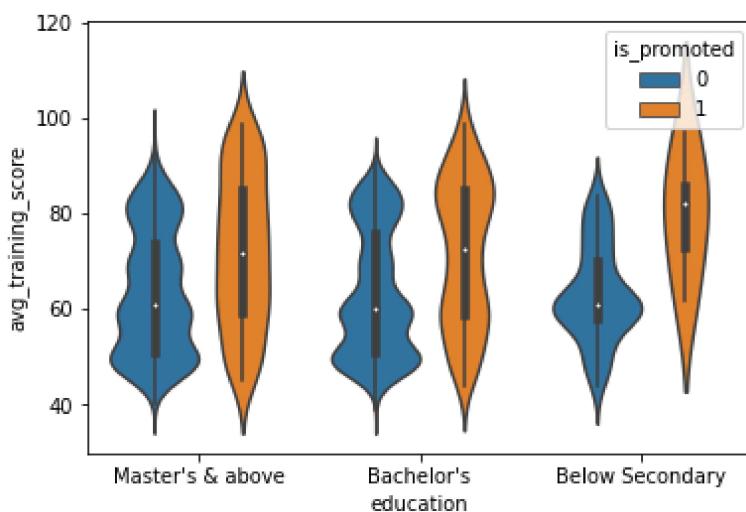
```
In [28]: 1 sns.boxplot(data=df1, x="education", y="avg_training_score", hue = 'is_promoted')
```

Out[28]: <AxesSubplot:xlabel='education', ylabel='avg\_training\_score'>



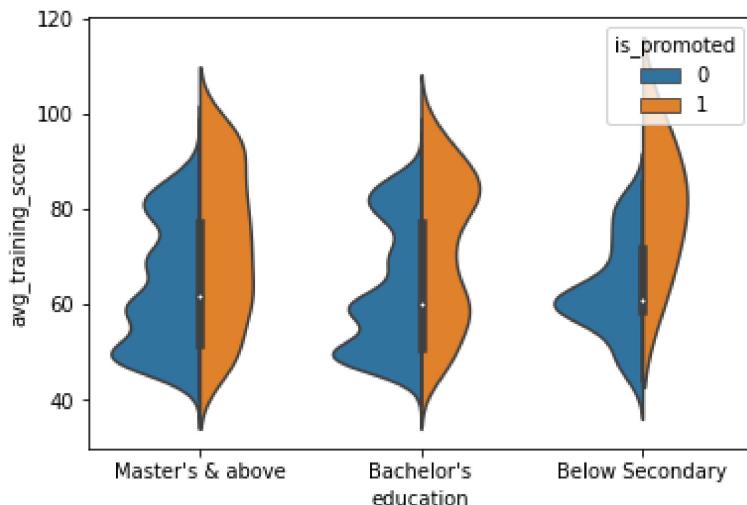
```
In [29]: 1 sns.violinplot(data=df1, x="education", y="avg_training_score", hue = 'is_promoted')
```

Out[29]: <AxesSubplot:xlabel='education', ylabel='avg\_training\_score'>



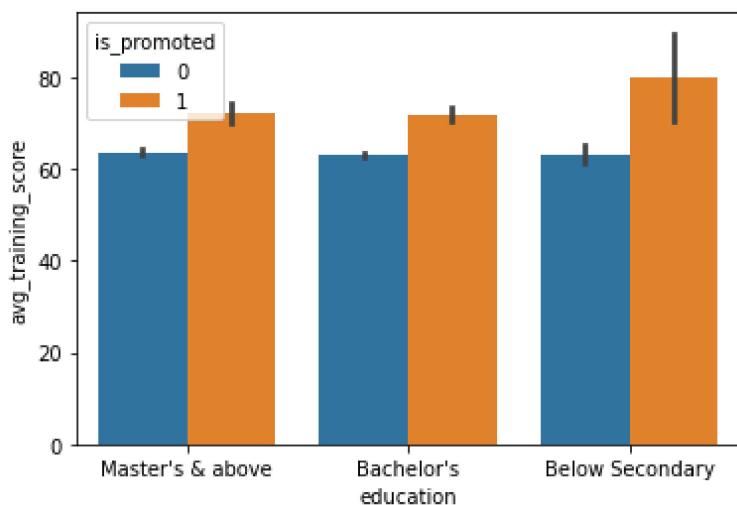
```
In [31]: 1 sns.violinplot(data=df1, x="education", y="avg_training_score", hue ='is_promoted')
```

```
Out[31]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



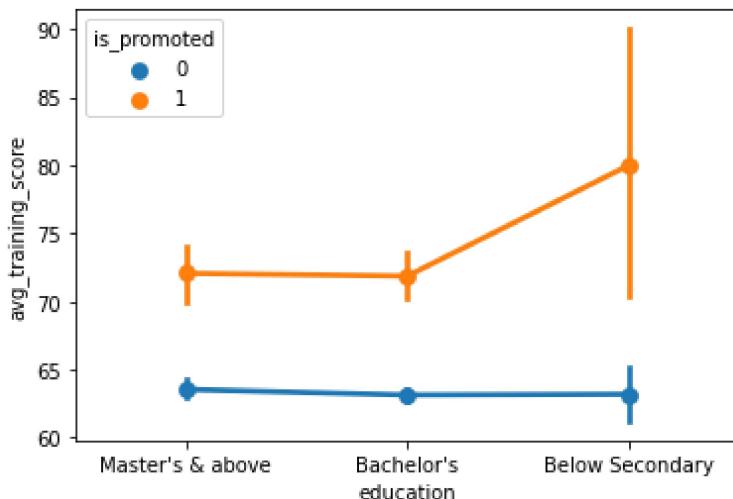
```
In [32]: 1 sns.barplot(data=df1, x="education", y="avg_training_score", hue ='is_promoted')
```

```
Out[32]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



In [33]: 1 sns.pointplot(data=df1, x="education", y="avg\_training\_score", hue ='is\_promoted')

Out[33]: <AxesSubplot:xlabel='education', ylabel='avg\_training\_score'>



In [34]: 1 sns.catplot(data=df1, x="education", y="avg\_training\_score", hue ='is\_promoted')

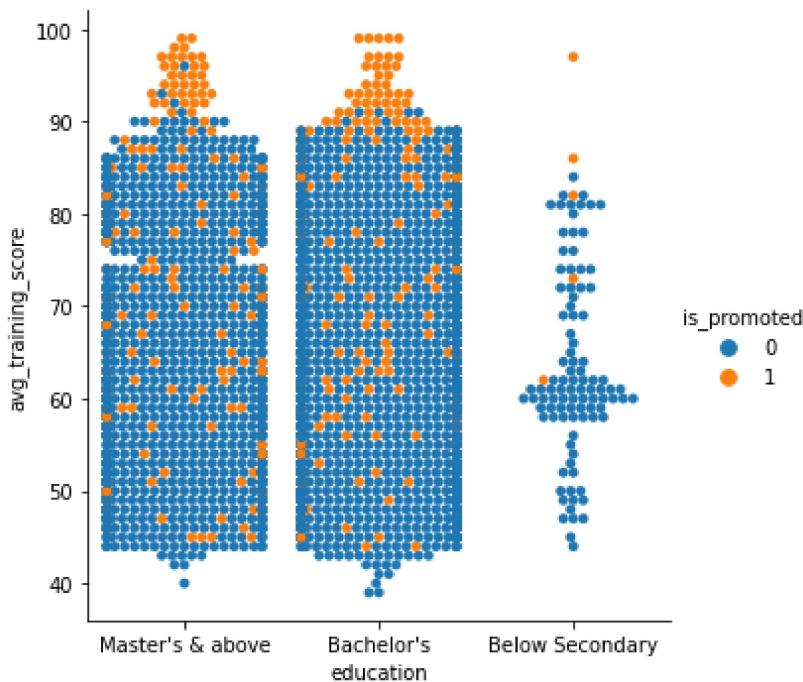
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 56.8% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.

warnings.warn(msg, UserWarning)

/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 81.5% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.

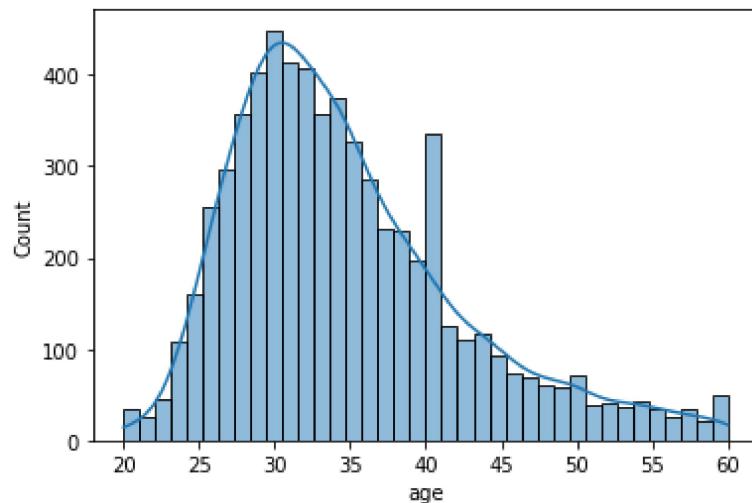
warnings.warn(msg, UserWarning)

Out[34]: <seaborn.axisgrid.FacetGrid at 0x7f151a776fd0>



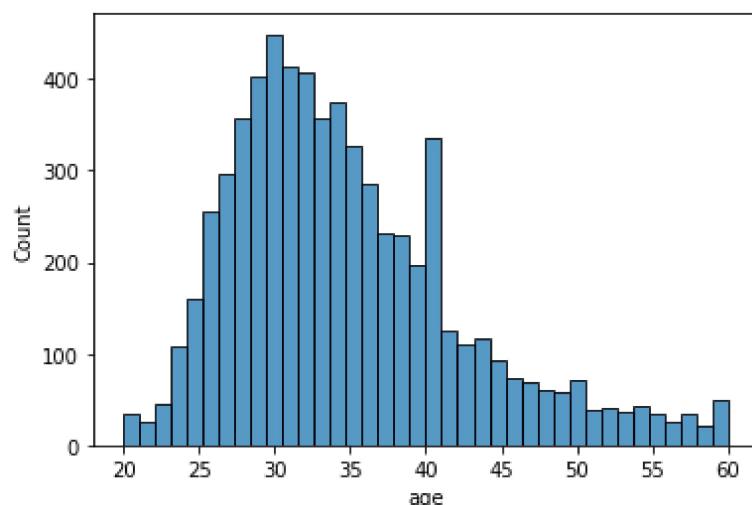
```
In [35]: 1 sns.histplot(x='age', data=df1, kde=True)
```

Out[35]: <AxesSubplot:xlabel='age', ylabel='Count'>



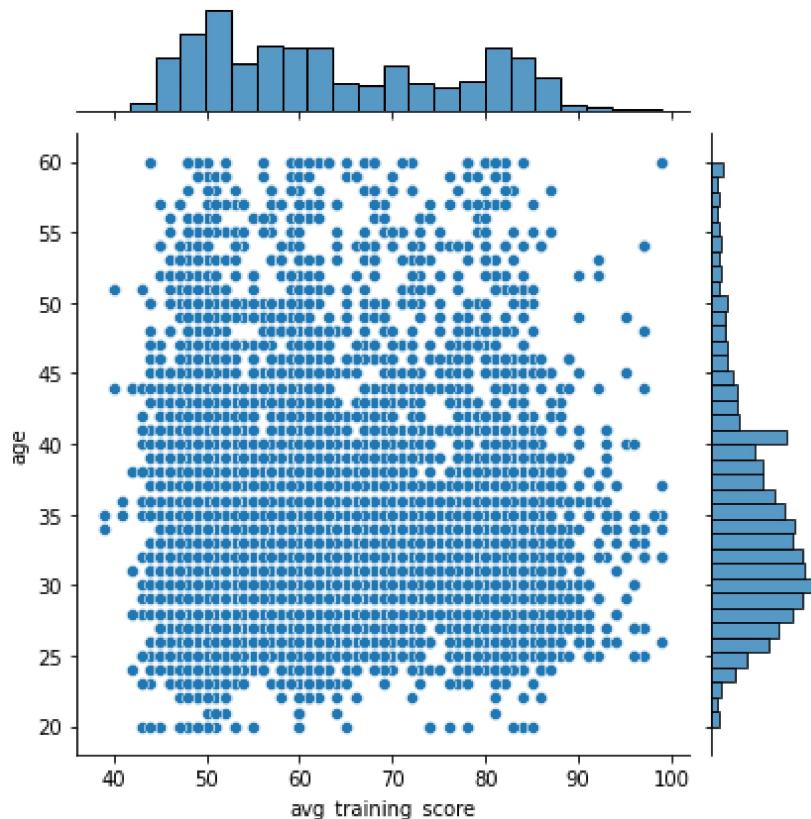
```
In [36]: 1 sns.histplot(x='age', data=df1)
```

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



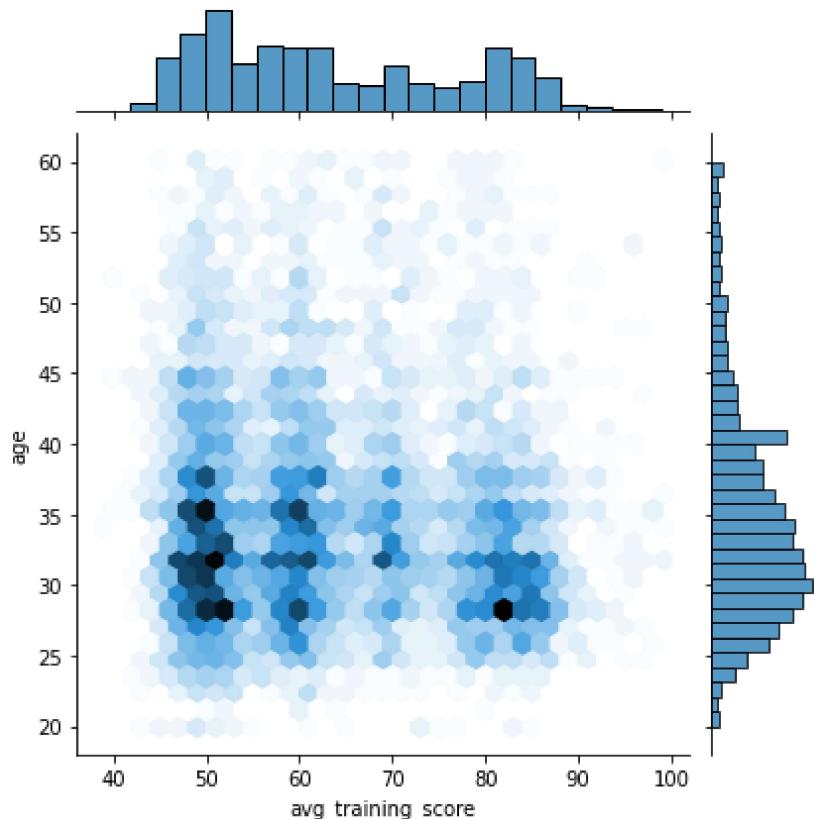
```
In [37]: 1 sns.jointplot(x='avg_training_score',y='age', data=df1 )
```

```
Out[37]: <seaborn.axisgrid.JointGrid at 0x7f151e590b80>
```



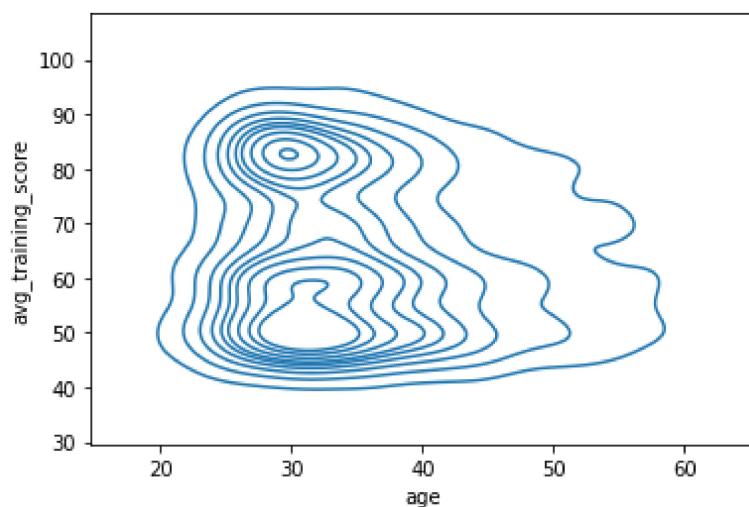
```
In [38]: 1 sns.jointplot(x='avg_training_score',y='age',kind='hex', data=df1 )
```

Out[38]: <seaborn.axisgrid.JointGrid at 0x7f151a58e100>



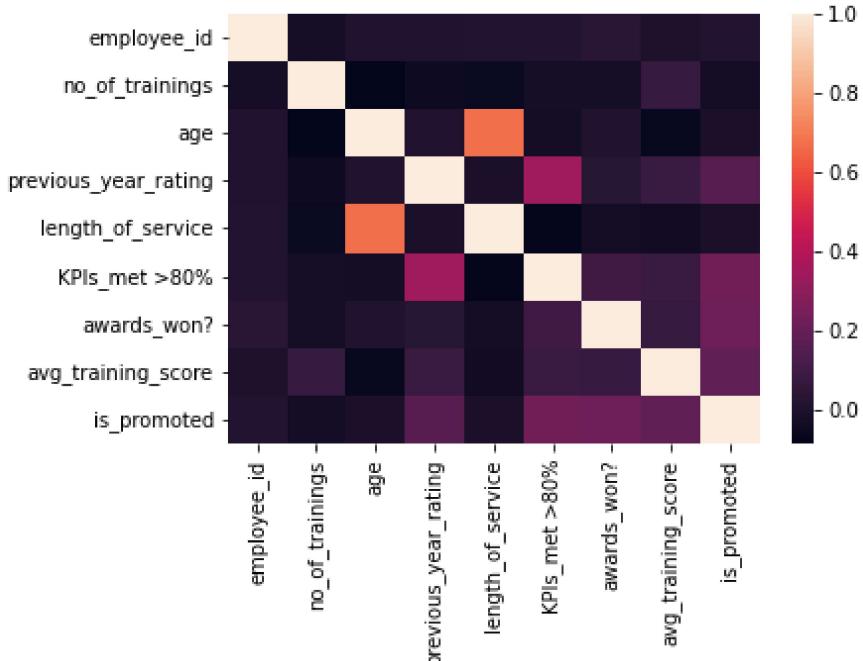
```
In [39]: 1 sns.kdeplot(x='age',y='avg_training_score', data=df1 )
```

Out[39]: <AxesSubplot:xlabel='age', ylabel='avg\_training\_score'>



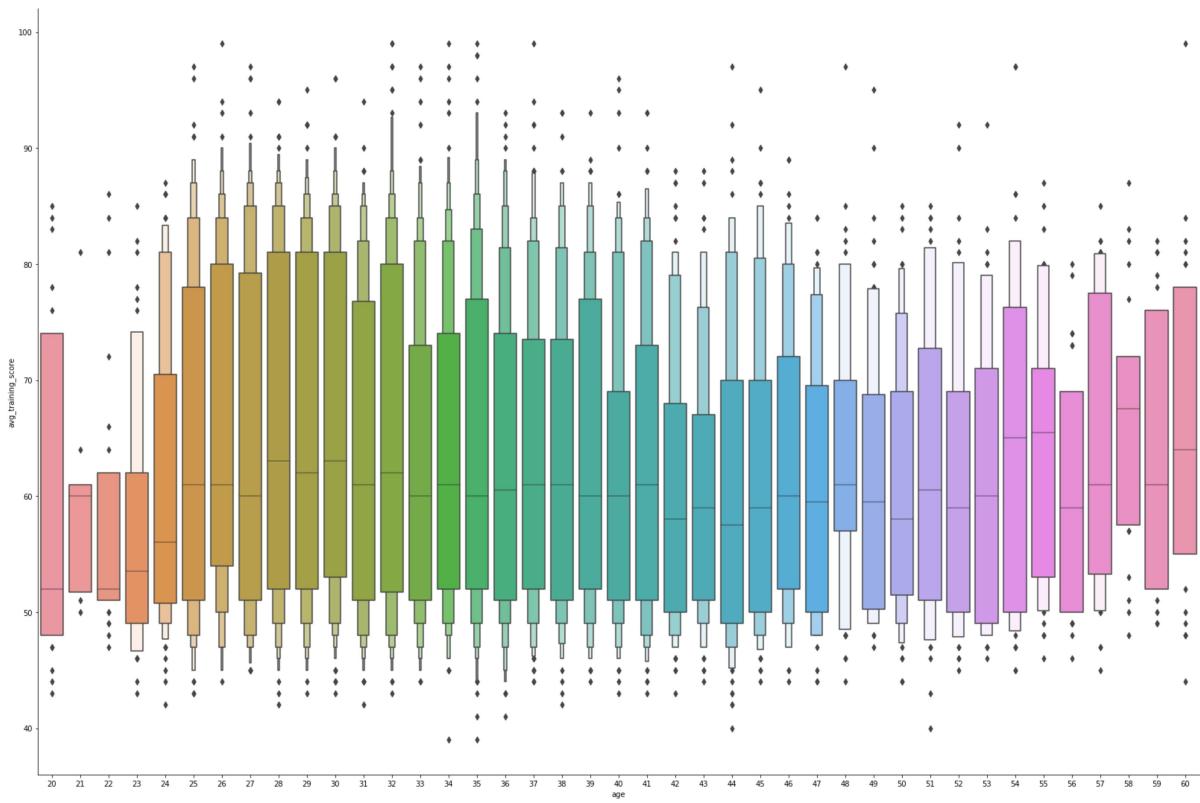
In [40]: 1 sns.heatmap(df1.corr())

Out[40]: <AxesSubplot:>



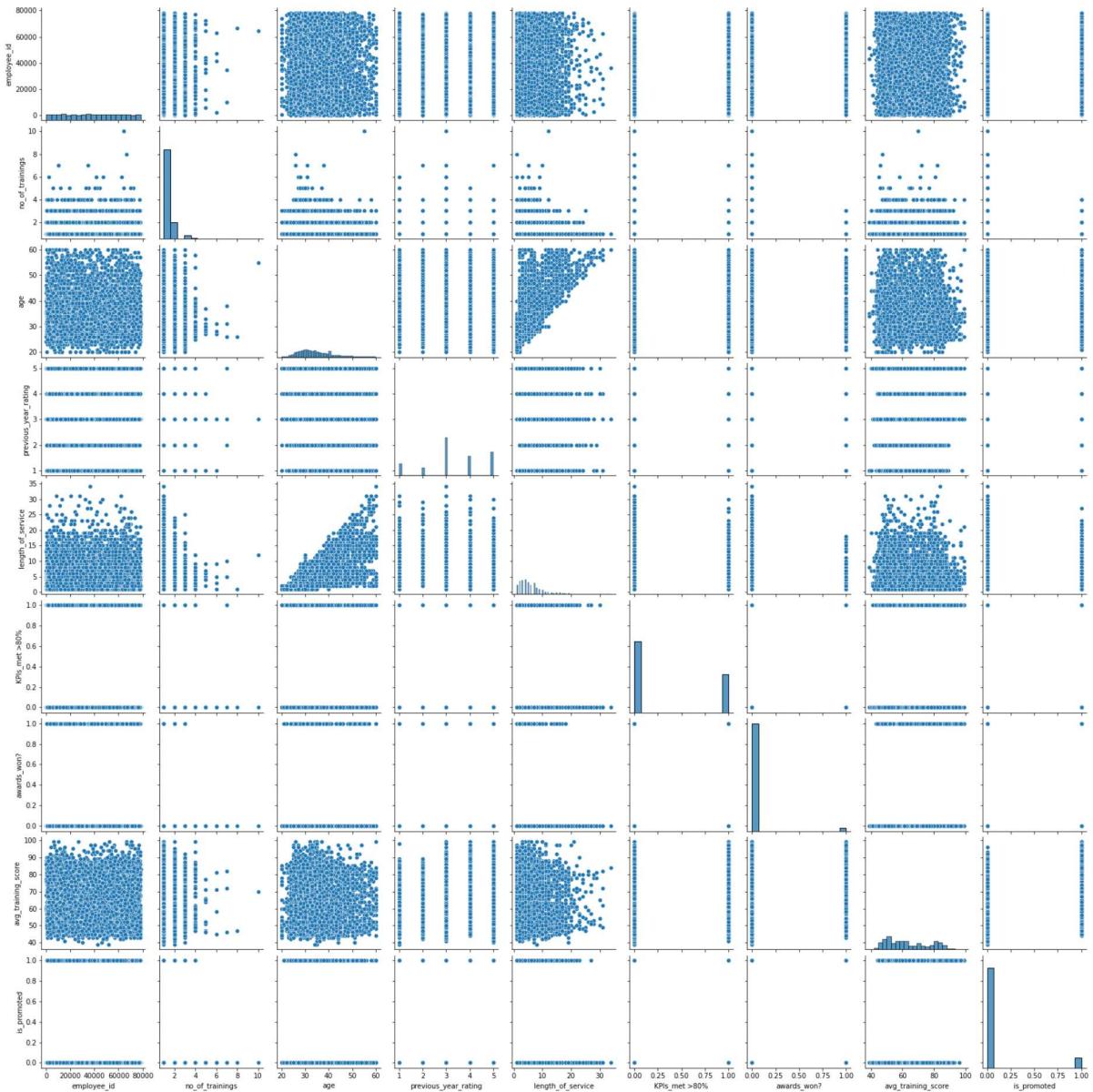
In [41]: 1 sns.catplot(x='age',y='avg\_training\_score',data=df1,kind='boxen',height=15,as\_box=False)

Out[41]: <seaborn.axisgrid.FacetGrid at 0x7f151a62be20>



In [42]: 1 sns.pairplot(df1)

Out[42]: <seaborn.axisgrid.PairGrid at 0x7f1519baa100>



In [ ]: 1