

Lab7. Data Visualization in Seaborn

SURIYA_225229140

```
In [2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [4]: df=pd.read_csv("train_upvote_mini.csv")
df.head()
```

Out[4]:

	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 [5]: df.size
```

Out[5]: 108080

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

Out[6]: (15440, 7)

```
In [7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15440 entries, 0 to 15439
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          --          --      
 0   ID          15440 non-null   int64  
 1   Tag         15440 non-null   object 
 2   Reputation  15440 non-null   float64 
 3   Answers     15440 non-null   float64 
 4   Username    15440 non-null   int64  
 5   Views       15440 non-null   float64 
 6   Upvotes     15440 non-null   float64 
dtypes: float64(4), int64(2), object(1)
memory usage: 844.5+ KB
```

```
In [8]: df.dtypes
```

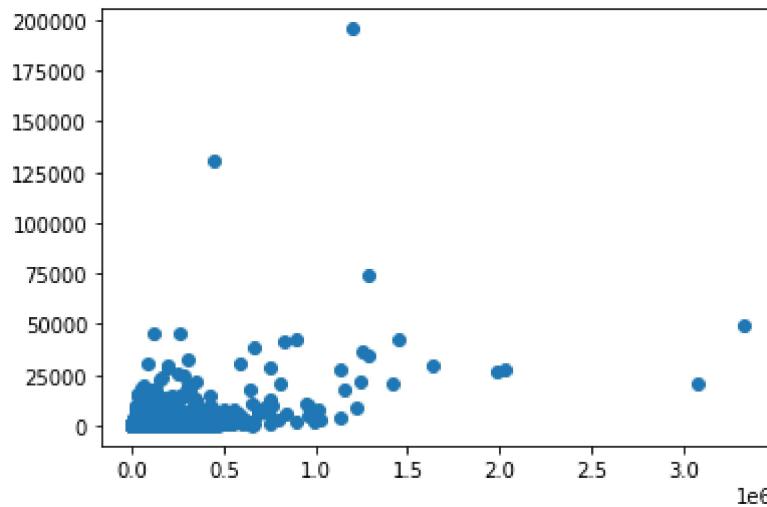
```
Out[8]: ID           int64
         Tag          object
         Reputation   float64
         Answers      float64
         Username     int64
         Views        float64
         Upvotes       float64
         dtype: object
```

```
In [9]: df['Tag'].unique()
```

```
Out[9]: array(['a', 'c', 'r', 'j', 'p', 's', 'h', 'o', 'i', 'x'], dtype=object)
```

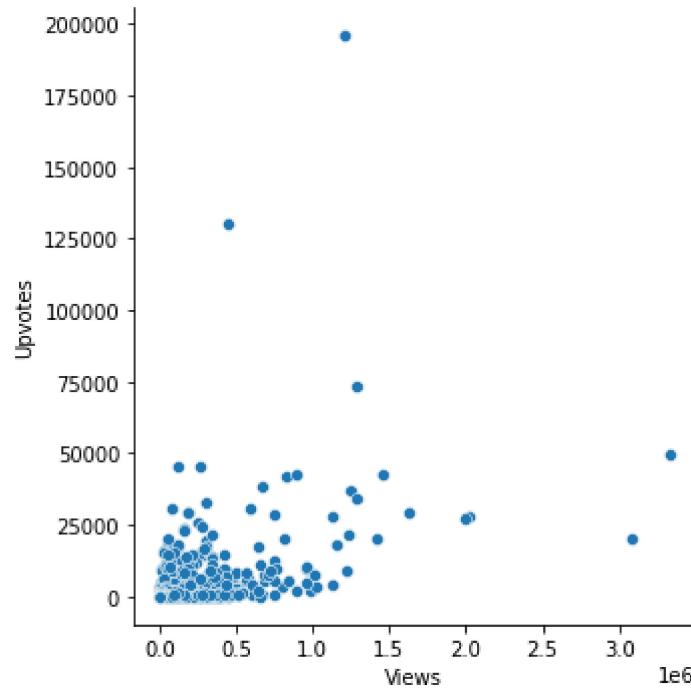
```
In [10]: plt.scatter(x=df['Views'],y=df['Upvotes'])
```

```
Out[10]: <matplotlib.collections.PathCollection at 0x7fe34011c280>
```



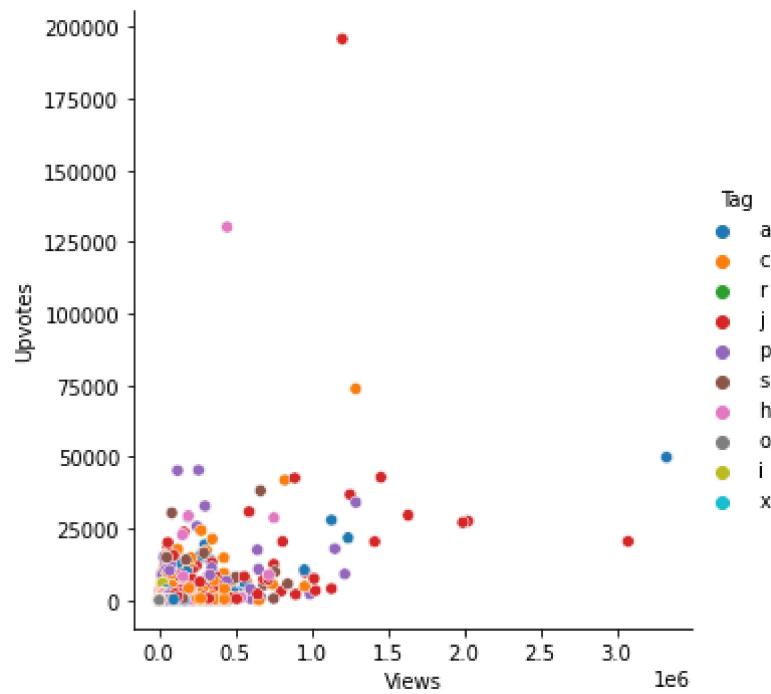
```
In [11]: sns.relplot(data=df,x='Views',y='Upvotes')
```

```
Out[11]: <seaborn.axisgrid.FacetGrid at 0x7fe3400fdbb0>
```



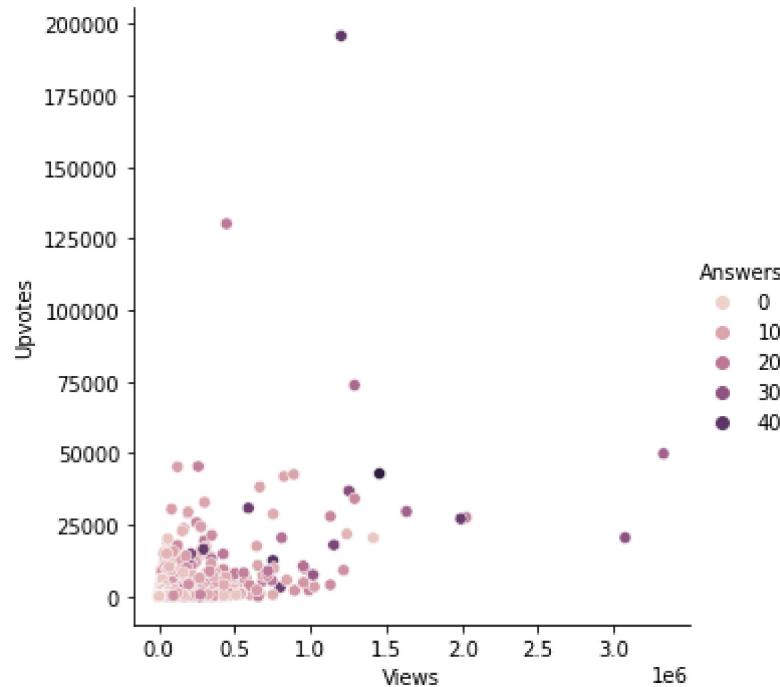
```
In [13]: sns.relplot(data=df,x='Views',y='Upvotes',hue='Tag')
```

```
Out[13]: <seaborn.axisgrid.FacetGrid at 0x7fe3400d7ee0>
```



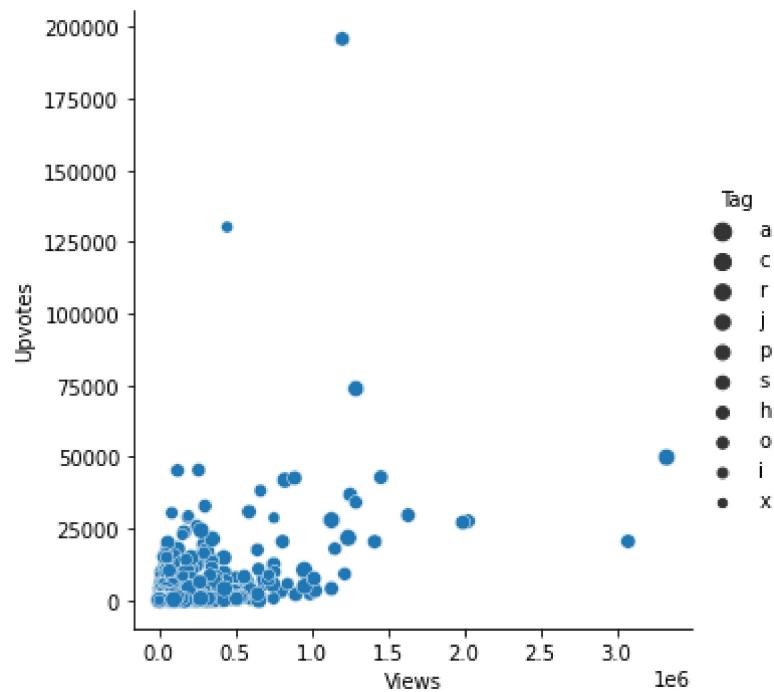
```
In [14]: sns.relplot(data=df,x='Views',y='Upvotes',hue='Answers')
```

```
Out[14]: <seaborn.axisgrid.FacetGrid at 0x7fe33d749b20>
```



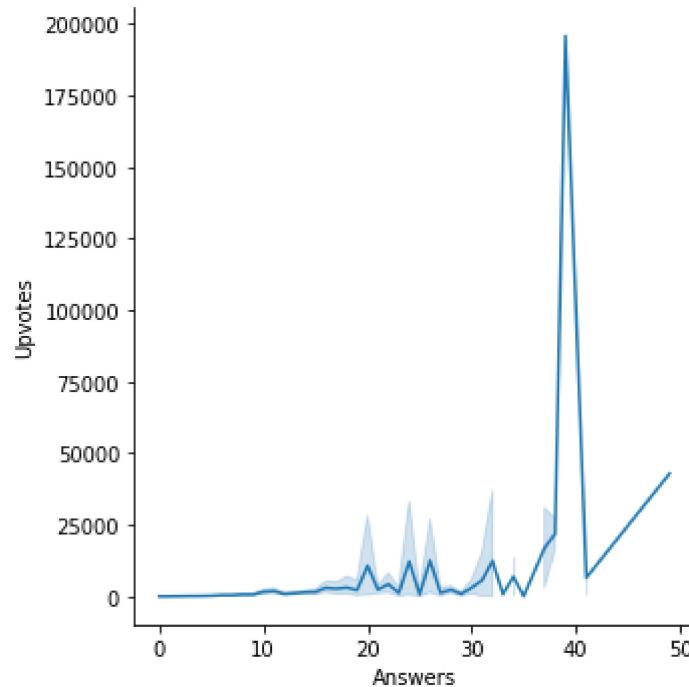
```
In [15]: sns.relplot(data=df,x='Views',y='Upvotes',size='Tag')
```

```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x7fe33d883880>
```



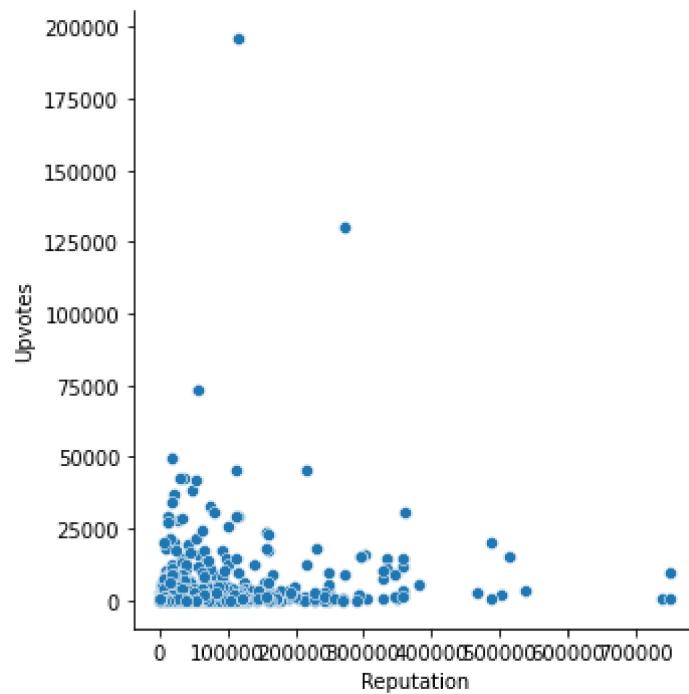
```
In [16]: sns.relplot(data=df, x="Answers", y="Upvotes", kind='line')
```

```
Out[16]: <seaborn.axisgrid.FacetGrid at 0x7fe37246c730>
```



```
In [17]: sns.relplot(data=df, x="Reputation", y="Upvotes")
```

```
Out[17]: <seaborn.axisgrid.FacetGrid at 0x7fe33d781430>
```



```
In [18]: df1 = pd.read_csv('train_hr_mini.csv')
df1.head()
```

Out[18]:

	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings
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

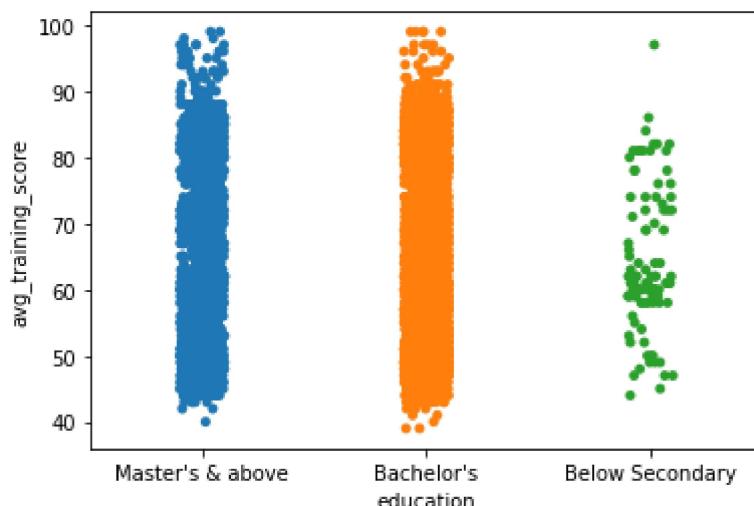


```
In [19]: df1.shape
```

Out[19]: (6397, 14)

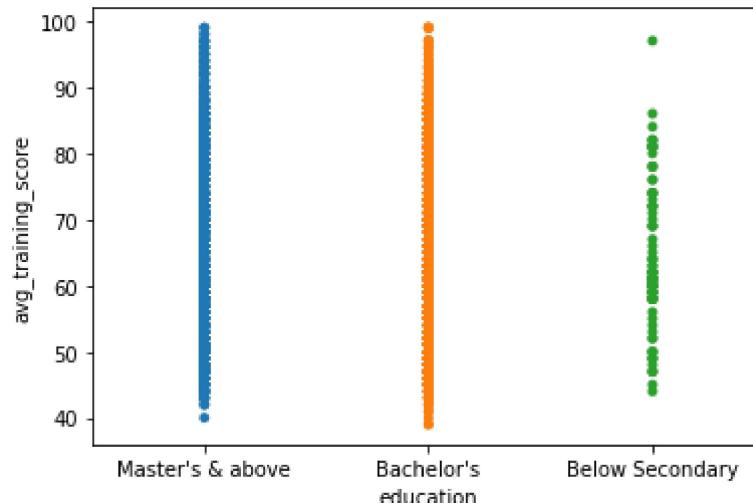
```
In [20]: sns.stripplot(data=df1, x="education", y="avg_training_score", jitter=True)
```

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



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

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

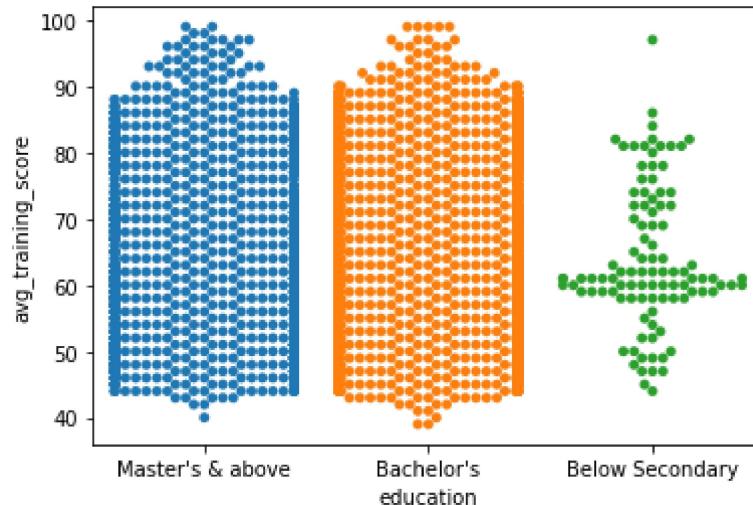


```
In [22]: sns.swarmplot(data=df1, x="education", y="avg_training_score")
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 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: 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)
```

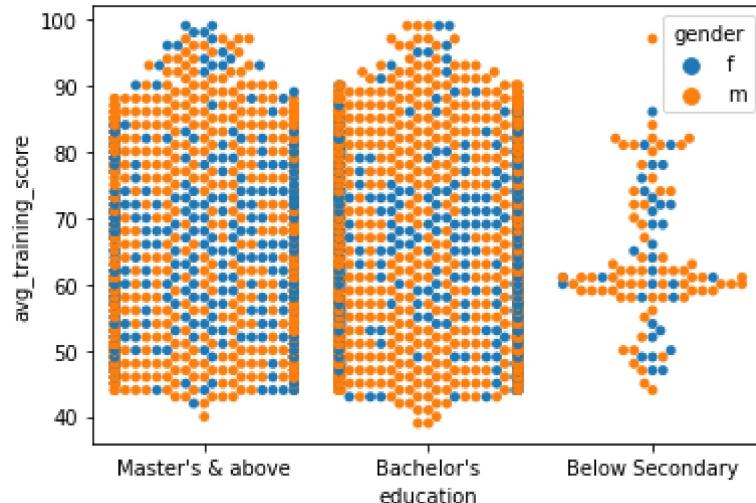
```
Out[22]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



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

```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 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: 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)
```

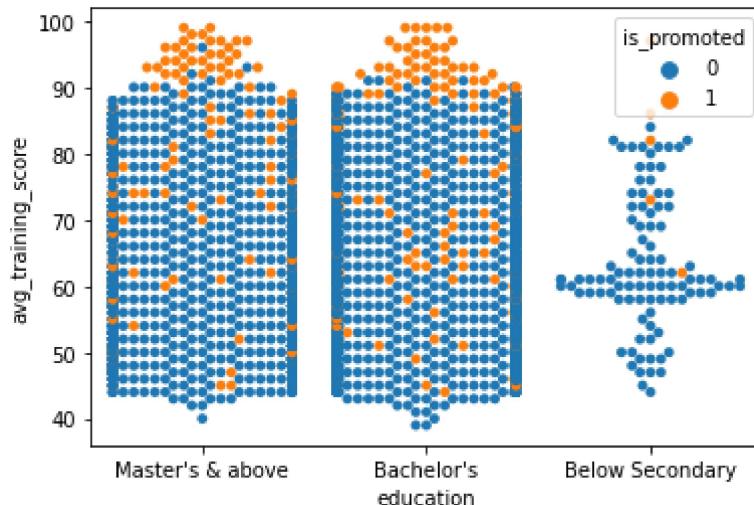
```
Out[23]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



In [24]: `sns.swarmplot(data=df1, x="education", y="avg_training_score", hue='is_promoted')`

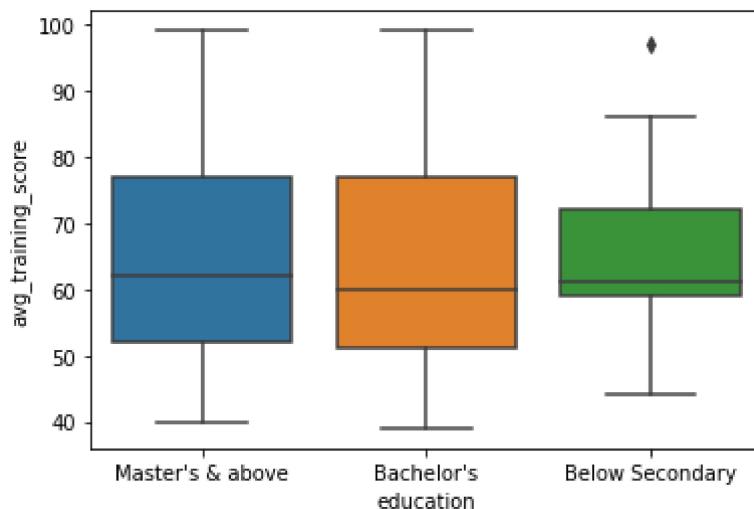
```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 74.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 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[24]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>



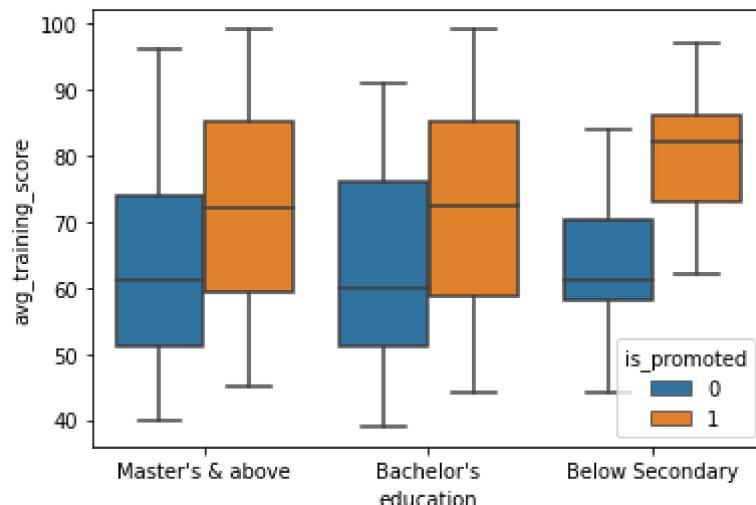
In [25]: `sns.boxplot(data=df1, x="education", y="avg_training_score")`

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



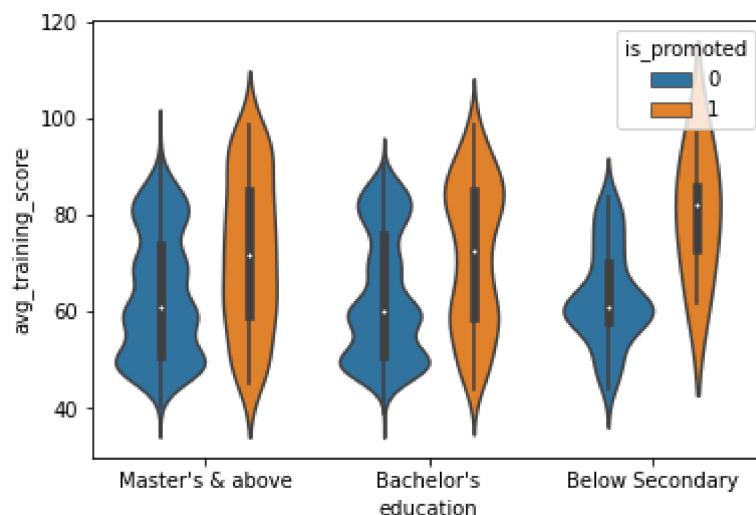
```
In [26]: sns.boxplot(data=df1, x="education", y="avg_training_score", hue ='is_promoted')
```

```
Out[26]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



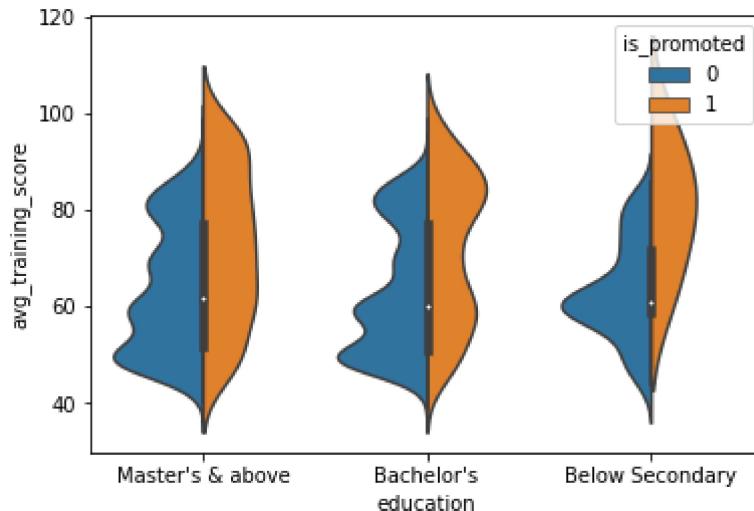
```
In [27]: sns.violinplot(data=df1, x="education", y="avg_training_score", hue ='is_promoted')
```

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



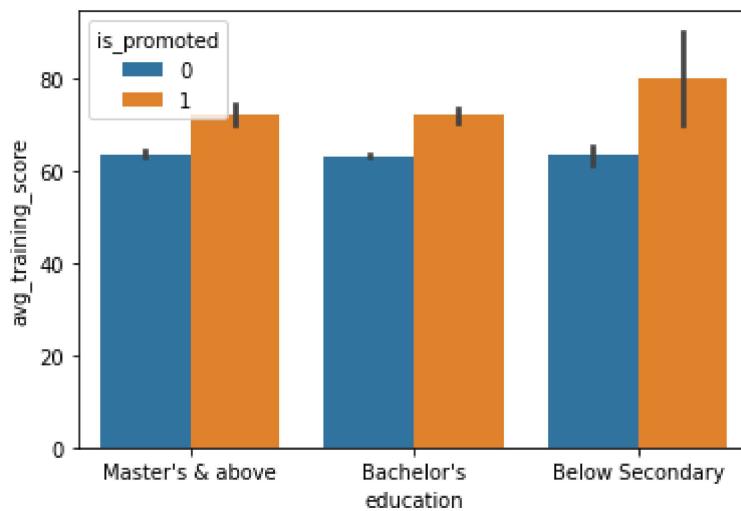
```
In [44]: sns.violinplot(data=df1, x="education", y="avg_training_score", hue = 'is_promoted')
```

```
Out[44]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



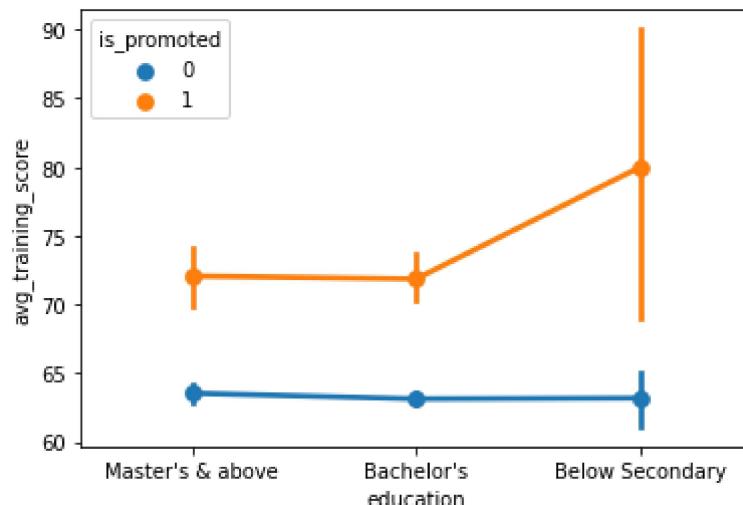
```
In [30]: sns.barplot(data=df1, x="education", y="avg_training_score", hue = 'is_promoted')
```

```
Out[30]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>
```



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

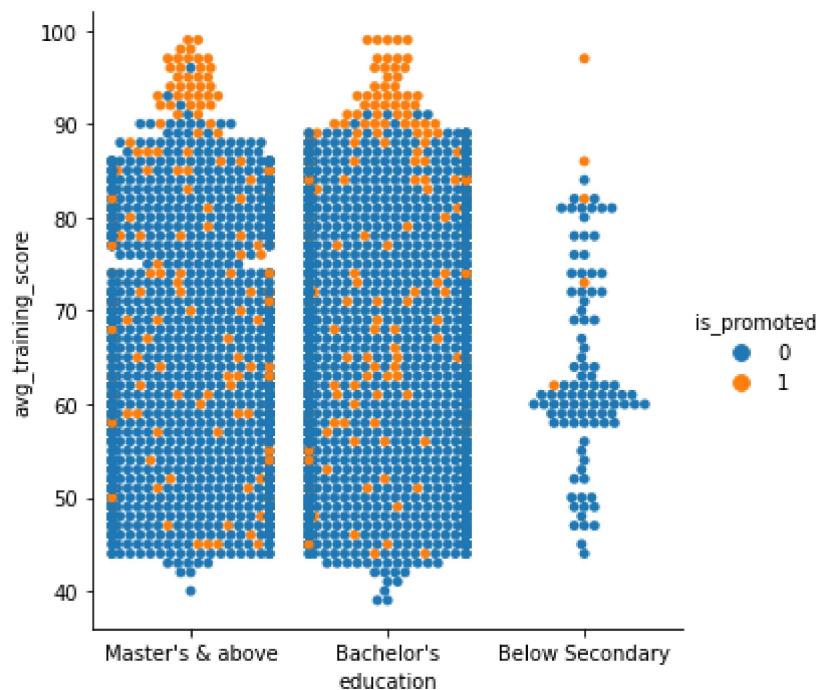
Out[31]: <AxesSubplot:xlabel='education', ylabel='avg_training_score'>



In [45]: `sns.catplot(data=df1, x="education", y="avg_training_score", hue = 'is_promoted')`

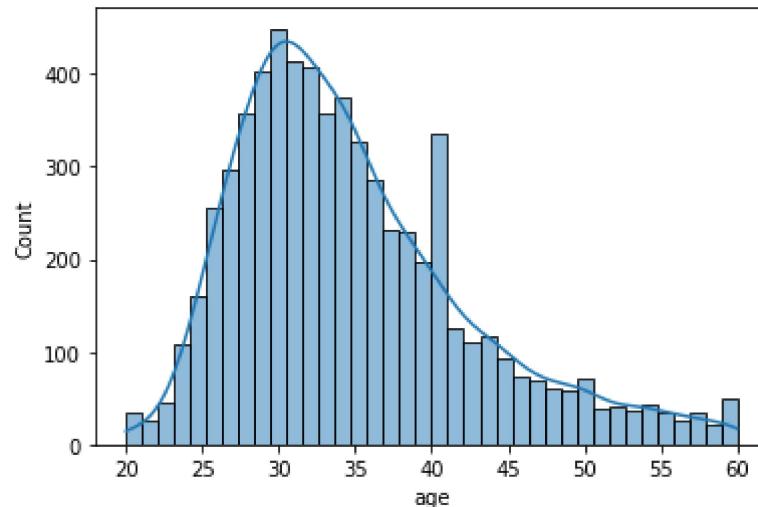
```
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 56.8% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
/usr/local/lib/python3.9/dist-packages/seaborn/categorical.py:1296: UserWarning: 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[45]: <seaborn.axisgrid.FacetGrid at 0x7fe335f69dc0>



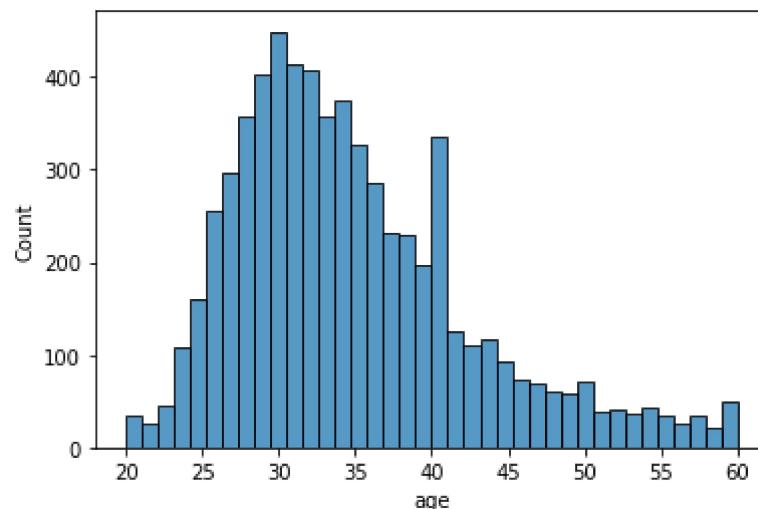
```
In [34]: sns.histplot(x='age', data=df1, kde=True)
```

```
Out[34]: <AxesSubplot:xlabel='age', ylabel='Count'>
```



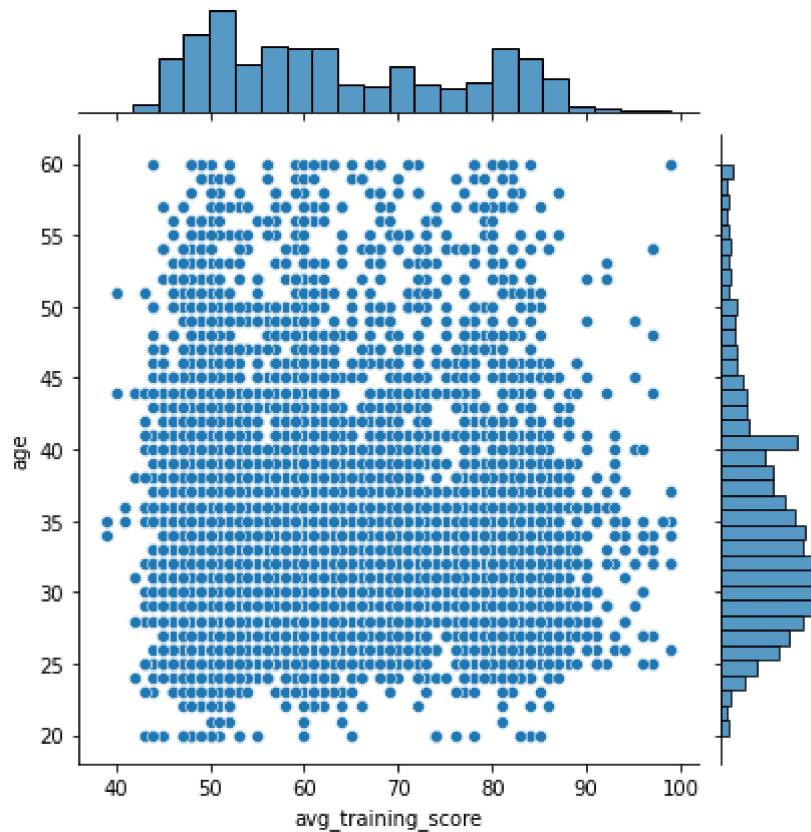
```
In [35]: sns.histplot(x='age', data=df1)
```

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



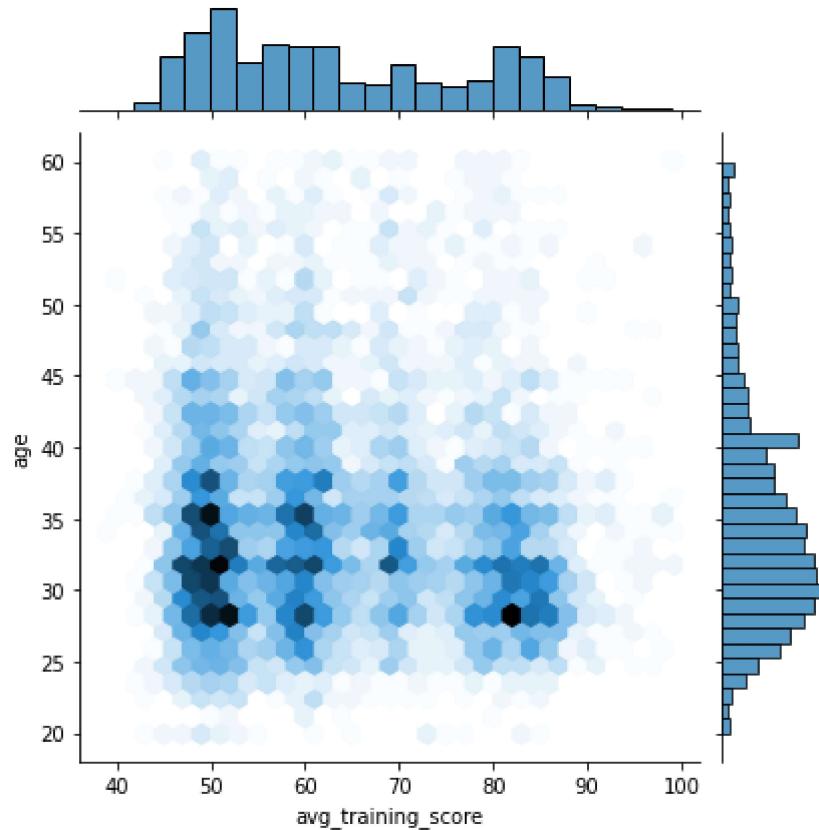
```
In [36]: sns.jointplot(x='avg_training_score',y='age', data=df1 )
```

```
Out[36]: <seaborn.axisgrid.JointGrid at 0x7fe3395207c0>
```



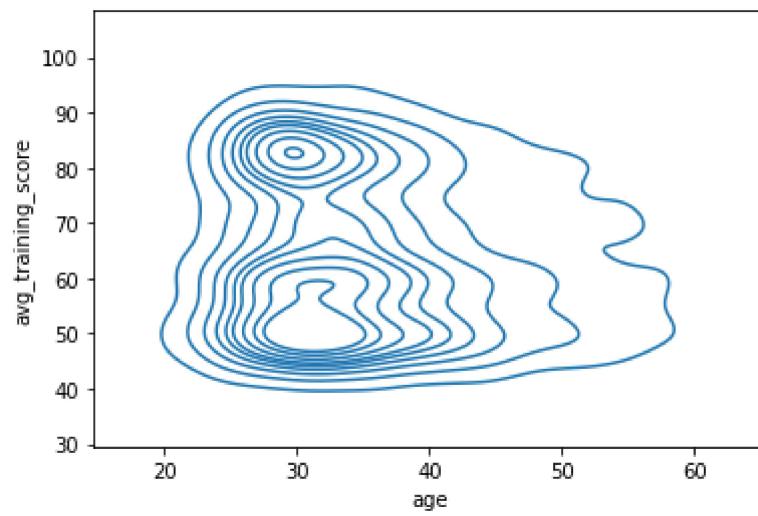
```
In [37]: sns.jointplot(x='avg_training_score',y='age',kind='hex', data=df1 )
```

```
Out[37]: <seaborn.axisgrid.JointGrid at 0x7fe33934a490>
```



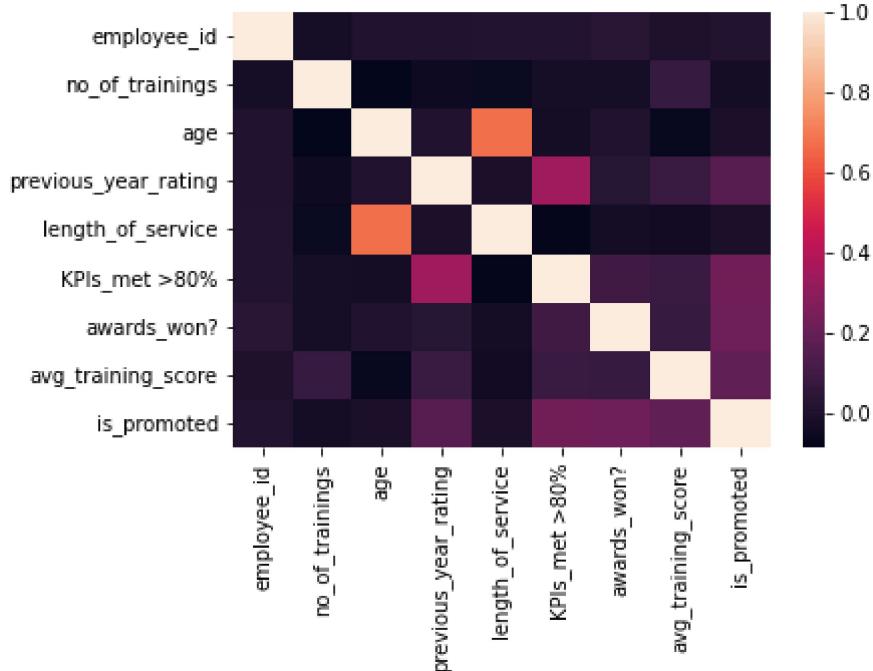
```
In [38]: sns.kdeplot(x='age',y='avg_training_score', data=df1 )
```

```
Out[38]: <AxesSubplot:xlabel='age', ylabel='avg_training_score'>
```



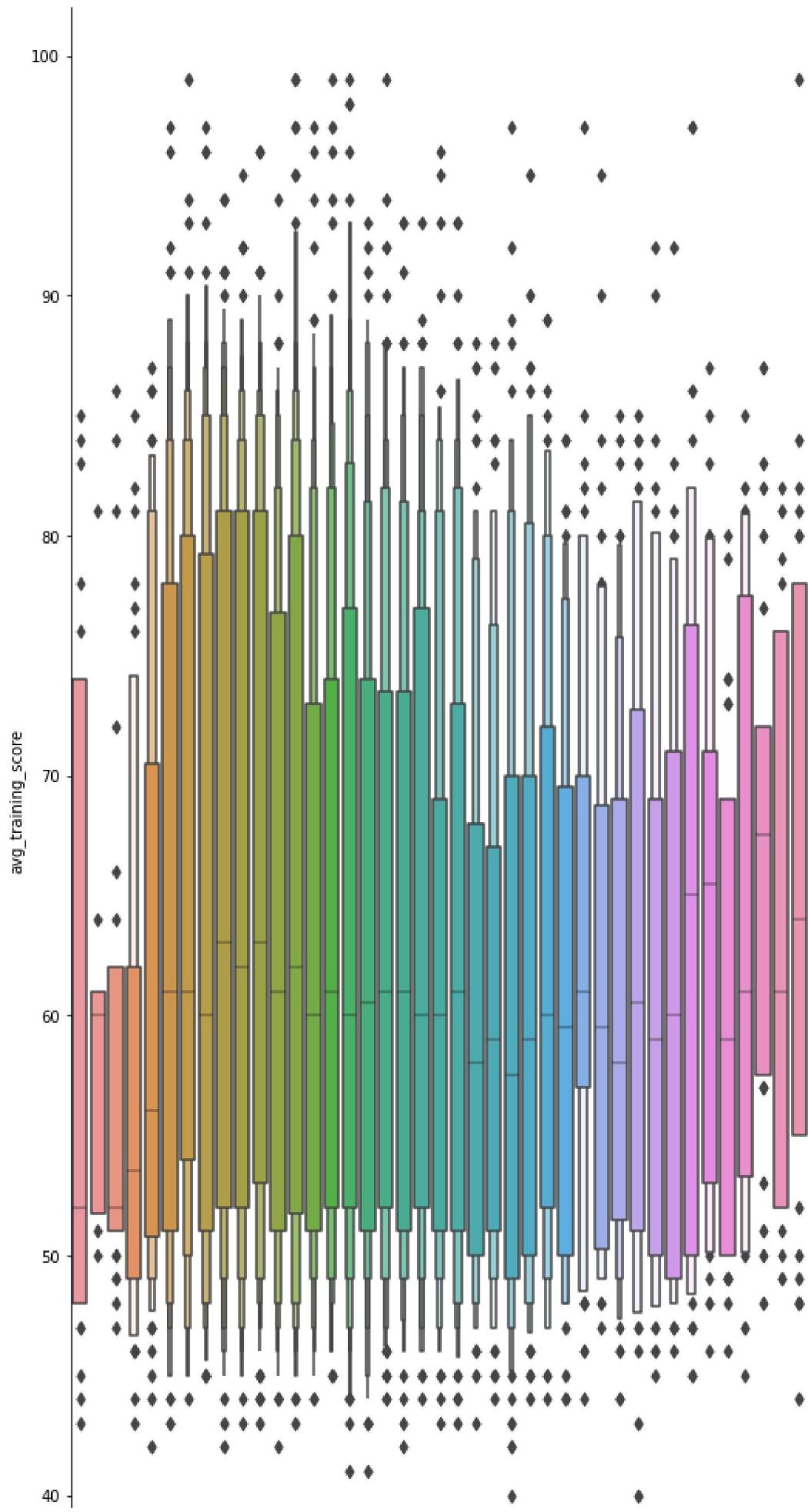
```
In [39]: sns.heatmap(df1.corr())
```

Out[39]: <AxesSubplot:>



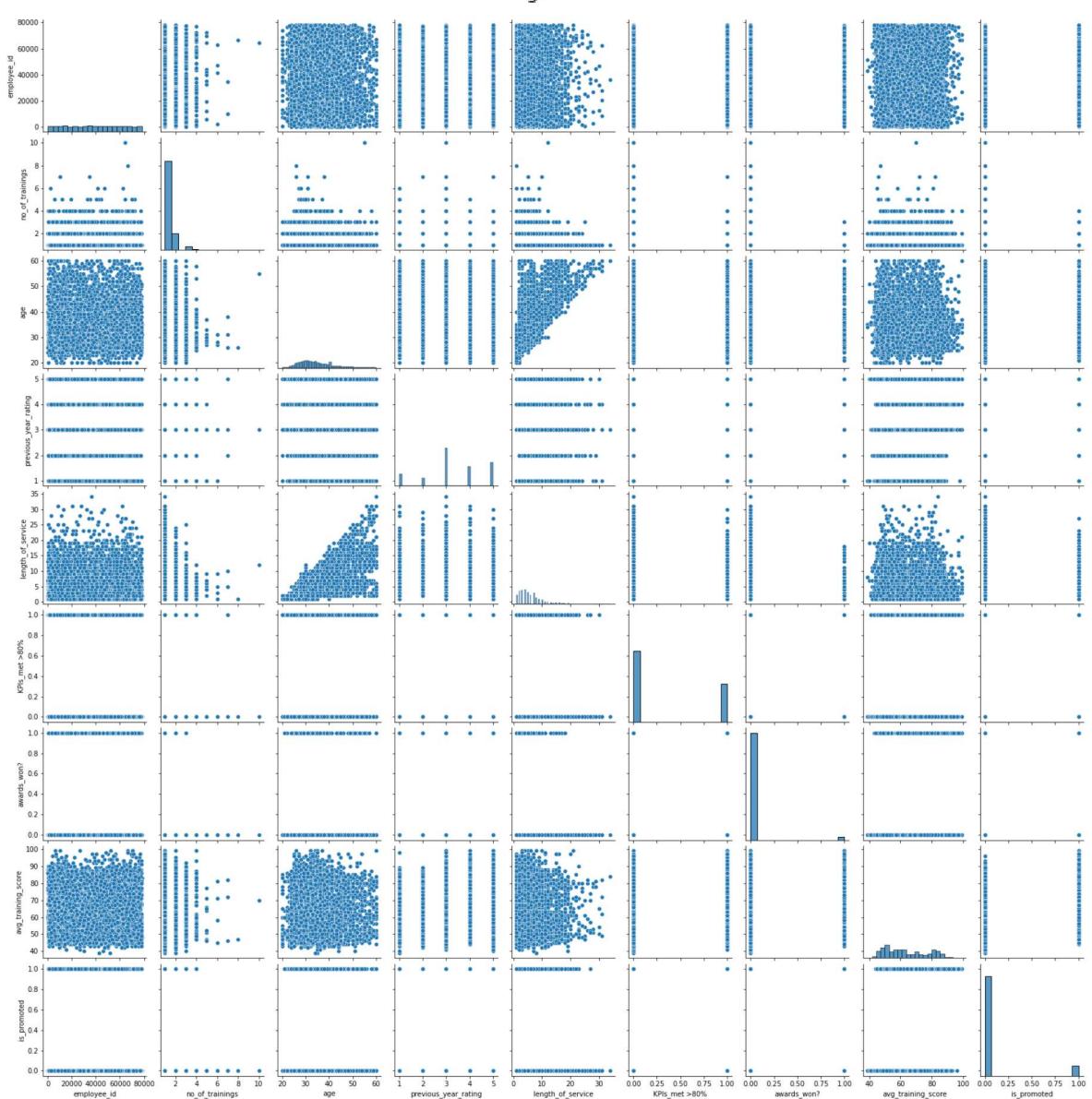
```
In [40]: sns.catplot(x='age',y='avg_training_score', data=df1, kind='boxen', height=15)
```

```
Out[40]: <seaborn.axisgrid.FacetGrid at 0x7fe338e8da00>
```

```
In [41]: sns.pairplot(df1)
```

```
Out[41]: <seaborn.axisgrid.PairGrid at 0x7fe338e043d>
```



```
In [ ]:
```