```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

df = pd.read_csv('/Users/aaryanbabuta/Documents/Prodigy DS Internship June 2024/bank+marketing/bank-additional/bank-additional.csv',delimiter=';')
df.head()
```

```
   age          job  marital           education default  housing loan  \
0   30  blue-collar  married           basic.9y      no      yes   no
1   39     services   single        high.school      no       no   no
2   25     services  married        high.school      no      yes   no
3   38     services  married           basic.9y      no  unknown unknown
4   47       admin.  married  university.degree      no      yes   no

     contact month day_of_week  ...  campaign  pdays  previous poutcome  \
0   cellular   may         fri  ...         2    999         0  nonexistent
1  telephone   may         fri  ...         4    999         0  nonexistent
2  telephone   jun         wed  ...         1    999         0  nonexistent
3  telephone   jun         fri  ...         3    999         0  nonexistent
4   cellular   nov         mon  ...         1    999         0  nonexistent

  emp.var.rate  cons.price.idx  cons.conf.idx  euribor3m  nr.employed  y
0         -1.8          92.893          -46.2      1.313       5099.1  no
1          1.1          93.994          -36.4      4.855       5191.0  no
2          1.4          94.465          -41.8      4.962       5228.1  no
3          1.4          94.465          -41.8      4.959       5228.1  no
4         -0.1          93.200          -42.0      4.191       5195.8  no
```

```
[5 rows x 21 columns]
```

```
df.head()
```

```
   age          job  marital          education default  housing
loan  \
0   30  blue-collar  married           basic.9y      no      yes
no
1   39     services   single        high.school      no       no
no
2   25     services  married        high.school      no      yes
no
3   38     services  married           basic.9y      no  unknown
unknown
4   47       admin.  married  university.degree      no      yes
no

     contact month day_of_week  ...  campaign  pdays  previous
poutcome  \
0   cellular   may         fri  ...         2    999         0
nonexistent
1  telephone   may         fri  ...         4    999         0
nonexistent
2  telephone   jun         wed  ...         1    999         0
nonexistent
3  telephone   jun         fri  ...         3    999         0
nonexistent
4   cellular   nov         mon  ...         1    999         0
nonexistent

  emp.var.rate  cons.price.idx  cons.conf.idx  euribor3m  nr.employed
y
0         -1.8          92.893          -46.2      1.313       5099.1
no
1          1.1          93.994          -36.4      4.855       5191.0
no
2          1.4          94.465          -41.8      4.962       5228.1
no
3          1.4          94.465          -41.8      4.959       5228.1
no
4         -0.1          93.200          -42.0      4.191       5195.8
no

[5 rows x 21 columns]
```

```
df.tail()
```

```
      age         job  marital    education default housing loan
contact  \
4114   30      admin.  married     basic.6y      no     yes  yes
```

```
cellular
4115   39      admin.  married  high.school      no     yes   no
telephone
4116   27     student   single  high.school      no      no   no
cellular
4117   58      admin.  married  high.school      no      no   no
cellular
4118   34  management   single  high.school      no     yes   no
cellular

     month day_of_week  ...  campaign  pdays  previous     poutcome  \
4114   jul         thu  ...         1    999         0  nonexistent
4115   jul         fri  ...         1    999         0  nonexistent
4116   may         mon  ...         2    999         1      failure
4117   aug         fri  ...         1    999         0  nonexistent
4118   nov         wed  ...         1    999         0  nonexistent

     emp.var.rate  cons.price.idx  cons.conf.idx  euribor3m
nr.employed   y
4114          1.4          93.918          -42.7      4.958
5228.1  no
4115          1.4          93.918          -42.7      4.959
5228.1  no
4116         -1.8          92.893          -46.2      1.354
5099.1  no
4117          1.4          93.444          -36.1      4.966
5228.1  no
4118         -0.1          93.200          -42.0      4.120
5195.8  no

[5 rows x 21 columns]

df.shape

(4119, 21)

df.columns

Index(['age', 'job', 'marital', 'education', 'default', 'housing',
'loan',
       'contact', 'month', 'day_of_week', 'duration', 'campaign',
'pdays',
       'previous', 'poutcome', 'emp.var.rate', 'cons.price.idx',
       'cons.conf.idx', 'euribor3m', 'nr.employed', 'y'],
      dtype='object')

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4119 entries, 0 to 4118
Data columns (total 21 columns):
```

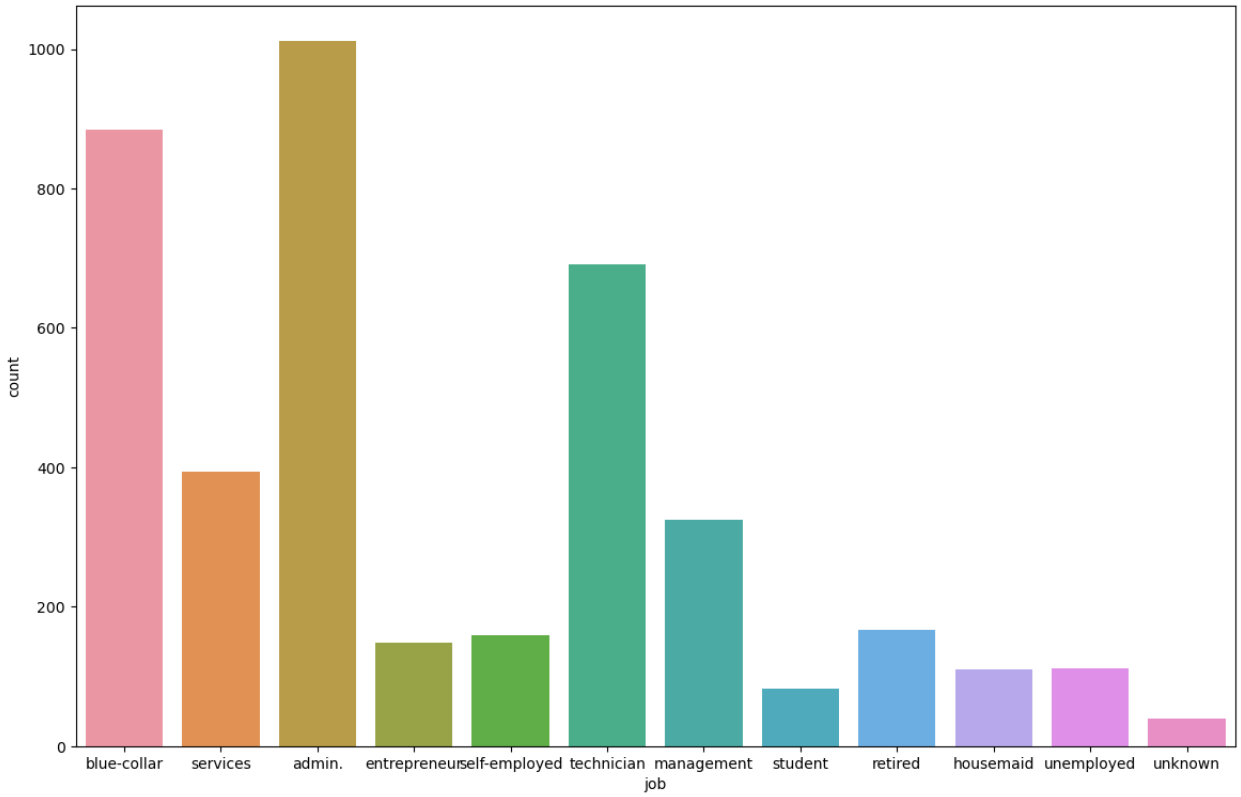
```
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   age              4119 non-null   int64  
 1   job              4119 non-null   object 
 2   marital          4119 non-null   object 
 3   education        4119 non-null   object 
 4   default          4119 non-null   object 
 5   housing          4119 non-null   object 
 6   loan             4119 non-null   object 
 7   contact          4119 non-null   object 
 8   month            4119 non-null   object 
 9   day_of_week      4119 non-null   object 
 10  duration         4119 non-null   int64  
 11  campaign         4119 non-null   int64  
 12  pdays            4119 non-null   int64  
 13  previous         4119 non-null   int64  
 14  poutcome         4119 non-null   object 
 15  emp.var.rate     4119 non-null   float64
 16  cons.price.idx   4119 non-null   float64
 17  cons.conf.idx    4119 non-null   float64
 18  euribor3m        4119 non-null   float64
 19  nr.employed      4119 non-null   float64
 20  y                4119 non-null   object 
dtypes: float64(5), int64(5), object(11)
memory usage: 675.9+ KB
```

```
df.isnull().sum().any()
```

```
False
```

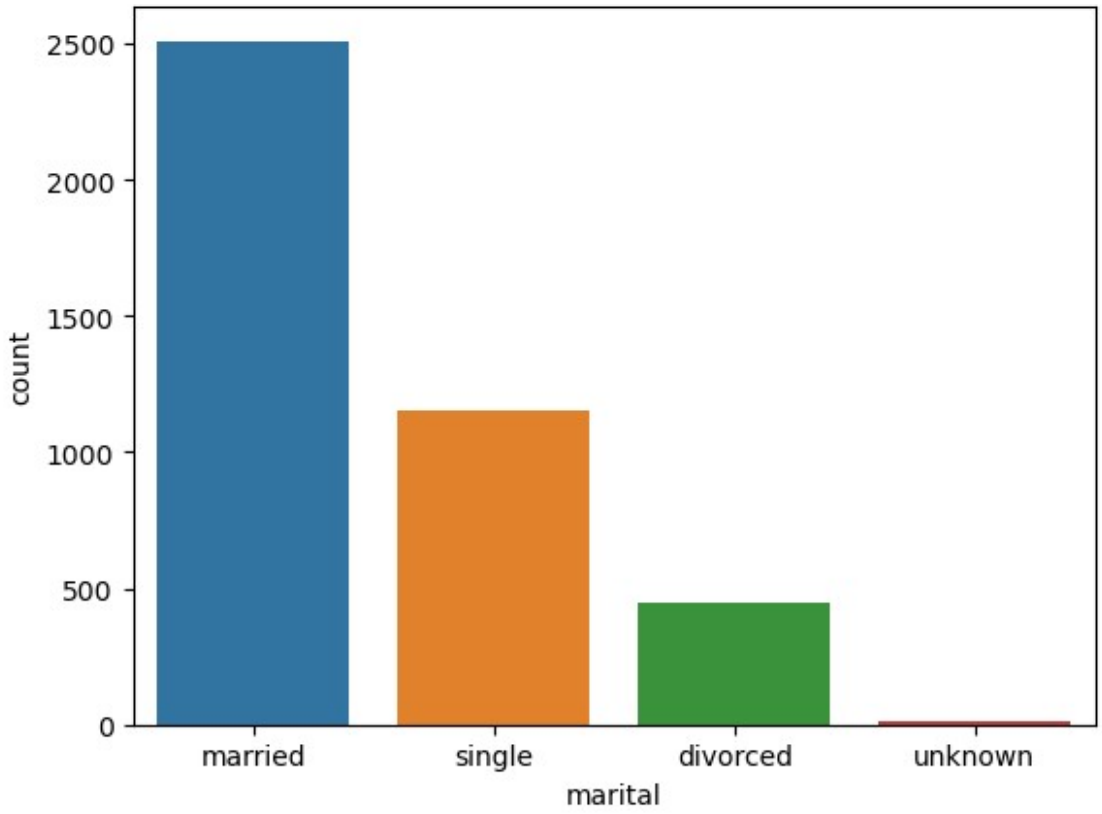
```
plt.figure(figsize = (14,9))
sns.countplot(x = "job",data = df)
```

```
<Axes: xlabel='job', ylabel='count'>
```



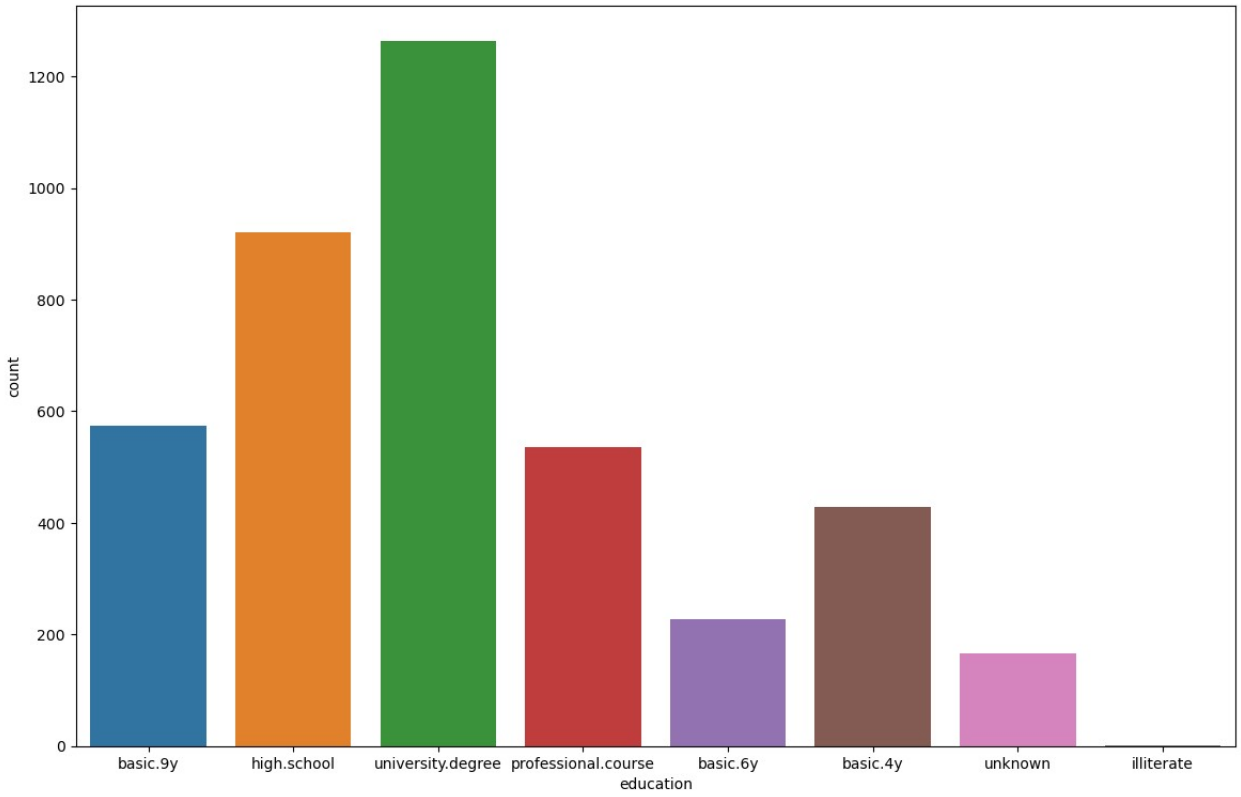

```
sns.countplot(x = "marital",data = df)
```

```
<Axes: xlabel='marital', ylabel='count'>
```



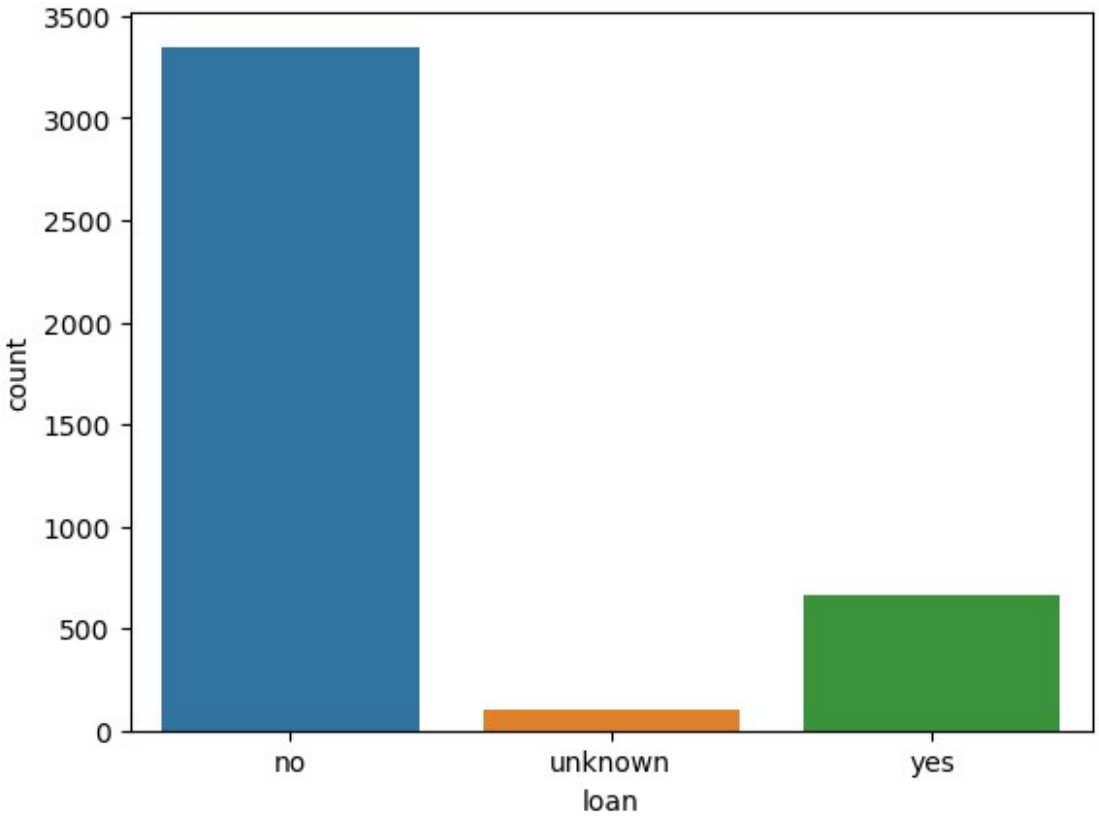

```
plt.figure(figsize = (14,9))
sns.countplot(x = "education",data = df)
```

```
<Axes: xlabel='education', ylabel='count'>
```



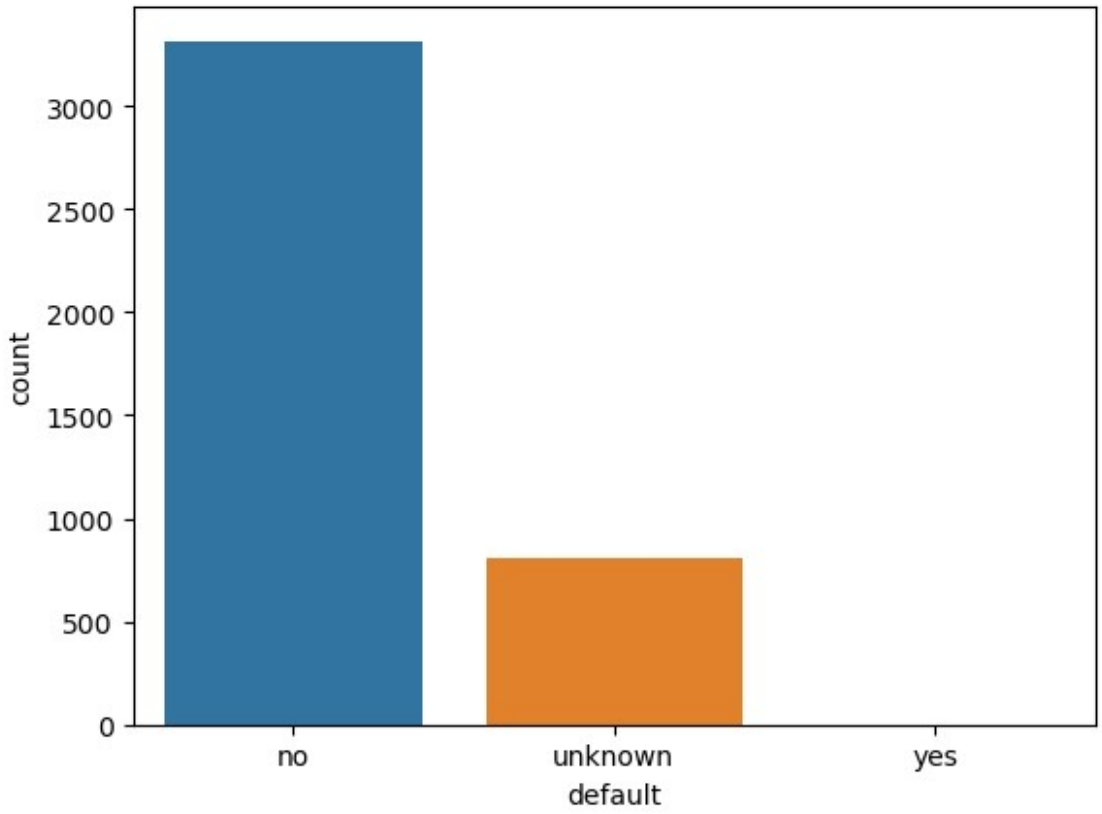

```
sns.countplot(x = "loan",data = df)
```

```
<Axes: xlabel='loan', ylabel='count'>
```




```
sns.countplot(x = "default",data = df)
```

```
<Axes: xlabel='default', ylabel='count'>
```




```
plt.figure(figsize = (16,9))
sns.pairplot(data = df,hue = "default")

my_df=df.select_dtypes(exclude=[object])
my_df.corr()

plt.figure(figsize = (16,9))
sns.heatmap(my_df.corr(),annot = True)

from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

df.drop(["pdays","previous","poutcome"],axis = 1)
df.head()
```