```
#Import Necessary Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:\\Users\\jayaraman\\OneDrive\\Pictures\\Desktop\\
Data science intership\\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.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.describe()
```

```
               age     duration     campaign        pdays     previous
\
count  4119.000000  4119.000000  4119.000000  4119.000000  4119.000000

mean     40.113620   256.788055     2.537266   960.422190     0.190337

std      10.313362   254.703736     2.568159   191.922786     0.541788

min      18.000000     0.000000     1.000000     0.000000     0.000000

25%      32.000000   103.000000     1.000000   999.000000     0.000000

50%      38.000000   181.000000     2.000000   999.000000     0.000000

75%      47.000000   317.000000     3.000000   999.000000     0.000000
```

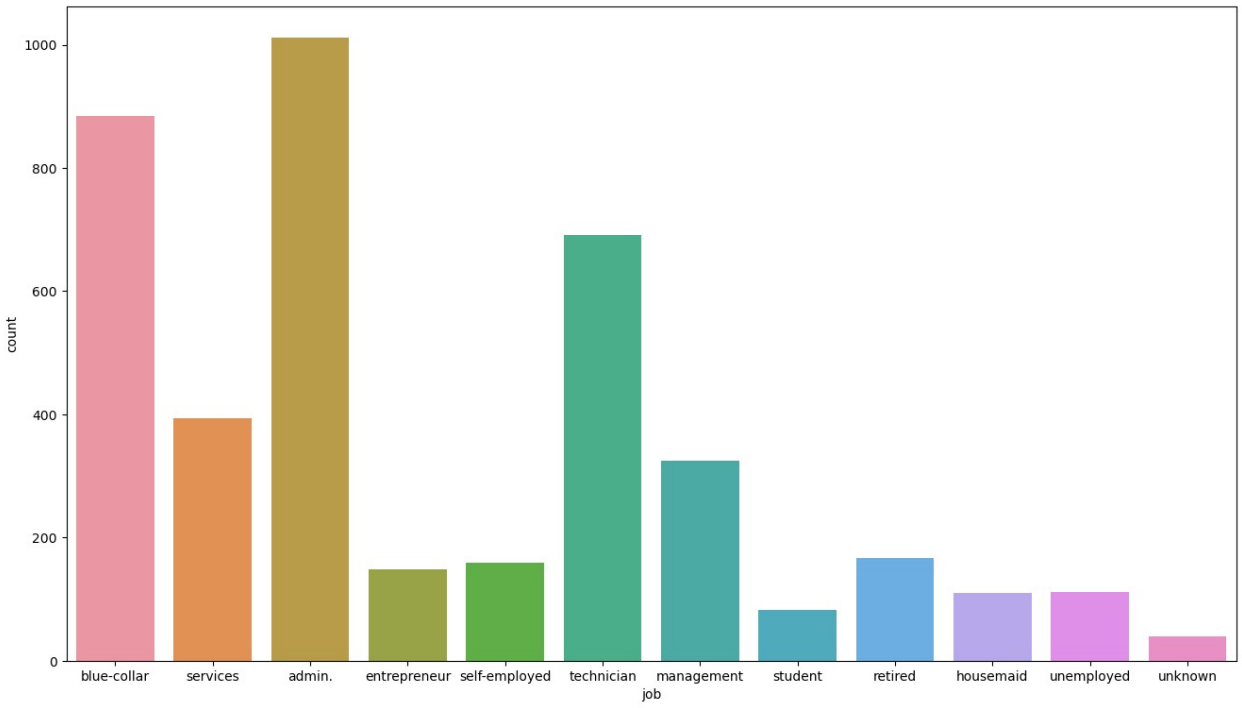
```
max       88.000000  3643.000000    35.000000   999.000000     6.000000


       emp.var.rate  cons.price.idx  cons.conf.idx    euribor3m
nr.employed
count   4119.000000     4119.000000    4119.000000  4119.000000
4119.000000
mean       0.084972       93.579704     -40.499102     3.621356
5166.481695
std        1.563114        0.579349       4.594578     1.733591
73.667904
min       -3.400000       92.201000     -50.800000     0.635000
4963.600000
25%       -1.800000       93.075000     -42.700000     1.334000
5099.100000
50%        1.100000       93.749000     -41.800000     4.857000
5191.000000
75%        1.400000       93.994000     -36.400000     4.961000
5228.100000
max        1.400000       94.767000     -26.900000     5.045000
5228.100000
```

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

```
age               0
job               0
marital           0
education         0
default           0
housing           0
loan              0
contact           0
month             0
day_of_week       0
duration          0
campaign          0
pdays             0
previous          0
poutcome          0
emp.var.rate      0
cons.price.idx    0
cons.conf.idx     0
euribor3m         0
nr.employed       0
y                 0
dtype: int64
```

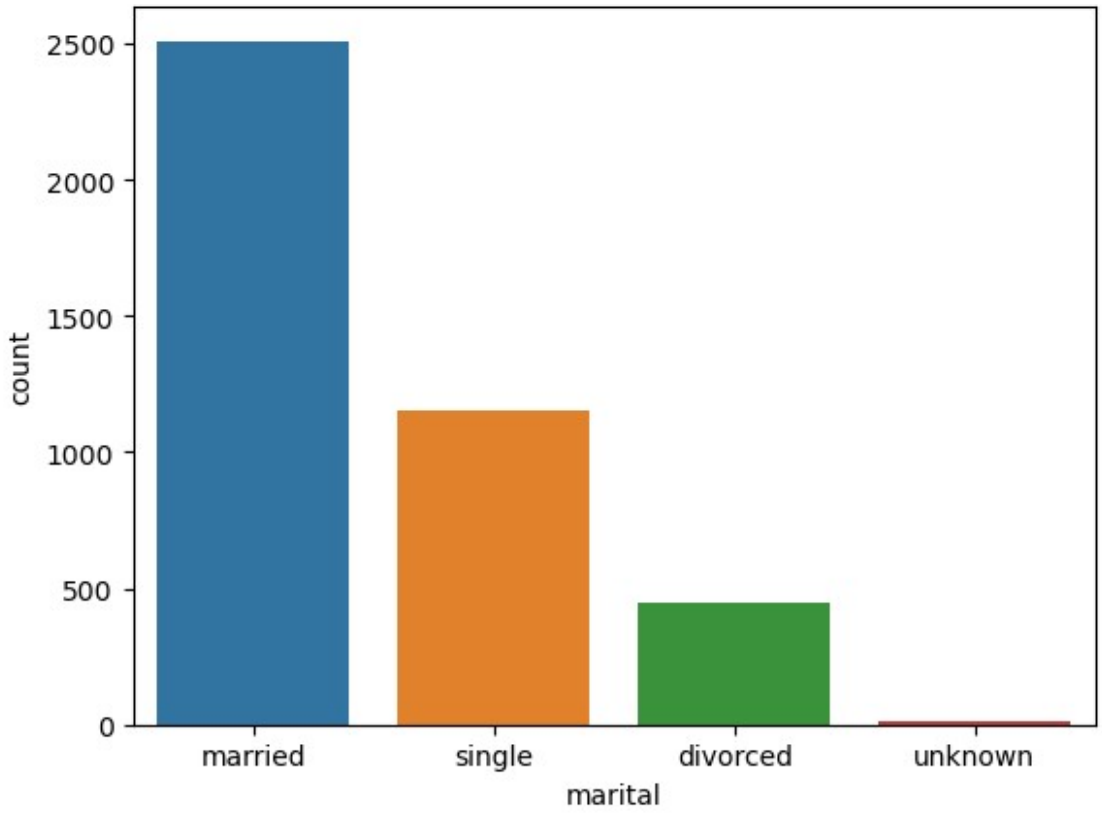
```
plt.figure(figsize = (16,9))
sns.countplot(x = "job",data = df)
```

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



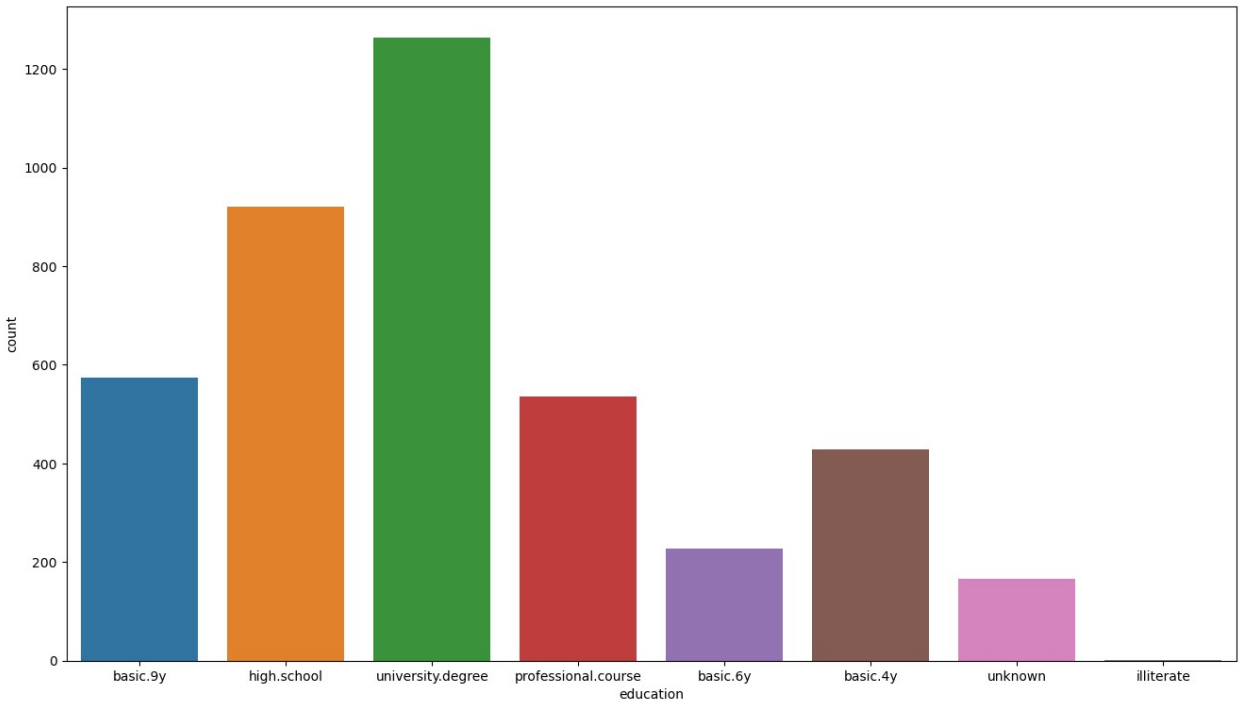

```
sns.countplot(x = "marital",data = df)
```

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



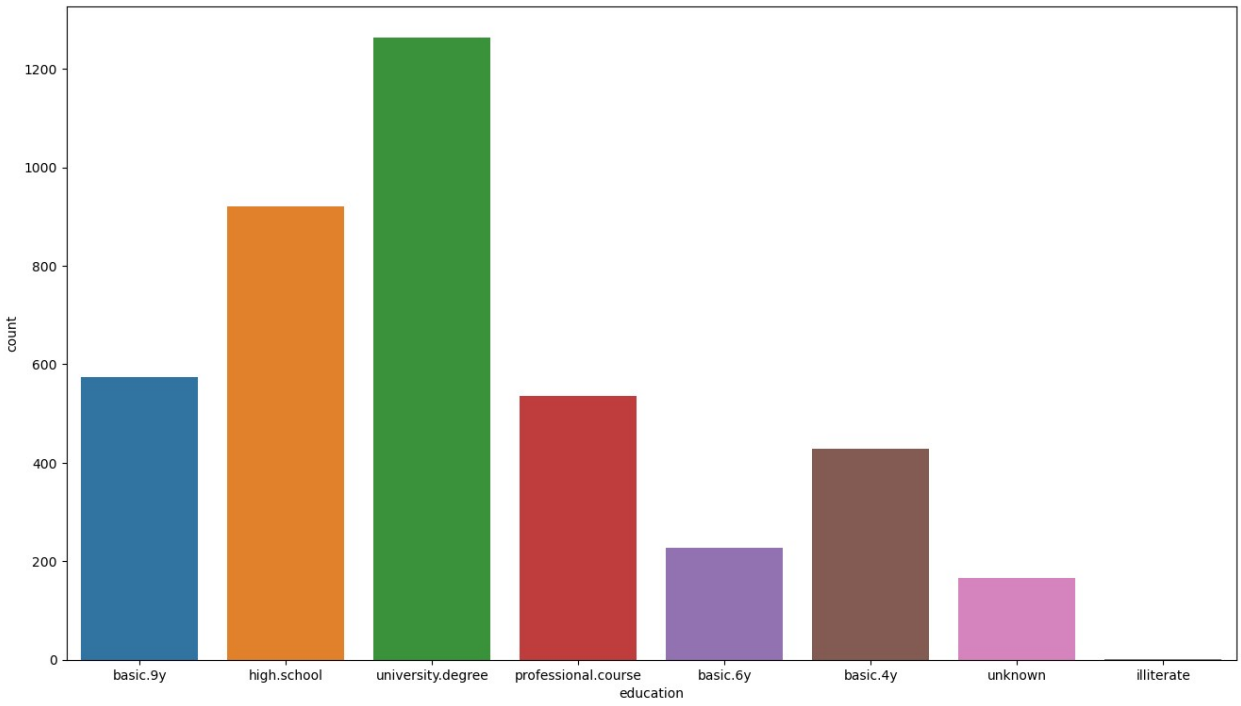

```
plt.figure(figsize = (16,9))
sns.countplot(x = "education",data = df)
```

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



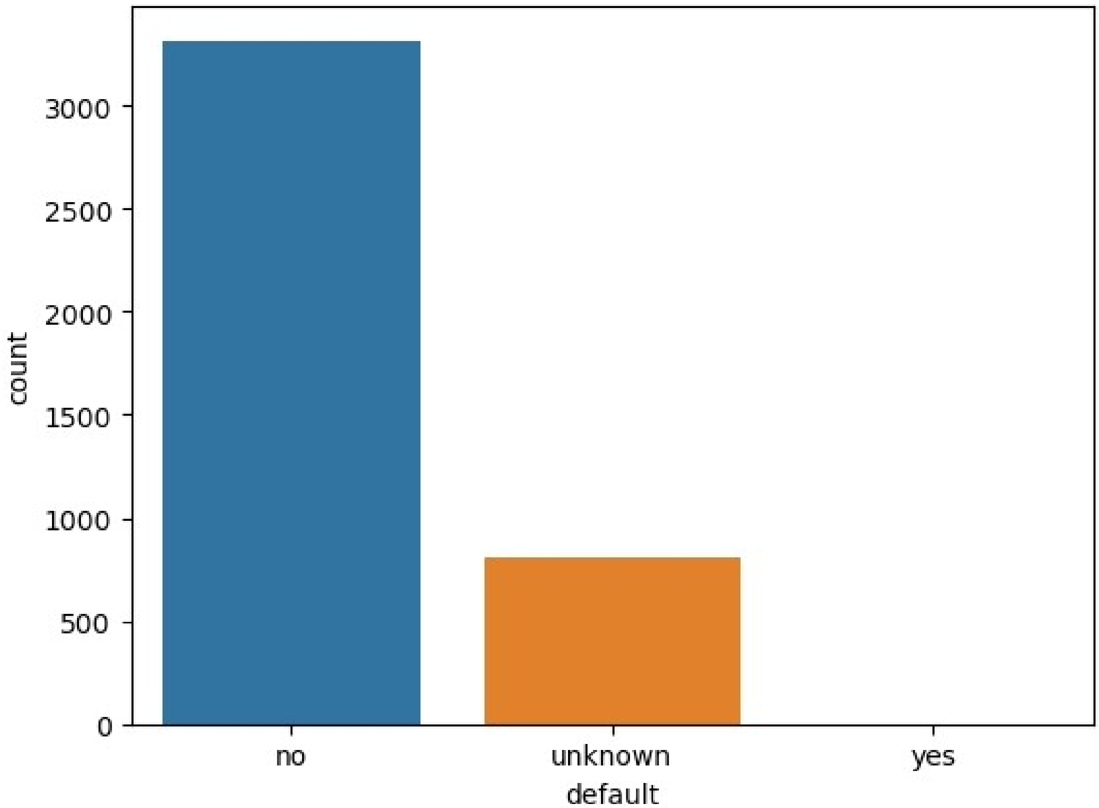

```
plt.figure(figsize = (16,9))
sns.countplot(x = "education",data = df)
```

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




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

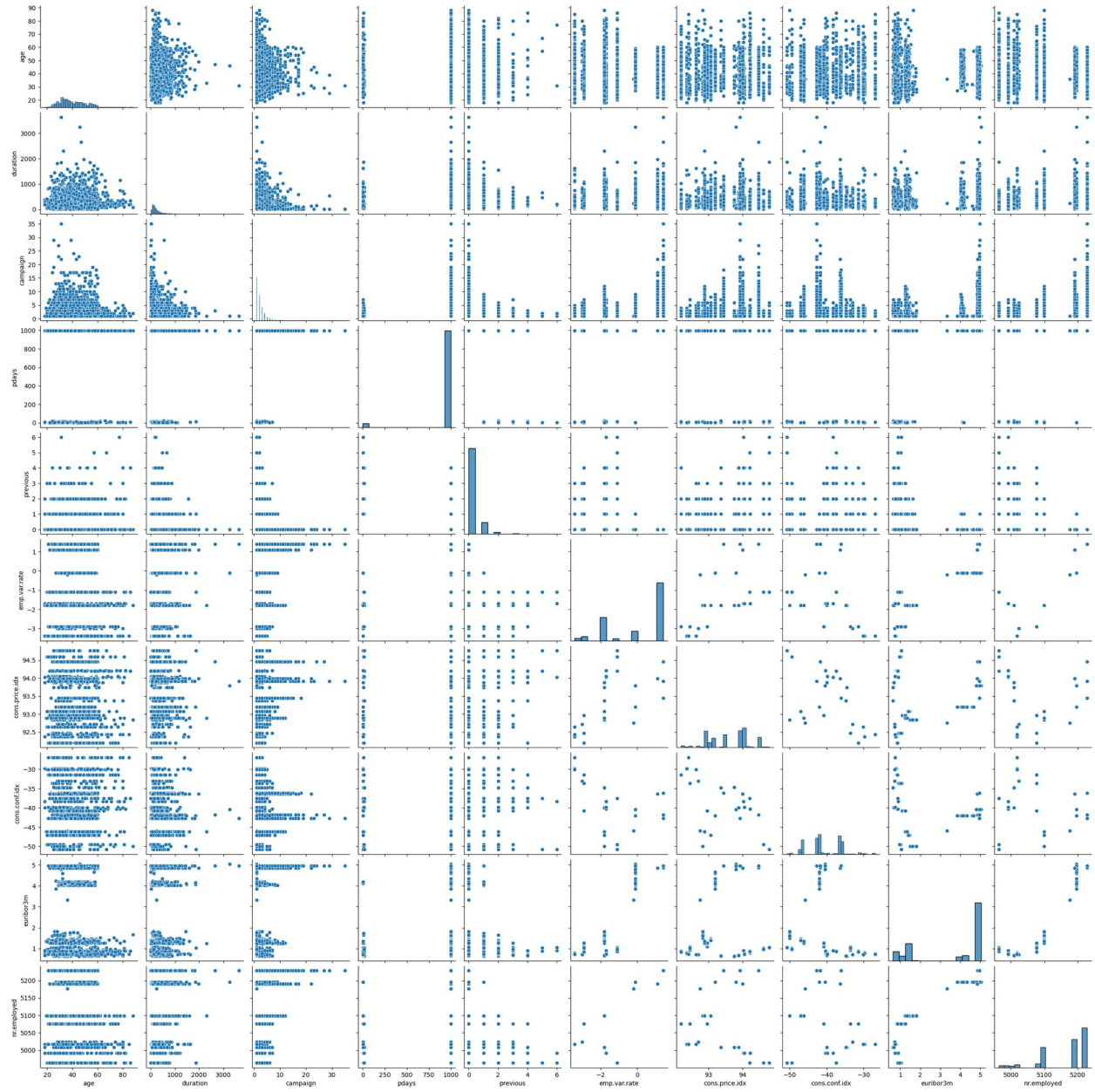
```
<Axes: xlabel='default', ylabel='count'>
```




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

```
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
```

```
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\jayaraman\anaconda3\Lib\site-packages\seaborn\
_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated
and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):

<seaborn.axisgrid.PairGrid at 0x1f6bd033190>

<Figure size 1600x900 with 0 Axes>
```



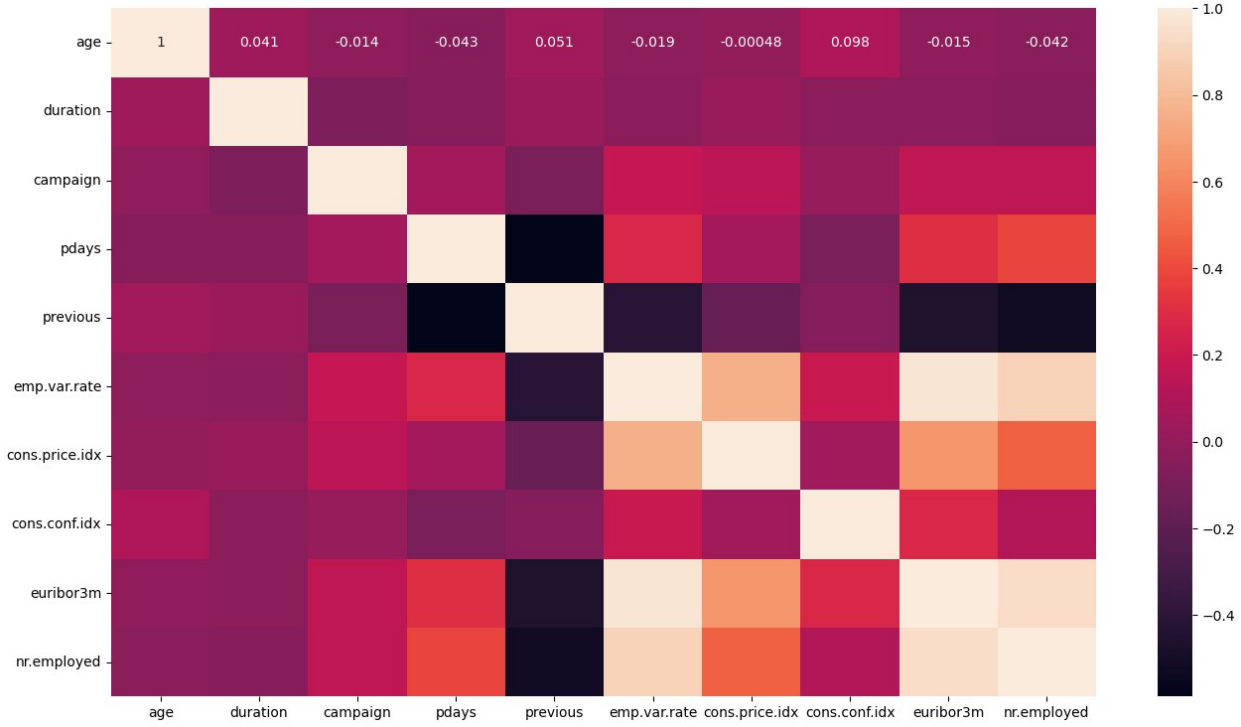

```
my_df=df.select_dtypes(exclude=[object])
my_df.corr()
```

	age	duration	campaign	pdays	previous	\
age	1.000000	0.041299	-0.014169	-0.043425	0.050931	
duration	0.041299	1.000000	-0.085348	-0.046998	0.025724	
campaign	-0.014169	-0.085348	1.000000	0.058742	-0.091490	
pdays	-0.043425	-0.046998	0.058742	1.000000	-0.587941	
previous	0.050931	0.025724	-0.091490	-0.587941	1.000000	
emp.var.rate	-0.019192	-0.028848	0.176079	0.270684	-0.415238	
cons.price.idx	-0.000482	0.016672	0.145021	0.058472	-0.164922	
cons.conf.idx	0.098135	-0.034745	0.007882	-0.092090	-0.051420	

```
euribor3m       -0.015033 -0.032329  0.159435  0.301478 -0.458851
nr.employed     -0.041936 -0.044218  0.161037  0.381983 -0.514853

                 emp.var.rate  cons.price.idx  cons.conf.idx  euribor3m
\
age                -0.019192       -0.000482       0.098135  -0.015033

duration           -0.028848        0.016672      -0.034745  -0.032329

campaign            0.176079        0.145021       0.007882   0.159435

pdays               0.270684        0.058472      -0.092090   0.301478

previous           -0.415238       -0.164922      -0.051420  -0.458851

emp.var.rate        1.000000        0.755155       0.195022   0.970308

cons.price.idx      0.755155        1.000000       0.045835   0.657159

cons.conf.idx       0.195022        0.045835       1.000000   0.276595

euribor3m           0.970308        0.657159       0.276595   1.000000

nr.employed         0.897173        0.472560       0.107054   0.942589


                nr.employed
age               -0.041936
duration          -0.044218
campaign           0.161037
pdays              0.381983
previous          -0.514853
emp.var.rate       0.897173
cons.price.idx     0.472560
cons.conf.idx      0.107054
euribor3m          0.942589
nr.employed        1.000000
```

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

```
<Axes: >
```




```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

df["job"] = le.fit_transform(df["job"])
df["marital"] = le.fit_transform(df["marital"])
df["education"] = le.fit_transform(df["education"])
df["default"] = le.fit_transform(df["default"])
df["loan"] = le.fit_transform(df["loan"])
df["contact"] = le.fit_transform(df["contact"])
df["poutcome"] = le.fit_transform(df["poutcome"])
df["housing"] = le.fit_transform(df["housing"])
df["month"] = le.fit_transform(df["month"])

df.head()
```

```
   age  job  marital  education  default  housing  loan  contact
month  \
0   30    1        1          2        0        2     0        0
6
1   39    7        2          3        0        0     0        1
6
2   25    7        1          3        0        2     0        1
4
3   38    7        1          2        0        1     1        1
4
4   47    0        1          6        0        2     0        0
7
```

```
  day_of_week  ...  campaign  pdays  previous  poutcome  emp.var.rate
\
0         fri  ...         2    999         0         1          -1.8

1         fri  ...         4    999         0         1           1.1

2         wed  ...         1    999         0         1           1.4

3         fri  ...         3    999         0         1           1.4

4         mon  ...         1    999         0         1          -0.1


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

[5 rows x 21 columns]
```

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

```
   age  job  marital  education  default  housing  loan  contact
month  \
0   30    1        1          2        0        2     0        0
6
1   39    7        2          3        0        0     0        1
6
2   25    7        1          3        0        2     0        1
4
3   38    7        1          2        0        1     1        1
4
4   47    0        1          6        0        2     0        0
7

  day_of_week  ...  campaign  pdays  previous  poutcome  emp.var.rate
\
0         fri  ...         2    999         0         1          -1.8

1         fri  ...         4    999         0         1           1.1

2         wed  ...         1    999         0         1           1.4

3         fri  ...         3    999         0         1           1.4

4         mon  ...         1    999         0         1          -0.1
```

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

[5 rows x 21 columns]
```