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

In [7]:
```
df = pd.read_csv(r'C:\Users\Lenovo\Desktop\SURAJ TASK DATASET\bank.csv',delimite
df.head()
```

Out[7]:

	age	job	marital	education	default	balance	housing	loan	contact	da
0	30	unemployed	married	primary	no	1787	no	no	cellular	
1	33	services	married	secondary	no	4789	yes	yes	cellular	
2	35	management	single	tertiary	no	1350	yes	no	cellular	
3	30	management	married	tertiary	no	1476	yes	yes	unknown	
4	59	blue-collar	married	secondary	no	0	yes	no	unknown	

In [8]:
```
df.tail()
```

Out[8]:

	age	job	marital	education	default	balance	housing	loan	contact
4516	33	services	married	secondary	no	-333	yes	no	cellular
4517	57	self-employed	married	tertiary	yes	-3313	yes	yes	unknown
4518	57	technician	married	secondary	no	295	no	no	cellular
4519	28	blue-collar	married	secondary	no	1137	no	no	cellular
4520	44	entrepreneur	single	tertiary	no	1136	yes	yes	cellular

In [9]:
```
df.shape
```

Out[9]:
```
(4521, 17)
```

In [10]:
```
df.columns
```

Out[10]:
```
Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housing',
       'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pdays',
       'previous', 'poutcome', 'y'],
      dtype='object')
```

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4521 entries, 0 to 4520
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   age        4521 non-null   int64 
 1   job        4521 non-null   object
 2   marital    4521 non-null   object
 3   education  4521 non-null   object
 4   default    4521 non-null   object
 5   balance    4521 non-null   int64 
 6   housing    4521 non-null   object
 7   loan       4521 non-null   object
 8   contact    4521 non-null   object
 9   day        4521 non-null   int64 
 10  month      4521 non-null   object
 11  duration   4521 non-null   int64 
 12  campaign   4521 non-null   int64 
 13  pdays      4521 non-null   int64 
 14  previous   4521 non-null   int64 
 15  poutcome   4521 non-null   object
 16  y          4521 non-null   object
dtypes: int64(7), object(10)
memory usage: 600.6+ KB
```

In [12]:
```
df.describe()
```

Out[12]:

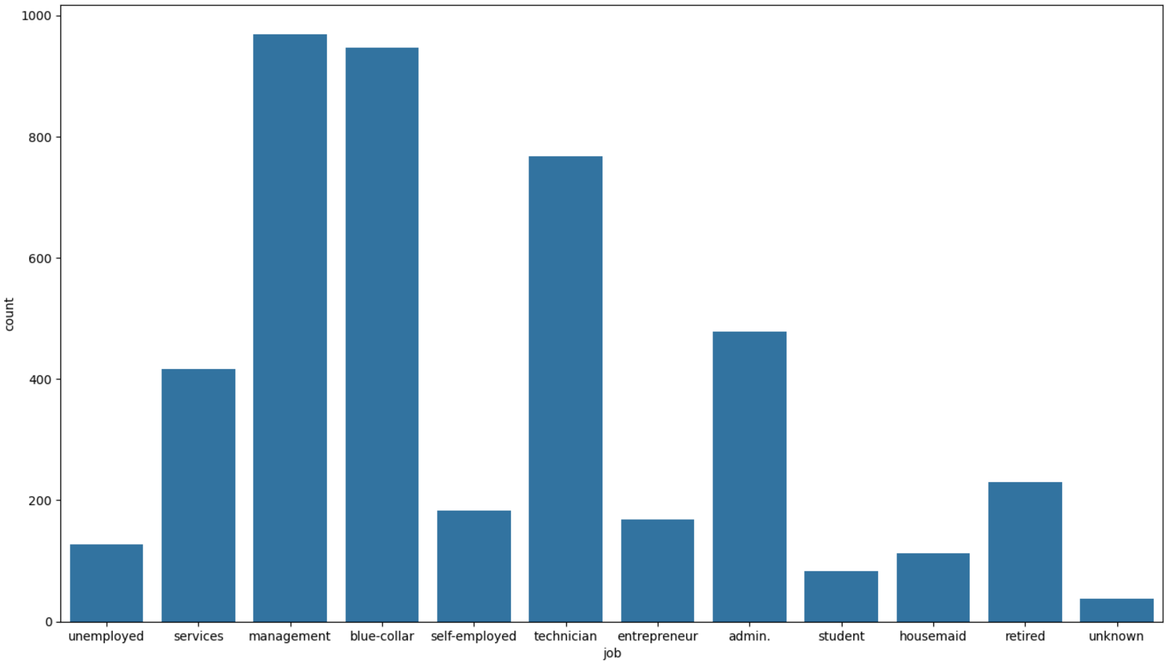
	age	balance	day	duration	campaign	pdays
count	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000	4521.000000
mean	41.170095	1422.657819	15.915284	263.961292	2.793630	39.766645
std	10.576211	3009.638142	8.247667	259.856633	3.109807	100.121124
min	19.000000	-3313.000000	1.000000	4.000000	1.000000	-1.000000
25%	33.000000	69.000000	9.000000	104.000000	1.000000	-1.000000
50%	39.000000	444.000000	16.000000	185.000000	2.000000	-1.000000
75%	49.000000	1480.000000	21.000000	329.000000	3.000000	-1.000000
max	87.000000	71188.000000	31.000000	3025.000000	50.000000	871.000000

In [13]:
```
df.isnull().sum()
```

```
Out[13]: age          0
         job          0
         marital      0
         education    0
         default      0
         balance      0
         housing      0
         loan         0
         contact      0
         day          0
         month        0
         duration     0
         campaign     0
         pdays        0
         previous     0
         poutcome     0
         y            0
         dtype: int64
```

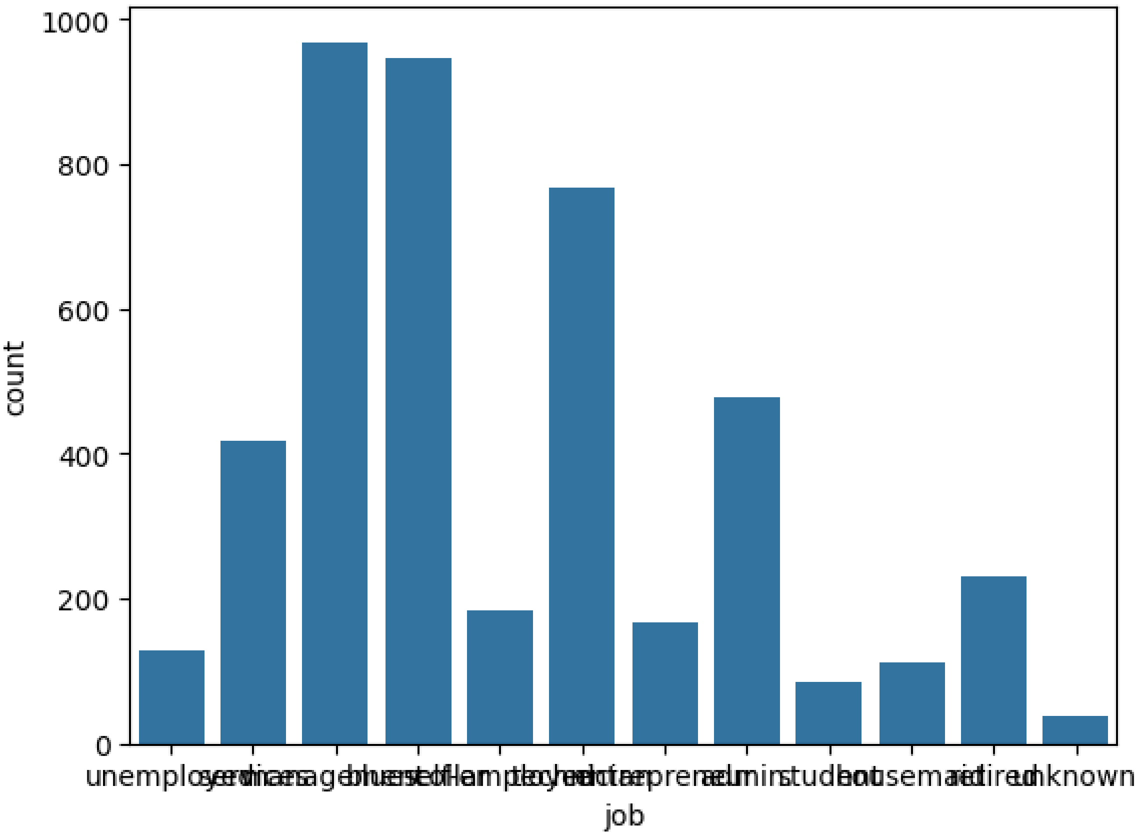
```
In [14]: plt.figure(figsize = (16,9))
         sns.countplot(x = "job",data = df)
```

```
Out[14]: <Axes: xlabel='job', ylabel='count'>
```



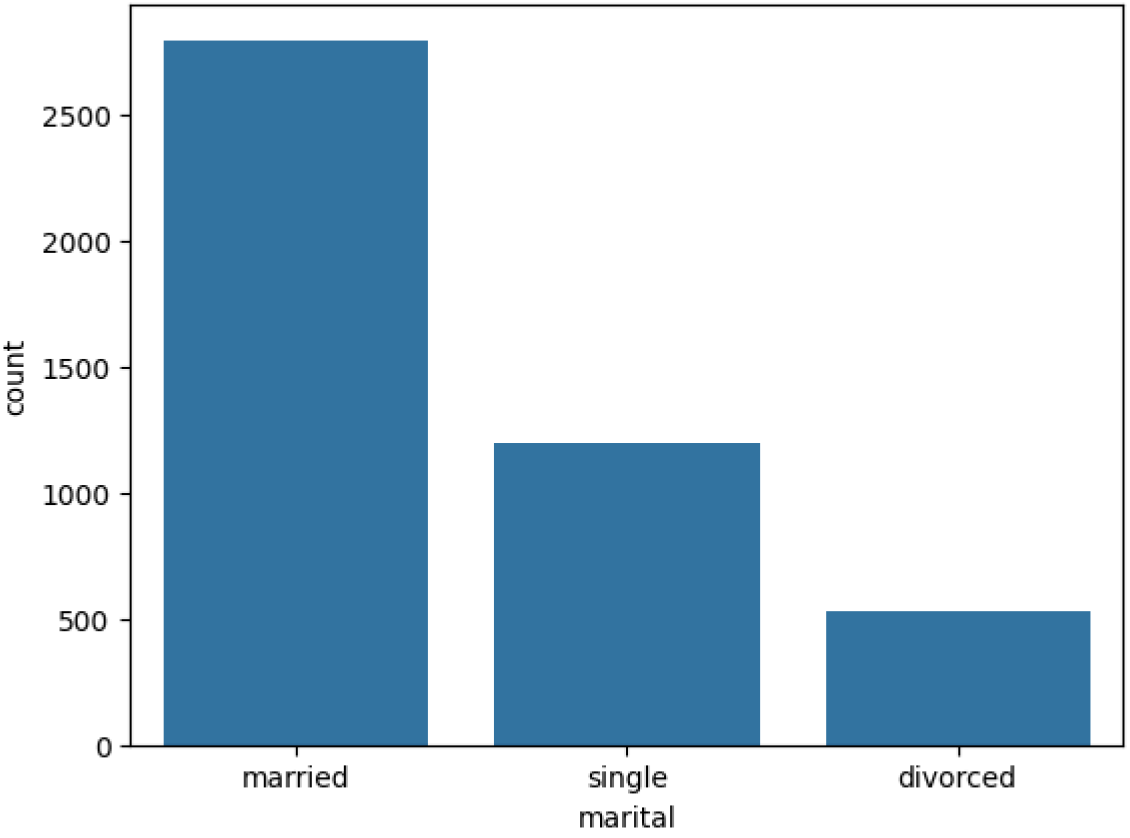
```
In [15]: sns.countplot(x = "job",data = df)
```

```
Out[15]: <Axes: xlabel='job', ylabel='count'>
```



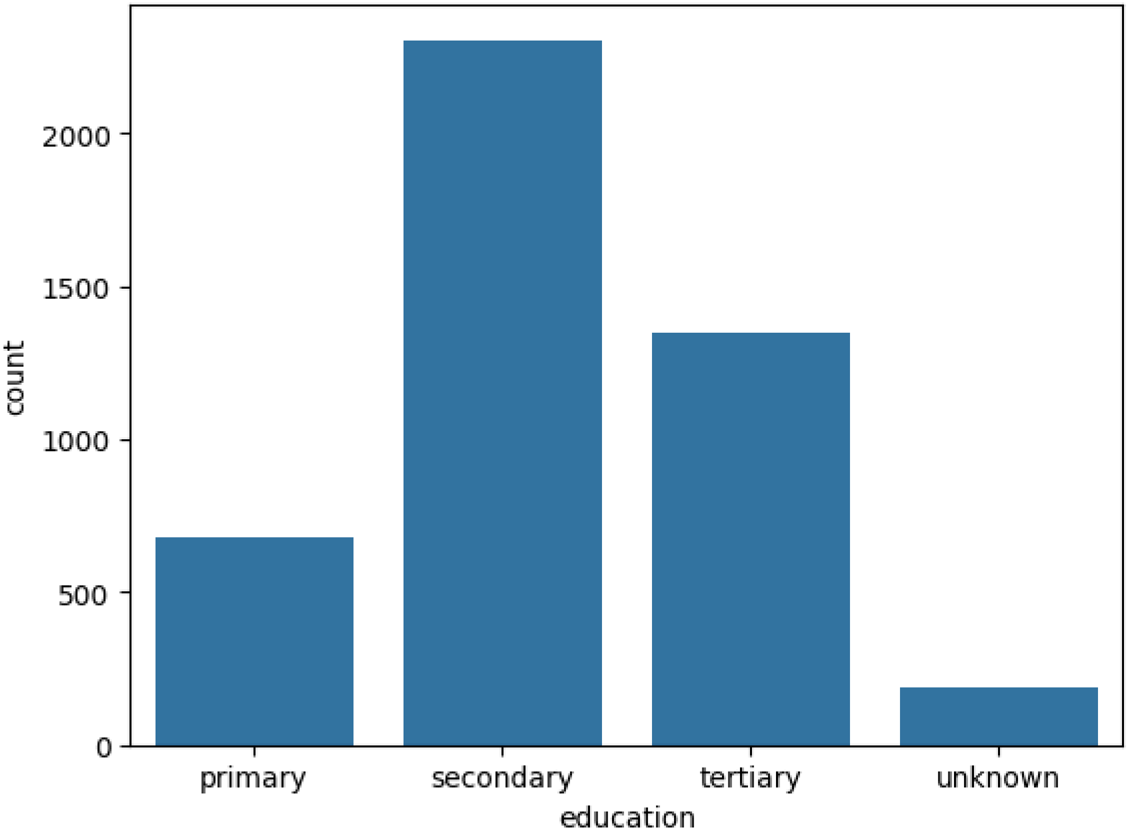
```
In [16]: sns.countplot(x = "marital",data = df)
```

```
Out[16]: <Axes: xlabel='marital', ylabel='count'>
```



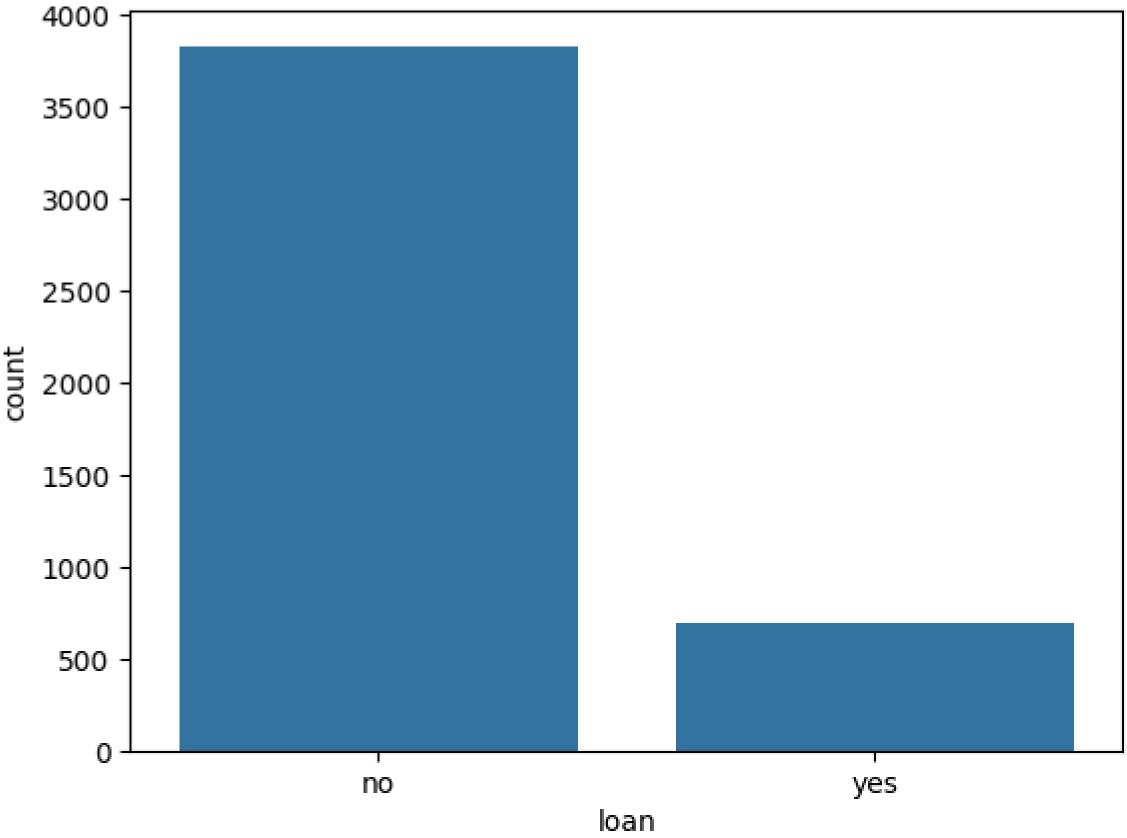
```
In [17]: sns.countplot(x = "education",data = df)
```

```
Out[17]: <Axes: xlabel='education', ylabel='count'>
```



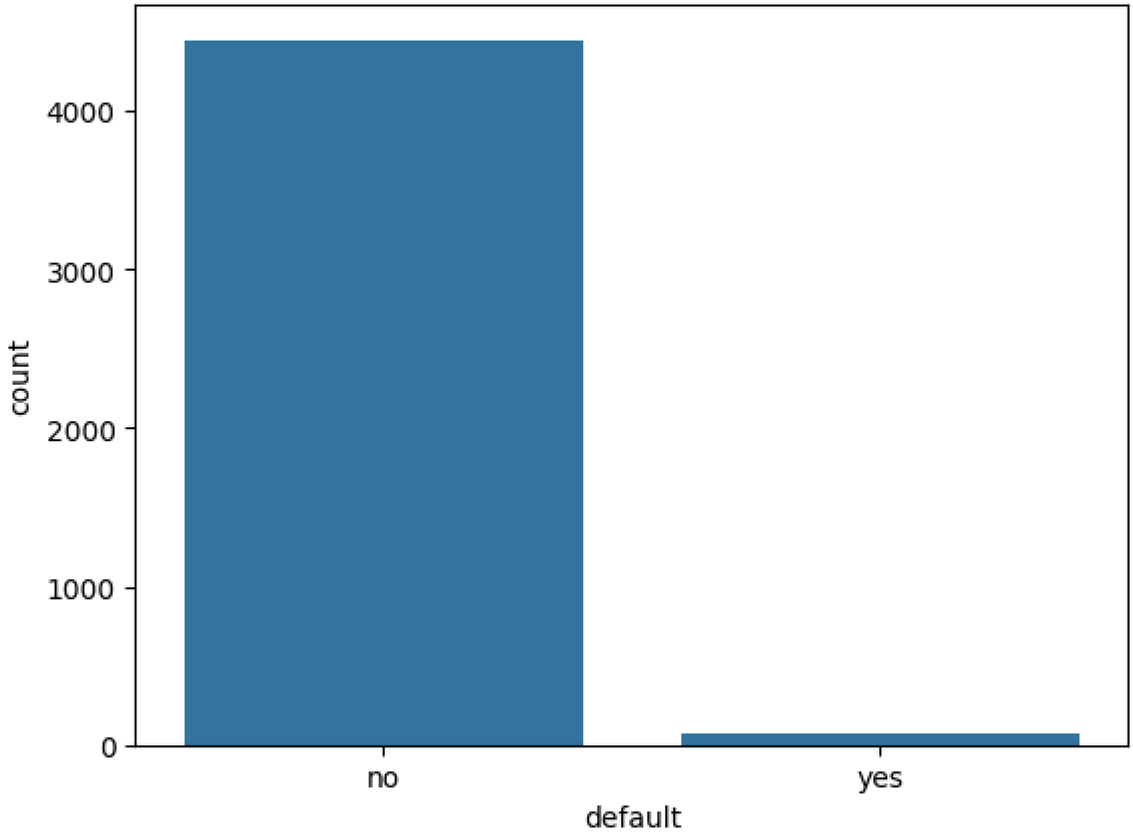
```
In [18]: sns.countplot(x = "loan",data = df)
```

```
Out[18]: <Axes: xlabel='loan', ylabel='count'>
```



```
In [21]: sns.countplot(x = "default",data = df)
```

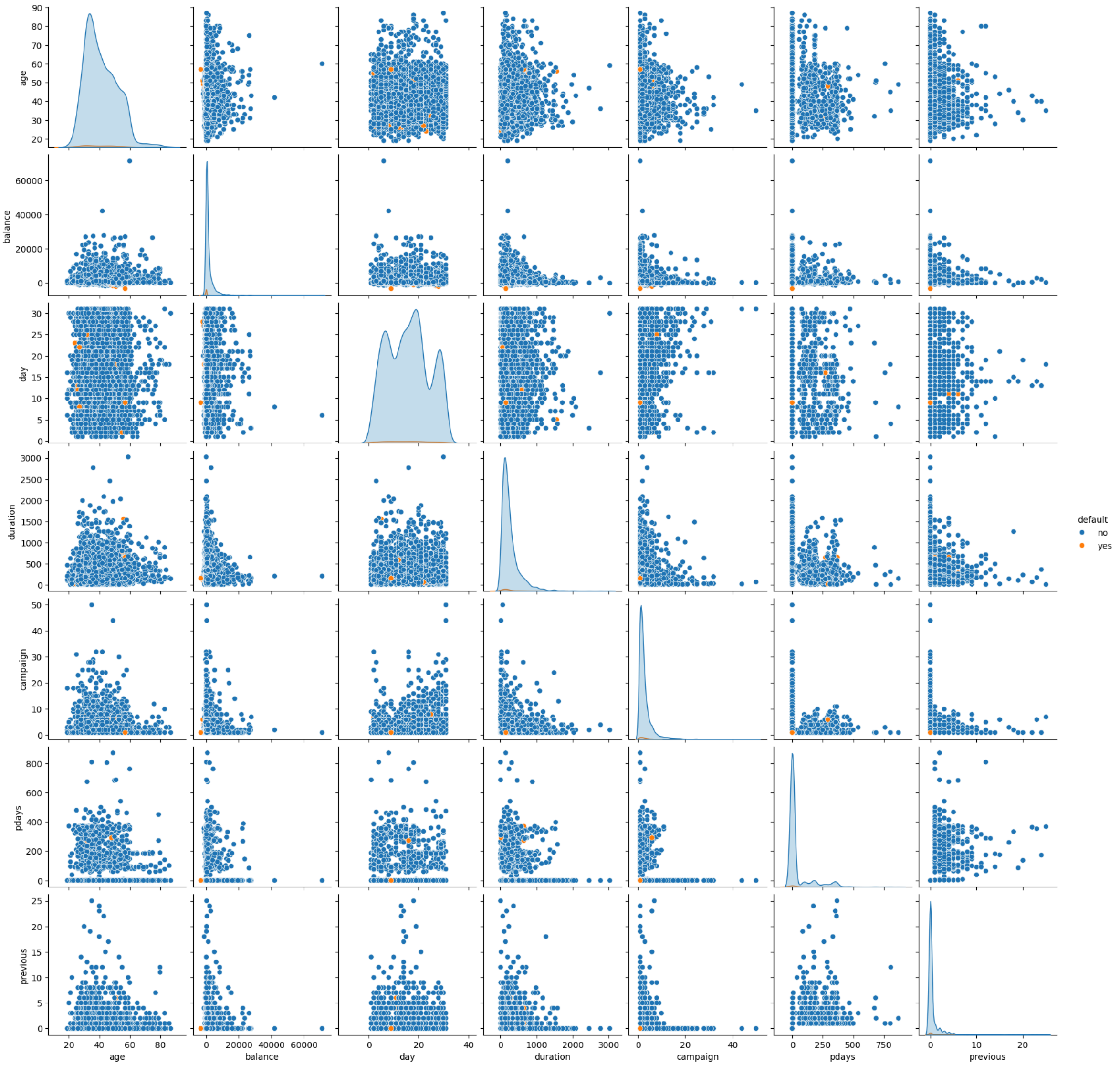
```
Out[21]: <Axes: xlabel='default', ylabel='count'>
```



```
In [23]: plt.figure(figsize = (16,9))
         sns.pairplot(data = df,hue = "default")

Out[23]: <seaborn.axisgrid.PairGrid at 0x1abb26f5e20>

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



In [24]:

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

Out[24]:

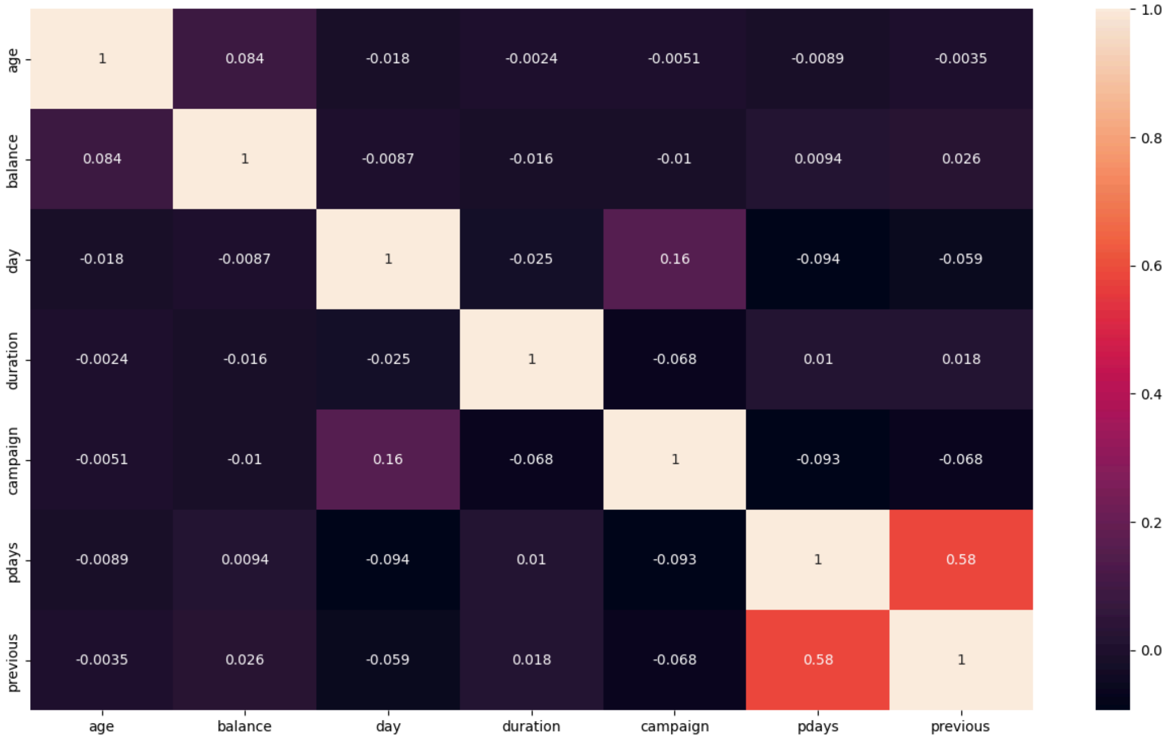
	age	balance	day	duration	campaign	pdays	previous
age	1.000000	0.083820	-0.017853	-0.002367	-0.005148	-0.008894	-0.003511
balance	0.083820	1.000000	-0.008677	-0.015950	-0.009976	0.009437	0.026196
day	-0.017853	-0.008677	1.000000	-0.024629	0.160706	-0.094352	-0.059114
duration	-0.002367	-0.015950	-0.024629	1.000000	-0.068382	0.010380	0.018080
campaign	-0.005148	-0.009976	0.160706	-0.068382	1.000000	-0.093137	-0.067833
pdays	-0.008894	0.009437	-0.094352	0.010380	-0.093137	1.000000	0.577562
previous	-0.003511	0.026196	-0.059114	0.018080	-0.067833	0.577562	1.000000

In [26]:

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

Out[26]:

```
<Axes: >
```



	age	balance	day	duration	campaign	pdays	previous
age	1	0.084	-0.018	-0.0024	-0.0051	-0.0089	-0.0035
balance	0.084	1	-0.0087	-0.016	-0.01	0.0094	0.026
day	-0.018	-0.0087	1	-0.025	0.16	-0.094	-0.059
duration	-0.0024	-0.016	-0.025	1	-0.068	0.01	0.018
campaign	-0.0051	-0.01	0.16	-0.068	1	-0.093	-0.068
pdays	-0.0089	0.0094	-0.094	0.01	-0.093	1	0.58
previous	-0.0035	0.026	-0.059	0.018	-0.068	0.58	1

In [29]:

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

In [31]:

```
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"])
```

In [32]:

```
df.head()
```

Out[32]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	30	10	1	0	0	1787	0	0	0	19	10
1	33	7	1	1	0	4789	1	1	0	11	8
2	35	4	2	2	0	1350	1	0	0	16	0
3	30	4	1	2	0	1476	1	1	2	3	6
4	59	1	1	1	0	0	1	0	2	5	8

In [34]:

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

Out[34]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	30	10	1	0	0	1787	0	0	0	19	10
1	33	7	1	1	0	4789	1	1	0	11	8
2	35	4	2	2	0	1350	1	0	0	16	0
3	30	4	1	2	0	1476	1	1	2	3	6
4	59	1	1	1	0	0	1	0	2	5	8

In [ ]: