```
In [1]:   import numpy as np
          import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns
```

```
In [2]:   df = pd.read_csv('C:\\Users\\lenovo\\Desktop\\courses\\Task3\\bank-additional.csv',
```

```
In [3]:   df.head()
```

Out[3]:

	age	job	marital	education	default	housing	loan	contact	month	day_of_
0	30	blue-collar	married	basic.9y	no	yes	no	cellular	may	
1	39	services	single	high.school	no	no	no	telephone	may	
2	25	services	married	high.school	no	yes	no	telephone	jun	
3	38	services	married	basic.9y	no	unknown	unknown	telephone	jun	
4	47	admin.	married	university.degree	no	yes	no	cellular	nov	

5 rows × 21 columns

```
In [6]:   df.tail()
```

Out[6]:

	age	job	marital	education	default	housing	loan	contact	month	day_of_w
4114	30	admin.	married	basic.6y	no	yes	yes	cellular	jul	
4115	39	admin.	married	high.school	no	yes	no	telephone	jul	
4116	27	student	single	high.school	no	no	no	cellular	may	
4117	58	admin.	married	high.school	no	no	no	cellular	aug	
4118	34	management	single	high.school	no	yes	no	cellular	nov	

5 rows × 21 columns

```
In [5]:   df.shape
```

Out[5]:   (4119, 21)

```
In [7]:   df.columns
```

Out[7]:   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', 'target'],
                dtype='object')

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

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
<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  target          4119 non-null   object
dtypes: float64(5), int64(5), object(11)
memory usage: 675.9+ KB
```

In [9]: `df.describe()`

Out[9]:

	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.id
count	4119.000000	4119.000000	4119.000000	4119.000000	4119.000000	4119.000000	4119.00000
mean	40.113620	256.788055	2.537266	960.422190	0.190337	0.084972	93.57970
std	10.313362	254.703736	2.568159	191.922786	0.541788	1.563114	0.57934
min	18.000000	0.000000	1.000000	0.000000	0.000000	-3.400000	92.20100
25%	32.000000	103.000000	1.000000	999.000000	0.000000	-1.800000	93.07500
50%	38.000000	181.000000	2.000000	999.000000	0.000000	1.100000	93.74900
75%	47.000000	317.000000	3.000000	999.000000	0.000000	1.400000	93.99400
max	88.000000	3643.000000	35.000000	999.000000	6.000000	1.400000	94.76700

In [10]: `df.isnull().sum()`

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Out[10]:
```
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
target             0
dtype: int64
```
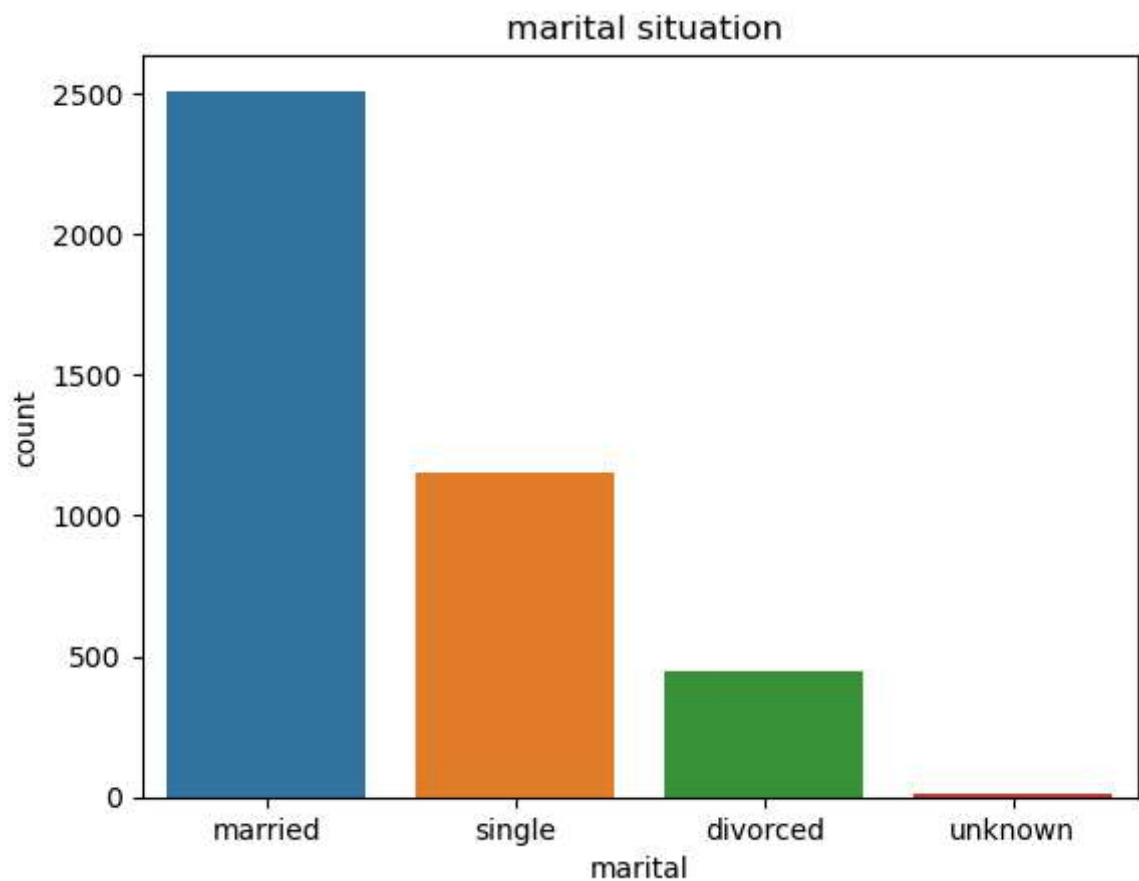
In [11]:
```python
plt.figure(figsize=(16, 9))
sns.countplot(x="job", data=df)
plt.title("Count of Jobs")
plt.show()
```
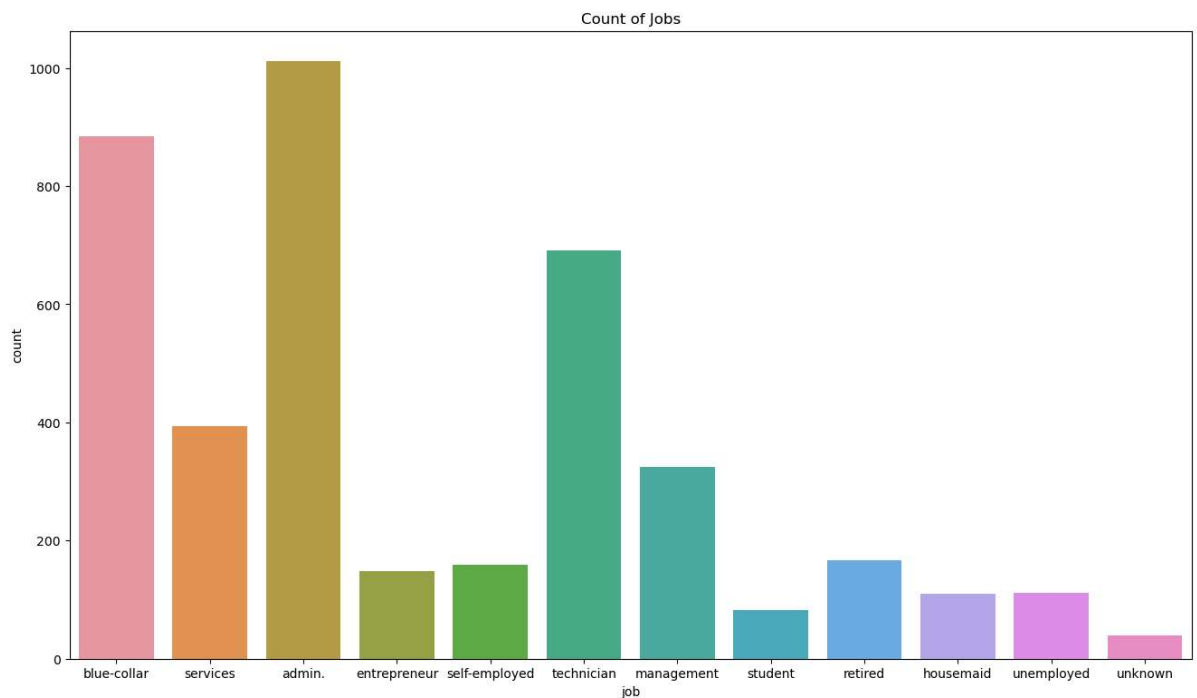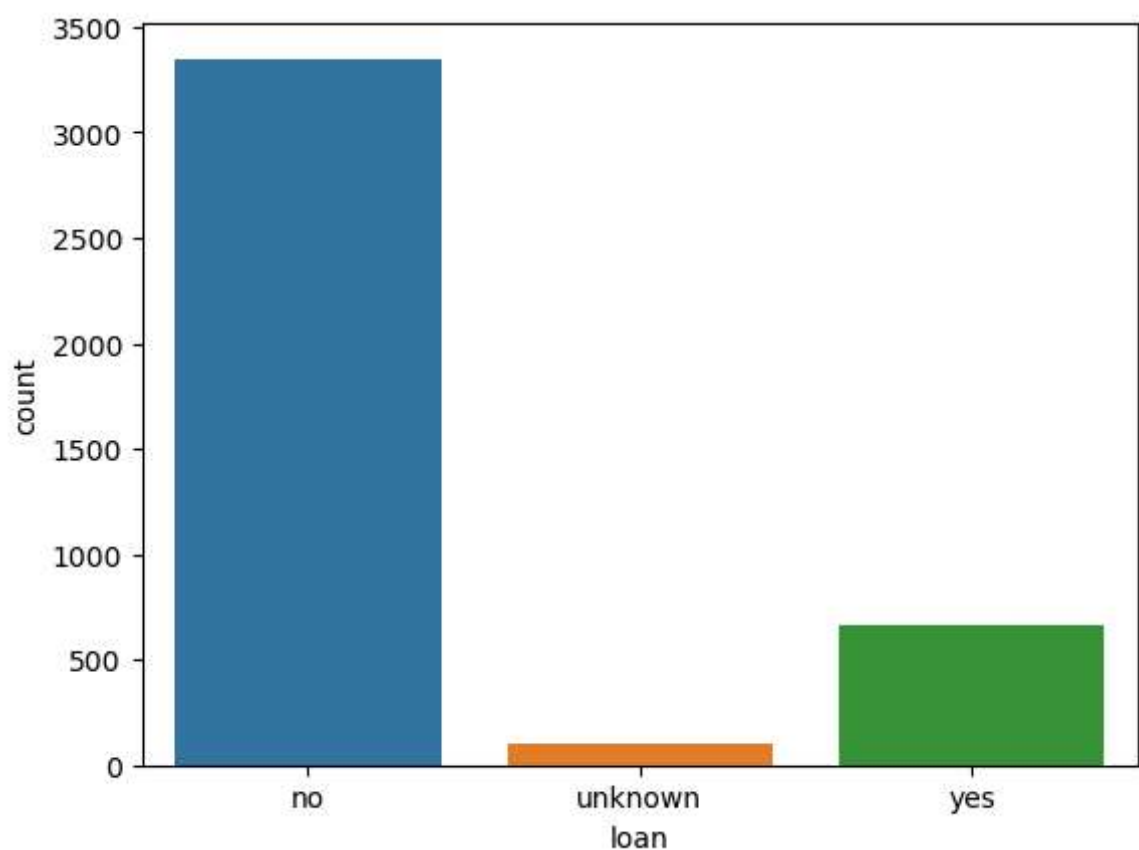

In [12]:
```python
sns.countplot(x = "marital",data = df)
plt.title(" marital situation")
```

Out[12]:
```
Text(0.5, 1.0, ' marital situation')
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

marital situation


```
In [13]:   sns.countplot(x = "loan",data = df)
```

```
Out[13]:   <Axes: xlabel='loan', ylabel='count'>
```
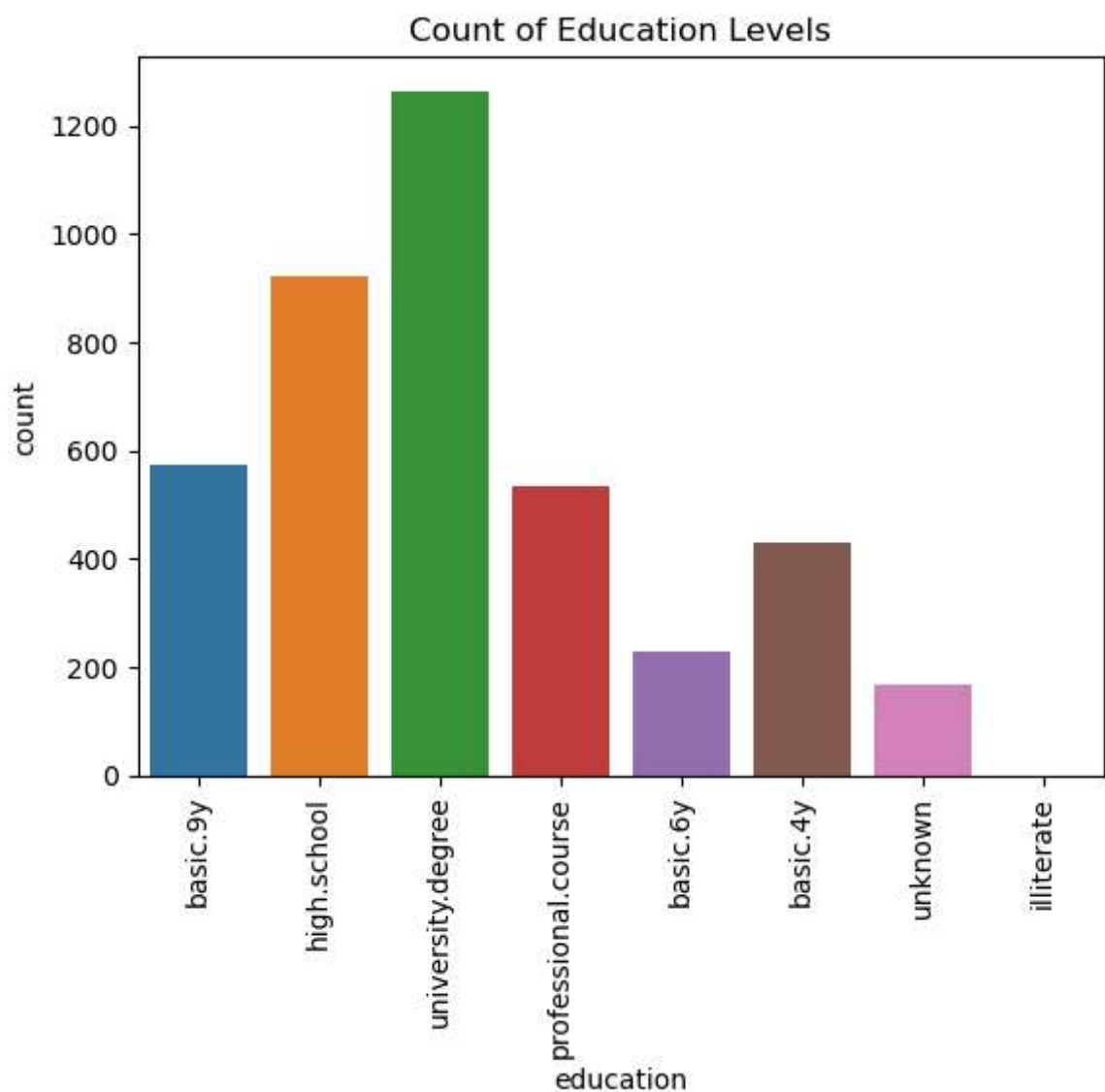


```
In [14]:   plt.xticks(rotation=90)
           sns.countplot(x = "education", data = df)
```
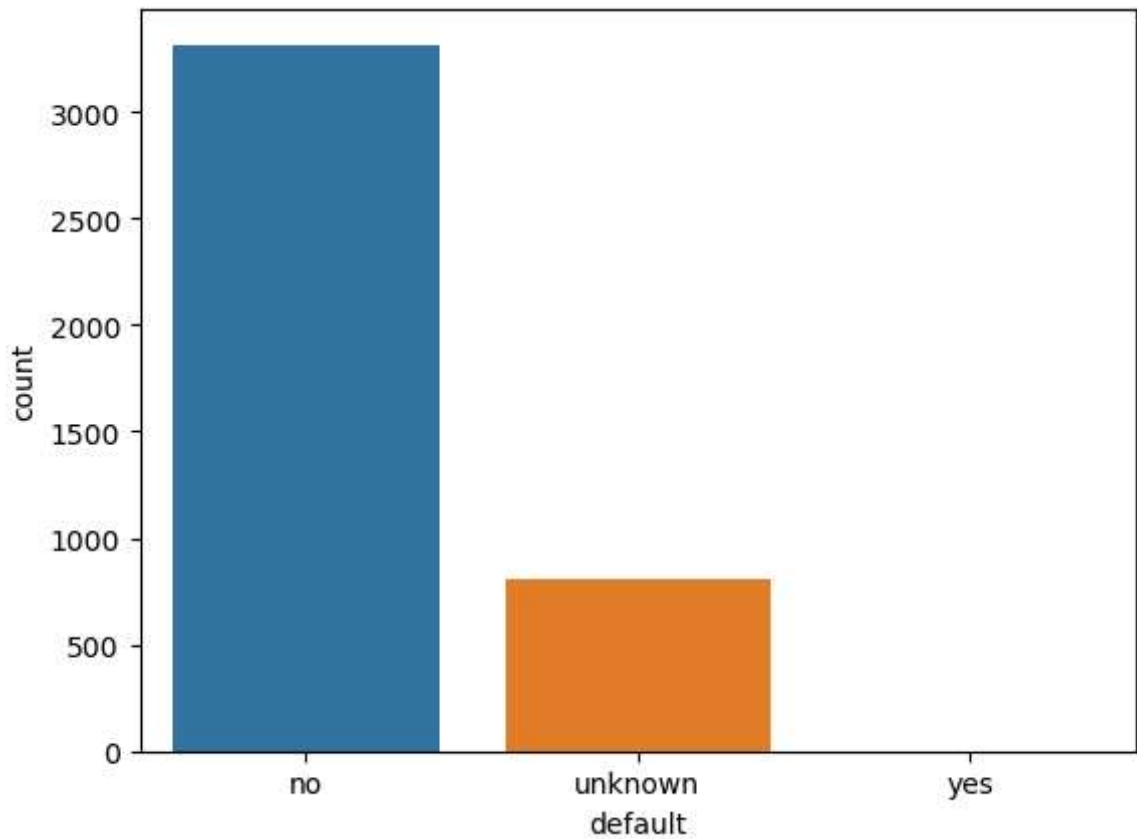
Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Out[14]:   `Text(0.5, 1.0, 'Count of Education Levels')`

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

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

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
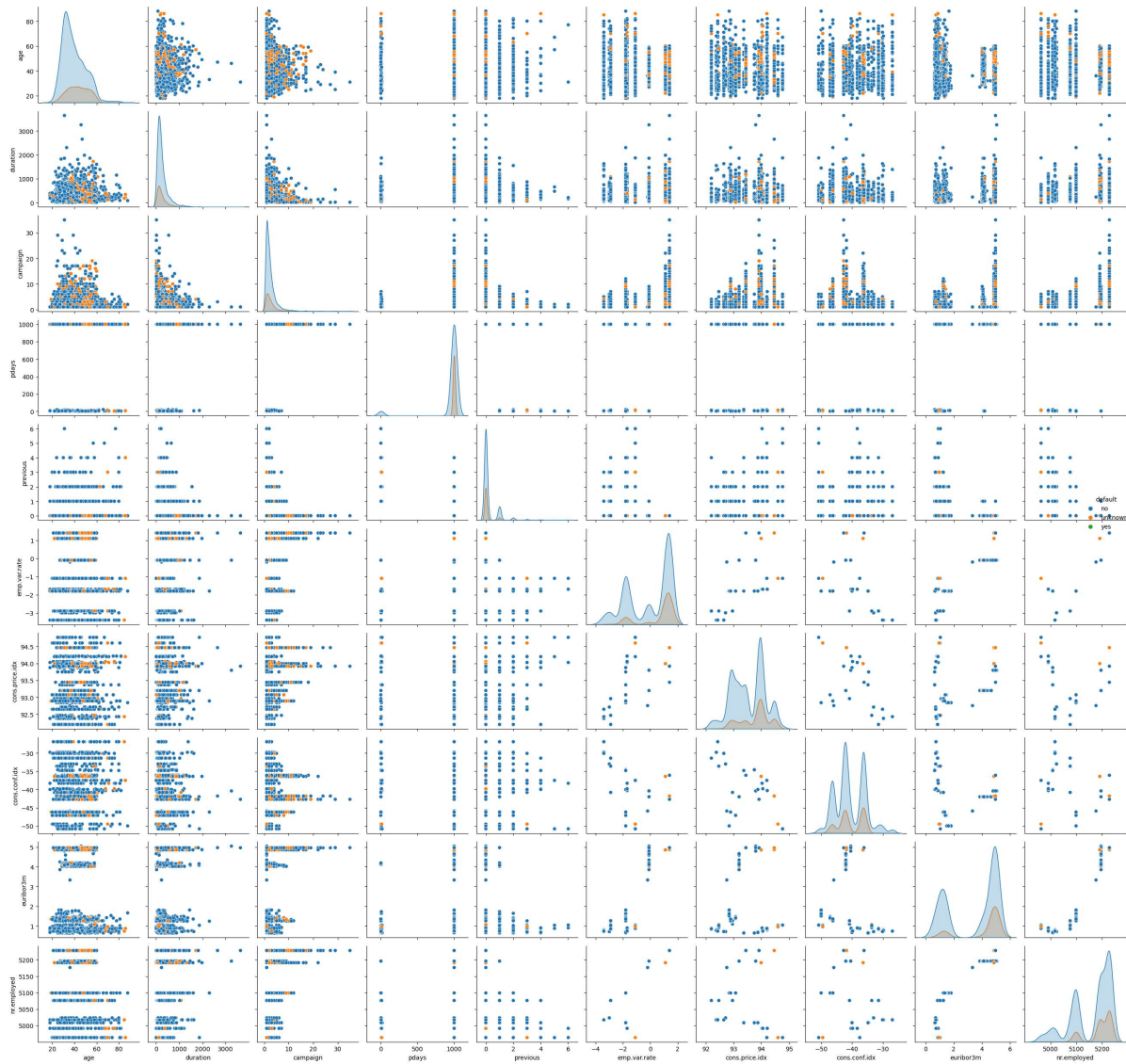In [16]:  plt.figure(figsize=(17, 9))
          sns.pairplot(data=df, hue="default")
          plt.suptitle("Pairplot of Variables with Default Hue", y=1.02)
          plt.tight_layout()
          plt.show()
```

```
C:\Users\lenovo\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning:
The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
C:\Users\lenovo\AppData\Local\Temp\ipykernel_3212\3805054004.py:4: UserWarning: Th
e figure layout has changed to tight
  plt.tight_layout()
<Figure size 1700x900 with 0 Axes>
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Pairplot of Variables with Default Hue


```
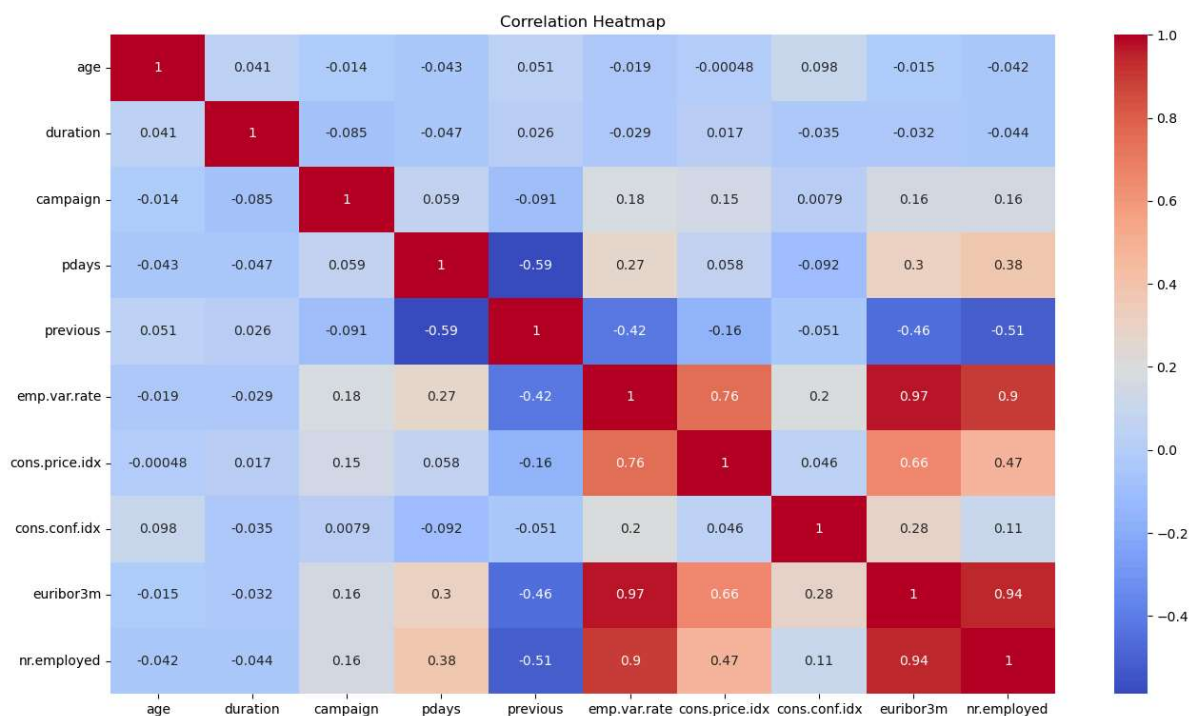In [17]: my_df=df.select_dtypes(exclude=[object])
         my_df.corr()
```

Out[17]:

	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.idx	c
age	1.000000	0.041299	-0.014169	-0.043425	0.050931	-0.019192	-0.000482	
duration	0.041299	1.000000	-0.085348	-0.046998	0.025724	-0.028848	0.016672	
campaign	-0.014169	-0.085348	1.000000	0.058742	-0.091490	0.176079	0.145021	
pdays	-0.043425	-0.046998	0.058742	1.000000	-0.587941	0.270684	0.058472	
previous	0.050931	0.025724	-0.091490	-0.587941	1.000000	-0.415238	-0.164922	
emp.var.rate	-0.019192	-0.028848	0.176079	0.270684	-0.415238	1.000000	0.755155	
cons.price.idx	-0.000482	0.016672	0.145021	0.058472	-0.164922	0.755155	1.000000	
cons.conf.idx	0.098135	-0.034745	0.007882	-0.092090	-0.051420	0.195022	0.045835	
euribor3m	-0.015033	-0.032329	0.159435	0.301478	-0.458851	0.970308	0.657159	
nr.employed	-0.041936	-0.044218	0.161037	0.381983	-0.514853	0.897173	0.472560	

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

```
sns.heatmap(my_df.corr(), annot=True, cmap="coolwarm")
```

```
plt.title("Correlation Heatmap")
plt.show()
```


Correlation Heatmap

In [23]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
print(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 target
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]
```

In [24]:
```
df.drop(["pdays","previous","poutcome"],axis = 1)
df.head()
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

Out[24]:

	age	job	marital	education	default	housing	loan	contact	month	day_of_
0	30	blue-collar	married	basic.9y	no	yes	no	cellular	may	
1	39	services	single	high.school	no	no	no	telephone	may	
2	25	services	married	high.school	no	yes	no	telephone	jun	
3	38	services	married	basic.9y	no	unknown	unknown	telephone	jun	
4	47	admin.	married	university.degree	no	yes	no	cellular	nov	

5 rows × 21 columns

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js