

In [1]:

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [5]:

```
sal=pd.read_csv('G:\Data Analysis\output.csv')
```

In [12]:

```
print (str(sal.shape[0])+" "+str(sal.shape[1]))
```

3498 22

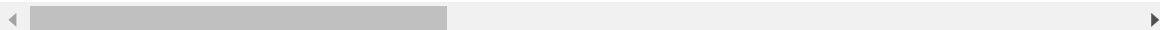
In [13]:

```
sal.head()
```

Out[13]:

	Registration Number	CTC	Joining_date	Leaving_date	Role	JobCity	Gender	DOB	10per
0	432376	4.20	01-06-2012 00:00	present	senior quality engineer	Bangalore	f	19- 02- 1990 00:00	
1	432383	5.00	01-09-2013 00:00	present	assistant manager	Indore	m	04- 10- 1989 00:00	
2	432390	3.25	01-06-2014 00:00	present	systems engineer	Chennai	f	03- 08- 1992 00:00	
3	432397	11.00	01-07-2011 00:00	present	senior software engineer	Gurgaon	m	05- 12- 1989 00:00	
4	432404	2.00	01-03-2014 00:00	01-03-2015 00:00	get	Manesar	m	27- 02- 1991 00:00	

5 rows × 22 columns



In [14]:

```
sal.describe()
```

Out[14]:

	Registration Number	CTC	10percentage	12graduation	12percentage	CollegeID
count	3498.000000	3498.000000	3398.000000	3498.000000	3442.000000	3498.000000
mean	444615.500000	3.082258	77.936004	2008.085477	74.403451	5301.473699
std	7069.509636	2.203398	9.839156	1.657165	10.938945	2780.436629
min	432376.000000	0.350000	43.000000	1995.000000	40.000000	1431.000000
25%	438495.750000	1.800000	72.000000	2007.000000	66.037500	2538.750000
50%	444615.500000	3.000000	79.000000	2008.000000	74.140000	5141.000000
75%	450735.250000	3.700000	85.745000	2009.000000	82.382500	7631.250000
max	456855.000000	40.000000	97.760000	2013.000000	98.700000	10874.000000

In [3]:

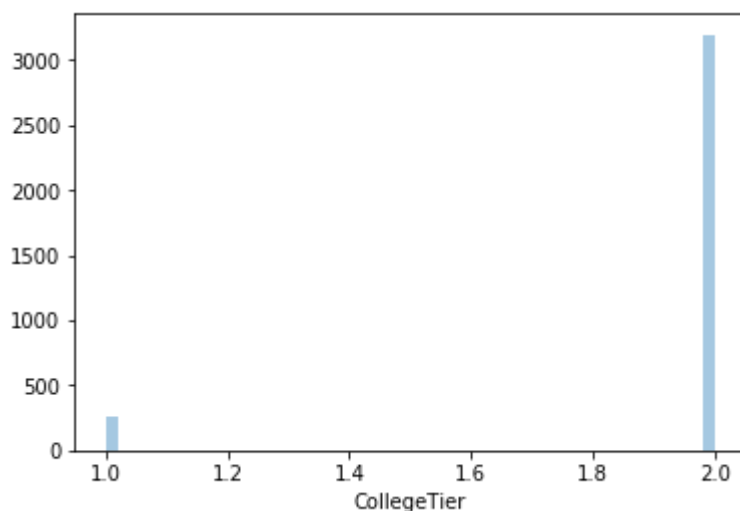
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [8]:

```
sal=pd.read_csv('G:\Data Analysis\Salary Data_Sharath.csv')
sns.distplot(sal.CollegeTier.dropna(),kde=False)
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x179231a8390>
```



In [10]:

```
sal=pd.read_csv('G:\Data Analysis\Salary Data_Sharath.csv')
sal.head()
```

Out[10]:

	Registration Number	CTC	Joining_date	Leaving_date	Role	JobCity	Gender	DOB	t10pe
0	432376	4.20	01-06-2012 00:00	present	senior quality engineer	Bangalore	f	19- 02- 1990 00:00	
1	432383	5.00	01-09-2013 00:00	present	assistant manager	Indore	m	04- 10- 1989 00:00	
2	432390	3.25	01-06-2014 00:00	present	systems engineer	Chennai	f	03- 08- 1992 00:00	
3	432397	11.00	01-07-2011 00:00	present	senior software engineer	Gurgaon	m	05- 12- 1989 00:00	
4	432404	2.00	01-03-2014 00:00	01-03-2015 00:00	get	Manesar	m	27- 02- 1991 00:00	

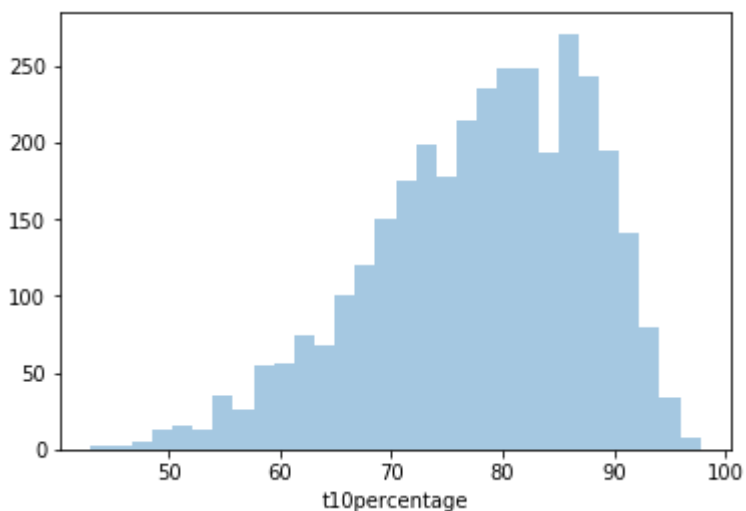
5 rows × 22 columns

In [41]:

```
sns.distplot(sal.t10percentage.dropna(),kde=False)
```

Out[41]:

<matplotlib.axes._subplots.AxesSubplot at 0x17923891518>

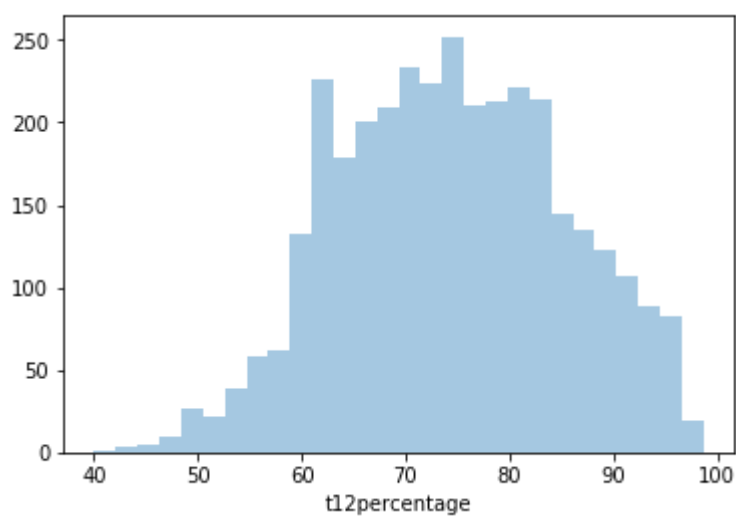


In [12]:

```
sns.distplot(sal.t12percentage.dropna(),kde=False)
```

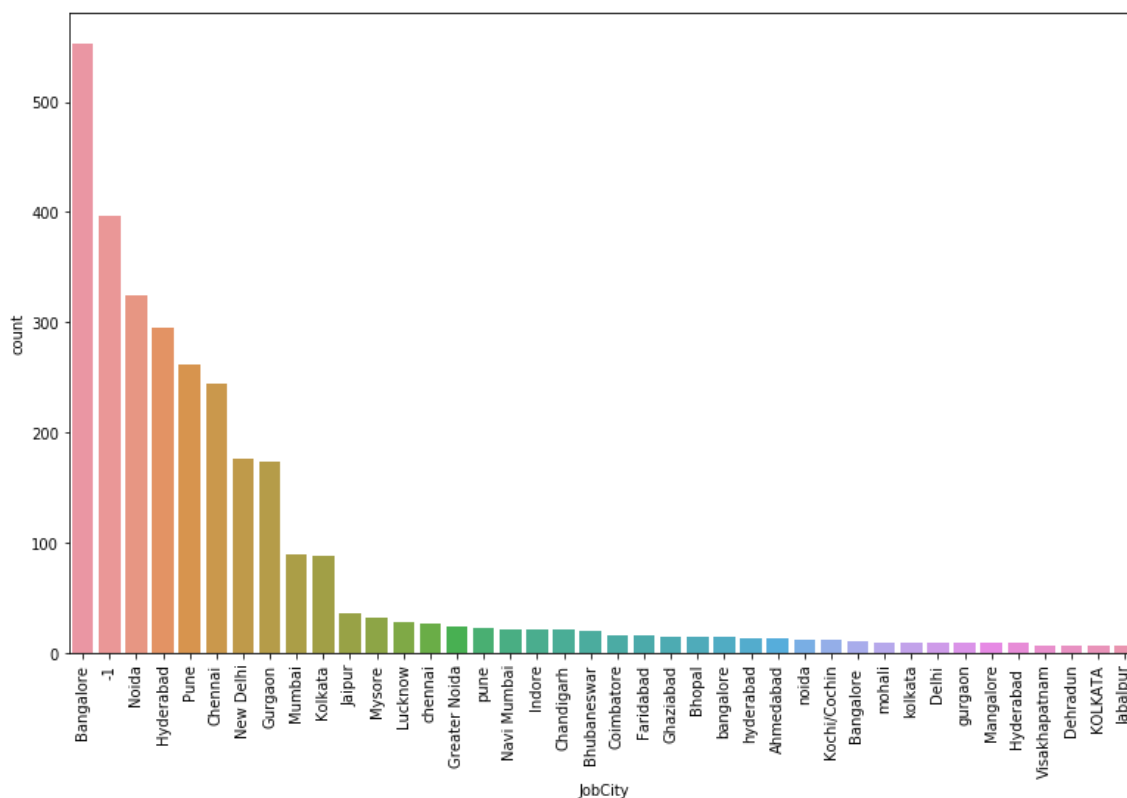
Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x17923250438>
```



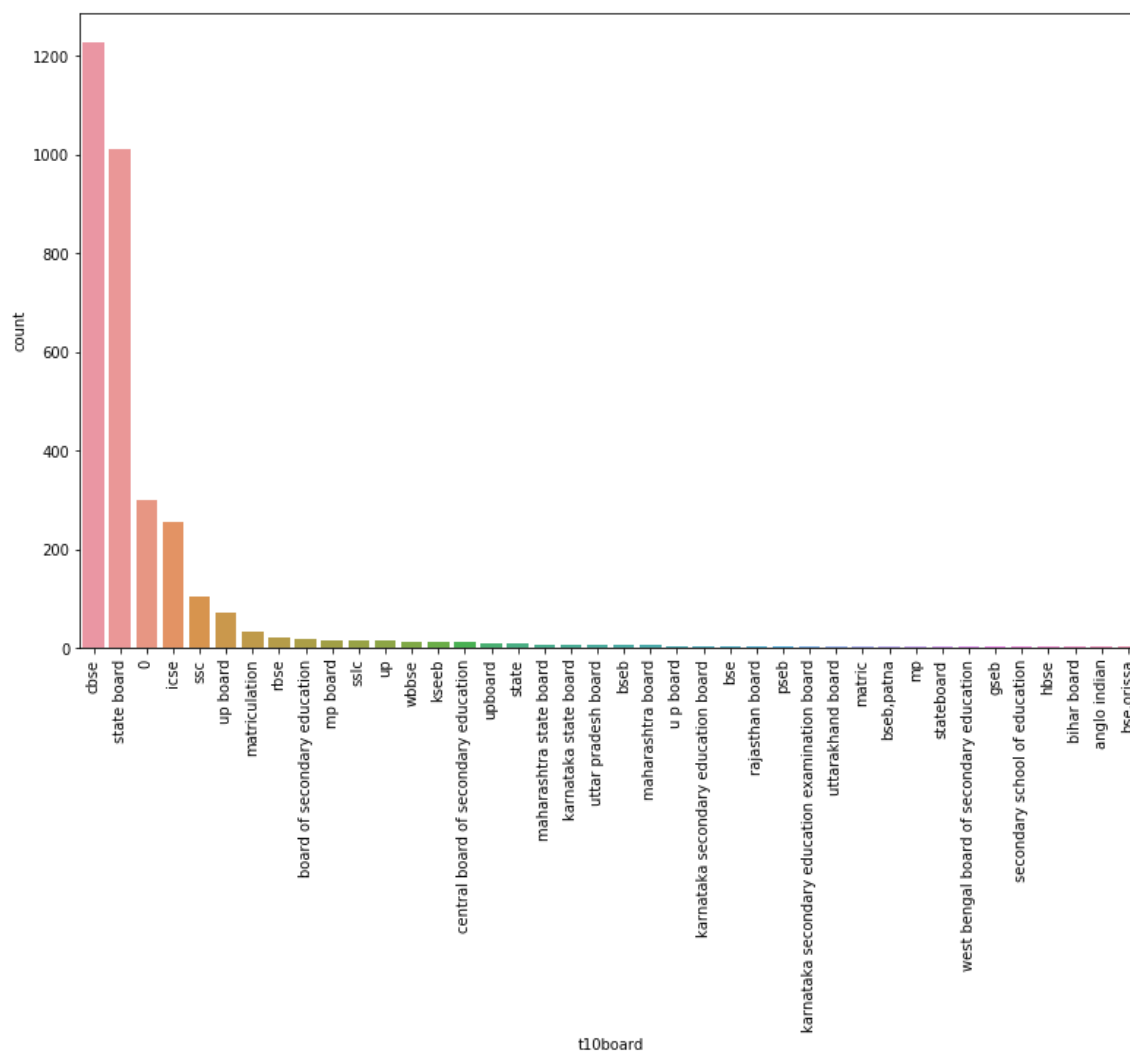
In [37]:

```
sal.describe()
plt.figure(figsize=(13, 8))
sns.countplot(sal.JobCity.dropna(),order=sal.JobCity.value_counts().iloc[:40].index);
plt.xticks(rotation=90);
```



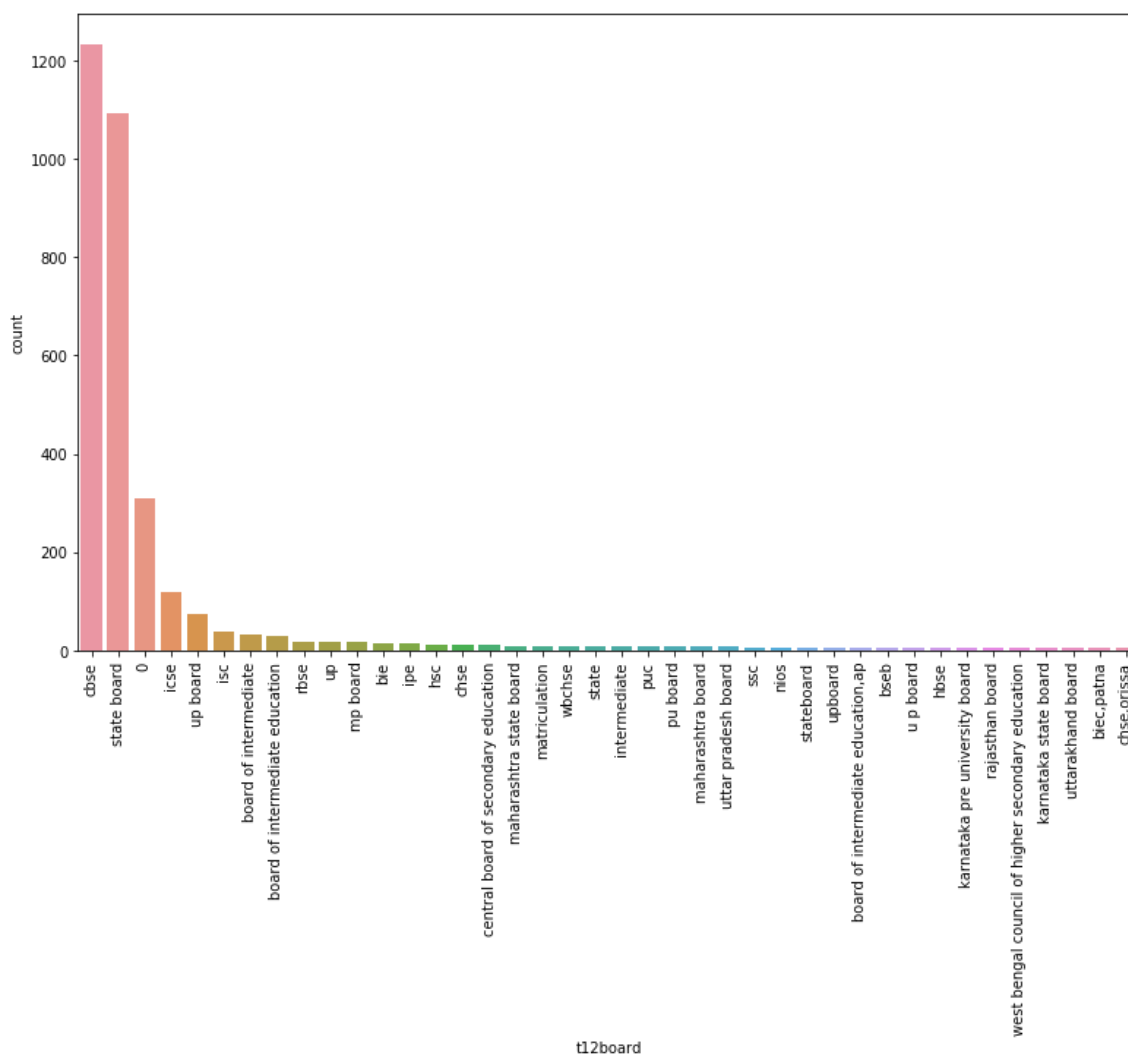
In [38]:

```
sal.describe()
plt.figure(figsize=(13, 8))
sns.countplot(sal.t10board.dropna(),order=sal.t10board.value_counts().iloc[:40].index);
plt.xticks(rotation=90);
```



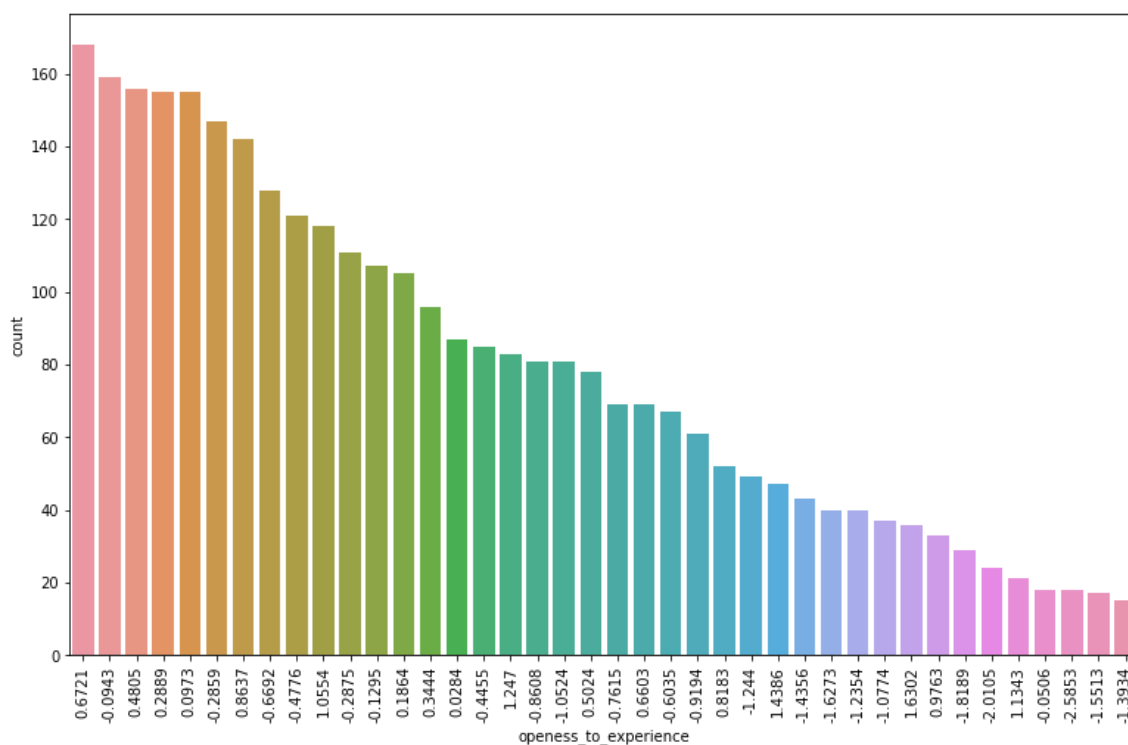
In [39]:

```
sal.describe()
plt.figure(figsize=(13, 8))
sns.countplot(sal.t12board.dropna(),order=sal.t12board.value_counts().iloc[:40].index);
plt.xticks(rotation=90);
```



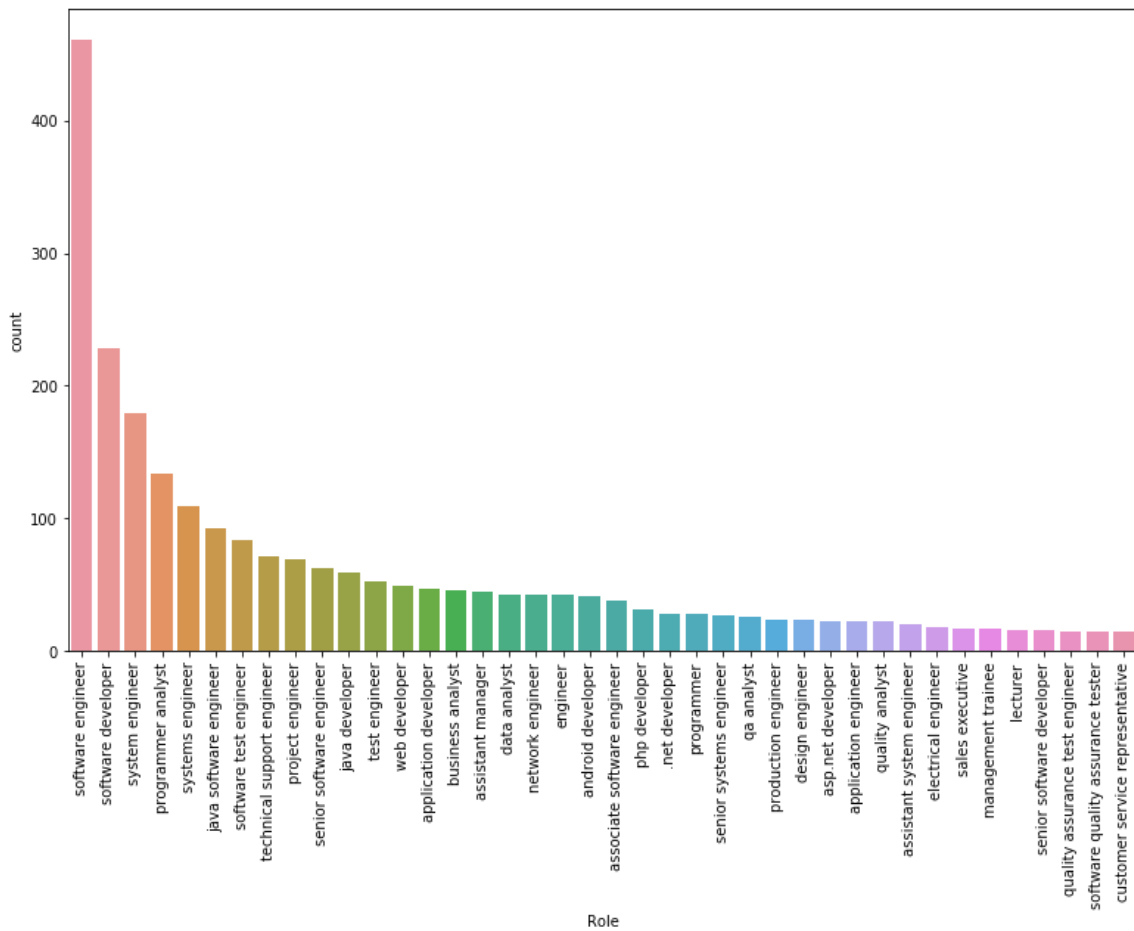
In [3]:

```
plt.figure(figsize=(13, 8))  
sns.countplot(pt.openess_to_experience.dropna(),order=pt.openess_to_experience.value_counts().iloc[:40].index);  
plt.xticks(rotation=90);
```



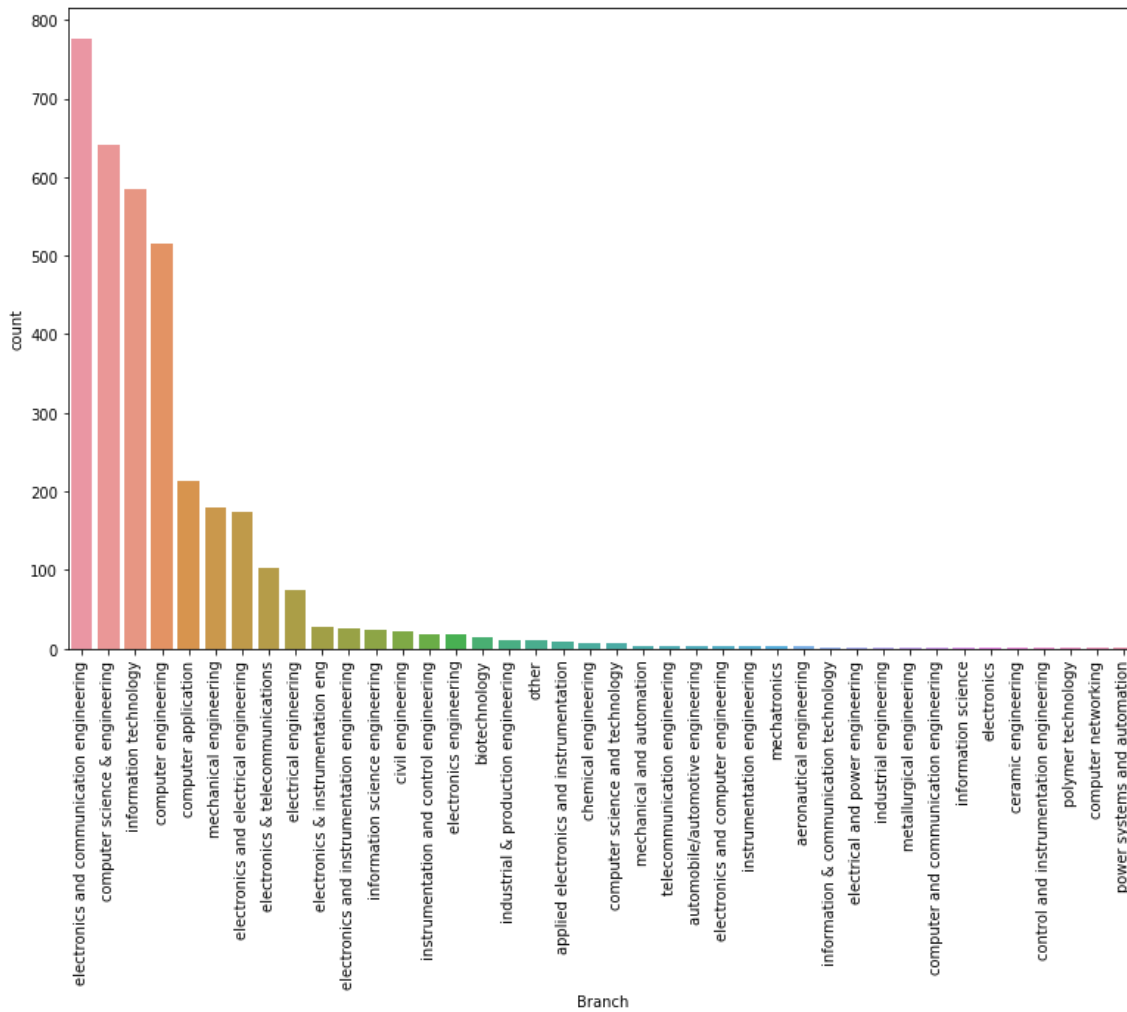
In [7]:

```
plt.figure(figsize=(13, 8))
sns.countplot(sal.Role.dropna(),order=sal.Role.value_counts().iloc[:40].index);
plt.xticks(rotation=90);
```



In [8]:

```
plt.figure(figsize=(13, 8))
sns.countplot(sal.Branch.dropna(),order=sal.Branch.value_counts().iloc[:40].index);
plt.xticks(rotation=90);
```



In []: