Loading necessary libraries and warnings

In [6]:

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

Importing the file and looking at the sample data

In [5]:

```
df = pd.read_csv(r'C:\Users\Lenovo\Downloads\bank1.csv')
df.sample(10)
```

Out[5]:

	age	job	marital	education	default	balance	housing	loan	contact	day
2044	44	management	married	tertiary	no	529	yes	no	unknown	4
3009	57	blue-collar	married	primary	no	93	yes	no	telephone	17
2459	46	housemaid	married	primary	no	0	no	no	telephone	2
3748	35	management	married	tertiary	no	2793	no	no	cellular	20
2318	53	admin.	married	secondary	no	225	yes	no	cellular	20
630	30	management	single	tertiary	no	2	yes	yes	cellular	25
3742	37	self-employed	married	tertiary	no	137	no	no	cellular	13
1689	28	management	single	tertiary	no	3238	yes	no	unknown	19
1644	33	management	divorced	tertiary	no	0	no	no	cellular	13
1891	32	management	single	tertiary	no	656	yes	yes	cellular	20

Statistical description of the dataset

In [38]:

```
df.describe()
```

Out[38]:

	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

The dataset has no null values

In [10]:

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

Out[10]:

```
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 [8]:

```
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 [11]:

```
df.columns
```

Out[11]:

```
Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housi
ng',
       'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pday
s',
       'previous', 'poutcome', 'y'],
      dtype='object')
```

The dataset has 4521 rows and 17 columns

In [13]:

```
df.shape
```

Out[13]:

```
(4521, 17)
```

In [15]:

```
df.dtypes
```

Out[15]:

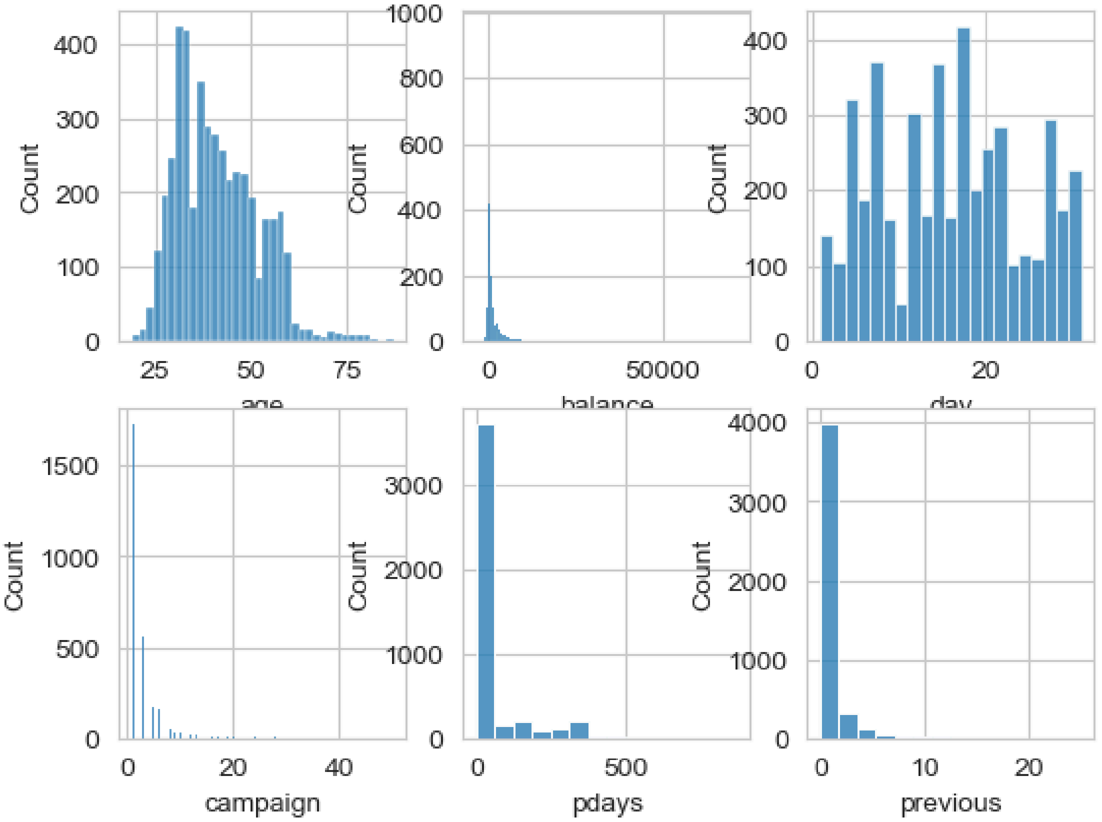
```
age              int64
job             object
marital         object
education       object
default         object
balance          int64
housing         object
loan            object
contact         object
day              int64
month           object
duration         int64
campaign         int64
pdays            int64
previous         int64
poutcome        object
y               object
dtype: object
```

In [26]:

```
fig,axes = plt.subplots(2,3)
sns.histplot(data = df, x ='age', ax = axes[0,0])
sns.histplot(data = df, x ='balance', ax = axes[0,1])
sns.histplot(data = df, x ='day', ax = axes[0,2])
sns.histplot(data = df, x ='campaign', ax = axes[1,0])
sns.histplot(data = df, x ='pdays', ax = axes[1,1])
sns.histplot(data = df, x ='previous', ax = axes[1,2])
```

Out[26]:

```
<Axes: xlabel='previous', ylabel='Count'>
```

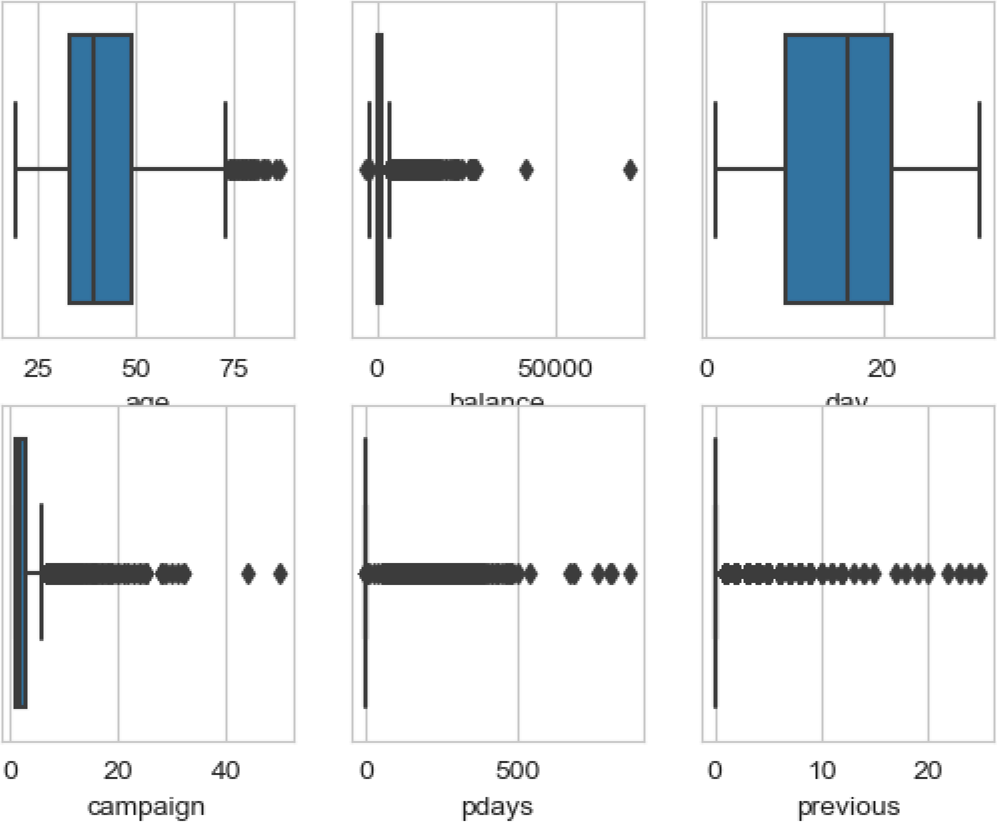


In [27]:

```
fig,axes = plt.subplots(2,3)
sns.boxplot(data = df, x ='age', ax = axes[0,0])
sns.boxplot(data = df, x ='balance', ax = axes[0,1])
sns.boxplot(data = df, x ='day', ax = axes[0,2])
sns.boxplot(data = df, x ='campaign', ax = axes[1,0])
sns.boxplot(data = df, x ='pdays', ax = axes[1,1])
sns.boxplot(data = df, x ='previous', ax = axes[1,2])
```

Out[27]:

```
<Axes: xlabel='previous'>
```



In [29]:

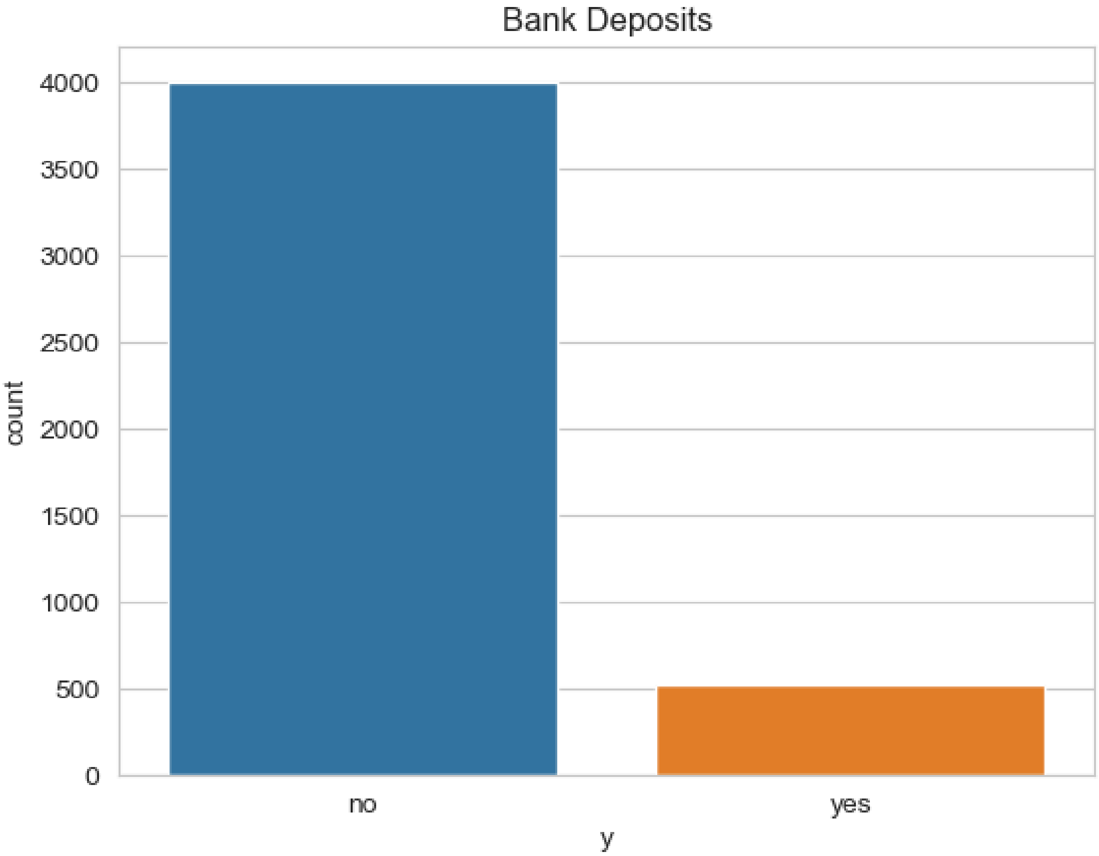
```
df.y.value_counts()
```

Out[29]:

```
y
no     4000
yes     521
Name: count, dtype: int64
```

In [31]:

```
x1 = sns.countplot(x = 'y', data = df)
plt.title('Bank Deposits')
plt.show()
```



Observations

Bank deposit products subscribed by 421 people out of 4521 people.

Bank deposit products not subscribed by 4000 people out of 4521 people.

In [33]:

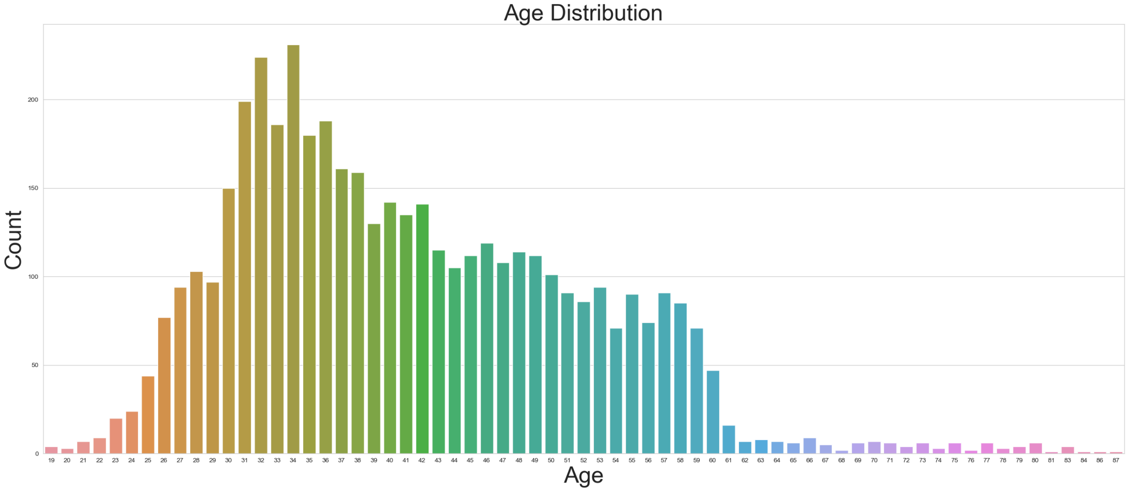
```
df.columns
```

Out[33]:

```
Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housi
ng',
       'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pday
s',
       'previous', 'poutcome', 'y'],
      dtype='object')
```

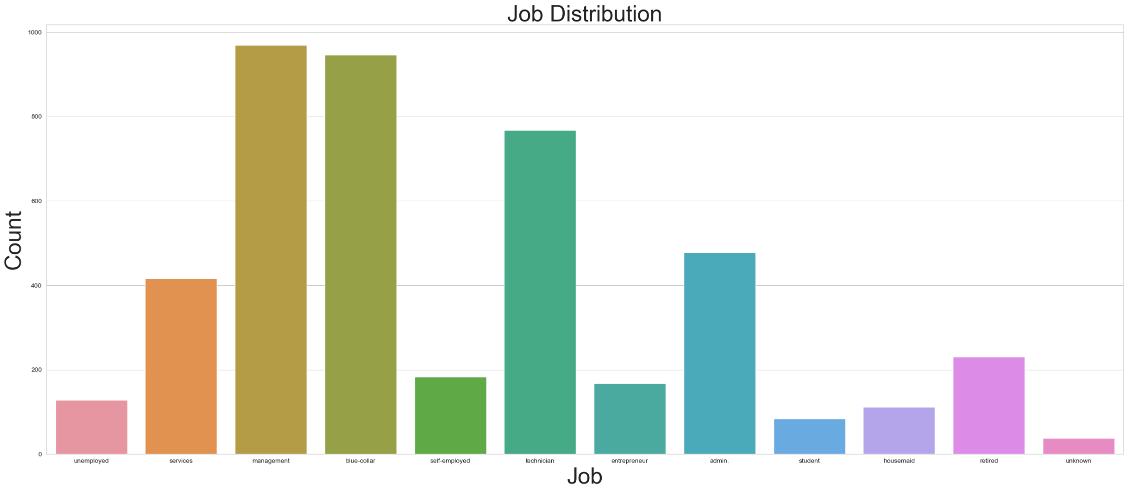
In [48]:

```
plt.figure(figsize = (30,12))
sns.countplot(x ='age', data = df)
plt.title('Age Distribution',fontsize = 35)
plt.xlabel('Age',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



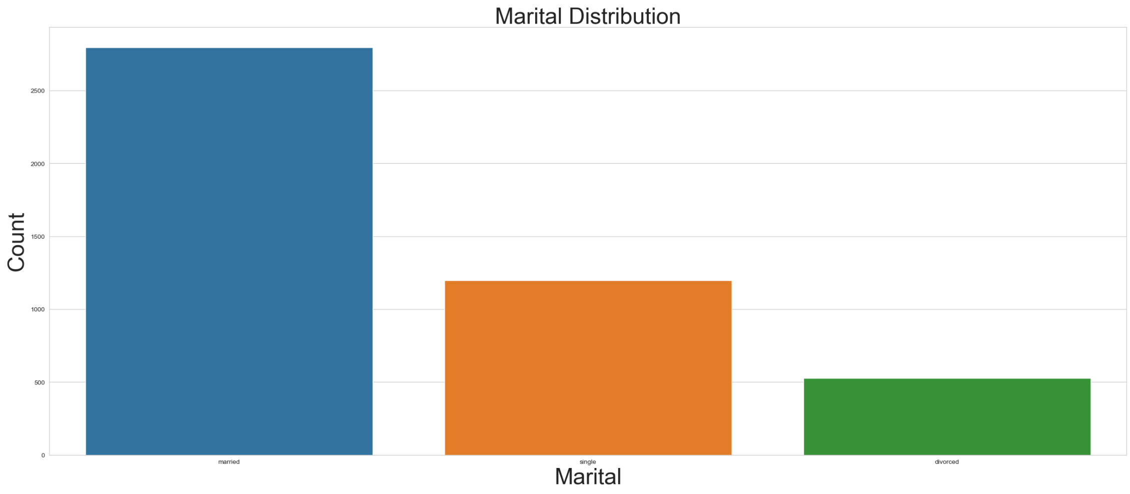
In [49]:

```
plt.figure(figsize = (30,12))
sns.countplot(x ='job', data = df)
plt.title('Job Distribution',fontsize = 35)
plt.xlabel('Job',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



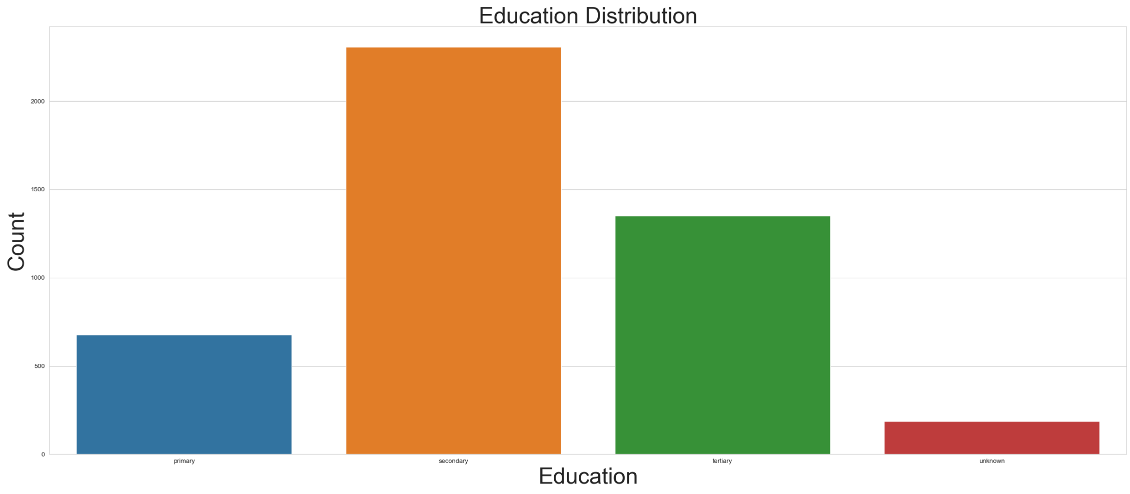
In [47]:

```
plt.figure(figsize = (30,12))
sns.countplot(x ='marital', data = df)
plt.title('Marital Distribution',fontsize = 35)
plt.xlabel('Marital',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```

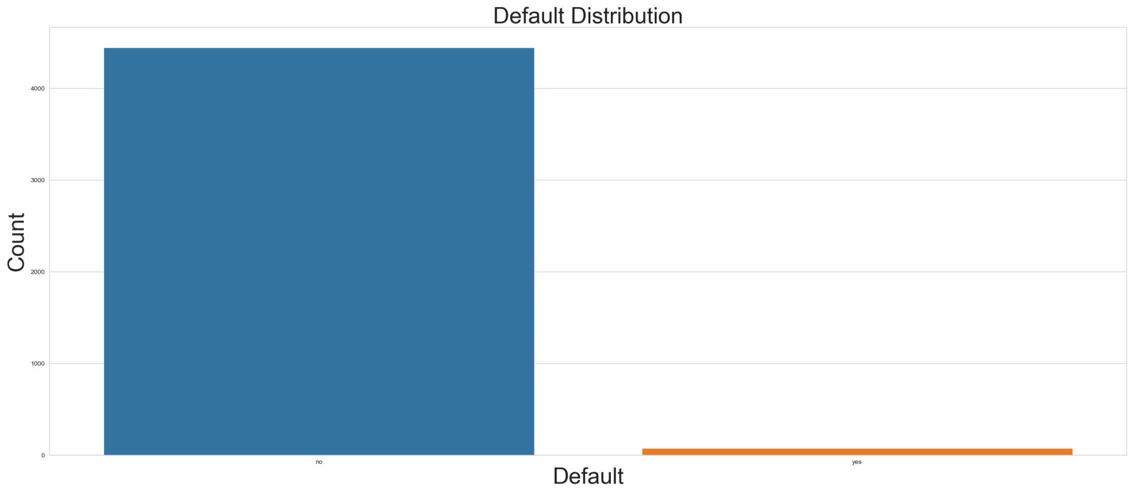


In [45]:

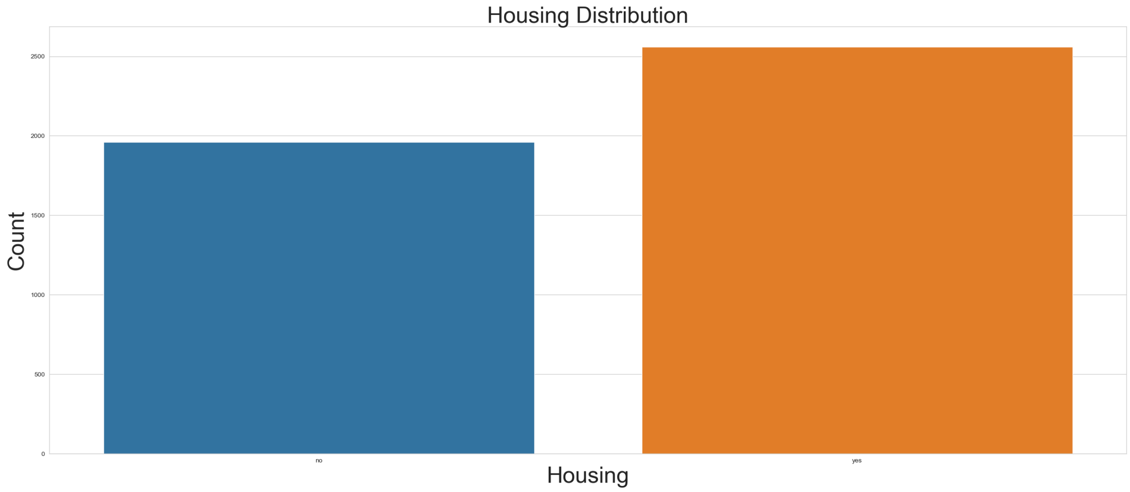
```
plt.figure(figsize = (30,12))
sns.countplot(x ='education', data = df)
plt.title('Education Distribution',fontsize = 35)
plt.xlabel('Education',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



In [50]:
```
plt.figure(figsize = (30,12))
sns.countplot(x ='default', data = df)
plt.title('Default Distribution',fontsize = 35)
plt.xlabel('Default',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



In [51]:
```
plt.figure(figsize = (30,12))
sns.countplot(x ='housing', data = df)
plt.title('Housing Distribution',fontsize = 35)
plt.xlabel('Housing',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



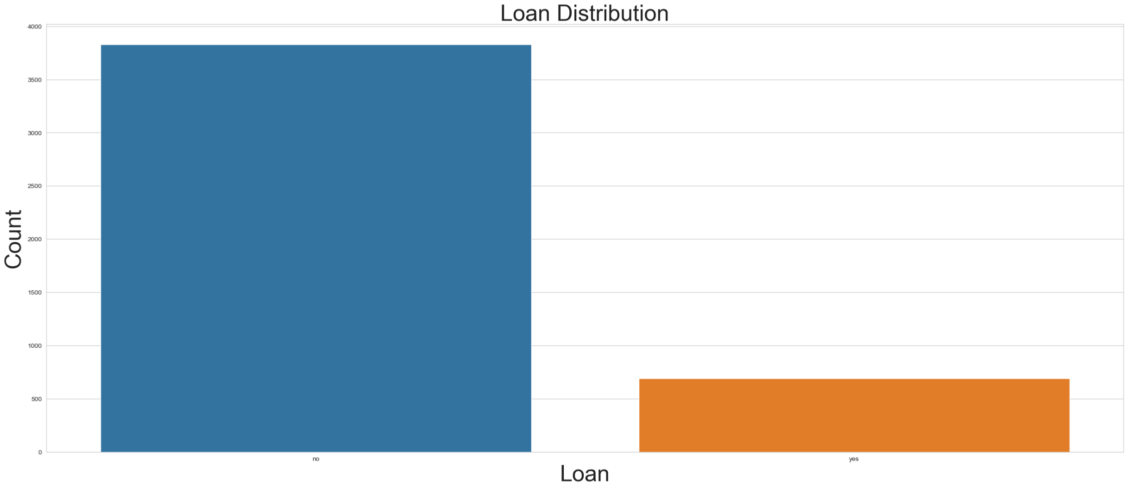
In [56]:
```
print('Housing:\n No Housing:' , df[df['housing']=='no'] ['age'].count(),
               '\n Yes Housing:' , df[df['housing']=='yes'] ['age'].count(
```

```
Housing:
 No Housing: 1962
 Yes Housing: 2559
```

The clients having housing loan is 597 more then clients who did not opted for housing loan.

In [57]:

```
plt.figure(figsize = (30,12))
sns.countplot(x ='loan', data = df)
plt.title('Loan Distribution',fontsize = 35)
plt.xlabel('Loan',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



In [59]:

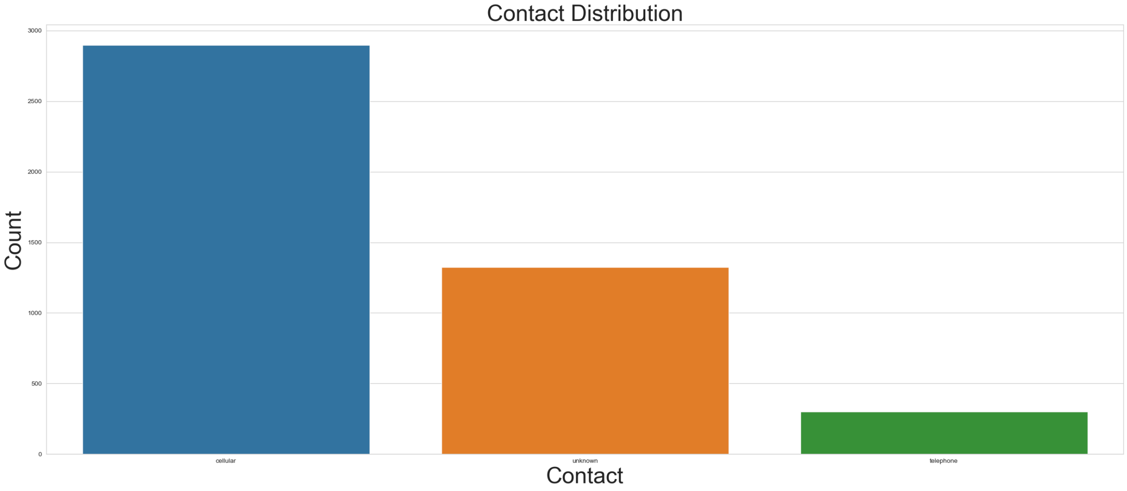
```
print('Loan:\n No Personal Loan:' , df[df['loan']=='no'] ['age'].count(),
               '\n Yes Personal Loan:' , df[df['loan']=='yes'] ['age'].cou
```

```
Loan:
 No Personal Loan: 3830
 Yes Personal Loan: 691
```

The clients having personal loan is 3139 less then clients who did not opted for personal loan.

In [60]:

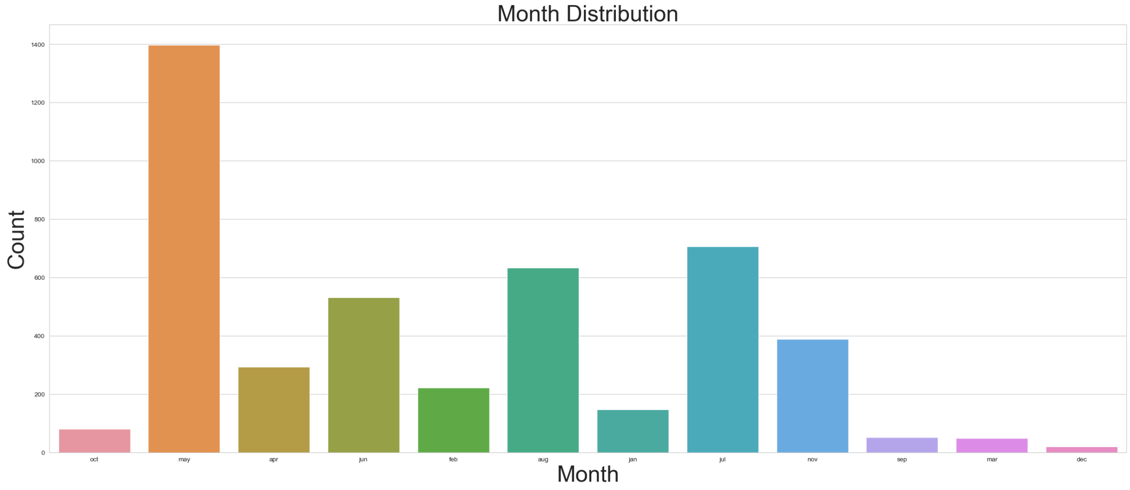
```
plt.figure(figsize = (30,12))
sns.countplot(x ='contact', data = df)
plt.title('Contact Distribution',fontsize = 35)
plt.xlabel('Contact',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



The number of customers who has mobile phones are more easily accessible for a connect than other categories.

In [61]:

```
plt.figure(figsize = (30,12))
sns.countplot(x ='month', data = df)
plt.title('Month Distribution',fontsize = 35)
plt.xlabel('Month',fontsize = 35)
plt.ylabel('Count',fontsize = 35)
plt.show()
```



Highest number of contacts to customers is in the month of May.

Conclusion

1. Number of outliers in the age column is less.
2. In customers, most of them are married.
3. Blue-collar jobs is most common in customers, Management jobs can also be noticed.
4. Most of the clients have passed secondary education.
5. Number of customers having by-default credit is less.
6. The clients having housing loan is 597 more then clients who did not opted for housing loan.
7. The clients having personal loan is 3139 less then clients who did not opted for personal loan.
8. The number of customers who has mobile phones are more easily accessible for a connect than other categories.
9. Highest number of contacts to customers is in the month of May.
10. Bank deposit products subscribed by 421 people out of 4521 people.
11. Bank deposit products not subscribed by 4000 people out of 4521 people
12. Highest number of customers of the bank is in their 30's.