In [17]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
import os
os.getcwd()
```

Out[2]:

```
'C:\\Users\\indhr'
```

In [7]:

```
data=pd.read_csv('Iris.csv.csv')
data
```

Out[7]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

In [8]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             150 non-null    int64  
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object 
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

In [9]:

```
data.isna().sum()
```

Out[9]:

```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

In [10]:

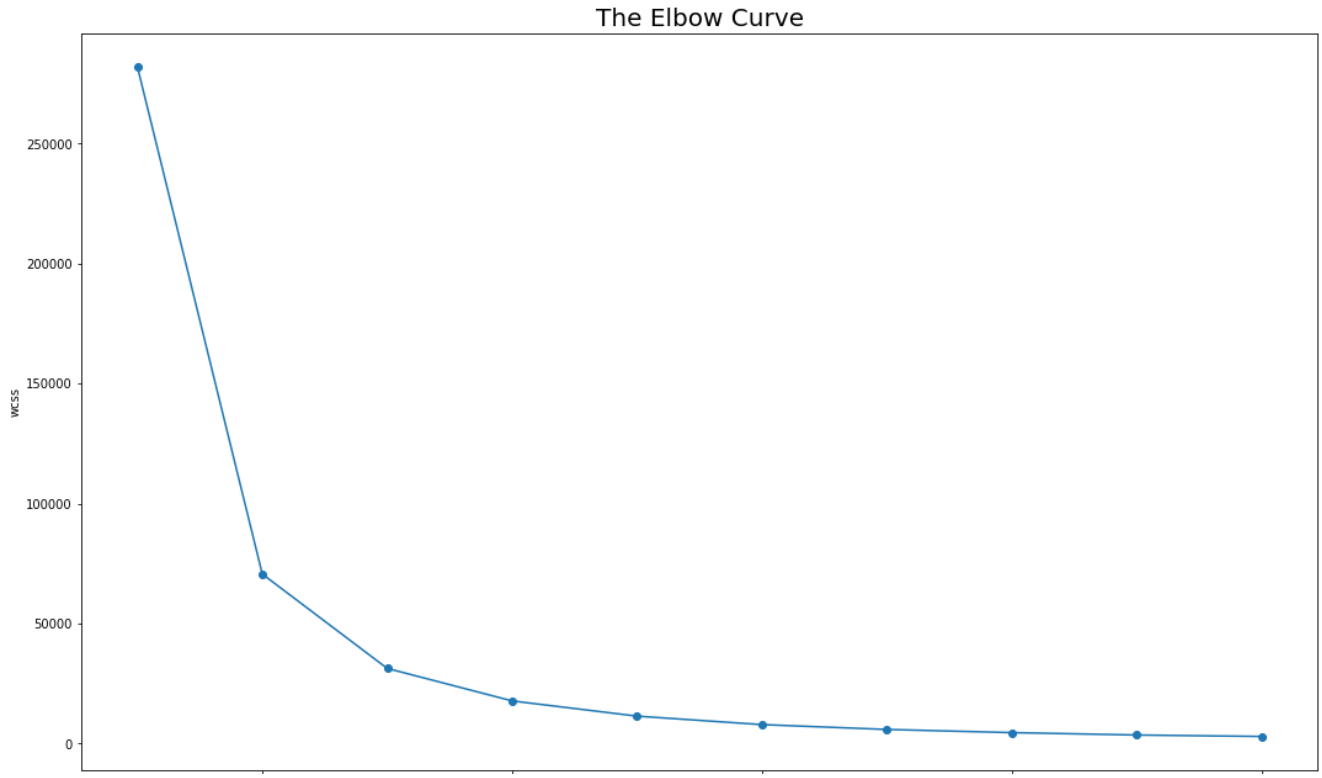
```
data.describe()
```

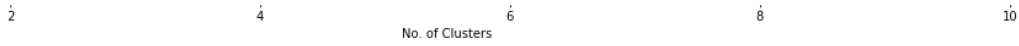
Out[10]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

In [19]:

```
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 11):
    model = KMeans(n_clusters = i,random_state = 0)
    model.fit(x)
    wcss.append(model.inertia_)

plt.figure(figsize=(18,11))
plt.plot(range(1, 11), wcss, marker = 'o')
plt.title('The Elbow Curve', fontsize = 20)
plt.xlabel('No. of Clusters')
plt.ylabel('wcss')
plt.show();
```





In [21]:

```
model = KMeans(n_clusters = 3, random_state = 0)
pred = model.fit_predict(x)
```

In [23]:

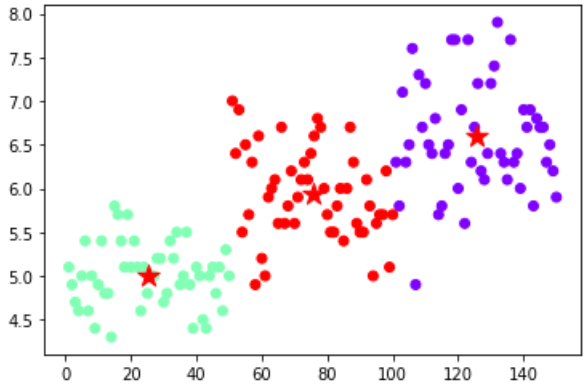
```
plt.scatter(x[:,0],x[:,1],c=model.labels_,cmap='rainbow')

plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:, 1], s = 200, c = 'red' , label =
'centeroid',
            marker='*')
```

Out[23]:

```
<matplotlib.collections.PathCollection at 0x20658a29ec8>
```

Out[23]:

```
<matplotlib.collections.PathCollection at 0x20659757488>
```



In [26]:

```
plt.scatter(x[pred == 0, 0],x[pred == 0, 1], s = 100, c = 'blue', label = 'Iris-setosa')
plt.scatter(x[pred == 1, 0], x[pred == 1, 1], s = 100, c = 'red', label = 'Iris-versicolour')
plt.scatter(x[pred == 2, 0], x[pred == 2, 1],s = 100, c = 'green', label = 'Iris-virginica')
plt.scatter(model.cluster_centers_[:, 0], model.cluster_centers_[:, 1], s = 100, c = 'yellow', labe
l = 'Centroids')
plt.legend()
plt.show()
```

Out[26]:

```
<matplotlib.collections.PathCollection at 0x206597cfe48>
```

Out[26]:

```
<matplotlib.collections.PathCollection at 0x206597da408>
```

Out[26]:

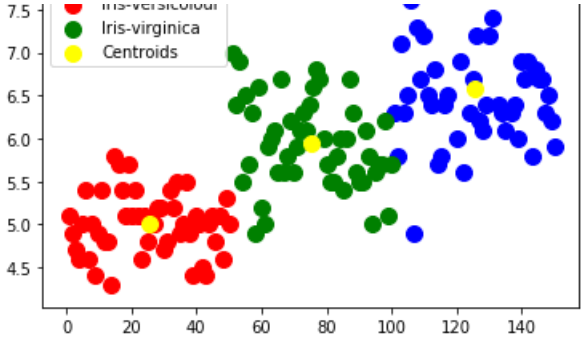
```
<matplotlib.collections.PathCollection at 0x206597d9948>
```

Out[26]:

```
<matplotlib.collections.PathCollection at 0x206597d9f48>
```

Out[26]:

```
<matplotlib.legend.Legend at 0x206597dadc8>
```


Iris-virginica
Centroids
7.5
7.0
6.5
6.0
5.5
5.0
4.5
0
20
40
60
80
100
120
140

In [ ]: