In [54]:

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

In [55]:

```
dataset = pd.read_csv("Pre_Processed-Zomato dataset.csv")
```

In [56]:

```
dataset
```

Out[56]:

	NAME	PRICE	CUSINE_CATEGORY	REGION	CUSINE TYPE	TIMING	RATIN
0	Hitchki	1200	Modern Indian,North Indian,Chinese,Momos,Birya...	Bandra	Casual Dining	12noon to 130am	
1	Baba Falooda	400	Desserts,Ice Cream,Beverages	Mahim	Dessert Parlor	2pm to 1am	V
2	Chin Chin Chu	1800	Asian,Chinese	Juhu	Casual Dining	12noon to 1am	V
3	Butterfly High	1000	Modern Indian	Bandra	Bar	12noon to 130am	V
4	BKC DIVE	1200	North Indian,Chinese,Continental	Bandra	Bar	1130am to 1am	V
...	...	...	...	...	...	...	
10542	Tirupati Balaji	500	Chinese,Fast Food,North Indian	Andheri	Casual Dining	8am to 11pm,12midnight to 115am	
10543	Hari Om Snack Bar	350	Fast Food,South Indian,Chinese	Kandivali	Quick Bites	11am to 230am	
10544	PitaBurg	400	Fast Food,Lebanese	Lower Parel	none	11am to 11pm	
10545	Uncha Otlawala	300	Desserts,Ice Cream	Kandivali	Dessert Parlor	9am to 1230AM	
10546	Mandarin Panda	400	Desserts,Chinese,Thai	Malad	none	12noon to 330pm,7pm to 1am	

10547 rows × 10 columns

In [57]:

```
X=dataset.iloc[:,[7,8]].values
```

In [58]:

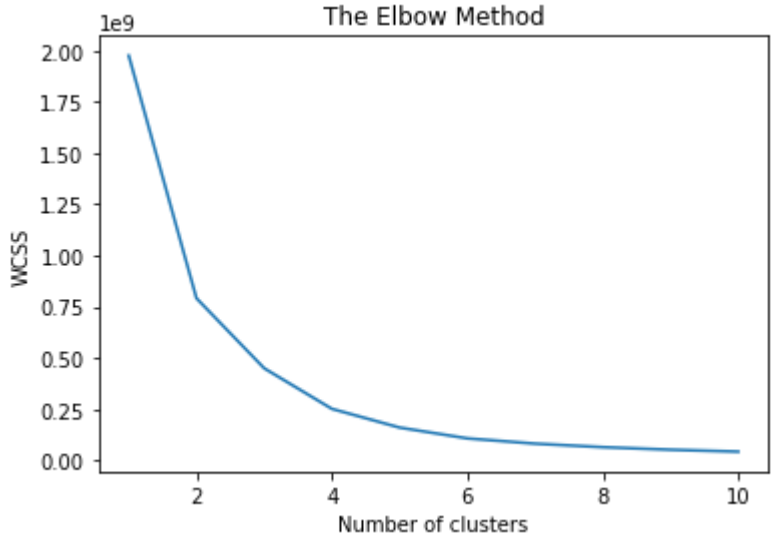
```
X
```

Out[58]:

```
array([[4.900e+00, 3.529e+03],
       [4.400e+00, 1.723e+03],
       [4.200e+00, 3.370e+02],
       ...,
       [3.400e+00, 9.900e+01],
       [3.500e+00, 2.900e+01],
       [3.700e+00, 1.210e+02]])
```

In [59]:

```
from sklearn.cluster import KMeans
list1 = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(X)
    list1.append(kmeans.inertia_)
plt.plot(range(1, 11), list1)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



In [60]:

```
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters = 4, init = 'k-means++', random_state = 42)
y_kmeans = kmeans.fit_predict(X)
```

In [61]:

```
y_kmeans
```

Out[61]:

```
array([3, 3, 2, ..., 2, 2, 2])
```

In [62]:

```
dataset["Cluster"]=y_kmeans
```

In [63]:

```
dataset
```

Out[63]:

	NAME	PRICE	CUSINE_CATEGORY	REGION	CUSINE TYPE	TIMING	RATIN
0	Hitchki	1200	Modern Indian,North Indian,Chinese,Momos,Birya...	Bandra	Casual Dining	12noon to 130am	
1	Baba Falooda	400	Desserts,Ice Cream,Beverages	Mahim	Dessert Parlor	2pm to 1am	V
2	Chin Chin Chu	1800	Asian,Chinese	Juhu	Casual Dining	12noon to 1am	V
3	Butterfly High	1000	Modern Indian	Bandra	Bar	12noon to 130am	V
4	BKC DIVE	1200	North Indian,Chinese,Continental	Bandra	Bar	1130am to 1am	V
...	...	...	...	...	...	...	
10542	Tirupati Balaji	500	Chinese,Fast Food,North Indian	Andheri	Casual Dining	8am to 11pm,12midnight to 115am	
10543	Hari Om Snack Bar	350	Fast Food,South Indian,Chinese	Kandivali	Quick Bites	11am to 230am	
10544	PitaBurg	400	Fast Food,Lebanese	Lower Parel	none	11am to 11pm	
10545	Uncha Otlawala	300	Desserts,Ice Cream	Kandivali	Dessert Parlor	9am to 1230AM	
10546	Mandarin Panda	400	Desserts,Chinese,Thai	Malad	none	12noon to 330pm,7pm to 1am	

10547 rows × 11 columns

In [64]:

```
dataset.to_csv("clustered.csv",index=False)
```

In [65]:

```
centroids=kmeans.cluster_centers_
```

In [66]:

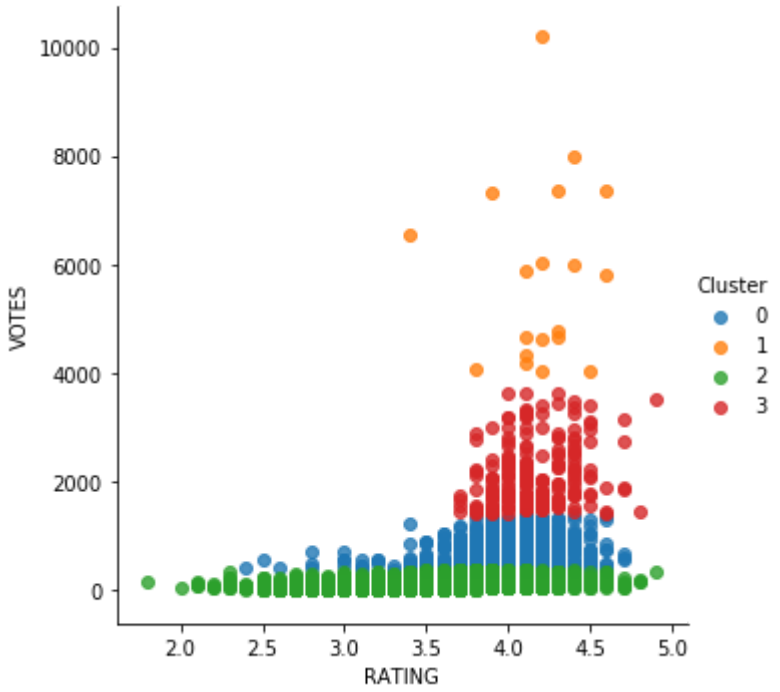
```
centroids
```

Out[66]:

```
array([[3.94019139e+00, 6.56618182e+02],
       [4.19473684e+00, 5.78152632e+03],
       [3.40510996e+00, 6.85789133e+01],
       [4.15458937e+00, 2.11337198e+03]])
```

In [67]:

```
import seaborn as sns
facet = sns.lmplot(data=dataset, x=dataset.columns[7], y=dataset.columns[8],hue=dataset.col
                   fit_reg=False, legend=True, legend_out=True)
```



From Cluster 0 - Rating varies from 2.3 to 4.7

Vote varies from 400 to 1600

From Cluster 1 - Rating varies from 3.5 to 4.5

Vote varies from 4000 to 10000

From Cluster 2 - Rating varies from 1 to 4.9

Vote varies from 1 to 400

From Cluster 3 - Rating varies from 3.7 to 4.8

Vote varies from 1600 to 3800

From the above graph analysis, its clear cluster 0 and 2 has to improve their votes and range of rating

Cluster 1 and 3 are in good range and better votes so no problem

In [79]:

```
cluster0=dataset[dataset["Cluster"]==0]
```

In [80]:

```
cluster0
```

Out[80]:

	NAME	PRICE	CUSINE_CATEGORY	REGION	CUSINE TYPE	TIMING	RA
3	Butterfly High	1000	Modern Indian	Bandra	Bar	12noon to 130am	
16	Ustaadi	1200	Asian,Biryani,Chinese,North Indian,Mughlai,Mid...	Mumbai Central	Casual Dining	12noon to 4pm,7pm to 4am	
17	Cafe Maaz	350	Chinese,North Indian,Mughlai,Biryani,Seafood	Bhandup	Quick Bites	8am to 1230AM	
18	Carter's Blue	900	North Indian,Lebanese,Fast Food,Chinese	Malad	Casual Dining	12noon to 1230AM	
19	Rajasthan	700	North Indian,Mughlai,Chinese,Fast Food	Khar	Casual Dining	12noon to 1am	
...	...	...	...	...	...	...	
10529	Gupta Sandwiches & Snacks	250	Fast Food,Juices,Street Food	Vashi	Quick Bites	10am to 11pm	
10531	Bismillah Catering	550	North Indian,Chinese,Mughlai	Airoli	Casual Dining	11am to 11pm	
10532	Hotel Deluxe	300	North Indian,Chinese,Kerala	Fort	Quick Bites	11am to 11pm	
10533	Square Pizza	600	Pizza,Fast Food	Marol	Quick Bites	11am to 11pm	
10536	Umame	3000	Japanese	Churchgate	Fine Dining	1230pm to 330pm,730pm to 11pm	

1045 rows × 11 columns

In [81]:

```
cluster2=dataset[dataset["Cluster"]==2]
cluster2
```

Out[81]:

NAME	PRICE	CUSINE_CATEGORY	REGION	CUSINE TYPE	TIMING	RATING_
hin Chin Chu	1800	Asian,Chinese	Juhu	Casual Dining	12noon to 1am	Very
ew Ajwa Family staurant	500	North Indian,Chinese,Mughlai,Kebab,Desserts	Marol	Casual Dining	10am to 4am	Very
Angrezi atiyalaa	1200	North Indian,Finger Food,American,Mexican,Chinese	Andheri	Casual Dining	12noon to 1230AM	Ex
readkart	400	Fast Food	Borivali	Quick Bites	4pm to 230am	
The Lemon Grass	300	North Indian,Chinese	Vasai	Quick Bites	12noon to 330pm,430pm to 2am	A
...	...	...	...	...	...	
Tirupati Balaji	500	Chinese,Fast Food,North Indian	Andheri	Casual Dining	8am to 11pm,12midnight to 115am	
Hari Om ack Bar	350	Fast Food,South Indian,Chinese	Kandivali	Quick Bites	11am to 230am	
itaBurg	400	Fast Food,Lebanese	Lower Parel	none	11am to 11pm	A
Uncha tlawala	300	Desserts,Ice Cream	Kandivali	Dessert Parlor	9am to 1230AM	
Mandarin Panda	400	Desserts,Chinese,Thai	Malad	none	12noon to 330pm,7pm to 1am	

× 11 columns

In [ ]:

In [ ]: