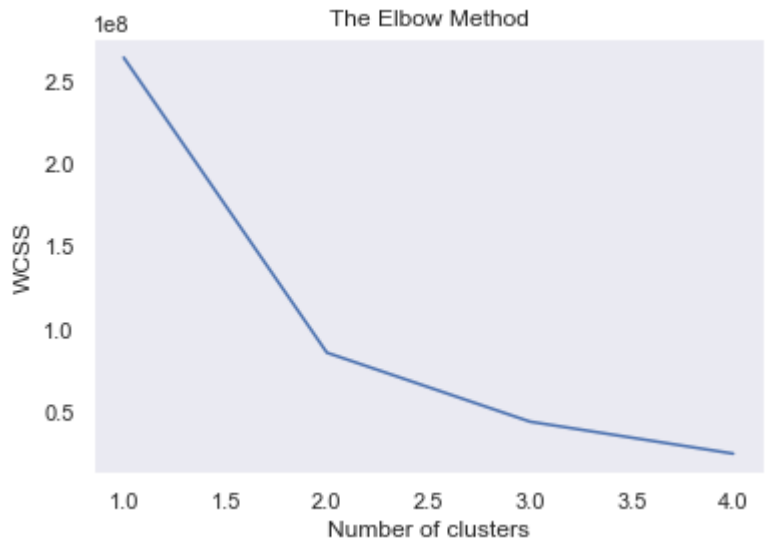
In [95]:
```
import sys
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(style="dark",color_codes=True)
import sklearn
from sklearn.cluster import KMeans
```

In [ ]:
```
# Before clustering, let's find out the number of clusters that we need to app
ly
```

In [96]:
```
# Finding optimum number of clusters using The Elbow Method

from sklearn.cluster import KMeans
clust = dataa.iloc[:, 35:37].values
wcss = []

for i in range(1,5):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init
= 10, random_state = 0)
    kmeans.fit(clust)
    wcss.append(kmeans.inertia_)
plt.plot(range(1,5),wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



In [97]:
```
#Clustering -
dataa = pd.read_csv("final.csv")
dataa = pd.DataFrame(dataa)

#print(dataa.head())

Data = pd.DataFrame(dataa,columns = ['Rent'])

import sklearn
from sklearn.cluster import KMeans

#print(Data)

# 3 clusters
kmeans = KMeans(n_clusters=3).fit(Data)
centroids = kmeans.cluster_centers_
print(centroids)
centroids = pd.DataFrame(centroids)
X= Data['Rent']


cluster_map = pd.DataFrame()
cluster_map['data_index'] = Data.index.values
cluster_map['cluster'] = kmeans.labels_

print(cluster_map[cluster_map.cluster == 0])
print(cluster_map[cluster_map.cluster == 1])
print(cluster_map[cluster_map.cluster == 2])

#plt.scatter(Data.index, X, c= kmeans.labels_.astype(float), s=50, alpha=0.5)
#plt.scatter(centroids.index,centroids, c='red', s=50)

plt.scatter(Data.index, X, c= kmeans.labels_.astype(float), s=50, alpha = 0.8)
plt.scatter(centroids.index,centroids, c='blue', s=50)
```

```
[[1209.26966292]
 [2237.66666667]
 [4105.40540541]]
     data_index  cluster
4             4        0
10           10        0
15           15        0
17           17        0
20           20        0
25           25        0
27           27        0
28           28        0
30           30        0
33           33        0
34           34        0
35           35        0
37           37        0
38           38        0
39           39        0
41           41        0
42           42        0
43           43        0
45           45        0
46           46        0
47           47        0
53           53        0
65           65        0
67           67        0
68           68        0
70           70        0
72           72        0
73           73        0
76           76        0
78           78        0
..          ...      ...
148         148        0
150         150        0
152         152        0
155         155        0
157         157        0
159         159        0
160         160        0
161         161        0
164         164        0
165         165        0
166         166        0
168         168        0
170         170        0
173         173        0
176         176        0
177         177        0
178         178        0
179         179        0
180         180        0
181         181        0
183         183        0
189         189        0
```

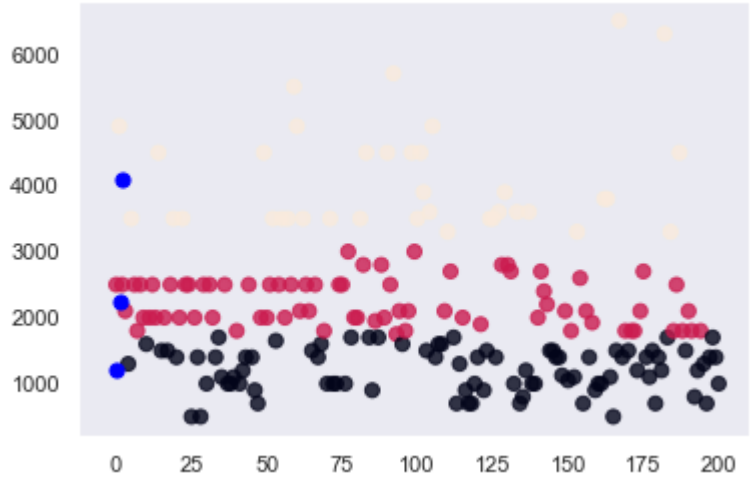
```
192         192        0
193         193        0
195         195        0
196         196        0
197         197        0
198         198        0
199         199        0
200         200        0

[89 rows x 2 columns]
     data_index  cluster
0             0        1
2             2        1
3             3        1
6             6        1
7             7        1
8             8        1
9             9        1
11           11        1
12           12        1
13           13        1
16           16        1
18           18        1
21           21        1
23           23        1
24           24        1
26           26        1
29           29        1
31           31        1
32           32        1
36           36        1
40           40        1
44           44        1
48           48        1
50           50        1
51           51        1
54           54        1
56           56        1
58           58        1
61           61        1
63           63        1
..          ...      ...
96           96        1
97           97        1
99           99        1
109         109        1
111         111        1
115         115        1
121         121        1
128         128        1
130         130        1
131         131        1
140         140        1
141         141        1
142         142        1
143         143        1
149         149        1
```

```
151         151        1
154         154        1
156         156        1
158         158        1
169         169        1
171         171        1
172         172        1
174         174        1
175         175        1
185         185        1
186         186        1
188         188        1
190         190        1
191         191        1
194         194        1

[75 rows x 2 columns]
     data_index  cluster
1             1        2
5             5        2
14           14        2
19           19        2
22           22        2
49           49        2
52           52        2
55           55        2
57           57        2
59           59        2
60           60        2
62           62        2
71           71        2
81           81        2
83           83        2
90           90        2
92           92        2
98           98        2
100         100        2
101         101        2
102         102        2
104         104        2
105         105        2
110         110        2
124         124        2
125         125        2
127         127        2
129         129        2
133         133        2
137         137        2
153         153        2
162         162        2
163         163        2
167         167        2
182         182        2
184         184        2
187         187        2
```

Out[97]: <matplotlib.collections.PathCollection at 0x207b397d898>


6000
5000
4000
3000
2000
1000
0
25
50
75
100
125
150
175
200

In [98]:
```
#Creating Dummy Variables for categorical features

l = ["EnrolmentTerm","BuildingType","RoomType","Furnished","SharedorPrivate",
"HouseAge","LocationWard"]
one_hot = pd.get_dummies(dataa[l])
dataa = dataa.drop(l,axis = 1)
dataa = one_hot.join(dataa)
print(dataa.head())
print(dataa.columns)
```

```
   EnrolmentTerm_Fall  EnrolmentTerm_Spring  EnrolmentTerm_Winter  \
0                   1                     0                     0
1                   1                     0                     0
2                   1                     0                     0
3                   1                     0                     0
4                   0                     0                     1

   BuildingType_Apartment  BuildingType_House  RoomType_Basement  \
0                       1                   0                  0
1                       1                   0                  0
2                       0                   1                  0
3                       0                   1                  0
4                       1                   0                  0

   RoomType_Master Room  RoomType_Other  RoomType_Regular Room  Furnished_No
\
0                     0               0                      1             0
1                     0               0                      1             0
2                     0               0                      1             0
3                     0               0                      1             0
4                     0               0                      1             1

   ...  AirConditioning  Laundry  Parking  Terrace   Security  Bedrooms  \
0  ...                1        1        1         0         0         5
1  ...                0        0        0         1         0         7
2  ...                1        1        1         1         0         5
3  ...                1        1        0         0         0         3
4  ...                0        0        0         1         0         1

   Bathrooms  MarketCommute  BusStopCommute  Rent
0          2             25               3  2500
1          3             25              18  4900
2          2             15               3  2500
3          1             25               8  2100
4          1             15               3  1300

[5 rows x 37 columns]
Index(['EnrolmentTerm_Fall', 'EnrolmentTerm_Spring', 'EnrolmentTerm_Winter',
       'BuildingType_Apartment', 'BuildingType_House', 'RoomType_Basement',
       'RoomType_Master Room', 'RoomType_Other', 'RoomType_Regular Room',
       'Furnished_No', 'Furnished_Yes', 'SharedorPrivate_Private',
       'SharedorPrivate_Shared', 'HouseAge_Middle ', 'HouseAge_New',
       'HouseAge_Old', 'LocationWard_1- Southwest ward',
       'LocationWard_2- Northwest ward', 'LocationWard_3- Lakeshore ward',
       'LocationWard_4- Northeast ward', 'LocationWard_5- Southeast ward',
       'LocationWard_6- Central-Columbia ward', 'LocationWard_7- Uptown war
d',
       'LocationWard_Central-Columbia ward', 'LocationWard_Southeast ward',
       'Internet ', 'Hydro', 'AirConditioning', 'Laundry', 'Parking',
       'Terrace ', 'Security', 'Bedrooms', 'Bathrooms', 'MarketCommute',
       'BusStopCommute', 'Rent'],
      dtype='object')
```

In [99]:
```
dataa.head()
```

Out[99]:

	EnrolmentTerm_Fall	EnrolmentTerm_Spring	EnrolmentTerm_Winter	BuildingType_Apartment	B
0	1	0	0	1	
1	1	0	0	1	
2	1	0	0	0	
3	1	0	0	0	
4	0	0	1	1	

5 rows × 37 columns

In [100]:
```
new_data = dataa.iloc[:, 0:32]
new_data.head()
```

Out[100]:

	EnrolmentTerm_Fall	EnrolmentTerm_Spring	EnrolmentTerm_Winter	BuildingType_Apartment	B
0	1	0	0	1	
1	1	0	0	1	
2	1	0	0	0	
3	1	0	0	0	
4	0	0	1	1	

5 rows × 32 columns

In [46]:
```
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
```

In [101]:
```
fme = apriori(new_data, min_support=0.1, use_colnames=True)
fme
```

Out[101]:

	support	itemsets
0	0.512438	(EnrolmentTerm_Fall)
1	0.283582	(EnrolmentTerm_Spring)
2	0.203980	(EnrolmentTerm_Winter)
3	0.696517	(BuildingType_Apartment)
4	0.303483	(BuildingType_House)
5	0.213930	(RoomType_Master Room)
6	0.671642	(RoomType_Regular Room)
7	0.184080	(Furnished_No)
8	0.815920	(Furnished_Yes)
9	0.273632	(SharedorPrivate_Private)
10	0.726368	(SharedorPrivate_Shared)
11	0.412935	(HouseAge_Middle)
12	0.343284	(HouseAge_New)
13	0.243781	(HouseAge_Old)
14	0.139303	(LocationWard_3- Lakeshore ward)
15	0.139303	(LocationWard_4- Northeast ward)
16	0.328358	(LocationWard_6- Central-Columbia ward)
17	0.109453	(LocationWard_7- Uptown ward)
18	0.776119	(Internet)
19	0.910448	(Hydro)
20	0.815920	(AirConditioning)
21	0.905473	(Laundry)
22	0.412935	(Parking)
23	0.298507	(Terrace)
24	0.134328	(Security)
25	0.338308	(EnrolmentTerm_Fall, BuildingType_Apartment)
26	0.174129	(EnrolmentTerm_Fall, BuildingType_House)
27	0.368159	(RoomType_Regular Room, EnrolmentTerm_Fall)
28	0.412935	(EnrolmentTerm_Fall, Furnished_Yes)
29	0.144279	(EnrolmentTerm_Fall, SharedorPrivate_Private)
...	...	...
3501	0.114428	(Internet , EnrolmentTerm_Fall, SharedorPrivat...
3502	0.109453	(BuildingType_Apartment, SharedorPrivate_Share...
3503	0.109453	(Internet , BuildingType_Apartment, SharedorPr...
3504	0.119403	(Internet , BuildingType_Apartment, SharedorPr...

	support	itemsets
3505	0.119403	(Internet , BuildingType_Apartment, SharedorPr...
3506	0.124378	(BuildingType_Apartment, SharedorPrivate_Share...
3507	0.114428	(BuildingType_Apartment, SharedorPrivate_Share...
3508	0.194030	(Internet , BuildingType_Apartment, SharedorPr...
3509	0.134328	(Internet , BuildingType_Apartment, Hydro, Roo...
3510	0.104478	(Internet , BuildingType_Apartment, Hydro, Loc...
3511	0.119403	(Internet , BuildingType_Apartment, SharedorPr...
3512	0.109453	(Internet , BuildingType_Apartment, SharedorPr...
3513	0.164179	(Internet , BuildingType_Apartment, SharedorPr...
3514	0.154229	(Internet , BuildingType_Apartment, SharedorPr...
3515	0.139303	(Internet , Parking, BuildingType_Apartment, S...
3516	0.109453	(Internet , BuildingType_Apartment, SharedorPr...
3517	0.114428	(Internet , BuildingType_Apartment, Hydro, Loc...
3518	0.129353	(Internet , Parking, BuildingType_Apartment, H...
3519	0.109453	(Internet , BuildingType_Apartment, Hydro, Ter...
3520	0.119403	(Internet , Parking, BuildingType_Apartment, H...
3521	0.114428	(Internet , Parking, BuildingType_Apartment, H...
3522	0.104478	(Internet , Parking, Hydro, Terrace , AirCondi...
3523	0.104478	(Internet , SharedorPrivate_Shared, Hydro, Hou...
3524	0.124378	(Internet , SharedorPrivate_Shared, Hydro, Roo...
3525	0.114428	(Internet , SharedorPrivate_Shared, Hydro, Loc...
3526	0.104478	(Internet , Parking, SharedorPrivate_Shared, H...
3527	0.109453	(Internet , Parking, SharedorPrivate_Shared, H...
3528	0.124378	(Internet , Parking, SharedorPrivate_Shared, H...
3529	0.109453	(Internet , Parking, Hydro, Terrace , AirCondi...
3530	0.109453	(Internet , BuildingType_Apartment, SharedorPr...

3531 rows × 2 columns

In [102]:
```
from mlxtend.frequent_patterns import association_rules

association_rules(fme, metric="confidence", min_threshold=0.7)

#http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/
```

Out[102]:

	antecedents	consequents	antecedent support	consequent support	support	confi
0	(EnrolmentTerm_Fall)	(RoomType_Regular Room)	0.512438	0.671642	0.368159	0.7
1	(EnrolmentTerm_Fall)	(Furnished_Yes)	0.512438	0.815920	0.412935	0.8
2	(EnrolmentTerm_Fall)	(SharedorPrivate_Shared)	0.512438	0.726368	0.368159	0.7
3	(EnrolmentTerm_Fall)	(Internet)	0.512438	0.776119	0.388060	0.7
4	(EnrolmentTerm_Fall)	(Hydro)	0.512438	0.910448	0.447761	0.8
5	(EnrolmentTerm_Fall)	(AirConditioning)	0.512438	0.815920	0.412935	0.8
6	(EnrolmentTerm_Fall)	(Laundry)	0.512438	0.905473	0.452736	0.8
7	(EnrolmentTerm_Spring)	(BuildingType_Apartment)	0.283582	0.696517	0.223881	0.7
8	(EnrolmentTerm_Spring)	(Furnished_Yes)	0.283582	0.815920	0.233831	0.8
9	(EnrolmentTerm_Spring)	(SharedorPrivate_Shared)	0.283582	0.726368	0.228856	0.8
10	(EnrolmentTerm_Spring)	(Internet)	0.283582	0.776119	0.233831	0.8
11	(EnrolmentTerm_Spring)	(Hydro)	0.283582	0.910448	0.278607	0.9
12	(EnrolmentTerm_Spring)	(AirConditioning)	0.283582	0.815920	0.243781	0.8
13	(EnrolmentTerm_Spring)	(Laundry)	0.283582	0.905473	0.273632	0.9
14	(EnrolmentTerm_Winter)	(Furnished_Yes)	0.203980	0.815920	0.169154	0.8
15	(EnrolmentTerm_Winter)	(Internet)	0.203980	0.776119	0.154229	0.7
16	(EnrolmentTerm_Winter)	(Hydro)	0.203980	0.910448	0.184080	0.9
17	(EnrolmentTerm_Winter)	(AirConditioning)	0.203980	0.815920	0.159204	0.7
18	(EnrolmentTerm_Winter)	(Laundry)	0.203980	0.905473	0.179104	0.8
19	(RoomType_Regular Room)	(BuildingType_Apartment)	0.671642	0.696517	0.507463	0.7
20	(BuildingType_Apartment)	(RoomType_Regular Room)	0.696517	0.671642	0.507463	0.7
21	(Furnished_Yes)	(BuildingType_Apartment)	0.815920	0.696517	0.572139	0.7
22	(BuildingType_Apartment)	(Furnished_Yes)	0.696517	0.815920	0.572139	0.8
23	(SharedorPrivate_Shared)	(BuildingType_Apartment)	0.726368	0.696517	0.537313	0.7
24	(BuildingType_Apartment)	(SharedorPrivate_Shared)	0.696517	0.726368	0.537313	0.7
25	(HouseAge_New)	(BuildingType_Apartment)	0.343284	0.696517	0.303483	0.8
26	(LocationWard_4- Northeast ward)	(BuildingType_Apartment)	0.139303	0.696517	0.109453	0.7
27	(LocationWard_6- Central-Columbia ward)	(BuildingType_Apartment)	0.328358	0.696517	0.248756	0.7
28	(Internet)	(BuildingType_Apartment)	0.776119	0.696517	0.547264	0.7
29	(BuildingType_Apartment)	(Internet)	0.696517	0.776119	0.547264	0.7
...	...	...	...	...	...	

	antecedents	consequents	antecedent support	consequent support	support	confi
27527	(Internet , BuildingType_Apartment, RoomType_R...	(SharedorPrivate_Shared, Hydro, AirConditioning)	0.154229	0.582090	0.109453	0.7
27528	(Internet , SharedorPrivate_Shared, Hydro, Roo...	(Laundry, Furnished_Yes, BuildingType_Apartment)	0.134328	0.517413	0.109453	0.8
27529	(Internet , SharedorPrivate_Shared, Hydro, Roo...	(Laundry, AirConditioning, BuildingType_Apartm...	0.134328	0.547264	0.109453	0.8
27530	(Internet , SharedorPrivate_Shared, Hydro, Roo...	(AirConditioning, Furnished_Yes, BuildingType_...	0.144279	0.487562	0.109453	0.7
27531	(Internet , SharedorPrivate_Shared, RoomType_R...	(Laundry, Hydro, BuildingType_Apartment)	0.134328	0.577114	0.109453	0.8
27532	(Internet , SharedorPrivate_Shared, RoomType_R...	(Hydro, Furnished_Yes, BuildingType_Apartment)	0.144279	0.532338	0.109453	0.7
27533	(Internet , SharedorPrivate_Shared, RoomType_R...	(AirConditioning, Hydro, BuildingType_Apartment)	0.144279	0.527363	0.109453	0.7
27534	(Internet , Hydro, RoomType_Regular Room, AirC...	(SharedorPrivate_Shared, Laundry, BuildingType...	0.154229	0.497512	0.109453	0.7
27535	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , Furnished_Yes, Laundry)	0.134328	0.621891	0.109453	0.8
27536	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , AirConditioning, Laundry)	0.144279	0.641791	0.109453	0.7
27537	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , AirConditioning, Furnished_Yes)	0.149254	0.572139	0.109453	0.7
27538	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , Hydro, Laundry)	0.134328	0.686567	0.109453	0.8
27539	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , Hydro, Furnished_Yes)	0.144279	0.616915	0.109453	0.7
27540	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , Hydro, AirConditioning)	0.149254	0.626866	0.109453	0.7
27541	(BuildingType_Apartment, Hydro, RoomType_Regul...	(SharedorPrivate_Shared, Internet , Laundry)	0.154229	0.512438	0.109453	0.7
27542	(SharedorPrivate_Shared, Hydro, RoomType_Regul...	(Internet , Laundry, BuildingType_Apartment)	0.139303	0.492537	0.109453	0.7
27543	(SharedorPrivate_Shared, Hydro, RoomType_Regul...	(Internet , Furnished_Yes, BuildingType_Apartm...	0.149254	0.462687	0.109453	0.7
27544	(SharedorPrivate_Shared, Hydro, RoomType_Regul...	(Internet , AirConditioning, BuildingType_Apar...	0.154229	0.477612	0.109453	0.7

	antecedents	consequents	antecedent support	consequent support	support	confi
27545	(SharedorPrivate_Shared, RoomType_Regular Room...	(Internet , Hydro, BuildingType_Apartment)	0.149254	0.507463	0.109453	0.7
27546	(Internet , BuildingType_Apartment, SharedorPr...	(Laundry, Hydro, Furnished_Yes, AirConditioning)	0.139303	0.636816	0.109453	0.7
27547	(Internet , SharedorPrivate_Shared, Hydro, Roo...	(Laundry, AirConditioning, Furnished_Yes, Buil...	0.144279	0.472637	0.109453	0.7
27548	(Internet , SharedorPrivate_Shared, RoomType_R...	(Laundry, Hydro, Furnished_Yes, BuildingType_A...	0.144279	0.487562	0.109453	0.7
27549	(Internet , SharedorPrivate_Shared, RoomType_R...	(Laundry, AirConditioning, Hydro, BuildingType...	0.144279	0.512438	0.109453	0.7
27550	(Internet , SharedorPrivate_Shared, RoomType_R...	(AirConditioning, Hydro, Furnished_Yes, Buildi...	0.154229	0.447761	0.109453	0.7
27551	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , AirConditioning, Furnished_Yes, La...	0.154229	0.557214	0.109453	0.7
27552	(BuildingType_Apartment, SharedorPrivate_Share...	(Internet , Hydro, Furnished_Yes, Laundry)	0.144279	0.587065	0.109453	0.7
27553	(BuildingType_Apartment, SharedorPrivate_Share...	(Laundry, Internet , Hydro, AirConditioning)	0.154229	0.611940	0.109453	0.7
27554	(SharedorPrivate_Shared, Hydro, RoomType_Regul...	(Internet , Laundry, Furnished_Yes, BuildingTy...	0.149254	0.422886	0.109453	0.7
27555	(SharedorPrivate_Shared, RoomType_Regular Room...	(Laundry, Internet , Hydro, BuildingType_Apart...	0.149254	0.467662	0.109453	0.7
27556	(SharedorPrivate_Shared, Internet , HouseAge_N...	(BuildingType_Apartment, Hydro, AirConditionin...	0.154229	0.442786	0.109453	0.7

27557 rows × 9 columns

In [ ]:
```
# End of this file
```