### **Clustering Result**

### **Applying Clustering Algorithms For Population Composition Subset:**

### In order to get the most accurate clustering result, three different n values are used for k-means and hierarchical algorithms. For dbscan method, three different eps values and minimum sample data values are used. The silhouette score was calculated after each method was applied in order to find the most accurate method for each subset of our dataset according to the silhouette score.

### For ‘Population composition’ subset, start with n equals to 2 for k-means and hierarchical algorithms. The average silhouette scores for these two methods are: 0.5428210594 and 0.522266955898. For n equals to 3, silhouette scores are 0.562044592089 and 0.499770744743. For n equals to 4, silhouette scores are 0.531996981475 and 0.530369189816. So the best clustering result for k-means and hierarchical come out when n equals to 3 and 4.

### Figure 1 and 2 shows the clustering result for k-means method with n equals to 3. Figure 3 and 4 shows the clustering result for hierarchical method with n equals to 4.

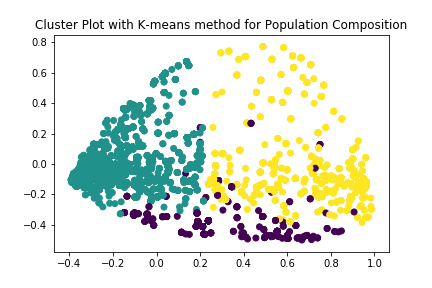


Figure 1 2D graph for k-means method for Population Composition with n=3

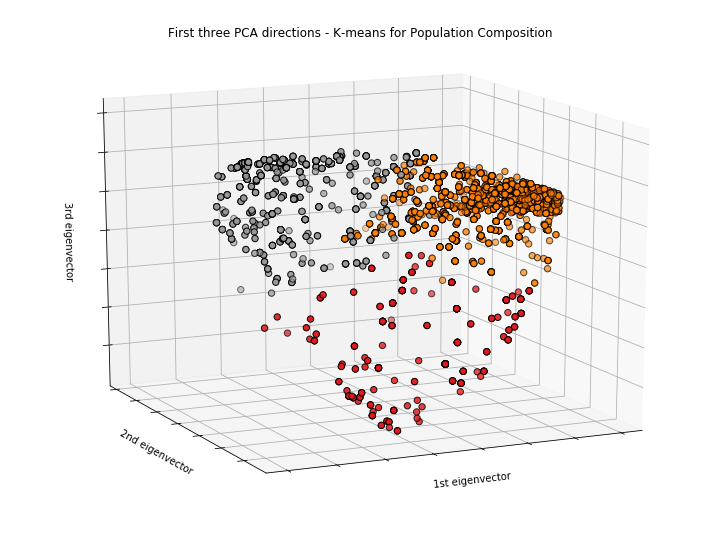


Figure 2 3D graph for k-means method for Population Composition with n=3

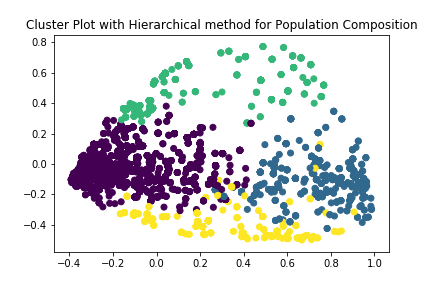


Figure 3 2D graph for hierarchical method for Population Composition with n=4

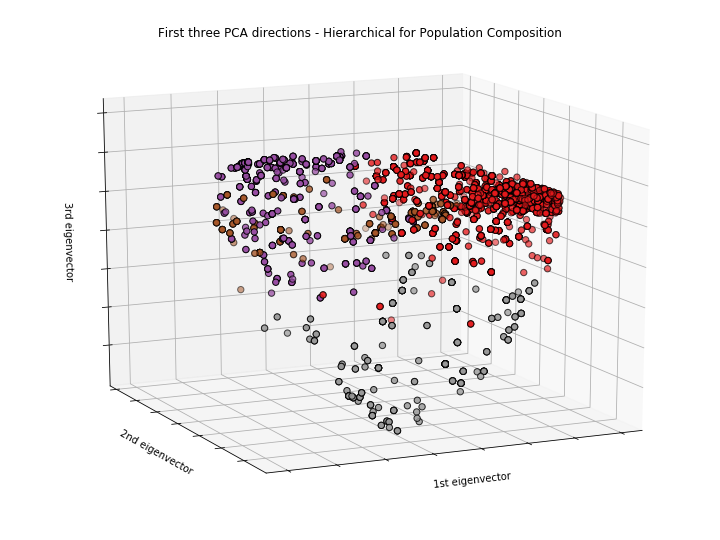


Figure 4 3D graph for hierarchical method for Population Composition with n=4

### For dbscan method, start with eps equals to 0.1, minimum sample data value equals to 130. The average silhouette score is 0.446518921327. When eps equals to 0.15, minimum sample data value equals to 100 the average silhouette score is 0.443423937684. When eps equals to 0.15, minimum sample data value equals to 150 the highest average silhouette score come out which is 0.497415211613. Figure 5 and 6 shows the clustering result for dbscan method with eps equals to 0.15, minimum sample data value equals to 150.

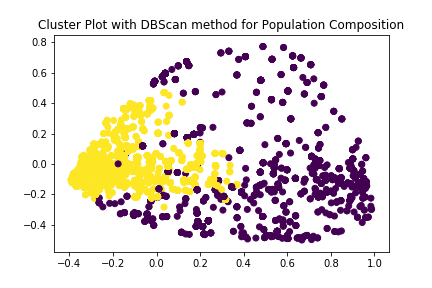


Figure 5 2D graph for dbscan method for Population Composition

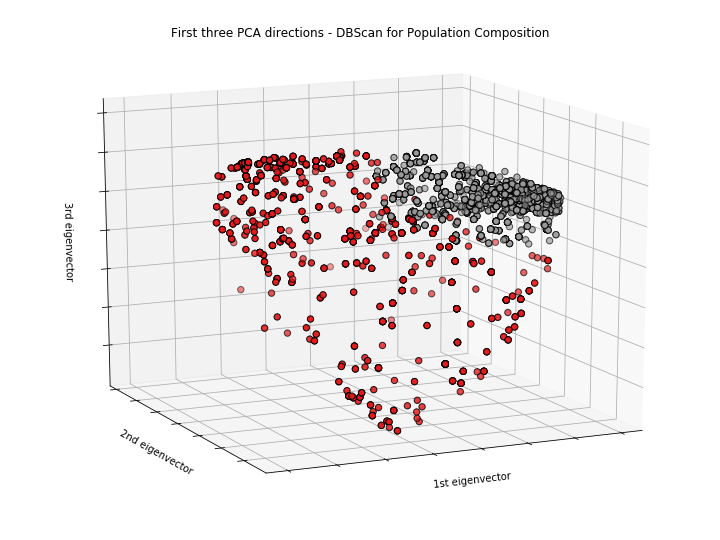


Figure 6 3D graph for dbscan method for Population Composition

### **Applying Clustering Algorithms For Neighborhood Subset:**

### For ‘Neighborhood’ subset both k-means and hierarchical algorithms are implemented with n=4, 6 and 8 respectively. For k-means algorithm, the average silhouette score raises from 0.1759 to 0.1775, and then decreases to 0.1708, so it’s apparent that n=6 is the most suitable value. For hierarchical algorithm, the average silhouette score drops from 0.1374 to 0.1370 and finally 0.1271, so n=4 should fit this algorithm best.

Figure 7 and 8 show the clustering results for k-means method with n=6. Figure 9 and 10 show the clustering results for hierarchical method with n=4.

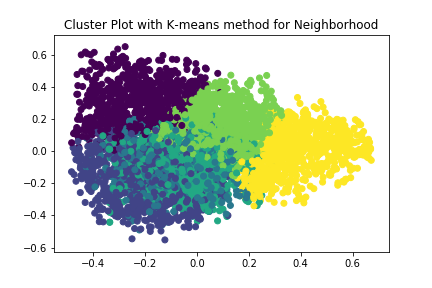


Figure 7 2D graph for k-means method for Neighborhood with n=6

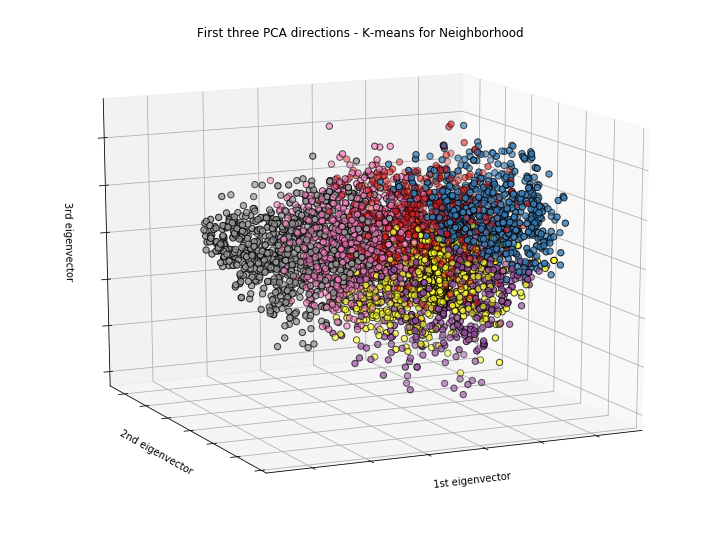


Figure 8 3D graph for k-means method for Neighborhood with n=6

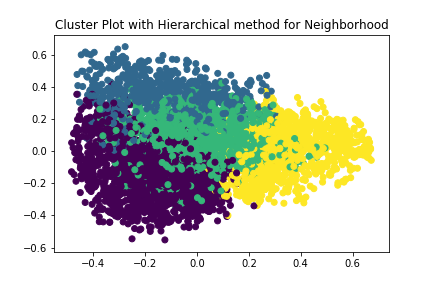


Figure 9 2D graph for hierarchical method for Neighborhood with n=4

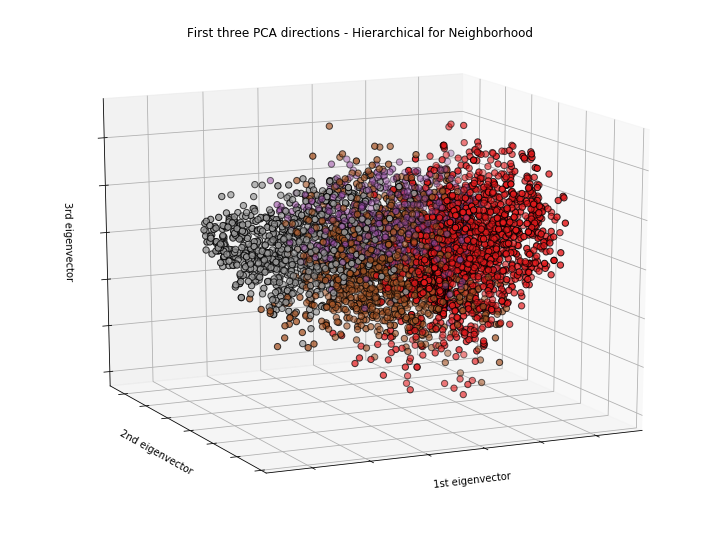


Figure 10 3D graph for hierarchical method for Neighborhood with n=4

### As for dbscan method, we calculate average silhouette score with different pair of eps values and msdv (minimum sample data values). When eps is 0.2, msdv is 100, average silhouette score equlas 0.0281; when eps is 0.25, msdv is 100, average silhouette score raises to 0.1880; when eps is 0.3, msdv is 100, average silhouette score increases to 0.2582. As a result, the third pair indicates the best clustering result. Figure 11 and 12 show the clustering result for dbscan methos with eps=0.3, msdv=100.

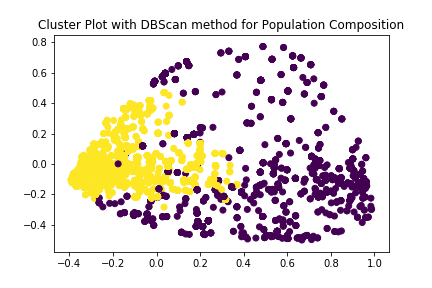


Figure 11 2D graph for dbscan method for Neighborhood

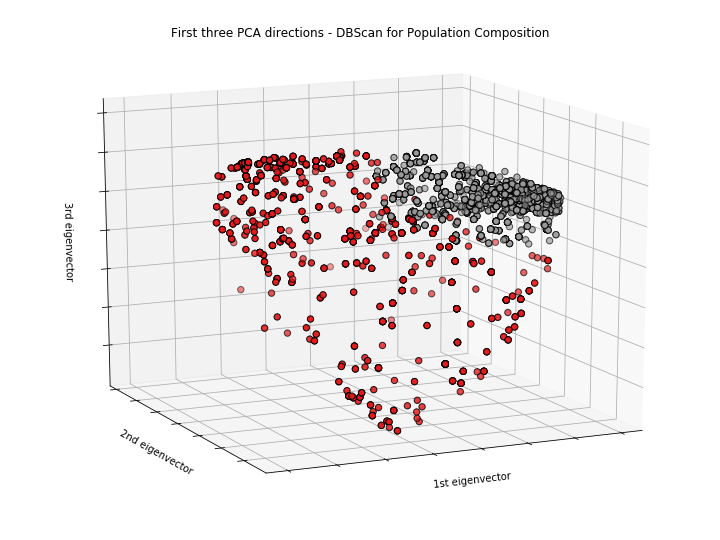


Figure 12 3D graph for dbscan method for Neighborhood

From all above analysis of the clustering results with three different result, we find the most reasonable and effective one is the result of k-means algorithm. Since we have eight columns for one sample, the clustering group shouldn’t be as small as two or four. Meanwhile, with six groups, we can clearly divide those restaurants into different levels and it’s a good way to evaluate the neighborhood’s environment.

### **Applying Clustering Algorithms For ﻿Internal Factors Subset:**

### For ‘﻿Internal Factors’ subset, start with n equals to 2 for k-means and hierarchical algorithms. The average silhouette scores for these two methods are: 0.562120857214 and 0.515396285078. For n equals to 3, silhouette scores are 0.669018440095 and 0.653200696932. For n equals to 4, silhouette scores are 0.80759697343 and 0.778485677777. So the best clustering result for k-means and hierarchical come out when n equals to 4.

### Figure 13 and 14 shows the clustering result for k-means method with n equals to 4. Figure 15 and 16 shows the clustering result for hierarchical method with n equals to 4.

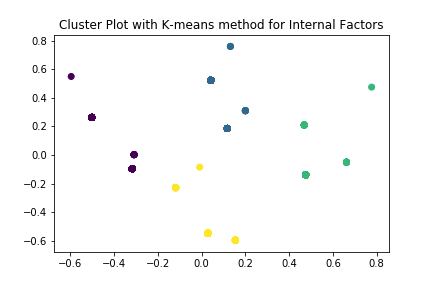


Figure 13 2D graph for k-means method for Internal Factors with n=4

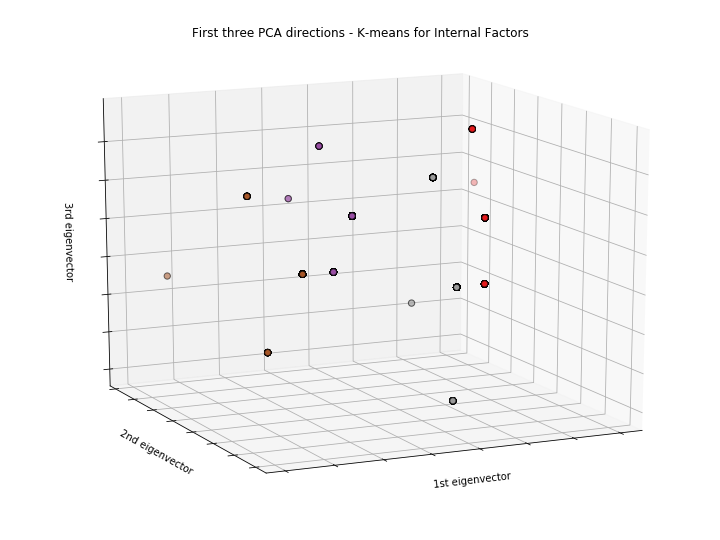


Figure 14 3D graph for k-means method for Internal Factors with n=4

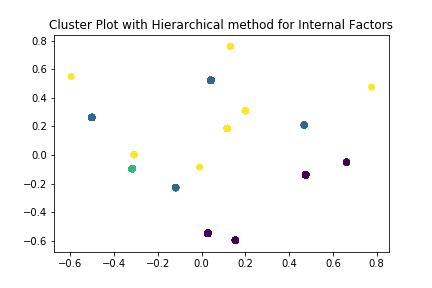


Figure 15 2D graph for hierarchical method for Internal Factors with n=4

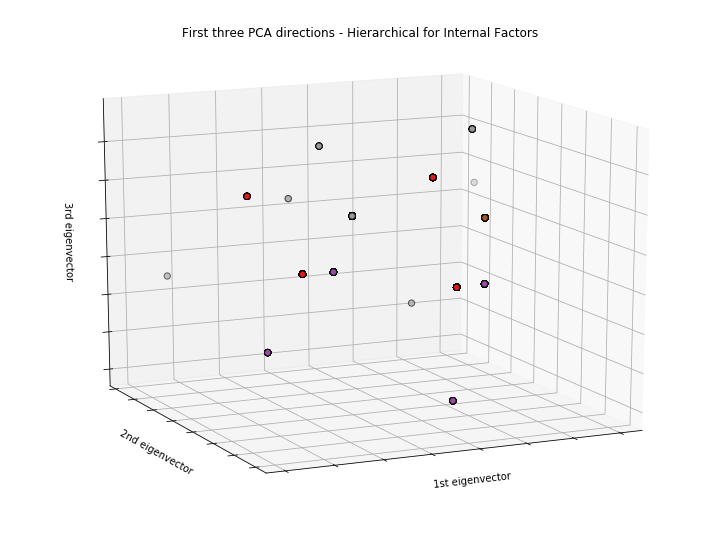


Figure 16 3D graph for hierarchical method for Internal Factors with n=4

### For dbscan method, start with eps equals to 0.1, minimum sample data value equals to 130. The average silhouette score is 0.96969164214. When eps equals to 0.2, minimum sample data value equals to 130 the average silhouette score is 0.96969164214. When eps equals to 0.05, minimum sample data value equals to 130 the average silhouette score is 0.96969164214. So the clustering results are same when applying different eps and minimum sample data value. Figure 17 and 18 shows the clustering result for dbscan method with eps equals to 0.1, minimum sample data value equals to 130.

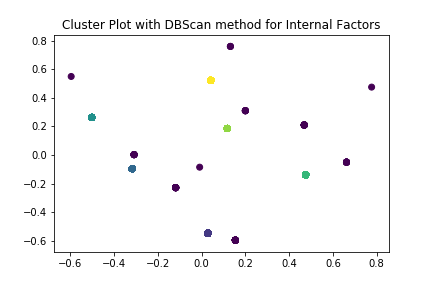


Figure 17 2D graph for dbscan method for Internal Factors

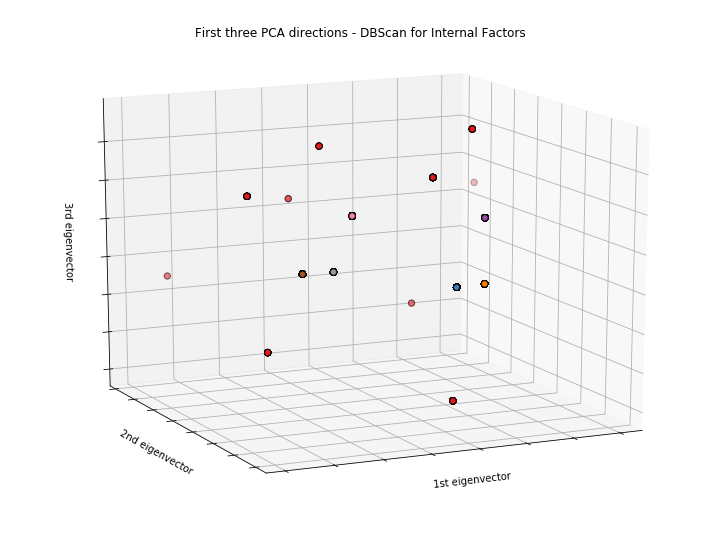


Figure 18 3D graph for dbscan method for Internal Factors

### **Applying Clustering Algorithms For ﻿** **﻿Internal Richness Subset:**

### For ‘﻿﻿Internal Richness’ subset, start with n equals to 3 for k-means and hierarchical algorithms. The average silhouette scores for these two methods are: 0.485197305752 and 0.45816872821. For n equals to 4, silhouette scores are 0.548061064904 and 0.513084367398. For n equals to 5, silhouette scores are 0.58864542381 and 0.564607876546. So the best clustering result for k-means and hierarchical come out when n equals to 5.

### Figure 19 and 20 shows the clustering result for k-means method with n equals to 5. Figure 21 and 22 shows the clustering result for hierarchical method with n equals to 5.

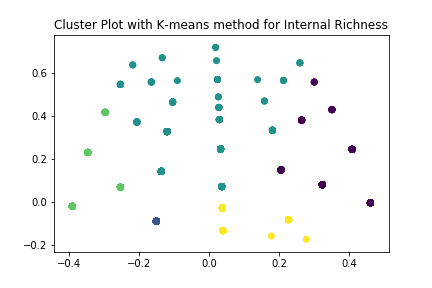


Figure 19 2D graph for k-means method for Internal Richness with n=5

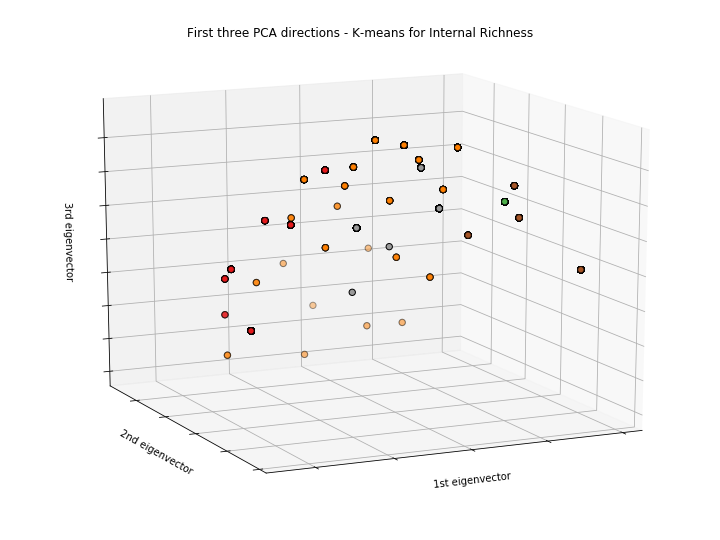


Figure 20 3D graph for k-means method for Internal Richness with n=5

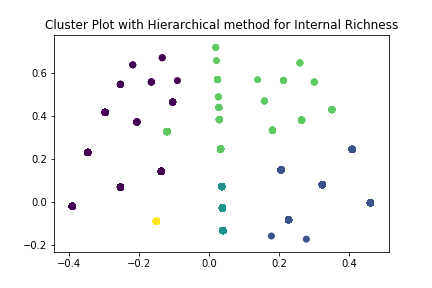


Figure 21 2D graph for hierarchical method for Internal Richness with n=5

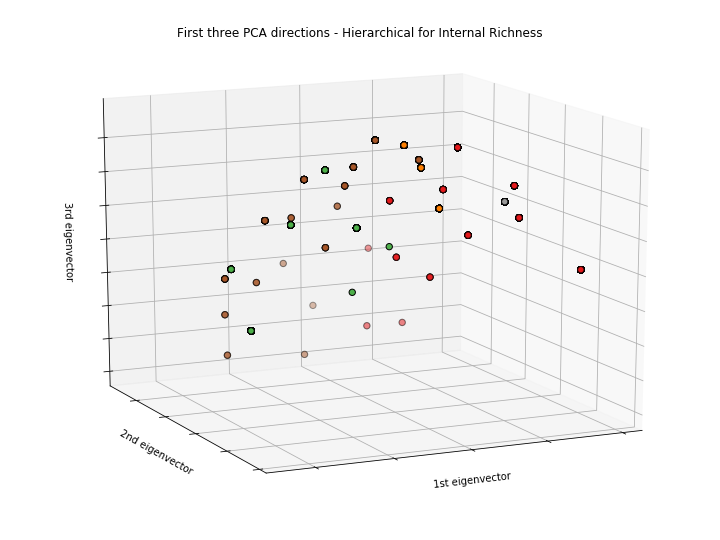


Figure 22 3D graph for hierarchical method for Internal Richness with n=5

### For dbscan method, start with eps equals to 0.1, minimum sample data value equals to 130. The average silhouette score is 0.851164854876. When eps equals to 0.15, minimum sample data value equals to 130 the average silhouette score is 0.579970616641. When eps equals to 0.2 and minimum sample data value equals to 130 the average silhouette score is 0.543111298669. So the clustering results comes out when eps equals to 0.1 and minimum sample data value equals to 130. Figure 23 and 24 shows the clustering result for dbscan method with eps equals to 0.1, minimum sample data value equals to 130.

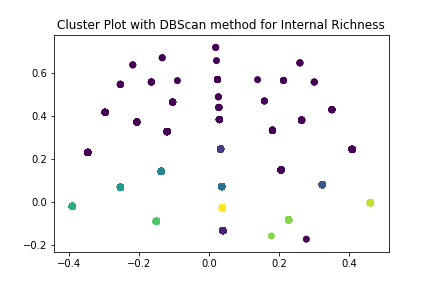


Figure 23 2D graph for dbscan method for Internal Richness

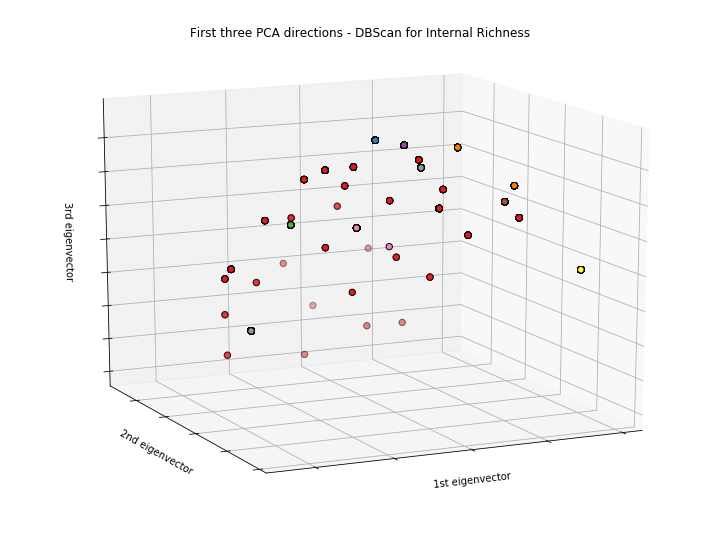


Figure 24 3D graph for dbscan method for Internal Richness

### **Applying Clustering Algorithms For ﻿Education Level Subset:**

### For ﻿‘Education Level’ subset, start with n equals to 3 for k-means and hierarchical algorithms. The average silhouette scores for these two methods are: 0.44461947209 and 0.396569829389. For n equals to 4, silhouette scores are 0.378366303973 and 0.314693459797. For n equals to 5, silhouette scores are 0.358719950757 and 0.290352300189. So the best clustering result for k-means and hierarchical come out when n equals to 3.

### Figure 25 and 26 shows the clustering result for k-means method with n equals to 3. Figure 27 and 28 shows the clustering result for hierarchical method with n equals to 3.

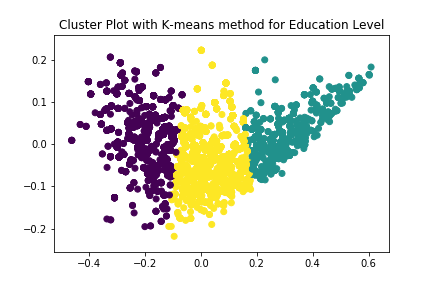


Figure 25 2D graph for k-means method for Education Level with n=3

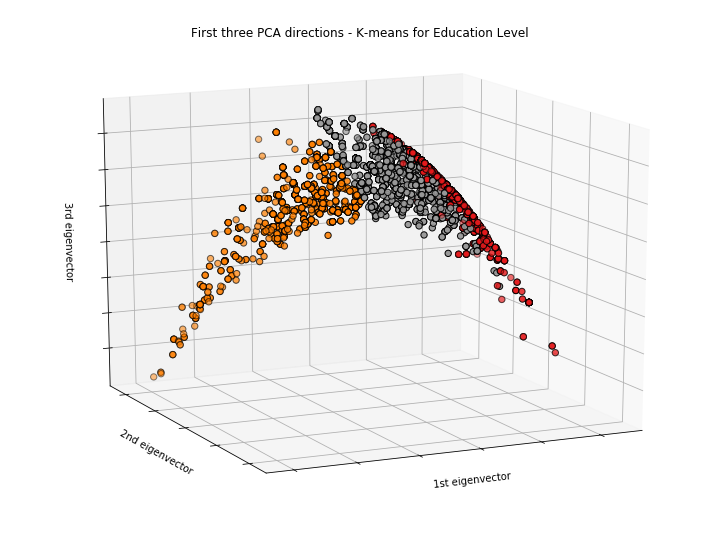


Figure 26 3D graph for k-means method for Education Level with n=3

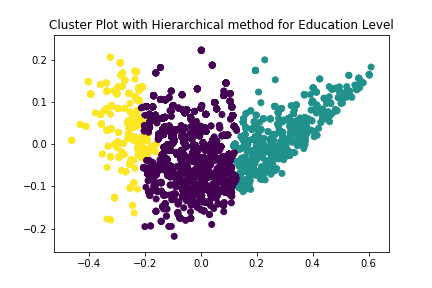


Figure 27 2D graph for hierarchical method for Education Level ness with n=3

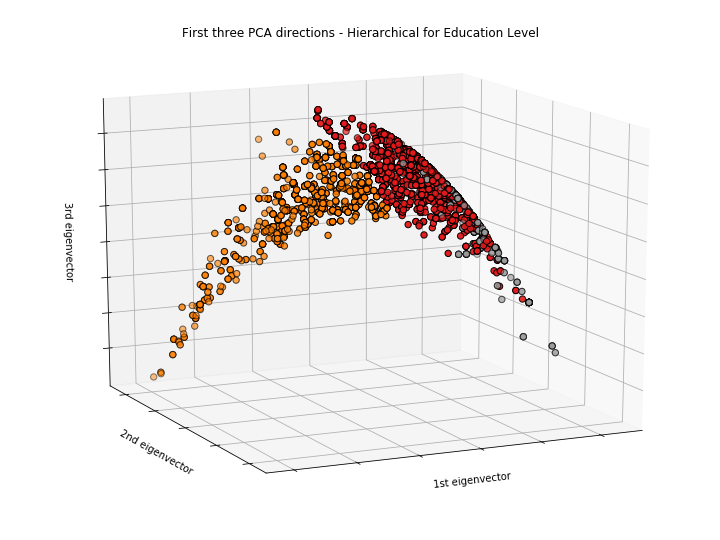


Figure 28 3D graph for hierarchical method for Education Level with n=3

### For dbscan method, start with eps equals to 0.1, minimum sample data value equals to 130. The average silhouette score is 0.545970518826. When eps equals to 0.05, minimum sample data value equals to 130 the average silhouette score is 0.37310461771. When eps equals to 0.02 and minimum sample data value equals to 130 the average silhouette score is -0.277423432552. So the clustering results comes out when eps equals to 0.1 and minimum sample data value equals to 130. Figure 29 and 30 shows the clustering result for dbscan method with eps equals to 0.1, minimum sample data value equals to 130.

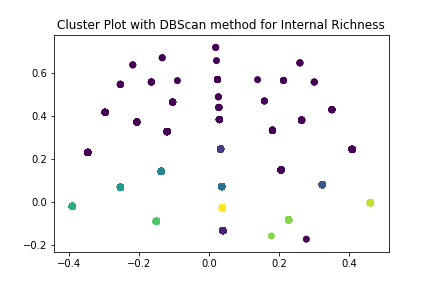


Figure 29 2D graph for dbscan method for Education Level

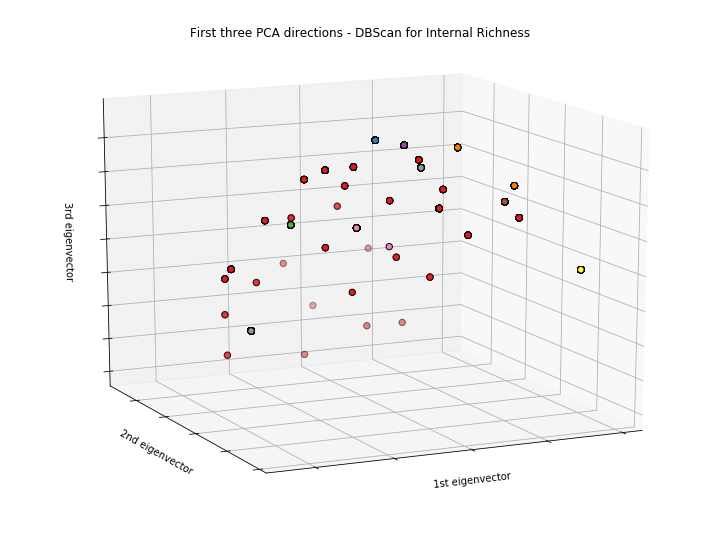


Figure 30 3D graph for dbscan method for Education Level