**스마트테크놀로지AIR HW – Clustering**

컴퓨터과학과

2018147563 주우진

1. 코드 설명

In [1] ~ [4] : 과제 설명에서와 같이 피벗 테이블을 생성하고 엑셀에서 진행한 피벗 테이블과 같은 형태로 만들었다. 각 values 들은 엑셀 피벗 테이블의 ‘개수’로 하였다.

In [5] ~ [8] : K=3, 4, 5에 대해 각각 K-means Clustering을 진행하였고, 각 clustering에서의 실루엣 값과 실루엣 값이 가장 높은 K값을 출력하였다.

In [9] ~ [11] : 실루엣 값이 가장 높은 K값에 대해 clustering된 표를 만든다.

In [12] ~ [15] : 엑셀의 ‘OfferInformation’ 시트를 불러와 clustering된 표와 합쳐서 정렬한다.

In [16] ~ [19] : 각 Cluster 별로 정렬하여 각 Cluster의 특징을 살펴본다.

1. 분석

K = 3, 4, 5에 대해 K-means Clustering을 진행하고 silhouette 값을 추출하였을 때 각 K에 대해 silhouette 값이 각각 약 0.1208, 0.1274, 0.1267로 나왔다. 이 중 silhouette 값이 1에 가장 가까운 K = 4가 적합하다고 판단하였고 이에 대한 평가를 하기로 하였다.

1. 평가

각 Cluster 별로 특징을 보이는지 살펴보았을 때, Cluster 1에선 대부분 주종이 ‘Pinot Noir’인 제안을 선호하는 것으로 나타났고 Cluster 2에선 Minimum Qty가 다른 주종보다 상당히 높은 제안을 선호하는 것으로 나타났다. 주종으로는 ‘Champagne’이 지배적이었고 ‘Prosecco’가 그 다음 순위로 많았다. Cluster 3에선, 대다수가 주종이 ‘Champagne’이거나 원산지가 ‘France’인 제안을 선호하였다. 하지만 다른 Cluster들에 비해 비교적 넓게 퍼져있어 특징을 단정짓기는 어렵다. 마지막으로 Cluster4는 4개의 Cluster 중 가장 밀집해있는 Cluster인데 이들은 주종이나 원산지와는 상관없이 Minimum Qty가 가장 작은, 6kg인 제안을 선호하는 것으로 나타났다.

4개의 Cluster에서 뚜렷한 특징을 보이지 않는 Cluster도 있었지만 대체적인 경향성을 각각이 가지고 있었다. Cluster 별로 주종과 Minimum Qty, 원산지에 대한 선호도는 보였지만 할인율에 대한 선호도는 대체로 비슷하였다.