

스마트테크놀로지AIR HW-Optimization

2017147535 하연안

1. Pivot Table 생성

Pivot Table은 Excel 파일의 Transactions 시트를 이용했다. 이 시트는 "Offer #", "Customer Last Name" 총 두 개의 열로 구성되어 있다. Python 코드의 In[3]에서 "Offer #"을 행과 값, "Customer Last Name"을 열로 갖는 pivotTable을 선언한다. 이후 NaN 값은 다 0으로 변환한 후 자료형을 int로 변환한다.

2. K-Means에서 K의 결정

Python 코드의 In[5]에서 3 이상 10 미만의 자연수 K를 각각 100번씩 실행한 후 평균 실루엣이 가장 높은 K를 찾는다.

| k | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| 평균 실루엣 | 0.231093 1328909 0733 | 0.252800 9860843 9683 | 0.243668 5581208 4232 | 0.239889 2520104 9935 | 0.241652 5274910 4627 | 0.242344 8299753 5298 | 0.245037 8912195 0423 |

3. Clustering된 dataframe

Python 코드의 In[7]에서 K개의 열을 갖는 list를 만들어 offer #에 따라 각 cluster의 개수를 구한다. 이후 In[9]에서 clustedDf를 선언한 후 최종적으로 In[10]에서 병합한다.

4. 분석

우선 K=4에 대하여 clus_0, clus_1, clus_2, clus_3 총 4개의 클러스터를 각각 내림차순으로 정렬한 후 분석을 진행했다.

- clus_0

와인의 Origin이 Chille, France일 때 구매가 잦았으며 대체로 Minimum Qty가 72kg일 때 구매가 잦았다.

- clus_1

와인의 Varietal이 Champagne일 때 구매가 잦았으며 Origin이 France일 때 구매가 잦았다.

- clus_2

와인의 Varietal이 Pinot Noir일 때 구매가 잦았다.

- clus_3

와인의 Minimum Qty가 6kg일 때 구매가 잦았다.