

# HW-Optimization

(스마트테크놀로지 AIR)

컴퓨터과학과 2015147533 유현석

## 1. 코딩 과정 및 중간 결과 출력

1. Pandas 안에 있는 read\_excel 함수를 이용하여 주어진 엑셀 데이터의 2번째 시트인 Transactions을 읽어왔습니다.

2. 그 후 count라는 column을 생성한 후 그 값을 1로 채웠습니다. 그 이유로는 Pivot table을 만들기 위해서 해당하는 Offer에 구매한 사람들의 데이터가 필요했기 때문입니다.

3. Pandas 안에 있는 pivot\_table 함수를 이용하여 table이라는 이름의 Pivot table을 만들었습니다.

4. Kmeans 함수를 이용하여 K = 3, 4, 5 일때의 값을 구하였고 그 데이터를 이용하여 비교하여 최적의 K값을 구하기 위해 Silhouette\_score라는 함수를 이용하여 값을 추출했습니다. 그 과정에서 각각의 K에 대해 100번의 반복을 실행하였고 그 값들 중 최고값을 추출하여 비교하였습니다.

( K=3, 4, 5에 대한 Silhouette\_score 값)

```
In [14]: print("3:",sil1, " 4:",sil2, " 5:",sil3)  
3: 0.12728440119354445  4: 0.2503738499502217  5: 0.21152679480717587
```

5. 가장 Silhouette\_score의 값이 높아 분류가 잘 되었다고 판단한 4개의 Cluster들에 대해 각각의 Cluster에 해당하는 Offer를 구매한 인원 수를 구하였습니다.

그리고 그 값을 마찬가지로 Pandas 안에 있는 read\_excel 함수로 읽어온 첫 번째 시트인 OfferInformation 옆에 4개의 Cluster column들을 만든 후 대입하였습니다. 그 결과로 나온 것이 아래의 사진이고 그 값들을 비교 분석하여 해당 Cluster들의 공통점을 골라서 비교 분석할 수 있게 되었습니다.

(OfferInformation 시트 옆에 해당 Cluster column에 값들을 대입 한 결과)

```
In [22]: df1
```

```
Out[22]:
```

	Offer #	Campaign	Varietal	Minimum Qty (kg)	Discount (%)	Origin	Past Peak	clu1	clu2	clu3	clu4
0	1	January	Malbec	72	56	France	False	0	5	1	4
1	2	January	Pinot Noir	72	17	France	False	0	4	6	0
2	3	February	Espumante	144	32	Oregon	True	0	6	0	0
3	4	February	Champagne	72	46	France	True	0	9	0	3
4	5	February	Cabernet Sauvignon	144	44	New Zealand	True	0	1	0	3
5	6	March	Prosecco	144	86	Chile	False	1	8	0	3
6	7	March	Prosecco	6	40	Australia	True	12	7	0	0
7	8	March	Espumante	6	46	South Africa	False	12	8	0	0
8	9	April	Chardonnay	144	57	Chile	False	0	7	0	3
9	10	April	Prosecco	72	52	California	False	1	3	1	2
10	11	May	Champagne	72	85	France	False	0	11	0	2
11	12	May	Prosecco	72	83	Australia	False	0	4	1	0
12	13	May	Merlot	6	43	Chile	False	6	0	0	0
13	14	June	Merlot	72	64	Chile	False	0	0	0	9
14	15	June	Cabernet Sauvignon	144	19	Italy	False	0	3	0	3
15	16	June	Merlot	72	88	California	False	0	4	1	0
16	17	July	Pinot Noir	12	47	Germany	False	0	0	7	0
17	18	July	Espumante	6	50	Oregon	False	13	1	0	0
18	19	July	Champagne	12	66	Germany	False	0	5	0	0
19	20	August	Cabernet Sauvignon	72	82	Italy	False	0	4	0	2
20	21	August	Champagne	12	50	California	False	1	3	0	0
21	22	August	Champagne	72	63	France	False	0	17	0	4
22	23	September	Chardonnay	144	39	South Africa	False	0	1	1	3
23	24	September	Pinot Noir	6	34	Italy	False	0	0	12	0
24	25	October	Cabernet Sauvignon	72	59	Oregon	True	0	4	0	2
25	26	October	Pinot Noir	144	83	Australia	False	0	0	12	3
26	27	October	Champagne	72	88	New Zealand	False	0	8	1	0
27	28	November	Cabernet Sauvignon	12	56	France	True	1	5	0	0
28	29	November	Pinot Grigio	6	87	France	False	16	1	0	0
29	30	December	Malbec	6	54	France	False	16	5	0	1
30	31	December	Champagne	72	89	France	False	0	15	0	2
31	32	December	Cabernet Sauvignon	72	45	Germany	True	0	4	0	0

## 2. 분석

1번 Cluster의 특징으로는 눈에 띄게 Minimum Qty가 6인 값들에 대해 높은 수치를 보여주고 있었습니다.

2번 Cluster의 특징으로는 나머지 Cluster중 대다수를 차지하였고 값이 제일 다양하고 많았습니다.

3번 Cluster의 특징으로는 varietal이 Pinot Noir이었을 때 그 해당하는 값들에 대해 높은 수치를 보여주고 있었습니다.

4번 Cluster Minimum Qty가 주로 72, 144인 값이 대다수를 차지하였습니다.