

PART II

연관분석

# Requirement 2-1

View 'category\_score' & category.csv

```
194 # TODO: Requirement 2-1. CREATE VIEW AND SAVE to .csv file
195 fopen = open('DMA_project2_team%02d_part2_category.csv' % team, 'w', encoding='utf-8')
196
197 cursor.execute('''
198 CREATE OR REPLACE VIEW category_score AS
199 SELECT c.category_id as category_id, c.category_name as category_name, num_business, num_reviews,
200 ctg_avg_stars, ctg_avg_stars*(num_reviews/num_business) AS score
201 FROM categories c,
202 (SELECT category_id, COUNT(*) AS num_business FROM business_categories GROUP BY category_id) AS bnum,
203 (SELECT bc.category_id, SUM(tnr) AS num_reviews, SUM(sr)/SUM(tnr) AS ctg_avg_stars FROM
204 business_categories bc,(SELECT business_id, COUNT(review_id) AS tnr,SUM(review_stars) AS sr FROM reviews
205 GROUP BY business_id) AS review WHERE bc.business_id=review.business_id GROUP BY category_id) AS r
206 WHERE c.category_id=bnum.category_id AND c.category_id=r.category_id
207 ORDER BY score DESC
208 LIMIT 30;
209 ''')
210 cursor.execute('SELECT * FROM category_score')
211 df1=pd.DataFrame(cursor.fetchall())
212 df1.columns=cursor.column_names
213 df1.to_csv('DMA_project2_team%02d_part2_category.csv' % team)
214 fopen.close()
```

}

business_id	tnr	sr
1	5	14
2	28	128
3	52	207
4	3	12
5	2	10
6	4	15
7	3	11
8	7	15
9	8	31
10	3	13
11	20	80
12	3	7
13	69	257
14	3	7
15	4	11
16	10	47
17	8	32
18	6	30
19	4	19
20	3	10
21	3	6

+

business_id	category_id
16	1
42	1
43	1
57	1
87	1
93	1
109	1
155	1
159	1
178	1
190	1
194	1
200	1
266	1
268	1
274	1
291	1
299	1
313	1
327	1
330	1

}

서브쿼리 review

Business categories

category_id	category_name
1	Active Life
2	Arts & Entertainment
3	Stadiums & Arenas
4	Horse Racing
5	Tires
6	Automotive
7	Fashion
8	Shopping
9	Department Stores
10	Women's Clothing
11	Men's Clothing
12	Accessories
13	Pet Services
14	Pet Boarding/Pet Sitting
15	Pets
16	Veterinarians
17	Mexican
18	Restaurants
19	Real Estate Services
20	Property Management
21	Home Services
22	Real Estate

categories

+

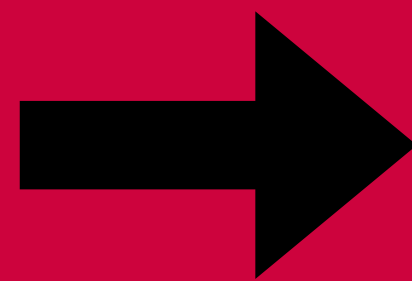
category_id	num_business
1	695
2	385
3	27
4	2
5	147
6	859
7	701
8	2499
9	223
10	227
11	118
12	121
13	159
14	65
15	308
16	101
17	822
18	5556

서브쿼리 bnum

+

category_id	num_reviews	ctg_avg_stars
1	7958	3.9910
2	11516	3.8367
3	1110	3.9360
4	29	3.6207
5	1163	3.8040
6	5361	3.7907
7	5901	3.6340
8	20492	3.7347
9	2251	3.3367
10	1854	3.7071
11	1179	3.6658
12	765	3.8627
13	1159	4.0604
14	390	4.3333
15	2085	4.1108
16	665	4.2120
17	28201	3.6615
18	224732	3.7587
19	96	3.9896
20	59	2.8475
21	3497	3.7383

서브쿼리 r



```
mysql> select * from category_score;
```

category_id	category_name	num_business	num_reviews	ctg_avg_stars	score
247	British	6	1500	4.2247	1056.17500000
526	Cambodian	1	161	4.3727	704.00470000
299	Botanical Gardens	2	278	4.5036	626.00040000
288	Szechuan	4	611	3.8936	594.74740000
572	Food Court	1	177	3.1977	565.99290000
584	Brasseries	1	127	4.3307	549.99890000
280	Climbing	2	249	4.4016	547.99920000
343	Zoos	2	251	3.8406	481.99530000
240	Pubs	57	6938	3.8687	470.89544912
147	Vegetarian	55	6497	3.9070	461.52325454
318	Brazilian	3	323	4.1331	444.99710000
234	Vegan	26	2802	4.0228	433.53406154
394	Breweries	31	3461	3.8388	428.58344516
509	Piano Bars	2	222	3.7477	415.99470000
382	Gastropubs	17	1727	4.0411	410.52821765
136	Wine Bars	64	6270	4.0769	399.40879688
517	Fondue	4	403	3.8635	389.24762500
364	Irish	17	1779	3.6869	385.82324117
135	German	6	597	3.8526	383.33370000
485	Airports	10	1142	3.3021	377.09982000
528	Tapas/Small Plates	8	759	3.8762	367.75447500
536	Scandinavian	3	252	4.2063	353.32920000
175	Soul Food	16	1388	4.0490	351.25075000
216	Latin American	32	2744	4.0160	344.37200000
419	Southern	19	1645	3.8340	331.94368421
207	Sushi Bars	124	10538	3.6607	311.10045645
502	Ethiopian	5	372	4.0995	305.00280000
125	American (New)	441	34085	3.8666	298.85047846
400	African	4	290	4.0931	296.74975000
493	Modern European	3	223	3.9686	294.99926667

```
30 rows in set (1.84 sec)
```



# Requirement 2-2.1

## View 'user\_category\_rating'

```
214 # TODO: Requirement 2-2. CREATE 2 VIEWS AND SAVE partial one to .csv file
215 # User category rating view
216 ~ cursor.execute('''
217     CREATE OR REPLACE VIEW user_category_rating AS
218     SELECT f.user_id, f.category_name, (LEAST(f.num,5)+2*IFNULL(s.onum,0)) AS rating
219     FROM (SELECT r.user_id, cs.category_id, cs.category_name, count(cs.category_id) AS num FROM reviews
220           r,category_score cs, business_categories bc where r.business_id=bc.business_id AND bc.category_id=cs.
221           category_id GROUP BY r.user_id, category_id ORDER BY user_id,category_id) AS f
222     LEFT JOIN
223     (SELECT r.user_id, cs.category_id, count(cs.category_id) AS onum FROM reviews r,category_score cs,
224           business_categories bc where r.business_id=bc.business_id AND bc.category_id=cs.category_id AND r.
225           review_stars>=4 GROUP BY r.user_id, category_id ORDER BY user_id,category_id) AS s
226     ON f.user_id=s.user_id AND f.category_id=s.category_id
227 ''')
```

서브쿼리 f

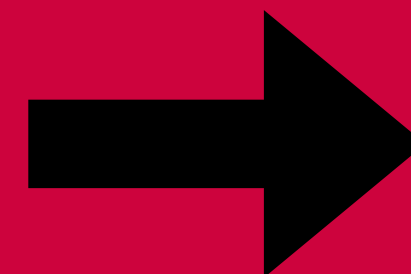
(LEFT JOIN)

서브쿼리 s

user_id	category_id	category_name	num
3	125	American (New)	8
3	136	Wine Bars	2
3	147	Vegetarian	1
3	175	Soul Food	2
3	207	Sushi Bars	2
3	216	Latin American	1
3	234	Vegan	1
3	240	Pubs	7
3	247	British	1
3	364	Irish	1
3	382	Gastropubs	1
3	394	Breweries	4
3	419	Southern	1
3	536	Scandinavian	1
19	207	Sushi Bars	1
20	394	Breweries	1
21	207	Sushi Bars	1
21	240	Pubs	1
21	517	Fondue	1
27	125	American (New)	1
27	147	Vegetarian	1
31	136	Wine Bars	1
33	136	Wine Bars	1

+

user_id	category_id	onum
3	125	7
3	136	2
3	175	2
3	216	1
3	240	5
3	247	1
3	364	1
3	382	1
3	394	4
3	419	1
3	536	1
20	394	1
21	207	1
27	125	1
27	147	1
31	136	1
33	136	1
33	240	1
33	247	1
35	382	1
35	419	1



```
mysql> select * from user_category_rating;
+-----+-----+-----+
| user_id | category_name | rating |
+-----+-----+-----+
| 3 | American (New) | 19 |
| 3 | Wine Bars | 6 |
| 3 | Vegetarian | 1 |
| 3 | Soul Food | 6 |
| 3 | Sushi Bars | 2 |
| 3 | Latin American | 3 |
| 3 | Vegan | 1 |
| 3 | Pubs | 15 |
| 3 | British | 3 |
| 3 | Irish | 3 |
| 3 | Gastropubs | 3 |
| 3 | Breweries | 12 |
| 3 | Southern | 3 |
| 3 | Scandinavian | 3 |
| 19 | Sushi Bars | 1 |
| 20 | Breweries | 3 |
| 21 | Sushi Bars | 3 |
| 21 | Pubs | 1 |
| 21 | Fondue | 1 |
| 27 | American (New) | 3 |
| 27 | Vegetarian | 3 |
| 31 | Wine Bars | 3 |
| 33 | Wine Bars | 3 |
```

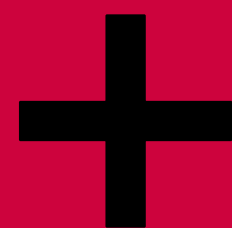
# Requirement 2-2.2

## View & csv 'partial\_user\_category\_rating'

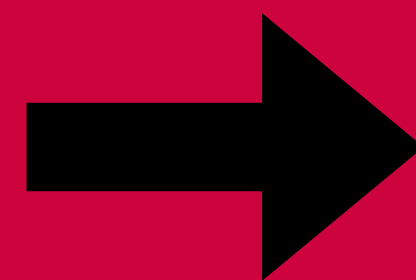
```
226 # Partial user category rating view
227 fopen = open('DMA_project2_team%02d_part2_UCR.csv' % team, 'w', encoding='utf-8')
228 cursor.execute('''
229     CREATE OR REPLACE VIEW partial_user_category_rating AS
230     SELECT ucr.user_id AS user, ucr.category_name AS category, ucr.rating
231     FROM user_category_rating ucr
232     RIGHT JOIN
233     (SELECT a.user_id, a.rating_num FROM (SELECT user_id, COUNT(*) AS rating_num
234     FROM user_category_rating GROUP BY user_id)a WHERE rating_num >= 10) AS rn
235     ON ucr.user_id=rn.user_id
236 ''')
237 cursor.execute("SELECT * FROM partial_user_category_rating")
238 df2=pd.DataFrame(cursor.fetchall())
239 df2.columns=cursor.column_names
240 partial_ucr=df2.set_index("user")
241 partial_ucr.to_csv('DMA_project2_team%02d_part2_UCR.csv' % team)
242 fopen.close()
```

user_id	category_name	rating
3	American (New)	19
3	Wine Bars	6
3	Vegetarian	1
3	Soul Food	6
3	Sushi Bars	2
3	Latin American	3
3	Vegan	1
3	Pubs	15
3	British	3
3	Irish	3
3	Gastropubs	3
3	Breweries	12
3	Southern	3
3	Scandinavian	3
19	Sushi Bars	1
20	Breweries	3
21	Sushi Bars	3
21	Pubs	1
21	Fondue	1
27	American (New)	3
27	Vegetarian	3
31	Wine Bars	3
33	Wine Bars	3
33	Pubs	3

(RIGHT JOIN)



user_id	rating_num
3	14
352	17
535	14
659	12
876	10
959	16
1135	10
1284	21
1463	10
1607	11
1718	10
1742	13
1771	10
1818	13
2160	12
2187	10
2297	17
2389	11
2445	11
2654	10
3171	11
3622	12
3665	11
3703	13
3789	10
3976	10
4210	10



```
mysql> select * from partial_user_category_rating;
```

user	category	rating
3	American (New)	19
3	Wine Bars	6
3	Vegetarian	1
3	Soul Food	6
3	Sushi Bars	2
3	Latin American	3
3	Vegan	1
3	Pubs	15
3	British	3
3	Irish	3
3	Gastropubs	3
3	Breweries	12
3	Southern	3
3	Scandinavian	3
352	American (New)	25
352	Wine Bars	11
352	Soul Food	9
352	Sushi Bars	5
352	Latin American	3
352	Vegan	3
352	Pubs	19
352	British	6
352	Climbing	3
352	Zoos	3
352	Irish	9
352	Gastropubs	1
352	Breweries	7
352	African	3
352	Southern	3
352	Airports	3
352	Ethiopian	3
535	American (New)	27
535	Wine Bars	12
535	Vegetarian	6
535	Soul Food	9

View 'user\_category\_rating'

서브쿼리 rn



# Requirement 2-3

## Horizontal View

## TA시간의 Code를 참고하여 query 저장

# DMA\_project2\_part2\_horizontal.pkl

## 파일로 저장

```

242 # TODO: Requirement 2-3. MAKE HORIZONTAL VIEW
243 # file name: DMA_project2_team##_part2_horizontal.pkl
244 # use to_pickle(): df.to_pickle(filename)
245 ucr=pd.read_csv('DMA_project2_team%02d_part2_UCR.csv' % team)
246 category_set=set(ucr.category.values)
247
248 lquery=[]
249 for category in category_set:
250     query='MAX(IF(category="{ }",1,0)) AS "{ }"'.format(category,category)
251     lquery.append(query)
252 jquery=', '.join(lquery)
253
254 cursor.execute('''
255 SELECT user,{ }
256 FROM partial_user_category_rating
257 GROUP BY user
258 '''.format(jquery))
259
260 df3=pd.DataFrame(cursor.fetchall())
261 df3.columns=cursor.column_names
262 df3=df3.set_index('user')
263 df3.to_pickle('DMA_project2_team%02d_part2_horizontal.pkl' % team)
264 # print output
265 print('Horizontal View')
266 with open('DMA_project2_team%02d_part2_horizontal.pkl' % team,'rb') as files:
267     print(pickle.load(files))

```

# Requirement 2-4

## Association.pkl & .csv

```
268 # TODO: Requirement 2-4. ASSOCIATION ANALYSIS
269 # filename: DMA_project2_team##_part2_association.pkl (pandas dataframe)
270 frequent_itemset=apriori(df3,min_support=0.15,use_colnames=True)
271 # print output
272 print('Support')
273 print(frequent_itemset)
274
275 rules=association_rules(frequent_itemset,metric='lift',min_threshold=3)
276 # write a pickle file
277 rules.to_pickle('DMA_project2_team%02d_part2_association.pkl' % team)
278 # print output
279 print('Association')
280 with open('DMA_project2_team%02d_part2_association.pkl' % team,'rb') as afiles:
281 |     print(pickle.load(afiles))
282 # write a csv file
283 rules.to_csv('DMA_project2_team%02d_part2_association.csv' % team)
284
285 cursor.close()
```

support 0.15 이상, lift 3 이상의 값들  
연관분석 시행

DMA\_project2\_part2\_association.pkl & csv  
파일로 저장



# Requirement 2-4

## 연관분석 결과 해설

1	support	itemsets					
2	0.9947644	frozenset({'American (New)'})					
3	0.94764398	frozenset({'Pubs'})					
4	0.94240838	frozenset({'American (New)', 'Pubs'})					
5	0.90314136	frozenset({'Sushi Bars'})					
6	0.89790576	frozenset({'Wine Bars'})					
7	0.89790576	frozenset({'American (New)', 'Wine Bars'})					
8	0.89790576	frozenset({'American (New)', 'Sushi Bars'})					
9	0.87958115	frozenset({'Vegetarian'})					
10	0.87434555	frozenset({'American (New)', 'Vegetarian'})					
11	0.85602094	frozenset({'Pubs', 'Wine Bars'})					
12	0.85602094	frozenset({'American (New)', 'Pubs', 'Wine Bars'})					
13	0.85078534	frozenset({'Pubs', 'Sushi Bars'})					
14	0.84554974	frozenset({'Pubs', 'American (New)', 'Sushi Bars'})					
15	0.82984293	frozenset({'Pubs', 'Vegetarian'})					
16	0.82460733	frozenset({'Pubs', 'American (New)', 'Vegetarian'})					
17	0.81937173	frozenset({'Sushi Bars', 'Wine Bars'})					
18	0.81937173	frozenset({'American (New)', 'Sushi Bars', 'Wine Bars'})					
19	0.80628272	frozenset({'Vegetarian', 'Wine Bars'})					
20	0.80628272	frozenset({'American (New)', 'Vegetarian', 'Wine Bars'})					
21	0.79842932	frozenset({'Sushi Bars', 'Vegetarian'})					
22	0.79319372	frozenset({'Sushi Bars', 'American (New)', 'Vegetarian'})					
23	0.77748691	frozenset({'Pubs', 'Sushi Bars', 'Wine Bars'})					
24	0.77748691	frozenset({'Pubs', 'American (New)', 'Sushi Bars', 'Wine Bars'})					
25	0.76439791	frozenset({'Pubs', 'Vegetarian', 'Wine Bars'})					
26	0.76439791	frozenset({'Pubs', 'American (New)', 'Vegetarian', 'Wine Bars'})					
27	0.7486911	frozenset({'Pubs', 'Sushi Bars', 'Vegetarian'})					
28	0.7434555	frozenset({'Sushi Bars', 'American (New)', 'Pubs', 'Vegetarian'})					
29	0.7382199	frozenset({'Sushi Bars', 'Vegetarian', 'Wine Bars'})					
30	0.7382199	frozenset({'Sushi Bars', 'American (New)', 'Vegetarian', 'Wine Bars'})					
31	0.72513089	frozenset({'Breweries'})					
32	0.72251309	frozenset({'Breweries', 'American (New)'})					
33	0.70157068	frozenset({'Breweries', 'Pubs'})					
34	0.69895288	frozenset({'Breweries', 'American (New)', 'Pubs'})					
35	0.69633508	frozenset({'Pubs', 'Sushi Bars', 'Vegetarian', 'Wine Bars'})					

antecedents	consequents	anteced	consequi	support	confide	lift	leverage	convicti
frozenset({'African'})	frozenset({'Ethiopian'})	0.188482	0.230366	0.172775	0.916667	3.979167	0.129355	9.235602
frozenset({'Ethiopian'})	frozenset({'African'})	0.230366	0.188482	0.172775	0.75	3.979167	0.129355	3.246073
frozenset({'African', 'Vegan'})	frozenset({'Ethiopian'})	0.183246	0.230366	0.167539	0.914286	3.968831	0.125326	8.979058
frozenset({'Vegan', 'Ethiopian'})	frozenset({'African'})	0.230366	0.183246	0.167539	0.727273	3.968831	0.125326	2.994764
frozenset({'African'})	frozenset({'Vegan', 'Ethiopian'})	0.188482	0.225131	0.167539	0.888889	3.94832	0.125106	6.973822
frozenset({'Ethiopian'})	frozenset({'African', 'Vegan'})	0.225131	0.188482	0.167539	0.744186	3.94832	0.125106	3.172299
frozenset({'African', 'American (New)'})	frozenset({'Ethiopian'})	0.167539	0.230366	0.151832	0.90625	3.933949	0.113237	8.209424
frozenset({'American (New)', 'Ethiopian'})	frozenset({'African'})	0.230366	0.167539	0.151832	0.659091	3.933949	0.113237	2.441885
frozenset({'African'})	frozenset({'American (New)', 'Ethiopian'})	0.183246	0.240838	0.172775	0.942857	3.914907	0.128642	13.28534
frozenset({'Ethiopian'})	frozenset({'African', 'American (New)'})	0.240838	0.183246	0.172775	0.717391	3.914907	0.128642	2.890052
frozenset({'African', 'Vegetarian'})	frozenset({'Ethiopian'})	0.17801	0.240838	0.167539	0.941176	3.907928	0.124668	12.90576
frozenset({'Vegetarian', 'Ethiopian'})	frozenset({'African'})	0.240838	0.17801	0.167539	0.695652	3.907928	0.124668	2.700823
frozenset({'African'})	frozenset({'Vegetarian', 'Ethiopian'})	0.162304	0.240838	0.151832	0.935484	3.884292	0.112744	11.76702
frozenset({'Ethiopian'})	frozenset({'African', 'Vegetarian'})	0.240838	0.162304	0.151832	0.630435	3.884292	0.112744	2.266708
frozenset({'African', 'Vegan', 'American (New)'})	frozenset({'Ethiopian'})	0.183246	0.235602	0.167539	0.914286	3.880635	0.124366	8.917976
frozenset({'Vegan', 'American (New)', 'Ethiopian'})	frozenset({'African'})	0.235602	0.183246	0.167539	0.711111	3.880635	0.124366	2.827225
frozenset({'African', 'Vegan'})	frozenset({'American (New)', 'Ethiopian'})	0.183246	0.21466	0.151832	0.828571	3.85993	0.112497	4.581152
frozenset({'African', 'American (New)'})	frozenset({'Vegan', 'Ethiopian'})	0.21466	0.183246	0.151832	0.707317	3.85993	0.112497	2.790576
frozenset({'Vegan', 'Ethiopian'})	frozenset({'African', 'American (New)'})	0.188482	0.209424	0.151832	0.805556	3.846528	0.11236	4.065819
frozenset({'American (New)', 'Ethiopian'})	frozenset({'African', 'Vegan'})	0.209424	0.188482	0.151832	0.725	3.846528	0.11236	2.950976
frozenset({'African'})	frozenset({'Vegan', 'American (New)', 'Ethiopian'})	0.188482	0.240838	0.172775	0.916667	3.806159	0.127381	9.109948
frozenset({'Ethiopian'})	frozenset({'African', 'Vegan', 'American (New)'})	0.240838	0.188482	0.172775	0.717391	3.806159	0.127381	2.871526
frozenset({'African', 'Vegan', 'Vegetarian'})	frozenset({'Ethiopian'})	0.183246	0.240838	0.167539	0.914286	3.796273	0.123407	8.856894
frozenset({'Vegan', 'Vegetarian', 'Ethiopian'})	frozenset({'African'})	0.240838	0.183246	0.167539	0.695652	3.796273	0.123407	2.68362
frozenset({'African', 'Vegan'})	frozenset({'Vegetarian', 'Ethiopian'})	0.188482	0.235602	0.167539	0.888889	3.77284	0.123133	6.879581
frozenset({'African', 'Vegetarian'})	frozenset({'Vegan', 'Ethiopian'})	0.235602	0.188482	0.167539	0.711111	3.77284	0.123133	2.809102
frozenset({'Vegan', 'Ethiopian'})	frozenset({'African', 'Vegetarian'})	0.167539	0.240838	0.151832	0.90625	3.762908	0.111483	8.097731
frozenset({'Vegetarian', 'Ethiopian'})	frozenset({'African', 'Vegan'})	0.240838	0.167539	0.151832	0.630435	3.762908	0.111483	2.252541
frozenset({'African'})	frozenset({'Vegan', 'Vegetarian', 'Ethiopian'})	0.188482	0.21466	0.151832	0.805556	3.75271	0.111373	4.038893
frozenset({'Ethiopian'})	frozenset({'African', 'Vegan', 'Vegetarian'})	0.21466	0.188482	0.151832	0.707317	3.75271	0.111373	2.772688

1. {African->Ethiopian}과 {Ethiopian->African}의 lift값은 동일하지만 confidence의 값은 {African->Ethiopian}가 더 크다.

-> 포함관계를 유추할 수 있다.

2. African, American(New), Ethiopian, Vegan, Vegetarian의 다섯 항목의 조합 사이에 큰 양의 상관관계가 존재.

3. 위 다섯항목 조합들의 support 값은 낮다.

-> 출현 빈도는 낮지만, 한번 나타나면, 다른 항목이 같이 나오는 경우가 매우 많다.

4. Support값이 0.9이상으로 높았던 Pubs, Sushi Bars의 항목은 많이 출현하지만 lift 값이 3이상인 경우에는 나타나지 않은 것으로 보아, 다른 항목과의 연관성은 상대적으로 적다.