

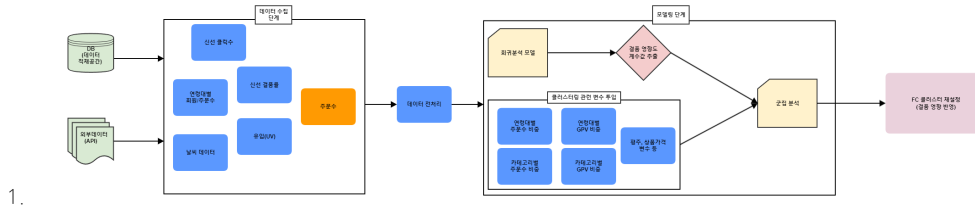
# 1.결품률 \* 주문수 영향도 계수 추정 모델링

## \*기획배경

- 결품률이 주문수에 주요하게 영향을 주는 PPC 파악
  - 결품률 주문 영향도에 대한 회귀 계수 값 산출 (회귀분석 모델)
  - 회귀 계수 값 산출 후 PPC 별 클러스터링을 통해 결품률의 영향도가 큰 PPC 산출 (클러스터링 분석\_은형님)

## \*FC 클러스터링 재설정 프로세스 로직

FC 클러스터링 모델 워크플로우



## \*활용 데이터

### \*데이터 예시

- 1) 대상 품목: 일자 \* PPC 데이터
- 2) 대상 기간: 2023-06-01 ~ 2023-12-31(12월 결품 Test 일정 & 검증데이터 셋 기간 고려하여 해당기간 활용)
- 3) 활용 변수: 주문수 예측에 영향을 주는 변수 고려

index	데이터 유형	컬럼명	컬럼 의미	데이터타입	비고
1	주문수	ord_dt	일자	datetime	
2		shop_nm	PPC명	varchar	
3		ord_cnt	주문수	integer	종속(예측)변수
4		ages_10_under_ord_cnt	10대이하 주문수	integer	
5		ages_20_ord_cnt	20대 주문수	integer	
6		ages_30_ord_cnt	30대 주문수	integer	
7		ages_40_ord_cnt	40대 주문수	integer	
8		ages_50_ord_cnt	50대 주문수	integer	
9		ages_60_over_ord_cnt	60대 이상 주문수	integer	
10	회원수	ages_10_under_mem_cnt	10대이하 회원수	integer	
11		ages_20_mem_cnt	20대 회원수	integer	

12		ages_30_mem_cnt	30대 회원수	integer	
13		ages_40_mem_cnt	40대 회원수	integer	
14		ages_50_mem_cnt	50대 회원수	integer	
15		ages_60_over_mem_cnt	60대 이상 회원수	integer	
16	클릭수	fruit_event	과일 클릭수	integer	
17		meat_event	정육/수산/계란 클릭수	integer	
18		veg_event	채소 클릭수	integer	
19	유입(UV)	UV	유입수	integer	
20		CVR	클릭률	float	
21	결품	lack_17	신선_17시 결품 상품수	integer	
22		lack_24	신선_24시 결품 상품수	integer	
23	날씨	temp_mean_divcd	평균기온 구분(6점 척도)	category	공공데이터 수집
24		rain_divcd	강수량구분	category	공공데이터 수집
25		wind_divcd	풍속구분	category	공공데이터 수집(결측값 다수로 산출 불가)
26	시간 정보	d	일	categorical	
27		week_day	요일	categorical	

4) 타겟 변수: 주문

\*학습데이터셋 예시

	ord_dt	shop_nm	ord_cnt	weekday	day	temp_mean_divcd	rain_divcd	lack_17	lack_24	fruit_event	...	ages_40_mem_cnt	ages_50_mem_cnt	ages_60_over_mem_cnt	ages_10_ord_mem_cnt	ages_20_ord_cnt	ages_30_ord_cnt	ages_40_ord_cnt	ages_50_ord_cnt	ages_60_over_ord_cnt	uv
0	2023-06-01	강서파곡점	637	3	1	3	0	210.0	340.0	837.0	...	212	86	29	8	301	430	216	89	32	11197
1	2023-06-01	성남점	374	3	1	3	0	276.0	373.0	436.0	...	108	54	19	7	223	236	111	55	20	6204
2	2023-06-01	용산한남점	391	3	1	3	0	303.0	406.0	428.0	...	118	42	11	1	169	299	119	43	11	3713
3	2023-06-01	대천유성점	352	3	1	3	0	275.0	357.0	460.0	...	86	37	5	17	276	203	87	37	5	7280
4	2023-06-01	인현청라점	357	3	1	3	1	283.0	373.0	396.0	...	128	39	15	8	159	256	132	40	16	4296
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
15162	2023-12-31	하남점	226	6	31	2	1	192.0	223.0	673.0	...	201	57	13	10	138	261	202	58	13	8042
15163	2023-12-31	수원점	748	6	31	2	0	168.0	293.0	1100.0	...	282	86	17	16	455	533	295	87	18	12629
15164	2023-12-31	용인점	310	6	31	2	0	209.0	246.0	907.0	...	293	122	14	11	165	318	301	124	17	10266
15165	2023-12-31	강서파곡점	631	6	31	2	0	118.0	210.0	1499.0	...	368	108	23	14	362	529	379	111	24	17140
15166	2023-12-31	은평점	486	6	31	2	0	197.0	258.0	1209.0	...	283	110	21	10	310	487	287	112	21	14155

feature\_dataset\_0207.csv

\*데이터 탐색

\*insight를 확인할 수 있는 주요 변수의 EDA 진행

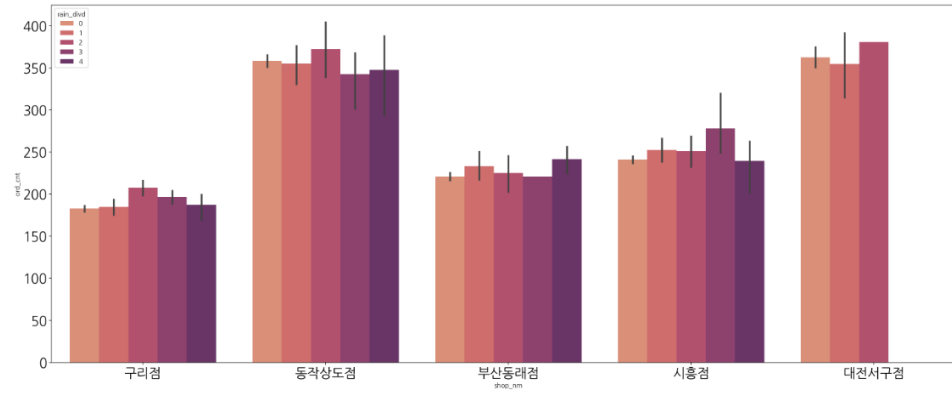
- 날씨 \* 주문수 데이터
  - 강수량 강도 구분
    - 강수량 < 5mm 이하
      - 5mm < 강수량 <= 20mm 이하
        - 20 < 강수량 <= 40mm 이하

3: 40 < 강수량 <= 60mm 이하

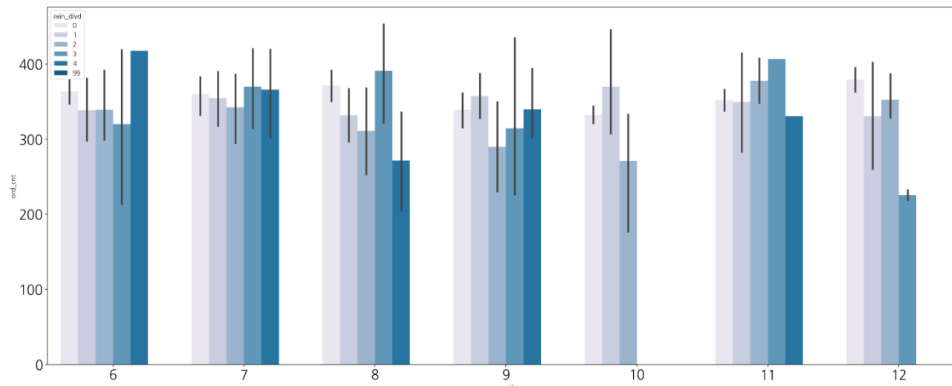
4: 60 <= 강수량

99: 알수없음

1-1) 결품대응 FC의 강수량 구분에 따른 주문수 변화



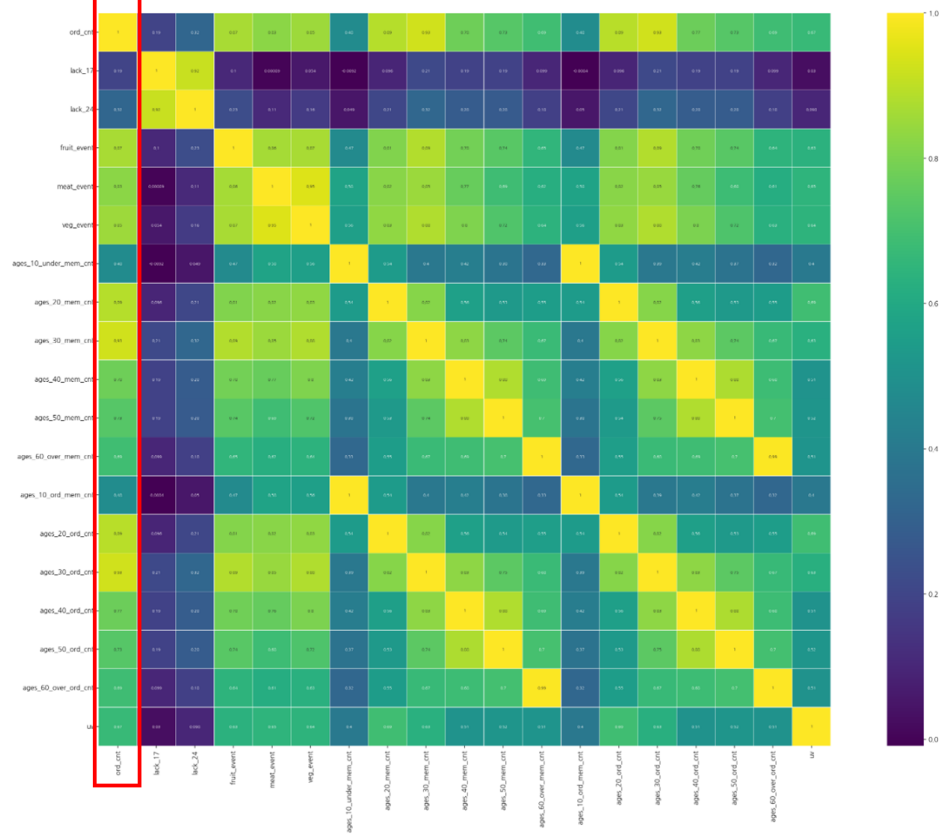
1-2) 월별 강수량 구분에 따른 주문수 변화



## 2. 결품 \* 주문수 데이터

## \*데이터 전처리

### 1. 상관관계 변수 제거



- 예측변수(주문수)와 독립변수(그외변수)의 상관관계가 전반적으로 높은 편이지만, 모델의 예측력 확보를 위해 주문수와 예측변수의 상관관계가 0.9이상인 케이스 제거 (제거 변수: 30대 회원수, 30대 고객 주문수)

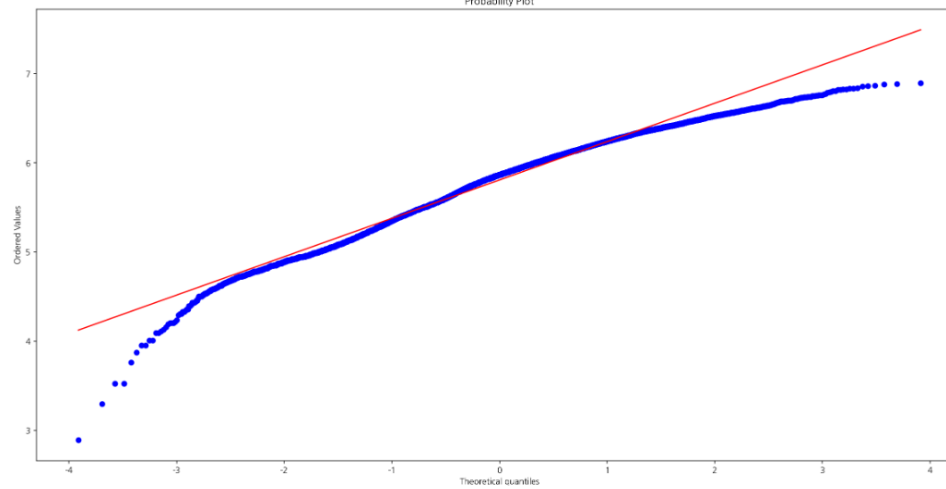
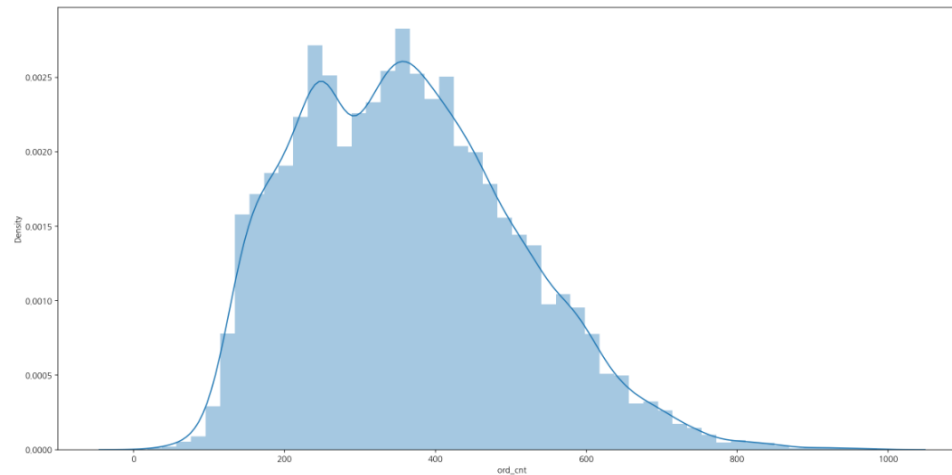
### 2. 다중공선성 변수 제거(VIF)

- 변수간에 상관관계를 볼수 있는 척도인 다중공선성이 높은 변수 제거

	<b>vif_factor</b>	<b>features</b>
0	0.736435	lack_17
1	0.851172	lack_24
2	0.305634	fruit_event
3	0.710719	meat_event
4	0.899930	veg_event
5	7.551757	ages_10_under_mem_cnt
6	113.739141	ages_20_mem_cnt
7	72.210545	ages_40_mem_cnt
8	20.675511	ages_50_mem_cnt
9	4.324017	ages_60_over_mem_cnt
10	7.530342	ages_10_ord_mem_cnt
11	113.814063	ages_20_ord_cnt
12	70.935788	ages_40_ord_cnt
13	20.344620	ages_50_ord_cnt
14	4.205402	ages_60_over_ord_cnt
15	0.123902	uv

- 독립변수들간의 상관관계도 높은 편으로 10이상의 변수들은 제거해야하지만, 모델의 예측력 확보를 위해 VIF 수치 100이상의 변수들 제거 (제거 변수 20대 고객수, 20대 고객 주문수)

### 3. 주문수 정규성 검증



- 회귀분석의 기본가정인 정규성 검증 진행
- P-value:  $6.4392e-42$
- Q-Q plot, distplot 시각화와 p-value 값이 아주작은 값을 가지는것으로 보아 주문수(예측 변수)가 정규성을 띤다고 볼 수 있음

## \*모델링 & 결과

### 1. 학습 방법

- 다중 회귀 예측 모델 활용  
blocked URLblocked URL

- 독립 변수에 예측변수의 변화량을 통해 예측 변수의 값을 예측하는 모형 (선형가정)
- b. FC별 모델 학습 진행

Dep. Variable:	ord_cnt	R-squared:	0.754
Model:	OLS	Adj. R-squared:	0.729
Method:	Least Squares	F-statistic:	30.28
Date:	Wed, 07 Feb 2024	Prob (F-statistic):	1.04e-44
Time:	14:22:20	Log-Likelihood:	-832.90
No. Observations:	197	AIC:	1704.
Df Residuals:	178	BIC:	1766.
Df Model:	18		
Covariance Type:	nonrobust		

OLS Regression Results

	coef	std err	t	P> t	[0.025	0.975]
const	168.1860	14.964	11.239	0.000	138.656	197.716
temp_mean_divd	-10.4100	2.062	-5.049	0.000	-14.479	-6.341
rain_divd	-0.1903	1.640	-0.116	0.908	-3.427	3.047
weekday	-1.6971	0.967	-1.754	0.081	-3.606	0.212
day	0.0158	0.153	0.103	0.918	-0.287	0.318
lack_17	-0.0805	0.052	-1.539	0.126	-0.184	0.023
lack_24	-0.0032	0.041	-0.077	0.938	-0.084	0.077
fruit_event	0.0185	0.021	0.891	0.374	-0.022	0.059
meat_event	0.0204	0.012	1.639	0.103	-0.004	0.045
veg_event	0.0412	0.025	1.656	0.099	-0.008	0.090
ages_10_under_mem_cnt	0.5565	3.597	0.155	0.877	-6.541	7.654
ages_40_mem_cnt	0.4319	0.990	0.436	0.663	-1.522	2.386
ages_50_mem_cnt	1.9486	1.899	1.026	0.306	-1.798	5.696
ages_60_over_mem_cnt	-2.3017	4.283	-0.537	0.592	-10.753	6.149
ages_10_ord_mem_cnt	1.0254	3.459	0.296	0.767	-5.800	7.850
ages_40_ord_cnt	0.1284	0.967	0.133	0.895	-1.780	2.037
ages_50_ord_cnt	-1.6323	1.837	-0.888	0.376	-5.258	1.994
ages_60_over_ord_cnt	2.2177	4.104	0.540	0.590	-5.881	10.316
uv	0.0035	0.002	2.011	0.046	6.57e-05	0.007

Omnibus:	0.928	Durbin-Watson:	2.151
Prob(Omnibus):	0.629	Jarque-Bera (JB):	1.037
Skew:	-0.118	Prob(JB):	0.596
Kurtosis:	2.734	Cond. No.	6.48e+04

- 각 FC별 학습결과 산출후 결품 영향도 계수 (lack\_17, lack\_24) 값 저장



## 2. 모델 학습 결과

평가지표	R_Squared(Train)	R2_Score(Test)
MAE	0.76	0.46

\*회귀 예측 모형에서 모델의 성능을 평가하는 지표중 하나로  $0 \leq R2\_Score \leq 1$  값을 가지며 1에 근접할수록 모델의 예측정확도가 높다고 볼수있음

\* Test dataset 기간: 23/12/15 ~ 23/12/31

\*PPC(71개의) R squared 값에 평균값으로 산출

- a. 테스트 데이터 셋에 대한 R2 Score 값이 높지않으나, 주문수 자체를 예측하는 것이 아닌 결품률이 주문에 영향을 끼치는 계수 추정을 목표로 하기 때문에 학습이 어느정도 되었다는 가정하에 결품계수값 추출

## 3. 계수 값 추정 결과

순서	PPC명	17시_결품률계수	24시_결품률계수	R2_train	R2_test	17시_결품률계수 유의여부(90%수준)	24시_결품률계수 유의여부(90%수준)
1	은평점	- 0.643	0.594	0.828	0.587	1	0
2	동작상도점	- 0.634	0.537	0.740	0.664	1	0
3	청주홍덕점	- 0.515	0.562	0.715	0.372	1	0
4	대전중구점	- 0.505	0.234	0.914	0.656	1	0
5	천안점	- 0.495	0.501	0.723	0.393	1	0
6	대전유성점	- 0.491	0.369	0.906	0.219	1	0
7	시흥점	- 0.479	0.435	0.725	0.416	1	0
8	대구달서점	- 0.469	0.253	0.686	0.662	1	0
9	부산동래점	- 0.464	0.286	0.692	0.262	1	0
10	부천점	- 0.428	0.309	0.812	0.561	1	0
11	송파장신점	- 0.411	0.332	0.797	0.667	1	0
12	대구중구점	- 0.410	0.291	0.696	0.438	1	0
13	송파신천점	- 0.408	0.341	0.764	0.560	1	0
14	수원점	- 0.401	0.254	0.746	0.260	1	0
15	관악신림점	- 0.389	0.144	0.829	0.630	1	0
16	부산사하점	- 0.377	0.286	0.778	0.482	1	0
17	파주점	- 0.370	0.293	0.668	0.589	1	0
18	중구명동점	- 0.363	0.312	0.743	0.457	1	0
19	영등포신길점	- 0.355	0.243	0.832	0.622	1	0
20	대구수성점	- 0.342	0.116	0.748	0.394	1	0
21	창원점	- 0.326	0.346	0.743	0.479	1	0
22	울산남구점	- 0.326	0.235	0.742	0.423	1	0
23	부산진구점	- 0.314	0.171	0.699	0.487	1	0
24	송파문정점	- 0.312	0.146	0.807	0.647	1	0
25	강북미아점	- 0.306	0.118	0.783	0.509	1	0
26	분당점	- 0.297	0.139	0.814	0.547	1	0
27	안산점	- 0.291	0.137	0.691	0.595	1	0
28	동대문제기점	- 0.273	0.256	0.806	0.535	1	0
29	인천청라점	- 0.268	0.150	0.704	0.210	1	0
30	강동천호점	- 0.267	0.202	0.821	0.389	1	0
31	구로신도림점	- 0.264	0.110	0.788	0.293	1	0
32	홍인점	- 0.261	0.103	0.812	0.315	1	0
33	안양점	- 0.255	0.075	0.718	0.668	1	1
34	강서가양점	- 0.250	0.103	0.777	0.549	1	0
35	마포홍대점	- 0.250	0.204	0.748	0.420	1	0
36	노원중계점	- 0.247	0.172	0.778	0.364	1	0
37	평택점	- 0.245	0.065	0.568	0.395	1	1
38	의정부점	- 0.243	0.089	0.694	0.276	1	1
39	서초반포점	- 0.235	0.108	0.717	0.670	1	0
40	광진자양점	- 0.228	0.130	0.759	0.561	1	0
41	서대문신촌점	- 0.218	0.130	0.763	0.594	1	0
42	강남양재점	- 0.215	0.037	0.800	0.688	1	1
43	도봉쌍문점	- 0.200	0.039	0.816	0.555	1	1
44	인천남부점	- 0.198	0.045	0.818	0.722	1	1
45	수원권선점	- 0.188	0.064	0.710	0.499	1	1
46	서초교대점	- 0.184	0.066	0.792	0.653	1	1
47	김해점	- 0.169	0.186	0.809	0.573	1	0
48	인천부평점	- 0.169	0.027	0.745	0.578	1	1
49	동탄점	- 0.157	0.057	0.769	0.021	1	1
50	김포점	- 0.153	0.062	0.558	0.545	1	1
51	강남압구정점	- 0.149	0.055	0.660	0.679	1	1
52	인천연수점	- 0.148	0.054	0.722	0.539	1	1
53	성동옥수점	- 0.146	0.036	0.778	0.506	1	1
54	강서화곡점	- 0.145	0.065	0.802	0.428	1	1

55	일산점	-	0.141	0.025	0.753	0.552	1	1
56	부산해운대점	-	0.139	0.029	0.697	0.243	1	1
57	강남논현점	-	0.138	0.039	0.825	0.634	1	1
58	군포점	-	0.136	0.146	0.695	0.384	1	0
59	용산한남점	-	0.121	0.052	0.772	0.560	1	1
60	구리점	-	0.110	0.061	0.619	0.177	1	1
61	성남점	-	0.105	0.041	0.819	0.132	1	1
62	중랑면목점	-	0.103	0.060	0.808	0.501	0	1
63	오산점	-	0.100	0.068	0.645	0.310	1	1
64	마포공덕점	-	0.097	0.049	0.800	0.305	1	1
65	하남점	-	0.081	0.003	0.754	0.347	0	1
66	서초내방점	-	0.056	0.019	0.829	0.463	0	1
67	인천동원점	-	0.050	0.034	0.882	0.372	0	1
68	대구북구점	-	0.049	0.047	0.580	0.315	0	1
69	금천독산점	-	0.049	0.015	0.798	0.311	0	1
70	대전서구점	-	0.000	0.072	0.919	0.515	0	1
71	원주점		0.031	0.013	0.783	0.793	0	1
72	청주상당점		0.034	0.025	0.558	0.310	0	1

- a. 17시 영향도 계수 값의 유의성은 전체중 89%로 대부분 PPC에서 17시 결품 상품수가 주문수에 유의한 영향을 끼치는 것으로 보임(유의수준 90% 이내)  
b. 24시 영향도 계수 값의 유의성은 전체중 45%로 17시 결품수 대비 유의성이 부족함. 즉 24시 결품이 주문수에 영향을 끼치는 PPC는 절반이하라고 볼 수 있음  
→ 17시 영향도 계수값을 클러스터링에 Feature로 사용하는것이 좋을 듯함  
c. 결품률이 주문수에 주요하게 영향을 주는 PPC 상위 5개는 은평점 > 동작 상도점 > 청주 흥덕점 > 대전 중구점 > 천안점 순임  
→ 동작상도, 대전 중구점등은 결품 대응 테스트 검토

output결과: [결품률\\_계수값\\_결과\\_0207.csv](#)

#### 4. 계수값 기반의 2차 군집화

- 다음과 같은 기준으로 PPC를 2차 군집화 함

계수 값 범위	RANK
$x \leq -0.6$	S
$-0.6 < x \leq -0.4$	A
$-0.4 < x \leq -0.25$	B
$-0.25 < x \leq -0.15$	C
$-0.15 \leq x$	D

- 결과

PPC	17시_결품률계수	P-val	Cluster	RANK	결품률 높은 신선 카테고리 1위			결품률 높은 신선 카테고리 2위		
은평점	-0.64257	0.000	수도권 장보기	S	B	채소	0.32	B	과일	0.24
통각상도점	-0.63380	0.000	서울 간식 장보기	S	B	채소	0.36	B	과일	0.25
청주흥덕점	-0.51547	0.000	지방 저성장	A	B	채소	0.36	A	과일	0.30
대전중구점	-0.50538	0.000	수도권 장보기	A	B	채소	0.30	B	과일	0.22
천안점	-0.49548	0.000	수도권 장보기	A	B	채소	0.32	B	과일	0.24
대전유성점	-0.49056	0.000	지방 저성장	A	B	채소	0.35	B	과일	0.27
시흥점	-0.47868	0.000	지방 저성장	A	B	채소	0.36	A	과일	0.29
대구달서점	-0.46939	0.000	지방 신규	A	S	과일	0.38	A	과일	0.25
부산동래점	-0.46397	0.000	지방 신규	A	S	과일	0.47	A	과일	0.28
부천점	-0.42779	0.000	수도권 장보기	A	B	채소	0.31	B	과일	0.23
송파잠실점	-0.41053	0.000	서울 프리미엄	A	B	채소	0.35	B	과일	0.23
대구중구점	-0.40997	0.000	지방 신규	A	S	과일	0.40	A	과일	0.27
송파신천점	-0.40809	0.000	서울 프리미엄	A	B	채소	0.40	B	과일	0.29
수원점	-0.40054	0.000	수도권 장보기	A	B	채소	0.31	B	과일	0.22
관악신림점	-0.38895	0.000	서울 간식 장보기	B	B	채소	0.33	B	과일	0.21
부산사하점	-0.37708	0.000	지방 신규	B	S	과일	0.45	A	과일	0.32
파주점	-0.37050	0.000	수도권 신선 장보기	B	B	채소	0.36	B	과일	0.29
충구명동점	-0.36318	0.000	서울 프리미엄	B	B	채소	0.36	B	과일	0.26
영등포신길점	-0.35525	0.000	서울 간식 장보기	B	B	채소	0.30	A	과일	0.19
대구수성점	-0.34232	0.000	지방 신규	B	S	과일	0.42	A	과일	0.28
창원점	-0.32630	0.000	지방 신규	B	S	과일	0.44	A	과일	0.28
울산남구점	-0.32605	0.001	지방 신규	B	S	과일	0.43	A	과일	0.28
부산진구점	-0.31435	0.008	지방 신규	B	S	과일	0.33	A	과일	0.24
송파문정점	-0.31207	0.000	서울 간식 장보기	B	B	채소	0.36	B	과일	0.26
강북미아점	-0.30569	0.000	수도권 장보기	B	B	채소	0.32	B	과일	0.22
분당점	-0.29669	0.000	수도권 장보기	B	B	채소	0.39	A	과일	0.30
안산점	-0.29114	0.000	수도권 장보기	B	B	채소	0.30	B	과일	0.23
동대문제기점	-0.27283	0.001	수도권 장보기	B	B	채소	0.29	B	과일	0.22
인천청라점	-0.26783	0.000	경인 고빈도 고품주 신선 장보기	B	B	채소	0.33	B	과일	0.26
강동천호점	-0.26708	0.001	서울 간식 장보기	B	B	채소	0.33	B	과일	0.24
구로신도림점	-0.26378	0.000	서울 간식 장보기	B	B	채소	0.34	B	과일	0.26
용인점	-0.26074	0.000	수도권 신선 장보기	B	B	채소	0.33	A	과일	0.23
안양점	-0.25523	0.000	수도권 신선 장보기	B	B	채소	0.33	B	과일	0.25
강서가양점	-0.25016	0.002	서울 간식 장보기	B	B	채소	0.32	B	과일	0.23
마포홍대점	-0.24951	0.010	서울 간식 장보기	C	B	채소	0.32	B	과일	0.24
노원중계점	-0.24749	0.007	수도권 신선 장보기	C	B	채소	0.34	B	과일	0.24
평택점	-0.24473	0.000	경인 고빈도 고품주 신선 장보기	C	B	채소	0.31	B	과일	0.23
의정부점	-0.24255	0.000	수도권 신선 장보기	C	B	채소	0.33	B	과일	0.26
서초반포점	-0.23494	0.000	서울 프리미엄	C	B	채소	0.38	B	과일	0.27
광진자양점	-0.22786	0.004	서울 간식 장보기	C	B	채소	0.29	B	과일	0.19
서대문신촌점	-0.21789	0.006	서울 간식 장보기	C	B	채소	0.34	B	과일	0.25
강남양재점	-0.21487	0.001	서울 프리미엄	C	B	채소	0.33	A	과일	0.24
도봉쌍문점	-0.20040	0.000	수도권 장보기	C	B	채소	0.35	B	과일	0.24
인천남부점	-0.19848	0.006	수도권 장보기	C	B	채소	0.30	B	과일	0.22
수원권선점	-0.18775	0.002	경인 고빈도 고품주 신선 장보기	C	B	채소	0.36	B	과일	0.27
서초교대점	-0.18355	0.001	서울 프리미엄	C	B	채소	0.36	B	과일	0.26
김해점	-0.16890	0.008	지방 신규	C	S	과일	0.48	B	채소	0.31
인천부평점	-0.16863	0.011	수도권 장보기	C	B	채소	0.30	B	과일	0.22
동탄점	-0.15678	0.020	수도권 신선 장보기	C	B	채소	0.33	A	과일	0.24
김포점	-0.15258	0.001	경인 고빈도 고품주 신선 장보기	C	B	채소	0.37	B	과일	0.28
강남압구정점	-0.14939	0.000	서울 프리미엄	D	B	채소	0.37	B	과일	0.27
인천연수점	-0.14785	0.027	수도권 신선 장보기	D	B	채소	0.36	B	과일	0.28
성동옥수점	-0.14575	0.014	서울 간식 장보기	D	B	채소	0.34	B	과일	0.25
강서화곡점	-0.14527	0.017	수도권 장보기	D	B	채소	0.28	B	과일	0.19
일산점	-0.14088	0.001	수도권 장보기	D	B	채소	0.39	A	과일	0.30
부산해운대점	-0.13896	0.017	지방 신규	D	S	과일	0.44	A	과일	0.29
강남논현점	-0.13819	0.046	서울 간식 장보기	D	B	채소	0.31	B	과일	0.22
군포점	-0.13609	0.010	수도권 신선 장보기	D	B	채소	0.40	B	과일	0.34
용산한남점	-0.12109	0.031	서울 프리미엄	D	B	채소	0.33	A	과일	0.24
구리점	-0.11015	0.050	수도권 신선 장보기	D	B	채소	0.37	B	과일	0.34
성남점	-0.10453	0.064	수도권 신선 장보기	D	B	채소	0.32	B	과일	0.25
중랑면목점	-0.10271	0.129	서울 간식 장보기	D	B	채소	0.31	B	과일	0.24
오산점	-0.09978	0.055	수도권 장보기	D	B	채소	0.34	B	과일	0.27
마포공덕점	-0.09713	0.069	서울 간식 장보기	D	B	채소	0.33	B	과일	0.23
하남점	-0.08053	0.126	수도권 신선 장보기	D	B	채소	0.35	B	과일	0.28
서초내방점	-0.05569	0.298	서울 간식 장보기	D	B	채소	0.35	B	과일	0.25
인천용현점	-0.05047	0.159	지방 저성장	D	B	채소	0.41	A	과일	0.31
대구북구점	-0.04927	0.577	지방 신규	D	S	과일	0.45	B	채소	0.31

금전독산점	-0.04889	0.531	서울 간식 장보기	D	B	채소	0.34	B	과일	0.25
대전서구점	-0.00015	0.998	수도권 장보기	D	B	채소	0.36	B	과일	0.27
원주점	0.03116	0.462	지방 저성장	D	B	채소	0.42	S	과일	0.37
청주상당점	0.03374	0.656	지방 저성장	D	B	채소	0.36	B	과일	0.30

서울 간식 장보기	고객수와 주문수가 높고, 기존고객의 비중이 높은 서울 지역내 FC. 30대 고객이 많으며 간식류와 가공 식품 구매가 많음. 자체 시간대의 주문수 비중이 높은 편.
서울 프리마켓	주문 빈도가 가장 높고, 기존 고객의 비중이 높은 서울 강남/송파 지역의 FC. 50대 이상 고연령층의 비중이 상대적으로 높음. 평주가 모든 클러스터로 가장 높으며 결품률이 가장 높은 클러스터로 관리가 필요함.
수도권 장보기	고객수와 주문수가 가장 많고, 기존고객수의 비중이 가장 높은 서울,경기, 인천 지역내 FC. 가공식품에 대한 구매가 높음. 단말대 비중은 비교적 고르게 나타남.
수도권 신선 장보기	정육수신/계란, 채소 등 신선 장보기 비중이 높은 서울, 경기 지역의 FC. 40~50대 고객의 비중이 상대적으로 높음.
경인 고빈도 고향주 신선 장보기	기존 고객의 주문 빈도가 가장 높고, 30~40대 중심의 고객의 비중이 높음. 평간 주문금액이 가장 높은 경인 지역 내 FC. 신선 카테고리에 대한 선호가 높으나 신선 카테고리외의 결품을 역시 높은편으로 관리가 필요함.
지방 신규	주문수, 고객수가 적으며 신규 고객의 비중이 높음. 10~20대, 60대의 비중이 높은 경상 지역 내 FC. 생활용품/잡화의 구매 비중이 높은편이며 결품률은 낮은 편.
지방 저성장	주문수, 고객수가 적으며 신규 고객의 비중이 높음. 10~20대 고객의 비중이 높은 대전, 충청 중심의 FC로 생활가전 및 즉석식품의 구매비중이 높음.

서울 간식 장보기	서울 프리마켓	수도권 장보기	수도권 신선 장보기	경인 고빈도 고향주 신선 장보기	지방 신규	지방 저성장
1	2	3	4	5	6	7
강남논현점	강남인구원점	강북미아점	구리점	김포점	김해점	대전유성점
강동천호점	강남장재점	강서화곡점	군포점	수원관산점	대구달서점	시흥점
강서가양점	서초교대점	대전서구점	노원중계점	인천청라점	대구북구점	원주점
관악신림점	서초반포점	대전중구점	동탄점	광복점	대구수성점	인천유현점
광진자양점	송파신천점	도봉청동점	성남점		대구중구점	창주상당점
구로신도림점	송파합성점	월대방제기점	안양점		부산동래점	창주용곡점
금천독산점	용산한남점	불안점	울안점		부산사하점	
동작상도점	중구명동점	분당점	의정부점		부산진구점	
마포공덕점		수원점	인천연수점		부산해운대점	
마포홍대점		안산점	파주점		울산남구점	
서대문신촌점		오산점	하남점		창원점	
서초내방점		은평점				
성동옥수점		인천남부점				
송파문정점		인천부평점				
영등포신길점		밀양점				
중앙안락점		천안점				
16	7	16	11	4	11	6

※ 다른 군집 대비 높은 수치를 보이는 변수들을 정리함.

Cluster	1	2	3	4	5	6	7
	서울 간식 장보기	서울 프리마켓	수도권 장보기	수도권 신선 장보기	경인 고빈도 고향주 신선 장보기	지방 신규	지방 저성장
고객수, 주문수	많음	가장 많음	가장 많음			적음	가장 적음
신규 고객수, 주문수	적음	가장 적음				가장 많음	많음
기존 고객수, 주문수	많음	가장 많음	가장 많음			가장 적음	적음
빈도	가장 높음	가장 높음	가장 높음		높음	가장 적음	적음
신규 고객 빈도	높음	높음	낮음	낮음		가장 낮음	낮음
기존 고객 빈도	30대	50대, 60대	40대, 50대	40대, 50대	30대, 40대	10대, 20대, 60대	10대, 20대
연령대	신규 낮음, 기존 높음	신규 가장 낮음, 기존 가장 높음			신규 가장 높음, 기존 낮음	신규 가장 높음, 기존 낮음	신규 높음, 기존 가장 낮음
시간대	21~24시	21~24시	1~8시	13~16시, 9~12시	13~16시, 9~12시	17~20시, 1~8시	17~20시, 1~8시
평균주문금액	높음	가장 높음	높음	가장 높음	가장 낮음	가장 낮음	낮음
신규고객 평균주문금액	가장 높음	가장 높음	가장 높음	가장 높음	가장 낮음	가장 낮음	낮음
기존고객 평균주문금액	가장 높음	가장 높음	가장 높음	가장 높음	가장 낮음	가장 낮음	낮음
지역	서울 중심	서울 강남, 송파 중심	서울 및 경기, 인천 중심	서울, 경기 중심	경기, 인천 중심	경상 중심	대전, 충청 중심
카테고리 (GPV 비중 top2)	- 간장/다이어트 - 과일 - 과자/사탕/초콜릿 - 라면/즉석면 - 생수/음료 - 아이스크림/얼음 - 부티/클렌징/선크림 - 편의점간식 선호	- 간장/다이어트 - 과일 - 과자/사탕/초콜릿 - 생수/음료 - 미역/즉석김 - 아이스크림/얼음 - 우유/유제품 - 부티/클렌징/선크림 - 편의점간식 선호	- 라면/즉석면 - 조리밥/소스/통조림	- 빵/제빵 - 요리/간주/반찬 - 우유/유제품 - 정육/수신/계란 - 채소	- 정육/수신/계란 - 조리밥/소스/통조림 - 채소	- 생활가전/용품/잡화	- 생활가전/용품/잡화 - 요리/간주/반찬
고단가 구매 카테고리	- 간장/다이어트 - 과자/사탕/초콜릿 - 우유/유제품 - 조리밥/소스/통조림 - 채소	- 간장/다이어트 - 과일 - 과자/사탕/초콜릿 - 빵/제빵 - 요리/간주/반찬 - 우유/유제품 - 조리밥/소스/통조림 - 채소 - 편의점간식	- 과일 - 라면/즉석면 - 시리얼	- 쌀/잡곡/면과 - 아이스크림/얼음 - 요리/간주/반찬 - 정육/수신/계란	- 부티/클렌징/선크림 - 빵/제빵 - 생수/음료/가리자 - 정육/수신/계란	- 라면/면/즉석면 - 생활가전/용품/잡화 - 빵/제빵/면과 - 아이스크림/얼음	- 부티/클렌징/선크림 - 생수/음료/가리자 - 생활가전/용품/잡화
결품률	높음		매우 높음			낮음	매우 낮음
결품률 높은 카테고리 (신선 중심)	- 우유/유제품	- 정육/수신/계란 - 채소	- 우유/유제품		- 과일 - 정육/수신/계란 - 채소	- 우유/유제품	- 과일
결품률 낮은 카테고리 (신선 중심)					- 과일 - 정육/수신/계란 - 채소		

분석 결과 상세 보기 :

- 은평점, 동작상도점의 경우 신선 카테고리 결품률이 주문수에 미치는 부정영향이 가장 큰 PPC로, 신선 카테고리의 결품을 관리가 다른 PPC 대비 매우 중요함
- PPC별도 결품률이 높은 신선카테고리는 상이하나, 다수의 PPC에서 B등급 채소, B등급 과일 카테고리의 결품률이 높게 나타나 해당 등급의 상품 결품을 개선이 필요해보임.
- 해당 결과를 기반으로, 결품을 개선 파일럿 PPC 테스트 진행시 클러스터간 유사성을 함께 고려하여 PPC를 선별할 수 있을것으로 보임.

