

# 주식상품 추천시스템을 위한 상품군집모델 및 성향분석모델 개발

# 목 차

- 1** 서론  
분석 배경 | 분석 목표
- 2** 데이터 소개  
데이터 설명
- 3** 상품군집모델  
전처리 | 모델링 | 분석 결과
- 4** 성향분석모델  
전처리 | 모델링 | 분석 결과
- 5** 결론  
추천시스템 | 기대효과 | 분석 한계

### | 프로젝트 배경



서론



데이터 소개



상품군집모델

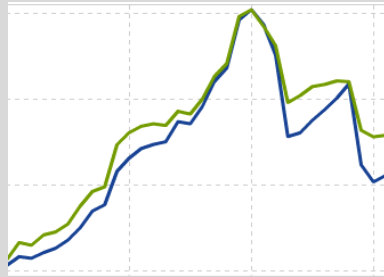


성향분석모델



결론

#### 단순한 상품분류 기준



단순한 상품분류 기준  
(상품종류, 종가, 산업 등)

#### 고객설문 기반 고객성향분류

“고객님의 투자성향을 어떤 유형이  
라고 판단하십니까?”

- 공격투자형
- 적극투자형
- 위험중립형
- 안정추구형
- 안정형

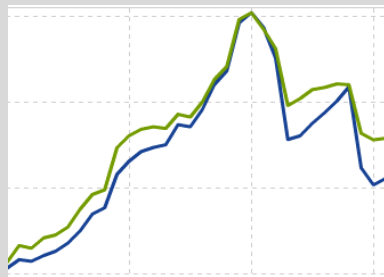
주관적 개인 설문응답 의존

#### 개인화되지 않은 추천시스템

유형	현도형	투자보존형
국내채권	현도형	최근 국내의 금리 변동에도 불구하고 안정적인 수익률을 실현하였으며, 단기채권 만일 및 리스크 관리로 변동성 적은 수익률을 기대
국내주식	현도형	시장 상승전망 및 대형주와 중소형주간의 과다폭소를 기대하며, 4차산업 관련주 등 시의적절한 투자 등으로 양호한 성과 실현
해외선진국주식	현도형	선진국 중심의 글로벌포트폴리오확충을투자자산다각화(투자)를 위하여
해외선진국주식	현도형	송지독자의 산업구조에 따라 1등 기업에만 집중투자하며, 해외주식자산 운용과 해외금융투자자의 과다지 농축의 경향을 통한 수익성 효과 기대
해외선진국주식	현도형	중국 내 시장 지배력이 높은 1등기업에 집중투자하며, 국내금융투자의 과다지농축의 경향을 통한 수익성 효과 기대

개인의 투자성향 관계없는  
일반적 추천시스템

## 단순한 상품분류 기준



단순한 상품분류 기준  
(상품종류, 종가, 산업 등)

## 고객설문 기반 고객성향분류

“고객님의 투자성향을 어떤 유형이  
라고 판단하십니까?”

- 공격투자형
- 적극투자형
- 위험중립형
- 안정추구형
- 안정형

주관적 개인 설문응답 의존

## 개인화되지 않은 추천시스템

유형	번호명	투자설명
국내채권	14-005(비대면)기업금융투자채권2년6월20년	최근 금리와 관리 갈등에서 볼 수 있듯, 안전적 수익자를 사형(외국)에 민간채권 편입 및 리스크 관리에 따른 변동성 수익을 기대
국내채권	14-001(비대면) Alt A등급중금리 증권투자채권2년6월20년	시장 상황변화 및 대환투자 중심상환으로 과다수익 기대. 4년4월 일관 투자 수익확보 투자 유도. 정당한 요구 시
채권신용국채주	한국투자글로벌채권비대면채권증권투자채권2년6월20년	신약국 중심의 글로벌채권 포트폴리오. 미국, 미국채권에 가세하는 것 인건비용과, 비대면채권 포트폴리오에 유망 시 A인
채권신용국채주	14-005(비대면)4년4월20년수출물리증권투자채권2년6월20년	수출물리 산업구조에 따라 1년 3개월에 집중투자, 4년4월20년 증권채권 비중 확대. 4년4월20년 수출물리 비중 확대. 4년4월20년 수출물리 비중 확대. 4년4월20년 수출물리 비중 확대.
채권신용국채주	14-005(비대면)수출물리증권투자채권2년6월20년	중국 내 1년 3개월에 집중투자, 4년4월20년 수출물리 비중 확대. 4년4월20년 수출물리 비중 확대. 4년4월20년 수출물리 비중 확대.

## 개인의 투자성향 관계없는 일반적 추천시스템

1

## 주식상품의 시계열 특성에 따른 상품군집

2

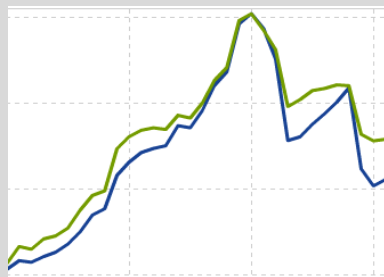
## 상품군집과 선호도에 따른 성향분석

3

## 고객 포트폴리오 성향에 따른 추천시스템 구현



## 단순한 상품분류 기준



단순한 상품분류 기준  
(상품종류, 종가, 산업 등)

## 고객설문 기반 고객성향분류

“고객님의 투자성향을 어떤 유형이  
라고 판단하십니까?”

- 공격투자형
- 적극투자형
- 위험중립형
- 안정추구형
- 안정형

주관적 개인 설문응답 의존

## 개인화되지 않은 추천시스템

유형	번호명	투자설명
국내채권	14-005(비대면)기업금융투자채권2년6월20년	최근 금리와 관리 갈등에서 볼 수 있듯, 안전적 수익자를 사형(외국)에 민간채권 편입 및 리스크 관리에 따른 변동성 수익률 기대
국내채권	14-001(비대면) Altman 신용등급을 충족하는 증권투자채권2년6월20년	시장 상황변화에 대한투자 중심채권으로 과거수익률 4.4% 이상 안정한 수익자획득 투자수익률 2.0% 이상 기대
채권신용국채주	한국투자글로벌채권비대면채권증권투자채권2년6월20년	신약국 중심의 글로벌채권으로 최근 수익성에서 기대되는 국가 신용평가에 따라, 비대면채권 편입에 따른 수익률 기대
채권신용국채주	14-005(비대면)4년12월15일수익률 2.0%증권투자채권2년6월20년	증권투자채권 신용국채에 따라 1년 12월간 집중투자하며, 수익률2.0% 이상수익자획득 투자수익률 2.0% 이상 기대
채권신용국채주	14-005(비대면)4년12월15일수익률 2.0%증권투자채권2년6월20년	증권 1년 12월간 집중투자하며 1년 12월간 수익자획득 투자수익률 2.0% 이상수익자획득 투자수익률 2.0% 이상 기대

## 개인의 투자성향 관계없는 일반적 추천시스템

1

## 주식상품의 시계열 특성에 따른 상품군집

시계열 특성 고려한 군집  
다양한 복합적인 요인 반영

2

## 상품군집과 선호도에 따른 성향분석

수치에 기반한  
객관적인 포트폴리오 성향분석

3

## 고객 포트폴리오 성향에 따른 추천시스템 구현

성향에 기반한 추천시스템으로  
고객만족도 증대



서론



데이터 소개



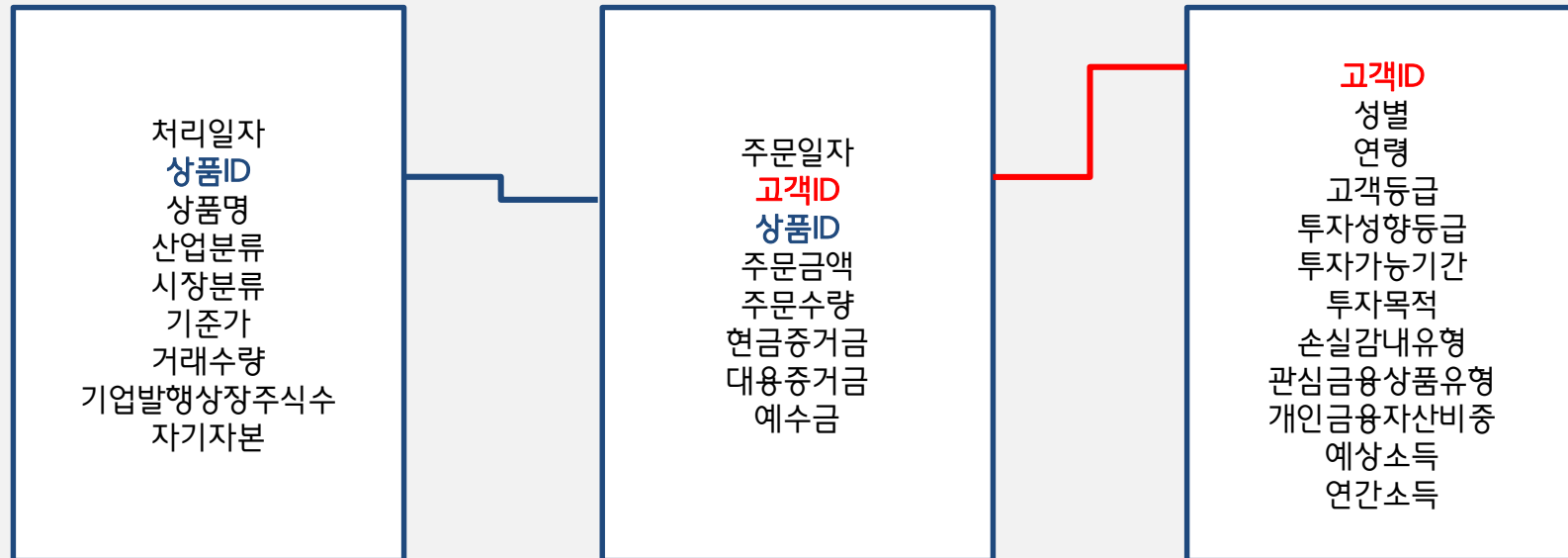
상품군집모델



성향분석모델



결론



	주식상품 마스터 테이블	주문내역 마스터 테이블	고객정보 마스터 테이블
목적	주식상품 군집 분석	선호도 기반 성향 분석	고객정보 확인 / 신규고객 성향 예측
테이블 크기	변수 - 8개 상품 수 - 6,963개	변수 - 6개 주문내역 - 17,191,031개	변수 - 11개 고객 수 - 58,389명
샘플링 기준	- 기간:170702-180122 - 한국거래소 운영 주식거래시장 - 주권, ETF, ETN, ELW - 거래정지 기간동안 일별내역 제거	- 기간:170702-180122 - 주식상품 마스터 테이블에 적재된 주식상품에 관한 주문내역	- 170702 이전 계좌개설 고객 - 개인투자자 (법인, 외국인 제외) - 정상계좌 보유 고객 - 최소 1번이라도 주문내역이 있는 고객 - 고객투자성향설문에 응답한 고객



서론



데이터 소개



상품군집모델



성향분석모델



결론

변수 탐색

이상치/결측치 처리

변수 선택

| 사전 조사

문헌/인터넷조사

전문가 의견조사

사례조사

| 변수 탐색

원장 DB 테이블 및 컬럼  
탐색

테이블 결합

변수 생성 및 가공

| 변수 목록

장세흐름	기준가의 평균	“이 주식상품의 평균 기준가는 얼마인가?”
	기준가 변화량의 분산	“이 주식상품의 기준가의 변동은 얼마나 한 방향으로 흐르며 추세적 인가?”
	기준가 변화량의 첨도	“이 주식상품의 기준가는 얼마나 변동이 큰가?”
	일별 기준가의 회귀직선 기울기	“이 주식상품의 기준가는 전반적으로 얼마나 상승/하락하는가?”
	상장주식수 대비 거래량 비율의 중앙값	“이 주식상품은 상장주식수 대비 얼마나 활발하게 거래되는가?”
업종모멘텀	일별 거래량의 회귀직선 기울기	“이 주식상품의 거래량은 전반적으로 얼마나 상승/하락하는가?”
	산업 시가총액의 회귀직선 기울기	“이 상장기업이 속한 산업에 대한 시가총액은 어떠한 상승/하락세를 보이는가?”
기업가치	PBR의 중간값	“이 상장기업의 현재가치는 얼마나 저/고평가 되었는가?”

- 고객이 주식상품 선택 시 고려하는 기준들을 측정가능 하도록 조작화 과정을 거침



서론



데이터 소개



상품군집모델



성향분석모델



결론

변수 탐색

이상치/결측치 처리

변수 선택

| 이상치 처리

- 액면분할
- 액면병합
- 자본감소

기업에 대해 기준가 전처리

| 결측치 처리

- 첩도  
추세에 대한 수치이기 때문에  
처리일자가 하루만 존재한다면  
결측값 가짐 -> 상품 제거



[ (주)오리온홀딩스 stock chart (액면분할 170707 상장일 중심) ]





서론



데이터 소개



상품군집모델



성향분석모델



결론

변수 탐색

이상치/결측치 처리

변수 선택

| 변수 간 상관분석

	상품ID	기준가 평균	기준가 기울기	거래수량 비율	기준가 분산	기준가 첨도	산업시총 기울기	PBR	거래수량 기울기	
	ROWNUM	MEAN	COEF_b1	QNT_RATIO	SPRC_VAR	SPRC_KUR	IDS_B1	PBR	QNT_B1	
상품ID	ROWNUM	1.00000000	0.052163419	0.014134522	0.09449681	0.126146525	0.09982257	0.087597673	0.16322174	-0.18138312
기준가 평균	MEAN	0.05216342	1.000000000	0.129200873	-0.01165984	-0.001264317	-0.03672198	0.032421342	0.27538067	0.02904488
기준가 기울기	COEF_b1	0.01413452	0.129200873	1.000000000	0.26583185	0.069993822	-0.19547979	-0.005775957	0.29169950	0.25171313
거래수량 비율	QNT_RATIO	0.09449681	-0.011659842	0.265831848	1.00000000	0.532893435	-0.03646104	-0.030644981	0.20818282	-0.04808596
기준가 분산	SPRC_VAR	0.12614653	-0.001264317	0.069993822	0.53289343	1.000000000	0.56278932	-0.015959899	0.25875668	0.12474728
기준가 첨도	SPRC_KUR	0.09982257	-0.036721984	-0.195479787	0.03646104	0.562789321	1.00000000	0.029753509	0.03267467	0.13879508
산업시총 기울기	IDS_B1	0.08759767	0.032421342	-0.005775957	-0.03064498	-0.015959899	0.02975351	1.000000000	0.01369978	-0.01511164
PBR	PBR	0.16322174	0.275380671	0.291699505	0.20818282	0.258756680	0.03267467	0.013699778	1.00000000	0.00346756
거래수량 기울기	QNT_B1	-0.18138312	0.029044881	0.251713132	-0.04808596	0.124747282	0.13879508	-0.015111635	0.00346756	1.00000000

[ 코스피 주권 시장 예시 ]  
ROWNUM은 상품ID를 나타내는 변수로서 무시 가능  
임계치 0.5 기준

기준가 분산 + 거래수량 비율

기준가 분산 + 기준가 첨도

코스피 시장 내 상품 군집 분석 시, “기준가 분산” 변수 제거  
즉, 시장 별 분석에 활용하는 변수가 상이할 수 있음



서론



데이터 소개



상품군집모델

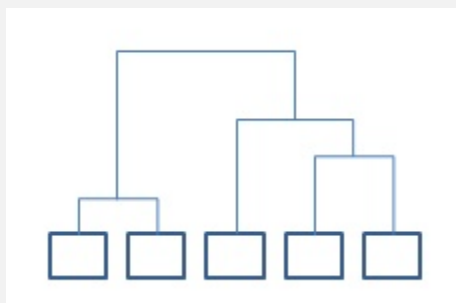


성향분석모델



결론

### | 군집분석모형 설명



- 계층적 군집분석

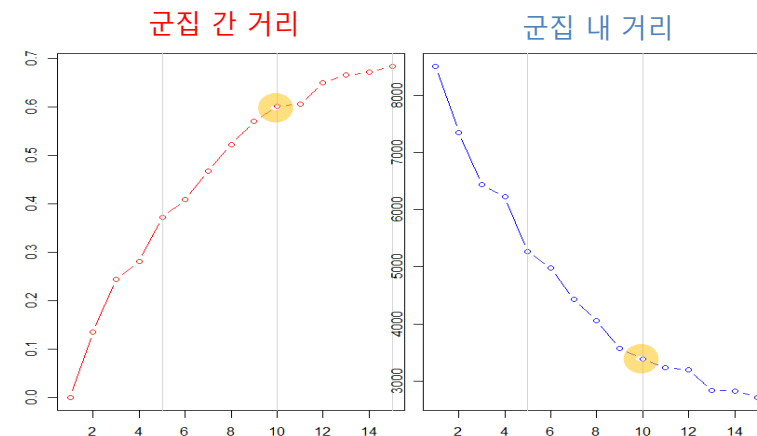


- 비계층적 군집분석

### K-평균 군집분석

: 군집의 개수와 초기 랜덤값만 지정하면, 알고리즘이 군집 내 임의의 중심점을 잡고 가까운 점들을 군집하고, 다시 중심점을 잡고 가까운 점들을 군집하는 과정을 반복하며 최적화시키는 군집 알고리즘

### | 군집 수 설정



- 적절한 군집 수는 10
- 너무 군집이 많은 경우 여러 군집이 비슷한 특성을 보일 수 있음
- 즉, 군집 간의 차이가 불명확할 수 있음

서론

데이터 소개

상품군집모델

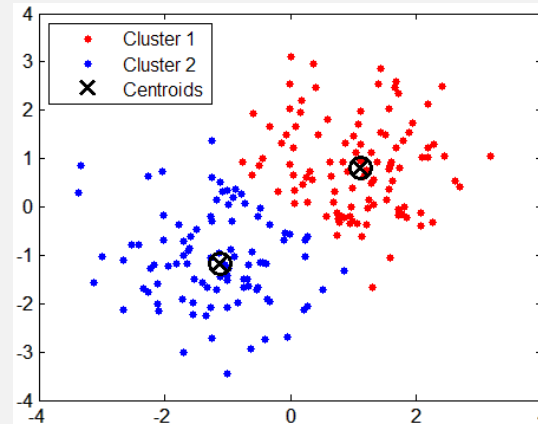
성향분석모델

결론

| 코스피 주권 시장 내 상품군집 결과

군집#	상품 수	별칭	대표 상품
10	6	황제우량주	삼성전자보통주, lg생활건강보통주
5	23	고평가우량주	네이버보통주, 오리온보통주
6	19	안정인기주	호텔신라보통주, Sk하이닉스보통주
1	118	안정성장주	엔씨소프트보통주, 현대로보틱스보통주
4	12	안정일반주	에스코보통주, Sk케미칼보통주
2	384	일반주	우리은행보통주, 하이트진로보통주
7	82	업종비인기주	태영건설우선주, 서흥보통주
8	58	일반저평가주	선진보통주, 현대약품보통주
9	134	위험테마주	삼호보통주, 대림산업1우선주
3	24	투기주의	아남전자보통주, 명문제약보통주

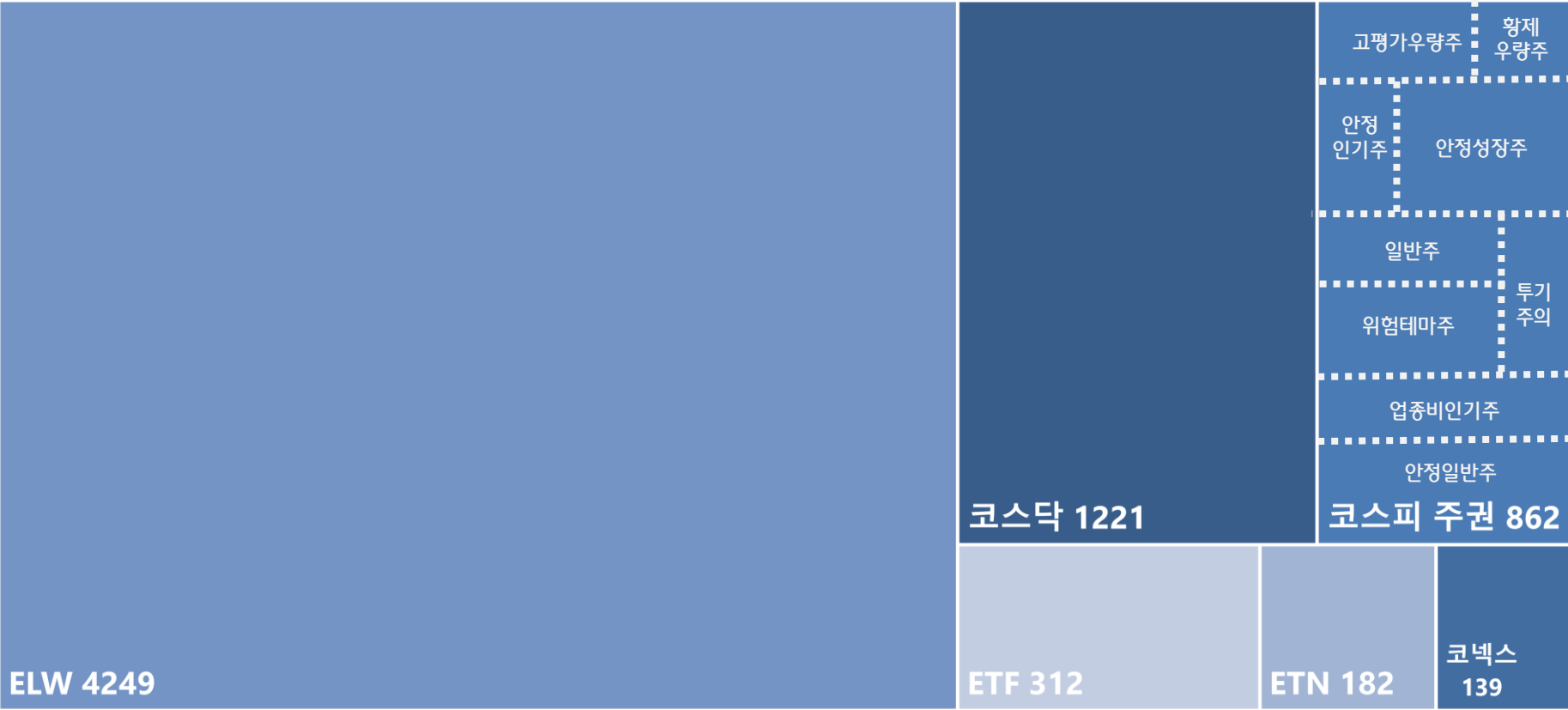
| 코스피 주권 상품군집 별 중앙값 (centroids)



	MEAN	COEF_b1	QNT_RATIO	SPRC_KUR	IDS_B1	PBR
1	0.303	1.013	0.066	-0.227	0.061	0.643
2	-0.166	-0.203	-0.254	-0.245	-0.127	-0.290
3	-0.232	1.151	4.100	0.117	-0.043	0.224
4	0.197	-0.743	0.123	6.556	0.315	0.585
5	0.528	0.493	0.081	-0.132	0.244	4.239
6	-0.140	1.104	0.441	0.063	-1.756	0.195
7	-0.163	-0.186	-0.302	-0.165	-1.718	-0.351
8	-0.089	-0.948	0.370	1.600	0.006	-0.308
9	-0.124	-0.107	-0.164	-0.272	1.526	-0.260
10	9.440	0.256	-0.415	-0.199	0.414	0.621

- 각 군집을 대표하는 센터가 각 변수에 대해 가지는 수치를 보고 군집 해석

| 코스피 주권 상품 군집분석 결과



총 6963개의 상품

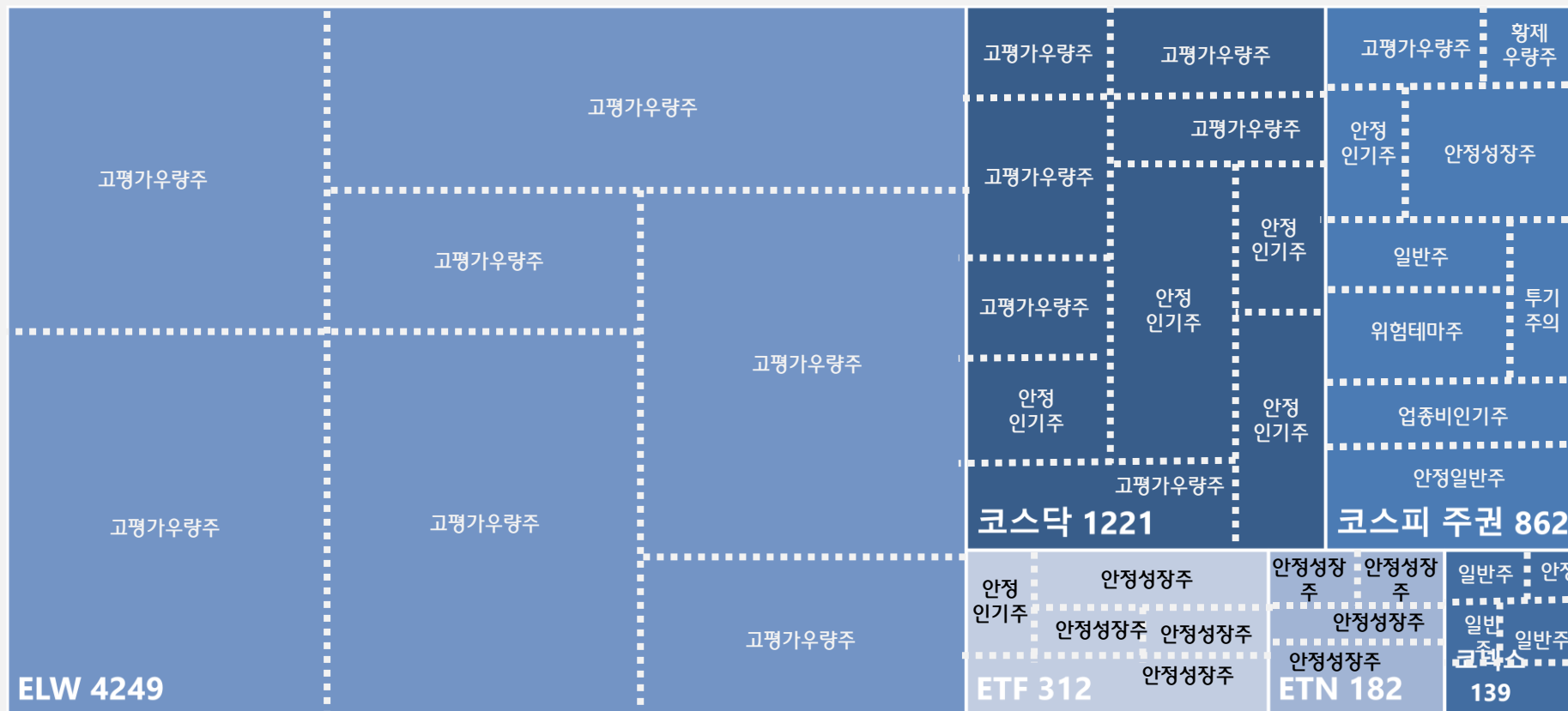
서론

데이터 소개

상품군집모델

성향분석모델

결론



총 6963개의 상품

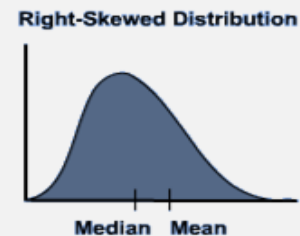


총 41개의 상품군집

### | 주문내역 테이블 가공



### | 주문내역 탐색적 분석



#### 1 고객 한 명당, 주문한 상품 개수

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	1.000	3.000	7.593	7.000	374.000

#### 2 고객 한 명당, 주문 빈도수

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.0	2.0	7.0	64.6	27.0	476977.0

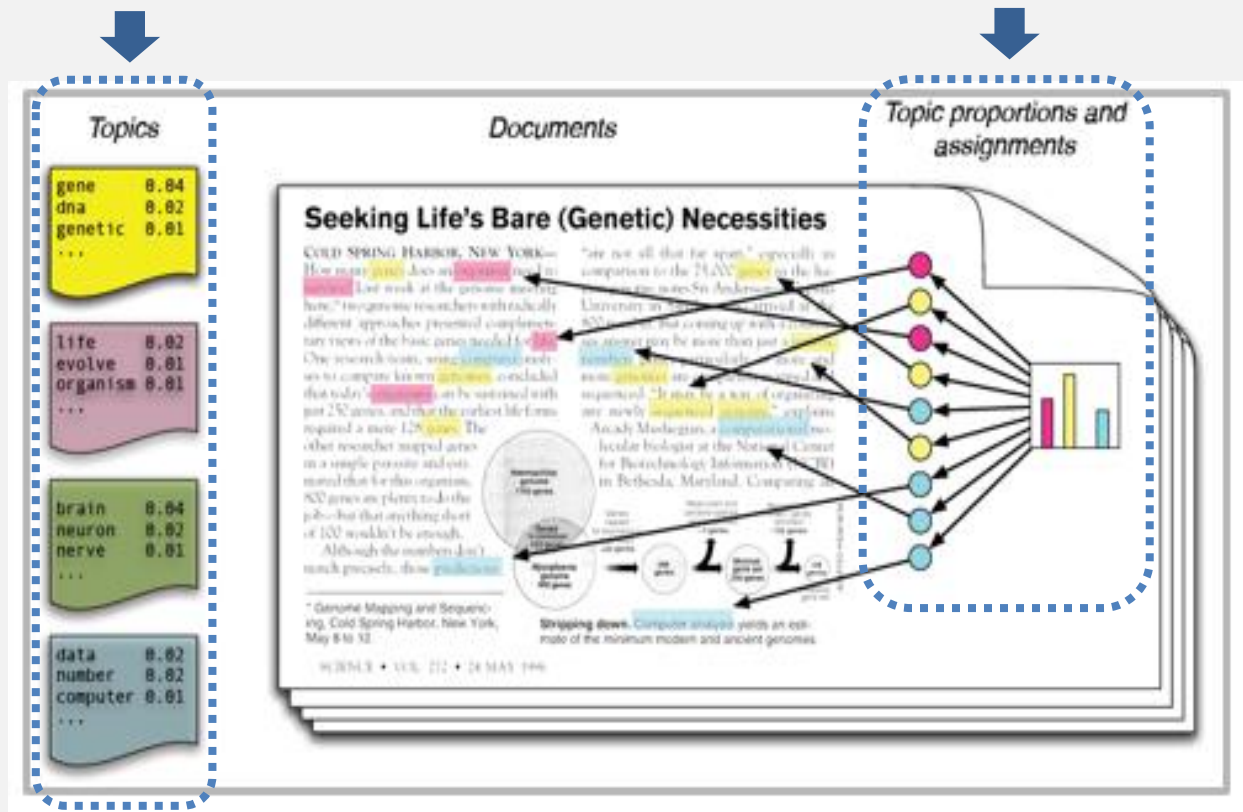
#### 3 상품 한 개당, 주문 빈도수 (모든 고객 기준)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	36	209	864	691	60664

| 성향분석모형 설명

- 어떤 단어들이 특정 주제 안에서 나올 확률
- 어떤 단어들이 함께 묶이는가

- 이 문서는 어떤 주제를 가질 확률이 가장 높은가



Topic Modeling LDA

: 특정 문서의 주제를 찾기 위한 모형



서론



데이터 소개



상품군집모델



성향분석모델



결론

### | 성향분석모형 설명



서론



데이터 소개



상품군집모델



성향분석모델



결론

상품

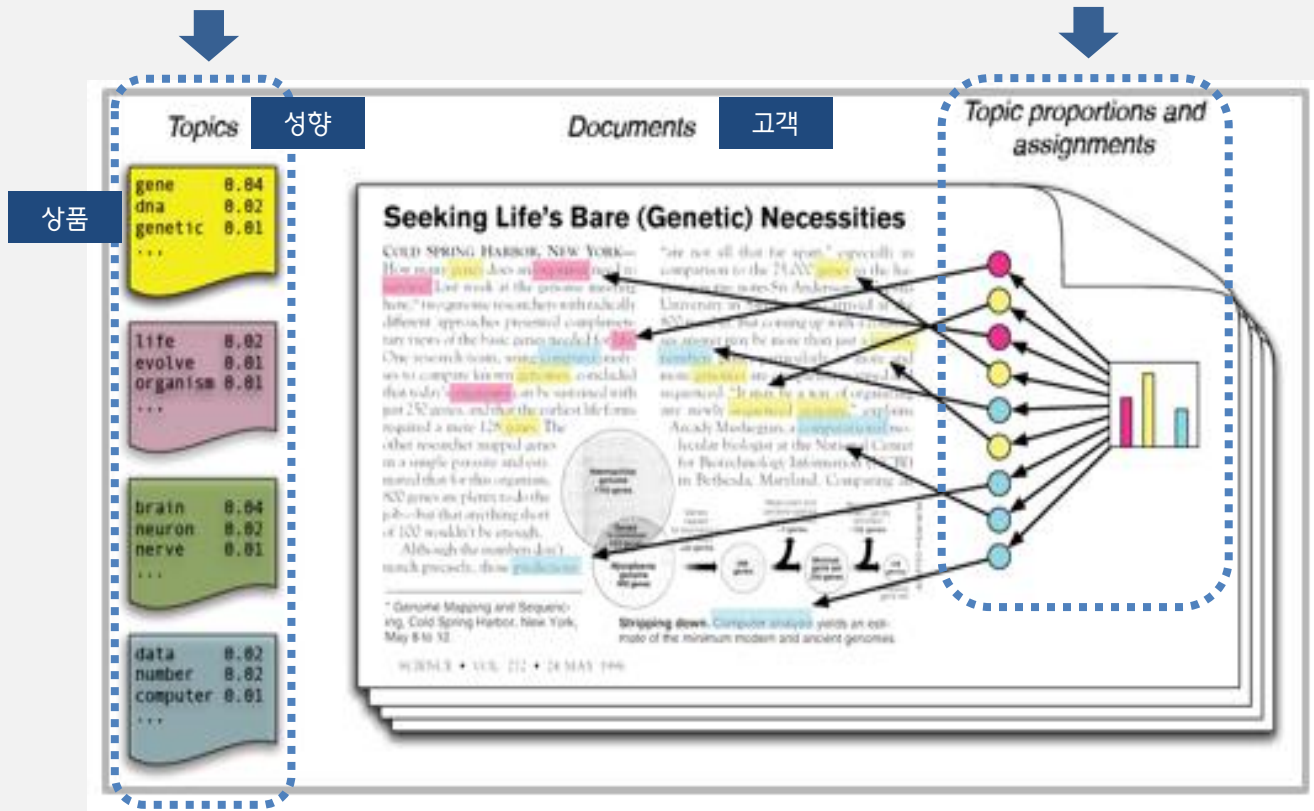
성향

- 어떤 단어들이 특정 주제-안에서 나올 확률
- 어떤 단어들이 함께 묶이는가

고객

성향

- 이 문서는 어떤 주제를 가질 확률이 가장 높은가



### Topic Modeling LDA

: 특정 문서의 주제를 찾기 위한 모형



특정 고객의 성향을 찾기 위한 모형

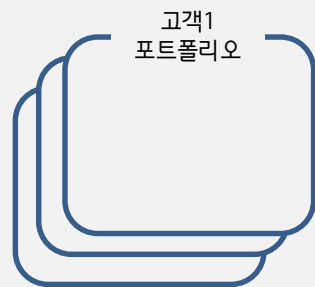
- 문서 = 고객
- 주제 = 성향
- 단어 = 상품



| Topic Modeling LDA 결과 테이블



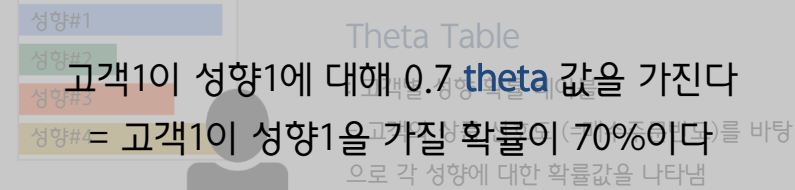
| Topic Modeling LDA 결과 테이블



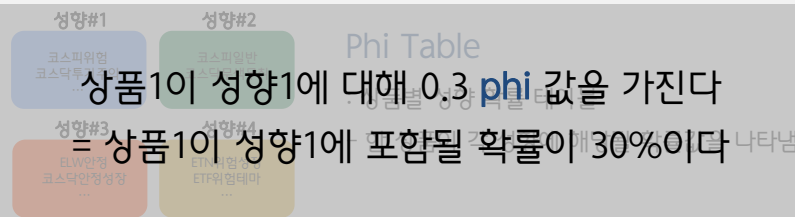
| Topic Modeling LDA 결과 테이블



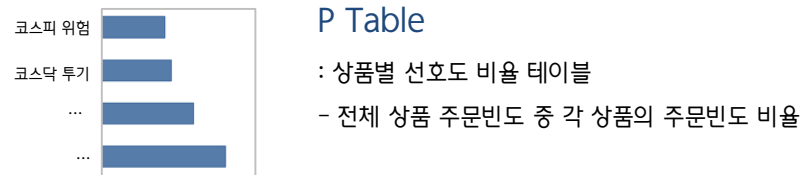
1



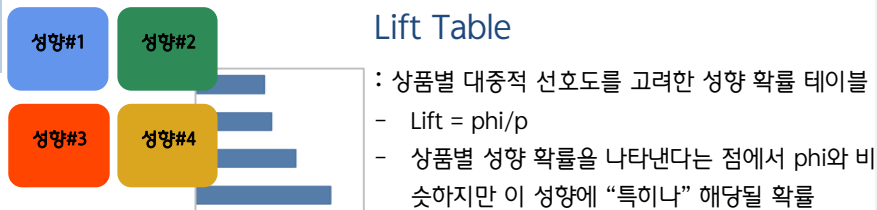
2



3



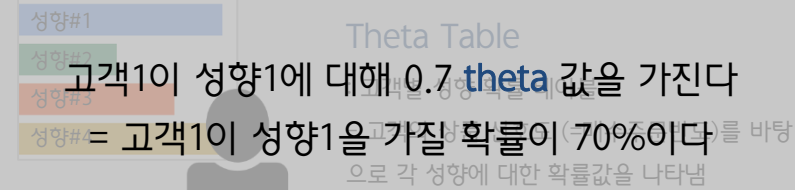
4



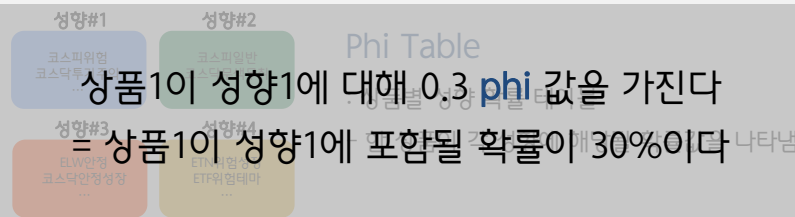
| Topic Modeling LDA 결과 테이블



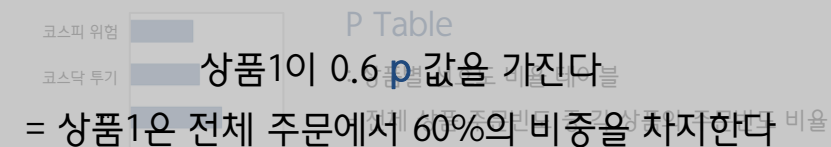
1



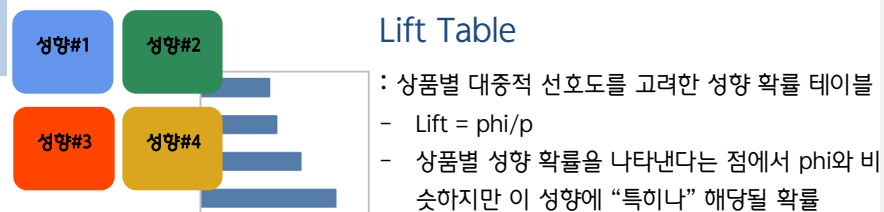
2



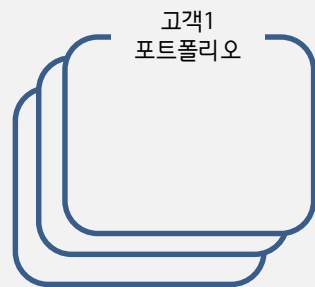
3



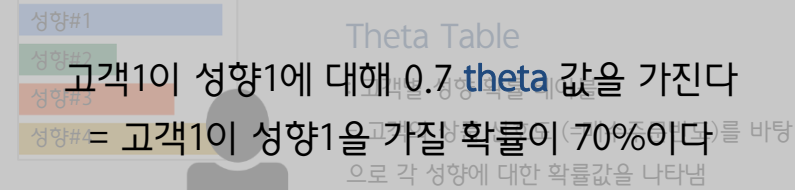
4



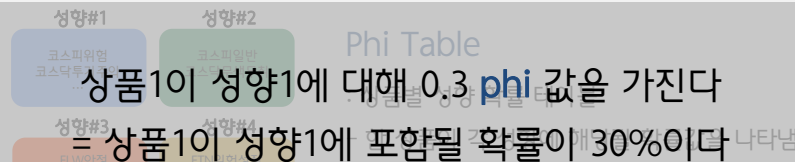
| Topic Modeling LDA 결과 테이블



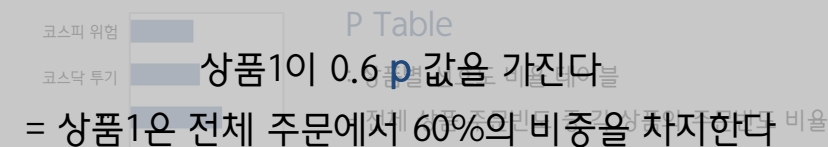
1



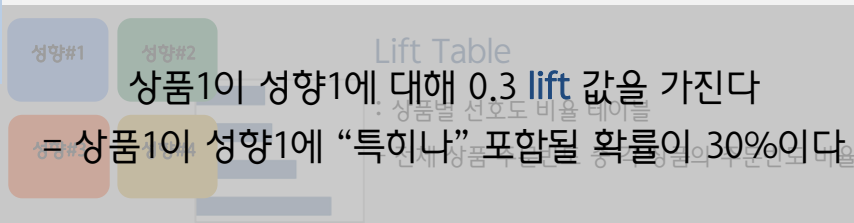
2



3



4



서론

데이터 소개

상품군집모델

성향분석모델

결론

### | Topic Modeling LDA 성향 해석

- 다음과 같은 4개의 기준으로 토픽별 상품 추출 후, 토픽별 성질을 해석
- 토픽별 성질을 기반으로 토픽에 이름 할당

1

Top Lift - Phi

- 리프트 기준 상위 10개의 상품 중 phi값이 높은 5개 추출
- 대중적 선호도는 낮지만 성향 내 선호도가 높음

2

Top Phi - Lift

- Phi값 기준 상위 10개의 상품 중 리프트가 높은 5개 추출
- 대중적 선호도가 높고 성향 내 선호도도 높음

3

Top Phi

- Phi값 기준 상위 10개의 상품
- 성향 내 선호도가 가장 높음

4

Bottom Lift

- 전체 상품의 phi값 기준 상위 10개 중 리프트 하위 5개 추출
- 대중적 선호도는 높지만 성향 내 선호도 낮음

성향#	Top lift- phi	Top phi- lift	Top phi	Bottom lift
1	27 19 10 22 33	22 10 19 16 11	11 19 9 16 13 10 7 40 18 22	28 3 30 40 5
2	1 13 25 17 16	1 17 13 25 19	1 13 11 19 4 20 25 17 34 2	3 32 33 34 35
3	19 11 10 1 20	10 1 11 19 16	19 11 10 16 13 18 1 14 7 9	32 33 34 35 36
4	11 10 19 1 20	1 19 16 14 13	11 10 19 16 18 9 13 14 1 7	28 29 3 30 31
5	19 16 10 14 17	10 16 19 2 11	19 11 16 10 13 4 18 17 20	18 26 28 29 3
6	24 1 13 10 17	1 10 17 13 19	11 19 1 13 34 4 10 18 17 20	28 3 30 32 33
7	19 1 10 16 17	10 1 19 16 13	19 11 13 1 10 18 16 4 7 2	28 3 30 31 32
8	41 1 4 17 10	4 41 1 19 25	31 13 19 41 1 4 11 25 20 2	28 3 30 31 32
9	19 16 10 1 17	10 19 16 1 11	19 11 16 13 10 18 7 2 4 1	28 29 3 31 32
10	1 13 10 17 25	13 19 11 16 1	19 11 1 13 10 18 16 4 34 17	28 30 32 33 34



서 론



데이터 소개



상품군집모델



성향분석모델



결 론

### | Topic Modeling LDA 성향 해석

- 다음과 같은 4개의 기준으로 토픽별 상품 추출 후, 토픽별 성질을 해석
- 토픽별 성질을 기반으로 토픽에 이름 할당

1

#### Top Lift - Phi

- 리프트 기준 상위 10개의 상품 중 phi값이 높은 5개 추출
- 대중적 선호도는 낮지만 성향 내 선호도가 높음

2

#### Top Phi - Lift

- Phi값 기준 상위 10개의 상품 중 리프트가 높은 5개 추출
- 대중적 선호도가 높고 성향 내 선호도도 높음

3

#### Top Phi

- Phi값 기준 상위 10개의 상품
- 성향 내 선호도가 가장 높음

4

#### Bottom Lift

- 전체 상품의 phi값 기준 상위 10개 중 리프트 하위 5개 추출
- 대중적 선호도는 높지만 성향 내 선호도 낮음

성향#	Top lift-phi	Top phi-lift	Top phi	Bottom lift	
1	27 19 10 22 33	22 10 19 16 11	11 19 9 16 13 10 7 40 18 22	28 3 30 40 5	의심형
2	1 13 25 17 16	1 17 13 25 19	1 13 11 19 4 20 25 17 34 2	3 32 33 34 35	YOLO
3	19 11 10 1 20	10 1 11 19 16	19 11 10 16 13 18 1 14 7 9	32 33 34 35 36	꿈꾸는소시민2
4	11 10 19 1 20	1 19 16 14 13	11 10 19 16 18 9 13 14 1 7	28 29 3 30 31	꿈꾸는소시민1
5	19 16 10 14 17	10 16 19 2 11	19 11 16 10 13 4 18 17 20	18 26 28 29 3	팔랑귀
6	24 1 13 10 17	1 10 17 13 19	11 19 1 13 34 4 10 18 17 20	28 3 30 32 33	무소
7	19 1 10 16 17	10 1 19 16 13	19 11 13 1 10 18 16 4 7 2	28 3 30 31 32	트렌드세터
8	41 1 4 17 10	4 41 1 19 25	31 13 19 41 1 4 11 25 20 2	28 3 30 31 32	가치투자자
9	19 16 10 1 17	10 19 16 1 11	19 11 16 13 10 18 7 2 4 1	28 29 3 31 32	소심한팔랑귀
10	1 13 10 17 25	13 19 11 16 1	19 11 1 13 10 18 16 4 34 17	28 30 32 33 34	모범생



서론



데이터 소개



상품군집모델



성향분석모델

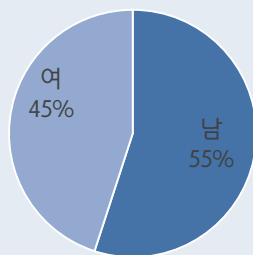


결론

### | 성향#2 'YOLO' 고객 분석

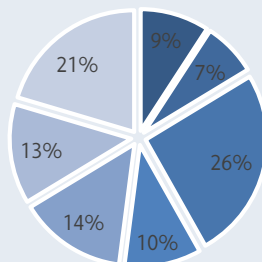
#### 고객기본정보

성별



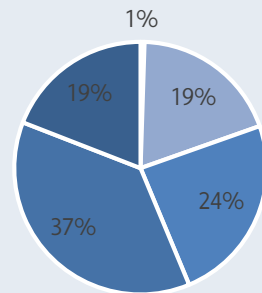
■ 남 ■ 여

고객등급



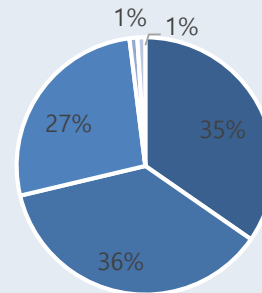
■ Diamond ■ Platinum ■ Vip  
■ Crystal ■ Gold ■ Silver  
■ Green

연령구간



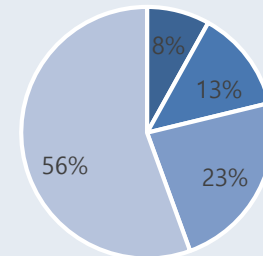
■ 만 19세 이하 ■ 20~40 세  
■ 41~50 세 ■ 51~64 세  
■ 65 세 이상

연간소득



■ 5천만원 미만  
■ 5천만원 ~ 1억원  
■ 1억원 이상  
■ 1억원 ~ 2억원 미만  
■ 2억원 이상

손실감내유형



■ 무슨 일이 있어도 투자 원금은 보전되어야 한다.  
■ 투자 원금에서 최소한의 손실을 감수할 수 있을 것 같다.  
■ 투자원금 중 일부의 손실을 감수할 수 있을 것 같다.

#### 고객 매매패턴

1일 평균 거래량

평균 보유상품개수

평균 예수금

평균 현금/대용 증거금

#### 선호 상품군집

코스피 안정성장주

코스닥 안정성장주

코스닥 무특징주

#### 특별선호 상품군집

코스피 안정성장주

코스닥 사양산업하락주

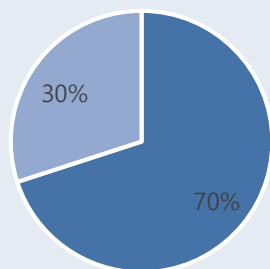
코스닥 안정성장주



### | 성향#8 '가치투자자' 고객 분석

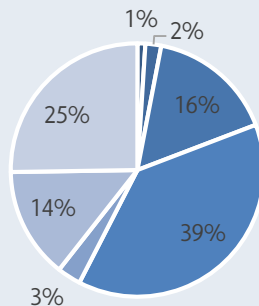
#### 고객기본정보

성별



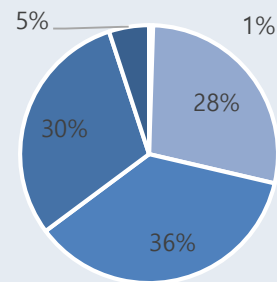
■ 남 ■ 여

고객등급



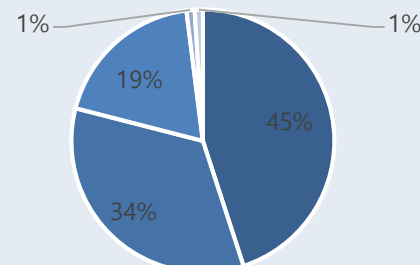
■ Diamond ■ Platinum ■ Vip  
■ Crystal ■ Gold ■ Silver  
■ Green

연령구간



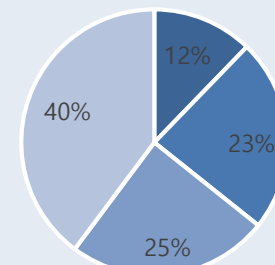
■ 만 19세 이하 ■ 20~40 세  
■ 41~50 세 ■ 51~64 세  
■ 65 세 이상

연간소득



■ 5천만원 미만  
■ 5천만원 ~ 1억원  
■ 1억원 ~ 2억원 미만  
■ 2억원 이상

손실감내유형



■ 무슨 일이 있어도 투자 원금은 보전되어야 한다.  
■ 투자 원금에서 최소한의 손실을 감수할 수 있을 것 같다.  
■ 투자원금 중 일부의 손실을 감수할 수 있을 것 같다.  
■ 기대수익이 높다면 위험이 높다고 상관하지 않겠다.

#### 고객 매매패턴

1일 평균 거래량

평균 보유상품개수

평균 예수금

평균 현금/대용 증거금

#### 선호 상품군집

ELW 안정인기주

코스닥 안정성장주

코스닥 초우량주

#### 특별선호 상품군집

ETF 위험성장주

코스피 안정성장주

코스피 안정일반주

### | 개인투자자 투자성향 유형



서론



데이터 소개



상품군집모델



성향분석모델



결론

YOLO

자금적 여력을 바탕으로 시장을 가리지 않고 투자 가치가 있다고 판단되는 상품에 잦은 변동을 감수하는 개인투자자

대담한 팔랑귀

기준가 변동이 잦아 위험 부담이 따름에도 불구하고 거래수량이 늘고 있는 상품군에 투자하며, 인지도가 높지 않은 상품도 가리지 않는 개인투자자

위험회피 팔랑귀

팔랑귀와 비슷한 성향을 보이거나 보다 위험회피적 성향이 강한 개인투자자

꿈꾸는 소시민

거래수량이 많지 않으면서 기준가 변동이 비교적 잦은 상품을 선호하는 소박한 개인투자자

정석 소시민

꿈꾸는 소시민과 비슷한 성향을 보이거나 보다 비주권 상품을 싫어하는 개인투자자

장기안목

기준가가 강한 상승·하락 추세에 있거나 혹은 예상되는 상품에 장기적 기대를 가지고 투자하는 개인투자자

트렌드팔로워

트렌드에 민감하며 산업군이 상승 모멘텀에 있어 기준가가 증가하거나 거래수량이 늘고 있는 상품에 투자하는 개인투자자

가치투자자

장기적 안목을 가지고 산업의 모멘텀이 상승 추세이거나 거래수량이 서서히 증가함에 따라 주가가 상승하는 상품에 투자하는 개인투자자

모범생

큰 수익을 바라지 않고, 비교적 잘 알려진 상품 중에서 기준가 상승이 예상되는 상품 위주로 투자하는. 말 그대로 '정석투자'를 추구하는 개인투자자

의심쟁이

코스피나 코스닥의 경우에는 시장 내 최고 우량주, 워런트의 경우에는 확실한 추세를 가지는, 손실 위험이 거의 없는 상품에 투자하는 절대안정형 개인투자자

### | 추천시스템 알고리즘



서론



데이터 소개



상품군집모델



성향분석모델



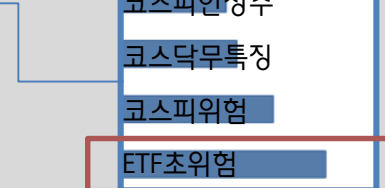
결론



매수주분 1회 이상  
기존고객



Theta테이블 통해  
어떤 성향이 가장 높은 확  
률을 가지는지 확인



phi테이블 통해  
어떤 상품군집이 가장 높  
은 확률을 가지는지 확인

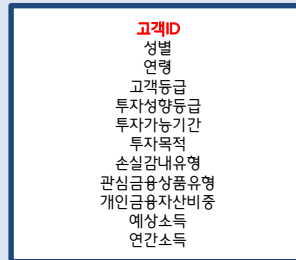


상품군집 안에서 예상수익  
률이 가장 높은 상품 추천

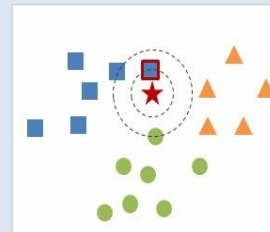


NEW

신규고객



기본 투자성향 정보 수합



KNN 알고리즘을 통해 가장 가까운  
거리에 위치한 기존고객의 성향으로  
신규고객 성향예측



기존고객과 동일



서론



데이터 소개



상품군집모델



성향분석모델



결론

추천시스템 구현 예시 웹페이지

로그인

회원가입

공인인증센터



회사소개



온라인지점

금융투자상품

주식파생상품

리서치센터

고객지원

상품추천

빅데이터가 추천하는

## 나에게 딱 맞는 주식상품

고객님의 개인정보 및 투자정보를 활용하여 최적의  
주식상품을 추천해드립니다.

고객님의 투자성향을 파악하고, 고객님의 비슷한 성  
향의 고객님의들이 선호하는 주식상품을 알아보세요!

추천상품 보러가기



서론



데이터 소개



상품군집모델



성향분석모델



결론

추천시스템 구현 예시 웹페이지

로그인

회원가입

공인인증센터

회사소개

하나금융투자

온라인지점

금융투자상품

주식파생상품

리서치센터

고객지원

상품추천

빅데이터가 추천하는

나에게 딱 맞는 주식상품

이름

투자성향

선택하세요 ▼

관심금융상품

선택하세요 ▼

생년월일

투자가능기간

선택하세요 ▼

손실감내유형

선택하세요 ▼

성별

☒ 남자
 ☐ 여자

투자목적

선택하세요 ▼

금융자산비중

선택하세요 ▼

연간소득

선택하세요 ▼

결과보기

추천시스템 구현 예시 웹페이지



로그인 회원가입 공인인증센터

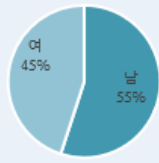
온라인지점 금융투자상품 주식파생상품 리서치센터 고객센터 상품추천

### 홍길동 고객님의 빅데이터 분석결과

고객님은 “가치투자자” 유형입니다.

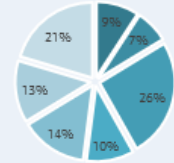
‘가치투자자’ 고객님의 인구통계/투자성향 정보입니다

성별



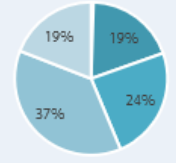
남 · 여

고객등급



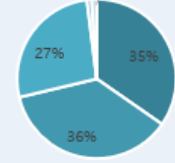
■ Diamond ■ Platinum ■ Vip  
 ■ Crystal ■ Gold ■ Silver  
 ■ Green

연령구간



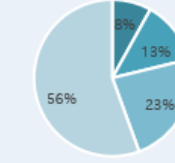
■ 만 19세 이하 ■ 20~40 세  
 ■ 41~50 세 ■ 51~64 세  
 ■ 65 세 이상

연간소득



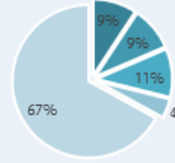
■ 5천만원 미만  
 ■ 5천만원 ~ 1억원  
 ■ 1억원 이상  
 ■ 10억원 이상  
 ■ 20억원 이상

손실감내유형



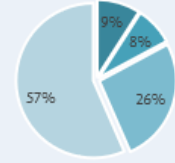
· 무슨 일이 있어도 투자 원금은 보전되어야 한다.  
 · 투자 원금에서 최소한의 손실을 감수할 수 있을 것 같다.  
 · 투자원금 중 일부의 손실을 감수할 수 있을 것 같다.

투자보유기간



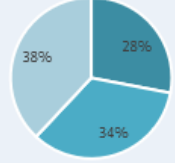
· 6개월 미만  
 · 6개월 ~ 1년  
 · 1년 ~ 2년  
 · 2년 ~ 3년  
 · 3년 이상

투자목적



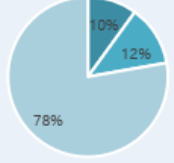
· 기존보유자산에 대한 위협 헤지  
 · 채권이자, 주식배당 정도의 수익실현  
 · 지장가격 변동 추이와 비슷한 수준의 수익실현  
 · 적극적 매매를 통한 수익 및 원금회복손실위험 감내 가능

금융자산비중



· 30% 초과  
 · 10% ~ 30%  
 · 10% 이하

예상소득



· 현재 일정한 수입이 없으며, 연금이 주 수입원임  
 · 현재 일정한 수입이 발생하나, 향후 감소하거나 불안정할 것으로 예상  
 · 현재 일정한 수입이 발생하며, 향후 현재 수준을 유지하거나 증가할 것으로 예상

‘가치투자자’ 고객님의 매매패턴

1일 평균 거래량	3.5회
평균 보유상품개수	17개
평균 예수금	4,000,000원
평균 현금/대출 증거금	현금: 1000만 / 대출: 0원

‘가치투자자’ 고객님의 선호하시는 상품

코스피 안정성장주	ETF 위험성장주
코스닥 안정성장주	코스피 안정성장주
코스닥 무특정주	코스피 안정일반주

홍길동 고객님의 위한 추천상품

코스피 하나금융지주 (086790)

상품정보 확인하기

| 추천시스템 알고리즘



서론



데이터 소개



상품군집모델



성향분석모델



결론

### 효율적인 상품군집 시스템

단순한 상품분류 기준  
(종가, 산업)



시계열 특성 고려한 군집  
다양한 복합적인 요인 반영

### 객관적 데이터 기반의 고객성향분류 시스템

주관적 개인 설문응답 의존



수치에 기반한  
객관적인 포트폴리오 성향분석

### 개인 성향에 맞는 추천시스템

개인의 투자성향 관계없는  
일반적 추천시스템



성향에 기반한 추천시스템으로  
고객만족도 증대

| 추천시스템 알고리즘



서론



데이터 소개



상품군집모델



성향분석모델



결론

1

테스트 DB의 제한적인 데이터 적재기간

: 6개월 기간에만 한정 (20170702-20180123)

2

제한적인 외부데이터 활용

: 기업의 재무정보, 실시간 정보, 거시경제 지표, 예상수익률 등의 데이터 활용 제약

3

불완전한 설문응답 데이터

: 설문응답 데이터 결측치 처리 과정(제거/예측)에서 분석 집단의 대표성 감소



감사합니다  
Q & A