

주식보유기간 예측을 활용한

# 투자성향 별 종목 추천

[ 빅데이터 분석을 통해 획득한 고객정보. 그리고 투자자를 유치하는 증권사 서비스 전략 ]



## 01 분석배경 및 목적

국내 개인투자자의 투자 현황을 바탕으로 NH증권사가 투자자를 유치하기 위해 요구되는 전략을 고민합니다

## 02 데이터 분석과정

데이터 추출 및 전처리 과정을 설명하고 예측에 대한 결과 추론과 예측 결과를 제시합니다

## 03 서비스 제안

예측된 결과를 바탕으로 신규투자자와 기존투자자의 마음을 사로잡을 서비스 전략을 제안합니다

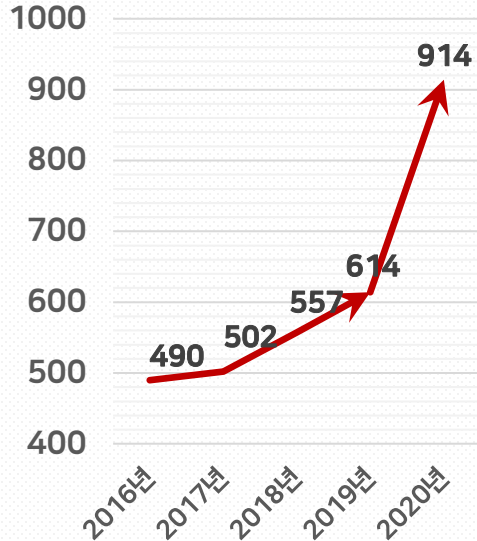
## 04 결론

제안된 서비스를 통해 기대되는 효과를 설명합니다

## 국내 주식투자의 관심은 증가하고 있지만 가계의 금융투자상품의 비중이 낮다

[ 국내 주식투자자 증감 추이 ]

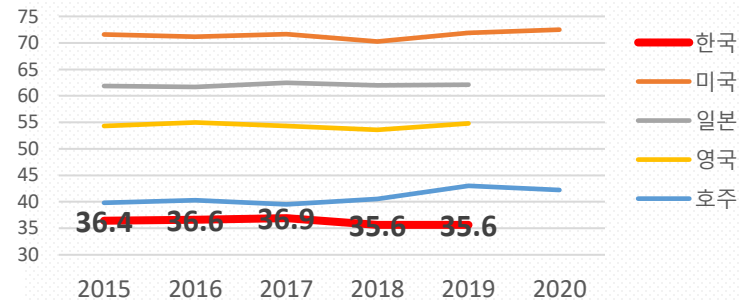
(단위=만명)



- ✓ 2016년부터 2020년까지 **국내 개인 주식 투자자 수 증가**
- ✓ 2020년 코로나로 인해 급격하게 투자자 수 증가

[ 국가 별 가계금융자산 비중 ]

(단위 %)



주요국 대비 한국 가계 자산 중 **금융자산 비중은 하위 수준**

[ 각국 금융자산 비중 구성 비교(2020) ]

	한국	미국	일본	영국	호주
현금·예금	43.4	13.4	54.7	25.5	22.1
금융투자상품	25.2	54.1	13.5	15.2	18.4
(주식)	19.4	36.2	8.6	9.4	17.4
(채권)	3.4	4.8	1.4	0.2	0.1
(펀드)	2.4	13.1	3.4	5.6	0.8
보험·연금	30.8	29.9	28.8	55.9	57.8
기타	0.7	2.6	3.0	3.4	1.7

현금/예금 투자 비중이 가장 높고, **금융투자상품 비중은 25.2%에 불과**

## 투자자를 유치하려는 증권 업계의 경쟁이 심화되는 상황에서 NH투자증권만의 고객 유치 전략이 필요하다

### 2021년 NH투자증권은 울고 웃고

“증권사 브랜드평판 2021년 10월 빅데이터 분석결과... 1위 삼성증권, 2위 미래에셋증권, 3위 키움증권”

- 2021년 10월 기준 NH투자증권은 브랜드 평판 6위
- 2020년 브랜드 평판 1위였던 것에 비해 5등급 하락

빅5 증권사 2분기 실적 순위 지각변동...미래에셋 선두 탈환·NH 2위로

- 2021년 2분기 NH투자증권 실적 순위 2위로 꺾춤
- 2021년 1분기 실적 순위 3위였던 것에 비해 1등급 상승

### 증권사의 로보어드바이저 서비스 전쟁

뉴스핌

충성 없는 '로보어드바이저' 전쟁...증권사, 상품 출시 각축전  
7일 금융투자업계에 따르면 최근 증권사들은 투자자들을 유치하기 위해 각종 로보어드바이저 서비스를 속속 출시하고 있다. 로보어드바이저는 로봇(robot)...

2021. 1. 8.

매일경제

키움증권, 투자자 성향 정밀분석해 AI비서 맞춤 자산관리

비즈니스워치

대신증권, 자녀 선물로 '로보어드바이저 투자' 어때요?

'대신 로보어드바이저'는 별도의 운용보수 없이 수익이 나면 수익금의 10%를 성과 보수로 받는 성과보수형 상품이다. 저비용, 저변동성 장...

2021. 4. 29.



증권사의 로보 어드바이저는 가입비를 내고 사용하고 싶어하는 **일부 고객**에 한해 서비스 제공하는 한계

그렇다면, 투자자 모두에게 개인화된 종목 추천을 하는 것은 어떨까?

# 01 분석목적

분석배경 및 목적

## NH투자증권에게 필요한 것은 “신규투자자 유입과 기존투자자 유지”

### 신규투자자

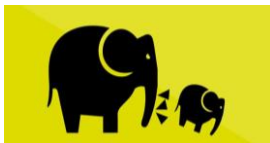


#### 주식 투자에 흥미를 유도하는 것이 필요하다

- 정교하게 예측된 수익을 제공하기 보다는 성별, 나이, 관심사 정도 만으로도 유추할 수 있는 **관심 기업, 업종**을 추천한다.

#### 넛지(Nudge) 효과

: 옆구리를 은근슬쩍 찌르는 행위.  
부드러운 개입을 통해 현명한 행동을 유도



### 기존투자자



#### 증권사에 고객충성도를 높이는 동인이 필요하다

- 증권사를 이용하는 투자자의 거래 내역, 보유기간, 수익률을 파악하여 투자성향에 적합한 종목을 추천한다.
- 추천 종목은 **각 고객의 투자성향과 동일한 투자성향을 가진 고객군의 실제 매수 종목**이다.

#### 밴드웨건(Band Wagon) 효과

:유행에 따라 상품을 구입하는  
소비현상으로 편승효과라고도 한다



## 01 분석목적

분석배경 및 목적

빅데이터 분석 기반 예측과 설문조사를 통해  
투자자의 투자성향에 적합한 주식상품 추천

타겟 별 접근 전략

로보 어드바이저 종목 추천 기술과 결합

타겟 별 추천

신규투자자

성별, 연령,  
관심사관심 종목  
세분화

기존투자자

개인투자자의  
주식보유기간3가지  
투자성향으로  
세분화

신규투자자

"20대 남자는 A종목  
20대 여자는 B종목"

기존투자자

"나와 같은 단기투자자는 이번주에  
A종목을 매수했어요"

신규투자자와 기존투자자에게 적합한 투자 상품을 추천하기 위해서, 신규투자자로부터는 투자 시작 시 설문조사를 통해 인구통계학 기준에 따른 추천 종목을 고려하고, 기존투자자로부터는 거래 데이터를 활용해 예측한 주식보유기간을 중요한 고객 정보로 활용한다.

NH 로보 EMP 자산배분 서비스는 실시간 시를 통해 실물, 심리 지표를 자동 수집하고 시장을 분석하여 현재 시장 상황에 유효한 자산 배분 포트폴리오를 구성한다.

해당 서비스의 기술을 통해, 예측되는 주식 종목의 예상 수익률, 리스크, 추천 보유기간 정보를 활용한다.

## 과거 주식거래 데이터



- 2016년 1월 1일부터 2020년 12월 30일까지의 데이터를 활용한다.
- 해당 기간의 데이터는 증권 시장의 영업일을 기준으로 한다

## 피쳐 선정

투자종목정보

연령

고객자산구간

총투자기간

잔고금액 및 수량

거래일자

시가총액 규모유형

시장구분

업종구분

## 학습



CatBoost

- Catboost 회귀분석 모델은 부스팅 계열 모델로 범주형 데이터가 많을 경우 유용한 분석 모델이다
- 선택한 설명변수에 범주형 데이터가 많기 때문에 Catboost 회귀분석 모델을 중점적으로 고려하였다

## 결과 및 활용

미래 주식 보유기간 예측

서비스 제안

# 02 데이터 전처리

데이터 분석과정

한국주식거래소 api를 활용해 **비상장주식** 종목코드 추출

출처: 한국거래소



2021.10.31 PM 07:57:17 (20분 지연 정보)

표준코드	단축코드	한글종목명
KR7098120009	098120	(주)마이크로컨텍솔루션
KR7131100000	131100	(주)스카이어앤엠
KR7009520008	009520	(주)포스코엠텍
KR7095570008	095570	AJ네트웍스보통주
KR7006840003	006840	AK홀딩스보통주

iem.csv

iem_cd	iem_krl_nm
878 A009520	포스코엠텍

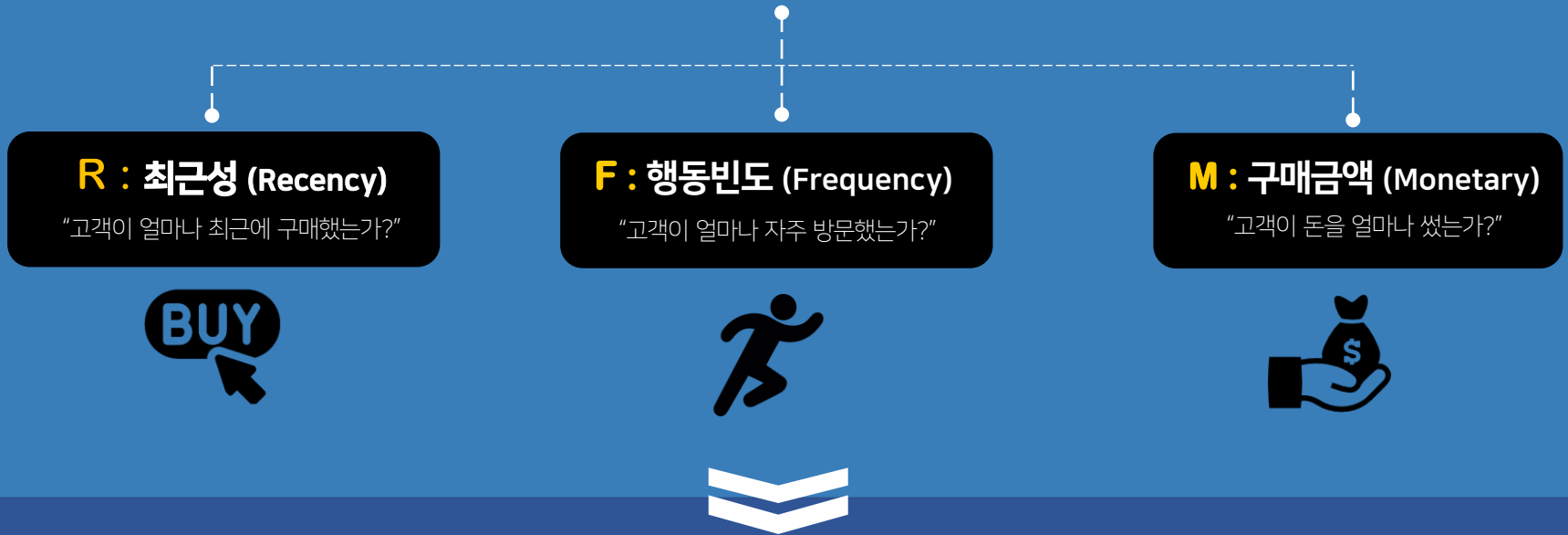
- 종목코드 맨 앞 'A' 제거를 한 6자리 숫자가 실제 종목코드와 동일
- Test 데이터 셋에 비상장 주식 153개가 포함된 것을 확인

**장외주식거래는 거래가 활발하지 않으므로  
미래 보유기간을 최대치로 가정**



# 'RFM 고객분석' 이란 ?

- ✓ CRM(Customer Relationship Management: 고객관계 관리)의 방법론
- ✓ 고객의 특정 행동을 **3가지 관점**으로 파악하여 **고객의 가치**를 분석



## RFM과 클러스터링 결합하기



## 02 데이터 전처리

데이터 분석과정

## 군집화(Clustering)을 활용한 RFM 고객분석

**R** 계좌ID 별 최근 매수일과 최초 매수일의 차이 값  
계좌ID R(최근성)

	계좌ID	R(최근성)
0	0ad104dbed99be0cd858aa772765ddedade554601a981bec9b177d8416085534	47
1	b6623032ed365b6c9a71c0344ebb8ca191f1ebc592777125215a8e7e85cf3395	36
2	356b35c423a191e1f5b8cc1dfa2d597c614ba00b52ccae64d50bc776adedf3e1	173
3	64aae8dd71e5c0761000db9f9a6779e504e7a4aa9dc097067e5b41a971afa1da	1821
4	5f7c3a8f37d9c268d06130ff0be5d32a1b9ef68c13049f804d2aae1393f7f6a2	14

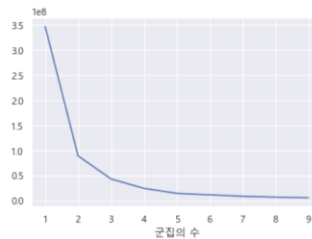
**F** 계좌ID 별 매수일의 개수  
계좌ID F(행동빈도)

	계좌ID	F(행동빈도)
0	0000c7a5a1db9c7e890da984b20d544c09cb6f51c5467ece45d5b35e99912a	23
1	0007f3d10282e6f9de1eb44a64eb1f13e9ed2308cfb9ab6e6efa399d0bdf31	32
2	0008d1bfac569bdb998897e2eb8519dec06b9277a686f4c6e06d582c9791b35	54
3	000a83a961cbcbce40673aed998effbd462fb10dcac8e64c6fa720f2150d3f5a	3
4	000b92ca83a96312a0909b1f489340f4d399c86555012ffe2c41952ae43dc741	3

**M** 계좌ID 별 잔고총액  
계좌ID M(구매금액)

	계좌ID	M(구매금액)
0	...bed99be0cd858aa772765ddedade554601a981bec9b177d8416085534	383002830.0
1	b6623032ed365b6c9a71c0344ebb8ca191f1ebc592777125215a8e7e85cf3395	0.0
2	356b35c423a191e1f5b8cc1dfa2d597c614ba00b52ccae64d50bc776adedf3e1	444872058.0
3	64aae8dd71e5c0761000db9f9a6779e504e7a4aa9dc097067e5b41a971afa1da	0.0
4	5f7c3a8f37d9c268d06130ff0be5d32a1b9ef68c13049f804d2aae1393f7f6a2	0.0

## 군집 수 결정



K-평균법의 엘보우 그래프를 통해  
4개의 군집이 적절하다 판단

**R'** 군집 별 R(최근성)

	count	mean	std	min	25%	50%	75%	max
R군집								
0	275.0	971.647273	139.452328	756.0	882.0	956.0	1047.0	1821.0
1	358.0	536.913408	112.957521	355.0	426.5	540.0	623.5	753.0
2	2368.0	174.613598	64.711630	102.0	123.0	153.0	214.0	354.0
3	6336.0	26.676136	26.177861	0.0	7.0	18.0	40.0	99.0

**F'** 군집 별 F(행동빈도)

	count	mean	std	min	25%	50%	75%	max
F군집								
0	8461.0	39.221369	42.059586	1.0	8.00	23.0	56.00	184.0
1	822.0	327.975669	138.967787	185.0	224.00	276.0	389.00	796.0
2	52.0	1285.250000	498.239720	816.0	950.00	1100.5	1493.25	2948.0
3	2.0	6595.500000	2362.443756	4925.0	5760.25	6595.5	7430.75	8266.0

**M'** 군집 별 M(구매금액)

	count	mean	std	min	25%	50%	75%	max
M군집								
0	9074.0	2.366307e+07	7.933599e+07	0.000000e+00	0.000000e+00	0.000000e+00	3.184375e+05	6.498726e+08
1	229.0	1.283375e+09	5.929755e+08	6.589930e+08	8.271400e+08	1.094551e+09	1.513917e+09	3.436940e+09
2	29.0	5.915942e+09	1.719276e+09	3.627865e+09	4.642435e+09	5.532752e+09	7.192333e+09	9.791030e+09
3	5.0	1.673380e+10	5.884106e+09	1.359097e+10	1.359582e+10	1.439324e+10	1.487503e+10	2.721394e+10

“R’에 0.1, F’에 0.4, M’에 0.5 가중치 조정을 통해 RFM 파생변수 생성”

# 02 데이터 전처리

데이터 분석과정

## 투자성향을 의미하는 파생변수 생성

### ① 투자자별 종목 별 평균 보유 기간

계좌ID 별 종목 별 보유기간  
거래 횟수

		bse_dt	total_hold	평균보유기간
act_id	iem_cd			
35e99912a	A002270	8	350.0	43.75
	A005070	2	19.0	9.50
	A005930	2	1.0	0.50
	A033270	4	75.0	18.75
	A036000	4	133.0	33.25

\*매수일을 거래 횟수로 파악

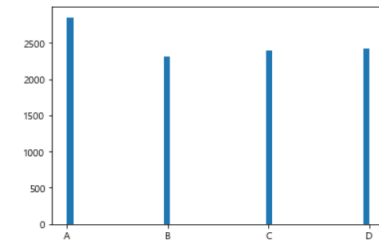
### ② 투자자 별 평균 보유 기간

투자자의 계좌 ID로 groupby 한 ①의 평균

	act_id	평균보유기간
390da984b20d544c09cb6f51c546...		20.575297
1eb445a64eb1f13e9ed2308cfbfa...		7.240785
3897e2eb8519dec06b9277a686f4...		11.780232
673aed998effbd462fb10dcac8e6...		259.130952
909b1f489340f4d399c86555012f...		8.227451

### ③ 구간화

빈도수 기준으로 4개의 구간으로 구간화하여  
새로운 파생변수 생성



“ 평균보유기간이 작을 수록 단기투자 성향이 강하고,  
평균보유기간이 클 수록 장기투자 성향이 강하다”

### 왜 기존의 투자성향 변수를 사용하지 않았는지?

기존의 투자성향 변수는 정성적 데이터로서 빅데이터 분석을 위한 데이터 셋으로는 적합하지 않다고 판단하였다. 또한 결측치가 40% 이상에 육박하였으므로 분석에 적합하지 않다고 판단하였다. 따라서, 새로운 투자성향 파생변수를 만드는 것이 필요하다고 보았다.



## Catboost Regressor 모델의 예측 및 분석결과

### Train 데이터와 Test 데이터의 Loss 점수 차이가 크지 않음

- 회귀분석으로 5년간의 데이터를 Training한 결과, 검증 데이터의 RMSE는 54.8530
- Public Score 점수 제출 결과 RMSE는 61.86549



## 예측 결과에 대한 오류 정정

### 일부 최대기준일 초과 데이터 발견

- 예측 결과 공모전 측에서 제시한 최대 보유기간(146일)을 넘는 기간이 발견되었다.
- 해당 예측값(hold\_d)에 대해, test 데이터 셋의 hist\_d 값에 Max 값인 146일을 더한 값으로 지정한다.

# 03 서비스 제안

서비스 제안

## 신규투자자

- ✓ 투자를 고민하는 개인에게
- ✓ 증권사 계좌 개설 전 부담 없이 제공되는
- ✓ 간단한 설문지 3항목



### [ 설문지 3항목 ]

- Q. 당신의 성별은 무엇입니까?
- Q. 당신의 연령은 무엇입니까?
- Q. 당신의 관심사는 무엇입니까?



### [ 종목 추천 이미지 팝업 ]

(예시)

20대 남성



20대 여성



## 기존투자자

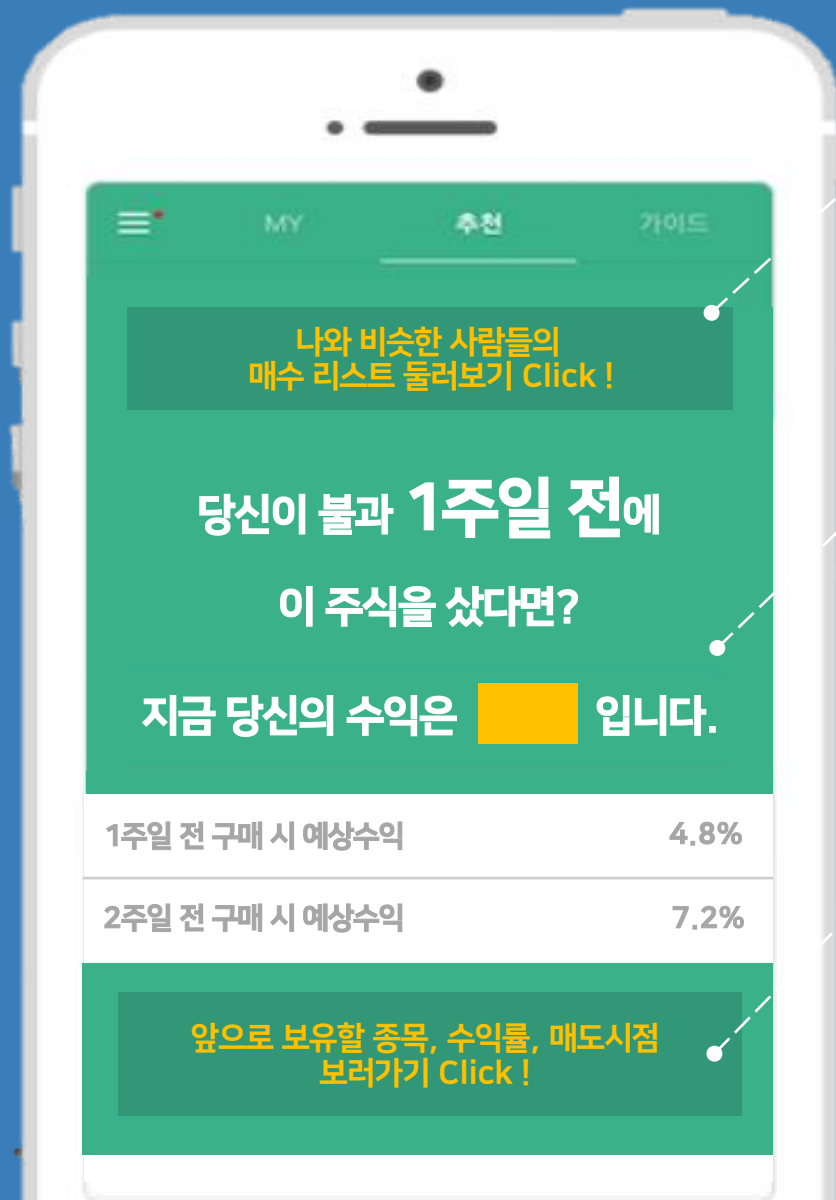
- ✓ 일정 거래횟수를 충족하여 데이터가 쌓인 고객의
- ✓ 주식 보유 기간에 따른 투자성향을 파악하여
- ✓ 팝업으로 알려주는 종목추천 서비스



### [ 3가지 포인트 ]

1. 당신과 같은 투자 성향을 가진 사람의 투자 정보를 알려드립니다.
2. 당신이 놓친 수익률을 알려드립니다.
3. 앞으로 구매하면 좋을 종목을 알려드립니다.

# 금융상품 추천 알고리즘



## 1 [ 같은 투자성향을 지닌 투자자들의 선택한 종목 추천 ]

- 한 번의 클릭으로 동일한 투자성향을 가진 투자자가 가장 많이 매수, 매도 내역을 보여줍니다.
- 투자자는 투자 선택에 해당 내역을 참고할 수 있습니다.

## 2 [ 내가 놓쳐버린 종목의 현재 수익률 구경하기 ]

- 추천종목을 구매했더라면 얻을 수 있었던 수익률을 날짜 별로 제공합니다.
- 투자자는 과거 선택에 따른 현재 결과를 파악함으로써, 미래 투자 시나리오를 구성할 수 있습니다.

## 3 [ 추후 나의 선택에 영향을 미치는 구체적인 정보 ]

- 나의 과거 거래내역 기반으로 나의 평균 수익률과 유사한 수익률을 낼 수 있는 종목, 매수 및 매도 시점을 알려줍니다.
- 투자자는 간접적으로 로보 어드바이저 기능을 경험할 수 있습니다.

## 04 기대효과

## 결론

NH투자증권사 고객은 데이터 기반의 **똑똑한 투자**를 합니다

- 저는 가입자만 대상으로 한다. 서비스 이용료를 지불해야 하므로 이용자가 충분하지 않다. (맞니?)
- 비스는 일정 주식거래 횟수만 충족한다면 나와 비슷한 투자성향을 가진 사람들이 선택한 주식을 추천 받을 수 있다
- 실적을 견인하는 가장 큰 동인이 개인 투자자이므로 이들을 이끌어오는 것이 관건
- 추천 서비스를 이용하는 진입장벽이 낮기 때문에, 초기 개인 투자자들을 끌어오는 데 효과적

NH투자증권사 고객은 무료로 **로보 어드바이저**를 간접 체험합니다

- 기업의 측면에서는 장기투자자 유치도 좋지만 단기 투자자 유치를 통해 자금의 회전율을 올리는 방향이 수익성 면에서 좋음. 따라서 기업분석이 부족한 로보 어드바이저를 사용하면, 장기투자의 리스크를 보여주면서 차라리 단기투자자로 유도하는 것이 수익률 면에서 유리하며 자금의 회전율이 올라갈 수 있음
- 매매수수료 수익 증대



## 04 기대효과

## 결론

“ NH투자증권은 **자금의 회전율**을 높입니다”



- 기업의 수익성은 장기투자자 유치도 좋지만, 단기 투자자 유치를 통해 자금의 회전율을 올리는 방향이 수익성 면에서 좋음. 따라서 기업분석이 부족한 저조한 수익률의 장기투자자라면, 장기투자의 리스크를 보여주면서 차라리 단기투자로 전환하는 것이 수익률 면에서 나을 수 있음을 보여주면, 자금의 회전율이 올라갈 수 있음
- 매매수익률의 증가

“ NH투자증권은 **로보 어드바이저** 서비스 이용자를 늘립니다”





# 05참고사항

기타

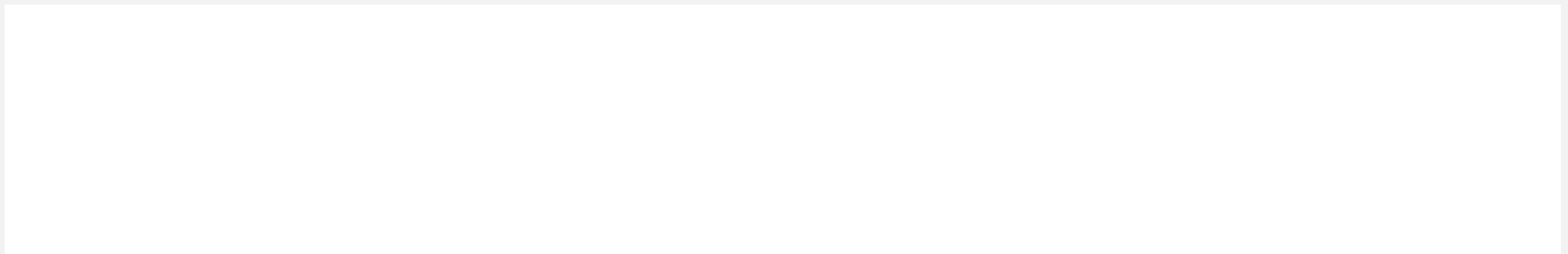
## 참고자료



## 분석도구



## 참고문헌



제 2회 NH 투자증권 빅데이터 경진대회

# 감사합니다





소비자에게 입히다



## 21일 챌린지 - 나의 프로페셔널을 위한 습관 만들기

왜 '21일' 일까?

21일동안 반복해야 습관기억세포가 형성된다 (중앙일보 2017.01)

습관을 바꾸려면 최소 21일은 계속해야 한다. '21일의 법칙'은 미국의 의사 존 맥스웰이 1960년대 그의 저서 '성공의 법칙'에서 처음 주장한 내용이다.



나의 professional을 위한 습관만들기 계정 생성

참여자들은 **해시태그**를 통해 참여 가능  
#21일 챌린지 #파크랜드 #wear the professional

SNS를 통한 바이럴 효과 창출  
파크랜드의 새로운 브랜드 가치인 professional을  
소비자가 자연스럽게 인지하고 직접 경험하게 함



당신의 21일 챌린지를 보여주세요

참여자들은 챌린지에 참여하는 **21일의 시간**을  
하나의 Vlog로 제작하여 올림

영상은 MZ세대가 가장 많이 이용하는 콘텐츠 유형  
더불어 인플루언서의 참여 기대  
인플루언서의 참여를 통해 더 큰 파급효과 창출

예) 내가 원하는 professional - 제2외국어 / 건강함 / 교양있는 / 따뜻한 마음  
그를 위한 습관 - 하루에 문장 세개 외우기 / 스쿼트 100개 / 독서 30분 / 칭찬 한마디

우수 참가자를 선정하여 자아도취 페스타에 초청하여 **시상식**을 진행한다.

상황 분석

문제 분석

타겟 분석

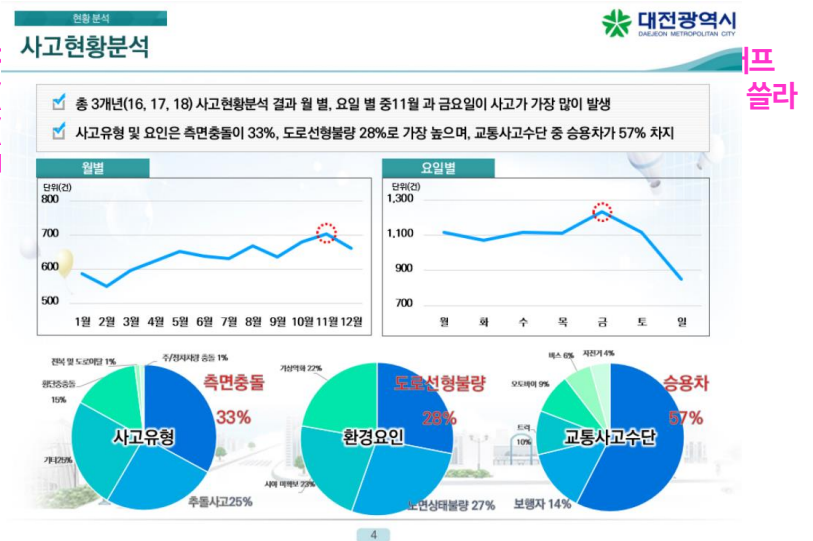
컨셉 도출

IMC 전략

스케줄링

- 시기 별 수익률 분포를 보면, 2016~2019년 보다 2020년 체결량이 크다
- 2020년 수익은 오히려 적음
- 20년 고객 매우 증가
- 남성이 여성보다 많은

전반  
있음  
했음  
데이



## 2. 기존 투자자 대상

- 은행에서 끌어오는 마이데이터는 신뢰할만 하지만, 증권사가 보유한 데이터는 품질이 좋지 않음.

=> (어떤 의미에서 안좋은 건지??)

따라서 고객의 성향을 유추하는 데에는 증권사의 고객정보를 제외하고, 거래내역/자산 등만으로 추론하는 것이 타당할 듯.

- RFM으로 군집화 시도 타당

- 기존 투자자의 거래내역이나 보유기간만으로는 부족함. 기존 투자자의 수익률 내역까지 필요. 수익률이 좋은 사람들에게는 굳이 투자 추천을 할 필요도 없고, 설득도 안될 듯.

=> 그쵸....

수익률이 나쁜 투자자들 중, 단기 투자자라면 다른 방향의 투자방식을 제안하거나, 수익률이 좋은 단기투자자들이 노린 종목들을 제시해주며 차이점을 보여줘도 될 듯.

- 수익률이 나쁜 장기투자자라면, 기업분석의 추이나 투자한 종목들의 동일업종의 데이터를 보여주는 형식은 어떨지

=> 기업분석의 추이가 어떤 걸 말하는 건가요??

거래 많이 하거나 크게 하면 매매 수수료.. 빈도가 클 수록,  
M도 클수록, R는 작을수록 좋은 거..

1. 증권사 데이터 품질 : 고객성향 쪽 신뢰도 없음
2. 수익률 나쁜 장기투자자라면, 본인이 매매한 종목에 대한 현재까지의 분석을 분기단위로 보여주거나, 유사업종의 분석까지 함께 보여주는 것
3. 관심을지는 모르겠는데, 증권사의 수익이 자산운용, 매매수수료, 단기매매 이자 등이 있어서 저희 차원에서 맞는 건 매매수수료 같아서요

그리고 기업에 대한 소개나 분석을 토스처럼 간략화할 필요가 있음

잔고총액이 같으면 빈도가 크면 총액이 분자로 가는데, 빈도가 클수록 이 사람은 단타로 됨  
빈도가 작을 수록 장타가 됨.

직관 그래프  
흘어진 내 자산정보를 한눈에

간단한 계좌등록  
공인인증서 등록만으로  
한번에 자동으로

맞춤 상품 추천  
내게 맞는 더 좋은 상품만  
알아서 한눈에



무료 신용조회  
매일 무료로 신용등급 확인

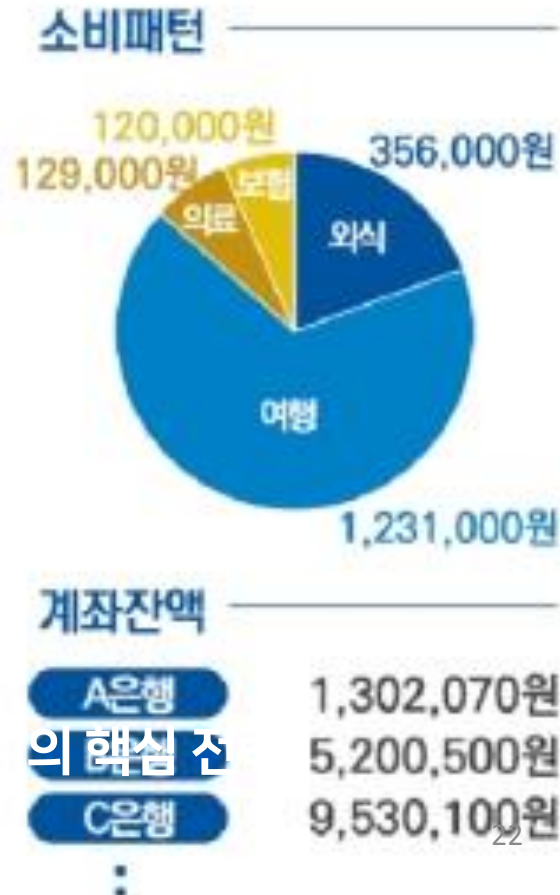
진짜 신용올리기  
비금융 정보로 확실한 신용상향

통합 상품조회  
은행 갈 필요없이 금융상품을 한 화면에

버려

마이 데이터(My Data)

투자자 유 의 핵심 전



# 04 분석 결과 활용 서비스 제안

## 금융상품 추천 알고리즘 제안

### 투자성향



[그림 3] 마이데이터 서비스: 통합조회 및 금융상품 추천

### 소비패턴



### 계좌잔액

A은행	1,302,070원
B은행	5,200,500원
C은행	9,530,100원
:	

### A 고객님에 대한 금융상품 추천

#### 신용카드

- (1위) A카드 : 연회비(30,000원), 월간 할인혜택(25,000원)
- (2위) B카드 : 연회비(20,000원), 월간 할인혜택(15,000원)

#### 정기예금

- (1위) A은행 00적금 : 연이율 2.1%, 1년예금 기준
- (2위) B은행 00적금 : 연이율 1.9%, 1년예금 기준

#### 재무상태 분석

A 고객님은 소득·연령 대비 여행 분야 지출이 많은 상태입니다. 입출금 통장 6개월 평잔 기준 약70%인 500만원을 정기예금에 예치하시면 약 10만원의 이자 혜택을 보실 수 있습니다.

1. 추  
: 7  
을

(시물: 자료: 금융위원회, 신용정보원, 금융보안원, 금융결제원, 한국투자증권

입



# 목차

- A. 아이디어 실현 가능성(30%)
  - 아이디어가 실제 실현 가능성이 있는가
  - 아이디어가 비즈니스적 임팩트를 기대할 수 있는가
- B. 논리적 일관성(30%)
  - 아이디어 전개 과정에서 탄탄한 자료로 이어져 있는가
  - 근거 있는 자료를 제시하고 적합하게 사용되었는가
  - 논리적으로 일관성이 있는가(20%)
- C. 주제에 대한 독창적인 아이디어와 창의적인 관점이 제시되어 있는가
  - 새로운 발상, 독특한 콘셉트 요소가 있는가
- D. 자료의 완결성(10%)
  - 보고서가 성의 있고 짜임새 있게 구성되어 있는가
  - 참가자의 노력이 잘 드러나는가
- E. 발표 능력(10%)
  - 발표가 자신감 있고 자연스러운가
  - Q&A에 대한 대응이 매끄럽고, 논리적으로 문제가 없는가



# 데이터 추출 및 전처리



결측 처리 및 오류 데이터 수정

## Life Stage 피쳐의 오류 수정

버려

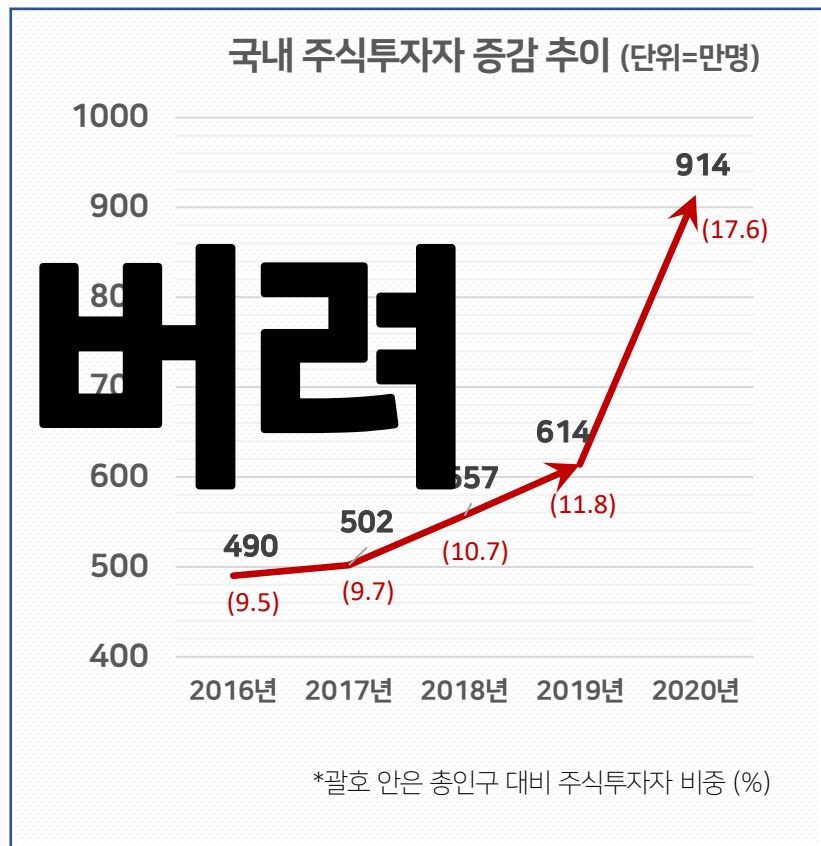
lsg_sgm_cd	2	3	4	5	6	7	8	9
sex_dit_cd								
1	1100	1926	0	1905	3	827	0	224
2	760	0	1164	688	848	0	400	155



lsg_sgm_cd	2	3	4	5	6	7	8	9
sex_dit_cd								
1	1100	1926	0	2593	0	827	0	224
2	760	0	1164	0	851	0	400	155

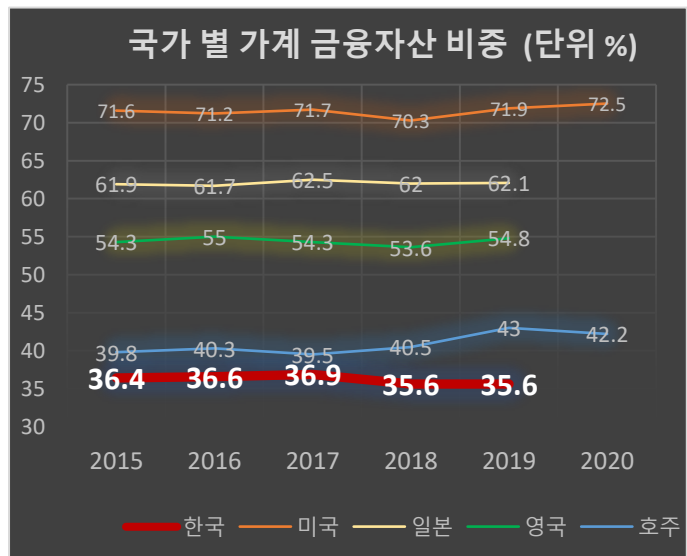
✓ 국내 주식투자 관심 **증가**

- ✓ 2016년부터 2020년까지 **국내 개인 주식 투자자 수 증가**
- ✓ **2020년 코로나**로 인해 급격하게 투자자 수 증가



✓ 가계 금융자산 투자 비중이 **적은** 상태

- ✓ 한국 가계 자산 중 **금융자산 비중은 주요국 대비 낮은 상황**



- ✓ **현금/예금 투자 비중이 가장 높고, 금융투자상품 비중은 25.2%에 불과**

각국 금융자산 비중 구성 비교 (2020)

	한국	미국	일본	영국	호주
현금·예금	43.4	13.4	54.7	25.5	22.1
금융투자상품	25.2	54.1	13.5	15.2	18.4
(주식)	19.4	36.2	8.6	9.4	17.4
(채권)	3.4	4.8	1.4	0.2	0.1
(펀드)	2.4	13.1	3.4	5.6	0.8
보험·연금	30.8	29.9	28.8	55.9	57.8
기타	0.7	2.6	3.0	3.4	1.7

빅데이터 분석 기반 개인투자자의 주식 보유 기간을 예측하여  
투자자의 투자성향에 적합한 주식상품 추천

투자종목정보

연령

고객자산구간

총투자기간

잔고금액 및 수량

거래일자



주식 거래 데이터 수집

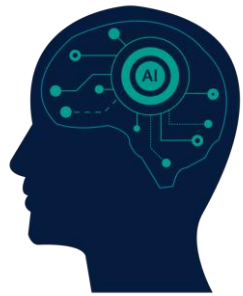


예측 모델을 활용해  
○○○○○○○○○○○○○○

주식 보유기간에 따른 투자성향

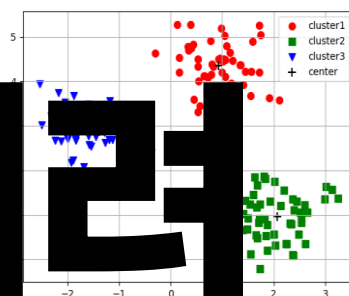
2016 ~ 2020 까지의 **과거** 주식 거래 데이터

2021 이후의 미래 주식 거래



머신러닝  
(Machine Learning)

**버려**



군집화(Clustering)

