

# 보험 상품 추천 알고리즘

eturn

SNU 4차산업혁명 아카데미

비즈니스 애널리틱스 과정 캡스톤 프로젝트

최종 발표

박 영 롱

지도 교수 : 유 병 준

담당 조교 : 석 민 성

# Contents

PART #1

# **INTRODUCTION**

01 프로젝트 소개

02 데이터 소개

PART #2

# **DATA ANALYSIS**

01 EDA 및 시각화

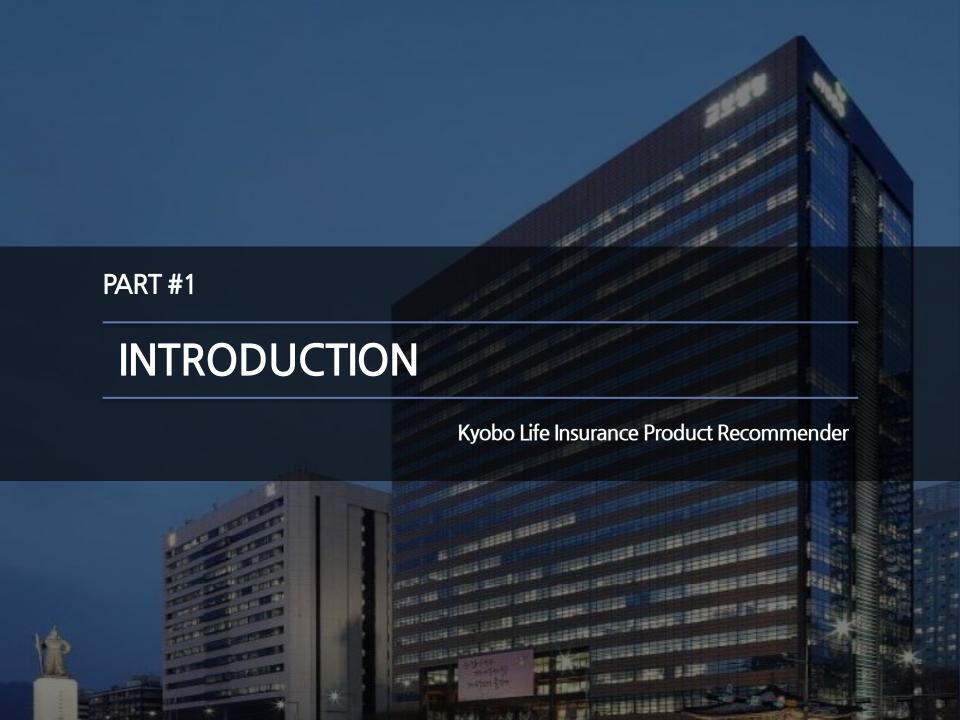
02 분석 과정 및 결과

PART #3

**CONCLUSION** 

01 적용 방안

02 보완점 및 제안사항



# 1-1. 프로젝트 소개



# 프로젝트 목적

## ✓ 우리나라 보험 소비자의 낮은 보험 이해도와 만족도 존재

# "소비자 보험 이해도 평가 점수는 61.7점"

송고시간 | 2015/05/20 13:48









### 소비자 44.5%는 "내 금융이해력 낮은 수준"

(서울=연합뉴스) 김수현 기자 = 소비자들의 보험 주요 개념에 대한 이해도가 100점 만점에 60점을 조금 넘는 수준인 것으로 나타났다.

보험연구원의 변혜원 연구위원과 이해랑 연구원은 20일 퍼낸 '소비자 금융이해결 강화 방안 : 보험 및 연금'이라는 보고서에서 "보험 이해도 평가 평균 점수는 100점 만점에 61.7점"이라. 고 밝혔다.

연구원은 지난해 4월 20일부터 5월 15일까지 제주도를 제외한 전국 만 20세 이상 성인 남녀 1천200명을 대상으로 설문조사를 해 이런 결과를 얻었다.

# 소비자 만족도 낮은 韓 보험... 30개국 중 꼴찌

음 박정식 기자 │ ② 승인 2016.06.28 13:14 │ ⊕ 댓글 0

캡제미니 30개국 평가... 한국 68.9점 받으며 지난해 25위→30위로 추락

우리나라의 보험 만족도가 최하위라는 평가 결과가 나왔다.

금융소비자연맹은 28일 글로벌 컨설팅업체 캡제미니(Capgemini)가 발표한 '2016 세계보험보 고서'를 인용해 보험소비자 경험평가지수(CEI)에서 한국이 평가대상 30개국 중 최하위인 30위 를 기록했다고 밝혔다.

캡제미니가 발표한 보고서에 따르면 보험 가입자들을 대상으로 설문조사를 실시한 결과 1위 는 79.7점을 받은 오스트리아로 나타났다. 이어 78.4점을 받은 미국이 2위, 77.4점을 받은 벨기 에가 3위를 차지했다.

### 반면 한국은 68.9점을 얻으며 지난해 25위에서 올해 30위로 떨어졌다.

CEI 평가에서 한국 소비자들이 답한 긍정적인 경험은 33.6%에 그쳤고 중립적인 경험이 53.8%, 부정적인 경험이 7.6%였다.

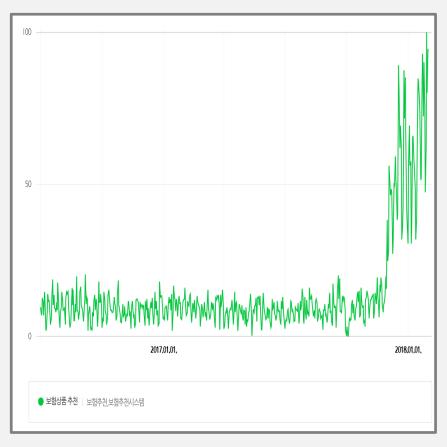
연맹은 한국 보험산업의 소비자 만족도가 최하위로 처진 이유로 △소비자정보 제공 부실 △ 보험상품의 부실판매 △빈발하는 보험금 지급 거부 △보험 민원의 늑장 처리 △부실한 금융 감독 등 5가지를 꼽았다.

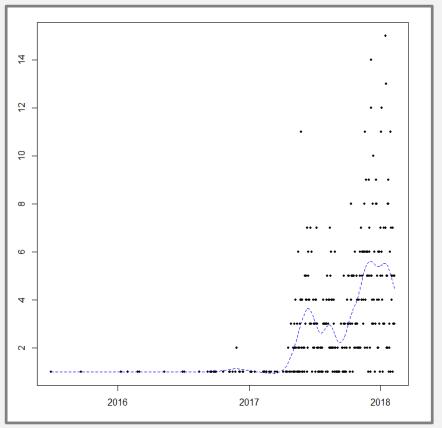
# 1-1. 프로젝트 소개



# 프로젝트 목적

✓ 보험 상품 추천에 대한 니즈와 관심 증가 추세





KEYWORD 검색량 추세

BLOG 언급 횟수 추세

# 1-1. 프로젝트 소개



# 프로젝트 목적

✓ 빅 데이터 기반 상품 추천을 통해 소비자의 만족도와 회사의 수익성 제고

### 현재의 문제점

■ 보험시장의 역선택 문제

- 설계사의 주관에근거한 현재의 판매방식
- 소비자의 낮은보험 이해도와 만족도

### 앞으로의 동향

- 빅데이터 기반맞춤형 상품 판매 활발
- 활용 가능한 데이터범위의 지속적인 확대
- 인슈어테크(InsurTech) 도입 및 발달 지속

### 분석의 기대 효과

■ 상품 판매의 객관성 강화

- 소비자의 다변화된니즈 파악 및 충족
- 상품 선택을 도움으로써소비자 만족도 향상

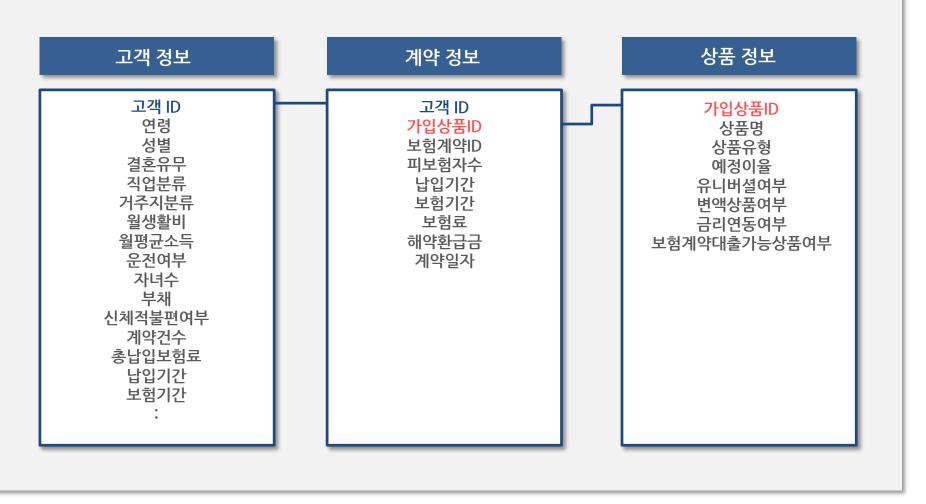


# 1-2. 데이터 소개



# 데이터 테이블 정의

✓ 데이터 마트를 바탕으로 분석에 필요한 데이터 테이블 및 변수 정의



# 1-2. 데이터 소개



# 분석 대상 고객

✓ 기준 년월: 17.04 기준 보험계약자/피보험자/계약 데이터

✓ 계약 수 : 약 850만 건

✓ 보험계약자 수:약 450만 명

✓ 피보험자 수 : 약 500만 명

▶ 분석 대상의 범위 축소 필요

# 자기 계약 대면 채널 계약 SAMPLING 보험 계약자와 주피보험자가 동일한 계약으로 한정 3,859,632 명 다면 채널 계약 SAMPLING 분석의 효율성을 위해 SAMPLE DATA 추출 및 활용 5,000 명

# 1-2. 데이터 소개



# 추천 대상 상품 : 총 37개 상품<sup>1)</sup> (보장성 상품 23개, 저축성 상품 14개)

상품명	구분
(무) 교보행복한준비보험	종신
(무) 교보LTC종신보험	종신
(무) 교보New종신보험	종신
(무) 교보노블리에종신보험	종신
(무) 교보프리미어종신보험!!!	종신
(무) 교보내게맞는종신보험	종신
(무) 교보베스트플랜변액유니버셜종신보험॥	종신
(무) 교보하이브리드변액종신보험	종신
(무) 교보하이브리드VIP변액종신보험	종신
(무) 교보알찬변액종신보험	종신
(무) 교보GI변액종신보험	종신
(무) 교보플러스정기보험	정기
(무) 교보생생플러스건강보험	질병
(무) 교보내생애맞춤건강보험	질병
(무) 교보곰두리보장보험 (장애인전용)	암
(무) 교보암보험	암
(무) 교보내게맞는건강보험	암
(무) 교보더든든보장보험	상해
더 든든한 (무) 교보우리아이보장보험 II	어린이

상품명	구분
(무) 교보베스트플랜CI보험Ⅲ	CI
(무) 교보미리미리CI보험	CI
(무) 교보미리미리여성CI보험	CI
(무) 교보실손의료비보험	실손
연금저축교보로연금보험	연금
연금저축교보연금보험	연금
개인연금저축교보로연금보험	연금
(무) 교보뉴플랜연금보험	연금
꿈을 이어주는 (무) 교보VIP연금보험	연금
꿈을 이어주는 (무) 교보연금보험Ⅱ	연금
(무) 교보바로받는웰스연금보험	연금
(무) 교보우리아이변액연금보험Ⅲ	연금
더 드림 (무) 교보변액연금보험Ⅲ	연금
더 드림 (무) 교보VIP변액연금보험 II	연금
미리 보는 내 연금 (무) 교보변액연금보험॥	연금
(무) 교보빅플러스저축보험	저축
(무) 교보사랑저축보험	저축
(무) 교보변액적립보험IV	저축





# 샘플데이터의 타당성 검증

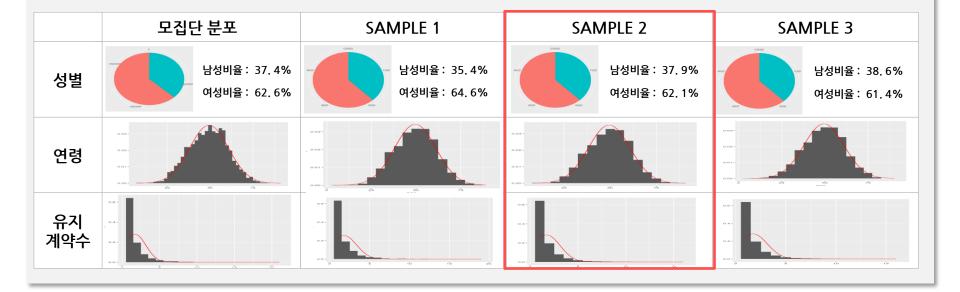
• 반복 추출



✓ 5000건의 샘플 데이터가 모집단을 대표할 수 있을까?

: 총 <mark>3번</mark> Ramdom sampling 반복 수행

• 분포 비교

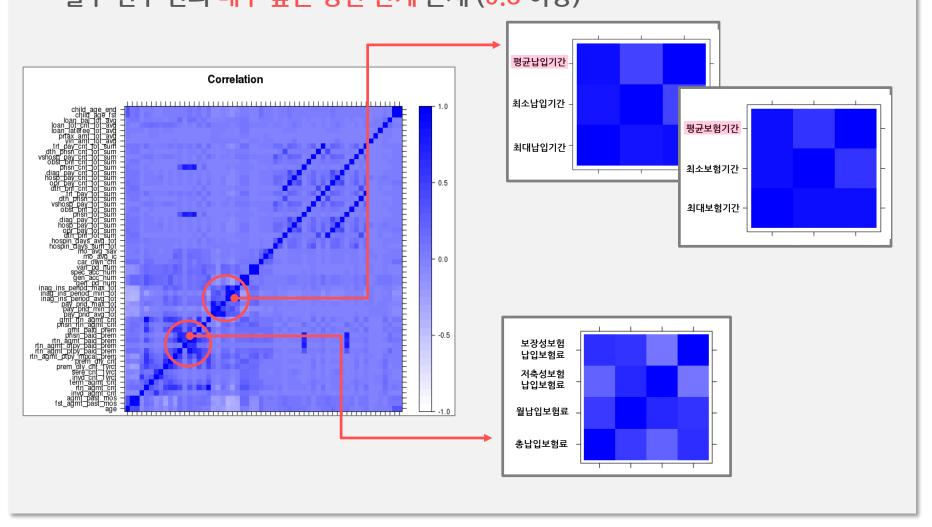


# 2-1. EDA 및 시각화



# 상관관계가 높은 변수 제외

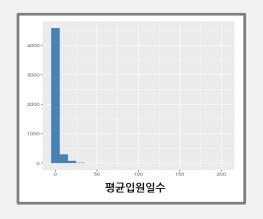
• 일부 변수 간의 <mark>매우 높은 상관 관계</mark> 존재 (0.8 이상)

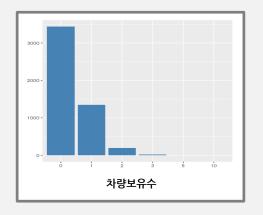


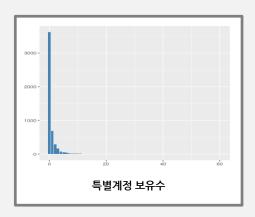


# 변수 분포 파악 : 불필요 변수 제거 / 파생변수 생성

• 0값의 비율이 높고 군집분류 / 특성파악에 불필요한 변수 제거







• 파생 변수 생성

자녀유무

첫째자녀나이 or 막내자녀나이 값 존재 시 1(있음)

직업대분류

직업을 7개로 범주화 (사무직/생산직/농임업/자영업/주부/학생/기타)

월소득대비 보험료

월납보험료 ÷ 월소득

월소득대비 저축

월저축 : 월소득



# 연속형 변수와 범주형 변수 동시 처리

• 군집 분류에 사용할 변수를 확정 (INPUT 고려)

변수명	변수형태
나이	연속형
성별	범주형
직업	범주형
교육수준	범주형
결혼여부	범주형
자녀유무	범주형
월소득대비 보험료	연속형
월소득대비 저축	연속형

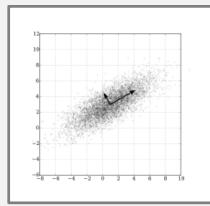
연속형 변수와 <mark>범주형 변수를</mark> 동시에 고려할 수 있는 분석 방법 필요

- PCA with mixed data
- Two steps clustering
- K-prototype Algorithm



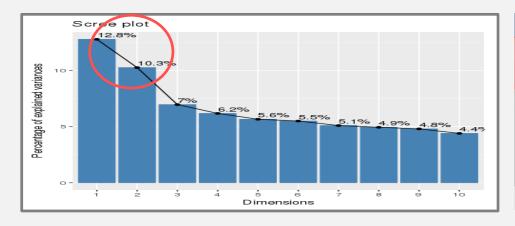
# 주성분분석 (Principal Component Analysis, PCA) with mixed data

### What is PCA?



- 다수의 변수를 전체 변동을 잘 설명할 수 있는 소수의 주성분으로 환원
- 두 개의 주성분을 이용하는 경우, 데이터를 2차원 평면에 표현 가능
- 범주형 변수에 대해 더미변수를 지정하여 연속형 변수와 함께 분석 가능

### • 적용 결과



Dimension	Cumulative variance percent
1	12.76%
2	23.01%
3	29.97%
4	36.14%
5	41.79%
6	47.28%
7	52.37%



# Two steps Cluster Analysis: Gower distance & Partitioning Around Medoids

• 군집 분류 프로세스

관측치 사이의 거리 계산

거리행렬 생성

K값 결정

클러스터링

• 거리 계산 : Gower distance

$$S_{ijk} = \begin{cases} 0 \ if \ X_{ik} = \ X_{jk} \\ 1 \ if \ X_{ik} \neq \ X_{jk} \end{cases}.$$

Categorical variable

$$S_{ijk} = \frac{|x_{ik} - x_{jk}|}{r_k}$$

Numeric variable

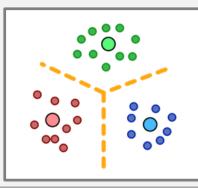
• 각 변수 유형에 따라 적합한 거리 매트릭이 다르게 적용 <sup>1)</sup>

• 적용된 거리들이 0과 1 사이의 비율로 조정

• 사용자 지정 가중치를 사용하여 최종 거리 행렬 생성

• Pros: 이해가 쉽고 직관적 계산이 가능 / Cons: 비정규성, 이상치에 민감

• 클러스터링: Partitioning Around Medoids (PAM)

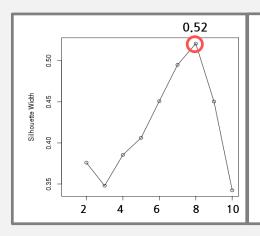


- K-Means Algorithm 과 유사
- 클러스터의 중심이 관측치 자체 (Medoid) 로 제한되는 차이점
- 노이즈나 이상값에 민감하지 않음



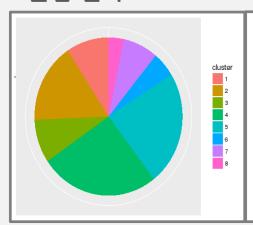
# Two steps Cluster Analysis: Gower distance & Partitioning Around Medoids

# • K값 결정



- Silhouette Width:  $s(i) = \frac{b(i) a(i)}{\max\{a(i), b(i)\}}$ 
  - a(i): i번째 관측치 기준 같은 군집 안의 다른 관측치와의 평균 거리 (Cohesion)
  - b(i): i번째 관측치 기준 다른 군집 과의 거리 중 가장 작은 거리 (Separation)
- 1에 가까운 값을 가질 수록 군집 분류가 잘 된 것으로 판단 가능
- 2~10 까지의 군집 개수의 Silhouette Width를 비교했을 때 K=8이 가장 적절

# • 군집 결과



- 8 개의 군집으로 분류를 하였으나 특정 변수에서 관측치 분류가 집중됨
- 해당 변수 제거 & K 값 조정으로 반복 수행 해봤으나 비슷한 문제 지속됨
- 새로운 거리 계산 방법, 분류 방법의 적용 필요성



# K-prototype Algorithm

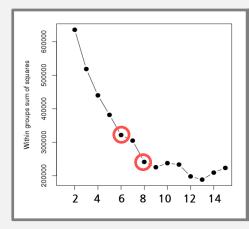
What is K-prototype?

$$d_2(X, Y) = \sum_{j=1}^{p} (x_j - y_j)^2 + \gamma \sum_{j=p+1}^{m} \delta(x_j, y_j)$$

Dissimilarity between two mixed type objects

- K- Means : Only 연속형 자료 & K-Mode : Only 범주형 자료
- 두 개를 혼합한 알고리즘 : 연속형/범주형 두 가지로 나누어서 거리 계산
- 분류 시 특정 변수 하나로 치우치지 않도록 가중치 부여

# • K값 결정

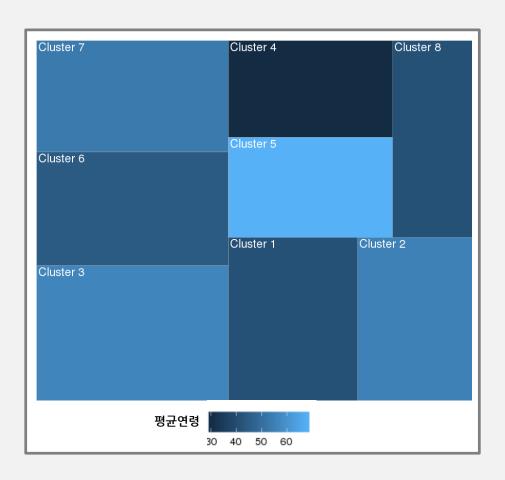


- Elbow Method:
  - 그룹 내 Sum of squares 값의 감소가 정체되는 시점의 군집 수를 K 값으로 설정
- 2~15 까지의 군집 개수의 Within groups sum of squares 를 비교했을 때
  - K = 6~8이 가장 적절하다고 판단할 수 있음
- K = 6, 7, 8 을 모두 적용해본 뒤 특성이 가장 잘 분류되는 K=8 을 최종 선정



# 각 군집의 빈도수에 따른 전체 군집 트리맵 (Treemap)

• K = 8 로 분류된 군집의 빈도수 및 평균 연령 비교



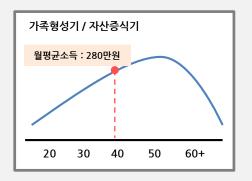
군집 번호	빈도수	평균연령
1	674명	41세
2	603명	55세
3	828명	56세
4	504명	29세
5	521명	68세
6	691명	44세
7	676명	53세
8	503명	41세



# 나눠진 군집에 대한 세부 정의

# 군집 1

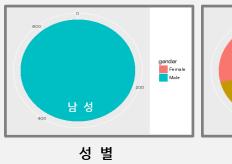
### 라이프 사이클



### 계약 정보

평균 1~2개의 상품 가입
유지계약 3건 이상 보유 비율 최대
보장성보험 가입 비율 높음
납입기간: 20년납 선호
보험기간: 종신 선호
일반보험 선호

### 인구통계학 정보







### 가입 상품 정보

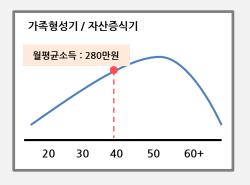
	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 다사랑 유니버셜 CI보험	보장성	질병/CI/암	일반	상세 확인
2	무배당 교보 변액 유니버셜 종신보험	보장성	종신/정기	변액	상세 확인
3	무배당 베스트라이프 교보 종신보험	보장성	종신/정기	일반	상세 확인
4	무배당 교보 가족사랑 통합 CI보험	보장성	질병/CI/암	일반	상세 확인
5	더 든든한 무배당 교보 통합 CI보험	보장성	질병/CI/암	일반	상세 확인
6	멀티 플랜 무배당 교보 통합 CI보험	보장성	질병/CI/암	일반	상세 확인
7	21C 슈퍼 골드 연금보험	저축성	연금	일반	상세 확인



# 나눠진 군집에 대한 세부 정의

# 군집 1 "사망도, 노후 대비도 미리미리!" 유형

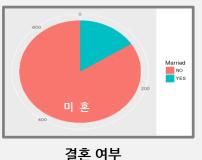
### 라이프 사이클



### 인구통계학 정보







계약 정보



가입 상품 정보

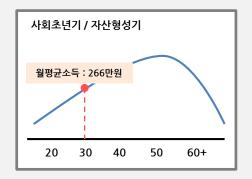
	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 미리미리 CI보험	보장성	질병/CI/암	일반	상세 확인
2	무배당 교보 하이브리드 변액 종신보험	보장성	종신/정기	변액	상세 확인
3	무배당 교보 내게맞는 종신보험	보장성	종신/정기	일반	상세 확인
4	무배당 교보 베스트플랜 CI보험	보장성	질병/CI/암	일반	상세 확인
5	무배당 교보 뉴플랜 연금보험	저축성	연금	일반	상세 확인
6	연금저축 교보 연금보험	저축성	연금	일반	상세 확인
7	무배당 교보 프리미어 종신보험	보장성	종신	일반	상세 확인



# 나눠진 군집에 대한 세부 정의

# 군집 4

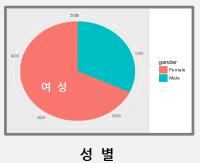
### 라이프 사이클



### 계약 정보

평균 1~2개의 상품 가입
유지계약 2건 보유 비율 최대
보장성보험 가입 비율 높음
납입기간: 20년납 선호
보험기간: 종신 선호
변액보험 선호

### 인구통계학 정보







### 가입 상품 정보

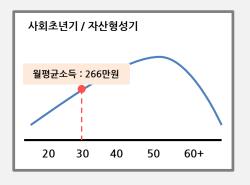
	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 다사랑 유니버셜 CI보험	보장성	질병/CI/암	비변액	상세 확인
2	무배당 교보 가족사랑 통합 CI보험	보장성	질병/CI/암	비변액	상세 확인
3	무배당 교보 여성 CI보험	보장성	질병/CI/암	비변액	상세 확인
4	더 든든한 무배당 교보 통합 CI보험	보장성	질병/CI/암	비변액	상세 확인
5	무배당 교보 변액 유니버셜 종신보험॥	보장성	종신/정기	변액	상세 확인
6	무배당 교보 변액 유니버셜 종신보험	보장성	종신/정기	변액	상세 확인
7	미리 보는 내 연금 무배당 교보 변액 연금보험॥	저축성	연금	변액	상세 확인



# 나눠진 군집에 대한 세부 정의

# 군집 4 "든든한 나의 미래, 투자도 틈틈이!" 유형

### 라이프 사이클



### 인구통계학 정보







계약 정보



가입 상품 정보

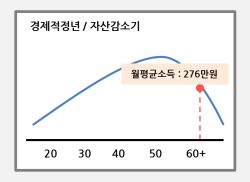
	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 미리미리 CI보험	보장성	질병/CI/암	일반	상세 확인
2	무배당 교보 베스트플랜 CI보험	보장성	질병/CI/암	일반	상세 확인
3	무배당 교보 미리미리 여성 CI보험	보장성	질병/CI/암	일반	상세 확인
4	무배당 교보 하이브리드 변액 종신보험	보장성	종신/정기	변액	상세 확인
5	미리 보는 내 연금 무배당 교보 변액 연금보험॥	저축성	연금	변액	상세 확인
6	무배당 교보 베스트플랜 CI보험	보장성	질병/CI/암	일반	상세 확인
7	더 드림 (무) 교보변액연금보험Ⅲ	저축성	연금	변액	상세 확인



# 나눠진 군집에 대한 세부 정의

# 군집 5

### 라이프 사이클



### 계약 정보

평균 1~2개의 상품 가입

유지계약 1건 보유 비율 높음

보장성보험 가입 비율 높음

납입기간: 10년납 선호

보험기간: 30년 선호

일반보험 선호

### 인구통계학 정보



### 가입 상품 정보

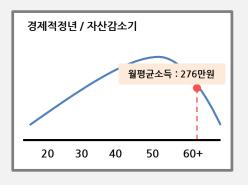
	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교통안전보험	보장성	상해/실손/기타	일반	상세 확인
2	무배당 참사랑효보험	보장성	상해/실손/기타	일반	상세 확인
3	무배당 내게맞는 종신보험	보장성	종신/정기	일반	상세 확인
4	무배당 교보 행복한 준비보험	보장성	종신/정기	일반	상세 확인
5	무배당 교보 다사랑 유니버셜 CI보험	보장성	질병/CI/암	일반	상세 확인
6	무배당 교보 베스트라이프 암보험	보장성	질병/CI/암	일반	상세 확인
7	무배당 생생 종합 건강보험	보장성	질병/CI/암	일반	상세 확인



# 나눠진 군집에 대한 세부 정의

# 군집 5 "아프지 말고, 다치지 말고 건강하게!" 유형

### 라이프 사이클



### 인구통계학 정보



### 계약 정보

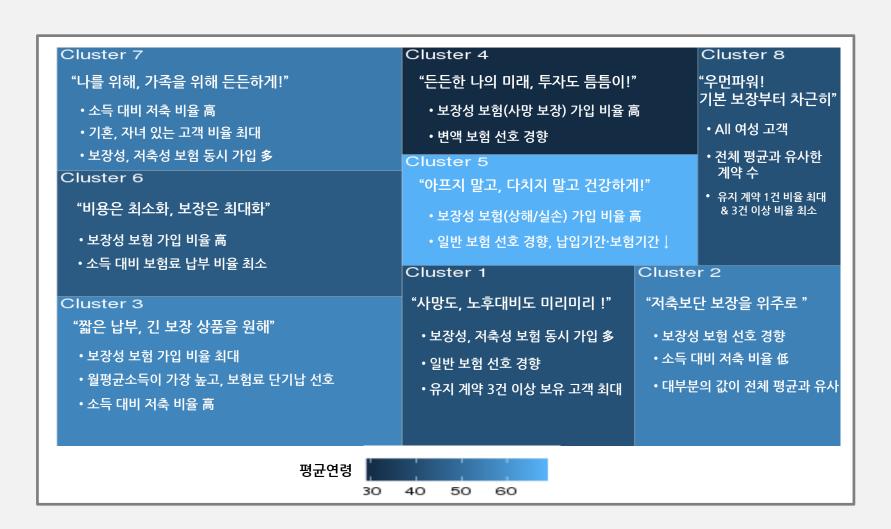


가입 상품 정보

	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 더 든든 보장보험	보장성	상해/실손/기타	일반	상세 확인
2	무배당 실손 의료비 보험	보장성	상해/실손/기타	일반	상세 확인
3	무배당 베스트라이프 교보 종신보험	보장성	종신/정기	일반	상세 확인
4	무배당 교보 행복한 준비보험	보장성	종신/정기	일반	상세 확인
5	무배당 교보 미리미리 CI보험	보장성	질병/CI/암	일반	상세 확인
6	무배당 교보 암보험	보장성	질병/CI/암	일반	상세 확인
7	무배당 생생 플러스 건강보험	보장성	질병/CI/암	일반	상세 확인



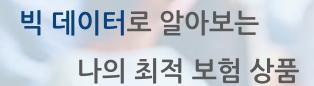
# 그 밖의 그룹들에 대한 특성 정리





나의 계약





알아 보기



















광화문에서 읽다 거닐다 느끼다



나의 계약

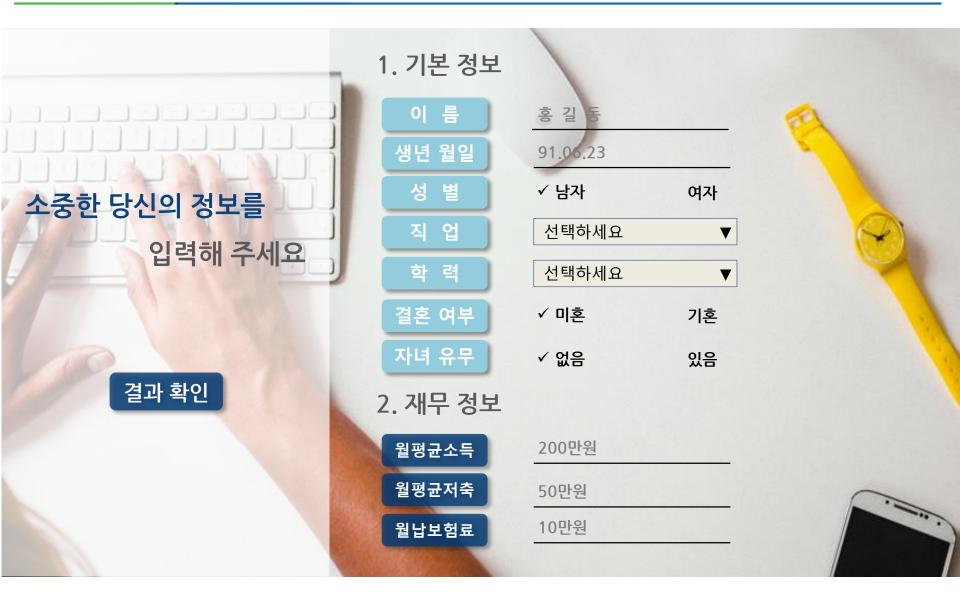
상품

든든라이프

고객센터

빅데이터 분석

퇴직연금 다이렉트



















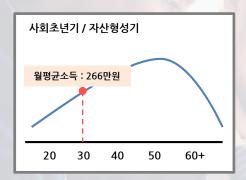
광화문에서 읽다 거닐다 느끼다



# 홍길동 고객님 빅 데이터 분석 결과

# "든든한 나의 미래, 투자도 틈틈이!" 유형 이십니다.

### 그룹의 평균 라이프 사이클



주로 이런 계약을 선호하십니다.

평균 1~2개의 상품 가입 유지계약 2건 보유 비율 최대 보장성보험 가입 비율 높음 납입기간: 20년납 선호 보험기간 : 종신 선호 변액보험 선호

### 주로 이런 분들이 속해있습니다.







퇴직연금 다이렉트

직 업

결혼 여부

### 다음 상품들을 추천해 드립니다.

	상품명	상품 유형	상품 종류	변액 여부	상품 정보
1	무배당 교보 미리미리 CI보험	보장성	질병/CI/암	일반	상세 확인
2	무배당 교보 베스트플랜 CI보험	보장성	질병/CI/암	일반	상세 확인
3	무배당 교보 미리미리 여성 CI보험	보장성	질병/CI/암	일반	상세 확인
4	무배당 교보 하이브리드 변액 종신보험	보장성	종신/정기	변액	상세 확인
5	미리 보는 내 연금 무배당 교보 변액 연금보험॥	저축성	연금	변액	상세 확인
6	무배당 교보 베스트플랜 CI보험	보장성	질병/CI/암	일반	상세 확인
7	더 드림 (무) 교보변액연금보험Ⅲ	저축성	연금	변액	상세 확인

















# 정확성 향상 및 신선함(Serendipity) 제공 필요

- 추천 시스템의 핵심 요건
  - ✓ 정확성 (Predictability) : 개인의 선호에 부합하는가?
  - ✓ 다양성 (Diversity) : 신선한 아이템을 추천할 수 있는가?
  - ✓ 확장성 (Scalability) : 확장 가능한 현실적인 시스템인가?
  - 앞선 추천 방법은 정확성과 다양성을 충분히 만족했다고 보기 어려움
- 해결 방안

