

8회 KCB 금융스타일 시각화

김하영, 배현주, 이승훈, 정회빈, 조윤영

- 1 8회 대회 참여 동기
- 2 8회 대회 소개
- 3 8회 대회 데이터 소개
- 4 분석을 하며 어려웠던 점
- 5 해결과정
- 6 대회를 마친 소감



- 이름: 정회빈
- 전공: 심리학/통계학(학사),
통계학(석사)
- KCB 인턴 (2019.08~2019.09)
- KCB 근무 중 (2019.10~)

1. 8회 대회 참여 동기

- 취업 준비를 하면서 공모전 수상의 필요성을 느낌
- 평소에 접하기 어려운 금융 데이터를 다뤄볼 수 있는 기회
- 시각화 공모전이어서 다른 공모전들과 차별화



허명회 (Huh, Myung-Hoe)

직위 Professor

연구분야 Exploratory Multivariate Data
Analysis, Data Visualization

지도교수님 연구분야가...

2. 8회 대회 소개



- KCB 데이터를 통한 금융 스타일 시각화
- KCB 전국민 금융 데이터의 새로운 인사이트 발굴
- 활용도 높은 결과물로 금융친화적 콘텐츠 발전에 기여

2. 8회 대회 소개

KCB

컨설팅
사업본부

디지털
사업본부

CB
사업본부



3. 8회 대회 데이터 소개

1. credit_card_data

- ▶ 전국민 카드 및 대출 이용통계 데이터
- ▶ 성별/연령대, 지역/연령대로 구분
- ▶ 신용등급, 카드이용금액, 대출업종별 대출금액, 거래형태별 대출금액, 신용/담보 대출금액

2. Jeju_financial_life_data

- ▶ 제주도의 우편번호 단위 지역민의 금융생활 통계 데이터
- ▶ 연소득, 소비액, 대출, 부동산, 직업

4. 분석을 하며 어려웠던 점

1. credit_card_data

- ▶ 각 변수의 고객 그룹별 평균값 데이터만 있기 때문에 변수의 정확한 분포를 알 수 없음
- ▶ 지역과 성별이 동시에 주어져 있지 않아서 '서울에 거주하는 20대 남성은 000한 특징이 있다' 같은 해석은 불가능

2. Jeju_financial_life_data

- ▶ 우편번호 단위별로 각 변수의 평균값이 제시되어 있어 특정 연령대, 성별 등을 기준으로 한 해석이 어려움

1. 고객 그룹의 전반적인 특성을 시각화 하는데 집중

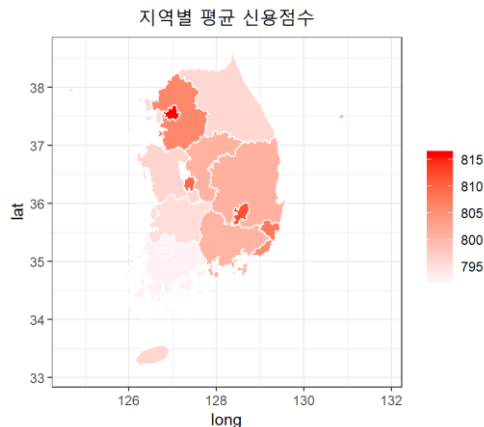
- ▶ 지도시각화, 지역x연령대/성별x연령대 특성 시각화, 주성분분석, 클러스터링
- ▶ 기본적인 것을 잘 하자

2. 실제 서비스에 활용될 수 있을 것 같은 결과물 제시

- ▶ 클러스터링을 기반으로 한 고객 페르소나 생성
- ▶ 금융 서비스 추천 시스템

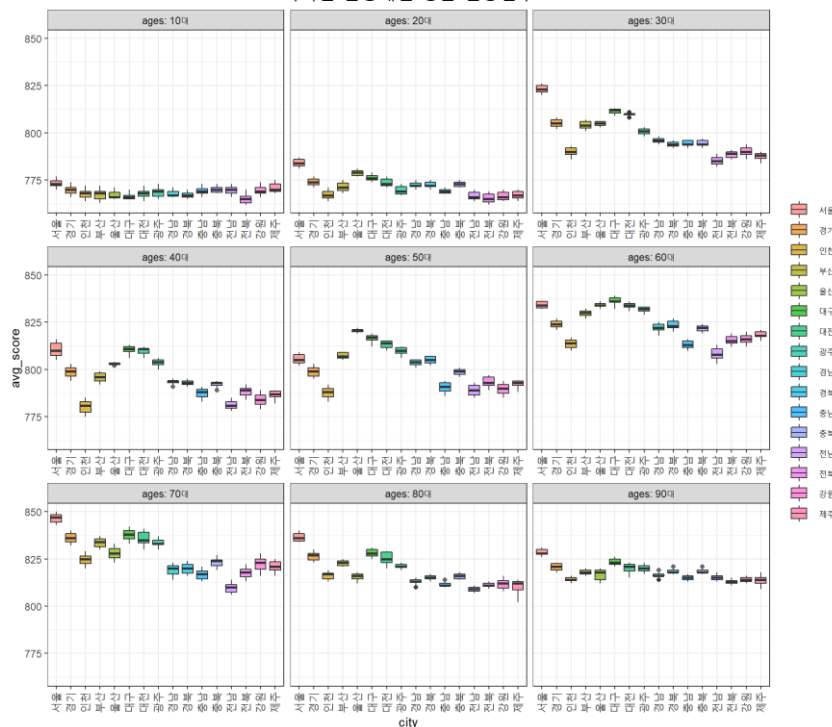
5-1. CREDIT_CARD_DATA : 변수 시각화

시각화 예시) 신용점수 변수 시각화



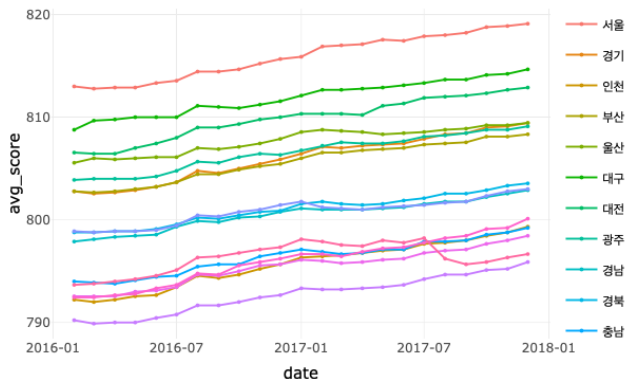
- 지도 시각화를 통해 지역별 평균값을 한눈에 비교해 볼 수 있음
- 상자그림을 통해 연령대별 편차와 지역별 편차를 한번에 볼 수 있음

지역별 연령대별 평균 신용점수

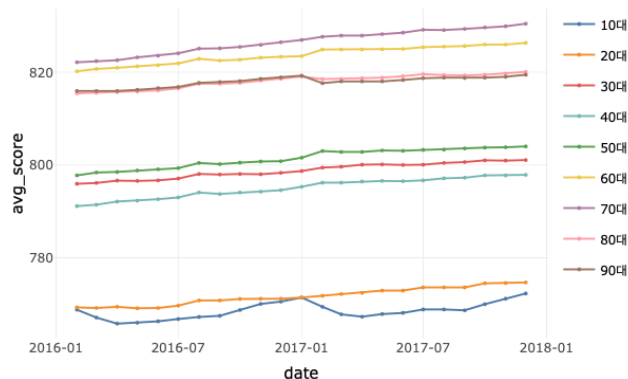


추세선을 통한 변수의 시각화

지역별 평균 신용점수 추세

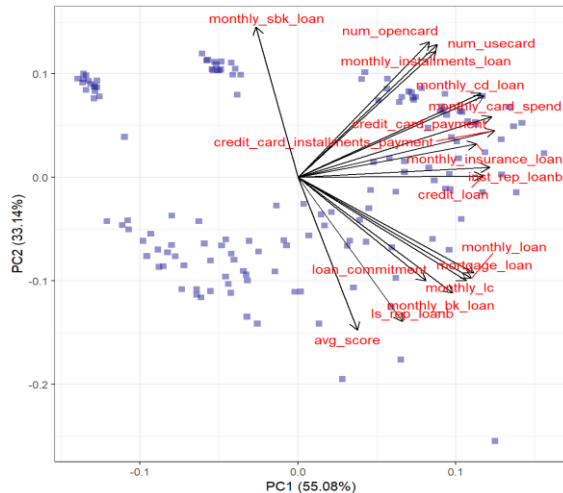


연령대별 평균 신용점수 추세

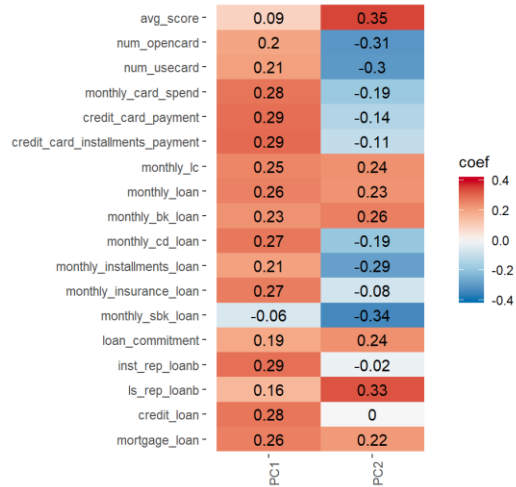


- 지역별, 연령대별로 추세선을 그려 시간에 따른 전반적인 패턴과 추세를 벗어나는 특이 패턴을 관찰할 수 있음
- 평균 신용점수의 지역별 평균값은 시간이 지남에 따라 증가하는 패턴
- 평균 신용점수의 연령대별 평균값은 시간이 지남에 따라 완만하게 증가하는 패턴
- 제주의 경우 2017년 7월 이후에 감소하는 패턴
- 10대에서는 감소와 증가를 반복하는 패턴

5-1. CREDIT_CARD_DATA : 주성분 분석

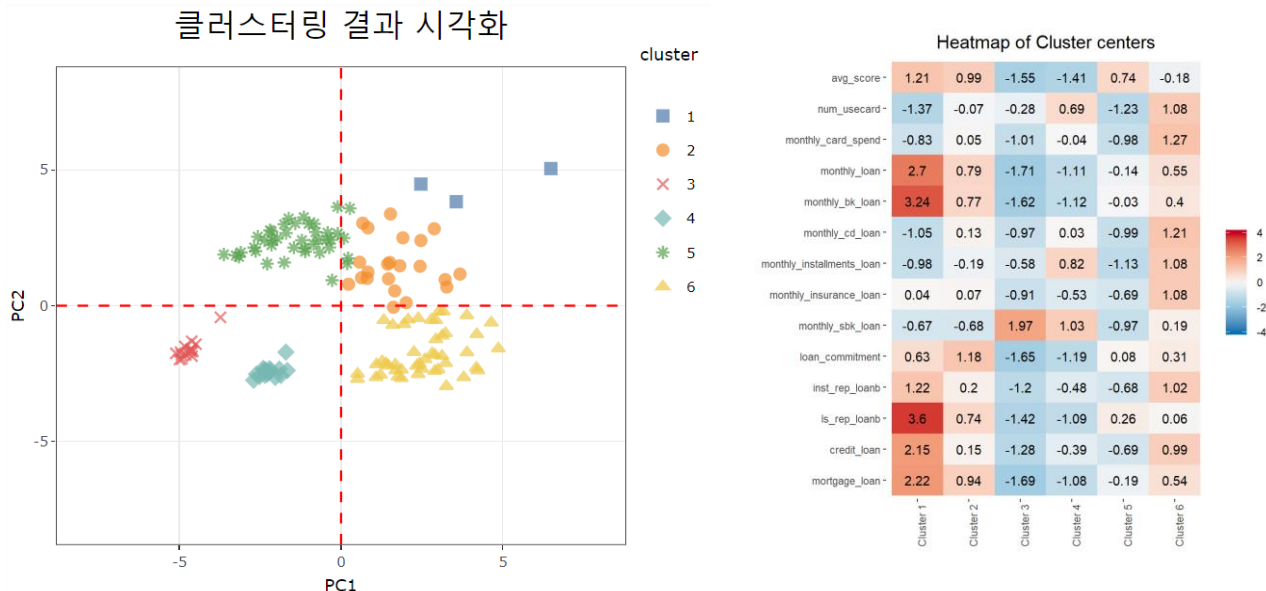


Heatmap of Principal Components



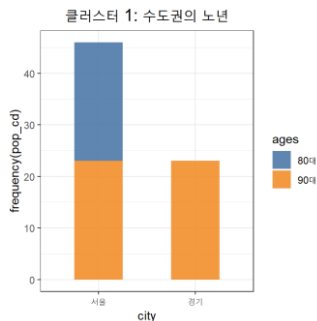
- Biplot(왼쪽)에서 화살표가 같은 방향일수록 양의 상관, 반대 방향일수록 음의 상관
- Heatmap(오른쪽)에서는 신용평점이 증가하는 방향으로 해석하기 위해 PC2의 부호를 바꾸어 표시
- PC1은 데이터를 전반적으로 설명해주는 축이고, PC2는 고객그룹을 신용평점에 따라 구분해주는 축

5-1. CREDIT_CARD_DATA : Clustering



- 클러스터는 지역보다는 연령대별로 나누어지는 모습을 보임(지역별, 연령대별 고객코드 시각화 참고)
- 클러스터 중심 값(클러스터 별 평균값)을 통한 클러스터 별 특징 해석 할 수 있음
- 평균이 0이 되도록 스케일링 후 클러스터링을 한 것이기 때문에 클러스터 중심 값이 0보다 크면 평균보다 큰 것으로 해석할 수 있고, 0보다 작으면 평균보다 작은 것으로 해석할 수 있음

5-1. CREDIT_CARD_DATA : Clustering



클러스터	신용	카드이용	대출	1금융 vs 2금융	저축은행	분할상환 vs 일시상환	신용 vs 담보
1	good	low	high	1금융	low	일시	both
2	good	low	high	1금융	low	일시	담보
3	bad	low	low	2금융(저축은행)	high	분할(평균이하)	신용(평균이하)
4	bad	high	low	2금융	high	분할(평균이하)	신용(평균이하)
5	good	low	soso	1금융(평균이하)	low	일시	담보(평균이하)
6	soso	high	high	2금융	middle	분할	신용(평균이하)

- 클러스터 별 중심값을 확인하여 금융변수의 특징을 확인한 후, 구성원들의 인구통계적 특징을 확인하여 최종적으로 명명

Cluster 1 수도권의 노년

서울, 경기 지역의 80, 90대가 포함되어 수도권 노년이라고 명명

Cluster 2 제2의 전성기

전국의 50대~70대가 포함되었고, 신용도가 좋으며, 1금융을 이용하는 것으로 판단되어 자녀를 독립시키고 제2의 전성기가 시작되는 클러스터로 명명

Cluster 3 서투른 10대

전국의 10대가 전부 포함되었으며, 저축은행 대출만이 높은 것으로 확인하여 아직 대출을 받기 어려운 서투른 10대로 명명

Cluster 4 전전금공20대

전국의 20대가 전부 포함되었으며, 아직 안정적인 금융생활에는 때 이른 전전금공 20대로 명명

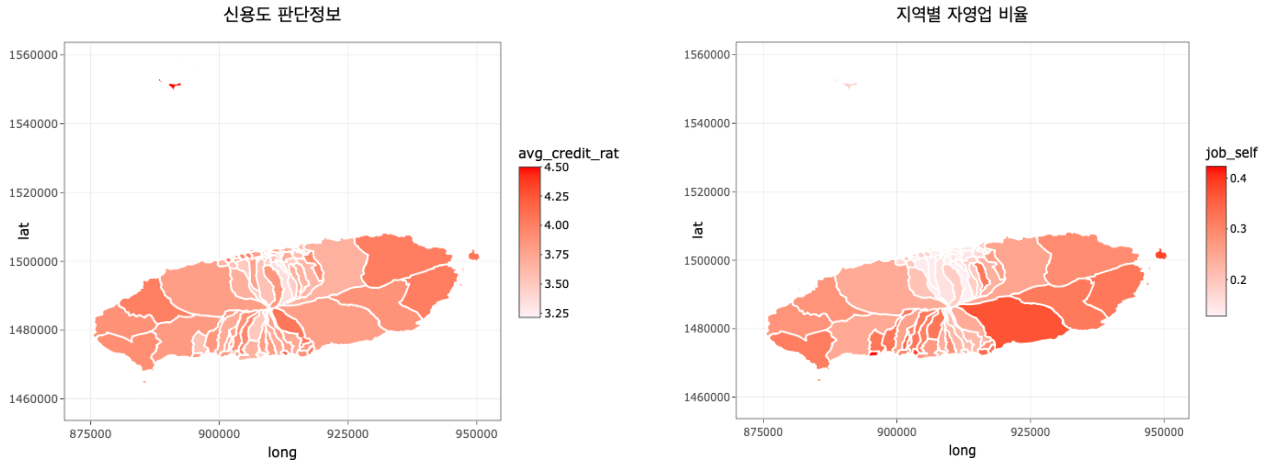
Cluster 5 비수도권의 노년

전국의 60대~90대가 포함되었고, 80, 90대가 수도권 노년과는 상이한 중심값을 보여 비수도권 노년이라고 명명

Cluster 6 일이 많은 중장년

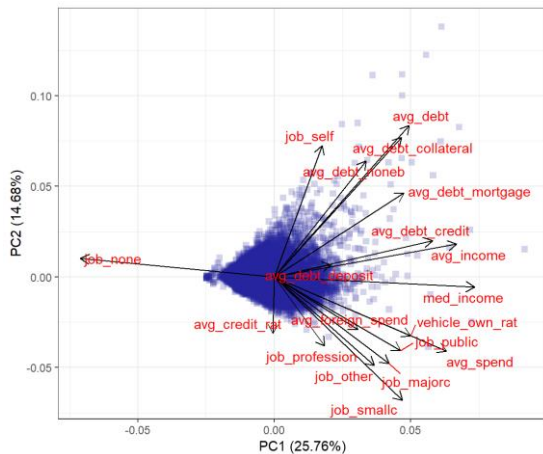
전국의 30대~50대가 포함되었으며, 결혼, 자녀 학비 등 인생의 중대사가 가장 많은 구간이라고 판단하여 일이 많은 중장년으로 명명

5-2. JEJU_FINANCIAL_LIFE_DATA : 변수 시각화



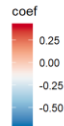
- jeju_financial_life_data.csv에 있는 zip_cd(우편번호)를 활용해 법정동명과 법정동코드를 새로운 데이터셋으로 생성한 후 읍면 우편번호 파일과 결합
- 생성된 데이터셋 기준으로 지역별 신용도 판단 정보, 소득, 소비, 대출, 주택 및 자가용 재산, 직업 변수의 평균을 지도 위에 시각화

5-2. JEJU_FINANCIAL_LIFE_DATA : 주성분 분석



Heatmap of Principal Components

	PC1	PC2	PC3	PC4
job_majorc	0.2	-0.23	0.03	-0.1
job_smallc	0.23	-0.33	0.11	-0.25
job_public	0.23	-0.2	-0.04	0.01
job_profession	0.09	-0.19	0.24	0.4
job_self	0.09	0.35	-0.5	-0.01
job_none	-0.35	0.05	0.3	0.04
job_other	0.18	-0.24	0.18	0.29
avg_income	0.32	0.09	-0.24	0.08
med_income	0.36	-0.03	-0.21	0.02
avg_spend	0.31	-0.2	0.08	0.12
avg_foreign_spend	0.15	-0.14	0.21	0.11
avg_debt	0.24	0.41	0.18	-0.04
avg_debt_credit	0.28	0.1	-0.21	0.11
avg_debt_noneb	0.16	0.31	0.35	-0.08
avg_debt_mortgage	0.23	0.22	0.33	-0.11
avg_debt_deposit	0.1	0.03	-0.05	0.17
avg_debt_collateral	0.23	0.37	0.29	-0.08
avg_credit_rat	0	-0.15	0.04	-0.7
vehicle_own_rat	0.24	-0.16	-0.04	-0.29



- 무직 비율(job_none)과 중위 연소득(med_incom)은 강한 음의 상관
- 평균 채무보유액(avg_debt)과 평균 연소득(avg_income)은 강한 양의 상관

PC3 값이 클수록

- 무직의 비율이 높고, 신용대출을 제외한 대출의 채무 평균 보유액이 높은 반면
- 평균 및 중위 연소득이 낮음

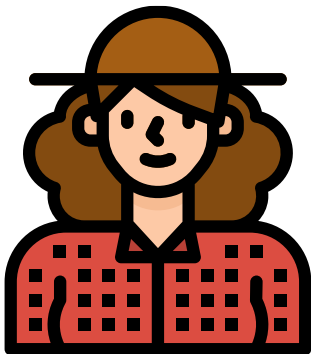
PC4 값이 클수록

- 전문직의 비율이 높은 반면
- 신용도 평균 관리지수와 중대형 차량 소유비율이 낮음

Persona 목적

- 분석을 바탕으로 credit_card_data의 각 클러스터 대표와 제주도 도민의 페르소나를 생성
- 각 페르소나는 클러스터의 가장 두드러지는 특징을 담아 고객별 특징을 깊게 이해하도록 함

Persona 예시



이름 : 김에스더 (만 38세)

거주지 : 제주도 남원읍

직업 : 70만 구독자 보유 유튜버 및 프리랜서 투어가이드

가족 : 이혼 후 자녀는 없이 리트리버 두마리와 함께 마당 있는 집 거주

고민 : 불규칙적인 수입, 노후준비, 콘텐츠 만드는데 비용 증가

목표 : 유기동물 보호소 설립

" 불규칙적인 수입을 어떻게 관리해야 할까? "

금융서비스 추천 시스템

- 1) 고객이 최소한의 정보를 제공하는 경우, 기존의 데이터에서 얻어진 클러스터 기반으로 추천 상품 제공
 - 고객은 본인의 기본 정보(연령과 지역) 제공
 - 고객이 가볍고 부담없이 본인에게 맞는 상품을 확인 가능
- 2) 고객이 더 자세한 정보를 제공하는 경우, 포트폴리오 제공
 - 고객은 자세한 정보 (본인의 금융 상황, 대출 이유와 원하는 대출 금액 등) 제공
 - 제공된 정보로 보다 구체적인 서비스 → 현 상황에 맞는 다양한 상품 추천, 자산 관리 방법 등을 제공
 - 고객이 속하는 클러스터 내에서의 고객의 분위수 제공
 - ✓ 분위수는 소득 및 자산(집, 자동차 등)을 기준으로 정해짐
 - ✓ 같은 클러스터에서 고객과 비슷한 상황에 있는 다른 사람들이 이용하는 상품 알림

예시)



5-3. 결론 : 올크레딧 서비스 화면

정희빈님

2019년 11월 리포트

- 나의상태 : 연령 20대, 소득 만원대, 신용평점
정희빈 님의 신용이용패턴을 신용평가사 분류하는 네가지 기준으로 분석한 결과입니다.
본인의 상태 및 신용 이용을 타인들과 비교·확인해보시기 바랍니다.

나 vs 유사그룹

신용거래의 종류 및 행태
(상품별 계좌건수, 월용비용 등)



신용형태 >

최대차이

일시불 이용 금액 비율

나 vs 상위그룹

채무의 적시 상환 여부 및 그 이력
(연체 정보 등)



상환이력 >

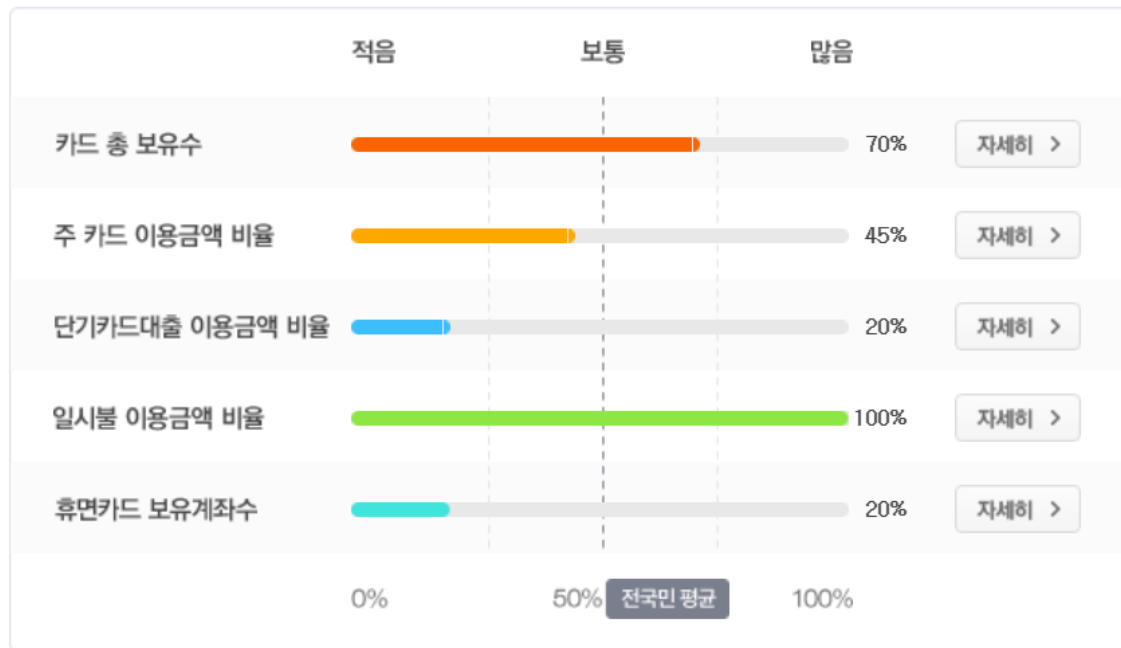
최대차이

총 연체건수

나 vs 전국민

5-3. 결론 : 올크레딧 서비스 화면

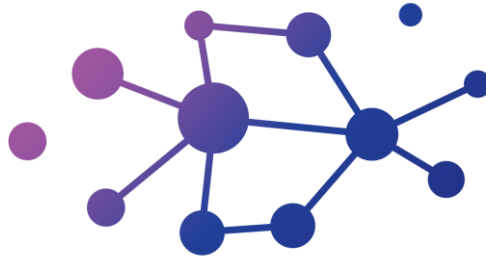
분석상세



6. 대회를 마친 소감

- ▶ 데이터 분석을 하려면 데이터가 있는 곳으로 가야 한다.
- ▶ 다양한 분야의 공모전에 참여하다 보면 나와 맞는 산업분야를 찾을 수 있다.
- ▶ 공모전 수상이 면접에서 분석 경험을 이야기 할 때 확실히 도움이 된다.
- ▶ 팀원을 잘 만나야 한다.

THANK YOU



THANK YOU