

# DACON

## KCB 데이터를 통한 금융스타일 시각화 경진대회

권홍욱 안재일 이태우 이한송

## 목차

### 1 전국 데이터

1-1. 전국 데이터 EDA

1-2. 전국 데이터 분석

### 2 제주 데이터

2-1. 제주 데이터 EDA

2-2. 제주 데이터 분석

### 3 요약 및 결론

## 개요 및 방향성

#### STEP 1

Top-down 방식으로 큰 feature에서 작은 feature 순으로 insight 확인

#### STEP 2

금융 데이터 활용하여  
고객 편의 및 특성 확인을 위해  
clustering 진행

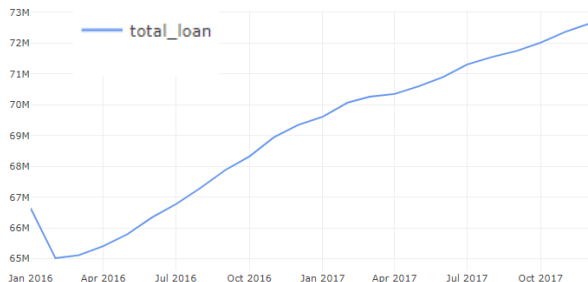
#### STEP 3

제주도 feature 확인 후  
지역별 issue 확인

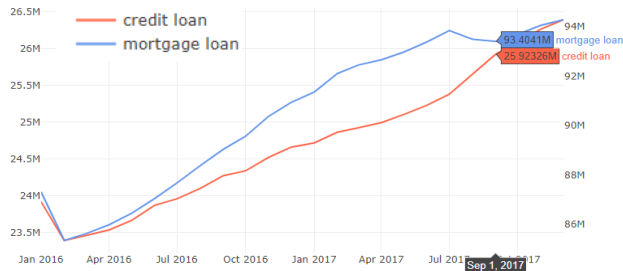
# 1-1. 전국 데이터 EDA

## 1) 총 대출 금액 경향 및 대출 유형별 시계열 분석

[ Figure1. 월별 총 대출 금액 변화]



[ Figure2. 월 평균 신용 대출 및 담보 대출 금액 변화]



### ✓ 총 대출금액 추세 분석 결과

- 2016년부터 2017년까지 전국민의 대출 금액이 증가하는 것을 확인

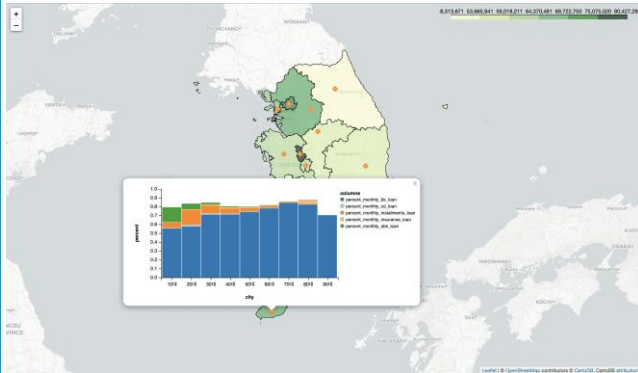
### ✓ 대출 상품별 추세 분석 결과

- 신용대출과 담보대출 모두 전반적으로 우상향
- 담보대출의 경우 2017년 7월부터 하락세를 보이며 두 달 뒤인 9월에 저점을 찍고 다시 상승하는 모습을 보임
- 이러한 경향은 지난 2017년 7월 3일부터 적용된 주택담보대출에 대한 대출 규제 강화에서 기인한 것으로 추정됨
- 담보대출이 감소한 기간동안 '풍선효과'로 인하여 신용대출의 증가율이 상승한 것으로 추정됨

# 1-1. 전국 데이터 EDA

## 2) 지역별 대출 금액에 대한 지도 시각화

[Figure3. 지역별 월별대출 평균 금액 및 연령대별 대출 이용 업종 비중]



지도 링크 (클릭해주세요)

- 지도상 색의 짙은 정도는 지역별 월별 대출 평균 금액을 표현
- 지역별 마커 클릭 시, 나타나는 그래프는 연령대별 대출 이용 업종 비중을 표현 (총 대출 금액을 대출 업종별로 차지하는 비중으로 나누어 확인)
- 지역별, 연령별로 세분화하여 확인한 결과, 모든 지역, 모든 연령에서 은행 업종 대출 금액이 가장 큰 비중을 차지하는 것으로 나타남

## 3) 지역별 대출금액 시계열 분석

[ Figure4. 지역별 월별 총 대출금액 변화 ]



- 지역별 총 대출 금액을 시계열 변화를 통해 확인
- 수도권 및 광역시 지역은 대부분 대출 금액 상위권에 위치한 반면, 도 단위 지역은 대부분 하위권에 위치한 것을 확인함.
- 이로 미루어 볼 때, 지역별 대출금액이 주택가격과 관련이 있을 것으로 예상됨
- 제주 지역의 대출 금액이 급증하는 것을 확인함

(“가계부채의 규모 중 상당 부분이 부동산과 연관이 있다”, 「주택담보대출 결정요인에 관한 연구: 개인특성과 지역특성간의 영향력 비교를 중심으로», 이찬영, 중앙대학교 대학원, 2017)

# 1-1. 전국 데이터 EDA

## 4) 지역별 분양가격과 대출금액의 상관관계

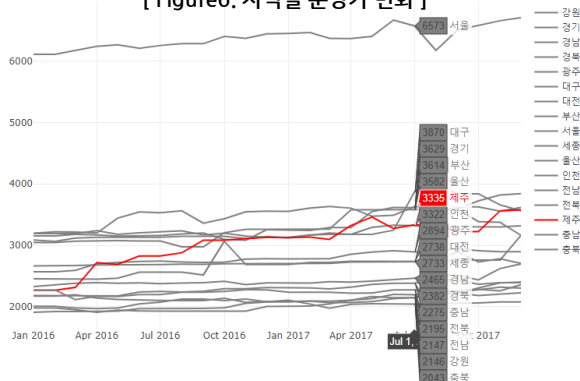
[ Figure5. 지역별 분양가격과 총 대출금액의 상관관계 ]

	서울	인천	경기	부산	대구	광주	대전	울산	강원	충북	충남	전북	전남	경북	경남	제주	평균
corr	0.81	0.53	0.76	0.84	0.58	0.88	0.75	0.79	0.46	0.37	0.61	0.88	0.94	0.94	0.68	0.93	0.734557

- 지역별 분양가격과 총 대출금액의 상관관계를 확인한 결과, 평균 약 0.73의 높은 상관관계를 나타냄. 특히 제주, 전북, 경남 등의 경우 매우 높은 상관관계가 나타남

## 5) 지역별 분양가

[ Figure6. 지역별 분양가 변화 ]

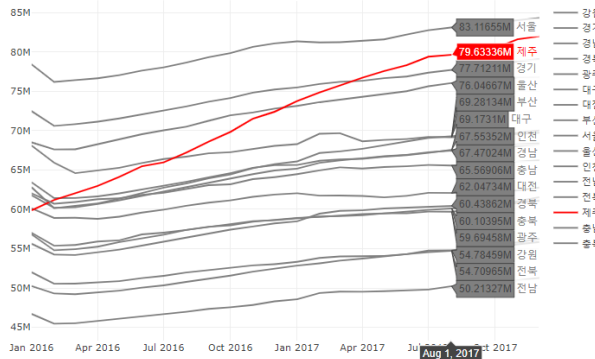


- 2016년 1월 대비 2017년 12월 분양가로 인상률을 확인한 결과 제주가 전국에서 가장 높은 인상률(157.76%)을 보임
- 제주 외 지역의 평균 분양가 인상률은 109.79%로 확인

데이터출처: 주택도시보증공사, 「민간아파트분양시장동향」, 지역별 규모별 m당 평균 분양가격

## 6) 지역별 대출금액

[ Figure7. 지역별 월별 총 대출금액 변화 ]



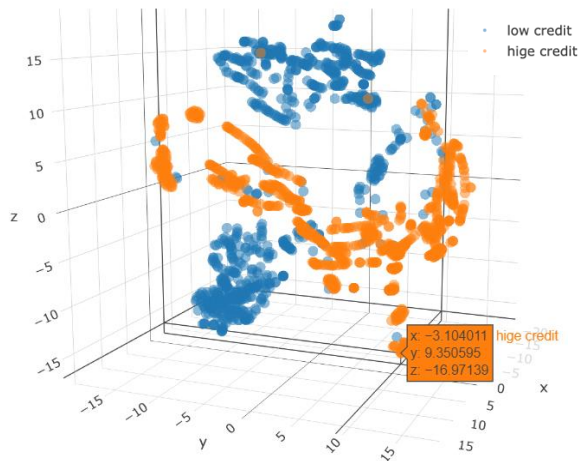
- 제주 지역의 월별 총 대출금액이 가장 두드러지게 증가한 것은 분양가격의 높은 인상률과 높은 상관관계가 있다고 볼 수 있음
- 따라서 분양가 변동이 예상되는 지역은 대출금액 또한 변화할 가능성이 있음을 나타냄

# 1-2. 전국 데이터 분석

## 1) 고신용/저신용 그룹 간의 비교 분석

- 고신용/저신용 그룹을 구분하는 데 가장 영향을 많이 미치는 변수를 확인하기 위해 신용점수에 따라 고신용(1분위수)과 저신용(3분위수)으로 분류

[ Figure8. 고신용/저신용 TSNE ]



- TSNE로 차원 축소하여 고신용과 저신용을 확인한 결과 군집화가 잘 되어있는 것을 확인 할 수 있음
- 따라서, 고신용 그룹과 저신용 그룹으로 나누어 그룹 간의 차이를 확인해 보기로 함

## 2) Feature Importance

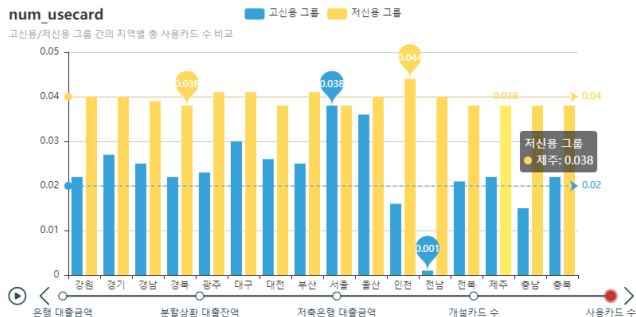
[ Figure9. RF의 변수 중요도 확인 ]

feature	importance
num_usecard	True
monthly_bk_loan	True
ls_rep_loanb	True
num_opencard	True
monthly_sbk_loan	True
population	False
loan_commitment	False
⋮	⋮
monthly_card_spend	False

- 고객그룹의 평균평점 (avg\_score)에 영향을 많이 미치는 변수가 무엇인지 확인하기 위해, Random Forest의 변수 중요도를 확인

- 종속 변수 : avg\_score

[ Figure10. 고객 그룹간 주요 변수 비교 ]

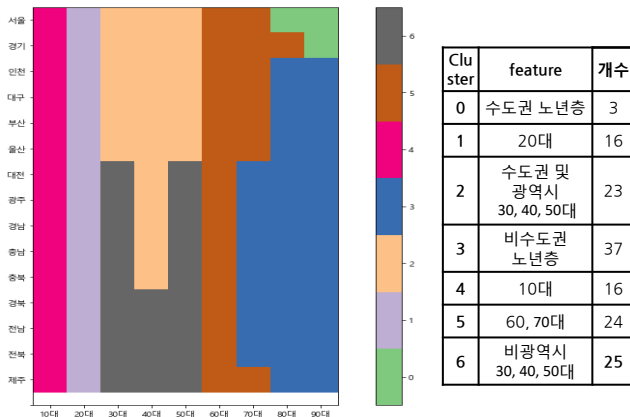


- 고신용 그룹이 저신용 그룹보다 은행 대출금액, 일시상환 대출잔액이 많음
- 저신용 그룹이 고신용 그룹보다 저축은행 대출금액, 개설 카드 수, 사용 카드 수가 더 많음

## 3) 지역별 연령별 군집 특성 비교 분석

- 지역별, 연령별 군집 특성을 확인하기 위해 K-Means 군집화를 함
- Elbow method를 통해 최적의 cluster 수는 7개로 확인

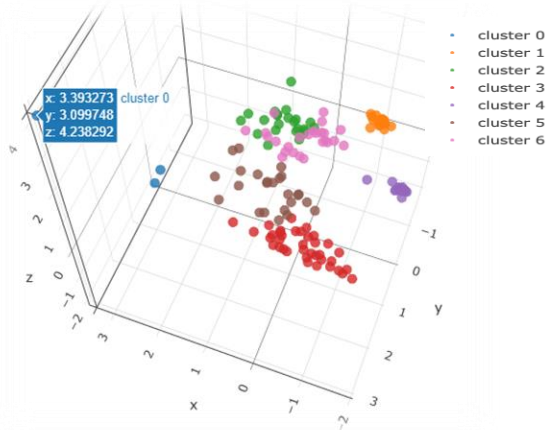
[ Figure11. K-Means clustering 결과 ]



- 연령대 뿐만 아니라 지역별 특성에서도 군집화가 이루어짐
- 따라서 군집별 금융활동 특성을 확인하면 각 군집에 적합한 금융 서비스를 제공할 수 있음
- 이에 Factor Analysis를 통해 변수를 3차원으로 축소하여 군집화를 시각화하고, 군집화에 영향을 미치는 변수들을 확인함

[ Figure12. Factor Analysis 결과 ]

- Factor analysis를 통해 얻은 각 축(x, y, z)에서 절댓값이 높은 값을 가진 변수를 바탕으로 추정해보면 x축은 대출, y축은 카드, z축은 신용정보로 이름을 붙일 수 있음

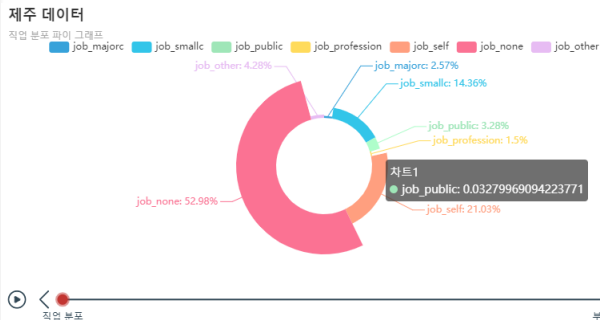


- Factor Analysis에서 K-Means 군집화가 잘 된 것을 확인
- 예를 들어, cluster0인 수도권 노년층 경우, 3개의 축에서 모두 높음. 즉, 대출(x축)과 양의 관계를 나타내므로 대출 금액이 많고, 카드(y축)과 음의 관계이므로 카드 사용이 낮은 신용정보(z축)에서는 신용이 낮음

## 2-1. 제주 데이터 EDA

### 1) 직업관련 및 부채관련 분포

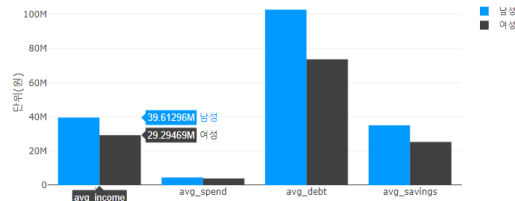
[ Figure13. 제주도의 각 직업군 및 부채 비율 비교 ]



- 직업관련 변수의 분포를 확인한 결과, 무직(52.98%), 자영업(21.03%), 중소기업(14.36%) 순으로 많은 비율을 차지함.
- 부채관련 변수의 분포를 확인한 결과, 물건담보대출(44.47%), 주택담보대출(31.84%), 비은행대출(14.09%) 순으로 많은 비율을 차지함.

### 2) 성별 소득/소비/부채

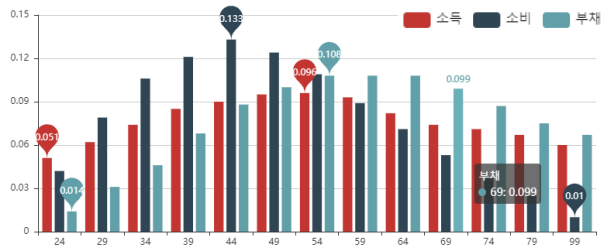
[ Figure14. 성별 소득, 소비, 부채 비교 ]



- 제주도 남성이 여성보다 소득/소비/부채/저축 모두 높음

### 3) 연령별 소득/소비/부채

[ Figure15. 연령별 소득, 소비, 부채 비교 ]



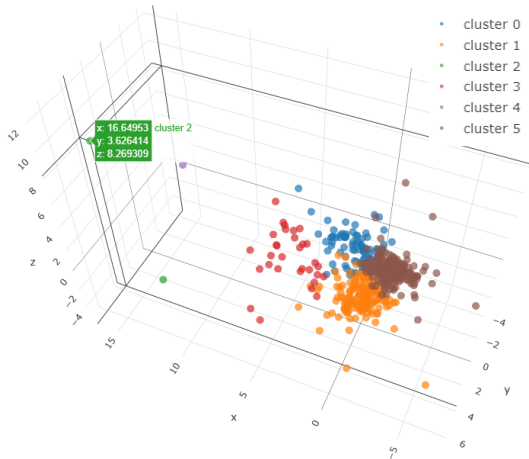
- 소득은 40, 50대가 높고, 소비는 30중반 ~ 40대가 높으며, 부채는 50대 ~ 60대중반이 높음



## 2-2. 제주도 분석

### 1) 제주 지역(우편번호)별 군집화

[ Figure16. 제주 K-Means clustering 및 PCA 결과 ]



- 총 474개의 우편번호(ZIP\_CD)를 기준으로 군집화 진행
- Elbow Method에 의거 최적의 클러스터 수를 6개로 K-Means Clustering을 진행한 결과, 군집화가 잘 이루어졌음을 확인할 수 있음

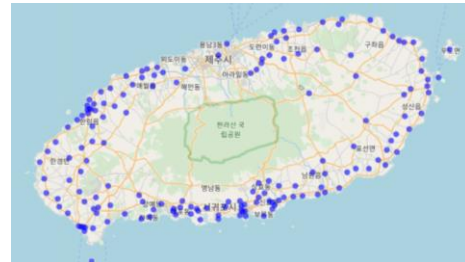
### 2) 주요 군집 시각화 및 특성 분석

[ Figure17. 제주 K-Means clustering의 Cluster 0 ]



- 제주시와 서귀포시 일부 지역에 집중된 것을 확인할 수 있음
- 공기업과 대기업 재직 비율이 높고, 국내/해외 소비가 평균 이상

[ Figure18. 제주 K-Means clustering의 Cluster 1 ]



- 제주시를 제외한 제주도 외곽지역에 분포함
- 0번 군집과는 반대로 자영업의 비중이 높고, 국내/해외 소비가 낮음

## 2-2. 제주도 분석

### 3) 인사이트 발굴

“제주농가 공급과잉으로 인한 부채규모 확대로

생산기반 붕괴 우려 지적 부채규모

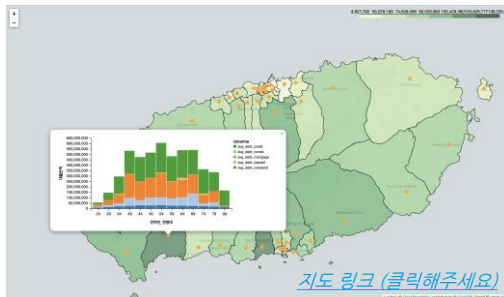
6,523만원으로 전국 평균대비 2.5배로 높아

농산물 브랜드화, 제주형 작물 관측시스템 고도화

및 자조금 사업 활성화 정책 강화 시급”

“제주농업 현황과 정책보완과제”보고서 “정책차론 제1호”  
제주특별자치도의회 정책연구실(2019.02)

[ Figure19. 지역별 평균 채무 보유액 및 농가집중지역의 부채 순위]



읍면동	경영체 수	부채순위
남원읍	4042.0	6위
애월읍	3857.0	15위
대정읍	2914.0	4위
구좌읍	2651.0	20위
한림읍	2554.0	11위
조천읍	2528.0	16위
성산읍	2391.0	14위
노형동	2137.0	10위
한경면	2086.0	3위
표선면	2050.0	13위

데이터출처 : 「농업경영체 현황(농가) : 지역별 농가 현황」, 농업경영체 등록정보 조회 서비스, 농림축산식품부

#### 1) 현황

- 제주농가는 과잉생산과 농산물가격의 약세로 인하여 소득의 불안정성이 높아지고 있음
- 이로 인해 부채규모가 확대되었고, 부채규모는 평균 6,523만원으로 전국 평균대비 2.5배로 높음

#### 2) 가설

- 농가의 부채규모가 확대되는 만큼, 농가가 많은 지역일수록 부채의 규모가 많을 것으로 예상됨

#### 3) 데이터 확인 결과

- 제주도의 전체 43개 행정동 중, 농가 수 기준 상위 10개 지역의 평균 부채 순위를 확인한 결과, 농가 수 상위 10개 지역 모두 부채기준 상위 그룹에 위치함
- 또한, 지역별 농가 수와 부채의 상관관계를 확인한 결과, 0.6860의 높은 상관관계를 보임
- 즉, 농가 수가 많은 지역일수록 부채가 많음을 확인할 수 있음

→ 이처럼 향후 대출을 포함한 금융 서비스, 또는 신용등급평가를 고려할 때도 위와 같은 대시보드를 통해 비슷한 인사이트를 얻을 수 있을 것으로 기대됨

## 전국 데이터 및 제주 데이터 요약 및 결론

- ✓ **전국민의 총 대출 금액과 대출 상품별 추세 분석 결과**
  - 2016년부터 2017년까지 전국민의 대출 금액이 증가
  - 일정기간 동안 담보대출에서 하락세가 나타났던 것은 주택담보대출에 대한 규제 강화에서 기인한 것으로 추정
  - 담보대출이 감소한 기간 동안 '풍선효과'로 인하여 신용대출의 증가율이 상승한 것으로 추정
- ✓ **지역별 분양가격과 지역별 대출금액의 상관관계 분석 결과**
  - 2016년 1월 대비 2017년 12월 분양가 인상률과 총 대출금액 인상률은 제주가 전국에서 가장 높음
  - 지역별 분양가격과 총 대출금액은 평균 약 0.73의 상관관계를 나타냄
  - 분양가 변동이 예상되는 지역은 대출금액 또한 변화할 가능성이 있을 것으로 추정
- ✓ **지역별 연령별 군집 특성 분석 결과**
  - K-means clustering 결과 지역별, 연령별 특성에 따라 7개의 군집으로 분류
  - Factor analysis 결과 3차원 상에서 각각의 축은 대출, 카드, 신용정보로 추정 가능
  - 예를 들어, 수도권 노년층의 경우 대출 금액이 많고, 카드 사용과 신용이 낮음
- ✓ **제주 지역(우편번호)별 군집 특성 분석 결과**
  - K-means clustering 결과 지역(우편번호)별 특성에 따라 6개의 군집으로 분류
  - 예를 들어, 제주시를 제외한 제주도 외곽지역은 자영업의 비중이 높고, 국내/해외 소비가 낮음
- ✓ **제주 지역별 채무 보유액과 농가 수의 상관관계 분석 결과**
  - 총 대출금액 인상률이 전국에서 가장 높은 제주에서 지역별 농가 수와 부채의 상관관계는 약 0.69로 높은편임
  - 농가 수가 많은 지역 채무 보유액이 많을 것으로 추정

“향후 금융 서비스 및 신용등급평가 등에서 위와 같은 분석을 통해 다양한 인사이트를 얻을 수 있을 것으로 기대”