SOL TO NEXT

금융 통합 어플의 개선 필요성과 증권의 기능적 발전 방향

신박한투자팀



1. 배경 및 필<u>요성</u>

- 현대 금융사들이 추구하는 공통된 방향을 생각
- 그룹사 별 교차거래 비교
- 컬럼 정리

2. 신사업 아이디어

- 데이터 분석 결과
- 아이디어 제안
- 실현 방안

3. 기대효과

- AUTO ML을 통한 기대효과 예측모델 실현
- 상관분석을 통한 분석 결과
- 증권거래 향상 예측

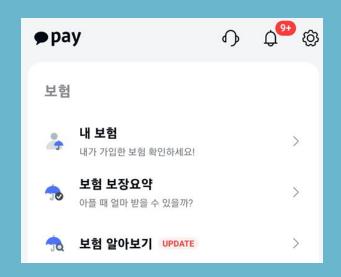
4. 보완가능성 & 회고

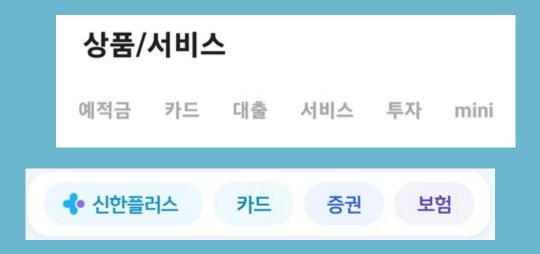
- 보완할점
- 회고

배경 & 필요성

현대 금융의 목표 : 통합 금융

삼성생명, 삼성화재, 삼성카드, 삼성증권이 함께 만든 모니모는 오늘도 이 가치를 펼치고자 합니다



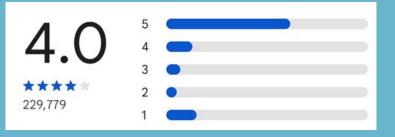


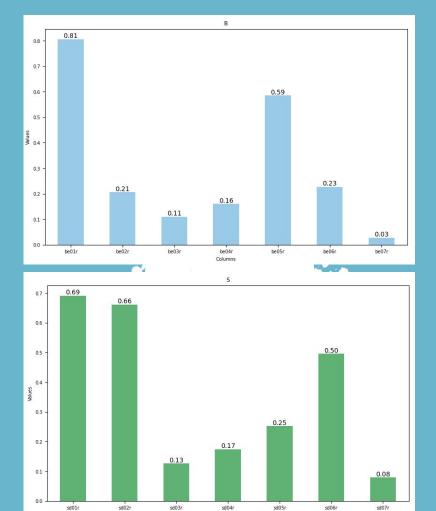
내 모든 금융 내역을 한눈에 조회하고 한 곳에서 관리하세요. 이제껏 경험 못 했던 쉽고 편리한 금융 서비스,











sd01r

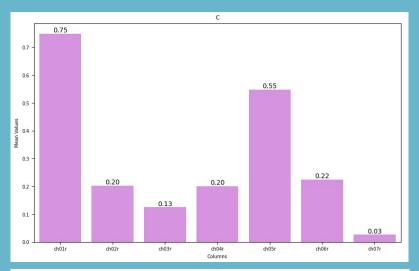
sd02r

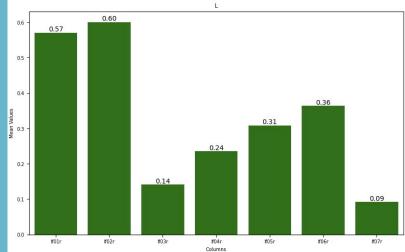
sd03r

sd04r Columns

sd05r

sd06r





교차 거래 분석 결과







4개의 케이스 모두 은행/카드/증권/라이프 전부를 활용하는 비율이 가장 적음. 은행과 카드가 비슷한 개형을 보이며, 서로를 같이 활용하는 경우가 많은 비율을 차지.

투자증권과 라이프가 비슷한 개형을 보이며, 투자증권의 경우 은행과 카드를 같이 활용하는 비율이 타 컬럼에 비해 높다.

투자증권이 성공적으로 개선된 통합 어플의 큰 주축이 될 가능성이 높음.

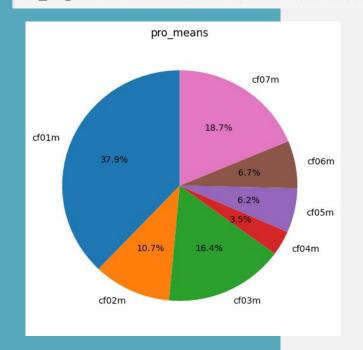
Column

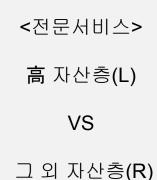
기준년월	pk1	거래경험 상품 개수	sa15r-as18r	
연령대	pk3	지역	pk4	
유통소비	cc01m-cc09m	종목별 거래경험	sa01r-sa14r	
대중교통	ce01m-ce05m	거래횟수	sb27r-sb32r	
레져	ch08m-ch14m	교차거래	be01r-be07r ch01r-ch07r sd01r-sd07r lf01r-lf07r	
고객유형	sc01r-sc05r			

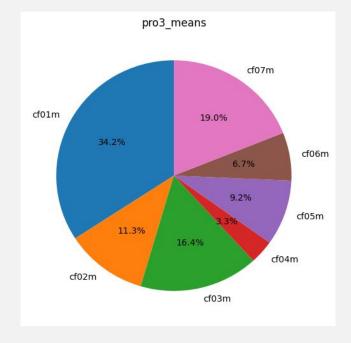


1. 高 자산층의 증권거래유도

```
df high = df.copy()
df_high = df[df['sb03m']>=df['sb03m'].mean()] 高 자산층 : 약 3만 rows
df_high # 고객가치등급을 예측을 위한 분류.
```







高 자산층 우위 비율

- · 매우 미미한 비율의 %point의 차이는 제외.
- 1. 유통소비: 백화점 0.2% / 이커머스 0.1% / 마트 0.4%

2. 생활서비스: 철도 0.8%

3. 레져: 실외골프장 1.4% / 스포츠센터 0.1%

4. 전문서비스: 교육(학원학습지) 3.7% / 치과 1%

高 자산층 증권 거래 경험 평균 비율

sa01r	0.780740
sa02r	0.768018
sa03r	0.072371
sa04r	0.391937
sa05r	0.118915
sa06r	0.079480
sa07r	0.035631
sa08r	0.026708
sa09r	0.012775
sal0r	0.031781
sallr	0.024411
sa12r	0.012007
sa13r	0.006802
sa14r	0.234105

- 高 자산층의 약 78%가 기준년월까지 주식 경험이 존재.
- top 5 상품 : 1. 국내 2. 공모주청약 3. 신용서비스 4. 수익증권(펀드) 5. 해외주식
- 아이디어 1: 큰 소득을 활용하는 높은 자산층이 많이 수요하는 분야와 항목들에 대해 일정 기간 동안 개인별 과거 신한카드 실적보다 많이 활용한 고객에 대한 MTS 차등 매매수수료를 적용. 개선된 통합어플의 활용성 증대를 위해 신한카드 그룹사와의 연계를 강조. 신한투자증권 모바일 주식 거래 수수료인 약 고정 0.189%, ETF/ETN/ELW 약 0.19% 포함 여러 수수료들을 과거 대비 초과 실적에 따른 동적 수수료를 적용하여, 수수료에 대한 부담을 줄이고 새로운 고객 유입을 도모.
- 고객가치등급 1등급, 2등급 이상의 고객 확보 및 충원을 목표.

2. 낮은 증권 거래량 지역에 대한 유인책

서울 ,부산,대구 등의 지역에 비해 울산, 경기, 인천 지역의 신규,활성 고객이 더 적다는것을 확인할 수 있다.

	신규	활성	비활성	무거래	이탈
서울	0.006334	0.204001	0.307366	0.044318	0.424708
부산	0.005392	0.155584	0.235013	0.041882	0.393739
대구	0.005012	0.148939	0.219050	0.038973	0.363932
인천	0.004513	0.135133	0.200057	0.037638	0.325805
광주	0.004553	0.137642	0.199273	0.038055	0.347029
대전	0.004988	0.141630	0.201174	0.033589	0.324281
울산	0.003399	0.141905	0.190929	0.035929	0.313536
경기	0.004238	0.139716	0.207343	0.049449	0.340968

2. 낮은 증권 거래량 지역에 대한 유인책

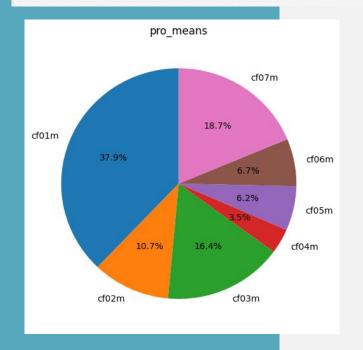
거래량을 보아도 서울, 부산, 대구 지역의 비율이 더 높고 상대적으로 인천, 울산, 대전 지역의 거래량이 낮게 나타남

아이디어2: 인천, 울산을 포함한 거래량이 낮은 지역에 대해 증권 거래 활성화를 위해 지역 지자체와 제휴하여 통합어플의 이름으로 행사후원을 하는 등 새로운 고객과 기존 고객에 대한 거래 복지 증가 방안을 도모.

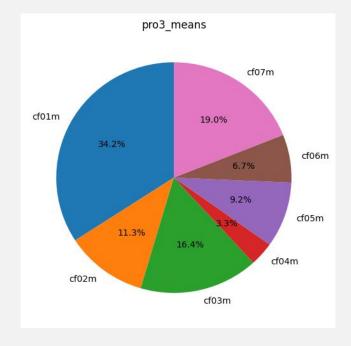
	0회	5회 이하	30회 이하	183회 이하	365회 이하	366회 이상
서울	0.421501	0.306402	0.140545	0.083041	0.016609	0.018754
부산	0.379029	0.246004	0.111108	0.067189	0.013360	0.015181
대구	0.359949	0.224590	0.103730	0.060772	0.012982	0.014015
인천	0.320528	0.190098	0.103014	0.063195	0.012616	0.013783
광주	0.342844	0.202239	0.096639	0.060506	0.011452	0.012771
대전	0.334743	0.189747	0.096377	0.058366	0.012516	0.014226
울산	0.322857	0.182780	0.101048	0.058690	0.010435	0.010369
경기	0.337302	0.198535	0.108299	0.067499	0.015232	0.014868

3. MTS 최적화: 증권분야 발전 방향

```
df high = df.copy()
df_high = df[df['sb03m']>=df['sb03m'].mean()] 高 자산층 : 약 3만 rows
df high # 고객가치등급을 예측을 위한 분류.
```



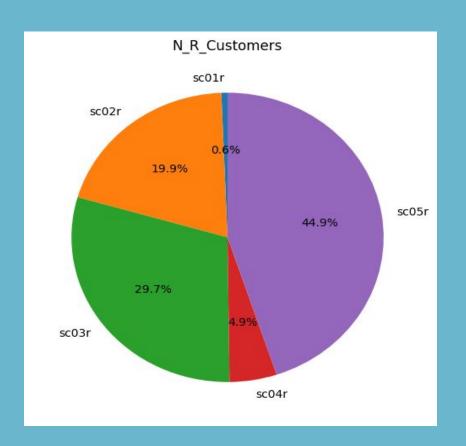
<전문서비스> 高 자산층(L) VS 그 외 자산층(R)



신규 및 복귀 고객에 대한 혜택

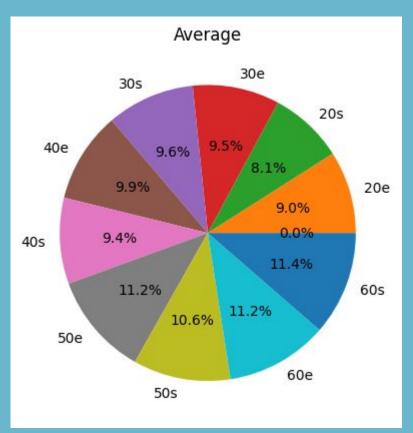
신규, 활성, 비활성, 무거래, 이탈

이탈 고객 즉, 증권 계좌를 개설하였지만 한번도 이용하지 않은 고객의 비율이 가장 높은데 해당 고객들에게 프로모션 이벤트를 진행하여 행사 기간 내에 거래를 시작하면 추첨을 통해 상품을 제공하는 이벤트를 진행.



나이대 별 창구거래 비율

60대후반, 60대 초반, 50대 초반, 50대 후반 연령대 분들의 창구거래 비율이 가장 높게 나타남을 볼 수 있다.



나이대 별 SOL이용 비율

20대 초반, 30대 후반, 30대 초반, 20대 후반 순으로 SOL이용 비율이 많았다.

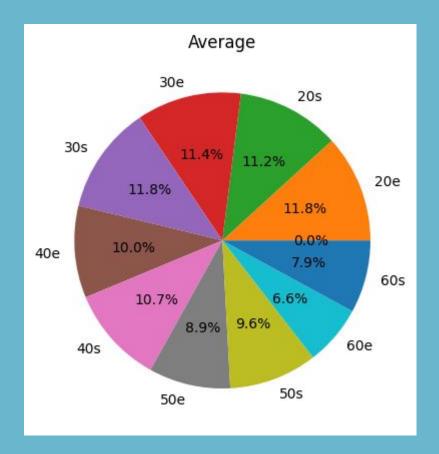
고향 좋은 상대적으로 어플이용자가 적기 때문에

직관적인 UI/UX 디자인:

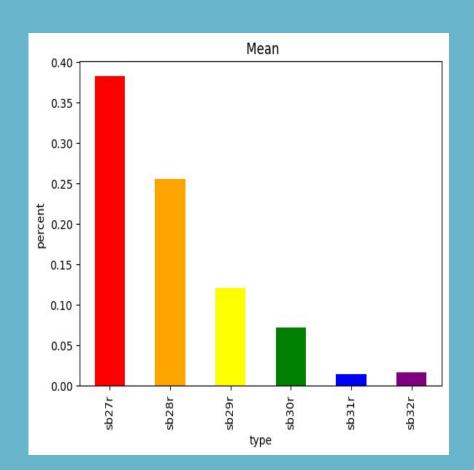
- 큰 글씨와 아이콘, 명확한 메뉴 구성으로사용의 편의성을 높입니다.
- 복잡한 기능은 최소화하고, 주요 기능에 초점을 맞춰서 디자인합니다.

도움말 및 튜토리얼 제공:

• 앱 내에 간단한 도움말이나 튜토리얼을 제공하여 초기 사용자가 앱을 쉽게 사용할 수 있게 도와줍니다.



12개월 증권거래 횟수



12개월동안 증권거래가 0회와 5회인 고객들이 가장 많은 비중을 차지함을 볼 수 있다.

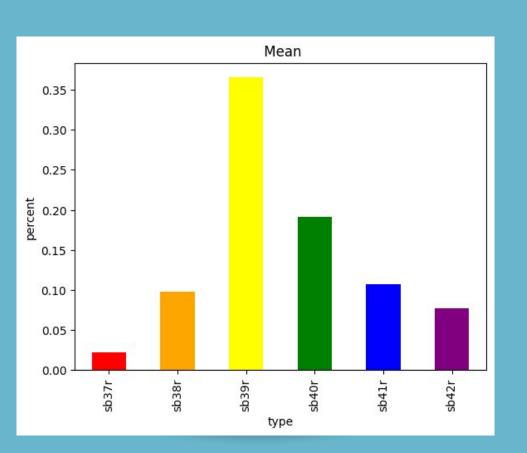
많은 고객들이 증권계좌를 개설 해놓고 자주 이용을 하지 않은 것을 확인할 수 있다.

해당 고객들은 투자에 대한 지식이 부족하거나, 시장에 대한 불안감을 가지고 있을 수 있기 때문에

다각화된 포트폴리오를 구성을 하여

 여러 종목이나 자산 클래스에 투자하여 리스크를 분산시킴으로써, 한 종목의 큰 손실이 전체 포트폴리오에 미치는 영향을 최소화할 수 있다.

거래유형



왼쪽부터

단타, 백화점포트폴리오형, 저자산 일반주 고객비율, 우량주 중심 안전형, cma중심 무거래, 초우량 고객 비율

저자산 일반주 고객

초보자용 투자 패키지: 쉽게 투자를 시작할 수 있도록 구성된 패키지나 포트폴리오를 제공합니다

우량주 중심 안전형 고객

우량주 리서치 리포트 제공: 주요 우량주에 대한 깊은 분석과 추천 리포트를 제공합니다.



```
dfdf_col = dfdf_temp[['pk3','pk4']]
 df = df[~df['pk1'].isin([
                                                               dfdf num = dfdf temp[num col]
           '# 085b1c27644a39070b400b519f4ef381 '.
           '# aecac25f95e764eceadfaa1e4efd04ca '.
                                                               from sklearn.preprocessing import RobustScaler
                                                               scaler = RobustScaler()
           '# 943655916746c71c58353309b4ca70af '])]
                                                               scaled dfdf num = scaler.fit transform(dfdf num)
 cat col = ['pk3','pk4']
 num col = ['bd20m',
          'cc01m', 'cc04m', 'cc02m', 'cc05m',
          'ce04m', 'ce03m', 'ce02m',
          'cf01m', 'cf02m', 'cf03m', 'cf07m', 'cf09m', 'cg10m',
          'sa01r', 'sa04r', 'sa05r', 'sa14r', 'sa16r', 'sa17r', 'sa18r',
          'sb01m', 'sb02m', 'sb03m', 'sb04m',
          'sb12r', 'sb13r', 'sb14r', 'sb15r',
          'sb28r', 'sb29r', 'sb30r', 'sb31r', 'sb32r', 'sb33r', 'sb34r', 'sb35r', 'sb37r', 'sb38r', 'sb39r', 'sb40r', 'sc02r', 'sc18r',
          'be02r', 'ch02r', 'ch06r', 'sd01r', 'sd02r', 'sd06r', 'lf03r',
         'sc06r']
                                                   Part I
X_train, X_test, y_train, y_test = train_test_split(lastdf, lastdf['sc06r'], test_size=0.2, random state=42)
model = RandomForestRegressor(n estimators=100, random state=42)
model.fit(X train, y train)
# Feature Importance 계산
feature importance = model.feature importances
cumulative importance = feature importance.cumsum()
# 설명력 70% 이상인 변수 추출
important features = data.columns[cumulative importance <= 0.7]</pre>
```

EDA -> 설명변수 채택(분석 방향의 기조를 유지) -> 결측치 및 이상치 정리 -> Robust Scaling 진행 -> Feature importance를 통해 설명력 70% 이상인 피쳐를 추출. -> Pycaret AUTO ML 진행.

		Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
	catboost	CatBoost Regressor	0.1796	0.1082	0.3288	0.8371	0.1520	0.3460	8.9300
	lightgbm	Light Gradient Boosting Machine	0.1781	0.1083	0.3289	0.8370	0.1517	0.3436	1.3350
	et	Extra Trees Regressor	0.1790	0.1088	0.3295	0.8364	0.1522	0.3477	13.6920
	gbr	Gradient Boosting Regressor	0.1810	0.1099	0.3313	0.8346	0.1540	0.3507	28.1290
	rf	Random Forest Regressor	0.1818	0.1107	0.3324	0.8335	0.1531	0.3541	52.1570
	lr	Linear Regression	0.1900	0.1127	0.3354	0.8304	0.1574	0.3752	0.1630
	ridge	Ridge Regression	0.1900	0.1127	0.3354	0.8304	0.1574	0.3752	0.1000
	br	Bayesian Ridge	0.1900	0.1127	0.3354	0.8304	0.1574	0.3752	0.2020
	xgboost	Extreme Gradient Boosting	0.1845	0.1145	0.3381	0.8278	0.1531	0.3567	0.5660

- Pycaret 결과 R-Squared, MSE, RMSE 가 가장 높은 catboost 모델을 선정 후 파라미터 튜닝을 진행하여 실제 데이터와 prediction에 대해 Regression 성능 평가지표 정리 시 R-squared = 0.841, MSE = 0.1053 으로 만족스러운 모델 완성.
- 이에 맞춰 아이디어 1 에 대해 과거 대비 얼마나 특정 산업에 카드 지출을 많이하고, 이에 맞는 차등 수수료 산정으로 평균적으로 몇 % 줄었냐에 따른 예상되는 1등급 비율의 상승을 추정할 수 있게 됨.

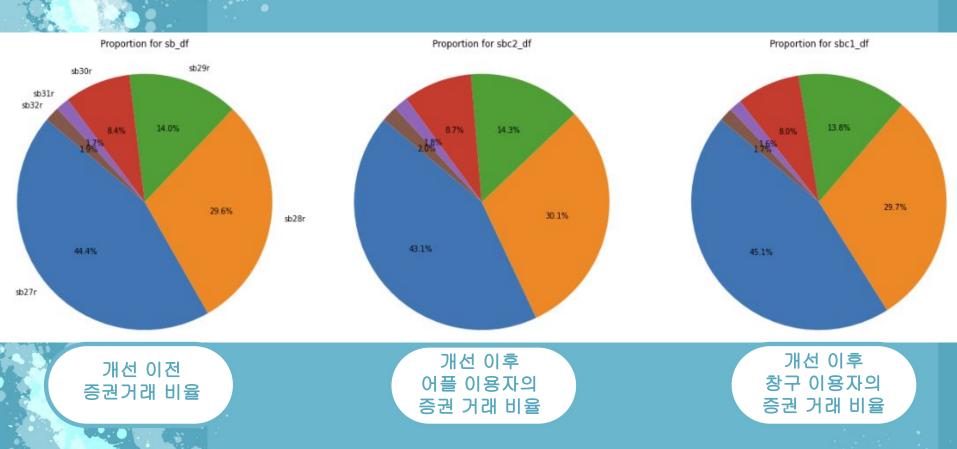
• 결측치에 대한 처리가 필요하며, 이번 #bc01r 참구거래 bco2r SOL 신한은행 어플 이용고객. #이 컬럼들에 대한 살관계수 분석. 경우는 타 컬럼값 대비 값이 작을 경우에 만들어 놓기에 0으로 대체 import scipy.stats as stats from scipy.stats import shapiro df sbbc = df[['sb27r','sb28r','sb29r','sb30r','sb31r','sb32r','bc01r','bc02r']] S 컬럼은 정규성을 만족하고, b 컬럼은 # 1. 결촉치는 0으로 대체. 만족하지 않는다. df sbbc copy = df sbbc.copy() for col in df sbbc copy.columns: 샘플의 수가 많고 이변량 정규분포를 df sbbc copy[col].fillna(0, inplace = True) 만족하므로 피어슨 상관계수를 측정하기 #2. 삼관분석을 위한 정규석 검정 위한 조건을 충족시킨다고 판단. for col in df_sbbc_copy.columns: 7 - 1.00 0.15 0.14 0.03 -0.01 -0.03 -0.09 0.25 , p value = shapiro(df[col]) alpha = 0.05- 0.8 0.15 1.00 0.29 0.19 0.08 0.08 0.13 0.14 if p value > alpha: 0.14 0.29 1.00 0.26 0.13 0.15 0.13 0.02 print(f"{col} 데이터는 정규분포를 따름.") - 0.6 else: s컬럼과 0.03 0.19 0.26 1.00 0.18 0.20 0.17 -0.11

#sb27r ~ sb32r 연간 평균 거래횟수에 대한것.

bc02r 데이터는 정규분포를 따르지 않음.

print(f"{col} 데이터는 정규분포를 따르지 않음.") b컬럼의 - 0.4 -0.01 0.08 0.13 0.18 1.00 0.13 0.10 -0.12 sb27r 데이터는 정규분포를 따름. 독립을 sb28r 데이터는 정규분포를 따름. 가정. -0.03 0.08 0.15 0.20 0.13 1.00 0.13 -0.16 - 0.2 sb29r 데이터는 정규분포를 따름. sb30r 데이터는 정규분포를 따름. -0.09 0.13 0.13 0.17 0.10 0.13 1.00 sb31r 데이터는 정규분포를 따름. - 0.0 sb32r 데이터는 정규분포를 따름. 0.02 -0.11 -0.12 -0.16 1.00 bc01r 데이터는 정규분포를 따르지 않음.

sb27r sb28r sb29r sb30r sb31r sb32r bc01r bc02r



통합어플의 개선은 기존 SOL 고객들을 이어받아 증권 거래에 있어 이탈자를 줄이고 평균 거래량을 올리는 효과를 기대할 수 있다.



보완가능성 및 회고



1. 컬럼에 대한 한계

내부인이 아니기에 컬럼에 대한 자세한 정보 파악이 힘들다는 한계점이 아쉬움.



2. 데이터 전처리

결측치 및 이상치에 대한 정보가 부족했음. 도메인의 영역에 따라 이를 처리해야 하지만 정보부족 및 데이터의 분포가 정규성을 따르기 힘든 태생적 한계가 있었음.



3. 데이터 셋의 문제

컬럼 값들에 대해 수치보다는 상대적인 비율이 많아 상대적으로 어느정도 많다 혹은 이정도의 파이를 차지하고 있다 이정도로 밖에 해석할 수 없는게 아쉬웠음.

외부인으로써 내부 데이터를 분석하는 데 있어 많은 노력과 한계를 느꼈으며 분석에 있어 여러 경험을 할 수 있던 소중한 경험.

