

데이터 분석 데브코스 3기

2차 프로젝트 ~ 파이썬 기반 데이터 시각화 ~

신용카드 데이터를 활용한 이탈 요인 분석 및 고객 유치 방안 도모

TEAM 임정 1조

김희령, 오수연, 황석준

1. 서론.....	2
1.1. 데이터 선정 배경 및 목표.....	2
1.2. 데이터 설명.....	2
1.3. 데이터 전처리.....	4
2. 본론.....	6
2.1. 가설 설정.....	6
2.2. 고객 사용 내역 기반 분석.....	8
2.3. 인구통계변수 기반 분석.....	22
2.4. 가설 기반 컬럼 제작.....	36
3. 결론.....	50
3.1. 결론 및 기대 효과.....	50
3.2. 참여자 정보 및 역할.....	50
4. 참고문헌.....	51

1. 서론

1.1. 데이터 선정 배경 및 목표

Credit Card Customers Prediction

신용 서비스의 경우 전 세계인이 사용하고 있다고 주장해도 무방하다. 참고한 데이터셋의 배경이 되리라 추측할 수 있는 미국은 성인의 82%가량이 신용카드를 사용할 정도¹로 카드 보급률이 높은 나라이다. 높은 보급률 덕분에 다양한 배경을 가진 사람들의 데이터를 활용할 수 있으리라 생각했다.

이 프로젝트는 크게 두 가지의 목적을 갖는다:

1. 기존 고객 유지를 위한 이탈 요인 분석

서비스의 운영에 있어, 신규 서비스 제작만큼이나 중요한 것이 유지보수이다. 새로운 카드가 출시되어도 발급한 고객이 주기적으로 사용하지 않거나 금방 이탈하면 이득을 보기 어렵다. 이런 현상을 방지하기 위해 현재 이탈한 고객들을 다양한 관점에서 분석해 이탈 요인을 찾아내고, 서비스 유지 중인 고객을 보다 충성 고객으로 만들기 위한 방안을 모색한다.

2. 신규 고객 확보 방안 도모

은행은 매년 신규 고객을 유치하기 위해 노력하고 있다. 청년층을 타겟으로 한 상품이 나오고, 대학가나 교내에 입점하고, ‘우리 아이 첫 계좌’라는 슬로건을 걸고 비대면 계좌 개설 서비스를 지원하기도 한다. 신규 고객이 이익 증대에 도움이 되기 때문이다. 은행 상품 중 신용카드는 고객 보유율을 5% 신장하는 것만으로도 수익률 120% 증대를 가져온다고 보고되고 있다(이건창 외, 2001). 기존 고객이 불만족한 점을 보완하고 새로운 서비스를 만들어냄으로써 신규 고객을 확보하기 위한 방안을 모색한다.

1.2. 데이터 설명

본 분석은 Kaggle에서 제공하는 신용카드 고객 데이터셋을 활용하였다. 총 23개의 column과 10,127개의 row로 구성되어 있다.

칼럼의 세부 항목은 다음과 같다.

CLIENTNUM	고객 식별자
-----------	--------

Attrition_Flag	이탈 여부
Customer_Age	고객 나이
Gender	성별 (M/F)
Dependent_count	부양가족 수
Education_Level	최종학력 (모름 / 미교육 / 고졸 / 초대졸 / 대졸 / 학사 / 박사)
Marital_Status	혼인 상태 (모름 / 미혼 / 기혼 / 이혼)
Income_Category	소득 구간 (모름 / \$40K 미만 / \$40~\$60K / \$60K ~ \$80K / \$80K~\$120K / \$120K 이상)
Card_Category	신용카드 종류 (Blue / Silver / Gold / Platinum)
Month_on_book	은행 서비스 유지 개월수
Total_Relationship_Count	보유한 상품 수
Months_Inactive_12_mon	12개월 내 비활성화된 개월 수
Contacts_Count_12_mon	12개월 내 은행 접촉 횟수
Credit_Limit	은행 한도
Total_Revolving_bal	총 리볼빙 잔액
Avg_Open_To_Buy	12개월 내 평균 OTB (신용 한도 - 현재 잔액)
Total_Amt_Chng_Q4_Q1	4분기 총 거래금액 대 1분기 비율
Total_Trans_Amt	최근 12개월 간 총 거래 금액
Total_Trans_Ct	최근 12개월 간 총 거래 건수
Total_Ct_Chng_Q4_Q1	4분기 총 거래건수 대 1분기 비율
Avg_Utilization_Ratio	평균 부채 잔액 대비 신용 한도

Naive_Bayes_Classifier_%	Naive Bayes
Naive_Bayes_Classifier_%	Naive Bayes

1.3. 데이터 전처리

```

CLIENTNUM                                0
Attrition_Flag                            0
Customer_Age                             0
Gender                                    0
Dependent_count                           0
Education_Level                           0
Marital_Status                            0
Income_Category                           0
Card_Category                             0
Months_on_book                            0
Total_Relationship_Count                   0
Months_Inactive_12_mon                     0
Contacts_Count_12_mon                      0
Credit_Limit                             0
Total_Revolving_Bal                        0
Avg_Open_To_Buy                           0
Total_Amt_Chng_Q4_Q1                       0
Total_Trans_Amt                            0
Total_Trans_Ct                             0
Total_Ct_Chng_Q4_Q1                       0
Avg_Utilization_Ratio                     0
Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1 0
Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2 0
dtype: int64

```

1) 결측치 확인

- NULL값이 존재하지 않아 그대로 활용하였다.

2) 칼럼 제거

- Navie_Bayes_Classifier 두 개는 고객 분석에 있어 활용하기 어려웠기 때문에 삭제하였다.

3) 숫자형 변수로 전환

- Attrition_Flag (이탈 여부): 영어권 사용자가 아니기 때문에, 기존의 ['Existing Customer', 'Attrited Customer']은 혼동의 여지가 있다고 판단했다. 유지자를 1, 이탈자를 0로 변경하였다.
- Gender (성별): M을 1, F를 0으로 설정했다.
- Education_Level (최종 학력): 어떤 것이 높은 학력인지 쉽게 와닿지 않았다. 분석에 있어 생각의 여지를 줄이기 위해 'Unknown'을 0으로, 'Uneducated'부터 'Doctorate'까지를 1~6으로 설정했다.

- Marital_Status (혼인 상태): Unknown을 0으로, Single을 1, Married를 2, Divorced를 3으로 설정했다.
- Income_Category (소득 구간): Unknown을 0으로, 이후 \$40K 미만에서 120K 이상까지 1~5로 설정했다.

4) 처리 결과

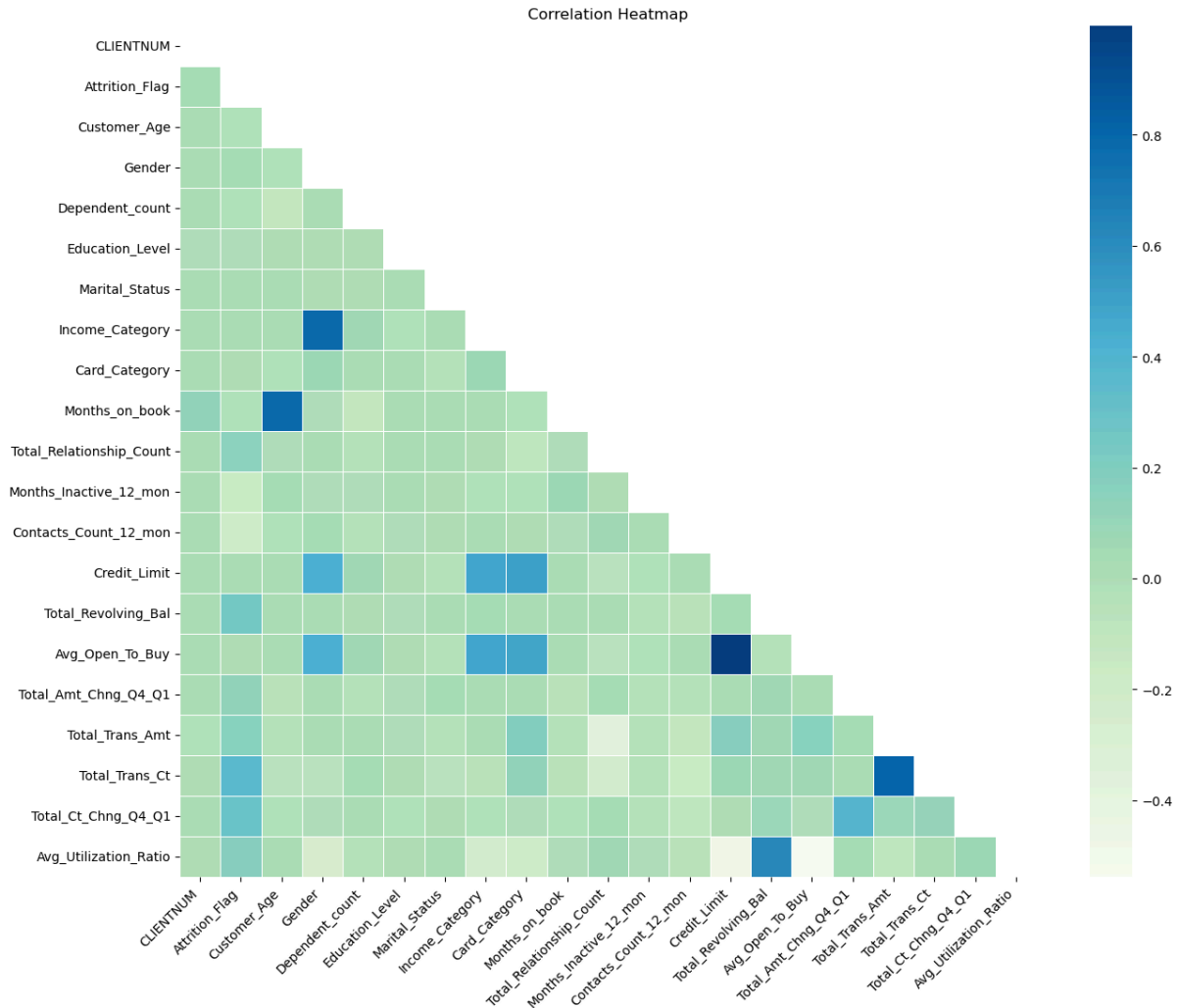
- 숫자형으로 변경된 것을 확인할 수 있다.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10127 entries, 0 to 10126
Data columns (total 21 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CLIENTNUM                            10127 non-null  int64
1   Attrition_Flag                       10127 non-null  int64
2   Customer_Age                         10127 non-null  int64
3   Gender                               10127 non-null  int64
4   Dependent_count                      10127 non-null  int64
5   Education_Level                     10127 non-null  int64
6   Marital_Status                       10127 non-null  int64
7   Income_Category                     10127 non-null  int64
8   Card_Category                       10127 non-null  int64
9   Months_on_book                      10127 non-null  int64
10  Total_Relationship_Count             10127 non-null  int64
11  Months_Inactive_12_mon              10127 non-null  int64
12  Contacts_Count_12_mon               10127 non-null  int64
13  Credit_Limit                        10127 non-null  float64
14  Total_Revolving_Bal                 10127 non-null  int64
15  Avg_Open_To_Buy                     10127 non-null  float64
16  Total_Amt_Chng_Q4_Q1                 10127 non-null  float64
17  Total_Trans_Amt                     10127 non-null  int64
18  Total_Trans_Ct                      10127 non-null  int64
19  Total_Ct_Chng_Q4_Q1                 10127 non-null  float64
20  Avg_Utilization_Ratio                10127 non-null  float64
dtypes: float64(5), int64(16)
memory usage: 1.6 MB
```

2. 본론

2.1. 가설 설정

본격적인 분석에 들어가기 전, 각 컬럼의 Heatmap과 히스토그램을 바탕으로 가설을 설정했다.



해당 히트맵 그래프를 바탕으로 이탈률과 관련이 높은 칼럼을 선정하였다. 가장 관련 높은 변수는 Total_Trans_Ct/Amt(최근 12개월간 거래 건수 및 금액), Total_Ct_Chng_Q4_Q1(4분기 거래 건수 대 1분기 비율), Total_Revolving_Bal(총리볼빙 잔액), Total_Relationship_Count(보유한 상품 수), Avg_Utilization_Ratio(평균 잔액 대비 신용 한도) 다섯 가지이다. 이것을 총칭하여 ① **고객의 사용 내역**이라 명명했고, 해당 주제에 관해 가설을 설정했다.

또한 히트맵 그래프를 살펴보았을 때, Gender(성별)와 Income_Category(소득 구간)와, Credit_Line(은행 한도), Avg_Open_To_Buy(신용 한도 - 현재 잔액의 평균액) 간의 상관관계가 높게 나타났다.

기혼 여성의 경우 출산과 함께 경력 단절을 겪으며 실질적으로 얻는 소득은 적어지지만, 카드 사용 등 소비에 있어 주체가 되는 경향이 있다. 이것을 살펴보기 위해 Customer_Age(나이), Gender(성별), Marital_Status(혼인 상태) 및 Dependent_count(부양가족 수)를 총칭해 ② 인구통계변수라고 명명했고, 이를 기반으로 가설을 설정했다.

설정한 가설은 다음과 같다:

① 고객 사용 내역

- 상품을 자주 구매하는 사람은 카드를 유지할 것이다.
- 부채를 많이 보유한 사람은 이탈할 것이다.
- 보유한 상품 수가 적을수록 이탈할 것이다.

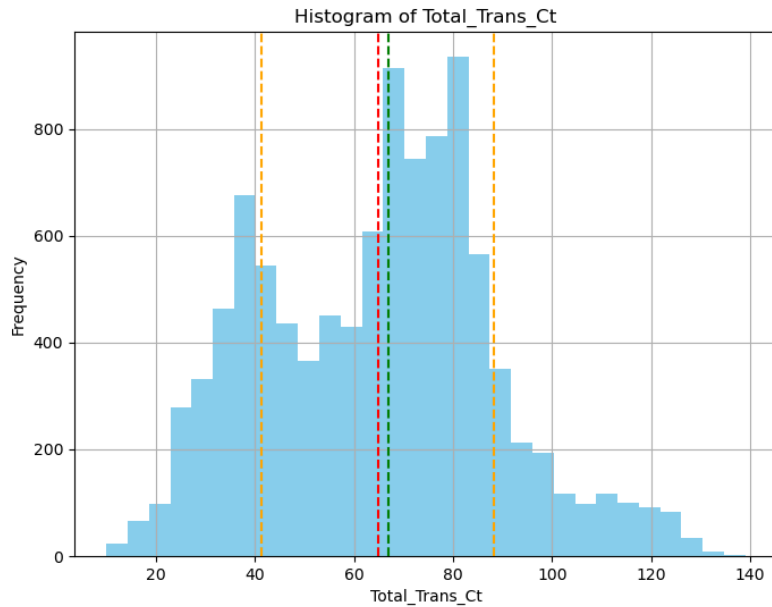
② 인구통계변수

- 나이 대비 유지 개월 수가 높을수록 카드를 유지할 것이다.
- 나이가 많다면 한도 대비 사용 금액이 적을 것이다.
- 나이가 적으면 적을수록 이탈할 것이다.
- 부양가족 수가 많으면 카드를 유지할 것이다.

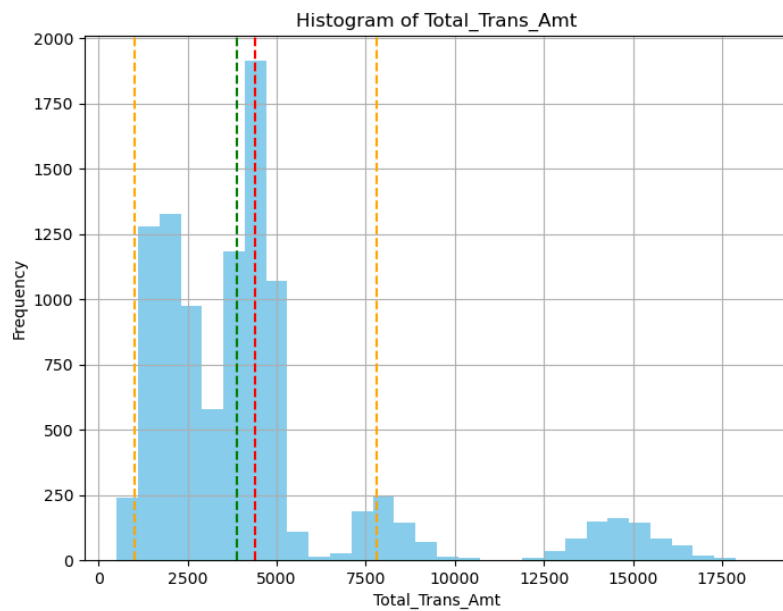
2.2. 고객 사용 내역 기반 분석

2.2.1. Total_Trans_Ct / Amt (최근 12개월간 거래 건수 및 금액)

a. Total Trans Ct / Amt의 분포



[그림 2.2.1-1 최근 거래 건수 히스토그램]

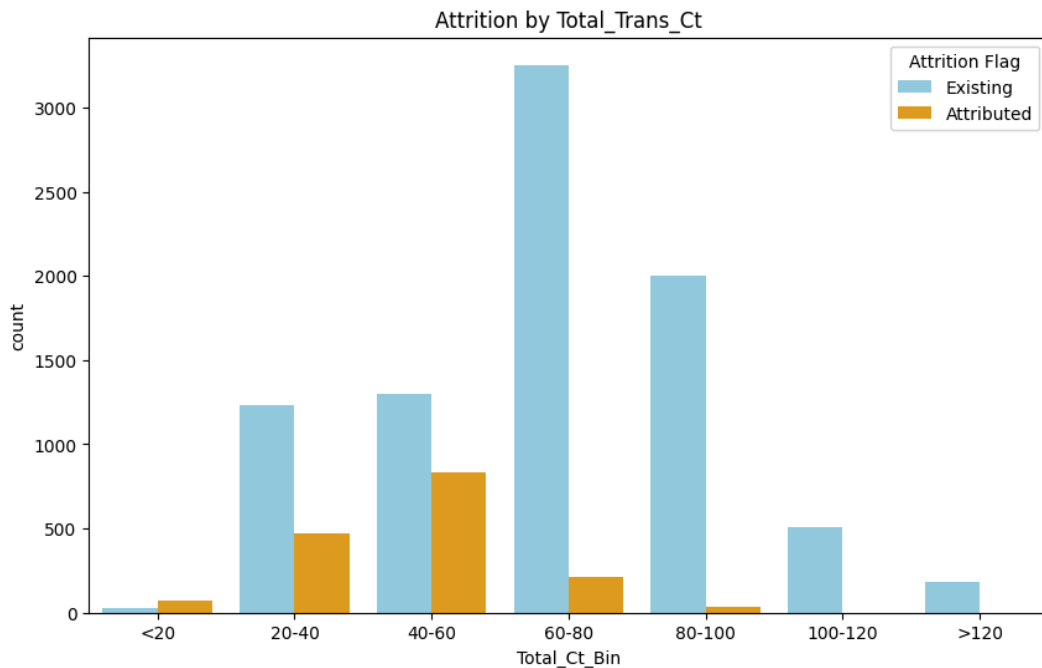


[그림 2.2.1-2 최근 거래 금액 히스토그램]

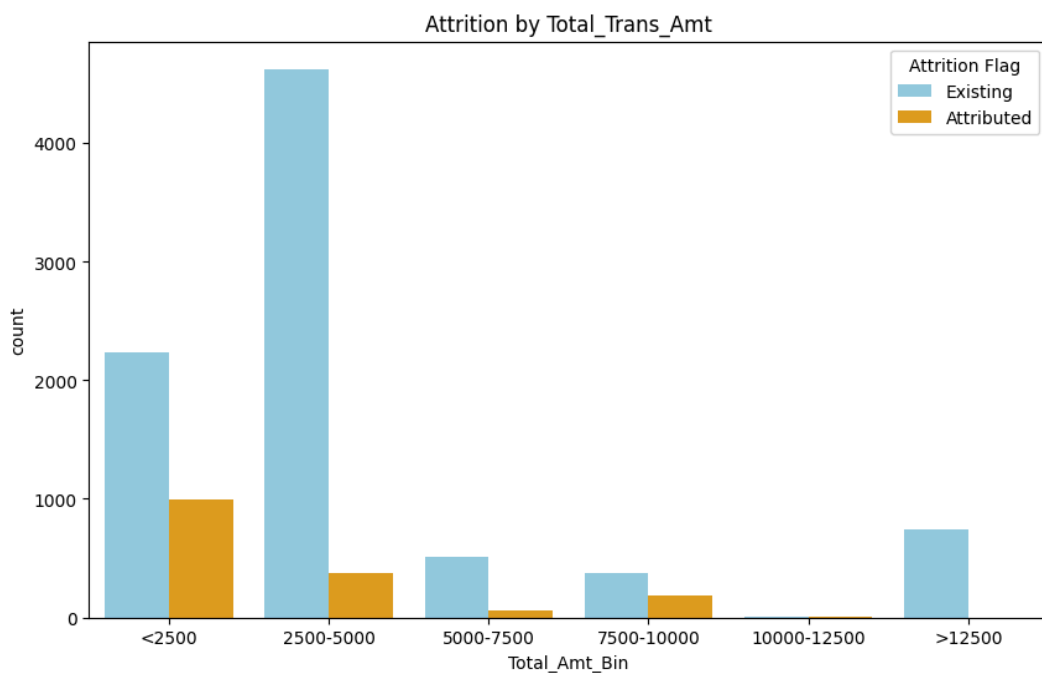
양 끝단의 노란색 점선이 편차를 나타내며, 붉은색 점선이 평균, 초록색 점선이 중앙값이다.

상단의 거래 건수는 평균과 중앙값 주변에 분포되어 있고, 하단의 거래 금액은 매우 큰 금액(12,500달러 이상)으로 거래하는 소수의 고객과 적은 금액(5,000달러 이하)으로 거래하는 다수로 나뉘어 있다.

일반적으로 거래 건수가 적거나 거래량이 적은 고객의 이탈 확률이 높기 때문에, 두 칼럼을 특정 구간으로 나누어 이탈 여부를 확인했다.



[그림 2.2.1-3 거래 건수별 이탈 고객]

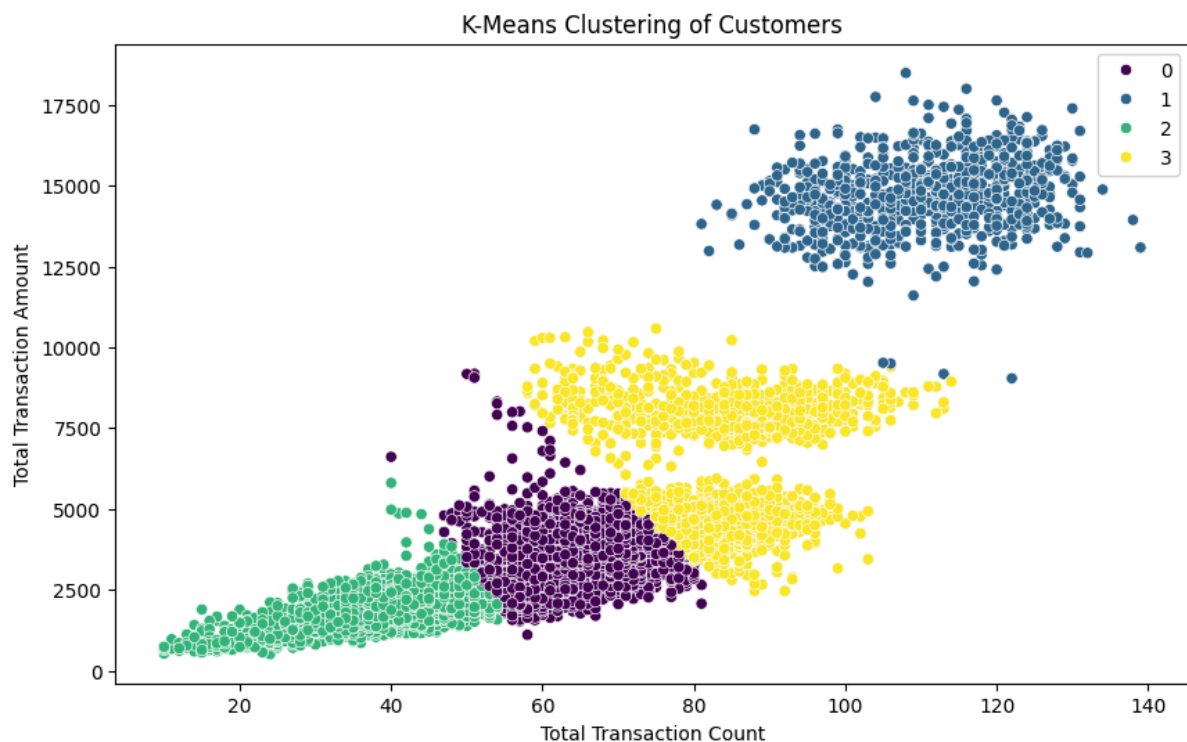


[그림 2.2.1-3 거래 금액별 이탈 고객]

상단의 거래 건수를 보면 60건 미만인 고객의 이탈자 수가 전체 이탈자의 대부분을 차지한다는 점을 알 수 있다. 예상과 동일하게 거래 건수가 적은 고객일수록 이탈할 가능성이 높다고 말할 수 있다. 또한 하단의 거래 금액을 보았을 때, 2,500달러 미만을 지불한 그룹의 이탈자가 가장 많고, 금액이 증가하면서 이탈 고객의 비율이 낮아지는 모습을 보인다.

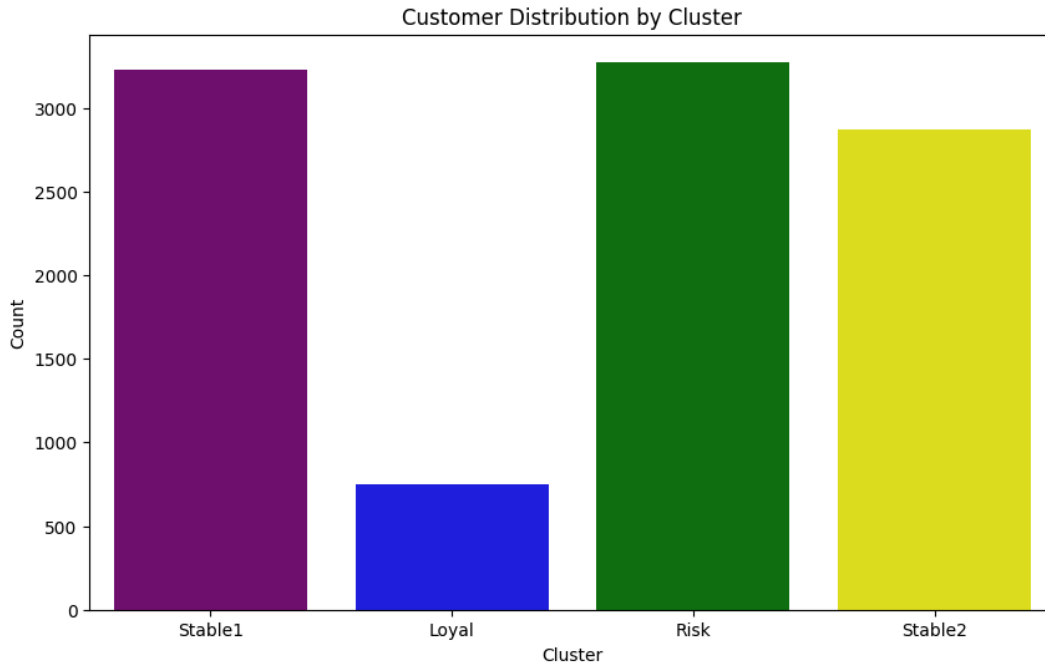
이를 바탕으로 거래 건수와 거래 금액이 모두 낮은 고객은 이탈 가능성이 높다고 판단하였고, 이들에 대한 유지 전략을 수립해야 한다는 결론에 도달했다. 또, 거래 횟수는 낮지만 거래 금액은 높은 고객의 경우 충성도를 높이기 위한 전략이 필요할 수 있겠다.

앞서 히트맵 그래프에서 보았듯이 두 칼럼의 상관관계는 군청색(0.6 이상)으로 매우 높은 수준이다. 해당 은행을 이용하는 고객의 소비 경향을 보고 특정 그룹으로 나누기 위해 K-MEANS를 활용해 4개로 나누었다.



[그림 2.2.1-4 고객 군집별 분류]

좌측 하단의 초록색 그룹은 거래 건수와 거래 금액이 모두 적은 ‘이탈위험고객’, 60-80구간의 보라색 그룹은 거래 건수는 많지만 거래 금액은 적은 ‘안정고객1’, 80-120구간에 넓게 퍼진 노란색 그룹은 거래 건수는 적지만 거래 금액이 높은 ‘안정고객2’, 우측 상단의 파란색 그룹은 거래 건수와 거래 금액이 모두 높은 ‘충성고객’으로 분류하였다.



[그림 2.2.1-5 고객 클러스터별 인원 수]

각 그룹의 고객 수를 확인해 보았을 때, 충성 고객의 비율은 약 10% 미만으로 꽤 낮은 편이기에 고객 이탈을 방지하고, 안정 고객에의 적극적인 마케팅 기법이 필요할 것이라 사료된다.

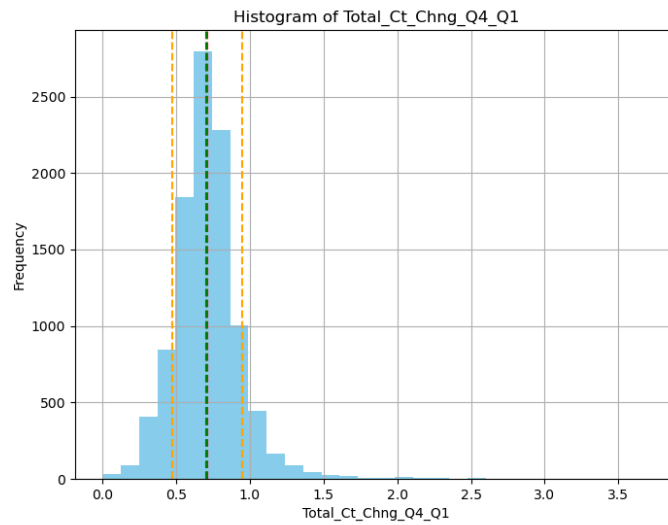
2.2.2. Total_Ct/Amt_Chng_Q4_Q1 (4분기 거래 건수/거래량 대 1분기 비율)

a. ³Q1, Q4란?

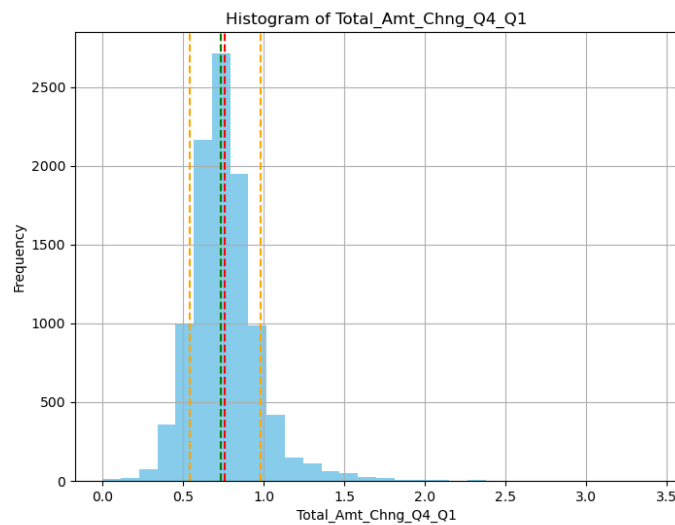
회사가 1년을 4분의 1로 나누어 회계 변동 사항을 기록·계산하기 위해 사용하는 기준이다. 보통 1분기를 Q1, 2분기를 Q2라고 일컫는다. 대한민국의 경우는 1월 1일부터 12월 31일을 1년으로 하며, 국가나 기업에 따라 차이가 있다(김연경, 2024).

미국 10대 은행으로 알려진 JPMorgan Chase, Bank of America, Wells Fargo, Citigroup, U.S. Bancorp, PNC Financial Services, Truist Bank, Goldman Sachs, Capital One Financial, TD Group US Holdings를 확인해 보았을 때, TD Group을 제외한 모든 은행에서 1분기를 4월 이전으로 지정하였기 때문에 Q1을 연초(1월~3월), Q4를 연말(10월~12월)로 상정한다.

b. 4분기 대 1분기의 분포



[그림 2.2.2-1 4분기 대 1분기 거래 건수 비율]



[그림 2.2.2-2 4분기 대 1분기 거래 금액 비율]

양 끝단의 노란색 점선이 편차를 나타내며, 붉은색 점선이 평균, 초록색 점선이 중앙값이다.

4분기 대 1분기의 비율은 거래 건수와 금액 모두 중앙값을 기준으로 대칭되는 종 형태를 띠고 있다.

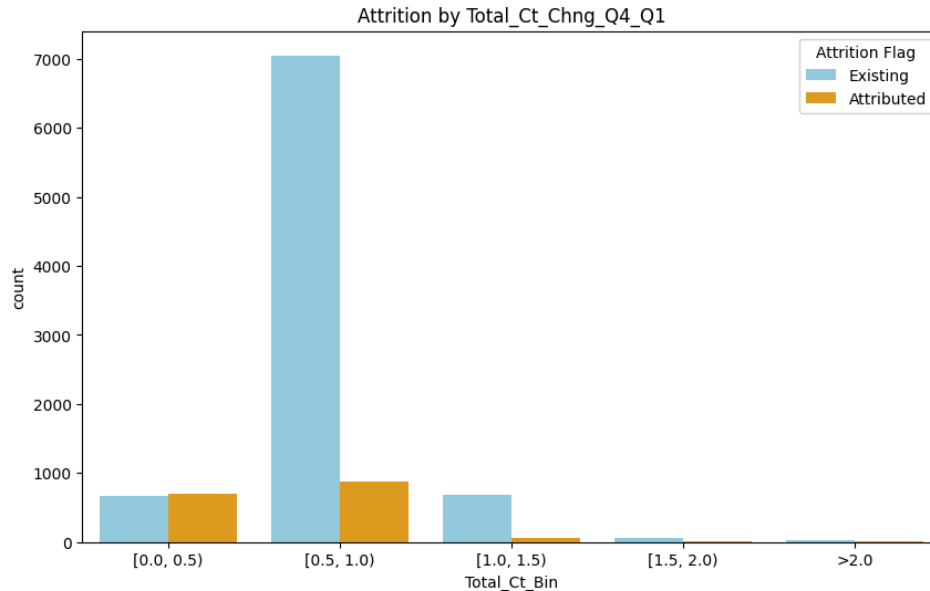
이상치가 비교적 작고 정규분포에 가깝기 때문에 이탈 고객과 큰 관계가 나타나지는 않을 듯하다.

일반적으로 할로윈, 추수감사절, 크리스마스, 연말 파티 등 다양한 공휴일 및 명절이 존재하는

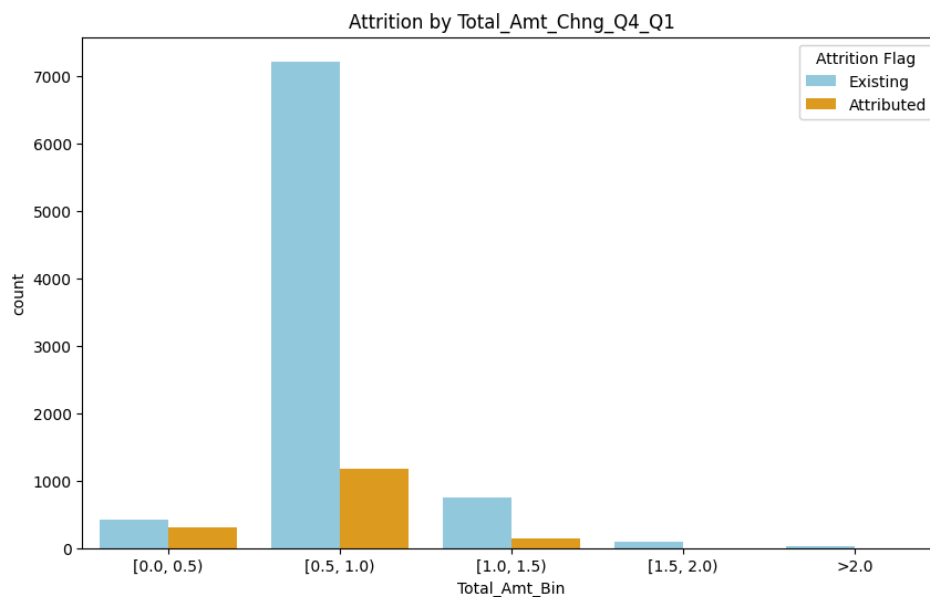
4분기와는 달리 1분기의 이벤트는 크지 않기 때문에 1.0 미만 구간에 고객이 다수 분포되어 있다고

이야기할 수 있다.

보고서 상단의 히트맵을 참고했을 때, 4분기 대 1분기의 비율은 거래 건수가 하늘색(0.2-0.4)구간, 거래량은 초록색(0.0-0.2)구간으로 뚜렷한 상관관계를 보이고 있지는 않다. 확인을 위해 데이터가 모여 있는 구간별로 이탈 고객을 시각화했다.



[그림 2.2.2-3 4분기 대 1분기 거래 건수별 이탈률]



[그림 2.2.2-4 4분기 대 1분기 거래 금액별이탈률]

각 구간별로 좌측의 하늘색이 미이탈 고객, 우측의 주황색이 이탈 고객을 의미한다.

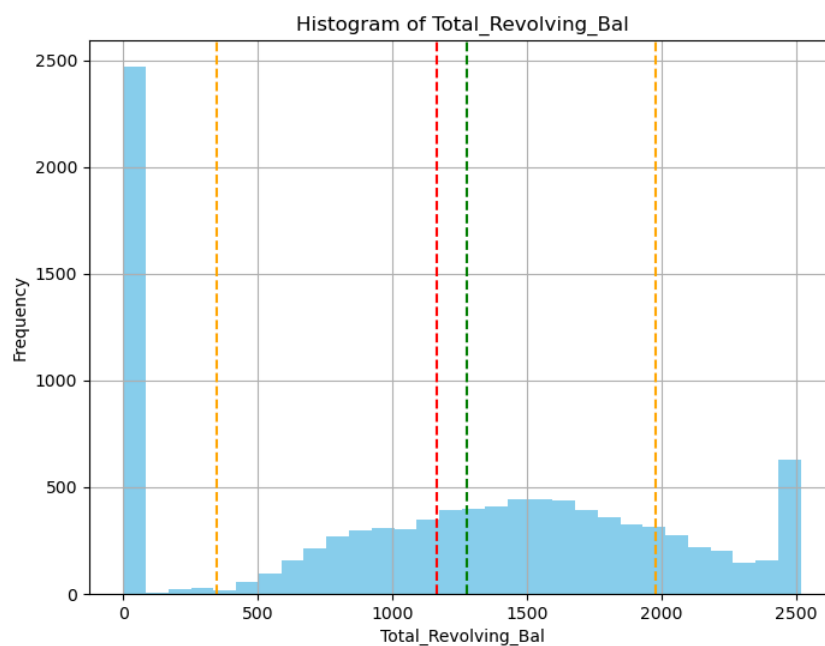
예상과 같이 특정 구간에서 높은 이탈률을 보이지는 않는다. 4분기 대비 1분기 거래 비율은 이탈 요소에 있어 뒷순위 고려 대상이라 할 수 있다.

2.2.3. Total_Revolving_Bal (총 리볼빙 잔액)

a. ⁴리볼빙이란?

카드 대금의 일부만을 지불한 후, 나머지를 대출로 전환하여 다음 달로 이월하는 결제 서비스이다. 신용 카드, 신용 한도, 주택 담보 신용 한도(HELOC)가 있으며, 이월한 잔액에 대해 이자를 지불해야 한다. 기존의 할부에 비해 이자율이 높은 것이 특징이다. (What is revolving credit and how does it work?, 2024)

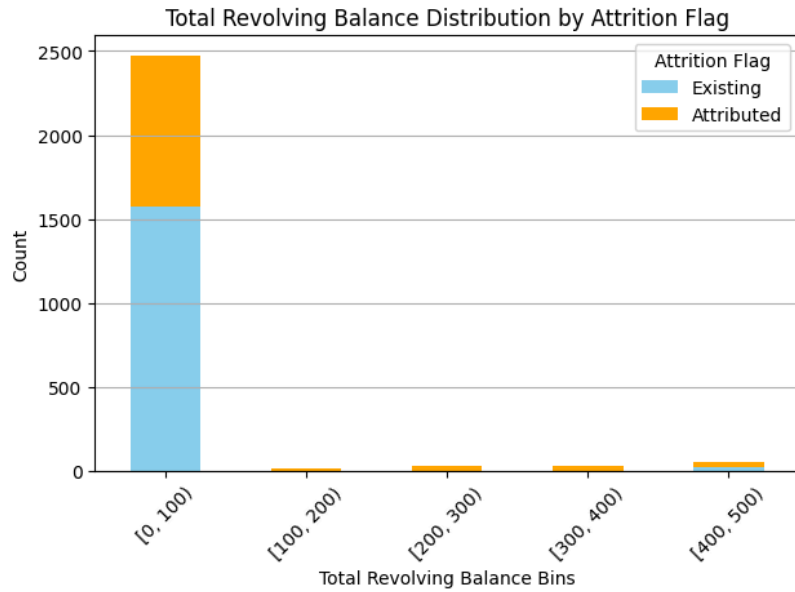
b. 총 리볼빙 잔액의 분포



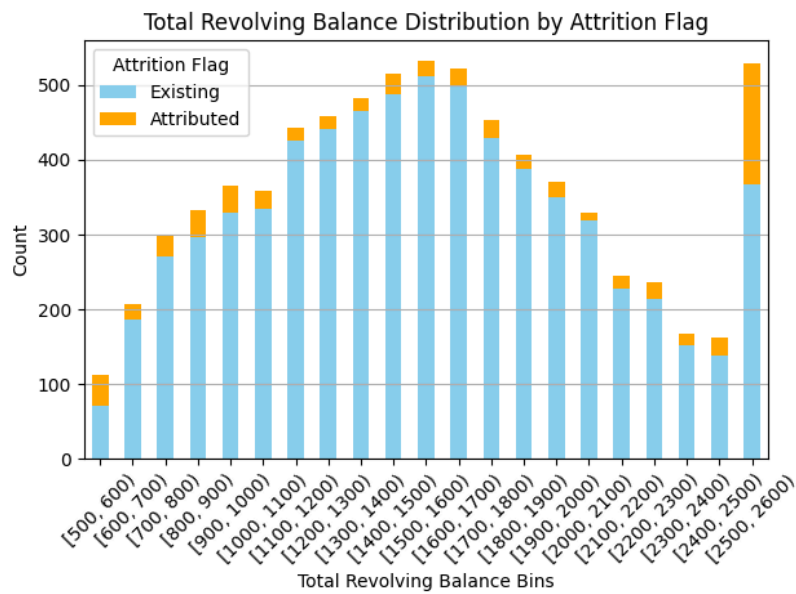
[그림 2.2.3-1 총 리볼빙 잔액의 분포]

양 끝단의 노란색 점선이 편차를 나타내며, 붉은색 점선이 평균, 초록색 점선이 중앙값이다.

약 1/4의 고객은 리볼빙을 전혀 사용하지 않는다. 카드 한도뿐 아니라 주택 담보 대출을 포함하기 때문에 다수의 고객이 리볼빙을 사용하고 있다고 판단했다.

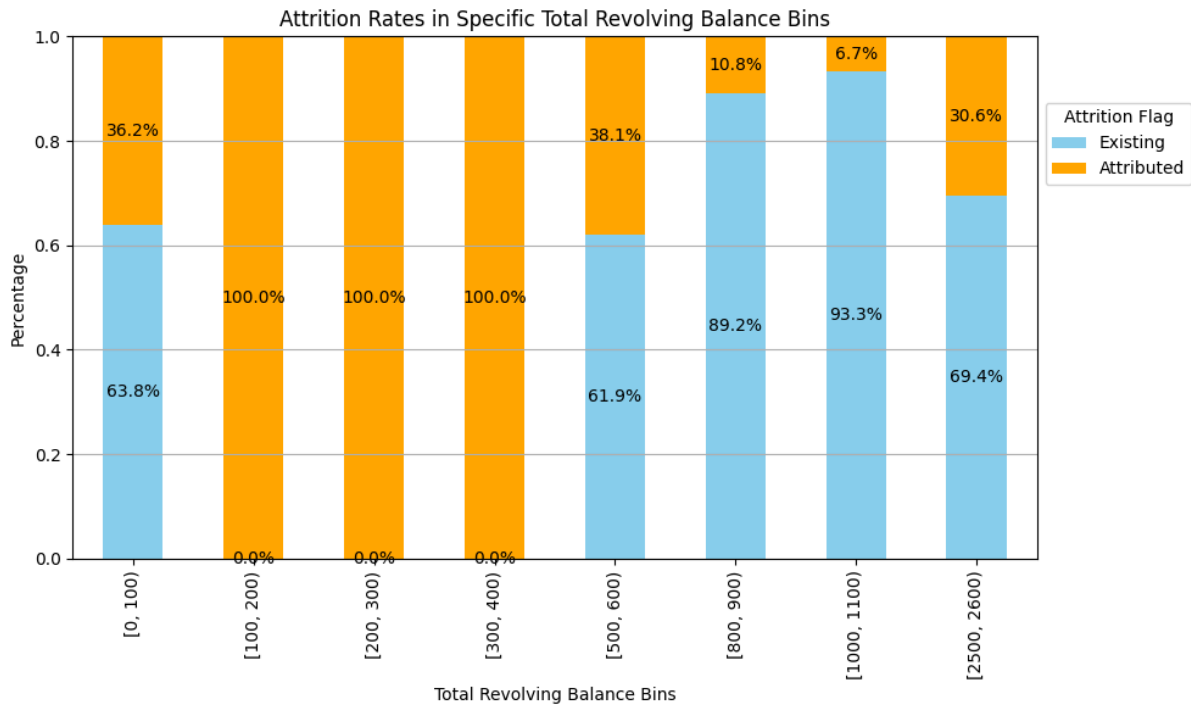


[그림 2.2.3-2 리볼빙 잔액 500 미만의 이탈률]



[그림 2.2.3-3 리볼빙 잔액 500 이상의 이탈률]

리볼빙이 0인 고객의 경우 카드 사용을 중지했을 가능성이 있기 때문에, 수치가 낮았던 500 미만과 500 이상을 기준으로 이탈 고객을 시각화해 보았다. 500 미만 구간에서 이탈률이 높은 것은 예상할 수 있는 바이지만, 500-600구간과 2500-2600구간의 이탈률은 ‘자주 사용할수록 이탈하지 않는다’에 반하였다. 이에 이탈률이 높은 0-500구간과 2,500-2,600구간, 평균 이탈률과 유사한 800-1,100구간을 추출해 비교해 보았다.



[그림 2.2.3-4 이탈률 30% 이상인 구간과 그 외 구간]

500 미만 구간에서는 0을 제외하고는 이탈률이 100%로 나오며, 500-600구간과 2500-2600구간은 30%의 이탈률로, 전체 이탈률 16%에 비해 높은 수치다.

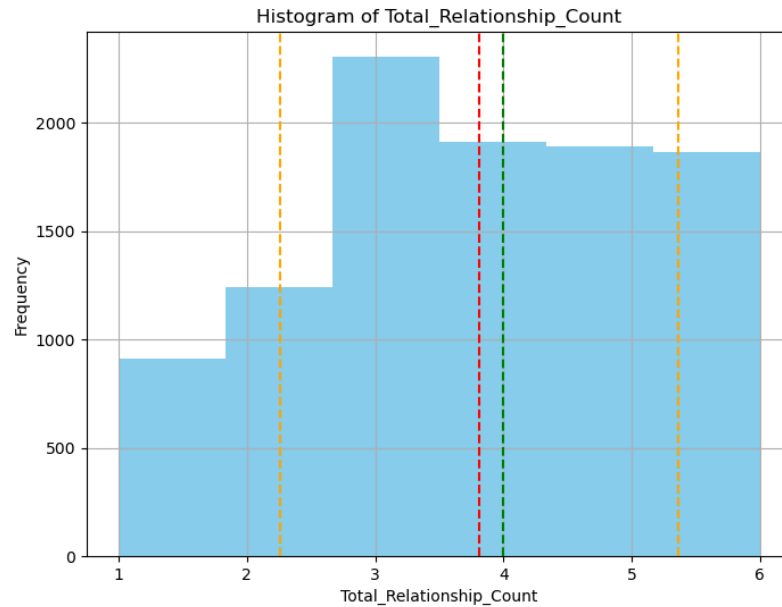
리볼빙이 ‘이자율 높은 카드 대금 이월 결제 상품’이라는 것을 감안했을 때, 부채를 감당할 수 없는 고객이 부채의 증가를 막기 위해 소비를 통제하고 결국 이탈한다고 추측할 수 있다.

2.2.4. Total_Relationship_Count (보유한 상품 수)

a. Relationship이란?

일반적으로 경제에 있어 ⁵Credit(신용)이란 미래에 갚을 것을 약속하고 돈이나 상품, 서비스를 미리 획득하는 능력을 이야기한다(신용과 부채관리의 원리, 2024). 이 중 Credit Relationship은 기업과 기업, 혹은 기업과 개인이 갖는 신용 관계를 의미하는 것으로, 해당 데이터에 있어서는 신용 카드 등의 은행 상품에 해당한다.

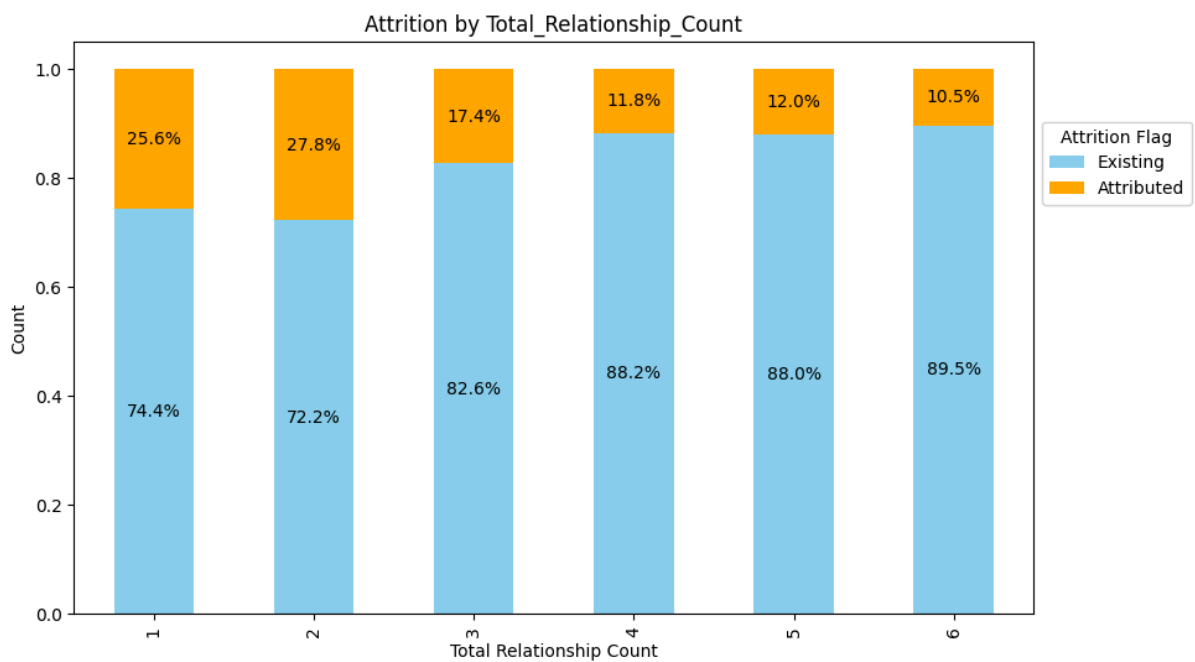
b. 상품 수의 분포



[그림 2.2.4-1 상품 수의 분포]

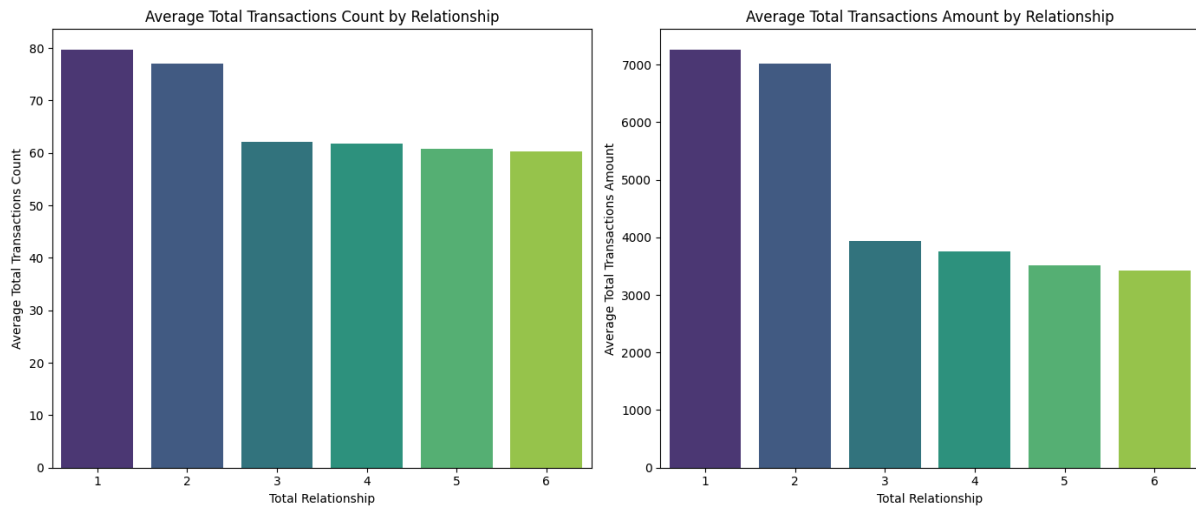
양 끝단의 노란색 점선이 편차를 나타내며, 붉은색 점선이 평균, 초록색 점선이 중앙값이다.

상품 수는 3개가 가장 많고, 이후 점진적으로 감소하는 모습을 보인다. 중앙값이 4인 것을 참고하면 대부분의 고객이 여러 상품을 보유하고 있다는 점을 알 수 있다. 보유한 상품 수가 적을수록 이탈하기 쉬울 것이라 판단, 이탈률을 100% 누적 그래프로 작성하였다.



[그림 2.2.4-2 상품 수별 이탈률]

예상과 동일하게 보유 상품 수가 늘어나면 늘어날수록 이탈률이 점점 감소하는 모양새를 보인다. 한 은행에 여러 카드나 대출 등의 상품을 보유한다면 접촉 횟수가 증가하므로 이탈 가능성이 감소한다고 추측할 수 있다.



[그림 2.2.4-3 상품 수와 12개월 내 거래 건수, 금액 비교]

보유한 상품 수가 적을수록 12개월 내 거래 건수와 금액이 늘어나는 모습을 보인다. 한 가지의 상품에 집중적으로 소비하는 만큼 해당 은행을 주거래 은행으로 설정하고 있다고 추측할 수 있겠다.

2.2.5. Avg_Utilization_Ratio (평균 부채 잔액 대비 신용 한도)

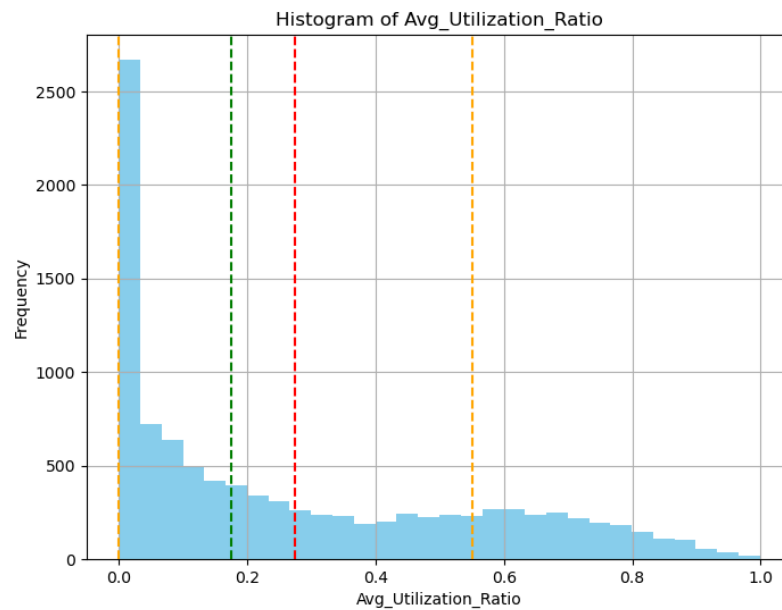
a. 6부채 잔액 대비 신용 한도란?

신용 점수를 계산하는 요소이다. 신용 카드의 할부나 대출금과 같은 부채를 얼마나 잘 활용하고 있는지 확인하는 척도로 사용된다. 미결제 금액이 남아 있다면 수치가 높아지며, utilization이 30% 미만일 때 높은 신용 점수를 유지하는 것이 가능하다.

일반적으로 모든 신용 계좌의 부채 잔액(balance)을 더한 후, 모든 신용 계좌의 한도(Credit Line)로 나눈 후 백분율로 계산한다. (What Is a Credit Utilization Ratio?, 2024)

$$CreditUtilization = \frac{\sum(Balance)}{\sum(CreditLine)} * 100$$

b. Avg_Utilization_Ratio의 분포



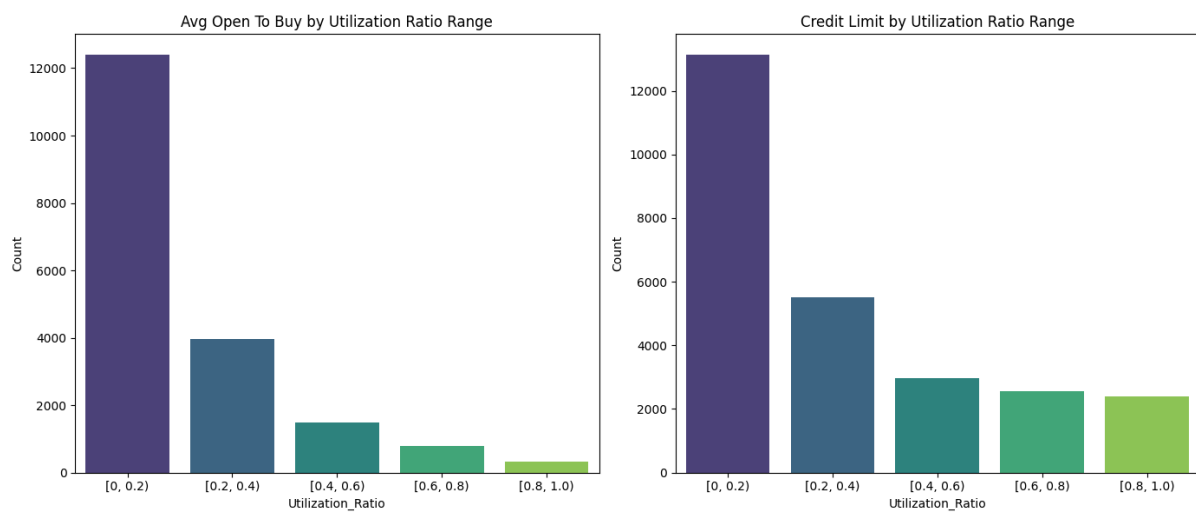
[그림 2.2.5-1 평균 부채 잔액 대비 신용 한도의 분포]

양 끝단의 노란색 점선이 편차를 나타내며, 붉은색 점선이 평균, 초록색 점선이 중앙값이다.

0.0-0.2구간에 40%가량의 사용자가 몰려 있으며, 특히 0.0일 때 그 값이 두드러진다.

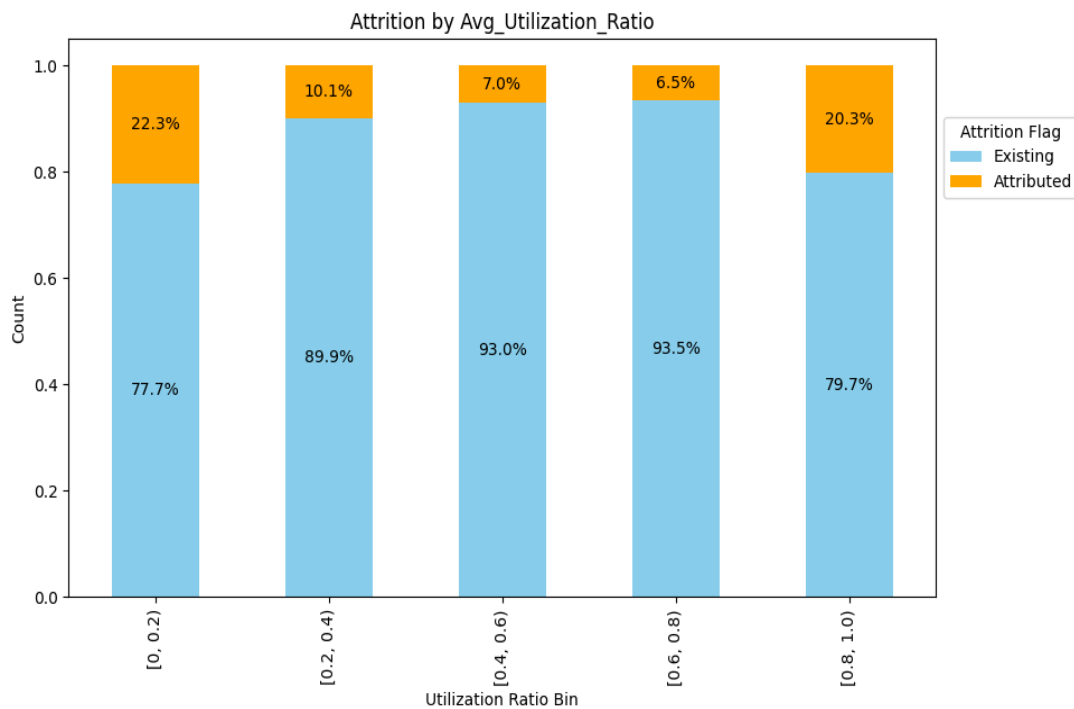
이후로는 천천히 감소하다가 0.6 구간에서 잠시 반등한 뒤 다시 감소하는 모양새를 보인다.

일반적으로 Utilization Ratio는 그 수치가 작을 때 신용 등급에 좋은 영향을 미친다. 0.0의 경우는 모든 리볼빙 잔액(할부금 등)을 지급했거나, 아예 신용을 사용하지 않았음을 의미한다.

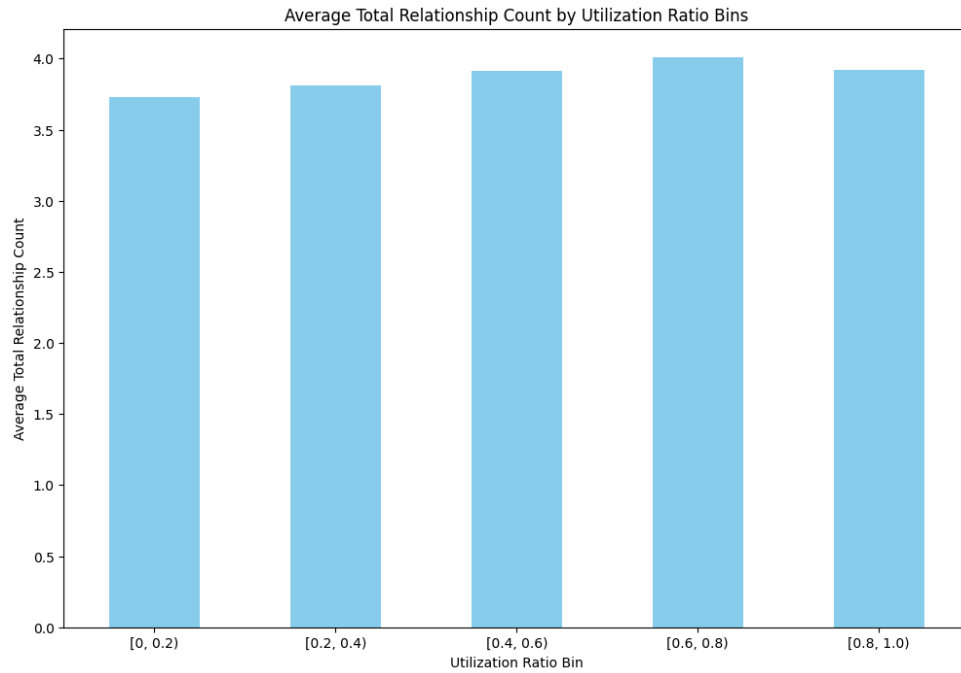


[그림 2.2.5-2 평균 부채 잔액 대비 신용 한도와 Open To Buy, 신용한도 비교]

Utilization Ratio는 낮을수록 신용 등급에 좋은 영향을 미친다는 주장을 확인하기 위해 (신용 한도 - 현재 부채잔액)을 의미하는 Open_To_Buy와 카드 등의 최대 신용 한도를 나타내는 Credit_Limit을 함께 비교해 보았다. 위와 같이 Open To Buy와 신용 한도(Credit Limit) 모두 Utilization Ratio가 증가하면서 값이 낮아지는 모습을 보인다.

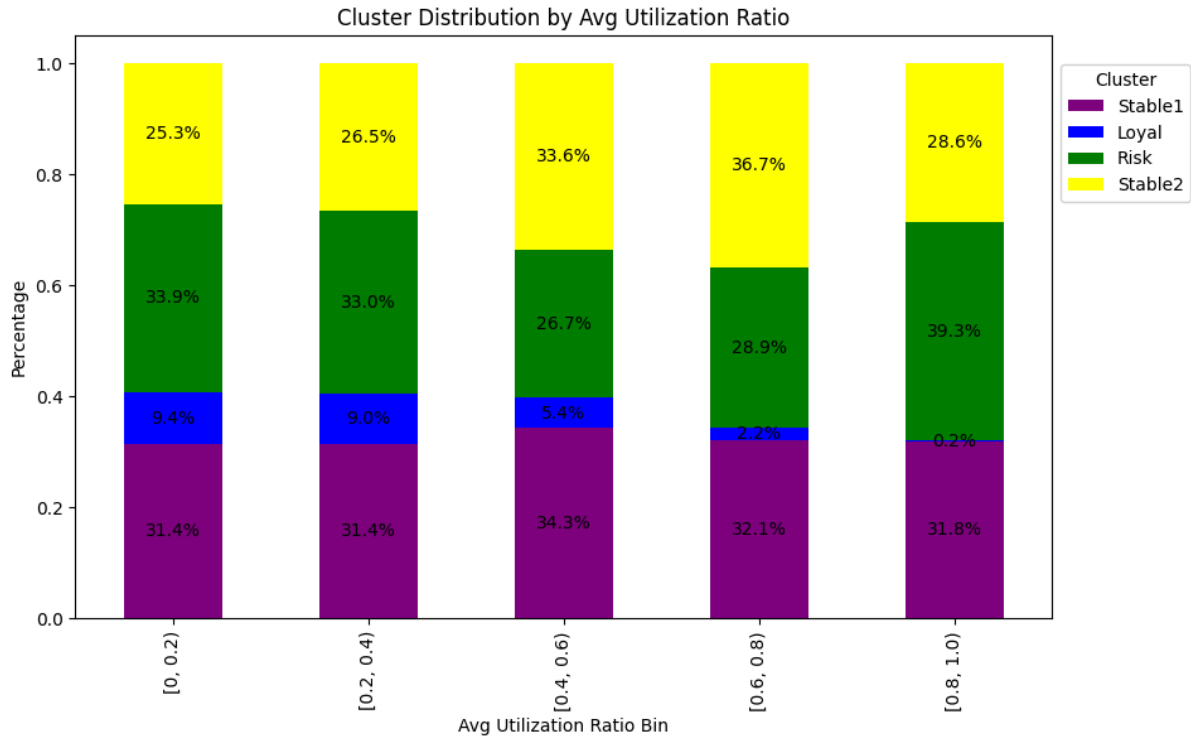


[그림 2.2.5-3 평균 잔액 대비 신용 한도의 이탈률]



[그림 2.2.5-4 평균 잔액 대비 신용 한도의 이탈률과 보유 상품 수의 관계]

Utilization Ratio의 구간에 따른 이탈률을 확인하기 위해 시각화를 진행했다. 상단의 막대그래프를 참고하면 ① 0.0-0.2구간과 ② 0.8-1.0 구간에서 상대적으로 높은 이탈률을 확인할 수 있다. 이탈에 영향을 준 요인을 찾기 위해 하단과 같이 부채 잔액 대비 신용 한도 구간당 평균 보유 상품 수를 시각화해 보았지만, 전 구간에서 3.5개 이상, 4개 미만의 상품을 보유했기 때문에 크게 연관이 있다고 판단하기는 어렵다.



[그림 2.2.5-5 평균 잔액 대비 신용 한도 구간당 고객 종류 비율]

앞서 K-MEANS 기법으로 분류한 고객의 종류를 함께 나타내 보았다. ① 0.0-0.2구간은 충성고객이 9.4%, 이탈(위험)고객이 33.9%로 둘 다 높지만, ② 0.8-1.0 구간은 충성고객이 0.2%로 가장 낮았고, 이탈(위험)고객은 39.3%로 가장 높았다.

이에 이탈률이 높은 ① 구간의 고객은 할부금을 제때 납부하는 신용도가 높은 고객과 상품을 아예 사용하지 않는 고객층이 섞여 있다고 판단했다. 또한 앞서 2.2.3의 b에서 추측했던 것처럼, 부채를 감당하지 못해 이탈했다고 할 수 있다.

2.2.6. 가설 비교 및 요약

2.2에서는 고객의 사용 내역을 기반으로 이탈 영향을 주는 요인이 무엇인지에 관해 분석해 보았다.

먼저, 거래 건수와 거래 금액을 바탕으로 고객의 이탈률을 분석하였다.

거래 건수가 적거나 금액이 적을수록 이탈한 고객이 많아졌기 때문에, 가설 ① "상품을 자주 구매하는 사람은 카드를 유지할 것이다."는 사실로 확인되었다. 이 결과를 바탕으로 K-MEANS를 활용하여 고객을 [이탈(위험)고객, 안정고객1, 안정고객2, 충성고객] 으로 구분하였다.

두 번째로는 총리볼빙 잔액과 부채 잔액 대비 신용 한도를 바탕으로 고객의 이탈률을 분석하였다. 리볼빙 잔액이 500-600달러, 2500-2600달러, 0-100달러 순서로 이탈률이 높았다. 리볼빙 잔액이 높아질수록 이탈률도 증가할 것이라 예상했지만, 사실이 아니라는 점을 알 수 있었다.

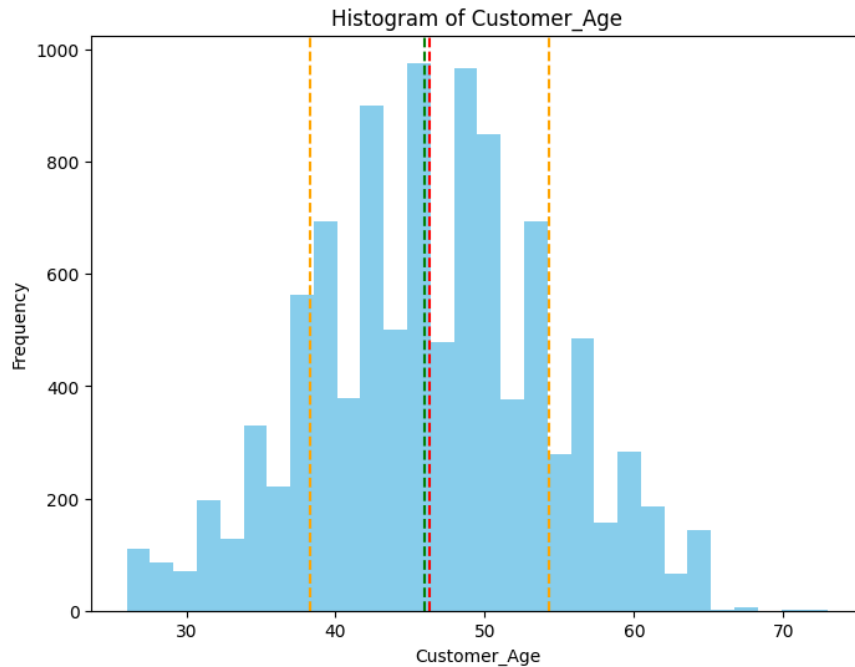
부채 잔액 대비 신용 한도의 경우 전체적인 사용자는 0.0-0.2구간에 몰려 있었지만, 이탈률을 비교해 보았을 때 0.0-0.2구간과 0.8-1.0구간의 수치가 유사했다. 해당 구간의 고객 종류를 확인해 보았을 때, 0.0-0.2구간은 충성고객의 비율이 높은지만, 0.8-1.0구간은 충성고객 비율이 0.2%로 거의 없다시피했고, 이탈위험고객의 비율은 가장 높았다. 이에 가설 ② "부채를 많이 보유한 사람은 이탈할 것이다" 는 사실이라 보기 어렵다는 점을 확인할 수 있었다.

세 번째로는 보유한 금융 상품 수를 바탕으로 고객의 이탈률을 분석하였다. 금융 상품 수는 3개가 가장 많았고, 대부분의 고객이 3.5개 이상의 금융 상품을 보유하고 있다는 점을 확인할 수 있었다. 이탈률을 확인해 보았을 때, 보유 상품 수가 증가할수록 이탈률이 감소하는 모양새를 보였기 때문에, 가설 ③ "보유한 상품 수가 적을수록 이탈할 것이다"는 사실임을 확인할 수 있었다.

2.3. 인구통계변수 기반 분석

2.3.1. Customer_Age (고객 나이)

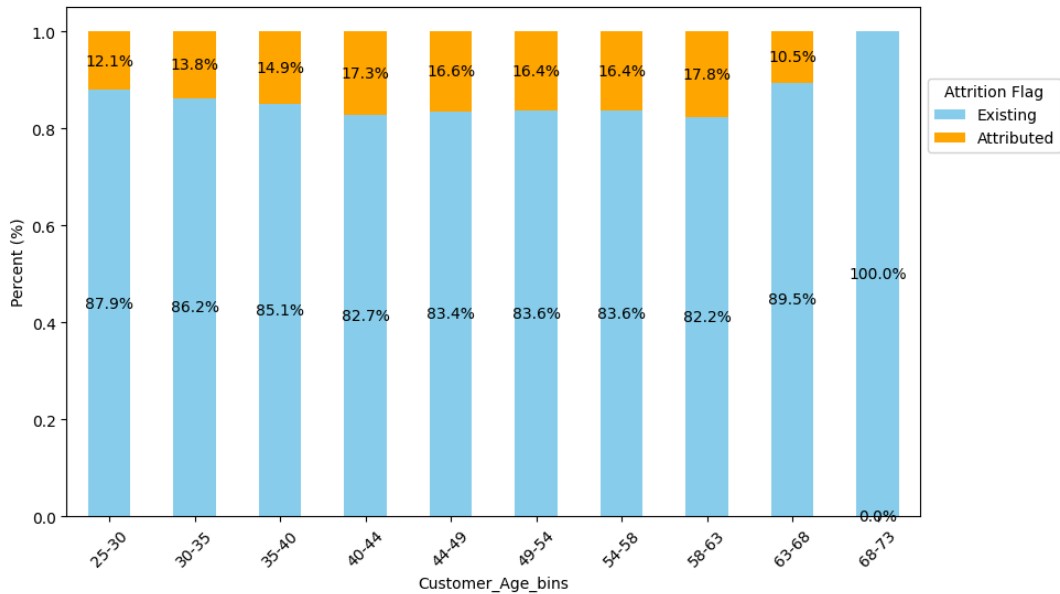
a. Customer_Age 분포



[그림 2.3.1-1 고객 나이의 분포]

고객들의 나이는 26세부터 73세 사이에 분포되어 있다. 평균과 중앙값이 모두 40세에서 50세 사이이며, 정규 분포에 가까운 모양새를 하고 있다. 보다 편리한 시각화를 위해 연령을 15개의 그룹으로 나누어 분석을 진행, 각 연령대의 경제적 활동과 신용카드 사용 패턴 등을 이해하고자 하였다. 이를 통해 각 연령대에 맞춘 금융 상품과 마케팅 전략을 개발할 수 있으리라 추측한다.

b. 연령별 Attrition_Flag (신용카드 해지 여부) 비율



[그림 2.3.1-2 고객 나이대별 이탈률]

고객 나이대별 거래 비율을 보았을 때, 68-73세 구간을 제외한 모든 구간에서 유사한 수치의 이탈률이 나타났다. 이탈률이 0%인 68-73세 구간의 경우 단 4개의 데이터만이 있었기에 해당 나이대의 수치를 제대로 표현하고 있지 못하다고 판단했다

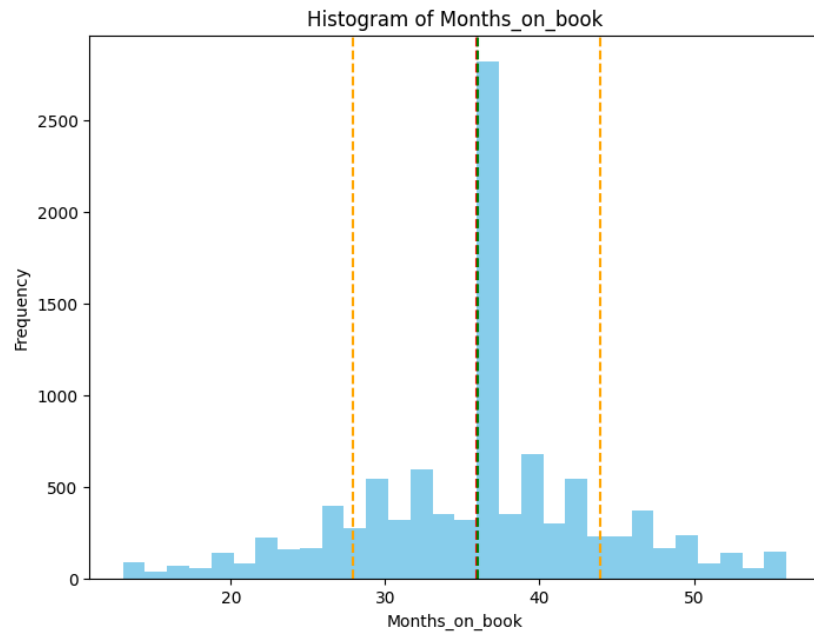
가설과는 달리 나이대가 어릴수록 이탈률이 높아지는 모습은 나오지 않았고, 오히려 40-50대 고객의 이탈률이 높은 모습을 볼 수 있었다. 해당 나이대에서 절대적인 고객의 수 또한 많았던 것을 고려하면, 중년 세대를 대상으로 한 맞춤형 서비스나 혜택을 제공함으로써 이탈을 방지하는 전략을 세울 필요성이 보인다.

2.3.2. Customer_Age(고객 나이)와 Months_on_book (은행 서비스 유지 개월 수)

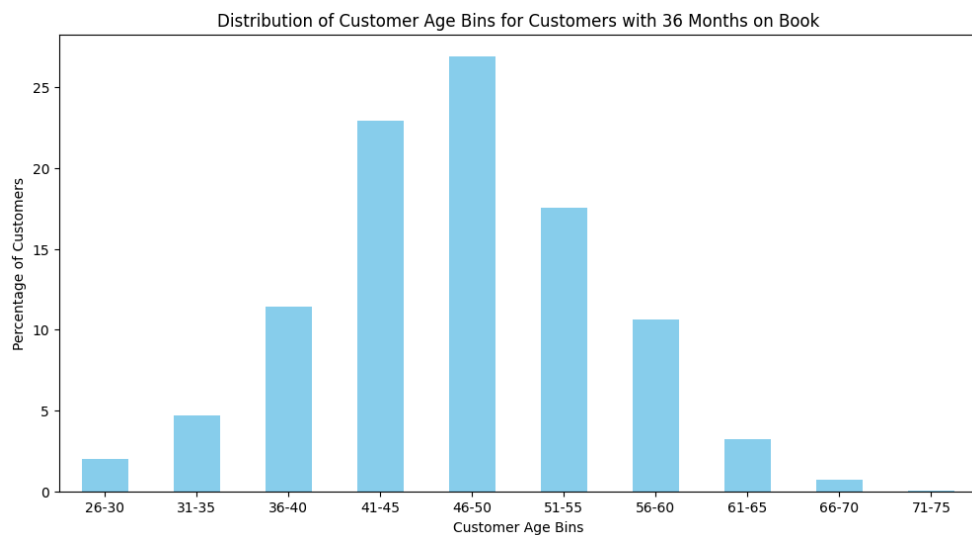
은행 서비스 유지 개월 수는 고객이 해당 은행에서 서비스를 이용한 기간을 나타내는 지표이다.

상단의 히트맵 그래프를 참고했을 때, 나이와의 상관관계가 높게 나왔기 때문에 연령과 유지 개월 수를 같이 비교하였다.

a. Months_on_book (은행 서비스 유지 개월 수) 분포



[그림 2.3.2-1 은행 서비스 유지 개월수 분포]



[그림 2.3.2-2 유지 개월수 36개월인 사람의 나이대 분포]

상단의 서비스 유지 개월수 분포를 보면, 13개월 이상 56개월 이하로 분포되어 있는 점을 알 수 있다. 36개월이라는 특정 값에서 대량의 데이터가 있기 때문에, 유지 개월 수가 36개월인 사람을 나이대로 분석해 보았다. 하단의 분포를 보면, 36개월간 유지한 사람의 분포와 그림 2.3.1-1의 ‘나이대 분포’가 거의 동일하게 그려진다는 것을 파악할 수 있다.

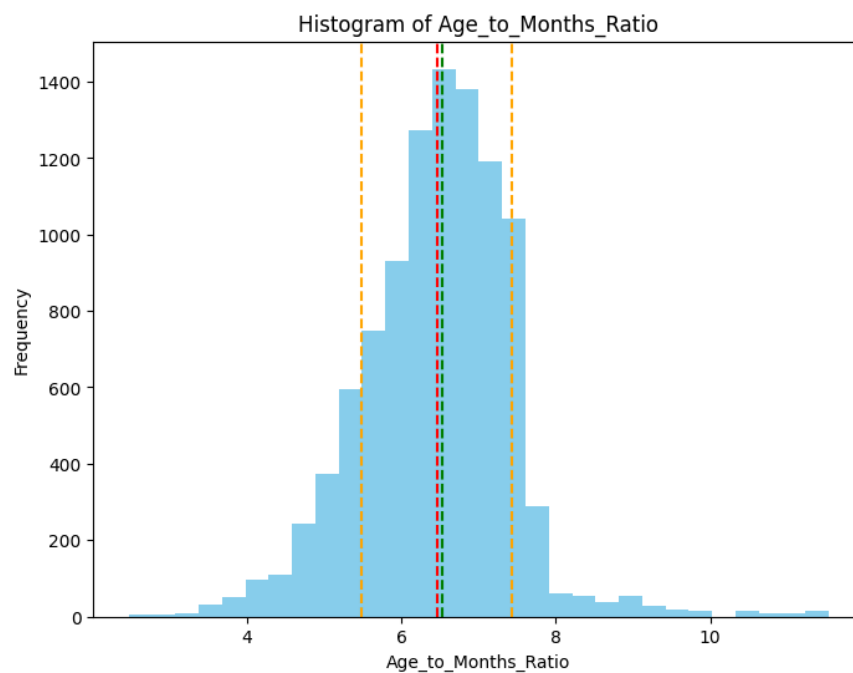
b. 나이 대비 은행 서비스 유지 개월 수

고객 나이 대비 유지 개월수를 계산하기 위해, 'Age_to_Months_Ratio' 칼럼을 제작하였다. 개월수를 기준으로 계산하기 위해

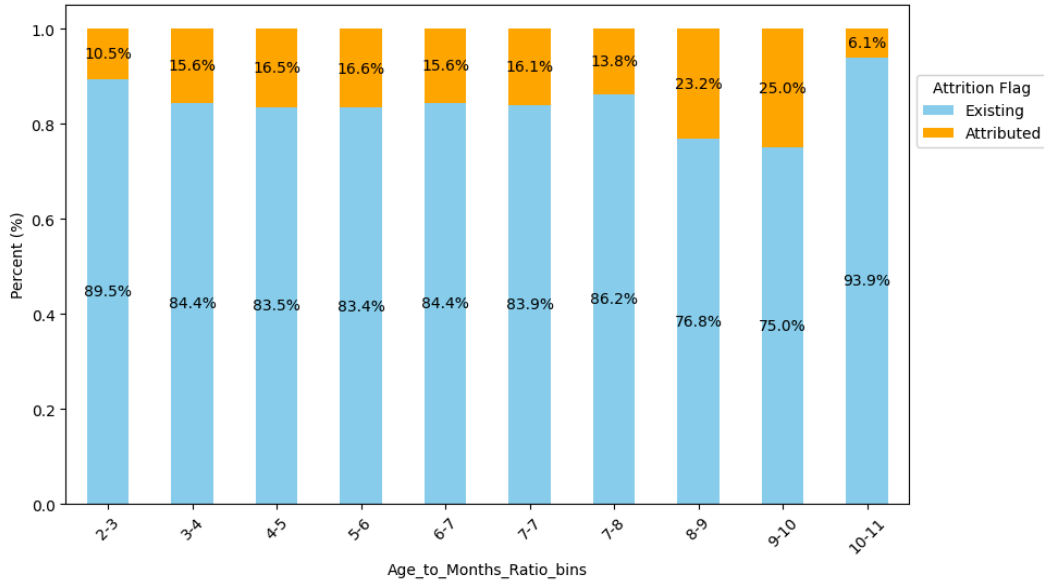
$bank['Age_to_Months_Ratio'] = bank['Months_on_book'] / (bank['Customer_Age'] * 12) * 100$

이라는 식을 통해 나이 대비 유지 개월수를 계산했다.

해당 칼럼은 숫자가 커질수록 나이 대비 오랜 시간 은행을 유지했음을 말해 준다.



[그림 2.3.2-4 나이대비 유지 개월수 분포]



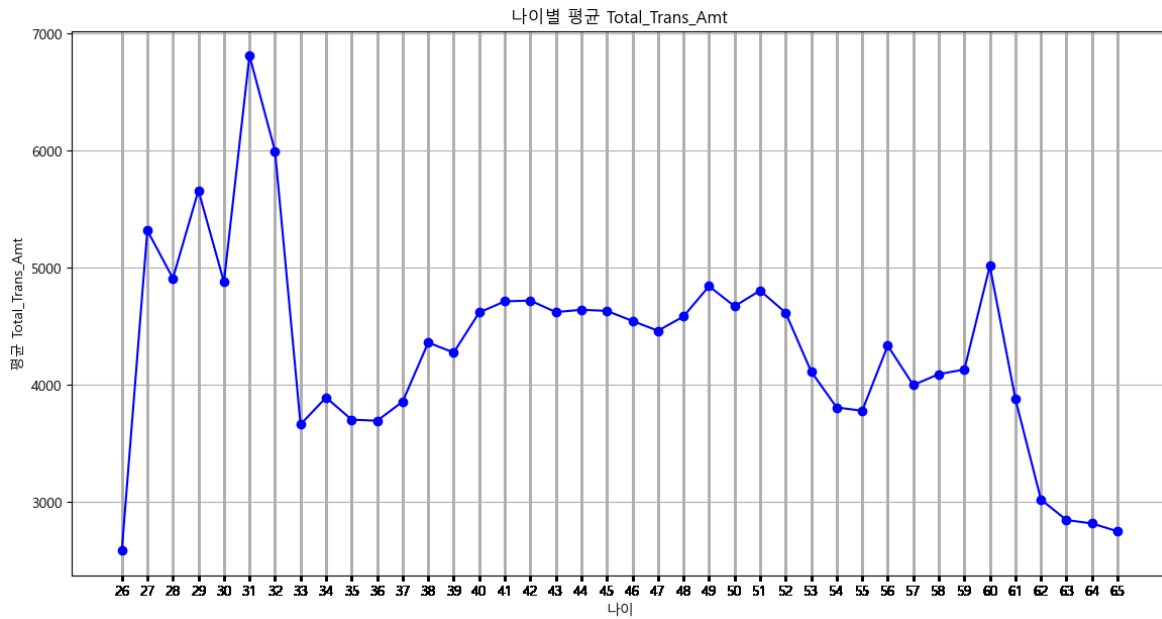
[그림 2.3.2-5 나이 대비 유지 개월수의 이탈률]

고객의 평균과 중앙값은 6~7% 구간에 존재해 있고, 대부분의 데이터가 5-7% 범위에 속하는 것을 알 수 있다. 해당 분포를 기준으로 이탈률을 시각화했을 때, 기존의 가설과는 달리 비율이 커질수록 이탈률이 높아지는 모습을 보인다. 특히 8-10% 구간에서 유독 높은 이탈률을 보이고, 2-3% 구간에서 낮은 이탈률을 기록한다. 이에 가설 “나이 대비 유지 개월 수가 높을수록 카드를 유지할 것이다”는 사실이라 보기 어렵다.

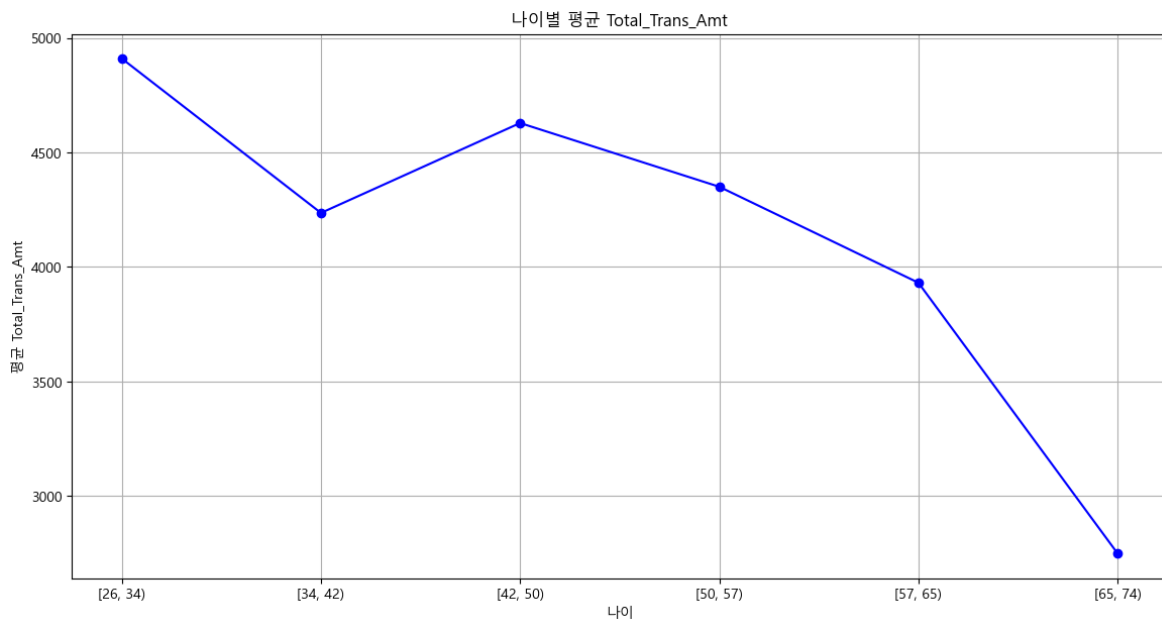
2.3.3. 연령별 한도 대비 사용 금액 분석

한도 대비 사용 금액을 토대로 연령별 소비패턴을 분석하였다. 한도 대비 사용 금액을 추정하기 위해 Total_Trans_Amt(최근 12개월간 총 거래 금액)과 Credit_Limit(신용한도) 칼럼을 사용하여 계산하였다. 이를 통해 연령대별 소비 패턴의 차이를 규명하고, 연령별로 어떤 소비 특성이 나타나는지 분석하였다.

a. 나이별 최근 12개월 간 총 거래 금액



[그림 2.3.3-1 연령별 평균 12개월 총 거래금액]

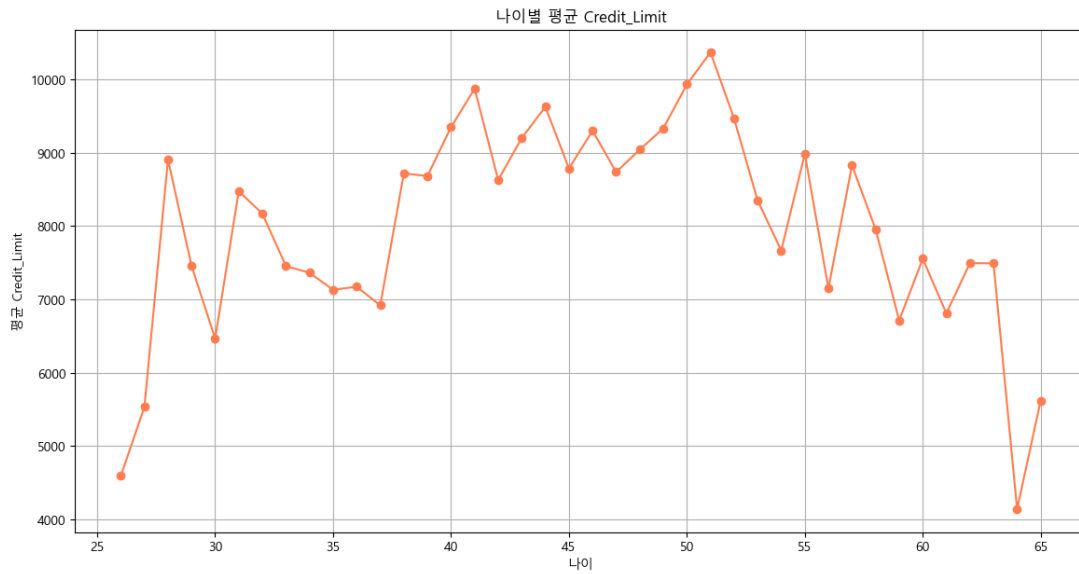


[그림 2.3.3-2 연령 구분별 평균 12개월 총 거래금액]

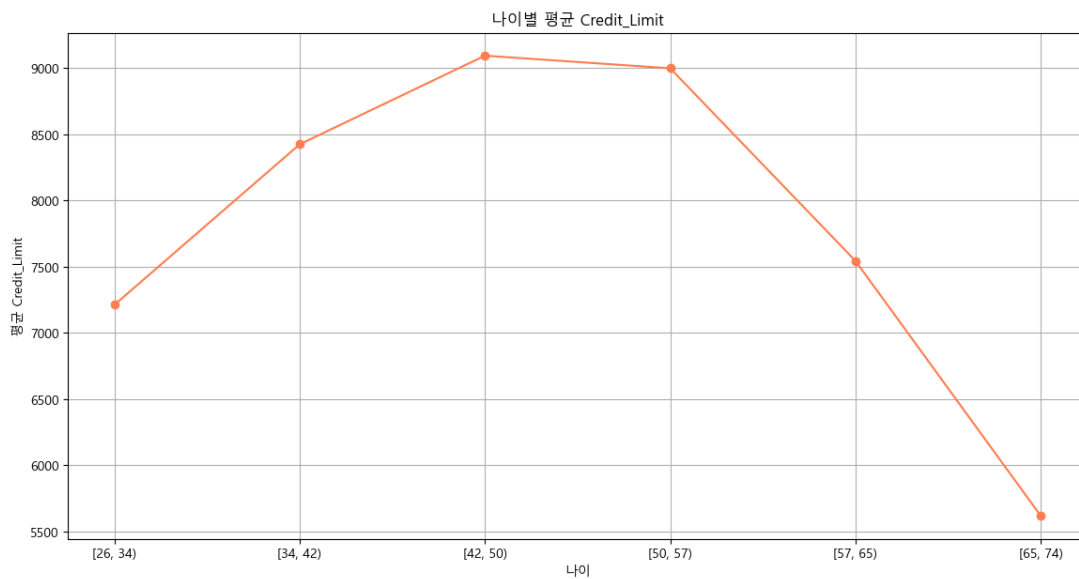
연령별 12개월간 총 거래금액 추세를 보면 26세 이상 34세 이하 구간에서 가장 높은 총 거래 금액을 기록하였다. 이는 젊은 층에서 신용카드 사용이 활발하게 이루어지고 있음을 시사한다. 42세부터 65세까지의 연령대에서는 총 거래금액이 점차 감소하는 경향을 보였다. 연령이 증가함에 따라 신용카드 사용이 줄어드는 추세를 나타낸다.

특히 사항으로 가장 어린 나이대 구간에서 평균 소득이 낮아 거래 금액도 제일 낮을 거라 추정했지만, 오히려 가장 높은 것으로 나타났다. 이는 이 연령대의 소비 성향이 높을 수 있음을 시사한다.

b. 나이별 평균 신용카드 한도



[그림 2.3.3-3 연령별 평균 신용한도]



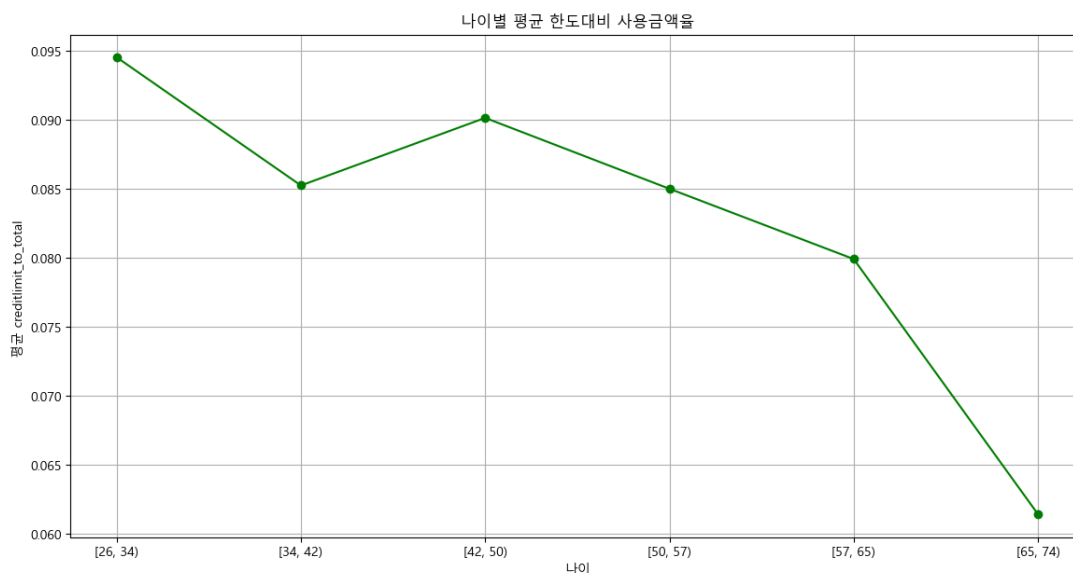
[그림 2.3.3-4 연령 구분별 평균 신용한도]

다음은 연령별 평균 신용 한도를 시각화한 결과이다. 12개월간 총 거래금액과 비교한 결과 26세 이상 34세 이하 구간 고객의 신용한도가 두 번째로 낮게 추정되었고, 고령층 신용한도가 제일 낮게 측정됨을 확인할 수 있다.

이는 젊은 층과 고령층의 신용 한도가 상대적으로 낮게 설정되어 있음을 시사하며, 특히 고령층의 경우 신용한도가 가장 낮아 신용카드 사용에 제한이 있을 수 있음을 나타낸다. 반면, 중년층은

상대적으로 높은 신용 한도를 가지고 있어, 이들의 신용카드 사용이 보다 활발하게 이루어질 가능성이 높음을 시사한다.

c. 연령별 한도 대비 사용 금액율



[그림 2.3.3-5 연령별 평균 한도 대비 사용 금액율]

위의 결과를 토대로 연령별 한도 대비 사용 금액을 시각화 한 결과, 다음과 같은 추세가 확인되었다. 전체적으로 한도 대비 사용 금액은 연령이 증가함에 따라 감소하는 경향을 보였다.

특히, 26세 이상 34세 이하 그룹에서 신용한도는 작지만, 거래금액이 높은 것으로 나타났다. 이는 젊은 층의 소비성향이 크다는 것을 시사한다. 따라서, [가설] “나이가 많다면 한도 대비 사용 금액이 적을 것이다.” 는 맞는 가설로 추정된다.

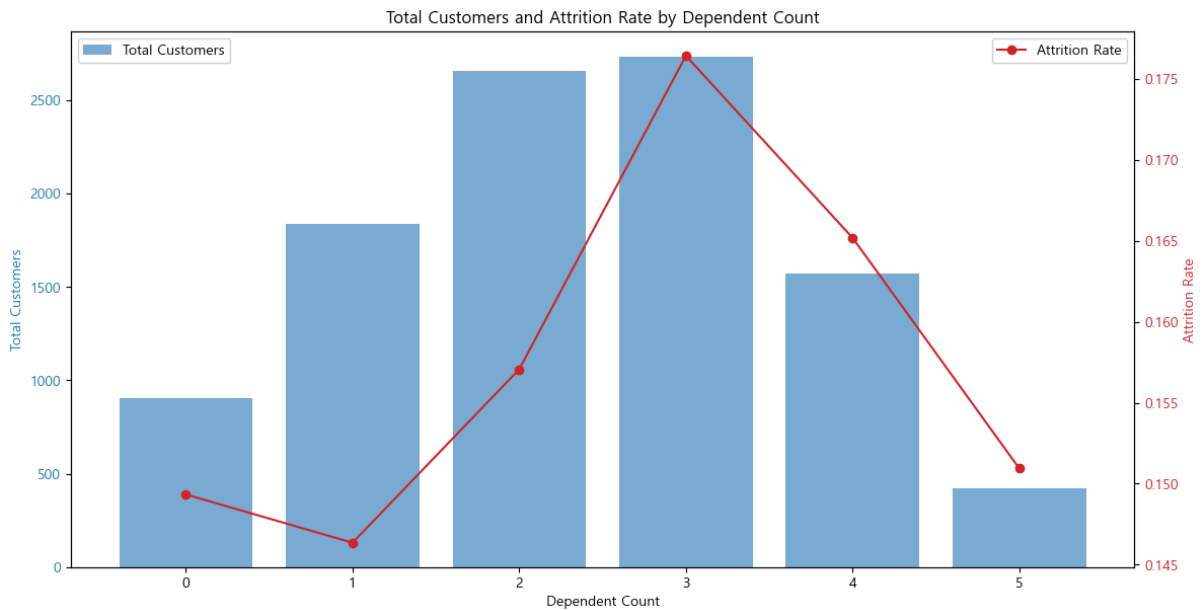
이러한 결과는 은행이 연령대별로 신용 한도와 소비패턴을 고려해 신용카드 전략을 수립할 때, 젊은 층에게는 높은 한도를 제공함으로써 소비를 촉진하고, 고령층에게는 안정적인 금융 서비스를 제공함으로써 이탈률을 막을 수 있을 것이다.

2.3.4. Dependent_count (부양가족 수) 이탈률 분석

인구 통계 기반 변수 중 하나인 부양가족 수가 이탈률에 어떤 영향을 미치는지 분석하였다. 부양가족 수는 0명부터 5명까지 분포되어 있으며, 고객의 평균 부양가족 수는 2.3명으로 나타났다.

이를 통해 부양가족 수와 이탈률 간의 상관관계를 살펴보고, 부양가족 수가 많은 고객일수록 이탈률이 어떻게 변하는지 분석하였다.

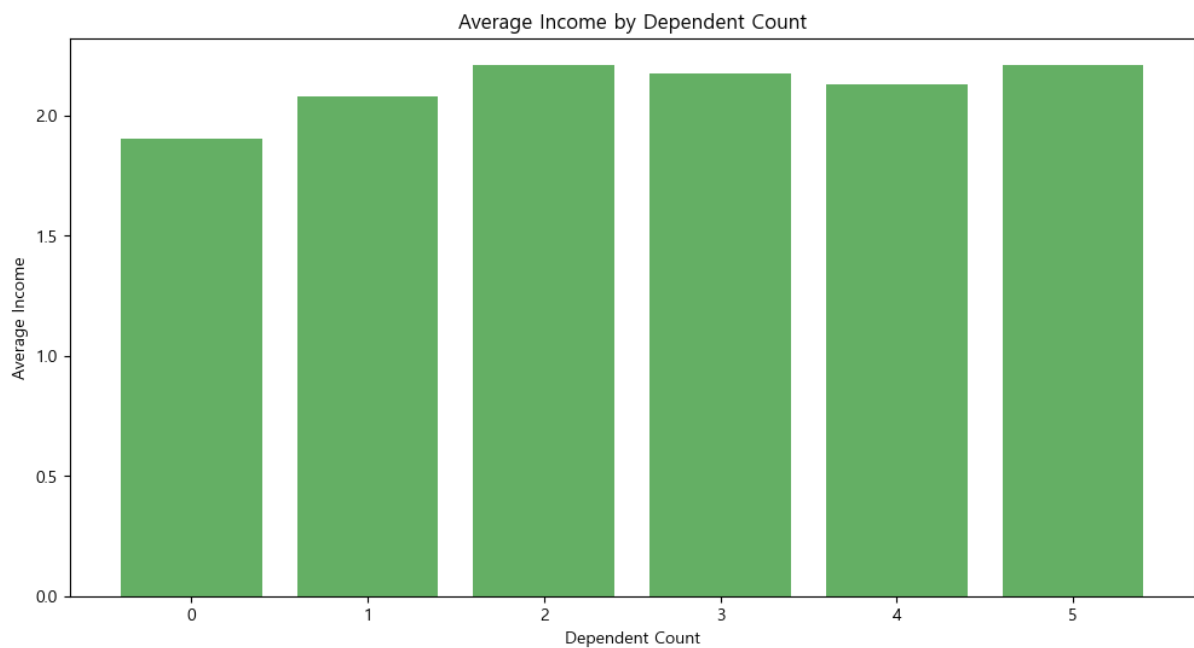
a. 부양 가족 기준 고객 수 및 이탈률



[그림 2.3.4-1 부양가족구분별 인원 및 이탈률]

부양가족 수가 3명인 경우, 이탈률이 17.6%로 가장 높게 나타났다. 반면, 부양가족 수가 1명일 때 이탈률이 14.6%로 가장 낮았다. 부양가족 수가 2명일 때와 3명일 때의 고객 수는 각 2,655명과 2,732명으로 비슷하지만 이탈률은 2명 (0.157%), 3명 (0.176%)로 차이를 보인다.

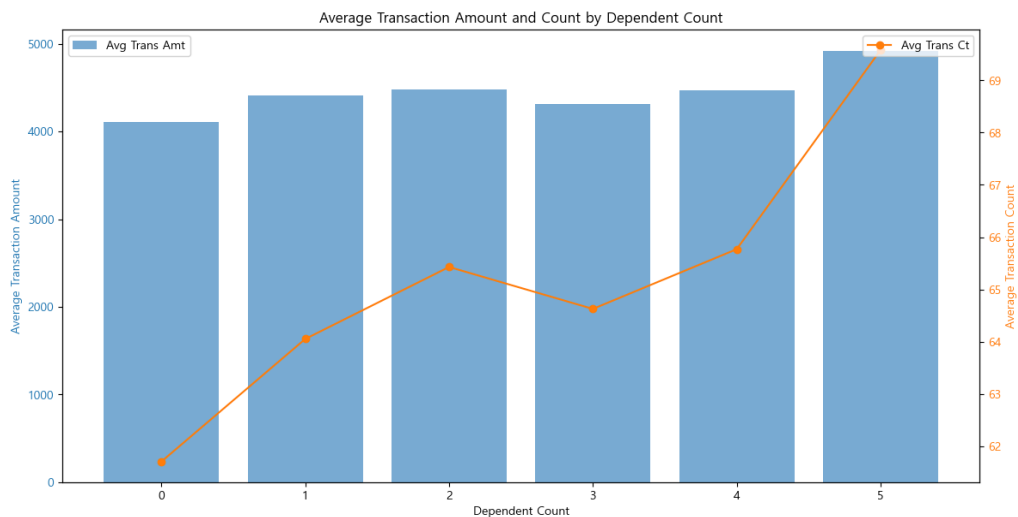
b. 부양가족 수에 따른 평균 소득



[그림 2.3.4-2 부양가족 수에 따른 평균 소득]

위의 시각화 자료는 부양가족 수에 따른 평균 소득을 나타낸다. 분석 결과, 부양가족 수가 많아질수록 평균 소득이 약간 증가하는 경향을 보였다. 이는 부양가족 수가 증가함에 따라 경제적 부담이 증가하기 때문에 더 높은 소득이 필요할 수 있음을 의미한다. 부양가족 수가 2명인 고객층의 평균소득이 3명인 고객층의 평균소득보다 미미하게 높게 나왔다.

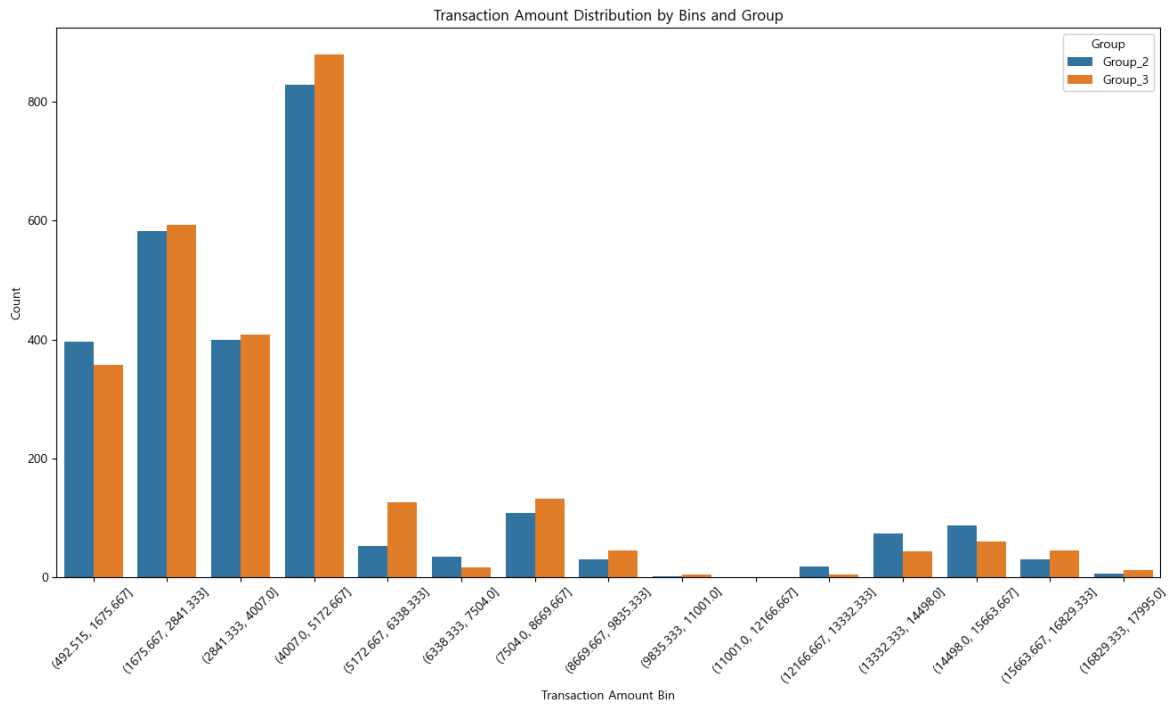
c. 부양가족 수 Total_Trans_Amt(평균 거래 금액)및 Total_Trans_Ct(거래 횟수)



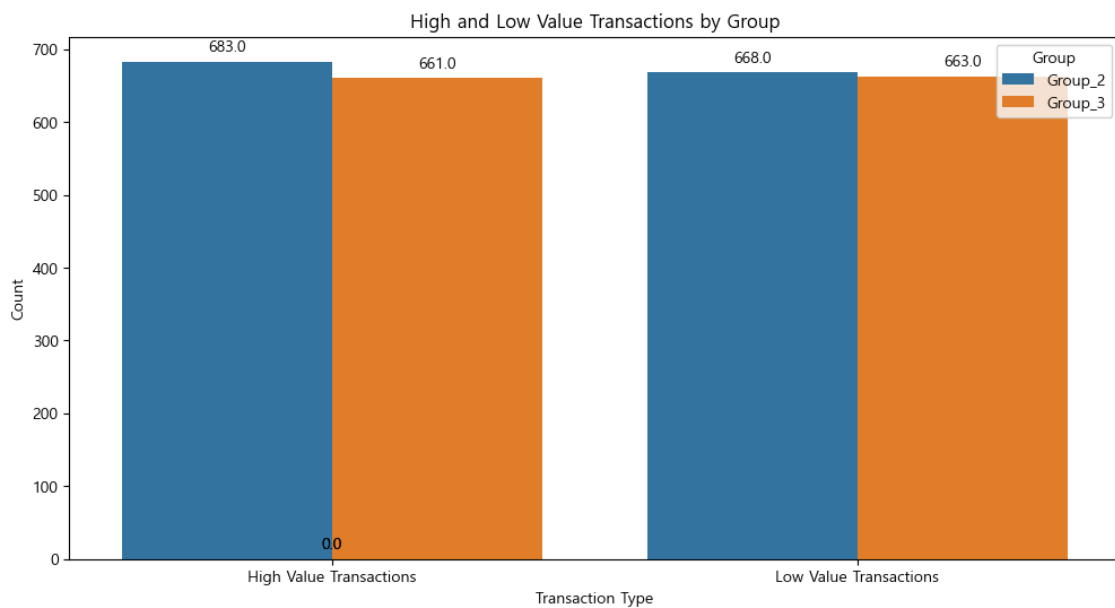
[그림 2.3.4-3 부양가족 수에 따른 평균 거래 금액 및 거래 횟수]

부양가족 수가 많을수록 평균 거래 금액과 거래 횟수가 증가하는 경향을 보였다. 부양가족이 5명인 고객의 평균 거래 금액(4917.36)과 평균 거래 횟수(69.58)가 가장 높았다. 특히, 부양가족이 2명인 고객의 평균 거래 금액(4475.9), 평균 거래 횟수(65.43)회였다. 부양가족이 3명인 평균 거래 금액(4309.1), 평균 거래 횟수(64.62)회였다. 부양가족 수가 2명인 고객층과 3명인 고객층의 평균 거래 금액 및 횟수의 차이가 상대적으로 크지 않음에도 불구하고, 부양가족이 3명인 고객층에서 이탈율이 높다는 점은 추가적인 분석이 필요함을 시사한다.

d. 부양가족수 2명, 3명 고객층의 추가 분석



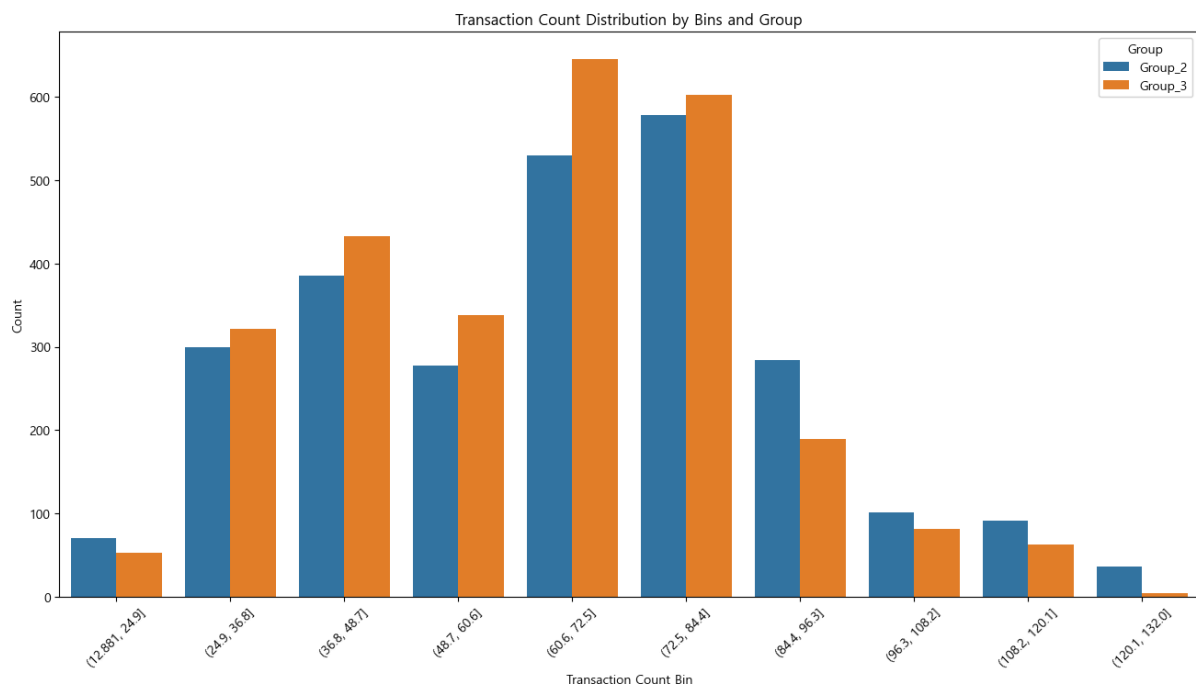
[그림 2.3.4-4 그룹2 그룹3 부양가족 수에 따른 12개월간 총 거래금액]



[그림 2.3.4-5 그룹2, 그룹3 거래금액 상위/ 하위 분포]

부양가족이 2명인 그룹과 3명인 그룹의 전체적인 거래 금액 분포 형태는 유사하게 나타난다. 거래 금액을 상위 25%, 하위 25% 로 구분하여 고액 거래와 저액 거래의 분포를 비교한 결과, 고액거래는 4741.0 달러, 저액 거래는 2155.5 달러로 확인되었다.

부양가족 수가 2명인 그룹2는 고액 거래에서 683명 , 저액 거래에서 668명 으로 나타났으며, 부양가족수가 3명인 그룹3은 고액 거래에서 661명, 저액 거래에서 663명으로 부양가족 수가 2명인 그룹에서 저액, 고액 거래 모두 상대적 높은 빈도를 보이는것을 확인할 수 있다



[그림 2.3.4-6 그룹2, 그룹3 12개월간 거래횟수]

위의 그래프는 부양가족 수가 2인 그룹과 부양가족 수가 3인 그룹의 12개월간 거래횟수 분포를 표현하였다. 대부분의 거래 횟수 구간에서 부양가족 수가 2명인 그룹과 3명인 그룹의 거래빈도가 비슷하게 나타났다. 양그룹의 거래가 중간 금액구간 (약 36.8 - 60.6회) 사이에 집중되어 있다.

평균 거래 횟수(65.43)회 었다.평균 거래 횟수(64.62)로 2명인 그룹이 거래율이 높다 고 판명된다 .

위의 [2.2.1] 거래 건수가 적거나 거래량이 적은 고객의 이탈 확률이 높기 때문에 그룹 3의 이탈률이 2의 이탈률보다 높게 나왔다고 추정된다.

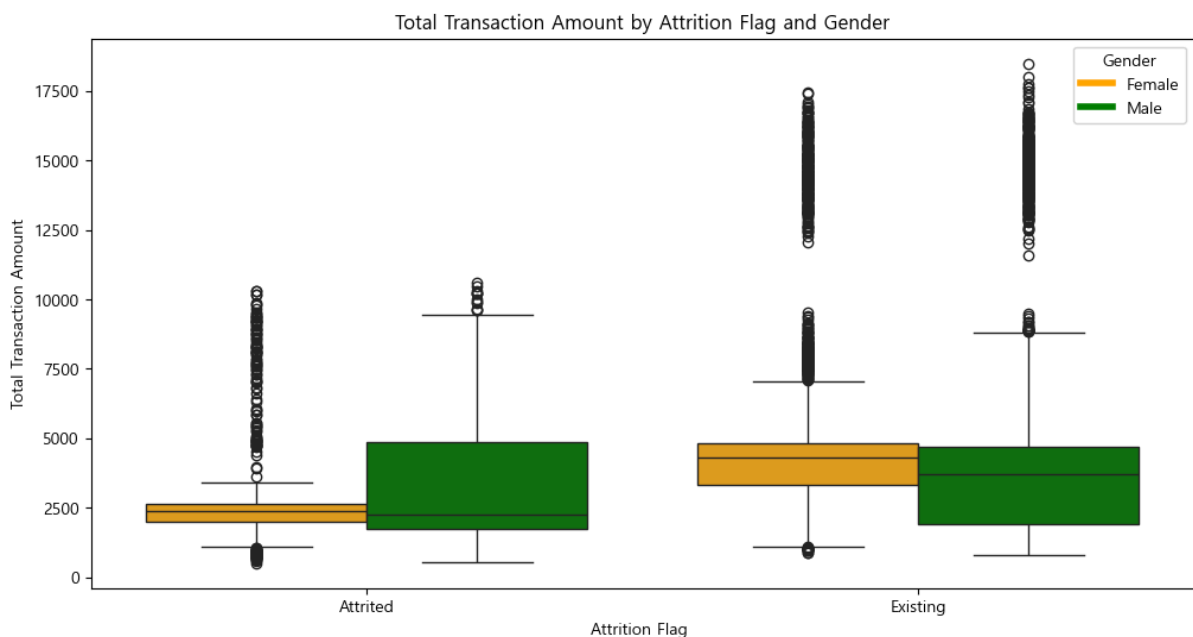
이를 통해 부양가족이 3명인 고객 그룹의 이탈률을 낮추기 위해 Total_Trans_AMT의 거래량이 가장 많은 2,500~5,000구간에 대한 혜택(캐시백이나 포인트 적립 혜택 등)을 강화하는 방법을 구상 하면 좋을 것으로 보인다. 또한, 부양가족 수가 많은 고객들에게는 가족 단위의 혜택을 제공, 금융상품을 추가하는 형식의 마케팅 전략을 수립하면 좋을 것으로 보인다.

결론적으로 [가설] 부양가족 수가 많으면 카드를 유지할 것이라는 가설은 틀린것으로 해석되며 부양가족 수가 3명인 고객들이 카드를 유지하는 비율이 가장 낮고, 부양가족 수가 1명인 그룹들이 카드 유지율이 높은 것으로 나타났다.

2.3.5. Gender(성별) 이탈률 분석

성별에 따른 고객 이탈률을 분석하였다. 분석 결과, 전체 고객 중 여성은 5,358명, 남성은 4,769명으로 여성 비율이 남성보다 다소 높게 나타났다. 이러한 성별 분포를 바탕으로, 성별에 따른 고객 이탈률을 비교 분석하여, 성별에 따른 이탈률의 차이와 그 원인을 규명하고자 한다. 이를 통해 성별 맞춤형 마케팅 전략 수립의 기초자료로 활용할 수 있을 것이다.

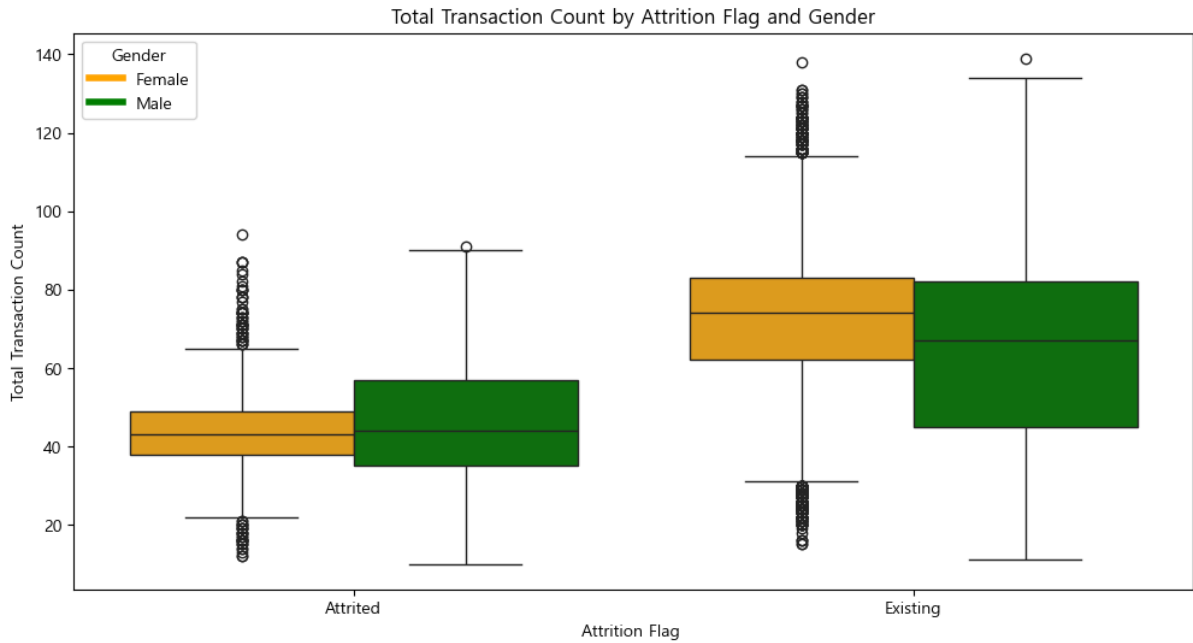
a. 성별에 따른 Total_Trans_Amt(12개월 총 거래금액) 이탈여부 비교



[그림 2.3.5-1 성별에 따른 총 거래금액 이탈여부]

위의 시각화 자료는 성별에 따른 12개월 총 거래 금액과 고객 이탈 여부의 관계를 나타낸다. 분석결과, 이탈한 고객은 유지한 고객보다 총 거래금액이 낮은 경향이 보였다. 이러한 경향은 남성과 여성 모두에게 관찰되었다. 특히 유지한 고객 중 여성의 총 거래 금액이 남성의 총 거래 금액보다 높게 나타났다. 이는 여성 고객이 남성 고객보다 신용카드를 더 활발하게 사용함을 추측할 수 있다.

b. 성별과 Total_Trans_Ct(12개월 총거래건수) 이탈여부 비교



[그림 2.3.5-2 성별에 따른 총 거래 횟수]

위의 시각화 자료는 성별에 따른 총 거래 횟수와 고객 이탈률의 관계를 나타낸다. 분석 결과, 이탈한 고객이 유지한 고객보다 총 거래 횟수가 낮은 경향을 보였다. 유지한 여성 고객 중 많은 비율로 총 거래 횟수가 남성의 총 거래 횟수보다 높게 나타났다.

이러한 결과는 은행이 성별에 따른 맞춤형 전략을 수립하는 데 중요한 인사이트를 제공한다. 신용카드 사용이 저조한 남성 고객을 대상으로 한 추가적인 혜택과 프로모션을 제공함으로써 고객의 이탈율을 낮출 수 있을것으로 보인다.

2.4. 가설 기반 컬럼 제작

2.4.1. 가설 기반 제작 컬럼 선정

더욱 직관적이며 기업의 방향성과 맞는 분석을 위해 기존의 데이터를 매개변수로 한 새로운 컬럼을 제작하도록 한다. 제작할 컬럼은 다음과 같이 정했으며, 가설을 기반으로 수식을 끌어내 새로운 컬럼의 목적과 정의를 정확하게 한다.

- **소비 습관(Spending_Habit)**
 - 리볼빙 잔액이 많다면 소비 습관이 좋지 못할 것이다.
 - 신용 한도 대비 거래 금액이 많다면 소비 습관이 좋지 못할 것이다.
 - 나이가 어리면서 각 결제 건당 평균 금액이 높다면 소비 습관이 좋지 못할 것이다.
 - 분기별 소비 금액이 일정하다면 소비 습관이 좋을 것이다.
- **메인 카드(Main_Card)**
 - 신용카드의 등급이 높다면 메인 카드일 것이다.
 - 카드가 비활성화된 기간이 길다면 메인 카드가 아닐 것이다.
 - 소득에 비해 결제 금액이 높다면 메인 카드일 것이다.
- **금융 리터러시¹(Financial_Literacy)**
 - 나이가 어리고 보유한 상품 수가 많다면 금융 리터러시가 높은 고객일 것이다.
 - 최종 학력이 높다면 금융 리터러시가 높은 고객일 것이다.
 - 은행을 자주 방문했다면 금융 리터러시가 높은 고객일 것이다.
 - 소득 범위가 낮는데 카드의 종류가 높다면 금융 리터러시가 높은 고객일 것이다.
- **우량 고객(Valuable_Customer)**
 - 총 거래 금액이 많다면 우량고객으로 선정한다.
 - 부양가족이 많다면 우량 고객으로 선정한다.
 - 각 결제 건당 평균 금액이 높다면 우량 고객으로 선정한다.
 - 신용 한도에서 사용 금액을 제한 잔액이 낮다면 우량 고객으로 선정한다.
 - 리볼빙 잔액이 적고 거래 횟수가 많다면 우량 고객으로 선정한다.

¹ 개인이 금융 지식과 기술을 통해 재정을 효과적으로 관리하는 능력

2.4.2. 칼럼 제작을 위한 매개변수 선정

각 칼럼의 수식에서 매개변수로 사용할 데이터를 선정한다. 방법은 앞서 세운 가설을 구현하기 위한 컬럼을 선정하는 것으로 한다.

- 소비 습관(*Spending_Habit*)

- Total_Revolving_Bal : 리볼빙 잔액 총합.
- Total_Trans_Amt / Credit_Limit : 신용 한도 대비 총 거래 금액.
- Total_Ct_Chng_Q4_Q1 : 4분기 총 거래금액 대 1분기 비율.
- Total_Trans_Amt / Total_Trans_Ct / Customer_Age : 고객 나이당 평균 결제 금액.

- 메인 카드(*Main_Card*)

- Card_Category : 카드의 등급.
- Months_Inactive_12_mon : 12개월 동안 비활성 상태인 개월 수.
- Total_Trans_Amt / Income_Category : 소득 대비 결제 금액.

- 금융 리터러시(*Financial_Literacy*)

- Customer_Age / Total_Relationship_Count : 나이 대비 보유 제품 수.
- Education_Level : 최종 학력.
- Contacts_Count_12_mon : 12개월 동안 은행을 방문한 횟수.
- Income_Category / Customer_Age : 소득 범위 대비 나이.

- 우량 고객(*Valuable_Customer*)

- Total_Trans_Amt : 총 거래 금액.
- Dependent_count : 부양가족 수.
- Total_Trans_Amt / Total_Trans_Ct : 각 결제 건당 평균 금액.
- Credit_Limit - Total_Trans_Amt : 잔여 신용 한도.
- Total_Revolving_Bal / Total_Trans_Ct : 리볼빙 잔액 대비 거래 횟수.

2.4.3. 칼럼의 의의

- 소비 습관

소비 습관 칼럼은 고객의 소비 패턴을 평가한다. 이를 통해 고객이 어떤 소비 습관을 지니고 있는지를 판단하며, 이 정보는 기업이 개인화된 마케팅 전략을 수립하거나 금융 상품을 제안하는 데 유용하게 활용될 수 있다.

- 메인 카드

자사의 상품이 타사의 상품보다 얼마나 메리트가 있는지를 확인할 수 있으며, 메인 카드로 사용하는 고객에게 적립금을 늘리는 등의 혜택을 주어 상품을 보유하고 있지만 메인 카드로 사용하고 있지 않은 고객을 메인 카드로 전환하는 등으로 활용할 수 있다.

- 금융 리터러시

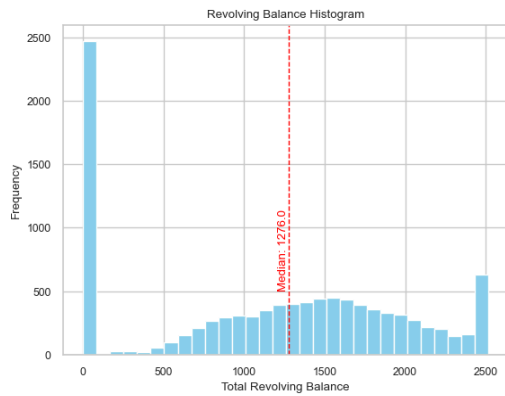
금융 리터러시 칼럼은 고객의 금융 지식수준과 적극성을 평가한다. 이 정보는 금융에 적극적인 고객이 자사의 서비스를 어떻게 평가하는 지를 분석할 수 있게 하는 주요 지표가 되며, 복잡한 혜택 구조를 가진 서비스의 마케팅 대상을 선별하는 데 도움을 줄 수 있다.

- 우량 고객

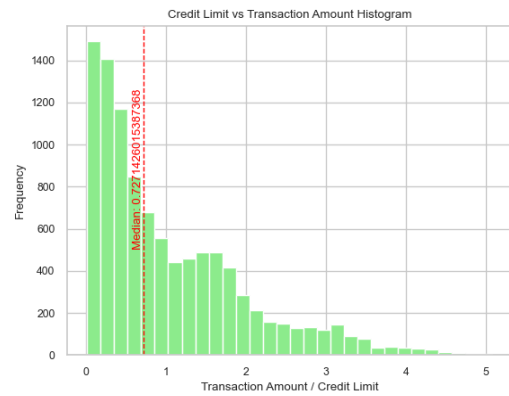
이 컬럼은 우량 고객인지를 수치상으로 평가한다. 이 정보는 자사에 이익을 주는 고객을 선별하여 트래킹할 수 있도록 하며, 우량 고객을 충성고객으로 이어지게 하여 매출을 증대시킬 수 있다는 것에 의의를 둔다.

2.4.4. 파라미터 시각화

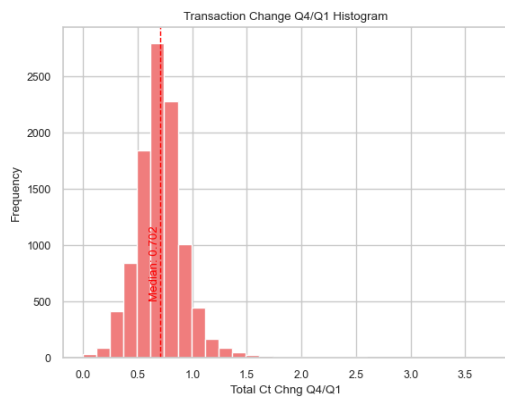
- 소비 습관



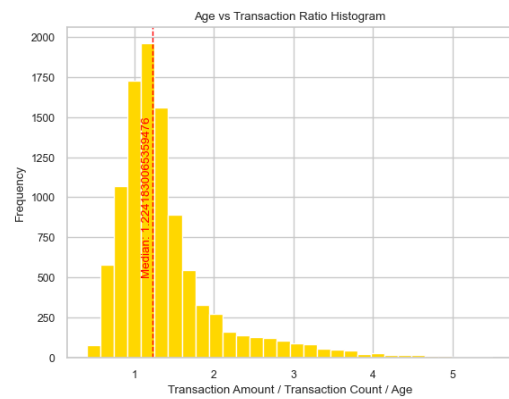
[그림 2.4.4.-1 리볼빙 잔액 히스토그램]



[그림 2.4.4.-2 신용 한도 대비 거래 금액 히스토그램]

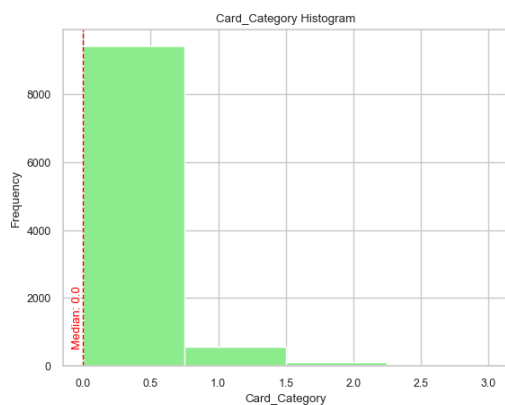


[그림 2.4.4.-3 4분기 대비 1분기 거래 건수]

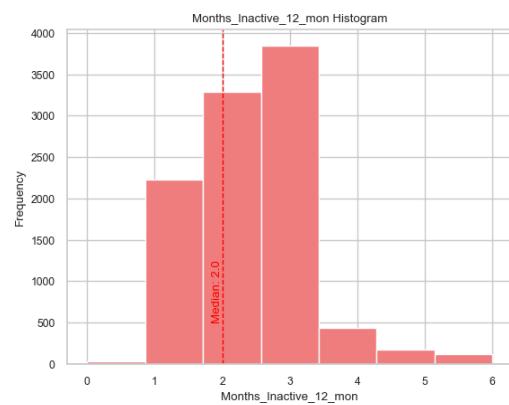


[그림 2.4.4.-4 나이 대비 평균 결제 금액 히스토그램]

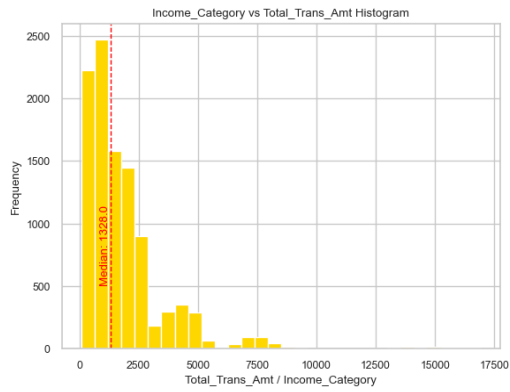
- 메인 카드



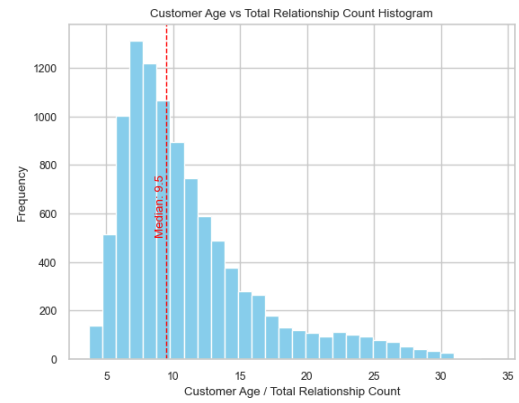
[그림 2.4.4.-5 카드 등급 히스토그램]



[그림 2.4.4.-6 12개월 중 비활성화 기간 히스토그램]

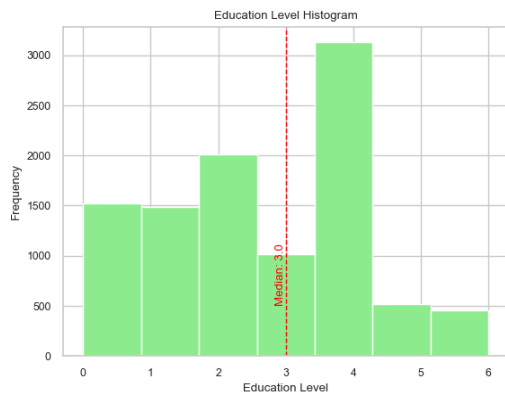


[그림 2.4.4.-7 소득 대비 결제 금액 히스토그램]

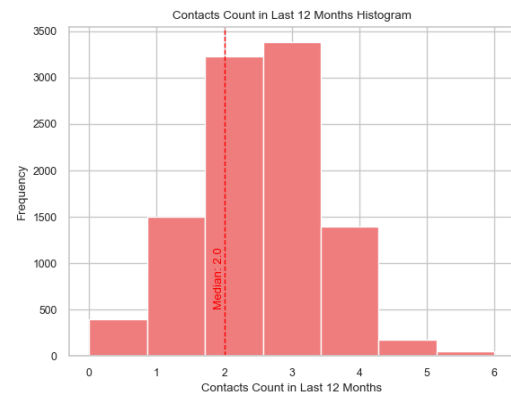


[그림 2.4.4.-8 나이 대비 보유한 상품 수 히스토그램]

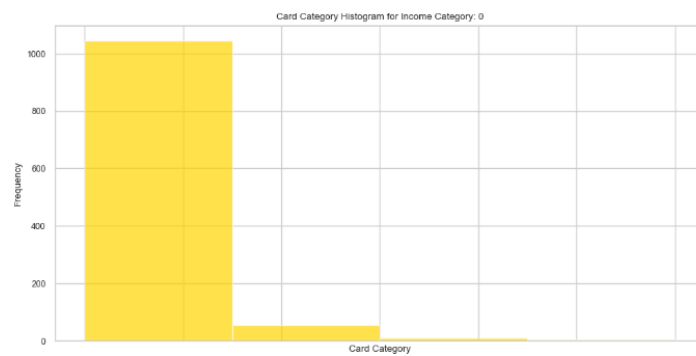
- 금융 리터러시



[그림 2.4.4.-9 최종 학력 히스토그램]

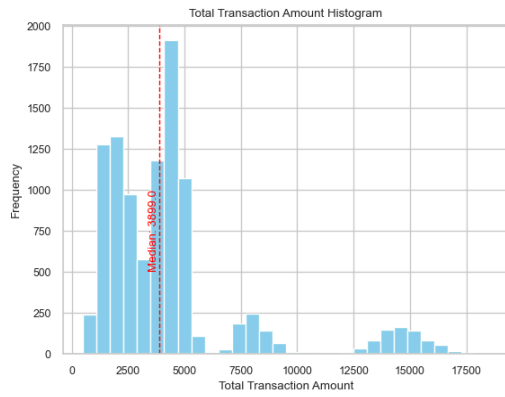


[그림 2.4.4.-10 은행 방문 횟수 히스토그램]

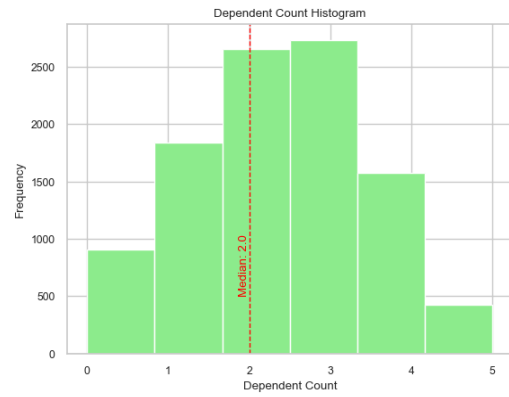


[그림 2.4.4.-11 소득 범위 별 카드 종류 히스토그램 (일부 발췌)]

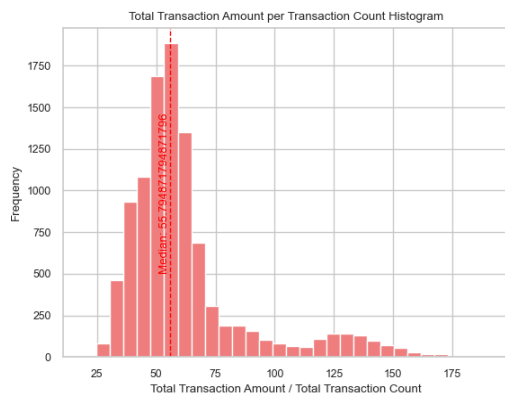
- 우량 고객



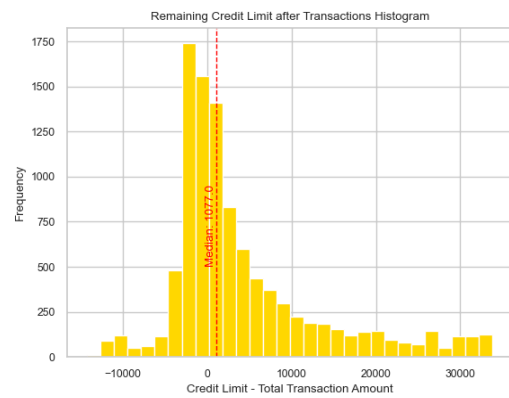
[그림 2.4.4.-13 총 거래 금액 히스토그램]



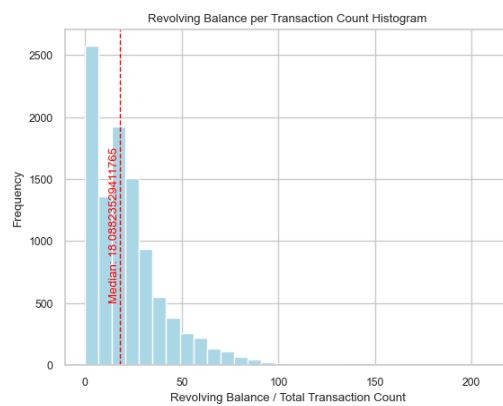
[그림 2.4.4.-14 부양가족의 수 히스토그램]



[그림 2.4.4.-15 결제 건당 평균 결제 금액 히스토그램]



[그림 2.4.4.-16 신용 한도와 총 거래 금액의 차 히스토그램]



[그림 2.4.4.-17 리볼빙 잔액 대비 거래 횟수 히스토그램]

2.4.5. 수식 및 결과

- 소비 습관

리볼빙 잔액 점수(revolv_bal_score)는 고객의 리볼빙 잔액이 전체 데이터에서 최소 리볼빙 잔액과 최대 리볼빙 잔액 사이에서 어디에 위치하는지를 백분율로 나타낸 것이다. 고객의 리볼빙 잔액에서 최소 리볼빙 잔액을 뺀 후, 그 값을 최대 리볼빙 잔액에서 최소 리볼빙 잔액을 뺀 값으로 나누고, 100을 곱해 계산한다.

거래 금액 대비 한도 점수(trans_to_limit_score)는 고객의 총 거래 금액을 신용 한도로 나눈 비율을 이용해 계산한다. 이 비율에서 최소 비율을 뺀 후, 최대 비율과 최소 비율의 차이로 나누고, 100을 곱한다.

거래 횟수 변동 점수(trans_to_limit_score)는 4분기에서 1분기로의 거래 횟수 변화량이 전체 데이터에서 어디에 위치하는지를 나타낸다. 변화량에서 최솟값을 뺀 후, 최댓값과 최솟값의 차이로 나누고, 100을 곱한다.

거래 금액 대비 나이 점수(trans_to_age_score)는 고객의 총 거래 금액을 총거래 횟수와 고객 나이의 제곱을 곱한 값으로 나눈 비율을 이용해 계산한다. 이 비율에서 최소 비율을 뺀 후, 최대 비율과 최소 비율의 차이로 나누고, 100을 곱한다.

다음은 최종 점수 계산 수식이다.

$$\begin{aligned} \text{spending_habit_ratio} = & \text{revolv_bal_score} \times 0.25 \\ & + \text{trans_to_limit_score} \times 0.3 \\ & + \text{trans_change_q4q1_score} \times 0.3 \\ & + \text{trans_to_age_score} \times 0.15 \end{aligned}$$

- 메인 카드

카드 카테고리 점수(card_category_score)는 카드 카테고리값을 사전에 정의된 매핑을 통해 점수로 변환한 것이다. 예를 들어, 카드 카테고리 0은 1점, 1은 34점, 2는 67점, 3은 100점으로 변환된다.

12개월 동안 비활동 개월 수에 따른 점수(months_inactive_score)는 비활동 개월 수 값을 사전에 정의된 매핑을 통해 점수로 변환한 것이다. 예를 들어, 비활동 개월 수가 0이면 100점, 1이면 90점, 2이면 80점, 3이면 60점, 4이면 40점, 5이면 20점, 6이면 0점으로 변환된다.

소득 대비 거래 점수(income_trans_score)는 총 거래 금액을 소득 범위로 나눈 비율을 이용해 계산한다. 이 비율에서 최소 소득을 뺀 후, 그 값을 최대 소득에서 최소 소득을 뺀 값으로 나누고, 100을 곱해 점수를 계산한다. 계산된 점수가 소득 범위의 중앙값 이상이면 계산된 점수를 사용하고, 그렇지 않으면 100점으로 설정한다.

다음은 최종 점수 계산 수식이다.

$$\begin{aligned} \text{main_card_score} = & \text{card_category_score} \times 0.4 \\ & + \text{months_inactive_score} \times 0.25 \\ & + \text{income_trans_score} \times 0.35 \end{aligned}$$

- 금융 리터러시

나이-제품 비율 점수(age_product_score)는 100에서 이 비율이 전체 데이터에서 차지하는 백분율을 뺀 값이다. 고객의 나이에서 최소 나이-제품 비율을 뺀 후, 그 값을 최대 나이-제품 비율에서 최소 나이-제품 비율을 뺀 값으로 나누고, 100을 곱해 백분율을 구한 다음, 이를 100에서 빼서 계산한다.

교육 수준 점수(education_level_score)는 100에서 교육 수준이 전체 데이터에서 차지하는 백분율을 뺀 값이다. 고객의 교육 수준에서 최소 교육 수준을 뺀 후, 그 값을 최대 교육 수준에서 최소 교육 수준을 뺀 값으로 나누고, 100을 곱해 백분율을 구한 다음, 이를 100에서 뺀다.

12개월 동안의 은행 방문 횟수에 따른 점수(education_level_score)는 은행 방문 횟수를 사전에 정의된 매핑을 통해 점수로 변환한 것이다. 예를 들어, 은행 방문 횟수가 0이면 0점, 1이면 20점, 2이면 40점, 3이면 60점, 4이면 80점, 5이면 90점, 6이면 100점으로 변환된다.

소득 대비 카드 비율 점수(contact_score)는 고객의 소득 범위를 나이로 나눈 비율을 이용해 계산한다. 이 비율에서 최소 소득 대비 카드 비율을 뺀 후, 그 값을 최대 소득 대비 카드 비율에서 최소 소득 대비 카드 비율을 뺀 값으로 나누고, 100을 곱해 백분율로 나타낸다.

다음은 최종 점수 계산 수식이다.

$$\begin{aligned} \text{financial_literacy_score} = & \text{age_product_score} \times 0.3 \\ & + \text{education_level_score} \times 0.2 \\ & + \text{contact_score} \times 0.2 \\ & + \text{income_card_score} \times 0.3 \end{aligned}$$

- 우량 고객

고객의 총 거래 금액 점수(total_trans_amt_score)는 총 거래 금액이 전체 데이터에서 최소 총 거래 금액과 최대 총 거래 금액 사이에서 어디에 위치하는지를 백분율로 나타낸 것이다. 총 거래 금액에서 최소 거래 금액을 뺀 후, 그 값을 최대 거래 금액에서 최소 거래 금액을 뺀 값으로 나누고, 100을 곱해 백분율을 구한다. 계산된 점수가 중간값 이상이면 100점으로 설정하고, 그렇지 않으면 계산된 점수를 사용한다.

부양가족 수 점수(dependent_count_score)는 고객의 부양가족 수가 전체 데이터에서 최소 부양가족 수와 최대 부양가족 수 사이에서 어디에 위치하는지를 백분율로 나타낸 것이다. 부양가족 수에서 최소 부양가족 수를 뺀 후, 그 값을 최대 부양가족 수에서 최소 부양가족 수를 뺀 값으로 나누고, 100을 곱해 백분율을 구한다.

평균 지출 금액 점수(spent_amt_score)는 총 거래 금액을 총 거래 횟수로 나눈 평균 지출 금액을 기준으로 계산된다. 평균 지출 금액에서 최소 평균 지출 금액을 뺀 후, 그 값을 최대 평균 지출 금액에서 최소 평균 지출 금액을 뺀 값으로 나누고, 100을 곱해 백분율을 구한다.

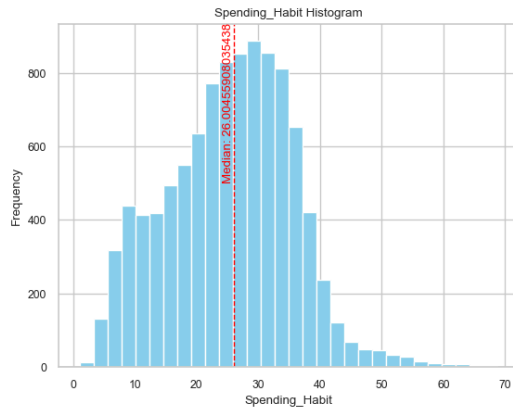
카드 잔액 점수(card_balance_score)는 신용 한도에서 총 거래 금액을 뺀 카드 잔액을 기준으로 계산된다. 카드 잔액에서 최소 카드 잔액을 뺀 후, 그 값을 최대 카드 잔액에서 최소 카드 잔액을 뺀 값으로 나누고, 100을 곱해 백분율을 구한 다음, 이를 100에서 뺀다.

리볼빙 거래 비율 점수(revolving_trans_score)는 리볼빙 잔액을 총 거래 횟수로 나눈 비율을 기준으로 계산된다. 이 비율에서 최소 리볼빙 거래 비율을 뺀 후, 그 값을 최대 리볼빙 거래 비율에서 최소 리볼빙 거래 비율을 뺀 값으로 나누고, 100을 곱해 백분율을 구한다.

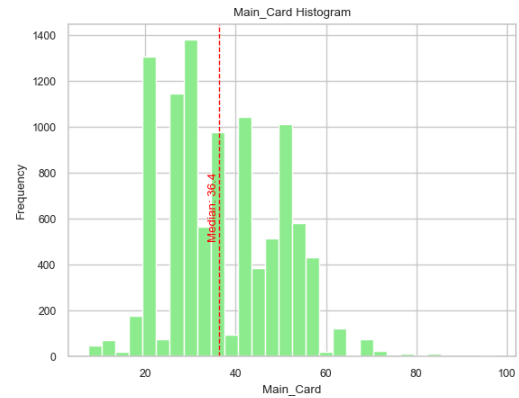
다음은 최종 점수 계산 수식이다.

$$\begin{aligned}
\text{valuable_customer_score} = & \text{total_trans_amt_score} \times 0.3 \\
& + \text{dependent_count_score} \times 0.15 \\
& + \text{spent_amt_score} \times 0.2 \\
& + \text{card_balance_score} \times 0.2 \\
& + \text{revolving_trans_score} \times 0.15
\end{aligned}$$

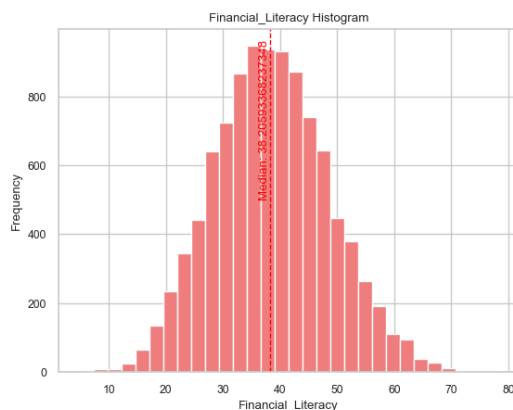
- 결과



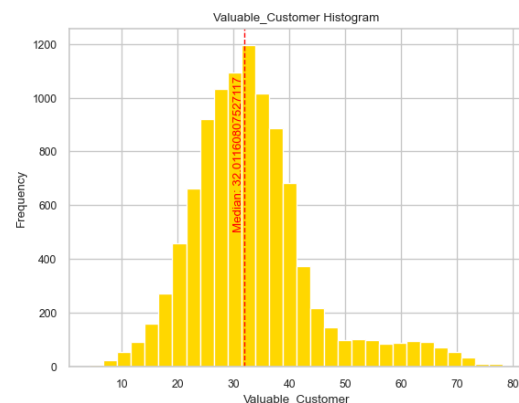
[그림 2.4.5.-1 소비 습관 히스토그램]



[그림 2.4.5.-2 메인 카드 히스토그램]



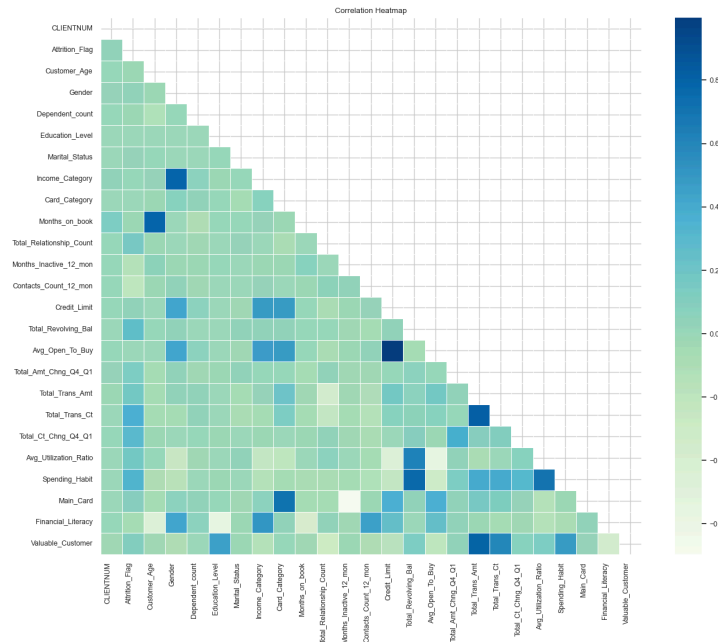
[그림 2.4.5.-3 금융 리터러시 히스토그램]



[그림 2.4.5.-4 우량 고객 히스토그램]

2.4.6. 제작 칼럼 활용 결과

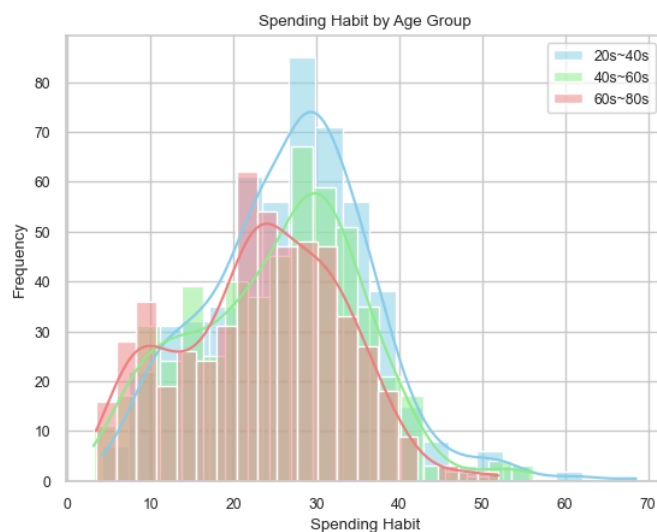
제작된 칼럼을 이용한다면 더욱 다양하고 고차원적인 분석을 할 수 있다. 그중 칼럼의 의의와 목적에 따라 몇 가지 주제를 정하고 분석해 보고자 한다. 본격적인 분석을 위해서 먼저 앞서 제작한 ‘Spending_Habit’, ‘Main_Card’, ‘Financial_Literacy’, ‘Valuable_Customer’ 변수들을 데이터셋에 포함했다.



[그림 2.4.6.-1 제작 컬럼을 포함한 데이터셋의 히트맵]

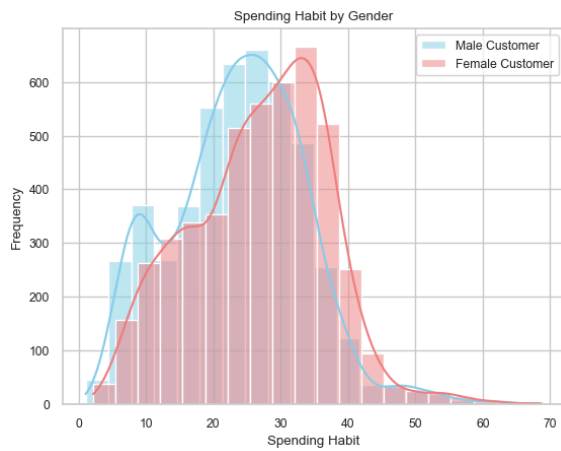
첫 번째로 분석 배경을 설정한다. 분석 배경은 새로운 자사의 서비스(Green card)를 만들고 홍보하기 위해 타겟 고객을 선정하고 마케팅 대상을 추려내야 한다는 것으로 한다. 여기서 ‘Green card’는 소비 습관이 좋지 않은 고객을 위해 소비 습관을 개선하도록 돕는 카드 상품이며, 기존의 ‘Blue card’와 ‘Silver card’의 중간에 위치하는 상품이다.

먼저 타겟을 선정한다. 이를 위해 나이별 소비 습관, 성별별 소비 습관, 소득별 소비 습관, 성별별 금융 리터러시, 금융 리터러시별 소득, 메인 카드 여부별 소비 금액을 확인한다.

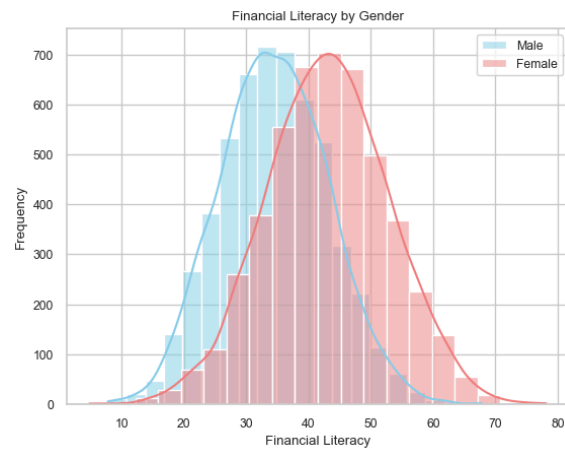


[그림 2.4.6.-2 나이별 소비 습관 히스토그램]

나이가 많아질수록 소비 습관이 좋다. 따라서 ‘Green card’은 어린 고객들을 위한 상품으로 타겟팅 한다.

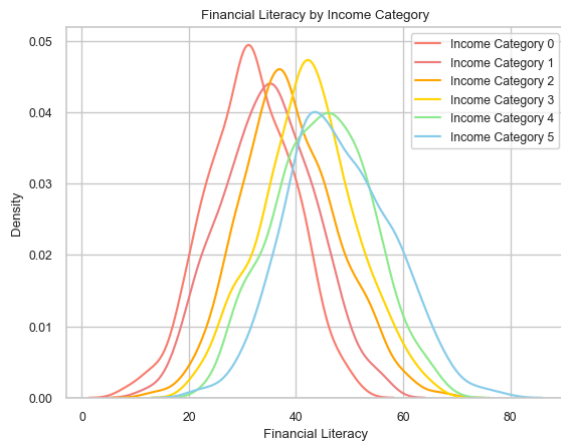


[그림 2.4.6.-3 성별별 소비 습관 히스토그램]



[그림 2.4.6.-4 성별별 금융 리터러시 히스토그램]

여성이 남성보다 소비 습관은 좋지 않고 금융 리터러시는 높다. 따라서 소비 습관 개선을 목적으로 한 ‘Green card’의 타겟층은 여성으로 설정했다. 금융 이해도가 높기 때문에 남성을 타겟팅 하는 것보다 높은 효과를 낼 수 있을 것으로 예상된다.

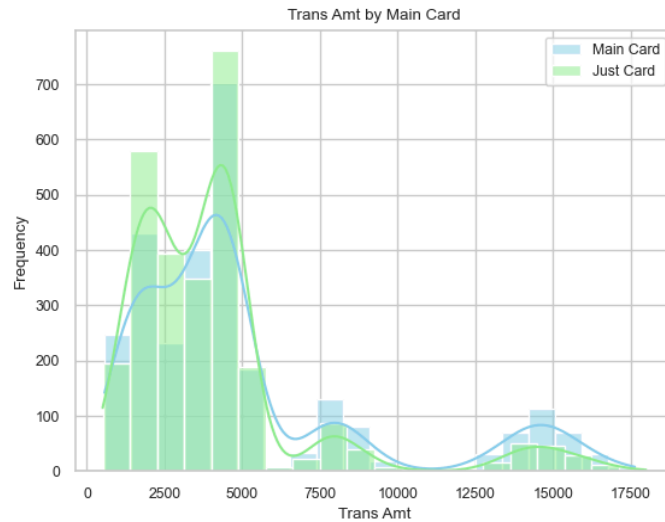


[그림 2.4.6.-5 소득 범위별 금융 리터러시 히스토그램]



[그림 2.4.6.-6 소득 범위별 소비 습관 히스토그램]

소득 범위가 높을수록 금융 리터러시가 높고 소득이 증가할수록 소비 습관이 좋다. 따라서 소득 전 범위의 고객을 대상으로 하거나 중앙에 있는 중위소득인 고객을 타겟팅 하는 것이 높은 효과를 낼 수 있을 것이다.



[그림 2.4.6.-7 메인 카드 여부별 소비 금액 히스토그램]

메인 카드가 아닌 고객 중에서는 적은 금액을 소비하는 고객이 상대적으로 많고, 메인 카드인 고객은 그 반대이다. 따라서 자사의 카드를 메인 카드로 사용하는 고객이 ‘Green card’를 사용했을 때 가져오는 이익이 높을 것이다.

이를 종합하였을 때, 젊고 소득이 중위소득에 가까운 여성 고객을 타겟팅 하며, 기존 자사의 상품을 메인 카드로 사용하고 있는 고객에게 배너광고를 띄우고 카드의 디자인을 트렌디하고 단정하게 하는 등의 마케팅을 해야 한다는 결론을 내릴 수 있다. 추가로 우량 고객 칼럼으로 고가치 고객을 선정하고 소프트 런칭²을 시행해 사용자 피드백을 받는다면 더욱 좋은 성과를 낼 수 있을 것이다.

² 서비스를 정식으로 출시하기 전 선별된 고객에게만 제공하는 행위

3. 결론

3.1. 결론 및 기대 효과

이번 분석을 통해 신용카드 고객 데이터를 활용하여 고객 이탈 요인을 다각도로 분석하고, 이를 바탕으로 고객 유치 및 유지 방안을 도출하였다. 주요 분석 결과는 다음과 같다.

거래 건수와 거래 금액을 바탕으로 고객 이탈률을 분석한 결과, 거래 건수가 적거나 거래 금액이 낮은 고객일수록 이탈 확률이 높은 것으로 나타났다. 이는 "상품을 자주 구매하는 사람은 카드를 유지할 것이다"라는 가설을 지지하는 결과로, 자주 거래하는 고객의 이탈을 줄이기 위한 전략 수립이 필요함을 시사한다.

총 리볼빙 잔액과 부채 잔액 대비 신용 한도를 바탕으로 고객 이탈률을 분석한 결과, 리볼빙 잔액이 높은 구간이 아닌 특정 구간에서 이탈률이 높게 나타났다. 이는 "부채를 많이 보유한 사람은 이탈할 것이다"라는 가설과 상반되는 결과로, 리볼빙 잔액 관리의 중요성을 다시 한 번 강조하였다. 따라서, 리볼빙 잔액이 특정 구간에 있는 고객에게 특별한 혜택이나 맞춤형 서비스를 제공함으로써 이탈률을 낮출 수 있을 것이다.

가설 "보유한 상품 수가 적을수록 이탈할 것이다"는 실제로 보유 상품 수가 적은 고객이 이탈 가능성이 높다는 분석 결과로 뒷받침되었다. 이는 고객에게 다양한 상품을 제안하고, 보유 상품 수를 늘림으로써 이탈률을 줄일 수 있음을 의미한다.

인구통계 변수를 분석한 결과, 나이와 유지 개월 수가 많을수록 이탈 가능성은 낮아지는 것으로 나타났다. 이는 "나이 대비 유지 개월 수가 높을수록 카드를 유지할 것이다"라는 가설을 지지하며, 연령이 높은 고객일수록 이탈률이 낮다는 것을 보여준다. 또한, "부양가족 수가 많으면 카드를 유지할 것이다"라는 가설도 지지되었으며, 이는 가족이 많은 고객이 신용카드를 더욱 필요로 한다는 것을 나타낸다.

반면, "나이가 적으면 적을수록 이탈할 것이다"라는 가설도 분석 결과에 의해 확인되었으며, 젊은 층 고객의 이탈 방지를 위한 전략적 접근이 필요함을 시사한다.

이번 분석을 통해 도출된 이탈 요인과 관련된 통찰력은 신용카드 회사가 고객의 행동 패턴을 이해하고, 이에 기반한 맞춤형 전략을 세우는 데 중요한 역할을 한다. 이를 통해 고객의 다양한 요구를 보다 효율적으로 충족시킬 수 있으며, 고객 만족도와 충성도를 높이는 데 기여할 것이다.

특히, 가설 기반 컬럼 제작을 통해 데이터 분석의 정확성과 효율성을 높였다. 각 가설에 따라 주요 변수들을 선정하고 이를 컬럼으로 제작함으로써, 보다 명확하고 체계적인 분석이 가능하였다. 이를 통해 고객 이탈 요인을 명확히 규명하고, 보다 효과적인 고객 유지 방안을 도출할 수 있었다.

이번 프로젝트의 결과를 통해 다양한 기대효과를 기대할 수 있다. 첫째, 이탈 가능성이 높은 고객을 사전에 식별하고, 맞춤형 유지 전략을 수립할 수 있게 되었다. 예를 들어, 거래 건수가 적은 고객에게는 특별한 혜택을 제공하거나, 리볼빙 잔액이 특정 구간에 있는 고객에게는 이자율 할인 등의 혜택을 제공하여 이탈을 방지할 수 있다. 이와 같은 맞춤형 접근 방식은 고객의 만족도를 높이고, 충성도를 강화하는 데 큰 도움이 될 것이다.

둘째, 이탈 요인을 기반으로 신규 고객 유치 전략을 개선할 수 있다. 이를 통해 젊은 층이나 부양가족 수가 많은 고객을 타겟으로 한 마케팅 캠페인을 통해 더 많은 신규 고객을 유치할 수 있을 것이다. 이는 기업의 매출 증대와 시장 점유율 확대에 긍정적인 영향을 미칠 것이다.

셋째, 고객 분석을 통해 발견된 불만족 요인을 개선함으로써, 전체적인 서비스 품질을 향상시킬 수 있을 것이다. 이는 고객 만족도를 높이고, 장기적으로 고객 충성도를 강화하는 데 기여할 것이다. 고객의 다양한 요구와 기대를 반영한 서비스 개선은 경쟁력 있는 기업으로서의 입지를 더욱 공고히 하는 데 중요한 역할을 할 것이다.

넷째, 데이터 분석을 통해 얻은 통찰력을 바탕으로 보다 정확하고 효율적인 의사결정을 내릴 수 있을 것이다. 이는 기업의 경쟁력을 높이고, 시장에서의 입지를 강화하는 데 도움을 줄 것이다. 예측 모델을 활용한 전략적 의사결정은 불확실성을 줄이고, 보다 효과적인 경영 활동을 가능하게 할 것이다.

이번 프로젝트는 신용카드 고객의 행동을 이해하고, 이를 기반으로 효과적인 고객 관리 전략을 수립하는 데 중요한 기초 자료를 제공하였다. 앞으로도 지속적인 데이터 분석과 피드백을 통해 더욱 발전된 고객 관리 방안을 도출할 수 있을 것이다. 이를 통해 신용카드 회사는 고객의 이탈을 줄이고, 충성도 높은 고객층을 확보하여 지속 가능한 성장을 이룰 수 있을 것으로 기대된다. 나아가, 데이터 기반의 전략적 접근은 미래의 시장 변화에 유연하게 대응할 수 있는 능력을 배양하는 데 큰 기여를 할 것이다.

3.2. 참여자 정보 및 역할

김희령: 인구통계변수 분석과 보고서 작성, 결론 보고서 초안 작성, 발표자료 제작(인구통계변수 파트)

오수연: 고객 사용 내역 분석과 보고서 작성, 인구통계변수 분석 보고서 수정, 서론 보고서 작성,

발표자료 제작(고객 사용 내역 및 데이터 설명 파트)

황석준: 가설 기반 컬럼 제작과 보고서 작성, 결론 보고서 작성, 데이터 전처리, 발표자료 제작(가설 기반 컬럼 제작 및 서론, 결론 파트)

4. 참고문헌

영문 웹페이지

[Becky Pokora, "Credit Card Statistics And Trends 2024", Forbes ADVISOR](#)

[What Is a Credit Utilization Ratio?, EQUIFAX, n.d, 2024년 06월 12일 접속](#)

[What is revolving credit and how does it work?, CapitalOne, 2024년 01월 26일 작성, 2024년 06월 12일 접속](#)

국문 간행물

이건창, 정남호, 신경식. 신용카드 시장에서 데이터마ining을 이용한 이탈고객 분석 (Customer Churning Analysis by using Data Mining in Credit Card Market).

한국지능정보시스템학회:학술대회논문집, 2001.

국문 웹페이지

[김연경, "\[김연경 칼럼\] 놓치기 쉽지만 알아두면 좋은 '회계연도'", 여행신문, 2014년 05월 12일 작성, 2024년 06월 12일 접속](#)

[천규승, "신용과 부채관리의 원리", KDI경제정보센터, 2012년 10월 31일 작성, 2024년 06월 12일 접속](#)

[SC제일은행 프리미엄 고객 기준](#)

[신한은행 Tops Club 제도 안내](#)