

정민영(Ryan)

Data Analyst

이름: 정민영 (Ryan Jeong)

생년월일 : 1990.02.09

연락처 : 010-7406-9949

이메일 : jeongmy9002@gmail.com

PORTFOLIO

Contents

- 1** [Internal Data / EDA]
사내 고객 데이터 분석을 통한 내부 평가지표 설계 및 이행
- 2** [이커머스 고객 Segmentation을 위한 RFM 분석 / EDA]
RFM분석을 이용한 고객 Segmentation 및 경영 전략 수립
- 3** [금융거래 데이터 분석 / ML]
신용카드 데이터를 통한 고객 이탈 예측 및 마케팅 전략 수립

Project 1

사내 고객 데이터 분석을 통한 내부 평가지표 설계

프로젝트 개요

프로젝트 일정 / 인원

- 2022 / 2023 사내 개인 프로젝트

문제정의

- 1년에 매니징 할 수 있는 감사 프로젝트는 1년 50개 이내
- 팀KPI 증진을 위해 능률적인 50개의 프로젝트 선정 필요
- 능률적인 프로젝트가 무엇인지 정의
- 각 프로젝트의 특징을 고려하여 50개의 연간 프로젝트 선정하여 KPI 달성 및 효율적인 팀을 운영

수행역할

- Excel과 Pivot table function을 통한 데이터 전처리
- Feature정의와 점수화를 통한 고객 세분화
- 고객 세분화에 따른 프로젝트 선정 및 팀 운영 전략 수립

GitHub

Internal Data

[2018-2022 자체 사내 고객 데이터]



Quant features

- 투입 된 인원수
- 총 일한 가격
- Budget
- 실제 받은 가격



Qual features

- 재무 기장 상태
- Tone at the top
- 회사와의 관계
- 실제 결제일수

Score & Ranking

분석 방법론

Definition

어떠한 프로젝트가 팀에게 도움이 되는 파일인지
가장 제일 금액의 프로젝트가 항상 좋은 파일은 아님
으로 능률적인 프로젝트 정의 자체 평가

Scoring & Ranking

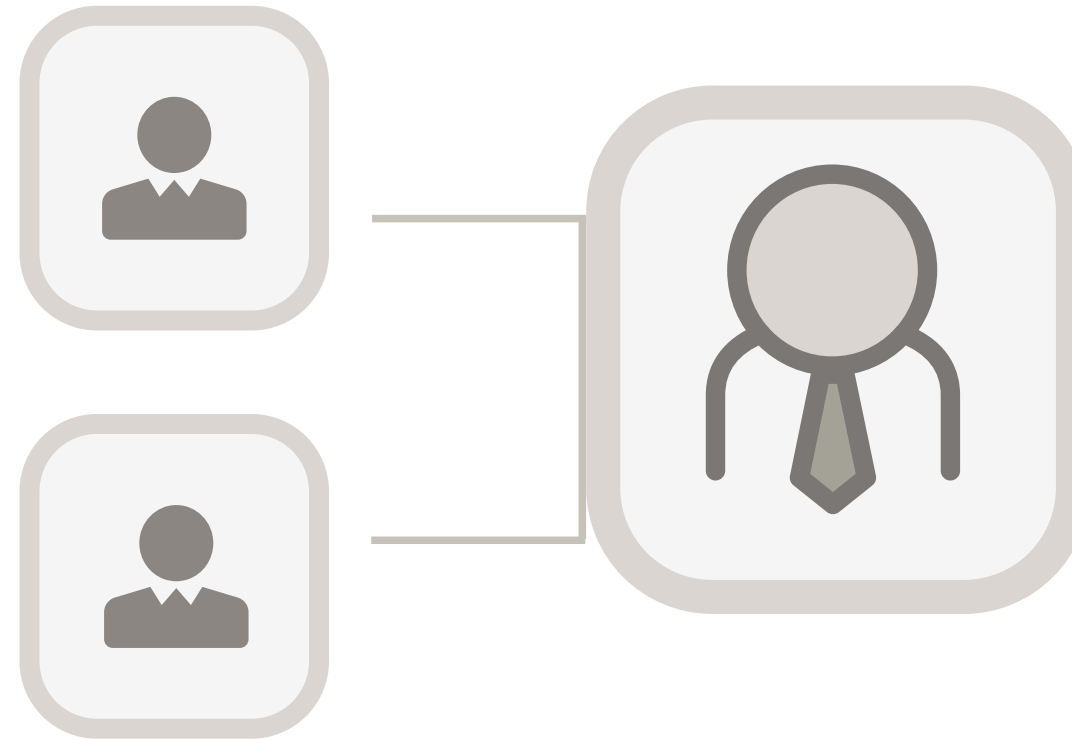
자체 기준에 따른 정량적/정성적 데이터에 가중치
를 부여, 해당 가중치로 점수를 곱한 후 합산하여
프로젝트 / 고객 순위 설정

Decision Making

설정된 평가지표로 팀내 토론 후 유지해야할 프로젝
트와 거절해야할 프로젝트 및 새로운 프로젝트 점수
지표 비교 후 연간 프로젝트 설정

KPI

- 연간 총 프로젝트 금액
- 실제 프로젝트 정산일
- 프로젝트 갯수
- 고객 유지 및 고객사 관리
- 신규 고객 유치



정의

능률적인 프로젝트란?

1. 금액 높다고 항상 좋은 프로젝트는 아님
2. Budget 대비 Cost(실제 비용) 적은 고객
3. 감사원들이 프로젝트에 대한 이해도가 높은 프로젝트
4. 결제가 빨리 이루어 지는 고객
5. 회계감사의 대한 고객의 태도

Scoring

Qualitative features

- $(\text{Budget} - \text{Cost}) / \text{Actual price}$
- 순위 1-50 부여후 1등은 50점, 50등은 1점 부여
- 순위에 따른 차등 점수 부여

Quantitative features

- 10가지의 정성적인 feature들을 활용
- 1 - 5까지의 점수 개별 평가 후 (5점 만점)
- Quant 50 + qual 50 점
- 합산된 점수로 최종 순위 부여
- 최하위 10개와
- 새로운 대기 중인 고객 10개 비교하여 결정

01

분석 결과 및 운영

- 2022년 대비 2023년 Collection amount 15% 상승
- 2022년 대비 2023년 매출 8% 상승
- 2023년 대비 2024년 매출 5% 상승
- 능률적인 연간 프로젝트 계획 및 운영으로 팀 KPI달성 및 팀의 매출 상승에 기여

배운점 및 발전 가능성

- 데이터 분석 직무가 아니더라도 데이터에 대한 기본적인 이해를 통한 기본적인 데이터 전처리에 대한 이해 상승
- 개인적인 KPI 달성을 위한 개인 프로젝트
- 2022년 대비 2023년 2024년 상승하는 성과 보여주었으나 내가 만든 지표가 계속 성과가 날 것 인지 나 뿐만 아님 다른 다른 팀원들 에게도 효과적인지 검증해보지 못한 아쉬움



Project 2

고객 Segmentation 및 RFM 분석을 활용한 경영 전략 수립

프로젝트 개요

프로젝트 일정 / 인원

- 2024 06.10 ~ 2024.06.24 (2주) / 개인

문제정의

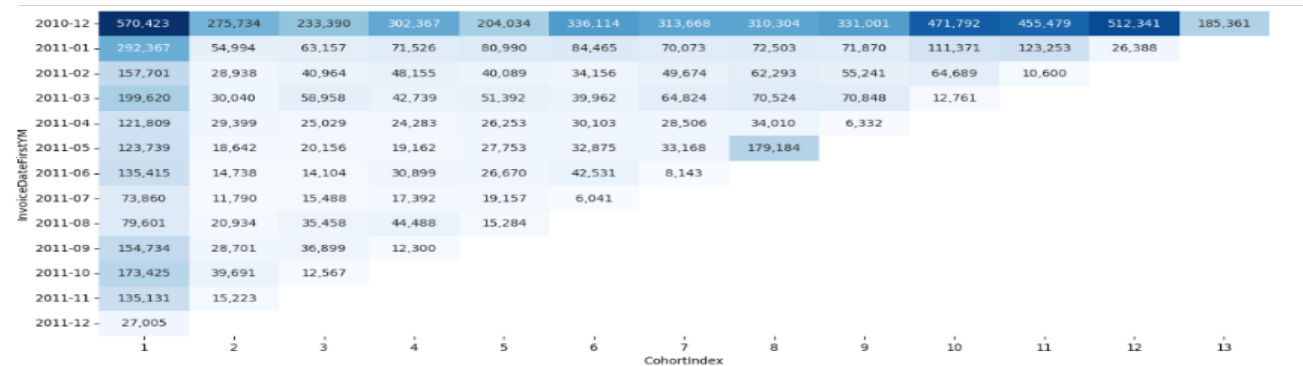
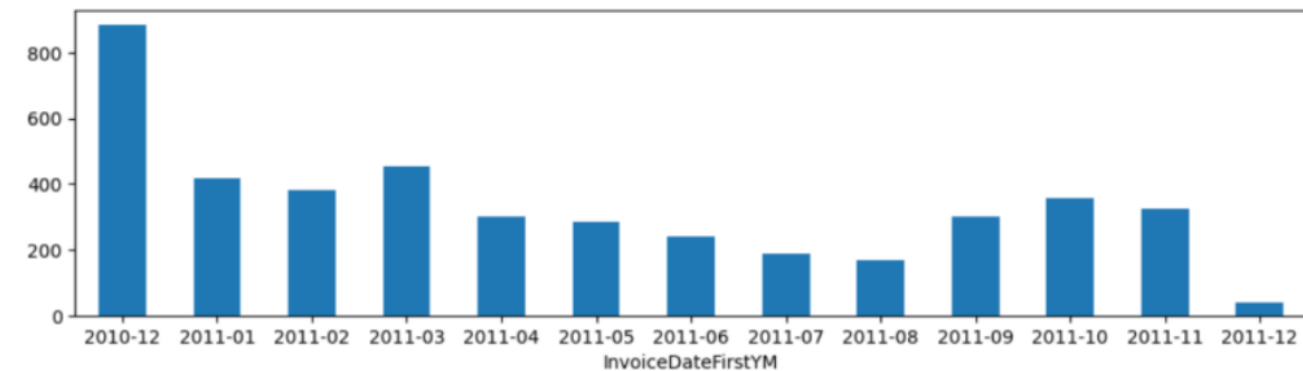
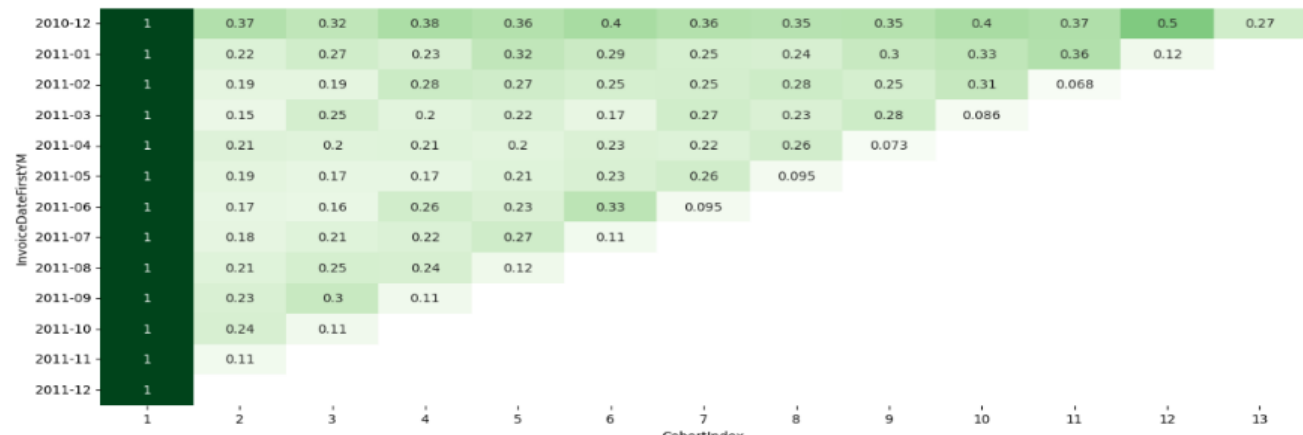
- 오픈마켓 런칭 이후 서비스 정체기로 인한 영업이익, 사용 고객수 감소
- 오픈마켓 플랫폼 고객 데이터를 활용하여 고객군을 나누고 각 고객군별 현재 고객들의 서비스 이용 현황 파악 및 경영 전략 수립

수행역할

주역할

- RFM 기법 활용 및 지표 정의를 통한 고객 Segmentation
- RFM 기반의 데이터 분석을 통한 인사이트 도출
- Retention Analysis를 통한 고객 수 변화 파악
- Python, numpy, pandas 활용한 데이터 EDA/전처리
- matplotlib, seaborn을 이용한 데이터 시각화

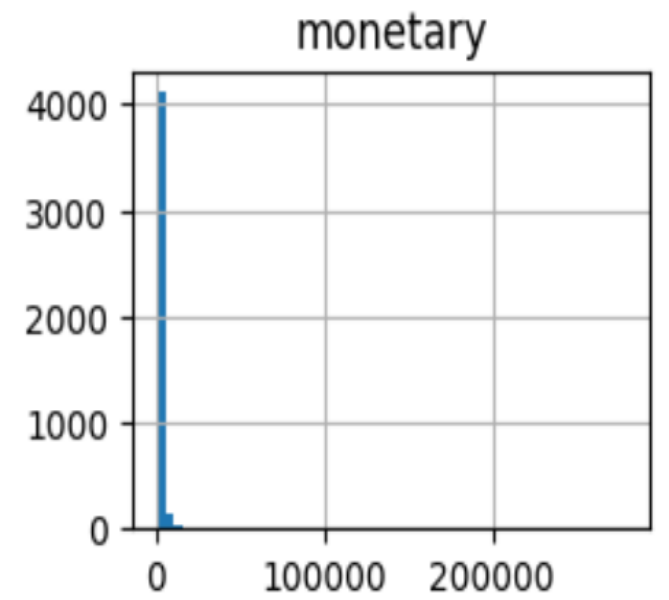
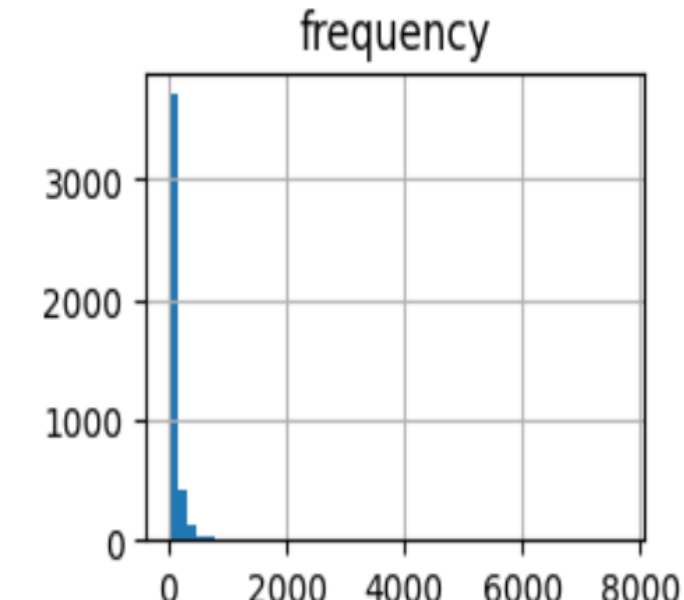
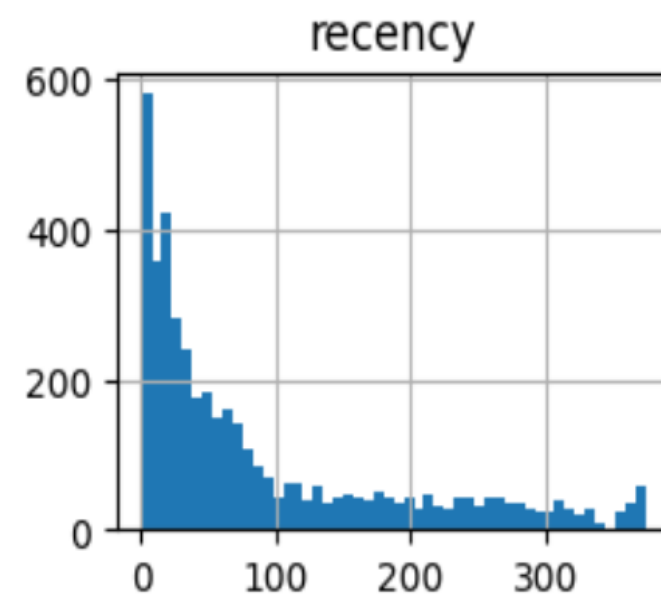
Retention Rate



Analysis

- 2010년부터 13개월간 5.4M 개의 이커머스 구매 데이터
- 매월 가입자를 기준으로 Cohort Analysis 진행
- 2010년 처음 오픈 후 새로운 고객의 유입이 꾸준히 감소
- 하지만 처음 유입된 고객층에 충성 고객층이 있음을 확인

RFM



RFM별 분포도 확인

- Recency : 최근에 구매를한 고객들의 수가 많음
- Frequency : 1회성으로 구매한 고객들의 수가 많음
- Monetary : 적은 금액을 구매한 고객들의 수가 많음
- Recency를 제외한 Frequency와 Monetary는 왼쪽으로 치우침이 심함

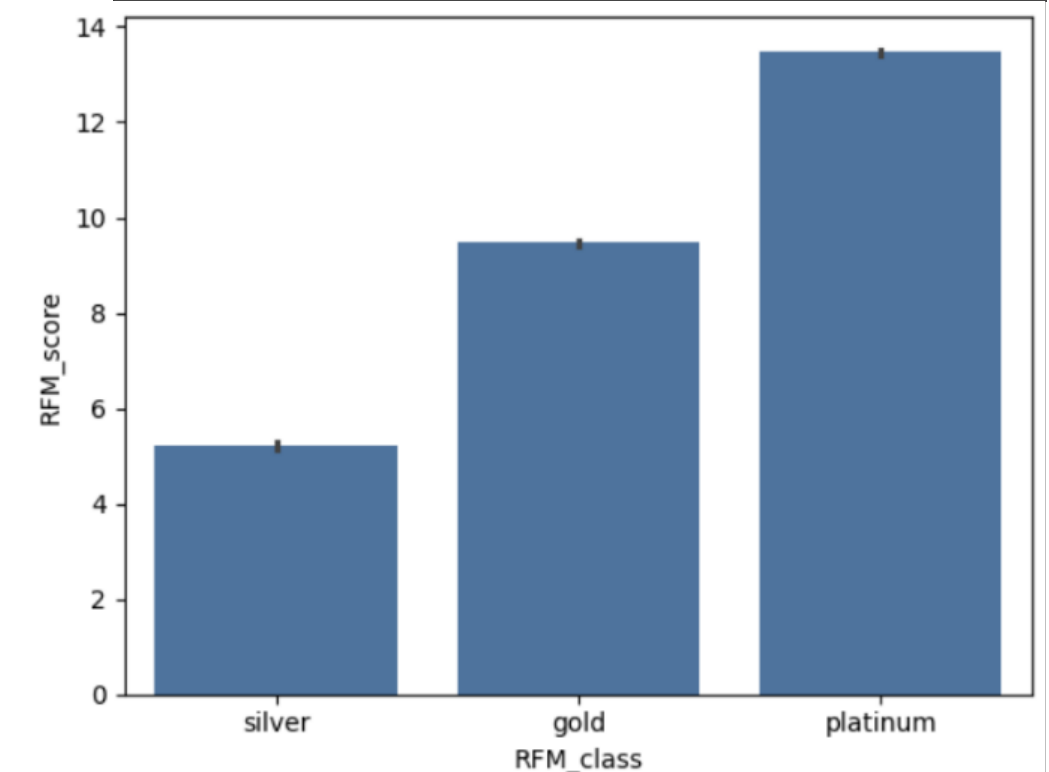
점수 구간화 (Scoring)

	recency	frequency	monetary	
	mean	mean	mean	sum
RFM_score				
3	278	7	138	37,309
4	204	11	200	52,138
5	183	16	295	108,729
6	126	20	371	142,014
7	103	26	898	345,746
8	87	36	628	227,483
9	70	46	858	309,676
10	59	62	1,123	392,016
11	45	80	1,445	487,016
12	35	108	1,794	592,058
13	23	140	3,080	973,379
14	16	230	4,797	1,467,897
15	5	439	11,596	3,583,278

- R, F, M의 각각의 점수를 합산
- RFM 값이 높을 수록 우수 고객
- RFM 값이 낮을 수록 이탈 / 휴면 고객
- RFM 점수가 높을 수록 최근에 높은 금액으로 구매를 한 정황이 확인됨

- Scoring : RFM 각 요인들을 5등급으로 분류
- Qcut을 이용하여 각요소에 20%를 적용
- Recency는 최근일수록 높은 점수를 부여하고, Frequency, Monetary는 값이 클 수록 높은 점수를 부여함
- RFM합산 점수를 Qcut을 이용 하여 Silver-Gold-Platinum으로 고객군 segmentation 기법 적용

	recency	frequency	monetary	R	F	M
CustomerID						
12346.0	326	1	77183.60	1	1	5
12347.0	2	182	4310.00	5	5	5
12348.0	75	31	1797.24	2	3	4
12349.0	19	73	1757.55	4	4	4
12350.0	310	17	334.40	1	2	2



RFM 분석결과 및 경영 전략

RFM_class	recency	frequency			monetary	
	mean	mean	mean	sum	count	
silver	170	17	411	685,936	1,668	
gold	66	56	1,005	1,416,192	1,409	
platinum	20	227	5,247	6,616,612	1,261	

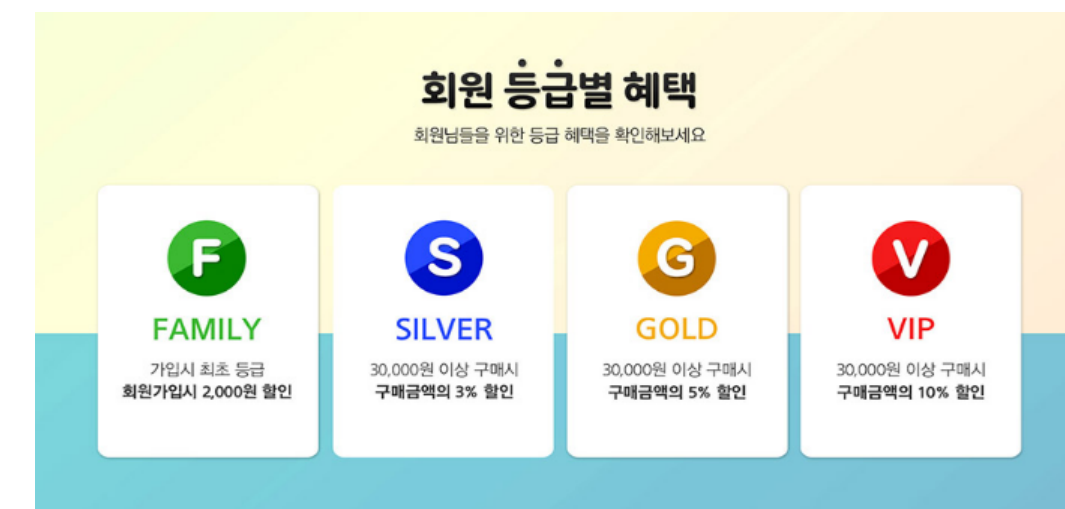
등급별 고객 군에 따른 마케팅 전략

- RECENCY, FREQUENCY, MONETARY 점수에 따른 차등 등급별 혜택 적용 필요
- 상품 구매시 다음 등급 혜택에 대한 프로모션 진행 및 기준 설명
- 다음 등급까지의 구매 횟수와 남은 금액을 알려주어 구매 유도

RFM 점수에 따른 마케팅 전략

- RECENCY(최근성) - R이 낮은 경우, COMEBACK 프로모션 알림 전송 및 빠른 구매시 혜택 증정
- FREQUENCY(빈도) - F가 낮은 경우, 방문을 자주하면 주는 포인트, 구매 혜택을 알림으로 사이트 자주 방문 유도
- MONETARY(구매금액) - M이 낮은 경우 추가적인 구매시 할인 혜택 부여, 더 높은 금액의 구매를 유도

[예시]



[예시]



배운점

RFM 점수에 따른 마케팅 전략

- Cohort Analysis / Retention rate 에 대한 이해
- RFM를 이용한 Segmentation에 대한 전반적인 이해도 향상
- Python, numpy, pandas에 대한 이해도 향상

Project 3

신용카드 데이터를 통한 고객 이탈 예측 및 마케팅 전략 수립

프로젝트 개요

프로젝트 일정 / 인원

- 2024 06.24 ~ 2024.07.20 (4주) / 4명

문제정의

- 신용카드 거래 데이터를 활용한 데이터 분석
- 리스크에 따른 고객군 정의 및 고객 분류
- 머신러닝을 이용한 패턴 도출
- 도출된 패턴을 활용한 맞춤형 마케팅 전략 수립

수행역할

주역할

- Python, numpy, pandas 활용한 데이터 EDA/전처리
- matplotlib, seaborn을 이용한 데이터 시각화
- 머신러닝 / Decision Tree를 사용한 분류 모델 구축
- 결과를 기반으로한 인사이트 도출 및 전략 수립

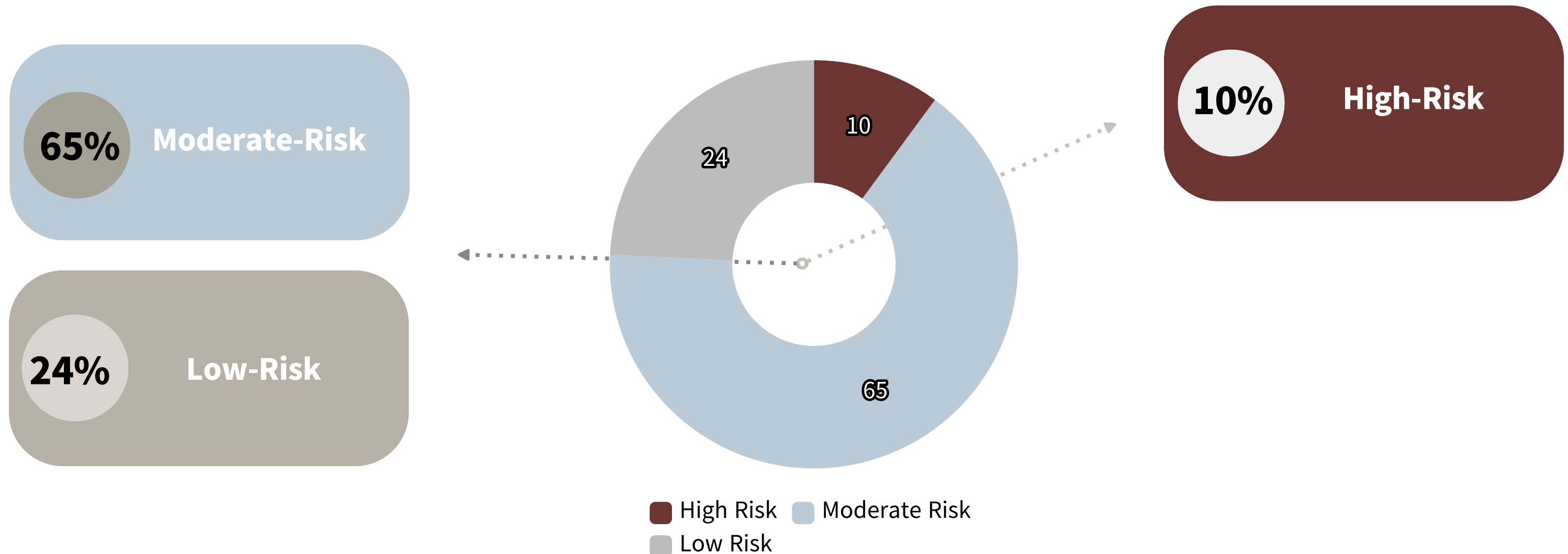
Risk 별 고객 분류

고객 Risk군 정의 : Income(수익) 대비 Revolving rate(채무) 값이 높으면 갚을 능력이 되지 않는다

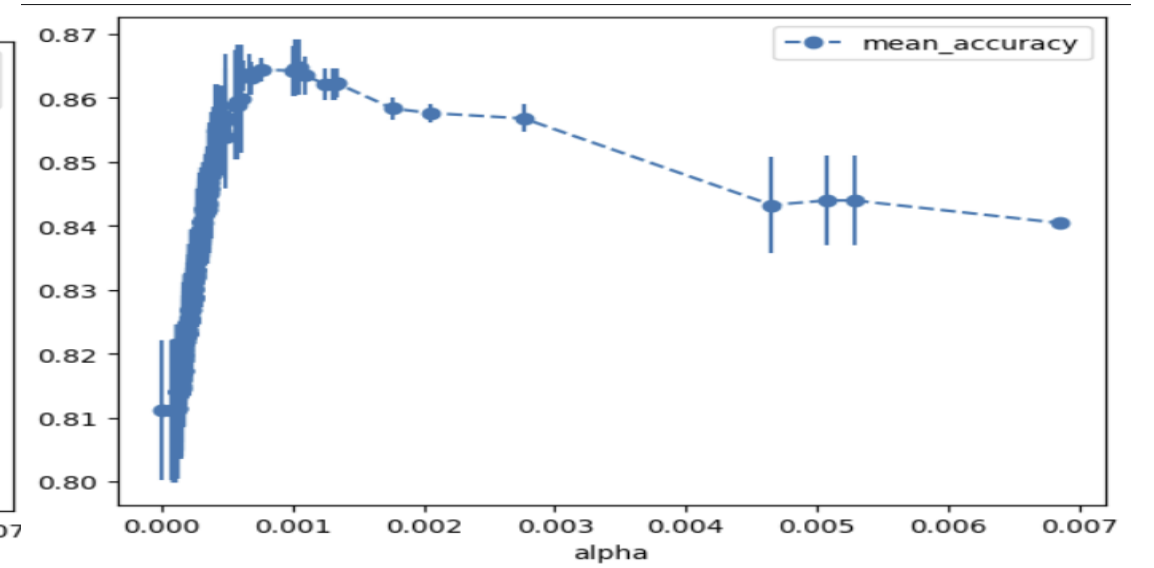
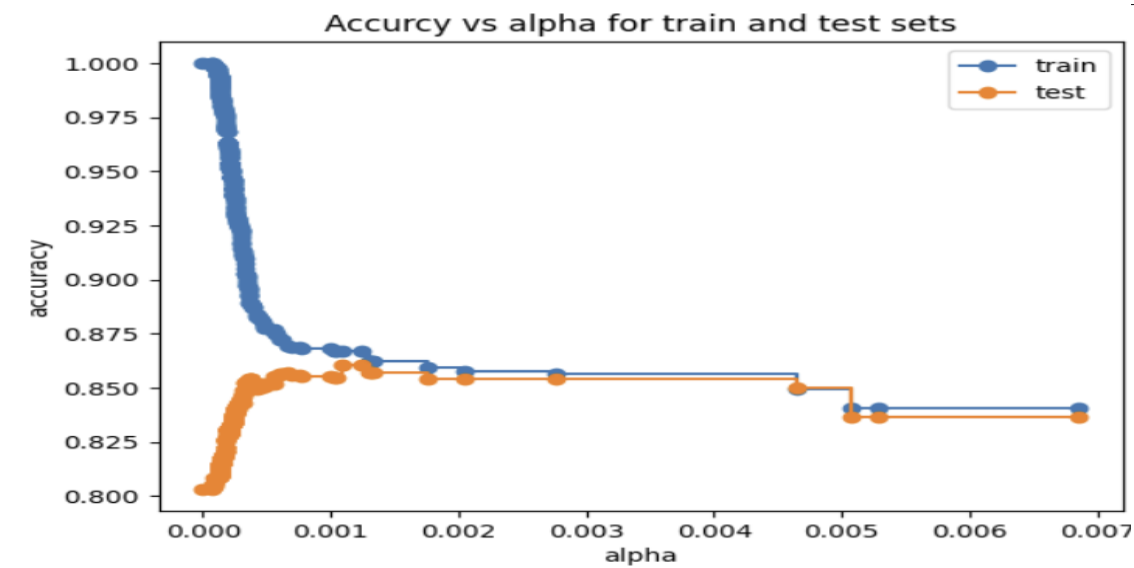
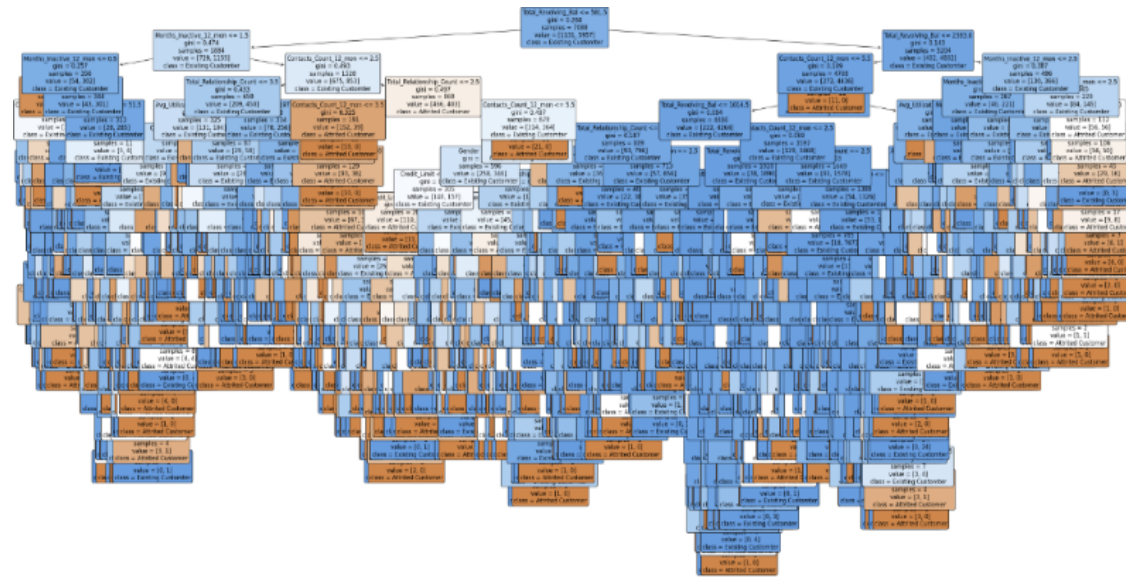
문제점 : 하지만 우리가 갖고 있는 data에 income은 숫자가 아닌 범위형 데이터로 되어있음

해결방안 : 일반적으로 신용한도는 income(수익)에 대비하여 credit_limit이 나오기 때문에 Income(수익) 대신 credit_limit(신용카드 한도)을 사용

신용 한도대비 채무가 70%넘는 고객들을 고위험군 채무가 0인 고객들을 저위험군으로 분류하기로 정의함



ML - Decision Tree

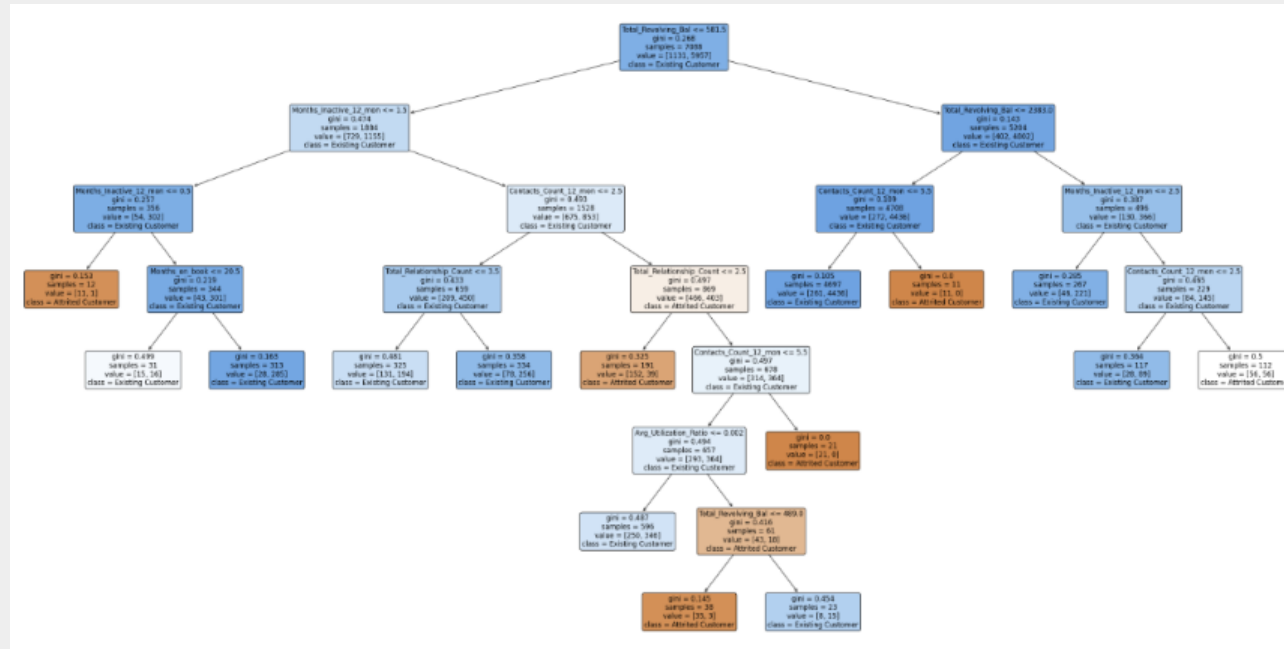


Machine Learning 분석 방법론

1. 모든 데이터를 이용 Decion Tree Machine Learning기법을 활용하여 모델링 실행
2. 결과를 이용하여 Decision Tree를 실제로 그려보고 과적합이 있다고 판단
3. Cost Complexity Pruning과 & Cross Validation기법을 이용하여 Decision Tree모델링에서 최적의 post-pruning(가지치기)를 하기위한 Alpha를 찾아냄
4. 찾아낸 최적화된 모델링을 이용하여 앞서 분류 했던 각 리스크별 Decision Tree를 그려 봄
5. Gini Index를 이용한 예측 모델 구축

Machine Learning

Low-Risk Clients



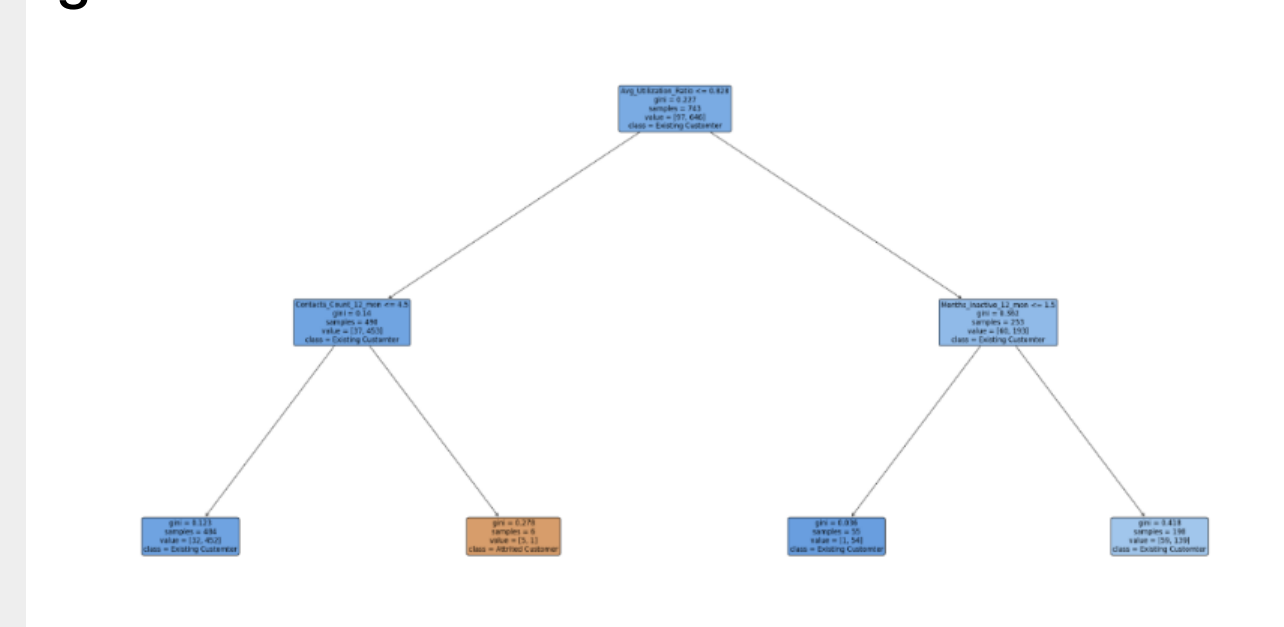
이탈 Low-Risk 고객군

- Low 리스크 고객들 중 이탈한 고객들은 1.5달 이상 카드를 사용하지 않으며 고객들이 신용카드 회사와 맺고 있는 상품의 수가 2.5개 이하인 경우가 많았음

이탈하지 않은 Low-Risk 고객군

- Low 리스크 고객들중 이탈하지 않은 고객들은 1.5 개월을 제외한 모든달에 카드 사용 기록이 있고, 2.5개 이상의 상품을 카드사와 맺고 있으며, 2.5번 이하로 연락하는 것을 선호했다

High-Risk Clients



이탈 High-Risk 고객군

- 신용 한도대비 사용한 비율이 0.828 이하이며 1년에 4.5 이상 연락을 하면 이탈하는 경향을 보임

이탈하지 않은 High-Risk 고객군

- 신용 한도대비 사용한 비율이 0.828 이상이며 1년에 1.5번 이하로 컨택을 하면 이탈하지 않는 경향을 보임

Marketing Strategy

Low-Risk 고객군

- Low Risk고객들은 고객들이 신용카드 회사와 맺고 있는 상품 또는 서비스 수가 높고 사용 빈도수가 잦을 수록 이탈하지 않는 패턴을 보임
- 매달 신용카드를 소비 할 수 있도록 하는 프로모션을 진행하고 고객들이 신용카드 회사와 맺고 있는 상품 또는 서비스를 더 이용 할 수 있는 프로모션을 제시한다면 Low-Risk 고객군의 이탈율을 줄일 수 있을 것 같음

High-Risk 고객군

- Decision Tree 마지막 node를 보니 4.5번 이하로 컨택 할 수 록 고객들이 이탈하지 않는 경향을 보임. 너무 잦은 연락은 이탈율을 높일 수 있으니 4.5번이하 즉 분기에 1번정도 연락을 하며 신용카드 한도 상향을 제한을 한다면 이탈율을 줄이며 고객들을 유지 할 수 있을 듯 함

배운점 및 발전 방향

- Python을 이용한 데이터 분석 인사이트
- Python을 이용한 머신러닝 수행 능력 향상
- 머신러닝 기반의 데이터 분석 및 예측 능력 습득
- 예측한 데이터를 바탕으로 결과 까지 실험 가능하다면 더욱더 배울게 많을 것 같음





