

고객 주문 데이터를 활용한 비즈니스 개선 방안

\*데이터 요약

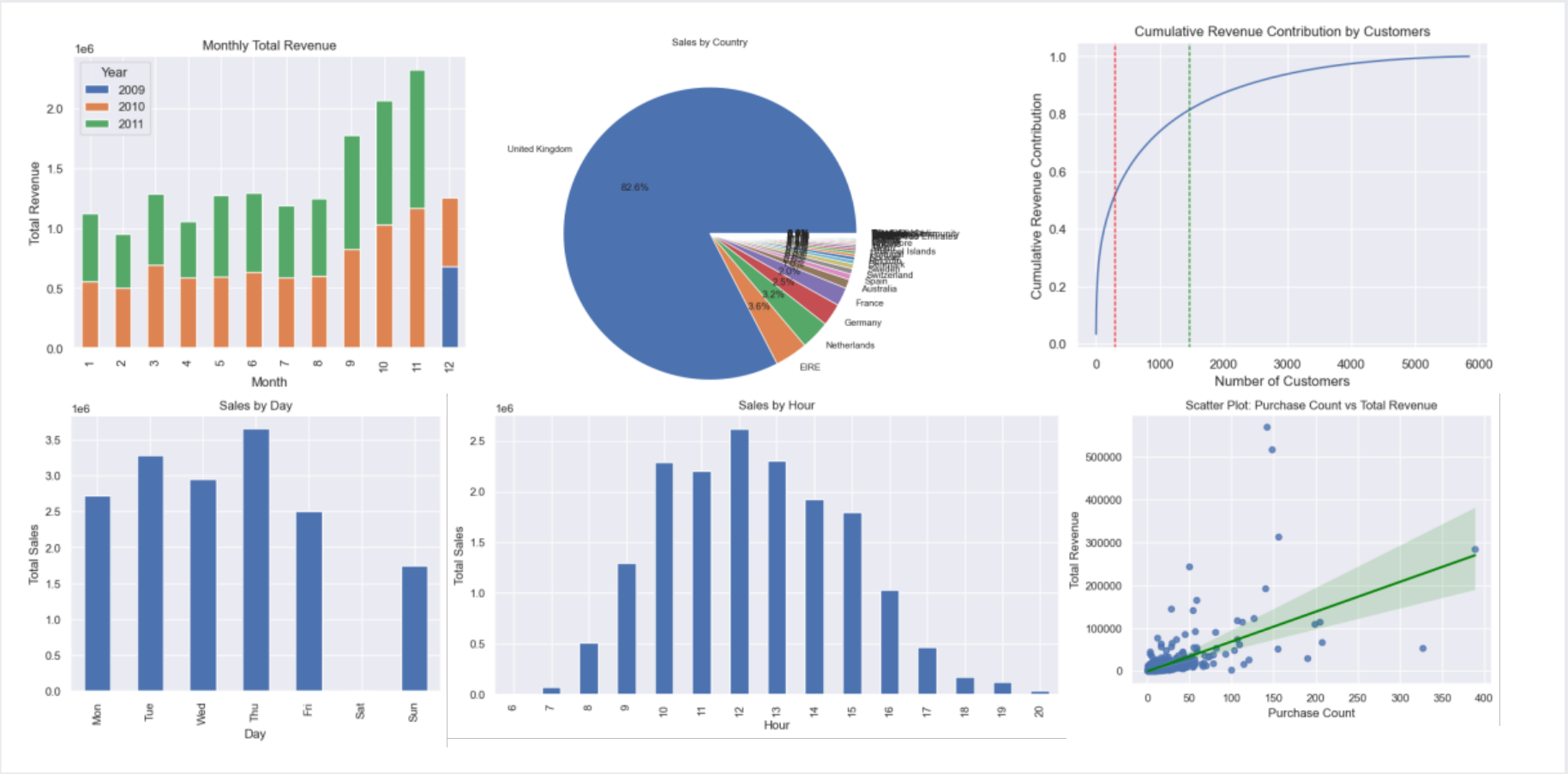
주문 데이터 : 주문번호, 상품코드, 상품명, 수량, 주문일자, 상품가격, 고객ID, 국가 (약 1백만건)  
데이터 개요 : 영국 소재 무점포 온라인 판매업. 도소매 모두 취급. 행사 및 선물용품 판매. 약 2년간 데이터

1. 데이터 전처리

[1] Null 데이터 및 중복 데이터 제거  
단, Null 데이터 중 다른 데이터를 참조해 채울 수 있는 경우는 보완

[2] 데이터 타입 변환

[3] 이원화 되어 있는 열을 계산하여 새로운 열 생성  
단위 가격 x 수량을 계산하여 자주 사용할 매출액 데이터 사전에 생성



2. [EDA] 데이터 탐색을 통한 인사이트 확인

- 1) 9, 10, 11월에 매출의 큰 상승을 확인. 매년 비슷한 패턴을 보이는 **주기성** 확인
- 2) 주문일자, 고객ID, 매출액 등 주요 변수를 기준으로 시각화를 수행하며 비즈니스 특성을 파악
- 3) 누적 매출 기여도를 시각화해 **상위 5% 고객이 약 50%의 매출을 차지**하고 있으며, 하위 75% 고객은 전체 매출의 20% 미만을 차지한 **다는 인사이트 도출**
- 4) 구매 횟수가 적지만 매출이 높은 일부 고객이 존재. 하지만 구매 횟수와 고객별 매출의 상관관계를 파악하고 시각화를 수행하여 0.65라는 양의 상관관계를 확인함

[문제 정의]

- 목표 : [1] 매출 대부분을 차지하는 핵심 고객의 관리 (이탈 방지, 재구매 유도)
- [2] 신규 고객에 대한 코호트 분석을 통해 초기 리텐션을 유지하여 지속적인 비즈니스 성장 도모

[분석 순서 및 개요]

1. 분석순서 : [사업 유형 정의] 전자상거래업은 연간 재구매율의 범위에 따라 사업 유형을 정의내릴 수 있으며, 사업의 초점을 달리 설정하여야 한다. (참고 도서) : 린 분석
- [RFM 분석] 사업 유형 정의에 따라 초점을 맞춰야하는 고객군이 달라진다. 이에 따라 고객군을 세분화하여 핵심 고객을 설정하고 지속적인 관리를 수행한다.
- [코호트 분석] 신규 고객에 대한 코호트 분석을 수행한다. (코호트 유형 : 획득 날짜, 코호트 크기 : 월 단위, 측정 항목 : 재구매율)

[사업 유형 정의]



1. 전자 상거래 - 고객 충성도 유형

- 린 분석에 따르면 다음과 같이 분류하고 있다.
- 1) 신규 고객 확보 유형 = 연간 재구매율(40%이하)
- 2) 하이브리드 유형 = 연간 재구매율(40~60%)
- 3) 고객 충성도 유형 = 연간 재구매율(60%이상)

본 프로젝트의 사업은  
연간 재구매율 계산 시, 61.83%로 고객 충성도 유형에 해당하였다.

2. 고객 충성도 유형 ?

- 충성도 높은 고객이 더 자주 구매하도록 만들어야 한다.
- 포인트 제도가 효과적
- 해당 유형은 10%로 대부분 성숙기에 도달
  - 전자상거래 진입은 얼마되지 않았으나, 1981년 부터 오래된 업력을 보유하고 있기 때문으로 추정

! 주의

- 신규 고객 확보를 등한시 해도 된다는 것은 아님
- 신규 고객을 확보하고 재구매를 유도하는 것은 비즈니스 성장에 필수적

3. 적용

- 1) 충성도가 높다!
  - 자주 구매한다.
  - 구매액이 크다.
- 2) 고객별 빈도수 편차 VS 매출액 편차
  - 행사 용품, 도매업자 포함이라는 비즈니스 특성, 빈도수는 적은 편
  - 빈도수 보다는 매출액이 고객 분류에 더 유용!

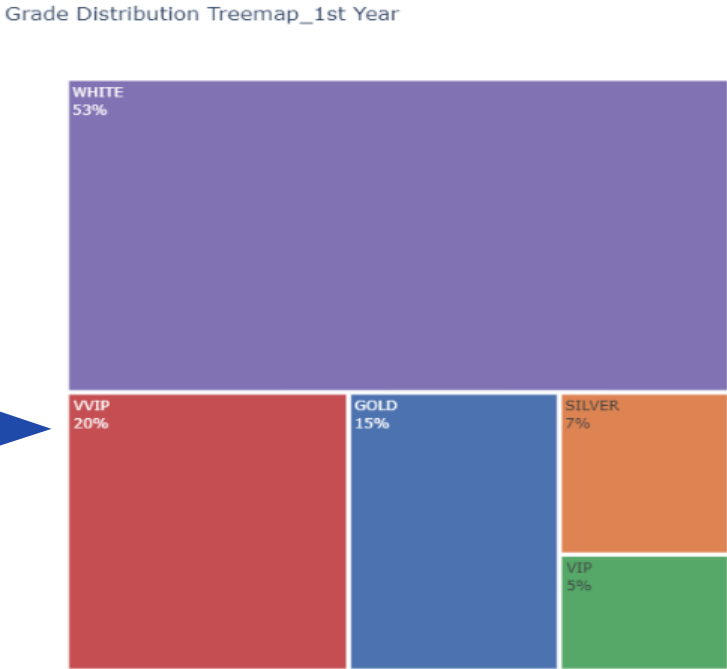
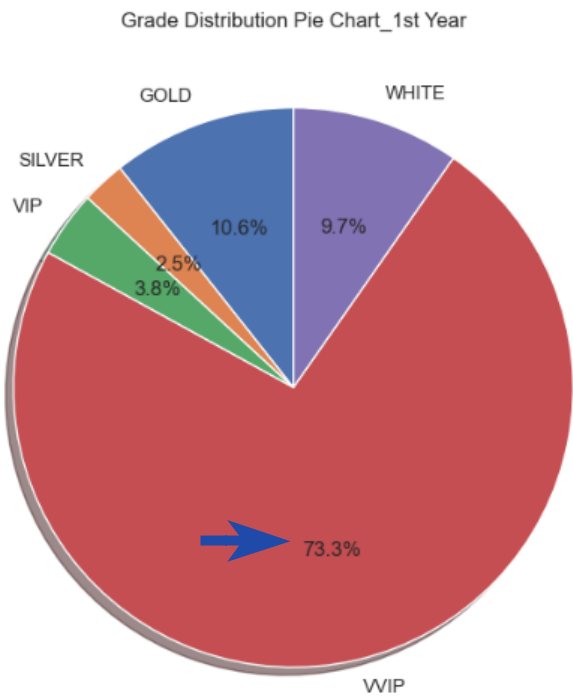
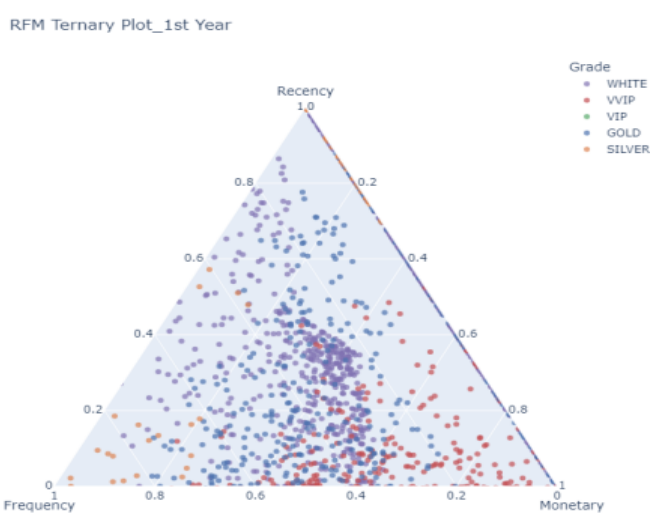
[RFM 분석]

RFM 분석이란?

- RFM은 **고객 관계 관리(CRM)**를 위한 데이터 드리븐 마케팅에서 활용되는 고객 세분화 모델
- 고객을 세분화 해 가치가 높은 고객을 발견하고 특성에 맞는 마케팅을 하여 비즈니스 개선을 도모함
- Recency(최근성), Frequency(빈도), Monetary(구매 크기) 세가지 지표로 고객을 분석

[등급 산정 기준 및 시각화] - 년 단위로 등급을 부여하는 것을 가정

	Recency	Frequency	Monetary
VVIP	-	-	4
VIP	R + M >= 9		3
GOLD	R + M < 9		3
SILVER	up 50%		< 3
WHITE	down 50%		< 3



[1] RFM 항목별 점수 부여

Recency : 년 단위 데이터 마지막 일자를 기준으로 고객별 가장 최근 구매일을 계산하고 5분위로 구분해 점수 부여

Frequency : 년 단위 구매 빈도를 측정한 뒤 5분위로 구분해 점수 부여

Monetary : 년 단위 총 매출을 측정해 5분위로 구분해 점수 부여

[2] 고객 세분화 RFM Score의 중요도 순

- Monetary ➡ Frequency = Recency
- VVIP의 매출이 약 3/4를 차지하므로 VVIP 이외 등급은 R, F를 동등하게 가중

[3] 각 등급별 특징

- VVIP : 매우 큰 매출을 일으키는 고객
- VIP : 많은 매출을 일으키면서 비교적 최근, 자주 방문한 고객
- GOLD : 많은 매출을 일으키지만 상대적으로 재구매가 뜸한 고객
- SILVER : 매출은 적지만 최근, 자주 방문한 고객
- WHITE : 매출, 최근성, 빈도 모두 낮은 고객

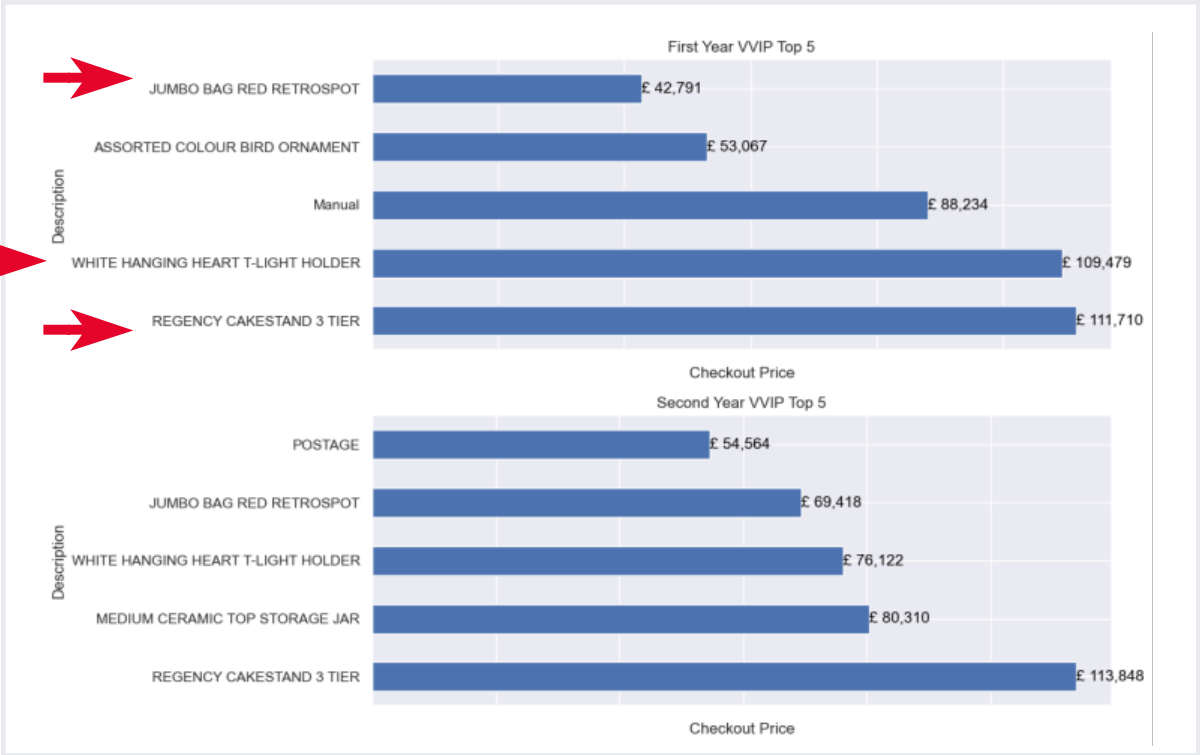
[4] 시각화 및 모니터링

- 1) 등급별 매출 비중 파이 차트 : 각 등급별 매출 비중을 확인 가능
- 2) 등급별 Tree Map : 각 등급의 인원 비율과 이에 대한 면적을 확인 가능
- 3) 고객별 Ternary Plot : 각 등급의 고객들이 어디에 위치해 있는지 확인 가능

[VVIP 대상 비즈니스 개선 방안]

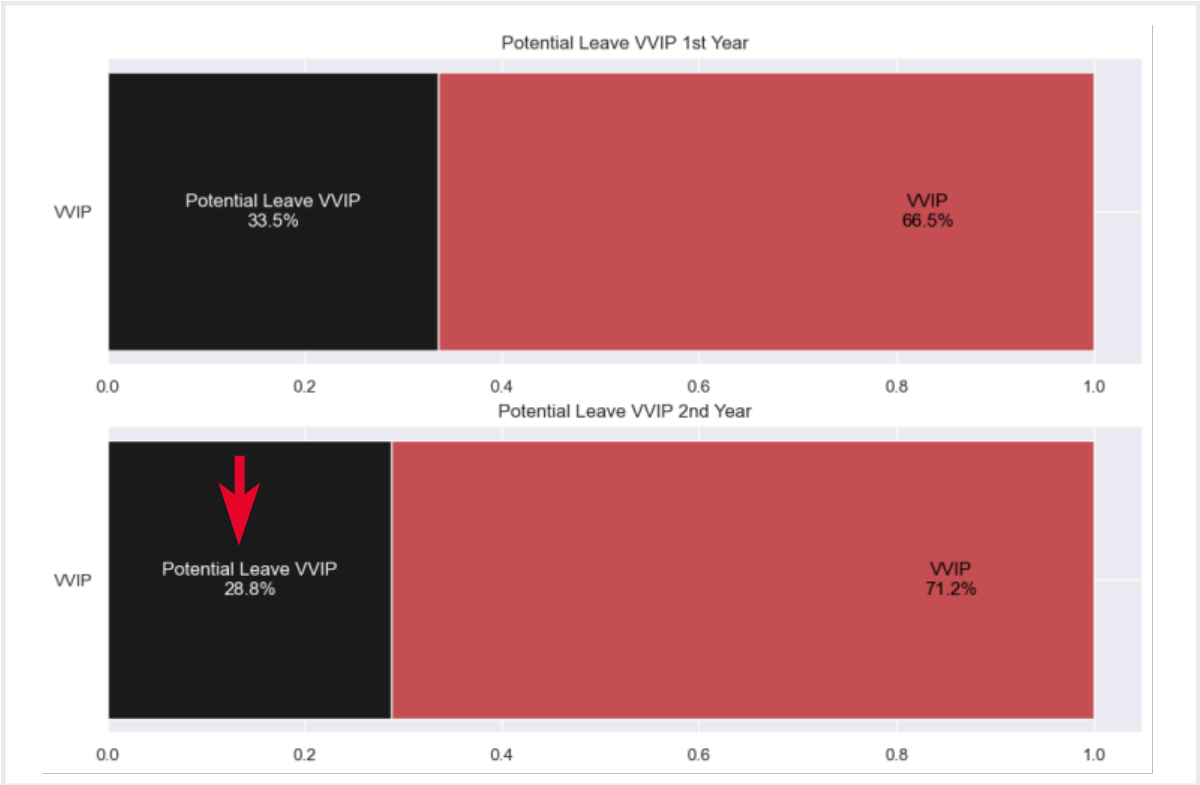
1. 등급별 차등 할인 및 포인트 지급 혜택 제공

- 등급에 따라 혜택을 제공시 수익도 고려해야 함
- 상품별 포인트 사용 한도 등을 적절히 두면 남은 포인트로 추가 상품 구매를 촉진할 수 있을것
- 잔여 포인트로 리마인드 알림을 통해 재구매를 유도해 볼 수 있음



2. 핵심 아이템 모니터링

- 각 년도별 VVIP 매출 TOP5 상품 비교
- 3가지 품목이 중복됨. 'Regency ~ 3 Ties', 'White Hanging ~ Holder', 'Jumbo ~ Retrosport'
- 도메인 지식에 따라 이들 상품을 미끼 상품, 핵심 경쟁 상품 등으로 사용할 수 있을 것
- 구매 빈도, 매출 모두 높은 상품들의 가격 경쟁력이 상실될 경우 VVIP 고객이 이탈할 수 있으므로 주의할 것



3. 잠재 이탈 가능 VVIP 고객 최소화

- 매출의 상당부분을 견인하는 VVIP 고객의 이탈은 큰 손해를 의미함

1) 이탈 우려 고객 분류 : VVIP 등급의 고객 중 Recency Score가 2 이하인 고객을 이탈 우려 고객으로 분류  
이를 통해 최근 구매하지 않은 고객을 파악해 모니터링 할 수 있다.

2) VVIP전체 고객 중 **이탈 우려 고객의 비율**을 **핵심 성과 지표(KPI)**로 정의하고 조직 달성 목표로 관리

- 첫 해의 비율은 33.5%였으나, 둘째 해의 비율은 28.8%로 감소

4. 알림을 통한 재구매 유도 및 로직 생성 - VVIP뿐 아니라 모든 고객에 적용 가능

- 지나친 알림 및 이메일 발송은 고객의 불쾌한 유저 경험을 생성할 수 있으므로 발송 주기를 주의해야함
- 고객별로 평균 재구매 기간 및 가장 많이 구매한 시간대를 계산하여 **개인화 알림**을 적용

로직 :

1) 최근 구매 이후, 개인별 평균 재구매 기간이 지났음에도 재구매가 발생하지 않을 경우 알림 발송

2) 재구매 기간 7일 미만은 7일로 고려(지나친 알림 방지)

+ ) 발송 시간은 = 개인별 가장 많이 구매한 시간

3) 일정 기간(ex.3개월)이상 경과할 경우 쿠폰 등 프로모션을 통한 재구매 유도



[코호트 분석]

코호트 분석?

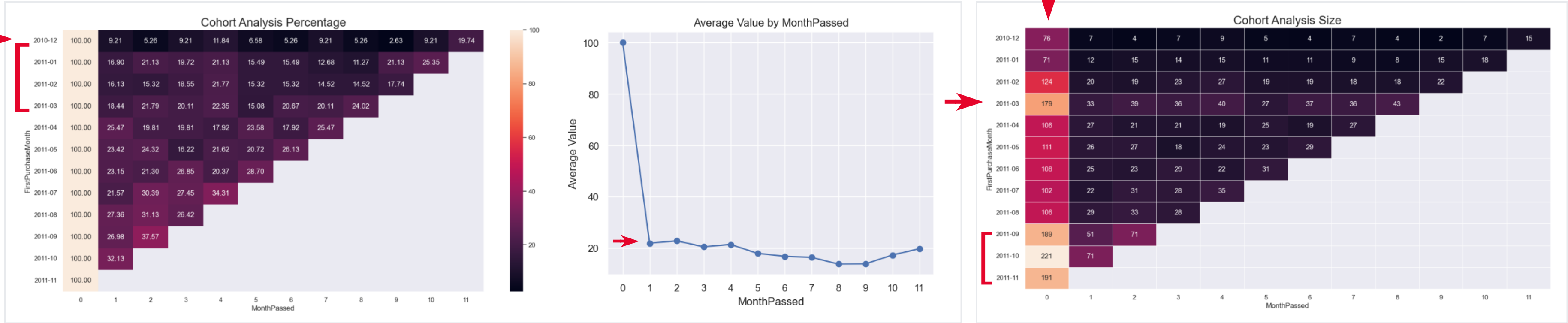
- 시간을 두고 비슷한 그룹을 비교하는 것
- 시간을 두고 새로운 서비스 출시, 테스트, 사업 모델 변경 등 여러 변화를 거치며 이것이 고객의 특정 행동(전환, 이탈, 매출 등)에 어떤 영향을 미쳤는지 알아내는 데 이용

1. 분석 대상 및 가정

[1] 신규 고객의 코호트별 재구매 유지율(Retention)관리 - 첫 구매 월을 기준으로 코호트 구분

- VVIP 고객 대상 비즈니스 개선 방안과 마찬가지로, 신규 고객 대상 재구매 유지율도 지속적인 비즈니스 성장을 위해서 관리해야 함

[2] 가정 : 2년간 데이터 중 첫 해에는 구매하지 않았고, 둘째 해에는 구매한 고객을 신규 고객으로 가정



2. 기간에 따른 유지율(%)

[1] 유지율 추이 - 첫 달 평균 재구매 유지율 : 21.88%

- 많은 경우에 초기에 유지율이 급격히 떨어지고 안정화 되는 구간을 가짐
- 첫 달에만 급격히 떨어지고 둘째 달 부터 바로 안정화되는 모습이 관찰됨

[2] 다른 코호트 간 비교

- 다만 초기 2010-12월 부터 2011-03월(4개월)까지의 신규 고객이 평균보다 더 급격히 하락하는 모습을 보임
- 특히, 2010-12월의 신규 고객에서 첫 달 경과 9.21%라는 매우 낮은 유지율을 발견함

3. 기간에 따른 규모

[1] 첫 달의 신규 고객 규모 관찰

- 2011-09 부터 2011-11 (3개월) 간 신규 고객 규모가 다른 기간에 비해 큼
  - ➡ 이는 EDA에서 관찰한 4분기에 매출이 늘어나는 주기성 덕분에 추정됨
- 2011-03의 규모가 갑작스레 커졌는데, 이 원인을 파악해 향후 신규 고객 유입에 참고

[코호트 분석]

4. 비즈니스 개선 방안

[1] 초기 유지율이 떨어지는 속도 늦추기

- 초기에 떨어지는 속도는 구매하는 단계의 영향을 많이 받을 것
- 2010-12 ~ 2011-03 (4개월) 중 특히 2010-12 코호트 그룹의 사용자 경험에 악영향을 미쳤을 요소를 분석해 개선
- 2011-03의 신규 고객 규모가 커진 내외적 요인을 검토하고 2011-04에는 해당 규모가 유지되지 못한 원인을 개선
  - 예상 내적 요인 : 일시적 이벤트 및 프로모션 ➡ 일시적 증가 요인이므로 '비용-수익' 평가 후 마케팅 지속여부 결정  
신규 제품 출시 혹은 서비스 개선 ➡ 해당 제품 및 서비스에 대한 나쁜 평가 및 홍보 부족 등 다각도로 검토
  - 예상 외적 요인 : 국가 기념일, 사회 지원금 ➡ 일시적인 증가 요인이나 주기성이 있을 경우 수요 예측에 활용

[2] 안정화 단계를 오래 유지하기

- 안정화 단계를 오래 유지하는 것은 사용자와의 관계를 잘 유지하는 것이 중요
- 장기적인 커뮤니케이션 플랜, CRM 마케팅 활동, 주기적인 프로모션, 장기 미사용 고객에 대한 복귀 리마케팅 등의 진행을 고려

[프로젝트 한계점 및 개선 아이디어]

[1] 고객당 수익을 고려할 수 있다면, 고객 세분화가 더욱 유용할 것

- 현재 : 고객당, 상품당 매출은 확인할 수 있지만, 해당 판매에서 할인율-포인트 적용, 유입 비용 등을 고려한 수익성에 대해 고려가 불가능
- 개선 방안 : 주문 내역 데이터에 상품당 마진, 할인율-포인트 적용여부 및 정보 등을 포함한 데이터셋을 관련 팀과 논의

[2] 유입-재방문 경로, 상품 검색 등을 고려할 수 있다면, CRM 마케팅 및 서비스 개선에 더욱 도움

- 현재 : 마케팅 캠페인을 통한 신규 유입, 프로모션을 통한 재구매 고객, 구매 전 검색 기록 등 디테일한 정보가 부족하여 마케팅 및 서비스의 평가와 개선 여부를 판단하기 어려움
- 개선 방안 : '비용-수익'을 고려하여 자체 Tool을 개발하거나 GA,GTM 또는 Amplitude와 같은 상용 애널리틱스 툴 도입을 고려