

# 고객 세분화 분석 보고서

\*RFM + Interpurchase Time 기반 정량적 군집화 (KMeans / Python)

## 1. 프로젝트 목적

### • 프로젝트 개요

이 분석은 영국 온라인 리테일사의 실제 거래 데이터를 바탕으로, 고객의 구매 행동을 기반으로 한 정량적 세분화를 수행하는 것을 목적으로 하고 있습니다. 기존의 RFM(Recency, Frequency, Monetary)분석에, Interpurchase Time(재구매 평균 주기)을 추가하여 고객의 구매 주기를 함께 고려해 보다 실질적인 타겟팅 전략 수립이 가능하도록 하였습니다.

### • 분석 방향 및 가정

- Interpurchase Time 을 결합하여, 고객의 구매 리듬을 반영한 정교한 세분화가 가능하다는 가정
- 군집 결과는 행동 특성에 기반하여 실질적인 마케팅 전략 수립에 활용 가능하다는 가정

## 2. 데이터 구성 및 전처리

- 데이터 출처: UCI Machine Learning Repository – Online Retail Dataset (약 8개월간, 총 541,909 건)
- 주요 변수: InvoiceNo, StockCode, Quantity, UnitPrice, InvoiceDate, CustomerID, Country 등
- 정제 과정:
  - 결측값 제거: CustomerID 가 누락된 행은 제거, Description 결측은 분석에 활용되지 않으므로 삭제
  - 이상치 제거: IQR 기반으로 Quantity 및 UnitPrice 의 이상치 제거, 클러스터링 안정성 확보

## 3. 분석 설계 및 모델링

| RFM + Interpurchase Time 확장 설계

- Recency: 기준일로부터 마지막 구매일까지의 경과 일수
- Frequency: 고객별 고유 거래 횟수
- Monetary: 누적 구매 금액
- Interpurchase Time:  $[(T_n - T_1) / (Frequency - 1)]$

→ **Interpurchase Time** 은 동일한 빈도를 가진 고객이라도 구매 간격에 따라 행동이 달라질 수 있기에 도입하였습니다. 이를 통해 고객의 반복 주기를 반영해 더 세밀한 세분화가 가능합니다.

| KMeans 클러스터링 적용

- 군집 수 결정: Elbow Method 로 왜곡 감소율이 완만해지는 지점, k=6 채택
- 군집 품질 평가: Silhouette Score = 0.6279 → 0.6 이상이면 클러스터 구분이 우수한 수준으로 간주됨

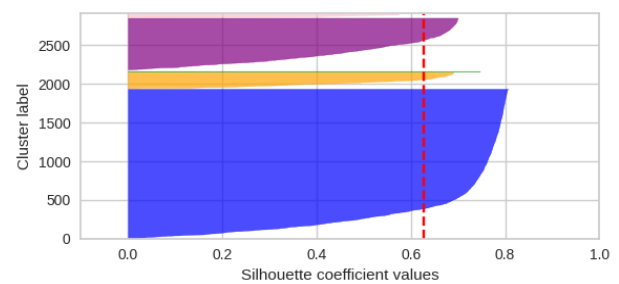
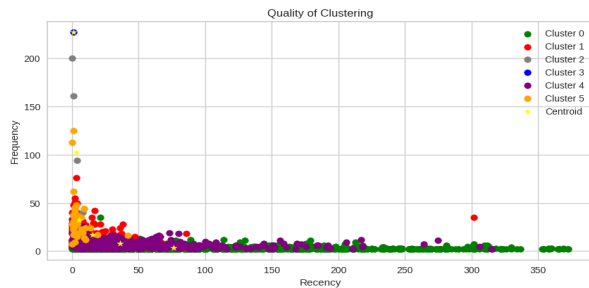
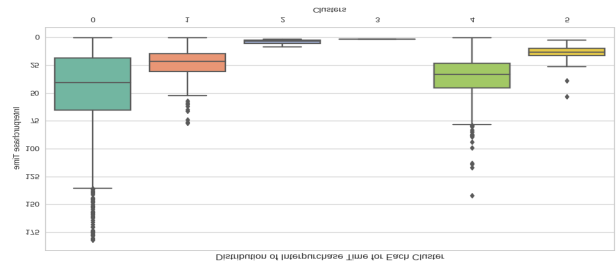
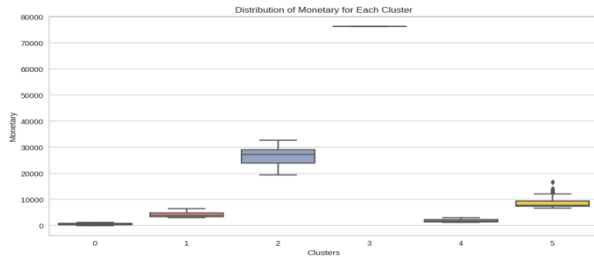
## 4. 분석 결과 요약

클러스터	주요 행동 특성	인사이트
Cluster 3	매우 높은 빈도 및 금액, 짧은 재구매 간격	프리미엄 타겟, 장기 유지 전략 적합
Cluster 2	높은 반복성 및 구매력	멤버십, 추천 프로그램 강화 대상
Cluster 1	중간 수준의 활동성	리텐션 관점의 맞춤 유지 전략 고려
Cluster 0, 4	낮은 빈도와 금액	재참여 유도형 리마케팅 우선 적용

→ 각 클러스터의 특성은 RFM + Interpurchase Time 평균값을 기반으로 해석되며, 행동 특성을 기반으로 한 실질적인 전략 수립이 가능합니다.

## 5. 시각화

- Boxplot (Frequency / Cluster): 군집별 구매 빈도의 분포 범위와 밀도를 비교하여 고활동 고객 식별
- Scatterplot (Recency vs Frequency): 활동성과 최근성의 상관구조를 통해 군집간 구조적 특성 시각화
- Silhouette Plot: 클러스터 품질(분리도와 응집도)을 시각적으로 확인



## 6. 적용 가능성 및 확장성

- 다양한 산업 및 공공 분야로의 확장:  
이 분석 프레임은 유통·커머스뿐 아니라, 공공 리텐션 전략(예: 복지 서비스 수혜자 분석, 기부자 유지 전략) 등에서도 활용할 수 있습니다.
- Interpurchase Time 의 장점:  
고객(또는 이용자)의 중장기 구매(이용) 패턴을 파악할 수 있어, 재구매(재방문) 유도 및 서비스 개선 전략을 구체적으로 설계하는 데 효과적입니다.

→ Interpurchase Time 을 반영하면, 각 고객이 다시 구매할 가능성이 높은 시점을 예측할 수 있어, 누구에게 언제 리마케팅을 해야 할지 우선순위를 정할 수 있습니다.

\*\*분석에 사용된 상세 코드와 처리 로직은 별도 요청 시 전달드릴 수 있습니다.