

고객 Segmentation의 정의/기법

[정의]

- 다양한 기준으로 고객을 분류하는 기법
- CRM이나 마케팅의 중요 기반 요소
- 보통 어떤 상품을 얼마나 많은 비용을 써서 얼마나 자주 사용하는가에 기반한 정보로 분류함
- 주요 목표: **타겟 마케팅**
 - 고객을 여러 특성에 맞게 세분화해서 그 유형에 따라 맞춤형 마케팅이나 서비스를 제공하는 것
 - 고객의 상품 구매 이력에서 시작함

[기법]

- RFM 기법
 - Recency(R): 가장 최근 상품 구입일에서 오늘까지의 시간
 - Frequency(F): 상품 구매 횟수
 - Monetary Value(M): 총 구매 금액

※ agg() 연산

- agg()에 인자로 대상 칼럼들과 aggregation 함수명들을 딕셔너리 형태로 입력하면 칼럼 여러 개의 서로 다른 aggregation 연산을 쉽게 수행할 수 있음

[코드 예시]

```
aggregations = {  
    'InvoiceDate': 'max',  
    'InvoiceNo': 'count',  
    'sale_amount': 'sum'  
}  
  
cust_df = retail_df.groupby('CustomerID').agg(aggregations)
```

왜곡된 데이터 분포도에서 군집화 수행

- 왜곡 정도가 매우 높은 데이터 세트에 K-Means 군집을 적용하면 중심의 개수를 증가시키더라도 변별력이 떨어지는 군집화가 수행됨
=> StandardScaler로 평균/표준편차를 재조정하여 왜곡 정도 완화 후 군집화 수행
- 데이터 값이 거리 기반으로 광범위하게 퍼져 있는 경우 군집 수를 계속 늘려봐야 해당 군집만 지속적으로 분리하게 되기에 의미 없는 군집화 결과로 이어진다.
- 개선: 전체 데이터에 **로그 변환** 적용