

# PORT FOLIO

# PROJECT 2

## RFM 분석 및 클러스터링 기반 고객 세분화를 통한 맞춤 상품 제안 및 세그먼트 활용 전략

### 프로젝트 개요

#### 기획 의도

기존의 일반적인 상품 추천보다 고객의 특성과 선호도를 고려한 개인화된 상품 추천 서비스를 제공하여,  
고객 만족도를 높이는 것이 목표

#### 문제 정의

RFM 분석과 K-means 클러스터링을 통해 고도화된 고객 세분화를 통해, 보다 고객특성에 맞는 맞춤 상품 제안

#### 분석 목표

고객 세분화 및 상품 추천 결과를 기반으로 한 효과적인 전략을 제시하여 ,매출 증대 및 이탈 방지

#### Language

Python

#### Tool

Pandas

## **PROJECT 02.** RFM 분석 및 클러스터링 기반 고객 세분화를 통한 맞춤 상품 제안 및 세그먼트 활용 전략

### 분석 프로세스

#### 데이터 탐색

- EDA 데이터 시각화

#### 데이터 전처리

- 결측치 처리 ⇒ 이상치 처리 ⇒ 변수 처리

#### RFM Analysis

- RFM 계산 ⇒ RFM 모형 ⇒ RFM segment

#### Clustering Analysis

- Cluster segment( K-Means clustering)  
⇒ 평가 (Elbow Method, Silhouette\_score)  
⇒ Cluster 특성 분석

#### Basket Analysis

- Basket Analysis

#### RecommendSystem

- RecommendSystem

# 데이터 준비

## 데이터 소개

### E-Commerce Data

Actual transactions from UK retailer



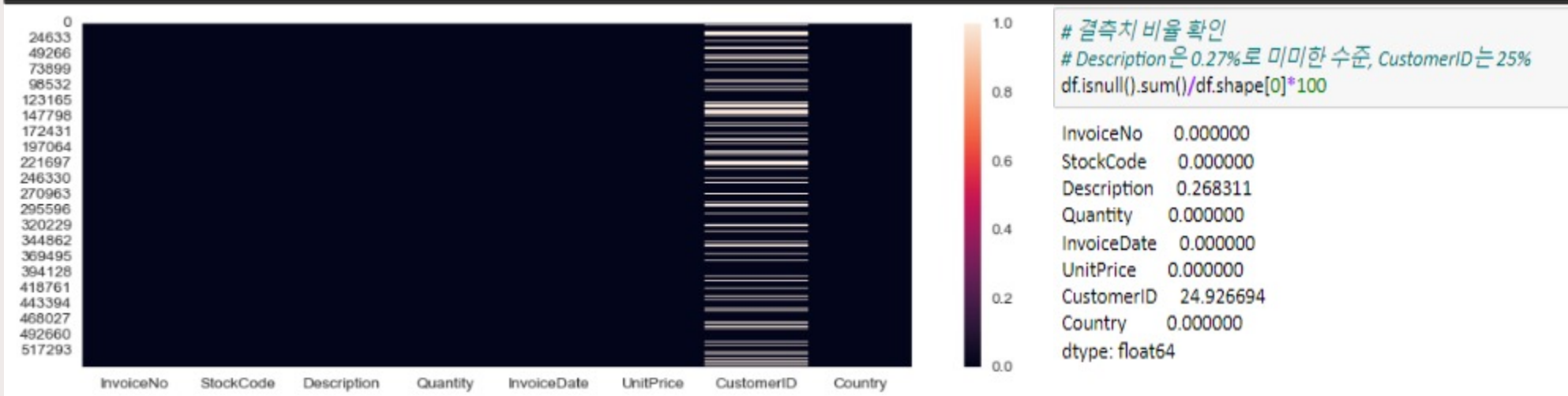
데이터 출처: Kaggle

541909 rows X 8columns

변수명	변수 설명	변수 유형	변수명	변수 설명	변수 유형
CustomerID	고객 고유아이디	Index	PreferredOrderCat	지난달 고객이 주문한 가장 선호하는 카테고리	Categorical
Churn	이탈 여부	Categorical	SatisfactionScore	만족도 점수	Categorical
Tenure	고객 사용 기간	Numerical	MaritalStatus	결혼 여부	Categorical
PreferredLoginDevice	선호하는 로그인 디바이스	Categorical	NumberOfAddress	등록한 주소 갯수	Numerical

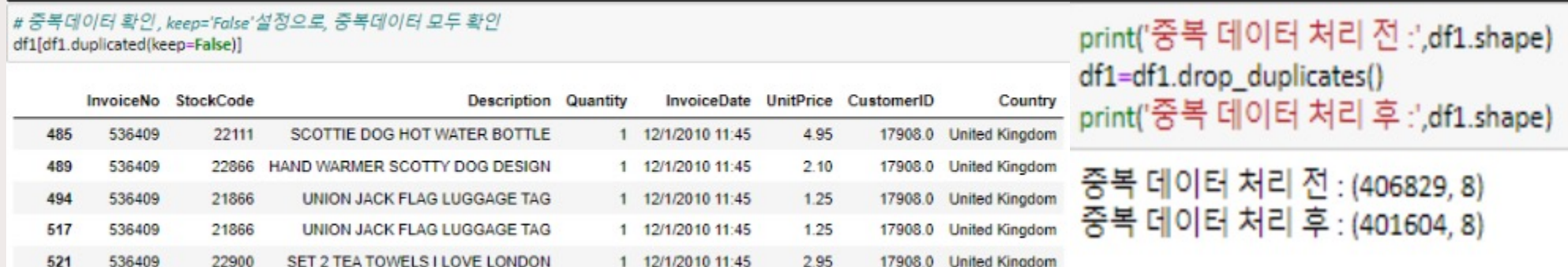
## 데이터 전처리

### 결측치 처리



CustomerID의 결측치는 약 25%, Description은 0.27%로, 분석에 편향이 발생할 수 있으므로, ⇒ 삭제

### 중복값 처리



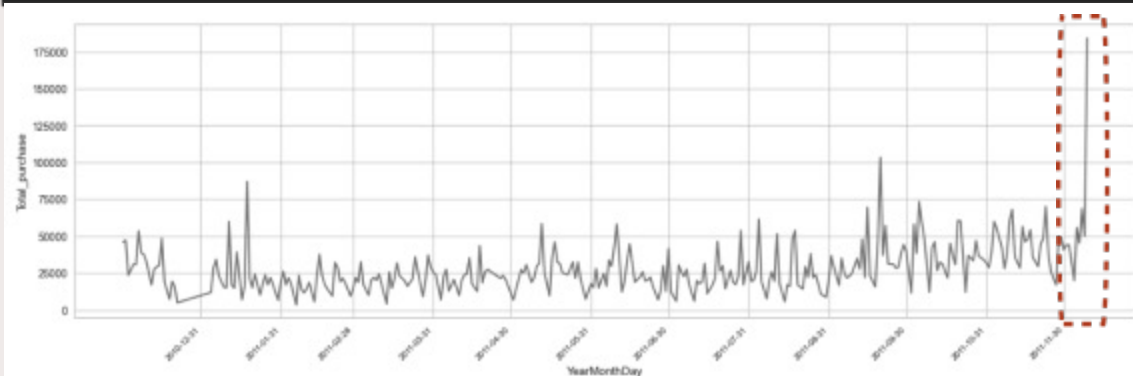
### 변수 처리

변수명	변수 설명	변수 유형	변수명	변수 설명	변수 유형
Year	주문 년도	Numerical	YearMonthDay	주문년도-월-일	Categorical
Month	주문 월	Numerical	YearMonth	주문년도-월	Categorical
Day	주문 일	Numerical	Hour	시간	Numerical
Day_of_week	주문 요일	Numerical	Total_purchase	제품단가 X 가격 ⇒ 총 구매금액	Numerical

## EDA

## 판매 트렌드

년/월/일 별 판매 금액 트렌드

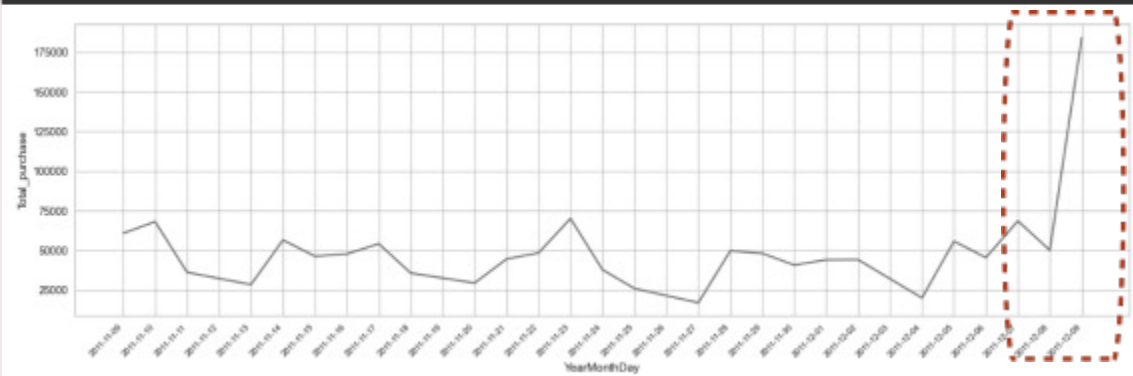


2010-12~2011-08까지는 안정적인 판매추세를 보이나,  
2011-09~10월에 판매량이 급증

'2011-12-09일' 최대 매출 184170\$ 달성

⇒ 평균 매출 대비 543% 상승 시기

최대 매출 발생 전 30일부터~ 최대 매출 발생일까지의 판매 금액 트렌드



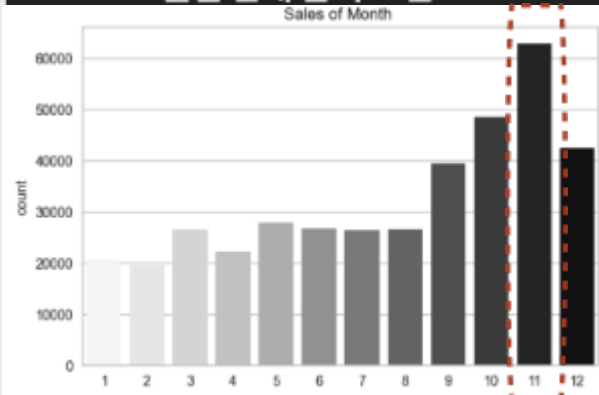
최대 매출 발생일 30일 전~ 최대 매출 발생일 까지의  
판매 추세 확인 시, 판매 증가가 두드러지게 증가

⇒ 해당 일자에 이벤트 시행했음을 추측할 수 있다.

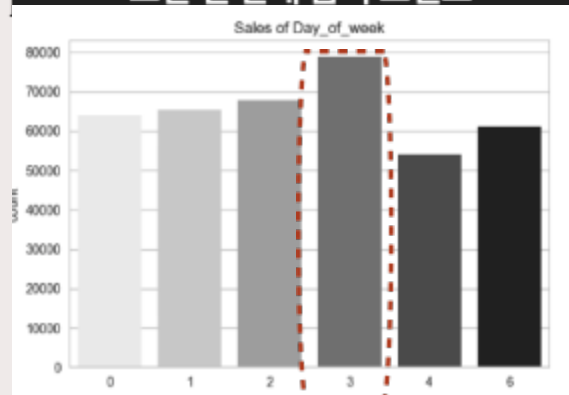
## EDA

## 판매 트렌드

월별 판매 금액 트렌드



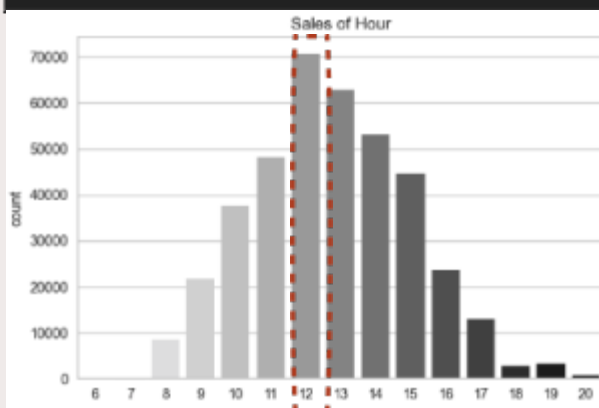
요일 별 판매 금액 트렌드



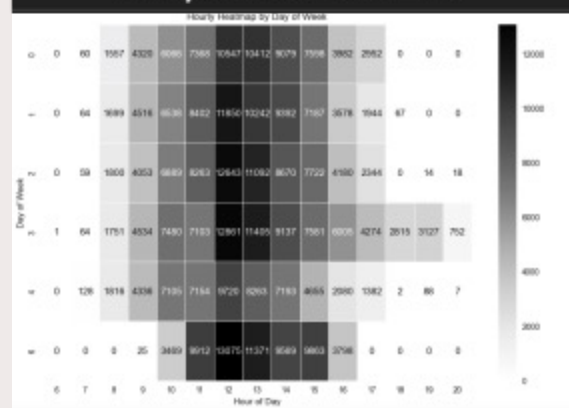
9월 기점으로 판매 금액 증가 추세

⇒ 11월 최대 판매 금액 달성

시간 별 판매 금액 트렌드



요일,시간별 판매 히트맵

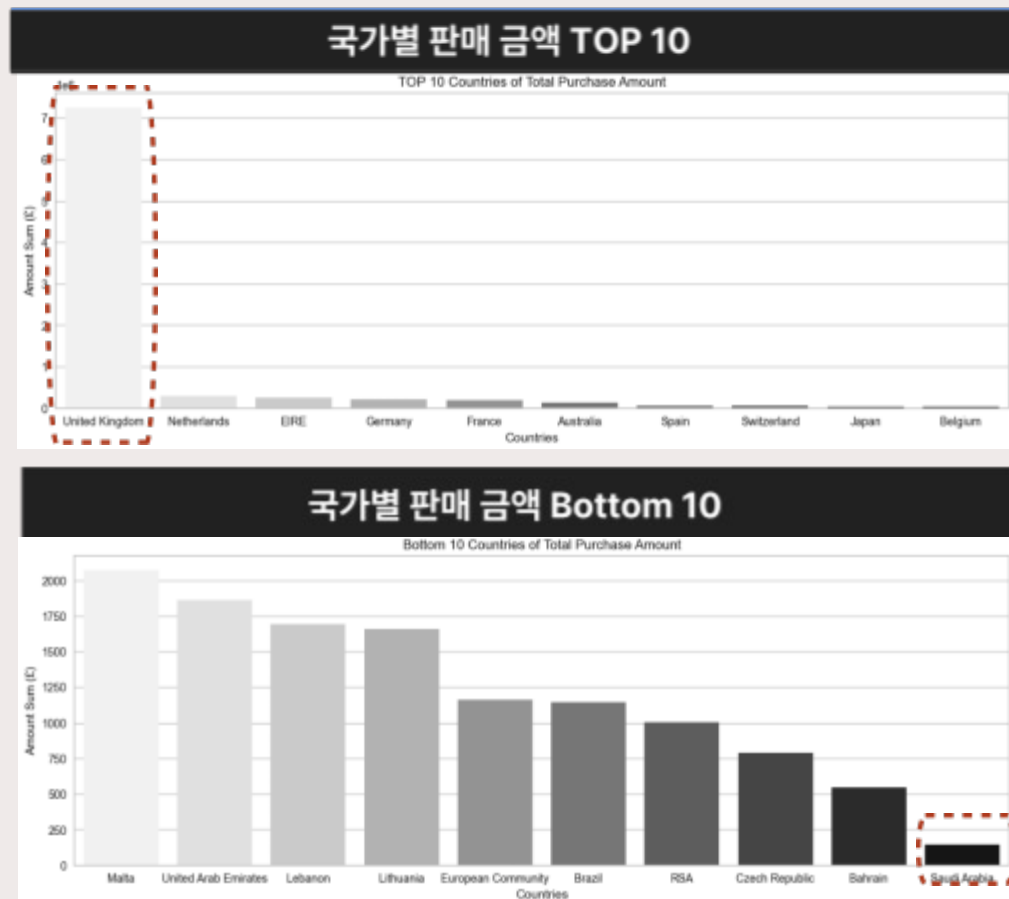


목요일 낮 12시에 최대 매출이 발생하며,  
주로 평일 화~목요일/ 낮 12시~오후 14시의  
판매량이 가장 우수하다.



## EDA

## 국가

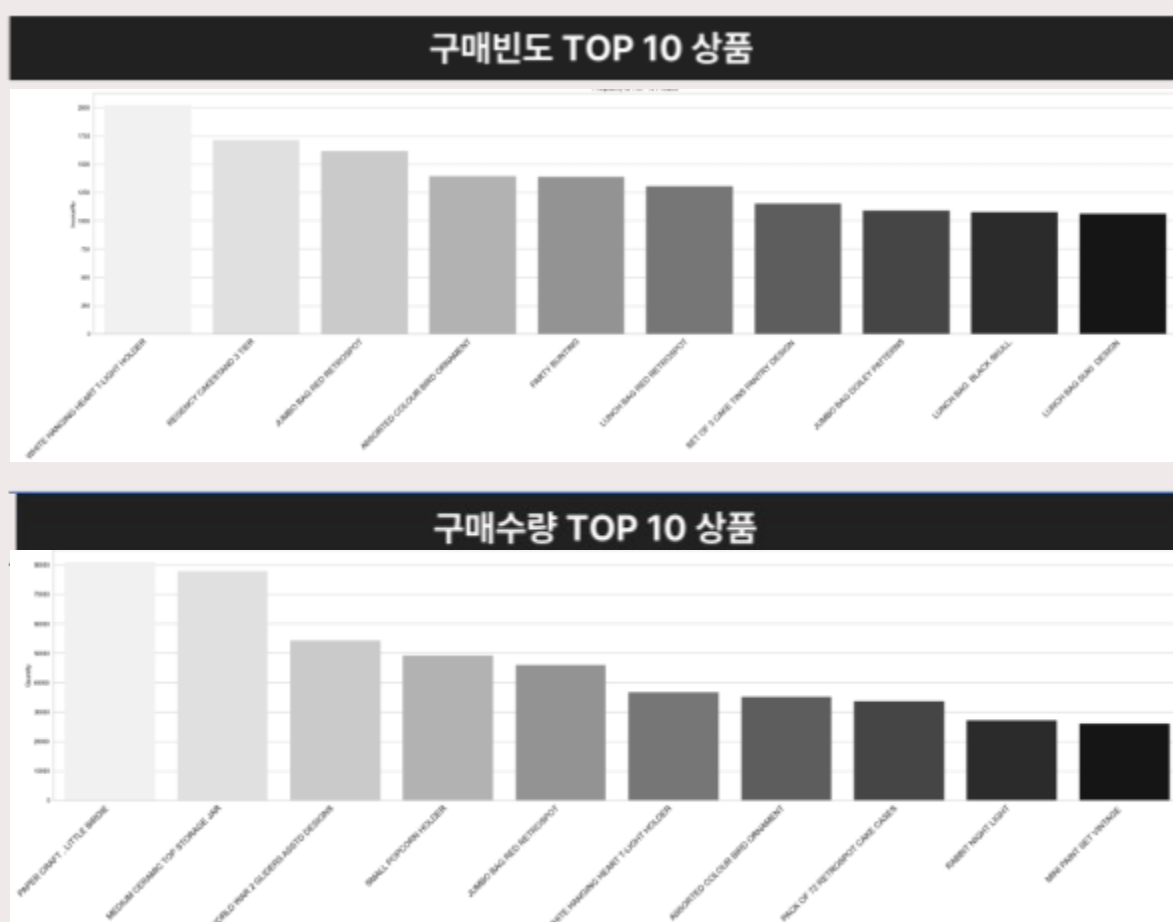


'United Kingdom'(영국)의 판매 점유율은 약 89%로, 과반수 이상의 매출을 점유하고 있다.

'Saudi Arabia' (사우디 아라비아)의 판매 금액이 가장 낮으며, 약 0.023%의 점유율

## EDA

## 상품



가장 자주 구매하는 상품은

'WHITE HANGING HEART T-LIGHT HOLDER'  
( 하얀색 하트 모양의 촛불 홀더)

가장 많은 수량 구매된 상품은

'PAPER CRAFT LITTLE BIRDIE'  
(종이로 만든 새모형)

⇒ 공통적으로 조명,장식품 등 상품이 상위 10위안에 랭크되어 있음을 알 수 있다.

# RFM 분석

## RFM 모형

RFM 계산			
CustomerID	Recency	Frequency	Monetary
12346.0	326	1	77183.60
12347.0	2	182	4310.00
12348.0	75	27	1437.24
12349.0	19	72	1457.55
12350.0	310	16	294.40

RFM 모형							
CustomerID	Recency	Frequency	Monetary	Recency_score	Frequency_score	Monetary_score	RFM_score
12346.0	326	1	77183.60	1	1	5	115
12347.0	2	182	4310.00	5	5	5	555
12348.0	75	27	1437.24	2	2	4	224
12349.0	19	72	1457.55	4	4	4	444
12350.0	310	16	294.40	1	2	2	122

### RFM 계산

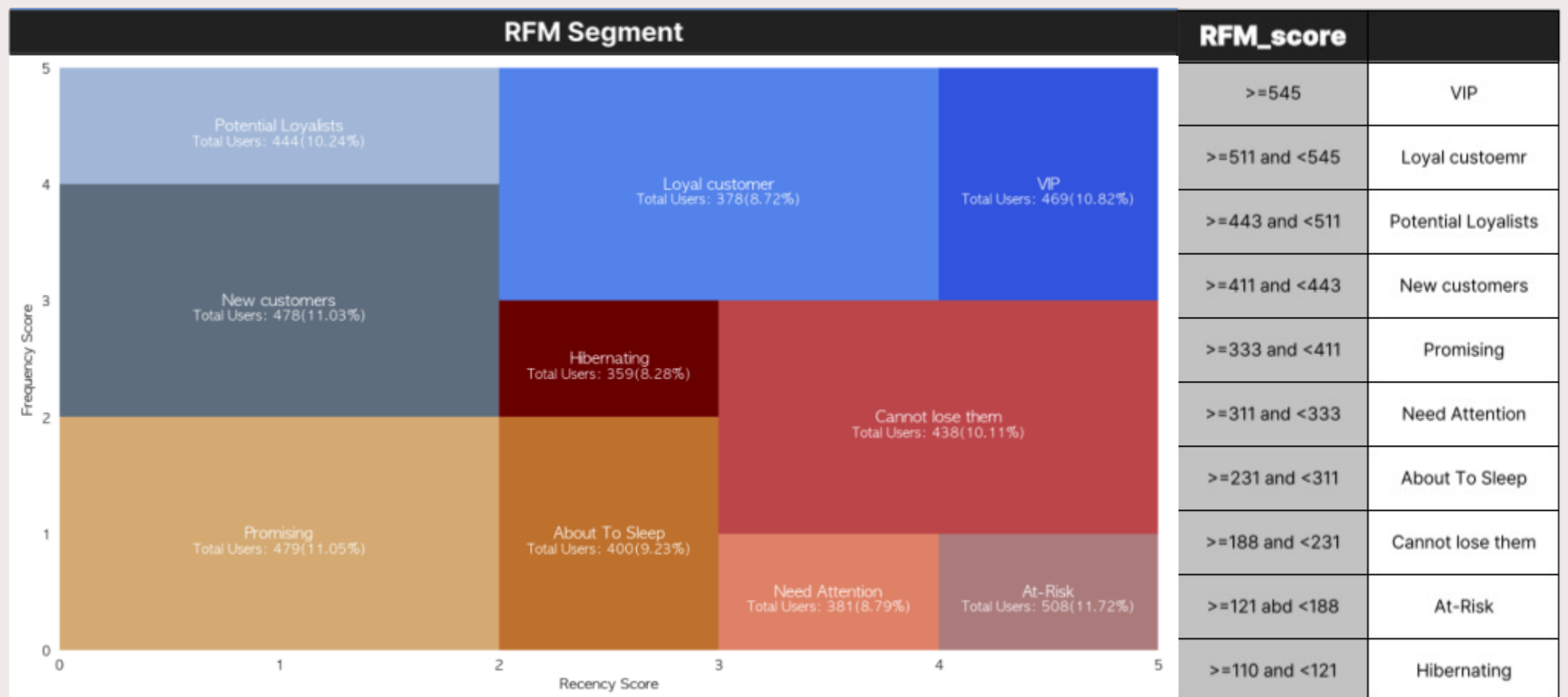
- Recency  
최근 거래 기준일 -고객별 최근 구매한 날짜
- Frequency  
고객별 구매 빈도수
- Monetary  
고객별 총 구매금액

### RFM 모형사용기법 : Scoring 기법 ⇒ RFM의 요인을 각각 5등급 간격으로 분류하는 방법

- Recency\_score: 최근 구매 날짜일수록, 높은 점수 부여
- Frequency\_score: 구매 횟수가 클수록, 높은 점수 부여
- Monetary\_score: 구매 총액이 클수록, 높은 점수 부여
- RFM\_score: RFM 요소에 대한 점수를 각각 100, 10, 1의 자릿수로 사용하여 하나의 점수로 통합  
⇒ RFM Segmentation 에 사용

# RFM 분석

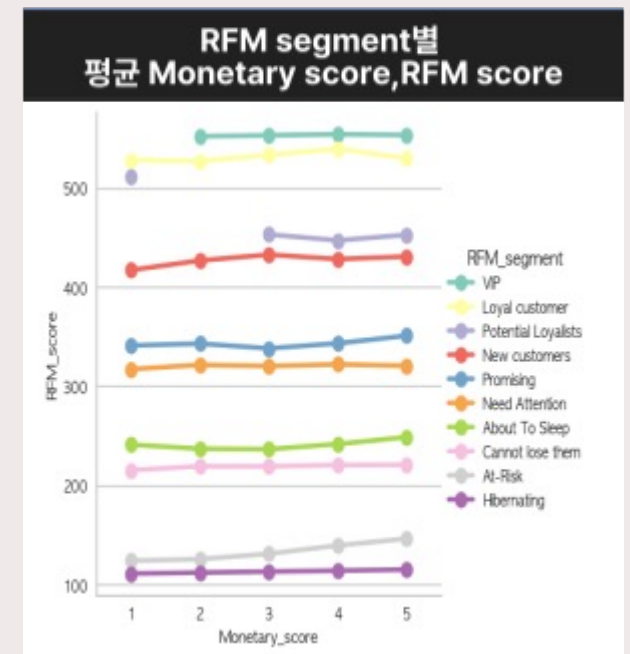
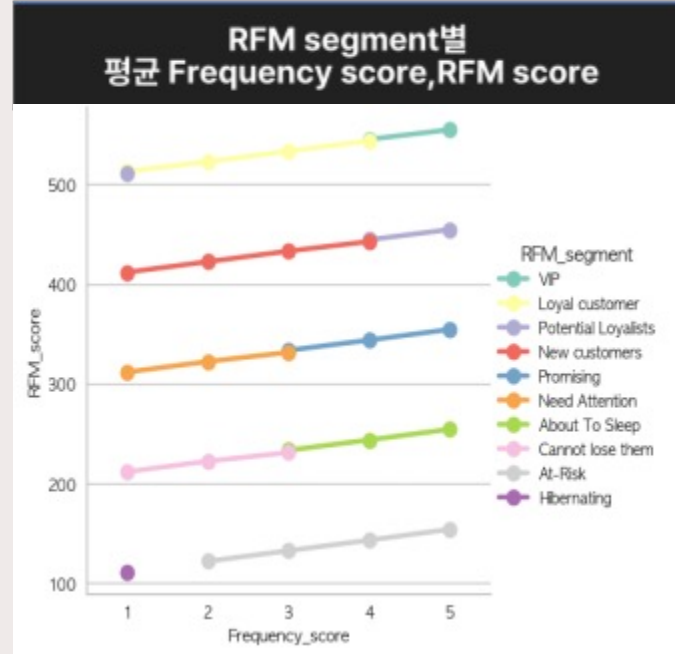
## RFM Segment



⇒ RFM score을 이용하여, segment 분류

# RFM 분석

## RFM Segment



# RFM 분석

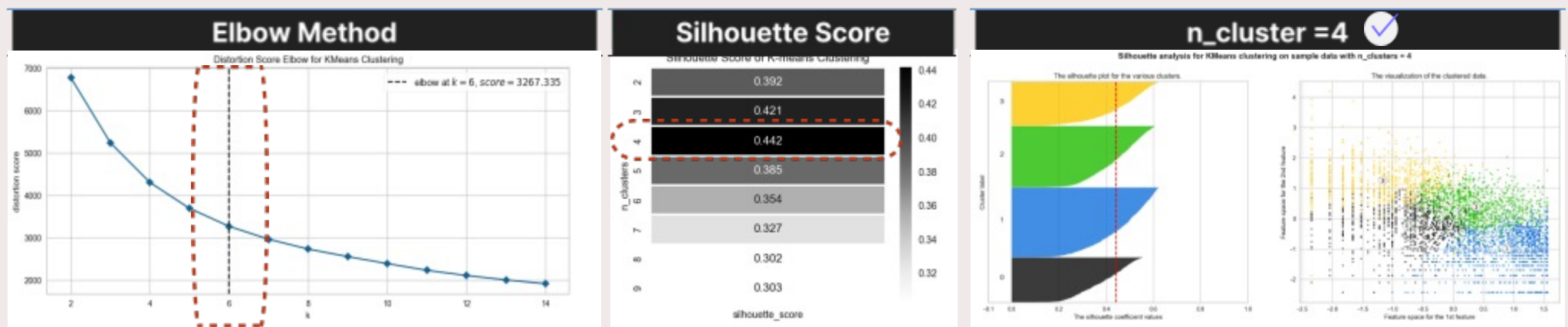
## RFM Segment

RFM Segment 통계 분석										
RFM_segment	Recency			Frequency			Monetary			
	mean	count	sum	mean	count	sum	mean	count	sum	
VIP	6.0	469	2794	344.0	469	161416	8568.0	469	4018366.0	가장 최근 활동/ 가장 높은 구매 빈도/가장 높은 구매금액 (높은 가치 창출 고객)
Loyal customer	7.0	412	2761	47.0	412	19561	1482.0	412	610440.0	최근 활동 /VIP에 비해 낮은 구매 빈도/ VIP에 비해 낮은 금액 소진 (충성도 높은 고객)
Potential Loyalists	23.0	452	10404	174.0	452	78527	3263.0	452	1475041.0	상대적으로 최근 구매 /높은 구매 빈도/ 높은 구매 금액 (잠재적 고객)
New customers	23.0	436	10171	30.0	436	13298	625.0	436	272699.0	최근 유입/ 낮은 빈도/낮은 구매 금액 (신규 고객)
Promising	52.0	479	24669	106.0	479	50719	1880.0	479	900498.0	평균적인 최근성/평균 빈도/평균 구매 금액 (가치를 높일 가능성 있는 고객)
Need Attention	53.0	381	20377	19.0	381	7364	435.0	381	165806.0	낮은 구매 빈도/낮은 구매 금액 (관심이 필요한 고객)
About To Sleep	114.0	417	47546	77.0	417	32140	1293.0	417	539044.0	활동은 있었지만, 더 이상 활발하지 않은 고객 (잠들기 직전 고객)
Cannot lose them	119.0	421	50048	14.0	421	6103	469.0	421	197435.0	평균 이하 최근성/낮은 구매 빈도/평균 구매 금액 값을 가진 고객(잃을수 없는 고객)
At-Risk	263.0	508	133679	38.0	508	19481	595.0	508	302045.0	낮은 최근성/낮은 구매 빈도/평균 구매 금액 값을 가진 고객 (이탈 위험 고객)
Hibernating	277.0	359	99326	7.0	359	2459	709.0	359	254627.0	가장 낮은 최근성/가장 낮은 빈도를 가진 고객



# K-means 클러스터링

## Cluster Segment



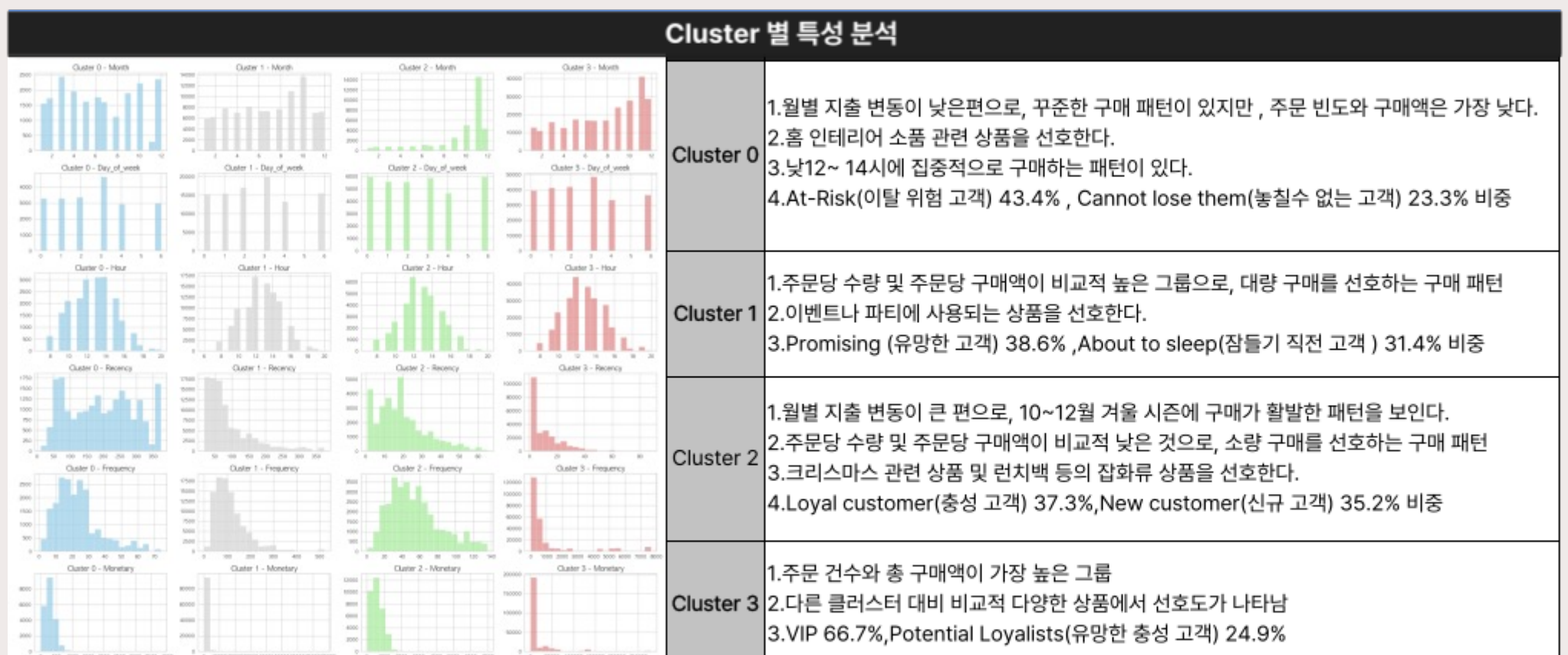
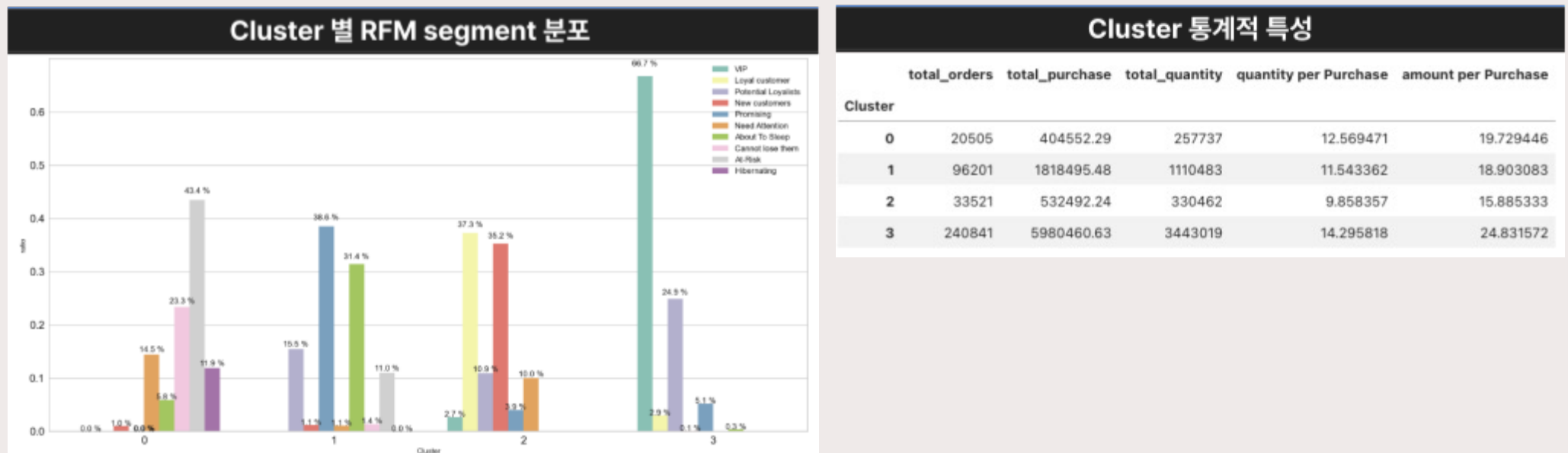
Elbow Method : 6 / Silhoutette Score: 4

두 지표를 함께 고려하여, 클러스터 개수 4 선택

⇒ 클러스터 4개로 분류했을때, 실루엣 계수는 0.442로 가장 높으며, 개별 군집의 평균 실루엣 계수 값이 비교적 균일하다.

# K-means 클러스터링

## Cluster 특성



# 클러스터 별 활용방안

## Cluster 0

### a. 고객 세그먼트 특성

- i. 월별 지출 변동이 낮은편으로, 꾸준한 구매 패턴이 있지만, 주문 빈도와 구매액은 가장 낮다.
- ii. 홈 인테리어 소품 관련 상품을 선호한다.
- iii. 낮12~ 14시에 집중적으로 구매하는 패턴이 있다.
- iv. 4.At-Risk(이탈 위험 고객) 43.4% , Cannot lose them(놓칠수 없는 고객) 23.3% 비중

### b. 고객 세그먼트 활용 전략

#### i. At-Risk (이탈 위험 고객):

- 1. 리워드 및 할인 제공: 추가 리워드나 할인을 제공하여 이탈 방지
- 2. 설문조사 및 피드백 요청: 고객의 불만,이탈 원인을 파악하기 위해 설문조사요청=>서비스나 상품 개선 시행.

#### ii. Cannot lose Them(잃을수 없는 고객)

- 1. 개인화된 혜택 제공: 할인, 멤버십 혜택, 또는 독점적인 프로모션을 통해 이탈 방지
- 2. 특별한 이벤트 및 선점 기회 제공:특별한 이벤트나 선점 기회를 제공하여 구매 유도

#### iii. 낮 12~14시에 특별한 프로모션 제공

### c. 장바구니 분석 +추천시스템 활용 전략

- i. 홈 인테리어 관련 상품에 대한 개인화된 추천 서비스 제공.
- ii. 이탈 위험이 있는 사용자에게는 이전에 관심을 보였던 상품, 카테고리를 중심으로 개인화된 상품 추천

## Cluster 1

### a. 고객 세그먼트 특성

- i. 주문당 수량 및 주문당 구매액이 비교적 높은 그룹으로, 대량 구매를 선호하는 구매 패턴
- ii. 이벤트나 파티에 사용되는 상품을 선호한다.
- iii. Promising (유망한 고객) 38.6% ,About to sleep(잠들기 직전 고객 ) 31.4% 비중

### b. 고객 세그먼트 활용 전략

#### i. Promising(유망한 고객)

- 1. 추가 혜택 및 프로모션 제공: 추가 혜택이나 독점적인 프로모션을 제공하여 더 많은 구매를 유도
- 2. 추가 서비스 및 제품 소개:신규 서비스나 제품을 소개하여 구매 유도

#### ii. About To Sleep (곧 잠들 예정의 고객 ):

- 1. 개인화된 알림 및 특별 제안 제공: 개인화된 알림이나 특별 제안을 통해 다시 관심을 유도
- 2. 재구매 유도 프로그램 도입: 재구매를 유도하는 CRM 메시지 전송 및 리텐션 마케팅 진행
- 3. 포인트 및 할인 유도 :추가 포인트나 할인을 제공하여 구매 유도

### c. 장바구니 분석 +추천시스템 활용 전략

- i. 추가 구매를 유도하는 번들 상품 제안:장바구니에 담긴 상품 간의 연관성을 고려하여,번들 상품을 제안
- ii. 개인화된 추천 시스템 : 이벤트 상품을 구매하는 경향으로 관련 상품 추천



# 클러스터 별 활용방안

## Cluster 2

### a. 고객 세그먼트 특성

- i. 월별 지출 변동이 큰 편으로, 10~12월 겨울 시즌에 구매가 활발한 패턴을 보인다.
- ii. 주문당 수량 및 주문당 구매액이 비교적 낮은 것으로, 소량 구매를 선호하는 구매 패턴
- iii. 크리스마스 관련 상품 및 런치백 등의 잡화류 상품을 선호한다.
- iv. 4.Loyal customer(충성 고객) 37.3%,New customer(신규 고객) 35.2% 비중

### b. 고객 세그먼트 활용 전략

#### i. Loyal Customer (충성 고객)

1. 추가 혜택 및 리워드 프로그램: 충성도를 인정하고 보상하는 리워드 프로그램을 제공하여 계속해서 구매 유도
2. 개별적인 감사 메시지: 고객에게 개별적으로 감사의 뜻을 전하는 메시지 전송

#### ii. New Customers (신규 고객)

1. 환영 혜택 및 할인 제공:초기 구매에 대한 환영 혜택이나 할인을 제공하여, 브랜드에 대한 긍정적인 경험을 제공
2. 계절 별 프로모션 및 할인:겨울 시즌에 특화된 할인 프로그램 도입하여 계절에 따른 구매를 유도.

### c. 장바구니 분석 +추천시스템 활용 전략

- i. 시즌별 상품 추천 :선호도 분석을 통해 겨울 시즌에 가장 선호하는 상품 식별 및 계절 상품 추천

## Cluster 3

### a. 고객 세그먼트 특성

- i. 주문 건수와 총 구매액이 가장 높은 그룹
- ii. 다른 클러스터 대비 비교적 다양한 상품에서 선호도가 나타남
- iii. VIP 66.7%,Potential Loyalists(유망한 충성 고객) 24.9%

### b. 고객 세그먼트 활용 전략

#### i. VIP 고객

1. VIP 멤버십 프로그램 운영: VIP와 Potential Loyalists 그룹에 대한 멤버십 프로그램 도입.
2. 선점적인 새로운 상품 소개:새로운 상품이나 서비스를 먼저 체험할 수 있는 기회를 제공하여 로열티 증대
3. 우선 배송 및 서비스 제공:빠른 배송 서비스나 24시간 고객 서비스 등을 우선적으로 제공

#### ii. Potential Loyalists (잠재적 충성고객)

1. 추가 혜택 및 가치 제시:높은 구매 빈도와 금액에 대한 추가 혜택을 제시하여, 실제 충성고객으로 성장하도록 유도
2. 추가 서비스 및 소식 제공 : 최신 상품이나 이벤트에 대한 정보를 제공하여, 고객의 참여를 유도

### c. 장바구니 분석 +추천시스템 활용 전략

- i. 최근 구매 이력 기반 추천:VIP 및 Loyal Customer의 최근 구매 이력을 기반으로 ,유사한 상품 추천
- ii. 고가 상품 및 프리미엄 아이템 추천:높은 구매 빈도와 금액을 갖는 이 세그먼트에게는 고가 상품이나, 프리미엄 아이템을 중심으로 추천

## 한계점

1. 고객 세분화를 위한 인구 통계학적 정보가 부족해서, 보다 다양한 측면의 인사이트 도출이 어려웠다.  
( ex). 성별,나이,월 수입..등)
1. 상품에 대한 평점 정보가 없어서, 추천 시스템을 다양하게 적용해보는데 어려움이 있었다.

# BASKET ANALYSIS

지지도 0.2이상,신뢰도 0.7이상 연관규칙 5가지 도출

Cluster 0										
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhanga_metric
17	(ROSES REGENCY TEACUP AND SAUCER, PINK REGENC...	(GREEN REGENCY TEACUP AND SAUCER)	0.022545	0.043636	0.021091	0.935484	21.436172	0.020107	14.823636	0.975344
3	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.032060	0.043636	0.027636	0.963636	19.791667	0.026240	7.013333	0.906061
1	(PAINTED METAL PEARS ASSORTED)	(ASSORTED COLOUR BIRD ORNAMENT)	0.029091	0.061818	0.023273	0.800050	12.941176	0.021474	4.690989	0.956375
18	(GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY...	(ROSES REGENCY TEACUP AND SAUCER )	0.027636	0.041455	0.021091	0.763158	18.409511	0.019945	4.047192	0.972558
6	(ROSES REGENCY TEACUP AND SAUCER )	(GREEN REGENCY TEACUP AND SAUCER)	0.041455	0.043636	0.029818	0.719296	16.483918	0.020609	3.407045	0.979959

Cluster 1										
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhanga_metric
238	(SET/10 BLUE POLKADOT PARTY CANDLES)	(SET/10 PINK POLKADOT PARTY CANDLES)	0.023837	0.030647	0.021566	0.904782	29.522046	0.020836	10.178207	0.989718
239	(SET/10 PINK POLKADOT PARTY CANDLES)	(SET/10 BLUE POLKADOT PARTY CANDLES)	0.030647	0.023837	0.021566	0.703704	29.522046	0.020836	3.294552	0.996672
101	(HAPPY STENCIL CRAFT)	(MONSTERS STENCIL CRAFT)	0.035187	0.030647	0.023837	0.677419	22.103943	0.022758	3.604694	0.989580
100	(MONSTERS STENCIL CRAFT)	(HAPPY STENCIL CRAFT)	0.030647	0.035187	0.023837	0.777778	22.103943	0.022758	4.341657	0.984945
63	(GLASS APOTHECARY BOTTLE PERFUME)	(GLASS APOTHECARY BOTTLE TONIC)	0.031782	0.031782	0.021566	0.678571	21.350765	0.020956	3.012234	0.984451

Cluster 3										
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhanga_metric
2098	(LUNCH BAG CARS BLUE, CHARLOTTE BAG SUKI DESIGN)	(RED RETROSPOT CHARLOTTE BAG)	0.023175	0.082271	0.023175	1.000000	12.154930	0.021268	inf	0.939502
2321	(JUMBO BAG SPACEBOY DESIGN, JUMBO BAG PINK VN...	(JUMBO BAG PINK POLKADOT)	0.024334	0.091541	0.023175	0.952381	10.403858	0.020947	19.077636	0.926425
3556	(LUNCH BAG RED RETROSPOT, LUNCH BAG APPLE DESI...	(LUNCH BAG BLACK SKULL...)	0.022016	0.133256	0.020857	0.947368	7.109382	0.017924	16.468134	0.878686
2574	(JUMBO BAG SCANDINAVIAN BLUE PAISLEY, LUNCH BA...	(JUMBO BAG RED RETROSPOT)	0.026651	0.165701	0.024334	0.913043	5.510185	0.019918	9.594438	0.840930
2568	(JUMBO BAG SCANDINAVIAN BLUE PAISLEY, JUMBO ST...	(JUMBO BAG RED RETROSPOT)	0.025492	0.165701	0.023175	0.909091	5.486332	0.018951	9.177289	0.839120

Cluster 4										
	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhanga_metric
203	(RED HANGING HEART T-LIGHT HOLDER)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.052923	0.187638	0.037159	0.714286	4.261884	0.028439	2.913295	0.867317
26	(CANDLEHOLDER PINK HANGING HEART)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.041288	0.187638	0.028076	0.680098	4.056552	0.021155	2.681156	0.785935
56	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.049546	0.061107	0.033031	0.666667	10.909910	0.030603	2.816680	0.955691
266	(REGENCY TEA PLATE GREEN )	(REGENCY CAKESTAND 3 TIER)	0.055986	0.156069	0.021470	0.694651	3.874246	0.015928	2.134648	0.769196
29	(CHARLOTTE BAG PINK POLKADOT)	(RED RETROSPOT CHARLOTTE BAG)	0.034882	0.049546	0.020644	0.595238	12.013889	0.018926	2.348181	0.949701

- ⇒ 두 제품 간의 연관성이 높으므로, 이를 함께 마케팅하거나 특별한 할인 혜택을 제공할 수 있다.
- ⇒ 이를 함께 판매하는등의 상품 번들구성을 기획할 수 있다.

# Recommend System

클러스터 기반 추천 시스템									
Description	4 PURPLE FLOCK DINNER CANDLES	50'S CHRISTMAS GIFT BAG LARGE	DOLLY GIRL BEAKER	I LOVE LONDON MINI BACKPACK	I LOVE LONDON MINI RUCKSACK	NINE DRAWER OFFICE TIDY	OVAl WALL MIRROR DIAMANTE	RED SPOT GIFT BAG LARGE	# 예시: 클러스터 2에서 사용자 12680.0 10개의 상품을 추천 cluster_id = 2 user_id = 12680.0 num_recommendations = 10 recommended_items_cluster_0 = recommend_items_in_cluster(cluster_id, user_id, num_recommendations) recommended_items_cluster_0
CustomerID									recommend_items
12346.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0 GIFT BAG BIRTHDAY
12347.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1 CARD GINGHAM ROSE
12348.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2 PACK OF 6 SANDCASTLE FLAGS ASSORTED
12349.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3 PAPERWEIGHT VINTAGE COLLAGE
12350.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4 PAINT YOUR OWN EGGS IN CRATE
									5 PACK OF SIX LED TEA LIGHTS
									6 PAPERWEIGHT VINTAGE PAISLEY
									7 DOORMAT NEW ENGLAND
									8 POPPY FIELDS CHOPPING BOARD
									9 POPPY'S PLAYHOUSE KITCHEN

## 클러스터 기반 추천 시스템 소개

사용자를 클러스터로 그룹화하여 취향 파악 ⇒ 각 클러스터에 대한 추천 제공  
이미 구매한 상품을 제외하고, 클러스터와 고객 기반으로 n개의 상품을 추천

# THANK YOU.