

매출 내역을 활용한 고객 및 재고 관리 효율성 전략 수립

활용 데이터

원본데이터 정보

- [캐글 데이터 | Superstore Sales Profit Discount Predict](#)
 - 슈퍼스토어 판매 데이터 : 가구, 전자제품, 사무용품을 주로 파는 미국 내 여러 지점이 있는 마트의 판매 데이터

EDA에 활용한 데이터

[Sample - Superstore_240830.csv](#)

- 300개 행이 제외된 데이터를 활용
 - 원본 데이터는 9994(컬럼행 미포함)개의 행으로 구성된 데이터였지만, mysql workbench에 데이터를 import 했을 때, 다수의 팀원들에게 공통적으로 300개의 행이 유실되어 9694(컬럼행 미포함)개의 행의 데이터를 활용하기로 함
 - 원본데이터에서 유식된 300개의 Row ID는 다음의 파일 참조

[check_RowID_240830.csv](#)

- 2014년 1월 4일 ~ 2017년 12월 30일 데이터로 구성되어 있음

데이터 상세 정보

컬럼명	데이터 상세	데이터 구성
Row ID	고유 행 번호	1~9994
Order ID	고유 주문번호	총 5,009건
Order Date	주문 일자	2014-01-03~2017-12-30
Ship Date	발송 일자	2014-01-07~2018-01-05
Ship Mode	배송 모드	총 3개 카테고리; Standard Class, Second Class, Other
Customer ID	고유 고객번호	총 793개
Customer Name	고객이름	총 793개

컬럼명	데이터 상세	데이터 구성
Segment	구매분류	Consumer, Corporate, Home Office
Country	국가	
City	도시	
State	주	
Postal Code	우편번호	
Region	지역	West, East, Central, South
Product ID	제품 번호	총 1,862개
Category	제품 카테고리	Office Supplies, Furniture, Other
Sub-Category	제품 세부 카테고리	
Product Name	제품명	
Sales	판매금액(매출)	
Quantity	주문 수량	
Discount	할인율	
Profit	순익	

1. 서론

1.1 배경

현대 비즈니스 환경에서 기업의 경쟁력을 유지하고 지속 가능한 성장을 이루기 위해서는 철저한 데이터 분석이 필수적이다. 특히, 월별 및 분기별 수익 데이터는 기업의 수익 구조를 이해하고, 시장 변화에 따른 대응 전략을 수립하는 데 중요한 역할을 한다. 기본적으로 월별 수익 데이터는 단기적인 수익 흐름을 세밀하게 분석할 수 있도록 하며, 분기별 수익 데이터는 장기적인 수익 패턴을 파악하는 데 도움이 된다.

'고객 관리 전략'을 세우기 위해서 탐색적 데이터 분석(EDA)을 통해 고객의 주문 패턴과 매출 관련 패턴을 파악하고자 하였다. 그 이후에 고객의 행동 패턴을 **코호트 분석**, **RFM 분석**, **고객생애가치분석**에 활용하여 고객 세분화(VIP 고객, VIP 잠재 고객, 이탈고객 등)를 하여, 고객 세분화 별 고객 관리 전략(마케팅)을 수립하고자 하였다.

실질적인 판매 과정에서 **재고 관리**는 필수적이다. 재고 관리의 효율성을 높이기 위해 판매 규모를 분석하고, 수요를 안정적으로 대비하기 위해 안전 재고의 개념을 정립하는 것이 좋다. 이를 위해 판매규모 EDA를 통해 판매 데이터를 분석하여 판매 패턴을 파악하고, 이를 바탕으로 필요한 재고를 예측하여 적정한 수준의 재고를 유지할 수 있도록 하는 과정이 요구된다. 감가 상각이나, 트렌드, 유통기한 등 시대의 흐름에 맞춰 수요 또한 변화하기에 분기 별로 서로 다른 재고 전략을 수립하는 것이 도움이 될 것이다.

1.2 분석 목표 및 진행 방향

이 프로젝트는 고객 관리 전략과 재고 관리 효율성을 수립하기 위해 진행되었다.

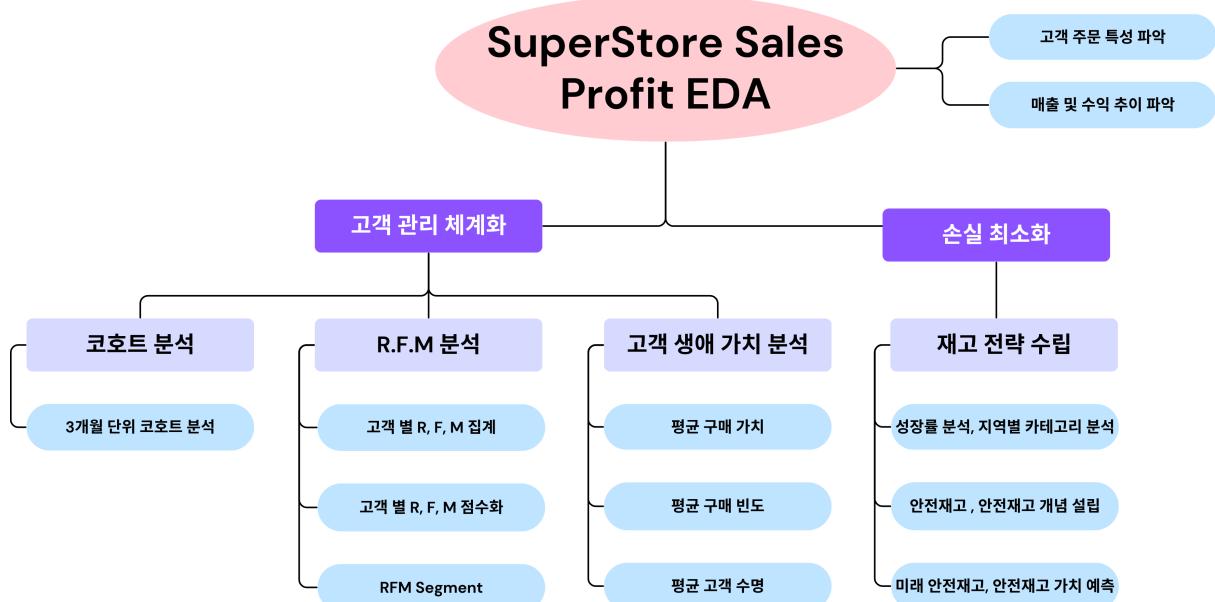
- 고객 관리 전략을 강화하기 위해 진행한 분석

1. EDA를 활용한 고객의 주문 특성 및 시기 별 매출 특성을 파악
2. 코호트 분석

- 3. RFM 분석
 - 4. 고객생애주기 분석
- 1~4를 통한 인사이트 도출 및 마케팅 전략 수립
- 재고 관리 전략을 강화하기 위해 진행한 분석
 1. 판매규모를 파악을 위한 EDA
 2. 안전 재고 개념 수립
 3. 필요한 재고 예측

→ 1~3를 통한 인사이트 도출 및 재고 예측

2. 필요한 데이터 정렬 및 정제와 가설 설립



2.1 EDA 분석

- 고객 주문 특성 관련 분석
 - 고유 고객수 및 고유 주문건수 관련 분석
 - 고유 고객수와 고유 주문건수의 변화를 살펴보면서 고객 기반의 성장 및 변동이 어떻게 되고 있는지 파악하여, 마케팅 및 고객 관리 전략을 세울 때 기준을 마련하고자 한다.
 - 유지기간 관련 분석
 - 고객이 얼마나 오랜 기간 동안 회사와 관계를 유지하는지 파악하며, 고객의 생애주기를 파악하고자 한다. 고객이 처음으로 주문한 시점부터 마지막으로 주문한 시점까지의 기간 및 재주문일까지의 소요기간을 파악하면서 고객 이탈을 방지하기 위한 마케팅 및 고객 관리 전략을 세울 때 기준을 마련하고자 한다.

- 매출 관련 특성 분석
 - 매출 분석
 - 총 매출 합계 분석을 통해 기업의 성장 지표를 확인하고, 장기적인 성공 가능성, 고객 유지 등을 거시적 측면에서 확인할 수 있다.
 - 분기 별 매출 합계 분석을 통해 해당 분기에 적용된 마케팅, 판매 전략 등이 수익에 어떻게 반영되었는지 볼 수 있다.
 - 지역 별 매출 합계 분석을 통해 사무 지역, 주거 지역 등 지리적 특성에 맞는 판매 전략 수립 및 재고 유지에 대해 계획 할 수 있다.

2.2 고객 관리 전략 수립을 위한 분석

(1) 코호트 분석

- 고객의 주문 특성을 분석하기 위해 2014년부터 2017년까지의 4년치 데이터를 활용하여 탐색적 데이터 분석(EDA)과 코호트 분석을 수행하였다. 먼저, EDA를 통해 고객의 재주문 주기, 이탈률, 고객당 총 구매 건수 등 다양한 지표를 분석하여 고객의 주문 패턴을 파악했다. 분석 결과, 재주문이 주로 3개월 단위로 이루어진다는 사실을 확인하고, 이를 바탕으로 코호트 분석을 3개월 단위로 구분하여 진행하였다. 이를 통해 각 고객군의 행동 변화를 시간에 따라 추적하고, 고객 관리 전략 수립에 필요한 인사이트를 도출하였다

(2) RFM 분석

- RFM 분석은 고객의 구매 행동을 기반으로 고객을 세분화해 그 가치를 평가하는데 유용한 도구이다. 특히, 분석에 활용한 Superstore와 같은 소매업체는 구매 빈도, 구매 금액 및 구매 시기 등에서 큰 차이를 보이는 등 다양한 고객 층이 존재하는 것이 특징인데, 고객 세분화를 통해 이들을 명확히 구분하고 각 고객층의 특성에 맞는 맞춤형 마케팅 전략을 수립하는 것은 중요한 과제일 것이다. 또한 마케팅 자원이 제한적인 소매업체의 배경에서 가치 있는 고객을 식별해 이들에게 집중적인 마케팅을 수행함으로써 자원을 효과적으로 활용하는 최적화된 방법을 제시할 것이다.

(3) 고객생애가치 분석

- 고객 생애가치 분석을 통해서 우리 회사에 더 가치가 높은 고객을 파악하기 위해 고객 생애가치 분석을 수행하였다. 각 고객이 주문한 총 매출액, 구매 횟수, 빈도 등을 활용하여 각 고객이 우리 회사에 어느정도의 가치를 가져올 수 있을지 파악하고, 고객 관리 전략 수립을 세우고자 한다.

2.3 비용 관리 분석 및 최적화를 위한 분석

- 재고 예측은 슈퍼스토어의 효율적인 재고 관리와 고객 만족도를 높이는 데 필수적이다. 이를 위해, 데이터를 기반으로 각 지역과 카테고리별 성장률을 분석하여 미래의 안전재고 수량과 가치를 예측한다. 안전재고는 평균 판매 수량과 표준 편차를 고려하여 계산하며, 이는 예기치 않은 수요 변동에 대비할 수 있도록 돋는다. 이러한 예측을 통해 물류 자원을 최적화하고, 재고 부족이나 과잉 문제를 사전에 예방할 수 있다.

3. 데이터 분석 결과

3.1 EDA

(1) 고객 주문 특성 파악

(1-1) 4년 동안 고객 및 주문수 (2014/01/04 ~2017/12/30)

- 고유 고객수 : 793명
- 고유 주문수 : 4,931건
- SQL 코드

```
-- 4년 동안 고객수 & 주문수 구하기
select count(distinct `customer name`), count(distinct `Order ID`)
from superstore_sales_240830;
```

(1-2) 연도/분기 별 고유 고객수 및 고유 주문건수

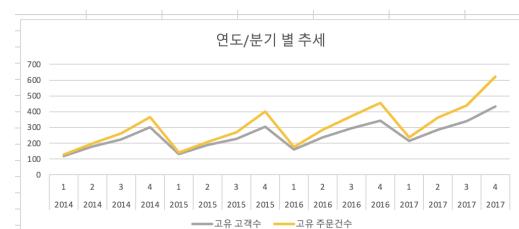
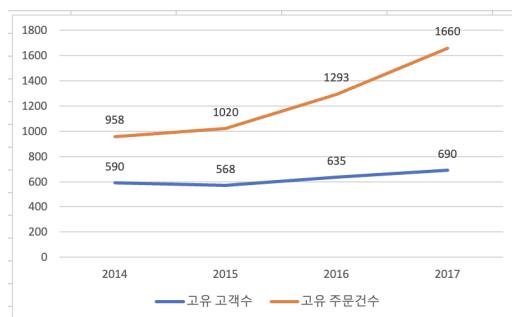
- SQL 코드

```
-- 연도 별 고객수 & 주문수 구하기
select year(order_date_new), count(distinct `customer name`), count(distinct
from superstore_sales_240830
group by year(order_date_new);

-- 연도/분기 별 고객수 & 주문수 구하기
select year(order_date_new) as year, quarter(order_date_new) as quat, count(d
from superstore_sales_240830
group by year, quat;
```

(1-3) 연도 별 성장률

- **고유 주문건수는 2014년 대비 2017년에는 약 73.7% 성장**
 - 2014년 대비 2015년: 약 6.48% 성장
 - 2015년 대비 2016년: 약 26.88% 성장
 - 2016년 대비 2017년: 약 28.35% 성장
- **고유 고객수는 2014년 대비 2017년에는 약 16.95% 성장**
 - 2014년 대비 2015년: 약 -3.73% 감소
 - 2015년 대비 2016년: 약 11.80% 증가
 - 2016년 대비 2017년: 약 8.66% 증가
- 연도 별 주문건수는 시간이 흐름에 따라 늘어났지만 고유고객수는 많이 늘지 않음
 - 연도 별 고유 고객수 및 고유 주문건수 추세
 - 연도/분기 당 추세



(1-4) 고객 당 구매건수

- SQL 코드

```
with order_num as (
    select `customer id`, count(1) as order_num
        from `sample - superstore_240830`
        group by `customer id`
        order by 2 desc),

first_order as(
    select `customer id`, min(str_to_date(`Order Date`, '%m/%d/%Y')) as first_o
        from `sample - superstore_240830`
        group by `customer id`),

first_order_details as (select a.* , year(a.first_order_date), month(a.first_o)
from(select o.* , f.first_order_date from order_num o
    left join first_order f on o.`customer id` = f.`customer id`) a),

total_order_num as(select order_num, count(order_num) as cnt
    from first_order_details
    group by order_num
),

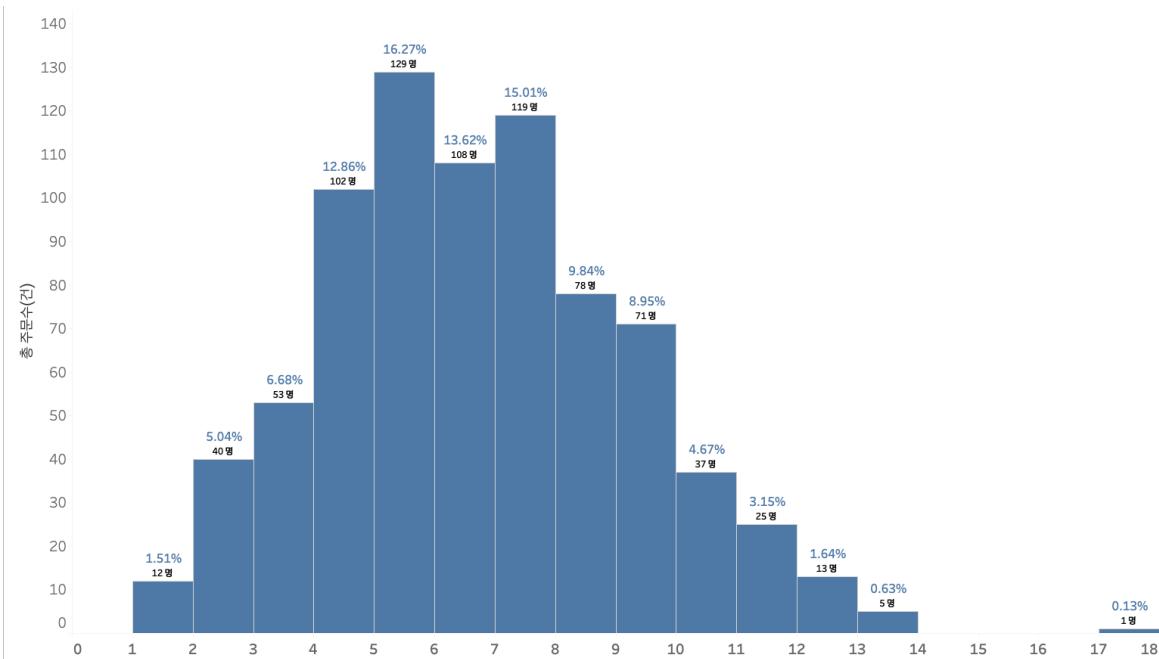
total_count AS (
    SELECT SUM(cnt) AS total_cnt
    FROM total_order_num
)

SELECT t.order_num,
    t.cnt,
    t.cnt / total.total_cnt * 100 AS percentage
FROM total_order_num t
CROSS JOIN total_count total
order by 3 desc;
```

- 구매건수 별 고객수 히스토그램(14~17년 전체)

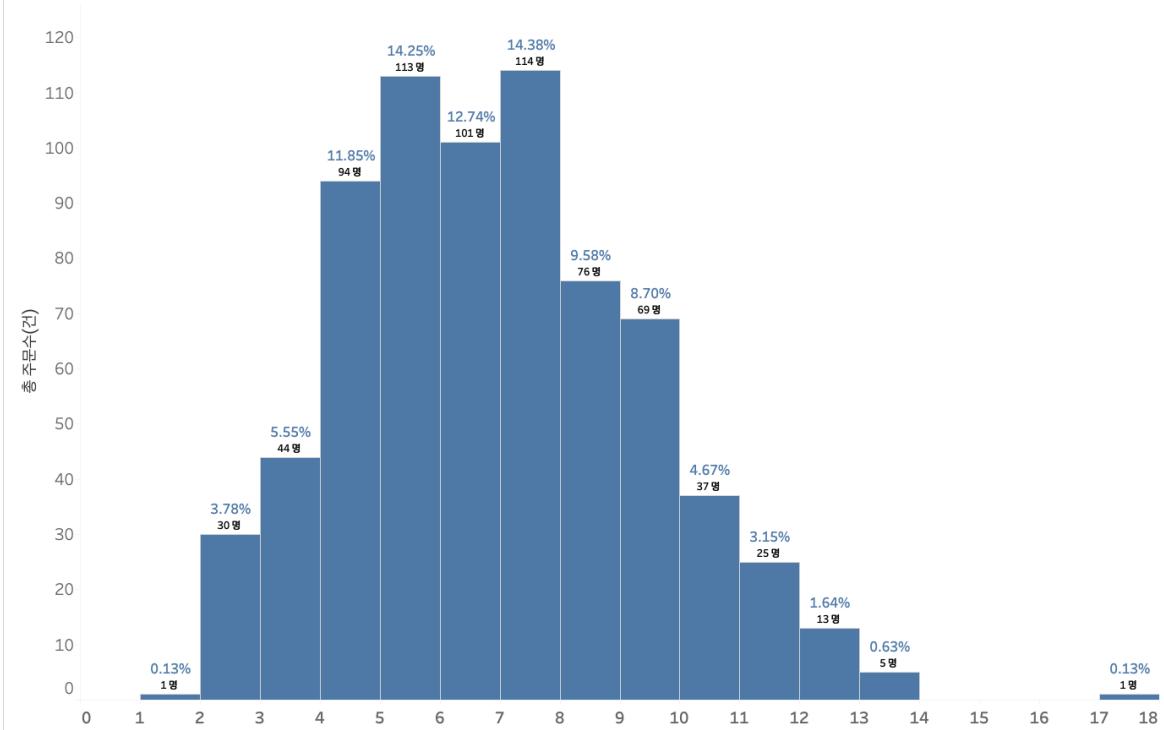
- SQL 코드

```
-- 구매건수 별 고객수 히스토그램(14~17년 전체)
select year(order_date_new), count(distinct `customer name`), count(distinct `customer id`)
    from superstore_sales_240830
    group by year(order_date_new);
```



- 약 70.79%의 고객들은 8번 미만 주문을 함
 - 약 57.76%의 고객들은 4번 이상 8번 미만 주문을 함
 - 약 13.23%의 고객들은 4번 미만 주문함
- (2014~2015년에 첫 주문을 시작한 고객들 대상) 구매건수 별 고객수 히스토그램
 - SQL 코드

```
-- 연도/분기 별 고객수 & 주문수 구하기
select year(order_date_new) as year, quarter(order_date_new) as quart, cou
from superstore_sales_240830
group by year, quart;
```



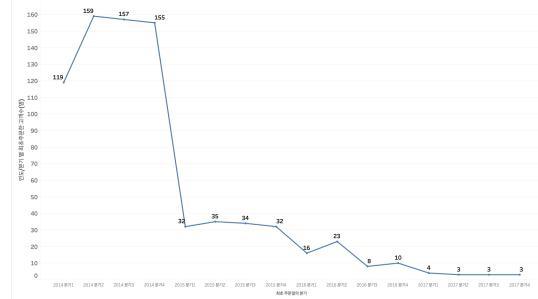
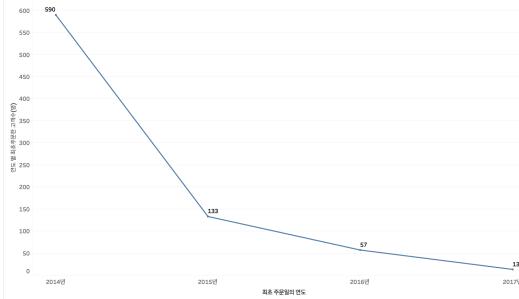
- 약 62.68%의 고객들은 8번 미만 주문을 함

(1-5) 고객 별 첫 주문일(연도 별 신규회원수)

- 연도 별 최초주문 고객수
 - SQL 코드

```
-- 고객 별 최초 주문일, 마지막 주문일, 최초 주문일과 마지막 주문일 차이
select `customer id`
, min(order_date_new) as first_order
, year(min(order_date_new)) as first_order_year
, month(min(order_date_new)) as first_order_month
, max(order_date_new) as last_order
, year(max(order_date_new)) as last_order_year
, month(max(order_date_new)) as last_order_month
, datediff(max(order_date_new), min(order_date_new)) as days_between_first
, timestampdiff(month, min(order_date_new), max(order_date_new)) as mont
from superstore_sales_240830
group by `customer id`;
```

- 연도 별 첫 주문고객수
- 연도/분기 별 첫 주문고객수



- 2014년 이후로 신규회원의 주문이 크게 줄어듦

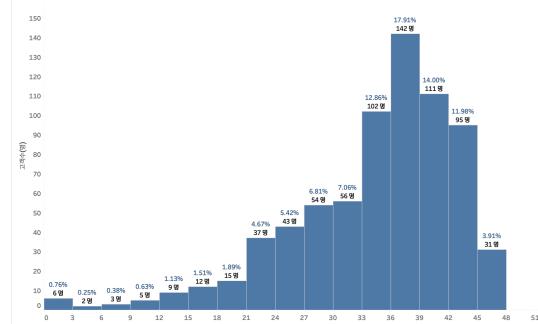
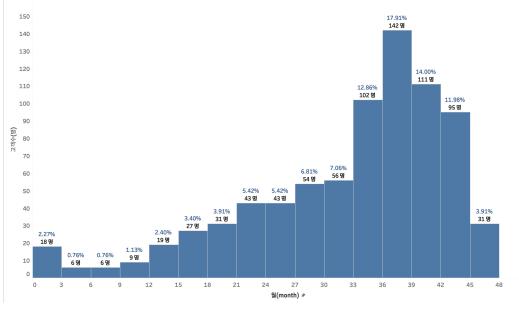
(1-6) 고객 유기기간(마지막 주문일 - 최초 주문일)

- SQL 코드

```
-- 고객 별 최초 주문일, 마지막 주문일, 최초 주문일과 마지막 주문일 차이
select `customer id`
, min(order_date_new) as first_order
, year(min(order_date_new)) as first_order_year
, month(min(order_date_new)) as first_order_month
, max(order_date_new) as last_order
, year(max(order_date_new)) as last_order_year
, month(max(order_date_new)) as last_order_month
, datediff(max(order_date_new), min(order_date_new)) as days_between_first_
, timestampdiff(month, min(order_date_new), max(order_date_new)) as months_
from superstore_sales_240830
group by `customer id`;
```

- 전체 고유고객 기준

- 2014~2015년에 주문하기 시작한 고객 기준



- 한 번 유입되면 장기고객(2.8년 ~ 3.6년 동안 주문)이 되는 경향으로 보이지만, 주문을 자주하지 않는다.

(1-7) 재주문일까지의 소요기간

- SQL 코드

```
-- 고객 별 다음 주문일 계산하기
with unique_order_date as(
    select distinct `customer id`, order_date_new
```

```

from superstore_sales_240830),

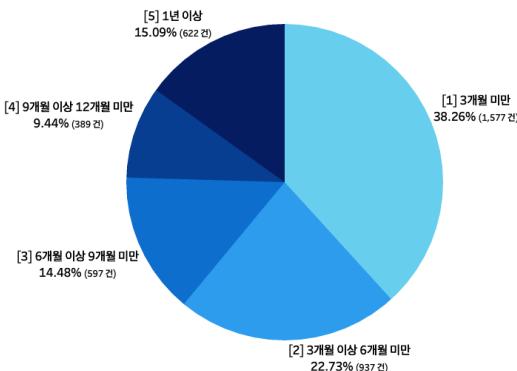
next_order_date as (SELECT
    `customer id`
    , `order_date_new`
    , LEAD(`order_date_new`, 1) OVER (PARTITION BY `customer id` ORDER BY `or
FROM unique_order_date

ORDER BY
    `customer id`, `order_date_new`)

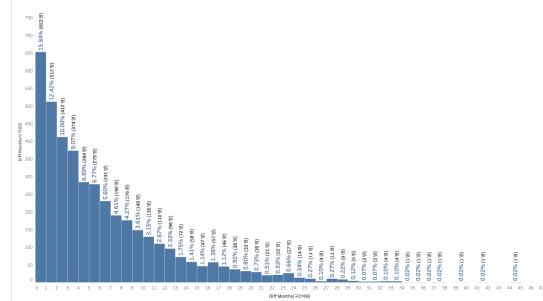
select *
    , datediff(next_order_date, order_date_new) as diff_days
    , TIMESTAMPDIFF(MONTH, `order_date_new`, `next_order_date`) AS diff_months
from next_order_date
where next_order_date is not null;

```

- 재주문 내역의 3개월 단위 별 원차트



- 재주문 내역의 1개월 단위 별 히스토그램



- 전체 주문의 약 38.26%는 3개월 이내 재주문(61%는 6개월 이내 재주문)

(2) 매출 및 수익 추이 파악

(2-1) 분기 별 총 수익 분석

연도 별 매출 성장 추세는 다음과 같이 설명할 수 있다.

- **2014년:** 총 매출은 약 481,764 달러를 기록한다. 이 해는 기준 연도 이므로, 성장을 계산에서 제외된다.
- **2015년:** 매출이 약 464,426 달러로 전년 대비 **3.6%** 감소한다. 이 해는 매출이 약간 감소하는 해이다.
- **2016년:** 매출이 약 601,265 달러로 **29.5%** 증가한다. 이 해는 매출이 큰 폭으로 성장하는 해로, 긍정적인 성장 추세를 보인다.
- **2017년:** 매출이 약 724,995 달러로 **20.6%** 증가한다. 성장세가 지속되며, 안정적으로 매출이 증가하는 해이다.



(2-2) 판매금액 합계가 높은 고객(customer_name) 30위

- 회사에 이득이 되는 것은 매출(sales)을 가장 높여주는 고객으로 판매금액 합계가 가장 높은 30명을 선별하여 합계 금액별 순위를 시각화 하였다.

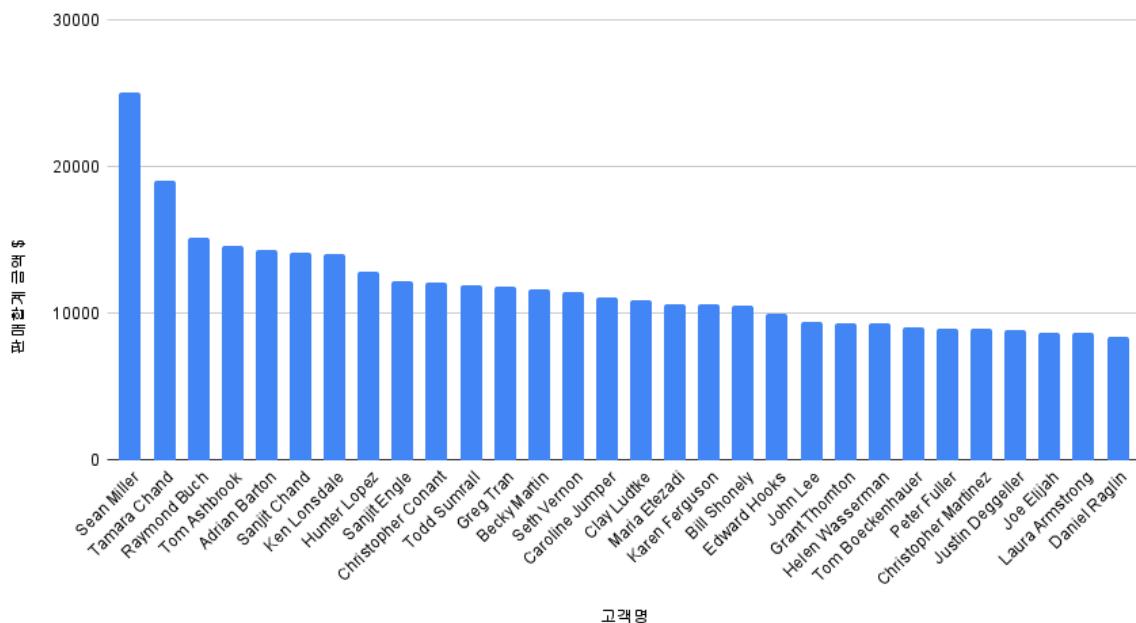
SQL query:

```
select customer_name, round(sum(sales), 2) sales
from superstore
group by 1
order by 2 desc
limit 30;
```

excel link :

[https://docs.google.com/spreadsheets/d/1U5KdBHEw6ea9iM4N4Ho7Si8z2jGUV3d8sZ7h6jGUpJk/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1U5KdBHEw6ea9iM4N4Ho7Si8z2jGUV3d8sZ7h6jGUpJk/edit?usp=drive_link)

판매합계 순위 30위까지 고객



- 분석 : 1순위, 2순위의 판매금액이 높았으며 그 이하의 순위는 완만한 차이를 보였다.

(2-3) 판매금액 합계가 높은 주(state) 30위

- 어느 주(state)가 판매금액(sales)을 높여주었는지 판매금액 30위 순위별로 시각화 하였다.

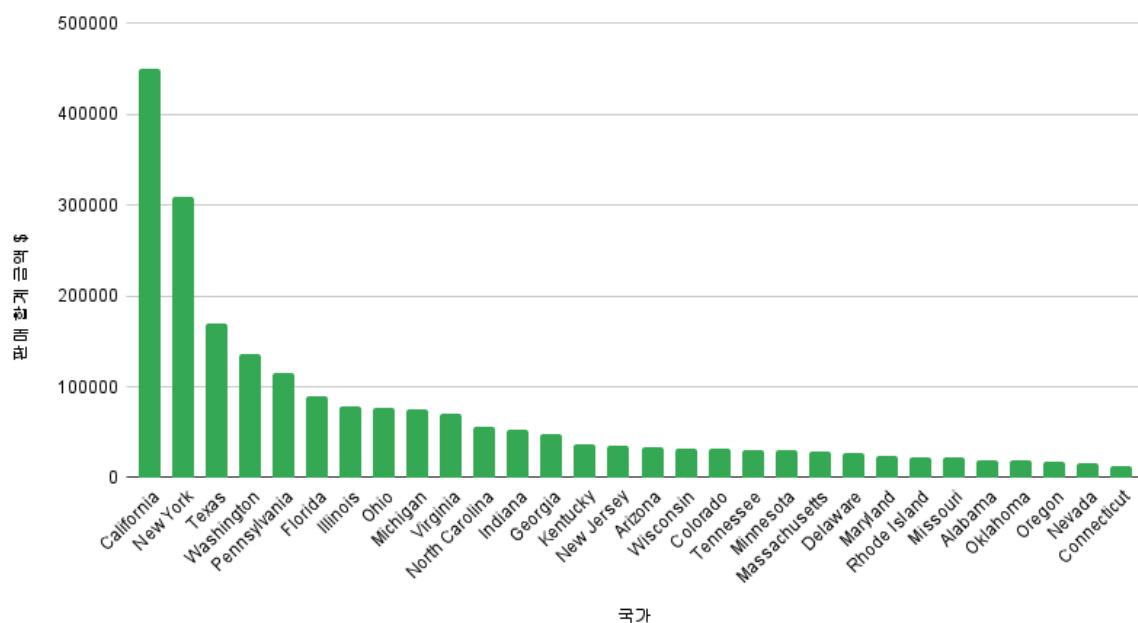
SQL query:

```
select state, round(sum(sales), 2) sales
from superstore
group by 1
order by 2 desc
limit 30;
```

excel link

[https://docs.google.com/spreadsheets/d/1z2eS0Y_UOt5RPlqTTPPJqY8xZHdXLz9WGVqO_0XPj-4/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1z2eS0Y_UOt5RPlqTTPPJqY8xZHdXLz9WGVqO_0XPj-4/edit?usp=drive_link)

주(state)에 대한 판매 합계 금액 30위



- 분석 : California 가 1위, New York 이 2위를 기록하였다. 6위 이하의 순위는 완만한 차이를 보인다.

(2-4) 판매금액 합계가 높은 도시(city) 30위

- 어느 도시(city)가 판매금액(sales)을 높여주었는지 판매금액 30위 순위별로 시각화 하였다.

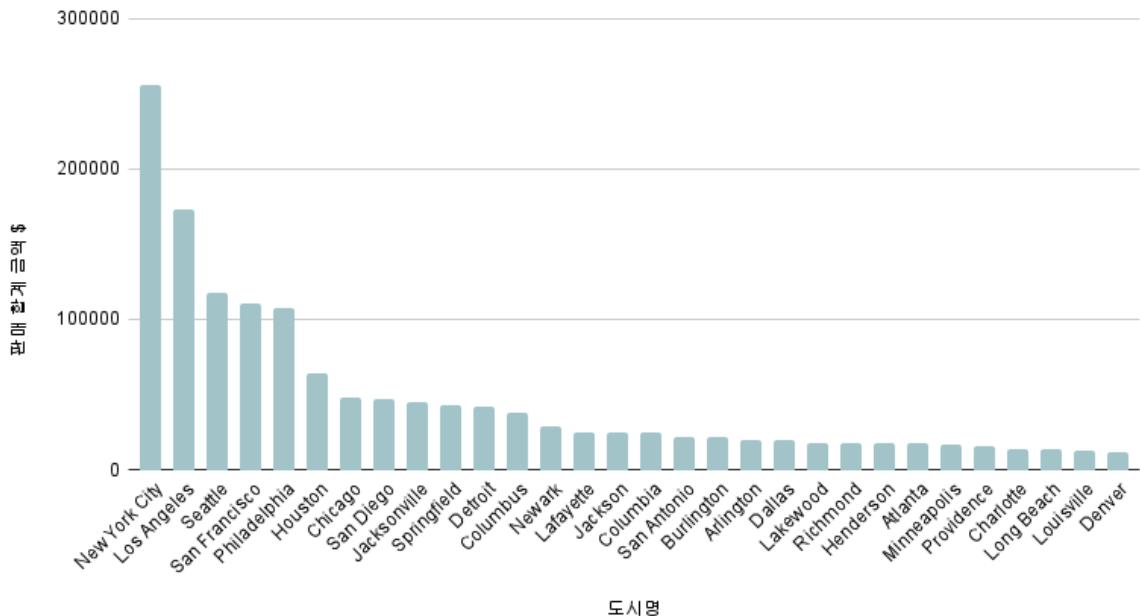
SQL query :

```
select city, round(sum(sales), 2) sales
from superstore
group by 1
order by 2 desc
limit 30;
```

excel link :

[https://docs.google.com/spreadsheets/d/108YSDfXkfo9htVO8cSGeWM2FA0onx6FU0fexjBha8ew/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/108YSDfXkfo9htVO8cSGeWM2FA0onx6FU0fexjBha8ew/edit?usp=drive_link)

30개 도시명에 대한 판매 합계 금액



- 분석 : New York City 가 1위, Los Angeles 가 2위를 기록하였다. 상위 5위까지가 매출의 큰 부분을 차지하고 있다.

(2-5) 판매금액(내림차순), 수익금액 별 선순위 도시(city) 30위

- 도시(city)의 판매금액(sales) 30위 순위별로 나열 하였으며, 동시에 수익(profit) 을 함께 시각화 하였다.

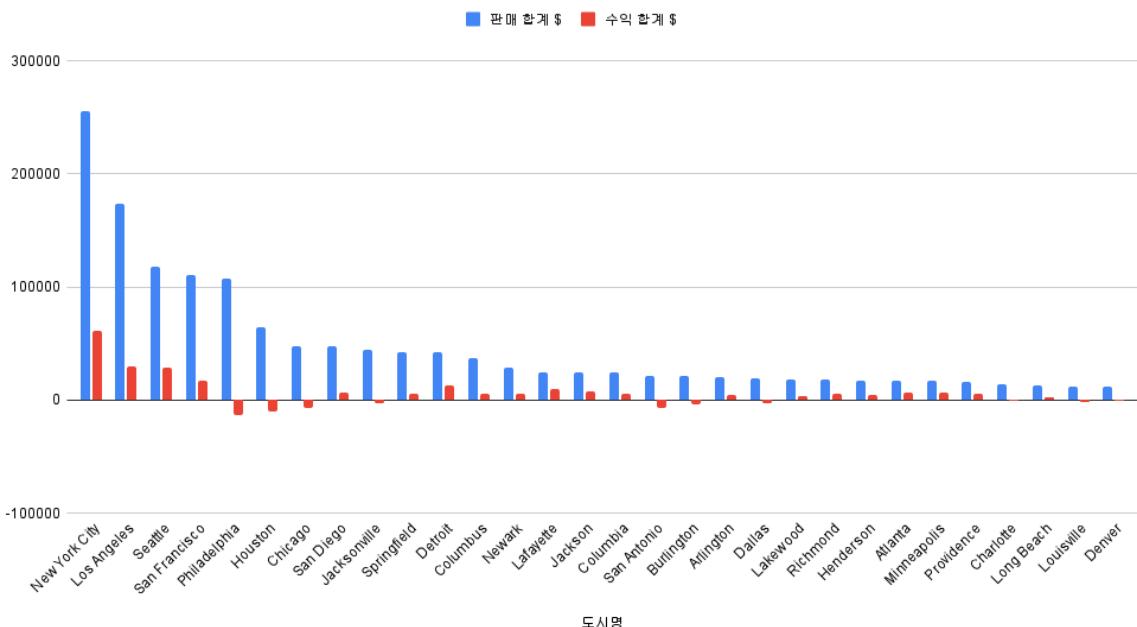
SQL query :

```
select city, round(sum(sales), 2) sales, round(sum(profit), 2) profits
from superstore
group by 1
order by 2 desc, 3 desc
limit 30;
```

excel link :

[https://docs.google.com/spreadsheets/d/1EbYc5rjAyP7LPZZe5hVZinKNW8BP1Ft0ev0w2u9LS8o/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1EbYc5rjAyP7LPZZe5hVZinKNW8BP1Ft0ev0w2u9LS8o/edit?usp=drive_link)

선순위 30개 도시별 판매금액, 수익금액



- 분석 : 상위 4위까지는 수익이 양수 이지만, 5위 부터는 수익이 음수를 나타내는 도시도 주목할 만 하다. 이는 discount 및 원가, 각종 프로모션의 영향으로 추론되며, 추후 다른 분석 및 가설 생성시 참고할 만하다.

(2-6) 카테고리별 매출, 수익, 판매수량

- 카테고리 3가지(Technology, Furniture, Office Supplies)의 총 매출(sales), 수익(profit), 판매수량(quantity)를 시각화 하였다.

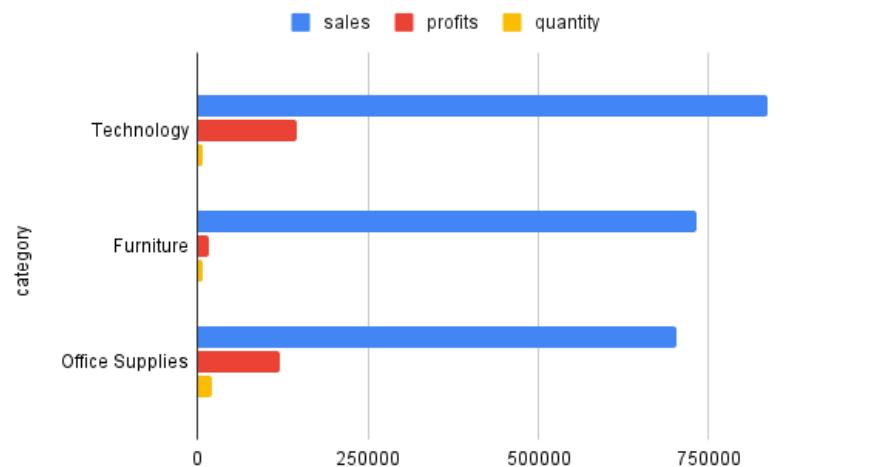
SQL query :

```
select category, round(sum(sales), 2) sales, round(sum(profit), 2) profits, sum(
from superstore
group by 1
order by 2 desc, 3 desc, 4 desc;
```

excel link :

https://docs.google.com/spreadsheets/d/1eRijPYNjpcI8kvOAXXp_VAs_aqGajCj_KZ1Y6rLZsrQ/edit?usp=drive_link

category 별 sales, profits, quantity



- 분석 : 가장 큰 매출과 수익을 내준 카테고리는 Technology 이지만 수익 대비 매출이 높은 것은 Office Supplies이다. 하지만 판매량을 고려하면 단가당 금액이 높은(수량대비 금액)것은 Technology 이다.

(2-7) 상품별 매출, 수익 10순위

- 상품별 매출(sales) 순위 10위를 시각화 하였으며, 수익(profit)을 함께 표시하였다.

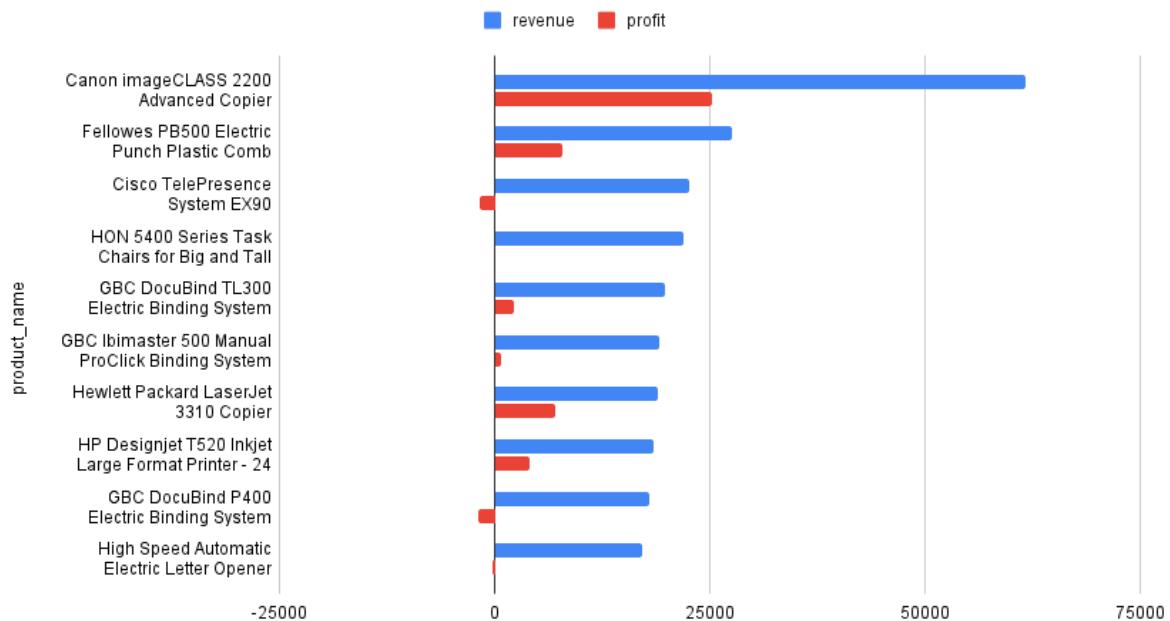
SQL query :

```
select product_name, round(sum(sales), 2) sales, round(sum(profit), 2) profits
from superstore
group by 1
order by 2 desc, 3 desc
limit 10;
```

excel link :

[https://docs.google.com/spreadsheets/d/1Plav6LZR_FHU1AmF8sYNxWpxkENVIOD1RjDupCEPOI8/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1Plav6LZR_FHU1AmF8sYNxWpxkENVIOD1RjDupCEPOI8/edit?usp=drive_link)

상품별 매출, 수익 순위 10



- 분석 : 매출과 수익의 1순위는 전자제품 이었으며, 10순위 안에 제품들은 대부분 전자제품이다. 특이하게 사무용 의자가 매출 4순위를 기록했지만, 수익은 0을 기록하고 있다.

(2-8) segment 별 매출, 수익

- 세그먼트 별 매출과 수익을 시각화 하였다.

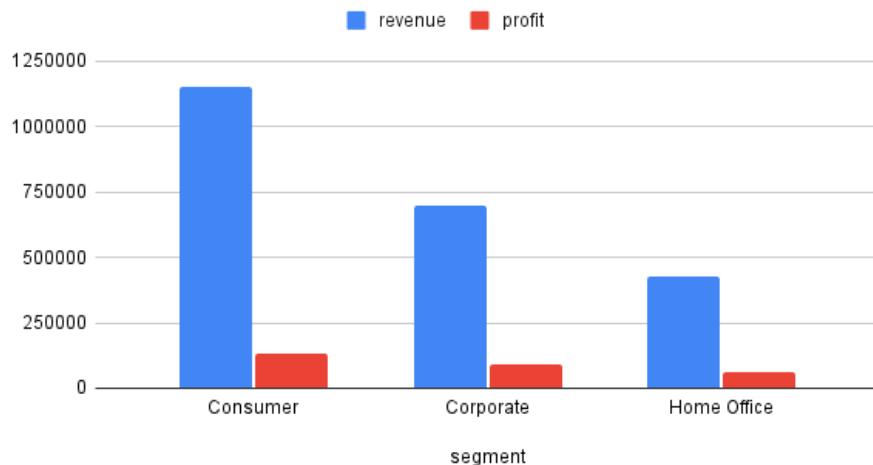
SQL query :

```
select segment, round(sum(sales), 2) sales, round(sum(profit), 2) profits
from superstore
group by 1
order by 2 desc, 3 desc;
```

excel link :

[https://docs.google.com/spreadsheets/d/1EhR3kqJBOkkdI2lFy21XUmRxhDiPDJGE_hRdpWtI0I/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1EhR3kqJBOkkdI2lFy21XUmRxhDiPDJGE_hRdpWtI0I/edit?usp=drive_link)

segment 별 매출, 수익



- 분석 : 매출 대비 수익이 높지는 않지만, 일반 소비자가 가장 많은 매출을 기록하고 있다.

(2-9) region 별 매출, 수익

4개의 지역(region/ West, East, Central, South)별로 매출(sales)과 수익(profit)을 시각화 하였다.

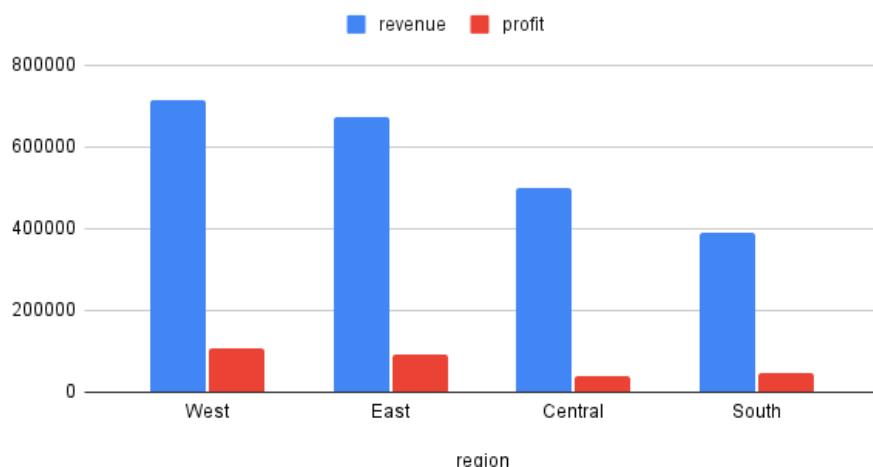
SQL query :

```
select region, round(sum(sales), 2) sales, round(sum(profit), 2) profits
from superstore
group by 1
order by 2 desc, 3 desc;
```

excel link :

[https://docs.google.com/spreadsheets/d/1lghddOfQVtAWbL9adXfV9wldZi_lv06BE7bmflq0jqE/edit?
usp=drive_link](https://docs.google.com/spreadsheets/d/1lghddOfQVtAWbL9adXfV9wldZi_lv06BE7bmflq0jqE/edit?usp=drive_link)

region 별 매출, 수익



- 분석 : 서부(West)가 가장 높은 매출과 수익을 기록 하였으며, 2순위로 동부(East)가 있다. 이는 위의 분석에서 주(state)별 1순위, 2순위 인 California, New York 이 각각 서부와 동부에 영향을 끼친 것으로 추론된다.

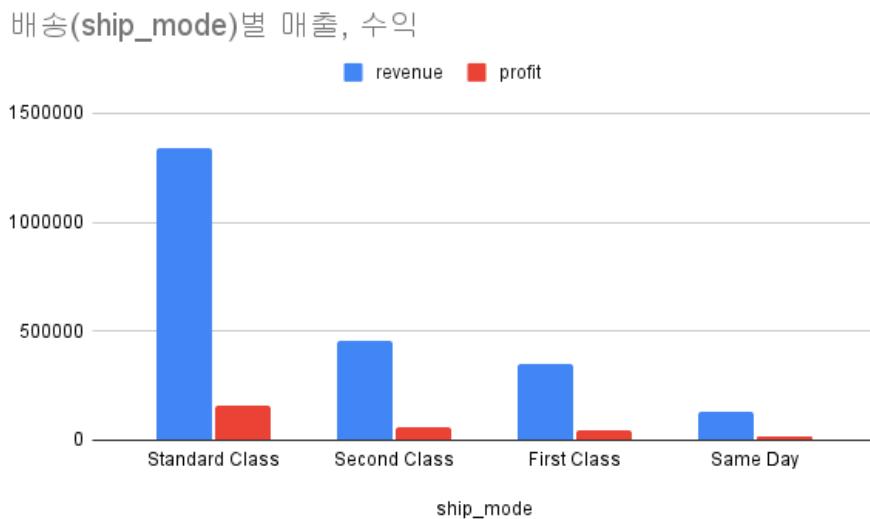
(2-10) 배송(ship_mode) 별 매출, 수익

- 배송별 매출(sales)과 수익(profit)을 시각화 하였다.

SQL query :

```
select ship_mode, round(sum(sales), 2) sales, round(sum(profit), 2) profits
from superstore
group by 1
order by 2 desc, 3 desc;
```

excel link : https://docs.google.com/spreadsheets/d/1MpSljG88nWbZ44foiW5-2biHRdpe3moiHfp0ORBKQbs/edit?usp=drive_link



- 분석 : 금일 배송(Same Day) 나 1등급 배송(First Class) 즉 좋은 옵션과 상관없이 일반적인 배송이 가장 많은 매출과 수익을 기록했다. 좋은 옵션의 비용이 상대적으로 높아서 소비자들의 다수가 일반적인 배송을 선택하였을 수 있지만, 소비자들은 배송방식에 큰 가중치를 두지 않는 것으로 추론된다.

kaggle notebook link : <https://www.kaggle.com/code/charles01dev/eda-by-sql-for-superstore-sales-dataset?scriptVersionId=194966014>

(2-11) 지역별 판매(sales) 지도

- 데이터는 미국 내 대형마트의 판매를 기록한 것으로, 주(state) 별로 매출(sales)을 한 눈에 볼 수 있도록 등치 지도(Choropleth Map)로 시각화 하였다.

python code :

```
state_df = pd.DataFrame(state, state_code) # Create a dataframe
state_df.reset_index(level=0, inplace=True)
state_df.columns = ['State Code', 'State']
```

```

sales = df.groupby(["State"]).sum("Sales").sort_values("Sales", ascending=False)
sales.reset_index(level=0, inplace=True)
sales.drop(columns='Postal Code', inplace = True)
sales= sales.sort_values('State', ascending=True)
sales.reset_index(inplace = True)
sales.drop(columns='index', inplace = True)
sales.insert(1, 'State Code', state_df['State Code'])

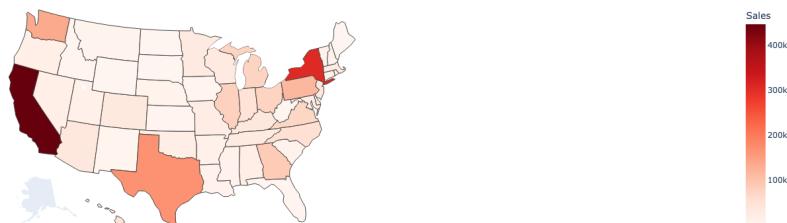
import plotly.graph_objects as go

sales['text'] = sales['State']
fig = go.Figure(data=go.Choropleth(
    locations=sales['State Code'],
    text=sales['text'],
    z = sales['Sales'].astype(float),
    locationmode = 'USA-states',
    colorscale = 'Reds',
    colorbar_title = "Sales",
))

fig.update_layout(
    title_text = 'Sales',
    geo_scope='usa',
)
fig.show();

```

Sales



(2-12) 지역별 수익(profits) 지도

- 데이터는 미국 내 대형마트의 판매를 기록한 것으로, 주(state) 별로 수익(profit)을 한 눈에 볼 수 있도록 등차 지도 (Choropleth Map)로 시각화 하였다.
- python code :

```

profits = df.groupby(["State"]).sum("Profit").sort_values("Profit", ascending=False)
profits.reset_index(level=0, inplace=True)

```

```

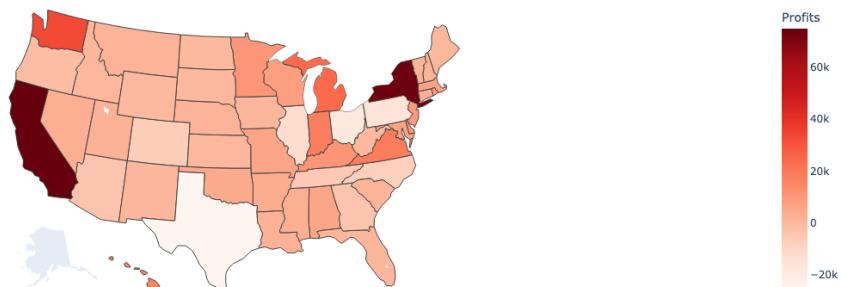
profits.drop(columns='Postal Code', inplace = True)
profits= sales.sort_values('State', ascending=True)
profits.reset_index(inplace = True)
profits.drop(columns='index',inplace = True)
# profits.insert(1, 'State Code', state_df['State Code'])

profits['text'] = profits['State']
fig = go.Figure(data=go.Choropleth(
    locations=sales['State Code'],
    text=sales['text'],
    z = profits['Profit'].astype(float),
    locationmode = 'USA-states',
    colorscale = 'Reds',
    colorbar_title = "Profits",
))

fig.update_layout(
    title_text = 'Profits',
    geo_scope='usa',
)
fig.show()

```

Profits



- 분석 : 지도에서 매출(sales)과 수익(profit)이 모두 주(state) 별 분석에서 나타난 것처럼 서부(West)의 California 와 동부(East) 의 New York 이 높게 나타났다. 주목할 점은 Texas 주에서 매출(sales)은 다소 높게 나온반면, 수익(profit)은 낮게 나타난 점이다.

3.2 고객 관리 전략 수립을 위한 분석 결과

(1) 코호트 분석

- SQL 코드(3개월 단위)

```

WITH first_purchase AS (
    SELECT `customer id`, MIN(order_date_new) AS cohort_day
    FROM superstore_sales_240830
    GROUP BY `customer id`
),
cohort_day AS (
    SELECT
        s.*,
        f.cohort_day,
        TIMESTAMPDIFF(MONTH, cohort_day, s.order_date_new) AS cohort_index,
        CONCAT(
            YEAR(cohort_day), '-',
            LPAD(((QUARTER(cohort_day) - 1) * 3) + 1, 2, '0'), '-01'
        ) AS cohort_group
    FROM
        superstore_sales_240830 s
    LEFT JOIN first_purchase f ON s.`customer id` = f.`customer id`
)
SELECT
    cohort_group,
    FLOOR(cohort_index / 3) AS cohort_index, -- Adjusting the cohort index for
    COUNT(DISTINCT `customer id`) AS customer_count
FROM
    cohort_day
GROUP BY
    cohort_group,
    FLOOR(cohort_index / 3)
ORDER BY
    cohort_group,
    cohort_index;

```

- 코호트 분석

Cohort Gro..	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Cohort Index															
2014년 1월	100.00	17.65	29.41	20.17	20.17	24.37	37.82	26.89	28.57	26.05	43.70	31.09	36.97	42.02	46.22	26.05
2014년 4월	100.00	26.42	32.08	20.75	27.04	40.25	32.70	23.27	30.82	44.65	38.99	35.22	31.45	50.31	32.70	.
2014년 7월	100.00	23.57	23.57	28.03	37.58	31.85	22.29	35.67	43.31	33.76	33.76	43.31	47.77	24.20	.	.
2014년 10월	100.00	21.94	22.58	35.48	30.32	25.81	30.32	43.23	31.61	36.13	36.77	52.26	36.13	.	.	.
2015년 1월	100.00	21.88	31.25	18.75	15.63	37.50	40.63	43.75	25.00	37.50	46.88	18.75
2015년 4월	100.00	37.14	34.29	22.86	22.86	62.86	34.29	22.86	37.14	62.86	28.57
2015년 7월	100.00	47.06	17.65	26.47	35.29	41.18	44.12	41.18	41.18	26.47
2015년 10월	100.00	12.50	28.13	34.38	25.00	37.50	34.38	40.63	40.63
2016년 1월	100.00	37.50	37.50	43.75	56.25	37.50	68.75	37.50
2016년 4월	100.00	56.52	30.43	30.43	39.13	47.83	34.78
2016년 7월	100.00	37.50	25.00	37.50	25.00	62.50
2016년 10월	100.00	20.00	.	60.00	60.00
2017년 1월	100.00	25.00	50.00
2017년 4월	100.00	66.67
2017년 7월	100.00	33.33
2017년 10월	100.00

- 평균적으로 3개월 단위 당 약 34.9% 수준으로 구매율이 유지됨

(2) RFM 분석

(2-1) RFM 분석이란?

RFM 분석이란 고객의 행동 패턴을 통해 마케팅 전략을 수립하는 방식 중 가장 범용적으로 사용되는 CRM(Customer Relationship Management) 기법 중 하나로, RFM 분석은 다음 세 가지 요소를 기준으로 한다.

1. Recency (최신성):

- 고객이 마지막으로 구매한 시점이 얼마나 최근인지를 측정한다.
- Recency가 낮은 고객은 최근에 구매한 고객으로, 장래에도 구매 가능성성이 높다고 판단한다.

2. Frequency (빈도):

- 고객이 일정 기간 동안 얼마나 자주 구매했는지를 측정한다.
- Frequency가 높은 고객은 해당 기간 동안 자주 구매한 고객으로, 장래에도 구매 빈도가 높다고 판단한다.

3. Monetary Value (금액):

- 고객이 일정 기간 동안 얼마나 많은 금액을 지출했는지를 측정한다.
- Monetary Value가 높은 고객은 해당 기간 동안 많은 지출을 한 고객으로, 장래에도 지출 금액이 많을 것으로 판단한다.

위 세 가지 지표를 집계한 뒤 고객군 별로 속성을 정의한다.

(2-2) RFM 분석의 의의

RFM 분석은 각 사용자들의 평소 구매 패턴을 기준으로 분류를 진행하기 때문에 각 사용자 그룹의 특성에 따라 차별화된 마케팅 메시지를 전달할 수 있다.

- 최근에 자주 구매하며 높은 금액을 지출한 고객을 VIP 고객으로 분류하고, 특별 혜택을 제공한다.
- 오랜 기간 동안 구매가 없었던 고객에게 재참여를 유도하는 프로모션을 제공한다.
- 이탈 위험이 높다고 판단되는 고객을 식별해, 그들에게 재참여를 유도하는 전략을 마련한다.

실질적인 데이터 기반의 의사결정 방식이므로 타겟 마케팅에 대한 각 고객군 별 변화를 쉽게 모니터링할 수 있어 해당 비즈니스 전략의 성과를 평가하고 향상시키는데 용이하다.

또한 복잡한 데이터 모델링 도구나 기술이 필요하지 않은 비교적 단순한 데이터 분석 기법이므로 비용 효율성이 높다는 장점이 있다.

(2-3) RFM 분석 및 시각화

Dataset

Kaggle - Superstore Sales Dataset을 7개의 컬럼으로 축소시켜 정제한 데이터셋

	Order Date	Order ID	Customer ID	Product ID	Category	Sub-Category	Sales
▶	11/8/2016	CA-2016-152156	CG-12520	FUR-BO-10001798	Furniture	Bookcases	261.96
	11/8/2016	CA-2016-152156	CG-12520	FUR-CH-10000454	Furniture	Chairs	731.94
	6/12/2016	CA-2016-138688	DV-13045	OFF-LA-10000240	Office Supplies	Labels	14.62
	10/11/2015	US-2015-108966	SO-20335	FUR-TA-10000577	Furniture	Tables	957.5775
	10/11/2015	US-2015-108966	SO-20335	OFF-ST-10000760	Office Supplies	Storage	22.368
	6/9/2014	CA-2014-115812	BH-11710	FUR-FU-10001487	Furniture	Furnishings	48.86
	6/9/2014	CA-2014-115812	BH-11710	OFF-AR-10002833	Office Supplies	Art	7.28
	6/9/2014	CA-2014-115812	BH-11710	TEC-PH-10002275	Technology	Phones	907.152
	6/9/2014	CA-2014-115812	BH-11710	OFF-BI-10003910	Office Supplies	Binders	18.504
	6/9/2014	CA-2014-115812	BH-11710	OFF-AP-10002892	Office Supplies	Appliances	114.9
	6/9/2014	CA-2014-115812	BH-11710	FUR-TA-10001539	Furniture	Tables	1706.184
	6/9/2014	CA-2014-115812	BH-11710	TEC-PH-10002033	Technology	Phones	911.424
	4/15/2017	CA-2017-114412	AA-10480	OFF-PA-10002365	Office Supplies	Paper	15.552

Table Name : super

Num. of Column : 7

Num. of Row : 9694

Order Date(date) : 주문 일자

Order ID(string) : 주문 ID

Customer ID(string) : 고객 ID

Product ID(string): 상품 ID

Category(string): 상품 카테고리

Sub-Category(string): 상품 하위 카테고리

Sales(integer) : 지출

Segmentation(분할) 수행

1. 사용자 별 RFM 수치 집계

```
1 •  SELECT `Customer Name`,
2      SUM(`Sales`) AS Monetary_Value,
3      COUNT(`Customer ID`) AS Frequency,
4      MAX(STR_TO_DATE(CONCAT('20', RIGHT(`Order Date`, 2), SUBSTRING(`Order Date`, 1, 6)), '%Y%m%d')) AS Last_Order_Date,
5      (SELECT MAX(STR_TO_DATE(CONCAT('20', RIGHT(`Order Date`, 2), SUBSTRING(`Order Date`, 1, 6)), '%Y%m%d')) FROM super) AS Max_Order_Date,
6      DATEDIFF(
7          (SELECT MAX(STR_TO_DATE(CONCAT('20', RIGHT(`Order Date`, 2), SUBSTRING(`Order Date`, 1, 6)), '%Y%m%d')) FROM super),
8          MAX(STR_TO_DATE(CONCAT('20', RIGHT(`Order Date`, 2), SUBSTRING(`Order Date`, 1, 6)), '%Y%m%d'))
9      ) AS Recency
10     FROM super
11     GROUP BY `Customer Name`;
```

	Customer Name	Monetary_Value	Frequency	Last_Order_Date	Max_Order_Date	Recency
▶	Claire Gute	1148.780000000002	5	2017-01-26	2017-12-30	338
	Darrin Van Huff	1119.483000000002	9	2017-12-11	2017-12-30	19
	Sean O'Donnell	2602.575500000001	15	2017-12-01	2017-12-30	29
	Brosina Hoffman	6255.351000000001	24	2017-12-07	2017-12-30	23
	Andrew Allen	1782.872	11	2017-04-15	2017-12-30	259
	Irene Maddox	4930.473999999999	21	2016-12-05	2017-12-30	390
	Harold Pawlan	1990.314	10	2017-11-30	2017-12-30	30
	Pete Kriz	8158.654	23	2017-12-21	2017-12-30	9
	Alejandro Grove	2565.257999999994	13	2017-03-27	2017-12-30	278
	Zuschuss Donatelli	1493.944	9	2017-06-11	2017-12-30	202
	Ken Black	2742.758000000003	22	2017-05-14	2017-12-30	230
	Sandra Flanaqan	763.545999999998	9	2017-11-01	2017-12-30	59

• Customer Name

고객의 이름을 나타내는 속성으로 각 고객별로 RFM 지표를 계산하기 위한 기준 열로 사용했다.

• Monetary_Value (SUM(Sales))

특정 고객이 발생시킨 총 매출 금액이다. Customer ID를 기준으로 Sales 열의 값을 SUM을 통해 합산해 각 고객이 얼마나 많은 금액을 지출했는지 계산했으며, RFM 분석에서 Monetary로 사용된다.

• Frequency (COUNT(Customer ID))

특정 고객이 구매한 총 횟수이다. Customer ID를 기준으로 각 고객의 구매 빈도를 계산할 수 있도록 COUNT 문을 사용했다. RFM 분석에서 Frequency로 사용된다.

- Last_Order_Date (MAX(STR_TO_DATE(...)))

특정 고객의 마지막 구매 날짜를 의미한다. Order Date 열을 날짜 형식으로 변환한 후, 해당 고객의 마지막 구매 날짜를 추출하는 방식으로 진행했다.

이 때, Order Date가 'MM/DD/YY' 형식의 문자열로 저장되어 있으므로, STR_TO_DATE 함수를 사용해 날짜 형식으로 변환했다. 또한 연도가 두 자릿수로 저장되어 있으므로, '20'을 추가하여 변환했다.

- Max_Order_Date (SELECT MAX(STR_TO_DATE(...)) FROM super)

전체 데이터에서 가장 최근의 주문 날짜로, 서브쿼리를 사용해 super 테이블의 모든 주문 중 가장 최근의 날짜를 계산했다.

도출된 Max_Order_Date는 각 고객의 Recency를 평가하기 위한 기준점으로 사용된다.

- Recency (DATEDIFF(..., MAX(STR_TO_DATE(...))))

기준이 되는 가장 최근의 주문 날짜(Max_Order_Date)와 각 고객의 마지막 구매 날짜(Last_Order_Date) 사이의 일수를 DATEDIFF 함수를 통해 계산했다.

일수 차이가 작을수록 해당 고객은 최근에 구매한 것으로 평가되며, RFM 분석에서 Recency로 사용된다.

2. 집계된 RFM 데이터 점수화

RFM 분석의 핵심은 각 사용자의 분포를 파악해 Recency, Frequency, Monetary Value에 점수를 매기는 과정에 있다. 분포의 경향성을 파악해 적절한 점수를 배분하는 것이 마케팅 전략을 최적화하는 것에 직결되기 때문이다. 점수를 매기는 방법은 여러 가지가 있지만, 가장 범용성 있는 세 가지를 소개한다.

- 분위(Quantiles) 기반 점수화

전체 데이터를 분위로 나누어 각 분위에 따라 점수를 할당하는 방법이다.

예를 들어, 데이터를 5개의 분위로 나눈다면, 상위 20%에 속하는 고객에게 5점을, 그 다음 20%에 속하는 고객에게 4점을 주는 방식이다.

모든 데이터 분포에 관계없이 고르게 점수를 분배할 수 있어, 비교적 간단하게 구현할 수 있다는 장점이 있으나 데이터의 분포가 심하게 불균형한 경우, 정밀한 점수화가 어려울 수 있다.

(예를 들어 0~100점을 기준으로 100 99 98 97 96는 모두 경계값 근처에 위치한 훌륭한 Value이지만 분위 기반 점수화로 할당된다면 각각 5, 4, 3, 2, 1의 점수를 부여받기 때문에 비즈니스 차원에서 유의미한 점수화로 보기 어렵다.)

- 정규분포(Standard Distribution) 기반 점수화

고객들의 R, F, M 값을 대한 표준 정규분포로 변환한 후, 해당 값에 따라 점수를 할당하는 방법이다. 즉, 평균을 기준으로 얼마나 떨어져 있는지를 평가해 점수를 매긴다.

분포의 평균과 표준편차를 토대로 Z-Score로 변환한 후, 일정한 구간을 정해 평균에 가까운 고객은 중간 점수(3점), 떨어진 정도를 고려해 현저히 높은 고객의 경우 5점을, 낮은 경우 1점을 부여한다.

- K-means 클러스터링 기반 점수화

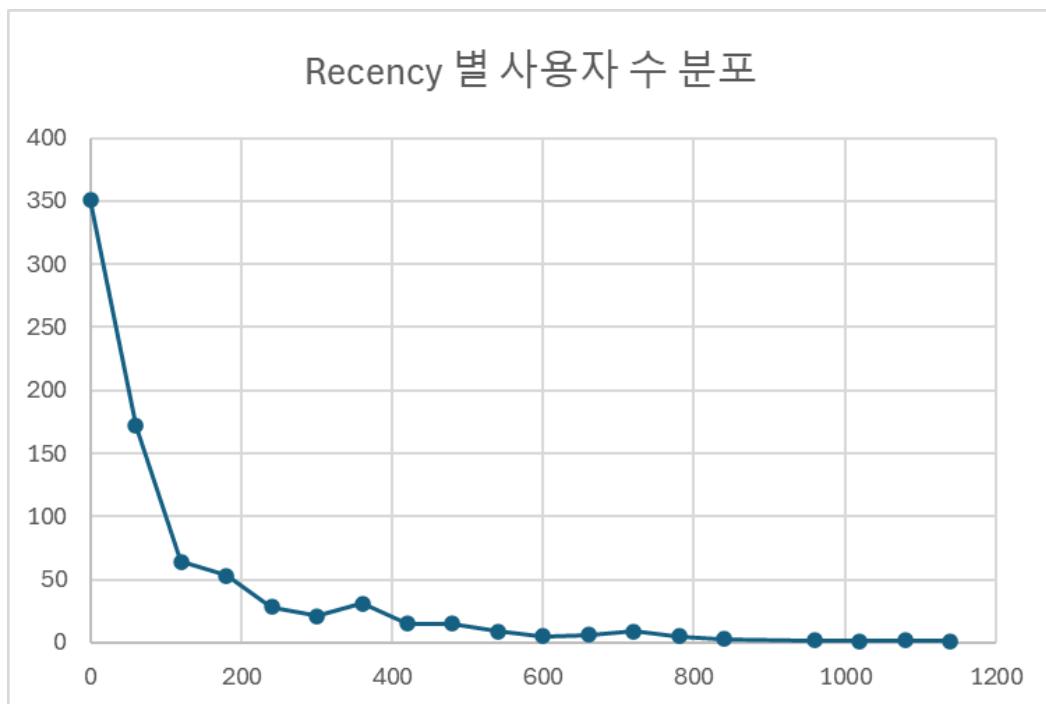
K-means와 같은 클러스터링 알고리즘을 사용하여 고객을 유사한 그룹으로 분류한 후, 각 그룹에 점수를 할당하는 방식이다. 데이터의 군집에 중심점을 임의의 centroid로 선정해 중심점으로부터 떨어진 거리에 따라 점수를 부여하며, 일반적으로는 R, F, M의 3차원 평면 상의 Euclidean distance를 통해 데이터 간 거리를 계산한다.

데이터에 내재된 자연스러운 그룹화가 가능하다는 장점이 있지만 수학적 클러스터링의 결과를 통한 점수화가 비즈니스 적으로도 유의미한 상관관계를 가지기 힘든 경우가 많다.

본 프로젝트에서는 고객을 세분화하여 마케팅 전략을 세우는 것에 RFM 분석의 의의를 두었다.

따라서 Recency, Monetary, Frequency 각각의 점수를 매기는 과정에서 데이터의 분포를 확인한 뒤 그 분포가 비교적 고른 편이라면 분위 기반 점수화, 불균형한 분포를 가진다면 정규분포 기반 점수화를 적용했다.

Recency



X : Recency , Y: Num. of Customer

Mean Value (Recency) : 150.63, SD : 191.53

60일 단위로 Recency를 그룹화해 각 구간마다 고객 수의 분포를 확인한 결과 0~180일 이내에 많은 고객이 모여있는 것을 확인했다.

불균형한 분포를 보정하기 위해 정규분포 기반 점수화를 진행했으며 쿼리 및 결과는 다음과 같다.

```

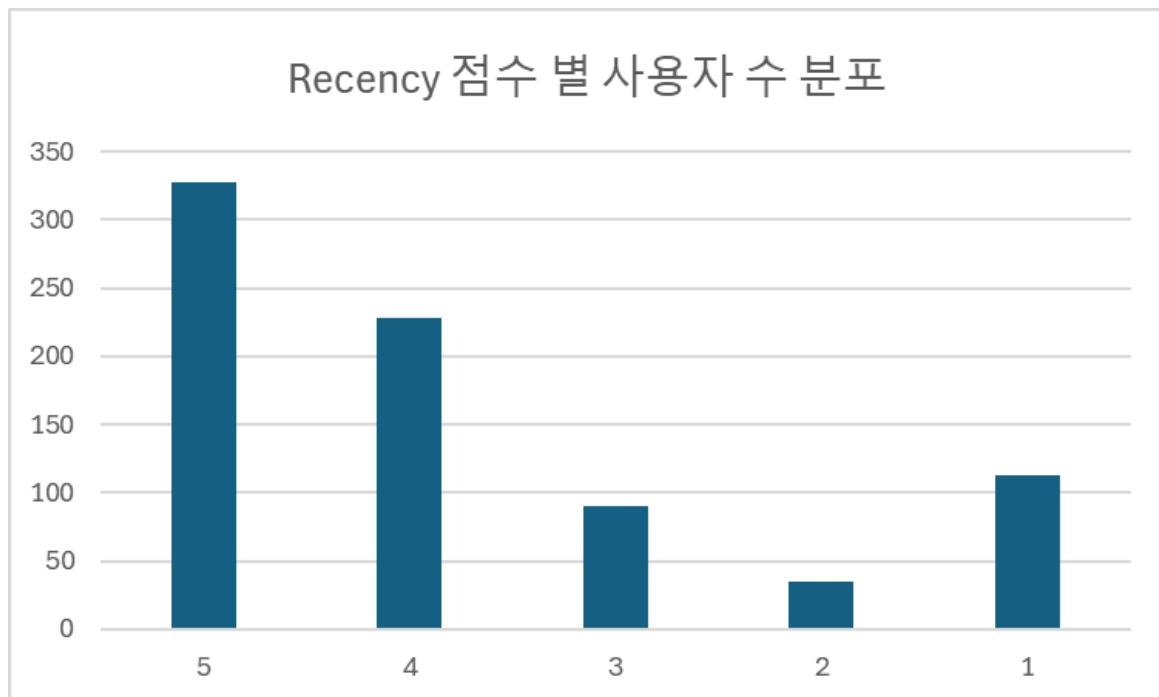
15    SELECT
16        `Customer Name`,
17        Recency,
18        ROUND((Recency - 150.63) / 191.53, 2) AS Z_Score,
19        CASE
20            WHEN ROUND((Recency - 150.63) / 191.53, 2) <= -0.5 THEN 5
21            WHEN ROUND((Recency - 150.63) / 191.53, 2) <= 0 THEN 4
22            WHEN ROUND((Recency - 150.63) / 191.53, 2) <= 0.5 THEN 3
23            WHEN ROUND((Recency - 150.63) / 191.53, 2) <= 1 THEN 2
24            ELSE 1
25        END AS Recency_Score
26    FROM Recency_Calculated;

```

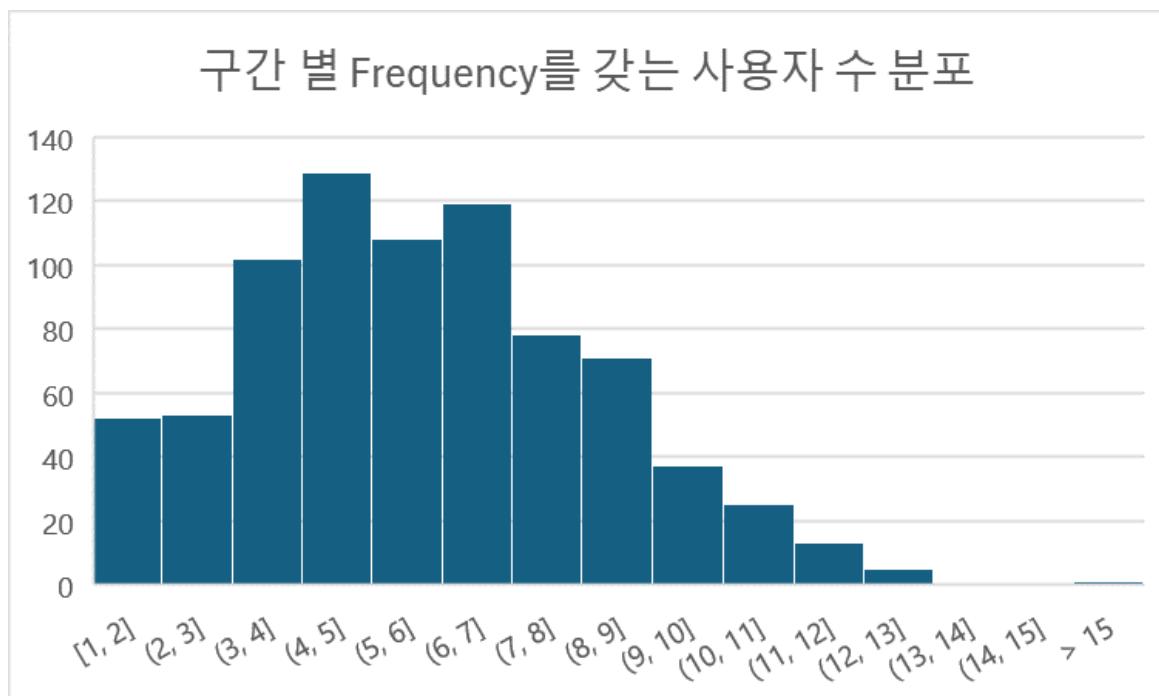
	Customer Name	Recency	Z_Score	Recency_Score
▶	Claire Gute	338	0.98	2
	Darrin Van Huff	19	-0.69	5
	Sean O'Donnell	29	-0.64	5
	Brosina Hoffman	23	-0.67	5
	Andrew Allen	259	0.57	2
	Irene Maddox	390	1.25	1
	Harold Pawlan	30	-0.63	5
	Pete Kriz	9	-0.74	5
	Alejandro Grove	278	0.67	2
	Zuschuss Donatelli	202	0.27	3
	Ken Black	230	0.41	3
	Sandra Flanagan	59	-0.48	4
	Emily Burns	100	-0.26	4

Z-score를 구간 0.5를 기준으로 분류해 -0.5 이하인 경우 최고점 5점, 1 초과인 경우 최하점 1점을 부여했다.

부여받은 점수에 따른 고객 수의 분포는 다음과 같다.



Frequency



2014~2018년 사이에 특정 사용자가 얼마나 많이 상품을 구매했는지를 나타내는 Frequency가 다음 구간 내 값을 가질 때, 그 사용자 수가 얼마나 되는지를 집계한 그래프이다.

Recency와는 구별되게 일정한 구간으로 Frequency를 분할했을 때, 사용자 수가 비교적 고른 분포를 가짐을 확인할 수 있다. 따라서 상위 20% 씩 일정 기준을 잡아 분할하는 분위 기반 점수화를 실시했다.

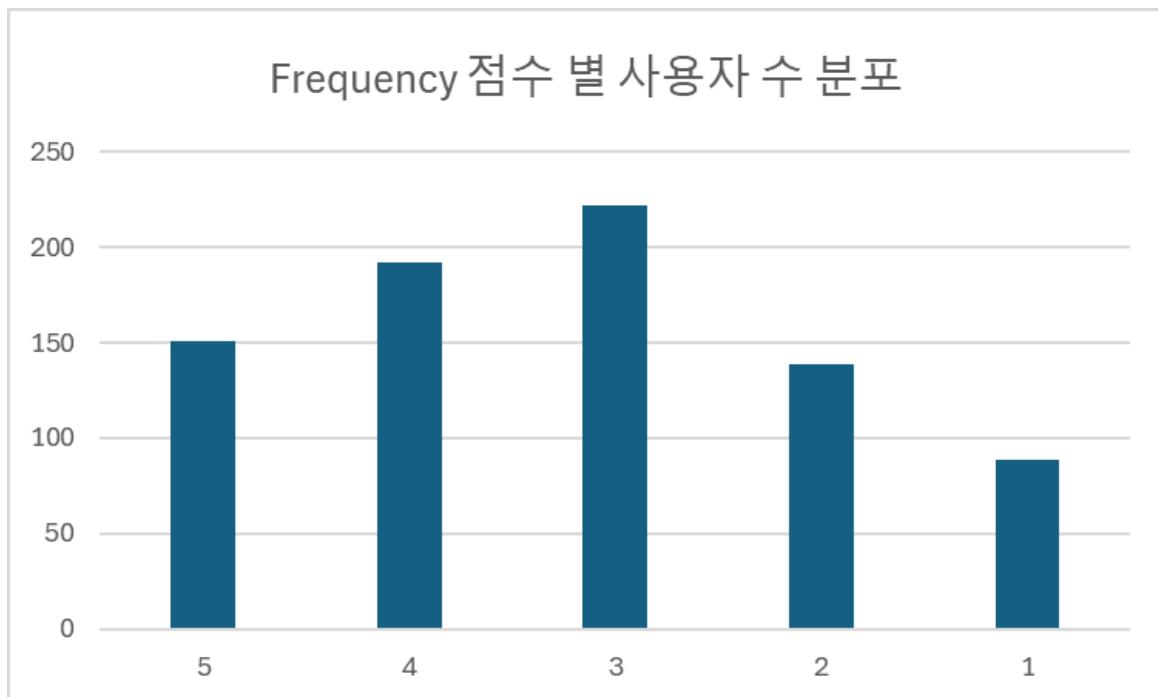
다음은 분위 기반 점수화를 실시한 쿼리 및 결과이다.

```
SELECT
    `Customer Name`,
    Frequency,
    Frequency_Rank,
    CASE
        WHEN Frequency_Rank <= (Total_Customers * 0.2) THEN 5
        WHEN Frequency_Rank <= (Total_Customers * 0.4) THEN 4
        WHEN Frequency_Rank <= (Total_Customers * 0.6) THEN 3
        WHEN Frequency_Rank <= (Total_Customers * 0.8) THEN 2
        ELSE 1
    END AS Frequency_Score
```

	Customer Name	Frequency	Frequency_Score
	Aaron Bergman	3	1
	Aaron Hawkins	7	4
	Aaron Smayling	7	4
	Adam Bellavance	7	4
	Adam Hart	10	5
	Adam Shillingsburg	9	5
	Adrian Barton	9	5
	Adrian Hane	7	4
	Adrian Shami	2	1
	Aimee Bixby	5	3
	Alan Barnes	8	5
	Alan Dominguez	7	4
	Alan Haines	4	2
	Alan Hwang	9	5

각 사용자가 갖는 Frequency 값이 1~18 사이의 작은 분포를 가지므로, NTILE 구문을 이용해 나누게 되면 같은 값의 사용자들이 다른 점수를 부여받는 일이 발생한다. 따라서 Rank를 통해 같은 값을 가진 사용자들은 적어도 같은 점수를 부여받도록 쿼리를 구성했다.

부여받은 점수에 따른 고객 수의 분포는 다음과 같다.



Monetary

Monetary_Value를 Customer ID를 기준으로 지출한 Sales의 총 합계로 정의했고, NTILE 구문을 통해 상위 20% 씩 분류해 5점부터 1점까지 부여했다.

```

26   WITH Monetary_Calculated AS (
27     SELECT
28       `Customer ID`,
29       SUM(`Sales`) AS Monetary_Value
30     FROM
31       super
32     GROUP BY
33       `Customer ID`
34   )
35   SELECT
36     `Customer ID`,
37     Monetary_Value,
38     NTILE(5) OVER (ORDER BY Monetary_Value DESC) AS Monetary_Score
39   FROM
40     Monetary_Calculated
41   ORDER BY
42     `Customer ID`;

```

Customer ID	Monetary_Value	Monetary_Score
AA-10315	5563.56	1
AA-10375	1056.39	4
AA-10480	1782.872	3
AA-10645	4455.895	1
AB-10015	886.156	5
AB-10060	7732.669999999999	1
AB-10105	14355.610999999997	1
AB-10150	966.709999999998	4
AB-10165	1113.838	4
AB-10255	898.084	5
AB-10600	1515.862	4
AC-10420	891.56	5
AC-10450	5527.8460000000005	1

RFM Segment

결론적으로 RFM 점수 산정 기준은 다음 랭크 테이블과 같다.

Rank	R	F	M
5	$Z_R \leq -0.5$	18회 이상	상위 20% 이내
4	$-0.5 < Z_R \leq 0$	13회 이상	상위 40% 이내
3	$0 < Z_R \leq 0.5$	9회 이상	상위 60% 이내
2	$0.5 < Z_R \leq 1$	6회 이상	상위 80% 이내
1	$1 < Z_R$	1회 이상	상위 100% 이내

R, F, M 각 항목마다 4~5점을 청신호 및 현상 유지, 1~3점을 개선 필요로 가정했을 때, 고객을 등급별로 나는 RFM Segment Table은 다음과 같이 구성할 수 있다.

	특징	R	F	M	cnt(고객 수)	매출 총액(\$)
VIP	최근에, 자주, 많은 금액을 지출한 고객	4~5	4~5	4~5	38	199968.03
VIP 잠재 고객(1)	지출액이 적으나 최근에, 자주 구매한 고객 (Monetary 개선 시 VIP 잠재성)	4~5	4~5	1~3	134	151423.64
VIP 잠재 고객(2)	자주 구매하지 않았으나 최근에, 많은 금액을 소비한 고객 (Frequency 개선 시 VIP 잠재성)	4~5	1~3	4~5	212	1066173.06
떠나간 VIP	자주, 많은 금액을 지출했지만 최근에 구매하지 않은 고객	1~3	4~5	4~5	17	93128.56
관심 필요 고객	모든 지표가 평균 이상인 고객	4 3 3	3 4 3	3 4 4	30	52632.01
신규 고객	최근에 왔으나 자주, 많은 금액을 지출하지 않은 고객	4~5	1~2	1~2	42	50861.98
이탈 고객	세 가지 지표 모두 낮은 고객	1~3	1~3	1~3	45	75081.35
기타 고객	뚜렷한 지표가 없는 나머지 고객				275	583181.22

CASE

```

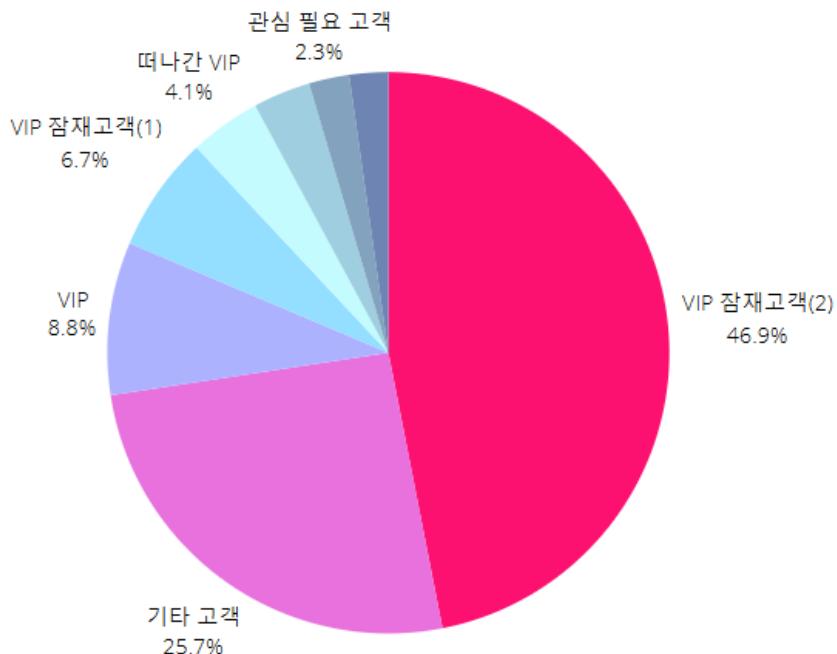
WHEN r.`Recency_Score` >= 4 AND f.`Frequency_Score` >= 4 AND m.`Monetary_Score` >= 4 THEN 'VIP'
WHEN r.`Recency_Score` >= 4 AND f.`Frequency_Score` >= 4 AND m.`Monetary_Score` <= 3 THEN 'VIP 잠재 고객(1)'
WHEN r.`Recency_Score` >= 4 AND f.`Frequency_Score` <= 3 AND m.`Monetary_Score` >= 4 THEN 'VIP 잠재 고객(2)'
WHEN r.`Recency_Score` <= 3 AND f.`Frequency_Score` >= 4 AND m.`Monetary_Score` >= 4 THEN '떠나간 VIP'
WHEN r.`Recency_Score` >= 4 AND f.`Frequency_Score` <= 2 AND m.`Monetary_Score` <= 2 THEN '신규 고객'
WHEN r.`Recency_Score` <= 3 AND f.`Frequency_Score` <= 3 AND m.`Monetary_Score` <= 3 THEN '이탈 고객'
WHEN r.`Recency_Score` = 4 AND f.`Frequency_Score` = 3 AND m.`Monetary_Score` = 3 THEN '관심 필요 고객'
WHEN r.`Recency_Score` = 3 AND f.`Frequency_Score` = 4 AND m.`Monetary_Score` = 3 THEN '관심 필요 고객'
WHEN r.`Recency_Score` = 3 AND f.`Frequency_Score` = 3 AND m.`Monetary_Score` = 4 THEN '관심 필요 고객'
ELSE '기타 고객'

```

산정된 점수를 기반으로 고객별 Segment를 분류할 때 위와 같이 CASE문을 활용해 분류했다.

위의 RFM Segment Table을 참고해,

최종적으로 전체 매출액에서 고객 세그먼트 별 매출 총액이 차지하는 비율을 시각화하면 다음과 같다.



RFM Segment 별 매출액 차지 비율

(2-4) 마케팅 전략 및 결론

분석 결과, VIP 잠재 고객(2) 가 전체 매출액 대비 46.9%를 차지하는 중요 segment임을 확인했다. VIP 잠재 고객(2)의 경우에는 자주 구매하진 않았으나 최근에 활발하게 많은 지출액을 소비하는 고객으로 향후 서비스 이용 빈도를 증가 시키면 VIP 고객으로의 잠재성이 큰 고객이다. 따라서 해당 segment 고객들을 대상으로 꾸준하게 서비스 이용 빈도를 높일 수 있는 프리미엄 멤버십 프로그램과 같은 회원제 프로그램에 가입하도록 유도하고, 가입 시 추가 혜택을 제공하는 방식이 효과적일 것으로 예상된다. 또한 해당 segment 고객들이 많이 구입하는 제품 카테고리에서 적극적인 1+1 프로모션 및 할인행사를 통해 계속적인 구매를 유도하는 것이 매출액을 극대화할 수 있는 방안이 될 것이다.

(3) 고객생애가치 분석

(3-1) 고객생애가치(Customer Lifetime Value; CLTV, CLV, LTV) 분석이란?

- 고객 생애가치 분석은 한 고객이 특정 기업과의 관계를 유지하는 동안 발생할 것으로 예상되는 총 수익을 측정하는 방법입니다. 이 분석은 고객이 기업에 얼마나 가치를 제공할 수 있는지를 평가하며, 고객의 장기적인 수익성을 예측하는 데 중요한 역할을 합니다. 이를 통해 기업은 고객 확보와 유지에 들어가는 비용을 최적화하고, 가장 가치 있는 고객층에 집중할 수 있습니다.

(3-2) 분석의 의의

- 고객 생애가치 분석은 고객 관리 전략에서 중요한 의의를 지닙니다. 이를 통해 기업은 고객 세그먼트를 보다 효과적으로 식별하고, 높은 생애가치를 가진 고객에게 집중적인 마케팅과 서비스 제공을 할 수 있습니다. 또한, 고객 유지율을 높이기 위한 맞춤형 전략을 수립하여 고객 이탈을 방지하고, 장기적으로 안정적인 수익을 창출하는 데 기여합니다.

니다. 궁극적으로, 고객 생애가치 분석은 기업이 자원을 효율적으로 배분하고, 마케팅 활동의 ROI를 극대화할 수 있도록 돋는 중요한 도구입니다.

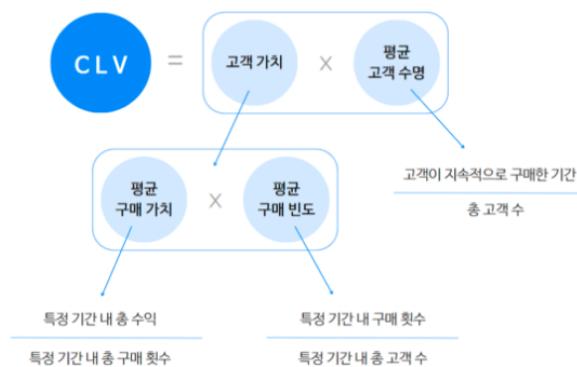
(3-3) 고객 생애가치 분석

[출처] <https://groobee.net/blog/고객-생애-가치/#:~:text=고객%20생애%20가치>

고객 생애 가치 = 고객 가치 X 평균 고객 수명

간단한 수식이지만 각 항목을 계산하기 위해서는 다음과 같은 과정을 거쳐야 합니다.

- 고객 가치 (=평균 구매 가치 X 평균 구매 빈도)
- 평균 구매 가치 (=특정 기간 내 총 수익 / 특정 기간 내 총 구매 횟수)
- 평균 구매 빈도 (=특정 기간 내 구매 횟수 / 특정 기간 내 총 고객 수)
- 평균 고객 수명 (=고객이 지속적으로 구매한 기간 / 총 고객 수)



고객 가치 = 평균 구매 가치 * 평균 구매 빈도

- excel link : <https://docs.google.com/spreadsheets/d/1qGSaPXeKTbxj3aWH4goB-E3iK2oANVVdr6bLG7svnA/edit?usp=sharing>

- 평균 구매 가치 = 특정기간 내 고객이 주문한 총 매출액/특정기간 내 총 구매 횟수

SQL query :

```
-- 평균 구매 가치 = 특정기간 내 고객이 주문한 총 매출액/특정기간 내 총 구매 횟수(order_id > 0)
select customer_id,
       customer_name,
       round(sum(sales), 2) '총 구매액',
       count(distinct(order_id)) '총 구매 횟수',
       round(sum(sales)/count(distinct(order_id)), 2) '평균 구매 가치'
from superstore
group by 1, 2
order by 5 desc;
```

- 평균 구매 빈도 = 특정기간 내 구매 횟수 / 특정기간 내 총 고객 수

SQL query :

```
-- 평균 구매 빈도 = 특정기간 내 구매 횟수 / 특정기간 내 총 고객수  
select count(distinct order_id) '구매 횟수',  
       count(distinct customer_id) '총 고객수',  
       round(count(distinct order_id)/count(distinct customer_id), 1) '평균 구매 빈도'  
from superstore;
```

평균 고객 수명 = 고객이 지속적으로 구매한 기간 / 총 고객 수

지속적 구매 총 기간 / 총 고객 수 = $25361 / 793 = 31.98$

- 고객이 지속적으로 구매한 총 기간 : 25361 (month)

SQL query :

```
-- 고객이 지속적으로 구매한 총 기간  
select sum(period.period_buy) '총 구매 기간'  
from (  
    select customer_id,  
           customer_name,  
           TIMESTAMPDIFF(MONTH, min(str_to_date(order_date, '%m/%d/%Y')), max(str_to_date(order_date, '%m/%d/%Y'))) as period_buy  
    from superstore  
    group by 1, 2  
    order by 1, 2  
) as period;
```

- 총 고객 수 : 793 (명)

(3-4) 고객 생애 가치 분석 산출물

- excel link : https://docs.google.com/spreadsheets/d/1tOz3JL7tLHB_tr8vK8-unYv8iRFqx0xBtuk5UsOyM/edit?usp=sharing

3.3 카테고리 별 재고 대비 예측

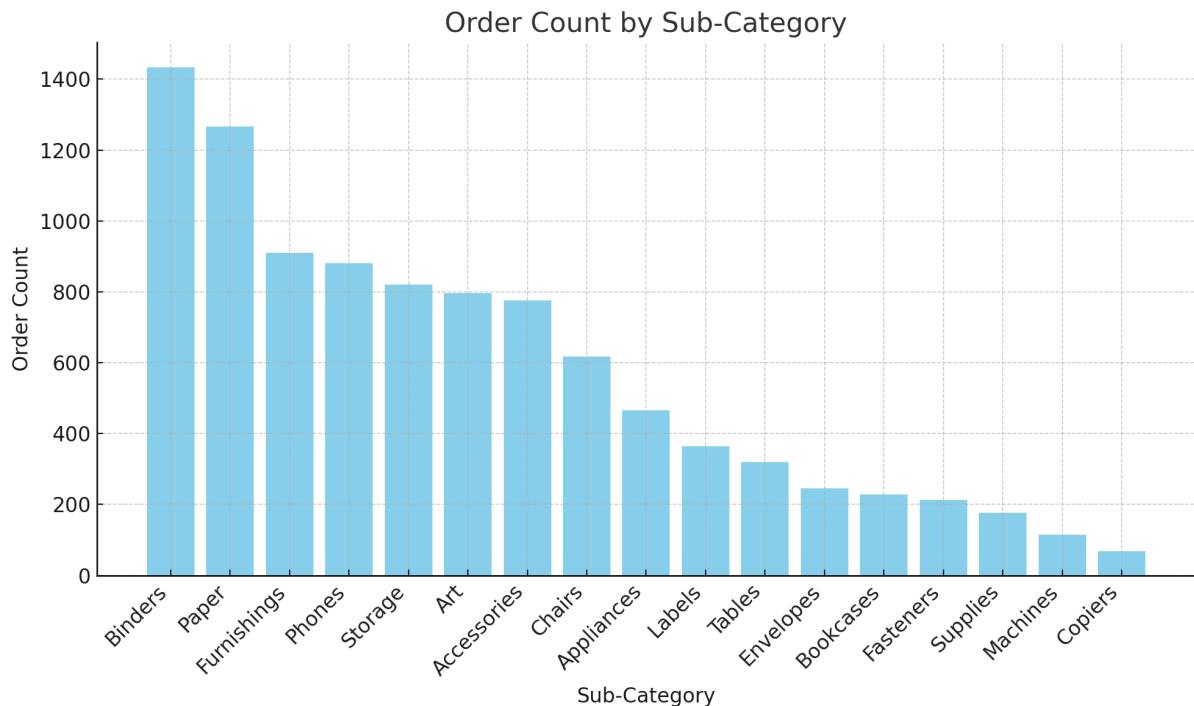
(1) 재고 예측

재고 예측은 기업의 재고 관리에서 필수적으로 고려해야 할 요소이다. 이를 효과적으로 관리하지 않으면 불필요한 재고 보관 비용, 감가 상각, 판매하지 못했을 때의 재고 손실 등의 문제가 일어날 수 있다. 또한 이 물품을 어느 지역군에 어느 분기에 어느 비용수준까지 재고로 남겨야 할지를 최적화 함으로써 재고손실을 효과적으로 관리할 수 있다.

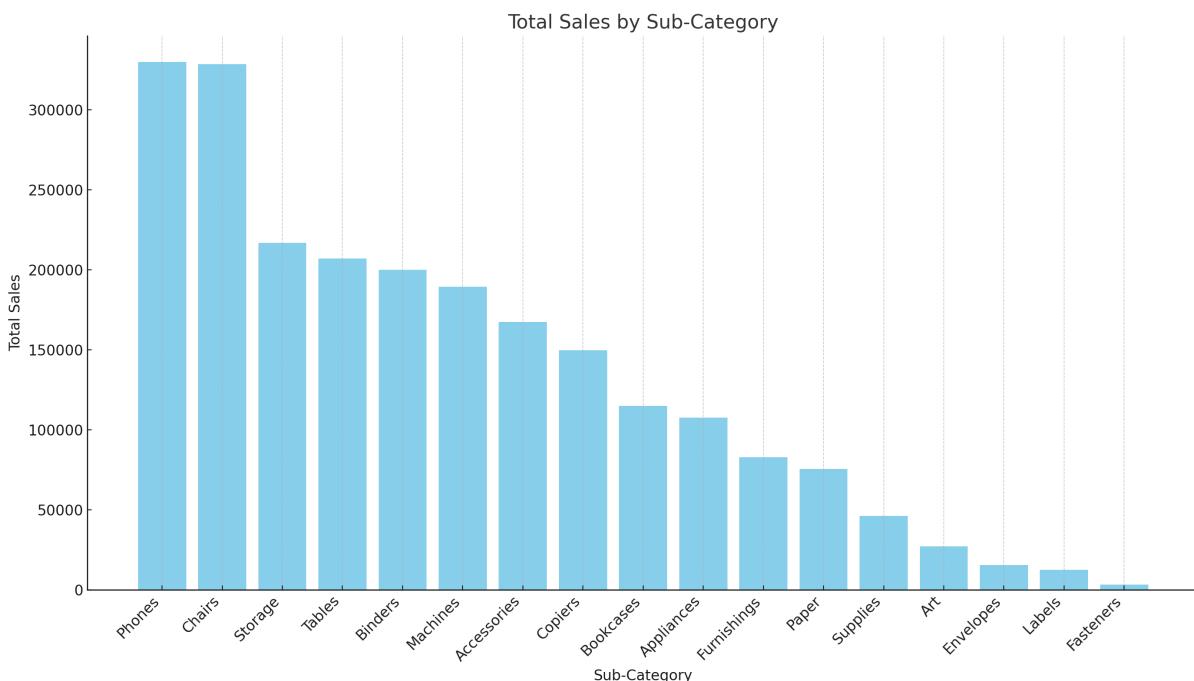
(2) EDA

(2-1) 총 매출(sales) 및 물량(counts) 분석

카테고리별 물품 현황을 통해 대략적인 총 판매 수를 볼 수 있다.



다만 판매를 보았을 때, 최대 count 는 binders 지만, 총 sales 는 phones가 차지한 것을 볼 수 있다.
따라서, sales 와 count 를 모두 고려하는 방향이 좋을 것이다.

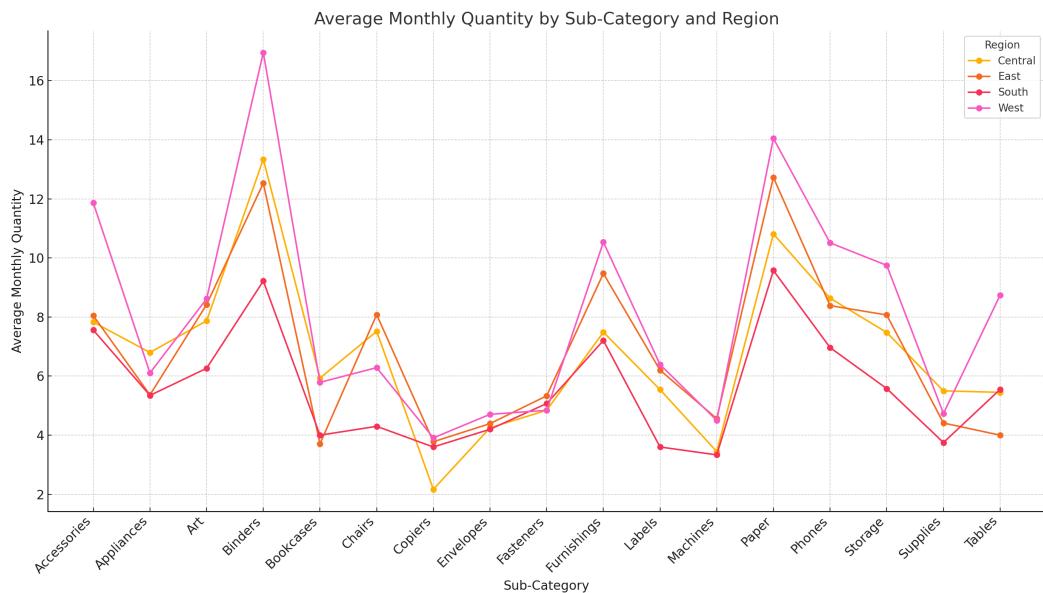


슈퍼스토어의 운영 특성상, 고객에게 제품을 직접 배송하기 위해 운송 회사를 통해 물품을 전달한다. 하나의 지점에서 모든 지역으로 물품을 배송하는 것은 비효율적이며, 물류 과정의 복잡성을 크게 증가시킨다.

따라서, 각 지역별로 재고 창고를 별도로 운영하여 물류 과정을 최적화한다. 지역별 판매 데이터를 분석하여 4개의 주요 지역으로 구분하고, 각 지역의 수요에 맞춰 물류 자원을 차등 배분함으로써 효율적인 배송 시스템을 구축한다.

(2-2) 지역 별로 물량 추세 확인

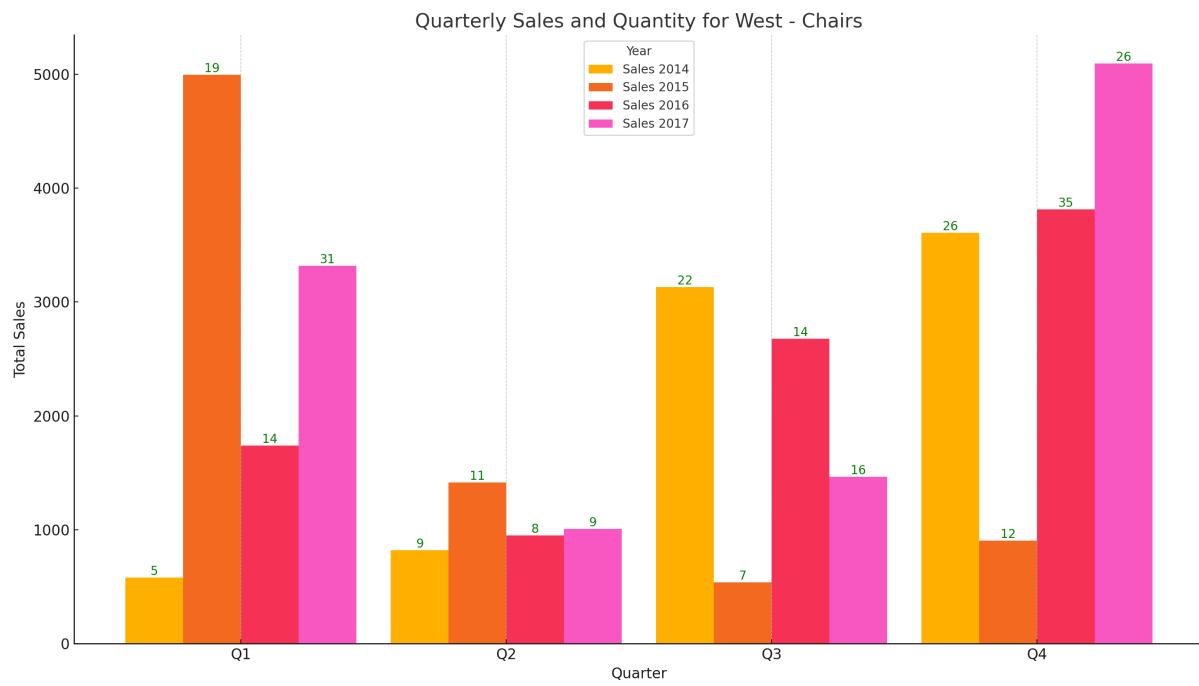
다양한 지역에서 각 소분류 제품의 연도별 구매 수량을 나타낸다. 각 지역은 서로 다른 도시들을 포함하며, 특정 도시에서 어떤 제품군이 더 많이 판매 되었는지를 시각적으로 비교할 수 있다.



그래프에서 볼 수 있듯이, 지역과 카테고리에 따라 매출과 수량이 많이 차이 나는 것을 볼 수 있고, 분석에도 이를 반영 해야 한다.

(2-3) 분기 별/ 지역 별 매출과 물량 분석

특정 분기에 대한 매출 및 물품 수량 데이터를 기반으로, 향후 수요를 예측하고 적절한 재고 수준을 유지하는 데 도움을 준다. 특히, 분기 별 수량 표시를 통해 과거 데이터와 비교하여 재고 부족이나 과잉 재고 문제를 미리 예방할 수 있다.



이 그래프는 “West” 지역에서 “Chairs” 카테고리의 연도별 분기 매출 물품 수량이다.

특정 분기에서의 성과 변화를 통해 시즌성 패턴을 보이고 있으며, 분기 별로 다른 재고 전략을 사용해야 함을 알 수 있다.

[Quarterly_Sub-Category_Growth_Rates_by_Region_with_Quantity.csv](#)

(3) 안전재고량(Safety Stock) 및 안전재고가치(Safety Stock Value) 분석

슈퍼스토어의 각 지역과 카테고리에 대한 분기별 매출 및 재고 관리의 최적화를 위해, 안전재고와 그 가치를 계산하였다. 이를 통해 적절한 재고 수준을 유지하고, 재고 부족으로 인한 문제를 방지하는 것을 목표로 한다.

안전재고량(Safety Stock) 계산

안전재고는 판매 수량의 평균 및 변동성을 기반으로 하여 계산한다. 안전재고는 예상 외의 수요 변동에 대응하기 위해 필요한 최소한의 재고 수준을 의미한다.

안전재고량 공식:

- Safety Stock=평균 판매 수량 (Avg_Quantity)+(1.65×판매 수량의 표준 편차 (Std_Quantity))
- 평균 판매 수량(Avg_Quantity): 해당 지역 및 소분류에서 분기별로 판매된 평균 물품 수량이다.
- 판매 수량의 표준 편차(Std_Quantity): 판매 수량의 변동성을 나타내는 값으로, 과거 분기별 수량 데이터의 표준 편차로 계산된다.
- 1.65: 이는 90% 서비스 수준을 의미하는 z-값이다. 즉, 90%의 확률로 재고가 부족하지 않도록 하는 기준이다.

안전재고의 계산 과정:

안전재고는 평균 판매 수량에 1.65배의 표준 편차를 더하여 계산한다. 이 값은 과거 데이터를 기반으로 변동성을 고려한 재고 수준을 제시하여, 예기치 않은 수요 급증에 대응할 수 있도록 한다.

안전재고 가치(Safety Stock Value) 계산

안전재고 가치는 안전재고의 매출 가치를 의미하며, 안전재고 수량에 해당 소분류의 평균 단가를 곱하여 계산된다. 이를 통해, 재고가 실제로 얼마나 큰 금전적 가치를 가지는지 평가할 수 있다.

안전재고 가치 공식:

- Safety Stock Value = 안전재고(Safety Stock) × 평균 단가(Avg_Unit_Price)
- Safety Stock Value = 안전재고(Safety Stock) × 평균 단가(Avg_Unit_Price)
- 평균 단가(Avg_Unit_Price): 해당 소분류의 평균 판매 단가로, 분기별 총 매출을 총 판매 수량으로 나누어 계산된다.

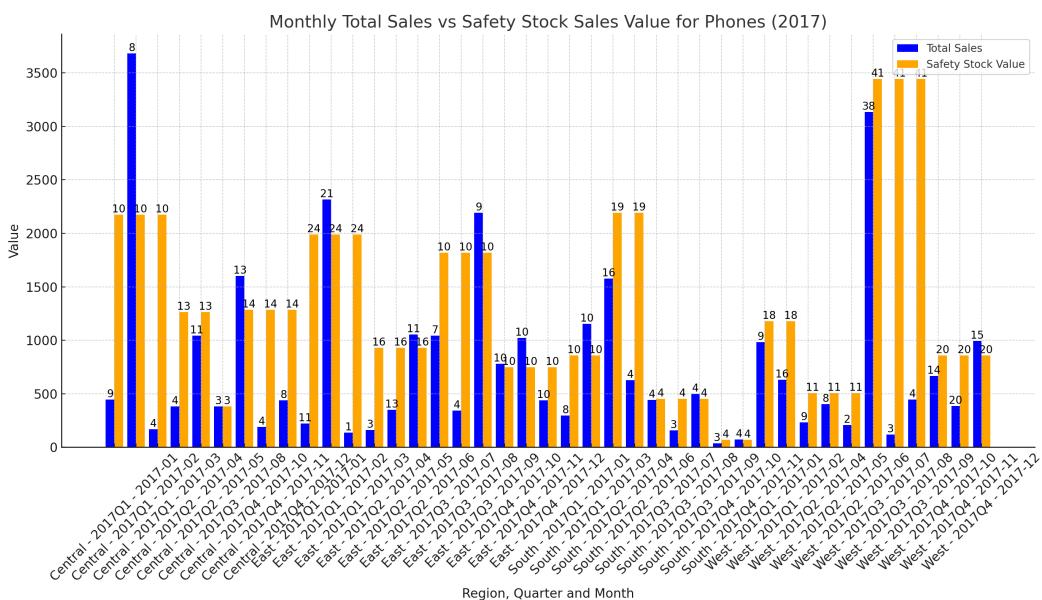
안전재고 가치의 계산 과정:

- 안전재고 수량을 해당 소분류의 평균 단가와 곱하여 계산하며, 이는 유지해야 할 재고의 금전적 가치를 나타낸다. 이를 통해, 재고 관리 시 필요한 비용을 더 정확하게 추정할 수 있다.

안전재고량 와 안전재고가치 반영

안전재고량과 안전재고가치를 실제 2014~2017년 데이터에 반영해 어느 수준으로 분기별로 재고를 준비했어야 하는지 알아볼 수 있다.

[WITH_Monthly_Sales_AS_SELECT_Region_Sub_Category_DATE_FORMAT_ST_202408301651.csv](#)

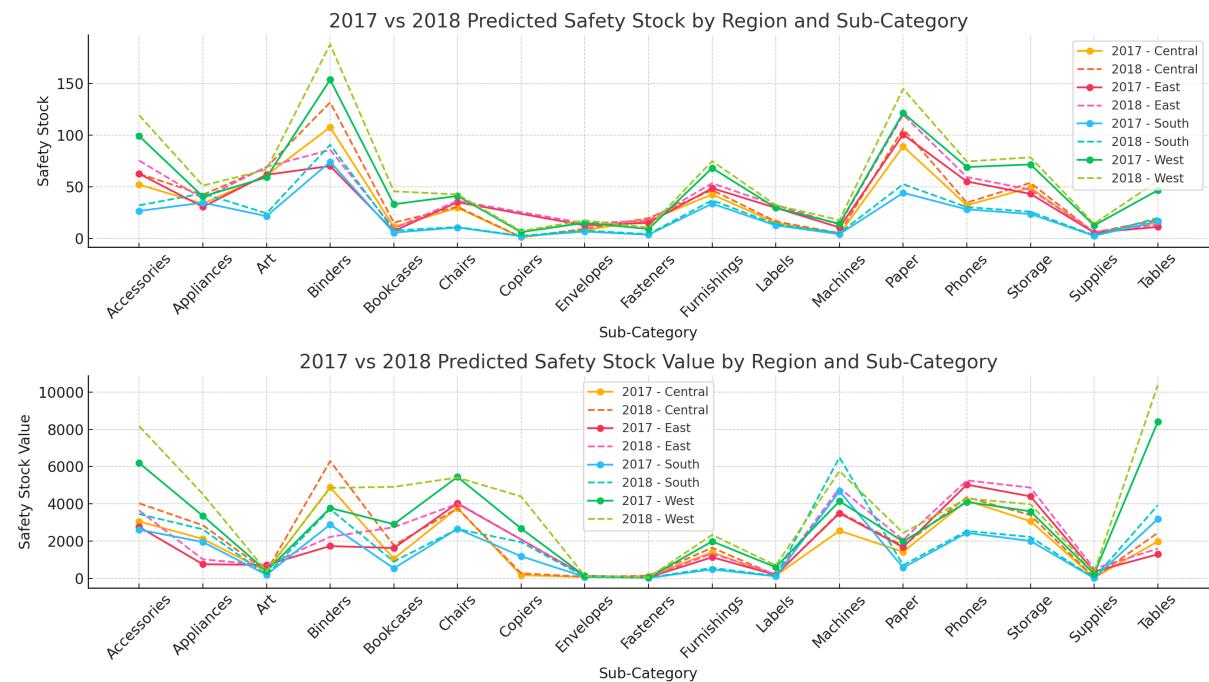


2014~2017년 성장률을 2018년 예측 재고에 반영

이제 2014~2017년 데이터를 활용해 안전재고량(counts)와 안전재고가치(sales)를 추측했으므로, 2014년~2017년의 성장률을 바탕으로 2018년의 안전재고량, 가치를 예측해 보도록 하겠다.

2014 ~ 2017 년의 성장률은

예를 들어, "Accessories"의 경우 2017년에 비해 안전 재고 counts 는 약 20.37% 성장할 것으로, 안전 재고 values는 약 32.01% 성장할 것으로 예측되었다.



각 카테고리와 지역별로 상이한 성장세를 가지고 있음을 알 수 있다.

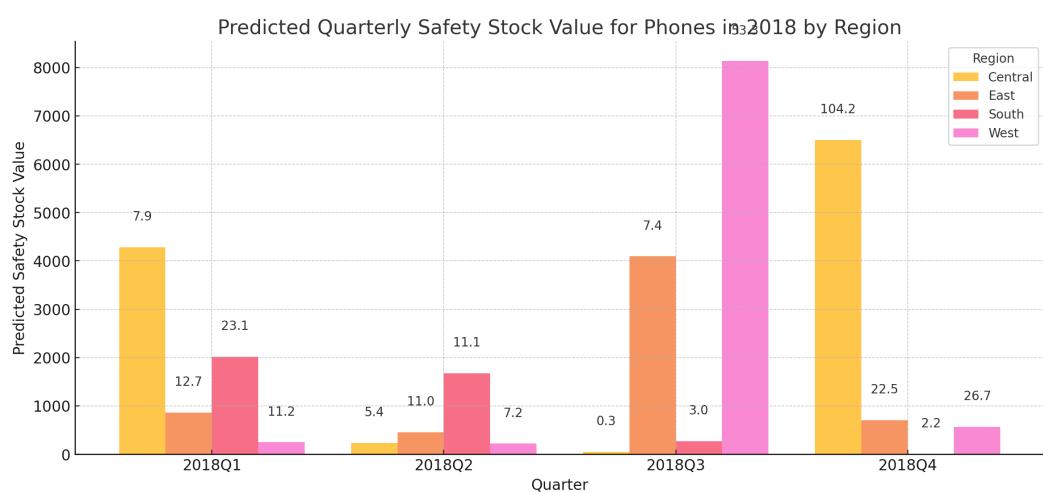
때문에, [카테고리/지역/분기] 의 분류로 별도의 성장을 추세를 예측한 후에

이를 바탕으로 2018년에 확보해야 하는 안전재고 물량 및 가치를 예측해 미리 대비할 수 있다.

우선 2014~2017년을 분기별, 지역별, 카테고리별로 성장률을 따로 계산해 볼 수 있다.

이 성장률을 2018년 예측 안전재고, 안전재고가치에 반영하면 다음과 같다.

Quarter	Region	Sub-Category	Predicted_Safety_Stock_2018	Predicted_Safety_Stock_Value_2018
2018Q1	Central	Accessories	11.03	1494.97
2018Q1	Central	Appliances	114.08	39712.28
2018Q1	Central	Art	22.62	101.54
2018Q1	Central	Binders	40.33	14369.9
2018Q1	Central	Bookcases	1	49.17
2018Q1	Central	Chairs	24.5	2821.35
2018Q1	Central	Envelopes	1.04	1.02
2018Q1	Central	Fasteners	25.6	271.28
2018Q1	Central	Furnishings	15.68	804.34
2018Q1	Central	Labels	2.25	18.6
2018Q1	Central	Machines	6	2058.49
2018Q1	Central	Paper	44.31	1407.72
2018Q1	Central	Phones	7.88	4285.95
2018Q1	Central	Storage	32.14	4376.4
2018Q1	Central	Supplies	2.67	1.18
2018Q1	Central	Tables	1.14	140
2018Q1	East	Accessories	13.09	642.08
2018Q1	East	Appliances	7.53	59.42
2018Q1	East	Art	24.14	60.27
2018Q1	East	Binders	19.56	344.26
2018Q1	East	Bookcases	2	552.45
2018Q1	East	Chairs	4.02	1048.33
2018Q1	East	Envelopes	7.03	29.05
2018Q1	East	Fasteners	4.65	13.55
2018Q1	East	Furnishings	11.84	419.72
2018Q1	East	Labels	10.67	45.24
2018Q1	East	Machines	48.17	4650.78
2018Q1	East	Paper	82.57	2840.36
2018Q1	East	Phones	12.66	864.08
2018Q1	East	Storage	2.28	82.9
2018Q1	East	Supplies	0.73	1186.35
2018Q1	East	Total	71.1	864.35



이를 위에서 사용한 안전재고 및 안전재고가치 계산을 통해
2018년에 얼마나 물품 확보를 해야 할지 예측 해볼 수 있다.

Updated_Predicted_Quarterly_Safety_Stock_and_Values_for_2018.csv

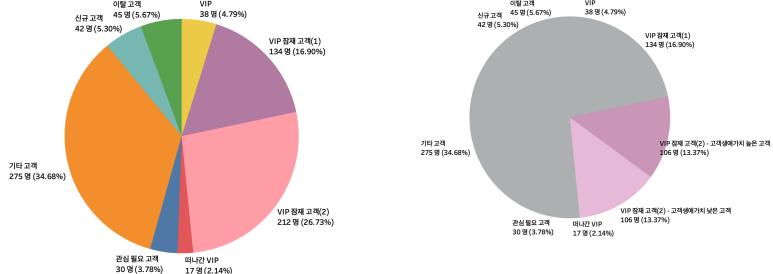
4 . 프로젝트 결론

본 프로젝트는 매출 내역을 활용한 고객 관리 및 재고 관리의 효율성 극대화를 목표로 진행되었다. 이를 위해 고객의 구매 패턴과 재고 수요를 심층 분석하여 최적의 관리 전략을 수립하고자 하였다.

첫 번째로, 고객 관리 측면에서는 다양한 분석 기법이 활용되었다. 먼저 EDA 분석을 통해 고객의 재구매가 3개월(전체 주문 건수의 약 38.26%), 6개월(약 22.73%), 9개월(약 14.48%), 12개월(약 9.44%) 단위로 유형화 할 수 있다고 판단하였다. 이를 통해 3개월 단위의 마케팅 전략을 세우고자 했다. 그 이후 3개월 단위의 코호트 분석을 진행한 결과, 평균적으로 약 34.9% 수준으로 재구매율이 유지된다는 것을 확인할 수 있었다. 일반적으로 코호트 분석을 했을 때, 시간이 지남에 따라 고객 이탈률이 커지는 데 본 기업의 매출 데이터를 토대로 진행한 코호트 분석에서는 일정 기간이 지난 후, 재구매율이 상승한다는 점이였다. EDA와 코호트 분석을 통해 파악할 수 있었던 것은 본 기업이 파는 제품(가전제품, 가구, 사무용품)의 특성 상 자주 구매하는 고객보다는 시간이 지난 후, 재구매가 이루어진다는 것을 확인할 수 있었다. 즉, 짧은 주기로 재주문을 하진 않지만 한 번 유입이 된다면 장기고객이 되는 경우가 많다는 것을 확인할 수 있었다.

전반적인 고객 관리 측면에서의 전략

- 코호트 분석에 따른 결론 → 3개 월 단위 마케팅 진행
- RFM 분석에 따른 결론 → 2014년~2017년까지 구매내역이 있는 고객들을 7개 단계로 나눔
- VIP 잠재고객(2)(46.9%)를 차지하므로 VIP 잠재고객(2) 중에서 더 가치 높은 고객을 선별하기 위해 고객생애가치 분석을 진행 함



RFM_Customer_ID_Segme nt12.csv

고객 그룹 별 상세 마케팅 전략

번호	고객 그룹	고객수(명) / 비율(%)	마케팅 전략
1	VIP 고객	38명 / 4.79%	<p>목표: 높은 고객 만족도 유지 및 추가 구매 유도</p> <ul style="list-style-type: none"> 고급 이벤트 초대: VIP 전용 행사나 이벤트를 개최하여 고객 로열티를 강화
2	VIP 잠재고객(1)	136명 / 16.90%	<p>목표: VIP로의 전환 유도</p> <ul style="list-style-type: none"> 맞춤형 혜택 제공: VIP로 승격될 가능성이 높은 고객에게는 맞춤형 혜택이나 보너스를 제공하여 그들의 구매를 촉진 고객 경험 강화: VIP 경험에 한 발짝 다가갈 수 있도록 개인화된 서비스와 지원을 제공
3	VIP 잠재고객(2) - 고객생애가치가 높은 고객	106명 / 13.37%	<p>목표: 고객 생애가치(LTV)를 극대화하고 VIP로 승격 유도</p> <ul style="list-style-type: none"> 장기 혜택 프로그램: 장기적인 관계를 구축할 수 있는 로열티 프로그램을 제안. 예를 들어, 누적 구매액에 따른 리워드를 제공하여 추가 구매 유도 프리미엄 콘텐츠 제공: 이 그룹에게는 고가의 프리미엄 제품이나 서비스를 소개하고, 이를 체험할 수 있는 기회 제공
4	VIP 잠재고객(2) - 고객생애가치가 낮은 고객	106명 / 13.37%	<p>목표: 구매 빈도와 금액 증대</p> <ul style="list-style-type: none"> 단기적인 혜택 제공: 이 그룹에게는 단기적으로 매력적인 프로모션을 제공하여 즉각적인 구매 유도 리마인더 및 추천: 과거 구매 데이터를 바탕으로 맞춤형 리마인더 이메일이나 푸시 알림을 보내어, 그들이 관심을 가질 만한 제품을 추천
5	떠나간 VIP	17명 / 2.14%	<p>목표: 고객 재유치 및 관계 회복</p> <ul style="list-style-type: none"> 이탈 원인 분석: 고객이 떠난 이유를 분석하여, 이를 해결할 수 있는 전략 수립. 예를 들어, 불만사항이 있었다면 이를 개선하고 그 내용을 공유 특별한 재유치 혜택 제공: 재방문 시 특별한 혜택을 제공하거나, 그들이 관심을 가질 만한 새로운 제품이나 서비스를 소개
6	관심 필요 고객	30명 / 3.78%	<p>목표: 관계 강화 및 추가 구매 유도</p> <ul style="list-style-type: none"> 주기적인 커뮤니케이션: 뉴스레터나 이메일 마케팅을 통해 새로운 제품, 프로모션, 이벤트 등을 지속적으로 알림 개인화된 메시지 발송: 고객에게 직접적인 개인화 메시지를 보내어, 그들이 다시 돌아오도록 유도
7	신규고객	42명 / 5.30%	<p>목표: 브랜드와의 첫 경험을 긍정적으로 만들고, 반복 구매를 유도하여 충성 고객으로 전환</p> <ul style="list-style-type: none"> 웰컴 캠페인: 신규 고객에게는 첫 구매를 기념하는 웰컴 이메일을 발송하고, 첫 구매 감사 혜택(예: 할인 쿠폰, 무료 배송, 무료 체험 기회) 제공 첫 구매 후 추가 구매 유도: 첫 구매 후 일정 기간 내에 사용할 수 있는 추가 할인 혜택이나 프로모션을 제공하여, 빠르게 두 번째 구매로 이어지도록 유도

번호	고객 그룹	고객수(명) / 비율(%)	마케팅 전략
8	이탈고객	45명 / 5.67%	<p>목표: 고객 재유치 및 관계 회복</p> <ul style="list-style-type: none"> 개인화된 연락: 이메일, 문자 메시지 또는 전화 등을 통해 개인화된 메시지를 보내, 그들이 돌아오기를 환영하는 마음 전달. 예를 들어, “당신을 그리워하고 있습니다”라는 메시지와 함께 돌아오면 받을 수 있는 혜택 소개 이탈 원인 분석 및 해결: 이탈 고객의 행동 데이터를 분석하여, 이탈 원인을 파악하고 그에 대한 해결책 제시. 예를 들어, 고객 서비스에 대한 불만이 있었다면, 이를 해결하고 개선된 서비스 강조
9	기타 고객	275명 (34.68%)	<p>목표: 관심도와 참여도 향상</p> <ul style="list-style-type: none"> 기초적인 혜택 제공: 이 고객군에게는 기본적인 할인 혜택이나 적립금을 제공하여, 브랜드에 대한 관심 유도 자동화된 마케팅: 이 그룹에는 자동화된 마케팅 캠페인을 통해, 정기적으로 정보를 전달하고 구매 유도 리마인더 및 추천: 과거 구매 데이터를 바탕으로 맞춤형 리마인더 이메일이나 푸시 알림을 보내어, 그들이 관심을 가질 만한 제품을 추천

두 번째로, 카테고리별 분석 결과, 판매량과 매출의 불일치를 확인할 수 있었다. 특히, 특정 제품군에 대해 판매량이 많지만 매출 기여도가 낮은 경우와 반대로 판매량은 적지만 매출 기여도가 높은 경우 각각에 맞는 재고 관리 방안을 수립하였다. 4개의 지역군으로 분리하여 분기별로 상이한 판매 패턴을 분석해 2018년 예측에 반영했다. 예시로, 18년도의 “Binders” 품목의 안전재고 가치 수준을 살펴보면 EAST 지역의 2분기와 4분기의 차이는 57.1 배 수준의 차이 (68.9 vs 3934.3) 를로 보이고 있으며, 이 추세는 단순히 분기별 판매 추이와 비슷해보인다. 하지만 16년 4분기 ~ 17년 4분기 매출 (6170 → 3756.3) 처럼 지역 단위의 매출 하락세에도 총 “binders”가 전반적인 상승세를 보이고 있음으로 3934.3라는 전체 성장률을 반영한 안전재고 가치를 설정했다. 이처럼 다각도에서 자료를 분석하고, 복합적인 요소를 반영할 수 있다면 보다 정교하게 대비할 수 있을 것이다. 기업에서는 안전재고량, 안전재고가치 개념을 수립해 재고 부족이나 과잉 문제를 미리 예방할 수 있고 재고 손실을 최소화 할 수 있을 것이다.

결론적으로, 본 프로젝트를 통해 기업은 데이터에 기반한 고객 및 재고 관리 전략을 수립함으로써, 운영 효율성을 극대화하고 장기적인 성장 전략을 마련할 수 있는 기틀을 마련하였다. 고객 측면에서는 고가치 고객을 식별하고 이들에게 집중적인 마케팅을 수행하여 매출을 극대화할 수 있었으며, 재고 관리 측면에서는 과거 데이터를 활용한 예측 모델을 통해 재고 관리의 비용을 최소화하고, 효율성을 극대화할 수 있었다. 이러한 접근은 현대 비즈니스 환경에서 경쟁력을 유지하고 지속 가능한 성장을 이루기 위한 필수적인 요소임을 확인하였다. 따라서, 이와 같은 데이터 기반의 전략 수립은 기업의 장기적인 성공과 경쟁력 강화를 위한 중요한 도구로 활용될 수 있을 것이다.