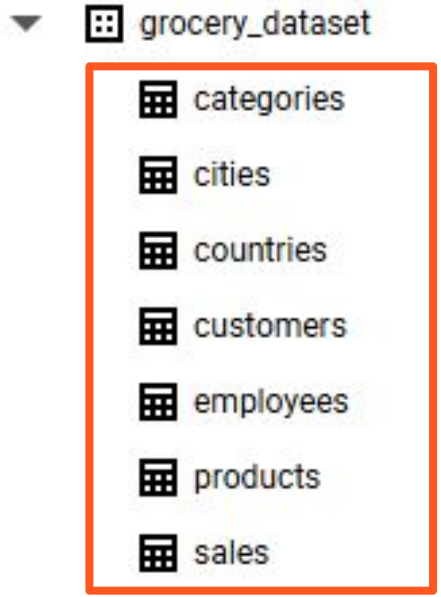
RevoGrocers

Week 4 & 5 - Intermediate & Advanced Assignment

Ligasyah Arnanda Primadana
Batch February 2025

Dataset Overview



📌 Kaggle Grocery Sales Dataset

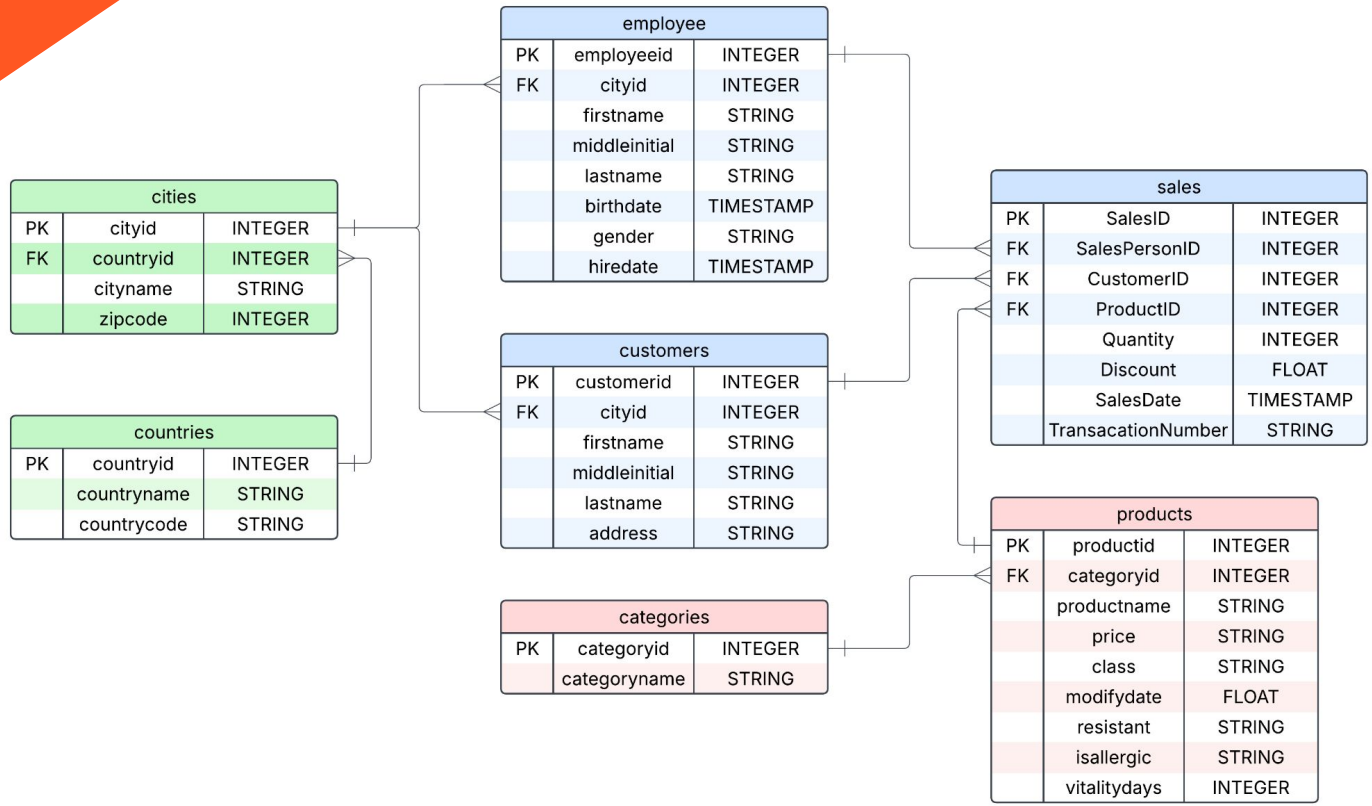
RevoGrocers is a fictional company created for analytical purposes.

Any insights or recommendations are based on this dataset and do not reflect real-world data.

RevoGrocers is a fictional grocery retail business that operates in multiple locations, offering a diverse range ofgrocery products to customers. The company aims to optimize sales strategies, enhance customer experience, andincrease revenue by leveraging data-driven decision-making.

Disclaimer: The data set and analysis on this deck only for educational purposes

Entity Relationship Diagram



01	**Specific**	Enhance sales performance across various product categories by focusing on marketing, customer engagement, and product development.
02	**Measureable**	Aim for a increase in overall revenue and a improvement in repeat purchase rates within the next fiscal year.
03	**Achieveable**	Allocate resources effectively to support marketing campaigns and product development
04	**Realistic**	10% increase in overall revenue and a 5% improvement
05	**Timebound**	within the next fiscal year.


Click for
The Queries
Breakfast
Tea

The product category that generates the highest revenue.

Product_Category	Total_Revenue
Confections	556930717.35

```sql
SELECT
  c.categoryname AS Product_Category,
  ROUND(SUM(s.Quantity * p.price * (1 - s.Discount)),2) AS Total_Revenue
FROM
  `fsda-sql-01.grocery_dataset.sales` s
JOIN
  `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
JOIN
  `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
GROUP BY
    c.categoryname
ORDER BY
    Total_Revenue DESC
LIMIT 1;
```

Dominance in Revenue Generation:

The Confections category stands out as the leading revenue generator for RevoGrocers, indicating a strong consumer preference for sweet products. This dominance suggests that confectionery items are a significant part of the shopping experience for customers.

The correlation between revenue and total units sold for each product category.

```sql
WITH Revenue_Units AS
(
  SELECT
    c.categoryname AS Product_Category,
    SUM(s.Quantity * p.price * (1 - s.Discount)) AS Total_Revenue,
    SUM(s.Quantity) AS Total_Units_Sold
  FROM
    `fsda-sql-01.grocery_dataset.sales` s
  JOIN
    `fsda-sql-01.grocery_dataset.products` p ON s.ProductID = p.productid
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
  GROUP BY
    c.categoryname
)

SELECT
    Product_Category,
    ROUND(Total_Revenue, 2) AS Total_Revenue,
    Total_Units_Sold,
FROM
    Revenue_Units
ORDER BY
    2 DESC;
```

Product_Category	Total_Revenue	Total_Units_Sold
Confections	556930717.35	11078474
Meat	492888844.69	9719292
Poultry	440025564.97	9159847
Cereals	427393431.91	8735296
Snails	372084885.21	7199409
Beverages	366515024.01	7393693
Produce	362861133.52	8174673
Dairy	354358156.77	6815143
Seafood	330527987.47	6996152
Grain	323879126.65	5433152
Shell fish	299598294.03	6983457

A correlation analysis would likely reveal a positive relationship between total units sold and total revenue for most categories. However, categories with high units sold but lower revenue may indicate pricing strategies that could be optimized.

The correlation between revenue and the number of unique customers for each product category.

```sql
WITH Revenue_Customers AS
(
  SELECT
    c.categoryname AS Product_Category,
    SUM(s.Quantity * p.price * (1 - s.Discount)) AS Total_Revenue,
    COUNT(DISTINCT s.CustomerID) AS Unique_Customers
  FROM
    `fsda-sql-01.grocery_dataset.sales` s
  JOIN
    `fsda-sql-01.grocery_dataset.products` p
    ON s.ProductID = p.productid
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c
    ON p.categoryid = c.categoryid
  GROUP BY
    c.categoryname
)

SELECT
    Product_Category,
    ROUND(Total_Revenue,2) AS Total_Revenue,
    Unique_Customers,
FROM
    Revenue_Customers
ORDER BY
    2 DESC;
```

Product_Category	Total_Revenue	Unique_Customers
Confections	556930717.35	98743
Meat	492888844.69	98701
Poultry	440025564.97	98679
Cereals	427393431.91	98651
Snails	372084885.21	98376
Beverages	366515024.01	98424
Produce	362861133.52	98601
Dairy	354358156.77	98308
Seafood	330527987.47	98334
Grain	323879126.65	97335
Shell fish	299598294.03	98338

A correlation analysis would likely reveal a positive relationship between total revenue and unique customers for most categories. Categories with high unique customer counts but lower revenue may benefit from targeted marketing efforts to increase purchase frequency.

The average price per unit for each product category.

```sql
SELECT
  c.categoryname AS Product_Category,
  ROUND(AVG(p.price),2) AS Average_Price_Per_Unit
FROM
  `fsda-sql-01.grocery_dataset.products` p
JOIN
  `fsda-sql-01.grocery_dataset.categories` c
  ON p.categoryid = c.categoryid
GROUP BY
  c.categoryname
ORDER BY
  2 DESC;
```

Product_Category	Average_Price_Per_Unit
Grain	61.4
Dairy	53.56
Snails	53.2
Meat	52.27
Confections	51.85
Beverages	51.04
Cereals	50.42
Poultry	49.46
Seafood	48.67
Produce	45.8
Shell fish	44.27

- Highest Average Price: The Grain category has the highest average price per unit at 61.40, indicating premium positioning.
- Competitive Pricing: Dairy (53.56) and Snails (53.20) suggest a focus on quality or specialty products.
- Confectionery Pricing: Confections at 51.85 balances significant revenue with competitive pricing.
- Mid-Range Pricing: Beverages (51.04) and Cereals (50.42) appeal to a broad customer base with mid-range pricing.

- Lower Average Prices: Produce (45.80) and Shell Fish (44.27) focus on volume sales for cost-conscious consumers.
- Strategic Implications: Higher-priced categories may benefit from quality-focused marketing, while lower-priced ones could emphasize promotions.

The correlation between the average price per unit and the number of buyers (unique customers) per category.

```sql
WITH Average_Price AS
(
  SELECT
    c.categoryname AS Product_Category,
    AVG(p.price) AS Average_Price_Per_Unit
  FROM
    `fsda-sql-01.grocery_dataset.products` p
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c
    ON p.categoryid = c.categoryid
  GROUP BY
        c.categoryname
),
Unique_Customers AS
(
  SELECT
    c.categoryname AS Product_Category,
    COUNT(DISTINCT s.CustomerID) AS Unique_Customers
  FROM
    `fsda-sql-01.grocery_dataset.sales` s
  JOIN
    `fsda-sql-01.grocery_dataset.products` p
    ON s.ProductID = p.productid
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c
    ON p.categoryid = c.categoryid
  GROUP BY
    c.categoryname
)
```

```sql
SELECT
    ap.Product_Category,
    ROUND(ap.Average_Price_Per_Unit,2) AS Average_Price_Per_Unit,
    uc.Unique_Customers,
FROM
    Average_Price ap
JOIN
    Unique_Customers uc ON ap.Product_Category = uc.Product_Category
ORDER BY
    2 DESC;
```

Product_Category	Average_Price_Per_U	Unique_Customers
Grain	61.4	97335
Dairy	53.56	98308
Snails	53.2	98376
Meat	52.27	98701
Confections	51.85	98743
Beverages	51.04	98424
Cereals	50.42	98651
Poultry	49.46	98679
Seafood	48.67	98334
Produce	45.8	98601
Shell fish	44.27	98338

The correlation between the average price per unit and the number of buyers (unique customers) per category.

- Positive Correlation: There appears to be a positive correlation between average price per unit and the number of unique customers, particularly in categories like Grain and Dairy, suggesting that higher prices may attract a loyal customer base.

- The evaluation of the correlation between average price per unit and unique customers reveals that while higher prices can attract loyal customers, effective pricing strategies across various categories can enhance overall customer engagement and revenue.

Categories contribute the most to overall revenue (percentage-wise)

```
WITH Category_Revenue AS
(
  SELECT
    c.categoryname AS Product_Category,
    SUM(s.Quantity * p.price * (1 - s.Discount)) AS Total_Revenue
  FROM
    `fsda-sql-01.grocery_dataset.sales` s
  JOIN
    `fsda-sql-01.grocery_dataset.products` p
    ON s.ProductID = p.productid
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c ON p.categoryid = c.categoryid
  GROUP BY
    c.categoryname
),
Overall_Revenue AS
(
  SELECT
    SUM(Total_Revenue) AS Total_Overall_Revenue
  FROM
    Category_Revenue
)
```

```
SELECT
    cr.Product_Category,
    ROUND(cr.Total_Revenue,2) AS Total_Revenue,
    ROUND((cr.Total_Revenue / ov.Total_Overall_Revenue) *
100,2) AS Revenue_Percentage,
    RANK() OVER (ORDER BY cr.Total_Revenue DESC) AS
Revenue_Rank
FROM
    Category_Revenue cr,
    Overall_Revenue ov
ORDER BY
    3 DESC;
```

Product_Category	Total_Revenue	Revenue_Percentage	Revenue_Rank
Confections	556930717.35	12.87	1
Meat	492888844.69	11.39	2
Poultry	440025564.97	10.17	3
Cereals	427393431.91	9.88	4
Snails	372084885.21	8.6	5
Beverages	366515024.01	8.47	6
Produce	362861133.52	8.39	7
Dairy	354358156.77	8.19	8
Seafood	330527987.47	7.64	9
Grain	323879126.65	7.48	10
Shell fish	299598294.03	6.92	11

Categories contribute the most to overall revenue (percentage-wise)

- **Top Contributor:** The Confections category leads with a total revenue of 55,693,071.35, contributing 12.87% to overall revenue, indicating its significance in the product mix.
- The analysis reveals that Confections, Meat, and Poultry are the top contributors to overall revenue, while other categories like Cereals and Snails also play significant roles. Understanding these contributions can help RevoGrocers optimize inventory, marketing strategies, and product offerings to enhance overall sales performance.

Categories contribute the most to overall revenue (percentage-wise)

```sql
WITH Customer_Purchases AS
(
  SELECT
    c.categoryname AS Product_Category,
    s.CustomerID,
    COUNT(s.CustomerID) AS Total_Purchases
  FROM
    `fsda-sql-01.grocery_dataset.sales` s
  JOIN
    `fsda-sql-01.grocery_dataset.products` p
    ON s.ProductID = p.productid
  JOIN
    `fsda-sql-01.grocery_dataset.categories` c
    ON p.categoryid = c.categoryid
  GROUP BY
        c.categoryname, s.CustomerID
)

SELECT
  Product_Category,
  COUNT(DISTINCT CustomerID) AS Unique_Customers,
  SUM(Total_Purchases) AS Total_Purchases,
  ROUND((SUM(CASE
             WHEN Total_Purchases > 1
             THEN Total_Purchases ELSE 0 END)
             / SUM(Total_Purchases)) * 100,2) AS Repeat_Purchase_Rate
FROM
    Customer_Purchases
GROUP BY
    Product_Category
HAVING
    COUNT(DISTINCT CustomerID) > 1
ORDER BY
    4 DESC;
```

Product_Category	Unique_Customers	Total_Purchases	Repeat_Purchase_Ra
Confections	98743	851979	99.98
Meat	98701	747762	99.95
Poultry	98679	704145	99.92
Cereals	98651	671771	99.88
Produce	98601	627923	99.83
Beverages	98424	569175	99.69
Snails	98376	553612	99.63
Seafood	98334	537576	99.59
Shell fish	98338	537486	99.58
Dairy	98308	523869	99.5
Grain	97335	417853	98.53

- Top Category: The Confections category has the highest repeat purchase rate at 99.98%, indicating exceptional customer loyalty and satisfaction.
- The analysis reveals that Confections leads in repeat purchase rates, followed closely by Meat and Poultry. This information is valuable for RevoGrocers to focus on maintaining customer satisfaction and loyalty in these high-performing categories.

Overall findings

- Top Revenue Generator: Confections leads with 55,693,071.35 in revenue, contributing 12.87% to overall sales.
- Strong Performance: Meat and Poultry follow with revenues of 49,288,844.69 and 44,002,556.97, respectively.
- High Repeat Purchase Rates: Confections has a repeat purchase rate of 99.98%, indicating strong customer loyalty.
- Average Price Insights: Grain has the highest average price per unit at 61.40, suggesting premium positioning.

- Correlation Observations: Higher average prices generally align with a significant number of unique customers, indicating effective pricing strategies.
- Sales Volume: Confections also lead in total units sold with 11,078,474, reflecting strong market demand.
- Diverse Product Range: The data highlights a diverse range of categories performing well, suggesting a balanced product portfolio.
- Strategic Opportunities: Insights indicate potential for growth in categories with lower revenue but high unique customer counts, such as Produce and Seafood.

Conclusions

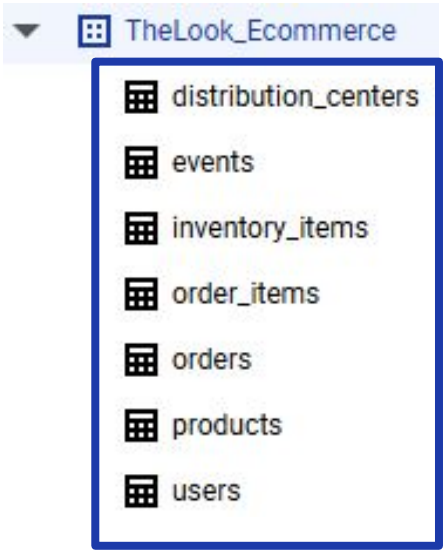
Overall, the analysis indicates that Confections is the standout category in terms of revenue and customer loyalty, while other categories like Meat and Poultry also perform strongly. The insights gained from this analysis can inform strategic decisions regarding inventory management, marketing initiatives, and pricing strategies to optimize sales performance across all product categories.

ADVANCED

Click for

THE QUERIES

Dataset Overview



theLook eCommerce

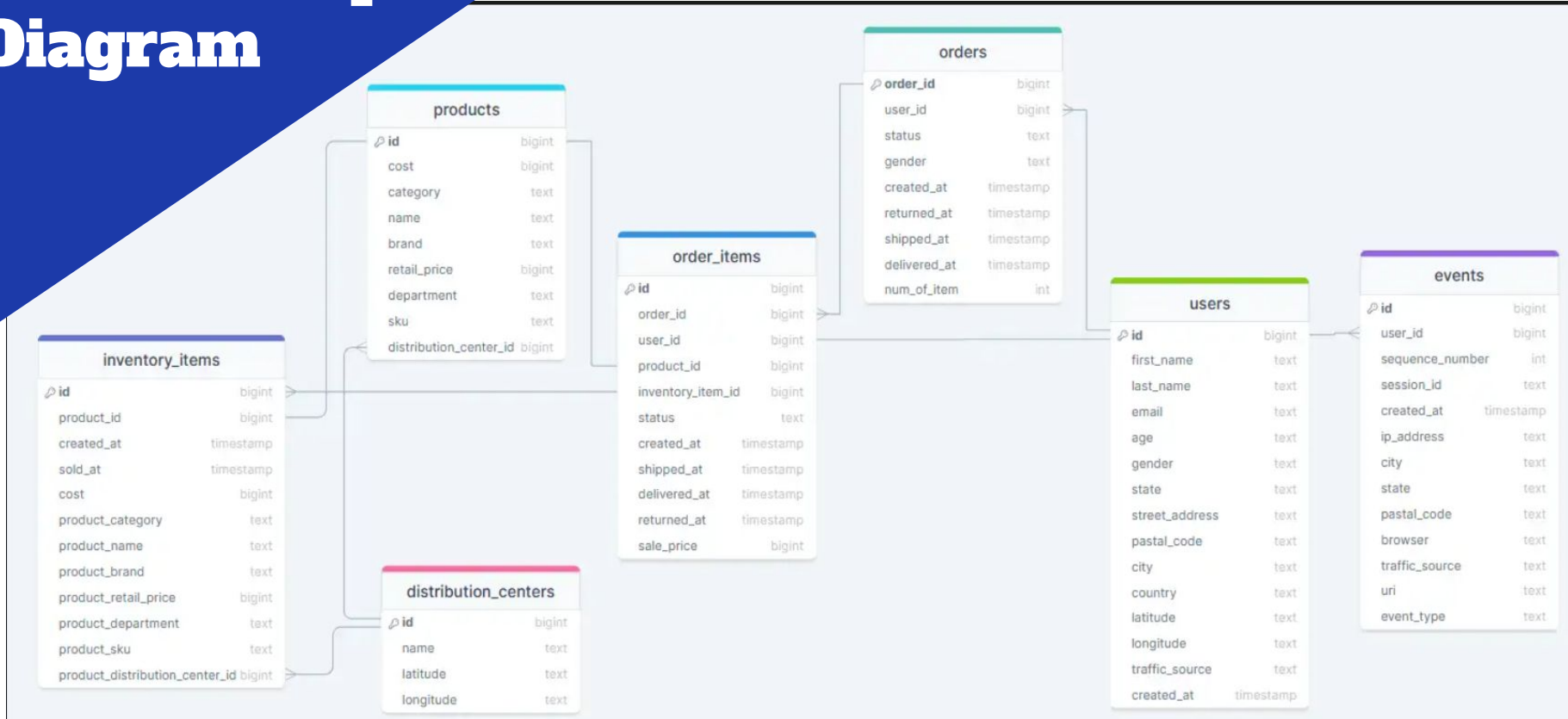
BigQuery Public Data

Synthetic eCommerce and Digital Marketing data

TheLook is a fictitious eCommerce clothing site developed by the Looker team. The dataset contains information about customers, products, orders, logistics, web events, and digital marketing campaigns. The contents of this dataset are synthetic and are provided to industry practitioners for the purpose of product discovery, testing, and evaluation.

Disclaimer: The data set and analysis on this deck only for educational purposes

Entity Relationship Diagram



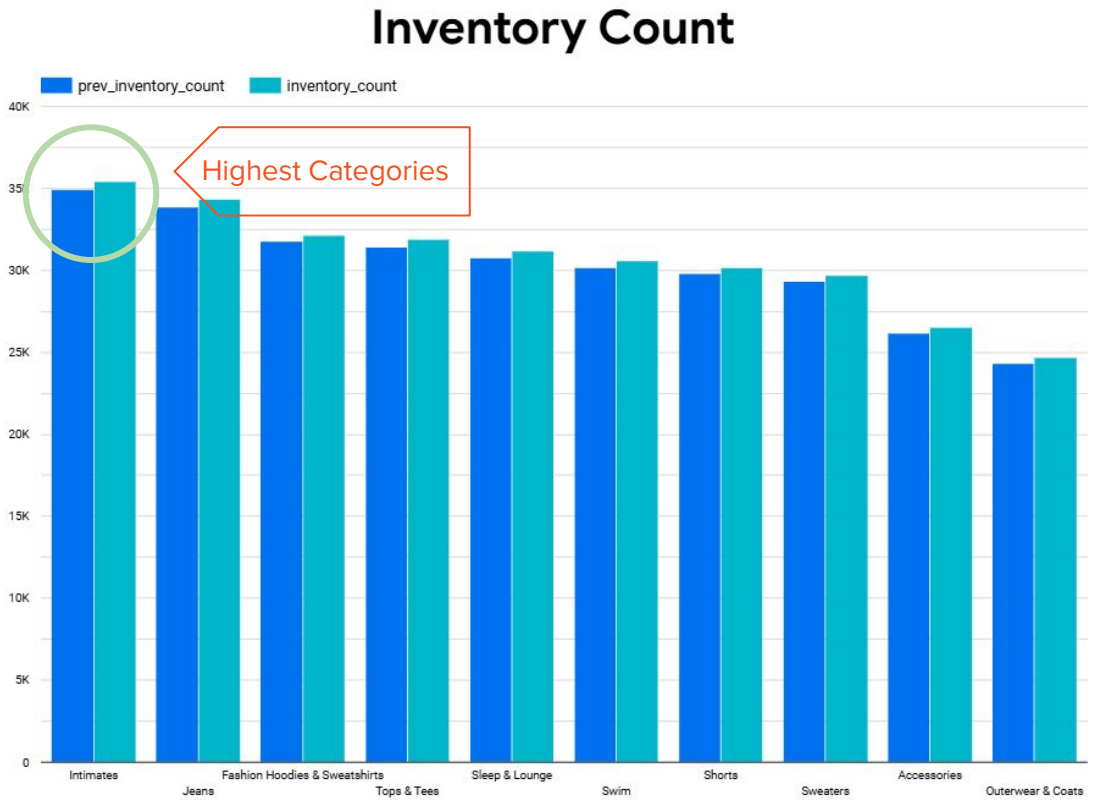
New and active buyers characteristics

01	**Specific**	Focuses on improving user retention and optimizing inventory management for an eCommerce clothing site.
02	**Measureable**	Retention Rates, Inventory, and Growth Trends
03	**Achieveable**	Implementing targeted marketing campaigns, Conduct ongoing cohort analysis and seasonal promotions planning
04	**Realistic**	increasing sales and customer loyalty 5 percent.
05	**Timebound**	Implement targeted marketing campaigns and inventory adjustments for one years.

Monthly growth of inventory in percentage breakdown by product categories

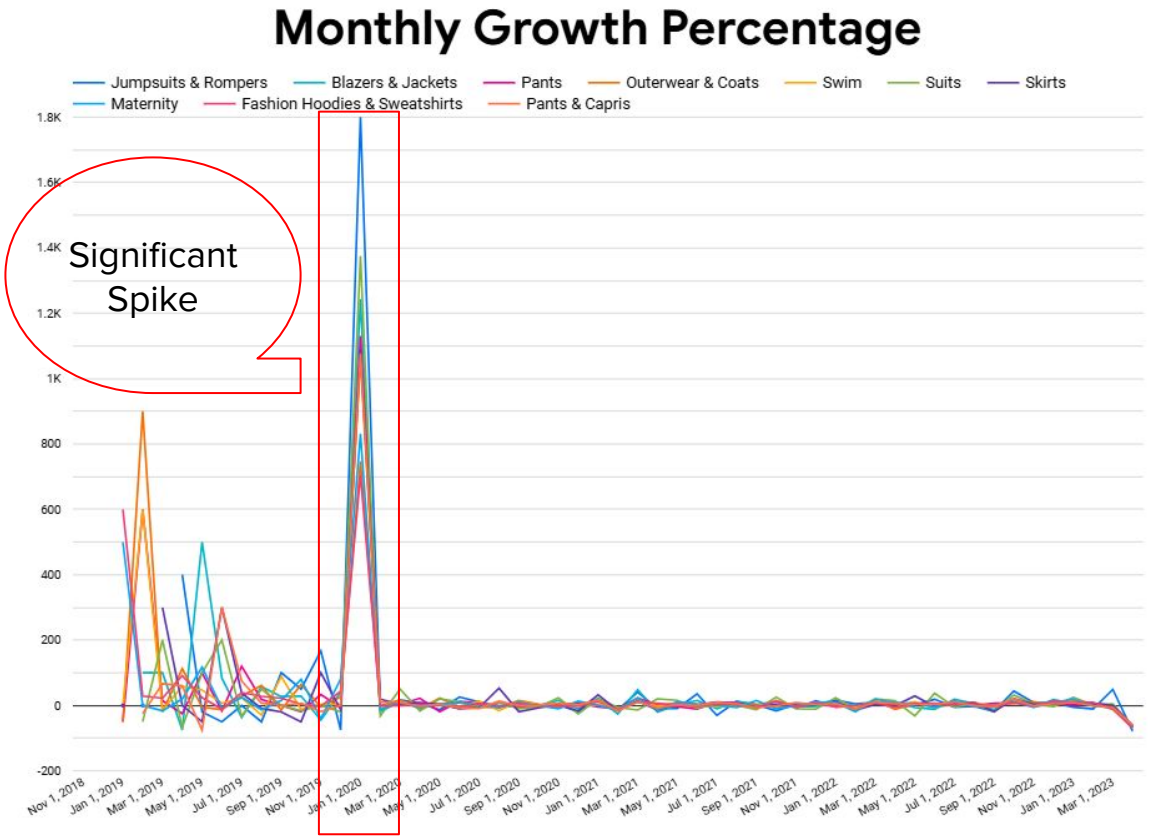
```
WITH MonthlyInventory AS (
    SELECT
        DATE_TRUNC(i.created_at, MONTH) AS month,
        p.category,
        COUNT(i.id) AS inventory_count
    FROM
        `fsda-sql-01.TheLook_Ecommerce.inventory_items` i
    JOIN
        `fsda-sql-01.TheLook_Ecommerce.products` p ON i.product_id =
p.id
    GROUP BY
        month, p.category
),
MonthlyGrowth AS (
    SELECT
        month,
        category,
        inventory_count,
        LAG(inventory_count) OVER (PARTITION BY category ORDER BY
month) AS prev_inventory_count
    FROM
        MonthlyInventory
)
```

```
SELECT
    DATE(month) AS Month,
    category,
    inventory_count,
    prev_inventory_count,
    ROUND(CASE
        WHEN prev_inventory_count IS NULL THEN NULL
        ELSE ((inventory_count - prev_inventory_count) /
prev_inventory_count) * 100
    END,2) AS monthly_growth_percentage
FROM
    MonthlyGrowth
ORDER BY
    1 DESC, 2;
```



On Every Categories Have Relatively same Increment

Monthly retention cohorts and retention cohort



Recommendations

- **Monthly Growth Trends:** The overall growth appears to stabilize after the initial spikes, suggesting a need for strategies to maintain or increase growth rates.
- **Inventory Count Analysis** Conversely, categories with lower inventory counts may indicate either lower demand or potential stockouts, which could affect sales. Also, observe the increasing stock in every categories
- **Focus on High-Growth Categories:** Analyze the categories that experienced significant growth during peak months. Consider implementing targeted marketing campaigns or promotions for these categories to capitalize on their popularity.

Monthly retention cohorts and retention cohort

```sql
WITH FirstPurchase AS
(
    SELECT
        user_id,
        DATE_TRUNC(MIN(created_at), MONTH) AS cohort_month
    FROM
        `fsda-sql-01.TheLook_Ecommerce.orders`
    WHERE
        EXTRACT(YEAR FROM created_at) = 2022
    GROUP BY
        user_id
),
MonthlyRetention AS (
    SELECT
        fp.cohort_month,
        EXTRACT(MONTH FROM o.created_at) AS purchase_month,
        COUNT(DISTINCT o.user_id) AS returning_users
    FROM
        FirstPurchase fp
    JOIN
        `fsda-sql-01.TheLook_Ecommerce.orders` o ON fp.user_id = o.user_id
    WHERE
        EXTRACT(YEAR FROM o.created_at) = 2022
    GROUP BY
        fp.cohort_month, purchase_month
),
CohortSize AS (
    SELECT
        cohort_month,
        COUNT(DISTINCT user_id) AS total_users
    FROM
        FirstPurchase
    GROUP BY
        cohort_month
)
SELECT
    DATE(r.cohort_month) AS Cohort_Month,
    r.purchase_month,
    r.returning_users,
    cs.total_users,
    ROUND(CASE
        WHEN cs.total_users = 0 THEN 0
        ELSE (r.returning_users * 100.0 / cs.total_users)
    END,2) AS retention_percentage
FROM
    MonthlyRetention r
JOIN
    CohortSize cs ON r.cohort_month = cs.cohort_month
ORDER BY
    1, 2;
```

Monthly retention cohorts and retention cohort

Retention Percentage	Purchase Month											
Cohort Month	12	11	10	9	8	7	6	5	4	3	2	1
1/1/2022	3.83	3.11	4.55	4.03	3.14	4.17	3.76	3.83	3.73	3.86	3.52	100
2/1/2022	4.21	3.83	3.98	3.91	4.1	4.4	4.36	3.83	4.43	4.13	100	
3/1/2022	3.93	3.63	3.99	3.43	4.52	4.39	3.96	4.06	4.62	100		
4/1/2022	3.78	4.05	4.42	4.52	3.72	4.81	3.92	3.69	100			
5/1/2022	4.98	4.42	4.58	4.45	4.17	4.48	4.51	100				
6/1/2022	5.03	4.81	4.97	5.25	5.19	4.55	100					
7/1/2022	5.64	4.34	4.93	4.66	5.59	100						
8/1/2022	5.7	5.2	5.28	4.9	100							
9/1/2022	5.74	5.8	5.97	100								
10/1/2022	7.05	6.83	100									
11/1/2022	6.99	100										
12/1/2022	100											
Grand Total	**13.07333333**	**13.27454545**	**14.267**	**15.01666667**	**16.30375**	**18.11428571**	**20.085**	**23.082**	**28.195**	**35.99666667**	**51.76**	**100**

- Consistent Retention Rates: Retention rates for users who made their first purchase in January 2022 show a gradual decline over the months, starting at 3.83% in January and dropping to 7.05% in October, indicating a slight recovery in retention as the year progresses.
- Overall Trend: The overall trend indicates that while initial retention rates are low, there is potential for recovery in later months, particularly for cohorts that are actively engaged through marketing strategies or promotions.

Google Sheets : Cohort Table Link

Recommendations

- **Retention Improvement Initiatives:** For older cohorts with declining retention rates, consider implementing targeted retention initiatives, such as personalized marketing, loyalty programs, or re-engagement campaigns.
- **Engagement Strategies:** The high retention rate for the November cohort suggests that effective engagement strategies were in place.
- **Seasonal Promotions:** The data indicates that certain months yield better retention rates. Planning seasonal promotions around these times could help boost sales and customer loyalty.


THANK YOU