**This is an demostration of SQL and Analysis skills along with data visualization skills as an emxaple:**

**Inquiry:**

Explore these tables and their underlying data, then write a query that properly joins these three tables. Please answer the following questions and provide a copy of your script along with the output in your answer document.

1. What was the total sales revenue in 2020, and how did this compare to the year prior? Are there any particular months that drove or hurt sales year-over-year? *Note that sales revenue is a derived field from the KPIs provided.*
2. What were the Top 5 Stores (based on revenue) in 2018 and what was their Inventory Sell-Through %? Please provide the store ID, store name, revenue, unit sales, inventory, and sell-through %. *Note that ST% is a derived field from the KPIs provided.*
3. What were the Top 5 Suppliers and each of their Top 5 Products (based on profit) in 2019? *Note that profit is a derived field from the KPIs provided.*

Please create a one-page dashboard with a couple of visualizations that provide key insights into the data.

**Answers for SQL Skills Part:**

Platform used: Google BigQuery

Preparation Period - Data Cleaning Process:

**Executive Summary**

The Data Cleaning process during the Preparation Period aimed to ensure the quality, accuracy, and consistency of our dataset. This report outlines the steps taken, issues identified, and actions performed to clean the data.

**Data Sources**

During this phase, I utilized data from three primary sources: Sales, Products, and Inventory tables.

**Data Cleaning Steps**

1. **Identifying Missing Values:**

I systematically examined each table for missing values. Fortunately, I found no missing data, ensuring that our dataset was complete.

1. **Handling Duplicate Values:**

I conducted a thorough check for duplicate values and successfully eliminated any duplicates encountered.

1. **Outlier Detection and Correction:**

I used plots in Excel to identify and assess outliers. Fortunately, I found no significant outliers in our dataset.

1. **Ensuring Data Consistency:**

I meticulously cross-referenced foreign keys in the Sales and Inventory tables with the primary keys in the Products and Inventory tables to ensure data consistency.

**Data Cleaning Results**

* No missing values were identified in any of the tables.
* No duplicates in any of the tables.
* No significant outliers were detected.
* Data consistency was confirmed across the tables.

Problem Solving Period - Writing Script and Analyze the Results :

1. What was the total sales revenue in 2020, and how did this compare to the year prior?  Are there any particular months that drove or hurt sales year-over-year?
2. The total sales revenue in 2020:

SELECT ROUND(SUM(UnitPrice \* Quantity), 2) AS Total\_Sales\_Revenue, 2020 AS Year

FROM `dluxury.Sales` as ds

LEFT JOIN `dluxury.Inventory` as di

ON ds.ProductId = di.ProductId

LEFT JOIN `dluxury.Products` as dp

ON dp.ProductId = ds.ProductId

WHERE EXTRACT(YEAR FROM Date) = 2020

UNION ALL

SELECT ROUND(SUM(UnitPrice \* Quantity), 2), 2019

FROM `dluxury.Sales` as ds

LEFT JOIN `dluxury.Inventory` as di

ON ds.ProductId = di.ProductId

LEFT JOIN `dluxury.Products` as dp

ON dp.ProductId = ds.ProductId

WHERE EXTRACT(YEAR FROM Date) = 2019;

Output:

|  |  |
| --- | --- |
| Total\_Sales\_Revenue | Year |
| 8329108.88 | 2020 |
| 8251369.7 | 2019 |

The total sales revenue in 2020 was **$8,329,108.88**. When compared to the year prior (2019), the total sales revenue in 2020 was **higher by $77,739.18** ($8,329,108.88 - $8,251,369.70).

Year-over-Year Comparison:

Percentage Change = ((Total Revenue in 2020 - Total Revenue in 2019) / Total Revenue in 2019) \* 100

Percentage Change = (($8,329,108.88 - $8,251,369.70) / $8,251,369.70) \* 100 ≈ 0.94%

1. Year-over-year monthly sales analysis:

select EXTRACT(YEAR FROM ds.Date) AS Year,EXTRACT(MONTH FROM ds.Date) AS Month, round(SUM(UnitPrice\*Quantity),2) as Revenue

from `dluxury.Sales` as ds

left join `dluxury.Inventory` as di on ds.ProductId = di.ProductId

left join `dluxury.Products` as dp on ds.ProductId = dp.ProductId

WHERE

EXTRACT(YEAR FROM ds.Date) IN (2019, 2020)

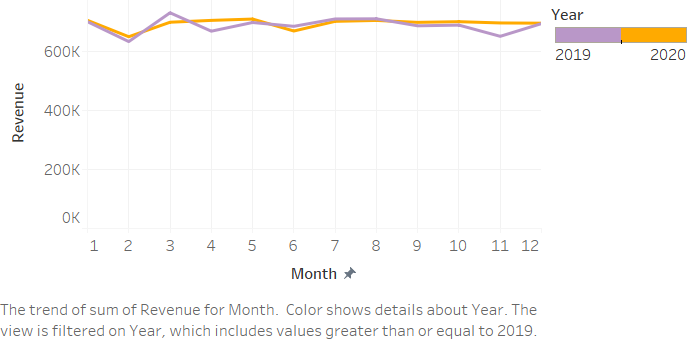
GROUP BY YEAR,MONTH

order by Year, Month

Output(modified to show clearly):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Month** | **Revenue** |  | **Year** | **Month** | **Revenue** |
| 2019 | 1 | 700105.3 |  | 2020 | 1 | 704686.64 |
| 2019 | 2 | 632782.2 |  | 2020 | 2 | 649212.25 |
| 2019 | 3 | 730295.12 |  | 2020 | 3 | 698286.49 |
| 2019 | 4 | 668014.3 |  | 2020 | 4 | 704552.06 |
| 2019 | 5 | 697403.37 |  | 2020 | 5 | 709063.18 |
| 2019 | 6 | 684484.45 |  | 2020 | 6 | 668601.55 |
| 2019 | 7 | 709409.29 |  | 2020 | 7 | 701346.89 |
| 2019 | 8 | 709912.7 |  | 2020 | 8 | 704410.02 |
| 2019 | 9 | 686199.31 |  | 2020 | 9 | 697858.89 |
| 2019 | 10 | 688543.59 |  | 2020 | 10 | 700165.45 |
| 2019 | 11 | 650603.99 |  | 2020 | 11 | 695854.95 |
| 2019 | 12 | 693616.07 |  | 2020 | 12 | 695070.51 |

To enhance clarity, I utilized Tableau to create the plot:



**Insights from the plot:**

The overall yearly comparison indicates a modest YoY increase in total revenue for 2020 compared to 2019. Several months experienced notable changes: **March** 2020 saw a significant decline in revenue, while **April** 2020 showed significant growth compared to April 2019. **November** 2020 had a substantial increase, contributing positively to the overall yearly performance.

February experienced a notable decrease in both of the two years, followed by a subsequent increase in March.

**Recommendations:**

* Given the YoY growth, it's crucial to continue analyzing monthly revenue trends to identify contributing factors.
* Investigate the reasons behind the March decline and consider strategies to mitigate similar fluctuations in the future.
* Capitalize on the positive growth months, such as April and November, by examining the factors that led to increased revenue and replicating successful strategies.

**Conclusion:**

The analysis of monthly revenue for 2019 and 2020 provides valuable insights into year-over-year performance. Understanding the trends and specific months of growth or decline can aid in decision-making and future revenue optimization strategies.

For more in-depth analysis and strategic planning, further examination of external factors and market conditions is recommended to gain a holistic view of revenue performance.

1. What were the Top 5 Stores (based on revenue) in 2018 and what was their Inventory Sell-Through %? Please provide the store ID, store name, revenue, unit sales, inventory, and sell-through %.

select

StoreName,

di.StoreId,

ROUND(SUM(UnitPrice \* Quantity), 2) as Revenue,

SUM(Quantity) AS Unit\_Sales,

SUM(di.QuantityAvailable) AS Inventory,

CONCAT(ROUND((SUM(Quantity) / (SUM(di.QuantityAvailable) + SUM(Quantity))) \* 100, 2), '%')

as Sell\_Through\_Rate

from `dluxury.Sales` as ds

LEFT JOIN `dluxury.Inventory` as di

ON ds.ProductId = di.ProductId

LEFT JOIN `dluxury.Products` as dp

ON dp.ProductId = ds.ProductId

where EXTRACT(YEAR FROM Date) = 2018

group by StoreName,di.StoreId

order by Revenue desc limit 5

Output:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **StoreName** | **StoreId** | **Revenue** | **Unit\_Sales** | **Inventory** | **Sell\_Through\_Rate** |
| Ben Franklin | 84879 | 374340.8 | 115649 | 15893 | 87.92% |
| Family Dollar | 71053 | 346013.82 | 96418 | 10879 | 89.86% |
| Burlington Coat Factory | 22745 | 342312.46 | 79011 | 9206 | 89.56% |
| Shopko | 22726 | 340826.5 | 97426 | 11641 | 89.33% |
| Renys | 22748 | 327836.39 | 83244 | 9749 | 89.52% |

**Insights:**

* Sell-Through Rate Variations:

Sell-through rates vary among the stores, with Family Dollar having the highest rate at 89.86% and Ben Franklin having the lowest at 87.92%. These differences indicate varying levels of inventory management and sales performance.

* Revenue Performance and Unit Sales Impact:

Ben Franklin generated the highest revenue of $374,340.80. This suggests that while Ben Franklin has a slightly lower sell-through rate, it compensates with higher revenue due to its larger inventory and competitive unit sales. This indicates that while its sell-through rate is slightly lower, it manages to move a significant number of units, resulting in higher overall revenue.

* Inventory Management:

Burlington Coat Factory maintains a relatively low inventory of 9,206 units, which is the lowest among the stores. This efficient inventory management contributes to its high sell-through rate.

**Recommendations:**

* Inventory Optimization:

Keep optimizing inventory levels to align better with unit sales. Maintaining an appropriate balance between inventory and sales can lead to improved sell-through rates.

* Marketing Strategies:

Explore marketing and promotional strategies to boost unit sales and, consequently, revenue. Identifying which products perform well and promoting them effectively can be beneficial.

* Benchmarking:

Benchmark the performance of each store against industry standards and competitors. This can help stores set realistic sell-through rate targets and improve their overall performance.

* Inventory Turnover:

Focus on improving inventory turnover rates, especially for stores with lower sell-through rates. Implement efficient inventory management practices to reduce carrying costs and avoid overstocking.

* Customer Engagement:

Invest in customer engagement initiatives to boost unit sales. Satisfied customers are more likely to make repeat purchases, leading to higher sell-through rates.

* Seasonal Planning:

Plan inventory and promotions according to seasonal demand patterns. Stores can take advantage of peak seasons to improve sell-through rates.

* Performance Monitoring:

Implement regular performance monitoring and reporting to track progress toward improving sell-through rates. Make data-driven decisions based on trends and insights.

1. What were the Top 5 Suppliers and each of their Top 5 Products (based on profit) in 2019?

SELECT

dp.Supplier AS Supplier,

ROUND(SUM(ds.UnitPrice \* ds.Quantity - dp.ProductCost \* ds.Quantity),2) AS Profit

FROM

`dluxury.Sales` AS ds

LEFT JOIN `dluxury.Inventory` as di

ON ds.ProductId = di.ProductId

LEFT JOIN `dluxury.Products` as dp

ON dp.ProductId = ds.ProductId

WHERE EXTRACT(YEAR FROM Date) = 2019

GROUP BY dp.Supplier

ORDER BY Profit DESC

LIMIT 5

Output:

**The top 5 Suppliers are:**

|  |  |
| --- | --- |
| **Supplier** | **Profit** |
| Ben Franklin | 160028.5 |
| Shopko | 151974.58 |
| Family Dollar | 144965.16 |
| Renys | 138228.15 |
| Burlington Coat Factory | 137484.6 |

**Top 5 products of the suppliers above:**

I started with Ben Franklin’s top5 products:

select Supplier, ProductName,

ROUND(SUM(ds.UnitPrice \* ds.Quantity - dp.ProductCost \* ds.Quantity),2) AS Profit

from

 `dluxury.Sales` as ds

LEFT JOIN `dluxury.Inventory` as di

ON ds.ProductId = di.ProductId

LEFT JOIN `dluxury.Products` as dp

ON dp.ProductId = ds.ProductId

WHERE EXTRACT(YEAR FROM Date) = 2019 and Supplier = "Ben Franklin"

group by

Supplier, ProductName

order by profit desc

limit 5

Then I changed the Supplier name in the script and got 5 results as follow:

|  |  |  |
| --- | --- | --- |
| **Supplier** | **ProductName** | **Profit** |
| Ben Franklin | Flour - Rye | 18374.92 |
| Ben Franklin | Soup - Campbells, Minestrone | 13474.8 |
| Ben Franklin | Wine - Red, Metus Rose | 8508 |
| Ben Franklin | Parsley Italian - Fresh | 7884.81 |
| Ben Franklin | Pepper - Cubanelle | 7606.4 |

|  |  |  |
| --- | --- | --- |
| **Supplier** | **ProductName** | **Profit** |
| Shopko | Chocolate - Liqueur Cups With Foil | 17021.21 |
| Shopko | Latex Rubber Gloves Size 9 | 11019.16 |
| Shopko | Wine - Cousino Macul Antiguas | 10269 |
| Shopko | Cheese - Pont Couvert | 9245.61 |
| Shopko | Bandage - Fexible 1x3 | 7790.55 |

|  |  |  |
| --- | --- | --- |
| **Supplier** | **ProductName** | **Profit** |
| Family Dollar | Cheese - Valancey | 10450.08 |
| Family Dollar | Onions - Cooking | 10424.4 |
| Family Dollar | Container - Foam Dixie 12 Oz | 9715.2 |
| Family Dollar | Coffee Cup 16oz Foam | 9429.37 |
| Family Dollar | Juice - Apple, 500 Ml | 8058.26 |

|  |  |  |
| --- | --- | --- |
| **Supplier** | **ProductName** | **Profit** |
| Renys | Bread - Italian Sesame Poly | 14318.53 |
| Renys | Vodka - Hot, Lnferno | 13200.66 |
| Renys | Pepper - Green, Chili | 11412.54 |
| Renys | Tarragon - Primerba, Paste | 7272.6 |
| Renys | Soup - Tomato Mush. Florentine | 6463.1 |

|  |  |  |
| --- | --- | --- |
| **Supplier** | **ProductName** | **Profit** |
| Burlington Coat Factory | Chips Potato Salt Vinegar 43g | 13145.63 |
| Burlington Coat Factory | Soup - French Onion, Dry | 9938.09 |
| Burlington Coat Factory | Oranges | 9767.52 |
| Burlington Coat Factory | Lid - 3oz Med Rec | 9246.53 |
| Burlington Coat Factory | Table Cloth 54x72 Colour | 8388 |

**Insights:**

* Diverse Product Mix:

The top-selling products from these suppliers encompass a wide range of categories, from food items like flour and cheese to specialty products like latex gloves and unique chocolates.

* Customer-Centric Approach:

The presence of everyday essentials and specialty items suggests a customer-centric approach, catering to both daily needs and niche preferences.

* Profitable Portfolio:

The profitability of these products indicates that these suppliers have identified items that resonate well with their customer base.

**Recommendations:**

* Inventory Optimization:

Continuously optimize inventory levels for these top products to prevent stockouts and reduce carrying costs.

* Marketing Strategies:

Develop effective marketing and promotion strategies for these items, highlighting their unique features and benefits.

* Supplier Relationships:

Strengthen relationships with suppliers to ensure a consistent and reliable supply of these profitable products.

* Product Diversification:

Explore opportunities to expand the product portfolio with items that align with customer preferences and market trends.

* Customer Engagement:

Engage with customers to gather feedback and insights, helping tailor the product mix to evolving preferences.

* Data-Driven Decisions:

Use data analytics to monitor performance and make informed decisions regarding inventory, pricing, and marketing efforts.