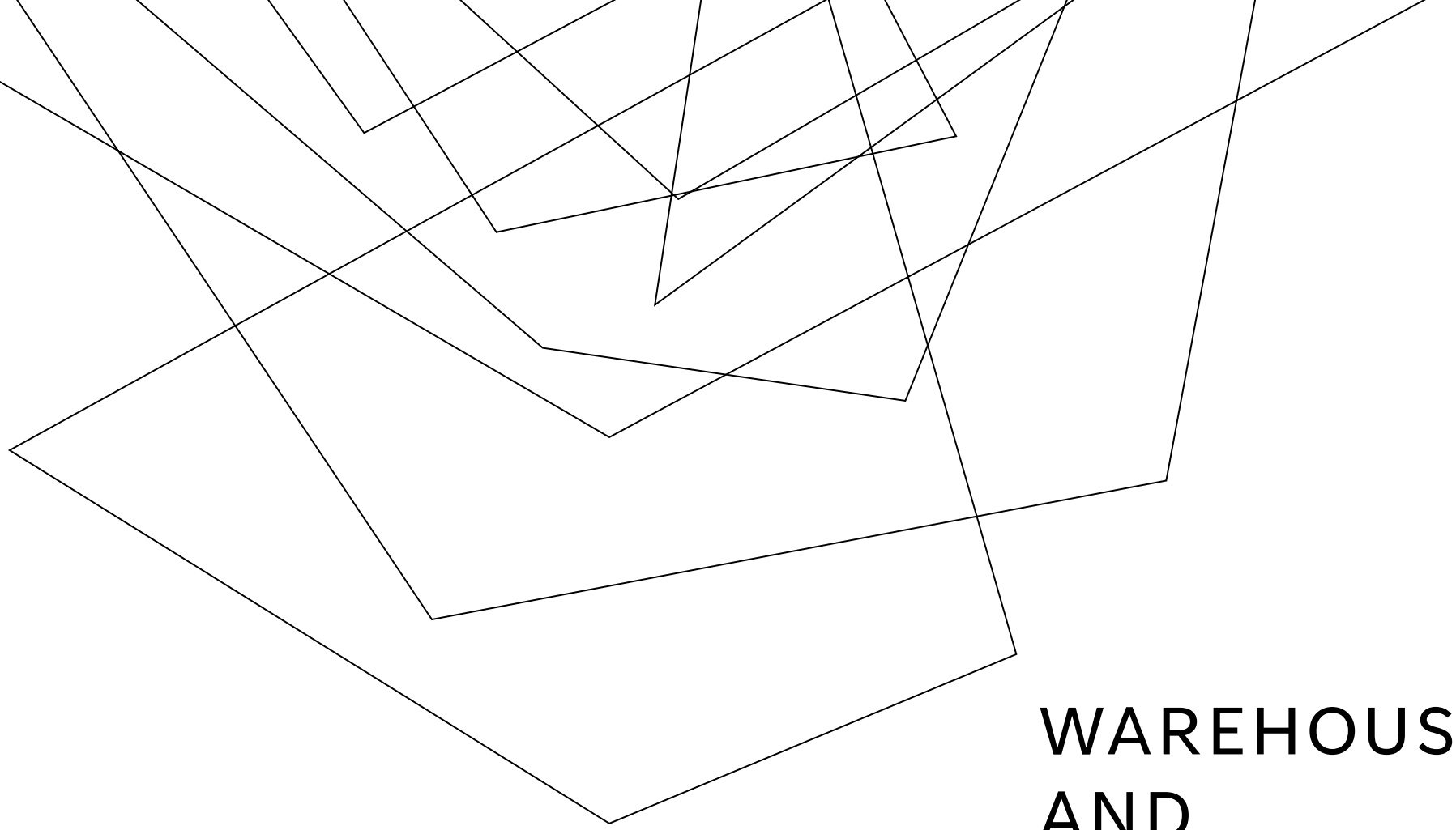
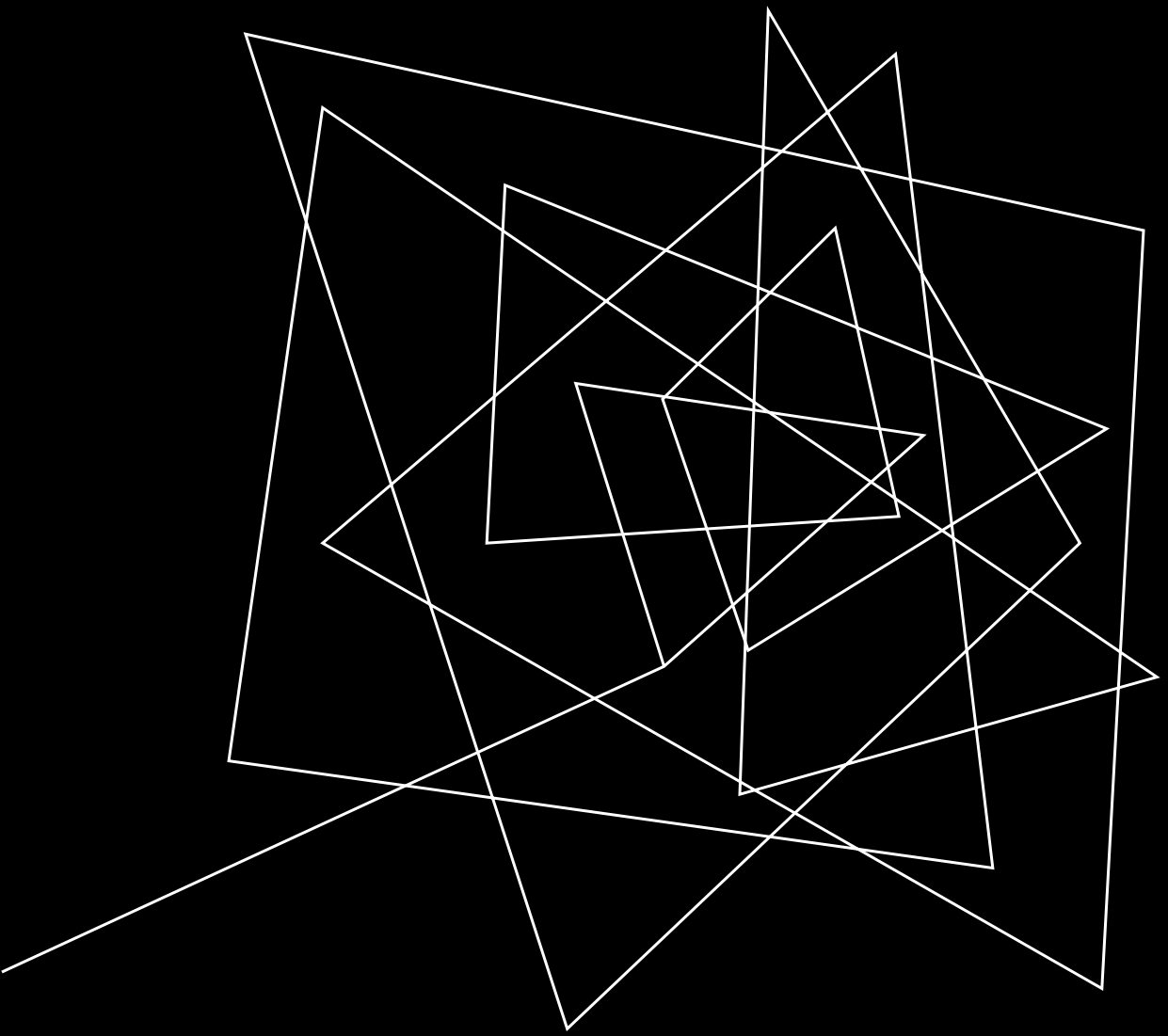


WAREHOUSE AND RETAIL SALES

Christina Gao



DATASET USED

Warehouse and Retail Sales

FROM Montgomery County of Maryland

INTRODUCTION

My mission is to transform this raw dataset into a structured and insightful resource, achieving this by leveraging the power of **SQL Server Management Studio (SSMS)**. Throughout this presentation, I'll explore the steps involved in database creation, data inspection, cleaning if necessary, and the formulation of meaningful queries. This presentation serves as a visual documentation of my process, shedding light on my data exploration adventure.

Additionally, I will be utilizing **Tableau** to create a dynamic and interactive visualization dashboard that will bring the data to life. This synergy of tools will not only enhance the depth of our analysis but also make our findings accessible and actionable for a wider audience, ultimately contributing to informed decision-making and data-driven strategies.

DATASET INFORMATION

Column Name	Description	Type
YEAR	Calendar Year	smallint
MONTH	Month	tinyint
SUPPLIER	Supplier Name	nvarchar(50)
ITEM_CODE	Item Code	int
ITEM_DESCRIPTION	Item Description	nvarchar(100)
ITEM_TYPE	Item Type	nvarchar(50)
RETAIL_SALES	Cases of product sold from DLC dispensaries	float
RETAIL_TRANSFER	Cases of product transferred to DLC dispensaries	float
WAREHOUSE_SALES	Cases of product sold to MC licensees	float

FIRST LOOK AT THE DATASET (FIRST 3 ROWS)

I downloaded the dataset as a Comma-Separated Values (CSV) file. When opening the file, we can take a glance at what the dataset contains. As shown below, there are 9 columns and a total of 307,645 rows.

YEAR	MONTH	SUPPLIER	ITEM CODE	ITEM DESCRIPTION	ITEM TYPE	RETAIL SALES	RETAIL TRANSFERS	WAREHOUSE SALES
2020	1	REPUBLIC NATIONAL DISTRIBUTING CO	100009	BOOTLEG RED - 750ML	WINE	0	0	2
2020	1	PWSWN INC	100024	MOMENT DE PLAISIR - 750ML	WINE	0	1	4
2020	1	RELIABLE CHURCHILL LLLP	1001	S SMITH ORGANIC PEAR CIDER - 18.7OZ	BEER	0	0	1

QUERY THE ENTIRE DATABASE

SELECT * FROM Warehouse_and_Retail_Sales;

Following the database creation in SSMS and importing the data as a flat file, I proceeded to execute the above query to ensure the successful data import. The resulting output, as displayed below, affirms the presence of 9 columns and a total of 307,645 rows in our dataset

Results Messages

	YEAR	MONTH	SUPPLIER	ITEM_CODE	ITEM_DESCRIPTION	ITEM_TYPE	RETAIL_SALES	RETAIL_TRANSFERS	WAREHOUSE_SALES
1	2020	1	REPUBLIC NATIONAL DISTRIBUTING CO	100009	BOOTLEG RED - 750ML	WINE	0	0	2
2	2020	1	PWSWN INC	100024	MOMENT DE PLAISIR - 750ML	WINE	0	1	4
3	2020	1	RELIABLE CHURCHILL LLLP	1001	S SMITH ORGANIC PEAR CIDER - 18.7OZ	BEER	0	0	1
4	2020	1	LANTERNA DISTRIBUTORS INC	100145	SCHLINK HAUS KABINETT - 750ML	WINE	0	0	1
5	2020	1	DIONYSOS IMPORTS INC	100293	SANTORINI GAVALA WHITE - 750ML	WINE	0.819999992847443	0	0
6	2020	1	KYSELA PERE ET FILS LTD	100641	CORTENOVA VENETO P/GRIG - 750ML	WINE	2.75999999046326	0	6
7	2020	1	SANTA MARGHERITA USA INC	100749	SANTA MARGHERITA P/GRIG ALTO - 375ML	WINE	0.0799999982118607	1	1
8	2020	1	BROWN-FORMAN BEVERAGES WORLDWIDE	1008	JACK DANIELS COUNTRY COCKTAIL SOUTHERN PEACH - 1...	BEER	0	0	2
9	2020	1	JIM BEAM BRANDS CO	10103	KNOB CREEK BOURBON 9YR - 100P - 375ML	LIQUOR	6.40999984741211	4	0
10	2020	1	INTERNATIONAL CELLARS LLC	101117	KSARA CAB - 750ML	WINE	0.330000013113022	1	2
11	2020	1	HEAVEN HILL DISTILLERIES INC	10120	J W DANT BOURBON 100P - 1.75L	LIQUOR	1.70000004768372	1	0
12	2020	1	BACCHUS IMPORTERS LTD	10123	NOTEWORTHY SMALL BATCH BOURBON - 750ML	LIQUOR	1.01999998092651	0	0
13	2020	1	BACCHUS IMPORTERS LTD	10124	NOTEWORTHY SMALL BATCH RYE 750ML	LIQUOR	0.680000007152557	0	0
14	2020	1	BACCHUS IMPORTERS LTD	10125	NOTEWORTHY SMALL BATCH HONEY BBN 750	LIQUOR	0.340000003576279	0	0
15	2020	1	MONSIEUR TOUTON SELECTION	101346	ALSACE WILLIAM GEW - 750ML	WINE	0	0	2
16	2020	1	THE COUNTRY VINTNER, LLC DBA WINEBOW	101486	POLIZIANO ROSSO MONTEPUL - 750ML	WINE	0	0	1
17	2020	1	THE COUNTRY VINTNER, LLC DBA WINEBOW	101532	HATSUMAGO SAKE JUN MAI SHU - 720ML	WINE	0.340000003576279	1	1
18	2020	1	ROYAL WINE CORP	101664	RAMON CORDOVA RIOJA - 750ML	WINE	0.159999996423721	0	2
19	2020	1	REPUBLIC NATIONAL DISTRIBUTING CO	101702	MANISCHEWITZ CREAM RED CONCORD - 1.5L	WINE	0	0	1

WHICH YEAR HAD THE HIGHEST TOTAL RETAIL SALES?

```
SELECT TOP 1 YEAR, SUM(RETAIL_SALES) AS Total_Retail_Sales
FROM Warehouse_and_Retail_Sales
GROUP BY YEAR
ORDER BY Total_Retail_Sales DESC;
```

The year **2019** had the highest total retail sales = **960,205.14 cases sold**

	YEAR	Total_Retail_Sales
1	2019	960205.139919708

WHO ARE THE TOP 5 SUPPLIERS IN TERMS OF RETAIL SALES?

```
SELECT TOP 5 SUPPLIER
FROM Warehouse_and_Retail_Sales
GROUP BY SUPPLIER
ORDER BY SUM(RETAIL_SALES) DESC;
```

The top 5 suppliers with the highest overall retail sales are:

E & J GALLO WINERY
DIAGEO NORTH AMERICA INC
CONSTELLATION BRANDS
ANHEUSER BUSCH INC
JIM BEAM BRANDS CO

	SUPPLIER
1	E & J GALLO WINERY
2	DIAGEO NORTH AMERICA INC
3	CONSTELLATION BRANDS
4	ANHEUSER BUSCH INC
5	JIM BEAM BRANDS CO

RETRIEVE ALL DATA FROM THE YEAR 2019 DURING MONTH 10.

```
SELECT *
FROM Warehouse_and_Retail_Sales
WHERE YEAR = 2019 AND MONTH = 10;
```

The query above outputs all the data from the year 2019, month 10. There are 13,086 total rows from this query.

	YEAR	MONTH	SUPPLIER	ITEM_CODE	ITEM_DESCRIPTION	ITEM_TYPE	RETAIL_SALES	RETAIL_TRANSFERS	WAREHOUSE_SALES
1	2019	10	AMERICAN BEVERAGE MARKETERS	166663	AGALIMA MARGARITA MIX - 1L	NON-ALCOHOL	18.4300003051758	18	0
2	2019	10	CHARM CITY BEVERAGE LLC	166665	GT'S CLASSIC SYNERGY KOMBUCHA DIVINE GRAPE 12/16OZ	BEER	0	0	3
3	2019	10	BARON FRANCOIS LTD	166666	LITTLE CRICKET GRUNER VELTLINER - 750ML	WINE	0	0	3
4	2019	10	SUTTER HOME WINERY INC	166668	CARMEN FRIDA KAHLO CABERNET - 750ML	WINE	0	0	4
5	2019	10	CAMPARI AMERICA LLC	166669	GRAND MARNIER CORDON ROUGE GLENCAIRN GLASS VAP	LIQUOR	5.09000015258789	31	0
6	2019	10	DEUTSCH FAMILY WINE & SPIRITS	166673	JOSH CELLARS NORTH COAST RSV CHARD - 750ML	WINE	0	3	0
7	2019	10	TAPWINES	166688	DANCING COYOTE PINOT NOIR - 750ML	WINE	0	0	1
8	2019	10	VINTAGE WINES INC	166691	CAMPARRON SELECCION - 750ML	WINE	0	0	5

FIND THE MONTH WITH THE HIGHEST RETAIL SALES FOR EACH YEAR.

```
WITH total AS (
   SELECT YEAR, MONTH,
   SUM(RETAIL_SALES) AS total_sales,
   RANK() OVER(PARTITION BY YEAR ORDER BY SUM(RETAIL_SALES) DESC) AS sales_rank
   FROM Warehouse_and_Retail_Sales
   GROUP BY YEAR, MONTH)

SELECT YEAR, MONTH
FROM total
WHERE sales_rank = 1;
```

	YEAR	MONTH
1	2017	12
2	2018	2
3	2019	11
4	2020	3

The query above utilizes a **common table expression (CTE)**. By ranking each summation of retail sales grouped by year, we can find the month with the highest retail sales for each year.

CALCULATE THE GROWTH RATE IN RETAIL SALES FOR 2018 - 2020.

```
WITH yearly_sales AS (
    SELECT YEAR, SUM(RETAIL_SALES) as total_sales
    FROM Warehouse_and_Retail_Sales
    GROUP BY YEAR),

    yearly_growth AS (
    SELECT t1.YEAR, t1.total_sales AS current_year_sales, t2.total_sales AS prev_year_sal
    ((t1.total_sales - t2.total_sales) / t2.total_sales) * 100 AS growth_rate
    FROM yearly_sales t1 JOIN yearly_sales t2 ON t1.YEAR = t2.YEAR + 1)

SELECT * FROM yearly_growth
ORDER BY YEAR;
```

	YEAR	current_year_sales	prev_year_sales	growth_rate
1	2018	153598.360052984	686760.020058304	-77.6343474333372
2	2019	960205.139919708	153598.360052984	525.140229093906
3	2020	360379.220276508	960205.139919708	-62.4685178933073

In order to calculate the growth rate, we must first query for the current year's sales and previous year's sales.

(Current Year – Previous Year) / (Previous Year) * 100 = Growth Rate

We can see that 2019 had the best growth rate

SEGMENT SUPPLIERS INTO CATEGORIES BASED ON THEIR SALES BEHAVIOR AND PERFORMANCE.

```
WITH total_sales AS (
    SELECT SUPPLIER, SUM(RETAIL_SALES) + SUM(WAREHOUSE_SALES) AS
total_retail_warehouse_sales
    FROM Warehouse_and_Retail_Sales
    GROUP BY SUPPLIER)


SELECT SUPPLIER, total_retail_warehouse_sales,
CASE
    WHEN total_retail_warehouse_sales >= 10000 THEN 'High Sales'
    WHEN total_retail_warehouse_sales >= 1000 THEN 'Average Sales'
    ELSE 'Low Sales'
END AS supplier_performance
FROM total_sales;
```

To categorized the suppliers, we first must calculate each supplier's total sales (retail + warehouse). We will then use a CASE statement to label each supplier based on their total sales.

If sales are greater than or equal to 10,000, then that supplier has **High Sales**.
If sales are greater than or equal to 1,000, then that supplier has **Average Sales**.
If sales are lower than 1,000, then that supplier has **Low Sales**.

	SUPPLIER	total_retail_warehouse_sales	supplier_performance
1	HEINEKEN USA	886035.369024925	High Sales
2	FLYING DOG BREWERY LLLP	154679.790092401	High Sales
3	BUCK DISTRIBUTING COMP...	50635.3800389916	High Sales
4	JORDAN VINEYARD	309.370001360774	Low Sales
5	DC BRAU BREWING LLC	27168.41000285	High Sales
6	BINDING BRAUEREI USA INC	1265.51000182331	Average Sales
7	POTOMAC SELECTIONS INC	669.20000128448	Low Sales
8	RAPP CAPITAL LLC	39.9800002127886	Low Sales

CALCULATE THE YEAR-OVER-YEAR GROWTH IN RETAIL SALES FOR EACH SUPPLIER.

```
WITH total AS (
    SELECT YEAR, SUPPLIER, SUM(RETAIL_SALES) AS total_sales
    FROM Warehouse_and_Retail_Sales
    WHERE SUPPLIER IS NOT NULL
    GROUP BY YEAR, SUPPLIER)

SELECT YEAR, SUPPLIER, total_sales,
LAG(total_sales) OVER (PARTITION BY SUPPLIER ORDER BY YEAR) AS prev_year_sales,
CASE
    WHEN
    LAG(total_sales) OVER (PARTITION BY SUPPLIER ORDER BY YEAR) IS NOT NULL AND
LAG(total_sales) OVER (PARTITION BY SUPPLIER ORDER BY YEAR) <> 0
    THEN (total_sales - LAG(total_sales) OVER (PARTITION BY SUPPLIER ORDER BY YEAR)) /
(LAG(total_sales) OVER (PARTITION BY SUPPLIER ORDER BY YEAR)) * 100
    ELSE NULL
END AS year_over_year
FROM total
ORDER BY SUPPLIER, YEAR;
```

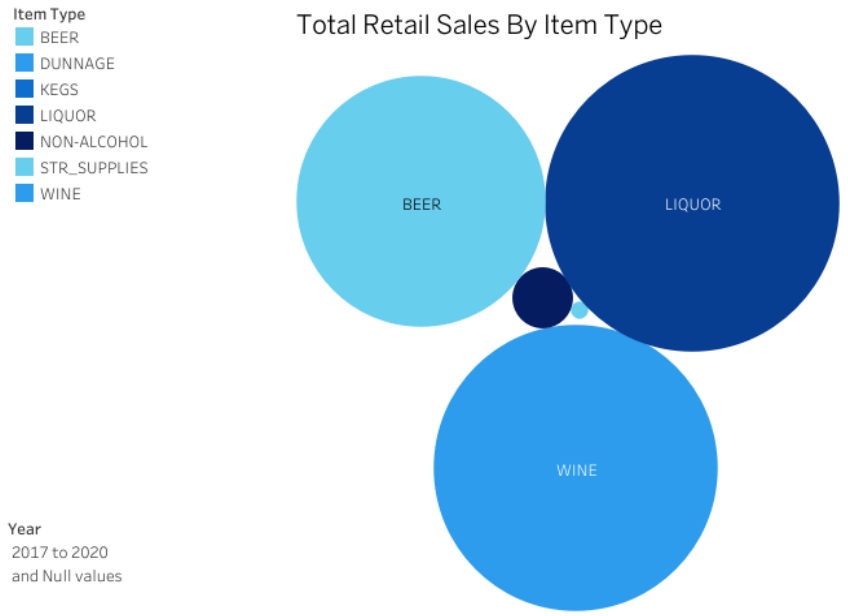
	YEAR	SUPPLIER	total_sales	prev_year_sales	year_over_year
1	2017	8 VINI INC	2.29000002145767	NULL	NULL
2	2018	8 VINI INC	0.239999994635582	2.29000002145767	-89.5196509874785
3	2017	A HARD...	0.559999987483025	NULL	NULL
4	2017	A I G WI...	10.1899998933077	NULL	NULL
5	2018	A I G WI...	2.05999998748302	10.1899998933077	-79.7841019720134
6	2019	A I G WI...	0.990000039339066	2.05999998748302	-51.9417453711405
7	2020	A I G WI...	0	0.990000039339...	-100
8	2017	A VINTN...	7021.61999166757	NULL	NULL

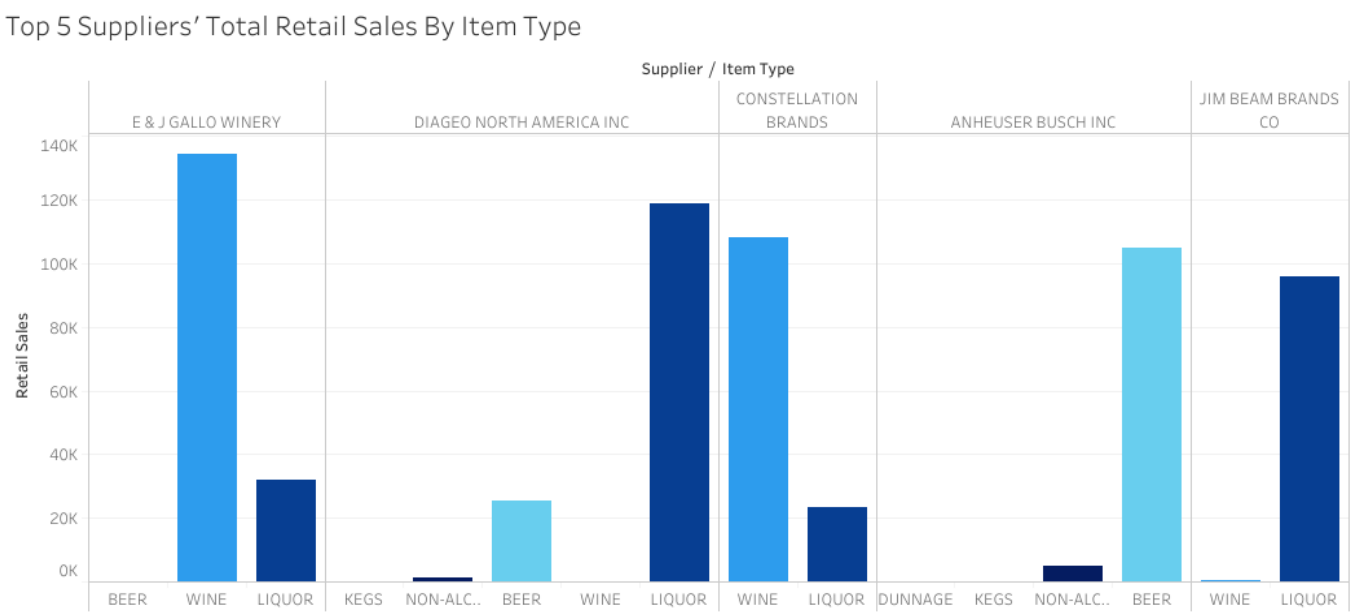
The query above calculates the year-over-year growth rate. It first checks if the previous year's sales are not null and not equal to zero to avoid dividing by zero. If these conditions are met, it calculates the growth rate as a percentage change.

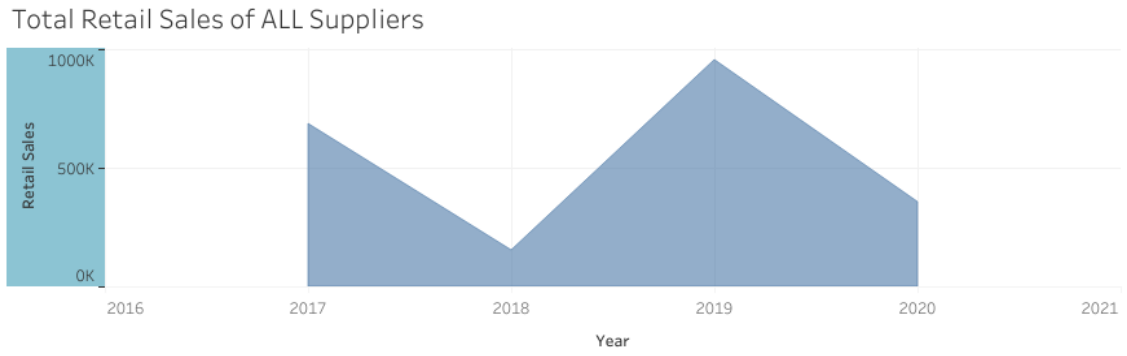
NULL VALUES : there's no previous year data resulting in no calculation

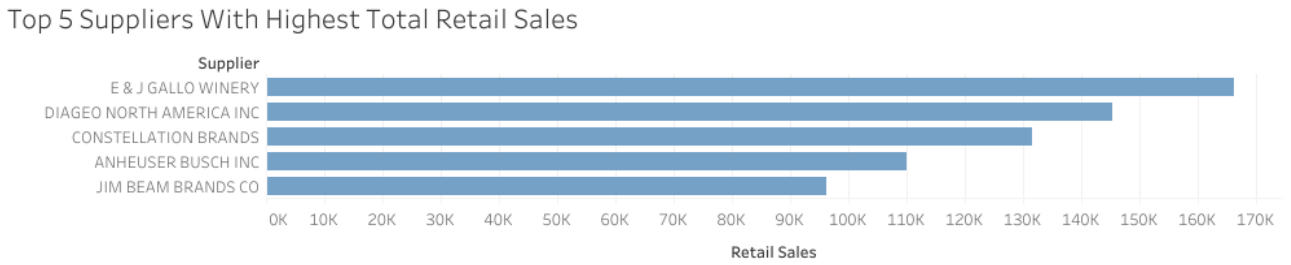
VISUALIZING THE DATASET UTILIZING TABLEAU

View on Tableau Public [HERE]







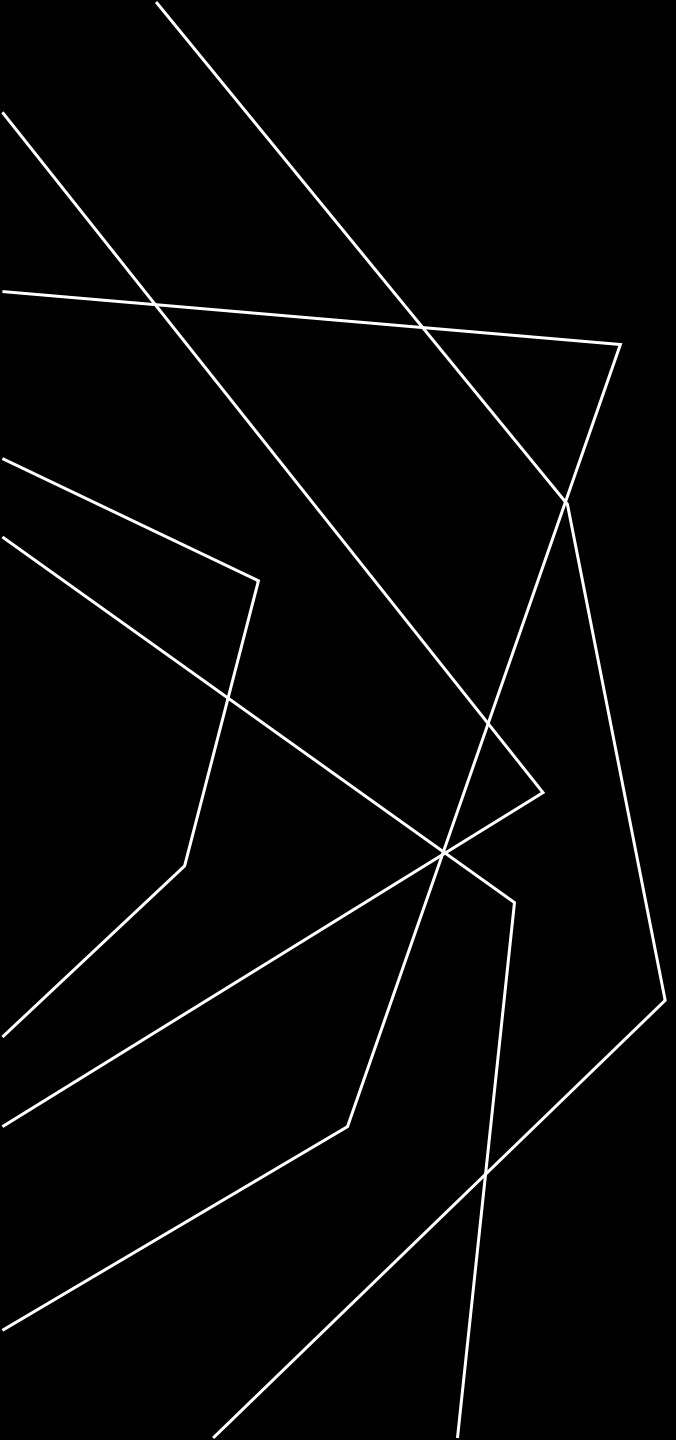


CONCLUSION

After completing this project, I have gained invaluable experience in data management, database creation, and data cleaning using SQL Server Management Studio. This journey has honed my skills in crafting meaningful SQL queries to extract insights from raw datasets.

Additionally, the integration of Tableau into our workflow has added a dynamic dimension to our analysis, enabling the creation of interactive visualizations that can transform data into actionable insights. This project has not only expanded my technical knowledge but also reinforced the importance of data-driven decision-making.

As I reflect on this experience, I am excited about the potential to apply these skills to future projects and contribute to informed decision-making and data-driven strategies. This journey has indeed been a stepping stone toward a more data-savvy and insightful future.



THANK YOU

Christina Gao

https://cwgao00.github.io

https://github.com/cwgao00