

01

# Smart Inventory Management in Retail and Warehousing

A Data-Driven Approach

Presented by:  
Team: Data Wranglers  
University of North Texas

02

# Table Of Content

- INTRODUCTION
- Problem Statement
- Data Lifecycle Steps and Tools Used
- Use Cases of Dataset
- Conclusion and Next Steps
- References

# 03



## Introduction:

In the dynamic world of retail and warehousing, efficient inventory management is crucial for operational effectiveness and customer satisfaction. The complexities associated with inventory management, such as overstocking, stockouts, and resource misallocation, pose significant challenges. This summary outlines a data-driven approach to tackle these issues, leveraging advanced tools and techniques across the data lifecycle.

# 04

## PROBLEM STATEMENT

Inefficient inventory management in retail and warehouse leads to issues like overstocking, stock outs and misallocation of resources. These challenges impact operational efficiency, customer satisfaction and ultimately profitability.



# Data Lifecycle Steps and Tools Used:

05

DATA CLEANING  
AND WRANGLING  
WITH OPENREFINE

STORING PRE-PROCESSED  
DATASET IN GOOGLE CLOUD  
PLATFORM (GCP) BUCKET

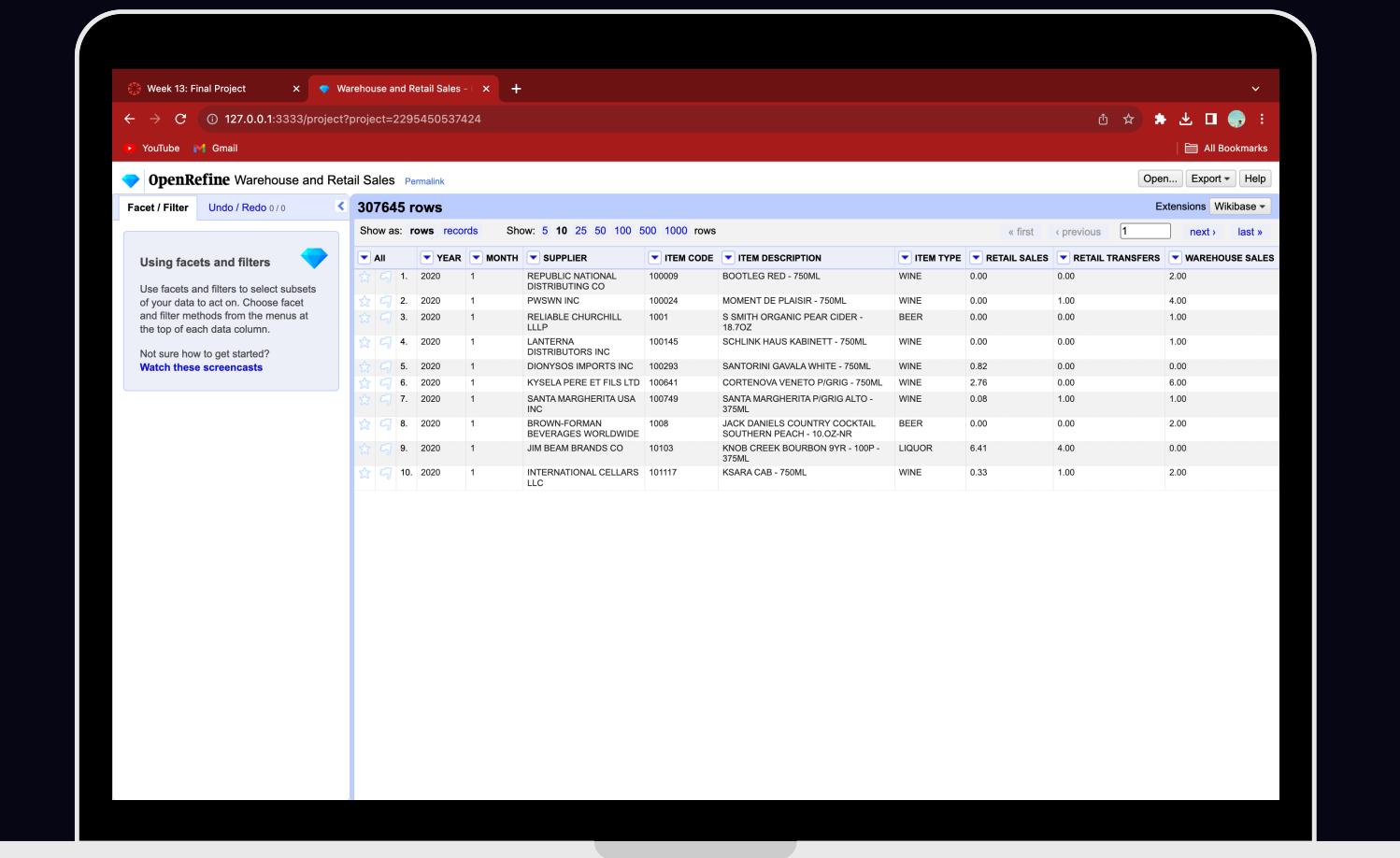
LOADING DATASET  
INTO HADOOP  
DISTRIBUTED FILE  
SYSTEM (HDFS)

ANALYSIS USING  
SQL QUERIES IN  
BIGQUERY

INTERPRETING  
ANALYSIS  
RESULTS

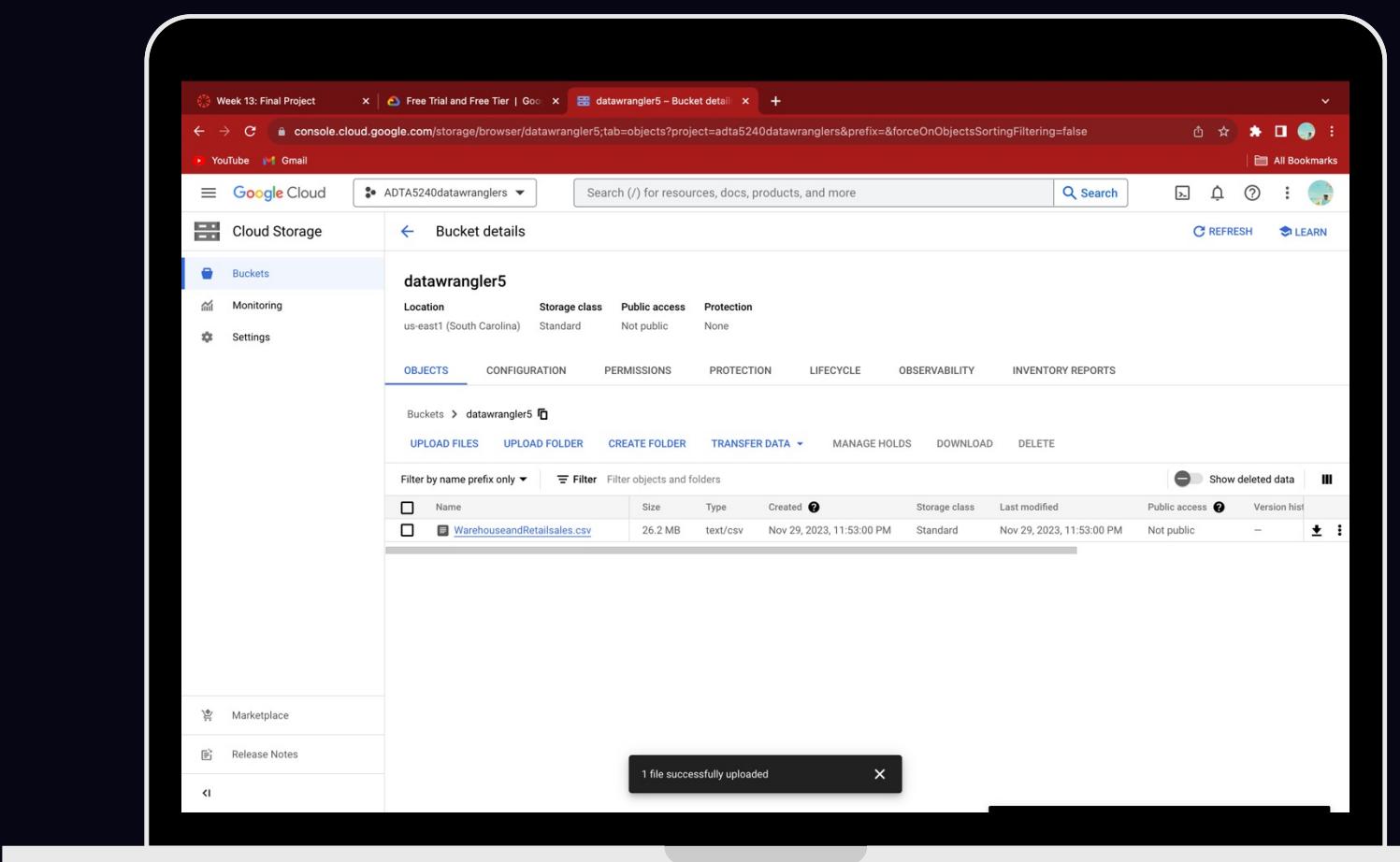
# 06

## Data cleaning and Data wrangling using Open refine



The screenshot shows the OpenRefine interface with a dataset titled "Warehouse and Retail Sales". The interface includes a facets and filters sidebar on the left and a main table view on the right. The table has 307645 rows and 8 columns: ITEM CODE, ITEM DESCRIPTION, ITEM TYPE, RETAIL SALES, RETAIL TRANSFERS, WAREHOUSE SALES, YEAR, MONTH, and SUPPLIER. The data includes various items like "REPUBLIC NATIONAL DISTRIBUTING CO", "PWSN INC", "RELIABLE CHURCHILL LLLP", etc., with details such as item descriptions like "BOOTLEG RED - 750ML", "MOMENT DE PLAISIR - 750ML", and "S SMITH ORGANIC PEAR CIDER - 18.7OZ".

## Storage of Dataset in GCP Bucket

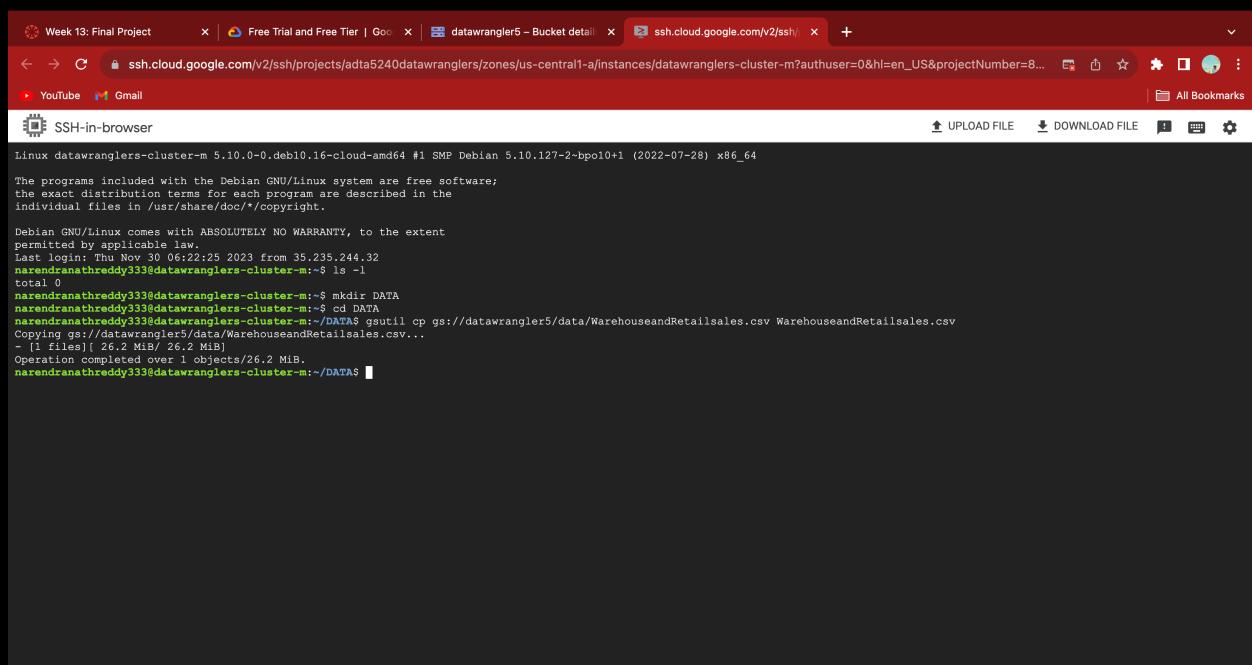


The screenshot shows the Google Cloud Storage interface with a bucket named "datawrangler5". The bucket details page shows a single object named "WarehouseandRetailSales.csv". The file was uploaded on Nov 29, 2023, at 11:53:00 PM, is 26.2 MB in size, and is a text/csv file. The storage class is Standard, and public access is set to Not public.

07

# Exploring Hadoop & Spark Cluster

# Loading of Dataset into Hadoop Distributed File System (HDFS)



The screenshot shows a web browser window with multiple tabs open. The active tab is titled "datawrangler5 - Bucket detail" and its URL is "ssh.cloud.google.com/v2/ssh". The browser interface includes a header with "Week 13: Final Project", "Free Trial and Free Tier | Goo", and "datawrangler5 - Bucket detail". Below the header are standard browser controls like back, forward, and search. The main content area displays a terminal session on a Linux system named "datawranglers-cluster-m". The session starts with a welcome message from the Debian 5.10.127-2-bpo10+1 kernel. It then shows the user navigating through a directory structure, creating a "DATA" folder, changing into it, and copying a file named "WarehouseandRetailSales.csv" from Google Cloud Storage into the folder. The terminal ends with a message indicating the operation completed over 1 object.

```
Linux datawranglers-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2-bpo10+1 (2022-07-28) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Nov 30 06:22:25 2023 from 35.235.244.32
narendranathreddy333@datawranglers-cluster-m:~$ ls -l
total 0
narendranathreddy333@datawranglers-cluster-m:~$ mkdir DATA
narendranathreddy333@datawranglers-cluster-m:~$ cd DATA
narendranathreddy333@datawranglers-cluster-m:~/DATA$ gsutil cp gs://datawrangler5/data/WarehouseandRetailSales.csv WarehouseandRetailSales.csv
Copying gs://datawrangler5/data/WarehouseandRetailSales.csv...
[1 files] [ 26.2 MiB / 26.2 MiB]
Operation completed over 1 objects/26.2 MiB.
narendranathreddy333@datawranglers-cluster-m:~/DATA$
```

# Use Cases of Dataset:

## Monthly Sales Summary:

1

Provides an overview of sales performance, highlighting trends and patterns monthly.

## Yearly Supplier Contribution:

6

Evaluates the annual performance and contribution of suppliers to the business.

## Supplier Performance Analysis:

2

Assesses the reliability and efficiency of suppliers, influencing procurement strategies.

## Item Type Sales and Stock Analysis:

3

Offers insights into the sales volume and stock levels of different item categories, aiding in inventory optimization.

## Identifying Top Selling Items:

4

Helps in recognizing the most popular items, guiding marketing and stock replenishment decisions.

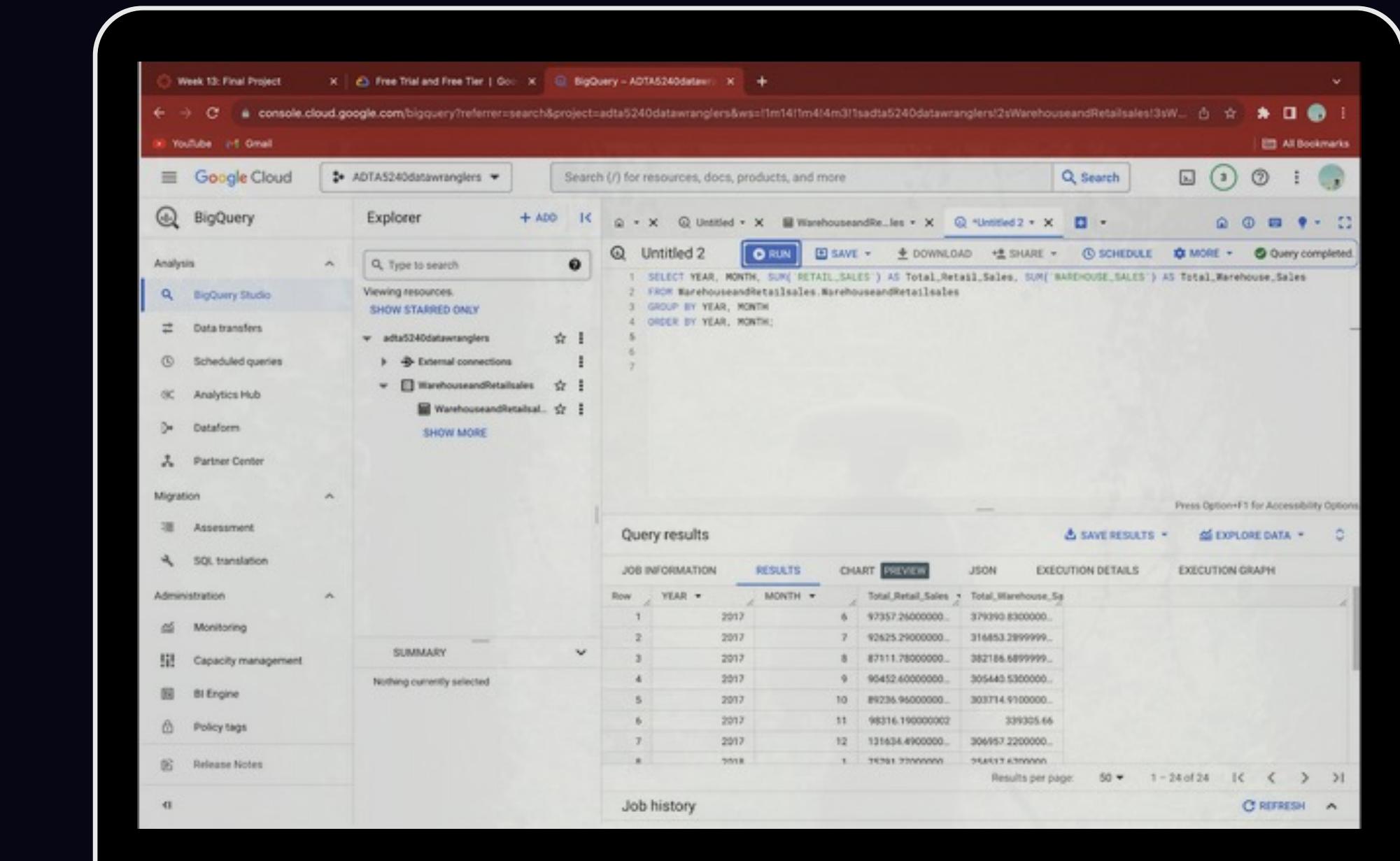
## Seasonal Sales Analysis:

5

Analyzes sales variations across different seasons, assisting in seasonal stock planning.

# 09

## Monthly Sales Summary:



The screenshot shows the Google BigQuery interface within the Google Cloud Platform. The sidebar on the left lists various services: Analysis (BigQuery Studio, Data transfers, Scheduled queries, Analytics Hub, Dataform, Partner Center), Migration, Administration (Monitoring, Capacity management, BI Engine, Policy tags, Release Notes). The main area displays a query titled "Untitled 2" with the following SQL code:

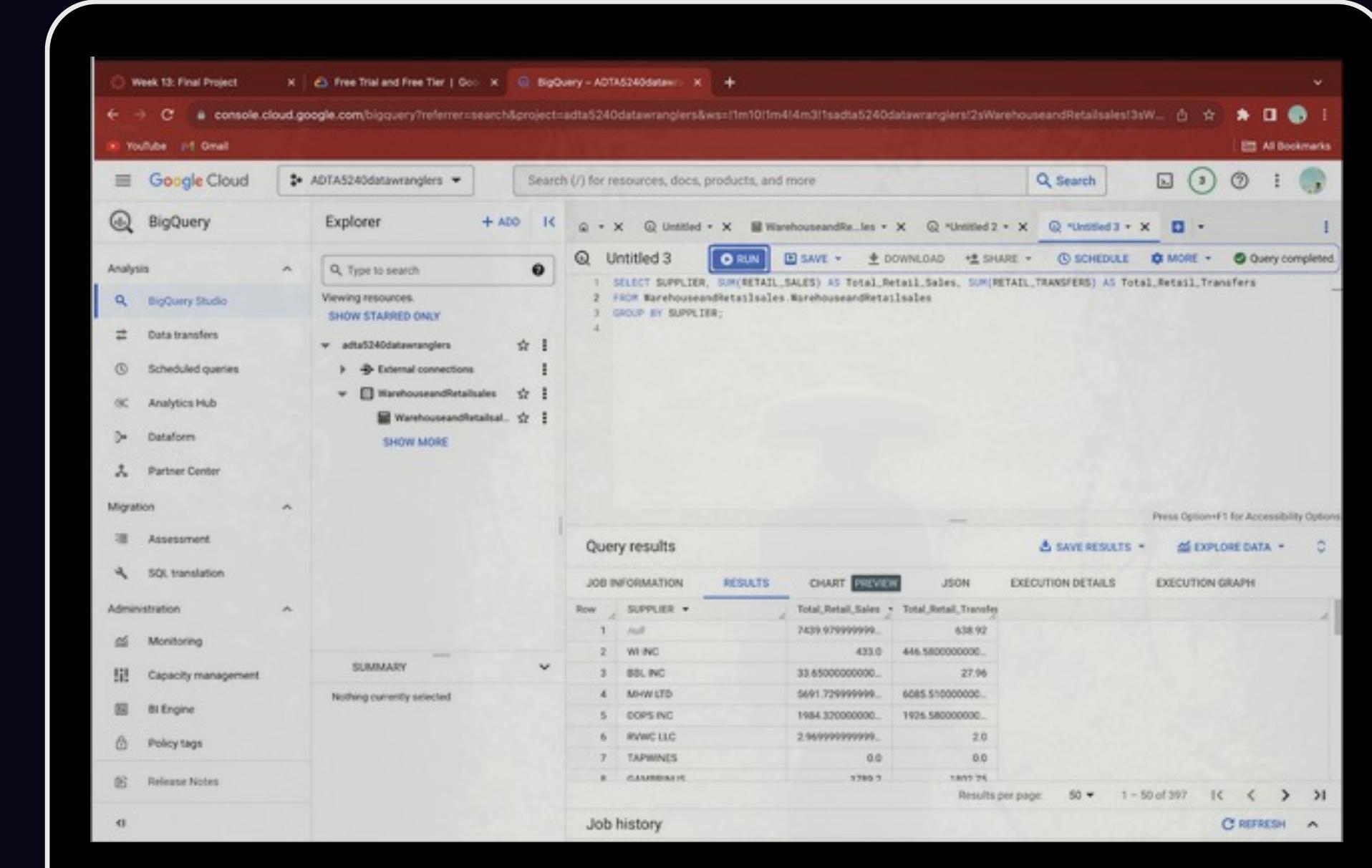
```
1 SELECT YEAR, MONTH, SUM( RETAIL_SALES ) AS Total_Retail_Sales, SUM( WAREHOUSE_SALES ) AS Total_Warehouse_Sales
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY YEAR, MONTH
4 ORDER BY YEAR, MONTH;
```

The "Query results" section shows a table with the following data:

Row	YEAR	MONTH	Total_Retail_Sales	Total_Warehouse_Sales
1	2017	6	97357.2600000...	379390.8300000...
2	2017	7	92625.2900000...	316853.2899999...
3	2017	8	87111.7800000...	382186.6899999...
4	2017	9	90452.6000000...	305449.5300000...
5	2017	10	89236.9600000...	303714.9100000...
6	2017	11	98316.1900000...	339305.66
7	2017	12	131634.4900000...	306957.2200000...
8	2018	1	76781.7700000...	316817.6700000...

# 10

## Supplier Performance Analysis:



The screenshot shows the Google BigQuery interface within the Google Cloud Platform. The left sidebar displays various services like BigQuery Studio, Data transfers, Scheduled queries, Analytics Hub, Dataform, and Partner Center. The main area shows a query titled "Untitled 3" with the following SQL code:

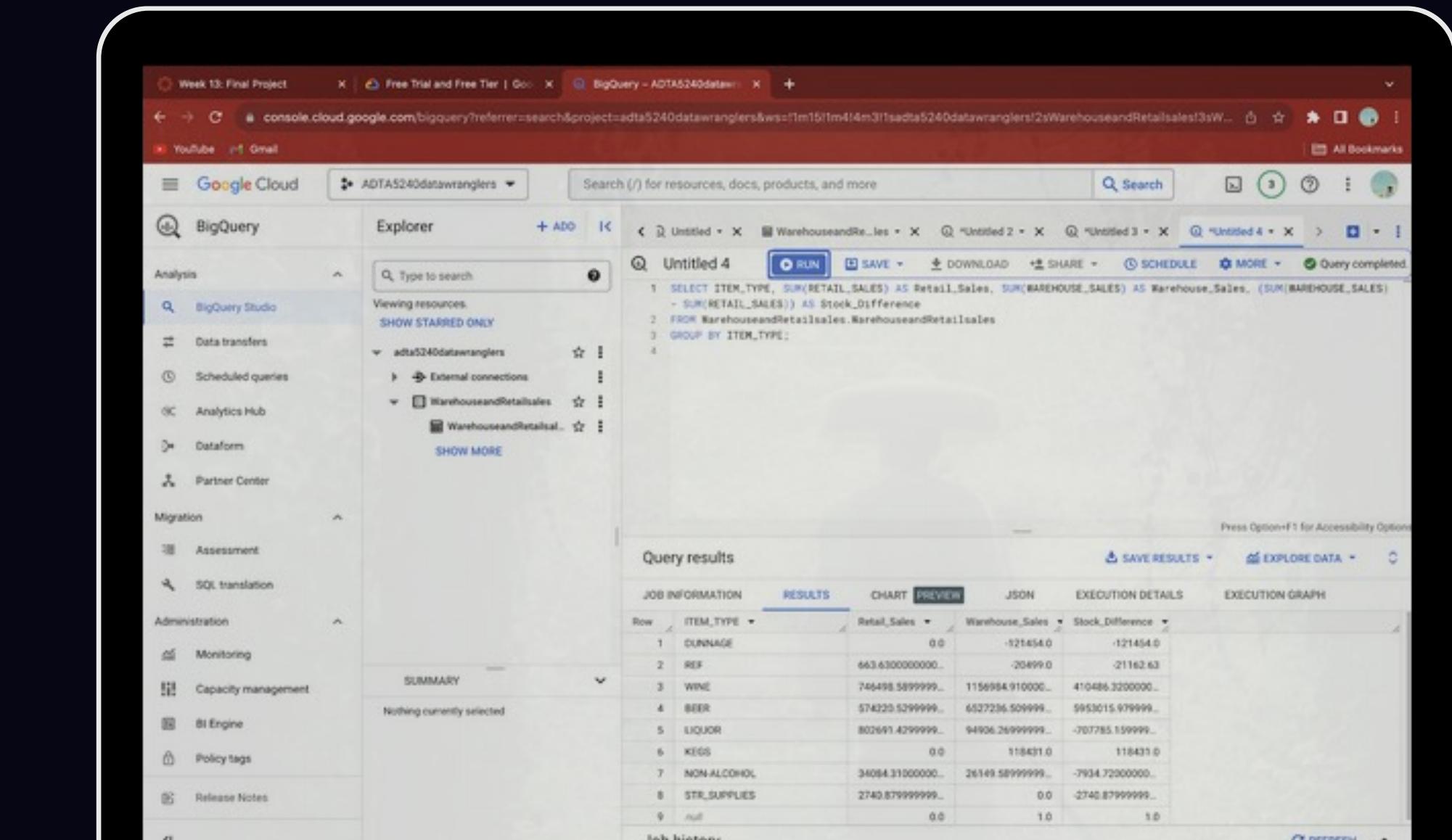
```
1 SELECT SUPPLIER, SUM(RETAIL_SALES) AS Total_Retail_Sales, SUM(RETAIL_TRANSFERS) AS Total_Retail_Transfers
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY SUPPLIER;
```

The "Query results" section displays the output of the query, showing data for 8 suppliers. The columns are Row, SUPPLIER, Total\_Retail\_Sales, and Total\_Retail\_Transfers.

Row	SUPPLIER	Total_Retail_Sales	Total_Retail_Transfers
1	null	7439.97999999...	638.92
2	WI INC	433.0	446.5800000000...
3	BBL INC	33.6500000000...	27.96
4	MHW LTD	5691.72999999...	6085.5100000000...
5	DOPS INC	1984.3200000000...	1926.5800000000...
6	IRVING LLC	2.969999999999...	2.0
7	TAPINNES	0.0	0.0
8	IGANBIRATE	1789.7	1897.76

# 11

## Item Type Sales and Stock Analysis:



The screenshot shows the Google Cloud BigQuery interface. On the left, the sidebar includes sections for Analysis, Data transfers, Scheduled queries, Analytics Hub, Dataform, Partner Center, Migration, Assessment, SQL translation, Administration, Monitoring, Capacity management, BI Engine, Policy tags, and Release Notes. The main area displays a query titled "Untitled 4" with the following SQL code:

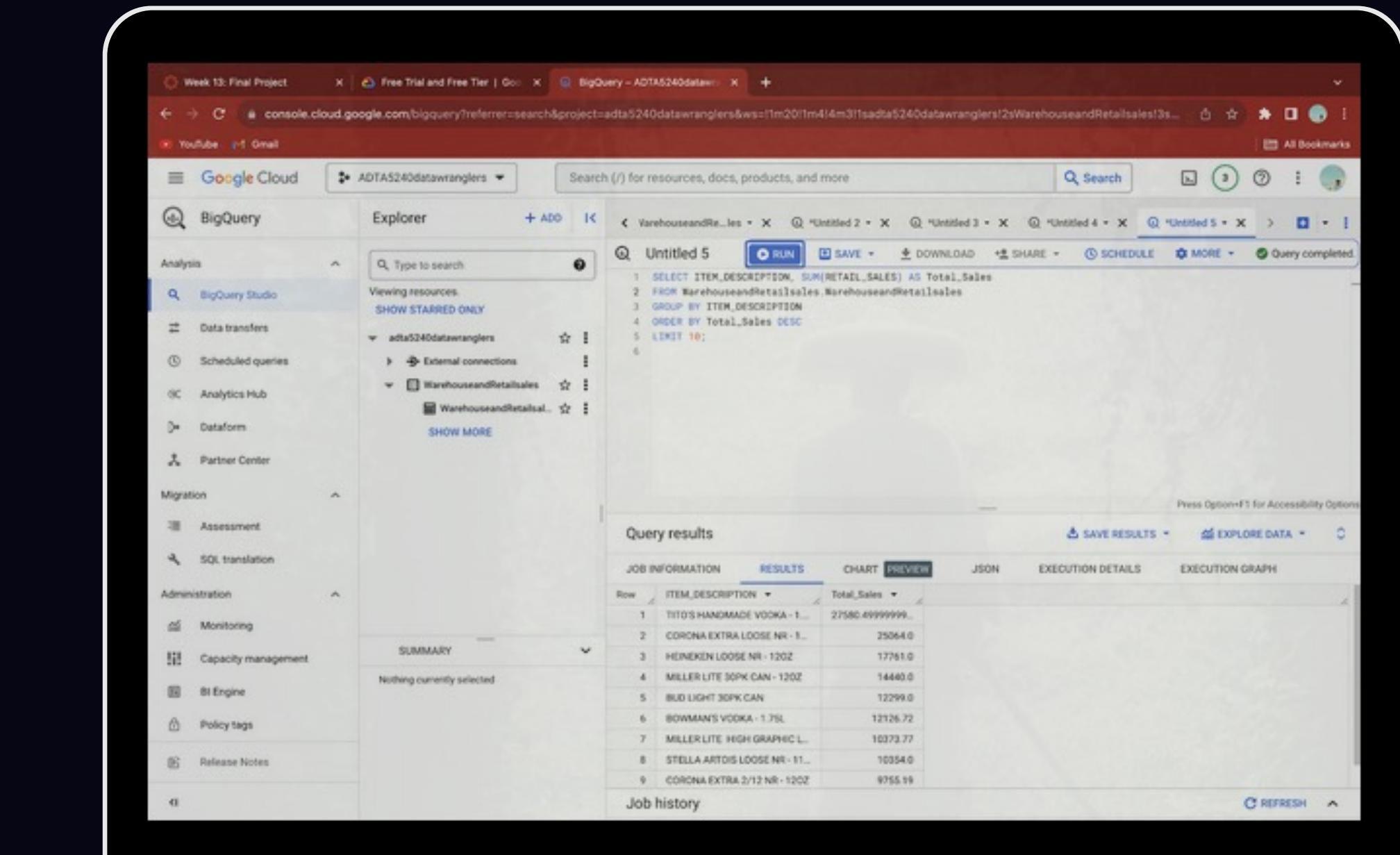
```
1 SELECT ITEM_TYPE, SUM(RETAIL_SALES) AS Retail_Sales, SUM(WAREHOUSE_SALES) AS Warehouse_Sales, (SUM(WAREHOUSE_SALES)
- SUM(RETAIL_SALES)) AS Stock_Difference
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY ITEM_TYPE;
```

The "Query results" section shows a table with the following data:

Row	ITEM_TYPE	Retail_Sales	Warehouse_Sales	Stock_Difference
1	CUNNAGE	0.0	-121454.0	-121454.0
2	REF	663.6300000000000	-20499.0	-21162.63
3	WINE	746498.589999999	1156984.910000000	410486.320000000
4	BEER	574220.529999999	6527236.50999999	5953015.97999999
5	LIQUOR	802691.429999999	94906.269999999	-707785.15999999
6	KEGS	0.0	118431.0	118431.0
7	NON-ALCOHOL	34084.310000000	26149.589999999	-7934.720000000
8	STR_SUPPLIES	2740.879999999	0.0	2740.879999999
9	null	0.0	1.0	1.0

# 12

## Identifying Top Selling Items:



The screenshot shows the Google Cloud BigQuery interface. On the left, the sidebar includes sections for Analysis, Data transfers, Scheduled queries, Analytics Hub, Dataform, Partner Center, Migration, Assessment, SQL translation, Administration, Monitoring, Capacity management, BI Engine, Policy tags, and Release Notes. The main area displays a query titled "Untitled 5" with the following SQL code:

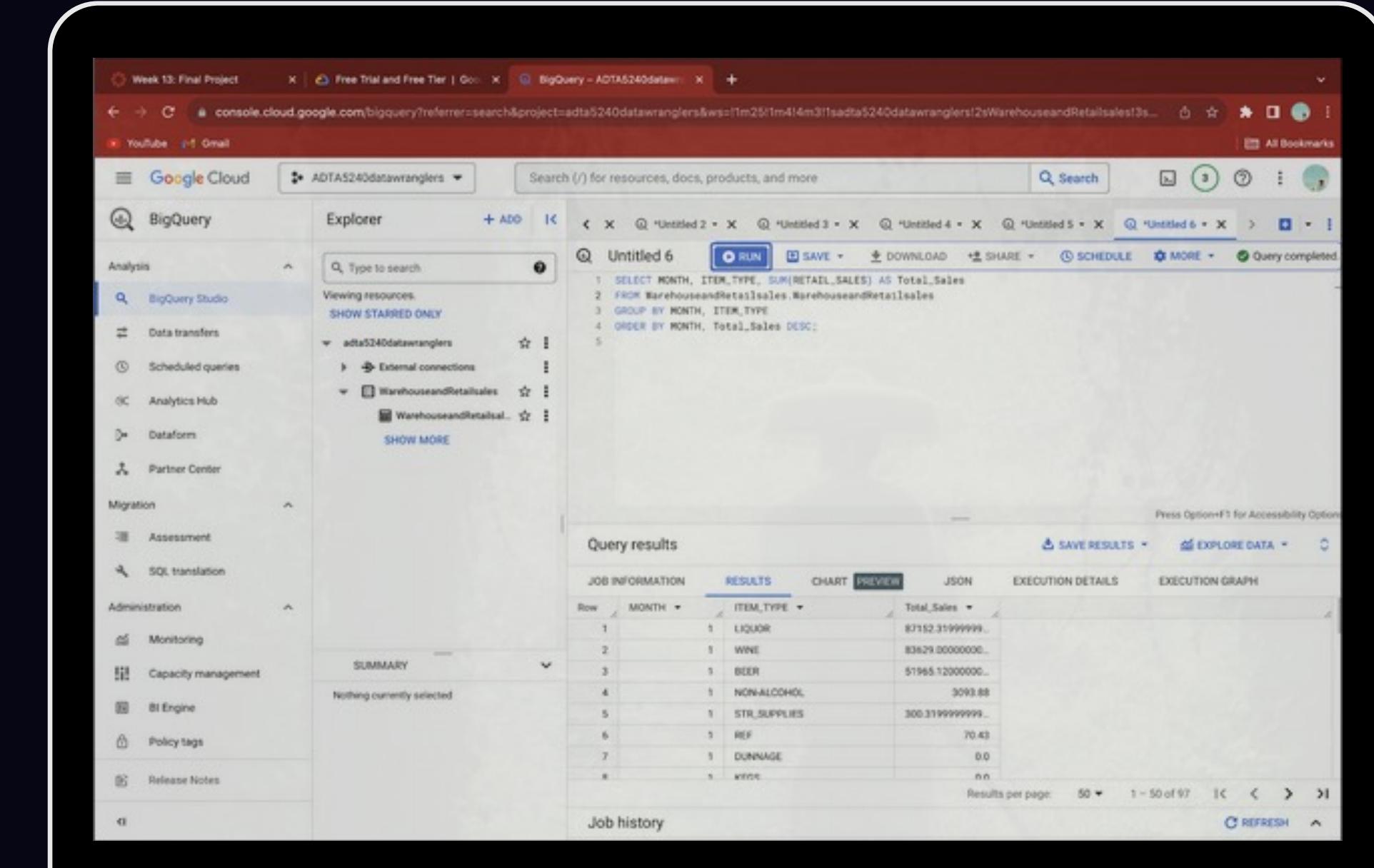
```
1 SELECT ITEM_DESCRIPTION, SUM(RETAIL_SALES) AS Total_Sales
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY ITEM_DESCRIPTION
4 ORDER BY Total_Sales DESC
5 LIMIT 10;
```

The "Query results" section shows the output of the query:

ITEM_DESCRIPTION	Total_Sales
TITO'S HANDMADE VODKA - 1L	27580.49999999999
CORONA EXTRA LOOSE NR - 1L	25064.0
HEINEKEN LOOSE NR - 12OZ	17781.0
MILLER LITE 30PK CAN - 12OZ	14440.0
BUD LIGHT 30PK CAN	12299.0
BOWMAN'S VODKA - 1.75L	12126.72
MILLER LITE HIGH GRAPHIC L...	10373.77
STELLA ARTOIS LOOSE NR - 1L	10354.0
CORONA EXTRA 2/12 NR - 12OZ	9755.19

# 13

## Seasonal Sales Analysis:



The screenshot shows the Google BigQuery interface within a web browser. The sidebar on the left lists various Google Cloud services, with 'BigQuery' selected. The main area displays a query titled 'Untitled 6'.

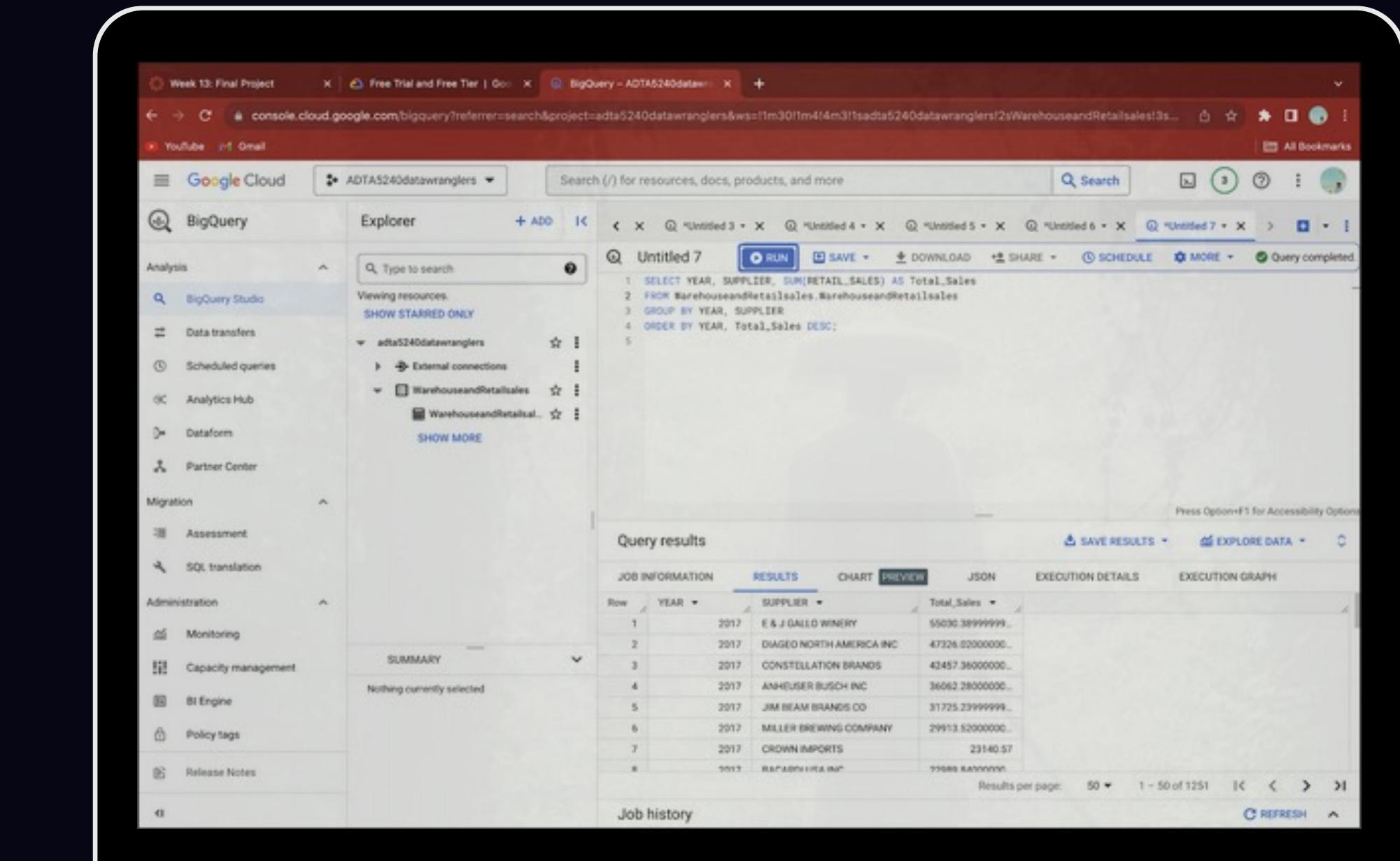
```
1 SELECT MONTH, ITEM_TYPE, SUM(RETAIL_SALES) AS Total_Sales
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY MONTH, ITEM_TYPE
4 ORDER BY MONTH, Total_Sales DESC;
```

The 'Query results' section shows the output of the query:

Row	Month	Item_Type	Total_Sales
1	1	LIQUOR	87152.31999999...
2	1	WINE	83629.00000000...
3	1	BEER	51965.12000000...
4	1	NON-ALCOHOL	3093.88
5	1	STR_SUPPLIES	300.3199999999...
6	1	REF	70.43
7	1	DUNNAGE	0.0
8	1	WATER	0.0

# 14

## Yearly Supplier Contribution:



The screenshot shows the Google Cloud BigQuery interface. On the left, the sidebar lists various services: Analysis, BigQuery Studio, Data transfers, Scheduled queries, Analytics Hub, Dataform, Partner Center, Migration, Assessment, SQL translation, Administration, Monitoring, Capacity management, BI Engine, Policy tags, and Release Notes. The main area displays a query titled "Untitled 7" with the following SQL code:

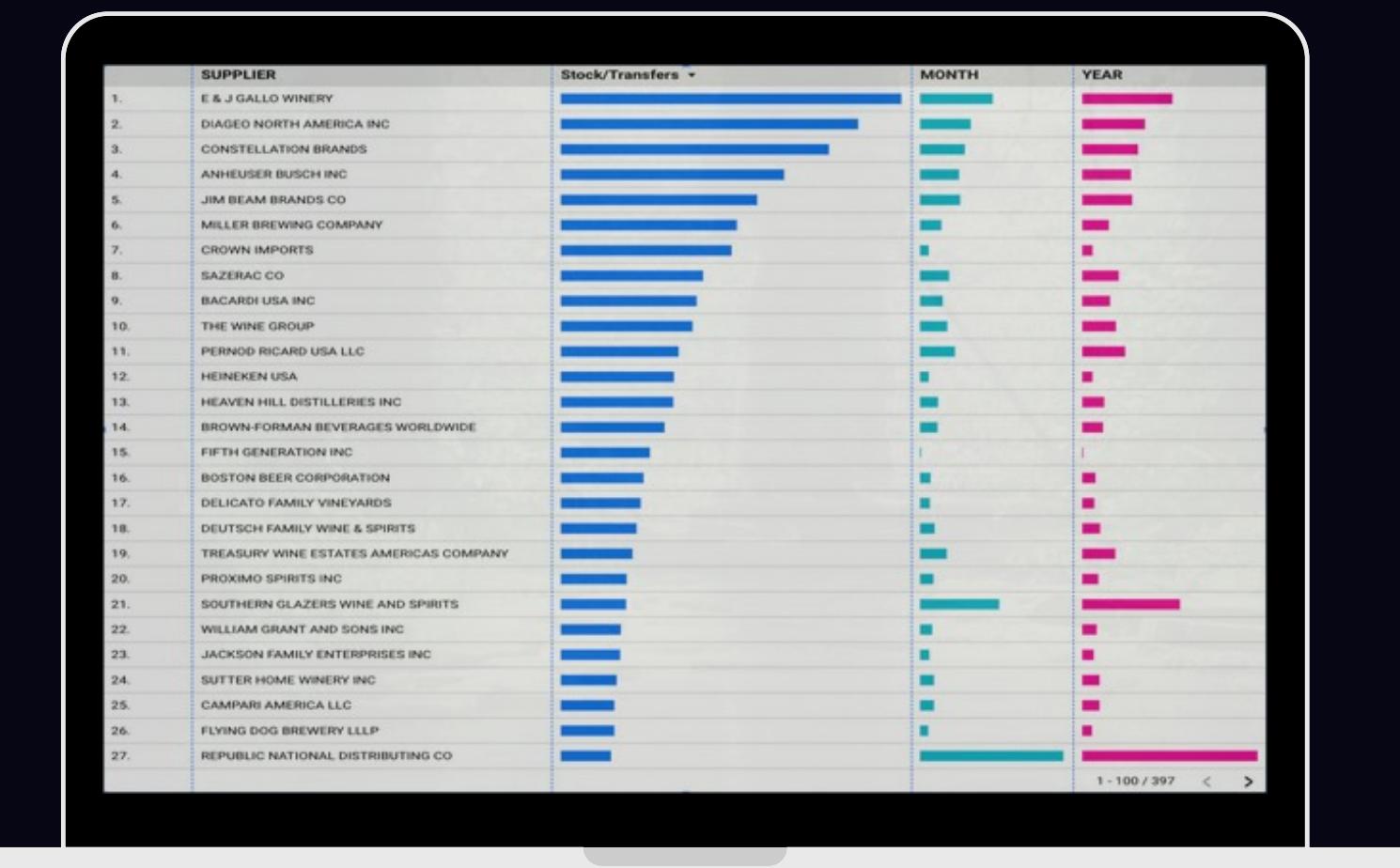
```
1 SELECT YEAR, SUPPLIER, SUM(RETAIL_SALES) AS Total_Sales
2 FROM WarehouseandRetailSales.WarehouseandRetailSales
3 GROUP BY YEAR, SUPPLIER
4 ORDER BY YEAR, Total_Sales DESC;
```

The "Query results" section shows a table with the following data:

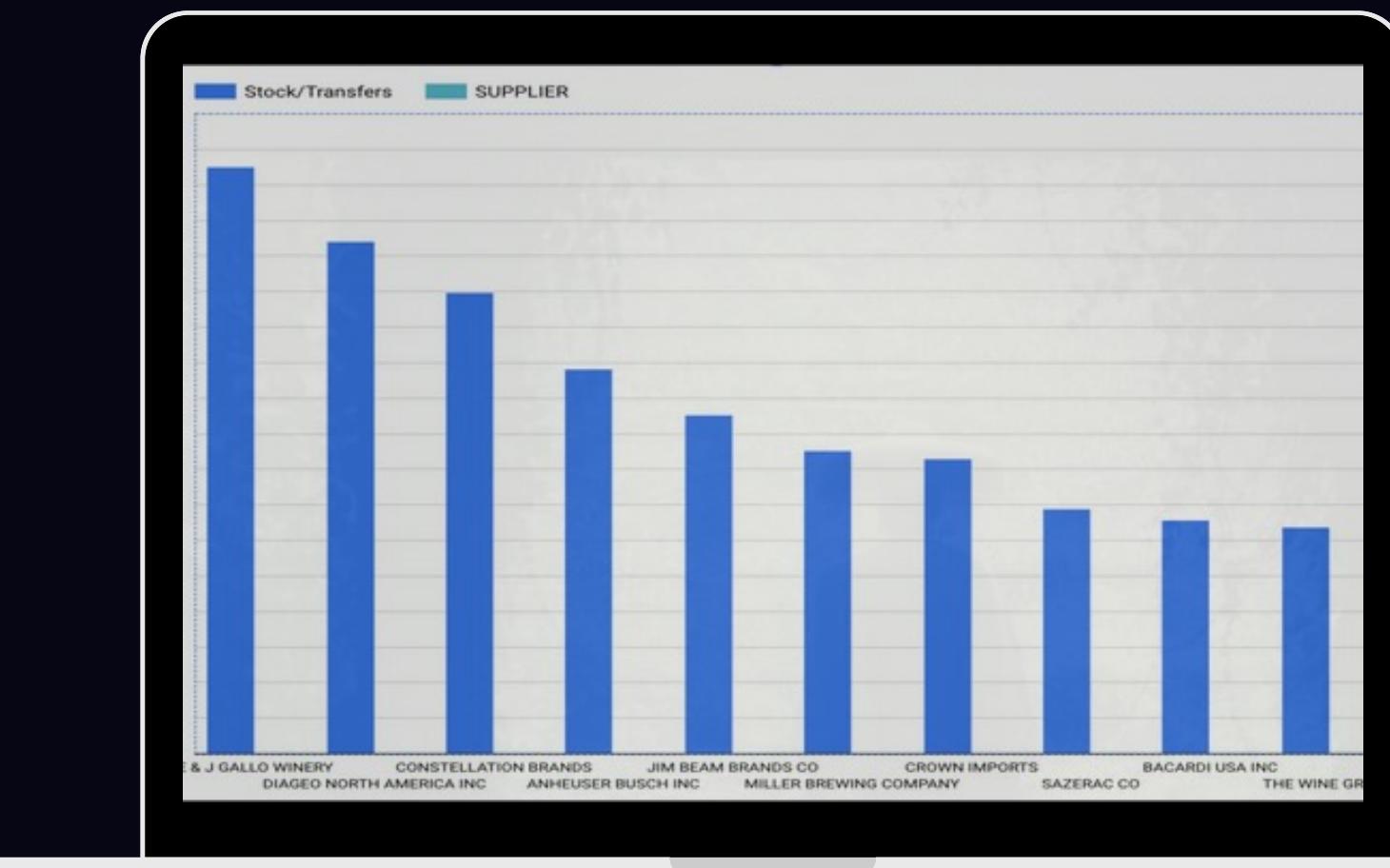
Row	YEAR	SUPPLIER	Total_Sales
1	2017	E & J GALLO WINERY	55030.38999999
2	2017	DIAGEO NORTH AMERICA INC	47326.02000000
3	2017	CONSTELLATION BRANDS	43457.36000000
4	2017	ANHEUSER BUSCH INC	36682.28000000
5	2017	JIM BEAM BRANDS CO	31725.23999999
6	2017	MILLER BREWING COMPANY	29913.52000000
7	2017	CROWN IMPORTS	23140.57
8	2017	BUCHANAN & CO INC	17988.84000000

# 15

## Monthly Sales Summary

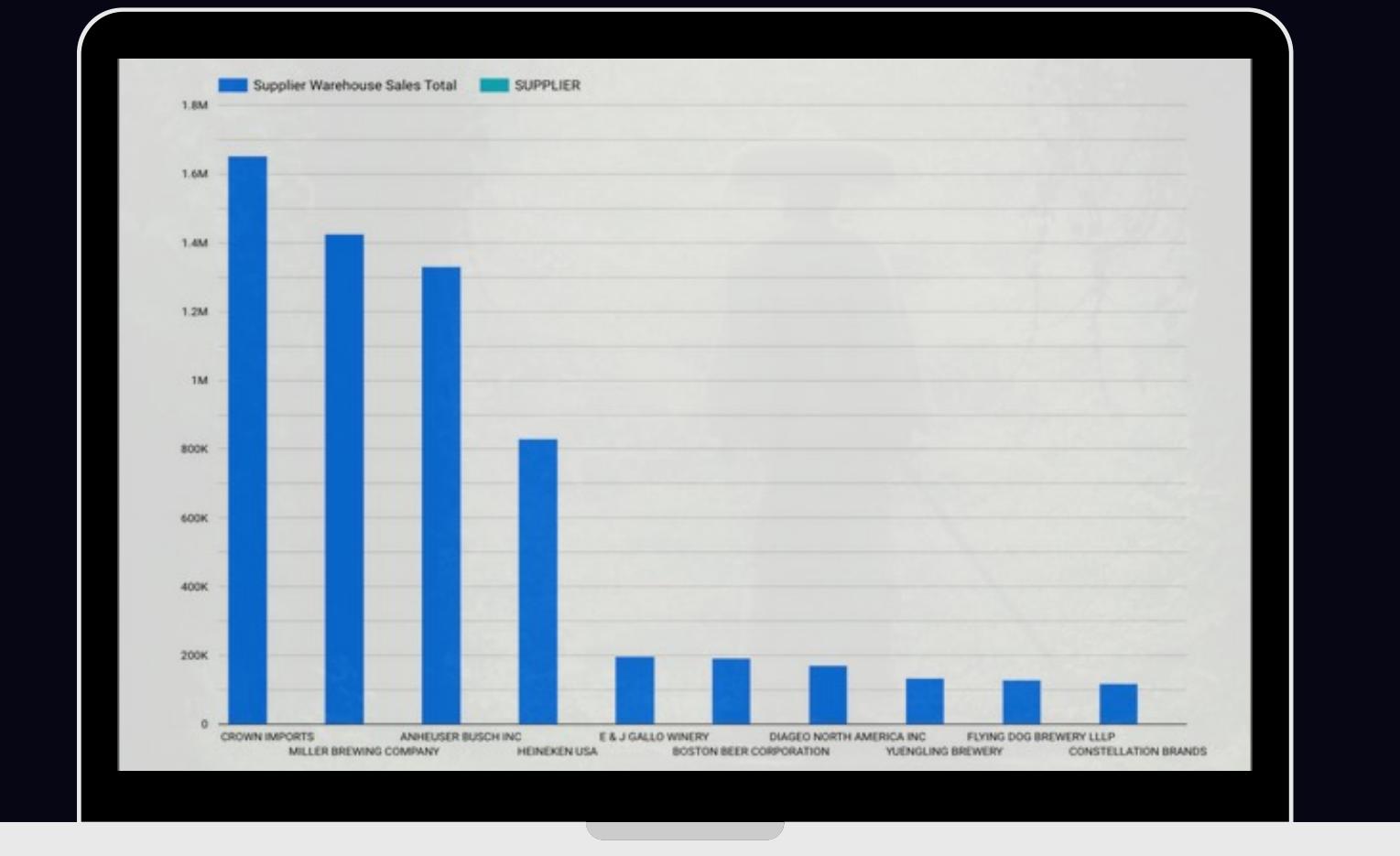


## Supplier Performance Analysis

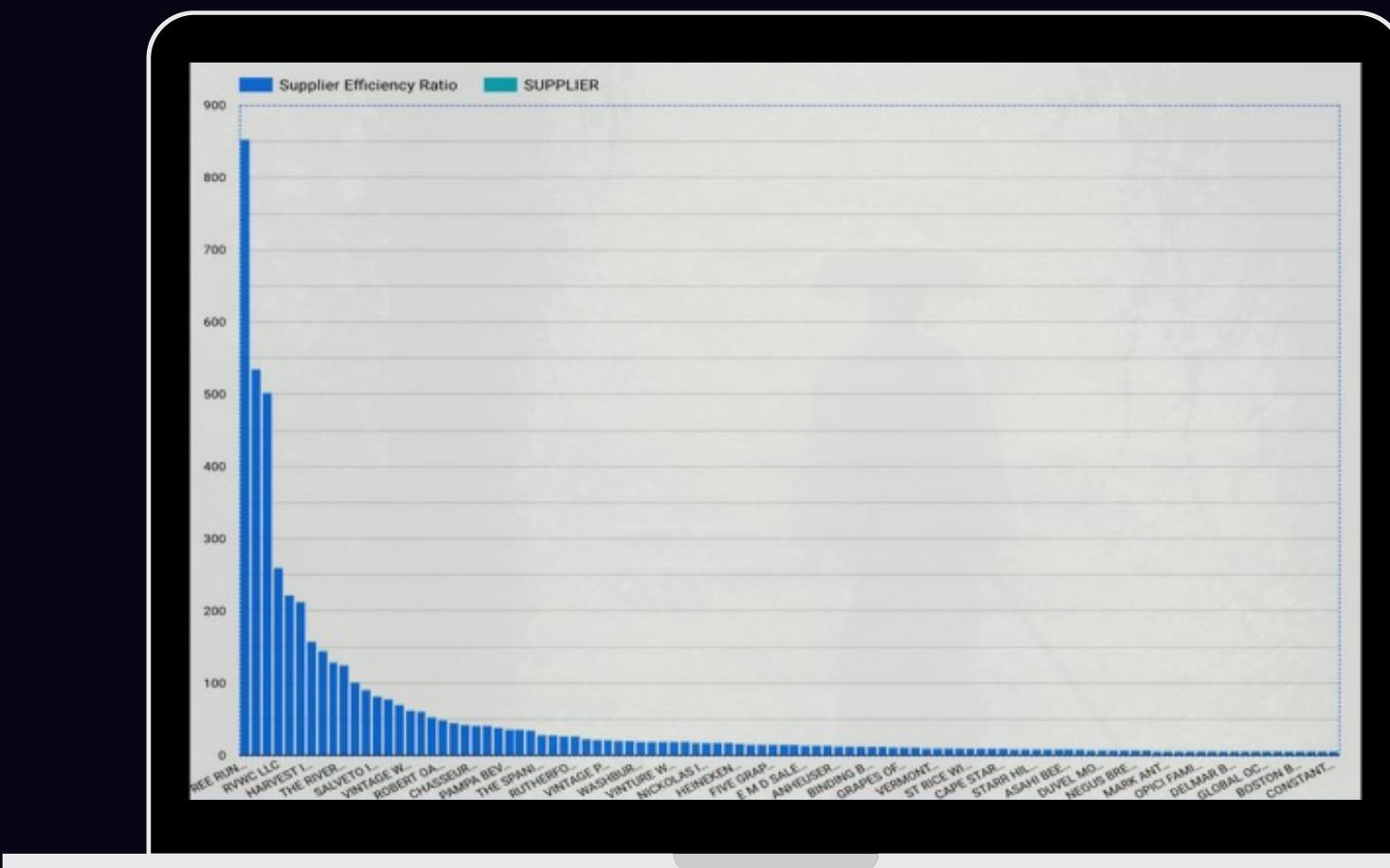


# 16

Supplier Warehouse  
Sales Total:

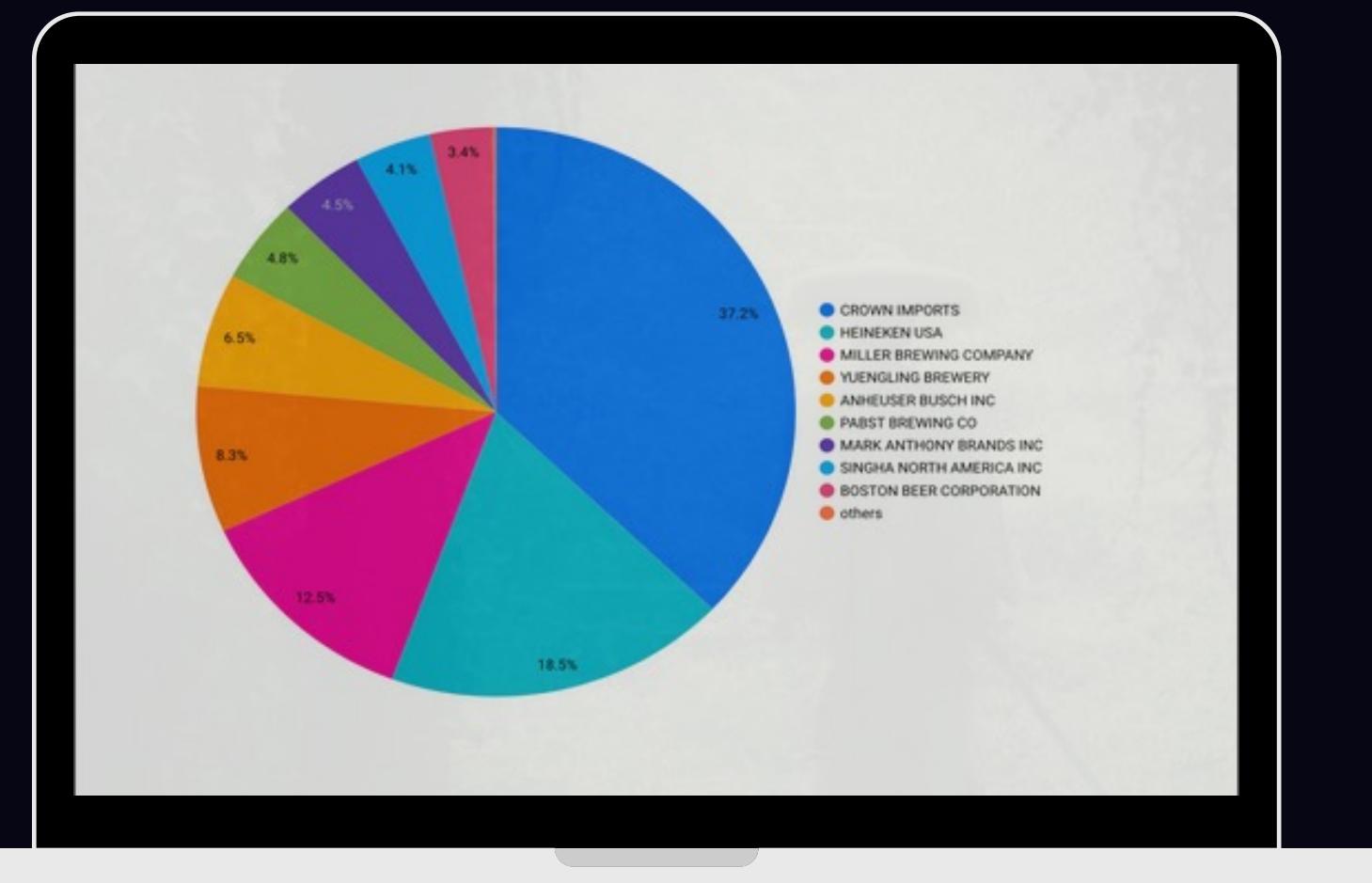


Supplier Efficiency  
Ratio:

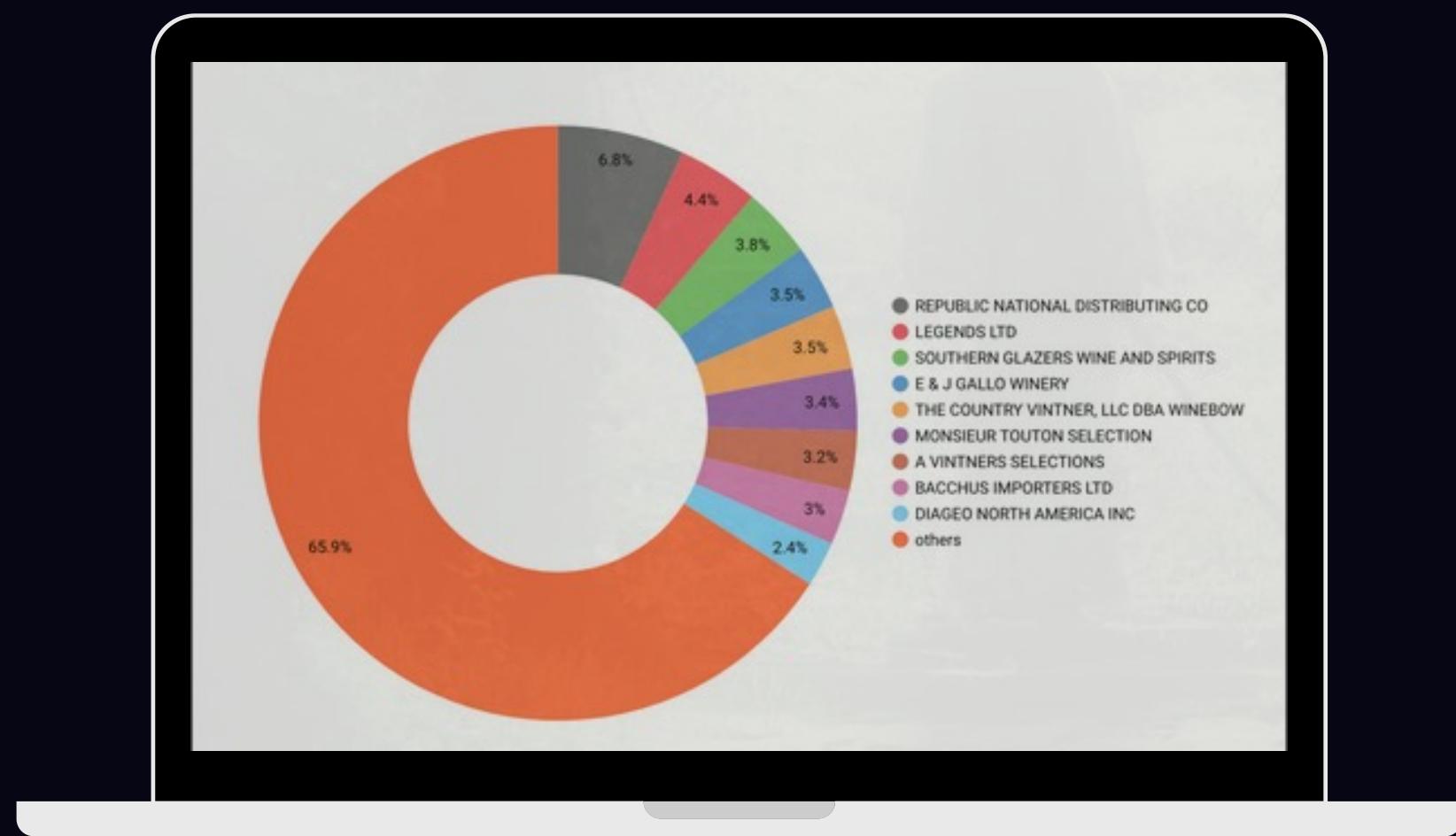


# 17

## Average Warehouse Sales per Supplier



## Average Monthly Transfers



# Interpretation of Analysis:

18

- Understanding Data Patterns
- Identifying Key Insights
- Evaluating Supplier Performance
- Analyzing Sales and Stock for Different Item Types
- Actionable Decision-Making
- Future Planning



## Conclusion AND KEY ACHIEVEMENTS:

### CONCLUSION

The implementation of this data-driven approach in inventory management represents a significant advancement in addressing the challenges faced by retail and warehousing sectors.

### KEY ACHIEVEMENTS

- Improved operational effectiveness.
- Enhanced customer satisfaction.

# Next Steps:

## CONTINUOUS MONITORING AND REFINEMENT:

- Ongoing efforts in monitoring and refining the analysis.
- Ensuring the relevance and accuracy of insights.

## REAL-TIME DATA ANALYTICS INTEGRATION:

- Focus on integrating real-time data analytics.
- Enables dynamic decision-making based on current trends.

## EXPLORATION OF ADVANCED ANALYTICAL TECHNIQUES:

- Focus on integrating real-time data analytics.
- Enables dynamic decision-making based on current trends.

# References:

- ZHEN, L., & LI, H. (2022). A literature review of smart warehouse operations management. *Front. Eng. Manag.* 2022, 9(1): 31-55. <https://doi.org/10.1007/s42524-021-0178-9>
- B., S. S. T., & S., N. (2018). Warehouse inventory management system using IoT and open-source framework. *Alexandria Engineering Journal* Volume 57, Issue 4, December 2018, Pages 3817-3823. <https://doi.org/10.1016/j.aej.2018.02.003>
- Liu, X., Cao, J., Yang, Y., & Jiang, S. (2018). CPS-Based Smart Warehouse for Industry 4.0: A Survey of the Underlying Technologies. *Computers* 2018, 7(1), 13;. <https://doi.org/10.3390/computers7010013>

22



THANK  
YOU