WAREHOUSE AND RETAIL LIQUOR SALES

This dataset contains a list of sales data of liquor items from Montgomery County of Maryland and has been obtained from Data.gov, which is the United States government’s open data website. It serves as a central repository for a wide range of datasets published by various federal agencies. The task here is to analyse this data to provide actionable insights that can help the company improve its operations and profitability.

Project Description:

The goal of this project is to analyze liquor sales data to understand trends, identify patterns, and make informed business decisions. We’ll focus on both retail (in-store) and warehouse (distribution centre) sales.

Approach:

The dataset consists of 9 columns and 307646 rows, with each column representing a unique attribute, namely:

'YEAR': Year of the sales record.

'MONTH': Month of the sales record.

'SUPPLIER': Name of the supplier.

'ITEM CODE': Unique code for each item.

'ITEM DESCRIPTION': Description of the item.

'ITEM TYPE': Type of the item.

'RETAIL SALES': Sales made through physical stores.

'RETAIL TRANSFERS': Transfers between stores.

'WAREHOUSE SALES': Sales made through the online warehouse.

The initial approach involves cleaning and pre-processing the data to ensure it is well-structured and usable. This process includes handling missing data points, correcting errors, and removing or replacing duplicate values. After data cleaning and preprocessing in Microsoft Excel, the next steps to derive insights typically include detailed data analysis using pivot tables and functions, creating visualizations like charts and graphs to identify trends and outliers, segmenting the data into meaningful groups, analyzing trends over time, examining correlations to understand relationships between variables, and compiling the findings into comprehensive reports or dashboards. These steps help transform raw data into valuable insights for strategic decision-making and improving process to drive up sales.

The modifications made during data cleaning and pre-processing is mentioned below:

Handling missing data: The missing data from the data table was treated as NULL value.  By converting missing data to null values, have maintained consistency in the dataset. It ensures that all missing values are represented uniformly.

Handling duplicate values: All duplicate identical rows have been removed

Handling error values: In column item type, ‘STR\_SUPPLIES’ stands for various novelty products such as paper bags, totes, shot glasses and beer mugs, etc. These do not necessarily contribute to liquor sales. Although this has been retained while conducting analysis, it is to be noted that it carries majorly null values as supplier. ‘Default’ has also been listed as supplier for this, which has been replaced with Null to keep the values consistent. Since they sell novelty products, their warehouse sales is consistently nill.

Secondly, sales records of a few months have been missed each year from 2017 to 2020. So, insights have been obtained based only on partial data provided by the dataset.

Tech-Stack Used:

* MySQL workbench

MySQL Workbench serves as a robust tool for managing MySQL databases. It allows data extraction, transformation, and exploration through SQL queries. Analysts can clean and structure data, create views, and export results for further analysis.

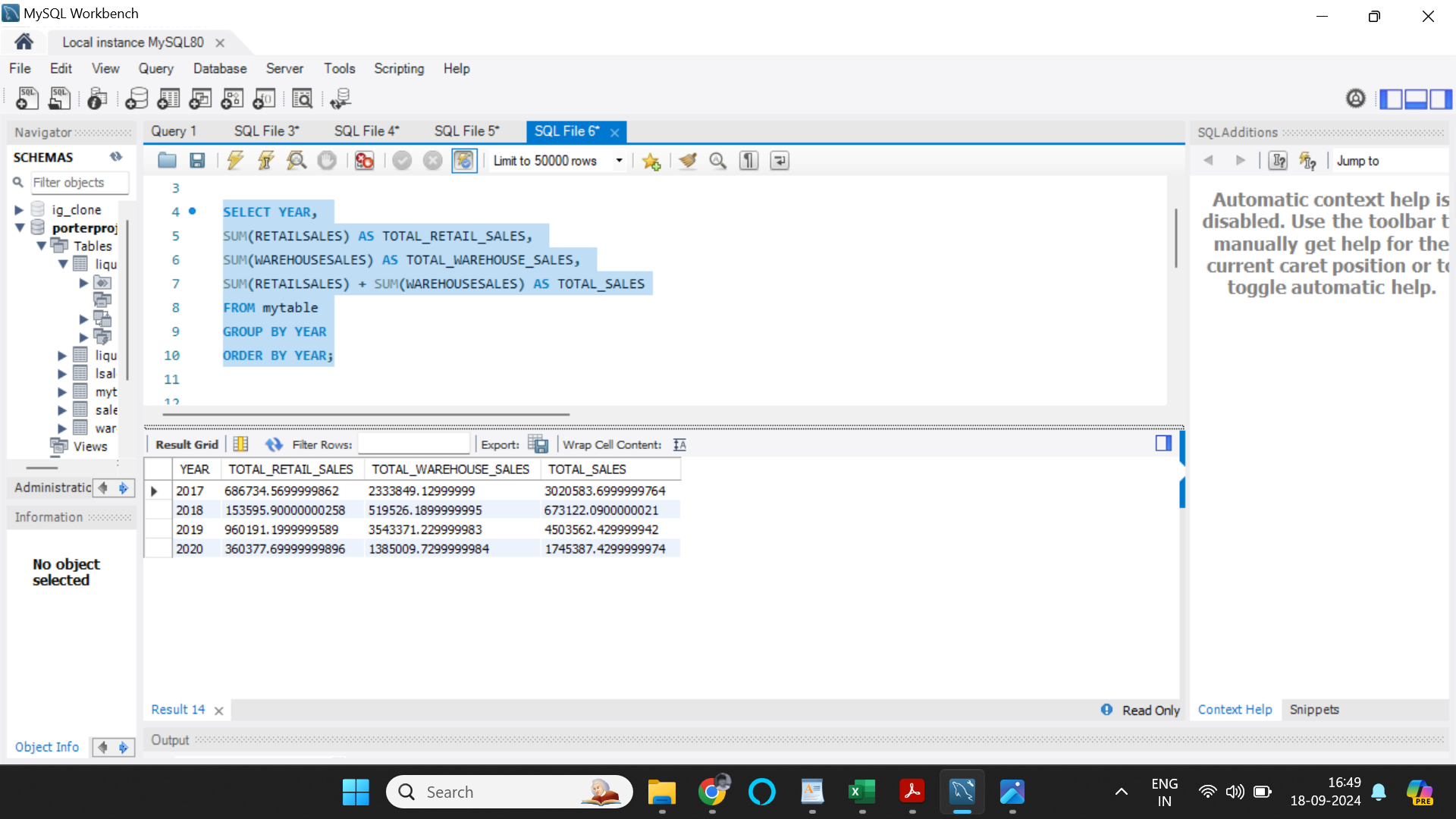
* Microsoft Excel

Microsoft Excel is a versatile spreadsheet application. It plays a crucial role in sales analysis by importing data, cleaning it, and performing basic calculations and to create charts and visualizations.

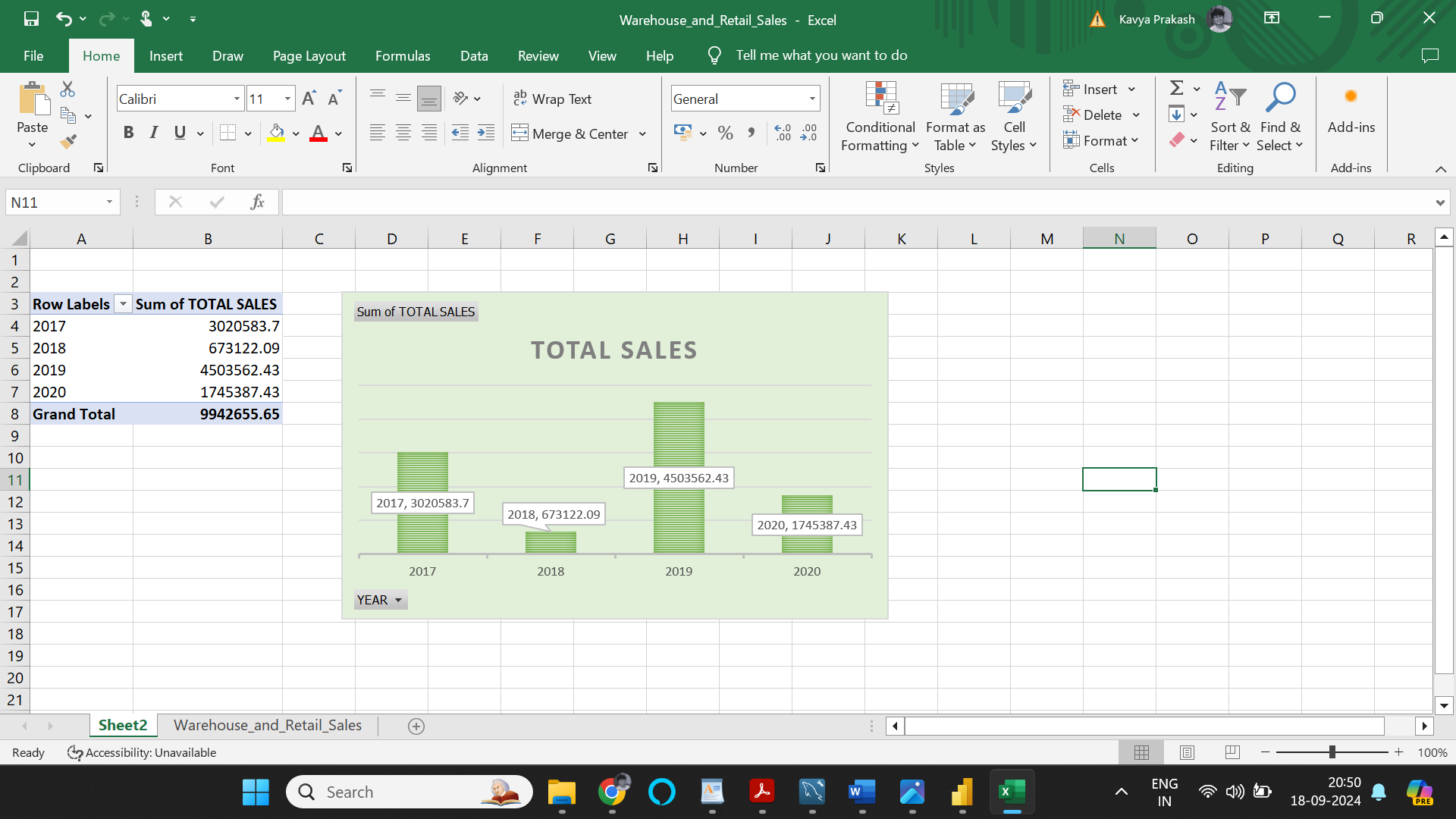
Exploratory data analysis:

EDA allows us to represent different sales trends graphically and visualize data related to best-selling product categories, buyer demographics, preferences, customer spending patterns, and units sold over a certain period.  By exploring the data, we can identify unusual sales spikes, and recurring patterns. For example, EDA can reveal whether certain liquor items consistently sell better during specific seasons or events.

1. Calculate and visualise each year's total sales (retail and warehouse):

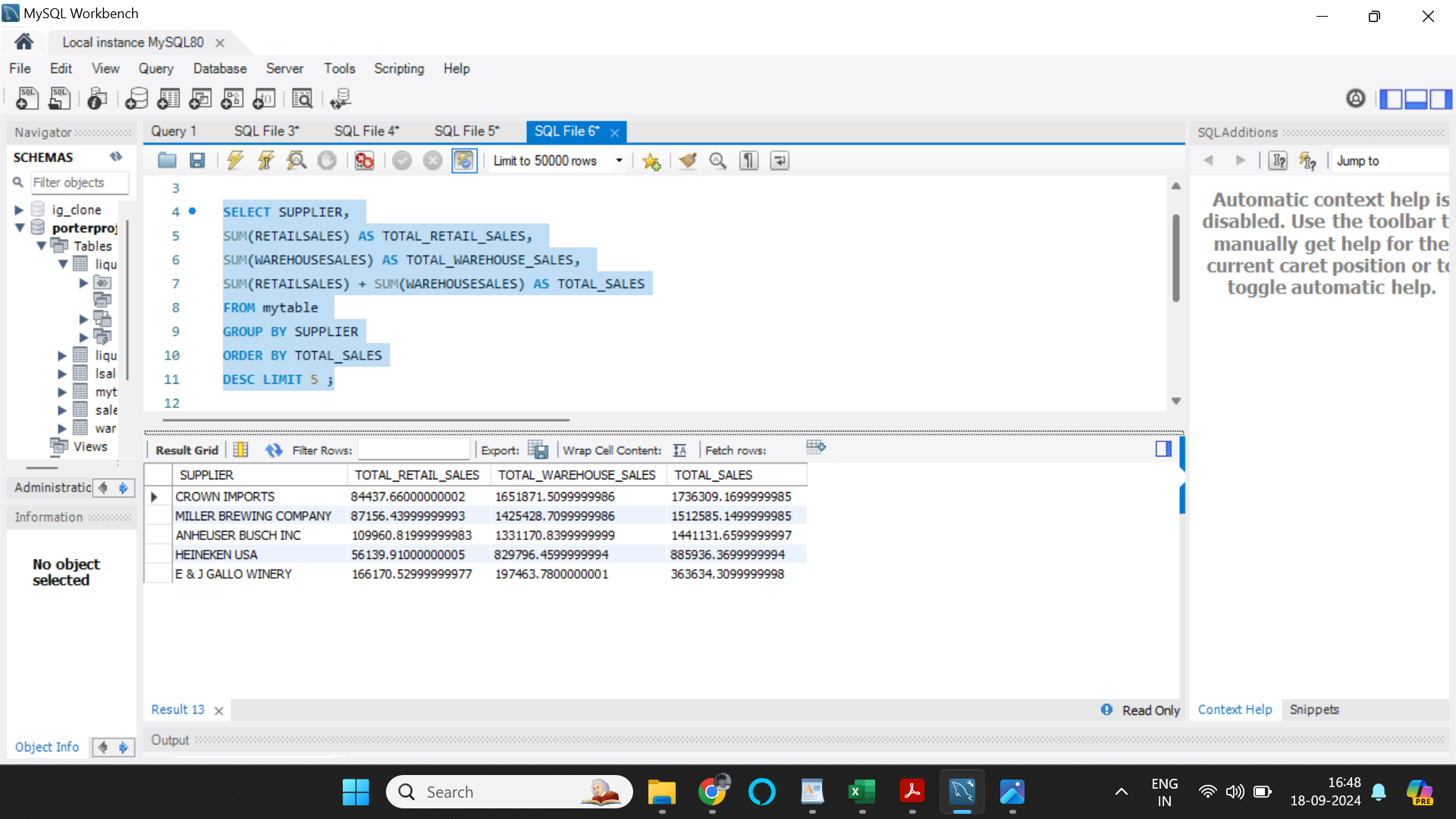


To visualize the data, I have created a pivot table using Microsoft Excel and have used the pivot table to create a bar chart depicting the total sales through the years, from 2017 to 2020.

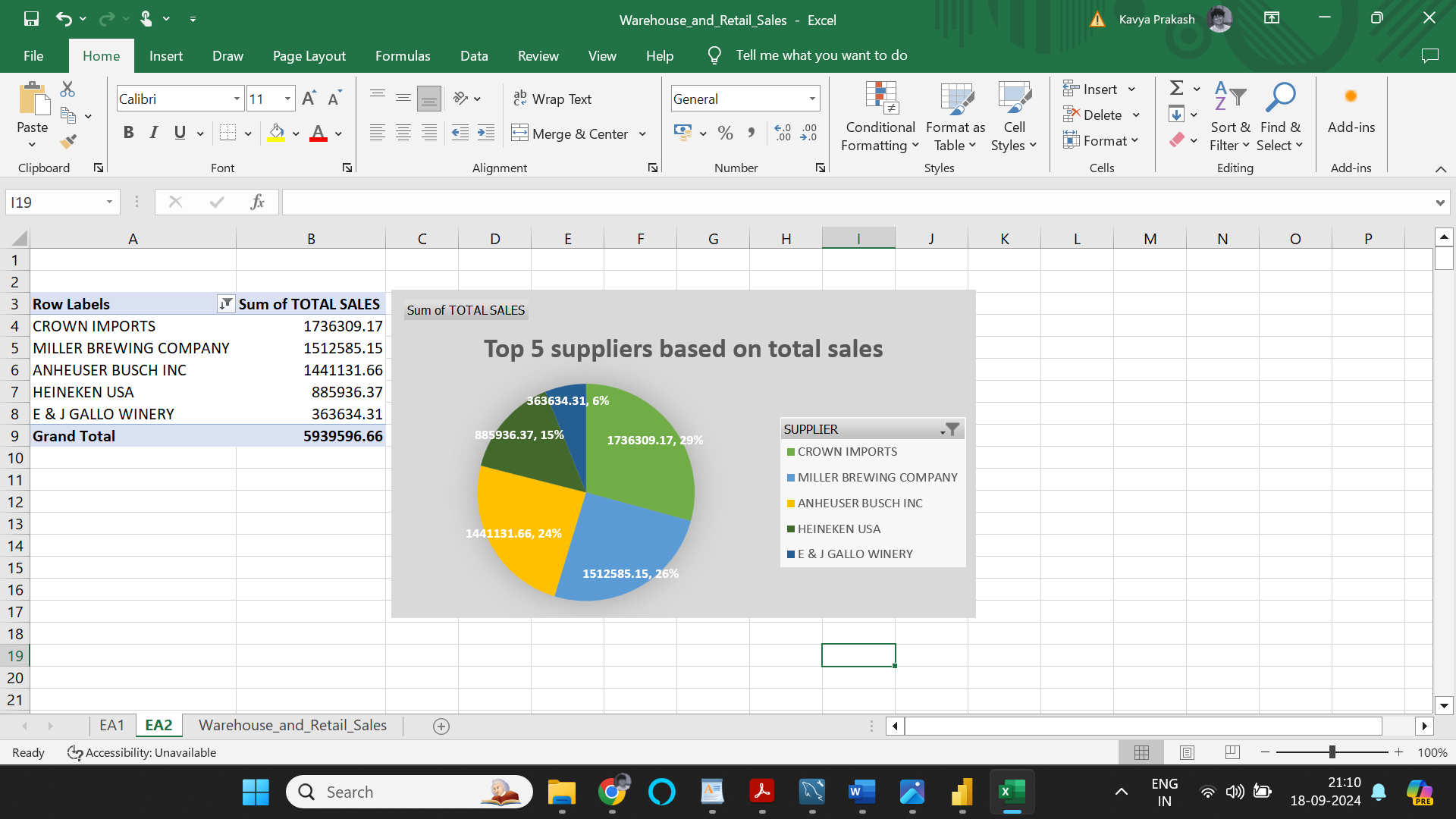


Insights: As per the data avilable, major sales has occoured in the year of 2019, while 2018 experienced lowest sales records. Unfortunately, in the year 2018, data of only two months are available in the dataset provided, whereas, 2019 had data from 11 months of the year

1. Determine the top 5 suppliers based on total sales (both retail and warehouse) for the entire dataset.

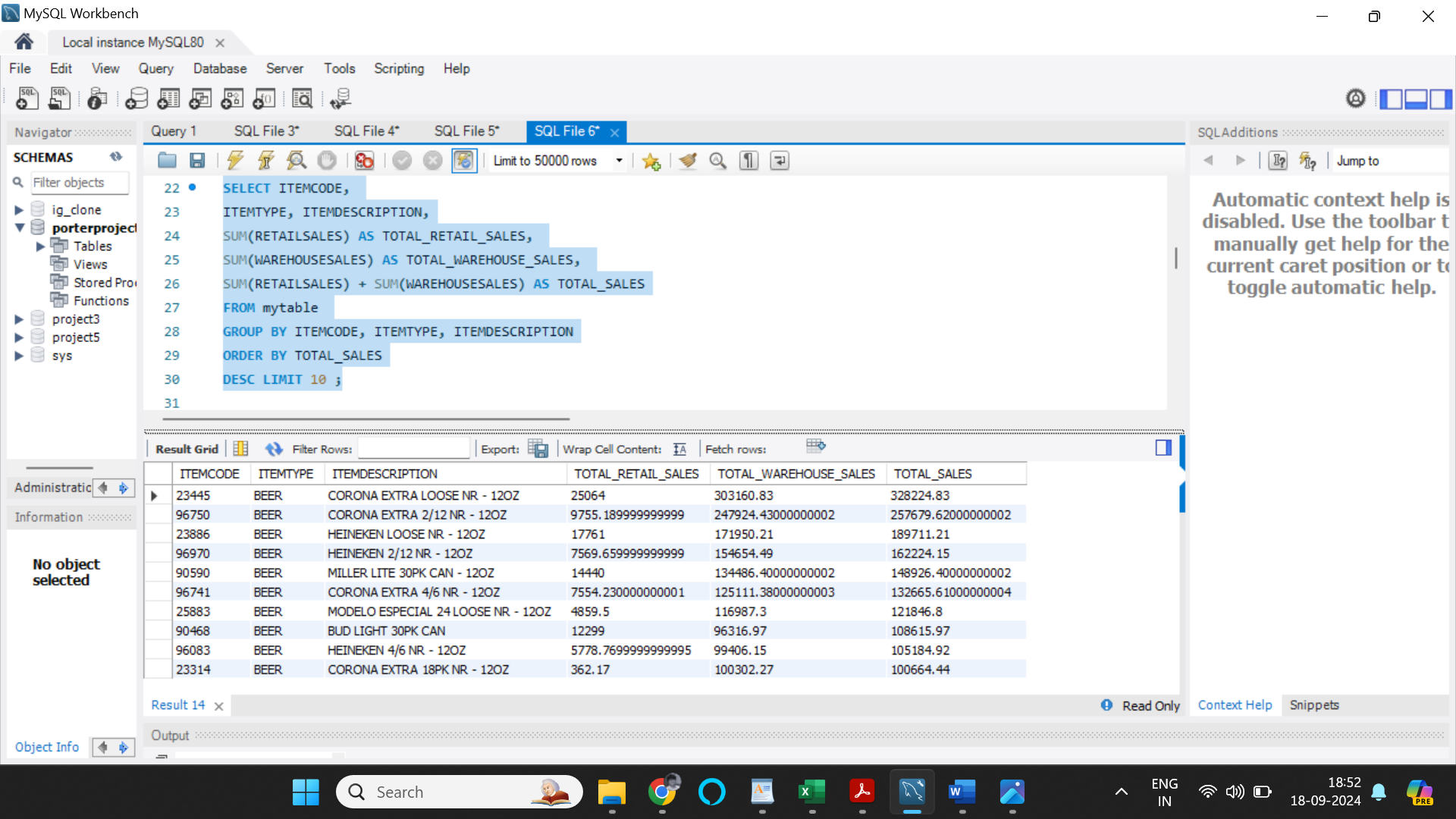


To visualize the data, I have created a pivot table using Microsoft Excel and have used the pivot table to create a pie chart depicting the total sales based on suppliers, and filtered to only show top 5 suppliers with highest sales records.

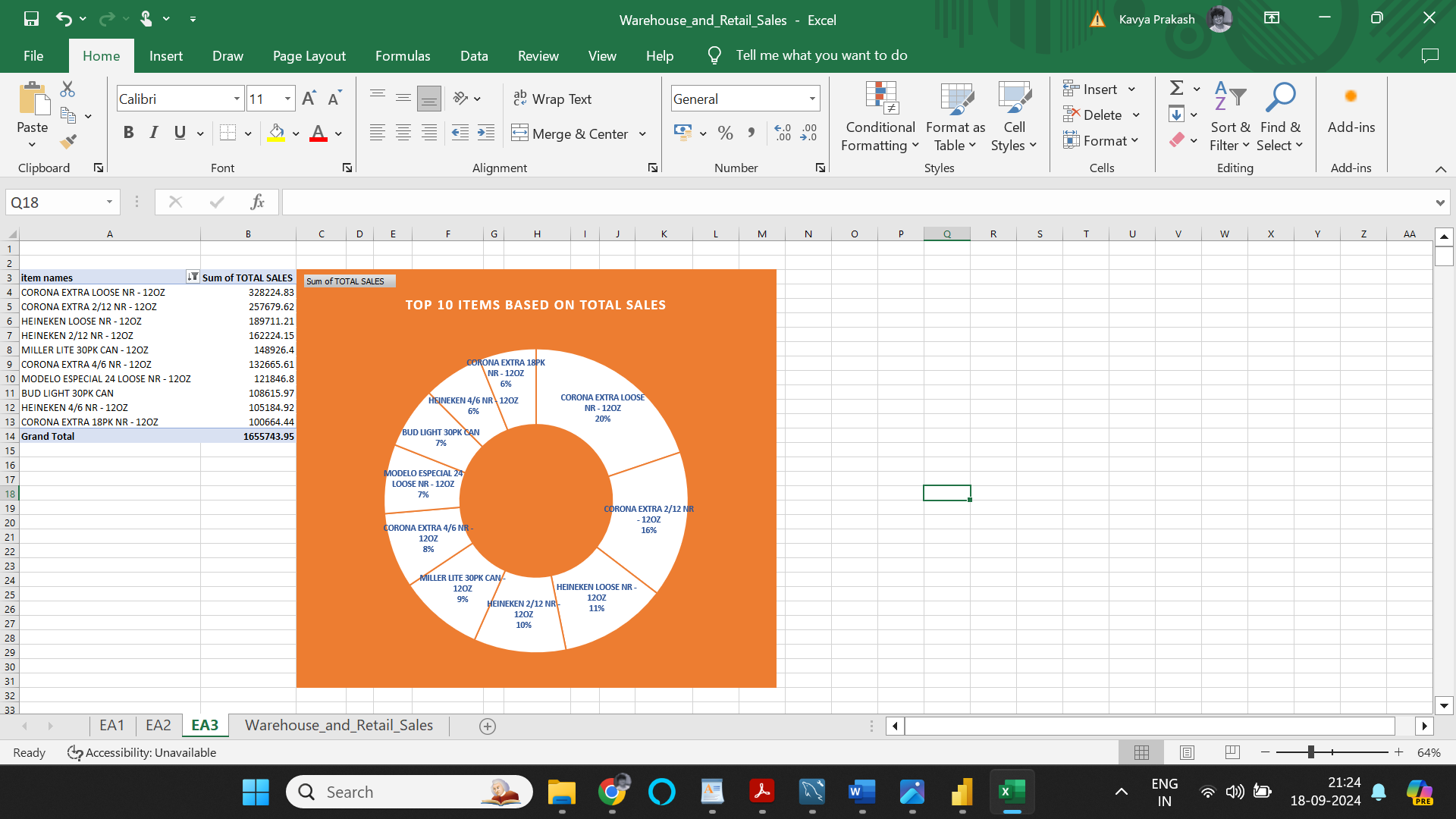


Insights: Clearly, major chunk of the sales was from ‘Crown Imports’ who made total sales of 1736309.17, which is roughly 29% of the total sales among the top 5 suppliers, closely followed by ‘Miller brewing company’ at 26% of total sales among the top 5 suppliers. Incidentally, both suppliers sell Beers and Kegs

1. Identify the top 10 best-selling items (based on total sales) and provide their descriptions and types



To visualize the data, I have created a pivot table using Microsoft Excel and have used the pivot table to create a pie chart depicting the total sales based on item type, and filtered to only show top 10 suppliers with highest sales records.



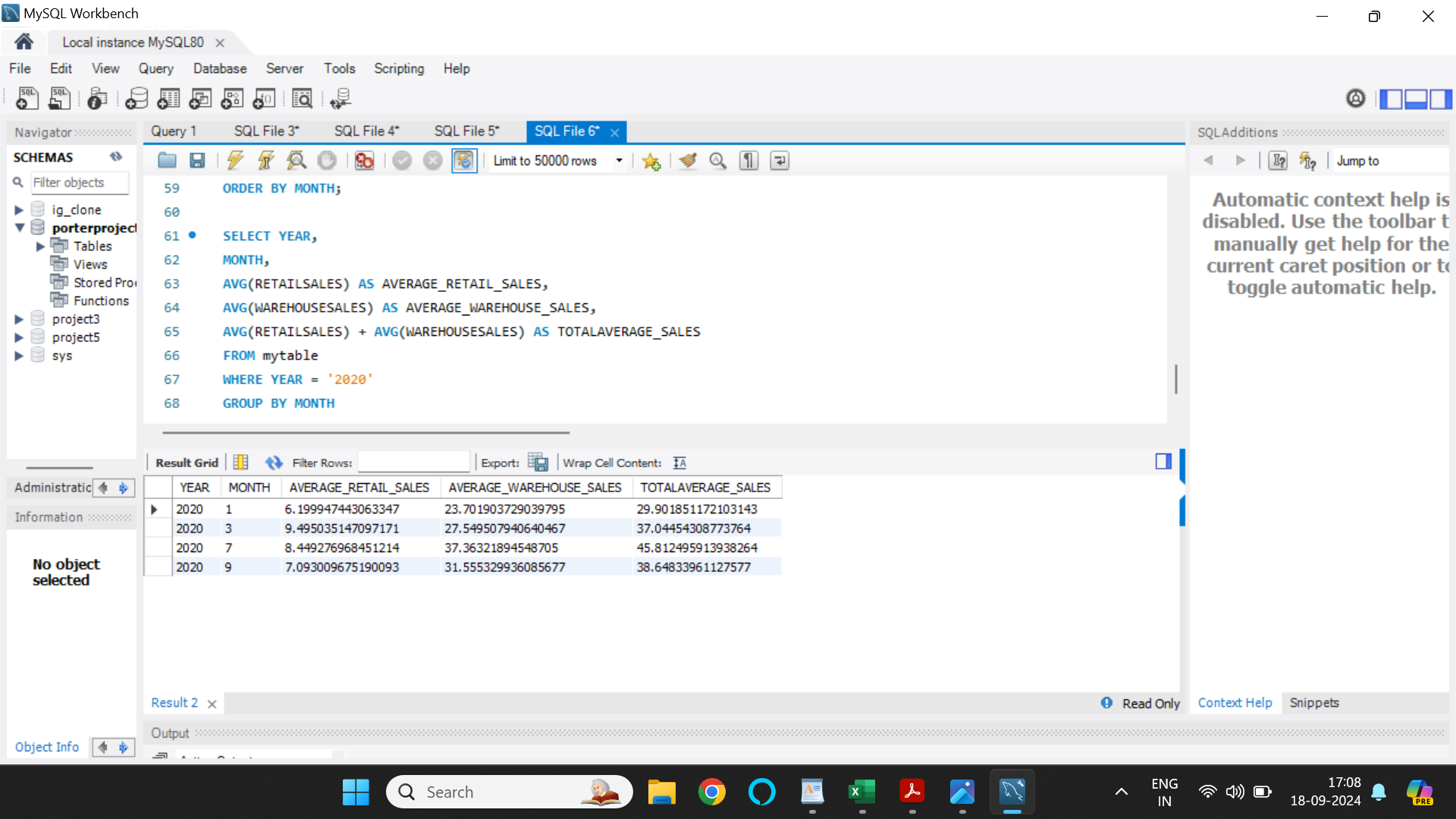
Insights: Clearly, a big chunk of the sales was from ‘Corona extra loose NR – 12OZ’ which brought in total sales of 328224.83, which is roughly 20% of the total sales among the top 10 items, closely followed by ‘Corona extra 2/12 NR – 12OZ’ at 16% of total sales among the top 5 suppliers. Incidentally, 4 item types from the top 10 sellers are from brands Corona (Supplied by Crown Imports), followed by Heineken(Supplied by Heineken USA) which both predominantly sell beers and kegs. Infact, allchunck of the top 10 sellers are item type ‘Beers’.

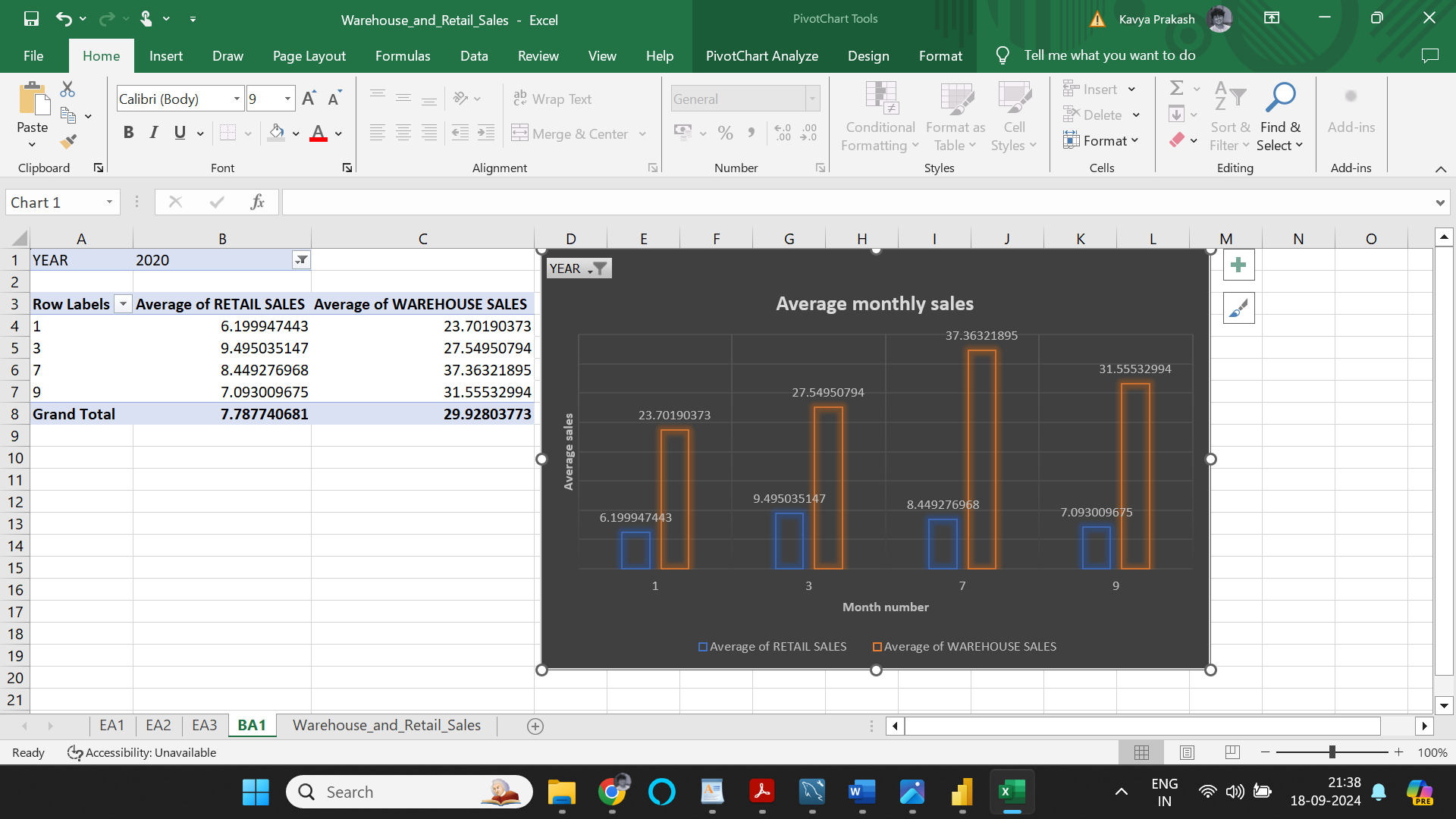
Business Analysis:

Calculate the monthly average retail sales and warehouse sales separately for each year. Analyse whether there are any seasonal trends in sales data. Provide visualisations to support your analysis.

To visualize the data, I have created a pivot table using Microsoft Excel and have used the pivot table to create a bar chart depicting the Average retail sales and Average warehouse sales based on month numbers, and filtered to only show values based on specific years.

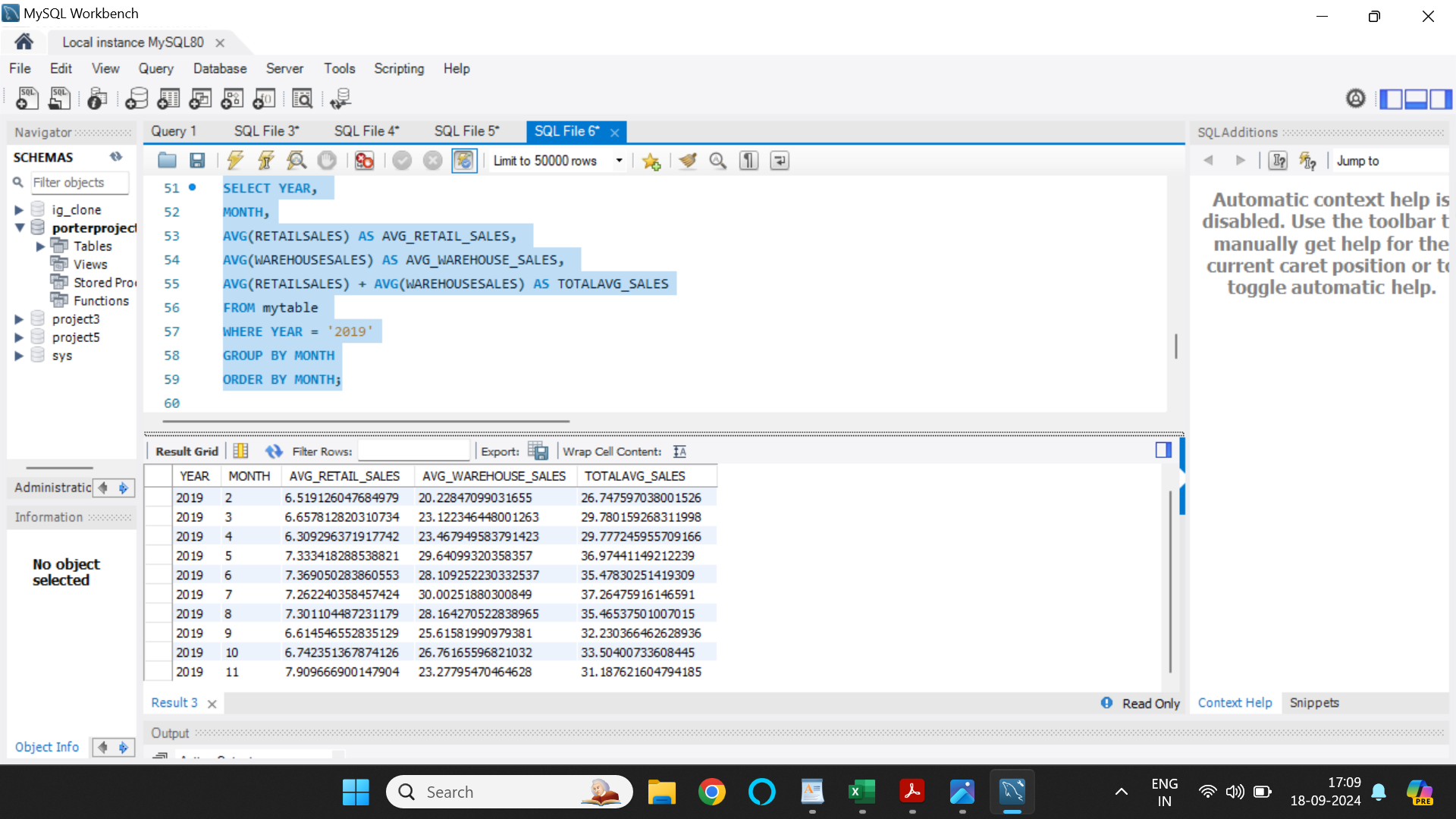
* **Year 2020:**

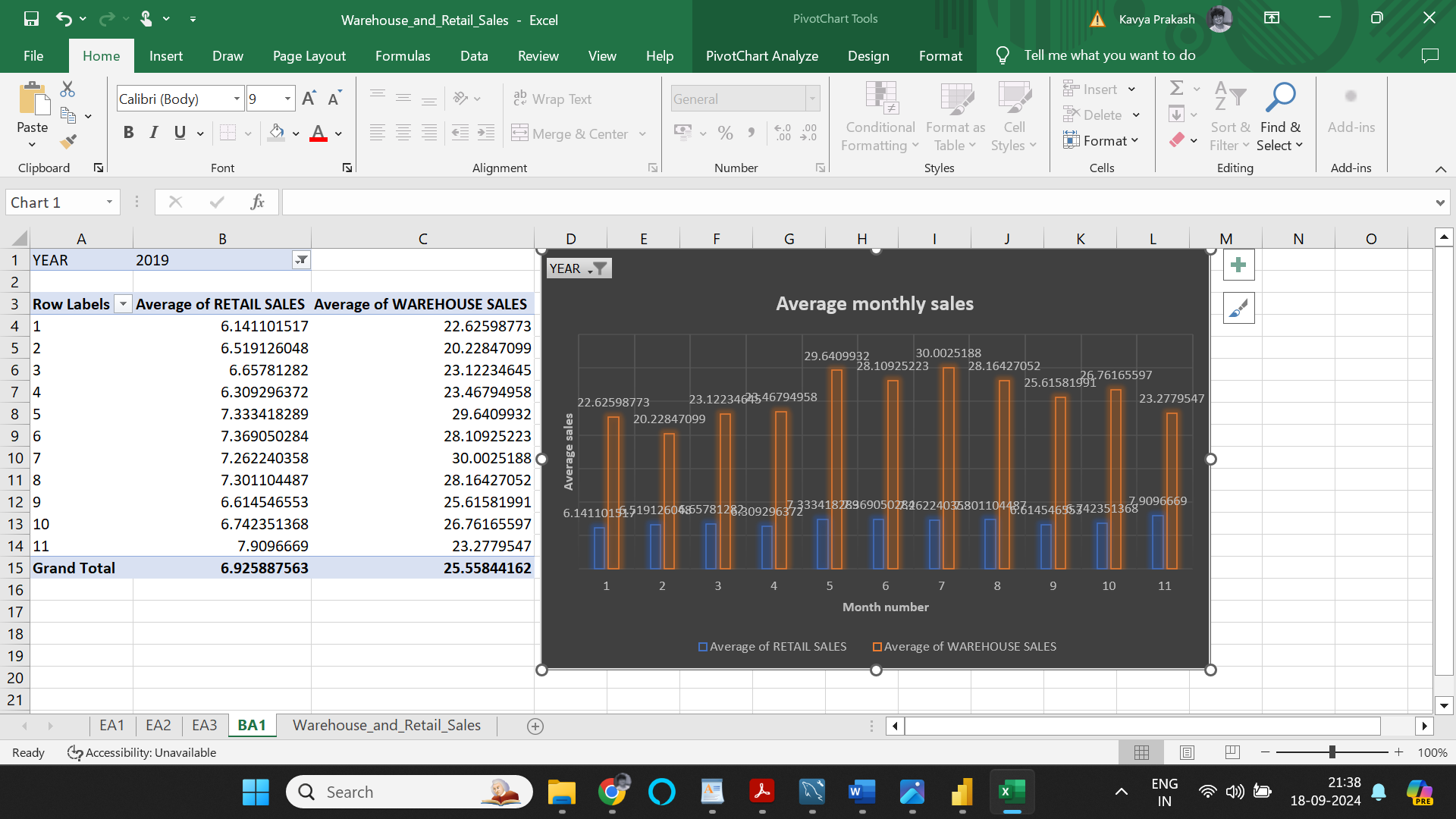




Insights: As per the data avilable from the year 2020, average warehouse sales is higher in the month of July, whereas, average retail sales is higher in the month of March. Unfortunately, in the year 2020, data of only four months are available in the dataset provided, which doesn’t help form reliable insights.

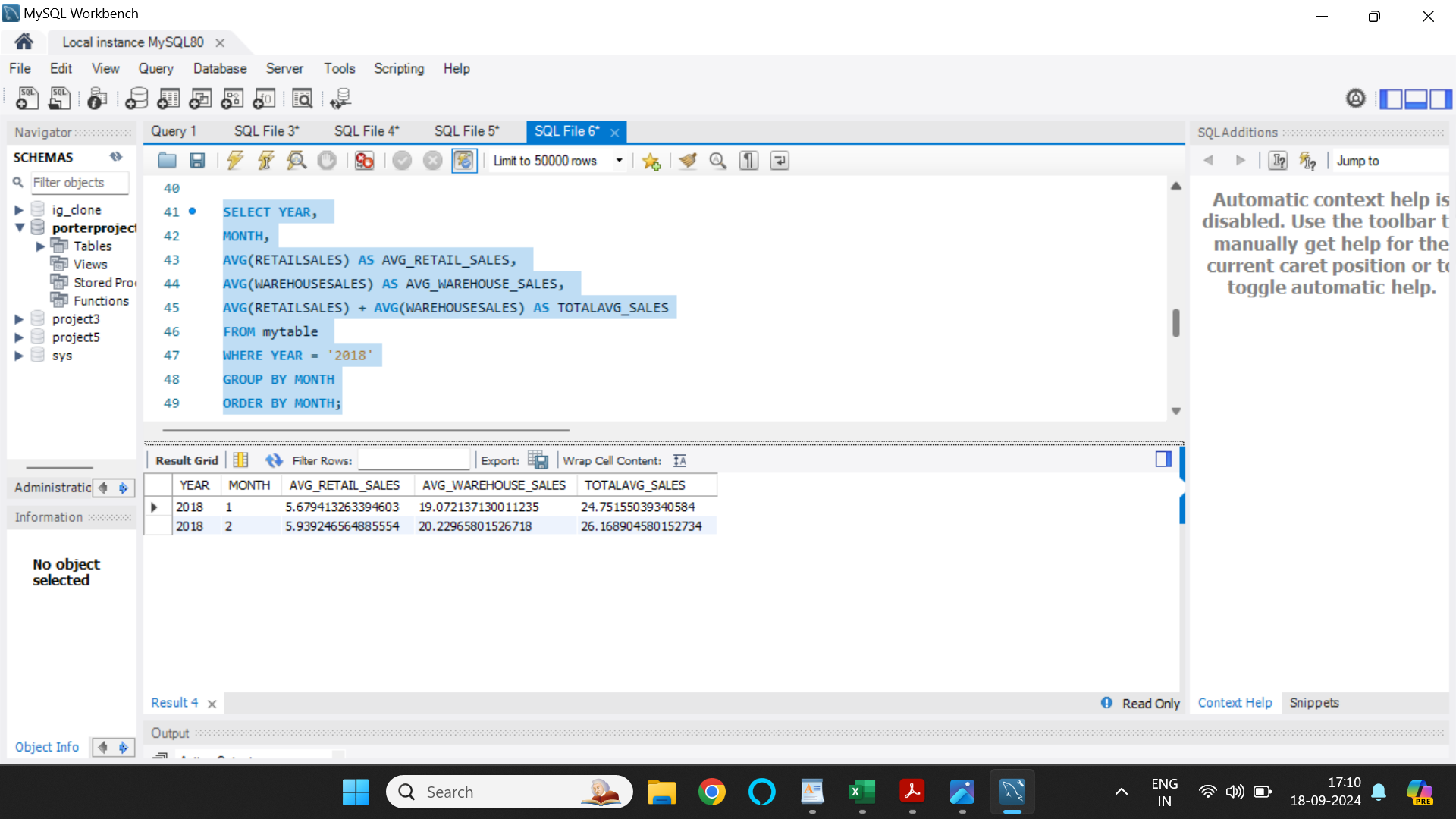
* **YEAR 2019:**

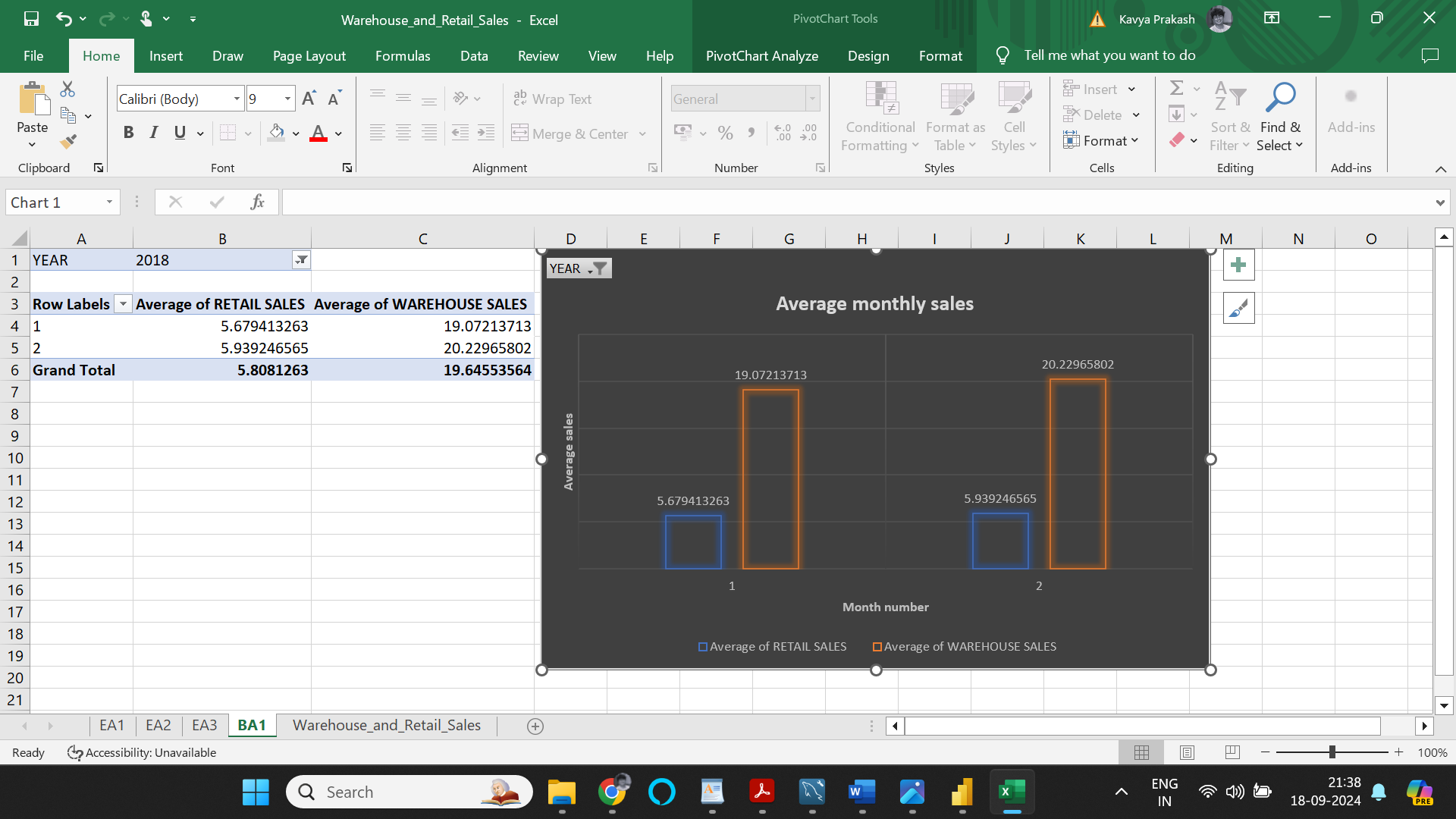




Insights: As per the data avilable from the year 2019, average warehouse sales is higher in the month of July, whereas, average retail sales is higher in the month of November. Unfortunately, in the year 2019, data of only eleven months are available in the dataset provided, which doesn’t help form reliable insights.

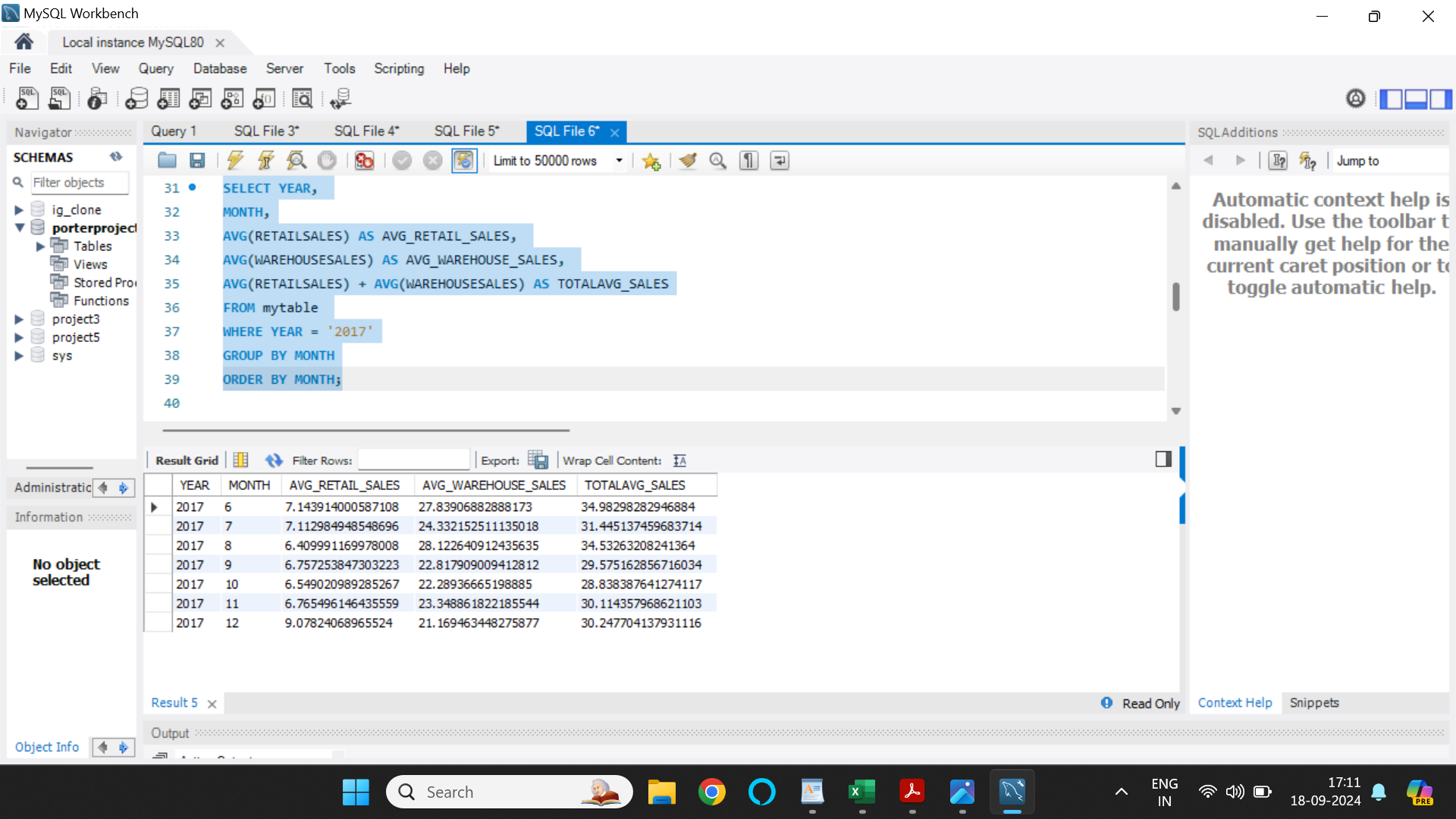
* **YEAR 2018:**

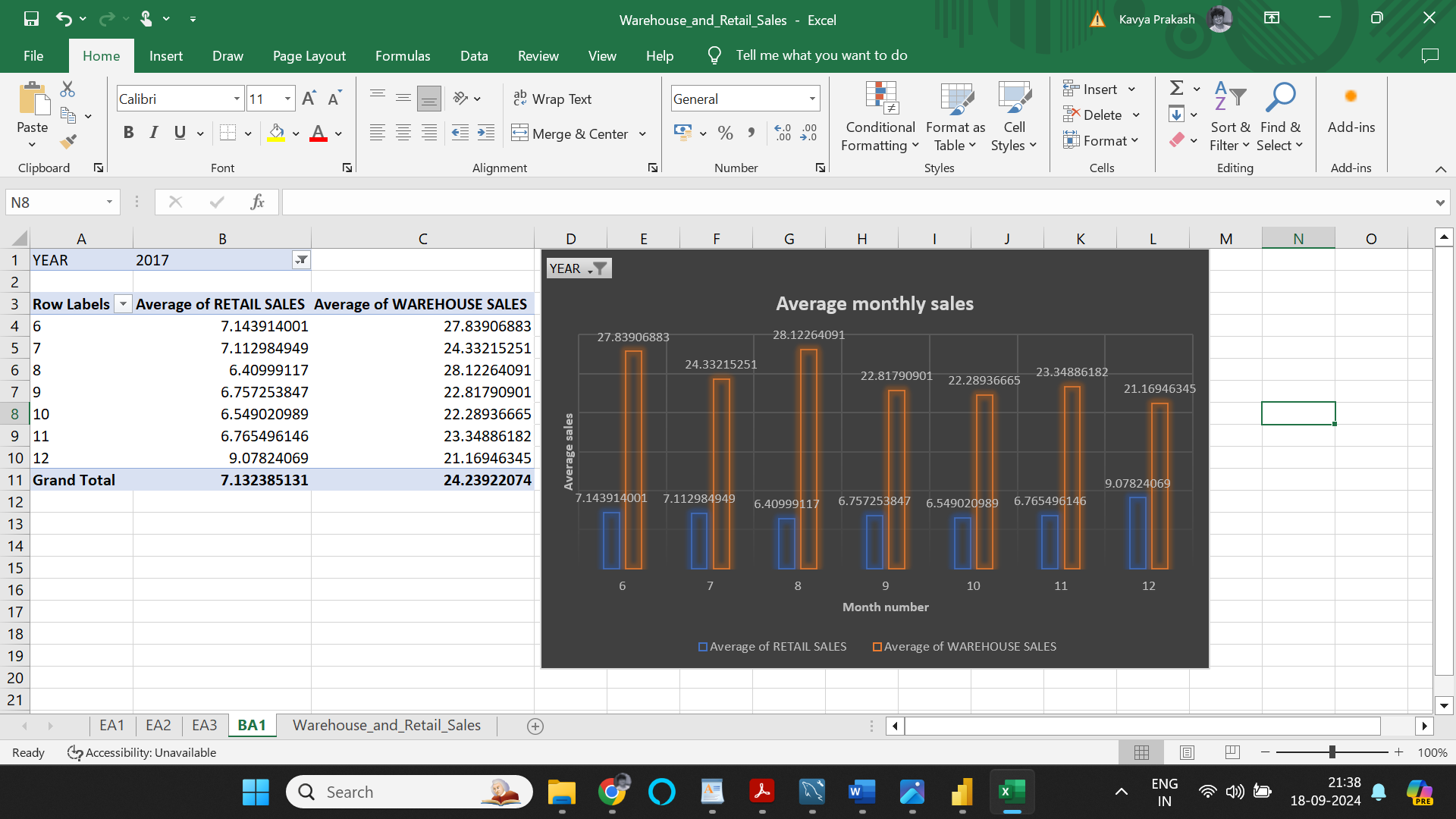




Insights: As per the data avilable from the year 2018, average warehouse sales and average retail sales is higher in the month of February. Unfortunately, in the year 2018, data of only two months are available in the dataset provided, which doesn’t help form reliable insights.

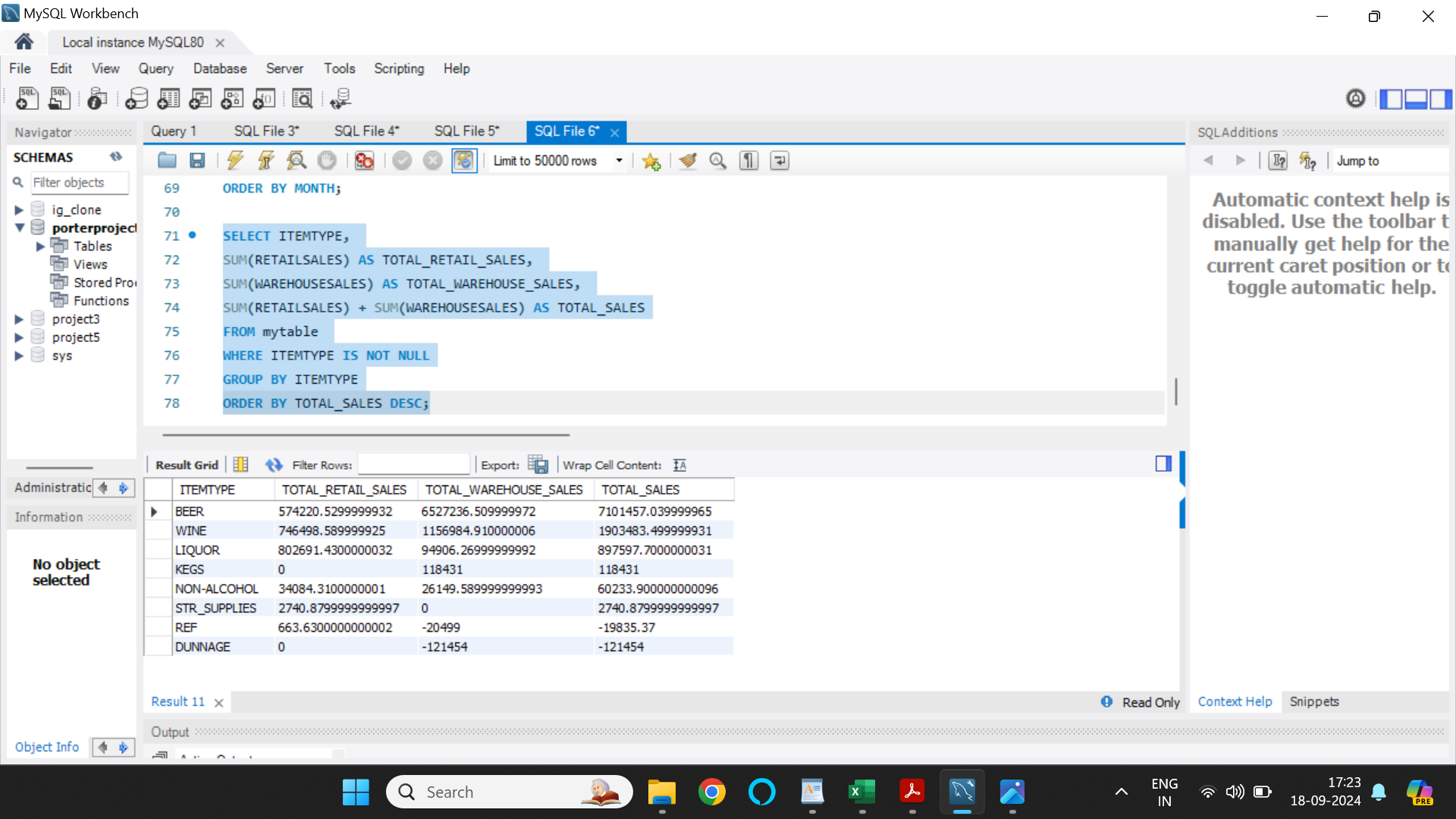
* **YEAR 2017:**



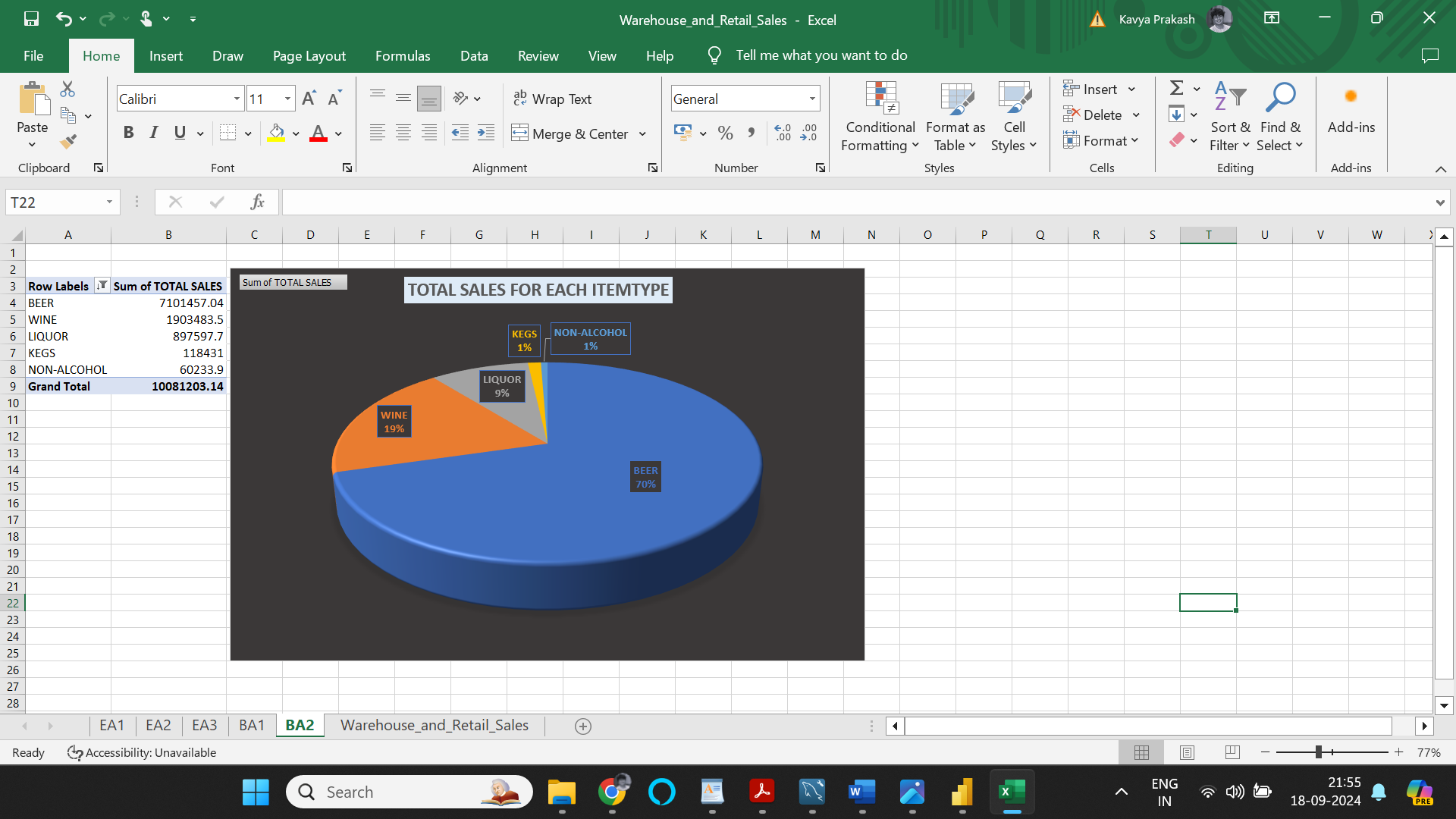


Insights: As per the data avilable from the year 2017, average warehouse sales is higher in the month of August and average retail sales is higher in the month of June. Unfortunately, in the year 2017, data of only seven months are available in the dataset provided, which doesn’t help form reliable insights.

Calculate the total sales for each item type and identify which item type contributes the most to overall sales.



To visualize the data, I have created a pivot table using Microsoft Excel and have used the pivot table to create a pie chart depicting total sales based on item type, and filtered to only show Top 5 sellers.



Insights: Clearly, major chunk of the sales was from ‘Beer’ which made total sales of 7101457.04, which is roughly 70% of the total sales among the top 5 item types, closely followed by ‘Wine’ at 19% of total sales among the top 5 item types.

Actionable Insights:

* The peak sales occurred in 2019, while 2018 had the lowest sales. However, keep in mind that the 2018 data only covers two months, so it’s not a complete picture. To capitalize on the trend seen in 2019, focus on understanding what contributed to that success. Were there specific marketing campaigns, seasonal factors, or product launches?
* ‘Crown Imports’ and ‘Miller Brewing Company’ stand out as major contributors to total sales. They collectively account for a significant portion of revenue. Strengthen relationships with these suppliers. Explore joint promotional efforts or negotiate better terms to enhance their impact on sales.
* ‘Corona extra loose NR – 12OZ’ and ‘Corona extra 2/12 NR – 12OZ’ are star performers among the top 10 items.  Leverage the popularity of these items. Consider bundling them with complementary products or running targeted promotions.
* Notably, all top 10 sellers belong to the ‘Beers’ category. This suggests a strong demand for beer-related products. Optimize your beer inventory. Ensure you have a diverse selection, including craft beers, seasonal brews, and popular brands.
* Since wine follows beer in sales numbers, enhance the wine section. Curate a diverse wine selection, offer tasting events, and educate customers about wine pairings.
* Since ‘Crown Imports’ dominates sales, explore ways to further boost their impact. Perhaps negotiate volume discounts or exclusive deals. Collaborate with ‘Crown Imports’ on joint marketing initiatives. Highlight their products prominently in-store and online
* Heineken, supplied by ‘Heineken USA,’ is also significant. Given its prominence, consider strategic partnerships or themed events around Heineken products. Cross-promote Heineken items with other popular brands to increase visibility.
* Plan seasonal campaigns (e.g., summer beer specials, holiday gift packs) aligned with customer preferences.
* While the dataset has limitations (e.g., incomplete data for certain years), we can still glean some insights. In 2020, average warehouse sales were higher in July, while average retail sales peaked in March. Investigate why these months stand out—were there promotions, events, or seasonal factors?
* With more complete data, we see that July remains a strong month for warehouse sales, and November shines for retail sales in 2019. Plan targeted marketing campaigns or discounts during these months to capitalize on existing trends.
* Dig deeper into the factors driving sales in February 2018. Was it related to holidays, special events, or product launches? Consider inventory adjustments based on seasonal patterns. Stock up on popular items before peak months.
* **Peak and Low Sales Trends**
  + Peak sales occurred in **2019**, while **2018** had the lowest sales. However, the 2018 data only covers two months, which does not provide a complete picture.
  + **Action:** Investigate the success factors from 2019, such as marketing campaigns, seasonal influences, or product launches, to replicate or build upon them.
* **Key Contributors to Sales**
  + **‘Crown Imports’** and **‘Miller Brewing Company’** are the top contributors to overall sales, collectively generating a significant portion of the revenue.
  + **Action:** Strengthen relationships with these suppliers. Explore joint promotional campaigns or negotiate better terms to maximize their impact on revenue.
* **Top-Performing Products**
  + The **‘Corona Extra Loose NR – 12OZ’** and **‘Corona Extra 2/12 NR – 12OZ’** stand out as star performers among the top 10 items.
  + **Action:** Leverage their popularity by bundling them with complementary products or running targeted promotions to boost sales further.
* **Category Insights**
  + All top 10 sellers fall under the **‘Beers’** category, indicating a strong demand for beer products.
  + **Action:** Optimize beer inventory by ensuring a diverse selection, including craft beers, seasonal brews, and popular brands.
* **Opportunities for Wine Products**
  + After beer, wine ranks second in sales.
  + **Action:** Enhance the wine section by curating a diverse selection, hosting wine-tasting events, and educating customers about pairings to drive sales in this category.
* **Supplier Collaboration Opportunities**
  + **Crown Imports** dominates sales performance.
  + **Action:** Negotiate volume discounts or exclusive deals and collaborate on joint marketing initiatives. Highlight their products prominently in-store and online.
  + **Heineken**, supplied by **Heineken USA**, is another significant contributor.
  + **Action:** Develop strategic partnerships or themed events centered around Heineken products. Use cross-promotion with other popular brands to enhance visibility.
* **Seasonal Trends**
  + Warehouse sales peaked in **July 2020**, while retail sales were highest in **March 2020**. In 2019, warehouse sales also peaked in **July**, and retail sales were strong in **November**.
  + **Action:** Plan targeted marketing campaigns and discounts aligned with these seasonal trends, such as summer beer specials or holiday promotions.
* **Sales Anomalies and Patterns**
  + February 2018 shows notable sales activity despite limited data.
  + **Action:** Investigate the factors behind this anomaly, such as holidays, events, or product launches, to identify replicable opportunities.
* **General Recommendations**
  + Use seasonal patterns to adjust inventory and stock up on popular items ahead of peak months.
  + Plan promotional campaigns around strong months like **July** for warehouse sales and **November/March** for retail sales.
  + Address dataset limitations by prioritizing better data collection and analysis in future projects.