A Presentation on Data Analysis Project Using SQL

Here are some business questions to be answered:

1) The Organization is planning to gift the best performing manager who made the best sales and want to know the region which the manager belongs to?
2) How many times was the delivery truck used as the ship mode?
3) how many orders were returned, and which product category got rejected the most?
4) Which Year did the company incurred the least shipping cost?
5) display the day of the week in which customer segment has the most sales?
6) The company wants to determine its profitability by knowing the actual orders that were delivered.
7) The Organization is eager to know the customer names and persons born in 2011?
8) What are the aggregate orders made by all the customers?
9) The company intends to discontinue any product that brings in the least profit, you are required to help the organization to determine the product?
10) What are the top 2 best selling items that the company should keep selling?

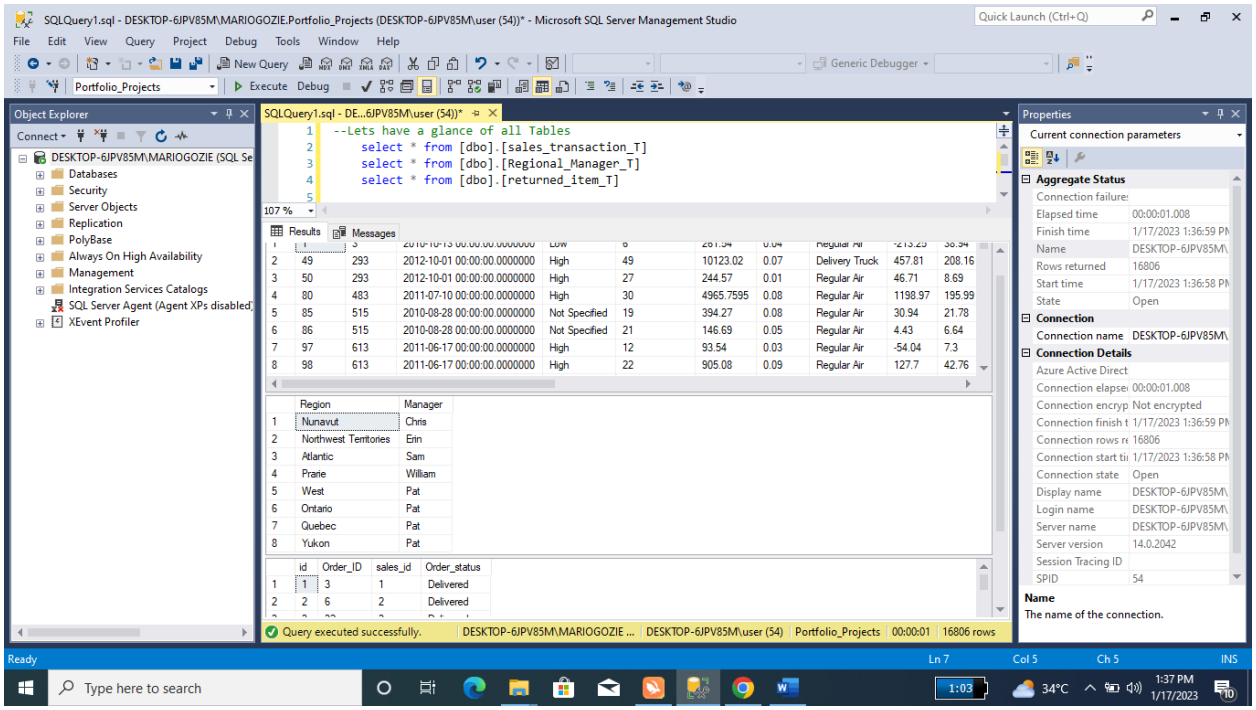
Solutions to the Tasks

Before I start, I need to view the three tables and Understand the data.

This was done using the Sql query

```
select * from [dbo].[sales_transaction_T]

select * from [dbo].[Regional_Manager_T]

select * from [dbo].[returned_item_T]
```

Below is what the Tables look like



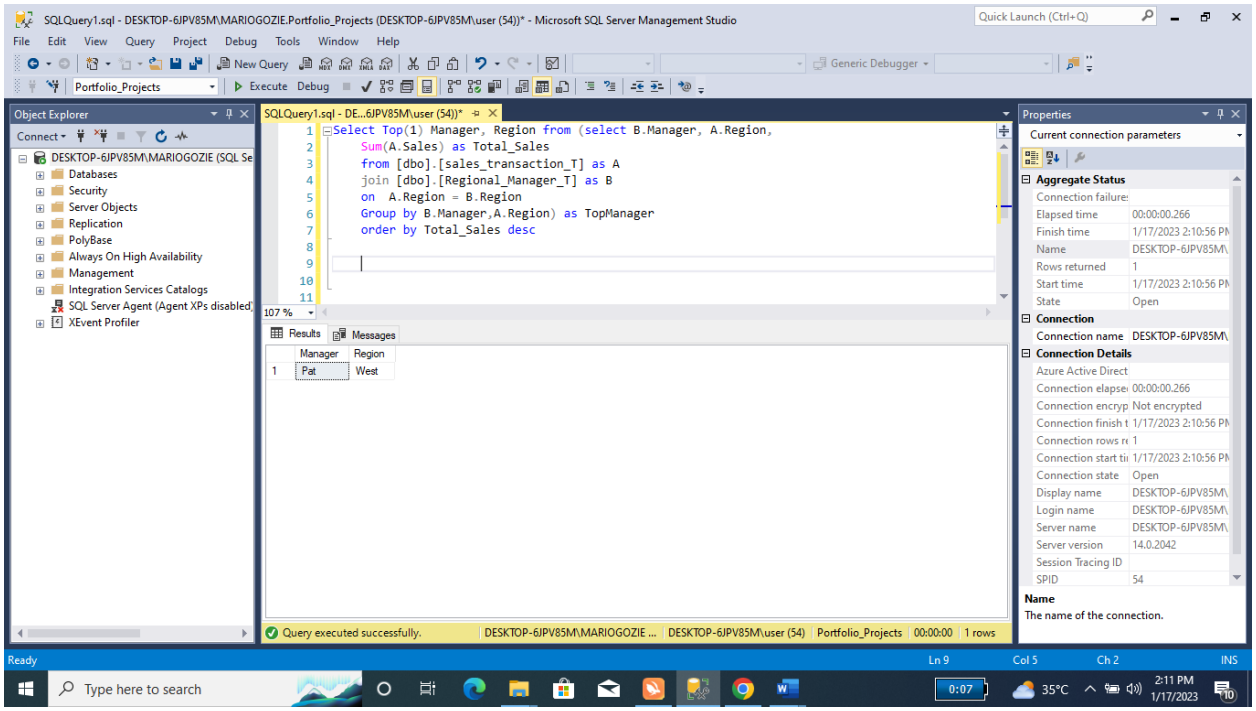
For Question 1

The Organization is planning to gift the best performing manager who made the best sales and want to know the region which the manager belongs to?

To carry out this task, I had to join two tables which are the sales table and regional managers table on regions using.

The query used for this task is :

```
Select Top(1) Manager, Region from (select B.Manager, A.Region,

        Sum(A.Sales) as Total_Sales

        from [dbo].[sales_transaction_T] as A

        join [dbo].[Regional_Manager_T] as B

        on  A.Region = B.Region

        Group by B.Manager,A.Region) as TopManager

        order by Total_Sales desc
```



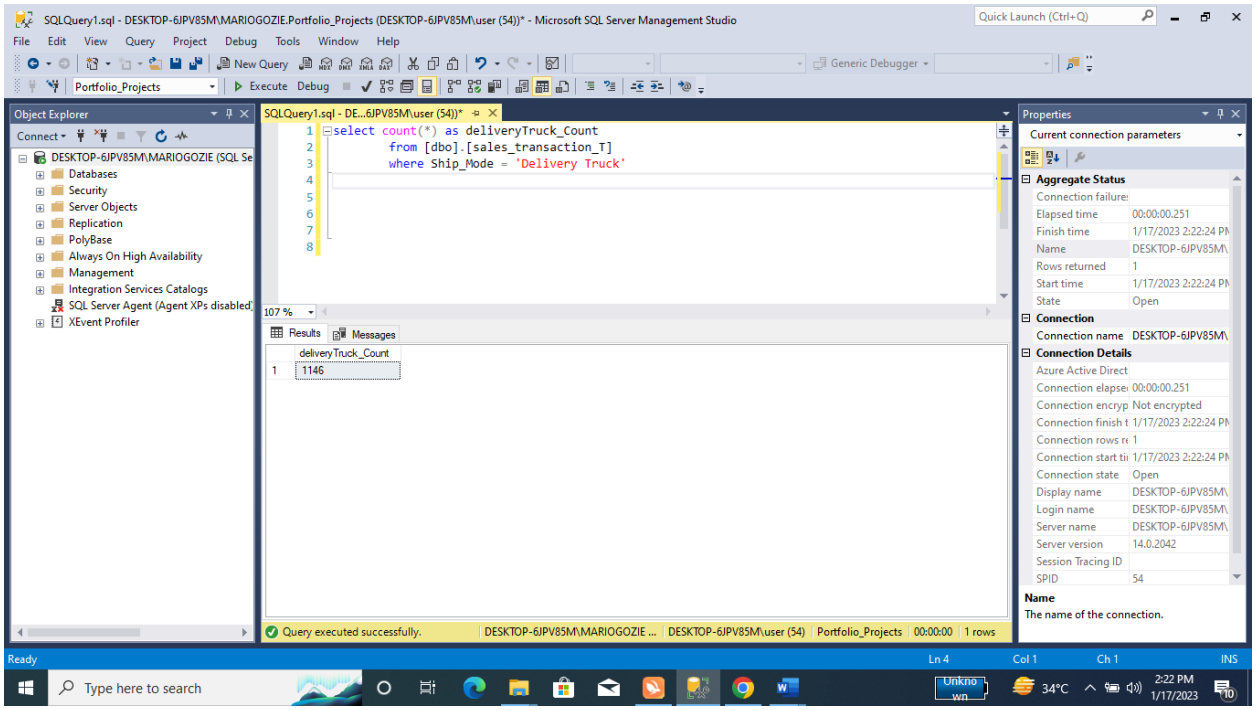
The query showed **Pat** of West **Region** as the manager with most sales.

Question 2 says:

How many times was the delivery truck used as the ship mode?

This task unlike the first one requires the use of one, which is the Sales_Transaction Table. Here, I counted all Transactions in which deliveries were made with delivery trucks using the code below.

```
select count(*) as deliveryTruck_Count

 from [dbo].[sales_transaction_T]

where Ship_Mode = 'Delivery Truck'
```



The result of the query showed that there were **1146** orders delivered with the **delivery Truck.**

For the Third Question:

how many orders were returned, and which product category got rejected the most?

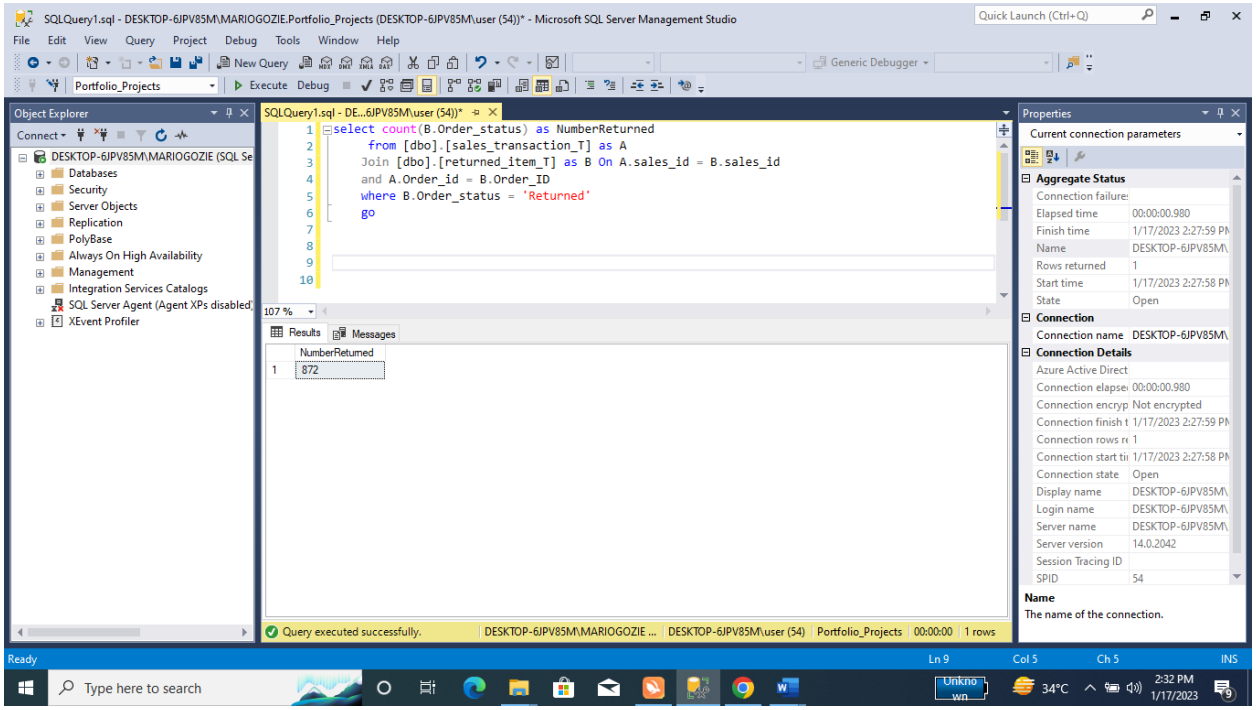
I answered this with two methods. For the first method (A), I had to join the sales Transaction table to returned_item table on sales_ID and order_id, then I counted the number of orders where sales status reads returned this gave me the total number of **872** returned orders . Afterwards, I counted orders returned with the tables joined and selected the top(1) after ordering by count in descending order. This gave me the most returned Product category which is **office supplies**.

Sql query for total order returned

```
select count(B.Order_status) as NumberReturned

        from [dbo].[sales_transaction_T] as A

      Join [dbo].[returned_item_T] as B On A.sales_id = B.sales_id
```

```
and A.Order_id = B.Order_ID
where B.Order_status = 'Returned'
go
```

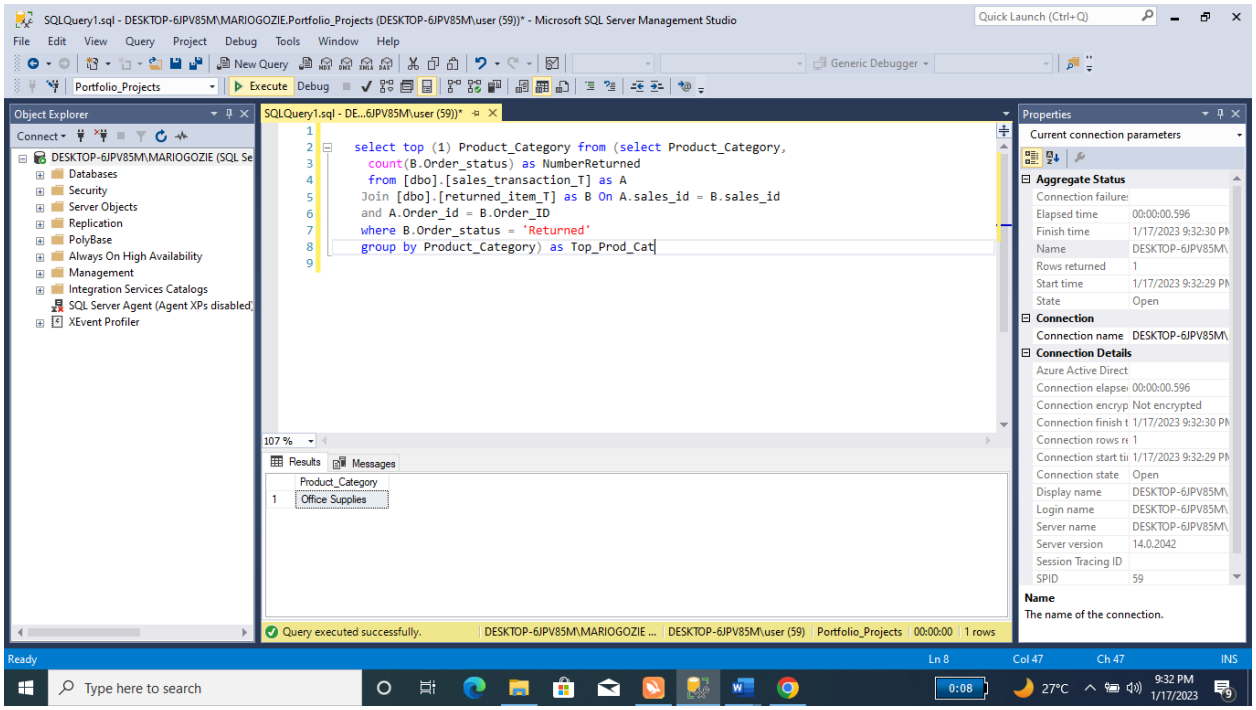
An image showing total orders retuned



The result show that 872 items were returned

Query to find the most returned Product Category

```
select top (1) Product_Category from (select Product_Category,
count(B.Order_status) as NumberReturned
from [dbo].[sales_transaction_T] as A
Join [dbo].[returned_item_T] as B On A.sales_id = B.sales_id
and A.Order_id = B.Order_ID
where B.Order_status = 'Returned'
group by Product_Category) as Top_Prod_Cat
```



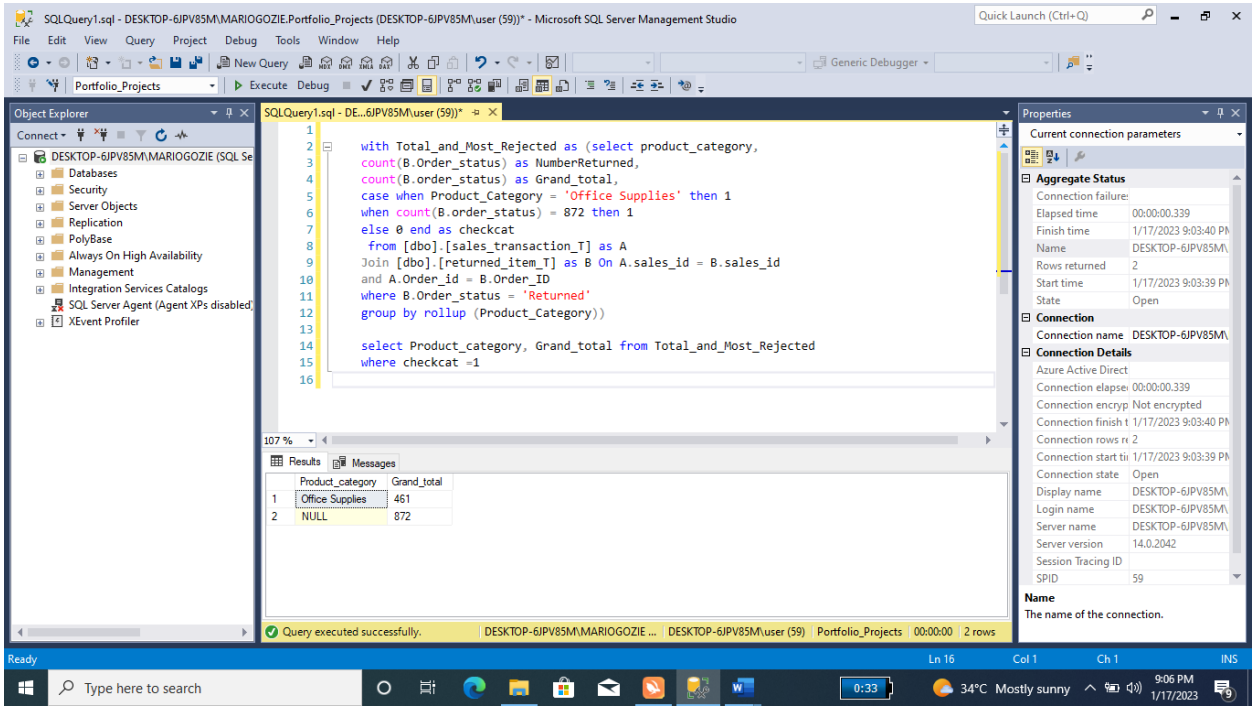
The image above shows **office supplies** as the most returned categories

For the second method (B), a common table expression was used alongside with an aggregating and a grouping function called rollup. This function presents the total as NULL. The main difference between this and the other method is that it will present both answers in one table.

Code used for the second method

```
with Total_and_Most_Rejected as (select product_category,
        count(B.Order_status) as NumberReturned,
        count(B.order_status) as Grand_total,
        case when Product_Category = 'Office Supplies' then 1
        when count(B.order_status) = 872 then 1
        else 0 end as checkcat
         from [dbo].[sales_transaction_T] as A
        Join [dbo].[returned_item_T] as B On A.sales_id = B.sales_id
        and A.Order_id = B.Order_ID
```

```
where B.Order_status = 'Returned'

group by rollup (Product_Category))


select Product_category, Grand_total from Total_and_Most_Rejected

where checkcat =1
```



The image above show that **Office supplies** is the most rejected category and the total reject item are **872**

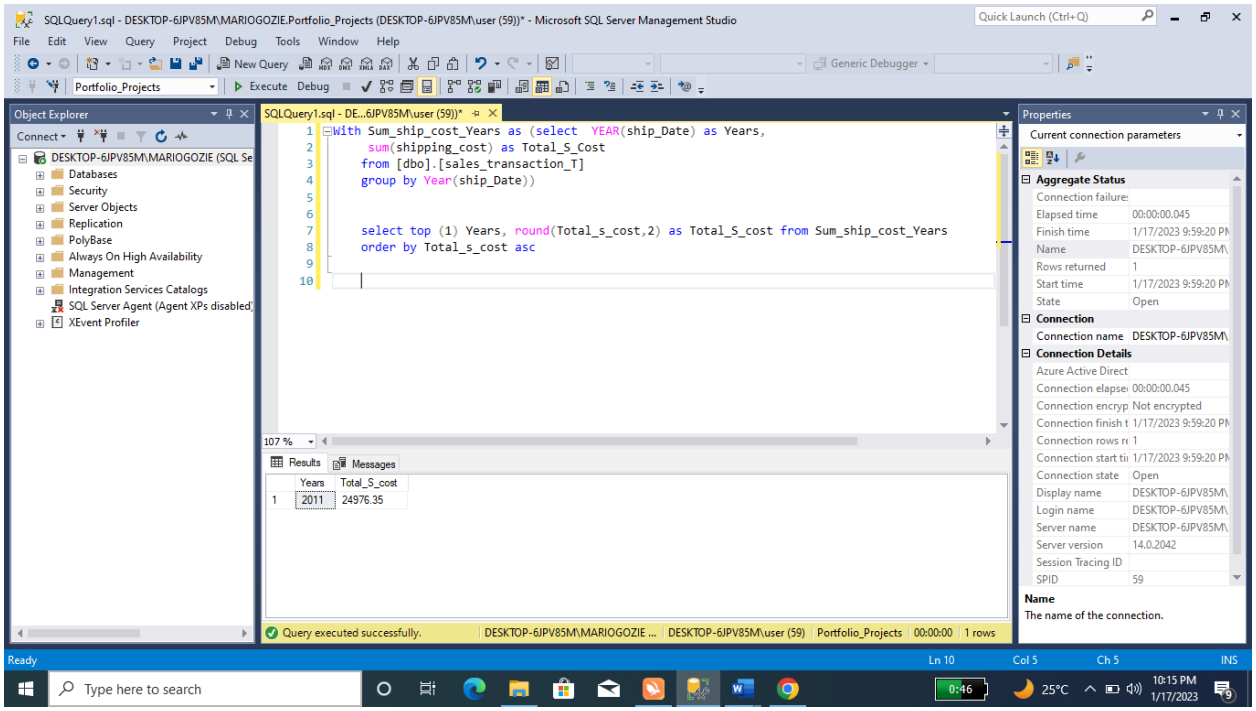
Moving on to the 4th question :

Which Year did the company incurred the least shipping cost?

In tackling this, I used a common table expression (CTE) to create a temporary table that grouped shipping cost and years, then selected the top (1) and ordering in ascending order of shipping cost.

Query used for this task

```
With Sum_ship_cost_Years as (select  YEAR(ship_Date) as Years,
        sum(shipping_cost) as Total_S_Cost
        from [dbo].[sales_transaction_T]
        group by Year(ship_Date))

        select top (1) Years, round(Total_s_cost,2) as Total_S_cost from     Sum_ship_cost_Years
        order by Total_s_cost asc

        go
```



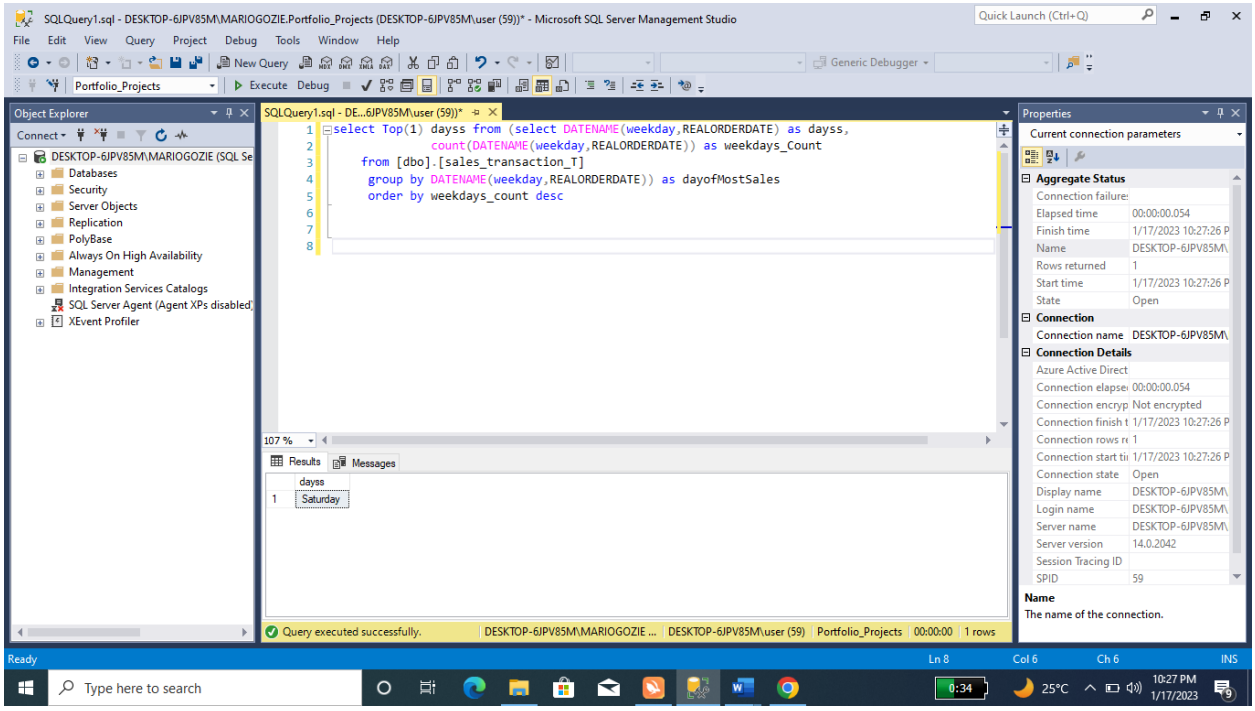
The picture show **2011** as the year with least shipping cost as well as the value (24976.35)

Question 5

display the day of the week in which customer segment has the most sales?

The task above was carried out by first making extracting days of the week and counting the number of orders made on each day. This was grouped into a subquery, then I selected the top(1) after ordering by count for weekday orders in a descending order. The result of the query showed that the biggest sales are made on saturdays.

```
select Top(1) dayss from (select DATENAME(weekday,REALORDERDATE) as dayss,
        count(DATENAME(weekday,REALORDERDATE)) as weekdays_Count
        from [dbo].[sales_transaction_T]
        group by DATENAME(weekday,REALORDERDATE)) as dayofMostSales
        order by weekdays_count desc
```



The picture above shows that most sales are made on **saturdays.**

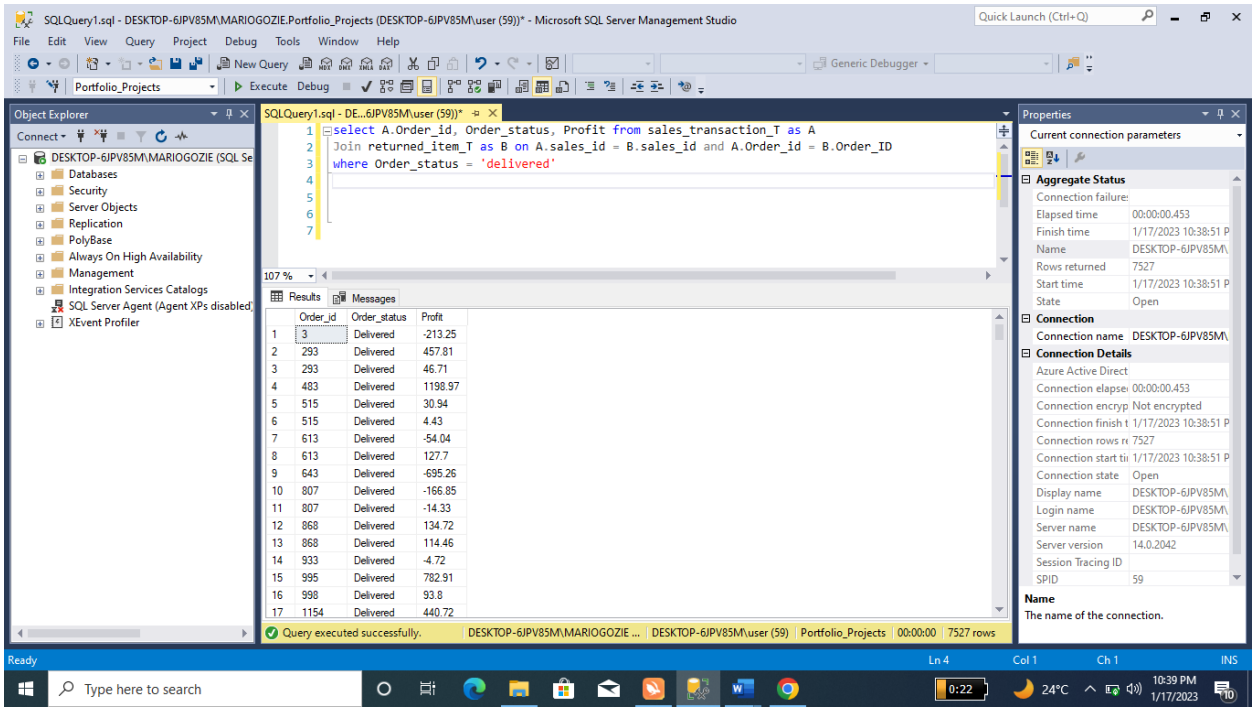
Question 6

The company wants to determine its profitability by knowing the actual orders that were delivered.

To show all orders that were delivered, I called Joined the sales table the returned item table on order id and sales id. The filtered the result with the word delivered on the returned item table.

Query used;

```
select A.Order_id, Order_status, Profit from sales_transaction_T as A

Join returned_item_T as B on A.sales_id = B.sales_id and A.Order_id = B.Order_ID

where Order_status = 'delivered'
```



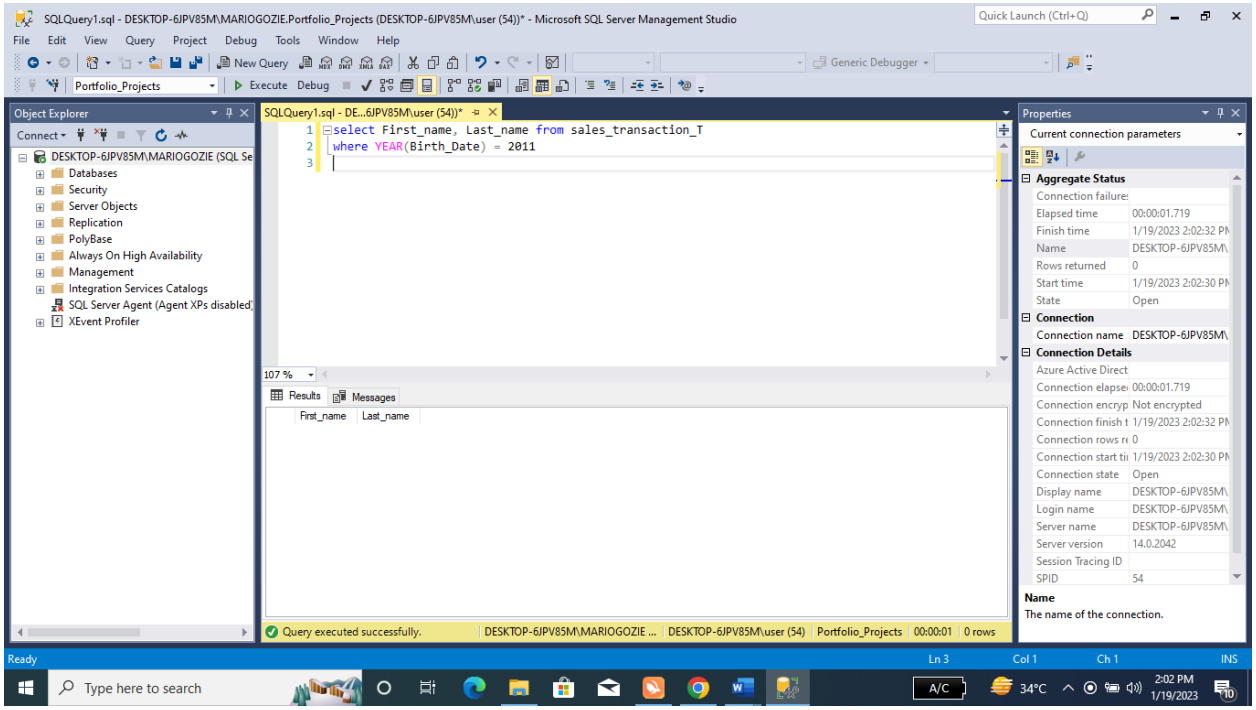
The image above show all orders that were delivered

Question 7

The Organization is eager to know the customer names and persons born in 2011?

To tackle this question, queried the first name and last name of customers on the Sales Transaction table and filtered for those born in 2011 using the year function for the birth date column. This showed that there was no customer born in 2011.

```
select First_name, Last_name from sales_transaction_T

where YEAR(Birth_Date) = 2011
```



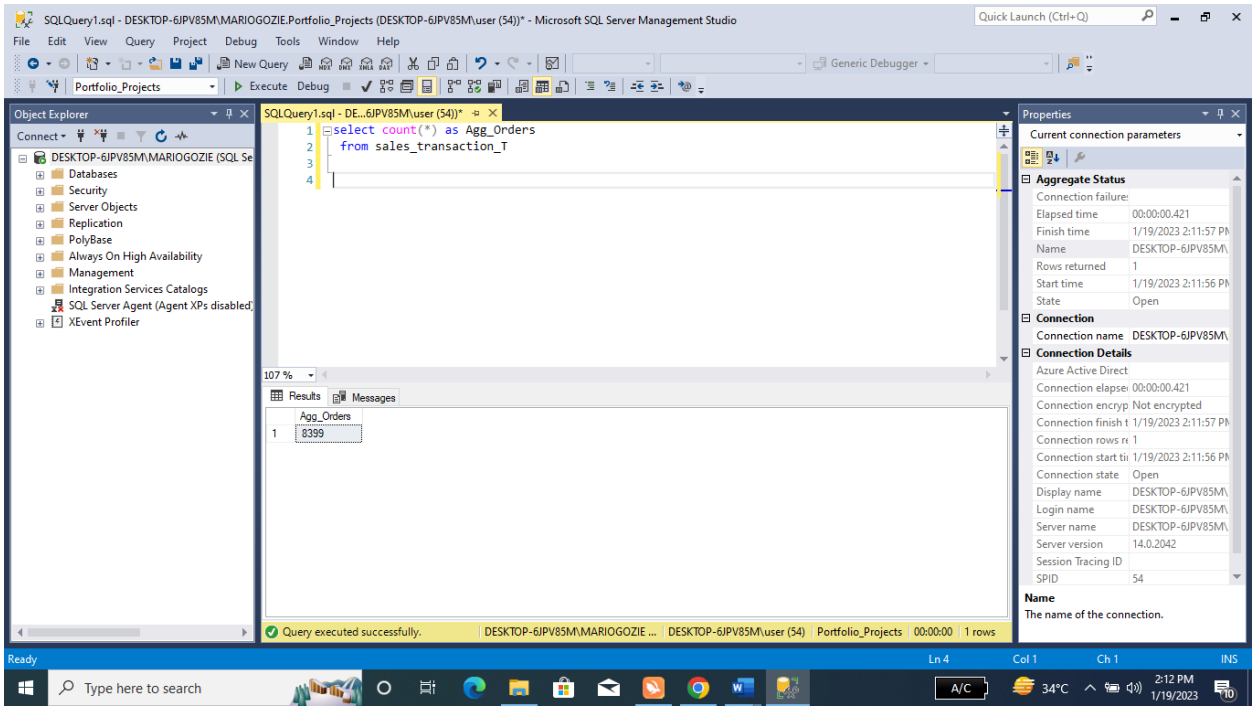
An image showing that no customer was born in 2011.

Question 8

What are the aggregate orders made by all the customers?

To tackle this question, we need to use the function count to find the aggregate of all orders by customers on the sales transaction table. The query showed that a total of 8399 orders were made.

```
select count(*) as agg_orders from sales_transaction_T
```



A picture showing the total number of orders to be 8399

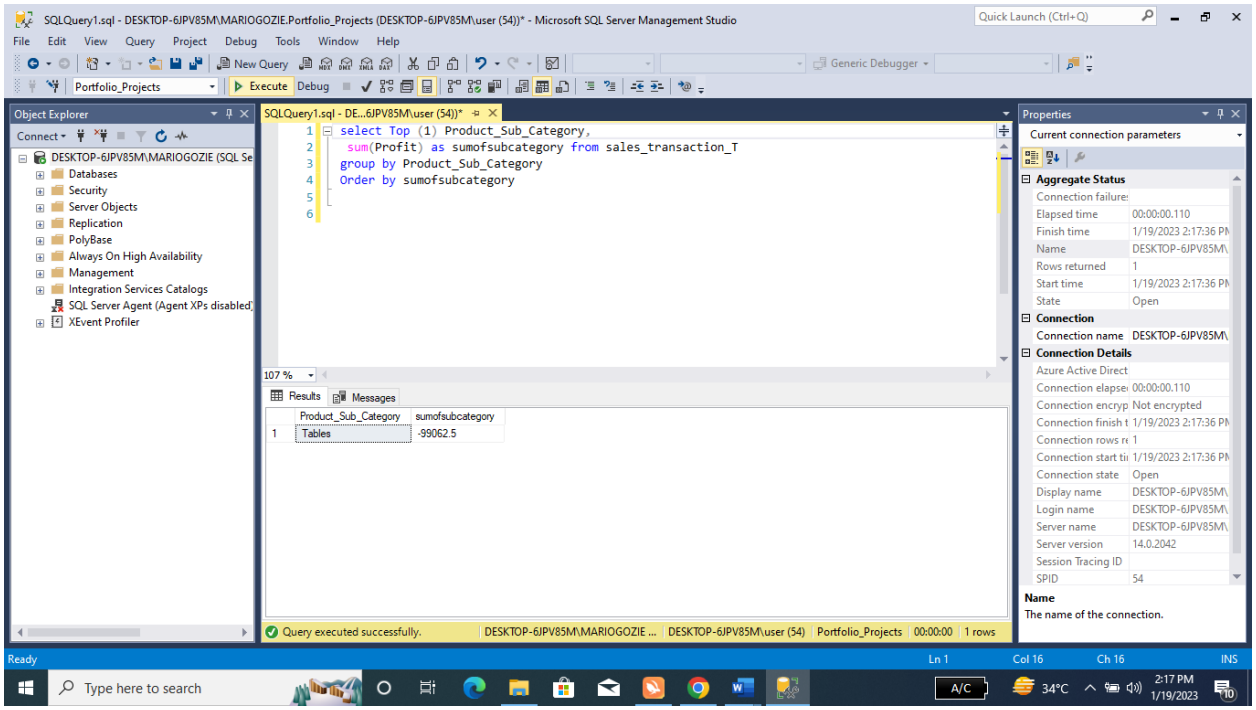
Question 9

The company intends to discontinue any product that brings in the least profit, you are required to help the organization to determine the product?

To take care of this task I selected and group sub categories of product, summed their profit and ordered them by sum of profit. This presented the data in ascending order of profit then I selected the top on the list which was Tables.

Query Used

```
select Top (1) Product_Sub_Category,
    sum(Profit) as sumofsubcategory from
    sales_transaction_T
     group by Product_Sub_Category
     Order by sumofsubcategory
```



The picture above shows that tables in the subcategory level are the least profitable product wit a profit of -99062.5

Question 10

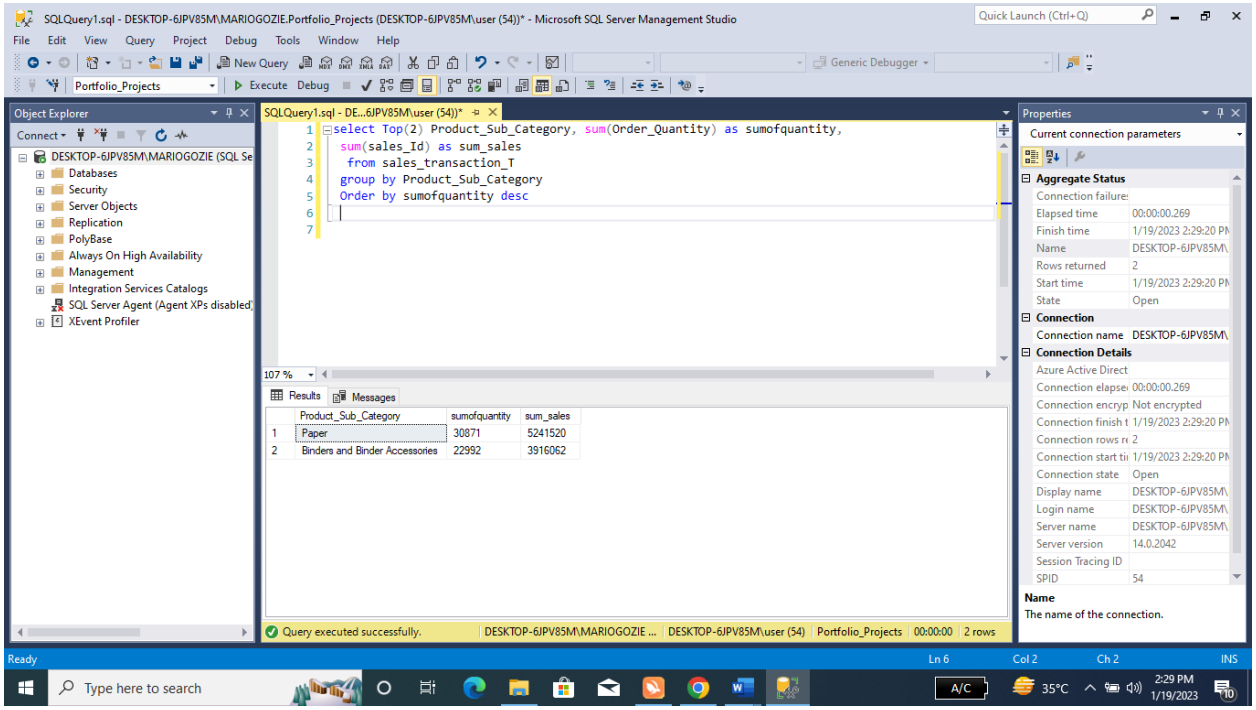
What are the top 2 best selling items that the company should keep selling?

To answer this question, I selected categories and grouped them while summing order quantity. Then I ordered them in descending order of quantity and picked the top 2 with papers topping the list while binders and binder accessories were next.

Code Used

```
select Top(2) Product_Sub_Category, sum(Order_Quantity) as sumofquantity,
sum(sales_Id) as sum_sales
 from sales_transaction_T
group by Product_Sub_Category
```

Order by sumofquantity desc



The Picture here shows that papers tops the list of top selling products while binders and binder accessories were next.