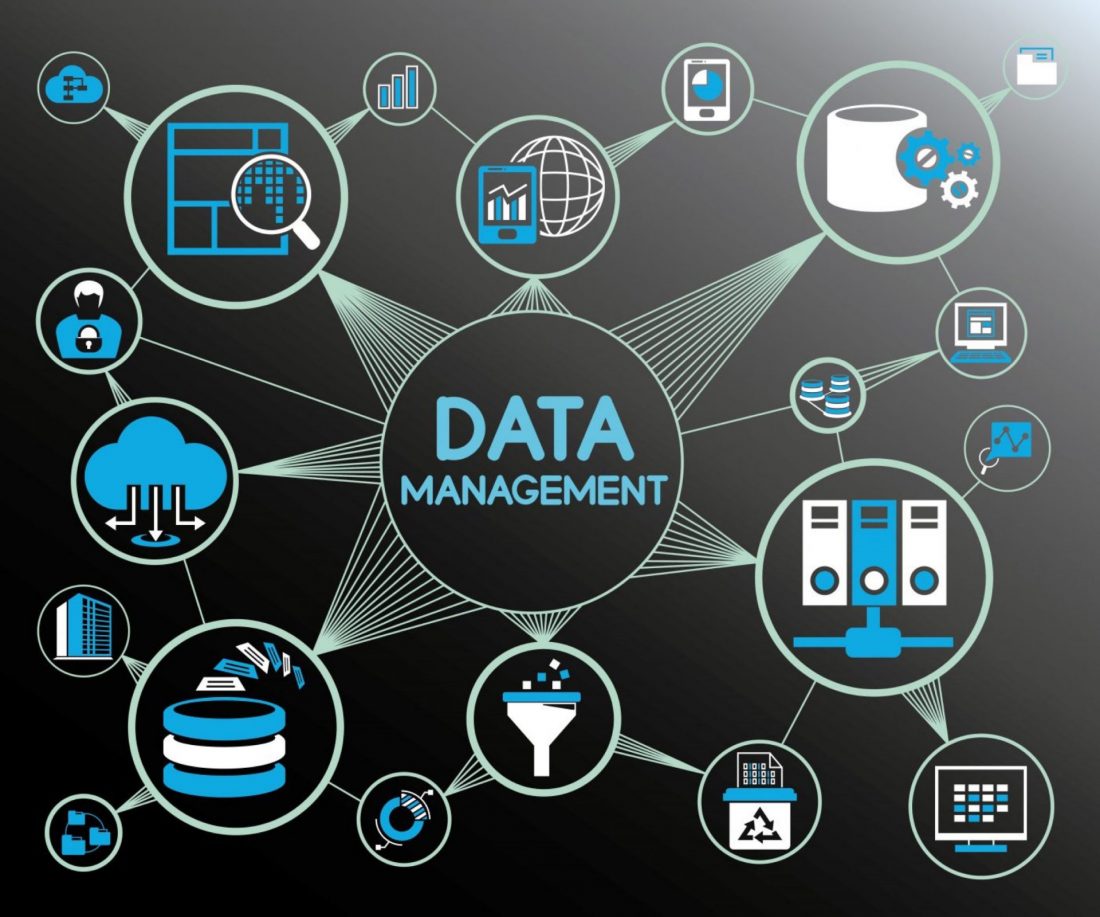
**Data Management**

Report On Coursework 2



Submitted By

Student ID: 4126927

Data Import – The appropriate star schema was designed using Ralph Kimball's Dimension Modelling Technique as specified in the coursework. The purpose of this model is to improve the database performance and efficiency by reducing query processing time and improving results. In order to import the data, I am using a CSV file provided in the module and following the specific steps for importing it into SSIS Tool in Visual Studio.

The Figure below shows how I merged the 12 csv files into one database

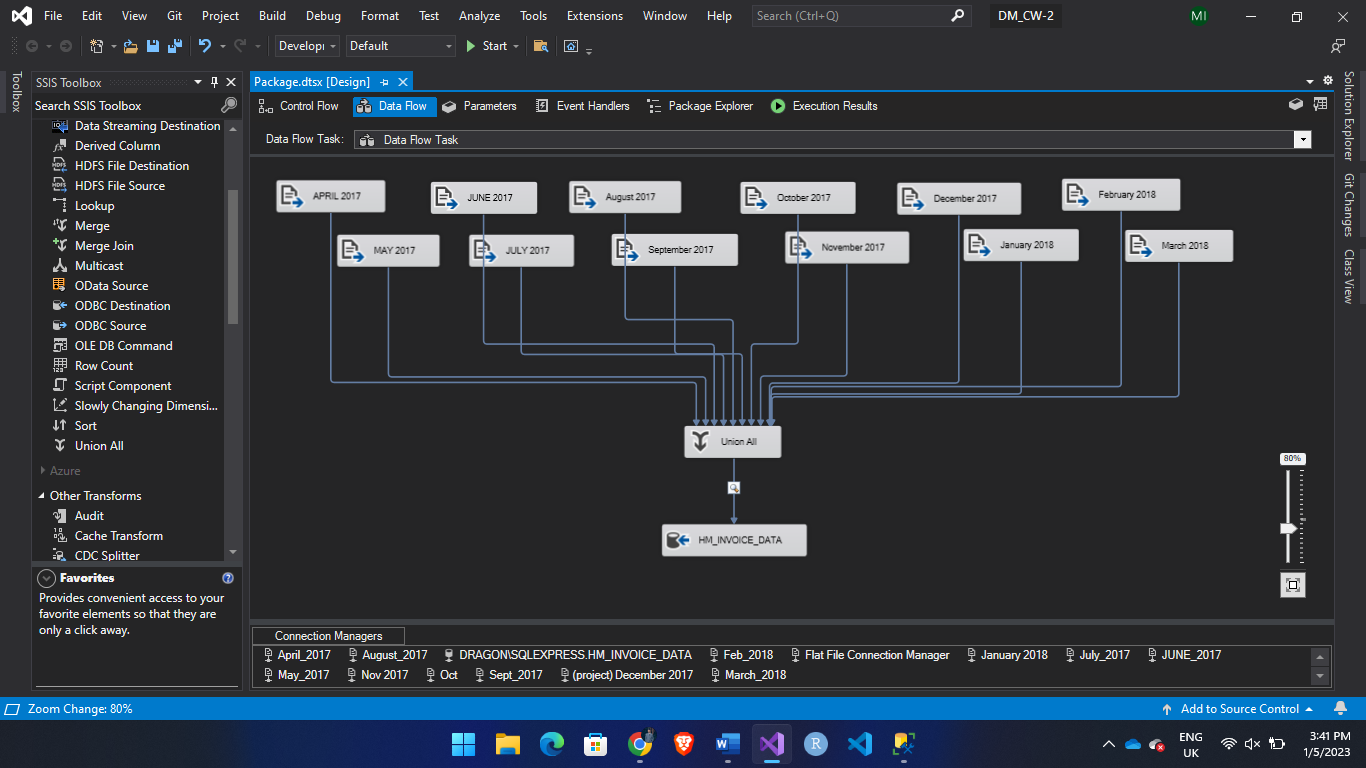


Figure :1 CSV File Import

1. First, I used Flat File Source to connect csv file
2. Then used union all to merge 12 csv file into one data file
3. Then stored the in a database using OLE DB Destination

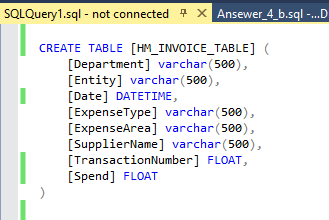


Figure 2: SQL Query in SSMS Tool

Figure 2 contains the query used for creating Table to store out data from csv file in OLE DB Destination.

SSMS Tools was used to perform queries in this project.

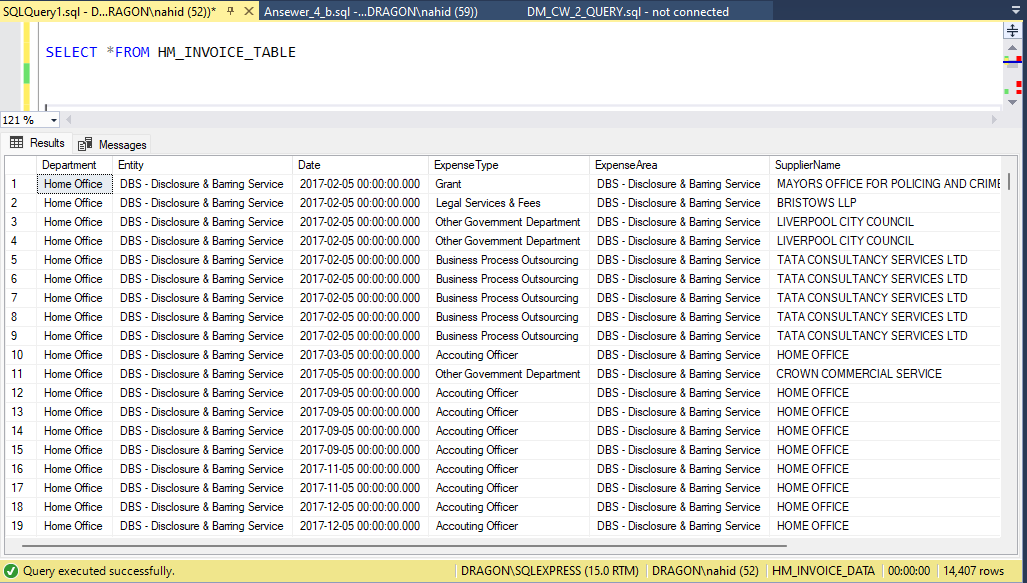


Figure 3: Merged Data Table From CSV (HM\_INVOICE\_TABLE)

In Figure 3 above we can see our data using SQL Query.

# Data Cleaning

For cleaning the data I used SSIS Tool to perform ETL on our data. Below Figure shows how ETL was use to clean the null values from each column.

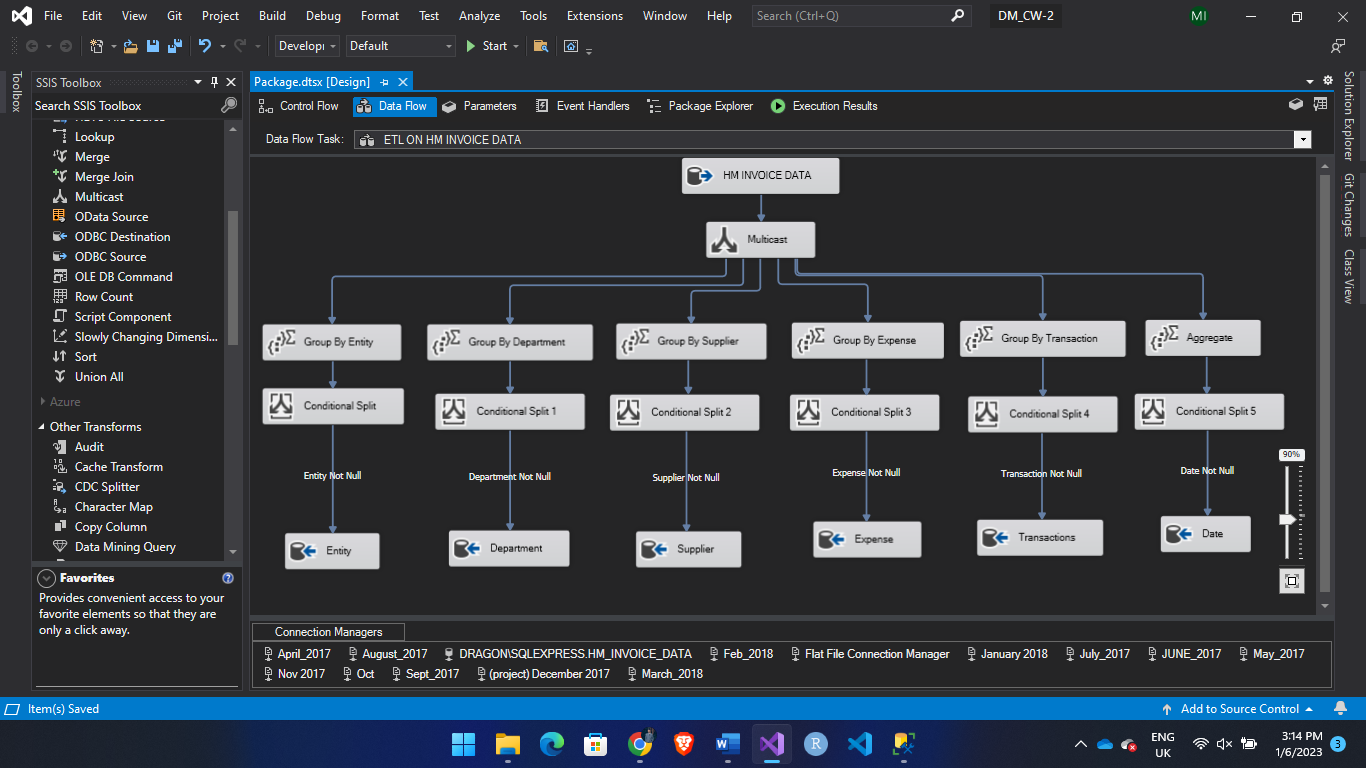


Figure 4: Data Cleaning in SSIS

In the VS SSIS Tool I used multicast to separate each table and used aggregate transformation to group each column. Figure 5 bellow displays some of the process.

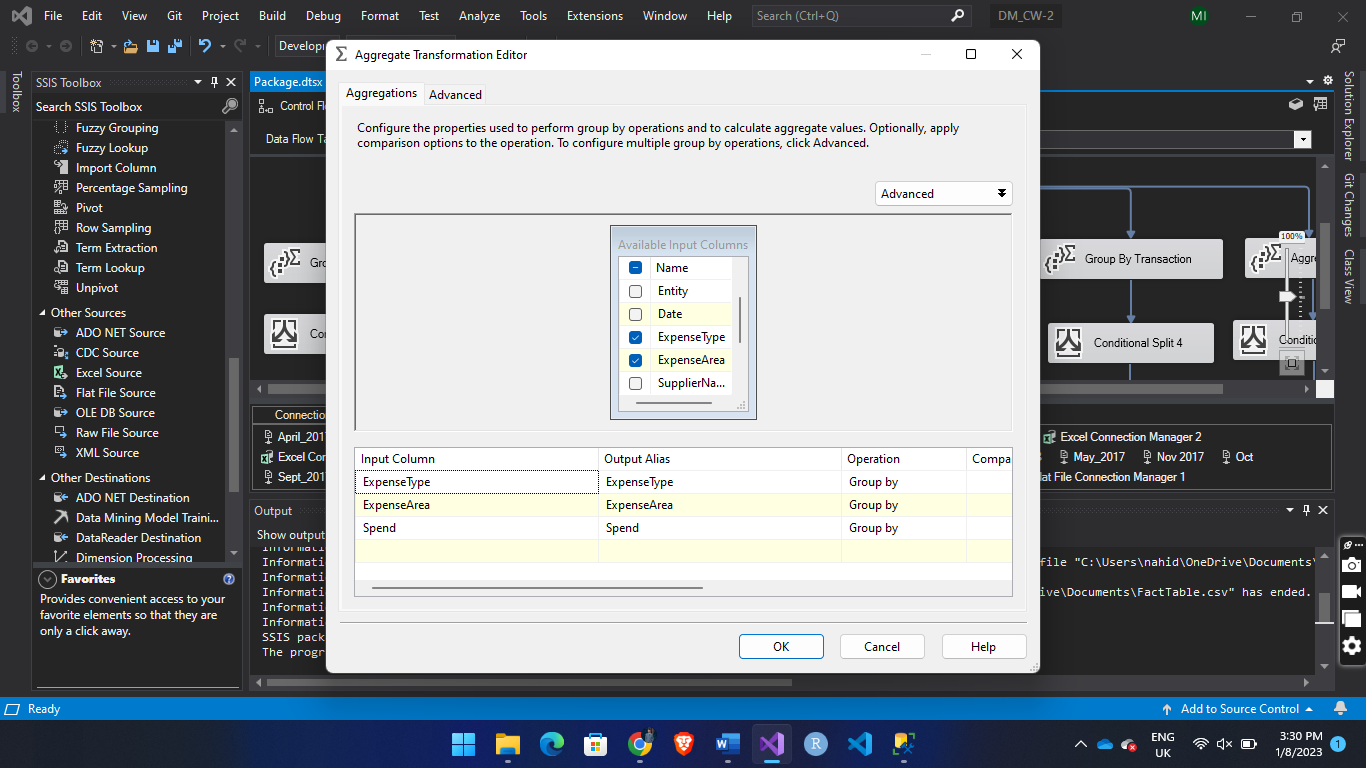
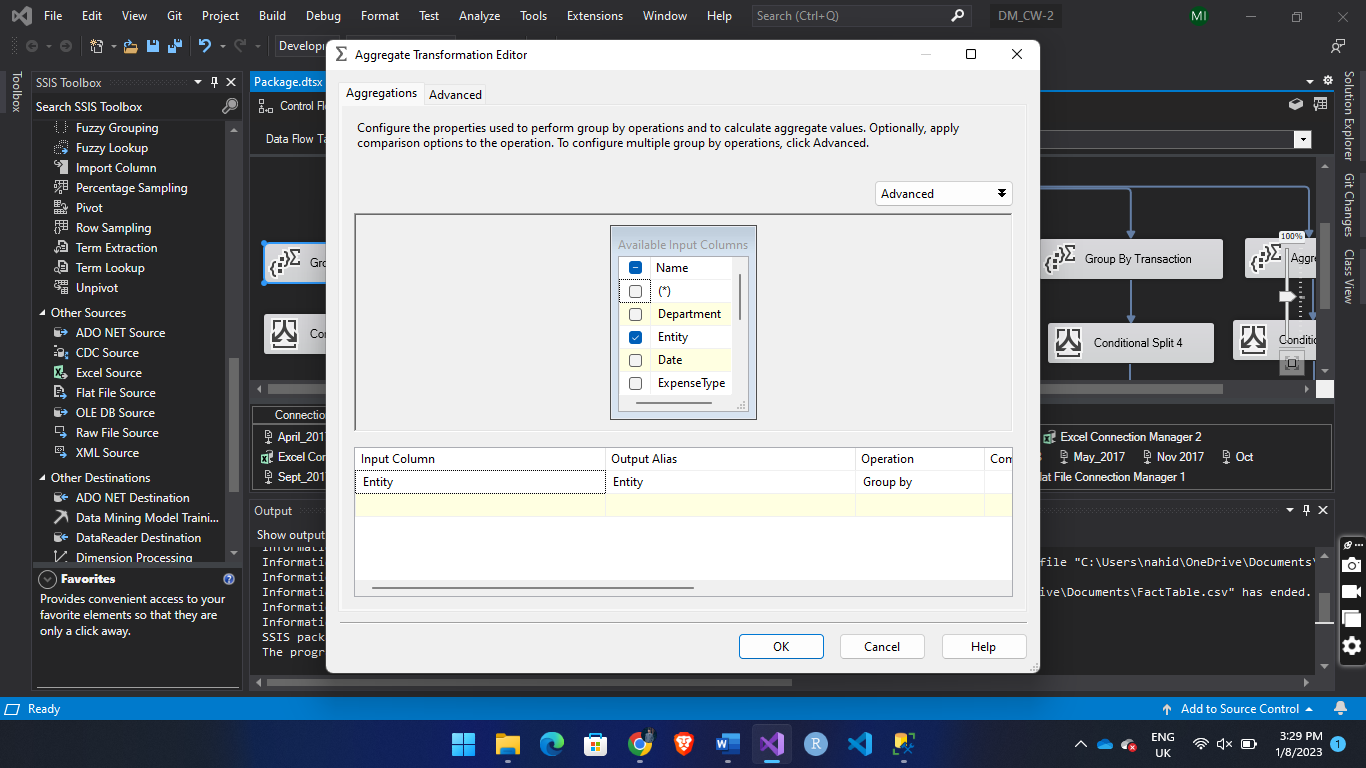


Figure 5: Grouping Each Table

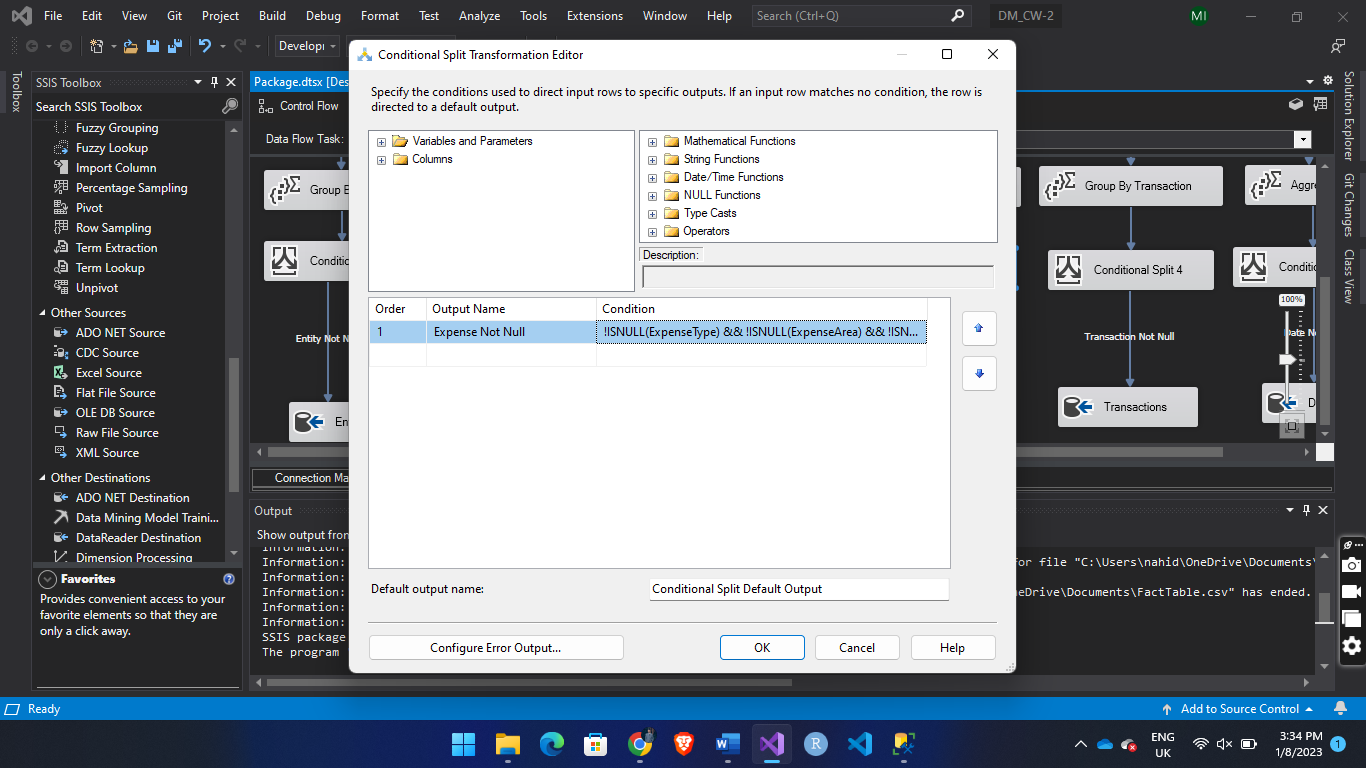
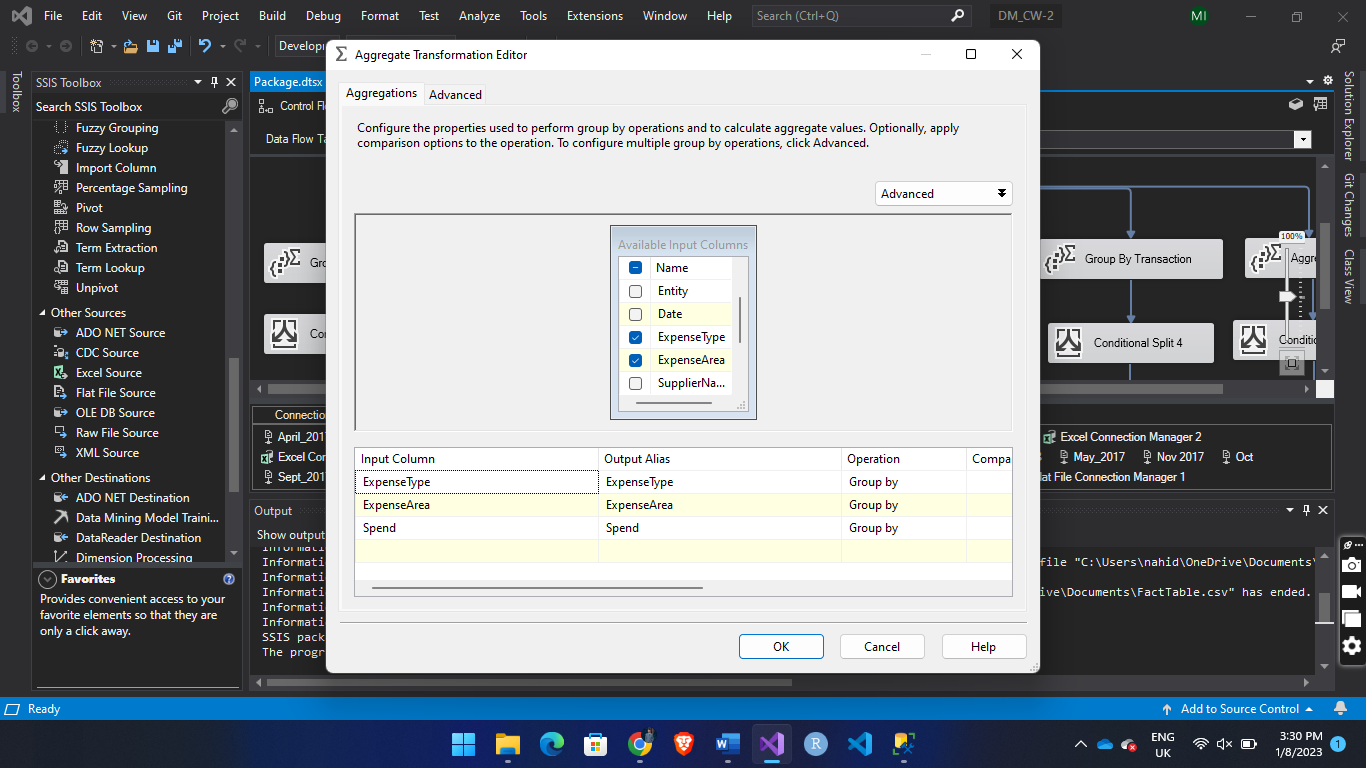


Figure 6: Cleaning Null values in SSIS

In figure 6 I used conditional split to remove null values from our data. The Following expression was used for cleaning null values

**!ISNULL(Entity)**

**!ISNULL(Department)**

**!ISNULL(Supplier)**

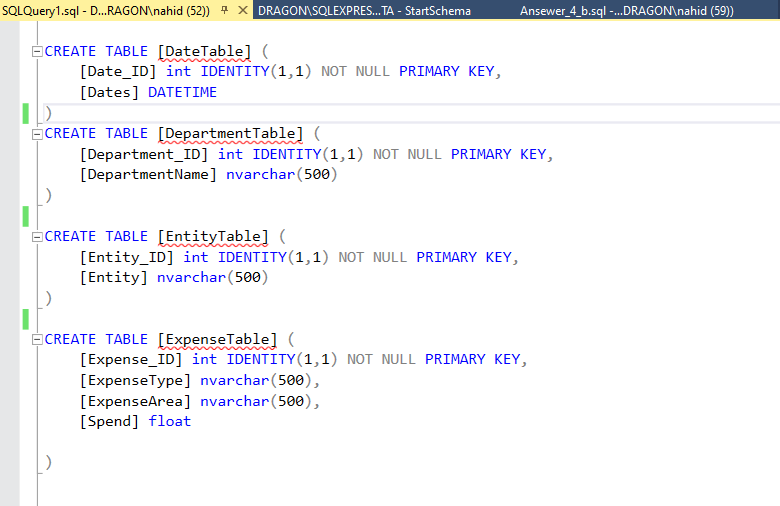
**!ISNULL(Date)**

**!ISNULL(Transacion)**

**!ISNULL(ExpenseType) && !ISNULL(ExpenseArea) && !ISNULL(Spend)**

Then the cleaned data was stored in 6 separate table which was created using SQL query

Below figures contains the SQL for creating **EntityTable , DepartmentTable , SupplierTable , DateTable , TransactionTable , ExpenseTable**

****

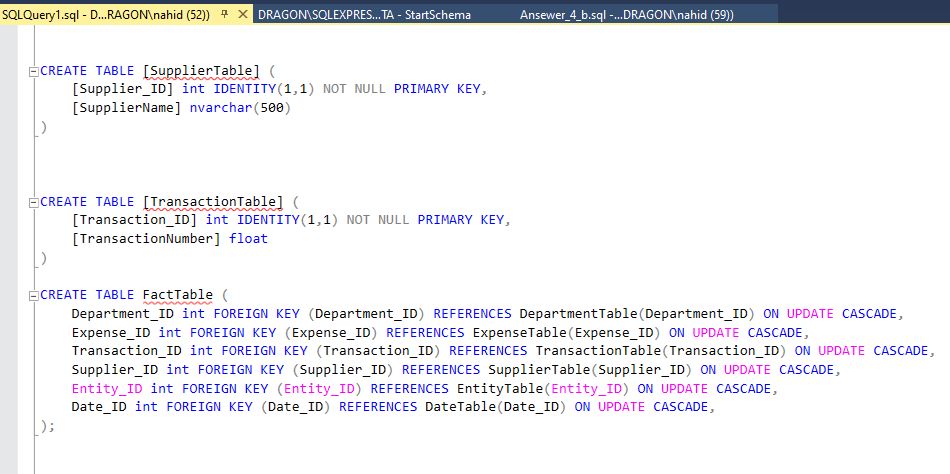
****

Figure 7: SQL Query to create Tables

After creating the tables along with the **Foreign Key** relationship on **Fact Table** a start schema was populated. Below picture shows the start schema.

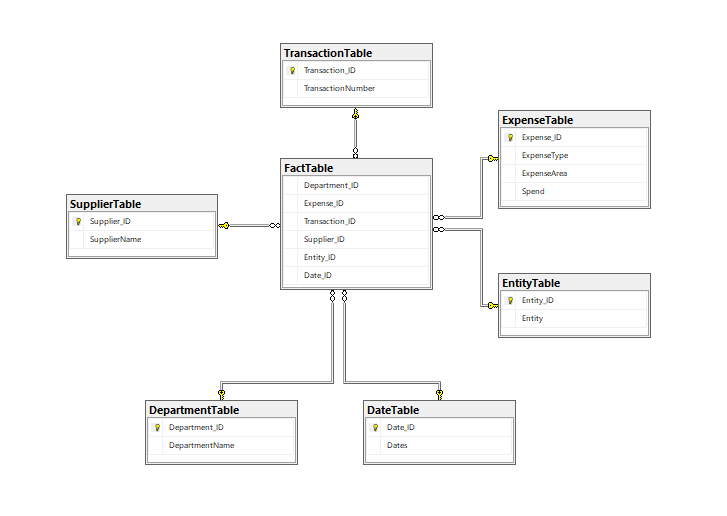


Figure 8: Star schema

After creating all these tables I inserted the cleaned data inside the table from SSIS using OLE DB Destination. Figure below displays the ETL where the cleaned data was inserted to our tables.

From above Figure we created 7 separate table.

Department Table, Supplier Table, Expense Table, Entity Table, Date Table, Transaction Table are connected to Fact Table in one-to-one relation.

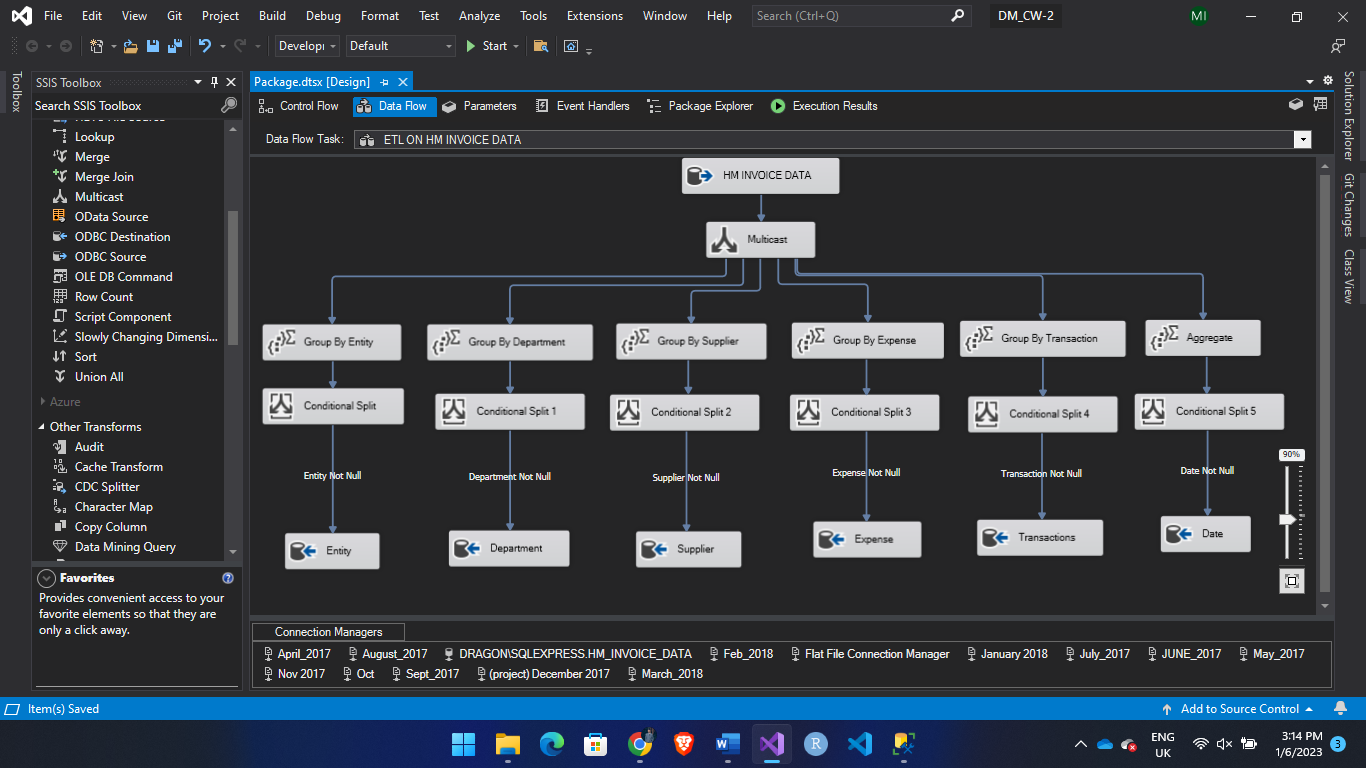


Figure 9: ETL

Critical Analysis – Transforming design to implementation

Merging 12 csv file inside the ETL was a long process and sometimes the values got mixed up with the other tables, however the issue was resolved later where I merged all the 12 files into single excel file and imported the data to the database using SSIS. Below figure demonstrate that.

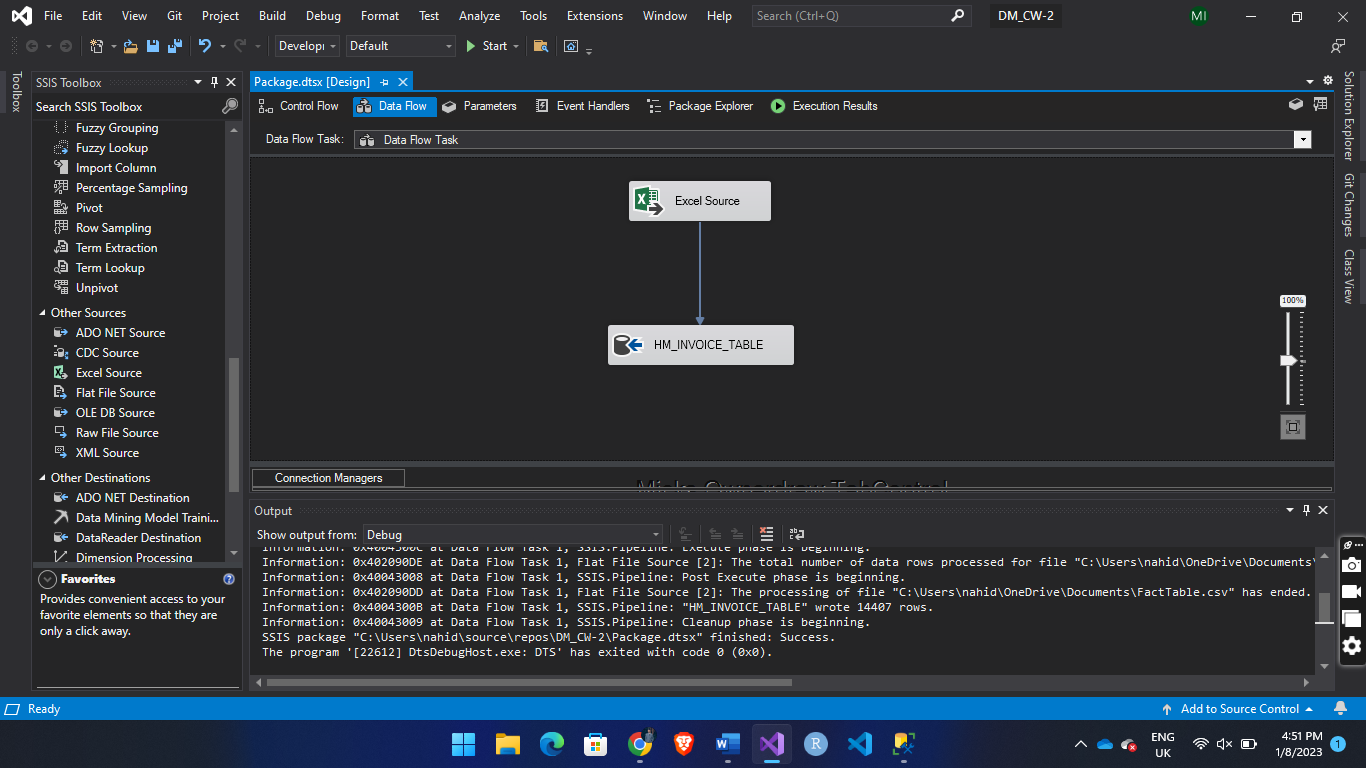
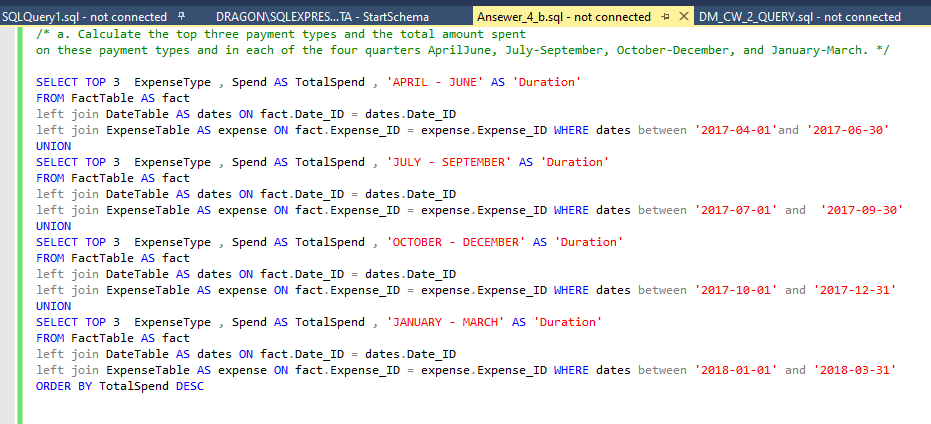
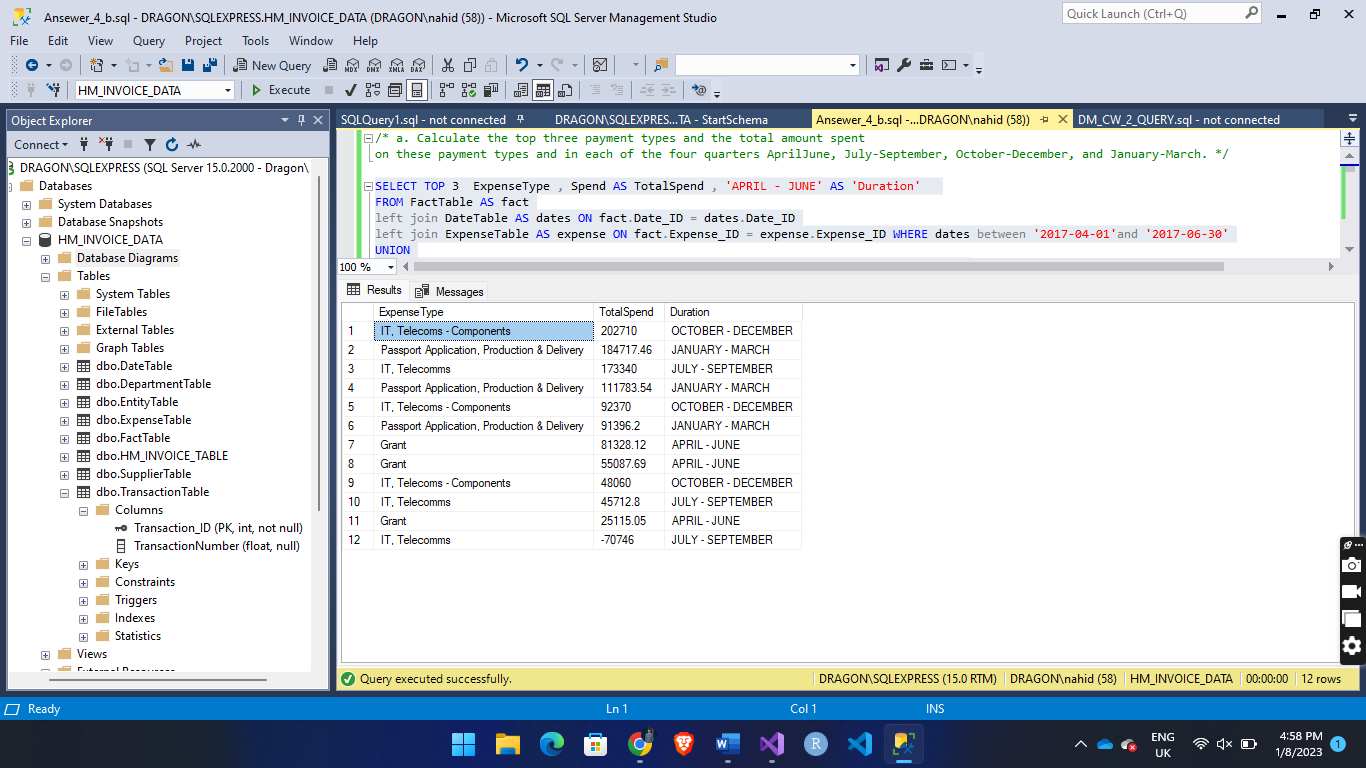


Figure 10: Importing data from excel source to OLE DB Destination

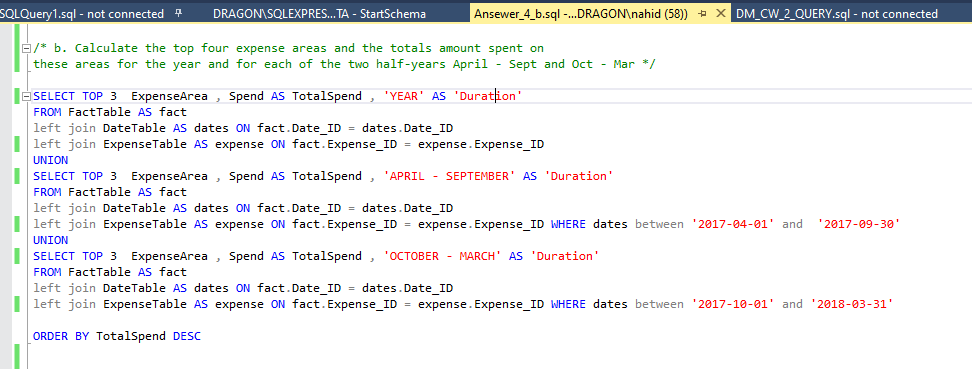
# Testing and Problem-Solving Skills

1. Calculate the top three payment types and the total amount spent on these payment types and in each of the four quarters April-June, July-September, October-December, and January-March.

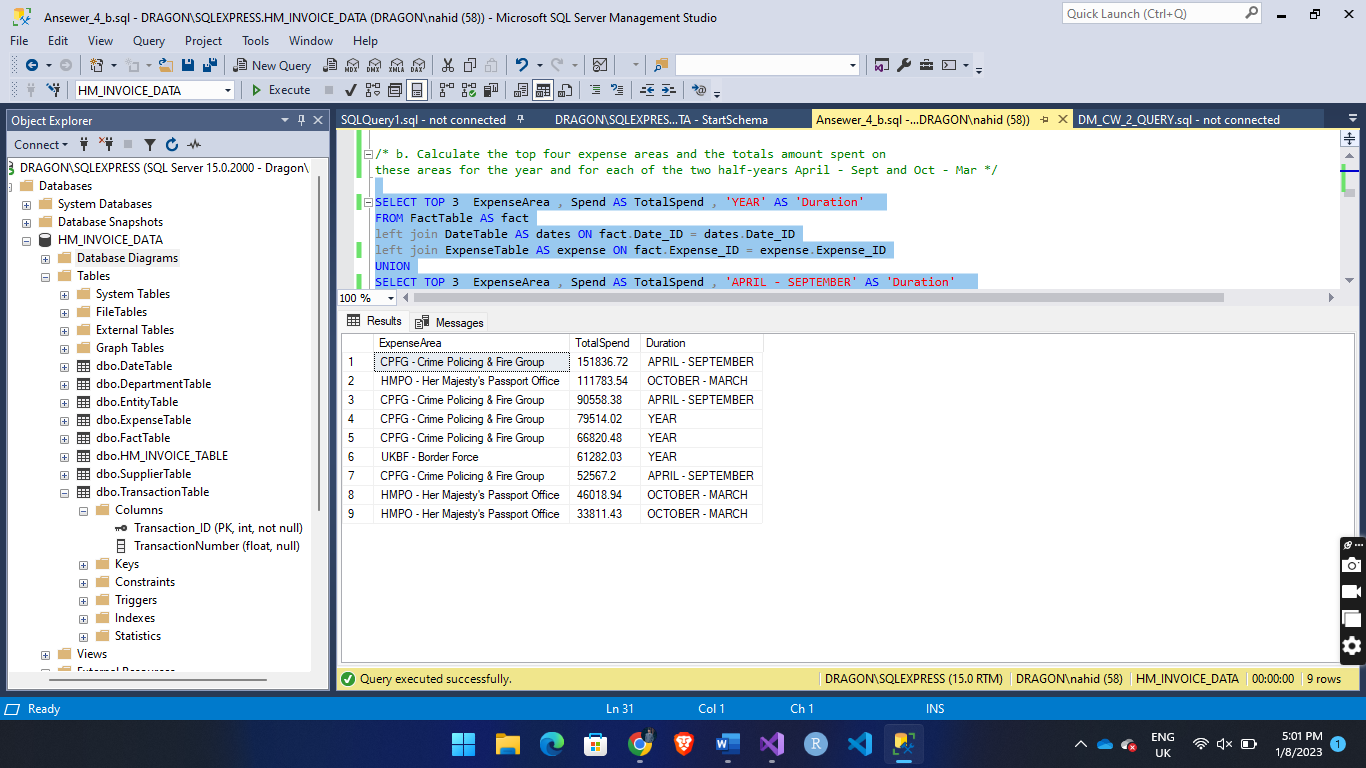


**Output**

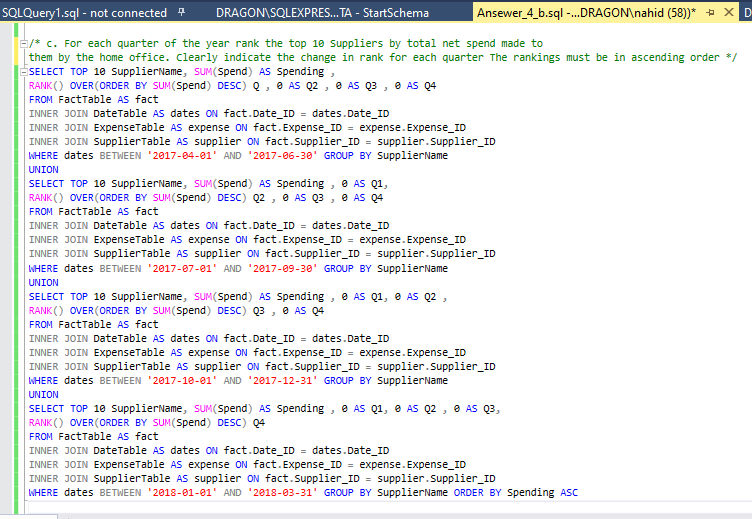
1. Calculate the top four expense areas and the totals amount spent on these areas for the year and for each of the two half-years April - Sept and Oct – Mar

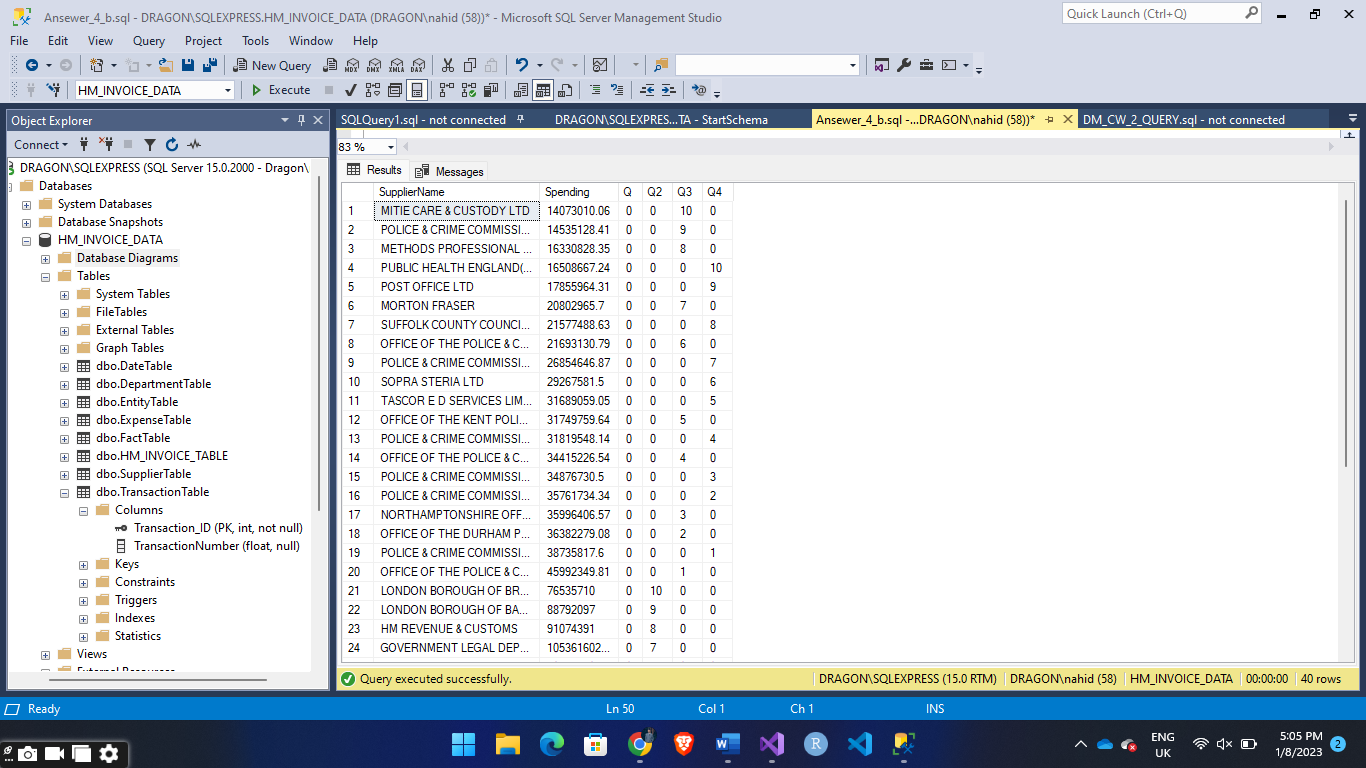


**Output**

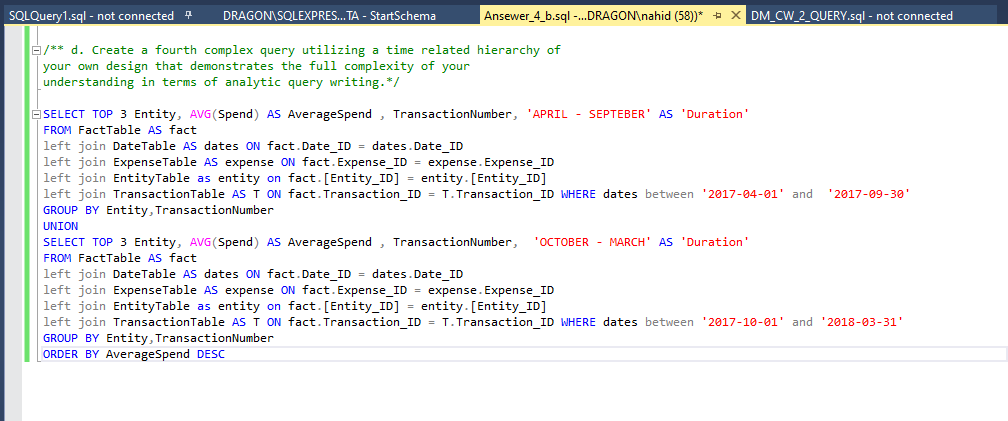
****

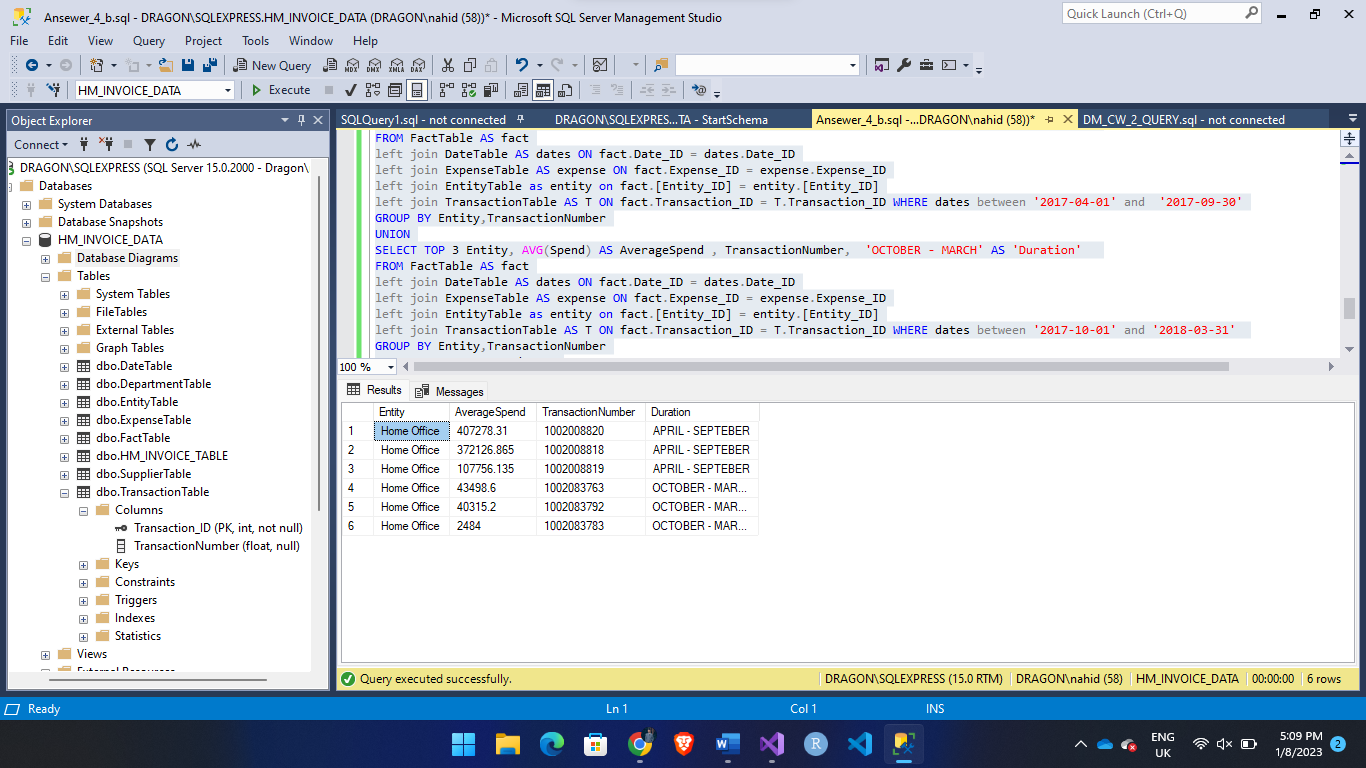
1. For each quarter of the year rank the top 10 Suppliers by total net spend made to them by the home office. Clearly indicate the change in rank for each quarter The rankings must be in ascending order



**Output**

1. Create a fourth complex query utilizing a time related hierarchy of your own design that demonstrates the full complexity of your understanding in terms of analytic query writing.

****

**Output: **

# Practical Competence – Dashboard

Connection Tableau with MS SQL SERVER to get the data

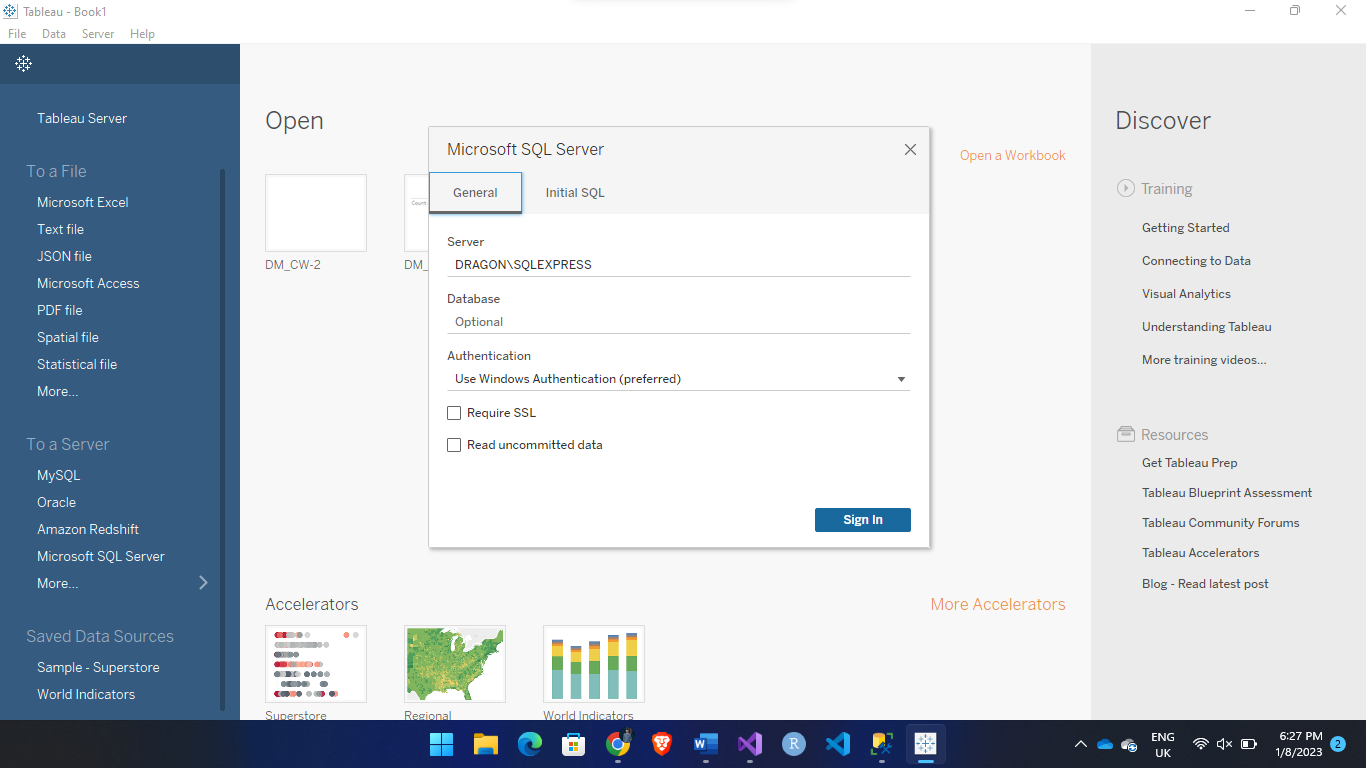
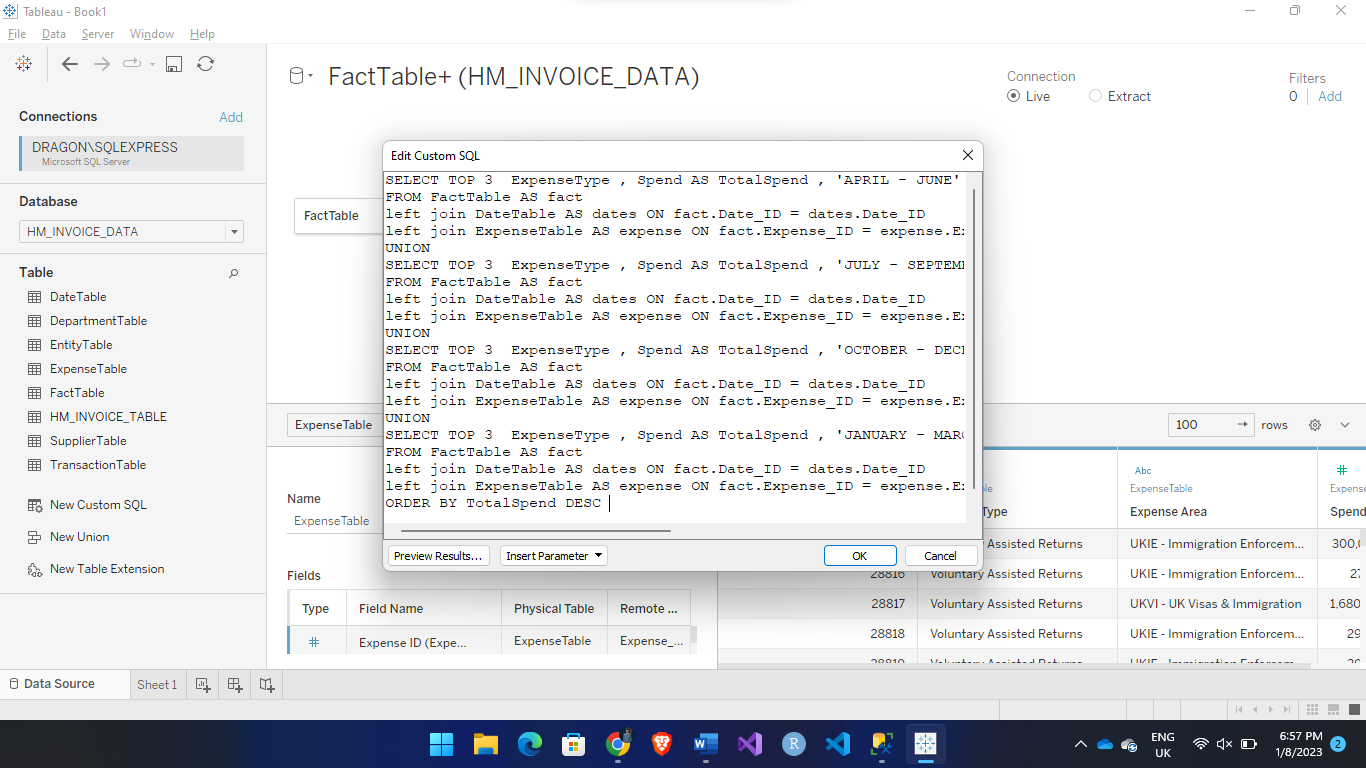
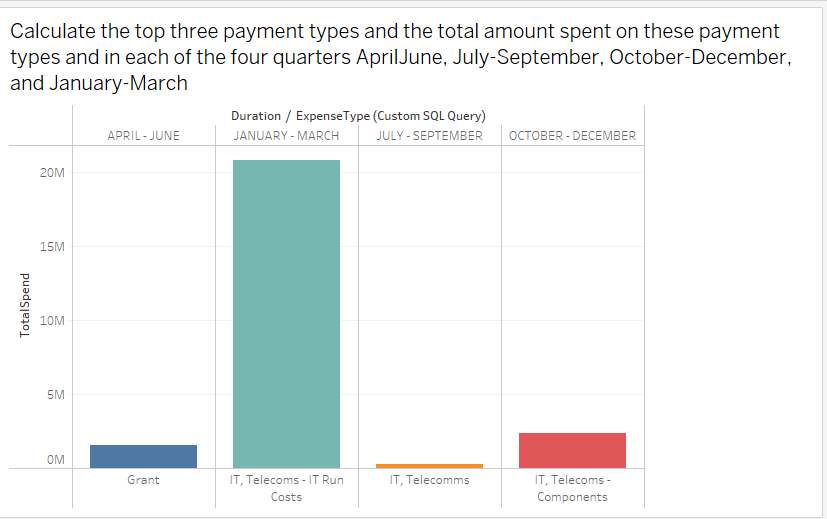
****

Figure Below shows hot create custom query in Tableau

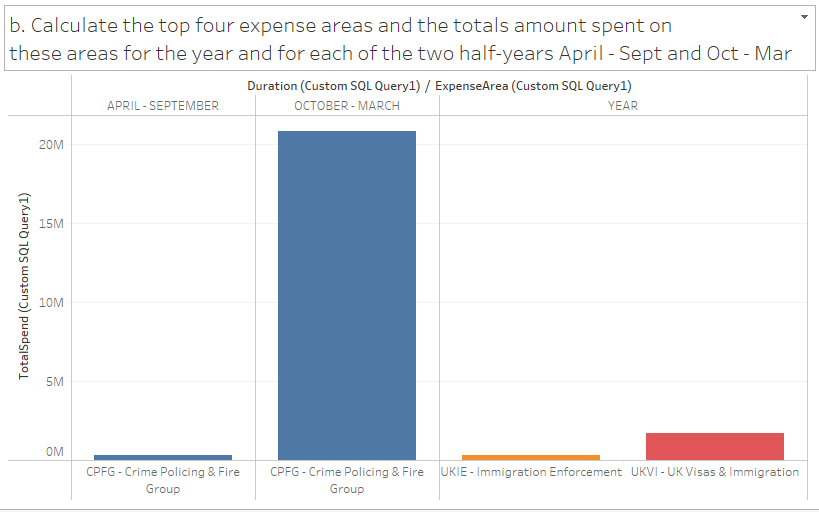
At the bottom of the left sidebar New Custom SQL button opens up a popup where we put our custom query to get the data and display in Tableau.

****

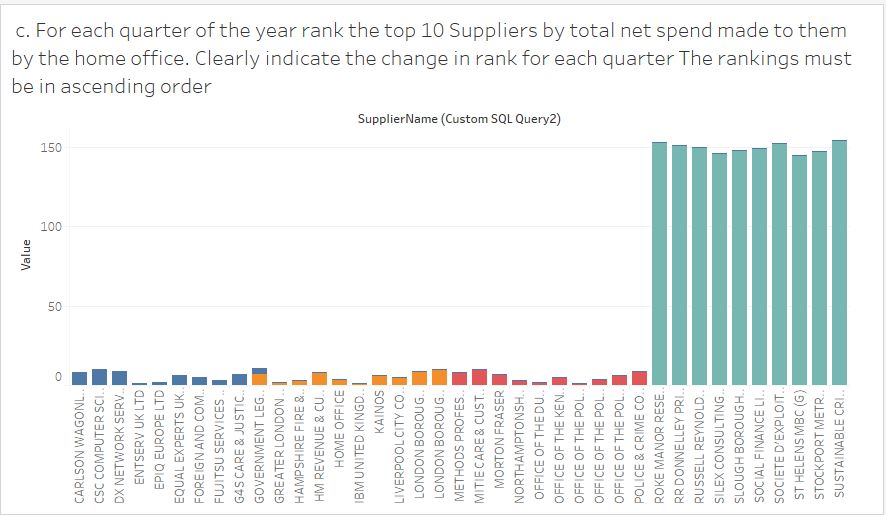
**Visualizing Frist Query**

****

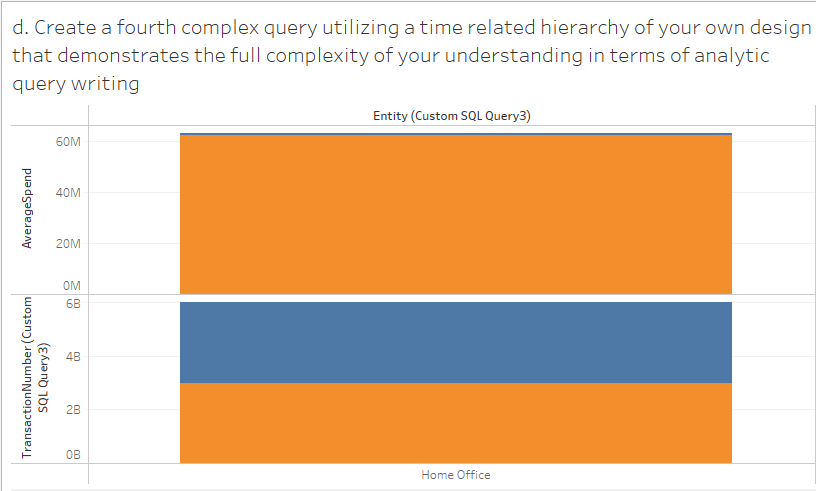
**Visualizing Second Query**

****

**Visualizing Third Query**

****

**Visualizing Fourth Query**

****

[**VIDEO DEMO**](https://stulsbuac-my.sharepoint.com/:v:/g/personal/islam124_lsbu_ac_uk/EYEuyMsn_qdDnVVbkB6z2scB5dj4MzBYp8WfSRyrpPGSFg?e=jzrRn9)