**This assignment is a data warehousing assignment and details of how the project was approached lies below;**

**I’ll first address the questions posed in the assignment.**

**Question 1.** Data Warehouse is developed to obtain business intelligence from data. Develop a proof of concept data warehouse/mart (using dimensional model) capturing data from an existing data source(s). Document your reasons for selecting the subject area(s), identify key stakeholders, formalise the business vision for developing the data warehouse from the given B9IS107 CA 2 data source. Also explain the insights that a company may attain from the given data. It should be reflected in the SSRS reports and Tableau visuals.

**Explanation 1:** here we are asked to develop a data warehouse. A data warehouse is a central point that holds data gathered from various sources together, a dimensional model or a dimensional data model is what the data warehouse is made up of, so the data warehouse is comprised of the dimensional data model. We are also asked in this number 1 question to show how we will achieve the building of this data warehouse, the steps involved and that is the proof.

Where a data warehouse is made up of the dimensional data model, one may ask ;what is a dimensional data model? **A dimensional model is just a data structure technique optimized for data storage in a data warehouse, It consists of** **dimensions** *(just like tables in a database but these are mainly qualitative data, like, ‘what’ – product, ‘where’ – location, ‘when’ – date , etc )* **and facts** *(this is also a kind of table, but a table that only contains primary keys of other tables and measurable quantities like profit, revenue, loss etc).*

So, a data warehouse is made up of dimensions *(qualitative data)* and facts *(primary keys with quantitative data/measurable quantities)* tables.

The key stakeholders or key performance indicators in the dataset we used are profit, quantity, sales and discount.

\*\* Here you will have to come up with a business vision and why the business needs a central data point *(a data warehouse).*

From the dataset *(https://github.com/mrarmsty/SSIS\_FILE/blob/main/data.csv)* we selected, we found insights like : that California, Texas, New York, Pennsylvania, and Michigan are the 5 states with the highest sales. However, Texas and Pennsylvania are not making profits but losses after discounts, despite the huge revenue they generate.

We also found that, among top 5 states with the highest good orders between 2014 – 2017, only 3 made profits after discounts were given.

And so on. Which the tableau visualizations will further represent.

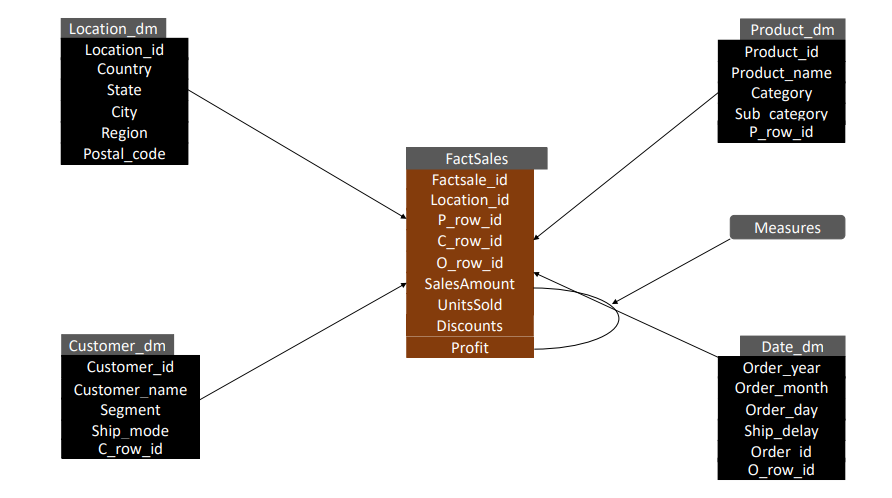
**Link to dataset and partitioned data**: https://github.com/mrarmsty/SSIS\_FILE

**Deliverable : MS word file.**

**Question 2.** Develop a dimensional model/star schema for developing Data warehouse.

**Explanation 2:** Here, we are asked to design a dimensional model of the star schema type. Let me say again that **a dimensional model has two major modeling types, the star model *(star schema)* and the snowflake model**, and in this project we are asked to develop a dimensional data model of the **star type,** so what this star contains is still dimensions and facts.The facts table is located at the center while the dimensions table surrounds the facts table to form a star-like design, just like the design I sent in PDF form. Remember why we need to build a dimensional data model? Without the dimensional data model, the data warehouse would not exist.

**Deliverable : Visio file** *(but Microsoft visio is a paid software, so I created the star schema in MS power point and saved as PDF).*



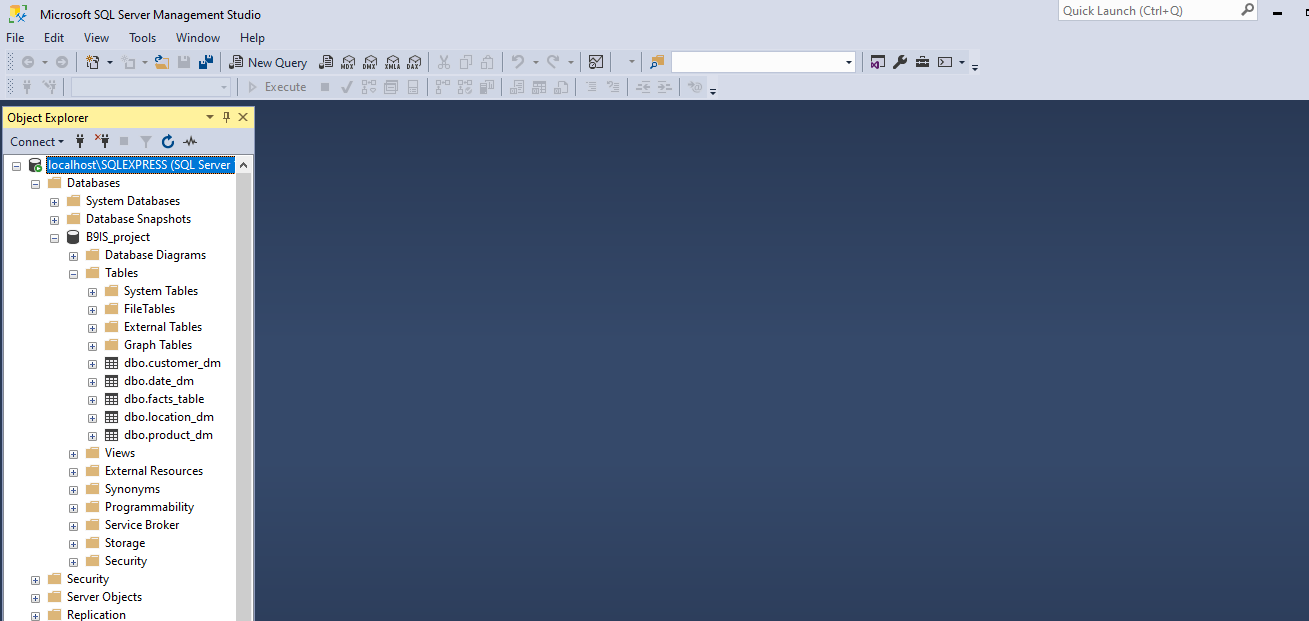
*Figure 1: Star schema*

**Question 3**. Implement the data warehouse in SQL Server by creating dimension and fact tables. Write SQL code for ETL or use an ETL tool to populate the data warehouse from operational database(s)/sources.

**Explanation 3:** I’ll list the tools used and what I did with each.

1. SSMS – SQL server management studio.
2. SSIS – SQL server integration services.
3. SQL server express.
4. MS Visual studio.

**SSMS** :

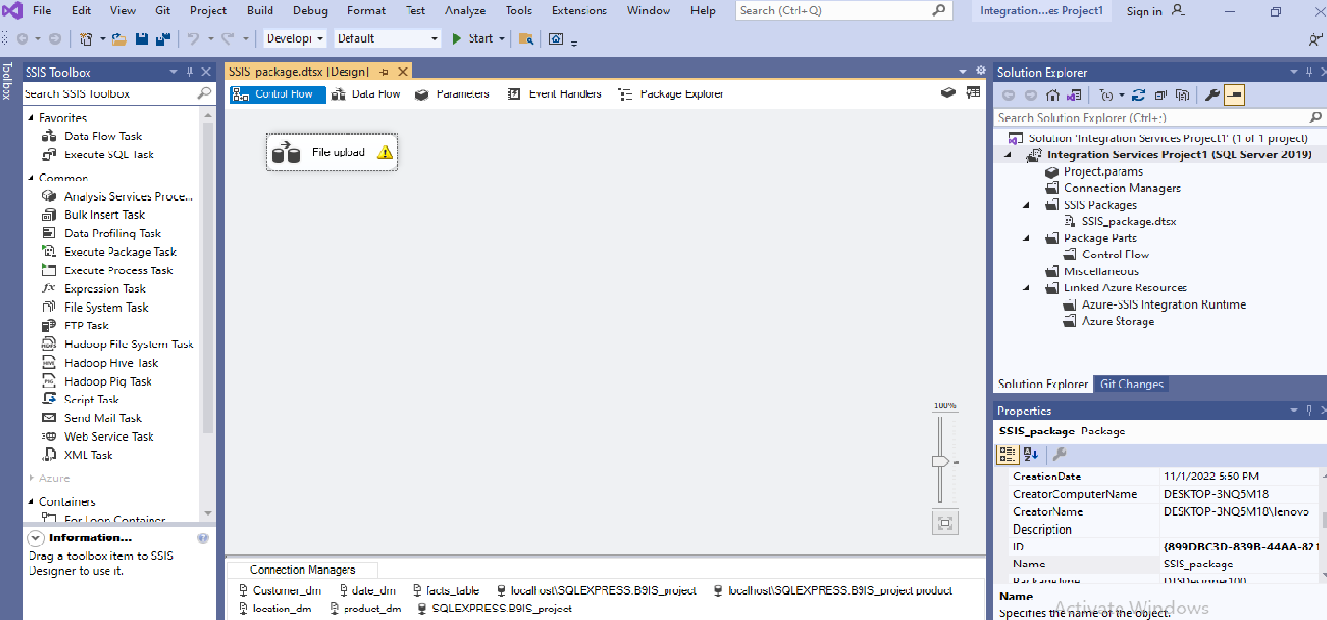
**

*Figure 2: SSMS interface*

This is a relational database management tool, like a software, it’s used to create and manage databases (and in this case – a data warehouse).

**SSIS** is a feature and part of Visual studio, you cannot perform integrations, i.e : movement of files from any data source to a database (or in this case – a data warehouse) without it.

**SQL server express** is just like a small server (software) that will be running locally on a computer to support other related software, and in this case – **SSMS and Visual studio**.

**MS visual studio:** 

*Figure 3: Data flow in SSMS*

This is a software that enables us to perform **ETL *(extract, transform and load)*** processes. It’s on this visual studio software that a SSIS file is created, it’s on this place that you push files into any database (in this case – data warehouse) whether the data warehouse is on a remote location or local *(on your computer)*.

Here I’ll state the whole process in the SSIS project file creation.

* I installed SSMS, visual studio and SQL server on my PC *(personal computer)*.
* Then I opened visual studio 2019 and installed SQL integration services, and restarted my PC.
* I started SQL server on my PC *(i.e ‘locally’),* then I also started SSMS. I took the server name on the display screen of the SQL server and filled the prompt on SSMS that requested for a server name. **The SSMS needs to be connected to a server before it can function, that’s part of the reasons why we installed SQL server express.**
* Then on SSMS I wrote SQL queries to create a database *(which will serve as our data warehouse)* and it’s associated tables *(which is our dimensions and facts tables)*, but I could not populate the database with just SQL queries, because it’ll possibly take upto a month of diligent effort before one could write enough code that’ll push the whole data *(1000+ rows of data)*  to the data warehouse, so, I used Visual studio code to perform the ETL *(the population of the data warehouse)* processes and hence, populated the data warehouse. All these things put together is the SSIS project.

*CREATE DATABASE B9IS\_project;*

*CREATE TABLE location\_dm(*

*location\_id INT PRIMARY,*

*country VARCHAR(40),*

*state VARCHAR(40),*

*city VARCHAR(40),*

*region VARCHAR(40),*

*postal\_code INT*

*);*

*CREATE TABLE product\_dm(*

*p\_row\_id INT PRIMARY,*

*product\_id VARCHAR(40),*

*product\_name VARCHAR(40),*

*category VARCHAR(40),*

*sub\_category VARCHAR(40)*

*);*

*CREATE TABLE customer\_dm(*

*c\_row\_id INT PRIMARY,*

*customer\_id VARCHAR(40),*

*customer\_name VARCHAR(40),*

*segment VARCHAR(40),*

*ship\_mode VARCHAR(40)*

*);*

*CREATE TABLE date\_dm(*

*o\_row\_id INT PRIMARY,*

*order\_id INT,*

*order\_year DATE,*

*order\_month INT,*

*order\_day INT,*

*ship\_delay INT*

*);*

*CREATE TABLE facts\_table(*

*facts\_id INT PRIMARY,*

*location\_id INT,*

*p\_row\_id INT,*

*c\_row\_id INT,*

*o\_row\_id INT,*

*sales\_amount INT,*

*units\_sold INT,*

discount FLOAT,

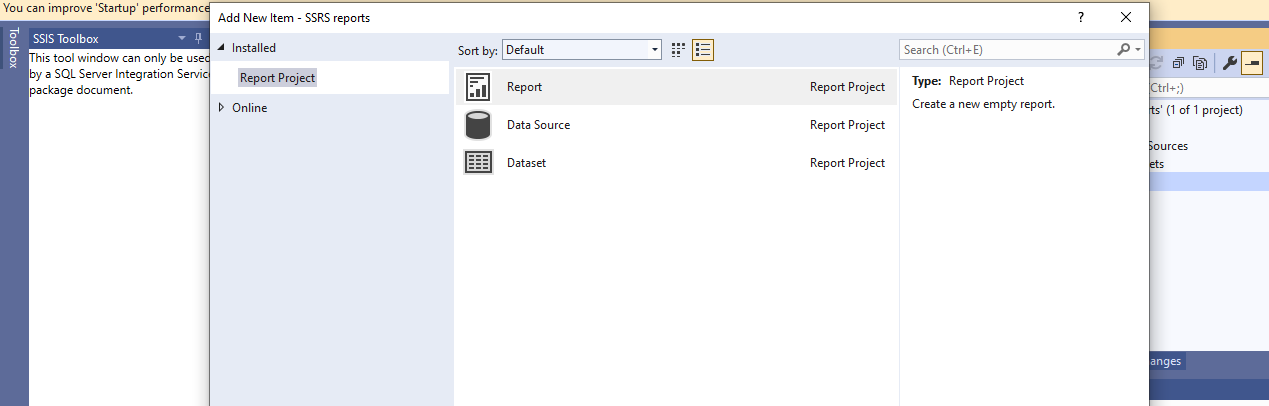
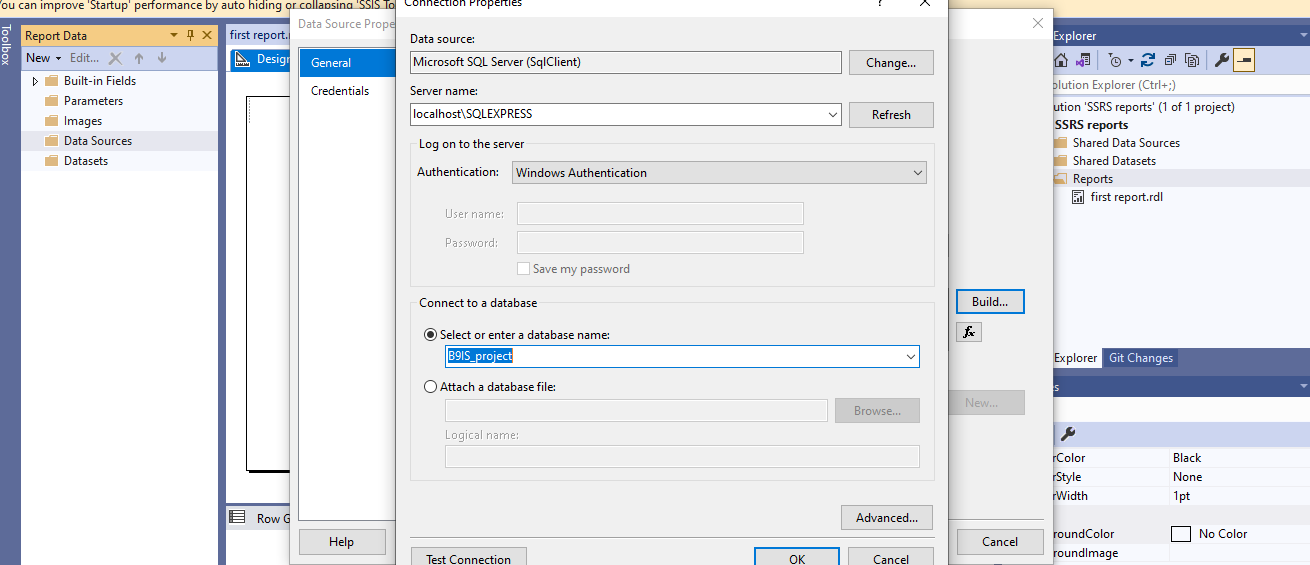
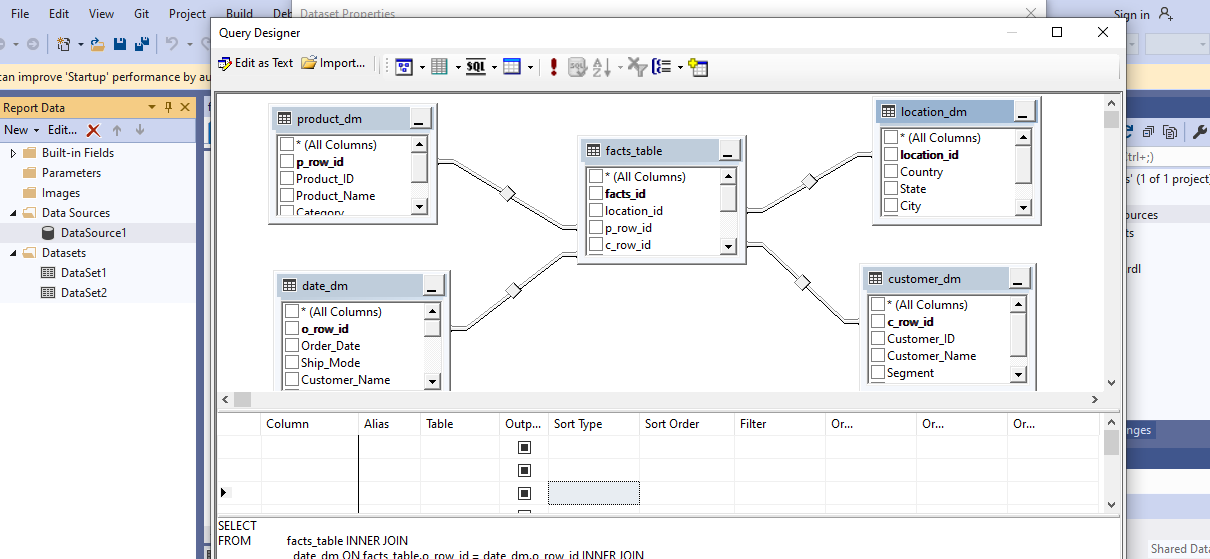
);

**File above ☝ :** SQL queries for creating the database *( used as a datawarehouse)* and the tables – which are the dimensions and facts tables.

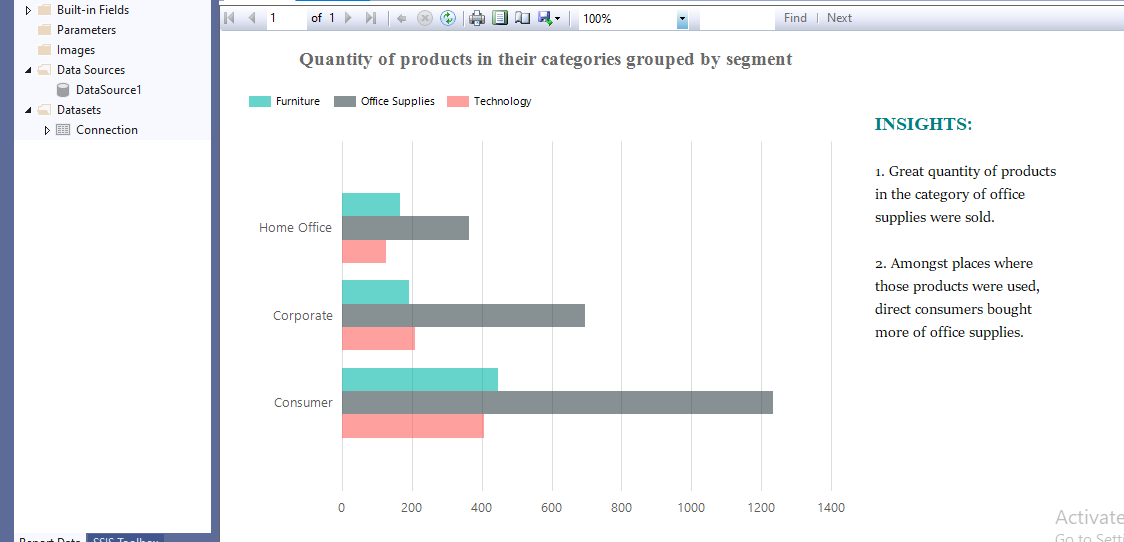
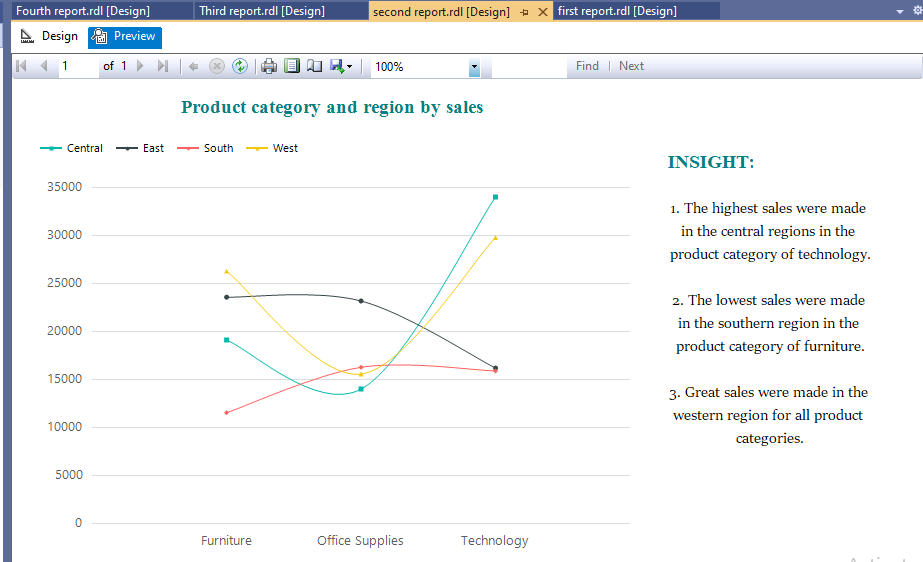
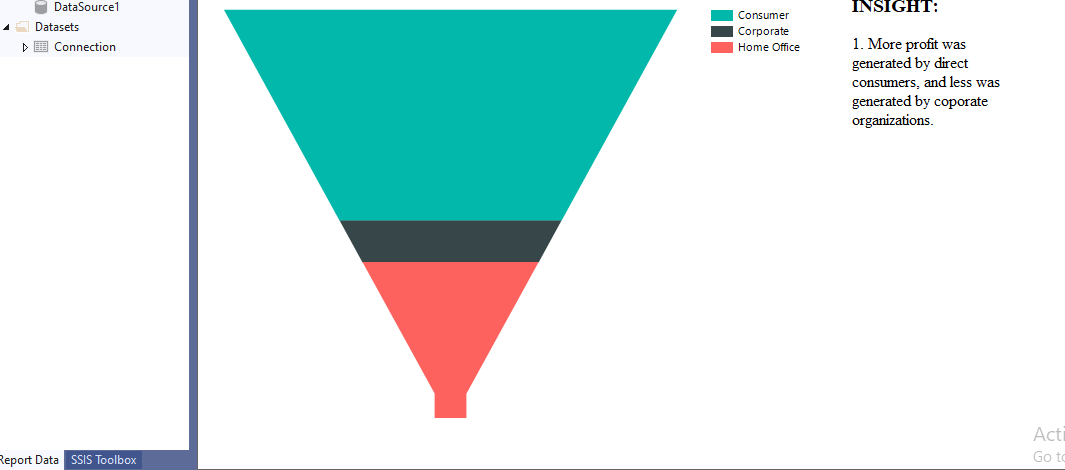
**Deliverable: SSIS project file.**

**Question 4.** Develop 4 reports using SQL Server Reporting Services (SSRS) demonstrating BI.

**Explanation 4:** In this case, SSRS is also, like SSIS, a feature and part of Visual studio, it has to be installed into visual studio before one can use visual studio to develop reports *(refers to findings/insights whether visual or not)*. So after it’s installation, connection to the already developed data warehouse through the server name has to be made, and after which the data in the data warehouse will be retrieved, so that analysis and subsequent reports can be developed/built.



*Figure 4,5 and 6: Procedures in setting up a data source and dataset connection in visual studio using SSRS.*

****

*Figures 7,8,9: Developed SSRS reports showing business intelligence in visual studio.*

**Deliverable**: **BI and Data analysis reports developed using SSRS.**

**Question 5.** Using your data warehouse, develop a dashboard in Tableau with four different multidimensional visualisations presenting data analysis/analytics of your data using several features such as colours, calculated fields, filters, trend lines, etc.

**Explanation 5:** Tableau is a business intelligence tool *(BI tool)* that aids decision making in business and other fields with its visualization/visuals *(pictorial/graphical representation of information/insight).* So The PDF file I sent contained the visuals, i.e the graphical representation of the insights drawn from the dataset, and all these with the data warehouse *(our already developed data warehouse)* as our data source.

**Deliverable: Visualisations.**

**Conclusion:** This project is just asking us to design *(dimensional modelling of the star type),* and map *(development)* a data warehouse, and state why we chose the subject/topic/dataset, key performance indicators or stakeholders of the subject or dataset, and explanation of insights that can be drawn from the subject/dataset after analysis, and to represent our findings visually.

We are also asked to return the whole process of developing the data warehouse *(I.e the SSIS project file).*

I think what this project is trying to do is to get the student to know SQL server management studio, how to design and map a data warehouse, how to create an ETL package using SSIS, how to retrieve data from a data warehouse, analyse and represent them visually using Business intelligence *(BI)* tools.

P.S: I uploaded the SQL code file, SSIS project files and SSRS files to GIThub *(https://github.com/mrarmsty/SSIS\_FILE)*. I separated the files and placed in different folders for easy navigation, also the SSRS reports are in .rdl format, this means that the files cannot be viewed without the right software.

Thanks,

Armstrong.