# DWMA Assignment 2 – Implementation for the case study requirements

# Note that this is extra guidance (Check marking schema and assessment guide for full details)

# Introduction

Set the scenario of your Report. Very brief and concise. This assignment submission will be via Turnitin.

# TASK1 – Data Integration (ETL) and Maintenance

## 1.1 Original datasets

* Create (upload – in this folder see recording ‘upload tables’) your datasets (tables) in your own Apex account for the original data sources (operational database) that are used in your SS. You have defined these previously in Assignment 1 – Task 1. Only those need to be included at this point. If you haven’t considered previously where will your data come from for your SS, now you have to do it?
* Evidence code, brief discussion, and screenshot that it is running successfully, making clear your own Apex account id in the screenshots.

## 1.2 Star Schema Implementation

* Have the QSEE diagram of your SS here.
* Create your star schema with your dimension tables and fact(s). Remember tables (or dimensions) are created using a command that refers to datasets. Here you can Forward engineer your tables. How this is done, watch the recording in this folder on ‘How to forward engineer your SS’. This will generate a sql script for your tables.
* Now you can consider surrogate key(s) in your dimension. Will you implement a sequence for your surrogate keys or use natural primary keys?
* Consider your measures for your reports, such as Avg of X per Y, or other calculations total, min… more aggregates…
* Decide on the Slowly Changing Dim Type you will be using… Implement it within your SS. Note: you can also wait with this until you are happy with your ETL without it.
* You may have done many of these tasks already in Assignment 1. Now you may review based on our feedback.
* Evidence code, brief discussion, and screenshot that it is running successfully, making clear your own Apex account id in the screenshots.

## Staging Area

* Are you going to use the Staging area approach? You may use Tableau prep for some parts of the task, but note that the assessment is asking for code done in Apex. If you going to extract and load data straight into your dimension table? If it is the latter one, which could be ok for some dimensions, not for all, you’ll not do this part. Note that for a good and excellent solution, we expect you to have done this.

### Extracting of data into a temp table?

* You’ll need to move data from one dataset to another dataset, so you can “prepare” data before you enter into the SS. E.g. Code: Insert into temp table as select x, y from original table

OR

* You can merge two datasets into one – concatenated data sets. These datasets are then cleaned in the next part. E.g. Code@ data x; set dataset1, dataset2, dataset3; run;
* Evidence code, brief discussion, and screenshot that it is running successfully
  + 1. Cleaning of data
* Show here what data were missing data, delete unnecessary data, and check if there is the same format in data?
* Produce an Error log, if any, is needed.
* Evidence this by presenting code, brief discussion, and screenshot that it is running successfully in your Apex account.
  + 1. Transforming of data
* Would you need to transform data? When? Why? How? Would you need to have a table that will hold error data? Null values, how will this be stored? Merged fields? Data type change? Have records stored in UPPER case in SS. **Design this ETL script. Have all considerations in a table.**
* Evidence code, brief discussion, and screenshot that it is running successfully
  + 1. Loading data into SS
* This is cleaned and transformed data!
* Clearly show the code for each dataset (dimension) population.
* Define surrogate keys for dim: e.g. Code: data x\_dim; set x\_dim; counter=\_N\_; run;
* You will need to consider

How will you populate the FACT table? What calculations of measures have been made? Plan it! Have a table to show your plan and if necessary assumptions you’ve made.

How will you populate your Time Dim?

NOTE: you may have chosen to populate dimensions directly from the original sources. In any case, you will need this part only.

* Evidence code, brief discussion, and screenshot that it is running successfully

1.3.5 Consider now - Ongoing Data Load (Data Maintenance (SCD)) implemented and evidenced as part of the solution. See the tutorial in week 7. You can wait with this until you complete tasks 2 and 3, before going back to implement SCD.

# TASK 2 - OLAP

* Decide on the technology to use to support the OLAP (pivot table) investigation – Access and Excel or Tableau?
* Decide on the model to use to support your OLAP. A detailed table or a star schema model (fact and dimensions).
* Design the OLAP investigations (dimensions – e.g. area: town, city, county and time: week, month, year)
* Create a query in Access (or other software) that can be imported into excel and provide a useful pivot table.
* Create screenshots of the data analysis provided in the pivot table (OLAP as far as this exercise is concerned).
* This exercise is about understanding, you shouldn’t need to significantly rework anything, you will probably provide just a subset of the data.
* Use literature to support your choices and arguments.
* Note: you do not need to evidence the steps undertaken e.g. Pivot table was created.

# TASK 3 – Findings and Reflection

Check to mark the scheme and reflect upon your findings, methodology, processes, methods, and possibly the architecture that you will find most appropriate to implement for your case study organization.

When you evaluate the current and potential **social and ethical issues for the case study organization**, consider some of the theories that will underpin your discussion, such as the PAPA framework, BigPapa, GDPR issue of centralized data stores, Theory of Stakeholders, etc…

The theory covered should reference back to the case study organization! Also, make sure text is paraphrased and citations included, as well as include references using up to date authoritative academic resources.

# Bibliography

# Appendixes – add of your APEX code here