Instructions:

Submit the following items:

1. **PDF report generated by PowerBI according to these 8 steps:**

https://www.youtube.com/watch?v=X9s1J\_pfSjY

1. Think about a few **business questions** that your data warehouse could help answer.
2. Draw a StarNet with the aim to identify the dimensions and concept hierarchies for each dimension. This should be based on the lowest level information you have access to.
3. Use the StarNet footprints to illustrate how the business queries can be answered with your design. Refine the StarNet if the desired queries cannot be answered, for example, by adding more dimensions or concept hierarchies.
4. Once the StarNet diagram is completed, draw it using software such as Microsoft Visio (free to download under [the Azure Education Link: https://aka.ms/devtoolsforteaching](https://aka.ms/devtoolsforteaching)) or a drawing program of your own choice. Paste it onto a Power BI Dashboard.
5. Implement a star or snowflake schema using SQL Server Management Studio (SSMS). Paste the database ER diagram generated by SSMS onto Power BI Dashboard.
6. Load the data from the csv files to populate the tables. You may need to create separate data files for your dimension tables.
7. Use SQL Server Data Tools to build a multi-dimensional analysis service solution, with a cube designed to answer your business queries. Make sure the concept hierarchies match your StarNet design. Paste the cube diagram to your Power BI Dashboard.
8. Use Power BI to visualise the data returned from your business queries.

**2. SQL Script File and CSV files for building and populating database**

**3. Solution project file (and it folder) of the SSDT (SQL Server Data Tools) Analysis service multi-dimensional project**

**4. Scripts for data cleaning for data transformation (shell script?)**

**5. PDF descripting the associate rule mining process and results**

* Explain the top k rules (according to importance or probability) that have the "crime" type (or other suitable columns) on the right-hand-side, where k>=1.
* Explain the meaning of the k rules in plain English.

So 1 (PowerBI Dashboard PDF) and 5 (Associate rule mining process based on these 2 dot-points) are PDFs submission.

First step:

Following four steps below of **dimensional modelling** (i.e.  Kimball's four steps), design a data warehouse for the dataset(s).

1. Identify the process being modelled.
2. Determine the grain at which facts can be stored.
3. Choose the dimensions
4. Identify the numeric measures for the facts.

Lab 2: Install the required softwares

* SQL Server Management Studio (SSMS)
  + SQL Server Data Tools (SSDT), extension required for this assignment
* Microsoft Visual Studio (Community Version)
* PowerBI Desktop

Lab 3: Learn how to use softwares to design your data warehouse

* Load a simple data warehouse (an SQL Script to run excel files to populate the data and construct a database). Using SSMS to construct a data warehouse.
* Generate Database Diagram and Schema using SSMS.
* Connect PowerBI Desktop

Lab 4: Understand how to design a new Data Warehouse based in these three steps:

1. Using SSMS to design a new data warehouse.
2. Build a data cube for drill-down and roll up analysis in SSDT.
3. Visualise data using PowerBI.

Lab 5: Processing Data using Microsoft Integration Services for Extraction, Transformation and Loading (ETL) aka data cleaning process if you know Python or R (OPTIONAL). Learn this tool to help you do the ETL process.

Lab 6: Understand how Rule Mining works using SQL Server Analysis Services.

[5 marks] Schema of each Dimension and Concept Hiearchies for each Dimension

[5 marks] Corresponding StarNet to illustrate query capabilities of the DW

[5 marks] At least 5 types of business queries that the StarNet can answer

[5 marks] Star/SnowFlake Schema (Fact Table) for DW design

[5 marks] Description of the ETL process for data transformation with code or screenshots

[5 marks] SQL Script file for building the database and loading the datasets

[5 marks] Power BI visualisation corresponding to the 5 business queries

[5 marks] Coherence between the design and implementation, quality and complexity of the solution, reproducibility of the solution

[5 marks] Association rule mining meaningful set up

[5 marks] Interpretation of top rules

Step 1: [5 marks] Description of the ETL process for data transformation with code or screenshots

Step 2: [5 marks] SQL Script file for building the database and loading the datasets

Create an SQL script for dimensions and fact tables.

Then figure out later how to insert the data.

Step 3: [5 marks] Star/SnowFlake Schema (Fact Table) for DW design

Step 4: [5 marks] At least 5 types of business queries that the StarNet can answer