Final Assignment: Group Project

Assignment 08, Assignment 09, and Assignment 10 is a group project worth 300 points. You must complete the final project to pass the course. The goal of the final is to demonstrate in a team setting you can:

Recommend platform: Power BI

* Work with an assigned dataset and chose another dataset to report on:
  + The assigned dataset is Bing Covid19 dataset: <https://github.com/microsoft/Bing-COVID-19-Data> (You can use the class dataset on our servers)
    - You cannot download this dataset from GitHub anymore. You will need download it from Azure. Which is here: [Bing COVID-19 - Azure Open Datasets | Microsoft Learn](https://learn.microsoft.com/en-us/azure/open-datasets/dataset-bing-covid-19?tabs=azure-storage).
    - We then need to scroll down to the .csv file located toward the bottom of the page: CSV: <https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/bing_covid-19_data/latest/bing_covid-19_data.csv>
    - When I downloaded the dataset it was: 560 mb
  + Chose an additional dataset from a website:
    - I used a dataset for Unemployment (professor’s example) <https://oui.doleta.gov/unemploy/claims.asp> then joined the dataset on State and Unemployment File Date. You should do something similar.
    - To upload the data to SQL Server I used the method we will demonstrate in class. I like to use the SQL Server interface on my local instance. Then backup the database as a ‘Export Data-tier Application’. This will create a .bacpac. Then all we do is ‘Import Data-tier Application’ . These settings are under Task.

A screenshot of a computer

Description automatically generated

Required Secondary Datasets

Last year we had a group struggle for weeks to find a secondary dataset. This is a data management course, not a business class.

Therefore, if you don’t pick a dataset and clearly explain your testing on it. You will be assigned a dataset below. An easier approach it to pick one from the Azure Open datasets.

[Example Jupyter notebooks using NOAA data - Azure Open Datasets | Microsoft Learn](https://learn.microsoft.com/en-us/azure/open-datasets/samples)

* Build a data warehouse in SQL Server:
  + Build a database on your local instance.
  + Export / Import that database to Azure SQL Server
  + Properly backup the databases and put in a folder
* Build a Data Visualization with data:
  + Build one dashboard.
  + Build two different reports using multiple data visualizations.
  + Demonstrate additional ways to manage the data using data management & maintenance techniques we discussed.
* Explain how the data could be refreshed in a production environment:
  + The requirement here is just how you would refresh the dataset. The Big Data Technologies course build more on this principle if you chose to enroll.
* Present your finding to the class:
  + Each team will have 30 minutes to present their findings.
  + Presentations will be Week 9 and Week 10
  + You must turn all parts of the assignment on time in canvas

This activity will take you about 10 to 15 hours, so plan accordingly!

**milestone 1: Confirm team**

# 25 points: Contact your team: Due by aPRIL 16, 2024

The recommended way to contact your team is to use Canvas and Zoom. However, if you would like to use a different portal like Slack, please feel free.

The requirement for the course is to message both instructors through Canvas with all your team members. A simple message will be sufficient. Below is a sample.

**Sample (with all Team Members on the message)**

*Hi Tim & Mary,*

*This is Team Black (your group color will go here). I am contacting you today that all team members have met virtually, and we look forward to presenting the first deliverable to you on April 29, 2021.*

*Thanks, Team Black (your Team Color)*

Note: If you are having an issue contacting a teammate and they are not responding please message Tim and he will reassign you a new teammate.

**milestone 2: Secondary dataset**

# 50 points: Due by aPRIL 30, 2024

**Step 1**: Create an excel document or just use mine. If you can think of a better way of providing the details. Please feel free.

**Step 2:** Find a secondary dataset. All students will need a DimDate table. At this point you have taken two ETL Courses. I am not going to explain why you need one, but you will have several examples in class on how to build one in DAX and SQL.

**Step 3:** Assign your teammates roles (optional). I know this can be difficult and if you are having an issue, please feel free to reach out to the teacher. I suggest the individual who sent the email will make an excellent project manager. Perhaps let this person assign the roles.

**Step 4**: After you complete the form, please use the following format to submit to canvas. Each Team Member must submit it as follows in the following folder structure:

|  |
| --- |
|  |

Graphical user interface, application

Description automatically generated