What we have:

1 - Three different sources of data from two different databases (MongoDB and SQL Server) that capture the same type of data.

|  |  |  |
| --- | --- | --- |
| **MongoDB** | **SQL Server** | **Description** |
| universitiesCollection.xlsx | universities\_table.csv | Contains university data. |
| studentFeedbackFormsCollection.xlsx | student\_feedback\_table.csv | Contains all the student feedback responses from classes from all universities. |
| classesCollection.xlsx | classes\_table.csv | Contains data from all the classes from their respective universities. |

2 - Using the data sources from the table below, we need to combine the three data sources somehow and have them each act as one unified source of truth for the data they’re capturing (i.e. **universitiesCollection.xlsx / universities\_table.csv** act as the single source of truth for universities).

3 - Documentation on your process and steps you took to solve would be ideal.

**With Python:**

1 -Create a **dashboard**, specifically using a **pivot table;**

2 - View the averages for:

* Pace
* Self-Mastery
* Support
* Instructor Clarity
* Instructor Engagement
* Instructor Knowledge

**for each class** and university.

3 - Create **two additional dashboards** in Jupyter notebook and/or other visualization tool showing the averages for each class and university.

Plan:

1 - Import all data to a MySQL database.

2 - Studying the tables and construct a workable dataset and taking notes:

* Read the dataset and understand what it is talking about.
* We have common topics for each table pair: University, Class, Feedback
* However, we also have data coming from 2 different environments with no coherence among them.
* Find data commonalities between each schema relations.
* Clean and combine them by data source first, i.e. combine mongoDB collections together and the same for SQL, then explore commonalities between these 2 big entities if possible.
* Side notes:
  + classescollection Join studentfeedbackformscollection On UniversityID

*We have 19 unique university IDs in classescollection and   
13 studentfeedbackformscollection*

3 - Create a dataset that has the clean data and group data by University, by Class and conduct analysis on the table attributes.