**Python Assignment 1 (Due Date august 13 2023)**

**Description:**

For this assignment we will use Pandas in Python to analyze data from some CSV or other source. A suggested assignment is analyzing data extracted from the DVD rental database in a jupyter notebook.

We should use:

-Lambda functions

-np.where()

-np.select()

-For-loops

-matplotlib to visualize the data

-different merge types in Pandas.

-Extraction methods to get specific data from the dataframe based on specific criteria.

-transformations to create new variables/columns based on other ones.

**Suggested Assignment:**

See if you can answer the following questions about the DVD rental database in a Jupyter Notebook:

1. Create a time-series plot showing daily total revenue for three different countries with three countries in one plot (you will want matplotlib for this)
2. Carry out a sentiment analysis on movie descriptions adding another column to the data frame that provides a sentiment classification. This can be based on a “[rules based system](https://en.wikipedia.org/wiki/Rule-based_system#:~:text=Rule%2Dbased%20programming%20attempts%20to,which%20lists%20execution%20steps%20sequentially.)” that you coded OR inference from a LLM such as using the API for chatgpt.
3. Create a visual in the notebook that shows how much revenue is coming from different countries to the DVD store (you will want matplotlib).
4. Do an analysis to show which actors are bringing in the most revenue (top 10 ranked)
5. Create a new data-frame where you classify some different countries into some categories that make sense to you (8 countries min, 20 max) in this df there will be two columns, one column will give the country name, the other column will give the classification category.
6. Use an appropriate merge to merge that new data-frame from the last step with a dataframe from the dvd db, that gives the countries, so now countries are connected with their classification. (no need to add this table back to the postgresql db, just doing it in the notebook is fine)

You can use pgadmin4 and Python if need be.

It’s for the best that you upload your queries as well so I can see them. Be warned it’s easy to make mistakes on this assignment.

**Python Assignment 2 (Due tbd) (tentatively)**

This assignment will be all about web-scraping. For this assignment you will pick a website that you’re interested in to scrape data from, please see a list of good websites for this in our module resources document.

You should carry out this assignment with:

* A Jupyter Notebook
* Or a Google Colab notebook

The assignment should have:

* A decent readme for github
* A decent amount of text cells to explain what you’re doing on both colab or github
* A good number of code cells where you can run the code
* Output below the code cells to show the project is already done and show what you got from it, to see what charts, metrics, and analysis can be printed out.
* The Jupyter notebook should “tell a story” about your analysis
* The Jupyter notebook will have a short conclusion at the end re: what you found with your analysis.

What you will do:

* Mine data just as we have seen in class from your web-page of choice. See:
  + <https://colab.research.google.com/drive/1Pm90WBXrdxCuapNBkQ30FvShfVikkm0l#scrollTo=_AcwpnX8oJ24>
  + <https://colab.research.google.com/drive/1Jz22RZ0et8ra3zRiyu4Wt5zCn6zFwD18>
* Structure the data in a pandas-dataframe and apply functions and analysis, as well as appropriate visualizations
* Make the Jupyter notebook or colab public and available to see.