SI 206 Data-Oriented Programming

Project Name: Data-Oriented programming with websites/APIs and Visualization

Project Objective:

Demonstrate the ability to:

- Create a fully-working program without any scaffolding
- Create and modify tables in a SQLite Database
- Utilize websites/APIs (including researching possible methods)
- Utilize visualization software (including researching options)
- Document your code
- Work in teams of 2 to 3 people

Deliverables and Submission Process:

1. **You must submit a plan by April 4th** – Tell us the following information:
    a. What is your group's name?
    b. Who are you working with?
    c. What websites/APIs will you be gathering data from?
    d. What data will you collect from each website/API and store in a database?
    e. What data will you be calculating from the data in the database?
    f. What visualization tools will you be using?
    g. What graphs/charts will you be creating?

2. **You must have all of your data collection done by April 22nd.** Students who demo their fully completed project by April 22 (in lecture) will receive 20 points of extra credit. The presentation should be 4-5 minutes long. We may not have time for everyone who volunteers to present.

3. **You must submit a report on your project and a zipped copy of all of your code by April 29th by 11:59pm.** Absolutely no late assignments will be accepted.

4. **You must demo your project** on either April 22 during the last lecture, May 1st from 1:30pm to 3:30pm (during the final exam time) or May 2nd from 9am to 11am or 1pm to 3pm. Pick your first and second choice for presenting at https://forms.gle/nDHPJeZUB12hDp2w7

Background:

In this assignment you will be using the skills learned from the course to gather data from at least 2 websites/APIs, store that data in a database in multiple-tables, calculate data from the database, create at least 2 visualizations from what you calculated, and print out a report with the calculated data and the visualizations. If you have 3 people in your group you will need to work with at least 3 websites/APIs and create at least 3 visualizations.

PART 1 – Submit your plan (10 points)

1. Submit your plan for your final project on Canvas by 11:59 pm on April 4th. Tell us the following information. You can earn a max of 2 points for c – g below.
    a. What is your group's name?
    b. Who are you working with?
    c. What websites/APIs will you be gathering data from?
    d. What data will you collect from each website/API and store in a database?
    e. What data will you be calculating from the data in the database?
    f. What visualization tools will you be using?
    g. What graphs/charts will you be creating?

PART 2 – Gather the data and save it to a database (100 points)

- Access 2 website/APIs of your choice (e.g. Facebook, GitHub, Gmail, Yelp, etc). **You can't use Twitter**. If there are three people in your group, you must access 3 website/APIs.
- Access and store at least 100 items in your database from each website/API. The database must have at least two tables in it. If you have three people in your group the database must have at least three tables in it.
- You must limit how much data you gather from the website/API to 20 or less items at a time. If the website/API doesn't allow you to limit the number of responses, you can limit how many items you store in the database at a time. Some websites automatically limit the amount of data you can get for a time. It is also good practice to not overwhelm a website/API with too many requests. The data must be stored in a SQLite database. You must have at least as many tables in your database as the number of websites/APIs you are using.

PART 3 – Process the data (50 points)

- You must select some data from all the tables in your database. You must calculate something from that data. It could be a count for how many items occur on a particular day of the week. It could be an average of the number of items per day.
- Write out the calculated data to a file. The file format can be JSON, csv, or txt.

PART 4 – Visualize the data (50 points)

- If you have 2 people in your group you must create at least 2 visualizations of the calculated data. If you have 3 people you must create at least 3 visualizations. You are free to choose any visualization tool/software that you can create with Python.
- You will not earn the full 50 points if your visualizations don't go beyond the examples you were given in lecture. If you use an example from lecture, you should change something from the example you were given in lecture. You could change the colors of the bars in a bar chart for example.

PART 5 – Report (100 points)

In addition to your API activity results, you will be creating a report for your overall project. The report must include:

1. The goals for your project.

2. The goals that were achieved.
3. The problems that you faced.
4. Your file that contains the calculations from the data in the database.
5. The visualization that you created (i.e. screen shot or image file).
6. Instructions for running your code.
7. Documentation for each function that you wrote. This includes the input and output for each function.
8. You must also clearly document all resources you used. The documentation should be of the following form:

Date	**Issue Description**	**Location of Resource**	**Result (did it solve the issue?**

You will be graded on clarity, completeness, and presentation (no typos, neatly formatted, etc.)

BONUS A - Add additional website/API sources (Max 30 points)

- Earn up to 30 points for an additional API. You have to gather 100 items from the website/API and store it in the database. You must calculate something from the data in the database. You must write out the calculation in a file.

BONUS B - Add additional visualizations (Max 30 points)

- Earn up to 15 points for each additional visualization.

Useful Links

Github API (https://developer.github.com/v3/)
Gmail API (https://developers.google.com/gmail/api/)

You have to use python-specific packages. For example, you might have to google "Gmail API for Python".

Further Examples of Visualizations

In Gmail, what percentage of emails are sent from github on Monday, on Tuesday, etc.
In Facebook, a scatter plot with length of post vs. number of likes.
In Spotify, for your five favorite bands, compare how many songs of theirs are in your playlists.

Tips

Start early - This project involves learning and using a new website/API. Planning ahead is important, and make sure to give yourself enough time to ask questions if stuck.

Learn online - There are many tutorials and helpful information online. Since this is the first time you are encountering a given website/API, you will probably make use of them (and we encourage you to make use of them!). Remember, though, that you must document all the resources you use.

Debugging and looking for help - Unlike past homework and projects, here you get to choose your own website/APIs. This means that likely the website/APIs you choose will not have been seen by the instructors of the course. They will try to help in any way they can, but more often than not, you will have to debug your own code. Once again, online resources and tutorials are useful!

Have fun! - This project is broad on purpose. Choose sites that you are genuinely interested in and extract the information you want to see! Working on a project that is interesting is 100x better than working on a dull, boring project.