Next

Module 3: Introduction to Python (Part III)
3.10 Semester Project

Print view Index of pages

3.10 Semester Project

PROJECT SPECIFICATIONS:

Pick a data set that you and your group find interesting. (Example source: UC Irvine Machine Learning Repository. Feel free to select your data from any other source as appropriate.)

Perform data pre-processing, data cleaning, outlier removal, and so on to sanitize your data, if necessary.

Save your data in a .csv file (or other format as appropriate for your data se and project scenario).

Read in data to your program from the .csv file.

(*Optional* - do as appropriate) Process the data or perform any calculations or statistics on it before storing the data into a data frame (see next step).

Save the data into one or more data frames (or other structures as appropriate) (*review lecture notes/examples*).

Once you have stored your data, query your data to reveal interesting/useful information based on your project scenario.
Query your data using at least 4 different queries.

Capture the results of the queries appropriately (either write results to a file, or store into another data structure, or do something else with the results as appropriate based on your project scenario).

(*Do as appropriate*) Process the results, or submit additional queries to the obtained results (if results were saved to a file or another data structure).

Display final results in a presentable way (use tables and/or visualizations). You will present this information during the Module 12 Live Session.

Perform adequate testing (TDD and/or unit testing). Submit this in a separate .py file.

Submit a write-up (as a Python file (.py) or as a Jupyter Notebook) along wit your project (see details in the next section below), in your write-up be sure to display the results and describe what you have learned, as well as how these results can be used by others (mention the relevance/significance of the results you've obtained).

Group size: Groups of 4. No individuals or pairs. (Let me know if you're having difficulty finding/forming a group, I'll be happy to help put individuals together. Just email me!)

WRITE-UP DETAILS:

Be sure to include the following in your write-up:

Introduction: Describe your project scenario. Starting out, what did you hope to accomplish/learn?

The Data: Describe your data set and its significance. Where did you obtain this data set from? Why did you choose the data set that you did? Indicate i you carried out any preprocessing/data cleaning/outlier removal, and so or to sanitize your data.

Experimental Design: Describe briefly your process, starting from where you obtained your data all the way to means of obtaining results/output.

Beyond the original specifications: Highlight clearly what things you did tha went beyond the original specifications. That is, discuss what you implemented that would count toward the extra-credit portion of this project (see section below).

Results: Display and discuss the results. Describe what you have learned and mention the relevance/significance of the results you have obtained.

Testing: Describe what testing you did. Describe the unit tests that you wrote. Show a sample run of 1 or 2 of your tests (screen captures or copy-and-paste is fine).

Conclusions: Summarize your findings, explain how these results could be used by others (if applicable), and describe ways you could improve your program. You could describe ways you might like to expand the functionality of your program if given more time.

EXTRA CREDIT OPPORTUNITIES FOR THOSE STUDENTS WHO WANT A CHALLENGE:

Web scraping to obtain your data set instead of downloading a ready-made data set from a source (*see Homework: Python and Web Scraper* ☺).

Have some user interaction where you may obtain some more data from a user (if appropriate).

Have some user interaction where the user may choose the kinds of queries to perform on the data. Retrieve/display only the appropriate resul

Use advanced queries or manipulate the data in another way (other data manipulation methods, etc.) and display the results. If you choose to do this, mention in your write-up how this goes beyond the basic/general queries you initially used.

If you think of something else you would like to do and feel it could be considered as extra credit, please let me know. If I agree, I'll add it here!

SUBMISSION:

Pay close attention to these submission requirements:

In one Zip file submit: (1) your code as Python file(s) (.py) or Jupyter Notebook; (2) write-up (as PDF); (3) your presentation visuals; (4) any other supplemental files or resources as appropriate. [Note: don't submit your video here]

Each group will need to create **a video presentation** of their Semester Project. Each student needs to present a section, so be sure to stitch together your presentations into one final video presentation. Your video presentation should be no more than 20 minutes (in total).

Post your video to the Semester Project discussion board.

Each student will **evaluate two (2) other groups** by posting a response to the Discussion Board. Your instructor will determine which groups you will evaluate.

Attend the live session. Your group will need to present your project abstracts and highlights. Please plan to speak about your project for at least 4 minutes and no longer than 7-8 minutes. There will be a few minutes of Q&A after each presentation.

Conduct peer evaluations within your group. Your instructor will post the link to the form as an announcement. Please complete this portion after the project (including live session) is complete.

Submission location: under Assignments on Collab

RUBRIC

Description	Possible Points	Earned Points	Comments
Introduction: Describe the project scenario (can be brief)	5		
An appropriate data set was used Explanation of data set/domain/where data was obtained Size (large enough)	10		
What data structure was the data stored in?	5		
Performed any data pre-			

processing? Such as (but not limited to): data cleaning outlier removal ..etc to sanitize the data	10			
Data Analysis / Data Processing At least four (different)	20			
Testing: Describe any test-driven development and/or unit tests	10			
Results displayed appropriately for each test	10			
Explanation of Results & Conclusions How these results can be used by others (relevance/significance) Ways to improve / ways to expand / adding or removing functionality	20			
Presentation skills and video Video presentation posted to the discussion board is no longer than 20 minutes and all group members presented aspects of the project				

All group members spoke in the video and were present at the live session The live session presentation was between 4-8 minutes The presentation was of good quality, clear and easy to understand	10		
Total Points	100		
Extra Credit: [Max 3%: 1% per item, up to 2% for single item if substantial] Web-scraping to obtain data (instead of downloading) User-interaction to retrieve/display/analyze certain results -or- User-interaction to obtain some more data Advanced queries / manipulated the data in another way	Max 3%:1% per item, up to 2% for single item		
Final Grade			

NEXT

- Gateway