Group Project

Data Programming in Python
Business Analytics Graduate Program MSCI:6040
Fall 2019 (Cedar Rapids)

This is a group project. Please form a group of 4–5 people to complete the project. If you cannot find a group, please talk to the instructor.

Summary of Deliverables and Deadlines

What?	When?	Where?
Project proposal	6:00 PM on Mon Oct 7	ICON Assignment
Project deliverables (Jupyter notebook file(s), data file(s))	11:59 PM on Sun Nov 10	ICON Assignment
Presentation	Mon Nov 11	in class

The goal of the group project is simple: *do something impressive (or interesting or fun) with the skills and tools you have learned this semester*. Please follow these steps to achieve this goal:

1. Identify a topic that interests you for which you can fairly easily collect a reasonable amount of data.

   It could be a situation at work or something involving a hobby such as sports, or if you need some ideas, let the instructor know. If the data comes from work or another private source, please make sure that you are allowed to use the data. Ideally, you should also be allowed to let the instructor see the raw data and to summarize the data for your classmates, but compromises can be worked out with the instructor if necessary.

   Do not choose one of the topics that have already been covered in the projects of other courses such as Data programming in R. Just redoing your previous R project in Python is probably not a good idea.

2. Determine a big question that you would like to investigate about the situation using the data you plan to collect.

   The question should be interesting in the sense that: (i) it does not have an obvious answer; (ii) the data is necessary for answering the question; and (iii) the answer will provide insight into the situation. For example, suppose you were interested in analyzing stock market trends. You might ask, "How big of an effect did the 2016 election have on the entire U.S. stock market?"

3. Prepare a one-page project proposal (in PDF format) containing the following items:

- Your group members.
- A short description of the situation you will investigate.
- A short description of the data you will collect including its source, how you will collect it (e.g., querying your company database or downloading from a public website), and your specific rights to use and share the data (e.g., limited rights for private data versus unlimited rights for public data).
- The big question you would like to answer about the situation using the data along with three small specific questions.

Have one group member submit your project proposal by the deadline and wait for the instructor to give feedback and approve the proposal.

4. Once your project proposal has been approved, follow these generic steps to complete the core work of the project using Python and Jupyter Notebook:

   - Get the data and import it into a Jupyter notebook.
   - Transform, clean, and prepare the data for analysis.
   - Analyze the data to answer your questions fully. Your analysis should include viewing the data graphically.

   During these steps, you must also write and use at least one Python user-defined function to somehow help your work.

   > For example, if your data cleaning includes handling text data in a particularly tricky way, you could separate that part into a function leaving the main Python code easier to read. Or if you would like to have the option of preparing different graphics based on different options, you could write a function to enable this. Overall, the function can be used to assist your handling and analysis of the data, or it could just an "add on" function that is auxiliary to your handling and analysis of the data.

   Throughout the process, try to use a single Jupyter notebook containing all the commands necessary to reproduce your analysis, including importing the data and producing any graphics. Please comment all of your code. Also, include the pip install commands in the notebook if you are using any external Python packages, so the instructor can easily build the same environment.

   You may divide the entire notebook into two separate notebooks, one is to cover only the data collection part and the other to cover the rest. This is recommended especially for those groups who would find it hard to reproduce the data collection process, for example, collecting data using APIs for several hours or several days.

Finally, keep track of any challenges you face along the way and how you are able to overcome them.

5. When preparing a project report, continue to use the Jupyter notebook(s) you have been working on for analysis discussing your entire project and its outcomes by creating markdown cells wherever needed. There is no limit on the length of the notebook, but try to make it as compact as possible. Your audience is the instructor and your classmates, and please work hard to make the report easy to read and digest.

   Please cover at least the following topics (preferably in marked sections in this order):

   - The situation, its data, and the questions to be answered.
   - Your analysis and the answers to the questions.
   - A summary of getting the data and preparing it for analysis.
   - A summary of the various obstacles you faced during the project and how you overcame them.
   - Documentation of the Python functions written to assist your analysis.

6. Install the latest RISE package on your computer, which is a Jupyter notebook extension for slideshow (refer to https://rise.readthedocs.io/en/maint-5.5/). Open the notebook(s) and set the types of the cells you would like to present to Slide or Fragment, so you can present them in the Slideshow mode.

7. Create a clean version of Jupyter notebook(s) by clicking Kernel on the menu bar and then Restart & Run All. Make sure that all the code cells in the notebook(s) have been run with no errors (warnings are okay). Double check if they display all the information in a desired manner.

   Have one group member submit by the deadline a single compressed Zip file containing the raw Jupyter notebook file(s) (.ipynb) and any data files used. Note that you do not have to make a separate document as a report. Everything should be contained in the Jupyter notebook(s).

8. Prepare a group presentation to be shared with the rest of the class on the date shown above. Note that you do not have to make presentation slides for your presentation. You will be using the RISE slideshow for presentation. During the 10 minute presentation, all group members should present for an equal amount of time.

Here are the evaluation criteria for the group project:

- Amount and depth of work
- Code management
- Answering questions
- Presentation