Bachelor of Data Science

Module
PGR107 Python Programming

Due date for submission
Autumn 2023 (see Wiseflow)

Module leader and e-mail
Huamin Ren | Huamin.Ren@kristiania.no

Teacher and e-mail
Huamin Ren | Huamin.Ren@kristiania.no

Learning outcomes

After successfully completing the course the student:

Knowledge

- Understand Python built-in data structures, functions and files, be able to use them for data manipulation
- Understand advanced computational functionality for larger datasets by using add-on libraries like Pandas and Numpy
- Understand the pipeline of loading data, understanding data, data preparation and transformation and data visualization.

Skills

- Be able to manipulate, process, clean in Python
- Be able to choose from Numpy and Pandas to perform large data loading and transformations
- Be able to draw line plots, scatter plots, visualize errors with a predefined or manually defined error function; know how to customize legends and styles; be able to generate multiple subplots
- For a given structured data file or user chosen data file, be capable of loading data with proper data structure with application needs and efficiency considered, as well as performing data manipulation and visualization

General competence

The student ...

- is able to demonstrate basic Python coding skills for data manipulation, data transformation, cleaning and visualization
- is able to critically assess the benefits of using Python data analytics programming
- is able to choose proper add-on libraries to assist large data loading and transformation
- is able to illustrate each processing, transformation step, and is able to explain the visualization results

Assignment specification

Group Size: 1-5 students
Code implementation with comments (note: use %% md for block-wise comments), should be saved and submitted as .ipynb (share directly on wise flow).
Referencing: Any acceptable academic style.

Please address the following questions in your submission.

1. Explore the data file (named as data_PGR107.csv), and then address the following questions. Please be noted you should add comments to the code in order to provide necessary explanations of your processing, or as a way of illustrating your discovery.
   a. Load the data
   b. Explore the data, and perform proper visualization to support your data exploration
   c. Perform data cleaning or transformation if necessary (note: if you perform any additional processing, please comment to explain the reason)
   d. Illustrate your discovery or conclusion after data exploration, and use visualization or analysis or support your results
2. Here is a list of hospitals located in Norway, see link here:

   https://en.wikipedia.org/wiki/List_of_hospitals_in_Norway. Perform the following task:

   a. Create a Series from Pandas, where information on Oslo University Hospital should be included, i.e., hospital name, major location, main campuses, description and/or other relevant information

   b. Choose 16 hospitals from the above list, and then create 16 Series, each of which includes information about that hospital

   c. Create a DataFrame from the Series generated from step b.

   d. Generate a new column, named 'Regional Health Authority'. For each row in DataFrame created from step c, assign a corresponding value for 'Regional Health Authority'.

   For example, for 'Oslo University Hospital', the value of 'Regional Health Authority' should be filled with 'Southern and Eastern Norway Regional Health Authority'.
3. A use case exploration.

   Nord Pool consists of the commercial power exchange function, Nord Pool, and the Market Coupling Operator function, Nord Pool European Market Coupling Operator. Nord Pool is Europe's leading power market and **offers trading, clearing, settlement and associated services in both day-ahead and intraday markets across 16 European countries**. Their product includes power price produced within their markets every hour, every day

   (https://www.nordpoolgroup.com/en/About-us/#:~:text=Nord%20Pool%20is%20Europe's%20leading,markets%20every%20hour%2C%20every%20day).

   3.1 Look into the webpage, choose several cities in Norway and several other European countries, and then download weekly data in October 2022: https://www.nordpoolgroup.com/en/Market-data1/Dayahead/Area-Prices/

   3.2 Look into the same webpage, choose a day in November 2022 and download the hourly data from the several countries and cities mentioned above

3.3 Write python code to properly load these data, use any relevant data handling methods (for example, data transformation, missing data handling, filling NaN values and so on) to convert the data to Series, DataFrames. Then perform analysis and visualization to compare the price changes between different countries in EU and different cities in Norway.

Assignment criteria*

Grade	Learning Outcome 1: Knowledge	Learning Outcome 2: Skills	Learning Outcome 3: Competence
A Excellent	Excellent and comprehensive understanding of concepts	Demonstrates excellent analytical, technical and writing skills	Outstanding degree of judgment and independent critical thinking
B Very good	Very good understanding of concepts	Demonstrates very good analytical, technical and writing skills	Sound degree of judgment and independent critical thinking
C Good	Good understanding of theory in most important areas	Demonstrates good analytical, technical and writing skills	Reasonable degree of judgment and independent critical thinking
D Satisfactory	Satisfactory understanding of theory, but with significant shortcomings	Demonstrates limited analytical, technical and writing skills	Limited degree of judgment and independent critical thinking
E Sufficient	Meets the minimum understanding of concepts	Demonstrates sufficient analytical, technical and writing skills	Very limited degree of judgment and independent critical thinking
F Fail	Fail to meet the minimum academic criteria.	No demonstration of analytical, technical and writing skills	Absence of judgment and independent critical thinking

*Adapted from The Norwegian Association of Higher Education Institutions

The assignment is worth 100 % of the grade of the course.
Question 1 is 40% of the overall grade.
Question 2 is 40% of the overall grade.
Question 3 is 20% of the overall grade.