

Assignment 1 Day 10 | 7th January 2021

For any doubts regarding the assignment, ask questions in the LetsUpgrade Advance Data Science Group in the Community.

Submit Assignments by **17th January 2021 11.59 PM.**

Assignment Submit Form: <https://forms.gle/2Pd6EnWbwRtxqaV17>

Submit assignments in Appropriate Dropdowns.

Questions 1:

Pandas Assignment

How to import pandas and check the version?

How to convert the index of a series into a column of a dataframe?

How to combine many series to form a dataframe?

How to assign a name to the series' index?

How to get the minimum, 25th percentile, median, 75th, and max of a numeric series?

What is qcut() in Pandas?

How to find the positions of numbers that are multiples of 3 from a series?

How to stack two series vertically and horizontally ?

How to get the day of month, week number, day of year and day of week from a series of date strings?

How to compute the euclidean distance between two series?

How to use apply function on existing columns?

Questions 2 :

Pandas Assignment

With respect to Earthquakes.csv "http://www.ldeo.columbia.edu/~rpa/usgs_earthquakes_2014.csv"

Use Pandas' read_csv function directly on this url to open it as a DataFrame

Re-read the data in such a way that all date columns are identified as dates and the earthquake id is used as the index

Use describe to get the basic statistics of all the columns.

Use sort_values to get the top 20 earthquakes by magnitude.

Find the 10 states / countries with the highest number of earthquakes.

Find the top 10 states / countries where the strongest and weakest earthquakes occurred.

Questions 3 :

Matplotlib Assignment

With respect to iris dataset,

Describing the data

Plot the frequency of the three species in the Iris Dataset.

Plot the Pie of the three species in the Iris Dataset.

Combining Box and Strip Plots for Species vs sepal length.

find out the correlation between different features in the dataset.



[Advanced Data Science] | [January and 2021]

Plot Venn diagram for sepal length vs sepal width.

Questions 4 :

Pandas+Matplotlib Assignment

Download the dataset from:

<https://www.kaggle.com/worldbank/world-development-indicators>

How many UNIQUE country names are there ?

Are there the same number of country codes ?

Are there many indicators or few ?

How many years of data do we have ?

What's the range of years?

ScatterPlot for comparing GDP against CO2 emissions (per capita)

FAQs

Q. How to upload a jupyter notebook as a part of an assignment?

- A. 1. Click the "File" option in a notebook
2. Go to "Download As" -> "Notebook(.ipynb)"
3. Upload the downloaded .ipynb file to GitHub and share the link in the google form.

Q. When do I submit the Assignments and how?

- A. The assignments for the week should be submitted by the weekend i.e. Sunday at 11:59 PM IST.
You can use Jupyter Notebook or python files or even Google Colab to Solve your Assignments

Q. Where do I get class links for the next session?

- A. All sessions will be Live on the Learning Management System. It will be also updated in the Community Group in the pinned post.
- B.

Q. I have some doubt, who do I ask?

- A. Post your Queries on the community, someone will help you out.

Q. Can we submit multiple .py or .ipynb assignment solution files for each question separately?

- A. Solve all assignments for a day in a single notebook. Make sure you are submitting a single file.