**Capstone Project: Data Science PGC**

**Project Overview:**
You are working as a Data Science Consultant for a data-driven analytics firm that helps organizations extract valuable insights from publicly available online data. Your task is to design and execute a complete end-to-end data science project using data scraped from the web. The project is open-ended, you can take data from various domains like ecommerce, movies, sports, books, crypto prices, IMDB, etc.

Your objective is to collect, clean, analyze, model, and visualize real-world data to derive meaningful insights and communicate them through an interactive dashboard. The goal is to demonstrate your ability to handle an unstructured, real-world data science workflow from raw data acquisition to actionable business intelligence.

The leadership is particularly interested in understanding:

- How effectively can real-world data from the web be transformed into structured, decision-ready insights?
- What trends, patterns, or anomalies can be uncovered through exploratory data analysis and visualization?
- Can predictive or classification models be built to forecast outcomes, identify key drivers, or segment users/items effectively?
- How can the findings be communicated through an intuitive, interactive **dashboard** that supports data-driven decision-making?

**Workflow/Steps to be followed:**

**Reference document: workflow template**

1. **Web Scraping Script (100 Marks)**– A reproducible Python script or notebook that extracts data from a chosen website or API.
2. **Data Cleaning & Preprocessing (50 Marks)** – Handling missing values, duplicates, inconsistent formats, and feature engineering.
3. **Exploratory Data Analysis & Visualization (50 Marks)** – Graphical and statistical summaries to uncover insights.

4. **Model Building & Evaluation (100 Marks)** – Develop a predictive model (clustering or regression/classification model)

5. **Interactive Dashboard (50 Marks)** – A dashboard (using Power BI, Tableau, or Python libraries like Streamlit/Plotly Dash) presenting key insights, trends, and model outcomes. Include at least two KPIs, trend charts, AI and custom visuals and important reporting features like RLS, sync slicer, bookmark, drill through, etc.

6. **Final Report / Presentation (50 Marks)**– Summarizing methodology, insights, model results, and recommendations.

This project encourages creativity and problem-solving, testing your ability to independently manage a real-world data project from **data acquisition to insight communication**.

## Web Scraping Guideline for Data Extraction

- **Extract meaningful data** (like product info, job listings, books, news, etc.) with clear columns.
- **Use only open-source / publicly accessible websites** that do not require login and allow scraping.
- **Collect at least 1,000 clean records** by scraping multiple pages or sections of the site.
- **Ensure each record has multiple useful fields** (e.g., title, price, rating, date, link).
- **The final dataset must be clean, consistent, and ready for analysis**

## Deliverables

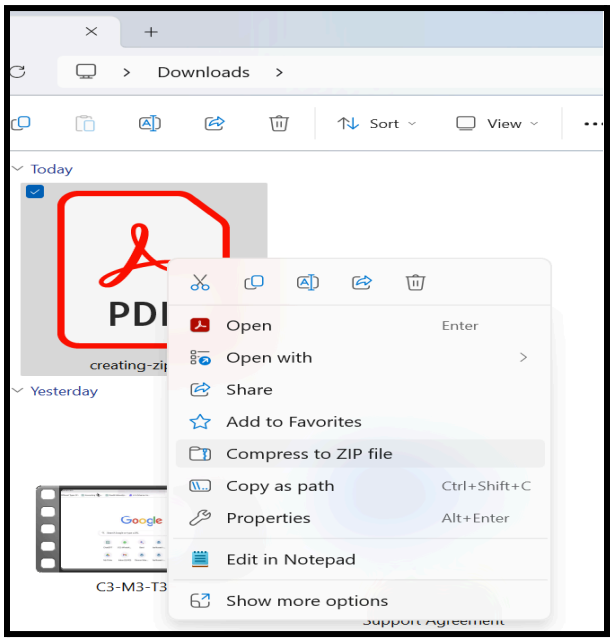
- Two code files are required, one should be for data extraction/acquisition with the help of a webscraping library/tool of your choice, and the other for data cleaning/preprocessing and model building.
- Clean .csv file extracted by you.
- A minimum of 5 minute long video presentation is required explaining the problem statement(which you will curate yourself as the data is subjective), approach, finding/insights and evaluation
- Final Report/Presentation converted into PDF with the business problem that you're solving, and summarizing methodology, insights, model results, and recommendations.

**Submission Guidelines:**

- Save the code files, and summary in a PDF and then convert it into a zipped (.zip) folder. **(Please note, the drivelink for the video created should also be added in the PDF itself.)**
- Upload the zipped folder on your respective dashboard.
- Failure to comply with submission guidelines will result in no grading/0 marks.

**How to ZIP a PDF file:**

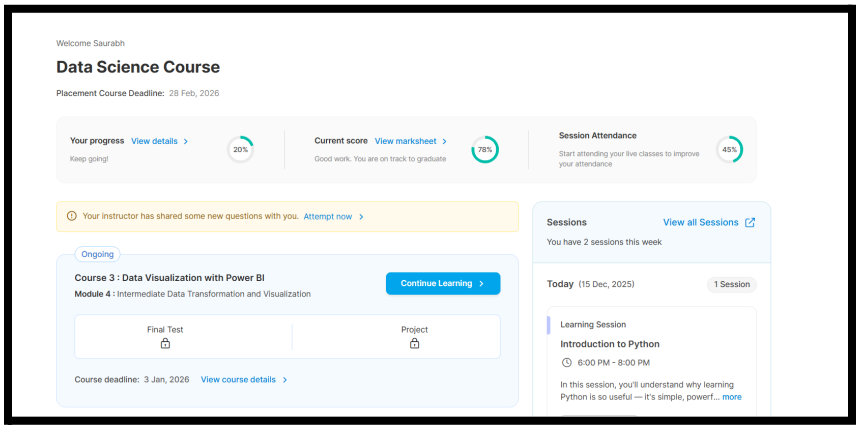
- Put all of the documents you want to compress (or just one) into a new folder.
- Right click on that folder.




- Select the “Compress to ZIP file” option and then click “Compressed (Zipped) folder.”
- A new .ZIP file will be created that contains your document(s).

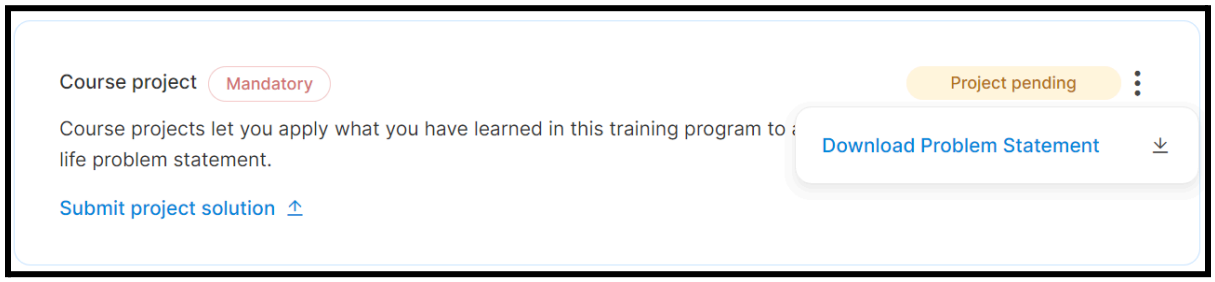
**In order to submit the projects please follow the following steps:**

1. Click on “ Your progress View details” after logging into your dashboard.



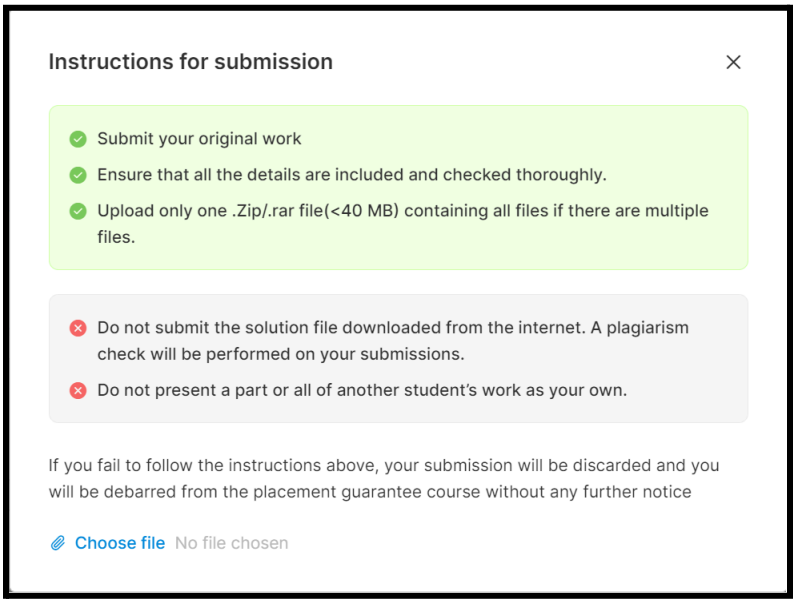

2. Next, click on the tab for Deep Learning. Then, scroll down to find the "**Course Project**" section.

3. Now, you will be able to see the option to ‘Submit project solution’.




4. Please follow the guidelines (screenshot is shared below) provided in the project to ensure correct submissions. Then, click on "**Upload Project Solution**" to submit

your work.


Instructions for submission
Submit your original work
Ensure that all the details are included and checked thoroughly.
Upload only one .Zip/.rar file(<40 MB) containing all files if there are multiple files.
Do not submit the solution file downloaded from the internet. A plagiarism check will be performed on your submissions.
Do not present a part or all of another student's work as your own.
If you fail to follow the instructions above, your submission will be discarded and you will be debarred from the placement guarantee course without any further notice
Choose file
No file chosen