

Project & Report Assignment

How-To Guide

This assignment represents 100% of the overall course grade.

Instructions

Develop a Python project to analyse real world scenarios and generate valuable insights by visualising information. The project aims to analyse data from different data sources, manipulate information and visualise to generate insights.

You can use any open-source dataset available online for analytics. Each bullet point for every learning outcome is a milestone to be achieved.

The project should be submitted on the Learn Site under the Assessments section. You will need to include two files, as described below.

There are three deliverables contained in two files:

1. Project ZIP
 - Create a ZIP file of your entire Python project along with all the code and data files and upload as part of your submission
 - The project should cover all milestones in each learning outcome to gain full marks (see below)
2. Project Report
 - A document containing between 1,500 and 2,000 words
 - Please use the template provided (see Assessments section to download)
 - The report describes your process, dataset, different sources, graphs and insights
 - Justify the use of each learning outcome concept, for example: Why did you use list over dictionary?
 - Upload the document file along with the ZIP file
3. GitHub repository URL
 - Create a new repository on GitHub as [UCDPA_yourname]
 - Keep committing to the repository
 - Remember to include the URL of your repository at the beginning of your Project Report document

The goal of the assignment is to demonstrate how you are thinking about putting course concepts and learning into practice to demonstrate the course learning outcomes:

1. Store and manipulate data in Python data structures, and understand key concepts of Boolean logic, control flow, and loops in Python
2. Visualise real data with Matplotlib's functions and get acquainted with data structures such as the dictionary and the pandas DataFrame
3. Understand various ways to import data into Python: from flat file such as .txt and

- .csv; from files native to other software such as Excel spreadsheets, Stata, SAS, and MATLAB files; and from relational databases such as SQLite and PostgreSQL
- 4. Create visualisations and generate insights for different kinds of datasets and be able to customise, automate, and share these visualisations using Matplotlib and Seaborn
- 5. Manipulate multiple DataFrames by combining, organizing, joining, and reshaping them using pandas

How You Will Be Assessed

The following list describes the areas being assessed, for a total of 150 points (points awarded are indicated in brackets).

1. Real-world scenario
 - The project should use a real-world dataset and include a reference of their source in the report (10)
2. Importing data
 - Your project should make use of one or more of the following: Relational database, API or web scraping (10)
 - Import a CSV file into a Pandas DataFrame (10)
3. Analysing data
 - Your project should include sorting, indexing, and grouping (10)
 - Replace missing values or drop duplicates (10)
 - Slicing, loc or iloc (10)
 - Looping, iterrows (10)
 - Merge DataFrames (10)
4. Python
 - Define a custom function to create reusable code (10)
 - NumPy (10)
 - Dictionary or Lists (10)
5. Visualise
 - Seaborn, Matplotlib (20)
6. Generate valuable insights
 - 5 insights from the visualisation (20)

The final grade is indicated by a scale as follows:

No attempt	Clear fail	Fail	Pass	Merit	Distinction
0 to 15	16 to 38	39 to 74	75 to 96	97 to 119	120 to 150