Combine the Assignment how to with the project report draft; in order to hit all measurables;

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.

\*using python already; make nice graphs

The project aims to analyse data from different data sources, manipulate information and visualise to generate insights.

\*Get data from anywhere; API nice but not necessary

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.

\*Exact URL: https://learn.ucdpa.ie/mod/assign/view.php?id=18313

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 \*I have the github repo fully stocked..
   * 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\*Literally the work of less than an hour.
   * Please use the template provided (see Assessments section to download) template is below;
   * 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? \*Why would one use list instead of dictionary? I must argue my case for all of these issues.
   * Upload the document file along with the ZIP file
3. GitHub repository URL
   * Create a new repository on GitHub as [UCDPA\_gavinhoran] \*done
   * https://github.com/gwavin/UCDPA\_gavinhoran/blob/main/README.md#ucdpa\_gavinhoran
   * Keep committing to the repository\*will do
   * 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, \*use one example of Boolean logic used;

control flow, \*one example of control flow;

and loops in Python \*one example of a loop; use a for loop

1. Visualise real data with Matplotlib’s functions and get acquainted with data structures such as the dictionary and the pandas DataFrame

\*Make certain the lines import pandas as pd

And import matplotlib as plt are included

Call words = pandas.read\_csv at least once

Plt.show() at least once.

1. Understand **various** ways to import data into Python: \*demonstrate all; try to get a date for all of them; so that they can be compared to one another; find one of each type from bitcoin; the metric could be to compare quoted prices across exchanges; see where they derive their prices from, and note if arbitrage could be engaged in; solely by selling bitcoin in different currencies;
2. 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 SQLLite and PostgreSQL (\*one of these two)

1. Create visualisations and generate insights for different kinds of datasets and be able to customise, automate, and share these visualisations using Matplotlib and Seaborn

\*Print these in the way the semmelweiss is printed;

1. Manipulate multiple DataFrames by combining, organizing, joining, and reshaping them using pandas\*merged two; removed all but Ireland;

# 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) \*bitcoin of various types; maternity data set from somewhere; nhs? done
   * Task 1; download ten files; csvs;
   * I made use of the Kaggle API; I created and downloaded the new API token which one would use to authenticate my access, this was saved as Kaggle.json in my .kaggle folder. I actually had to expire and refresh this because it failed to grant me access at one point. I used this to download several sets of data. One being the cdc/national-health-and-nutrition-examination-survey.
   * This contained a range of files. 10
2. Importing data
   * Your project should make use of one or more of the following: Relational database, API or web scraping (10) \*one line, easy; done 10
   * Import a CSV file into a Pandas DataFrame (10) \*one line, easy; done 10
3. Analysing data
   * Your project should include sorting(3), indexing(4), and grouping(3) (total 10)
   * \*need to do this; change the date to the index; group by country sort; by highest level of tests; idea for indexing; choose an arbitrary value for sorting; index at that; sort first
   * Perhaps see if you can do a pivot table
   * Replace missing values or drop duplicates (10) \*test for duplicates and drop them
   * \*replace missing values in the sheet I have there.
   * Slicing, loc or iloc (10) \*revise the iloc chapter, I strolled through this.
   * Iloc is the location by the index location
   * Looping, iterrows (10)
   * Merge DataFrames (10) \*have done this.
4. Python
   * Define a custom function to create reusable code (10) \*have done this
   * NumPy (10) what could NumPy do for me?
   * Dictionary or Lists (10)
5. Visualise
   * Seaborn, MatPlotlib (20) \*have done this
6. Generate valuable insights
   * 5 insights from the visualisation (20)
   * \* I can shit these out;
   * he 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 |

Project Report

GitHub URL

https://github.com/gwavin/UCDPA\_gavinhoran

Abstract

(Short overview of the entire project and features)

I have a casual interest in the price of Bitcoin. I am not alone in this. I would also not be alone if I had an all-consuming obsession with it. I decided to have a look for various datasets covering the historical price of bitcoin. Each location that offers these lists of prices is not doing so for the good of their health. Bitcoin is unusual in that people can freely trade in them, and so the prices offered in the various exchanges are very influenced by the market. But, there is such a thing as price ignorance, and I suspect that this is somethingthat would be the case for most users.

I went looking for a diversity of prices for the price of bitcoin over the last five years. These should be very similar to one another, but I expect that there will be differences.

The goal of this is to demonstrate competence in the following; it will be dressed up to look like I’m trying to share some useful information.

\*Goal one; download several sets of bitcoin datasets, generate several typs of graph with these.

Try to lay them over one another.

Introduction

(Explain why you chose this project use case)

Case 1: I am interested in Bitcoin.

I want to compare the Bitcoin price over time with the Ethereum price over time. I would also be interested in drawing in some other data from other sources, with the goal of finding a lagging or leading indicator.

These could be used to design an investment strategy.

Case 2: I am interested in Healthcare Datasets, and Healthcare analytics, with emphasis on maternity and neonatal information. I wished to see if I could locate an insight that might be derived; do I think that there is a scenario or a piece of information that might be associated with a given trait that might make it more likely that a negative outcome or medical event might be associated with them, eg. PPH;

Dataset

(Provide a description of your dataset and source. Also justify why you chose this source)

1. Bitcoin dataset: chosen because it was reputable and can be updated and also I wished to be able to demonstrate that I can use an API.
2. Relevant/trusted, also offered an API; I wished to demo this twice.

Implementation Process

(Describe your entire process in detail)

Results

(Include the charts and describe them)

Insights

(Point out at least 5 insights in bullet points)

1 bitcoin has been increasing since X, bitcoin prices rise on Fridays and go down on Mondays.

2 Ethereum is tightly bound to bitcoin, and lags, but not in a useful way. (demo by using correlation)

Racial group A has a higher chance of outcome C; this tallies with information observed from info source X,

References

(Include any references if required)