Python project:

Certificate in Introductory

Data Analytics

Assignment: Project + Report

Instructions for completion:

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

A student 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 student portal using Turnitin. The instructions are present on your student portal.

There are three deliverables:

1. Project Zip

● Create a zip file of your entire Python project along with all the code and data files and upload on a Turnitin link.

● The project should cover all milestones in each learning outcome to gain full marks.

2. Project Report

● 1500-2000 Words

● Describing your process, dataset, different sources, graphs and insights.

● Justify the use of each learning outcome concept for eg.. Why did you use list over dictionary?

● Include GitHub repository URL in the beginning.

● Submit on another Turnitin link

3. GitHub repository URL

● Create a new repository on GitHub with name [UCDPA\_yourname]

● Keep committing to your repository

● Include the URL of your repository in the Report

.

Project Report: 25% Marks

Project code: 75% Marks

Title: Data Analysis and Visualization on real-world Dataset

Submission date: 26th April 2021

% of overall marks: 100%

Pass mark for individual component: 50% (50 marks)

Required Length of assignment: Overall word count not to exceed 2,000 words.

Learning Outcomes assessed: 1) Real World Scenario

• Project should use a real world dataset and include a reference of their source in the report [1]

2) Importing data

• Your project should make use of one or more of the following, [1]

• Relational Database or API or Web Scraping

• Import a CSV file into a Pandas DataFrame. [1]

3) Analyzing data

• Your project should include sorting, indexing, grouping. [1]

• Replace missing values or dropping duplicates. [1]

• Slicing, loc or iloc. [1]

• Looping, iterrows [1]

• Merge dataframes [1]

4) Python

• Define a custom function to create reusable code. [1]

• Numpy. [1]

• Dictionary or Lists. [1]

5) Visualize

• Seaborn, Matplotlib [2]

6) Generate Valuable Insights

• 5 insights from the visualization. [2]

See overleaf for Marking Criteria

I downloaded "Attractions.csv" from https://data.gov.ie/dataset/attractions and saved it locally to demonstrate

reading a csv file into a dataframe.

For the attivities and accommodation datasets i made a call to the APIs https://failteireland.azure-api.net/opendata-api/v1/accommodation

https://failteireland.azure-api.net/opendata-api/v1/activities

Using SQLite Developer I created a demo database "sqlliteDB\_ucdproj.db with a table COUNTY\_PROVINCE\_LINK to demonstrate importing data from a Relational Database - see "data" folder.