Independent Project

Due date: 18th January 2023 by 4pm

Instructions

You will have the opportunity to test your Python skills by developing a ”Crime Analysis Interface”. The purpose of this interface is to retrieve a information on crimes by neighbourhood during a specified year and month. The interface will also produce statistical summaries, data visualisation, and timeseries breakpoints on crime locations by neighbourhood searched.

Answer all questions, paying attention to the instructions on file and function names. In addition to the requested code files, some questions ask for text answers such as short explanations. Include these, indicating clearly which answers are relevant to which questions, in a single file answers.txt or answers.docx. If you would like to create a more professional report you are welcome to include your graphs in this file as well. [illegible] and hand in that zip file, using your candidat[illegible]`0012_Project.zip`).

You are allowed to w[illegible] the same code but your written answers m[illegible]nning of the project who you are working wi[illegible]ou must indicate in your report who you wo[illegible]o ensure anonymity.

Although in pairs it [illegible] by one person, it is important that both the[illegible]ir own.

Plagiarism

If your code for a questi[illegible]n the Internet or to the code of your own pre[illegible]r student not in your group, we will consider [illegible]mportant differences between the previous as[illegible]o easily recognise if code is copied by previo[illegible]

1 Downloadin[illegible]

1.1 Downloading[illegible]

We will analyse the crimes recorded over two years in the boroughs of London. The website implements APIs to download the different datasets automatically. The URL for this dataset is `https://data.police.uk/api/outcomes-at-location?date=2022-01&lat=51.5098656&lng=-0.118092`.

Once you have saved all the data downloaded in a variable, print the number of crimes in the dataset and add as a comment below a description of the structure of the data. If you need further print (or pprint) statements to understand the structure of the dataset, use them and comment the code in an appropriate manner, but the print statement containing the number of crimes in the dataset needs to be the last one.

Create a script named downdata.py. Your first task is to download the data at the link above and print the keys of the json file including the type of the paired value. You will then save the data in a pickle file called crimedata.p.

1.2 API requests based on location [15 marks]

The records saved in the previous exercise use the latitude and the longitude identifying London. Use the locations.csv file from Moodle to import the data of the different locations around the UK. Create a function in the script userlocation.py that asks a user to input up to 5 locations and create a nested dictionary containing the location as a key and the latitude and longitude as items of the nested dictionary related to that location. The nested dictionary will be the output of your function and you will use it to submit the requests using the API, one request per each location by inserting the correct values of latitude and longitude. Print the nested dictionary and the number of crimes per each location at the end of the script.

2 Data Parsing [10 marks]

This section of the project consists of exploring the data and retrieving interesting information. You will use the pickle file you saved as crimedata.p

2.1 Import data [3 marks]

Create a script called dataparser.py and start by reading your crimedata.p pickle file.

2.2 Select the type of crime [7 marks]

The data contains different types of crime. Create a list with all the different type of crimes (e.g. violent-crime, theft-from-the-person) that have been committed and another list with the location subtypes (e.g. road, nightclubs, etc) according to the dataset.

Print the two lists.

Make sure to take t [illegible] input to lowercase, string, re-asking user in [illegible] will serve as search parameters in the follow [illegible]

3 Graphical r [illegible]

3.1 Bar charts [1 [illegible]

Use the lines of code fr [illegible] file and create two bar graphs.

The first one will ha [illegible] the crimes as value for each bar.

The second bar plot [illegible] ubtypes occurrences for the violent-crime da [illegible]

Remember that the [illegible] in a research work, therefore it is important [illegible]

First questions to be [illegible]

1. Which are the m [illegible] his result something you have found in other [illegible]

2. Which are the lo [illegible] etation you can give to justify this result?

3.2 Line graph [15 marks]

In the same script (use comments to separate the tasks), create the following line graph using the months and years as ticks on the X axis. While the API string specifies the date of the last case update, the data contains the crime date as well. Plot four lines:

- One line for all the crimes in the dataset
- One line for the theft from person type of crime
- One line for the other theft
- One line for the violent crimes

Third question to be answered in the answers file:

3. Which differences do you notice among the lines in the plot?

4 Folium: plotting on a map [20 marks]

The data we are analysing has precise locations around the city that can be visualised on a map. To do this, we will use the library called folium. The instructions on how to install the library are at the following link: `https://pypi.org/project/folium/`.

The documentation explaining the library is available at `https://python-visualization.github.io/folium/`. There are several commands in the library, we will indicate which ones we suggest you to use in the two subsections of this part. If you find another command more useful for how you envisage the graph, there is no problem at all in using it.

4.1 A point for each crime [10 marks]

Write a mappoint.py script for this task. Import the data saved in crimedata.p and use the code written previously to load and parse the dataset in the way you prefer.

Use folium.Map to create the map of London and folium.CircleMarker to indicate the point where each crime has been committed according to the latitude and longitude values in the dataset.

Graph and save the resulting map.

4.2 Cluster crimes on the map [10 marks]

In the same script, create another graph where the crimes are clustered in groups. Use folium.Map to
load the map of London
o create the clusters
according to the data yo
et using the accurate
values of latitude and lo

Graph and save the

Fourth question to b

4. Are there hotspot
he graphs and which
graph is more effective i

5 Handing in

Remember to submit al
Candidate Number
(`ABCD1_SECU0012_Proj`

- all the .py scripts
- the pickle and csv
- the answers.docx (
s to the questions in
the last sections
- if working in a pai
ur partner

You may have noticed that the sections only arrive to 85 points in total. The remaining 15 points are assigned based on the answers that are in the answers.docx file.

Remember to comment your files and use functions where you feel it is appropriate.