Cardiff School of Technologies

Assessment Brief

Module Code

DAS7003

Module Title

Geospatial Analysis

Academic Year

24-25

Semester

2

Module Leader email

ashahaab@cardiffmet.ac.uk

Content

Assessment Details

Assessment title	Abr.	Weighting
Applying Geospatial Techniques to real world dataset and presenting the findings.	PRAC1	80%

Pass marks are 40% for undergraduate work and 50% for postgraduate work unless stated otherwise.

Task/assessment brief:

Visualisation and Analytics of Geospatial Data

Task 2.1: Application of Python-based geospatial visualisation tool (e.g., GeoPandas) on a real- world dataset

According to UNHCR, 122.6 million people are forcibly displaced worldwide, as of mid-2024. This is a result of persecution, conflict, violence, human rights violations or events seriously disturbing public order.
This task requires you to use the Refugee population dataset. The source of the dataset is the World Bank. The dataset contains information about forcibly displaced populations spanning more than 70 years of statistical activities. It covers displaced populations such as refugees, asylum-seekers and internally displaced people, including their demographics. Stateless people are also included, most of who have never been displaced.
The dataset is available on Moodle under the Assessment folder. You can also access all the data related to this assignment here. Use a Python-based visualisation tool (such as GeoPandas) to plot a set of choropleth maps representing the Refugee population of the world for the years 2000 and 2023 respectively.
You must perform data cleaning and EDA before plotting the choropleth maps.

Provide a short discussion on your interpretation of the two plots.

The solution should be in a Jupyter notebook (.ipynb), wherein all the functions, libraries and coding steps should be explained in a lucid manner. Major steps for generating the choropleths would typically involve, importing the datasets using appropriate Python libraries, data cleaning, geospatial operations, and plotting. The Jupyter Notebook should be able to reproduce the choropleth maps without any error.

Task 2.2: Analysis of geospatial datasets

In this task, you are required to use one more dataset (conflict dataset), the UCDP Georeferenced event dataset in addition to the Refugee Population dataset used in the previous task. This UCDP's dataset covers individual events of organised violence (phenomena of lethal violence occurring at a given time and place). These events are sufficiently fine-grained to be geo-coded down to the level of individual villages, with temporal durations disaggregated to single, individual days.

Both datasets are available on Moodle under the Assessment folder. All the choropleths and plots must be generated using appropriate Python-based tools.

For further information on the UCDP datasets, please refer to the following:
   • Davies, Shawn, Garoun Engström, Therese Pettersson & Magnus Öberg (2024). Organized violence 1989-2023, and the prevalence of organized crime groups. Journal of Peace Research 61(4).

• Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50(4).

Before proceeding to attempting the following tasks, you must perform EDA on the datasets and minimise / aggregate them, so that you can answer the following tasks in a lucid manner. You will need to merge/join both datasets. There are multiple ways it can be done, it is at your desecration if you use pandas or geopandas capabilities to achieve the set task.

Task 2.2.1

For the year 2020, generate choropleth maps of the refugee population for the countries having more than 5000 conflicts since 1993. Very briefly interpret the generated map.

Task 2.2.2
For the year 2020, generate choropleth maps of the conflicts from conflict countries for only the countries from region Middle East, having total refugee population more than 50,000. Very briefly interpret the generated map.

Task 2.2.3
For the period of 1993 until 2023, plot the top 10 countries with most numbers of total refugees over the period. What is the total number of conflicts for these countries over the same period? Please provide a brief discussion.

Task 2.2.4
Plot (scatter or line plot) the percentage change in Refugee population from 2003 to 2023, for the country having the highest number of conflicts since 2010. In this question, you must consider the Refugee population for each year between 2003 and 2023. Very briefly interpret the generated plot.

Task 2.2.5
Present a scatter plot between the mean Refugee population of each country and the mean conflicts from the year 1990 until 2023. Very briefly interpret the generated plot, particularly looking for any correlation (if present) among the plotted variables. In this question, you must consider each year between 1990 and 2023 to find the mean conflicts and mean Refugee population.

NOTE:
The solution for Task 2.2 should be presented through a Jupyter Notebook (.ipynb). All the functions, libraries and coding steps should be explained in a lucid manner. The notebook should run without any error and all the results should easily be reproducible. Your brief interpretation of the generated plots should also be contained in this Jupyter Notebook.

Geospatial Sentiment Analysis Using Social Media Data

In this part, you will apply geospatial sentiment analysis to Twitter data using the Python library, TextBlob. Data consisting of tweets relevant to ChatGPT chatbot launch is provided. The dataset can be found on Moodle under the Assessment folder.

Task 3.1: Data Pre-processing

Using a set of suitable Python libraries, randomly retrieve 500 tweets **where user locations are available**. The user **MUST be a verified user and have atleast 100 followers**. You should also filter out the irrelevant characters, symbols, hashtags, URLs etc. from the tweets to avoid any possible masking of the actual sentiment associated with the tweets. From this point onward you should use the processed tweet data for all the subsequent analyses.

Task 3.2: Geocoding

Geocode on all the 500 tweets retrieved and filtered in the previous step. To perform geocoding, you must be using a Python-based tool. Once the geocoding is performed then augment the tweet data set with two extra columns. One column should contain latitude and the other one should contain longitude information corresponding to a tweet.

Task 3.3 Polarity analysis

Calculate the polarity values of all the tweets. For a given geographical location, if you have more than one tweet then find the average polarity value taking into consideration all the tweets generated from the same location. Using a suitable plot type (such as a geographical map), perform a geospatial visualisation of the polarities corresponding to all the tweets. Whilst you are free to choose a plot type, the visualisation must be clear and easy to understand/interpret.

Task 3.4 Subjectivity Analysis

Calculate the subjectivity values of all the tweets. For a given geographical location, if you have more than one tweet then find the average subjectivity value taking into consideration all the tweets generated from the same location. Using a suitable plot type (such as a geographical map), perform a geospatial visualisation of the subjectivities corresponding to all the tweets. Whilst you are free to choose a plot type, the visualisation must be clear and easy to understand/interpret.

Task 3.5 Storify/Interpretation

In this task, use your geospatial data analytical skill to storify (in not more than 500 words) the results obtained in the preceding two tasks. Imagine yourself as an advisor to a leading competitor of ChatGPT. Your task is to update about the public sentiment related to ChatGPT across different parts of the world. You may try to answer some of these example questions – How is the public opinion about Chat GPT? Which locations have positive views about this IA bot and where can you see a vast amount of negativity? Despite having positive/negative/mixed sentiment about intelligent AI bot, will you take these tweets very seriously? Are the messages loud and clear? Please note that these are only suggestive questions. You are strongly recommended to not constrain your sentiment analytical skills only within these questions. Remember, a good data scientist should be able to retrieve every possible information buried within the data!
The solution for the Task 3.1 to 3.5 should include a Jupyter notebook (.ipynb) describing all the major steps performed during the analysis. All the functions, libraries and coding steps should be explained in a lucid manner. The notebook should run without any error and all the results should easily be reproducible.

Word count (or equivalent):	Please see the individual task for any applicable word limit.

This a reflection of the effort required for the assessment. Word counts will normally include any text, tables, calculations, figures, subtitles and citations. Reference lists and contents of appendices are excluded from the word count. Contents of appendices are not usually considered when determining your final assessment grade.

Academic or technical terms explained:

Geospatial – The word geospatial is used to indicate that data that has a geographic component to it.

GeoPandas – This is an open-source project to make working with geospatial data in Python easier.

Sentiment analysis – It is the use of natural language processing to systematically identify, extract, quantify, and study affective states and subjective information.

UCDP - Uppsala Conflict Data Program

Key Bloom elements:

Application – Application of Python-based tools such as Pandas and GeoPandas to arrange data, pre-process data, and prepare data for analysis. Application of Python-based GIS tools to geocode, generate choropleth maps etc.

Analysis – Various stages of sentiment analysis such as polarity analysis, subjectivity analysis, and geospatial analysis of public sentiments.

Synthesis – Aggregating the major components of sentiment analysis like polarity and subjectivity values to storify the overall public sentiment related to cryptocurrency in different parts of the world.

Artificial Intelligence Models – Guidance for this assessment:

Artificial Intelligence (AI) models can be a powerful tool to support your learning. The University has provided some resources to support you in its appropriate usage:
- Library Services AI Hub
- Student Guide to AI and Assessment
- Code of Conduct for Students on the use of AI
- Cite Them Right resource on citing materials relating to AI (if permitted)

As per the academic regulations (Academic Handbook Ah1_08), in all cases you must submit work that is your own, acknowledging any part of it that has been informed by another source – including that which is AI generated. Upon submission of work, you will be asked to confirm the following statement:
> *I confirm that this assignment is my own work, except where I have acknowledged the use of works from other sources, including the use of any artificial intelligence (AI) tools, in accordance with what is allowable as described in the assessment brief.*

Please note the following:
- AI should not be used as a substitute for your own knowledge, and you should never include any material that you do not understand and could not explain if asked.
- Not being able to explain your work when asked is likely to be a key factor when considering cases of academic misconduct related to AI.

The following information provides specific guidance for this assessment about what level of AI use is appropriate for this assessment. Remember that in all cases you must submit work that is your own, acknowledging any part of it that has been provided by another source.

NO USE OF GENERATIVE AI EXPECTED • Your assignment should be produced using information sourced by you from your learning materials and academic sources and cited appropriately.	☑

• AI tools for checking spelling, grammar and referencing may be used.	
AI ACKNOWLEDGED • You can use AI tools to learn about your topic, as part of your study, or in preparing initial guidance on assignments (e.g. headline structure, suggestions for inclusion of topics). • Any materials that you have sourced from AI should be rewritten or reconfigured and integrated into your own work and referenced appropriately. It is recommended that this is confirmed by a relevant academic source. • Any support gained from AI should be acknowledged in a statement at the end of the assignment, making clear what the support was, and how you used it and developed it for your own work. Example statements are available in the Student Code of Conduct [link].	
AI EMBEDDED • Use of AI is an integral and expected part of the assessment. • The explicit inclusion of AI within the assessment means that instructions on the expected use will be part of the assessment brief. • Your assessment brief will describe how you should acknowledge the way in which you used AI tools.	

Submission Details

Submission Deadline:	Check Moodle for submission dates	**Estimated Feedback Return Date**	This will normally be 20 working days after initial submission.
Submission Time:	By 4.00pm on the deadline day.		
Moodle/Turnitin:	Any assessments submitted after the deadline will not be marked and will be recorded as a non-attempt unless you have had an extension request agreed or have approved mitigating circumstances. See the School Moodle pages for more information on extensions and mitigating circumstances.		
File Format:	The assessment must be submitted as a pdf document (save the document as a pdf in your software) and submit through the Turnitin submission point in Moodle. **Your assessment should be titled with your:** **student ID number, module code and assessment ID, e.g. st12345678 BHL5007 WRIT1**		
Feedback	Feedback for the assessment will be provided electronically via Moodle. Feedback will be provided with comments on your strengths and the areas which you can improve. View the guidance on how to access your feedback. All marks are provisional and are subject to quality assurance processes and confirmation at the programme Examination Board.		

Assessment Criteria

Learning outcomes assessed

Learning Outcomes		Assessment Criteria
[LO1] Demonstrate an understanding of concepts underlying geospatial analysis and apply them appropriately. [LO2] Carry out forms of social analytics, applying appropriate techniques on social information. [LO3] Design, prototype and implement geospatial analysis framework. [LO4] Identify and describe emerging technologies and research areas relevant to geospatial analytics.		Task 2.1 Application of geospatial visualisation tool (e.g., GeoPandas) on a real-world dataset (LO1, LO3) Task 2.2 Analysis of geospatial datasets (LO1, LO3) Task 3.1 Data pre-processing (social analytics) (LO2) Task 3.2 Geocoding (LO1, LO2) Task 3.3 Polarity analysis (LO2) Task 3.4 Subjectivity analysis (LO2) Task 3.5 Storifying (LO1, LO2)

Other skills/attributes developed

This includes elements of the Cardiff Met EDGE (Ethical, Digital, Global and Entrepreneurial skills) and other attributes developed in students through the completion of the module and assessment. These will also be highlighted in the module guidance, which should be read by all students completing the module. Assessments are not just a way of auditing student knowledge. They are a process which provides additional learning and development through the preparation for and completion of the assessment.

Ethical	Ethical issues regarding the usage of geospatial and social media data are considered.
Digital	The assignment deliverable will consist of executable Python code for digital representation of geospatial data.
Global	The assignment is heavily based on the World Bank data. In addition, global issues like cryptocurrency are considered.
Entrepreneurial	The solution to part 2 involves applications of geospatial data in gauging public sentiment which could be translated into entrepreneurial activities such as market research, think tanks etc.

Marking/Assessment Criteria

Assessment Criteria	100%
Task 2.1 (LO1, LO3) Application of geospatial visualisation tool (e.g., GeoPandas) on a real-world dataset	15%
Task 2.2 (LO1, LO3) Analysis of geospatial datasets	35%
Task 3.1 (LO2) Data pre-processing (social analytics)	10%
Task 3.2 (LO1, LO2) Geocoding	10%
Task 3.3 (LO2) Polarity analysis	10%
Task 3.4 (LO2) Subjectivity analysis	10%
Task 3.5 (LO1, LO2) Storifying	10%

Further Information on assessment, referencing and grading can be found in the Module Handbook (on Moodle)

Marking/Assessment Criteria

80 – 100%	An excellent theme is developed. Solutions to tasks 1 and 2 demonstrate excellent usage of Python-based GIS tools. The solution to task 2 demonstrates an excellent usage of sentiment analysis tools. The solution to task 3 demonstrates very good research and articulation skills. Excellent structure to the report. Very few mistakes. Clear punctuated sentences. The writing style is very clear and informative without being verbose. Demonstrates an in-depth understanding of the techniques used for geospatial analysis, social media information analysis and the importance of geospatial analysis in different areas. Evidence of a wide range of appropriate, quality sources used, including appropriate academic journals and books. Clear ability to evaluate the quality of sources. All sources are cited with no errors. Standard referencing style used with no (or minimal) errors/omissions. A fully working and publication standard application that demonstrates an excellent understanding of Geospatial analysis and social analytics techniques with excellent justification. Able to utilise geospatial analysis tools for plotting, and analysing data, with an excellent understanding of modelling of real-world problems with social information integration.
70 – 79%	A very good theme is developed. Solutions to tasks 1 and 2 demonstrate very good usage of Python-based GIS tools. The solution to task 2 demonstrates a very good usage of sentiment analysis tools. The solution to task 3 demonstrates very good research and articulation skills. Well- formulated structure. Few mistakes. Clear punctuated sentences. The writing style is clear and informative without being verbose. Demonstrates a very good understanding of the techniques used for geospatial analysis, social media information analysis and the importance of geospatial analysis in different areas. Evidence of a wide range of appropriate, quality sources used, including appropriate academic journals and books. Ability to evaluate the quality of sources. Relevant sources cited with no errors. Standard referencing style used with no (or minimal) errors/omissions. A fully working application that demonstrates a very good understanding of the Geospatial analysis and social analytics techniques with excellent justification. Able to utilise geospatial analysis tools for plotting, and analysing data, with an excellent understanding of modelling of real-world problems with social information integration.

60-69%	A good theme is developed. Solutions to tasks 1 and 2 demonstrate good usage of Python-based GIS tools. The solution to task 2 demonstrates a good usage of sentiment analysis tools. The solution to task 3 demonstrates good research and articulation skills. Well-written, clearly designed, and contains some grammatical mistakes. Demonstrates an understanding of the techniques used for geospatial analysis, social media information analysis and the importance of geospatial analysis in different areas. Evidence of reading a minimal number of relevant publications (books/journals/websites) of appropriate quality. Shows some ability to evaluate the quality of sources. Citations were attempted; most sources
40-49%	Minimal structure to the work. Minimal discussion of the techniques used for geospatial analysis, social media information analysis and the importance of geospatial analysis in different areas. Minimal level of discussion with significant errors and omissions. Numerous spelling mistakes in the report. Contains some significant errors. Evidence of some research. Sources mainly websites and/or class notes. Limited ability to evaluate the quality of sources. Citations attempted; with errors/ omissions. Incorrect referencing standard used but with minor errors. A partially working application that demonstrates some understanding of the Geospatial analysis and social analytics techniques with or without justification. Able to utilise geospatial analysis tools for analysing data, with minimal understanding of modelling of real-world problems with social information integration.
20-39%	Unclear structure. Does not show a full understanding of the issues. The writing style is unclear. Many spelling/grammatical mistakes. Limited understanding of requirements. Missing or no evidence of research. Copyright restrictions infringed. Few or no sources are cited in the text. Standard referencing style not used. Non-working or poor application that demonstrates little understanding of the geospatial analysis and social analytics techniques with poor justification. Not Able to utilise geospatial analysis tools for analysing data, with little or no understanding of modelling of real-world problems with social information integration. No learning outcomes are met in full although there may be minimal attainment in relation to one or two.
1-19%	No learning outcomes are met in full. The report indicates a very poor understanding of the module and assignment's requirements. Any code (if attempted) is either non-relevant or non-working. Likewise, any application of Geospatial attempted and/or reviewed is either vague or incorrect.

Further Information

Who can answer questions about my assessment?

Questions about the assessment should be directed to the staff member who has set the task/assessment brief. This will usually be the Module Leader. They will be happy to answer any queries you have.

Staff members can often provide feedback on an assignment plan but cannot review any drafts of your work prior to submission. The only exception to this rule is for Dissertation Supervisors to provide feedback on a draft of your dissertation.

Referencing and independent learning

Please ensure you reference a range of credible sources, with due attention to the academic literature in the area. The time spent on research and reading from good quality sources will be reflected in the quality of your submitted work.

Remember that what you get out of university depends on what you put in. Your teaching sessions typically represent between 10% and 30% of the time you are expected to study for your degree. A 20-credit module represents 200 hours of study time. The rest of your time should be taken up by self-directed study.

Unless stated otherwise you must use the **HARVARD** referencing system. Further guidance on referencing can be found in the Study Smart area on Moodle and at www.citethemrightonline.com (use your university login details to access the site). Correct referencing is an easy way to improve your marks and essential in achieving higher grades on most assessments.

Technical submission problems

It is strongly advised that you submit your work at least 24 hours before the deadline to allow time to resolve any last minute problems you might have. If you are having issues with IT or Turnitin you should contact the IT Helpdesk on (+44) 2920 417000. You may require evidence of the Helpdesk call if you are trying to demonstrate that a fault with Moodle or Turnitin was the cause of a late submission.

Extensions and mitigating circumstances

Short extensions on assessment deadlines can be requested in specific circumstances. If you are encountering particular hardship which has been affecting your studies, then you may be able to apply for mitigating circumstances. This can give the teachers on your programme more scope to adapt the assessment requirements to support your needs. Extensions and mitigating circumstances policies and procedures are regularly updated. You should refer to your degree programme or school Moodle pages for information on extensions and mitigating circumstances.

Unfair academic practice

Cardiff Met takes issues of unfair practice **extremely seriously.** The University has procedures and penalties for dealing with unfair academic practice. These are explained in full in the University's Unfair Practice regulations and procedures under Volume 1, Section 8 of the Academic Handbook. The Module Leader reserves the right to interview students regarding any aspect of their work submitted for assessment.

Types of Unfair Practice, include:

Plagiarism, which can be defined as using without acknowledgement another person's words or ideas and submitting them for assessment as though it were one's own work, for instance by copying, translating from one language to another or unacknowledged paraphrasing. Further examples include:

- Use of any quotation(s) from the published or unpublished work of other persons, whether published in textbooks, articles, the Web, or in any other format, where quotations have not been clearly identified as such by being placed in quotation marks and acknowledged.
- Use of another person's words or ideas that have been slightly changed or paraphrased to make it look different from the original.
- Summarising another person's ideas, judgments, diagrams, figures, or computer programmes without reference to that person in the text and the source in a bibliography/reference list.
- Use of assessment writing services, essay banks and/or any other similar agencies (NB. Students are commonly being blackmailed after using essay mills).
- Use of unacknowledged material downloaded from the Internet.
- Re-use of one's own material except as authorised by your degree programme.

Collusion, which can be defined as when work that that has been undertaken with others is submitted and passed off as solely the work of one person. Modules will clearly identify where joint preparation and joint submission are permitted, in all other cases they are not.
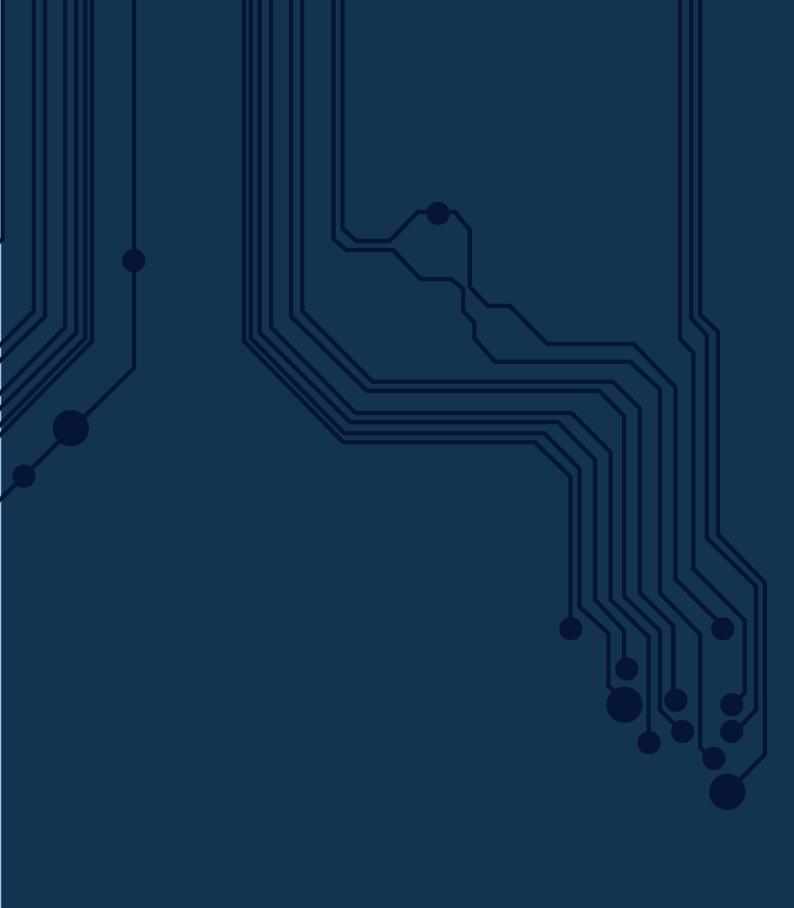
Fabrication of data, making false claims to have carried out experiments, observations, interviews or other forms of data collection and analysis, or acting dishonestly in any other way.

How is my work graded?

Assessment grading is subject to thorough quality control processes. You can view a summary of these processes on the Assessment Explained Infographic.

Grading of work at each level of Cardiff Met degree courses is benchmarked against a set of general requirements set out in Volume 1, Section 4 of our Academic Handbook. A simplified version of these Grade Band Descriptors (GBDs) with short videos explaining some of the academic terminology used can be accessed for Foundation, 1st year, 2nd year and 3rd year undergraduate and MSc programmes.

We would strongly recommend looking at the Study Smart area of Moodle to find out more about assessments and key academic skills which can have a significant impact on your grades. Always check your work thoroughly before submission

Cardiff Met

Met**Caerdydd**