Cardiff School of Technologies

Assessment Brief

Module Code	Module Title
CIS7030	Geospatial Analysis
Academic Year	**Semester**
2022-23	I

Module Leader email

AAnupam@cardiffmet.ac.uk

Content

Assessment Details

Assessment title	Abr.	Weighting
Mapping and Sentiment Analysis of Geospatial Data	WRIT1	100%

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

Task/assessment brief:

PART 1 – Visualisation and Analytics of Geospatial Data

Task 1.1: Application of Python based geospatial visualisation tool (e.g., GeoPandas) on a real-world dataset

This task requires you to use the dataset, renewable electricity output (% of total electricity output). The source of the dataset is the World Bank. The dataset is available on Moodle under the Assessment folder. Use a Python based visualisation tool (such as, GeoPandas) to plot a set of choropleth maps representing the world renewable electricity output for the years 2000 and 2015 respectively. 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 1.2: Analysis of geospatial datasets

In this task you are required to use one more dataset, world total population (source: World Bank) in addition to the renewable electricity output dataset used in the previous task. Both the datasets are available on Moodle under the Assessment folder. All the choropleths and plots must be generated using appropriate Python based tools.

Task 1.2.1

For year 2005, generate choropleth maps of renewable electricity output for only the countries having population less than 9029575. Very briefly interpret the generated map.

Task 1.2.2

For year 2005, generate choropleth maps of renewable electricity output for only the countries having population greater than 295516595. Very briefly interpret the generated map.

Task 1.2.3

For year 2005, generate choropleth maps of renewable electricity output for only the countries having population between 138865014 and 1147609925. Very briefly interpret the generated map.

Task 1.2.4

Plot (scatter or line plot) the percentage change in renewable electricity output from 2000 to 2010, for the country having the highest population in 2010. In this question, you must consider the renewable electricity output for each year between 2000 and 2010. Very briefly interpret the generated plot.

Task 1.2.5

Present a scatter plot between mean population of each country and mean renewable electricity output considering from year 2000 until 2010. Very briefly interpret the generated plot, particularly look for any correlation (if present) among the plotted variables. In this question, you have to consider each year between 2000 and 2010 to find the mean population and mean renewable electricity output.

NOTE:

The solution for the Task 1.2 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 about the generated plots should also be contained in this Jupyter notebook.

PART 2 – Geospatial Sentiment Analysis using Social Media Data

In this part, you will apply geospatial sentiment analysis to Twitter data using the Python library, TextBlob. A data consisting of tweets relevant to the 2021 United Nations Climate Change Conference, more commonly referred to as COP 26 is provided. The dataset can be found on Moodle under the Assessment folder.

Task 2.1: Data Pre-processing

Using a set of suitable Python libraries, randomly retrieve 500 tweets where user locations are available. You should also filter out the irrelevant characters, symbols, hashtags, URL 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 2.2: 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 2.3 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, 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 2.4 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, 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 2.5 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. The UN COP is held on a yearly basis to discuss the issues related to climate change especially global warming and the steps required to mitigate the impact of climate change. One of the plausible solutions to reduce the global warming is switching to the renewable or green sources of energy. The tweet dataset is collated during the COP 26 event and hence contains public opinion regarding transition to green energy for net carbon zero. Imagine

yourself as policy advisor to the UK government whose job is to update about the public sentiment related to energy transition across different parts of the world. You may try to answer some of these example questions – How is the public opinion about energy transition? Which locations are having positive views about this issue and where can you see a vast amount of negativity? Despite having positive/negative/mixed sentiment about energy transition, will you take these tweets very seriously (HINT: if the tweet originates from outside the UK, then it may not affect the government policies!)? Are the messages loud and clear? Please note that these are only suggestive questions. You are strongly recommended to not constrain your sentiment analysis 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 part 2 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.

PART 3 – Critical review

Task 3.1 In this task, you are required to perform critical review of the article – Geospatial analysis of misinformation in COVID-19 related tweets. The article is also available on Moodle under the Assessment folder. Critical review can be achieved by the following suggested steps: (a) Read the paper thoroughly from Abstract until Conclusion (b) summarise the basic principle and applications of the key techniques (such as, Geographically Weighted Regression) employed in the paper (c) write down three major contributions of this paper (d) write down three critical points as observed by you while reading this paper. The deliverable for this task should include adequate citations to the relevant sources.

Please note that the maximum word limit for the Task 3.1 is 500. The word limit is applicable only to the actual review and the references/bibliography are excluded from this word limit.

Task 3.2 Report on the importance of geospatial analysis in different areas and how these areas are benefitting from geospatial analysis

Among the prominent areas of applications for geospatial analysis – choose any TWO areas. You can find the details about these areas of applications through the URL, https://www.esri.com/en-us/industries/index. Write a report, explaining current state, what are the challenges, what solutions are available and the future direction. Like the previous task, the sources must be cited appropriately. Please note that the maximum word limit for the Task 3.2 (including both the areas) is 500. The word limit is applicable only to the main content and the references/bibliography are excluded from this word limit.

Word count (or equivalent):	Look at 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.

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 energy transition in different parts of the world.

Submission Details

Submission Deadline:	16 December 2022	**Estimated Feedback Return Date**	This will normally be 20 working days after initial submission.
Submission Time:	By 4.00 pm 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 using **TWO** files through the Turnitin submission point on Moodle. The **FIRST** file should be a zip/compressed file within which all the separate files (.pdf/ .ipynb etc.) corresponding to all the tasks (Part 1, Part 2, and Part 3) will be merged as a single zip file. The **SECOND** file should be in a .pdf format corresponding to the Part 3 (Task 3.1 and 3.2).

Your assessment should be titled with your:

student ID number, module code and assessment ID,
e.g. st12345678 CIS7030 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
[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.

Assessment Criteria	**100%**
Task 1.1 (LO1, LO3) Application of geospatial visualisation tool (e.g., GeoPandas) on a real-world dataset	10%
Task 1.2 (LO1, LO3) Analysis of geospatial datasets	30%
Task 2.1 (LO2) Data pre-processing (social analytics)	5%
Task 2.2 (LO1, LO2) Geocoding	5%
Task 2.3 (LO2) Polarity analysis	10%
Task 2.4 (LO2) Subjectivity analysis	10%
Task 2.5 (LO1, LO2) Storifying	10%
Task 3.1 (LO1, LO2, LO4) Critical review	10%
Task 3.2 (LO4) Report on the importance of geospatial analysis in different areas.	10%

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.

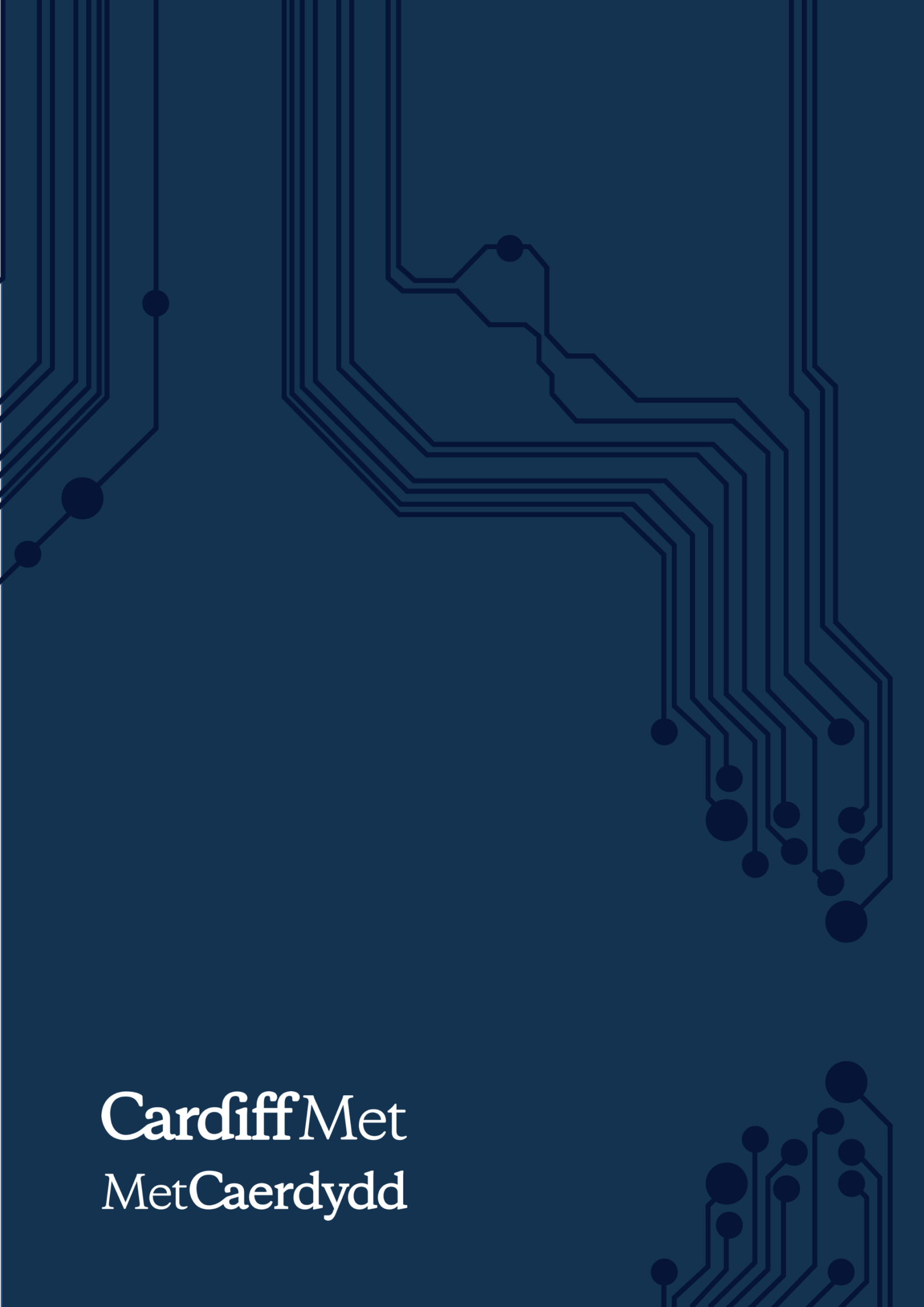
Ethical	Ethical issues regarding usage of geospatial and social media data are considered.
Digital	The assignment deliverable will consist of executable Python code for digital representation of geospatial data.
Global	The assignment is heavily based on the World bank data. In addition, the global issues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

Marking/Assessment Criteria

70 – 100%	A strong theme is developed. Well formulated structure. Very few mistakes. Clear well punctuated sentences. Writing style clear and informative without being verbose. Demonstrates an in depth understanding of the techniques used for geospatial analysis, social media information analysis and importance of geospatial analysis in different areas. Evidence of wide range of appropriate, quality sources used, including appropriate academic journals and books. Clear ability to evaluate quality of sources. Quotations used appropriately and sparingly. All sources cited with no errors. Standard referencing style used with no (or minimal) errors/omissions. Fully working application that demonstrates excellent understanding of the Geospatial analysis and social analytics techniques with excellent justification. Able to utilise geospatial analysis tools for plotting, analysing data, with an excellent understanding of modelling of real-world problems with social information integration.
60-69%	A clear theme is developed. Well written, clearly designed, containing some spelling mistakes. Demonstrates an understanding of the techniques used for geospatial analysis, social media information analysis and importance of geospatial analysis in different areas. Evidence of reading minimal number of relevant publications (books/journals/web sites) of appropriate quality. Shows some ability to evaluate quality of sources. Citations attempted; most sources cited appropriately in the text. Some errors in use of standard referencing style. Fully working application that demonstrates a good understanding of the Geospatial analysis and social analytics techniques with a good justification. Able to utilise geospatial analysis tools for analysing data, with a good understanding of modelling of real-world problems with social information integration.
50-59%	Clear structure to work. Suggests some understanding of the techniques used for geospatial analysis, social media information analysis and importance of geospatial analysis in different areas. However, it is accompanied with significant omissions. Some spelling mistakes. Basic sentence construction rules followed. Evidence of reading minimal number of relevant publications (books/journals/websites). Some attempt made at evaluating quality of sources. Citations attempted; key sources cited; may include some errors. Correct referencing standard attempted, but with many errors. Clear message presented but contains errors that would have a significant impact. A working application that demonstrates some understanding of the geospatial analysis and social analytics techniques with a good justification. Able to utilise geospatial analysis tool for analysing data, with some understanding of modelling of real-world problems with social information integration.

40-49%	Some structure to the work. Minimal discussion of the techniques used for geospatial analysis, social media information analysis and 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 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 tool for analysing data, with minimal understanding of modelling of real-world problems with social information integration.
0-39%	Unclear structure. Does not show full understanding of the issues. Writing style unclear. Many spelling/grammatical mistakes. Limited understanding or requirements. Missing or no evidence of research. Copyright restrictions infringed. Few or no sources 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 tool 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.

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


CardiffMet
MetCaerdydd