Cardiff Metropolitan University | Prifysgol Metropolitan Caerdydd

School of Technologies

Assessment Brief

Module Code	Module Title
CMP7005	Programming for data analysis

Academic Year	Semester
2025-2026	2

Module Leader email

aprasad@cardiffmet.ac.uk

Contents

Assessment Details

Assessment title	Abr.	Weighting
From Data to Application Development	PRAC1	70%

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

Task/Assessment brief:

Context & Background:

Air pollution refers to the presence of harmful chemicals and particulate matter suspended in the atmosphere, posing significant risks to both public health and the environment (Lim et al., 2020). Prolonged exposure to polluted air markedly increases the likelihood of cardiovascular diseases, strokes, chronic respiratory illnesses, and lung cancer (Brauer et al., 2021; Li et al., 2019).

Over the past two decades, Beijing (39°26′–41°03′ N, 115°25′–117°30′ E), the capital of China, has experienced persistently poor PM2.5 air quality. This degradation is largely attributed to rapid economic expansion and growing emissions within the city and its surrounding regions (Xu & Zhang (2020); Batterman et al., 2016). Hence, both monitoring and emission controls have been ramped up in recent years.

Although air quality has improved in recent years, Beijing continues to face substantial challenges in managing pollution amid ongoing economic development (Li et al., 2024). More broadly, air pollution remains a critical global concern with far-reaching implications (Sokhi et al., 2022).

Given the seriousness of this challenge, it is critical to build data-driven systems capable of monitoring and forecasting air pollution. In this assignment, the air quality context provides a real-world dataset; however, the primary aim is to assess your ability to design and implement software solutions in Python. You will be expected to write modular, reusable, and well-documented code to ingest and process datasets, perform exploratory analysis, construct predictive models, and deliver an interactive platform for data exploration and forecasting. The assessment also evaluates your ability to apply version control and collaboration tools to manage your programming project effectively.

Dataset Information

You are provided with hourly air quality data collected from **12 nationally controlled monitoring stations** in Beijing:(Gucheng, Wanshouxigong, Tiantan, Guanyuan, Dongsi, Nongzhanguan, Wanliu, Aotizhongxin, Shunyi, Changping, Dingling, and Huairou). The datasets include two major categories:

1. **Air-quality pollutants:**
   - Particulates: PM2.5, PM10
   - Gas pollutants: SO2, NO2, CO, O3
2. **Meteorological conditions:**
   - Wind speed (Wspd)
   - Rainfall (Rain)
   - Temperature (Temp)
   - Dew Point (Dewp)
   - Pressure (Pres)

The datasets range from **1 March 2013 to 28 February 2017**.

Task 1: Data Selection & Handling

You are required to construct a combined dataset for analysis by selecting:

- **Select 2 inner (urban) and 2 outer (suburban) stations.** *(You may read the articles such as Xu& Zhang(2004) and/or Yao et al. (2015) for station categorisation and provide short justification for your selection)*
- Download the four selected datasets.

- Import them into your development environment.
- Merge them into a single unified dataset for analysis (ensure proper handling of timestamps and station identifiers).

Task 2: Exploratory Data Analysis (EDA)

2.1. Data Understanding

- Provide an overview that may include the following, but not limited to:
  - Number of rows and columns
  - Column descriptions
  - Data types
  - Missing values
  - Statistical Summary
  - Initial observations & interpretation

2.2. Data preprocessing:

- Perform the necessary data preprocessing steps, including but not limited to handling missing values, removing duplicate entries, feature engineering (e.g., datetime components, AQI levels), and overall data cleaning on the main dataset.

2.3. Statistical/Computational Analysis & Visualisation

- Perform the necessary steps such as univariate (distribution of pollutants & meteorological variables), bivariate(e.g. relationships such as PM2.5 vs. Temp, NO2 vs. O3 but not limited to these), and multivariate analysis (correlation, heatmaps, pairplots), statistical summary, and visualizing the data (Various charts and graphs, such as bar charts, line charts and scatter plots) that will help in understanding relationships between variables and to gain important insights from data. Interpret the key results to demonstrate understanding generated from statistical and visual analysis.
- Explore the dataset however you find meaningful. You may examine different variables, compare stations, investigate temporal behaviours, or analyse interactions between pollutants and meteorological factors. Choose the approaches that you believe best help you understand and interpret the dataset, and present the insights you consider most relevant

Task 3: Model Building:

- After completing all the tasks listed under Task 1 and Task 2, identify and implement the best practices to build a suitable machine-learning model (e.g., feature scaling, encoding techniques, variable selection, and parameter optimization).
- Justify your modelling decisions and evaluate model performance using appropriate metrics.

Task 4: Application Development

Develop an interact*ive application with a graphical user interface (GUI)*. *The application should include multiple sections/pages that allow users to explore*

- The dataset section,
- Visualization section, and
- Model outputs section.

You may design the structure in any way you find appropriate, but it should enable clear navigation between the key components of your workflow.

Task 5: Version Control:

- Use GitHub for version control.
- Commit changes regularly with clear, descriptive messages, for example, added PM2.5 prediction model", "Created correlation heatmap," etc.
- Maintain an organised repository structure and include screenshots of:
  - GitHub commit history
  - GitHub project repository layout

These screenshots should be submitted in your final report to demonstrate your version control usage.

Word count (or equivalent):	2800 words

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:

EDA- Process of analysing and investigating data sets and summarizing their main characteristics, often employing data visualization methods.

GUI - Graphical User Interface (GUI) applications are software programs that display an interface on the screen with which users can interact.

Univariate Analysis- Univariate analysis explores each variable in a data set, separately.

Bivariate Analysis - Bivariate analysis is the analysis of exactly two variables.

Multivariate Analysis- Multivariate analysis is the analysis of more than two variables.

Comprehension – Understanding how to access datasets and perform modifications for data analysis
Analysis - Evaluate commonly used libraries for data analysis tasks to develop solutions on a real-world dataset.
Synthesis- Integrating different programming concepts and techniques to develop functional software
Evaluation- Assessing the effectiveness of data analysis for effective decision-making
Application- Implement functional software and apply version control in practice

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.	☑
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:	This will be provided on the Moodle submission point.	**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 CMP7005 PRAC1

You need to maintain your GitHub repository with all the required files such as your data, your. ipynb working file, and any other files you think are necessary.

Late Submission Window Eligibility

Where submissions are eligible for the late-submission window this will be communicated in the relevant assessment submission point within Moodle.

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

[LO1] Develop software through programming using a suitable high-level general-purpose programming language.
[LO2] Manage file I/O handling in order to access and modify datasets for further analysis.

[LO3] Critically evaluate commonly used libraries used for data analysis tasks such as machine learning and visualisation.
[LO4] Critically evaluate, develop, and implement solutions for processing datasets and solving complex problems in various environments using relevant programming paradigms.
[LO5] Manage version control systems and collaborative programming and project management tools, and their application to real-world problems.

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	Students will be encouraged to consider the ethical implications of data handling and analysis and will be made aware of the university and/ or other global ethics processes.
Digital	Develop strong digital skills by focusing on programming, software development, and the use of industry-standard tools relevant to the industry, enhancing their digital competence.
Global	Through international research, literature, and real-world case studies, students will be able to build a strong awareness/ appreciation of the importance of being able to think outside the box and find solutions.
Entrepreneurial	Foster an entrepreneurial mindset by requiring students to work on real-world problems. Through collaborative programming and project management, students will learn to take initiative, be innovative, and apply professional practices in developing data-driven solutions that could have commercial applications.

Marking/Assessment Criteria

Assessment Tasks	Marks
Task 1: Data Handling [LO2]	**5%**
Task 2: EDA [LO2, LO3, LO4]	**50%**
a. Fundamental Data Understanding [LO2, LO4]	10%
b. Data pre-processing [LO3, LO4]	10%
c. Statistical/computational-based analysis and Visualisation [LO3, LO4] and inferences drawn from analysis	30%
Task 3: Model Building [LO3]	**15%**
Task 4: Application development [LO1]	**20%**
Task 5: Version control [LO5]	**10%**
Total	**100%**

Staff note – Please use this area to provide your marking criteria. Where possible this should give a breakdown of what would be expected for each element of the work and/or expectations of work from each of the different grade brackets (using the full range of marks). Grading of assessments should reflect the learning outcomes for the module and be aligned with Cardiff Met's Generic Band Descriptors for *Level 3*, *Level 4*, *Level 5*, *Level 6* and *Masters Level*.

Marking/Assessment Criteria

Tasks	1-19%	20%-39% (Clear Fail)	40%-49% (Narrow Fail)	50%-59% (pass)	60%- 69% (Merit)	70%-79% (Distinction)	80%-100% (Distinction+)
Task 1: Data Handling	No data imported or significant errors in the import and merging process that make further analysis impossible. Demonstrates no understanding of data importing techniques and handles data inefficiently.	Data imported with major errors, or merging failed, making the dataset unusable for the next steps. Demonstrates little to no understanding of data importing techniques and handles data with limited to no inefficiency.	Data imported with moderate issues; merging is attempted, but with noticeable mistakes or partial success. Demonstrates a limited understanding of data importing techniques and limited data handling efficiency.	Data is imported correctly, but there are minor errors in dataset merging or missing small parts of data. Demonstrates a moderate understanding of data importing techniques and moderate data handling efficiency.	Data is imported and merged correctly, with a few minor issues. Demonstrates a good understanding of data importing techniques and handles data with good efficiency.	Data is imported and merged correctly and efficiently with no issues. Demonstrates a very good understanding of data importing techniques and handles data with a very good efficiency.	Data is imported and merged flawlessly. Demonstrates a deep understanding of data importing techniques with a highly efficient data handling process.
Task 2: EDA	No attempt or incorrect understanding of the dataset. No insight into data structure, types, or missing values. No data pre-processing. No attempt or incorrect statistical/computation-based analysis. No visualizations to provide insight into the variables and the data. Shows no analytical judgement in selecting meaningful ways to explore the dataset and interpreting results with no clarity and depth	Little to no understanding of the dataset. Incorrect identification of key features, such as missing values or incorrect data types. Minimal or flawed attempt at data pre-processing. Minimal statistical/computation-based analysis with major errors. Poor visualizations that lack clarity in providing insight between variables and the data. Shows Little to no analytical judgment in selecting meaningful ways to explore the dataset	Limited understanding of data with significant gaps in knowledge. Some errors in identifying key features, like missing values. Basic pre-processing done. Basic statistical/computation-based analysis. Visualizations lack clarity in providing insight between variables and about the data. Shows minimal analytical judgement in selecting meaningful ways to explore the dataset and interpreting results with some clarity and depth	Adequate understanding of dataset structure, types, and missing values, but lacks depth or clarity in insight. Adequate pre-processing of data. Adequate statistical/computation-based analysis with minor errors. Visualizations are basic but functional, offering some insights between variables and about the data. Shows some analytical judgement in selecting meaningful ways to explore the dataset and interpreting results with limited depth.	A clear understanding of the dataset, its structure, types, and missing values. Minor gaps in insights. Good data pre-processing. Good statistical/computation-based analysis with minor gaps. Visualizations provide clear insights into relationships between variables and about the data. Shows clear analytical judgement in selecting meaningful ways to explore the dataset and interpreting results with clarity and depth	Very good understanding of the dataset, identifying all key features, types, and missing values correctly. Very good data pre-processing. Very good statistical/computation-based analysis and clear, insightful visualizations that demonstrate relationships between variables and about the data. Shows strong analytical judgement in selecting meaningful ways to explore the dataset and interpreting results with clarity and depth.	Excellent understanding of the dataset, identifying all key features, types, and missing values correctly. Excellent data pre-processing. Excellent, error-free statistical/ computation-based analysis with highly effective and professional visualizations that offer deep insights between variables and about the data. Shows outstanding analytical judgement in selecting meaningful ways to explore the dataset and interpreting

		and interpreting results with Little to no clarity and depth					results with great clarity and depth
Task 3: Model Building	No model built or entirely incorrect. No attempt at feature scaling or selecting variables. Poor model performance.	Minimal or incorrect model building. Major issues in feature scaling, variable selection, Poor model performance.	Basic model built, but with significant issues in feature scaling. Basic model performance.	Model built and functional with minor issues in scaling or variable selection. Moderate model performance.	Good model with mostly correct feature scaling and variable selection. Good model performance.	Very good model with correct feature scaling and well-selected variables. Very Good model performance.	Excellent model with perfect feature scaling, variable selection, and outstanding performance. Models are optimized and well-tuned.
Task 4: Application Development	No application developed or entirely non-functional. No GUI pages or sections created.	Minimal application with major functionality or design issues. The GUI is incomplete or non-functional.	Basic application with significant issues in functionality or usability. GUI is present but poorly designed.	Functional application with a basic GUI. Minor issues in design or navigation, but all sections are present.	Good application with a functional GUI and logical navigation. Some minor improvements are needed in the design.	Very good application with a well-designed GUI. All sections are functional and well-integrated.	Excellent application, fully functional with a highly polished GUI. Sections are well-organized and intuitive, demonstrating advanced skills.
Task 5: Version Control	No use of version control with no commit messages, or errors in documentation.	Minimal version control with poor commit messages or errors in documentation.	Basic use of version control, but with errors in documenting progress. Commit messages are unclear.	Adequate use of version control with minor issues. Commit messages are basic but sufficient to track progress with adequate documentation.	Good version control usage with clear commit messages and an organized workflow.	Very good version control usage with clear, detailed commit messages and a well-structured workflow.	Excellent version control usage, with highly detailed commit messages. Workflow is exceptionally organized and well-documented.

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 using Cite Them Right (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.

Late submission and mitigating circumstances

If you are experiencing personal difficulties which are impacting your ability to engage with assessment, there are a range of support options available, including via your Personal Academic Tutor, the Mitigating Circumstances procedure and the Support to Study procedure.

The Mitigating Circumstances policy and procedure sets out when and how students can:

- Submit eligible assessments late but have it recorded as 'on time', up to four times per level of study.
- Self-declare significant exceptional circumstances in order to defer the assessment to the next submission opportunity.

More information about Late Submission and Mitigating Circumstances is available on MetCentral.

Students with a disability or chronic illness should contact the University's Student Wellbeing Team to discuss any appropriate reasonable adjustments which may be made to their learning and assessments.

Academic Misconduct

Cardiff Met takes issues of academic misconduct **extremely seriously.** The University has procedures and penalties for dealing with academic misconduct. These are explained in full in the University's Academic Misconduct 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 Academic Misconduct, 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

MetCaerdydd