Kingston University Assignment Brief

School of Computer Science and Mathematics

Module Code	CI7340
Module Title	Applied Data Programming
Assessment Title	Research Report and Visualisation
Element Label	Coursework 1 (E1)
Type	Summative – 40% of module mark
Set by	
Assessment due date and time	23rd Nov 2025
Formal feedback due date	

All assignments must be submitted by the date and time specified above.

You are required to submit an electronic copy of your completed assignment, in the file format(s) specified by the module team (e.g., Word, PDF, programme code files), via the Assignments section of Canvas and follow any specific instructions provided. Any change to this instruction will be advised via Canvas.

If files are shared outside of Canvas (where specified by the module team), you must ensure that the files are accessible and available for staff to access without the need to request additional access privileges.

In line with University Regulations coursework submitted up to a week late will be capped at **50%**.

Coursework submitted after this time will receive 0%.

In case of illness or other issues affecting your studies please refer to the University Mitigating Circumstances and Extensions Regulations. Please note that once you have submitted your work you have judged yourself fit to undertake the assessment and cannot usually claim mitigating

circumstances retrospectively. Please refer to the Mitigating Circumstances Regulations for more information.

Guidance on avoiding academic assessment offences such as plagiarism and collusion can be found in the Digital Learning and Tools module on Canvas – see Academic Integrity.

PURPOSE OF THIS ASSESSMENT / WHAT IS EXPECTED

Note: This is a group work

The Case Study

In this coursework, you are provided with a dataset that represents Montgomery County of the USA's crime statistics. The data showed in the file is derived from reported crimes classified according to the National Incident-Based Reporting System (NIBRS) of the Criminal Justice

Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program and documented by approved police incident reports. This data presented allows the public to access the Montgomery County Police Department's database of founded crime.

The information contained includes founded crimes reported between 2018 and 2022. The data are entered utilizing Uniform Crime Reporting (UCR) rules. As per the Montgomery County's UCR rules multiple offenses may appear and may link as part of a single founded reported incident, and each offense may have multiple victims. Also note that these crime reports are based on preliminary information supplied to the Police Department by the reporting parties.

The information presented to you will help understand the pattern of crimes occurred in the Montgomery County, which will give you opportunity to decide the types of crimes, frequencies of the occurrence and may help suggest the local government to mitigate crime rate in the county. Thus, it will help support local government and afterwards nationally to minimise certain types of crime and predict pattern of future crimes.

Given the dataset you are required to prepare a research report with visualisation that includes the following:

- Technically explain the dataset (e.g., define your own problem statement) and its scope of analysis.
- Assess the quality of the dataset and write how the dataset can be improved for missing values. Mention different ways to handle missing values in the dataset.
- Practical implementation using Python and its libraries to show different approaches to handle missing values.
- Explain possible ways you can use to interpret the dataset and what method you can use to interpret the data. Try to find inconsistencies (e.g. noise) in the dataset and come up with your own proposition/solution/assumption to make the dataset better. For example, by eliminating what proportion or percentage of the noise the dataset will be better.
- Once you have understood the dataset, please formulate your own 10 research questions that will address the case study dataset using relevant information. Each of your research questions need to be supported by two visualisations that will answer the questions. We are expecting 20 visualisations in-total throughout the coursework (10 questions * 2 visualisations).

In this report, you are asked to use the dataset provided to perform a research analysis and produce a report consisting of the following:

Introduction (20 marks)

This section should include an introduction to the data provided to you in data science context.

Based on the dataset, form a problem statement that will help better understand the data. The problem statement should include relevant explanation of the approaches you will take to solve the or predict solution to the problem. Identity scope and research questions (10 questions) that you plan to investigate as part of your research report. These questions will formulate the base of your intended research method, techniques and expected outcome to aid the solution. Describe your workflow supported with figures/tables/graphs(visualisations) etc.

Preliminary Data Analysis (20 marks)

- **Dataset(s) (5 marks)**
  In this part, discuss the dataset: its source, format, type of data, the format used (structured/unstructured), etc. (Hint: Read the dataset and discuss the columns, type of data in the columns, etc.).
- **Data Quality initial assessment (15 marks)**

Discuss about the quality of data that you have. Are there many missing values, data is clean, etc.). Are different transformation methods required for your further analysis? Make sure that you discuss the different methods of data cleaning, different methods to account for missing data (Null values), pre-processing, and transform features. What are the different types of data wrangling skills (e.g., extraction, merging, and/or construction of analytical data sets) that can be used?

Exploratory Data Analysis (60 marks)

- **Introduction to EDA (15 marks)**

  Discuss briefly about EDA. Review your discussion from preliminary data analysis section and explain if further data cleaning or transformation is required? What are the different approaches and which ones do you think will be suitable for the data in hand.

- **Descriptive Statistics (5 marks)**

  In this section, discuss the various statistical methods that you think will be relevant to understand your data, help answer your research questions and why? You need to present a detailed discussion on various methods and their suitability for your use case.

- **Data Visualisation (35 marks)**

  Discuss different form of data visualisation techniques (e.g., graphs, heatmap etc). Indicate which visualisation method(s) will be appropriate for your research questions and why? (Hint the 10*2=20 Visualisation happens here!)

- **Conclusion and references (5 marks)**

  Summarise your work and learnings. Refer work taken from external sources/ borrowed/copied from books/journals. Any referencing method can be used (IEEE preferred). Refer to Research Methods (Library talk), SASC or Library Canvas & tools like “Cite them Right” for referencing guidance.

Starting Guide

1) There will be a discussion on the problem statement and how to approach it. Make sure to attend the lecture and take notes, as appropriate.

Note: You are welcome to start with Jupyter Notebook and data analysis and submit Jupyter (.ipynb) output file for the coursework.

Instructions for submission

You need to submit a Microsoft Word document (.doc,.docx) or pdf file only. Any other file format is NOT acceptable and will be considered a non-submission.

At the start of the report, add a section (including names and KU IDs of ALL the group members, and title of the report), a ToC (Table of Contents) and Table of Figures, etc. Make sure that you have informed the module leader of your team and got approval before starting to work together on the report.

Assessment and Feedback

You will be assessed primarily on (but not necessarily limited to) the following: completeness of the task (report content, see above), quality of the work (research questions, reasoning, etc.), presentation style (report structure, logical flow, grammatical and spelling errors, sentence structure etc.), appropriate references using an appropriate style. You will be provided feedback on your work by the deadline mentioned on the Canvas Assessment and Feedback Schedule page here.

Rubric

The above description clarifies the marking structure. Below is just a summary copy of the marks for the research report. In the case of any discrepancies, the above description supersedes this one below.

Section		Marks
Introduction		20
Preliminary Data Analysis	Dataset(s)	5
	Data Quality initial assessment	15
Exploratory Data Analysis	Introduction to EDA	15
	Descriptive Statistics	5
	Data Visualisation	35
Conclusion and references		5

Please note that as part of this group report, ALL group members will initially receive the SAME marks unless stated differently in the contribution part(table) while submitting the coursework.

University Policy and Regulations

Any submission found/suspected of academic misconduct including but not limited to collusion, plagiarism and purchasing or commissioning will NOT be marked, and all the members of the group will be reported to the University for further action. Make sure you read the below guidelines available here:
Academic Misconduct (including Plagiarism)

To apply for extensions and mitigating circumstance, please refer to the guidelines here:
Extensions and Mitigating Circumstances Policy

Viva: We reserve the right to withhold your marks and invite you to a viva where you will be asked questions about your work to demonstrate ownership and understanding of your own submission.

Whether you are asked to viva or not will depend on:

- Your engagement during scheduled classes
- Your engagement with the team
- Your formative submissions (i.e. the work you do for learning purposes rather than marks)
- The work you submit (if it's an interesting submission and we want to ask some questions – this is a GOOD thing, as you may end up with more marks!)

If you are asked to do a viva, then your final mark will be subjected to a multiplier factor as follows:

Multiplier	Viva assessment
1.2	Ownership and understanding of the artefact submitted was excellent, and gave an insight into not only their own work as-is but also possible alternative approaches, demonstrating more understanding of development techniques than the work alone provides.
1	Ownership and understanding of the artefact submitted is clearly established. The student responded to questions about their work with authority.
0.6	The student was able to answer most questions about their work, but there were some gaps in understanding of the artefact submitted.
0.3	The student was able to answer some questions and demonstrate only a limited understanding of their own submission. However, there was a clear disconnect between the quality of artefact and the students' understanding of the technology and principles encapsulated within.
0	The student was not able to provide any meaningful responses to questions about their work. Extremely limited, or no understanding of the artefact submitted was apparent. Alternatively, the student did not respond to the invitation to viva and/or did not attend the viva at the appointed time.

LEARNING OUTCOMES

The following module learning outcomes and professional body learning outcomes are tested in this assessment:

- Apply data preparation and transformation methods ready for data analysis, selecting appropriate methods to transform structured and unstructured data between platforms.
- Retrieve structured data and use exploratory data mining for the purposes of data summary and dimensionality reduction, whilst selecting and evaluating appropriate algorithms.
- Interpret information from large data sets and identify and illustrate relationships within them.
- Program in an appropriate language, environment and infrastructure for Data Science.

- Demonstrate an appreciation of real-time applications and data-driven problems through the production of an artefact based on real world data.

Can I use Generative AI (GAI) as part of this Task?

Default use of GAI: you are permitted to use Generative AI for the following purposes:

- Support spelling, punctuation and grammar.
- Support ideation.
- Support research for the assignment (identifying sources, search) with appropriate citation.
- Take the role of a constructive critic.
- Aid understanding.
- Perform basic image / media editing encompassing cropping, noise reduction, sharpening, enlarging, compression, changing format type and adjusting lighting.

Please note: All of the core writing, creativity, arguments, analysis and reasoning must be your own

Student Guide to GAI at Kingston University 2025/6 in the Digital Learning and Tools module in Canvas (Generative AI section).

Do I need to declare my use of GAI tools?

Yes, if you use Generative AI for any part of your assessment, you must declare this. This applies to all assessments including those in the default and explicit categories.

For this assignment the declaration should be provided at the end of the submission with the heading 'Acknowledgement of GAI Contribution'. This declaration should include a statement on the use of generative AI including the extent of use, and how it was used as part of all stages in creating the final submission.

For assessments that fall into the explicit category (does not apply to the purposes listed in the Default category), any GAI content included in the assignment, e.g., a quoted paragraph of text or an image, should be properly cited as with any non-GAI source.

Further guidance on completing this acknowledgement is provided in the Digital Learning and Tools module in Canvas (Generative AI section).

The module team may also provide additional advice on the specific details required, depending on the nature of the GAI tool used.

You will also need to read and accept the similarity declaration when submitting an assignment in Canvas.

FURTHER INFORMATION ABOUT THIS ASSESSMENT

1) **Coursework Template**

   Please download and use the official *Coursework 1 Template* available in your CI7340 module **Canvas**. The template includes a complete **Table of Contents (TOC)** with the required section headings aligned with the assessment rubric. It is strongly recommended that you use this template to structure your report. The completed coursework must be submitted via **Turnitin** as your final submission. Email submission or any other form of submission is not permitted.

2) **Word Count**

   The expected length of the coursework is **4,000–5,000 words**, with an allowable variation of **±10%** buffer excluding the references, image/figure/table captions**.**

3) **Group Contribution Form:**

   Each group member must complete and submit the following **Group Contribution Form** along with the coursework. This form should appear at the **beginning of the coursework report** to clearly indicate each member's contribution (in %).

Group Members and Participation

Student Name	Student ID	Group Assignment Development Contribution Percentage(%)

MARKING CRITERIA (DETAIL RUBRIC)

Assessment of your submission will be based on the following weighted assessment criteria as given below which relate to the specified module and PSRB learning outcomes. Assessment criteria are reproduced in Canvas in a rubric.

Criteria	Ratings	Pts
Introduction This section should include an introduction to the field of data science, industry/application/domain that the dataset is representative of. You are required to formulate a problem statement and explain the proposed approach you will take to solve the identified problem. Identify different (research) questions (10 questions) that you plan to investigate as part of your research report. Define the objectives, intended research methodology and expected outcomes. Describe your workflow supported with figures.	This area will be used by the assessor to leave comments related to this criterion.	20 pts
Dataset(s) In this part, discuss the dataset - its source, format, type of data, the source of the dataset and format used (structured/unstructured), etc. (Hint: Read the dataset and discuss the columns, type of data in the columns, etc.).	This area will be used by the assessor to leave comments related to this criterion.	5 pts
Data Quality initial assessment Discuss about the quality of data that you have. Are there many missing values, data is clean, etc.). Are different transformation methods required for your further analysis? Make sure that you discuss the different methods of data cleaning, different methods to account for missing data (Null values), pre-processing, and transform features. What are the different types of data wrangling skills (e.g., extraction, merging, and/or construction of analytical data sets) that can be used?	This area will be used by the assessor to leave comments related to this criterion.	15 pts
Introduction to EDA In this section you need to discuss EDA. Also make sure to include discussion about if further data cleaning or transformation is required? What are the different approaches and which ones do you think will be suitable for the data in hand.	This area will be used by the assessor to leave comments related to this criterion.	15 pts
Descriptive Statistics In this section, discuss the various statistical methods. Discuss which ones you think will be relevant to understand your data, help answer your research questions and why? You need to present a detailed discussion on various methods and their suitability for your use case.	This area will be used by the assessor to leave comments related to this criterion.	5 pts

Data Quality initial assessment Discuss about the quality of data that you have. Are there many missing values, data is clean, etc.). Are different transformation methods required for your further analysis? Make sure that you discuss the different methods of data cleaning, different methods to account for missing data (Null values), pre-processing, and transform features. What are the different types of data wrangling skills (e.g., extraction, merging, and/or construction of analytical data sets) that can be used?	This area will be used by the assessor to leave comments related to this criterion.	15 pts
Introduction to EDA In this section you need to discuss EDA. Also make sure to include discussion about if further data cleaning or transformation is required? What are the different approaches and which ones do you think will be suitable for the data in hand.	This area will be used by the assessor to leave comments related to this criterion.	15 pts
Descriptive Statistics In this section, discuss the various statistical methods. Discuss which ones you think will be relevant to understand your data, help answer your research questions and why? You need to present a detailed discussion on various methods and their suitability for your use case.	This area will be used by the assessor to leave comments related to this criterion.	5 pts
Data Visualization What is data visualization, how and why are they used. Discuss about the best practices. What are the different types of graphs (pie-chart/scatter plots/bar plots/histogram, etc.-you can use sample ones from the official website) and when are they used? Which visualization methods do you think will be appropriate to your research questions and why?(10*2=20 Visualisation ?!)	This area will be used by the assessor to leave comments related to this criterion.	35 pts
Summary and Conclusion Summarize your work and learnings.	This area will be used by the assessor to leave comments related to this criterion.	3 pts
References Anything borrowed/copied from any external source needs to be referenced. Any referencing method can be used (IEEE preferred). Refer to Research Methods (Library talk), SASC or Library Canvas & tools like "Cite them Right" for referencing guidance.	This area will be used by the assessor to leave comments related to this criterion.	2 pts
		Total points: 100

ACADEMIC SKILLS SUPPORT

For help and advice on this assessment please contact the module lecturers. For advice on academic writing and referencing please contact the SASC.