ISYS5050 - KM Systems

Group Project (35%)
Maximum Score: 35 marks
(To be completed in groups of 4 or 5, due in Week 13)

Instructions

1. Provide **a written report** answering each of the following questions. There is no limit on the total number of pages but keep it less than 25 pages if you can. References are not considered.
2. This is a group assessment and, as such, you must be part of a group to be able to complete this assessment. Having said that, **all group members must be across all the questions and have an understanding of the details of the approach taken for each question**.
3. Please make sure that you follow the guidelines and University policy on plagiarism (check the week 1 lecture notes if you are not sure what the guidelines are). If you draw on particular references, please ensure that you do not copy-and-paste or copy text verbatim and **provide full citations** in an academic style (such as APA). We will check for plagiarism using Turn-it-in and other tools, both among the student submissions and external sources.
4. Your work must be submitted on Canvas before the deadline. You may lose marks for late submissions.

Dataset Description

Please refer to the flood dataset from Dartmouth Flood Observatory[1] which includes details on all major flood events recorded globally between 1985 and 2021 with over 5000 such events. Some of the key attributes in the dataset include:

- Country or countries affected
- Detailed location
- Relevant dates (start and end dates)
- Number of people dead or displaced
- Severity Class
    - Class 1: large flood events: significant damage to structures or agriculture; fatalities; and/or 1-2 decades-long reported interval since the last similar event.
    - Class 1.5: very large events: greater than 20 year but less than 100 year recurrence interval, and/or a local recurrence interval of at 10-20 yr.
    - Class 2: Extreme events: with an estimated recurrence interval greater than 100 years
- Main cause

---

[1] *G.R. Brakenridge. Global Active Archive of Large Flood Events. Dartmouth Flood Observatory, University of Colorado, USA. http://floodobservatory.colorado.edu/Archives/.*

Questions

1. Pre-processing, data preparation (5 marks)

   - Identify and describe possible quality issues with the data.
   - Provide a description of the ETL processes and data cleaning steps that you performed (if any) to address each of the issues.
   - Provide before and after screenshots of the most important changes and your ETL processes.
   - Specify any ETL and BI tools as well as any programming or scripting language that are used as part of the project.

2. Climate scientists define Flood Magnitude as a measure of "how severe" a flood is and is calculated using LOG(Duration x Severity Class x Affected Area). Calculate the Flood Magnitude index using the available data and produce a report of the top five countries in terms of the total number of floods of different levels of magnitude (5 marks).

3. Perform analyses to discover seasonal patterns in flood data over time (e.g. monthly/weather seasons etc.) and present a summary of your findings using relevant charts, visualisation and description (5 marks).

4. Carry out a range of analyses to understand the damage and impact of floods on countries and people worldwide. Present how the impact of floods changed overtime and identify countries and areas most affected by the floods. You are required to define or adopt an accurate measure of impact and explain and justify metric(s) used. You can use any, all or a combination of the available measures (10 marks).

5. We are interested to gain insights into the driving factors of floods and understand the effects of climate change and global warming on severity (magnitude) and number of floods over time. Despite the relatively short period of the given dataset (around 37 years), it can provide some insights on certain trends. Perform a range of analyses to identify causes of floods worldwide. In addition to using the given data, you are encouraged to adopt and/or incorporate relevant datasets to conduct your analysis (10 marks).

Suggested Structure for the Final Report

- The report should have a professional look and feel.
- Suggested structure: cover page, summary/abstract, introduction, one section for each question, conclusions, references (if applicable).
- For each question, provide a brief introduction to the question and to your approach followed by any tables, charts, visualisations, and a description of the key findings, inferences and justification.
- Please include only relevant tables and charts that you have created. Each visualisation should be supported by appropriate discussion of the results and the details of how you arrived at your findings.

Notes

- You are free to use any ETL or BI tool or any programming or scripting language. Please specify them.
- You can use external data sources in addition to the given dataset. Provide justification and explanation of the dataset and how it is incorporated into your analyses. You must provide full citation of the dataset in the references section and include a link.
- You can refer to credible official or academic reports and studies that support or contradict your findings. You must provide full citation in an academic style (such as APA) in the references section.

Marking Criteria[2]

- For each question (90%)
    - Quality of visualisations
    - Approach and methods used
    - Overall analysis, accuracy and completeness of the findings
    - Sufficient explanation and discussion
    - Flow, structure and organisation
- Overall quality of the report (10%)
    - Organisation
    - Style of report preparation and presentation
    - Language etc.
    - References and citations (if any)
- Too long or too many charts and visual elements (-5%)

---

[2] These criteria are provided as an overall guideline, are not exhaustive and may change.