# King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

| | |
|---|---|
| **Degree Programmes** | MSc |
| **Module Code** | 7CCSMSDV |
| **Module Title** | Simulation and Data Visualization |
| **Examination Period** | Summer 2021 (Period 2) |
| **Time Allowed** | Submission Deadline April 15, 11PM (BST) |
| **Rubric** | ANSWER ALL QUESTIONS Part 1, 2 and 3. Part 1 carries FIFTEEN marks. Part 2 carries FIFTEEN marks. Part 3 carries THIRTY marks. |

# Project Based Exam

## Introduction

This exam is divided into 3 parts: Analytics, Design and Prototyping, Implementation. Please read carefully the instructions related to each part.
Given the current pandemic situation, we as Data Scientists are called to put our skills to help humanity to mine the large and diverse amount of data available to support knowledge discovery and to battle the Covid-19 pandemic.
To this aim the assignment has two main themes (or tracks). You are required to choose one.
Proposed tracks:

- **Track 1:** Covid-19 Data Analysis.
- **Track 2:** Covid-19 Mutations Data Analysis.

**Data.** We are providing you with:

- Two starting datasets reporting: (i) up-to-date data on the geographic distribution of Covid-19 cases worldwide; (ii) up-to-date data on the Covid-19 mutations in the UK.
  At least one of the two provided datasets should be used for the Exam, depending on which track you choose to work on. [Dataset available on module Keats page]
- A series of links to Covid-19 datasets currently available. These are given as a starting point, it is not mandatory to use them and you are invited to look for any other datasets available on the web that you may find interesting. [List provided on module Keats page]

**Visualizations.** We are providing you with links to existing visualizations, again these are given as a starting point and for inspiration, it is not mandatory to follow them and you are invited to look for other visualizations that are available on the web and that you may find inspirational. [List provided on module Keats page]

**Research questions.** You are provided with an initial research question for each track:
*Q1 - Track 1: "Analyze the spread trend of this virus all over the world. What is the spread over time?"*
*Q1 - Track 2: "Analyze the spread of virus mutations in the UK. What is the spread over time?"*

## Part 1. Analytics [15 marks total]

Based on the track you chose and the respective data provided in the Data section and/or other datasets you may have encountered or may find online, you are required to:

**a.** Propose two or more exploratory research questions (non-trivial questions) beyond Q1, label them as Q2, Q3, etc. [Note: You can propose questions that might require access to datasets beyond the one provided, these can be questions complementary to Q1, investigating other aspects than the ones investigated by Q1, or refinements/expansions of Q1, you are free to decide.].

[5 marks]

**b.** Explain what type of data, beyond the one provided, could be used to answer Q1 and each one of the questions you proposed in **a.**. Assess the fitness of each dataset(s), between the ones we provided or other resources you may have found (please provide links to the latter), you would potentially be using to answer the questions.

[5 marks]

**c.** Explain if and how datasets, described in **b.** and including the one provided, could be correlated.

[5 marks]

## Part 2. Design and Discussion [15 marks total]

Based on visualization approaches surveyed in class and in recommended readings:

**a.** Propose and design a minimum of 3 visualizations that would answer the research questions proposed in **Part 1** (creativity will be rewarded). By design we mean drawing/sketching a prototype. Design can be hand-drawn on paper or using a tool of your choice e.g. PowerPoint, Sketch, Illustrator, D3, Tableau, etc.

[6 marks]

**b.** Each visualization should be accompanied by a maximum of 500 words describing the design rationale, which question(s) your design would help answer and if/how your design may improve upon existing examples.
By **design rationale** we mean: the process and principles followed in choosing the specific visualization. You should provide a rigorous rationale for your design decisions, e.g. visual encodings used and why they are appropriate for the data. These decisions include the choice of visualization type, size, colour, scale, mark and channels and other visual elements, as well as the use of sorting or other data transformations. Consider how these decisions facilitate analysis and/or communication.

[9 marks]

Note: In this part we are only asking to design possible visual layouts not to implement them. If you are hand-drawing your designs take pictures and add them to your document as figures. If you are using a tool to develop your designs save them as images and add them to your document.

## Part 3. Implementation [30 marks]

Out of the visualizations proposed in **Part 2** implement one in D3 as a webapp. You shall use either the data provided or other data sources you foraged yourself.

Your visualization shall support answering one of the research questions, therefore:

- it shall be accompanied by a short description of how data are being processed (and acknowledgement of your data source(s)). [10 marks]
- it can include either a composition of linked/related simple visual layouts or a more sophisticated single visual layout. [10 marks]
- it shall allow some level of user interaction. [10 marks]

Note: You are allowed to use D3 Example code available on the web as long as it is adapted to your data and you explicitly acknowledge the source of the original code.
The implementation is to be done using D3, visualizations done with other software and embedded in an html page will not be accepted.