MATH 70096 - Exploratory Data Analysis & Visualisation

Spring 2022 - Assessed Coursework 3

Deadline: 8 March 2022, 23:59 (UK time)

Each group should submit a PDF report and the source file for an accompanying R Markdown notebook. Each individual should also submit an MP4 file containing their presentation. These files should all be submitted, via the Imperial College Blackboard VLE, by the deadline stated above; the MP4 file should be uploaded to the Panopto folder linked on the Blackboard submission page. Further submission guidelines are included below; read these carefully.

Each individual will receive a mark out of 20: this will comprise a group mark out of 15, and an individual mark out of 5.

This coursework counts for 15% of your total mark for EDAV.

Plagiarism: Your submission should be *your own work*. Note that software tools are used for plagiarism detection.

Question

Suppose you have been hired as consultants for a small energy provider in the US, and that you have been asked to provide an exploratory analysis of monthly residential energy consumption in the US, since January 2000. You are provided with a multivariate time series dataset, detailing (US) residential energy consumption patterns, from 1949 onwards. This data is contained in the accompanying file, `MER_ResEnergyConsumption.csv`, and was downloaded from Table 2.2 of the US Energy Information Administration's online data browser (https://www.eia.gov/totalenergy/data/browser/).

Your boss (who is also a data science consultant) will be responsible for discussing the results of your analysis with the client. You are required to produce a report for your boss, which provides:

- univariate EDA and visualisation for one of the time series variables in the dataset;
- multivariate EDA and visualisation for at least two of the time series variables in the dataset.

For both of these analyses, you are free to use any time series variable(s) from within the provided dataset, and your analysis should be interpreted in the context of the data. Your analysis should also be communicated using well-designed data visualisation and any visualisation design decisions that are made to improve the communication of the data should be briefly explained.

You are also required to provide the (.Rmd) source file for an accompanying R Markdown notebook, in which the analyses are implemented, and visualisations constructed.

Further Instructions / Guidelines

This is a group assessment; you have been placed into groups of 3 or 4. You are expected to contribute equally to the group analyses and report preparation; it is up to you to coordinate group meetings and arrange division of work.

I have not given a guideline for the amount of time you need to spend on this. The time you devote to this is something that you should discuss in your groups. I have, however, given a limit to the length of the report, which you might use to guide the required amount of work.

Each group should submit one PDF report, detailing the analyses, the motivation for these analyses and the corresponding results, and including a selection of visualisations to aid the communication of these analyses. You should also provide the source file for an accompanying R Markdown notebook. ***The group output will be marked out of 15.***

- The (PDF) report should be well-formatted (with correct mathematical typesetting, e.g. using LaTeX or MS Word), concise and no more than 1200 words (approximately two A4 pages of 11pt text); the total report (*including* visualisations) should be no more than four A4 pages in length.
- Visualisations should be chosen judiciously – i.e. don't just plot everything you can think of! Wherever possible, visualisations should also be constructed using the `ggplot2` library in R.
- The group (PDF) report should not simply be the knitted version of the .Rmd notebook - the report should be a more concise communication of the analyses performed, the motivation for these analyses, and their results.
- The R markdown document should reproduce the analyses and visualisations provided in the report; this notebook may contain supplementary material (e.g. further visualisations), but will not necessarily contribute to the final mark.
- The .Rmd file should list its dependencies at the top of the document, and 'knit' successfully when each of these are installed in the appropriate location; you do not need to submit the knitted PDF/HTML.

In addition, each individual should record a short video presentation, and submit this as an MP4 file. ***The individual presentations will be marked out of 5.***

- Each individual should record a 2-minute presentation on one of the following three aspects of the group project:
  - univariate EDA
  - multivariate EDA
  - visualisation
- Taken together, the 3 (or 4) individual presentations should cover all three of the above aspects of the project – i.e. in groups of 3, each aspect should be presented by exactly one group member; in groups of 4, one aspect should be covered twice.
- Each presentation should be accompanied by a small number of well-chosen slides; the presentation can be delivered either with or without the camera on – the choice is entirely yours.

- One option might be to have a group meeting in Teams, with the meeting 'organiser' recording each presentation in turn. Recordings are then automatically shared in the chat, and can subsequently be downloaded as MP4 files from MS Sharepoint.
- Alternatively, you could make use of the recording features in MS Powerpoint to narrate each slide, and save the resulting narrated presentation as an MP4 file. Further guidance on how to do this is included on the Blackboard submission page.
- I recommend that you prepare, record and upload your presentations well in advance of the deadline, in case you encounter technical difficulties and (in the worst case) need to re-record.
- Each MP4 file should contain one presentation only (i.e. if you're recording in a group Teams meeting, make sure you turn the recording off/on between each presentation!)
- Your MP4 file should be uploaded to the required Panopto folder by the deadline; the filename should use the following naming convention:

  `{your surname}_{your CID}_Assessment3.mp4`,

  e.g. `Martin_00123456_Assessment3.mp4`. Instructions for accessing and uploading to Panopto are included on Blackboard.