Faculty of
Information
Technology

FIT3152 Data analytics – 2024: Assignment 1

Your task	• Analyse the country level predictors of pro-social behaviours to reduce the spread of COVID-19 during the early stages of the pandemic. • This is an individual assignment.
Value	• This assignment is worth **25%** of your total marks for the unit. • It has 40 marks in total.
Suggested Length	• 8 – 10 A4 pages (for your report) + extra pages as appendix (for your R script, clustering table, and Generative AI statement if required). • Font size 11 or 12pt, single spacing.
Due Date	**11.55pm Monday 15th April 2024**
Submission	• Submit a single PDF file **and** single video file on Moodle. • Note that submission of a video report is a hurdle requirement. • Use the naming convention: *FirstnameSecondnameID.{pdf, mp4, mov etc.}* • Turnitin will be used for similarity checking of all written submissions.
Generative AI Use	• In this assessment, you can use generative artificial intelligence (AI) in order to <u>search for R functions and examples to perform tasks that you specify</u> only. Any use of generative AI must be appropriately acknowledged ([see Learn HQ](#)).
Late Penalties	• 10% (4 mark) deduction per calendar day for up to one week. • Submissions more than 7 calendar days after the due date will receive a mark of zero (0) and no assessment feedback will be provided.

Instructions

Address each of the research questions below and report the results of your analysis and your interpretation of those results.

You are expected to include at least one high quality multivariate graphic summarising key results. You may also include other simpler graphs and tables. Report any assumptions you've made in modelling and include your R code as an appendix. Your R code must be machine readable text as the university requires all student submissions to be processed by plagiarism detection software. You must include a declaration on the use of Generative AI at the beginning of your written report and if you used Generative AI, a statement on how it was used, as an appendix.

There are two options for compiling your written report:
(1) You can create your report using any word processor with your R code pasted in as machine-readable text as an appendix, and save as a pdf, or
(2) As an R Markup document that contains the R code with the discussion/text interleaved. Render this as an HTML file and save as a pdf.

Your video report should be less than 100MB in size. You may need to reduce the resolution of your original recording to achieve this. Use a standard file format such as .mp4, or mov for submission.

Software

It is expected that you will use R for your data analysis and graphics and tables. You are free to use any R packages you need but must document these in your report and include in your R code. You may use other software, such as Excel, to create the table of clustering data for Question 3(a).

Use of Generative AI

In this assessment, you can use generative artificial intelligence (AI) in order to <u>search for R functions and examples to perform tasks that you specify</u> only. Any use of generative AI must be appropriately acknowledged (<u>see Learn HQ</u>).

If you do use Generative AI for your assignment then you must include the statement "Generative AI was used in this assignment." In the introductory/first paragraph of your report. You must also include the following information as an appendix in your report: (1) the technology you used (e.g. ChatGPT), (2 the information that was generated (e.g. R code fragments), (3) the prompts used (i.e. the questions you asked), and (4) how the output was used in your work.

If you did not use generative AI in your assignment, then include the statement "Generative AI was not used in this assignment." In the introductory/first paragraph of your report.

Questions

During the early stages of the COVID-19 pandemic, researchers surveyed participants around the globe. A baseline study was conducted with the aim of identifying the most important predictors of pro-social COVID-19 behaviours, that is, actions that would reduce the spread of the virus. You can read a more detailed description of the research and results in Van Lissa (2022), see references.

The aim of this assignment is to understand country-level differences in predictors of pro-social behaviours, reported by participants as: "<u>I am willing to:</u>
- <u>help others</u> who suffer from coronavirus." **(c19ProSo01)**
- <u>make donations</u> to help others that suffer from coronavirus." **(c19ProSo02)**
- <u>protect vulnerable groups</u> from coronavirus even at my own expense." **(c19ProSo03)**
- <u>make personal sacrifices</u> to prevent the spread of coronavirus." **(c19ProSo04)**

Your task is to analyse the baseline survey data overall, with a focus on the country you have been assigned. You may make use of any additional data you require to answer the following questions.

1.   Descriptive analysis and pre-processing. (6 Marks)

(a) Describe the data overall, including things such as dimension, data types, distribution of numerical attributes, variety of non-numerical (text) attributes, missing values, and anything else of interest or relevance.

(b) Comment on any pre-processing or data manipulation required for the following analysis.

2.   Focus country vs all other countries as a group. (12 Marks)

(a) Identify your focus country from the accompanying list (FocusCountryByID.pdf). How do participant responses (attributes) for your focus country differ from the other countries in the survey as a group?

(b) How well do participant responses (attributes) predict pro-social attitudes `(c19ProSo01,2,3 and 4)` for your focus country? Which attributes seem to be the best predictors? Explain your reasoning.

(c) Repeat Question 2(b) for the other countries as a group. Which attributes are the strongest predictors? How do these attributes compare to those of your focus country?

3.   Focus country vs cluster of similar countries. (10 Marks)

(a) Using a collection of social, economic, health, political or other indicators from external sources, identify at least 5 countries in the baseline data that are similar to your focus country using clustering. Van Lissa (2022) refers to several indicators you might consider, among others. Some of these are listed in the references, but these are not exhaustive. State the indicators used and describe how you calculated/identified similar countries. Copy and paste the table of values you used for your clustering into your report as an Appendix.

(b) How well do participant responses (attributes) predict pro-social attitudes `(c19ProSo01,2,3 and 4)` for this cluster of similar countries? Which attributes are the strongest predictors? How do these attributes compare to those of your focus country? Comment on the similarity and/or difference between your results for this question and Question 2(c). That is, does the group of all other countries 2(c), or the cluster of similar countries 3(b) give a better match to the important attributes for predicting pro-social attitudes in your focus country? Discuss.

4.   Video Presentation: (Submission Hurdle and 4 Marks)

Record a short presentation using your smart phone, Zoom, or similar method. Your presentation should be approximately 5 minutes in length and summarise your main findings for Sections 1 – 3, as well as describing how you conducted your research and any assumptions made. Pay particular emphasis to your results in Questions 2(c) and 3(b)

5   Overall considerations (8 Marks)

This includes: the quality and clarity of your reasoning and assumptions; the strength of support for your findings; the quality of your writing in general and communication of results; the quality of your graphics throughout, including at least one high-quality multivariate graphic; the quality of your R coding.

Data

The data for this assignment is a reduced version of that collected for the PsyCorona baseline study, Van Lissa et al. (2022). The filename is "PsyCoronaBaselineExtract.csv". The data includes ordinal data coded on a numerical scale. For this assignment assume it is reasonable to treat these responses as numerical.

Create your individual data as follows:

```
rm(list = ls())
set.seed(12345678) # XXXXXXXX = your student ID
cvbase = read.csv("PsyCoronaBaselineExtract.csv")
cvbase <- cvbase[sample(nrow(cvbase), 40000), ] # 40000 rows
```

Locate your focus country using the accompanying document FocusCountryByID.pdf.

Selected references and web links

C. J. Van Lissa, et al., (2022) Using machine learning to identify important predictors of COVID-19 infection prevention behaviors during the early phase of the pandemic. Patterns 3, 100482.
https://doi.org/10.1016/j.patter.2022.100482

The World Bank Data Collections (and Governance Indicators)
https://datacatalog.worldbank.org/collections
http://info.worldbank.org/governance/wgi/

Organisation for Economic Co-operation and Development (OECD)Data
https://data.oecd.org/

Global Health Security Index: Reports and Data
https://www.ghsindex.org/report-model/

World Health Organization
https://www.who.int/

Data fields and brief descriptor (note AD = Agree/Disagree). See PsyCorona Codebook Extract for full description.)

Concept	Variable Name	Label
Employment Status	employstatus_1	Which best describes your employment status during the last week (multiple may apply)?-Employed, working 1-24 hours per week
	employstatus_2	… Employed, working 24-39 hours per week
	employstatus_3	… Employed, working 40 or more hours per week
	employstatus_4	… Not employed, looking for work
	employstatus_5	… Not employed, not looking for work
	employstatus_6	… Homemaker
	employstatus_7	… Retired
	employstatus_8	… Disabled, not able to work
	employstatus_9	… Student
	employstatus_10	… Volunteering
Isolation offline	isoFriends_inPerson	In the past 7 days, how much social contact have you had with friends or relatives who live outside your household? (1:7)
	isoOthPpl_inPerson	In the past 7 days, how much social contact have you had with people in general who live outside your household? (1:7)
Isolation online	isoFriends_online	In the past 7 days, how much online (video or voice) contact have you had with friends or relatives who live outside your household? (1:7).
	isoOthPpl_online	In the past 7 days, how much online (video or voice) contact have you had with people in general who live outside your household? (1:7).
Loneliness	lone01	During the past week, did you feel lonely?
	lone02	During the past week, did you feel isolated from others?
	lone03	During the past week, did you feel left out?
Life Satisfaction	happy	In general, how happy would you say you are?
	lifeSat	In general, how satisfied are you with your life?
	MLQ	Agree or disagree: "My life has a clear sense of purpose."
Boredom	bor01	AD - "I wish time would go by faster."
	bor02	AD - "Time is moving very slowly."
	bor03	AD - "I feel in control of my time."
Conspiracy	consp01	AD - "I think that many very important things happen in the world, which the public is never informed about."
	consp02	AD - "I think that politicians usually do not tell us the true motives for their decisions."
	consp03	AD - "I think that government agencies closely monitor all citizens."
Rank Order Life	rankOrdLife_1	Rank order the following in terms of their importance in life in the cells below (A = Very

		important, F = Not very important) A:Beauty; B:Achievement; C:Victory; D:Friendship; E:Love; F:Empathy.
	rankOrdLife_2	
	rankOrdLife_3	
	rankOrdLife_4	
	rankOrdLife_5	
	rankOrdLife_6	
Corona Personal Behavior	c19perBeh01	AD - "To minimize my chances of getting coronavirus, I..."-...wash my hands more often."
	c19perBeh02	AD - "...avoid crowded spaces."
	c19perBeh03	AD - "...put myself in quarantine."
Corona RadicalAction	c19RCA01	AD - "I would sign a petition that supports..."-...mandatory vaccination once a vaccine has been developed for coronavirus."
	c19RCA02	AD - "...mandatory quarantine for those that have coronavirus and those that have been exposed to the virus."
	c19RCA03	AD - "...reporting people who are suspected to have coronavirus."
Corona Proximity	coronaClose_1	Do you personally know anyone who currently has coronavirus? (click all that apply)-Yes, myself
	coronaClose_2	Yes, a member of my family
	coronaClose_3	Yes, a close friend
	coronaClose_4	Yes, someone I know
	coronaClose_5	Yes, someone else
	coronaClose_6	No, I do not know anyone
Gender	gender	What is your gender?
Age	age	What is your age?
Education	edu	What is your highest level of education?
Country Self Report	coded_country	In which country do you currently live in?
Corona ProSocial Behavior	**c19ProSo01**	**AD - "I am willing to help others who suffer from coronavirus."**
	c19ProSo02	**AD - "I am willing to make donations to help others that suffer from coronavirus."**
	c19ProSo03	**AD - "I am willing to protect vulnerable groups from coronavirus even at my own expense."**
	c19ProSo04	**AD - "I am willing to make personal sacrifices to prevent the spread of coronavirus."**