

DATA471__A2_Part1

kakauchad_300212228

2022-08-12

DATA471 Assignment 2 Part 1

1. What is your working question?

The current working question is: *what have been the impacts of the New Zealand Government's decisions to control the spread of COVID on society.*

The question started as a general interest in the Police data and wanting to look for some patterns in the data. Initial thoughts were around looking at changes over time (rather than between groups or categories) to look for trends and we thought of what effects that COVID may create. We thought that there might be noticeable changes in things like family violence, alcohol-related offending as the Government introduced controls on movement. We also thought about how the social atmosphere has changed as the country (and the world) become more accustomed to COVID and the effective precautions.

From there we thought about looking for relationships between timing of government controls and offending, but also at other indicators (like well-being, employment/wages, and confidence in institutions etc.). We still need to refine the question because it's still a bit massive...

2. What are the SMART components of the project?

The SMART components of the project are set out below as our key outputs. We have also done an initial plan for the group portion of the project, which includes setting some specific outputs with their own SMART elements and have allocated initial tasks (dataset collection and initial pre-processing).

The key outputs are: 1. a nicely refined question. 2. a collection of nice clean datasets to work from. 3. a completed initial EDA. 4. a completed full EDA.

Output 1 S - We will have a refined question that incorporates the fundamental underlying objective and limits to the specific datasets of interest

M - one question, acceptable to the supervisor

A - is mostly done and will be influenced by the first phase of pre-processing datasets R - very relevant to the project, essential to make sure that we can focus activity ... because this is the point of the activity

T - final question due by end of August, revision dates: 16 August, 23 August

Output 2

S - To check for trends we need time-series data, with aggregation to the same basis. Key datasets are:

- NZ Government decisions and timeline (include date of decision and indicator of type)

- Family violence offending/victimisations (include region for geographical analysis, possibly by sex and/or age)

- Employment data (salary/wages, employment rate) (include region, possibly by sex, age) - Well-being data (include region, sex, age)

- Public confidence data (include region, sex, age)

M - presence of dataset - on a time basis (ideally monthly, no less than quarterly for periodic)
A - datasets are available, most have already been downloaded but not yet pre-processed to confirm suitability
R - will form the basis of the analysis, but also the limitations, so is fundamental to achieving the objective
T - currently overdue - extend to 16 August for report back to team.

Output 3

S - Initial EDA, as defined by DATA471 course requirements
M - must meet defined requirements
A - all tools and skills are held by project members and individuals
R - essential driver of the project
T - due 6 September 2022

Output 4

S - Initial EDA, as defined by DATA471 course requirements
M - must meet defined requirements
A - all tools and skills are held by project members and individuals
R - essential driver of the project
T - due Exam week 2022

3. How did you model the 5-Ws and an H of the project?

These questions were generated in discussion and are broken down into basic elements we can use to organise relevant tasks. Some of the questions were based around our area of interest, some around what we need to do and how we need to do it:

We used a google chrome extension to generate a mind map but it is a bit difficult to share - the extension purports support for collaboration through google drive, but it isn't very intuitive. It does support a number of mindmap formats including as a text file, but we are still figuring that out.

The mindmap can be accessed here: .mup format: https://github.com/ctkakau/DATA471_A2/blob/main/Impact%20of%20COVID.mup .mymind format: [https://github.com/ctkakau/DATA471_A2/blob/main/Impact%20of%20COVID%20\(1\).mymind](https://github.com/ctkakau/DATA471_A2/blob/main/Impact%20of%20COVID%20(1).mymind)
.txt format: https://github.com/ctkakau/DATA471_A2/blob/main/DATA471_A2_MM.txt

We are still coming to grips with the Kanban, the concept seems useful, but we are still trying to develop the workflow and responsibilities. The trello page has been setup, but not much uploaded while we have been still trying to figure out the direction of the project.

the trello page is accessible at: https://trello.com/invite/data471_group22/88ce03d29f07da416592b8a3c7942666

4. How did you model your team?

We have a small team of two people, from different backgrounds, but we seemed to get on ok. it has been a bit tough meeting in person, but we have had a couple of meetings to discuss direction and responsibilities. We are still in the initial data collection and review phase, focussing on separate datasets. Because our team is small, we will likely share responsibilities for many tasks, but in terms of capabilities and areas for development, my preference is to work with data that can also be relevant to my work in law enforcement (this is why I preferred the Police dataset); and I am interested in developing a wide repertoire of visualisation techniques for communicating messages from data to a broad audience base (which is why I'm keen to look at the effects of Government decisions in managing COVID because there was a real change in atmosphere over the course of the last two years).

The specific areas I am hoping to develop is working on disparate datasets (generally, data provided during academic have been manicured) and trying to bring that data together to answer my own questions. That means identifying potential data sources, refining the question based on what the data can tell, applying some of the statistical analysis tools and models to the data to determine the answers, and finally, communicating those findings by way of visualisations and reports.

5. Show and describe your project's timeline as a project.

We have a general plan, it isn't set in stone, and just covers things at a high level (e.g. main phases: collection, pre-processing, Initial EDA, with the addition of some weekly reporting). We are still deciding on the platform, but have produced an initial project plan using ProjectLibre. The project has initial timelines through to early October, but there is no contingency built into this plan. The current timeline is built on an estimate of the minimum amount of time to complete tasks to a reasonable standard and we assume there is likely to be over-runs, especially as multiple assessments, work-pressures and social engagements increase over time.

The project plan can be accessed here: https://github.com/ctkakau/DATA471_A2/blob/main/DATA471_EDA.pod

The project has around two weeks for question development and refinement and overlaps with initial data collection (Police, Stats NZ for official statistics; MoH and COVID websites for key dates - likely in written language format). The two biggest phases of the project will be the pre-processing stage (cleaning and transforming data for comparable analysis) and the initial exploration of the data to determine features and any trends within individual datasets and then between different groups of datasets. These pre-processing and initial exploration phases are expected to take two to three weeks. Once the initial analysis is done, preparation for submission should be fairly quick, as much of the analysis pre-processing and analysis will likely be done in markdown or notebook format.

The project will be refined at weekly intervals to identify any significant barriers or challenges, but team members are expected to contact several times per week to maintain cohesion.

This document and associated support files can be accessed at https://github.com/ctkakau/DATA471_A2