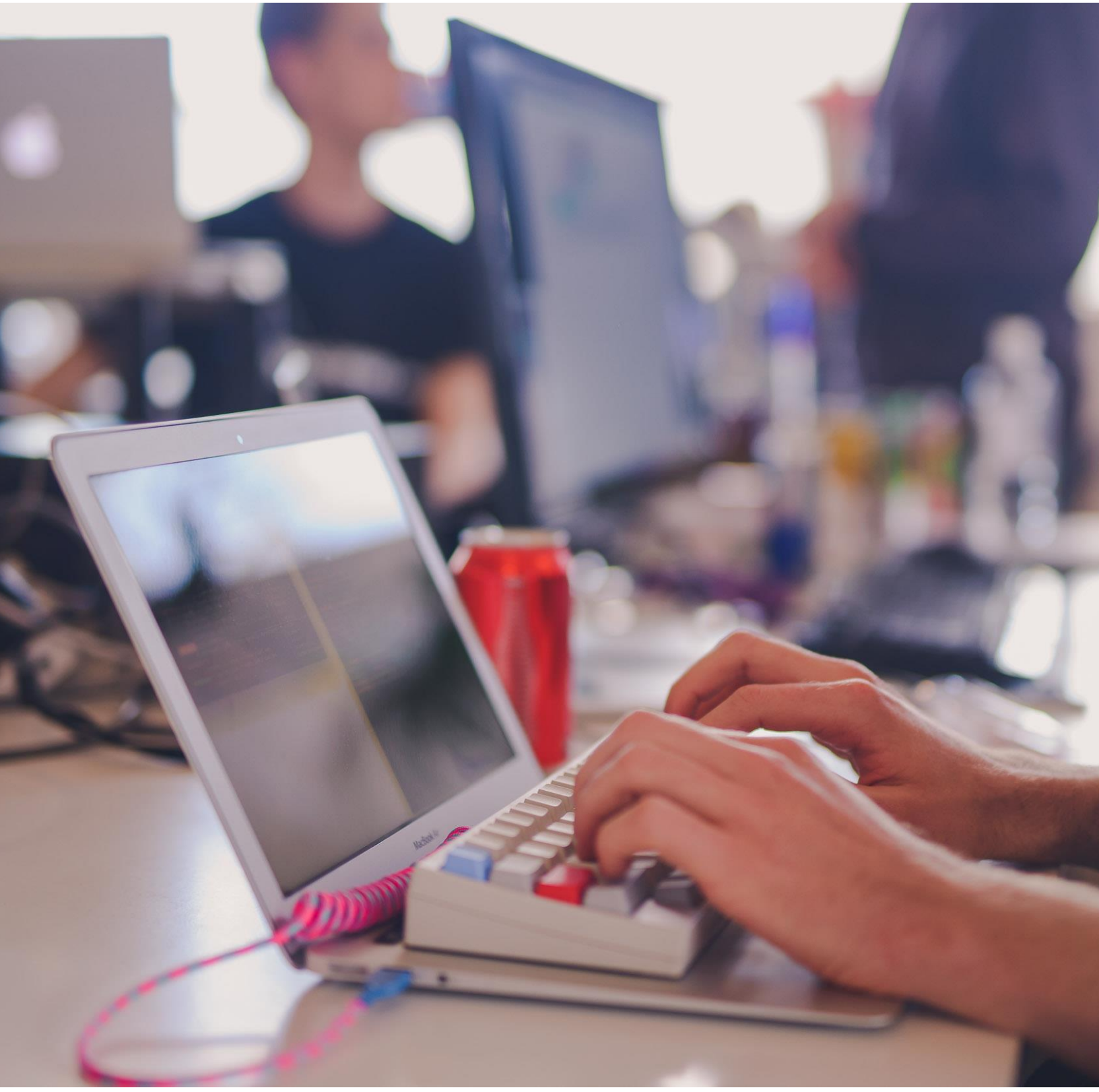

Quantify

the university datathon

Getting started guide

New to a datathon? Don't despair, Quantify is beginner friendly! We will be running 3 tutorials over the event – one about loading the data in and exploring it, one about building models and one about data visualisation.

To participate in these tutorials, you should make sure there are a few things installed on your computer first.

Quantify is an open competition – so you can use whatever tools you would like!

Communication

On the datathon weekend, and during the week leading up to the datathon weekend, we will be using Slack to communicate.

You can install Slack from here:

https://slack.com/

And then click here to get an invitation in your email: http://tinyurl.com/quantifyslack

Anaconda, Python and R

Please install Anaconda and Jupyter notebook, with both the R kernel and the Python kernel to run our demonstration scripts.

1. Download Anaconda with Python 3.6 version here: https://www.anaconda.com/download/
2. Download R from here: https://cran.ms.unimelb.edu.au/
3. Run the Anaconda prompt and type:
   `conda install -c r ipython-notebook r-irkernel`
4. Now launch notebook using the Anaconda Navigator, and you should have the option of starting both a new Python notebook or a new R notebook.

Cheatsheets

Please install Anaconda and Jupyter notebook, with both the R kernel and the Python kernel to run our demonstration scripts. Here are some cheat sheets to use for coding in both Python and R.

Python	R
Numpy	RStudio Cheatsheets https://www.rstudio.com/resources/cheatsheets/

https://www.datacamp.com/community/blog/python-numpy-cheat-sheet#gs.dScbsCQ	
Pandas https://www.datacamp.com/community/blog/pandas-cheat-sheet-python#gs.QI1cg=4	ggplot2 https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
Scipy https://www.datacamp.com/community/blog/python-scipy-cheat-sheet#gs.286pSOA	Diplyr and Tidyr https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf
Matplotlib https://www.datacamp.com/community/blog/python-matplotlib-cheat-sheet#gs.glw1WFk	Leaflet https://github.com/rstudio/cheatsheets/blob/master/source/pdfs/leaflet%20cheat%20sheet.pdf
Scikit-Learn https://www.datacamp.com/community/blog/scikit-learn-cheat-sheet#gs.F1OSros	Shiny Cheat sheet https://shiny.rstudio.com/articles/cheatsheet.html

Process

Understand the data

High-level data exploration.	Look at the data, write down all the fields that are included in the data.	Read about how the data was collected and the background of the data.	Do a little bit of research about current issues already identified.	Timeline hint: Do this as soon as you have the data.

Define the problem statement

Use ours or make your own.	What is the problem? Who has the problem? Where does the problem occur?	When and how often does the problem occur? What causes the problem?	What is the problem impact?	Timeline hint: Get this done first thing in the morning on Saturday 30th of September.

Explore the data

Deep dive data exploration.	Load the data into your favourite tool.	Start plotting. Visualise the data in whichever way you want.	Start looking at statistical summaries. Start looking for patterns.

Brainstorm solutions

Work with your team to look at different solutions for the problem you have identified.	No ideas are bad ideas.	Brainstorm as many as you can and pick the best bits from each of them to include in your final solution.	Timeline hint: Have your solution identified by about 1 or 2pm on Saturday the 30th of September.

Build

Code, code code!	Test and code some more!	Always check that your solution matches your problem statement identified.	If you need some help with coding - ask around! Don't be scared to talk to the mentors walking around the floor.	Timeline hint: Have your solution almost finalised by 8pm on Saturday the 30th of September.

Present

Prepare your presentation.	Practice your presentation with your mentor on Sunday morning and get any feedback before the final judging.	Timeline hint: Start preparing your presentation while you are finalising your solution as soon as you get in on Sunday.

Making your slide deck for presentation

Please make sure that you've submitted a slide deck about your project before judging. This is due on Sunday, 12PM.

The slide deck at a minimum should:

1. Contain a title a description about the project as well as a screenshot of the project. This presentation should be no more than 10 slides long.
2. On the first page of the presentation, please include your team member names and email.

Please email your slide deck to club.mdss@monsu.org with the filename as <YOUR_TEAM_NAME>.pdf.

To get your entry judged, you need to prepare a 3-minute presentation using Google slides, Prezi, or PowerPoint.

You will need to demonstrate your implementation and your code separately to the judges. Please do not include any code in your final presentation.

Prezi: https://prezi.com/

Google Slides: https://www.google.com.au/slides/about/

Final presentation checklist

Check	Task
	Has the problem statement been checked by a mentor?
	Does the presentation slide deck include the problem statement?
	Does the presentation slide deck include a summary of how you decided on your solution?
	Does the presentation slide deck include screenshots of your solution?
	Have you practiced your presentation to make sure its under 3 minutes?
	Have you prepared your code in a manner so that it can be viewed by the judges?