Projects

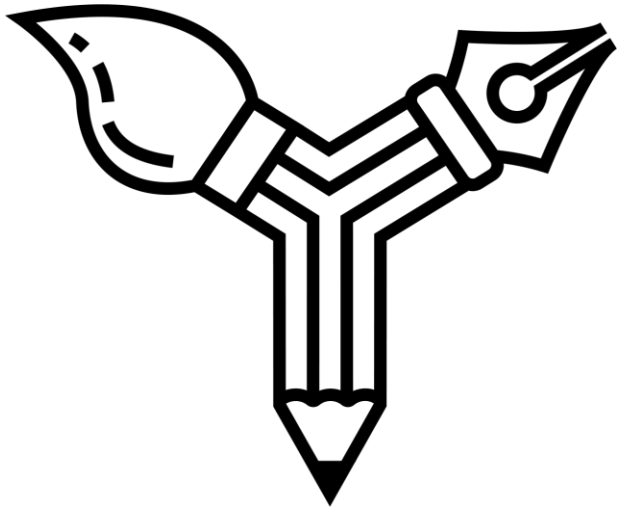
Data Boot Camp

Lesson 7.1

Project Week! (This Week)

Day 1:

- Form groups
- Outline project ideas
- Initial data exploration
- Begin research of datasets
- Submit project proposal for approval



Day 2:

- Hardcore development

Day 3:

- Hardcore development

Project Week! (Next Week)

Day 4:

Hardcore development

Day 5:

Hardcore development

Presentation prep

Day 6:

Presentations

Time to divide into teams!

Teams

Group 1	Group 2	Group 3	Group 4	Group 5	Group6	Group7
Matt Vella	Helen Aminalolama	Fern Bradder	Alvin Lucero	Jack Pan	Rajesh Ravindranathan Nair	Dale Currigan
Josh Cullen	Stephanie Salvona	Srichakra Vegunta	Alysha Snowden	Simona Suko	Amin Muhammad Anwer Ali	Jimmy Faccioli
Hideaki Kaneko	Heriawan Muhamad	Petra Moyle	Rebecca Gould	Ahmad Makintha Brany	Narisara Kantanong	Adam Lever
	Samantha Van Wyngaarden	Pravakar Neupane	Ernest Bondi	Ray Camo	Fang Xuan Foo	Vijay Kasina

Project Requirements

Development Requirements

Use Pandas to clean and format your dataset(s).

- Use Pandas to clean and format your dataset(s).
- Create a Jupyter Notebook **describing the data exploration and cleanup** process.
- Create a Jupyter Notebook **illustrating the final data analysis**.
- Use Matplotlib to create a total of 6–8 visualisations of your data (ideally, at least 2 per ‘question’ you ask of your data).
- Save PNG images of your visualisations to distribute to the class and instructional team, and for inclusion in your presentation.
- (Optional) Use at least one API, if you can find an API with data pertinent to your primary research questions.
- Create a write-up summarising your major findings. This should include a heading for each ‘question’ you asked of your data and a short description of your findings, and any relevant plots.

Presentation Requirements

You will also be responsible for preparing a formal 10-minute presentation that covers:

- Questions you found interesting and what motivated you to answer them
- Where and how you found the data you used to answer these questions
- The data exploration and cleanup process (accompanied by your Jupyter Notebook)
- The analysis process (accompanied by your Jupyter Notebook)
- Your conclusions including a numerical summary and visualisations of the summary
- The implications of your findings: what do your findings mean?

Marking rubric

SECTION	WEIGHT	REQUIREMENT
Presentation	2.5%	Describe the core message or hypothesis for your project.
	2.5%	Describe the questions you and your group found interesting, and what motivated you to answer them
	2.5%	Summarize where and how you found the data you used to answer these questions
	5.0%	Describe the data exploration and cleanup process (accompanied by your Jupyter Notebook)
	5.0%	Describe the analysis process (accompanied by your Jupyter Notebook)
	10.0%	Summarize your conclusions. This should include a numerical summary (i.e., what data did your analysis yield), as well as visualizations of that summary (plots of the final analysis data)
	10.0%	Discuss the implications of your findings. This is where you get to have an open-ended discussion about what your findings "mean".
	12.5%	Tell a good story! Storytelling through data analysis is no different than in literature. Find your narrative and use your analysis and visualization skills to highlight conflict and resolution in your data.
Technical	5.0%	Use Pandas to clean and format your data set(s)
	15.0%	Perform data preparation
	10.0%	Create a Jupyter Notebook describing the data exploration and cleanup process
	17.50%	Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)
	2.5%	Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation
	Bonus 5%	Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions

Suggested Data Sources

Suggestions for Data Sources

Feel free to ask us (the instructional staff) for input, but our general advice is to stick to data sources that:

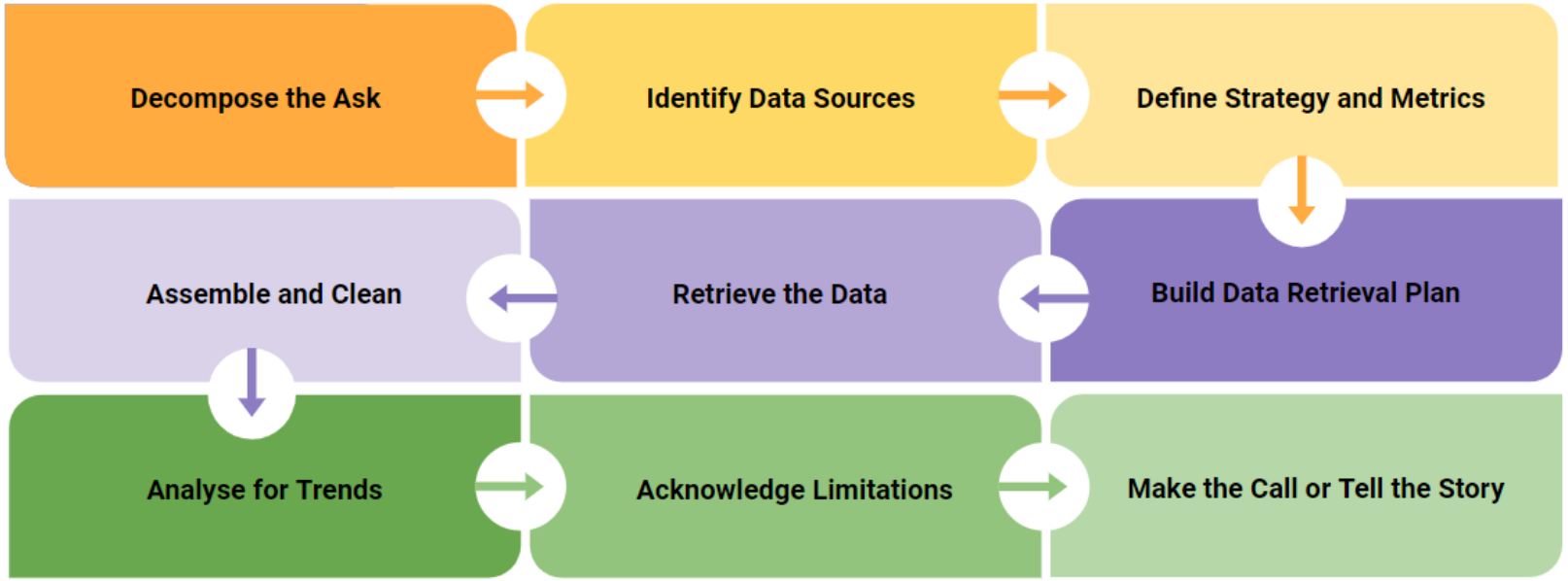
- Are sufficiently large
- Have a consistent format
- Ideally, contain more data than needed
- Are well-documented

Suggested data sources

Description	URL
Australian Government Open Source Datasets	https://data.gov.au/
Kaggle Open Source Datasets	https://www.kaggle.com/datasets
Public APIs	https://github.com/public-apis/public-apis
Australian Bureau of Statistics	https://www.abs.gov.au/
World Bank Open Data	https://data.worldbank.org/
World Health Organization Data	https://www.who.int/data/gho/
FiveThirtyEight (news agency open data)	https://data.fivethirtyeight.com/
Google dataset search	https://toolbox.google.com/datasetsearch

Approach

Suggested approach



Example Project Ideas

Private Investigator

01

Use aggregate crime data from different police precincts in a city to uncover patterns in criminal activity.

02

Most crime in NYC takes place in the summer.

Can you uncover similar patterns in your city?

03

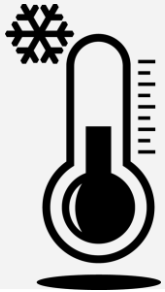
What do your results suggest about how police should plan their patrols?

What do your results suggest about how best to distribute law enforcement resources over the calendar year?

Uber Rides and Weather

01

No one likes to walk in subzero temperatures *or* scorching heat. Do people use Uber more when the weather is uncomfortable?



02

Using Uber ride data from Kaggle and data from a weather API, find out if people take Uber more during summer and winter, and if there are relationships between daily temperature and ride frequency.

03

What do the results tell you about surge pricing strategies and commuter habits?

Bullying and Crime Rates

01

Bullying and violent crime seem like they should be related. Can we find a correlation between frequency of bullying and rates of violent crime?

02

Using aihw.gov.au's data on bullying and data from police districts of your choice, investigate relationships between bullying and violent crime frequency, and location (postal code, city, etc.).

03

Are these two activities correlated?

What do the results suggest about society and public policy?

Today’s Focus

By the End of Today's Class:

- Brainstorm possible project ideas
- Begin data research
- Write a description of the scope of your research

Create a short, one-page project proposal that covers the following:

- Project title
- Team members
- Project description/outline
- Research questions to answer
- Datasets to be used
- Rough breakdown of tasks

Questions?


TRADING PLATFORM
ETH USD LTC
190.34 +3.44 +1.88%
199.31 +0.31 +0.05% 19:55 Jun 01.06
SELL
BUY
Sales
UP 23%
DOWN 14%
Month
190.34