DATA IN CONTEXT

This resource is to support the Data in Context virtual workshop session, giving you an overview of the key points and ideas discussed during the session, along with the relevant links for your own investigation..

Work your way through the sections, thinking about the ideas and concepts raised in each. After going through this resource, arrange a call with your mentor to discuss your data proposal and your understanding of the topic.

DATA BREACHES

Perhaps the most publicly-visible examples of the importance of careful data management are the times when it goes wrong: data breaches expose huge amounts of personal & financial information to the public, causing significant harm.

Explore the [World's Biggest Data Breaches visualisation](). Look for interesting ones, industry-relevant ones, ones you personally were affected by. Are there patterns to the problems or the consequences?

ETHICS

Data science is a powerful tool, and like powerful things, it's important to consider the ethical dimension. As you learn new techniques and apply them to your work, you'll need to have an awareness of the potential consequences and trade-offs you're making; whether or not you **should** do something is just as important as whether or not you **can**.

Read the following articles, thinking about the ethical considerations and implications discussed.

- [The ethical problem of the ethical dilemma of driverless cars]()
- [Can you sue a robocar?]()
- [Rise of Autonomous Military Drones Raises Number of Ethical and Legal Questions]()
- [China 'social credit': Beijing sets up huge system]()

With other powerful and potentially life-threatening disciplines, such as [medicine](), society has developed ethical codes and sets of principles to ensure that harm is minimised. Although many standards have been proposed (such as the [Oxford-Munich Code of Conduct]()), there is currently no single, agreed upon code of ethics for data scientists.

Research ethical codes for data scientists, looking at the various proposals available online; generate your own set of principles for ethical data science: what are the red lines that should not be crossed? What standards should data scientists hold themselves to?

DATA SOURCES

Before you can do anything with data science, you need to have data. Your organisation will have its own internal data storage & sources, but it's worth taking some time to explore some of the other options; there are many public data sources that could be relevant to problems you want to solve.

The following sites host a range of datasets; explore the various options and look for data that could be useful to you.

- [Kaggle](#)
- [UK Datastore](#)
- [Google Datasearch](#)
- [Decoded Datastore](#)
- [Quandl](#)

All the datastores above have one thing in common: they focus primarily on static data: data that is not (or only occasionally) updated with new information. Static data is absolutely useful, and you'll work with it a lot throughout the Academy, but the ideal is **live** data: data that is up-to-date, constantly increasing in volume, and relevant to today.

Live data can be accessed through [API](#) calls, and (just as with static data) there are many open sources of live data. Spend some time investigating the data sources below, looking for data that interests you and links to your day-to-day role.

- [GitHub list of public APIs](#)
- [RapidAPI](#)
- [ProgrammableWeb](#)

PROJECTS

Throughout the course, you'll submit several projects; these serve both as an opportunity to develop your skills, and as a chance to apply your learning in practice, exploring a business-relevant problem and working towards a solution.

Before beginning any actual analysis of data, there are several things that need to be established:

1. The project aim
2. The business requirements
3. The stakeholders for the project

The **project aim** is a simple, high-level statement of what the project is intended to accomplish: the end goal. This should be directly related to and support the aims of your organisation.

The business **requirements** are the targets or standards that the project must meet in order to be considered successful. We can divide them into two main categories: **functional** (what the project does) and **non-functional** (how well it does it)

The **stakeholders** of a project include any person or group who will be affected by the project, as well as any person or group who can/will affect the project. Each stakeholder will have their own separate requirements; although these should link to the business requirements, they are separate, and will generally be more concerned with **how** a project meets its objectives.

PROPOSAL

Think of a problem you or your organisation currently faces that could be solved with data. Create a brief proposal for your project, explaining what you want to do, and why it is necessary.

Your proposal should contain:
1. A clear aim
2. A set of specific requirements
3. A list of relevant stakeholders, and how they are involved
4. The sources of data you want to use
5. Consideration of ethical/security issues involved in your project

Your proposal does not need to be incredibly detailed; at this stage, you don't need to consider specific technologies or tools. This is an initial, brief proposal focused on a high-level understanding of the context of your problem.

FEEDBACK

Once you've completed the above activities, please fill out the feedback form for this workshop. Your feedback not only confirms your attendance at the session for compliance purposes, but helps us improve our workshops in the future.

Your workshop ID code is **5508**.