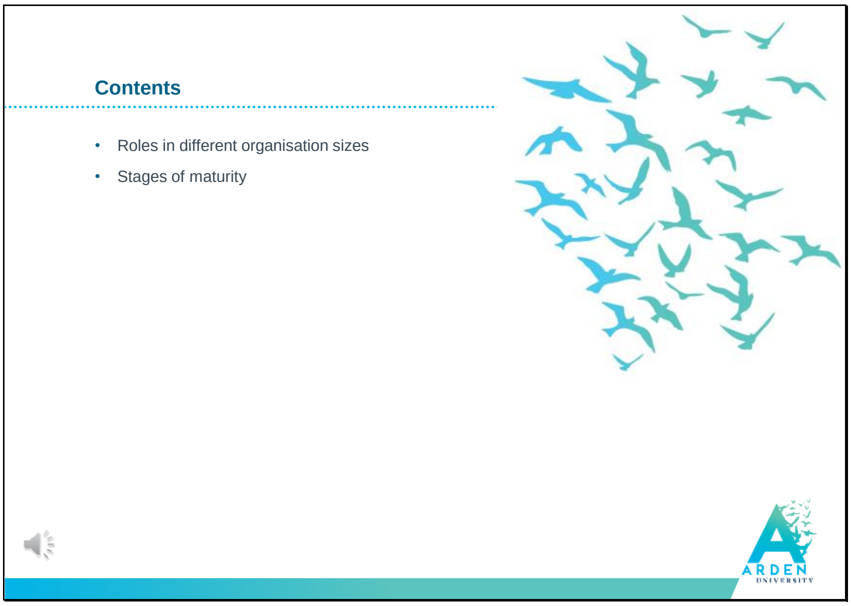
Slide 1


DATA ANALYTICS ROLES
AND MATURITY STAGES
ARDEN
UNIVERSITY

Slide 2



In this presentation, we will discuss two main topics:
First, we will discuss the different roles participating in a data analytics project, and the implications of the organisation’s size.
Second, we will discuss the stages of maturity an organisation can go through in terms of its ability to be data driven and use data analytics as an inherent part of the business.

Slide 3

Slide 4

It takes a village to raise a child

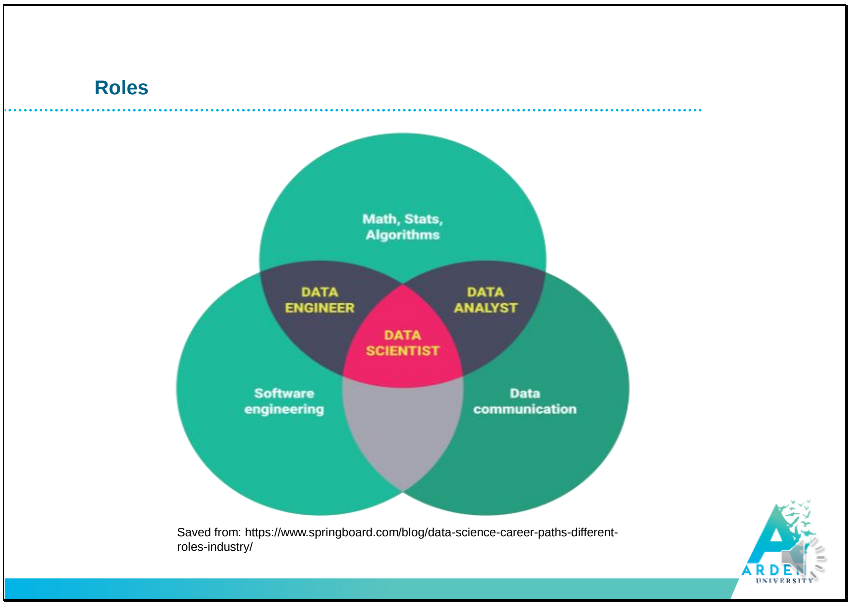
Phase	Time	Resources	Risks
Business understanding	1 week	All analysts	Economic change
Data understanding	3 weeks	All analysts	Data problems, technology problems
Data preparation	5 weeks	Data mining consultant, some database analyst time	Data problems, technology problems
Modeling	2 weeks	Data mining consultant, some database analyst time	Technology problems, inability to find adequate model
Evaluation	1 week	All analysts	Economic change, inability to implement results
Deployment	1 week	Data mining consultant, some database analyst time	Economic change, inability to implement results

Source: IBM
https://www.ibm.com/support/knowledgecenter/en/SS3RA7_17.1.0/modeler_crispdm_ddita/clementine/crisp_help/crisp_project_plan_ex.html

Throughout this module, we have been moving around different phases of a 'typical' data analytics project. Each phase requires a slightly different set of skills, perspectives and even interests.
If you remember, IBM's example project plan from Lesson 2 of this module, you'll probably recall that different roles are associated with different phases.
For example, the business understanding phase is likely to include a domain expert, the data understanding phase will probably include a data analyst, while the data preparation phase is very likely to be based on a person whose expertise is in data management and databases. If the business question is around predictions, the modelling phase is likely to be based on a machine learning specialist, where if it is about descriptive or inferential statistics, it might be using the help of a statistician, perhaps in addition to a data visualization consultant as an example.
The deployment phase will assumingly require a major involvement of someone who is very familiar with the users and their needs, an IT person and perhaps a domain expert again. Throughout all this iterative process, there is also the managerial role, who needs to oversee the process, make sure the work in different phases and roles is communicated and interacted smoothly and that risks are anticipated and handled in a timely manner.

It takes a village to raise a child, and the organisational reality is that it takes a very diverse set of people and roles to bring a data analytics project to a successful deployment.

Slide 5



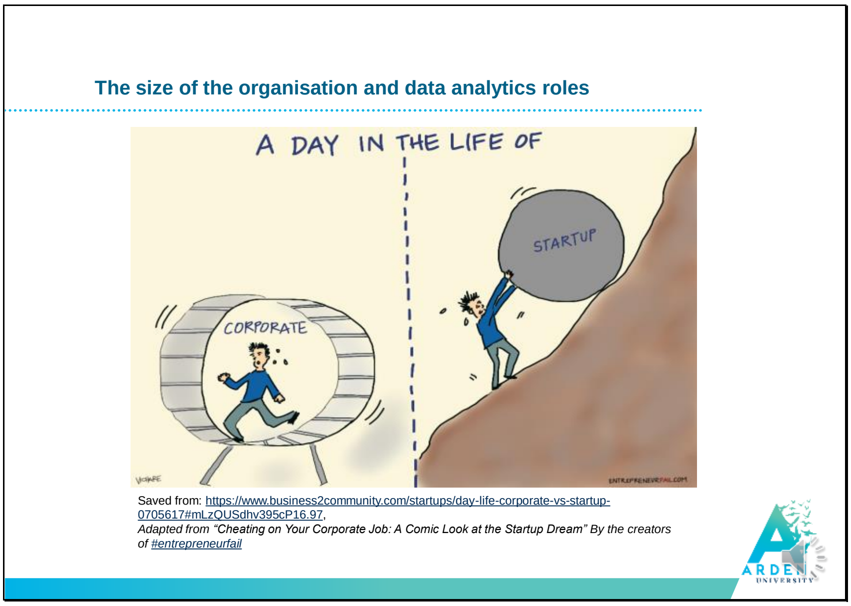
If you recall, in the first lesson of this module we discussed what data science is, and what a still evolving and interdisciplinary domain this is.
The three roles in this image are usually considered the main roles at the heart of data analytics projects.
As you can see, skills like software engineering, statistics and communication are shared with varying intensity.
Dependent on the scale and the aim of a data analytics project, these roles can be employed by the same person, or by a different people, coming from diverse background and perhaps even working in different organisations or departments.

A **Data Engineer** is usually more oriented towards systems and programming. The data engineer usually comes from the discipline of ICT, software engineering or computer science. They are best informed about the required infrastructure (e.g., databases, hardware, cloud solutions), software development at scale (in order to develop routines and scripts of extracting data, cleaning and integrating different datasets coming from different sources) and managing the technical considerations of the deployment solution.

A **Data Scientist** is also no stranger to these topics, but is more oriented towards the modelling and evaluation phases. Even when preparing and characterizing data, the data scientist will be more focused on analysis requirements more than on scalability and deployment considerations for example. A data scientist will often have a strong statistical or mathematical background, will have the experience in various modelling techniques and tools and will be best informed to adapt the right technique to the business question and available data.

A **Data Analyst** is kind of a ‘bridge role’. A data analyst should be familiar with both the technical, statistical, algorithmic language, as well as the business domain language and be able to translate between the two. The analyst is typically very central in the first phases, where requirements and risks should be clearly described. This role is focused on communicating business to the data world and vice versa. Obviously, it is very important that the analyst will be very familiar with the organisation.

Slide 6



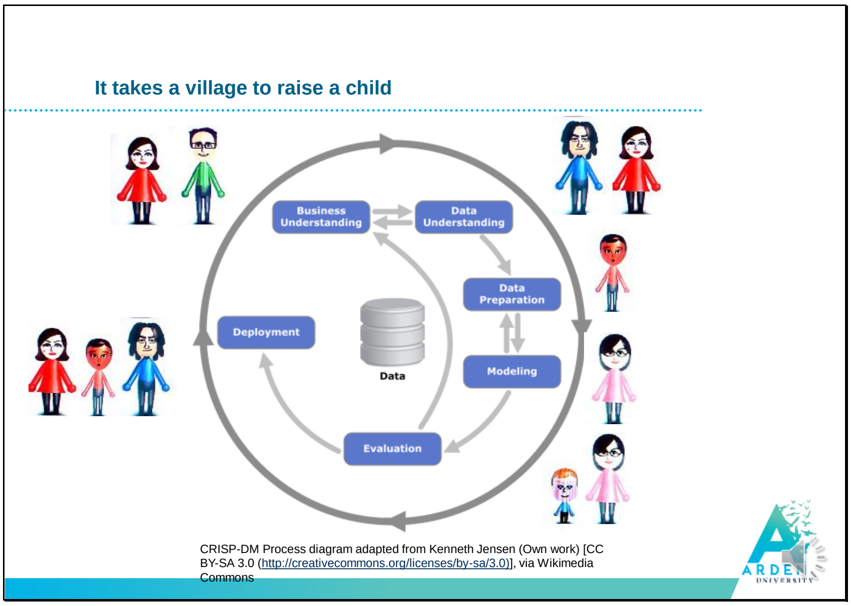
Whether a data analytics project requires all three roles or can settle with just one, it is pretty much dependent on the scale of the project and the size of the organisation.

In small organisations such as startups, or in the very first steps of establishing a data analytics team, we may be more focused on setting an infrastructure for data extraction and collection (e.g., setting up a data warehouse), since without data, there is no use in any other step. Thus, a small organisation may start with a person which is oriented towards data engineering, but still has some expertise as a data scientist and analyst.

A mid-size organisation, will hopefully be out of the way of infrastructure worries, but still won’t have the resources to employ a big data analytics team. In this case, again, multidisciplinarity of the person employed is key.
Alternatively, the organisation can consider employing data science external consultants or contractors.

Large organisations may be able to afford employing full teams more easily. This scaling up will require also adding a manager and a data analyst to help in coordinating between the data analytics team and different stakeholders in the organisation.

Slide 7

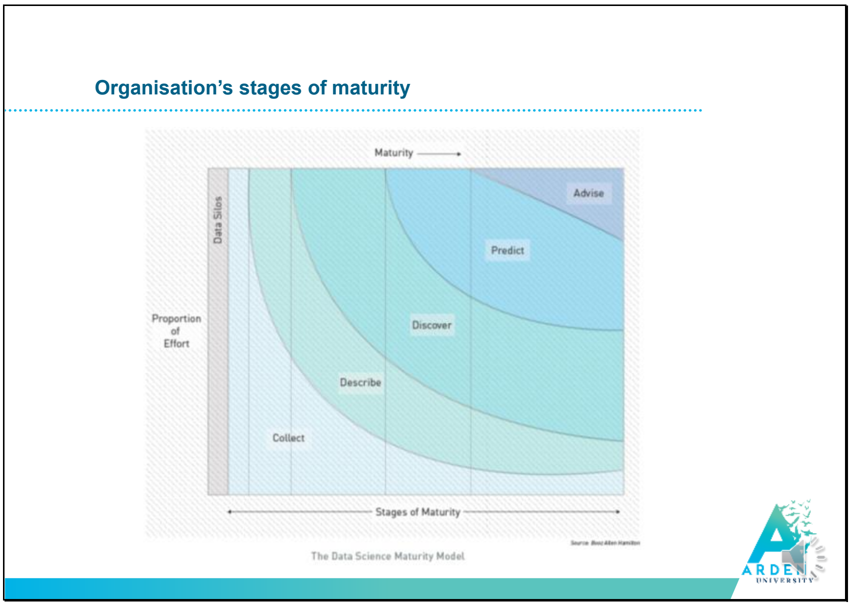


In reality, the data engineer, data scientist and data analyst can hardly be the only stakeholders of a successful data analytics project.
Collaboration and support from and to other stakeholders in the organisations are crucial, especially those of decision-makers and domain experts from within and outside the organisation.
Not less importantly, the role of a data analytics team in a data-oriented organisation does not end within the CRISP-DM cycle.
Rather, they should be involved in other matters of the organisation via consulting, collaborating, teaching and empowering others to use data.

Slide 8

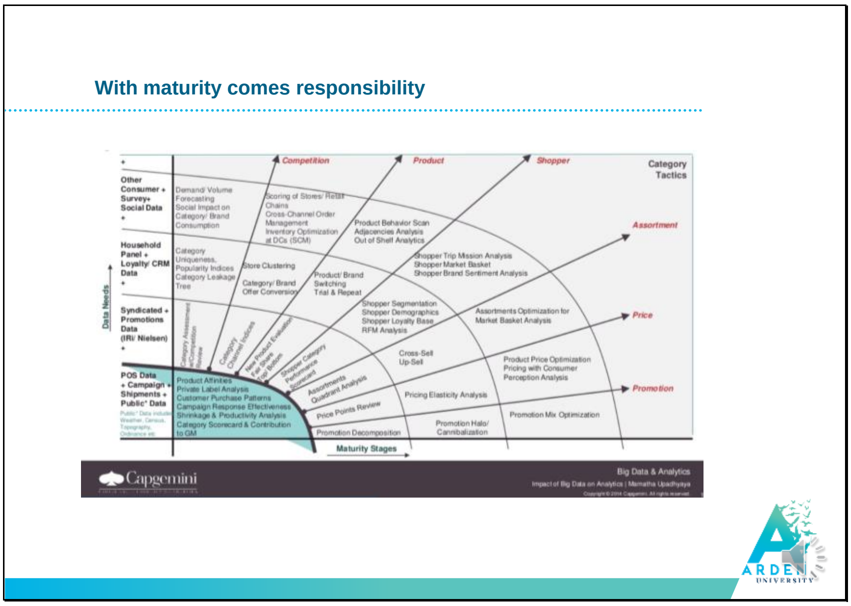
The CRISP-DM model merely provides a simplified view of a single data analytics project. Organisations will repeat this cycle with each new project. As time goes by, and the organisation becomes more experienced in this type of project, skills, data, tools and knowledge will become more and more available to more and more people. As these resources become more available, the organisation, becoming more mature, is able to shift its focus to new and advanced challenges.

Slide 9



Let us have a look at the Booz Allen Hamilton's Data Science Maturity model.
At the first maturity phase, an organisation will typically start from a position not minded to the need for collected data. Even if data had been collected, it would probably be scattered in data silos.
Thus, as an organisation is in its first steps of becoming data-oriented, will typically invest most of the time and budget on collecting data and thinking about canonical representations.
As an organisation becomes more mature in managing data analytics projects, it will be able to scale up from investing most resources in describing and collecting the data into more advanced data analytics applications such as running experiments, clustering and predicting. Data will become available and accessible to more people in the organisation, methodologies (e.g., Agile) and work processes will become a de-facto standard, collaboration around data will become more frequent, new and more innovative questions will be asked, execution will become faster, communicating results will be easier, the value of data analytics outputs will become more measurable and data analytics will become a common tool for decision-making and business understanding.
Data analytics projects, following the CRISP-DM model, or any other process model, could be carried out in any maturity stage. However, it will face a different focus on different challenges as shown in the maturity model.
In addition, higher management needs to be aware and make sure the organisation makes its maturity progress between one project and another.

Slide 10

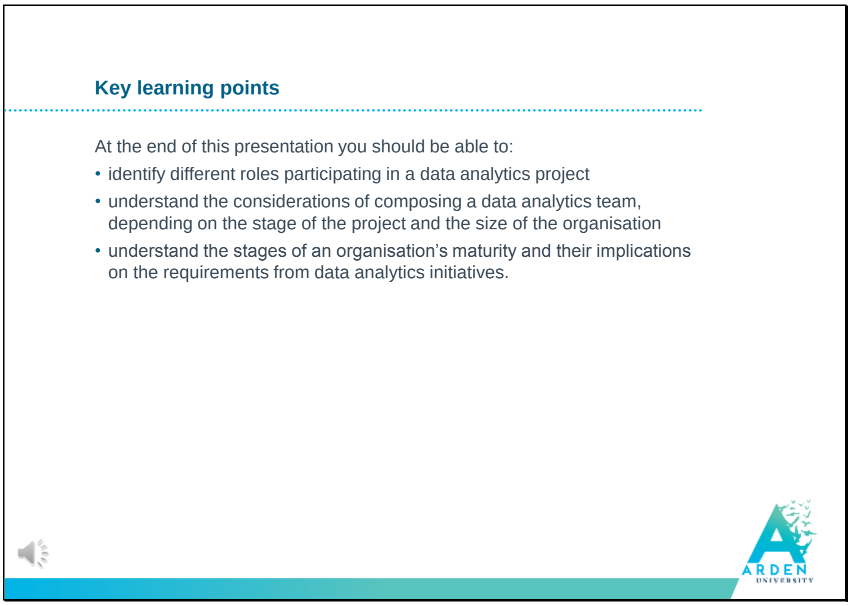


With maturity comes more available data. More innovation and more complexity is required from data analytics.

As complex as the challenge gets, we need more fine-grained data sources, and that in turn will require that we will adapt our analytics techniques, storage capacities and collection methods.

In addition, as the data we are collecting is more complex and fine grained, we need to be much more cautious about the sensitivity of the data, and be aware to all kinds of ethical considerations that we might not have needed to attend to in the first maturity phases. For example, if we go on to shopper segmentation, we will need to find demographic data, which is naturally very sensitive. As we go on to analyze the behaviour of our customers, we need to collect more data about their actions, and that might jeopardize their privacy.

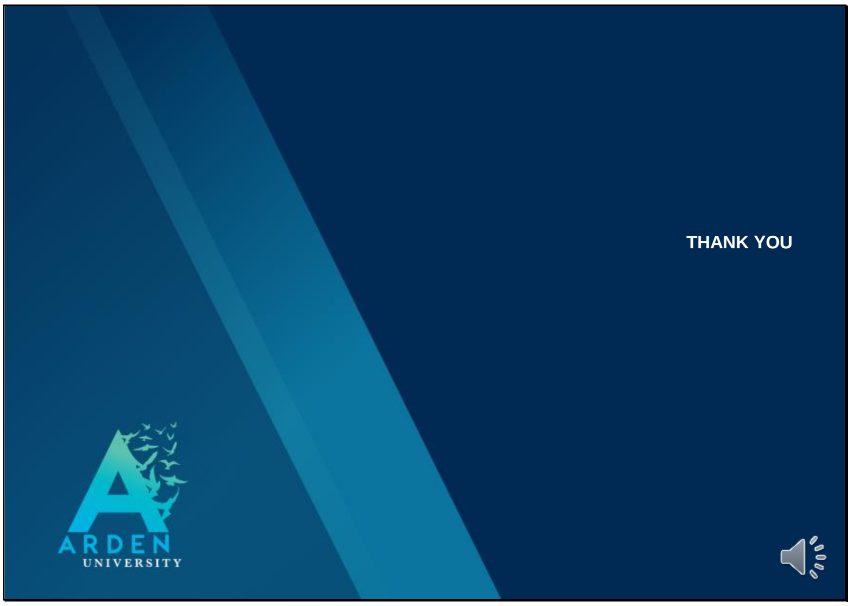
Slide 11



At the end of this presentation you should be able to:

- identify different roles participating in a data analytics project
- understand the considerations of composing a data analytics team, depending on the stage of the project and the size of the organisation
- understand the stages of an organisation’s maturity and their implications on the requirements from data analytics initiatives.

Slide 12



Thank you for listening.