Data Science Fundamentals

WOMEN IN DATA ACADEMY

TECH TALENT ACADEMY

Session Content

What is data science?

Data science lifecycles

Data science roles and responsibilities

Cloud Computing

What is Data Science?

Data has long been used in research environments to gather insight. For example, scientists would gather data manually to help them glean knowledge using the scientific method.

In modern terms, data science is the umbrella term used to describe the application of scientific methods to extract information from raw data using computing.

The field combines programming skills, business/domain expertise, machine learning and other tools and technologies.

Raw data is messy. Data science aims to construct three-dimensional pictures to extract meaning from the data.

Why is it so important?

The connected nature of the internet (The Internet of Things), has meant that we have never been able to capture data on the scale that we can today. Owing to this, data science is now used frequently in both research **and** commercial business environments. By using data, we can gather insights that would not ordinarily be available; these insights can then be used for key benefits, such as:

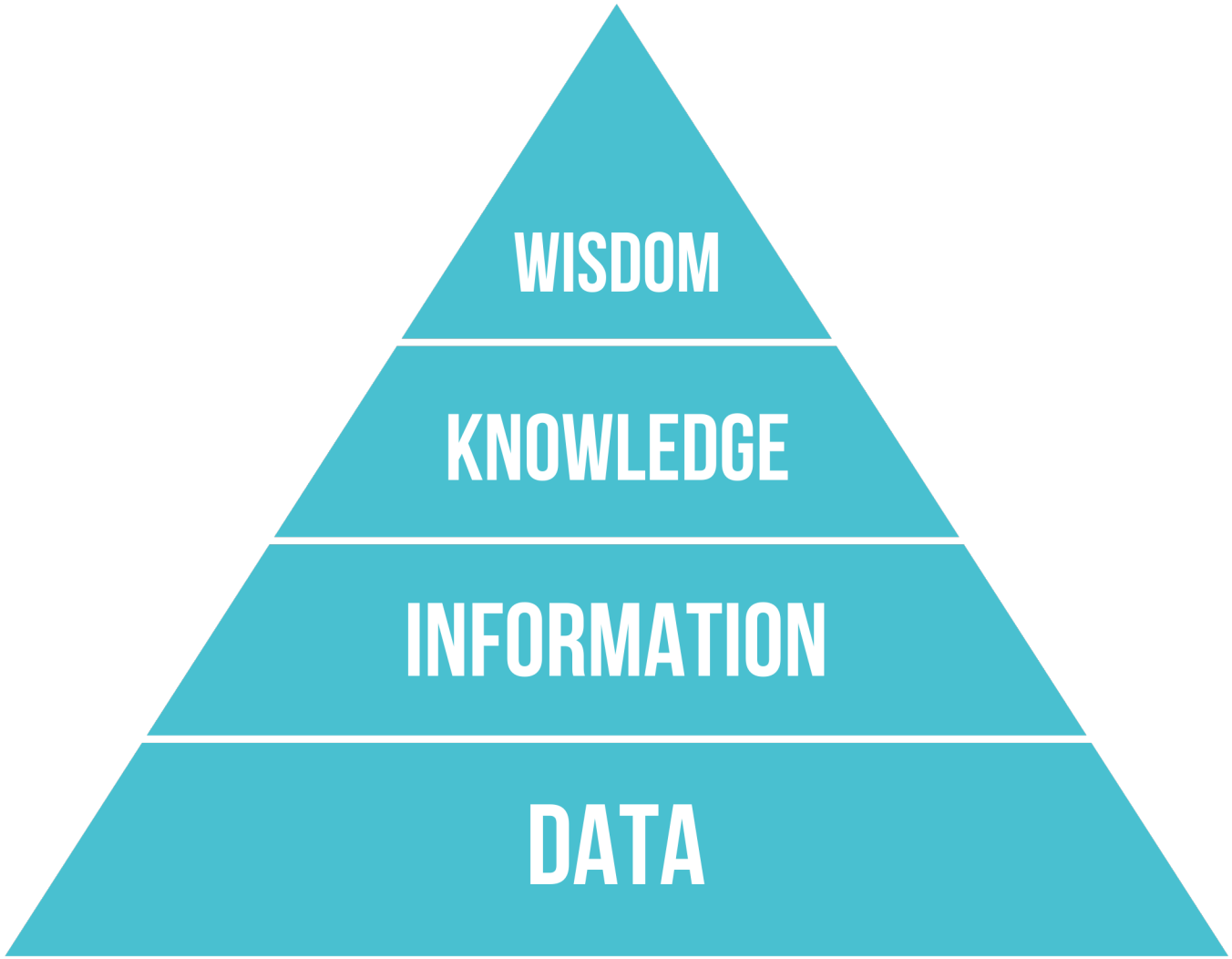
- Improving user experience (UX)
- Better understanding customers
- Successful marketing
- Predicting or increasing financial performance
- Helping to ensure competitive advantage

With such a multitude of applications, both *The Economist* and *Forbes* have argued that data is now more valuable than oil:

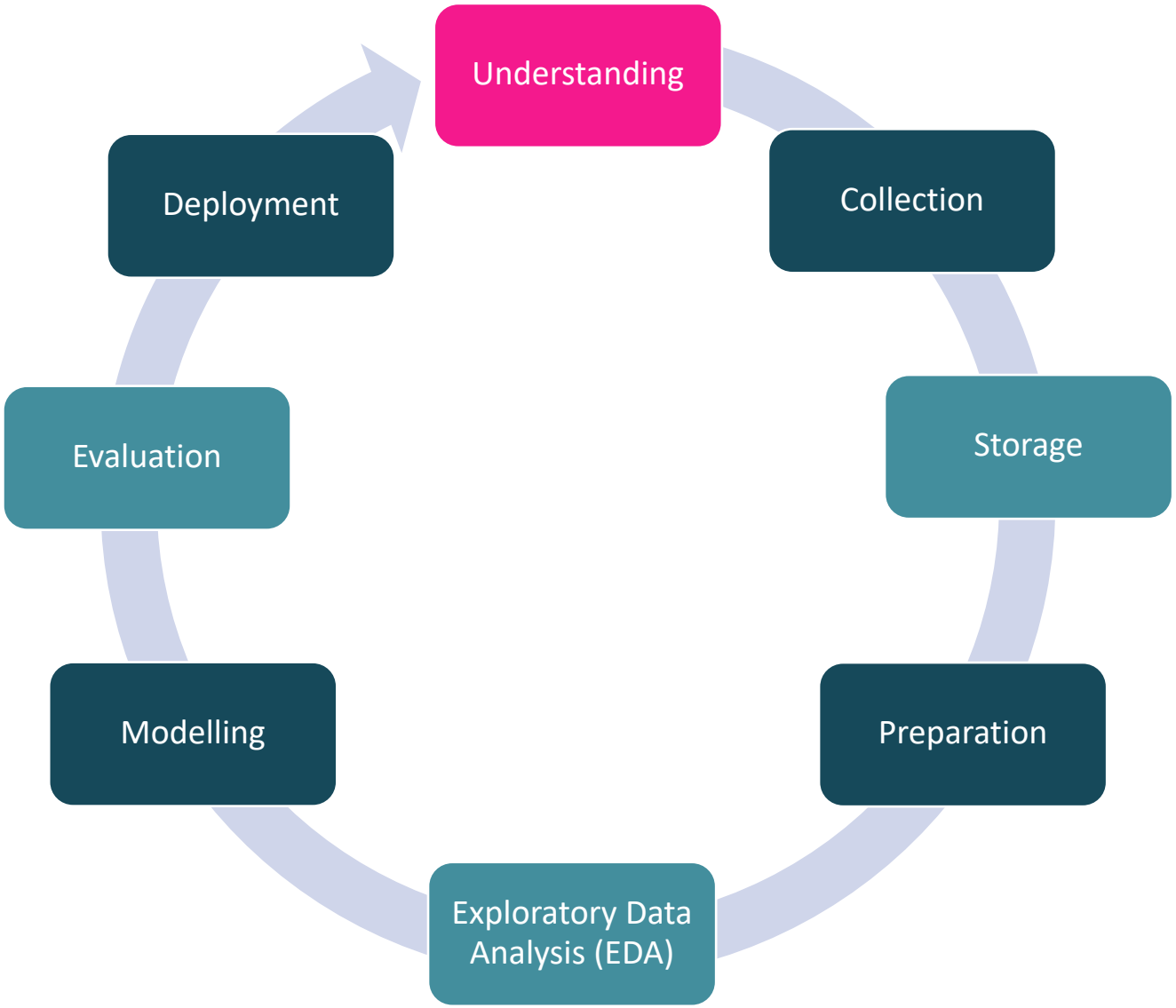
https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (Paywalled)

https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=343fc13d7304

The Data Pyramid



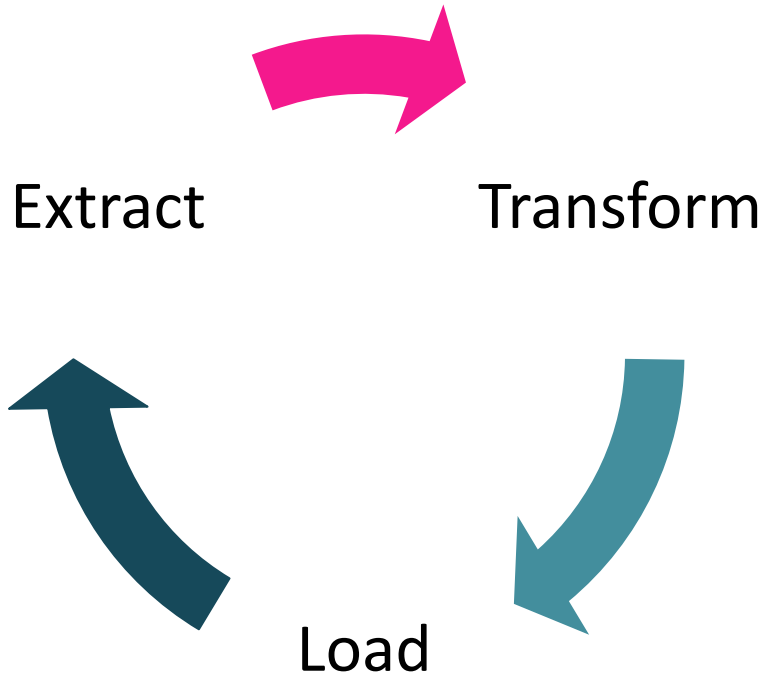
Data Lifecycle



TECH TALENT ACADEMY

Data Lifecycle and Pipelines

The term “data pipeline” refers to the process of setting up the necessary infrastructure that allows data science to take place. This includes the processes of capturing data, storing and transforming the data, and feeding this into the relevant tools for analysis, model creation and visualisation.



WOMEN IN DATA ACADEMY

TECH TALENT ACADEMY

Working in Data Science

There are a multitude of different roles available with the field of data science. The most famous of these is *Data Scientist*, but this is not the only data focussed role within many organisations.

Depending on your area of interest, you might be well suited to other roles.

WOMEN IN DATA ACADEMY

Data Scientist

What do they do?

- Focused on finding meaning in data.
- Builds, tests and evaluates machine learning models
- Data visualization
- Looking to the future and creating future-proof models and insights

Key Skills:

- Python
- R
- Statistical knowledge
- Databases (SQL/No SQL)
- Data lakes

Data Engineer

What do they do?

- Provide reliable data infrastructure
- Set up pipelines to gather, store and deliver data
- Analysis is focused on the best methods to move data along these data pipelines. Does not create ML models.
- Data access and structures

Key Skills:

- Understanding data warehousing (what is stored and where)
- Cloud platforms (AWS/Azure/Google Cloud)
- Databases (SQL/No SQL)
- Data lakes
- Python

Data Analyst

What do they do?

- Collect and capture data using multiple sources and work with Engineers to set up pipelines
- Data cleaning, validation and standardisation
- Reporting and data visualisation
- Dashboard creation
- Solve problems with existing data

Key Skills:

- Python
- R
- SQL
- PowerBI/Tableau
- Excel
- Cloud platforms (AWS/Azure/Google Cloud)

Business Analyst

What do they do?

- Work with clients and/or the business to ascertain problems that need to be solved.
- Use existing data to highlight problems.
- Communicate requirements to team.
- Propose data/IT solutions to solve problems.
- Entity relationship modelling.

Key Skills:

- Excel
- SQL
- Access
- PowerBI/Tableau
- Communication/Presentation Skills

Machine Learning Engineer – focussed on integrating ML code into software. Also monitors, updates, or retires models in production environments.

Data Quality Engineer – concerned with ensuring that data is extracted, transformed and loaded in such a way that data quality (how clean, complete or ordered it is) is prioritised.

BI Developer – concerned with developing data feeds into business intelligence software (PowerBI/Tableau) for business and data analysis and visualisation.

Data Product Manager – a management focussed role. Usually less technical, and more focused on features that are needed in data driven systems and dealing with the product.

What is “The Cloud”?

Cloud computing utilises a network of remote servers to offer computing services such as, data storage, processing and analytics.
These can then be accessed via the internet.
Various third-party providers operate and maintain these remote servers (data centres). This reduces costs and provides economies of scale to businesses.

Amazon Web Services (AWS)

Microsoft Azure

Google Cloud Platform

Further Information

Roles in Data Science Teams
https://www.youtube.com/watch?v=m5hLUknIi5c

Data Scientist V Data Engineer
https://www.youtube.com/watch?v=fUpChfNN5Uo

Business Analyst V Data Analyst
https://www.youtube.com/watch?v=G4syHs3M82E

Data Scientist Role
https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html

https://www.youtube.com/watch?feature=oembed&v=_Wk9T_G-u4o

WOMEN IN DATA ACADEMY

TECH TALENT ACADEMY