



**Institute** of  
**Data**

---

2023



# Data Science and AI

## Module 0

---

## Introductions, objectives & overview

---



# Logistics

- Zoom and Slack
- Breaks
- Questions



# Agenda of Module 0

- Introductions
- The Data Scientist role
- Objectives
- Overview of the course
- Hands-on labs and homework



# Introductions

- Please share with the class:
  - Current role and background
  - Why you are here?
    - Your **objectives and expectations** of attending the course
  - Your current skill levels in:
    - **Programming**
    - **Mathematics**
    - Other related areas (if applicable to you):
      - Information Management
      - Software Engineering
      - Business domain knowledge
  - Your **experience completing the prerequisites**



# What is data scientist's job

In simple terms, analysing data for actionable insights.

Specific tasks include:

- **Identifying** the **data-analytics problems** that offer the greatest opportunities to the organisation
- Determining the **correct data sets** and **variables**
- **Collecting** large sets of structured and unstructured data from disparate sources
- **Cleaning** and validating the data to ensure accuracy, completeness, and uniformity
- **Devising and applying models and algorithms** to mine the stores of big data
- Analysing the data to **identify patterns and trends**
- **Interpreting the data** to discover solutions and opportunities
- **Communicating** findings to **stakeholders using visualisation** and other means



# Skills of various roles in Data Science and AI

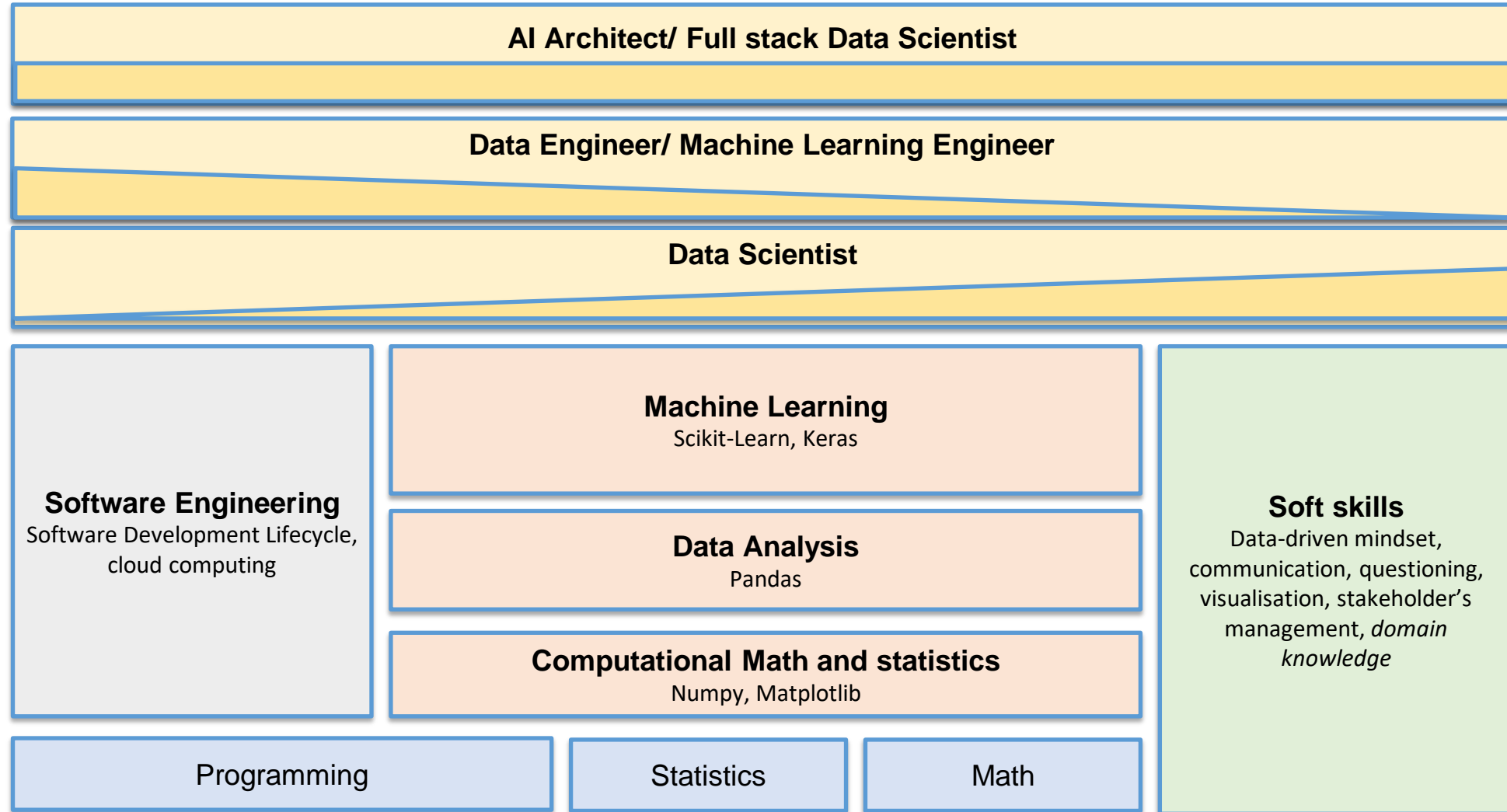
- There are a number of variations of roles that are required to deliver Data Science/AI projects.
- Some can be considered closer to business while others being more technical.
- There is a growing demand for Data Scientists to be able to contribute directly to systems in 'production'.

	Data Engineer	ML/AI Engineer	AI Architect	Data Scientist	Business Analyst
'Soft Skills' Data-driven mindset, Communication, Collaboration, Critical Thinking, Creativity					
Business Domain Knowledge					
Software Engineering & Information Management					
Programming					
Math Linear Algebra, Calculus, Statistics					

← Not Important      → Very Important

← Technical → Business

# Data Science skills for industry





## Foundational skills

- Programming for Data Science (Python)
- Maths and Statistics for Data Science

## Core Data Science and AI skills

- Exploratory Data Analysis (EDA) and data wrangling
- Data visualisation
- Database access
- Application Programming Interfaces (APIs)
- Supervised learning (Regression and Classification)
- Unsupervised learning (Clustering and Dimensionality reduction)
- Deep learning
- Natural Language Processing (NLP)
- Artificial Intelligence
- Cloud computing
- Machine learning deployment
- Data science industry practices

## Applying Data Science in industry

- Applying data science on different data structures and domains
- Defining a data science project
- Designing a data science project
- Delivering data science project
- Optimising machine learning model algorithms
- Overall end-to-end solution
- Presenting to stakeholders and obtaining buy-in
- Capstone project

## Soft skills

Consulting, Questioning, Critical Thinking, Problem Solving, Documenting, Presenting

## Learning how to learn effectively framework

Minimal Viable Learning (MVL), Multimodal learning, Learn-Create cycle



# Approach and principles of the course

This course aims particularly to prepare students to **get a role as a data scientist and perform well in this role in industry**. This aim shapes the curriculum and the delivery of the course through the following principles:

- Emphasis on **practical skills** for succeeding as a data scientist in industry
- **Workshop-style, highly interactive** and **collaborative** teaching techniques
- Use of **computational math and statistics** rather than theoretical aspects
- Priority on **doing than remembering**
- **Minimal Viable Learning** (MVL) approach



# Tips for succeeding in the course

- **Attend and be engaged** – have your camera on if online, minimise distractions, ask questions, take notes, seek clarification when there are doubts
- **Learn with others** – practice explaining concepts in simple (but precise) terms, seek feedback
- **Do self-study where needed** – use online resources (videos, blog posts, study the code of others)
- **Learn to use online searches effectively** – it's perfectly acceptable to debug by looking up error messages online
- **Document** frequently used code and learn from common mistakes
- Avoid going too deep into an area at the expense of falling behind
- **Pace yourself** and take breaks – avoid too many marathon sessions to catch up, this can impact your health



# Objective of Data Science and AI course

By the end of the Data Science and AI program you will be able to:

*Help business to make effective data-driven decisions and track their effectiveness using the appropriate combination of the following tasks:*

- Collect, extract, query, clean, and aggregate **data** for advanced analytics purposes
- Perform **statistical and visual analysis** on data using Python and its libraries and tools
- Build, implement, and evaluate advanced analytics problems using appropriate **machine learning models** and algorithms
- Use data visualisation tools to **communicate** findings
- Create clear **and reproducible** reports for stakeholders
- Use **business consulting** skills and frameworks in data science to assist managers and stakeholders understand the application of AI technology
- Identify **big data** problems in businesses and understand how computing technologies are solving these challenges
- Apply **hypotheses testing, modelling, and validation problem-solving** processes to datasets from different industries in order to provide insight into real-world problems and solutions

# Data Science skills for industry

- **Foundational skills** that are required to learn Data Science:
  - Programming
  - Maths, Statistics
  - Basic software engineering
  - Soft skills

# Data Science skills for industry

- **Core** Data Science skills
  - Computational maths and statistics
  - Data Analysis
  - Machine Learning
- **Complementary** Data Science skills
  - Business domain knowledge
  - Software Engineering
  - Soft skills
    - Questioning
    - Critical Thinking
    - Communication and presentation
    - Problem solving



# Programming Data Science in Python

## Programming is:

- the **process of creating a set of instructions** that tell a computer how to perform a task.
- thinking **systematically and critically**
- breaking a task into steps. Examples include: a recipe, directions to a destination and mathematical problem solving.

Python has a very **active community** with a vast selection of **libraries**, especially in scientific computing, data analysis and visualisation which makes it **very suitable for Data Science**.

There are a number of tools available to support the development of Python.  
**Jupyter notebook** has emerged as an effective way to develop and share Data Science projects.  
Visual Studio Code (VSC) is an alternative for developing reusable software modules.

Programming (**computational mathematics and statistics**) can be crucial for developing deep mathematical and statistical knowledge and skills.

# Why is Statistics important for a Data Scientist?

- **Statistical Thinking** is an essential component of a data-driven mindset which is crucial for a Data Scientist
  - Statistical analysis must start with the appropriate **data** (sample)
  - Statistical Inference (reasoning) should start with measurement, ideally, via **controlled experiments**
  - Statistics uses samples (a small subset of the population) and therefore always has a degree of **uncertainty**
  - Sampling must be **random, and preferably, independent**
- The best way to learn statistics is by **experimenting with data using Python code and visualisation**





# Hands-on labs and homework

- The course focus on the practical aspects of Data Science to prepare for a real-life role.
- You will need around 6 hours/week for homework
- Programming environment
  - We will use Jupyter Notebook with Anaconda for coding on your own machine



# Questions?



# End of Presentation!