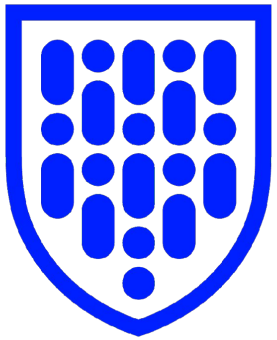


Institute of Data

2019

Data Science and AI

Module 0

Introductions, objectives & overview

2

Institute of Data

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:
    - **Mathematics**
    - **Programming**
    - Other related areas (if applicable to you):
      - Information Management
      - Software Engineering
      - Business domain knowledge
  - Your experience completing the prerequisites

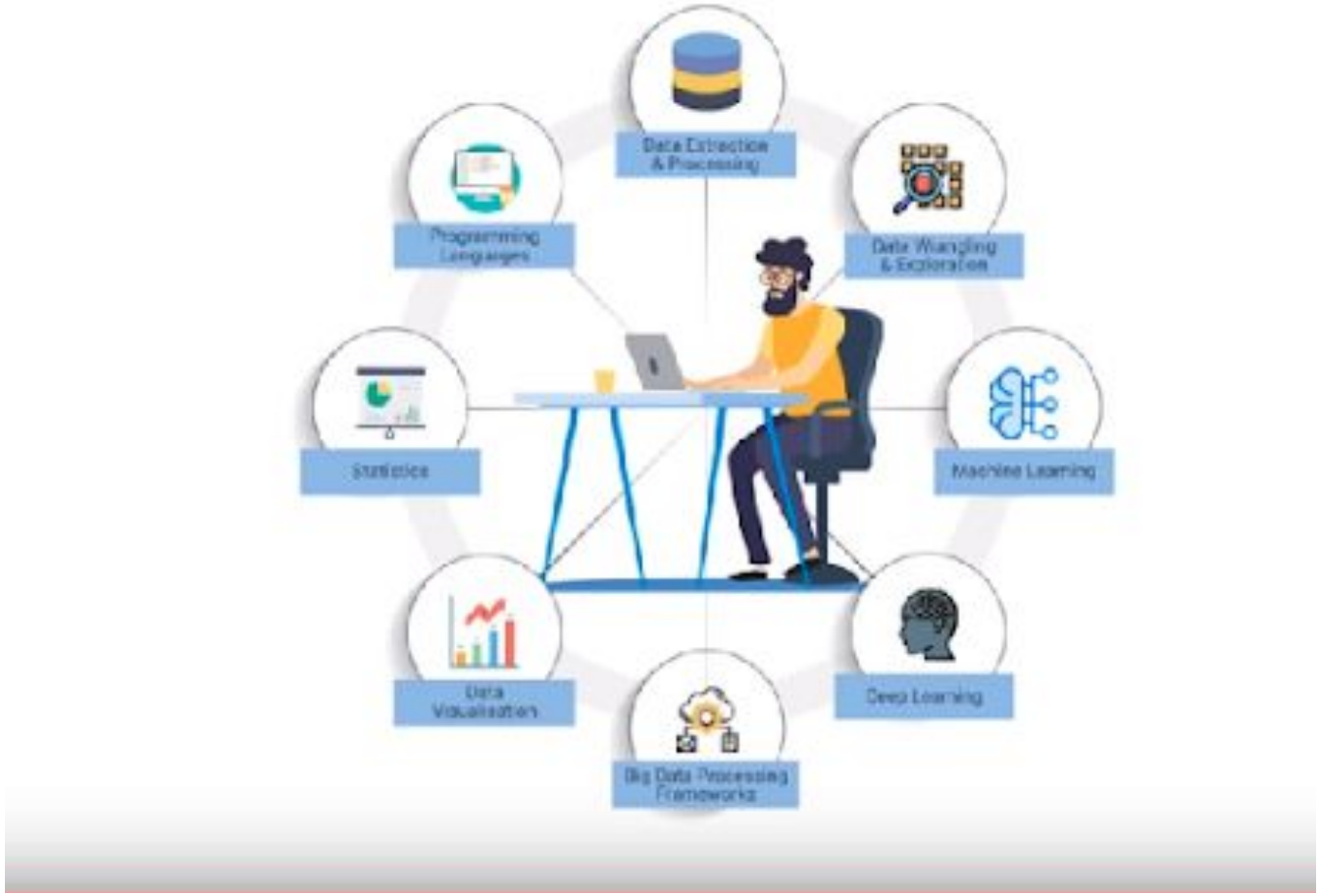
What is data scientist's job

In simple terms, Analyze data for actionable insights.

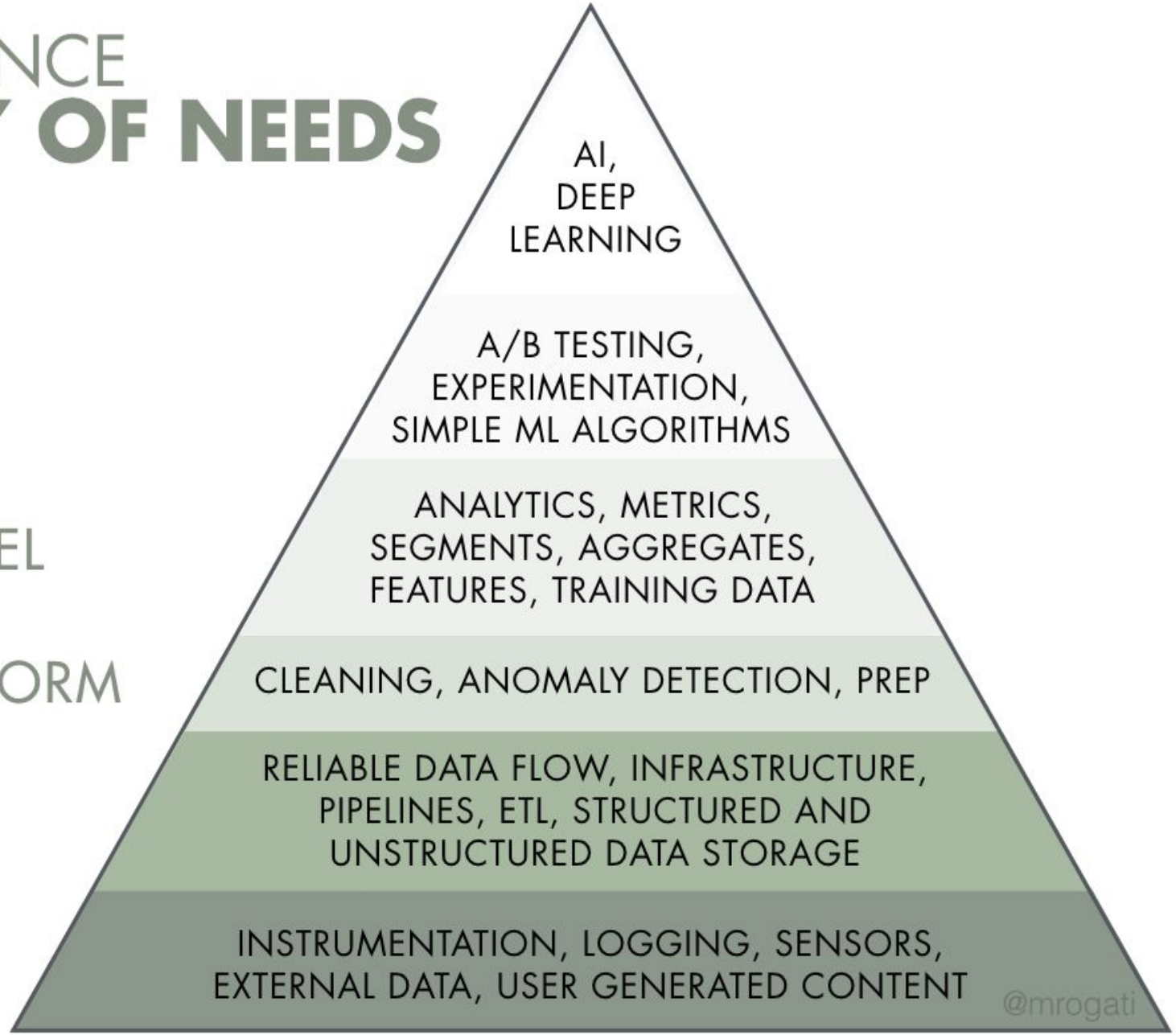
Specific tasks include:

- **Identifying** the **data-analytics problems** that offer the greatest opportunities to the organization
- 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
- Analyzing the data to **identify patterns and trends**
- **Interpreting the data** to discover solutions and opportunities
- **Communicating** findings to **stakeholders using visualization** and other means

Data Science Skills



THE DATA SCIENCE HIERARCHY OF NEEDS



LEARN/OPTIMIZE

AGGREGATE/LABEL

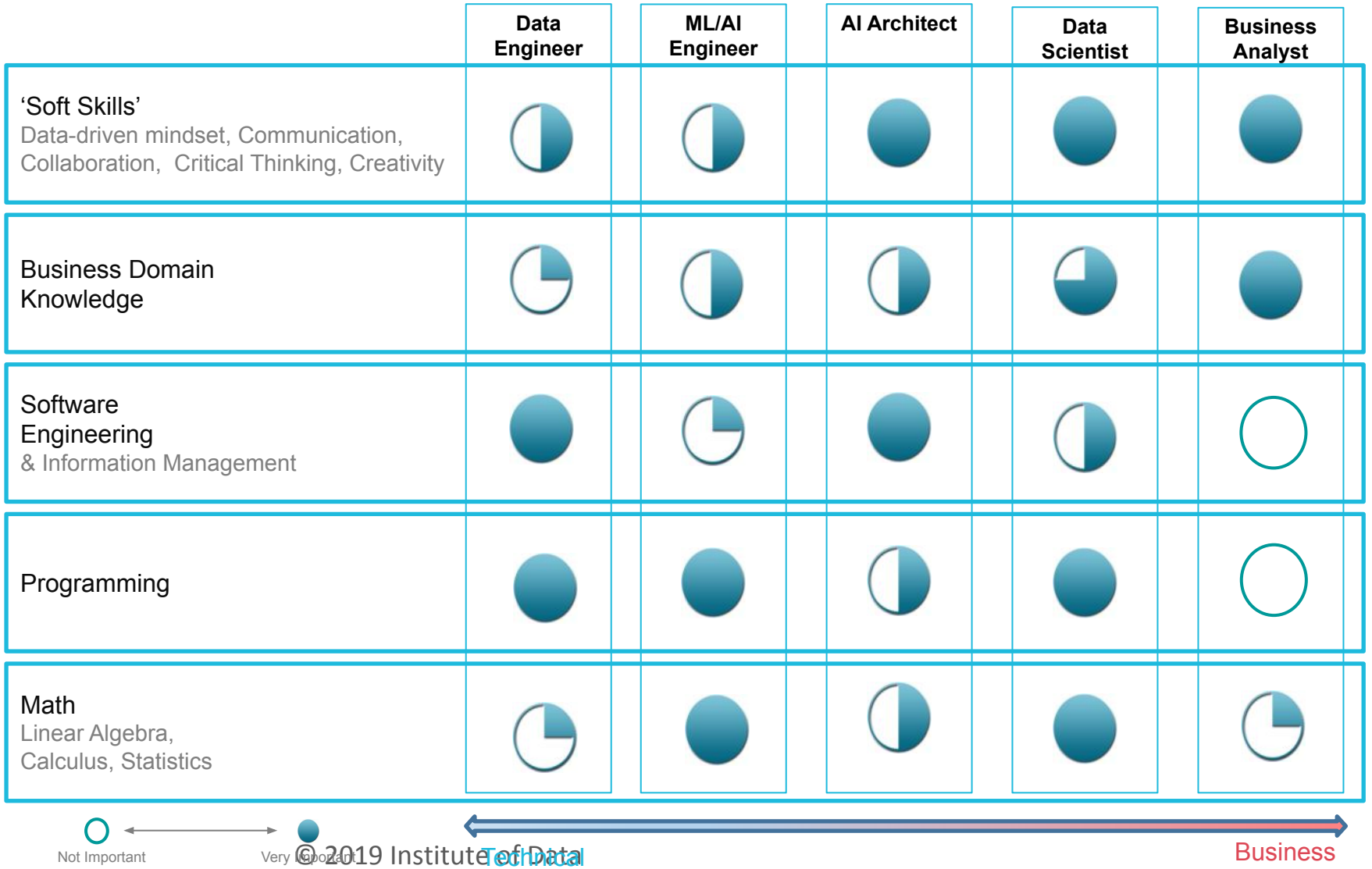
EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

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’.



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

Course overview

Foundation	Algorithms	Practical Applications
• Math and statistics • Python Programming • SQL and Databases • Exploratory Data Analysis (EDA)	• Introduction to Machine Learning • Supervised classification • Clustering and unsupervised classification • Classification and regression • Ensemble models • Network analysis • Text analytics • Artificial Intelligence	• Data Science leading practices • Case studies • Capstone project

Hands-on labs and homework

- The course focus on the practical aspects of Data Science to prepare for real-life role.
- You will need around 6 hours/ week for homework
- Programming environment
  - **We will use Google Colaboratory (Colab) for *coding and sharing* Notebooks**
    - Colab is a free Jupyter notebook environment that requires no setup and runs entirely in the cloud.
    - With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.
  - **We will use Jupyter Notebook with Anaconda for coding on your own machine**

Questions?

End of Presentation!