INTELLIGENT SYSTEMS LABORATORY

Lesson 1
Data Science

Mathematics and Statistics for Data Science

Tapanan Yeophantong
Vincent Mary School of Science and Technology
Assumption University

In this Lesson …

- Fundamental Concepts in Data Science
- Applications of Data Science in the Real World

Data Science

Why do we need to know this?

Our Goal

- Competitiveness
- Collect and use the data about our customers (audience), and what's trending, to discover insights (patterns) that can help us improve the way we market ourselves.

ACTi
Connecting Vision

Smart Retail Solution

Offline

Online

O2O Customer ID Mapping

From Stores to Record Labels

- How do we define the audience of our artists?
- Can we define similar touchpoints and journey in our audience's experience?
- How can we capture their activities and behaviour, and from where?

What is Data Analytics?

- A systematic analysis of data for discovery, interpretation, and communication of meaningful patterns.
- It focuses on explaining *why* something happens and *what will happen* in the future.
- It involves computer skills, math, and statistics.

Objectives in Data Analytics

- To understand what we're dealing with?
- To confirm our beliefs?
- To predict what's to come?
- To decide on what to do knowing what's to come?

4 STAGES OF DATA ANALYTICS MATURITY

Descriptive Analytics

Diagnostic Analytics

Predictive Analytics

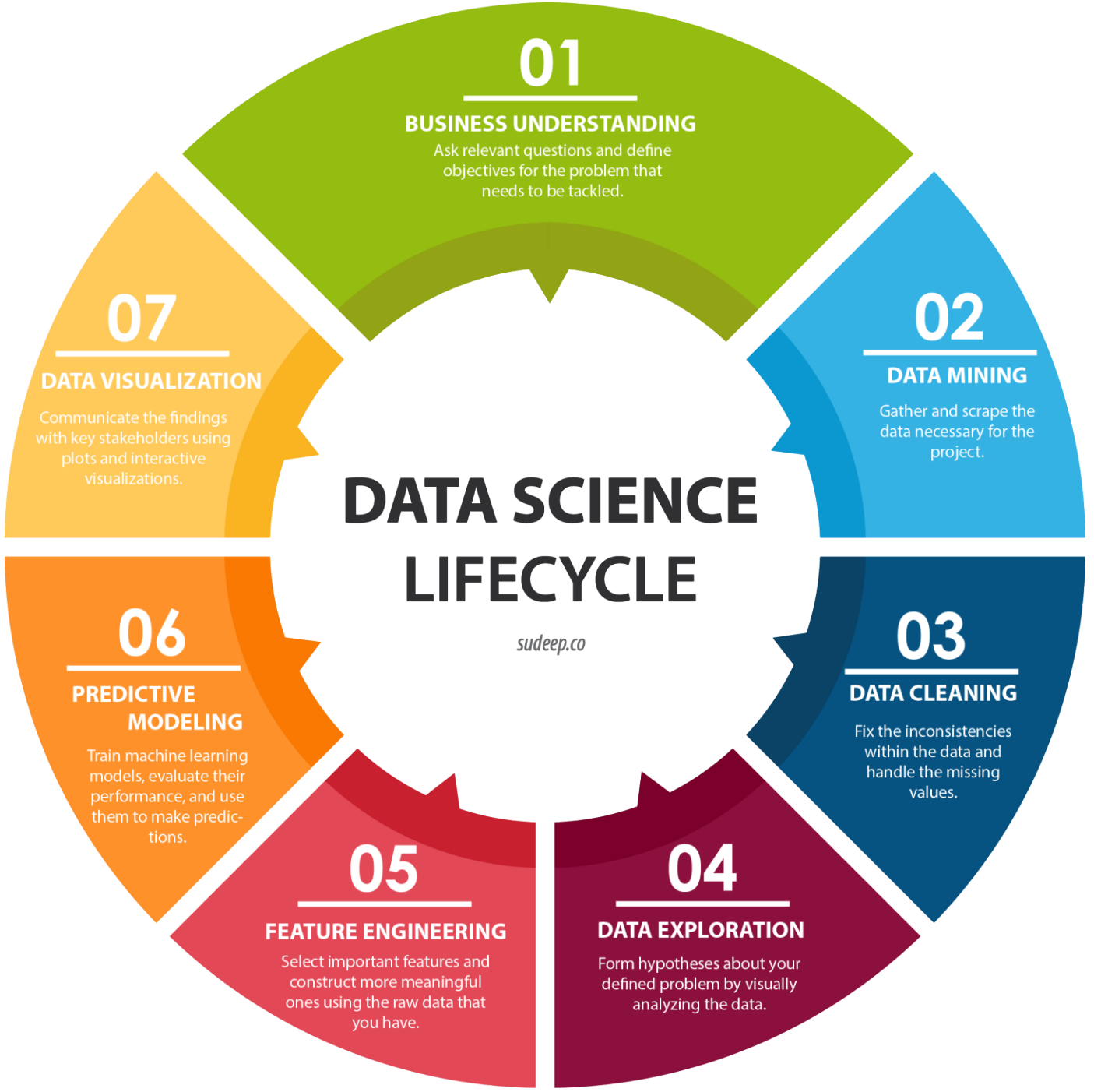
Prescriptive Analytics

Making Decisions

Data science always begin with **Business Understanding**.

Data Science Lifecycle

- Define
- Collect
- Process
- Extract
- Model
- Use



Collect

- After you've set your goal, the next step is to collect data.
- Data can be:
  - Structured (customer records, sales transactions), or
  - Unstructured (images, videos, text messages)
- Traditionally, we obtain data explicitly through questionnaires and surveys.
- Today, data is EVERYWHERE.

Filtering

- The idea is to make data more understandable (both for the human users and the machine).
- Typical tasks:
  - Filtering. Cleaning up noise and stuff irrelevant to your question.
  - Transformation. Making your data more meaningful.

Extract

- A very important task most often neglected.
- This involves exploring and extracting important elements from the data to reduce complexity.
- Typical tasks:
  - Selection. Choose which features to use.
  - Aggregation. Combining two or more features to form a new one.
  - Decomposition. Performing math & statistical techniques to convert the data into something more meaningful.

Model

- Apply statistical or machine learning techniques to get the data to answer your questions.
- Three common ways:
  - Clustering
  - Classification
  - Regression

Use

- Finally, we use the results of our hard work to meet our goals.
- A common usage is in visualization.
- But today, we can also use the results to automate tasks, like:
  - Recommend content to users
  - Monitor and improve customer engagement
  - Alert on important activities or events
  - Take actions on behalf of the users