Course Title:

Data Science (数据科学)

(Semester: Fall 2021/2022)

Dr. Oluwarotimi W. SAMUEL

Research Center for Neural Engineering
Shenzhen Institutes of Advanced Technology
Chinese Academy of Sciences

Contact: (Email: samuel@siat.ac.cn & timitex92@gmail.com)

Phone: +86-15814491870

(2020.09.16)

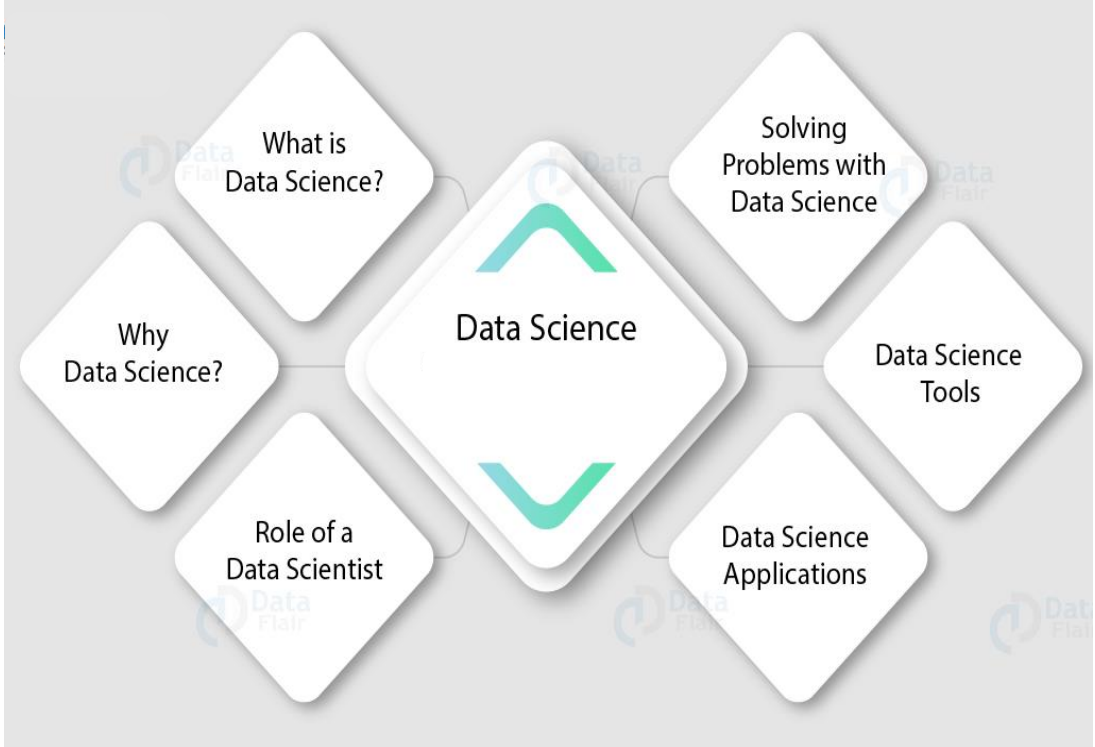
Introduction to Data Science

- **Objective:** This lecture will introduce:
  - ✓ the concept of Data Science (**DS**),
  - ✓ its importance in modern day life, and
  - ✓ key areas where the concept of **DS** have been successfully deployed to solve challenging real-life problems.
- **Expectation:** At the end of this lecture, students are expected to have an ***Overview*** of the concept of **DS** and its wide range of relevance in the modern day life**.**

Introduction to Data Science

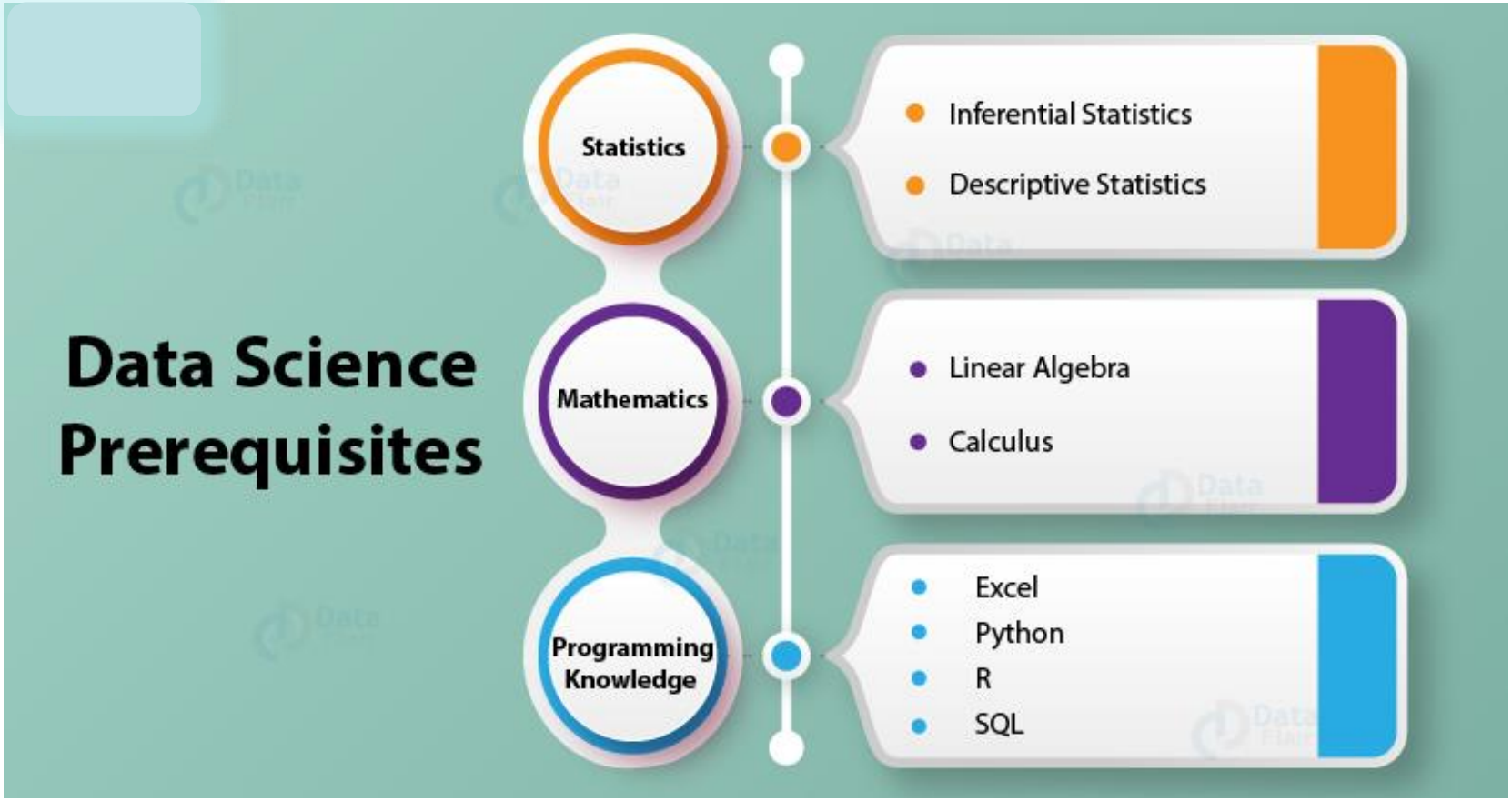
❑ Outline for today's Lecture

- ✓ What is Data Science
- ✓ Why Data Science
- ✓ Role of Data Scientist
- ✓ Solving Problems with Data Science
- ✓ Data Science Tools
- ✓ Data Science Applications



Introduction to Data Science

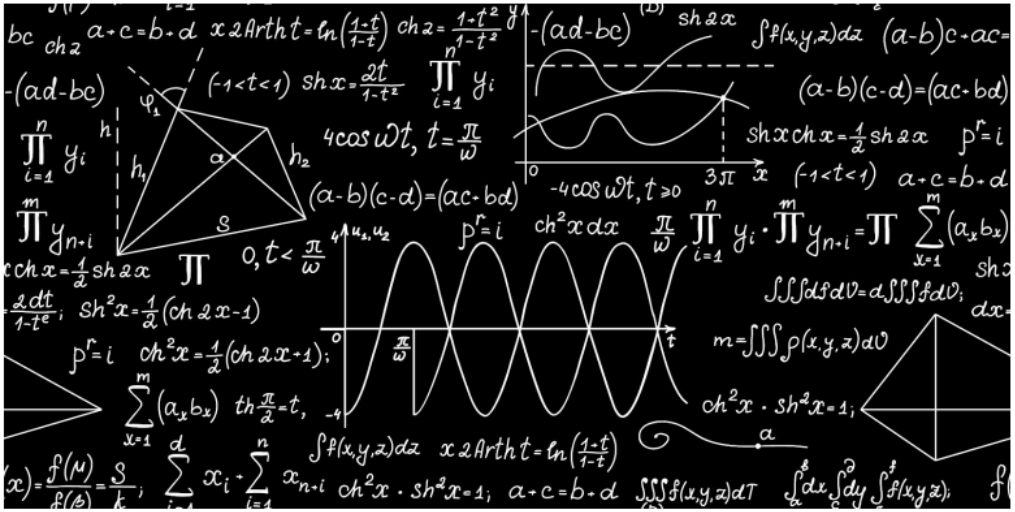
❑ Data Science Prerequisites

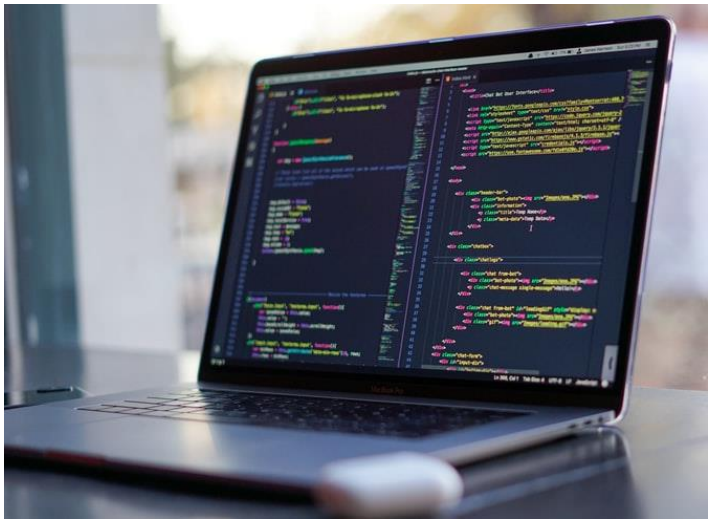


Introduction to Data Science

❑ Data Science Prerequisites

Statistics, Mathematics (Linear Algebra & Calculus), & Programming





❑ What is Data Science

"The future belongs to the companies and people that turn data into products"

Introduction to Data Science

Data Science (DS): is the science which uses computer science, statistics and machine learning, visualization and human-computer interactions to:

- ✓ collect,
- ✓ clean,
- ✓ integrate,
- ✓ analyze,
- ✓ visualize, and

interact with data to create data products.

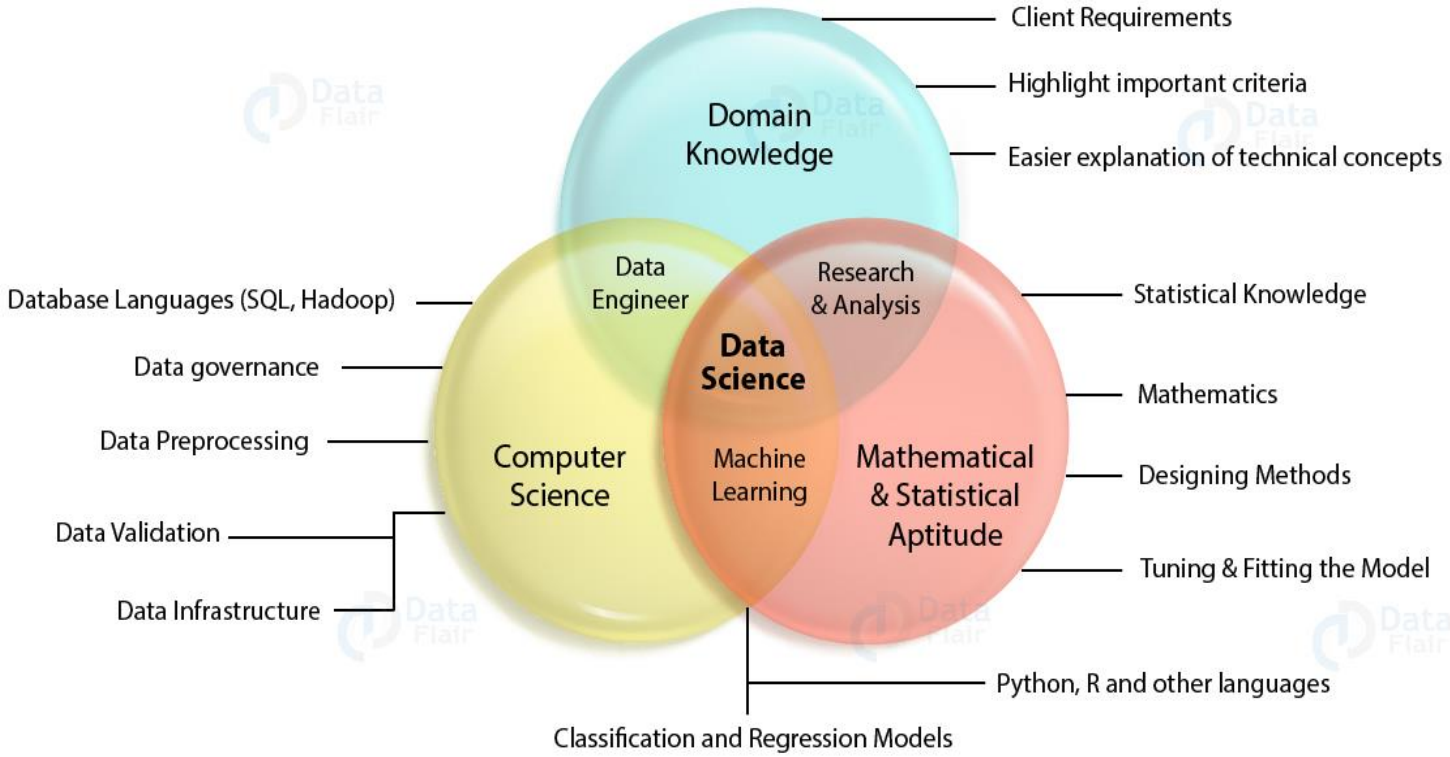
Introduction to Data Science

DS: The application of data-centric computational, and inferential thinking to understand the world and solve problems.

-Joseph Gonzalez, University of California, Berkeley

Introduction to Data Science

"**DS** is about extraction, preparation, analysis, visualization, and maintenance of information. It is a ***cross-disciplinary field*** which uses scientific methods and processes to draw insights from data. "



❑ Why Data Science?

Before the advent of DS, we have the traditional programming concept:

- ✓ firstly understand the problem
- ✓ develop an algorithm to solve the problem
- ✓ transform the algorithm into program (C++, Python, Java)
- ✓ run the program using ***structured inputs*** to obtain an output

Introduction to Data Science

❑ Why Data Science?

- ✓ About **80%** of the data gathered by any company is likely to be ***unstructured***, and analyzing such ***unstructured data*** is not as simple as analyzing ***structured data***.
- ✓ In other words, with the emergence of new ***technologies***, there has been an exponential increase in volume of data.
- ✓ This has created an opportunity to ***analyze*** and ***derive*** meaningful insights from data.

❑ Why Data Science?

- ✓ Therefore, we need DS to extract meaningful information from such data.

- ✓ Handling such data requires special expertise of a '***Data Scientist***' who can use various statistical & machine learning tools to understand and analyze data.

Introduction to Data Science

❑ Goals of Data Science

The main goals of DS include:

- ✓ to gain insight into some problem in the real world
- ✓ finding hidden patterns from ***raw/unstructured data***
- ✓ turning ***raw data/unstructured data*** into data products

❑ Roles of Data Scientists

Data Scientists' have unique Roles compared to other professionals:

Data Scientists

- Optimize data processing
- Define metrics
- Establish collection methods
- Work with enterprise systems

Data Engineers

- Optimize data flow
- Mainly, from input to output

❑ Roles of Data Scientists

Data Scientists' have unique Roles compared to other professionals:

Data Scientists	Statisticians
• Data collection	• Surveys
• Data cleaning	• Polls
• Create machine learning models	• Experiments
• Implement algorithms	• Improve upon simple model

❑ Roles of Data Scientists

Data Scientists' have unique Roles compared to other professionals:

Data Scientists

- Automate reports

Business Analysts

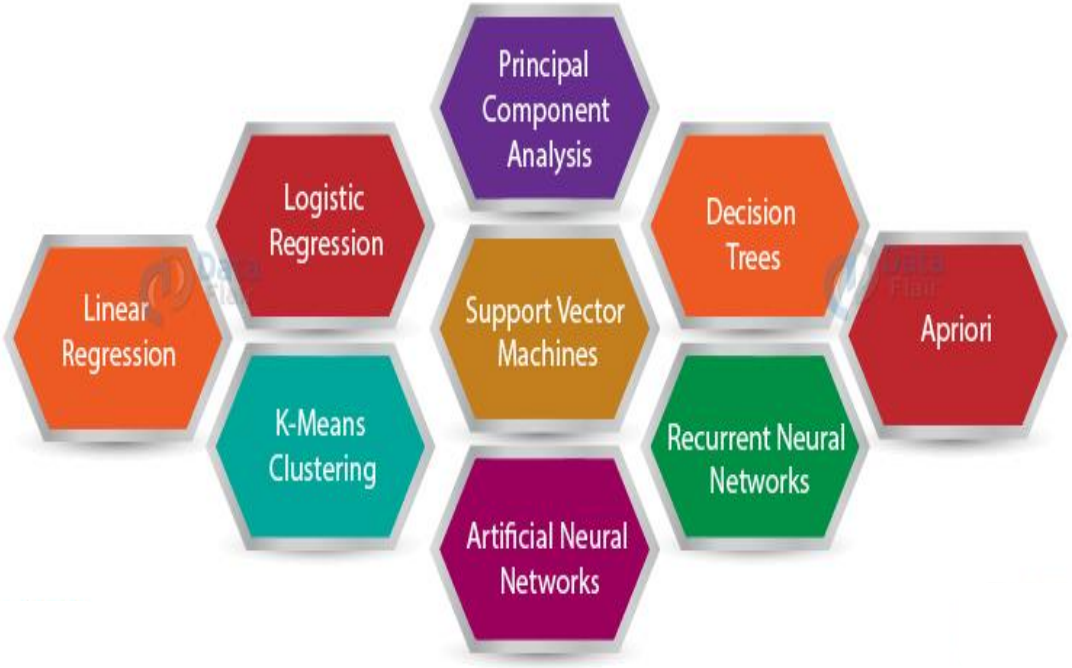
- Database design
- ROI assessment
- Finance planning
- Optimization
- Risk management

"Data sets in the right hands can help predict and shape the future"

❑ Top Data Science Algorithms

Here is the list of **top Data Science Algorithms** that you must know to become a data scientist.

- ✓ Linear Regression
- ✓ Logistic Regression
- ✓ K-Means Clustering
- ✓ Principal Component Analysis
- ✓ Support Vector Machines
- ✓ Artificial Neural Networks
- ✓ Decision Trees
- ✓ Recurrent Neural Networks
- ✓ Apriori Algorithm



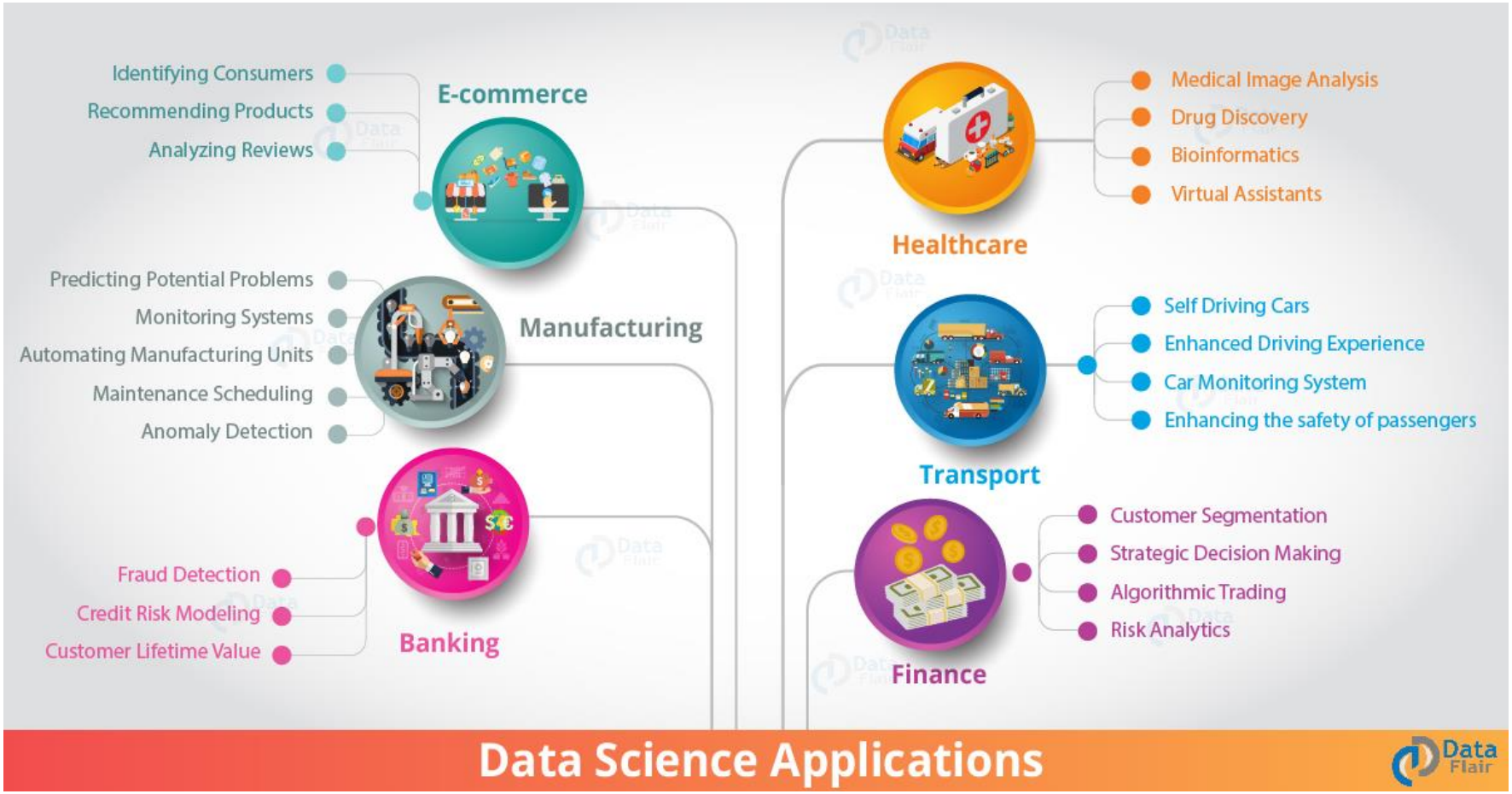
❑ Data Science Applications

The following are some typical areas where Data Science have been successfully applied:

- ✓ E-Commerce
- ✓ Manufacturing
- ✓ Healthcare
- ✓ Banking
- ✓ Transportation
- ✓ Finance
- ✓ Agriculture

Introduction to Data Science

❑ Data Science Applications



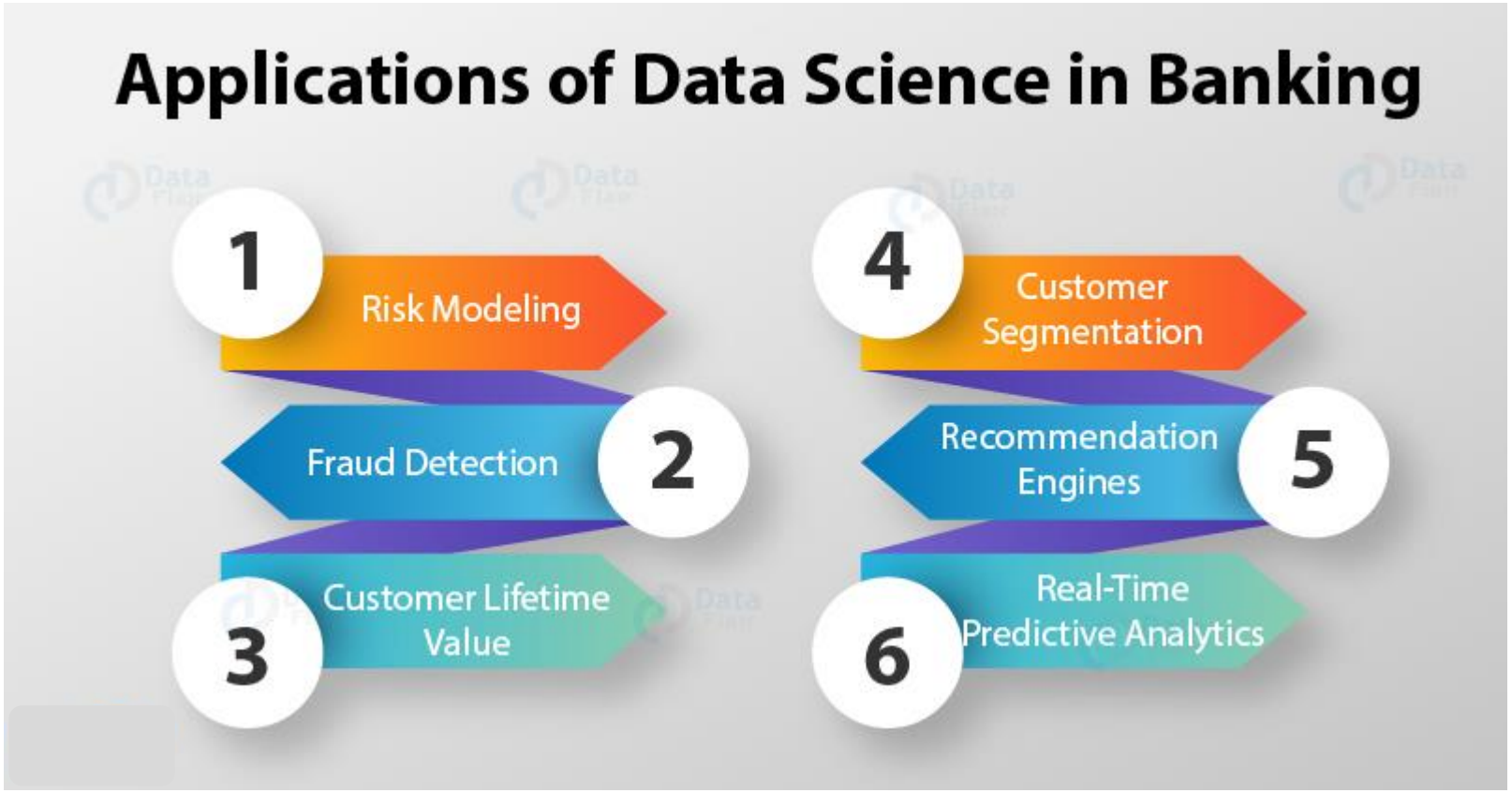
❑ Data Science Applications (Agriculture)

Applications Data Science in Agriculture Include:

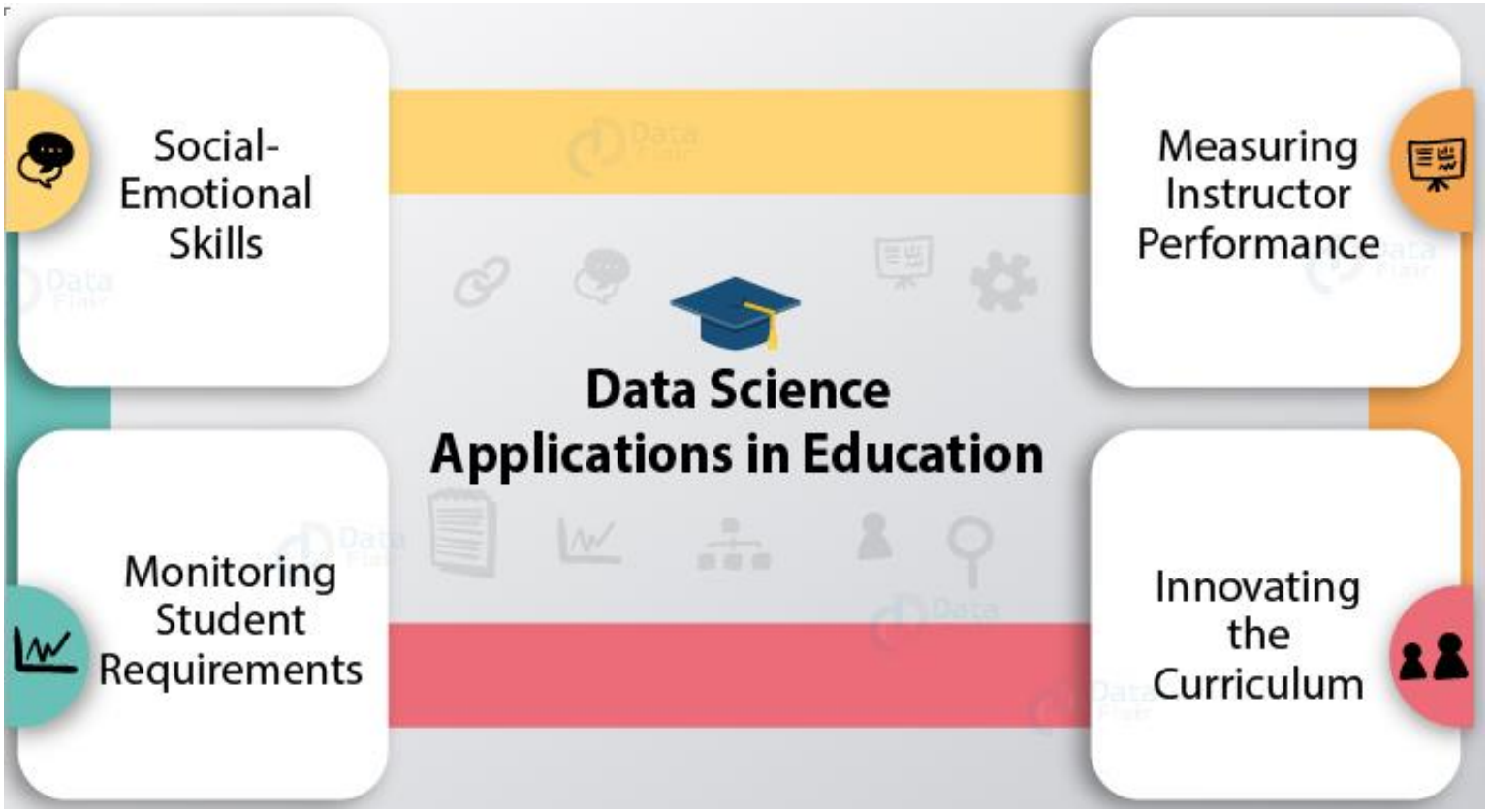
- ✓ Digital Soil & Crop Mapping
- ✓ Weather Prediction
- ✓ Fertilizer Recommendation
- ✓ Disease Detection
- ✓ Disease Management
- ✓ Adaptation to Climate Change
- ✓ Automated Irrigation System

Introduction to Data Science

❑ Data Science Applications (Banking)

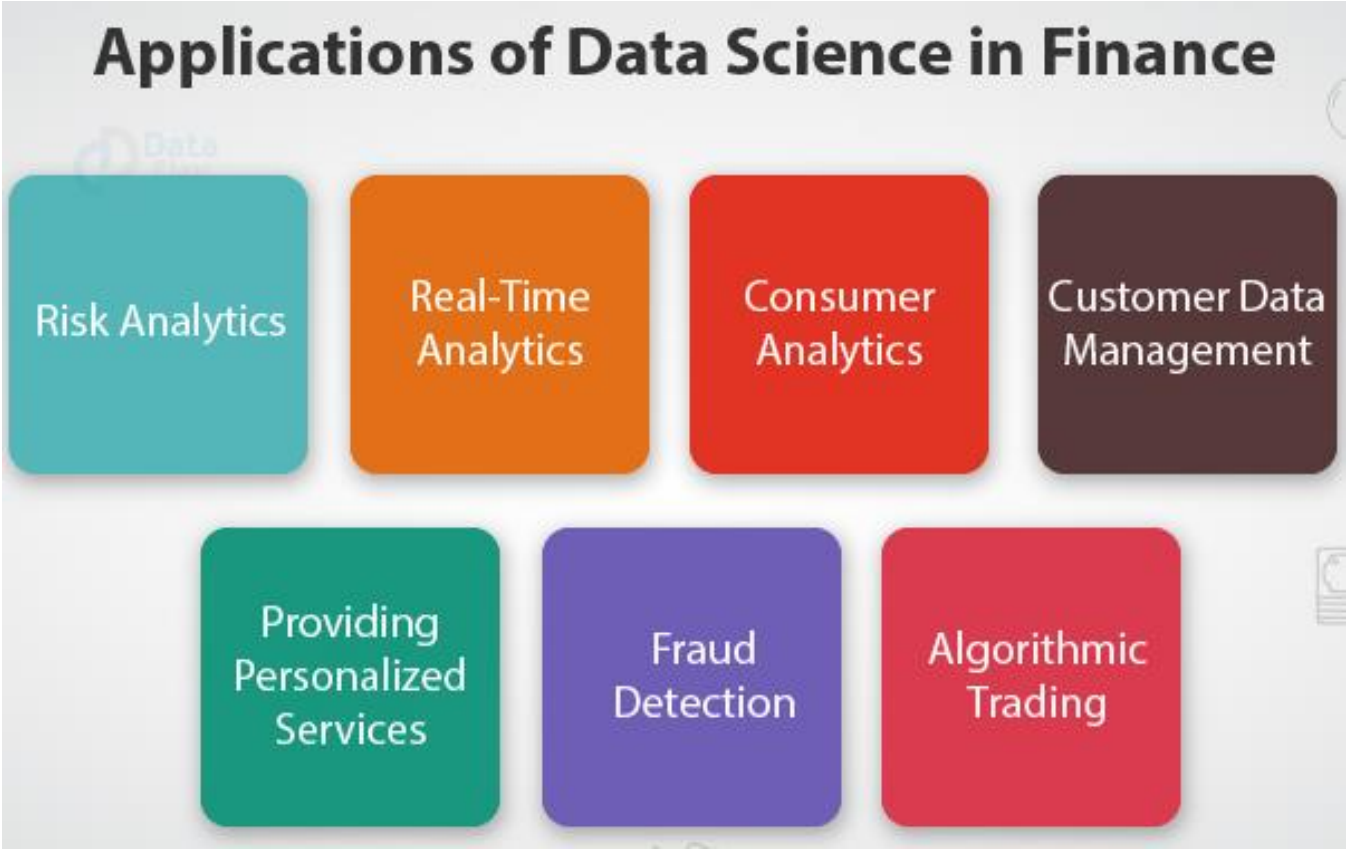


❑ Data Science Applications (Education)

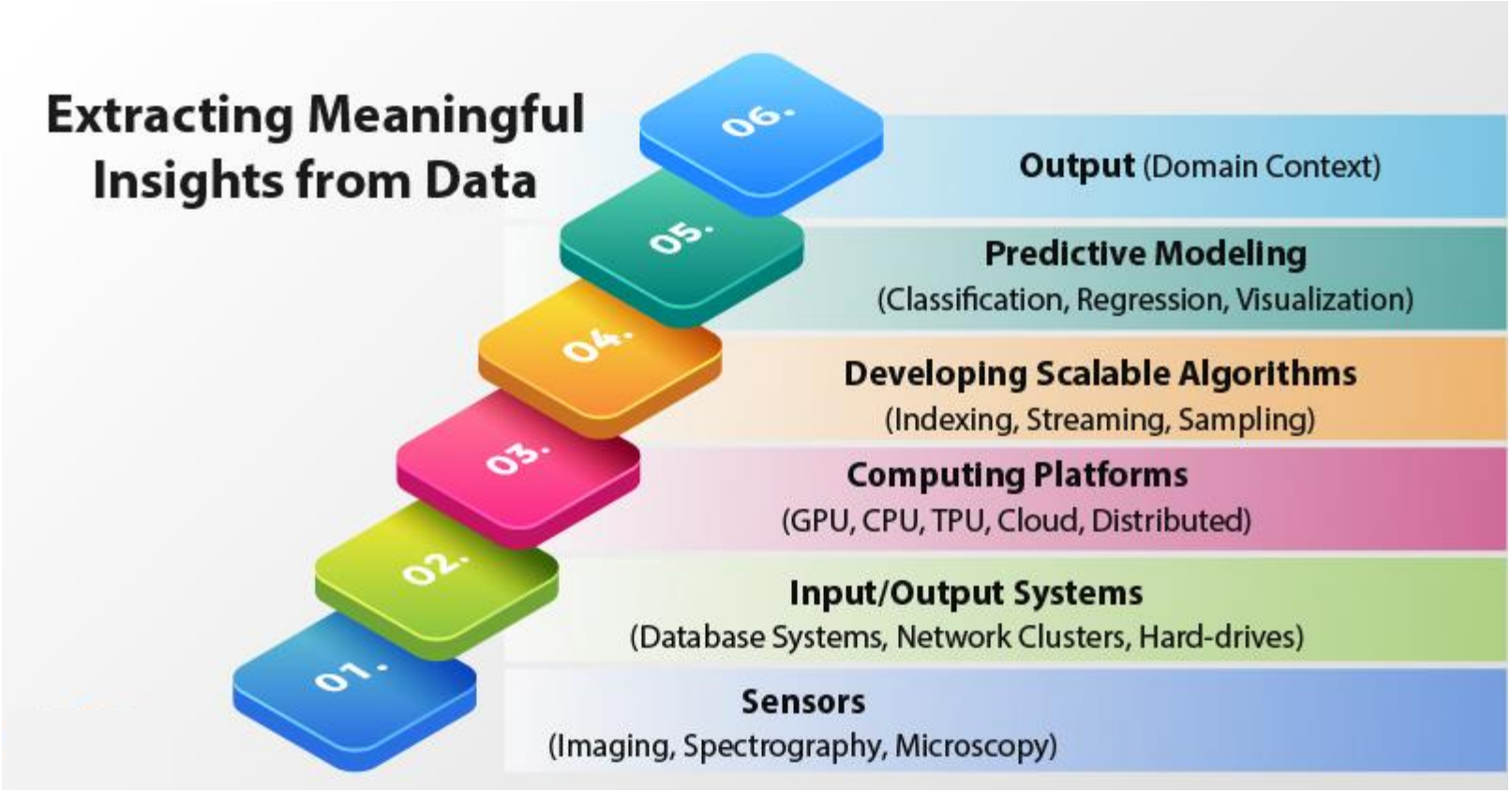


Introduction to Data Science

❑ Data Science Applications (Finance)



❑ Data Science Applications (Healthcare)

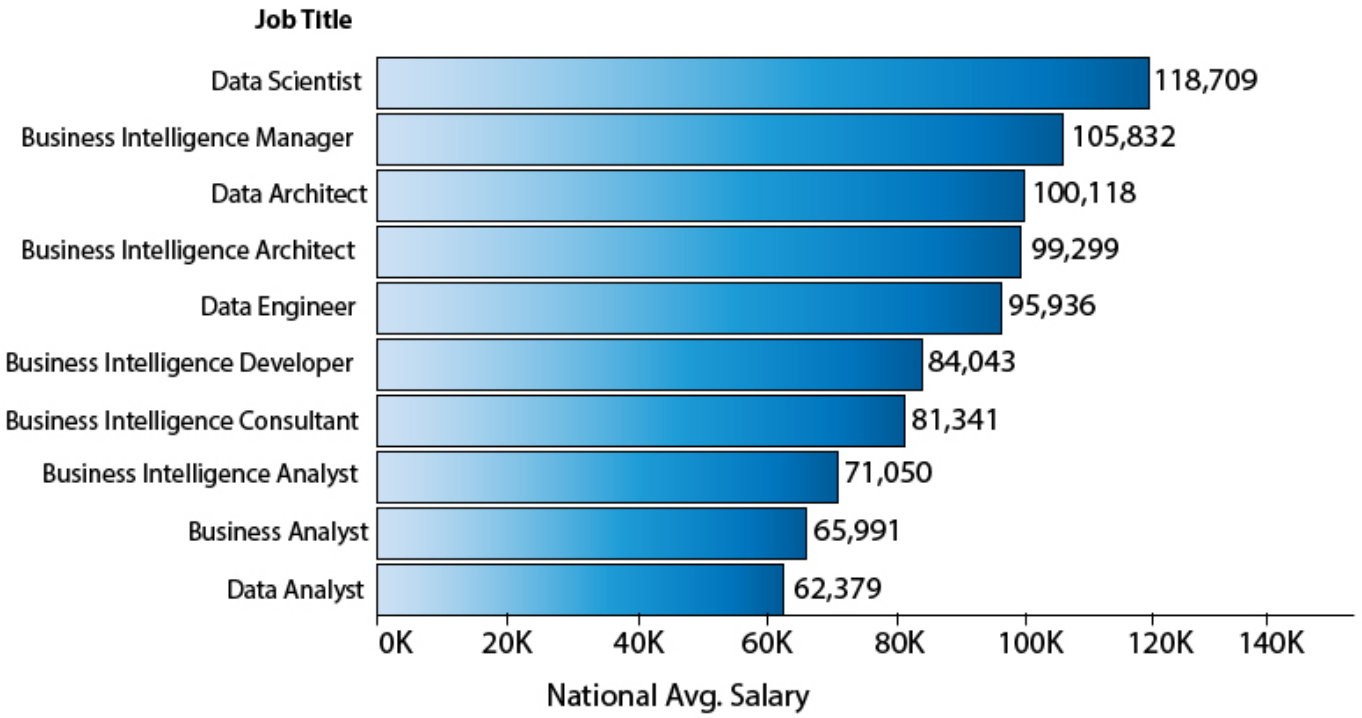


Introduction to Data Science

❑ Prospects of DS as a discipline:

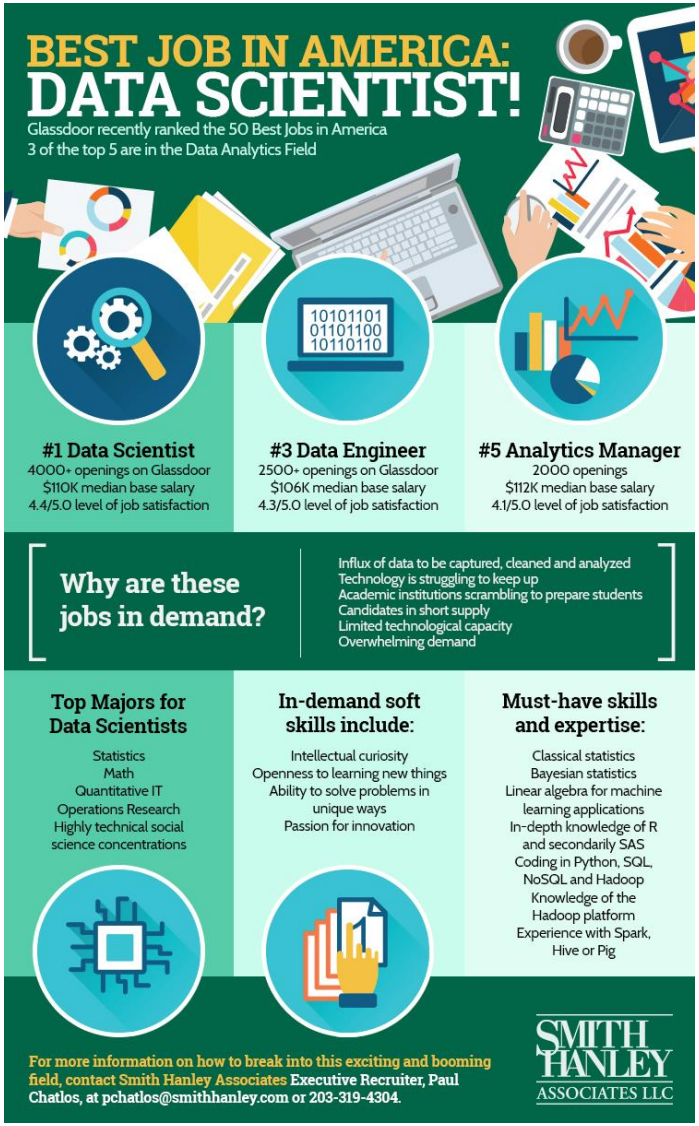
Introduction to Data Science

❑ Prospects of DS as a discipline:



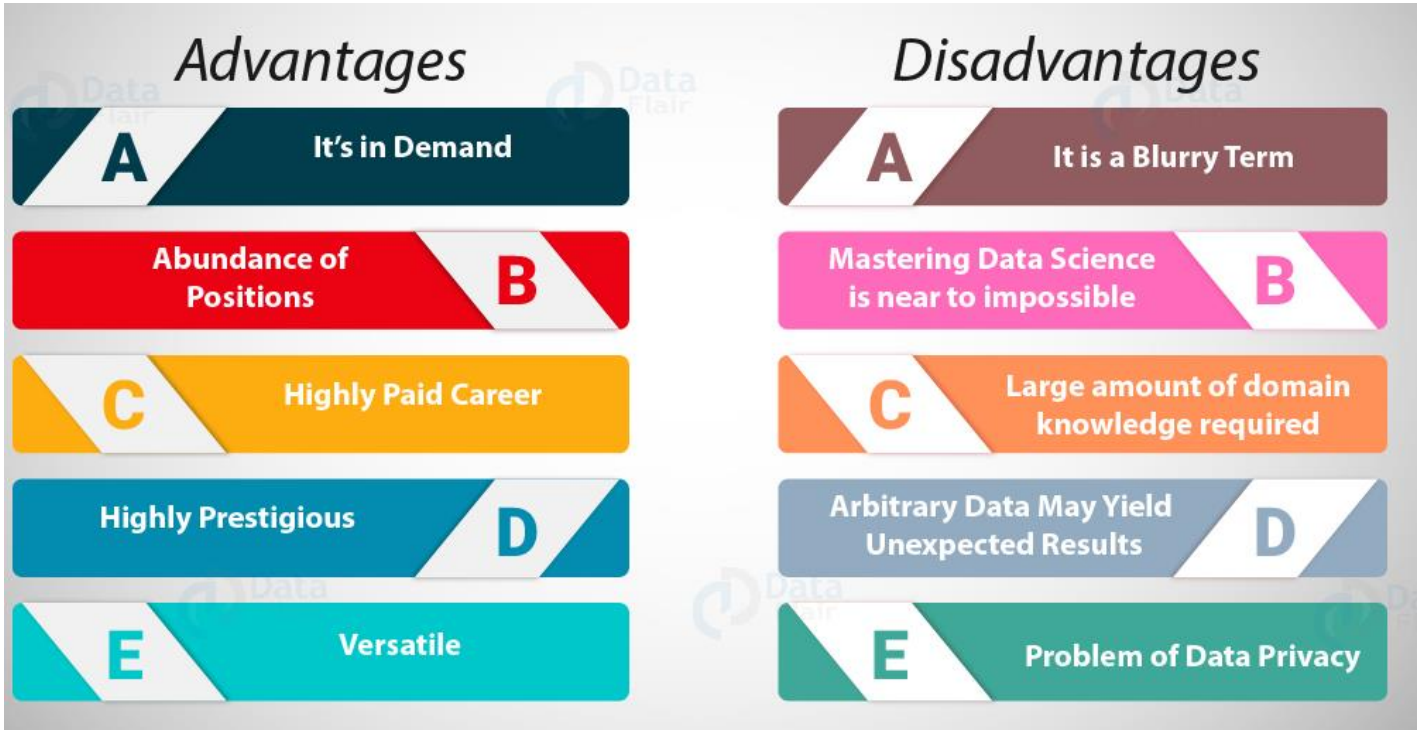
Introduction to Data Science

- **Prospects of DS as a discipline:**



Introduction to Data Science

❑ Advantages and Disadvantages of Data Science



❑ Summary for today's lecture

- ✓ We have learned about Data Science concept, its importance, core algorithms used by Data Scientists,
- ✓ Different applications areas of Data Science was covered as well as opportunities that are open to Data Scientists.
- ✓ The advantages and disadvantages of choosing Data Science as career were highlighted.

Questions and Comments!