

# INTRODUCTION TO DATA SCIENCE

FOR-IAN V. SANDOVAL

Asst. Professor II

Laguna State Polytechnic University

# LEARNING OBJECTIVES

---

- Apprehend the field of Data Science impact and importance in the society
- Reflect on its applications, importance and advantages

# CONTENTS

---

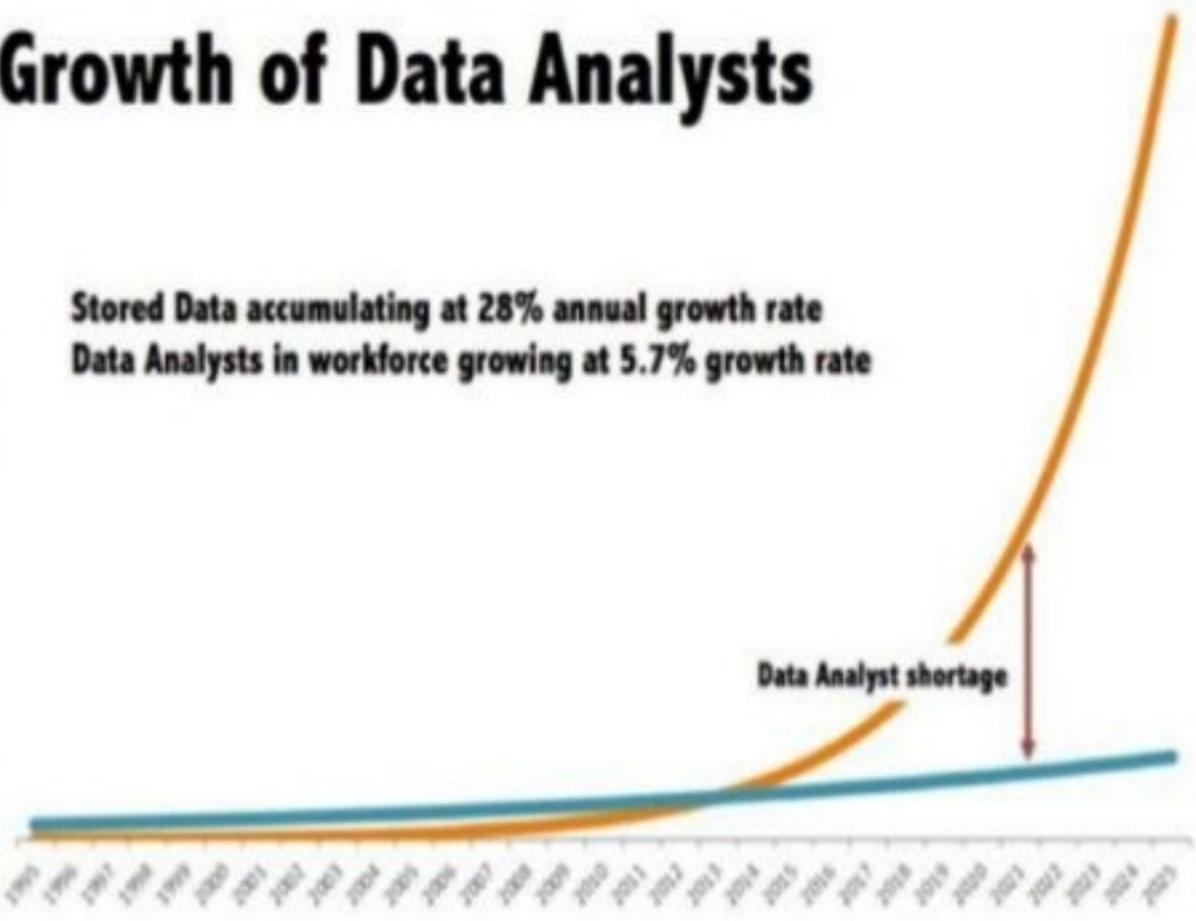
- Why should study Data Science?
- How Does Data Science Impact Organizations?
- Application and Competitive Advantage of Data Science in Organization
- Importance of Data Science to Society
- Road to Become a Data Scientist

# WHY WE ARE TALKING ABOUT DATA SCIENCE?



## Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate  
Data Analysts in workforce growing at 5.7% growth rate



# WHAT IS DATA SCIENCE?

- *"Data Science is a new term. But in the same sense as Columbus was discovered NEW Continent 1000 years ago."*



- Hector Garcia-Molina  
Professor in the Departments  
of Computer Science and  
Electrical Engineering of  
Stanford University

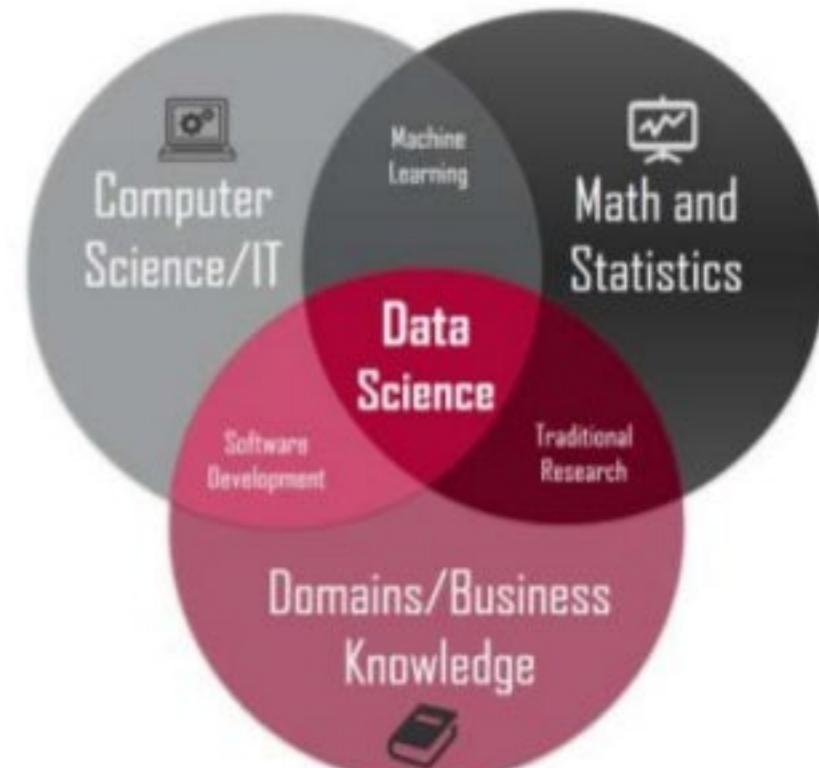
# WHAT IS DATA SCIENCE?

- a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.

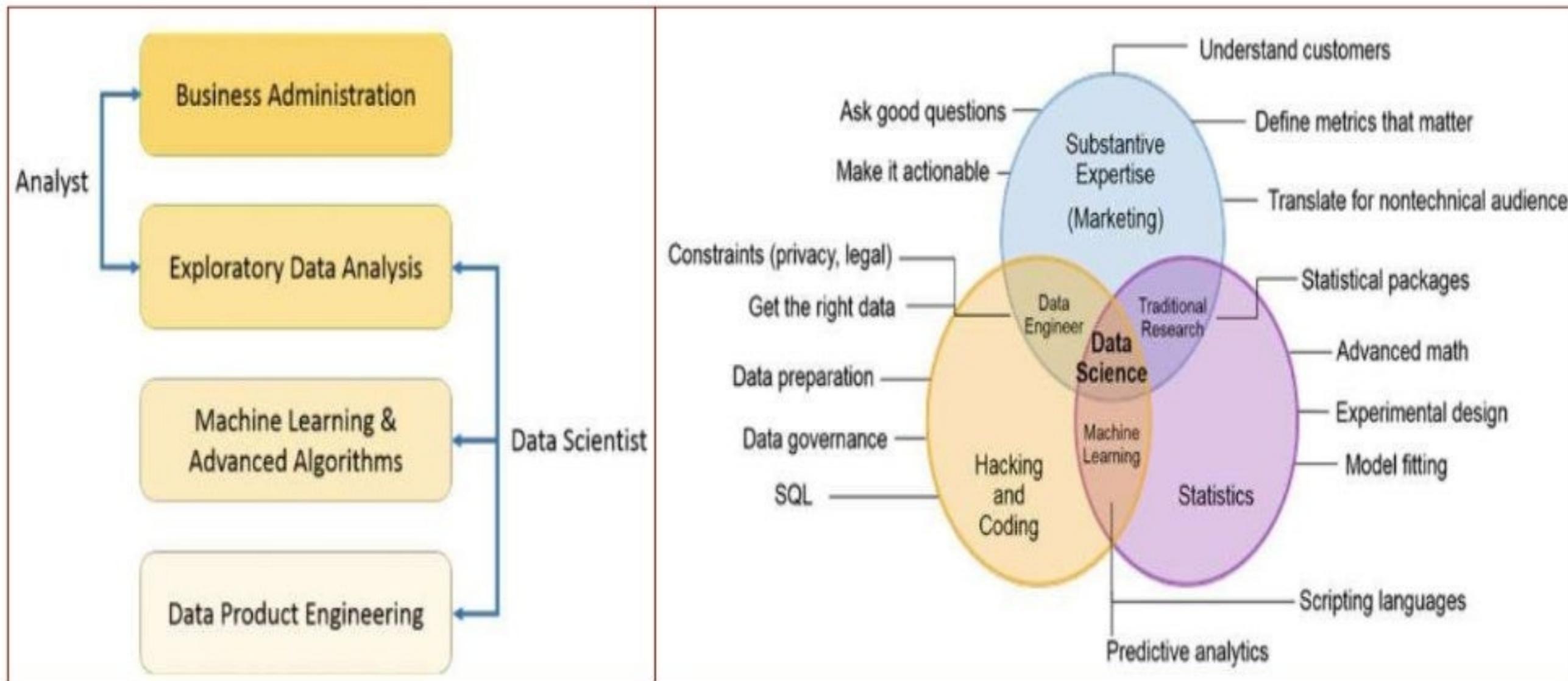


# WHAT IS DATA SCIENCE?

- a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data.
- employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science.



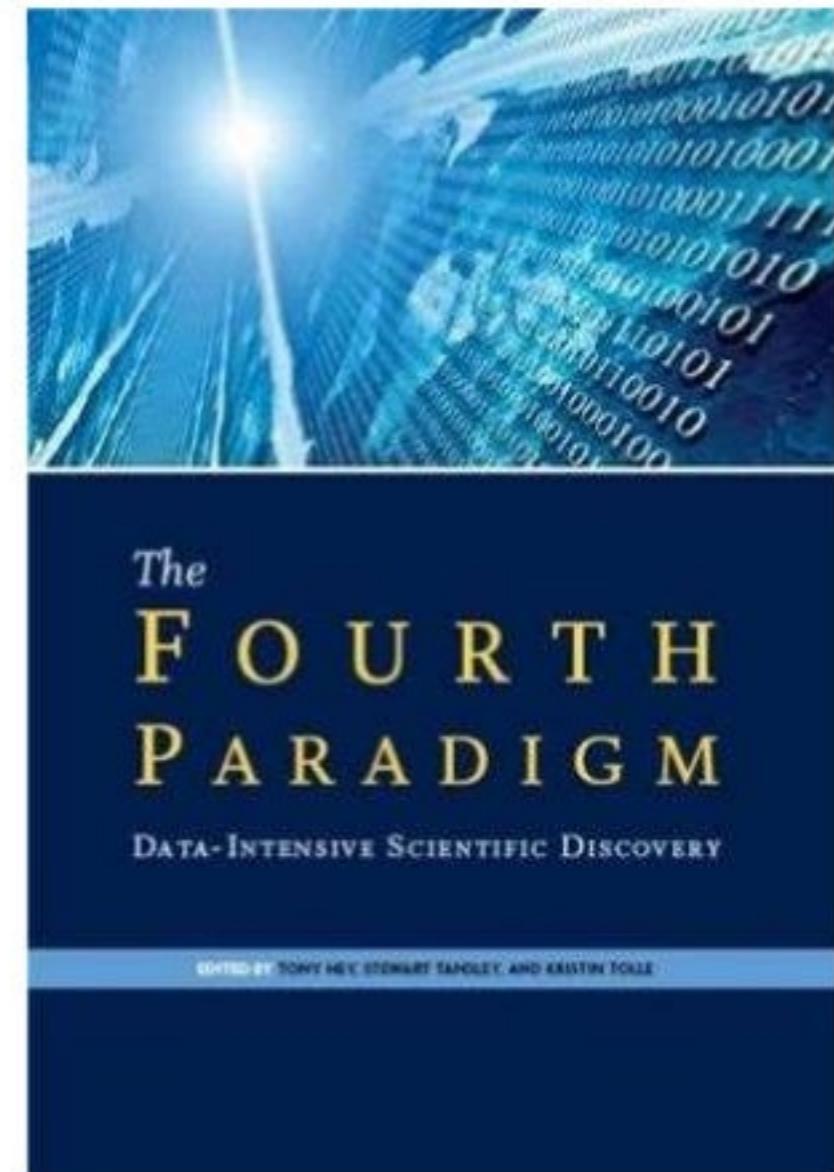
# WHAT IS DATA SCIENCE?



# WHAT IS DATA SCIENCE?

## *Fourth Paradigm of Science*

- Thousand of years
  - Empirical
- Few hundred of years
  - Theoretical
- Last fifty years
  - Computational
  - "Query the world"
- Last twenty years
  - eScience (Data Science)
  - "Download the world"



# WHAT IS DATA SCIENCE?

## Data Science and others

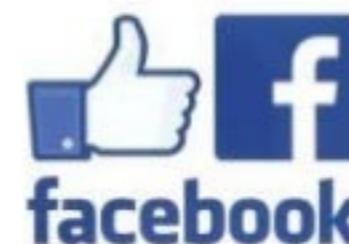
- Statistics
- Big Data Analytics
- Business Analytics
- Business Intelligence
- Data(base) Management
- Visualization
- Machine Learning
- Data Mining
- Artificial Intelligence
- Predictive Modelling



# WHAT IS DATA SCIENCE?

## Big Data Science Tasks

- Facebooks
- Amazon
- Google
- LinkedIn
- Netflix
- Rozetka
- Microsoft



# WHAT IS DATA SCIENCE?

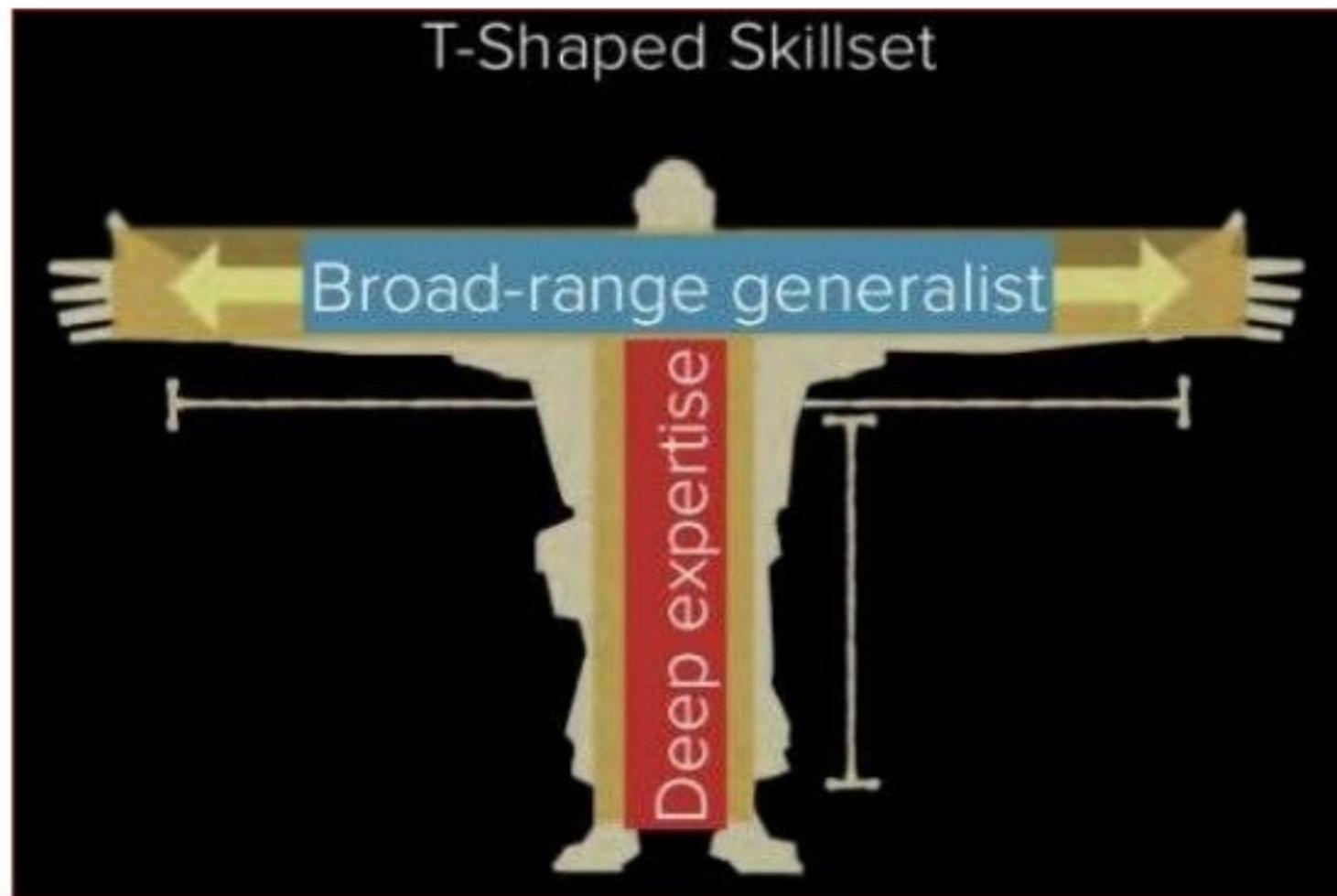
---

## Regular Data Science

- Data Analysis
- Modelling Statistics
- Engineering / Prototyping

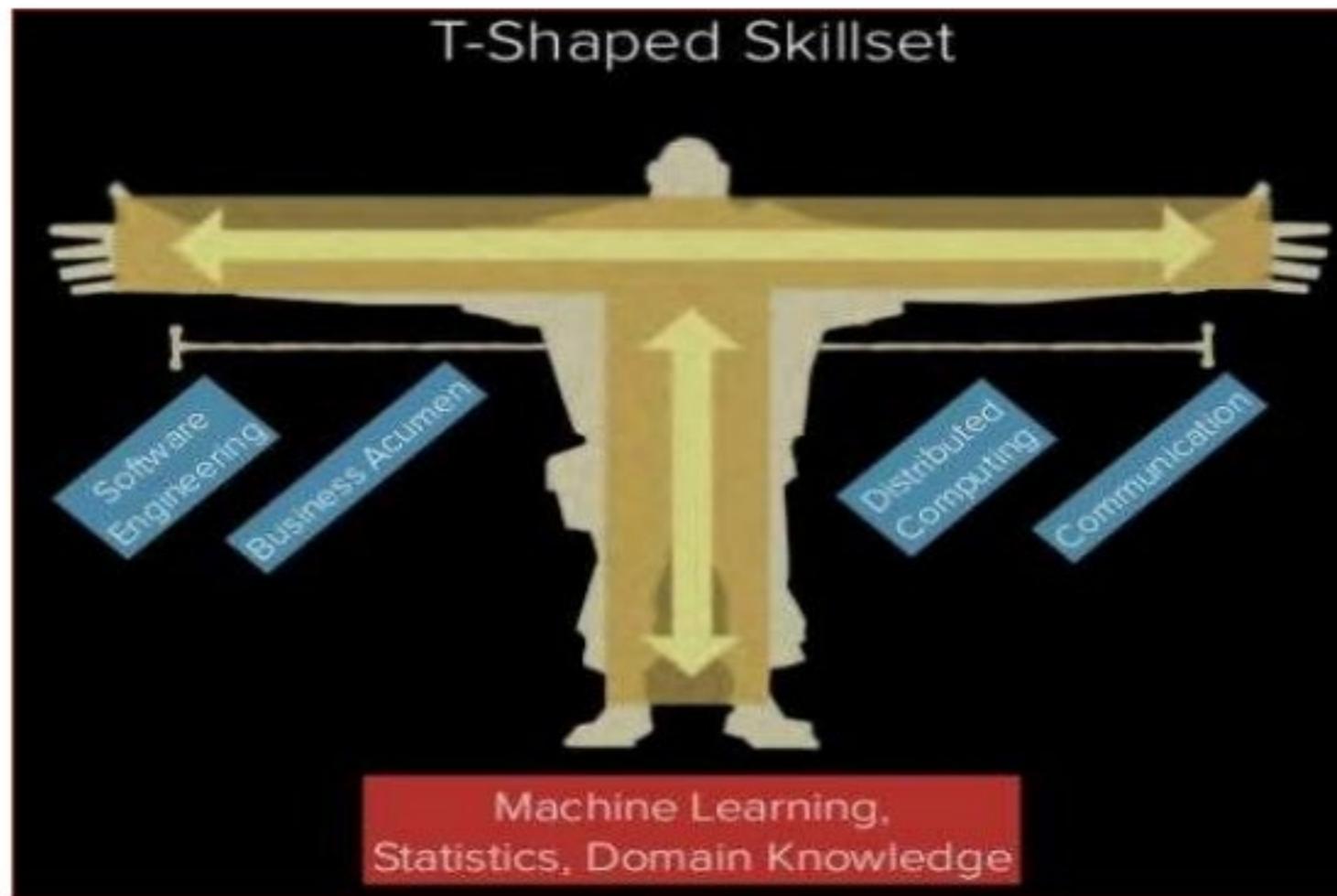
# WHAT IS DATA SCIENCE?

What do people look for in a data scientist?



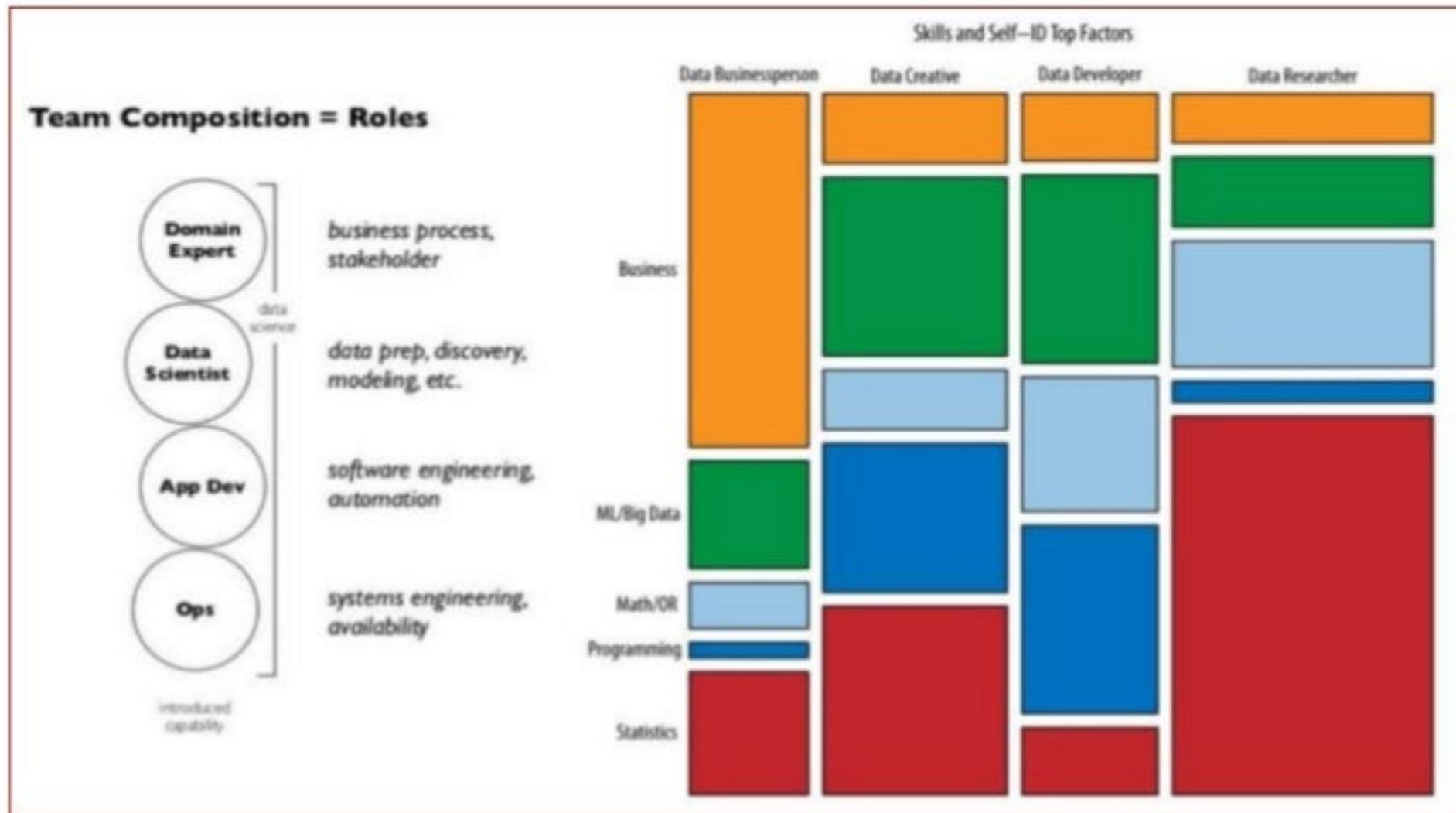
# WHAT IS DATA SCIENCE?

What do people look for in a data scientist?



# WHAT IS DATA SCIENCE?

## Data Science Roles



# WHAT IS DATA SCIENCE?

## Roles Required in Data Science Project

- Prove / disprove hypotheses.
- Information and Data gathering.
- Data wrangling.
- Algorithm and ML models.
- Communication.

**Data  
Scientist**



**Data  
Engineer**



- Build Data Driven Platforms.
- Operationalize Algorithms and Machine Learning models.
- Data Integration.

**Visualization  
Expert**



- Storytelling.
- Build Dashboards and other Data visualizations.
- Provide insight through visual means.

**Process  
Owner**



# WHAT IS DATA SCIENCE?

How to become a data scientist?

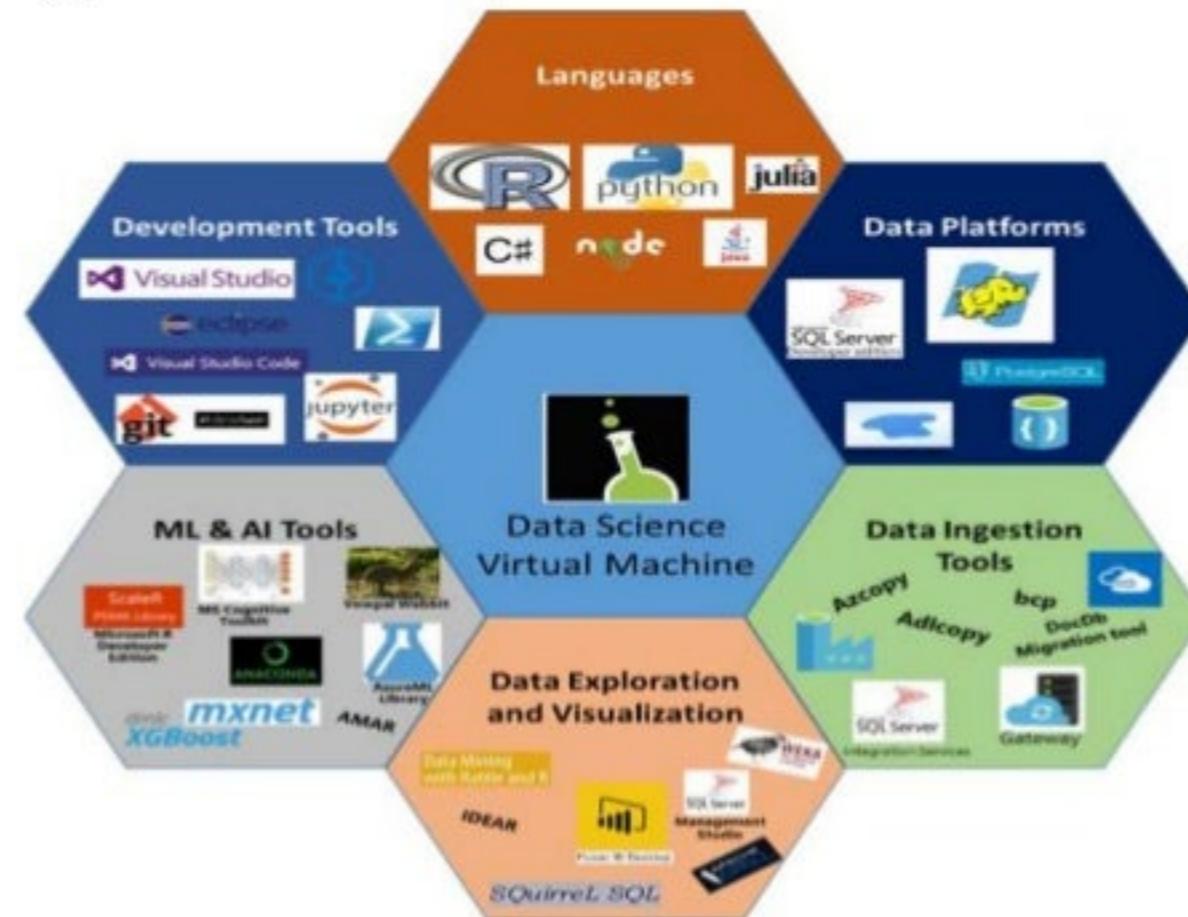
- Data Scientists need to know how to “CODE”



# WHAT IS DATA SCIENCE?

How to become a data scientist?

- Other languages, tools, platforms and visualization



# WHAT IS DATA SCIENCE?

## Learning Data Science with Python - Libraries



### NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.



A free software machine learning library that features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, and k-means and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.



### Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

# WHAT IS DATA SCIENCE?

## Learning Data Science with Python - Libraries



A plotting library for the Python programming language and its numerical mathematics extension NumPy



TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.



Keras is an open source neural network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or MXNet. It was developed with a focus on enabling fast experimentation

# WHAT IS DATA SCIENCE?

## Learning Data Science with Python - Tools



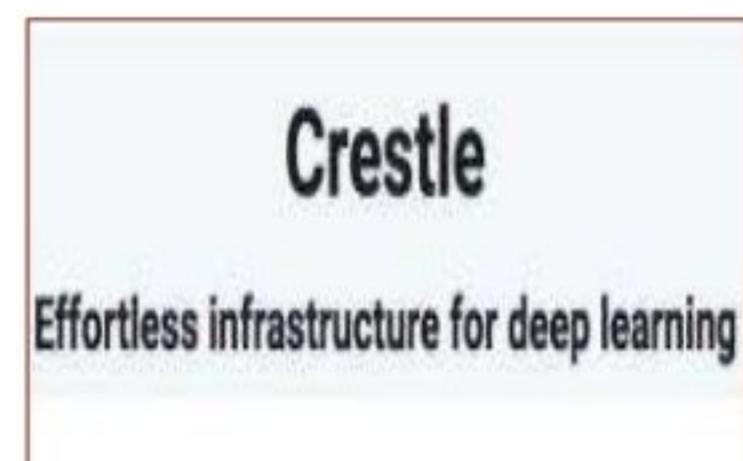
Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text

<http://jupyter.org/>



Similar to Jupyter Notebook, but with the added benefit of "google doc" type sharing and collaboration

<https://colab.research.google.com>



**Crestle**

Effortless infrastructure for deep learning

Crestle is your GPU-enabled Jupyter environment in the cloud.

<https://www.crestle.com/>

# WHAT IS DATA SCIENCE?

How to become a data scientist?

- Learn to code

*Codecademy*



treehouse®

coursera

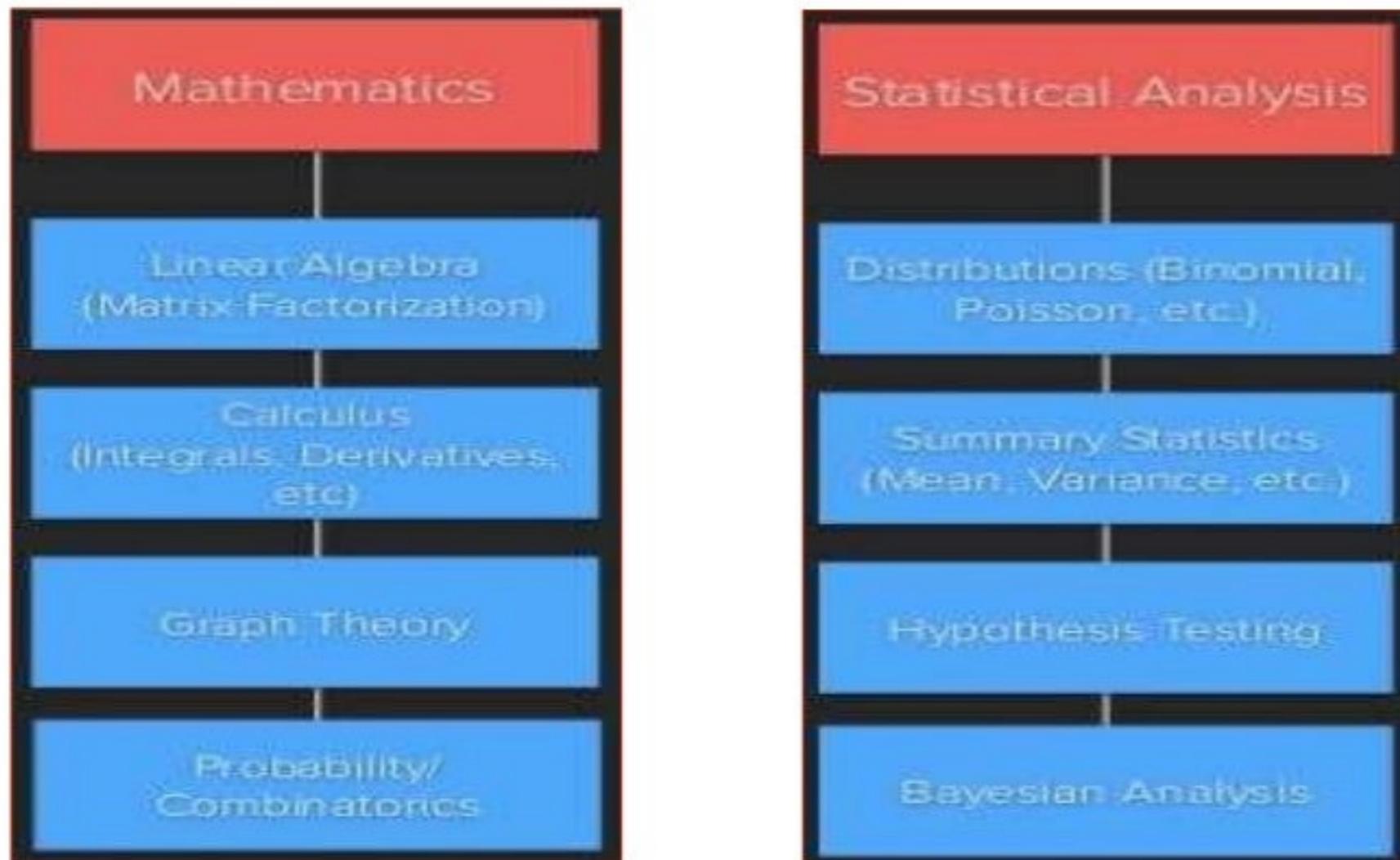
HACKERRANK



Project Euler  
.net

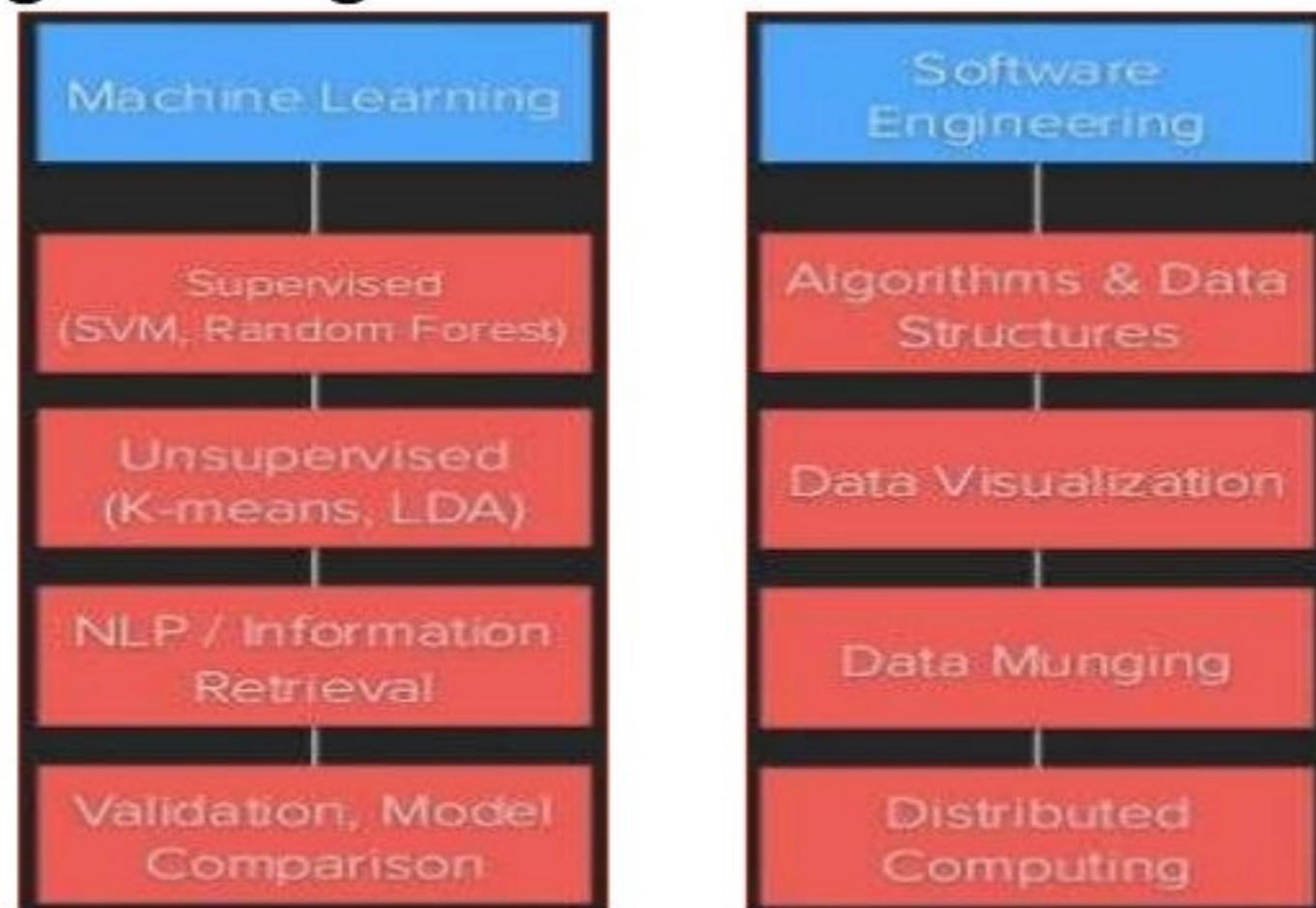
# WHAT IS DATA SCIENCE?

Data Scientist need to comfortable with:



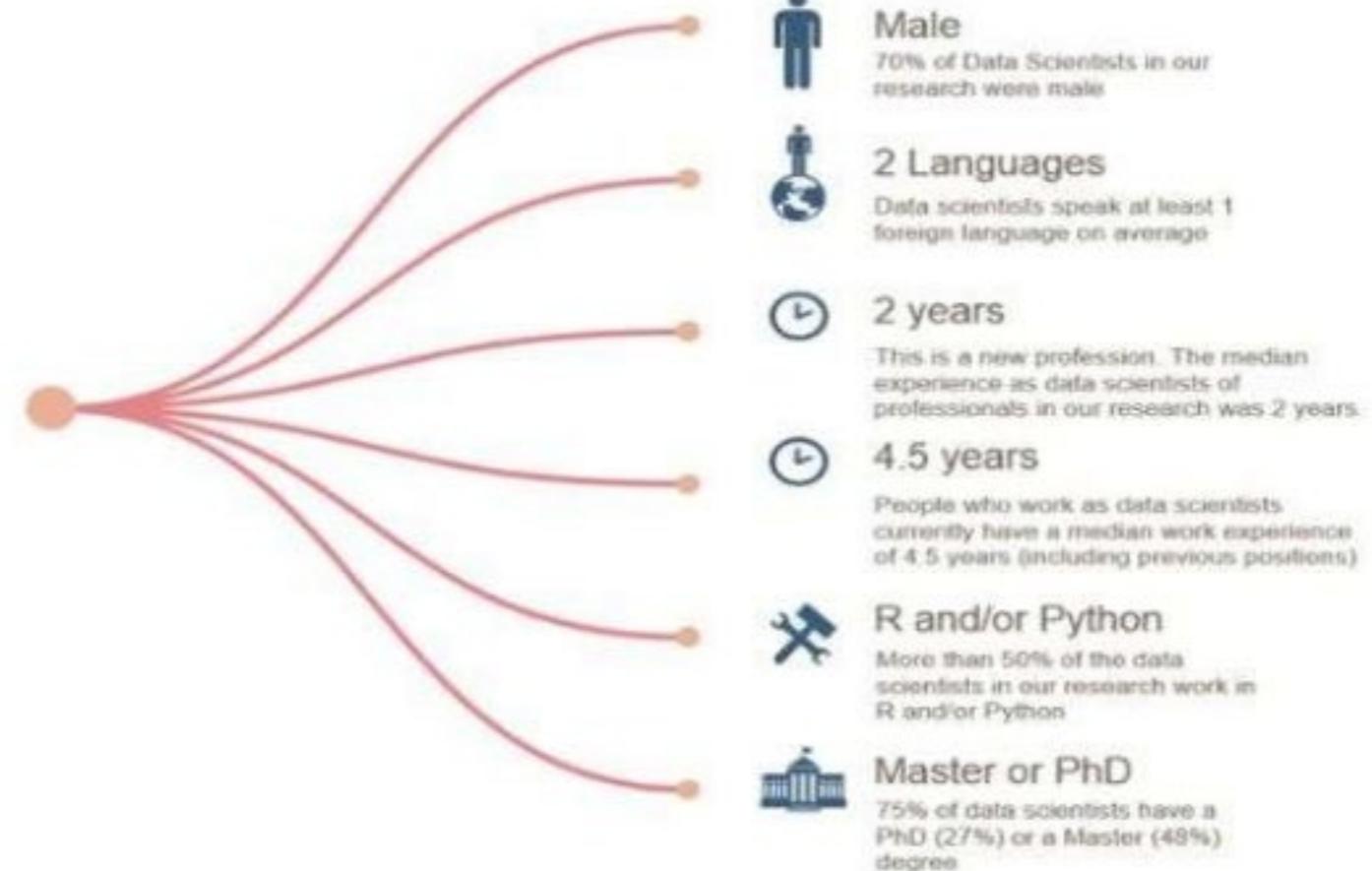
# WHAT IS DATA SCIENCE?

Data Scientist need to learning machine learning & software engineering



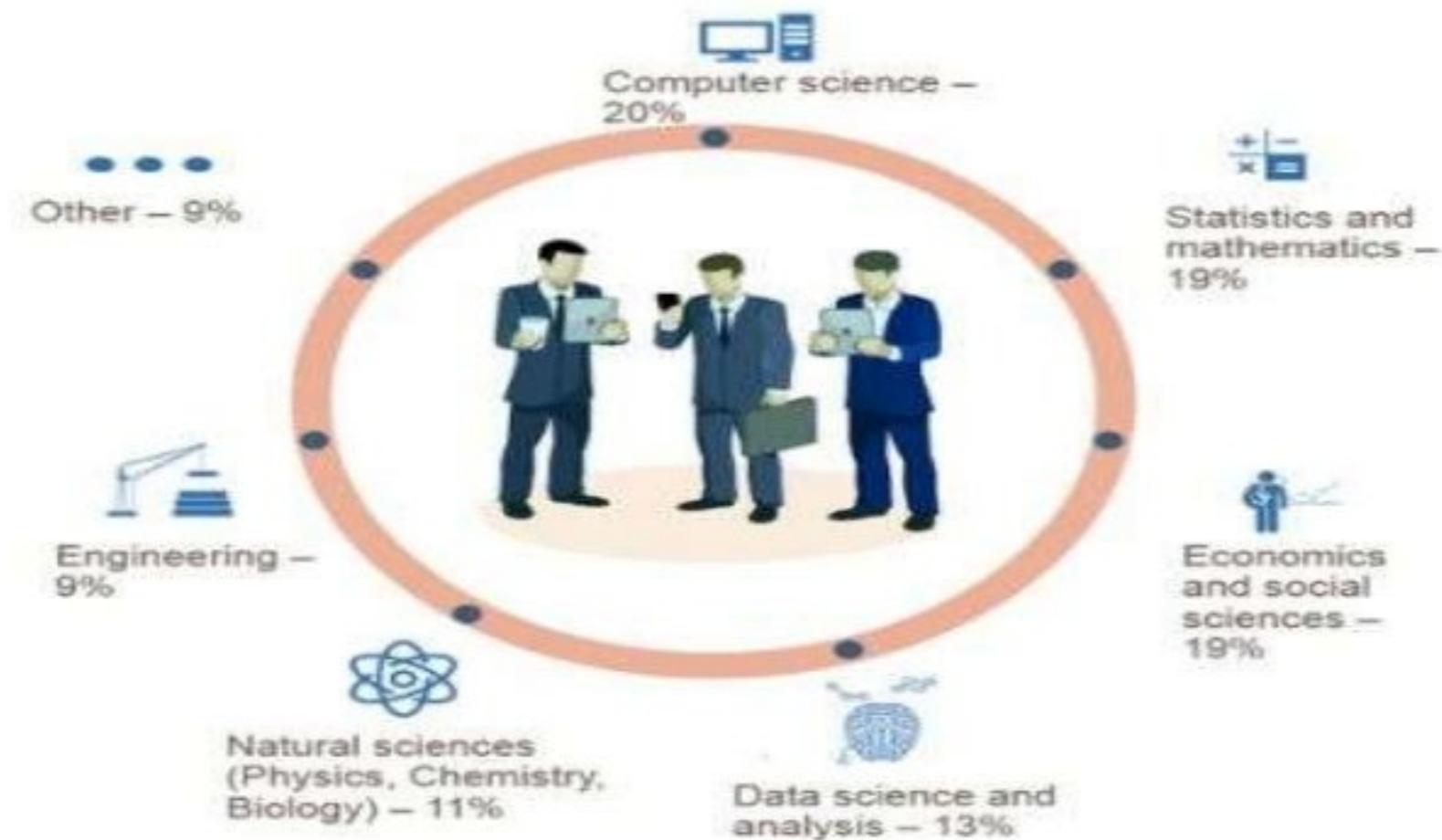
# WHAT IS DATA SCIENCE?

## Who are the Data Scientist?



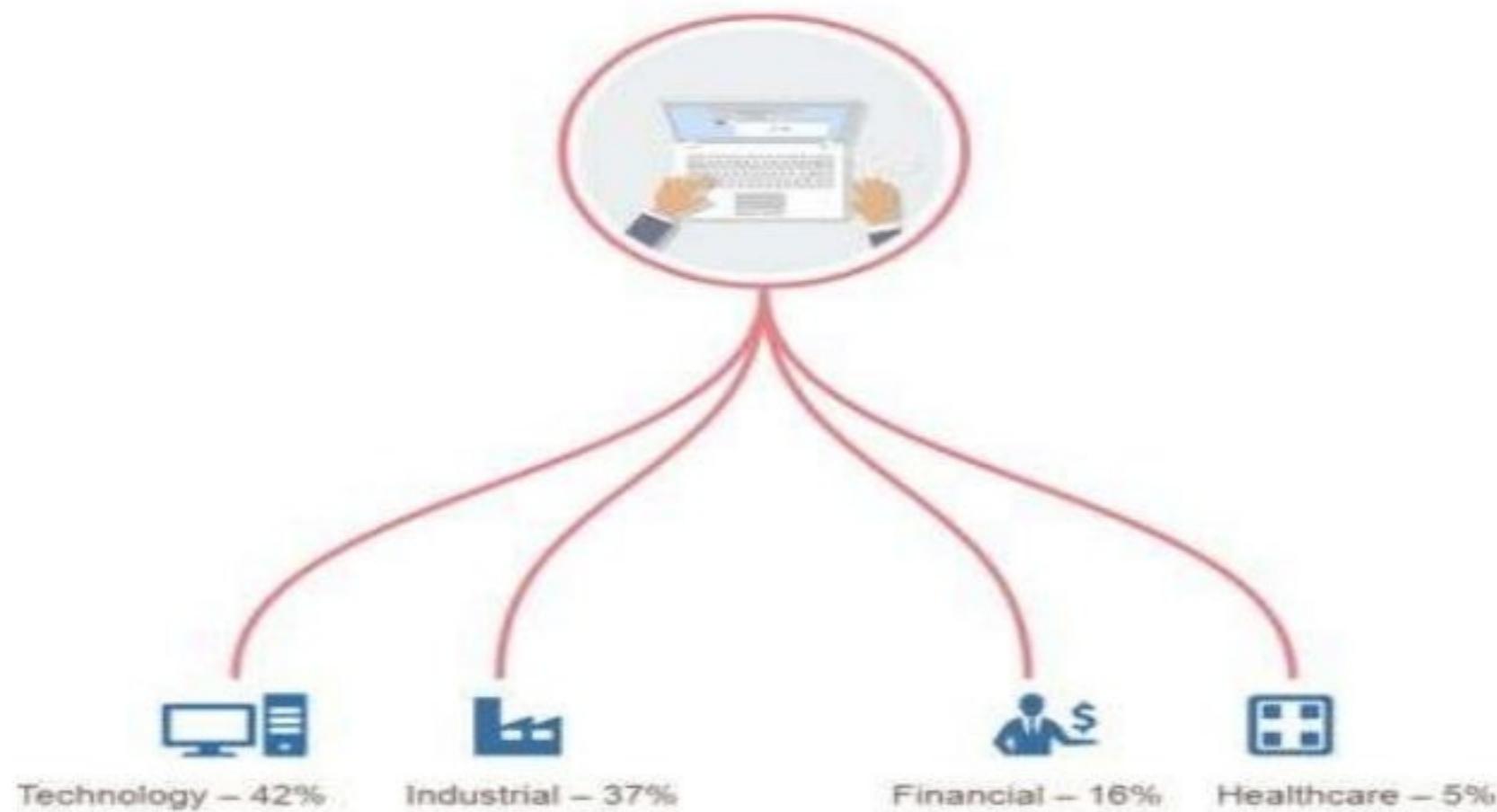
# WHAT IS DATA SCIENCE?

## Who are the Data Scientist?



# WHAT IS DATA SCIENCE?

Who are the Data Scientist?



<https://www.datascience.com/blog/data-scientist-skills>

# WHAT IS DATA SCIENCE?

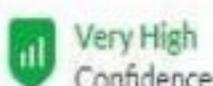
<https://bit.ly/2P4eyl0>



# WHAT IS DATA SCIENCE?

## Data Scientist Salaries in United States

4,354 Salaries Updated Aug 17, 2019



Average Base Pay

**\$117,345** /yr



Additional Cash Compensation ⓘ

Average \$11,530

Range \$3,933 - \$26,784

How much does a Data Scientist make?

The national average salary for a Data Scientist is \$117,345 in United States. Filter by location to see Data Scientist salaries in your area. Salary estimates are based on 4,354 salaries submitted anonymously to Glassdoor by Data Scientist employees.

### Salaries for Related Job Titles

Data Analyst \$67K

Data Scientist Intern \$96K

Quantitative Analyst \$116K

Senior Data Scientist \$137K

# WHAT IS DATA SCIENCE?

## Average Data Scientist Salary in Philippines

PHP 720,000

Avg. Salary

PHP 105,000

BONUS

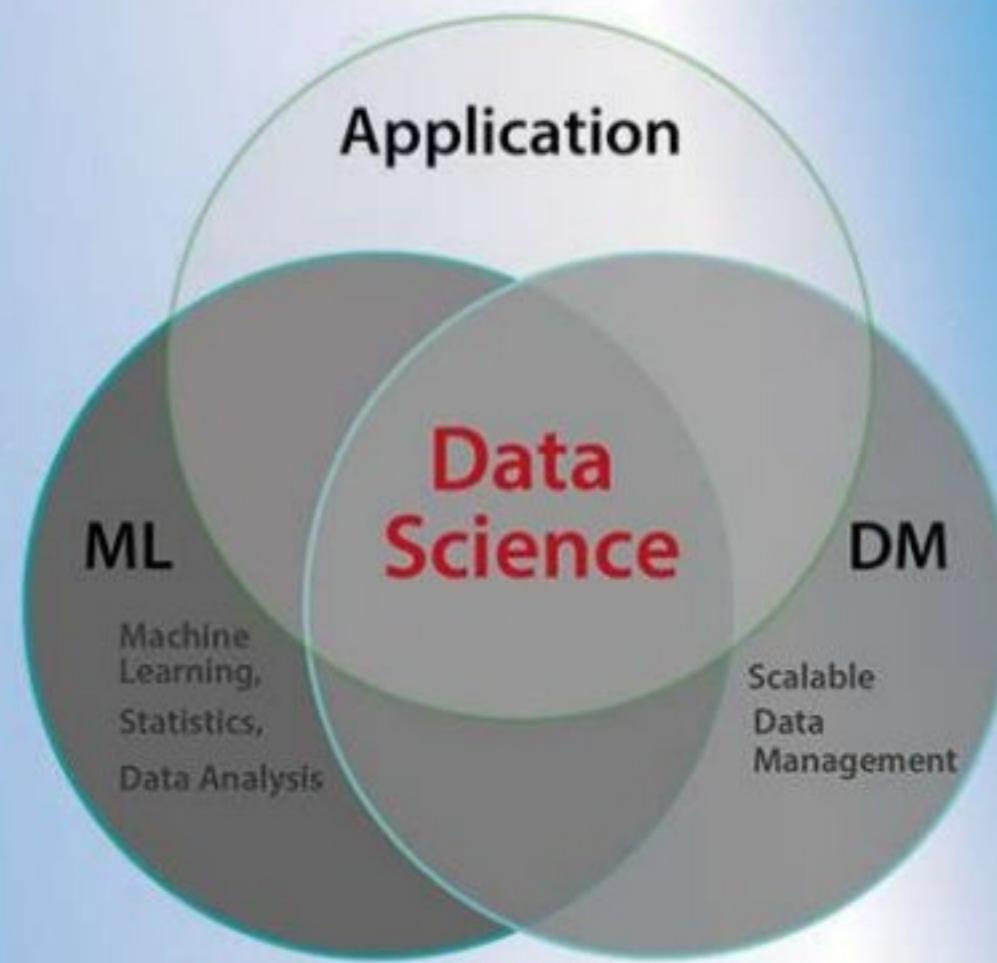
PHP 72,500

PROFIT SHARING

The average salary for a Data Scientist in Philippines is PHP 720,000.



# **Applications of Big Data & Data Science**



## **APPLICATION OF DATA SCIENCE**

# APPLICATIONS OF DATA SCIENCE

- Security



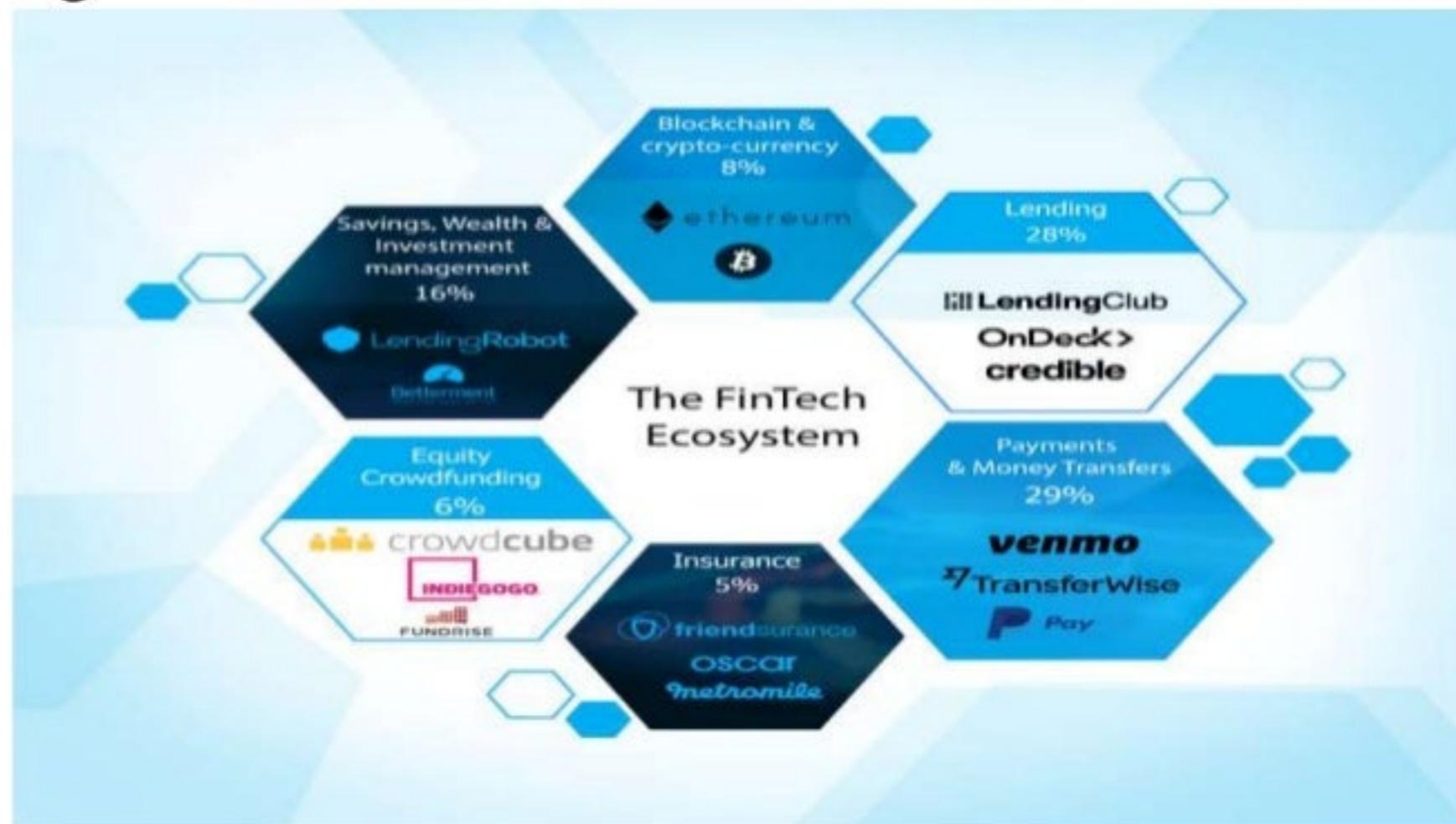
# APPLICATIONS OF DATA SCIENCE

- Sports



# APPLICATIONS OF DATA SCIENCE

- Banking and Finance



# APPLICATIONS OF DATA SCIENCE

- Internet Search

The screenshot shows a Google search results page. The search bar at the top contains the query "how do i know if i am intelligent". Below the search bar, there are tabs for "Web", "News", "Images", "Videos", and "More". A search tools button is also present. The first result is a link titled "20 Scientifically-Proven Signs You're Smarter Than Average..." from www.businessinsider.com. The snippet below the title discusses scientific studies on intelligence. The second result is a link titled "How do I know if I'm smart? - Quora" from https://www.quora.com. The snippet discusses the concept of intelligence. The third result is a link titled "How to know if I am intelligent - Quora" from https://www.quora.com. The snippet discusses what it means to be intelligent. The fourth result is a link titled "Are you Intelligent? - GoToQuiz.com" from www.gototquiz.com. The snippet discusses a quiz about intelligence. The fifth result is a link titled "Am I Dumb Test - Intelligence Test - How Smart Are You?" from www.am-i-dumb.com/. The snippet discusses an intelligence test.

Google how do i know if i am intelligent

Web News Images Videos More Search tools

About 12,80,90,000 results (0.86 seconds)

**20 Scientifically-Proven Signs You're Smarter Than Average...**  
www.businessinsider.com/20-scientifically-proven-signs-youre-smarter-t... •  
Dec 18, 2010 · Many scientific studies have been conducted to determine the cause of high IQs, and a lot of it boils down to genetics. According to researchers ...

**How do I know if I'm smart? - Quora**  
https://www.quora.com/How-do-I-know-if-I'm-smart  
If you ever stop and think, "Hold on a minute, am I actually smart?" ... Among highly intelligent people, there are two kinds of minds, the sharp and the soft.

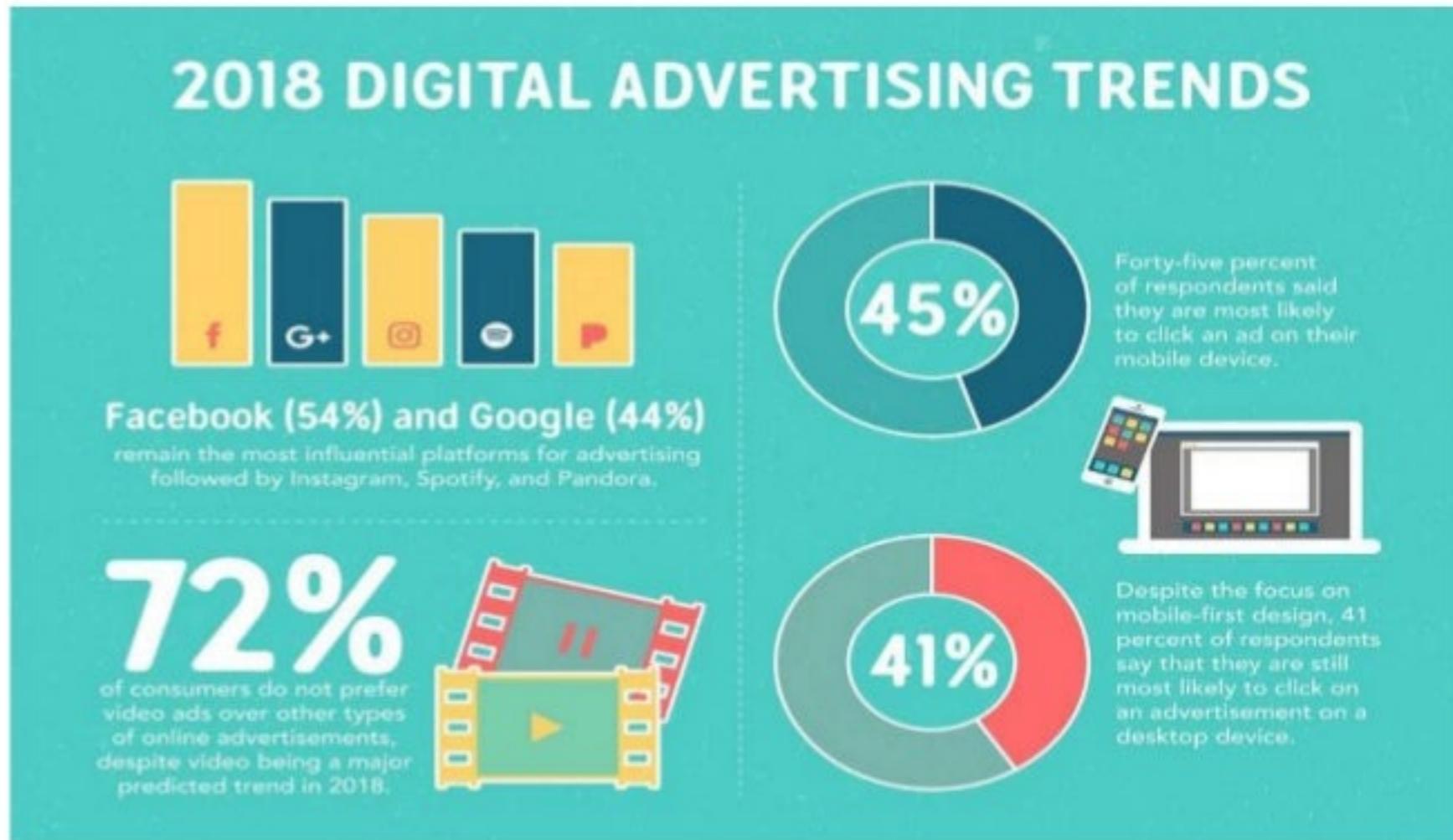
**How to know if I am intelligent - Quora**  
https://www.quora.com/How-do-I-know-if-I-am-intelligent  
First, read up on what it means to be intelligent, the good and the bad. ... How do I know if I am actually smart and my intelligence is not a delusion resulting from ...

**Are you Intelligent? - GoToQuiz.com**  
www.gototquiz.com/are\_you\_intelligent\_3 •  
Or are you the type of person who doesn't even know what the alphabet is? Or maybe ... If you have 5 apples and take away 2, how many do you have? 5. 7. 3. 2

**Am I Dumb Test - Intelligence Test - How Smart Are You?**  
www.am-i-dumb.com/ •  
The Am I Dumb Test is a free intelligence test that will reveal your true ... Find out how much you're worth on the open human market. See if you're a ... Take the intelligence test and we'll tell you how you compare to the rest of the world.

# APPLICATIONS OF DATA SCIENCE

- Digital Advertisements



# APPLICATIONS OF DATA SCIENCE

- Recommender System

The image displays two side-by-side screenshots of digital platforms using recommender systems.

**Left Screenshot (Netflix):**

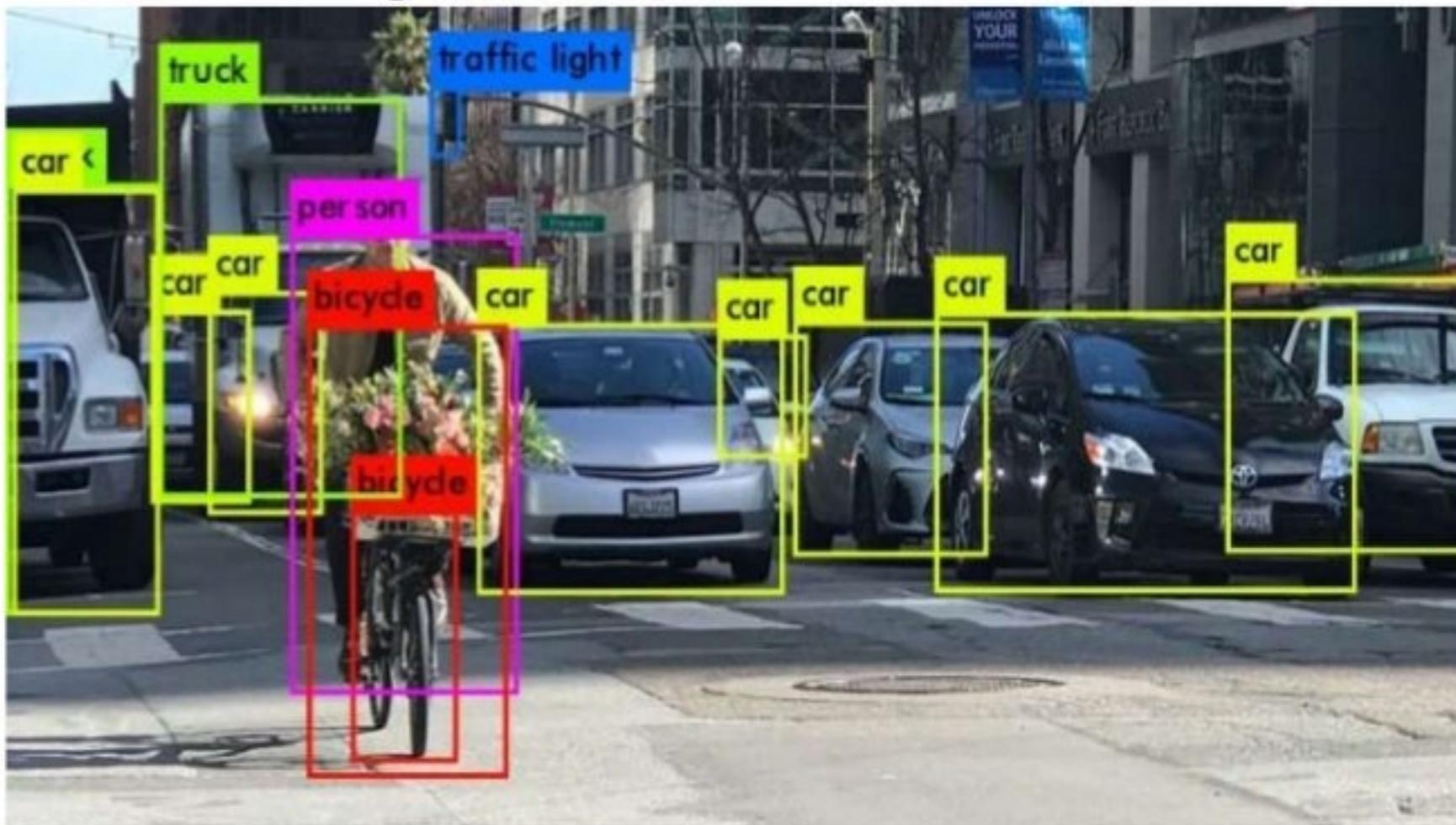
- The top section shows a "Congratulations!" message: "Movies we think You will ❤️".
- The main content area displays a grid of movie thumbnails, including "Spider-Man 2", "The Hobbit: Desolation of Smaug", "The Hobbit: The Desolation of Smaug", "Mad Max: Fury Road", "Lee Daniels' The Butler", "The Last Stand", "Star Wars: Episode III - Revenge of the Sith", and "Star Trek: Into Darkness".
- The bottom section features a sidebar titled "Events" with a search bar and a list of birthday reminders: "Birthdays: Kunal Vohra, Nate McRae".
- A red arrow points to the "People You May Know" section, which lists "Harry Zamacois Perte" and "Gina Spahn", each with a profile picture, mutual friend count (85 and 103 respectively), and an "Add as friend" button.

**Right Screenshot (YouTube):**

- The top right corner has an "Autoplay" toggle switch.
- The "Up Next" section lists several video thumbnails:
  - "YouTube Rewind 2014: Behind the Scenes" by YouTube Spotlight (3,212,337 views)
  - "REWINDPICTURES" by YouTube Spotlight (106,428,885 views)
  - "Rewind YouTube Style 2012" by YouTube Spotlight (154,722,053 views)
  - "Girls' Generation "I Got a Boy" Wins Video of the Year - Live at the YTMAs" by YouTube Spotlight (350,421 views)
  - "Mind the Gap: The Making of unreal Episode 1" by Teton Gravity Research (Recommended for you)
  - "BEST NEWS BLOOPERS 2014 Part 2" by NewsBiteFunny (1,010,904 views)

# APPLICATIONS OF DATA SCIENCE

- Image Processing



# APPLICATIONS OF DATA SCIENCE

- Speech Recognition



# APPLICATIONS OF DATA SCIENCE

- Gaming



# APPLICATIONS OF DATA SCIENCE

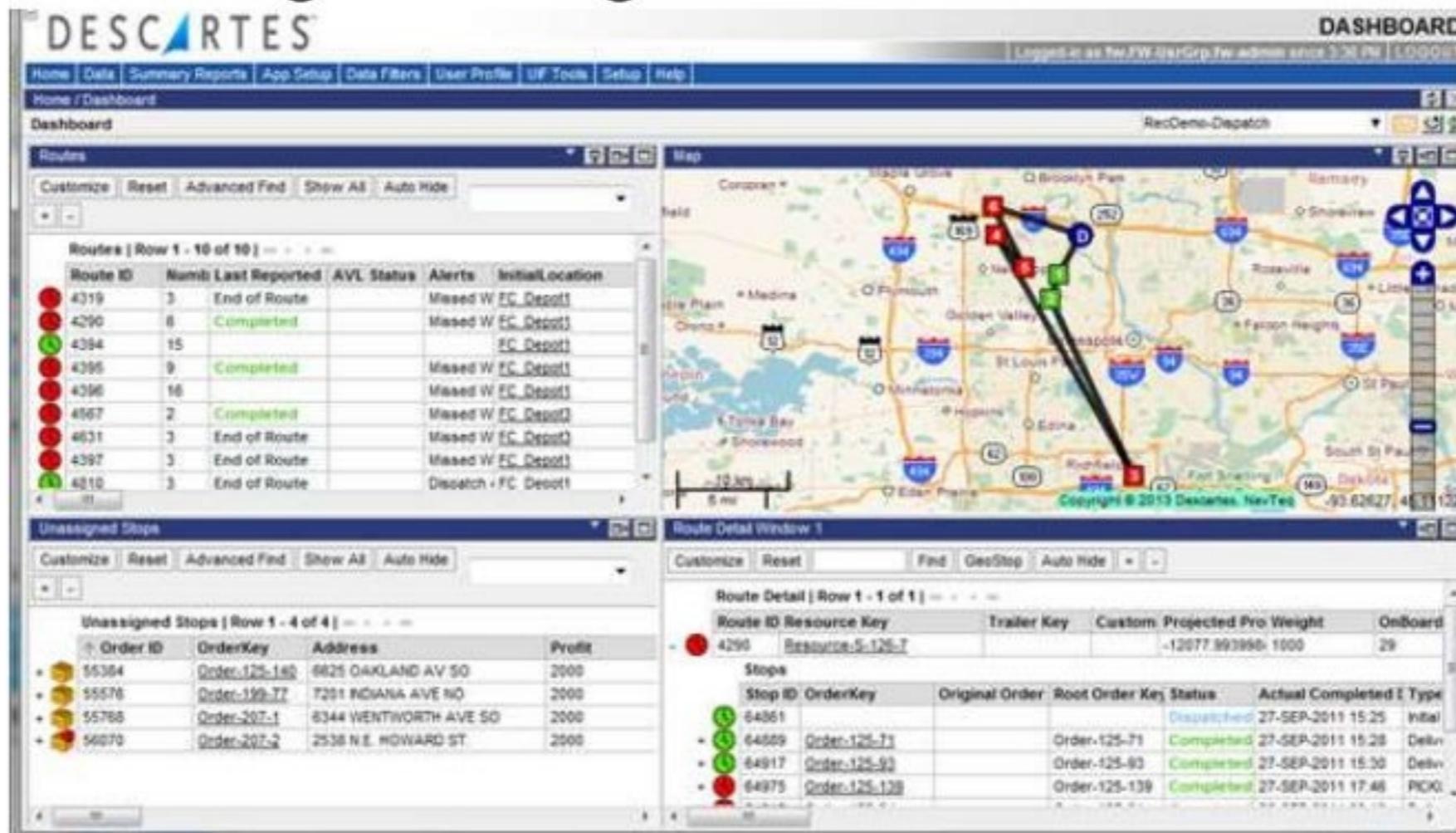
- Price Comparison Websites

## 10 Price Comparison Websites



# APPLICATIONS OF DATA SCIENCE

- Airline Routing Planning



# APPLICATIONS OF DATA SCIENCE

- Fraud and Risk Detection



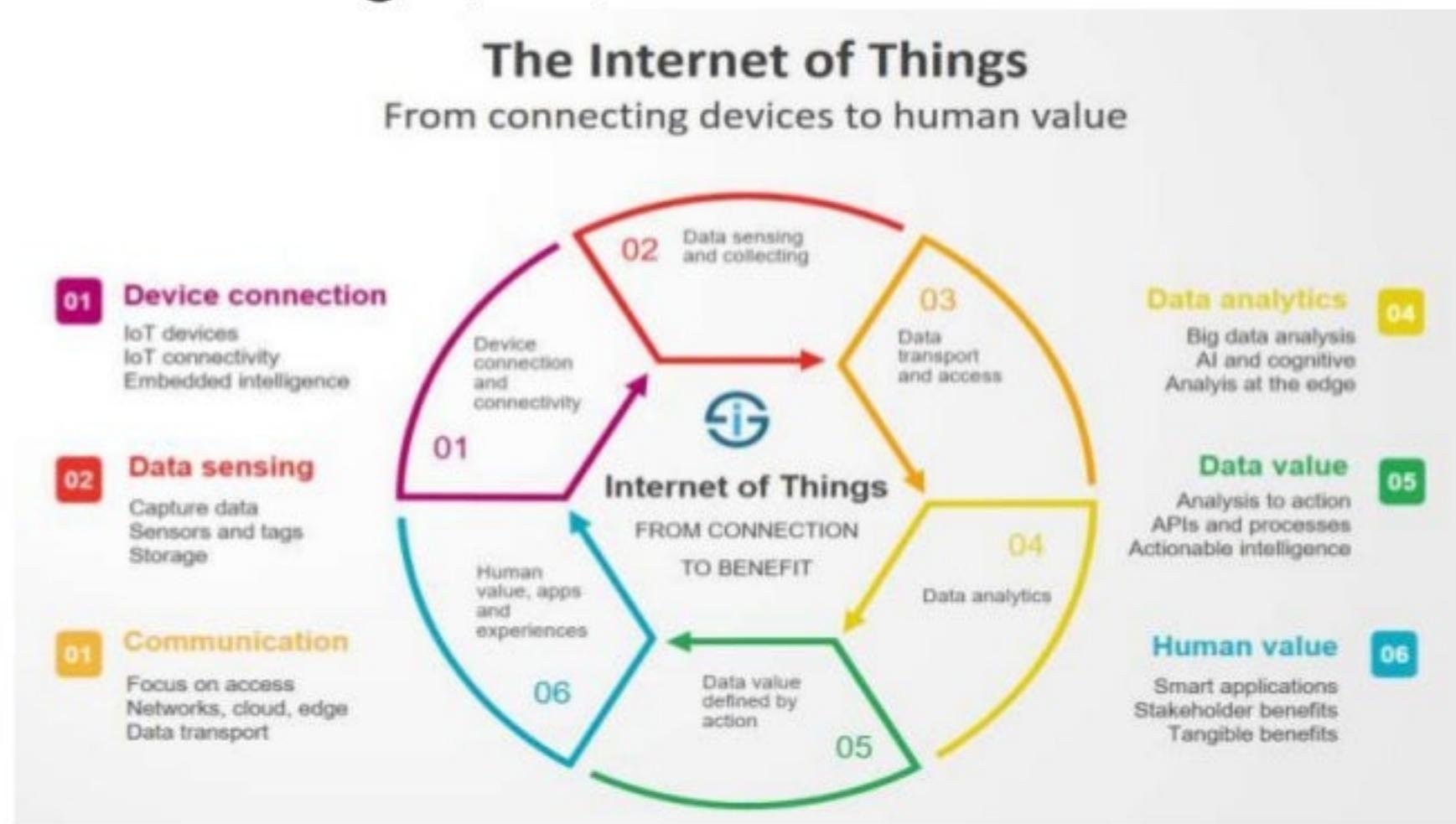
# APPLICATIONS OF DATA SCIENCE

- Delivery Logistics



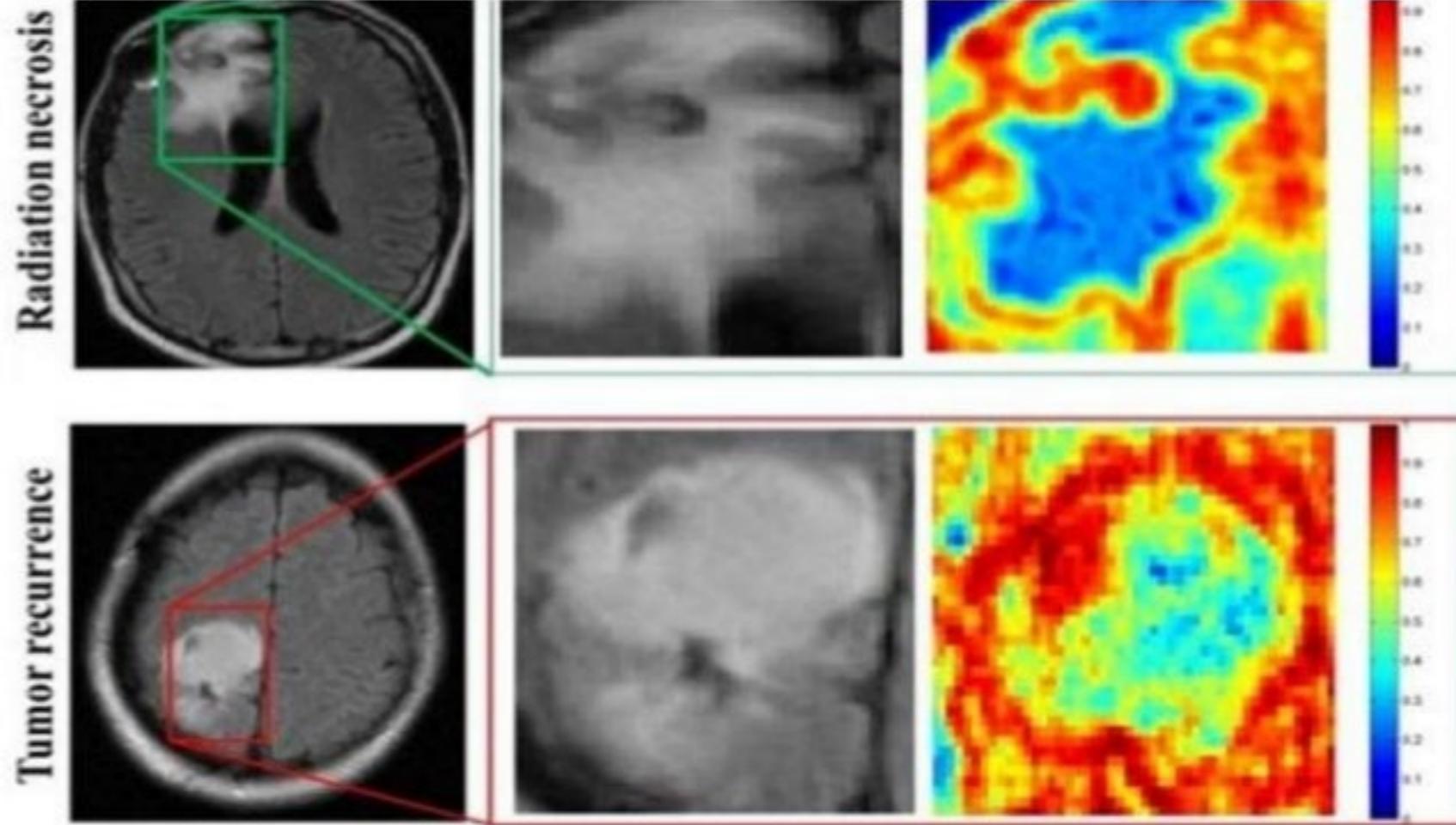
# APPLICATIONS OF DATA SCIENCE

- Internet of Things (IoT)



# APPLICATIONS OF DATA SCIENCE

- Health Care



# APPLICATIONS OF DATA SCIENCE

- Augmented Reality



# APPLICATIONS OF DATA SCIENCE

- Self-Driving Cars



# APPLICATIONS OF DATA SCIENCE

- Robots



# USING DATA SCIENCE TO BUILD ACTIONABLE INSIGHTS



## IMPACT OF DATA SCIENCE ON SOCIETY

# IMPACT OF DATA SCIENCE ON SOCIETY

- Saving Energy



# IMPACT OF DATA SCIENCE ON SOCIETY

- Data-Driven Hospitals



# IMPACT OF DATA SCIENCE ON SOCIETY

- A Cleaner Environment



# IMPACT OF DATA SCIENCE ON SOCIETY

- Volunteer with a socially-oriented data science program/organization



## **DATA SCIENCE SYMPOSIUM 2019**

***October 10 & 11***

A photograph of a modern, multi-story building with a large glass facade and a steel frame. The building has a distinctive angular design. In the foreground, there is a paved area with some small trees and a metal fence. Overlaid on the left side of the image is a white rectangular box containing a bulleted list of event details.

- **2 Days of Data Science**
- **3 Keynotes**
- **16 Tech Talks**
- **Python, R, Machine Learning, Data Visualization, Data Management, AI ...**
- **Special Attendee Reception**

# IMPACT OF DATA SCIENCE ON SOCIETY

- Contribute via competitions



# IMPACT OF DATA SCIENCE ON SOCIETY

- Consider solutions to real-world problems that you encounter



# IMPACT OF DATA SCIENCE ON SOCIETY

- Be thoughtful in professional work





# IMPORTANCE OF DATA SCIENCE

# IMPORTANCE OF DATA SCIENCE

---

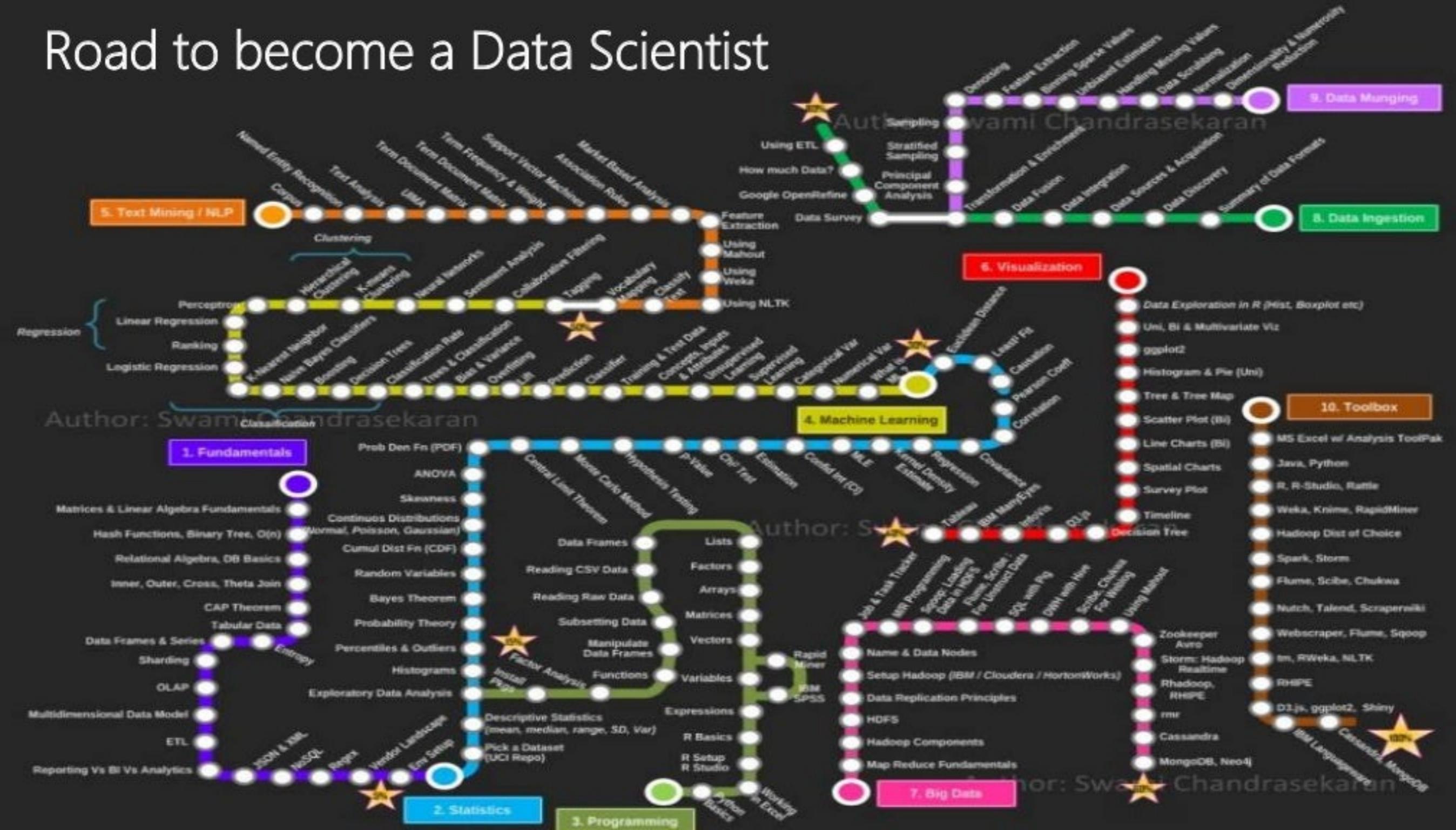
1. Data science helps brands to understand their customers in a much enhanced and empowered manner.
2. It allows brands to communicate their story in such an engaging and powerful manner.
3. Big Data is a new field that is constantly growing and evolving.

# IMPORTANCE OF DATA SCIENCE

---

4. Its findings and results can be applied to almost any sector like travel, healthcare and education among others.
5. Data science is accessible to almost all sectors.

# Road to become a Data Scientist



# REFERENCES

---

- <https://slideplayer.com/slide/10398517/>
- <https://www.slideshare.net/ryanorban/how-to-become-a-data-scientist>
- Dhar, V. (2013). "Data science and prediction". Communications of the ACM. 56 (12): 64–73. doi:10.1145/2500499.
- Hayashi, Chikio (1 January 1998). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann; Ohsumi, Noboru; Tanaka, Yutaka; Baba, Yasumasa (eds.). Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1\_3. ISBN 9784431702085.
- Davenport, Thomas H.; Patil, DJ (October 2012), Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review
- Jeff Leek (12 December 2013). "The key word in "Data Science" is not Data, it is Science". Simply Statistics.
- <https://www.analyticsvidhya.com/blog/2015/09/applications-data-science/>
- <https://www.edureka.co/blog/data-science-applications/>
- <https://dutchdatascienceweek.nl/2018/04/05/the-impact-of-data-science-on-society/>
- <https://www.educba.com/data-science-and-its-growing-importance/>