<> Developer Student Clubs

An Introduction to Data Science

Updated 10.22.19

Google Developers

An Introduction to Data Science

"Data is now considered as Electricity for industries and corporations, so those who will not become data driven will face a black out"
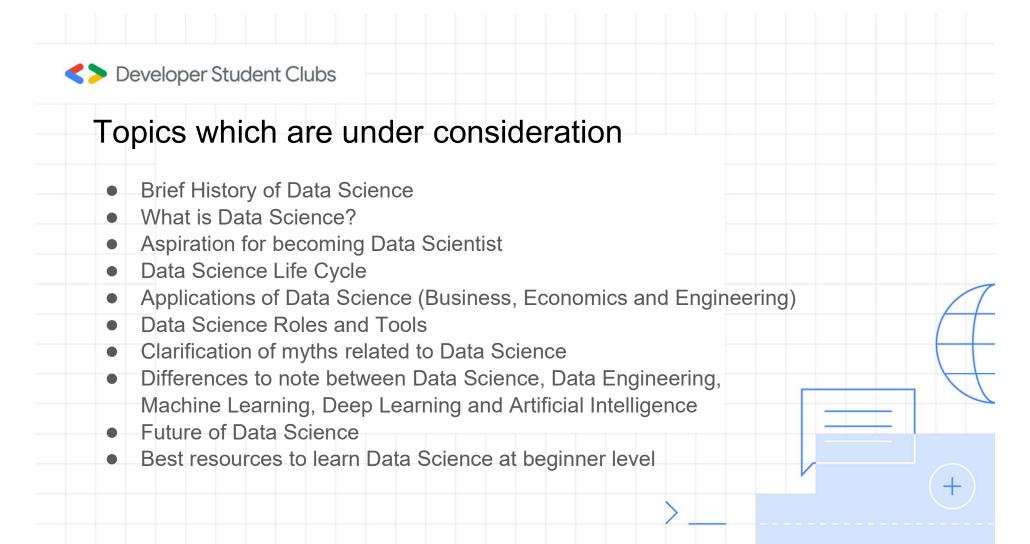~MUB

Information Session by:

Muhammad Umair Baig
MS Data Science in Progress -NED
B.E Electrical Engg –NED
Founder of Engineering Digitization (working for democratizing digital revolution)

Developer Student Clubs

Topics which are under consideration

- Brief History of Data Science
- What is Data Science?
- Aspiration for becoming Data Scientist
- Data Science Life Cycle
- Applications of Data Science (Business, Economics and Engineering)
- Data Science Roles and Tools
- Clarification of myths related to Data Science
- Differences to note between Data Science, Data Engineering, Machine Learning, Deep Learning and Artificial Intelligence
- Future of Data Science
- Best resources to learn Data Science at beginner level

Brief History of Data Science

- Data originates from the **Latin** word, "**datum**," which means a "something given" The expression "data" has been utilized since 1500s, however, the modern practice began during the **1940s** and **1950s**.

- 1962 – JOHN WILDER TUKEY
  John Wilder Tukey expounded on a move in the world of statistics in 1962. He alluded to the converging of **statistics** and **PCs**, when factual outcomes introduced in hours, as opposed to **hand-rolled work** that may take a long time of several days or a month.

- 1974 – PETER NAUR
  In 1974, Peter Naur wrote the **Concise Survey of Computer Methods**, utilizing the expression "**Data Science**," over and over.

- 1977 – JOHN WILDER TUKEY
  John Wilder Tukey composed a second paper titled "**Exploratory Data Analysis**" in Jan-1977

Developer Student Clubs

- **1996 – INTERNATIONAL FEDERATION OF CLASSIFICATION SOCIETIES**

  Members of this society hold a conference with title "**Data Science, Classification and Relative Methods**", this title was used first time in the history for any conference.

- **1997 – C.F. JEFF WU**

  He gave **introductory lecture** in which he talked about **Statistics** and called it as a branch of "Data Science".

- **2001 – WILLIAM S. CLEVELAND**

  William S. Cleveland of **Bell Labs** published a paper, "**Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics**" in which he talked about increasing **Computing Power** and **Data**. It was published in **International Statistical Analysis** .

- **2002 – THE INTERNATIONAL COUNCIL FOR SCIENCE**

  Council started to publish **Data Science Journals** in which main concentration was related to **highlighting problems** which can be solved in **Data Science domain**. Main topics were; explanation of **Data systems**, publishing **Data systems on internet**, **Data Science Applications** and **Law issues** related to them.
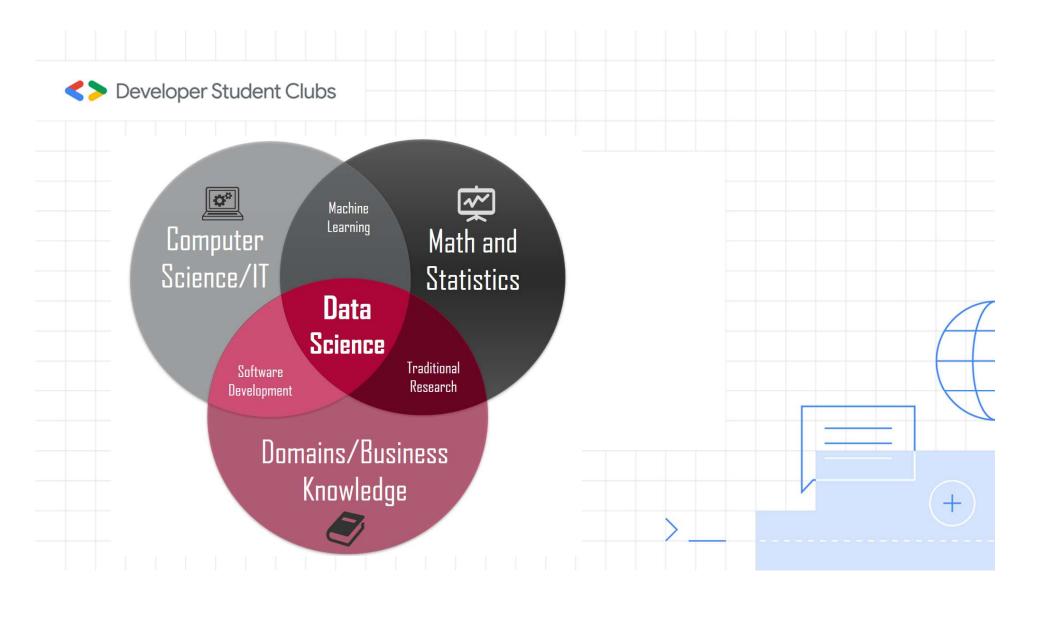
- 2005 – NATIONAL SCIENCE BOARD

  They wrote descriptive essay on "**Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century**" in which they clearly defined role of a Data Scientist as well as work of Data Science.

- 2009 – TITLE DATA SCIENTIST

  The title, "**Data Scientist**" turned into a trendy expression and in the long run a piece of the language. **Jeff Hammerbacher** and **DJ Patil** of **Facebook** and **LinkedIn** are given acknowledgment for starting its utilization as a trendy expression.
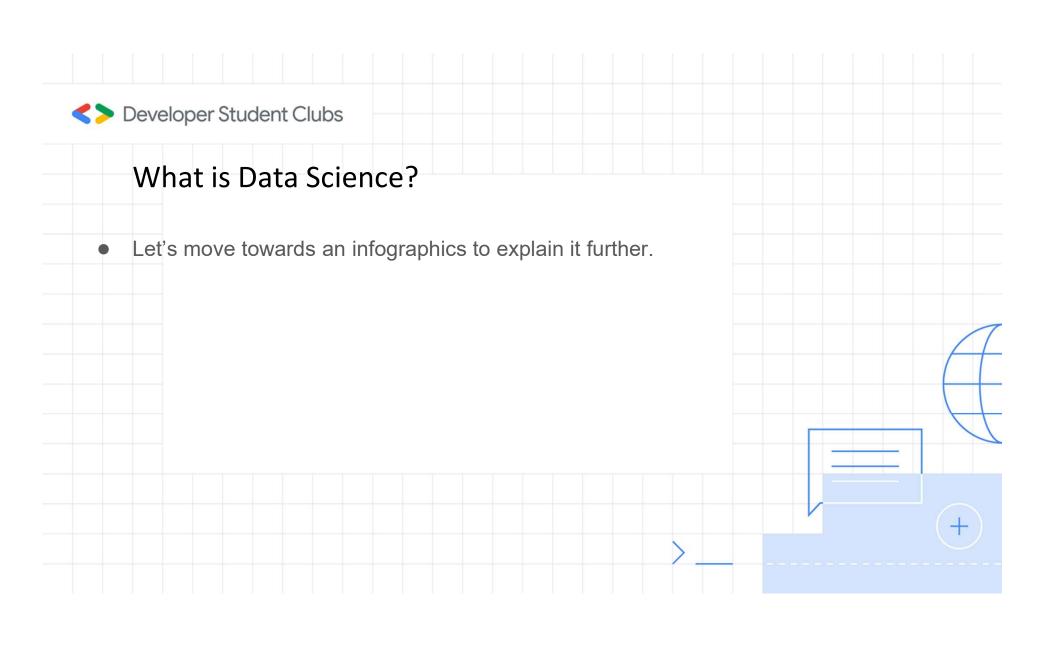
- FINAL WORDS

  The concept of **Data Lakes** was promoted by **James Dixon** in **2011**. According to his concept, at the point of entry, **Data Warehouse pre-categorizes the data**, whereas, **Data Lake accepts the data by utilizing a non-relational Database i.e. NoSQL** without categorizing the data by stores it, which save **time** and **energy** as well. According to the statistics shared by the **IBM** in **2012**, about **90%** of the **digital data** had been produced in past two years.

What is Data Science?

- Data Science, despite its hype-laden veneer, is perhaps the best label we have for the cross-disciplinary **set of skills** that are becoming increasingly important in many applications across industry and academia.
(JakeVanderPlas, Python Data Science Handbook)

- Think Data Science not as new domain of knowledge to learn, but as a new set of skills that you can apply within your **current area of expertise**.
(JakeVanderPlas, Python Data Science Handbook)

- It is about extracting **useful insights** from the data in order to add business value or to solve complex problems.

- It is about uncovering hidden information that may be useful to help companies make **smarter choices** for their business.

Developer Student Clubs

What is Data Science?

- Let's move towards an infographics to explain it further.

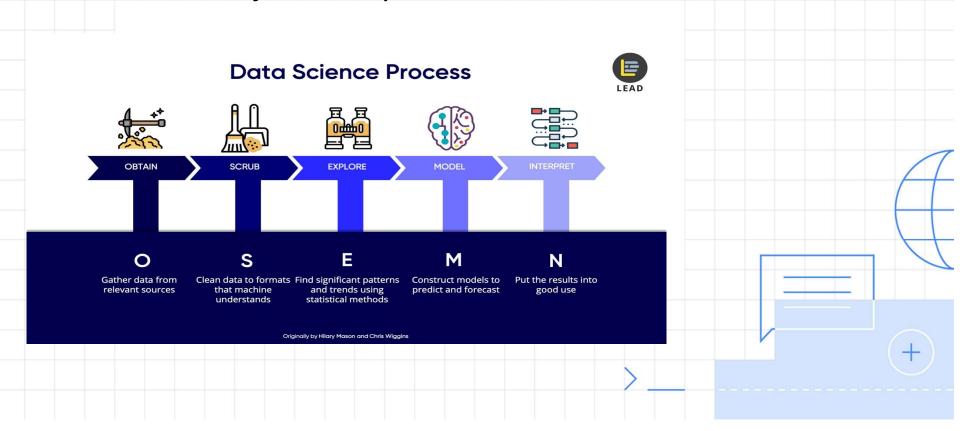# Aspiration For Becoming Data Scientist

- **90%** of the data in the world today has been created in the last two years. (IBM,2012)

- Data-driven organizations are three times more likely to report significant improvements in decision making (PwC Global Data and Analytics Survey), but less than **0.5%** of data is analyzed (MIT Tech Review). The challenge is to rapidly scale a company's ability to work with data to drive better decisions.
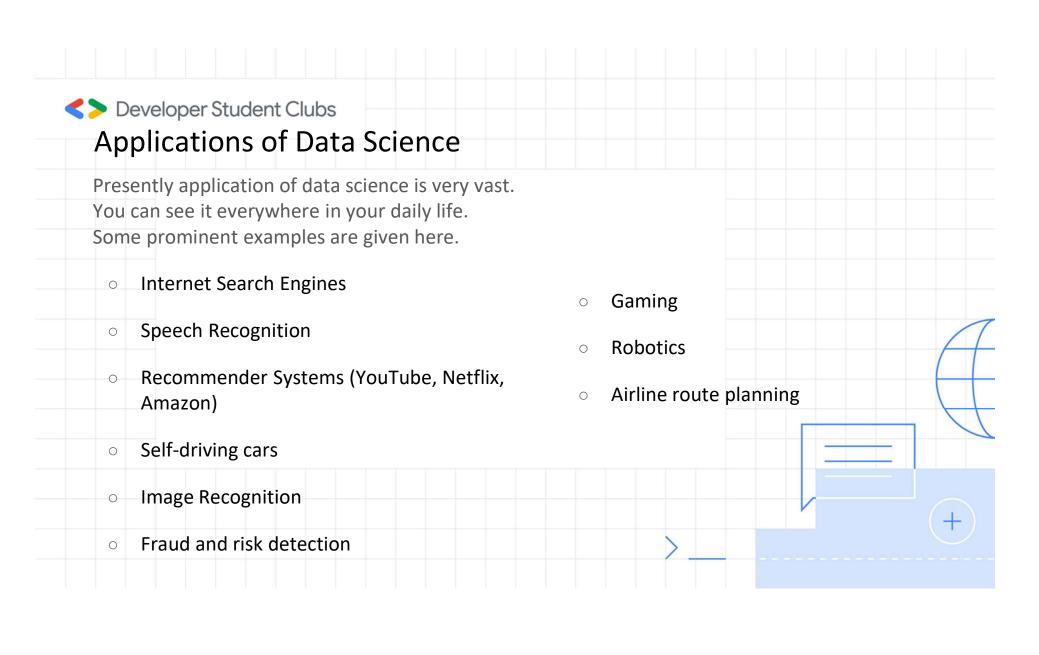 (Whitepaper published on DataCamp as "Democratizing Data Science in Your Organization")

- Data Science job is considered as the sexiest(**most attractive**) job of the **21st century**.
  -Harvard Business Review

- We will be reaching in range of **zettabytes** generation of data in coming years and organizations won't want it to lose without gaining useful insights.
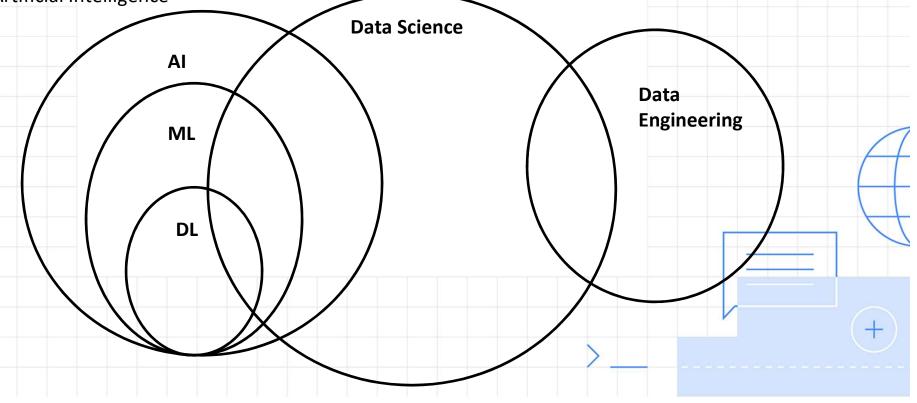
Data Science Project Life Cycle

Applications of Data Science

Presently application of data science is very vast.
You can see it everywhere in your daily life.
Some prominent examples are given here.

- Internet Search Engines

- Speech Recognition

- Recommender Systems (YouTube, Netflix, Amazon)

- Self-driving cars

- Image Recognition

- Fraud and risk detection

- Gaming

- Robotics

- Airline route planning

Developer Student Clubs
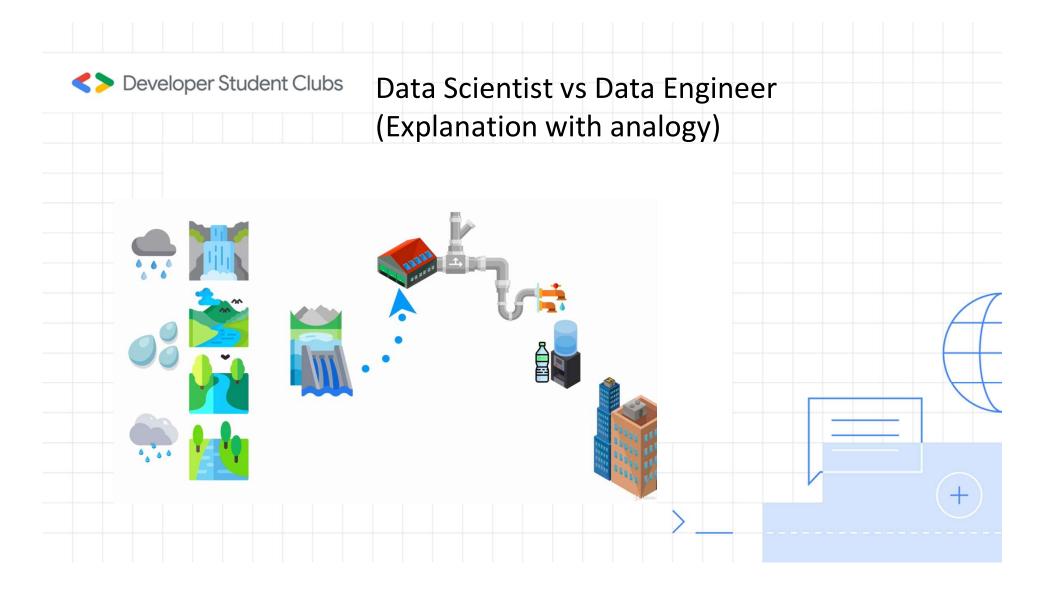
Clarification of Myths Related to Data Science

- Data Science is **not** a part or a subset of **Software Engineering** and **Computer Science**, however they have **intersections** with each other as there are many concepts and tools related to Computer Science and Software Engineering which are being used in Data Science.

- Data Science is **not** about making beautiful **infographics**.

- Data Science is a name of **Statistics** solely, is also **not** a good explanation of Data Science because Data Science goes **beyond** to that.

- Data Science is not about **Data Management**, **Data Quality**, **Data Encapsulation**, **Data Encryption** **solely.**

- Data Science is neither a subset of **Machine Learning**, **Artificial Intelligence** and **Data Engineering**, however Data Science has **subsets** of all these domains.
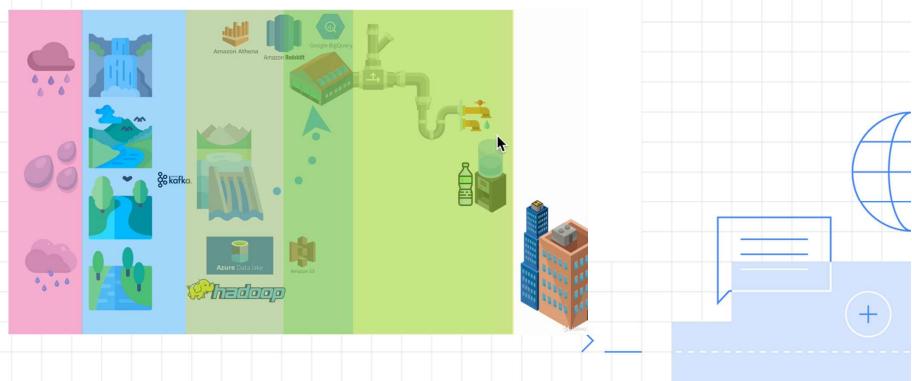
Differences to Note Between Data Science, Data Engineering, Machine Learning, Deep Learning and Artificial Intelligence

Data Science

AI
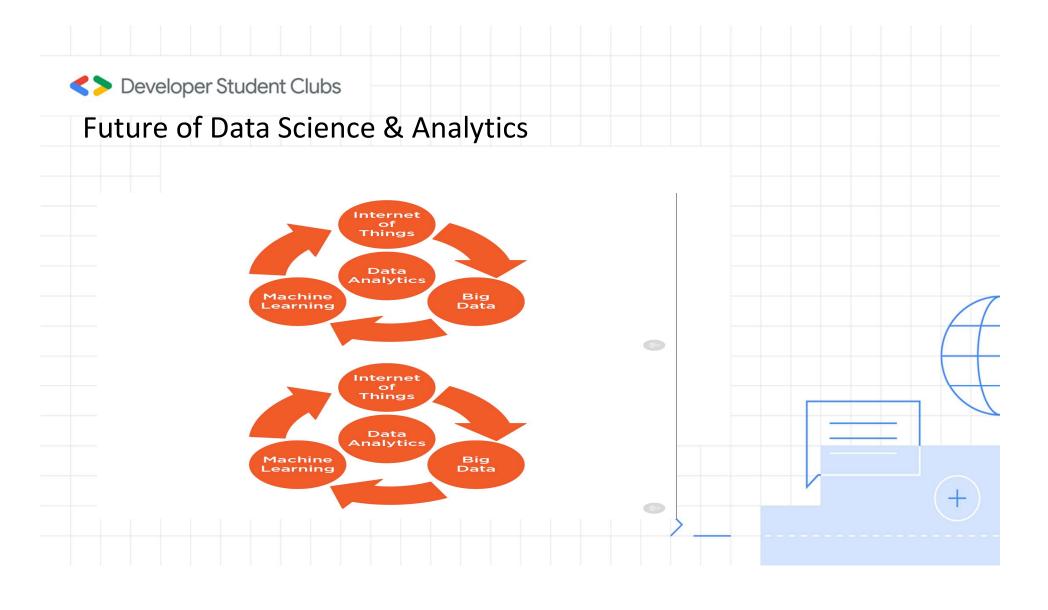
ML

DL

Data Engineering

Data Scientist vs Data Engineer
(Explanation with analogy)

Data Scientist vs Data Engineer (Explanation with analogy)

Future of Data Science & Analytics
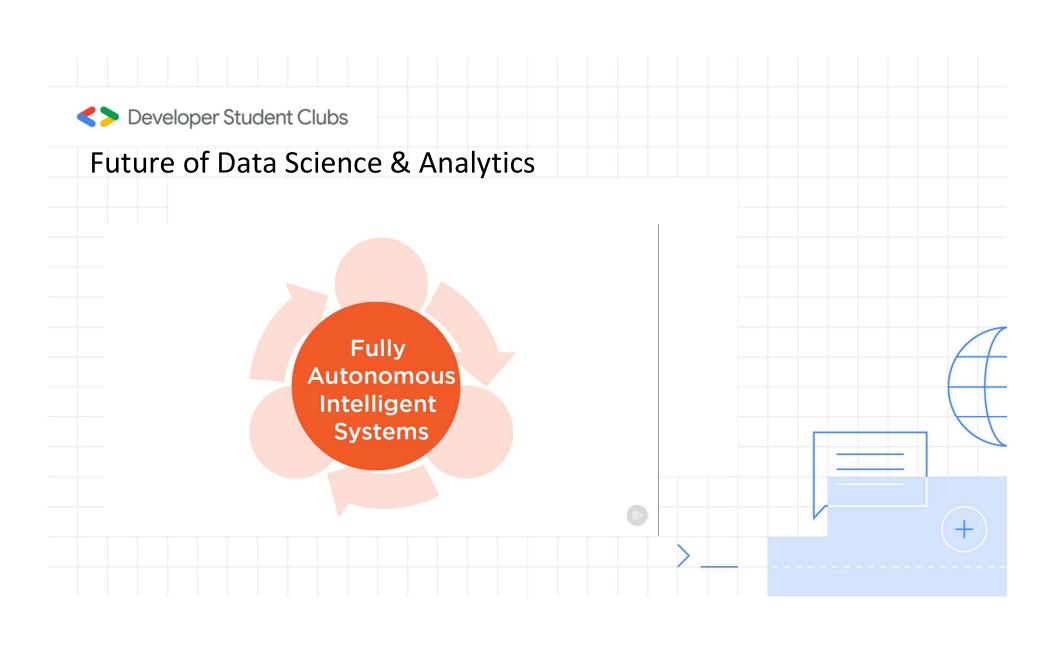
Industrial Revolution 4.0

Industrial Revolution will not replace humans but it seems clearly that it will replace those who will not have skills set required to adjust in the revolution.

Smart systems
Cloud robotics
Cyber-physical systems

Smart systems
Cloud robotics
Cyber-physical systems

Best Resources To Learn Data Science at Beginner Level

- I recommend to read and understand these two books at Beginner level

Developer Student Clubs

References

- Python Data Science Handbook by Jake VanderPlas
- An Introduction to Data Science by Dr Zeeshan ul Hasan Usmani
- https://www.datacamp.com/community/
- https://365datascience.teachable.com/
- https://www.pluralsight.com/
- https://towardsdatascience.com/data-science-life-cycle-101-for-dummies-like-me-e66b47ad8d8f

Developer Student Clubs

Thank you for showing patience

- One can like my Facebook page named as "Data Science Portal"
- One can join my Facebook Group named as "Engineering Digitization" where I usually share free resources for online learning

Thank you!

Developer Student Clubs

Google Developers