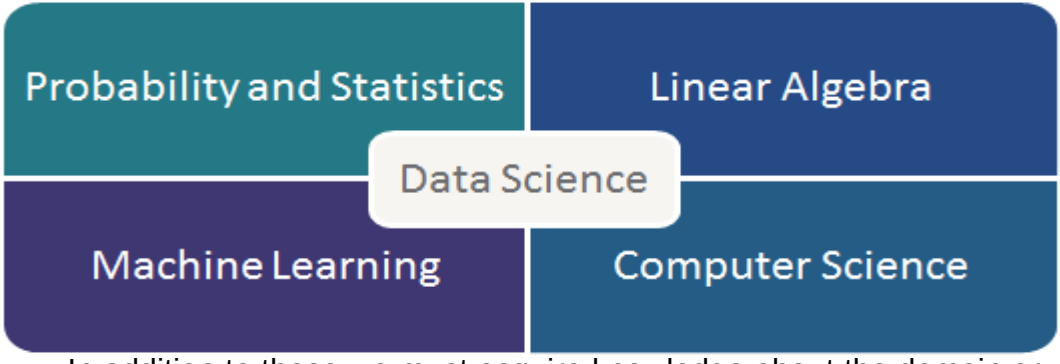
UNIT I	**Introduction to Data Science** **Data Science Fundamentals:** Data, Data Science Process,Components of Data Science, Data Scientist roles and responsibilities, Introduction to R and R Studio, Variables and Datatypes in R, Data frames, Recasting and Joining of Data frames,Arithmetic, Logical and Matrix Operations in R, **Advanced Programming in R** : Functions, Data Visualization in R Basic Graphics.	**06 - Hrs**

Recap...!!!!

- Data? Need of Data?
- What is Data Science & Why Data Science?
- Some important facts about Data Science?
- Case Study: Instant Insurance and claims mgt.
- Benefits of Data Science.
- Data Science Life Cycle Process?
- Tools for data science.
- Applications of data science.

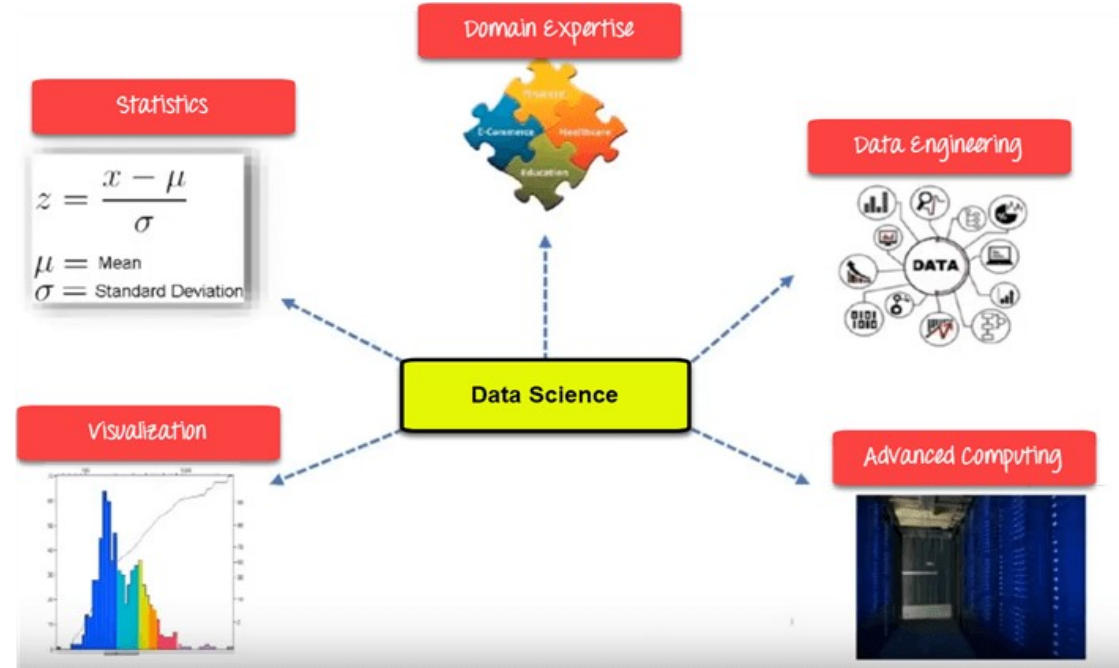
Components of Data Science

- Following are the various components of data science which act like tools to enable a data scientist to draw meaningful insights from data.



- In addition to these we must acquire knowledge about the domain or industry vertical in which we plan to apply Data Science, such as retail, banking & finance, healthcare, e-commerce, life sciences, telecom etc.

Components of Data Science

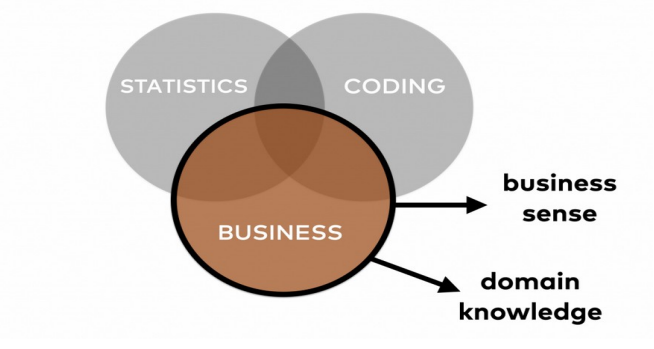


Statistics

- Statistics is the backbone of Data Science, It helps the Data Scientists to understand the underlying patterns present in the data.
- Through statistics, the user can understand the relationship between various **variables** that helps in providing a better picture of the given data. Apart from this advanced exploratory data analysis(EDA).
- Statistics is the most critical unit of Data Science, it is the method or science of collecting and analyzing numerical data in large quantities to get useful insights.

Domain Expertise

- Domain knowledge is often referred to as a general discipline or field to which data science is applied to.
- Domain expertise implies knowledge and understanding of the essential aspects of a specific field of inquiry.
- data science is at the intersection of three big fields:



Data Engineering

- Data engineering is the complex task of making **raw data** usable to data scientists and groups within an organization.
- Data engineers focus on the applications and harvesting of big data. Their role doesn't include a great deal of analysis or experimental design.
- **Why is Data Engineering important?**

  -> If your company lacks a fundamental data engineering strategy, the data that is collected is essentially **useless**. Data engineering is a **vital aspect of company growth**, network interactions, and predicting future trends.

Visualization

- Data visualization is meant by representing data in a visual context so that people can easily understand the significance of data. Data visualization makes it easy to access the huge amount of data in visuals.
- Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from.
- **Why is data visualization important?**

  --> Data visualization provides a quick and effective way to communicate information in a universal manner using visual information.

Job in Data Science

- Data Scientist
- Data Analyst
- Data Architect
- Machine learning expert
- Data Engineer
- Data Administrator
- Business Analyst
- Machine Learning Engineer
- Statistician
- Data and Analytics Manager
- Business Intelligence Developer

Data Scientist

- Data scientists examine which questions need answering and where to find the related data.
- A data scientist is a professional who works with an enormous amount of data to come up with very exciting business insights through the deployment of various tools, techniques, methodologies, algorithms, etc.
- A data scientist uses data to understand and explain the phenomena around them, and help organizations make better decisions.
- **Skill required:** To become a data scientist, one should have technical language skills such as R, SAS, SQL, Python, Hive, Pig, Apache-spark, MATLAB. Data scientists must have an understanding of Statistics, Mathematics, visualization, and communication skills.

Role and Responsibilities of Data Scientist

- Extracting usable data from valuable data sources
- Using machine learning tools to select features, create and optimize classifiers(Undertake preprocessing of structured and unstructured data)
- Enhancing data collection procedures to include all relevant information for developing analytic systems.
- Processing, cleansing, and validating the integrity of data to be used for analysis
- Developing prediction systems and machine learning algorithms.
- Presenting results in a clear manner.
- Propose solutions and strategies to tackle business challenges.
- Collaborate with Business and IT teams.

Introduction to R and R Studio