



Data

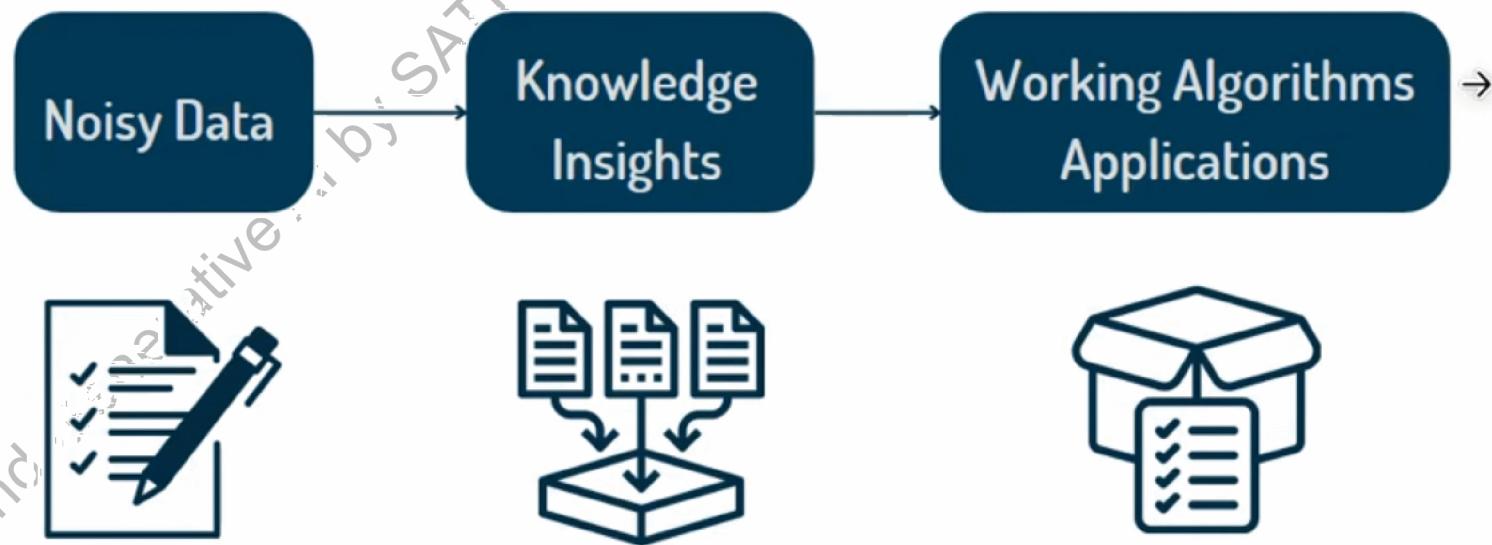
Data Science

“Data science is the scientific study of data to gain knowledge and make informed decisions.”

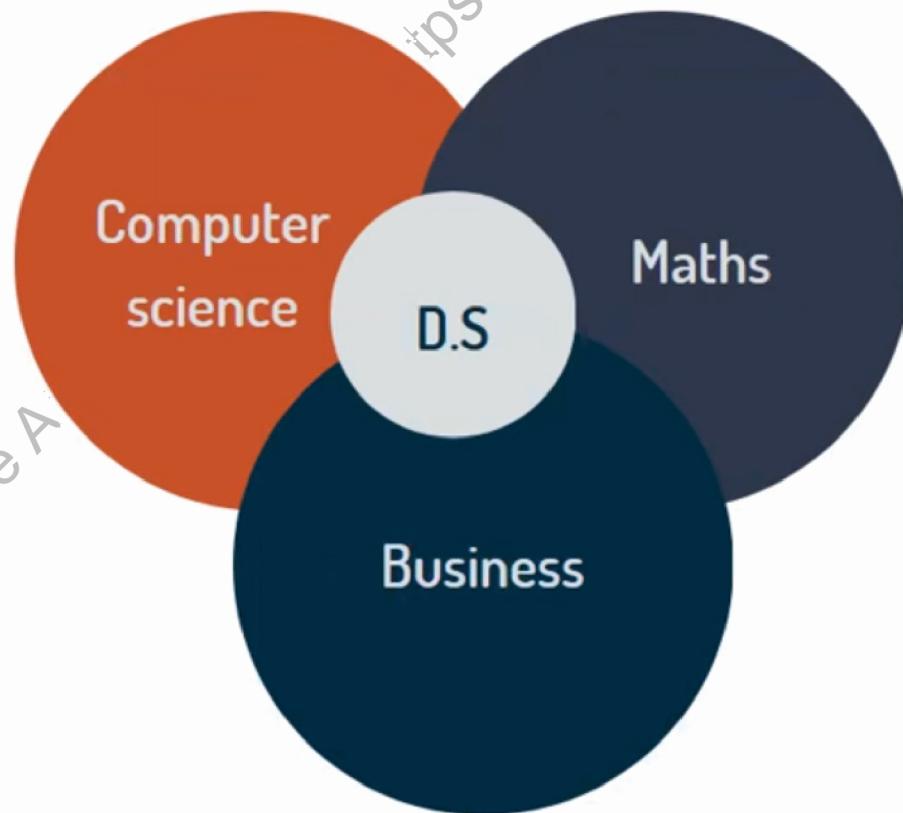
It's a multidisciplinary field that combines principles and practices from mathematics, statistics, artificial intelligence, and computer engineering.

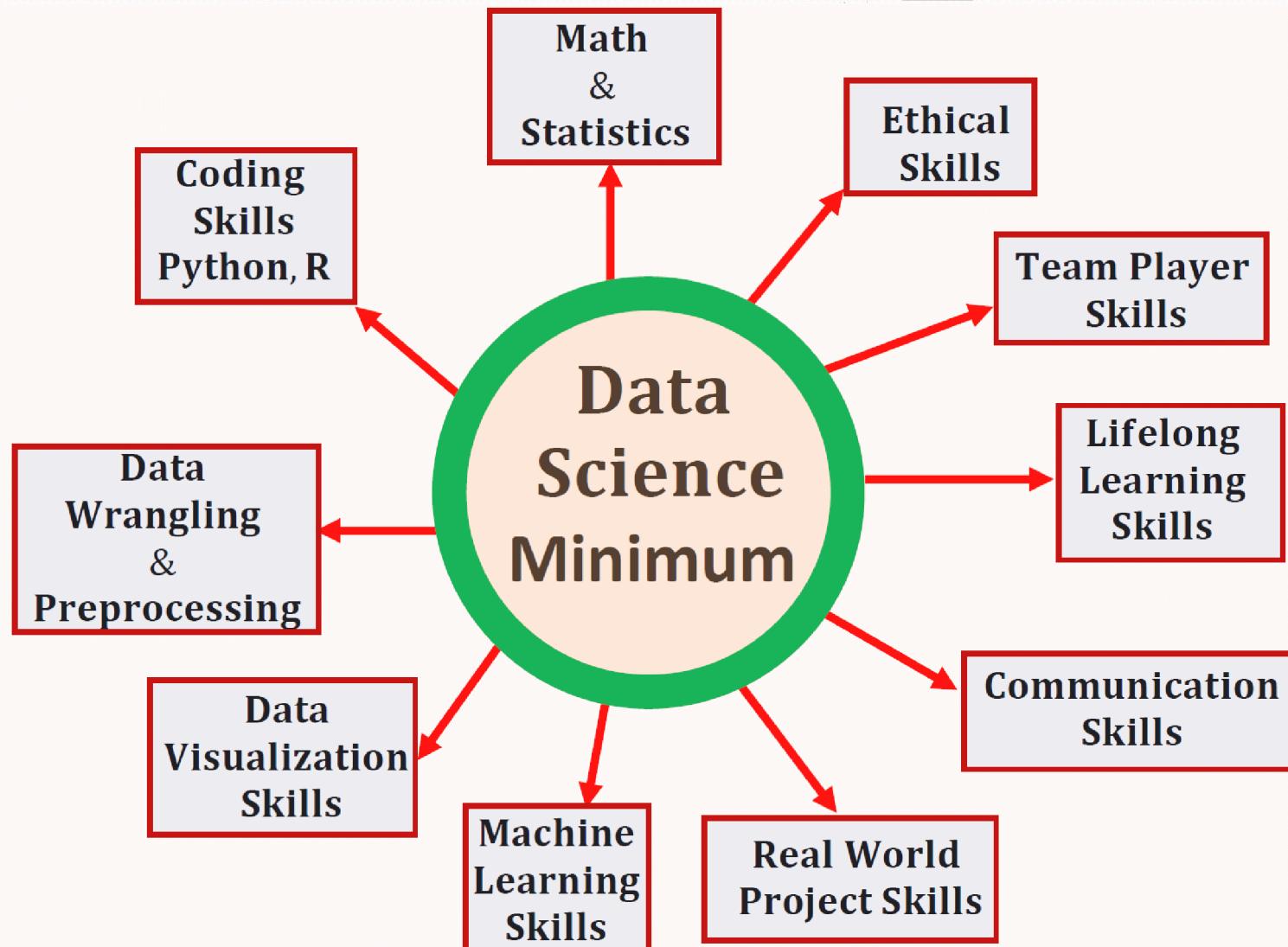
Data Science

What is Data Science



Data Science





Data Science

SKILLS REQUIRED

01 FRAME THE PROBLEM

- Domain Knowledge (needs)
- Product Intuition (metrics)
- Business Strategy (priorities)
- Teamwork (people & resources)

02 COLLECT RAW DATA

- Database Management
 - Systems: MySQL, PostgreSQL, Oracle, MongoDB
- Querying Structured Databases
 - SQL
- Retrieving Unstructured Info
 - Informational Retrieval / Text Mining
- Distributed Storage
 - Hadoop HDFS, Spark, Flink

03 PROCESS THE DATA

- Scripting Language
 - Python or R
- Data Wrangling & Cleaning
 - Python "Pandas" library
- Distributed Processing
 - Hadoop MapReduce / Spark

04 EXPLORE THE DATA

- Scientific Computing
 - Python: numpy, matplotlib, scipy, pandas
- Inferential Statistics
 - hypothesis testing
 - correlation vs. causation
- Experimental Design
 - A/B tests, controlled trials

05 PERFORM IN-DEPTH ANALYSIS

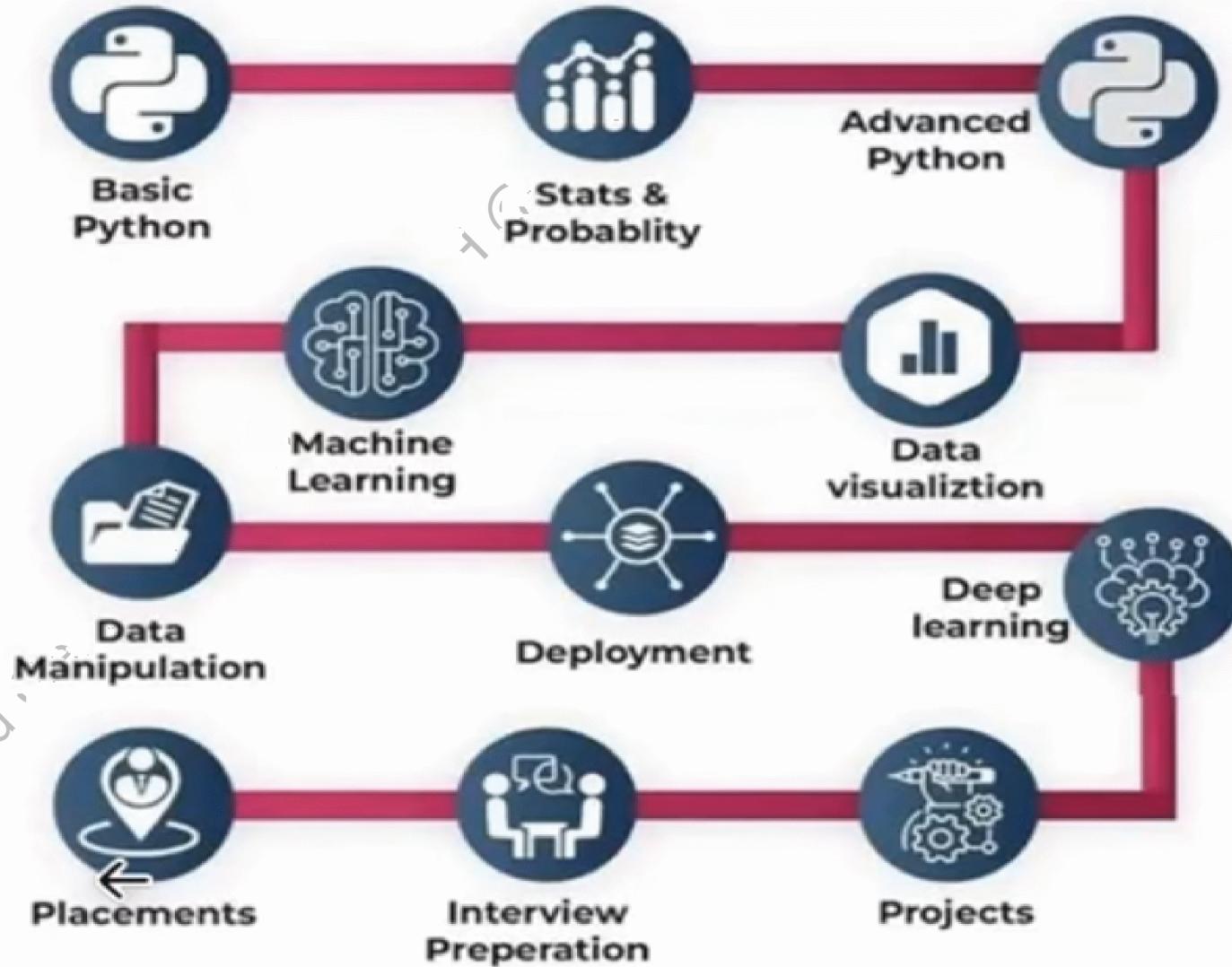
- Machine Learning
 - Supervised / Unsupervised algorithms
 - Contextual pros/cons
- ML Tools Library
 - Python: scikit-learn
- Advanced Math
 - Linear Algebra & Multivariate Calculus

06 COMMUNICATE RESULTS

- Business Acumen
 - Non-technical terminology
- Data Visualization Tool(s)
 - Tableau, D3.js, Google visualize, matplotlib, ggplot, seaborn
- Data Storytelling
 - presenting & speaking
 - reporting & writing

Data Science Skill	Includes
Python The most popular language used in data science for data wrangling, analysis, ML, visualization, deep learning, image processing, computer vision, NLP	Python libraries such as: - pandas - NumPy - scikit-learn - TensorFlow
Data Wrangling Before any analysis, data often needs to be cleaned and transformed	Data Cleaning Identifying and correcting errors or inconsistencies in data to improve its quality Data Transformation Converting data from one format into another
Statistical Analysis Understanding and applying statistical tests to data with the purpose of interpreting it	- Hypothesis testing - Probability - Descriptive statistics - Inferential statistics
Machine Learning Using supervised and unsupervised learning algorithms	- Linear regression - Logistic regression - Decision trees - Neural networks - Clustering methods
Data Visualization Effectively visualize data to communicate the findings clearly	Libraries for static visualizations in Python: - Matplotlib - seaborn Libraries for interactive plots in Python: - Plotly - Bokeh

Data Science Road Map



Advantages of Data science

Improved decision-making: Data science can be used to analyze large amounts of data and extract valuable insights that can inform business decisions and improve organizational performance.

Predictive modeling: Data science can be used to build predictive models that can forecast future events and outcomes, such as sales or customer behavior.

Automation: Data science can be used to automate repetitive tasks, such as data cleaning, feature engineering, and model selection, which can save time and resources.

Advantages of Data science

Personalization: Data science can be used to personalize experiences for customers, such as recommending products or tailoring advertising campaigns.

Cost reduction: Data science can be used to identify inefficiencies and reduce costs in various industries, such as supply chain management and healthcare.

Fraud Detection: Data science can be used to analyze large amounts of transaction data and identify fraudulent activities, which can reduce financial losses.

Advantages of Data science

Improved customer service: Data science can be used to analyze customer data and understand their needs, preferences and behavior which can improve the overall customer service.

Improved product innovation: Data science can be used to analyze data from research and development, customer feedback, and market trends to identify new product opportunities.