Google Developer Student Clubs

Introduction to Data Science

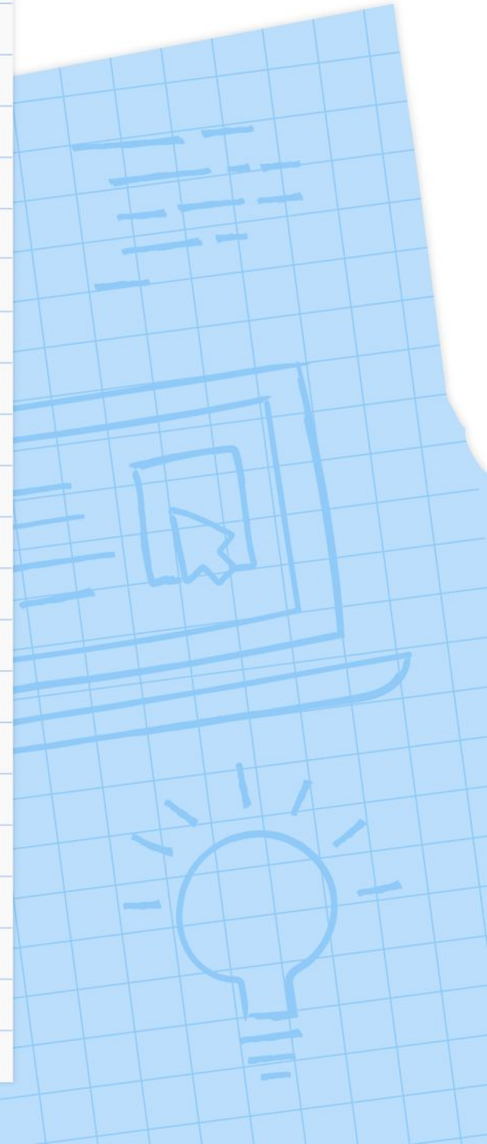
Ahmed Bahaaeldin
Sr. Data Scientist

Agenda

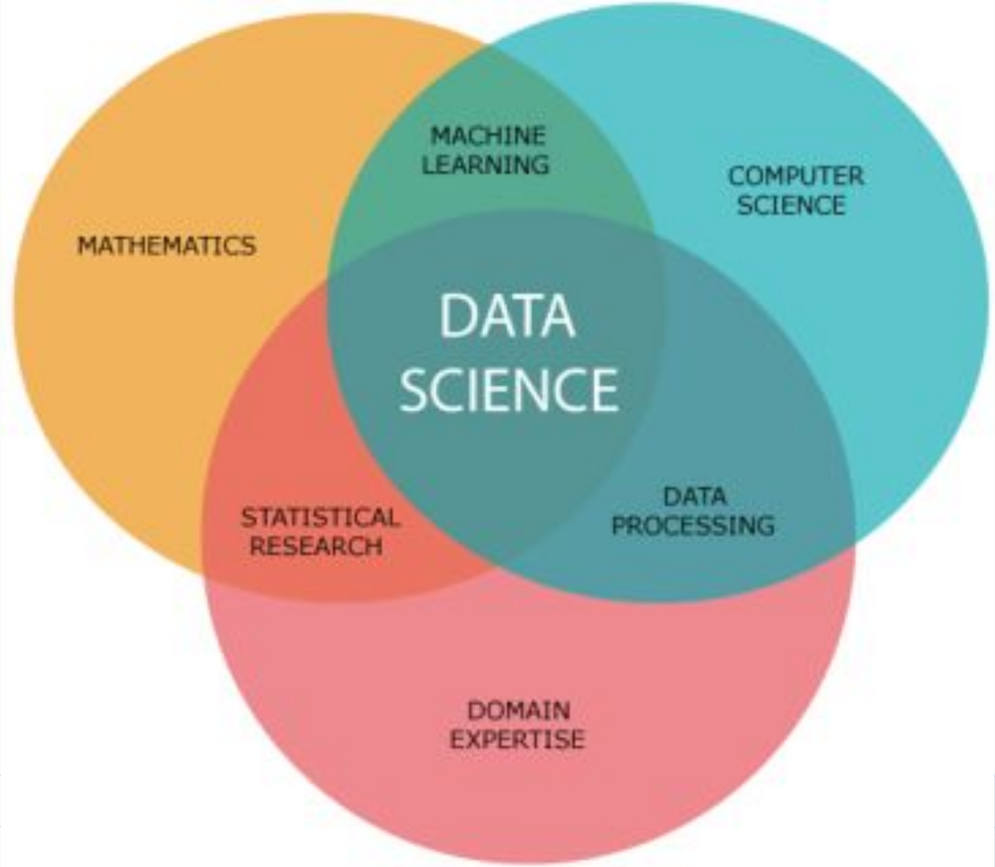
- What is Data Science?
- Skillset of a Data Scientist.
- How to be a Data Scientist.
- Our Plan for the year.
- Tips on how to gain the maximum from this year.

What is Data Science?

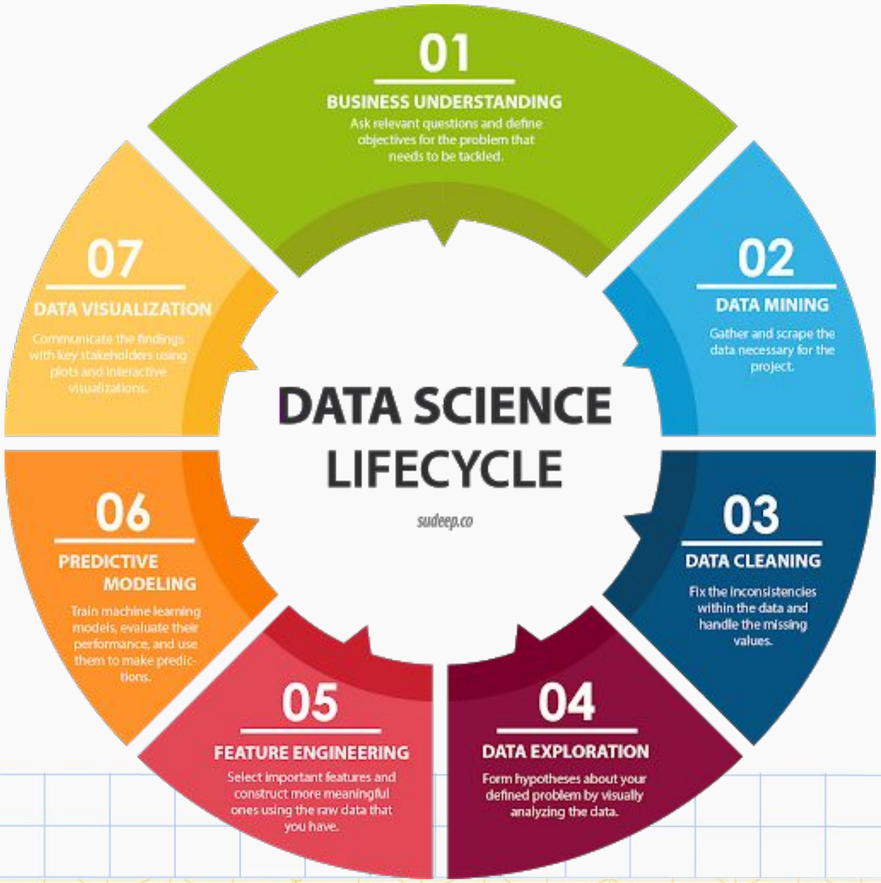
It is the practical activity done on Data that study the structure and aim to solve problems through observation and experiments.



Data Science Skillset



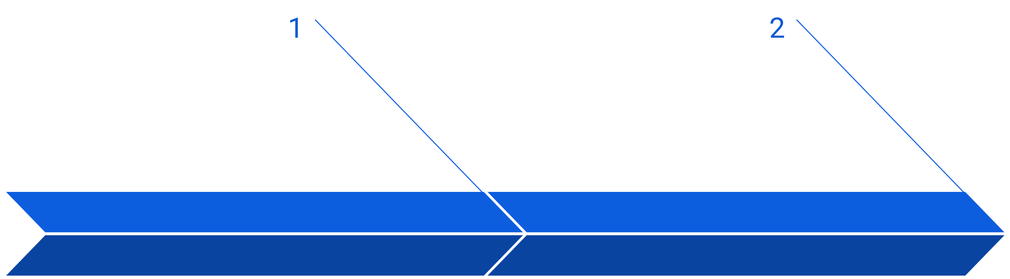
Data Science Life-cycle



How to become a Data Scientist?

1. Top-Down Approach:
   Start with the Practical Side and then end up learning the mathematics and science behind it.
2. Bottom-Up Approach:
   Start with the mathematics and science then end with the Practical Side.

Best approach according to surveys

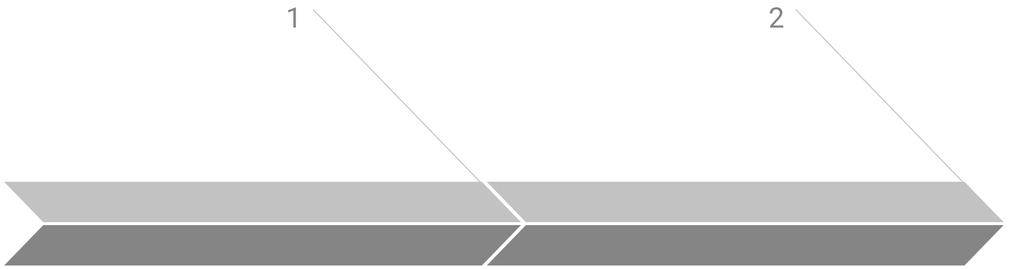


Statistics

The most important science branch in Data Science.

Probability theory

Most of ML models depend on probability and statistics combined together.



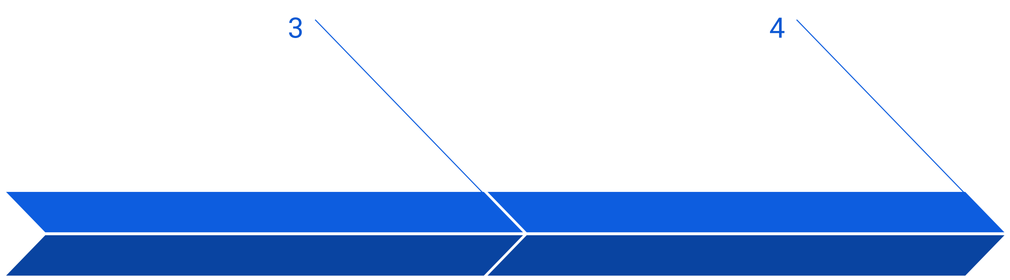
Python

Most of ML and DS libraries are coded in python.

Data Analysis

Exploration of Data is one of the most important steps in DS.

Best approach according to surveys

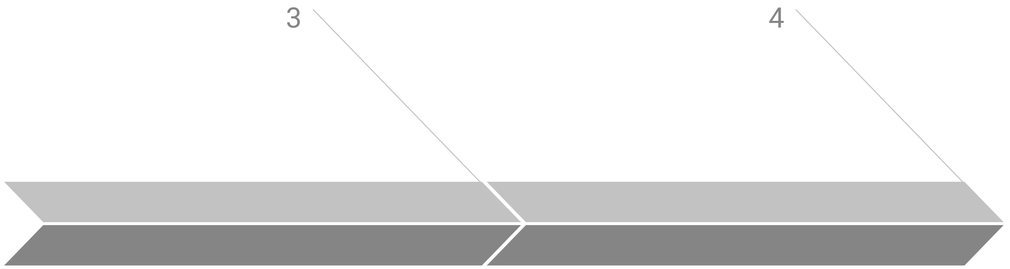


ML Model theory

Understand how the black box work.

Advanced Statistics inference and hypothesis

Validation of our models.



ML Models Library

Sklearn is one of the most famous libraries in the field.

Hands-On-Full Project

To implement what you have learned and gain sense of achievement.

All the crucial skills

Skills	Usage
Statistics	Gain insights about the use-case with better understanding
Probability Theory	Understanding the odds and interpret the results of your solution
Python	Most famous programming language for ML&DS.
DS related libraries (sklearn,numpy,pandas,matplotlib)	An important tool-box for any DS.
Machine Learning Theory	Help interpret the results and understand which solution is relevant to the problem.
Deep Learning	State of the art solution for nearly every AI related problem
Research Mindset	That not the same solution will work every time.

Plan

Revision on basics

This course will cover basic statistics, python and probability theory with a flavor of EDA (exploratory data analysis).

Classification and Regression

How to deal with specific problems with a defined scope like classification and regression end-to-end.

1

2

3

4

How to deal with Data

How to sample from your dataset to gain information about the whole population ? How to validate your outcome results ?

Advanced Part

Explaining Deep Learning approaches and how to integrate it with what was explained before.

Google Developer Student Clubs

Tips

- Always keep studying from external resources before we explain the concepts.
- Dont ever copy code, always write it yourself.
- Try to finish all the assignments.
- Always ask if you face a blocker.