**python**

**Day: 1**

* **What is Data Science?**

Definition from Wikipedia:

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured similar to data mining.

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

**Data:** Data is a set of values of subjects with respect to [qualitative](https://en.wikipedia.org/wiki/Qualitative_property) or [quantitative](https://en.wikipedia.org/wiki/Quantitative_data) [variables](https://en.wikipedia.org/wiki/Variable_(research)).

**Qualitative properties** are properties that are observed and can generally not be measured with a numerical result. They are contrasted to [quantitative properties](https://en.wikipedia.org/wiki/Quantitative_property) which have numerical characteristics.

**Or**

Quantitative data is any data that is in numerical form such as statistics, percentages, etc.

**Quantitative Research** In [natural](https://en.wikipedia.org/wiki/Natural_science) and [social sciences](https://en.wikipedia.org/wiki/Social_science), and sometimes in other fields, **quantitative research** is the systematic [empirical investigation](https://en.wikipedia.org/wiki/Empirical_research) of observable [phenomena](https://en.wikipedia.org/wiki/Phenomenon) via statistical, mathematical, or computational techniques.[[1]](https://en.wikipedia.org/wiki/Quantitative_research#cite_note-Given_2008-1) The objective of quantitative research is to develop and employ [mathematical models](https://en.wikipedia.org/wiki/Mathematical_model), [theories](https://en.wikipedia.org/wiki/Scientific_theory), and [hypotheses](https://en.wikipedia.org/wiki/Hypothesis) pertaining to phenomena. The process of [measurement](https://en.wikipedia.org/wiki/Measurement) is central to quantitative research because it provides the fundamental connection between [empirical](https://en.wikipedia.org/wiki/Empirical) [observation](https://en.wikipedia.org/wiki/Observation) and mathematical expression of quantitative relationships.

2. Definition from datarobot .com

Data science is the field of study that combines domain expertise, programming skills, and knowledge of math and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems that perform tasks which ordinarily require human intelligence. In turn, these systems generate insights that analysts and business users translate into tangible business value.

* **What does Data Scientist do?**
* **Data Science Process**
* Ask an interesting question or address a problem
* Get the data
* Explore the data
* Model the data
* Communicate & visualize your result
* **Data Science is multidisciplinary**
* The Scientific method
* Programming
* Databases
* Statistics
* ML
* Domain Knowledge
* **O-S-E-M-N Things (pronounced as ‘awesome’)**
* **O**btain Data
* **S**crub Data
* **E**xplore Data
* **B**uild Models
* i**N**terpret results
* **Common Tasks**
* Data Munging/scraping/sampling/cleaning in order to get an informative, manageable data set.
* Data storage & management in order to be able to access data quickly reliably during subsequent analysis
* Exploratory data analysis to generate hypotheses and intuition about the data
* Prediction based on statistical tools such as regression, classification, clustering, forecasting & optimization
* Communication of results through visualization, stories & interpretable summaries.