
Аналіз даних в Python та R

Data science with Python and R

Осінь 2023/2024

Викладач:

Розова Людмила Вікторівна

Лекція 2

Аналіз даних – це процес дослідження, фільтрації, перетворення та моделювання даних з ціллю отримання корисної інформації з них, створення прогнозів, та прийняття рішень

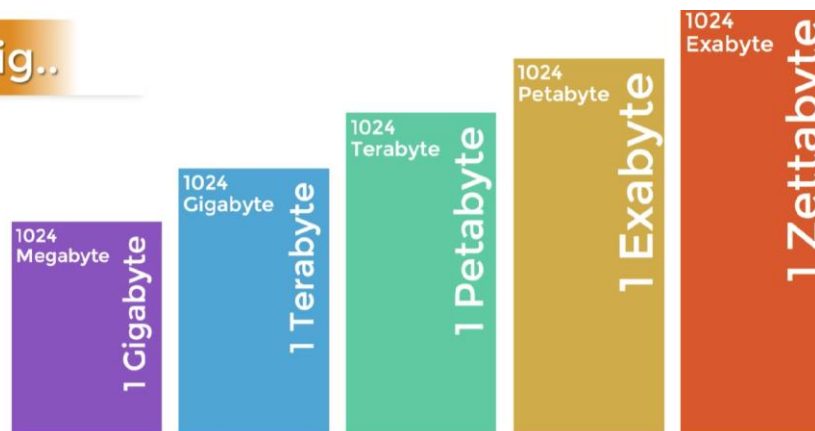
Кожен день
К 2025 року щорічно

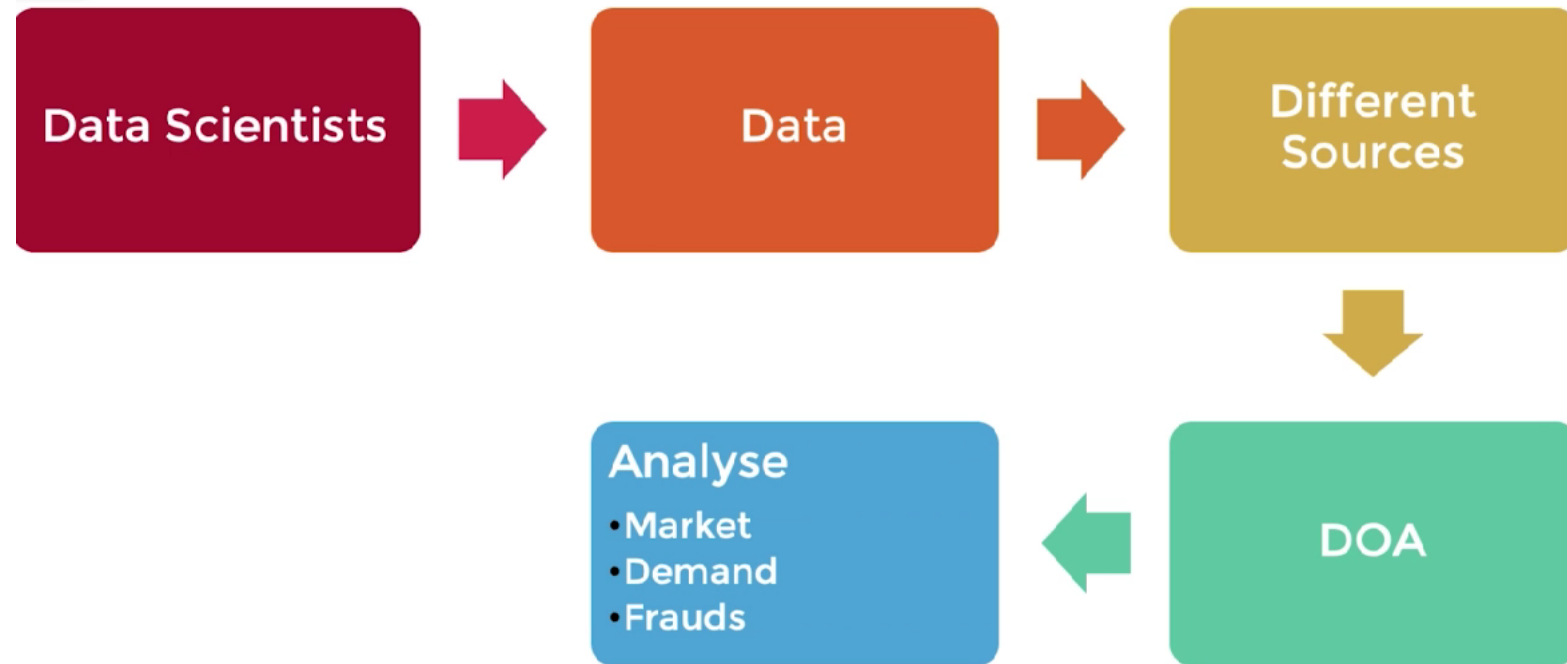
2.5 exabytes

175 zettabytes

90% існуючих даних були створені за останні роки

How big..





Derive meaningful information

Data literacy— це здатність відокремлювати корисну інформацію

Complexity

Потребує конкретні знання з математики та статистики, особливо при роботі з Big Data

Specialist with analytical skills

Багато компаній потребує спеціалістів, які розуміються на Data literacy. Це Data scientistes або вимагає від працівників знання з тієї області

Crucial to a business' success

Інформаційна грамотність / Data Literacy

Data Literacy передбачає :

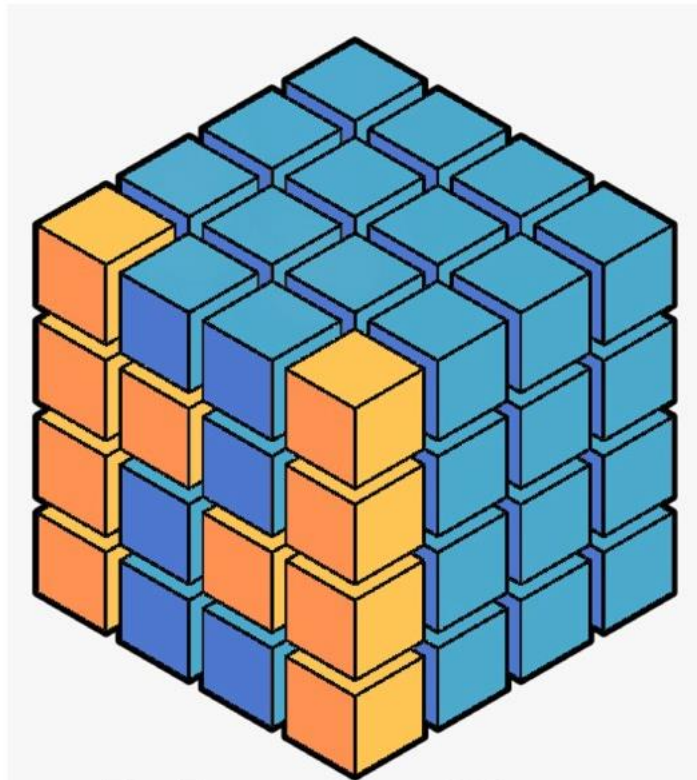
- Розуміння які дані є в наявності, як їх використати для конкретної цілі.
- Інтерпретація візуалізації даних (графів, діаграм, ...)
- Критичний підхід до інформації, яка отримана за допомогою аналізу даних
- Розуміння інструментів та методів аналізу даних, де і коли їх застосовувати
- Визначення де і які дані відсутні
- Створення інформаційного забезпечення для загального пояснення

Data Literacy \neq Statistical Literacy

Розуміння, що означають дані

Інтерпретація статистичних звітів

Бібліотека NumPy



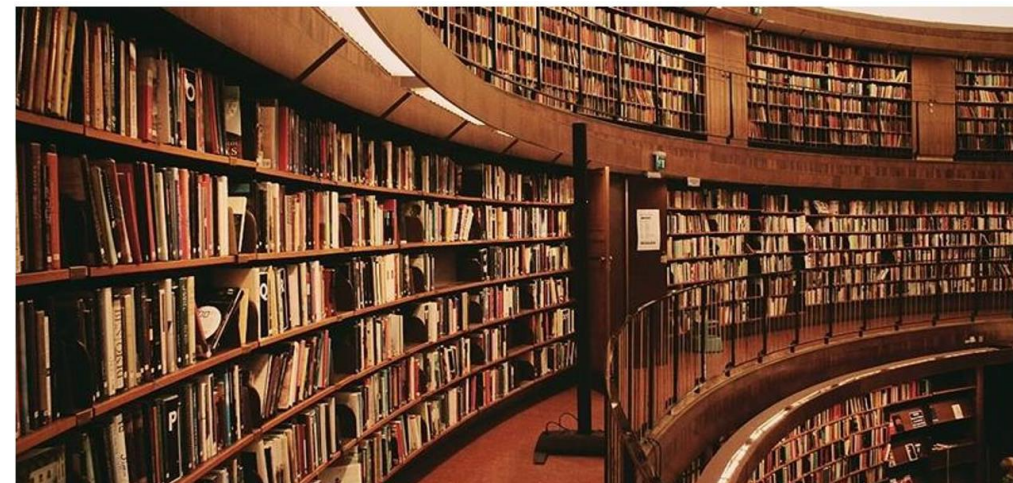
NumPy



NUMERICAL



PYTHON



NumPy дозволяє реалізовувати велику кількість науково-математичних обчислень ефективно, швидко та гнучко. Це дуже суттєво для операцій аналізу даних.



NumPy застосовується в

DATA MINING

MACHINE LEARNING

DATA ANALYSIS

**COMPUTATIONAL
STRUCTURES**

Numpy, Numpy Pyhon

- NumPy Library

<https://numpy.org/>

- Universal Function in NumPy

<https://numpy.org/doc/stable/reference/ufuncs.html>

Бібліотека NumPy

- **NumPy** це математична бібліотека, яка використовує одномірні так і багатомірні масиви даних (**NumPy arrays**). Так звані, «less costly data», відрізняються від списків швидкістю їх обробки.
- Також містить інструменти обробки цих масивів. Побудовані на основі об'єктно-орієнтованого програмування.
- Це фундаментальна бібліотека для проведення наукових обчислень в Python.
- В NumPy реалізована лінійна алгебра, перетворення рівнянь, робота з випадковими числами тощо.
- Робота з багатомірними контейнерами згенерованих даних
- Можливості інтегрування C, C++, Fortran коду.

Функції Numru для роботи з масивами

Створення масивів . Функції

Array()

Zeros()

Ones()

Full()

Arange()

Eye()

Random()

Властивості масивів. Функції

Reshape()

Max(), Argmax()

Min(), Argmin()

Concatenate()

The Split()

Vsplit, Hsplit()

Дякую за увагу!
