



# Data analysis of Amazon top 50 selling books

# What is Data?

---

- Data in a general term refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing.
- Data can be measured, collected and reported, and analyzed.
- You can visualize the data using graphs, images or other analysis tools.



# Life cycle of data science



# Important statistics terms in data science

---

- What is Statistics?
- What is population?
- What is parameter?
- What is sample?
- What is mean?
- Types of analysis in statistics

# What is Statistics?

---

Statistics is a part of integrated applied mathematics which deals with data

- It helps to collect data and analyze them properly
- With the help of statistics we can read the data and organize them in order to get the hidden information from them
- In data science domain statistics concepts are used to process the complex data to get the insights from them using mathematical computations

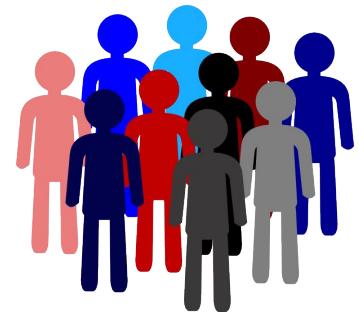
# What is population?

---

Population terms in statistics use to refer the total set of observations

Suppose,

If we want to study a diabetes dataset to understand the symptoms and the other factors then the whole dataset is referred as population

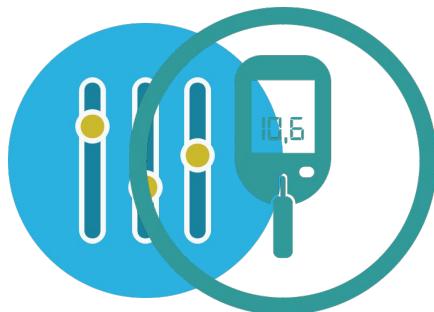


# What is parameter?

---

Parameters are referred to characteristics which describes the population

- Parameters are like average or percentage which helps to describe the entire population
- Mean and the standard deviation are two common parameters of population
- Example: Average age for being diabetic is the parameter for whole diabetes dataset population



# What is Sample?

---

Sample is basically a small part or portion of the large population

Suppose,

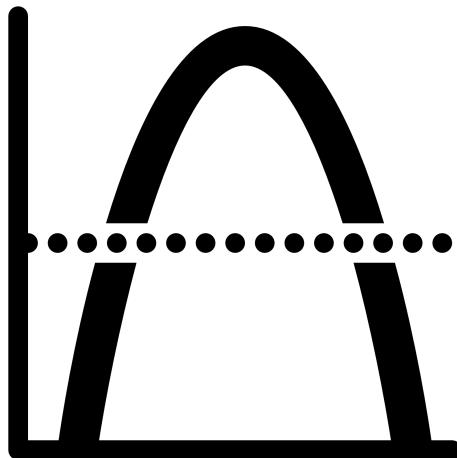
From the whole diabetes dataset you picked 100 rows of information to do the analysis, that 100 rows of information will be referred as Sample



# What is mean?

---

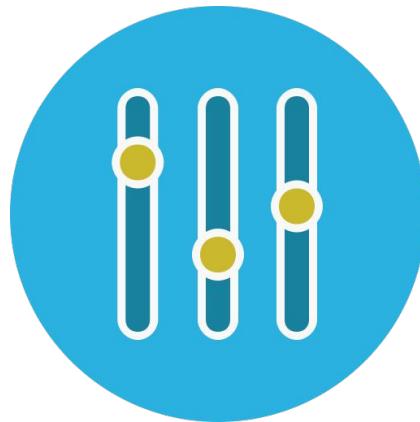
Mean term referred as average value of the whole population



# What is median?

---

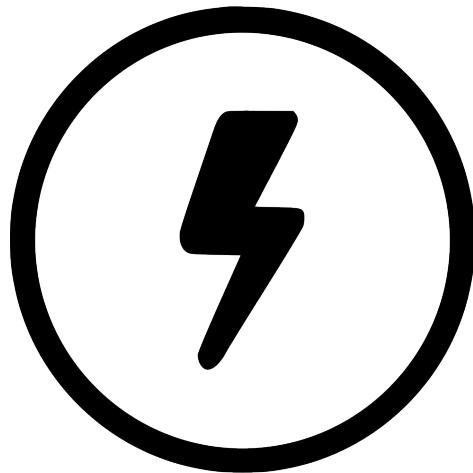
Median is the middle value of the data when your data is sorted in manner



# What is mode?

---

Mode stands for the most occurring element in the dataset



# Types Of Analysis In Statistics

---

## Descriptive statistics

It helps to describe the data in mathematical or graphical way

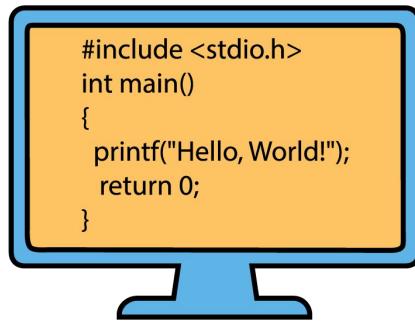
## Inferential Statistics

Inferential statistics split the data into samples and applies probability to arrive to the conclusion

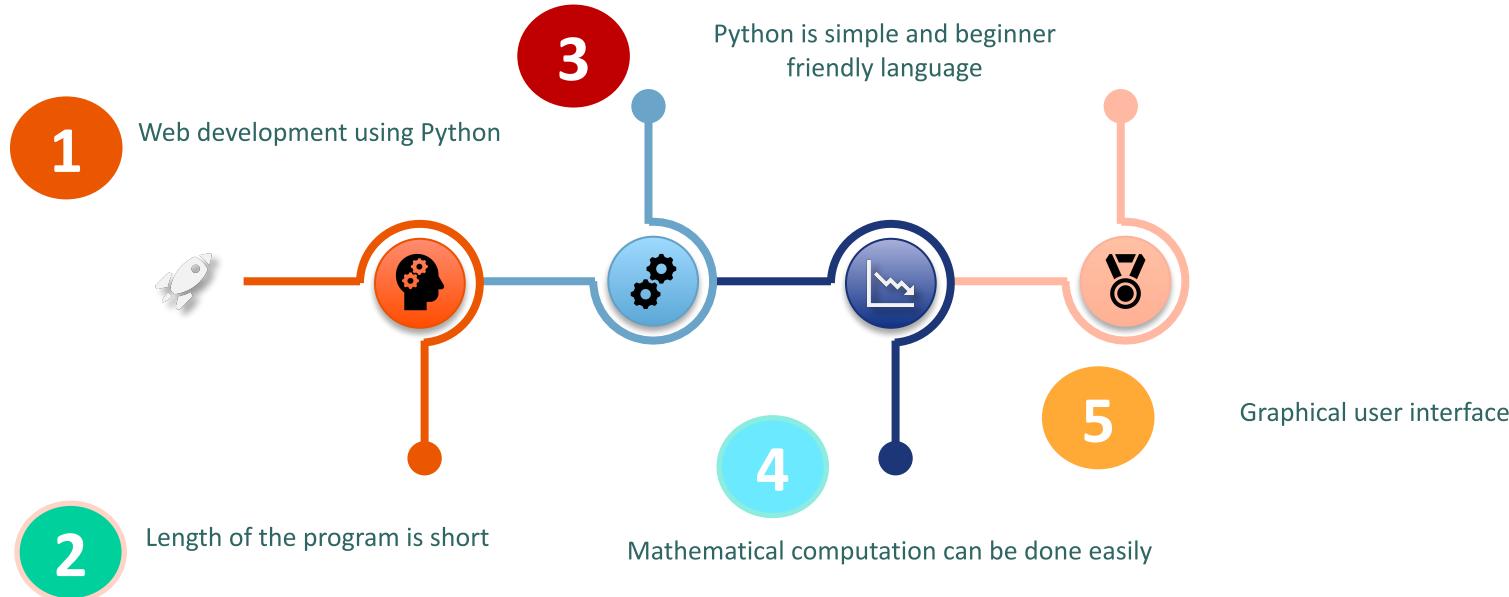
# What is Python?

Python is a popular high level, object oriented and interpreted language

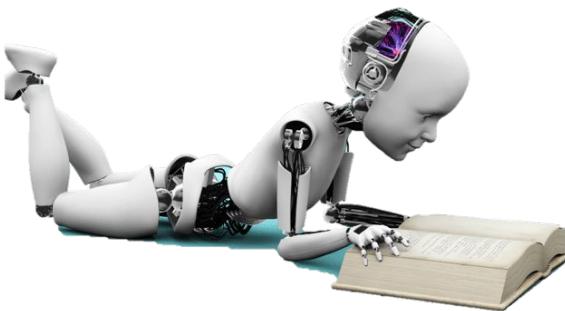
- High level
- Object oriented
- Interpreted



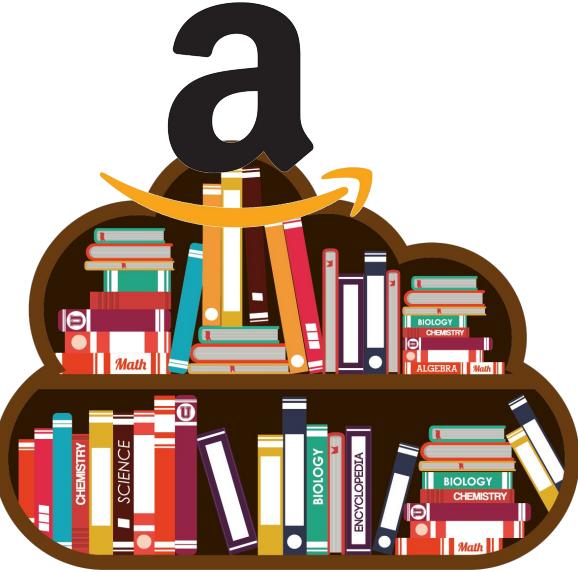
# Benefits of Python



# Important libraries in Python



- Numpy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn



## Demo: Data analysis of Amazon top 50 selling books

A thin horizontal bar at the top left of the slide, divided into two segments: a teal segment on the left and an orange segment on the right.

# Thank You!