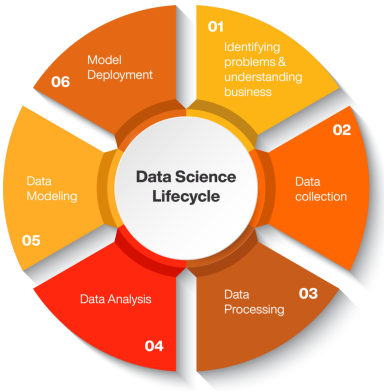
Data Science Life Cycle



What is Data Science?

The Study of data in order to ro make more meaningful conclusions

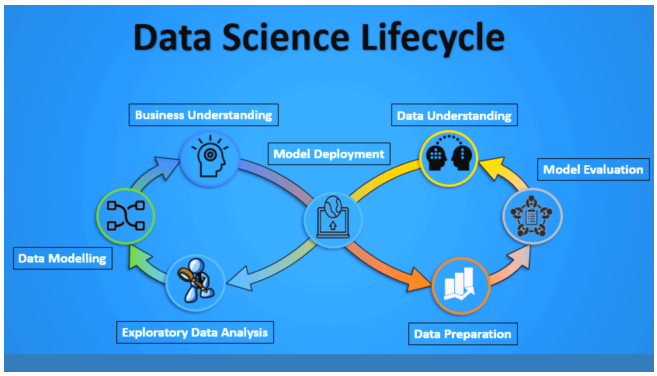
The difference between Data and Information

Data: Is unprocessed raw input

Information: Is processed data that can be understood

Data science Life Cycle steps

1. Question
2. Collection
3. Wrangling Data
4. Analyze
5. Visualize
6. Communicate



Python Fundamentals

Collection Types

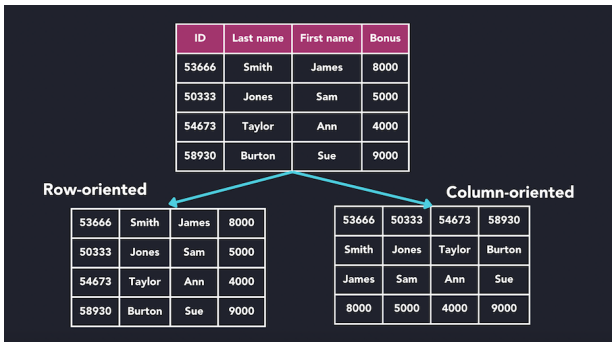
Very useful in data science because datasets are the collection of data

Python collection types:

- Dictionary
  - Ordered
  - Changeable
  - Duplicates not allowed
- List
  - Ordered
  - Changeable
  - Duplicates allowed
- Tuple
  - Unordered
  - Unchangeable
  - duplicates not allowed
- Set
  - Unordered
  - Unchangeable
  - duplicates allowed

Code Representation Datasets

- Column- Oriented:
  - Grouping by features
- Row-oriented:
  - Grouping by a single observations



Indexing

- To access values, we need to INDEX

Type	Indexing Pattern
List	name[index]
Dictionary	name[key]
Set	For loop (next slide)
Tuple	Name [index}

Iteration

- You can repeat processes with loops or recursion in python

Python loop types: (https://www.w3schools.com/python/python_ref_dictionary.asp)

❖ For loop

```
for thing in collection:
  statements
```

❖ While loop

```
while condition:
  statements
```

Useful Methods

Dictionaries:

- values()
- items()
- keys()

Lists:

- len()
- append()
- sort()

Other:

- range()
- print()
- split()
- type()
- int()
- str()

Next Heading