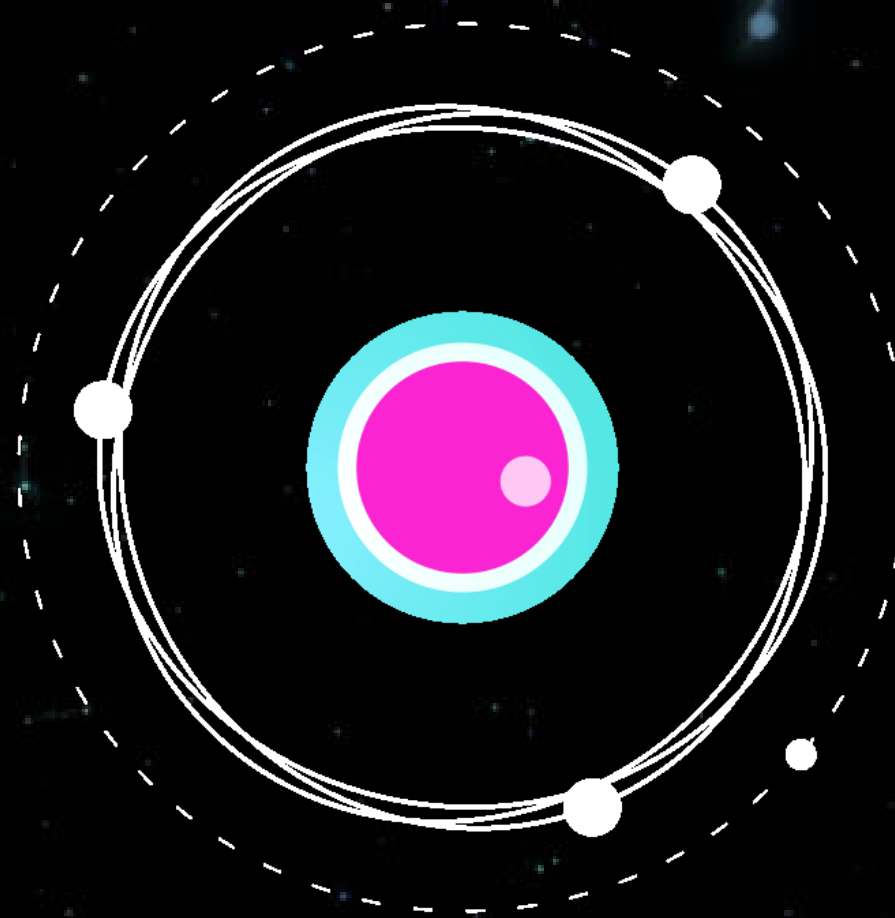


# DATA CELL TRAINING ROADMAP

INPT

CIT



MADE BY:

TAHA ELGHABI





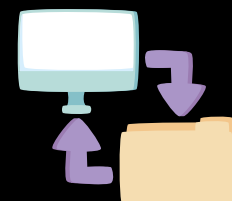
PLAN



OVERVIEW, DATA ROADMAP



DATA COLLECTION



EXPLORATORY DATA ANALYSIS



CLEANING, FEATURE ENGINEERING



VISUALIZATION



# Data Science buzzwords



Artificial Intelligence

Business Intelligence

# Data Science

Case Study

Big Data

Machine Learning

Data Analysis

Business Analytics

Statistics

Deep Learning

SDOTM



**What the heck is Data Science**



# What is Data?



Data is a bunch of facts,  
such as numbers, words,  
measurements...

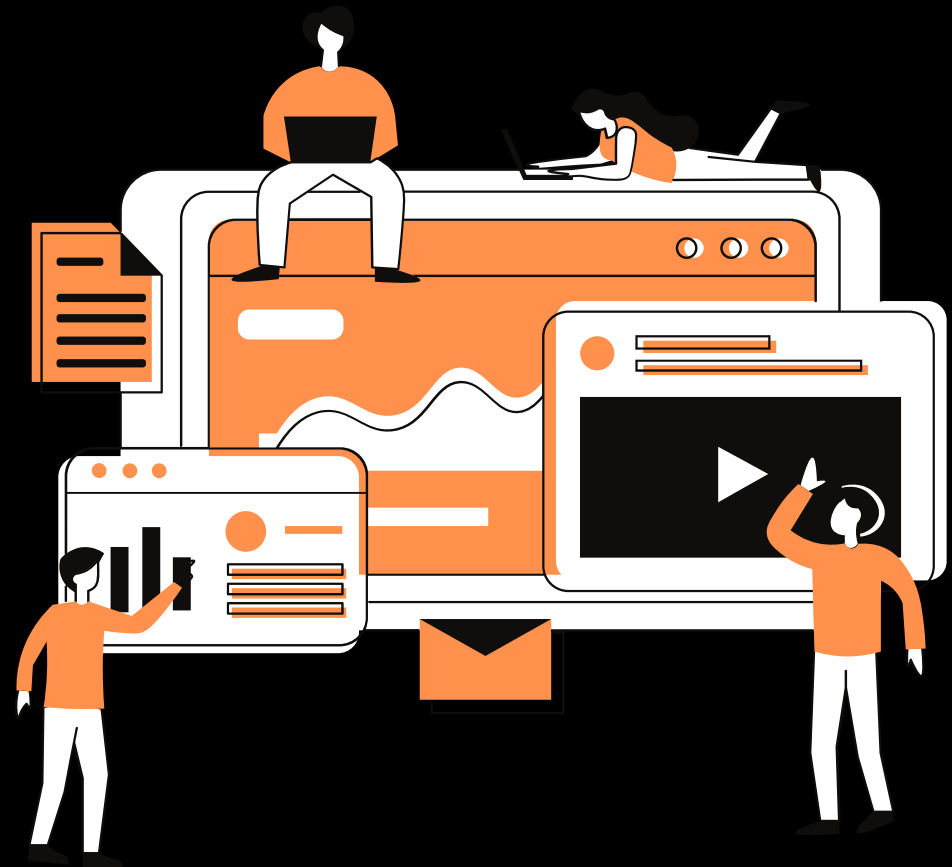
or could be a descriptions  
of things **used to better  
understand a subject**

# Types of Data

Structured



Unstructured



# "It all starts with a **Dataset**"

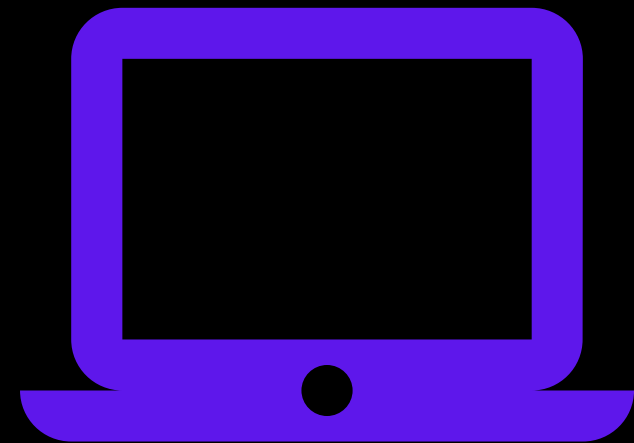
- Unknown Data Scientist

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

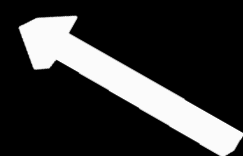
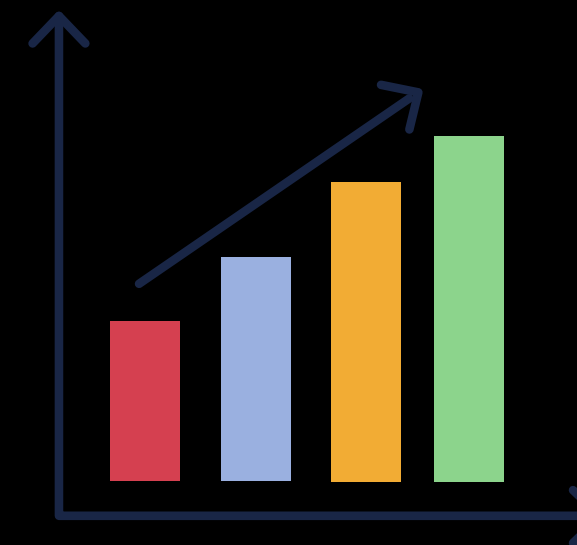




Computer Science



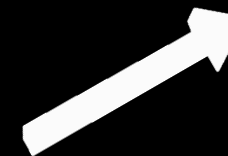
Math & Statistics



Data

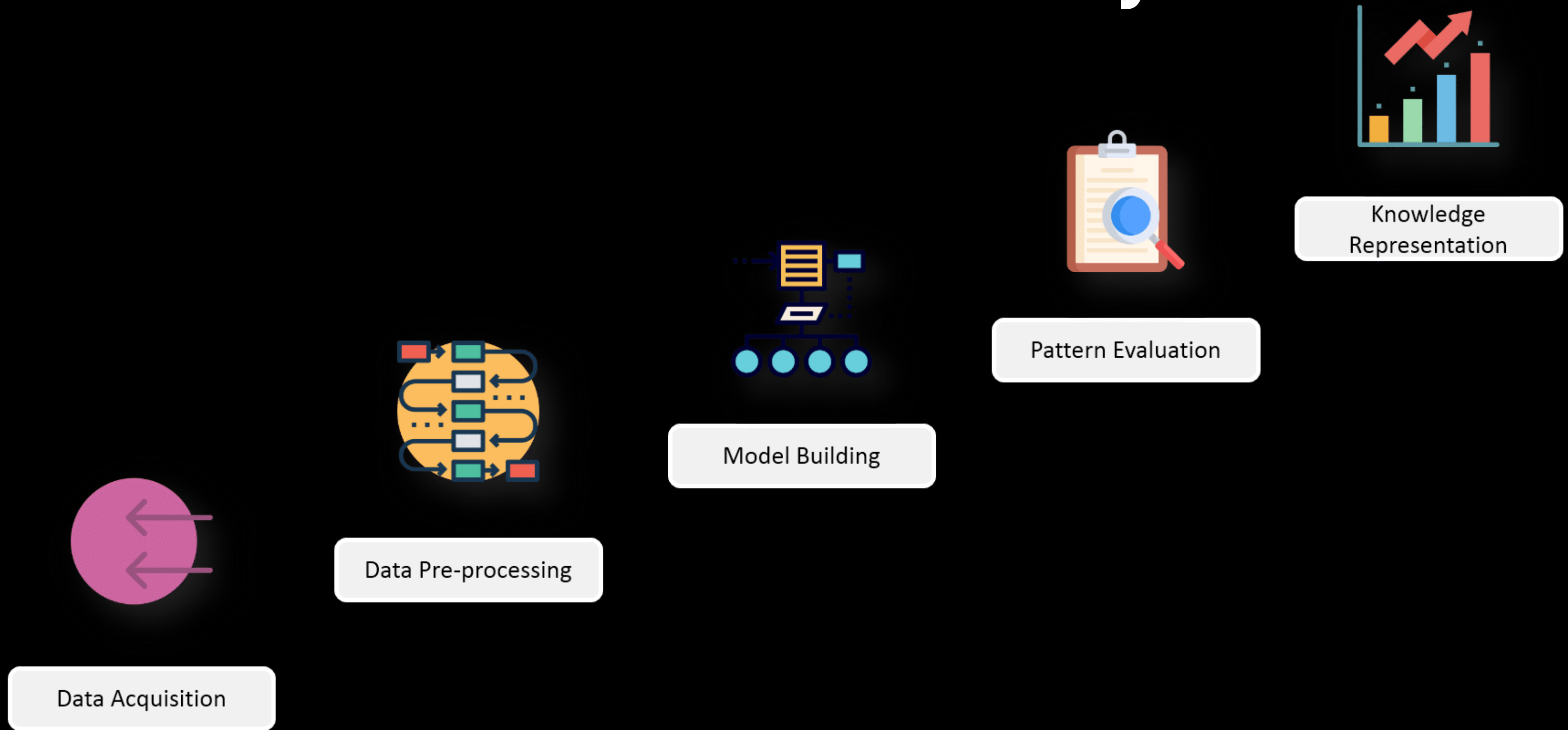


Science



Domain Knowledge

# Data Science Lifecycle



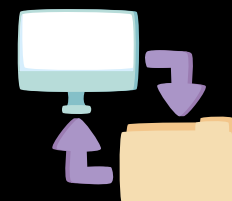


PLAN

OVERVIEW, DATA ROADMAP



DATA COLLECTION



EXPLORATORY DATA ANALYSIS



CLEANING, FEATURE ENGINEERING



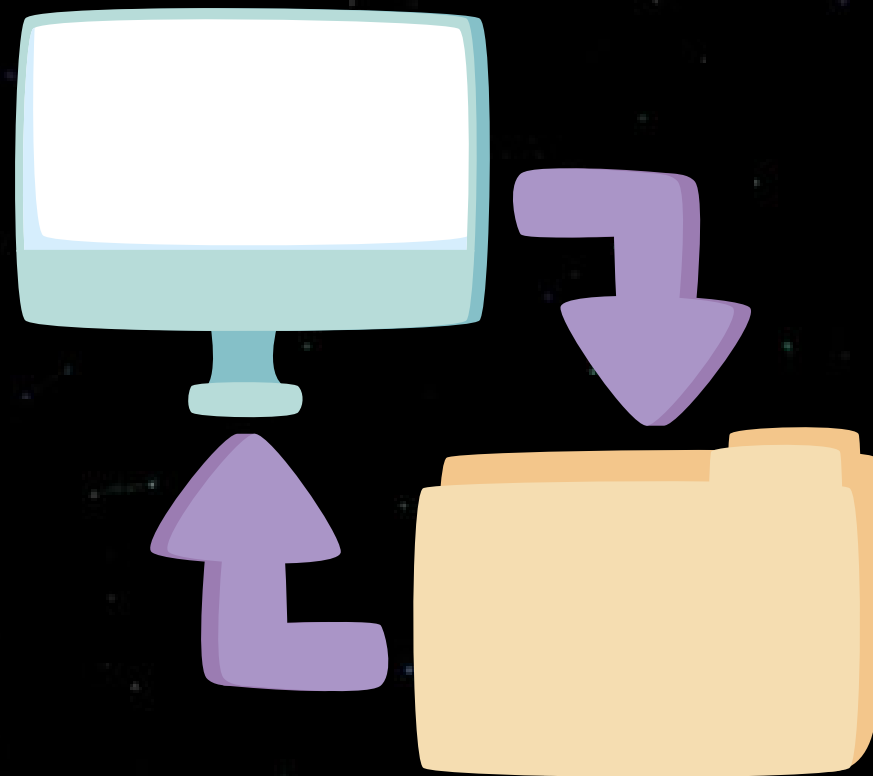
VISUALIZATION



# DATA CELL TRAINING

## S2: DATA COLLECTION

INPT



CIT





# TABLE OF CONTENT

- Where is the Data ?

- What is Web Scraping



- How do I read this HTML nonsense



- Some tips and resources



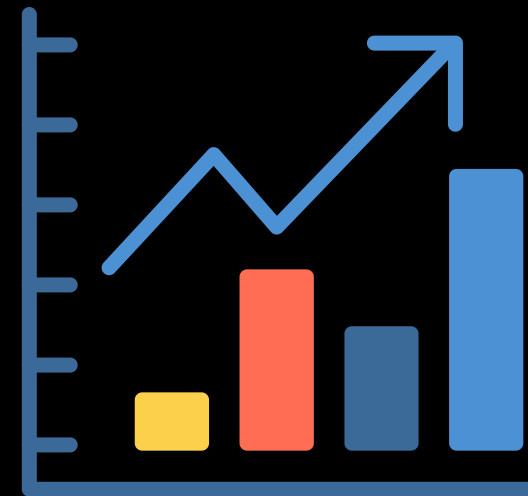
# Where is the Data



# Data can be everywhere



Surveys



Sales reports



INTERNET

Company Data

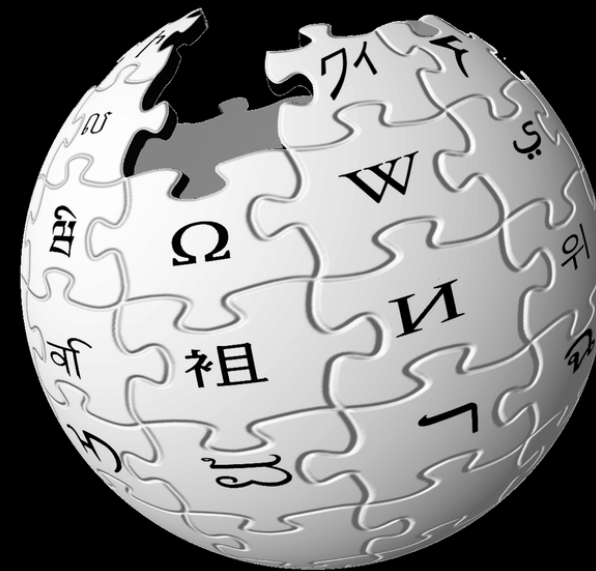


There is a lot of the Data here!



Google

INTERNET ARCHIVE  
WayBack Machine



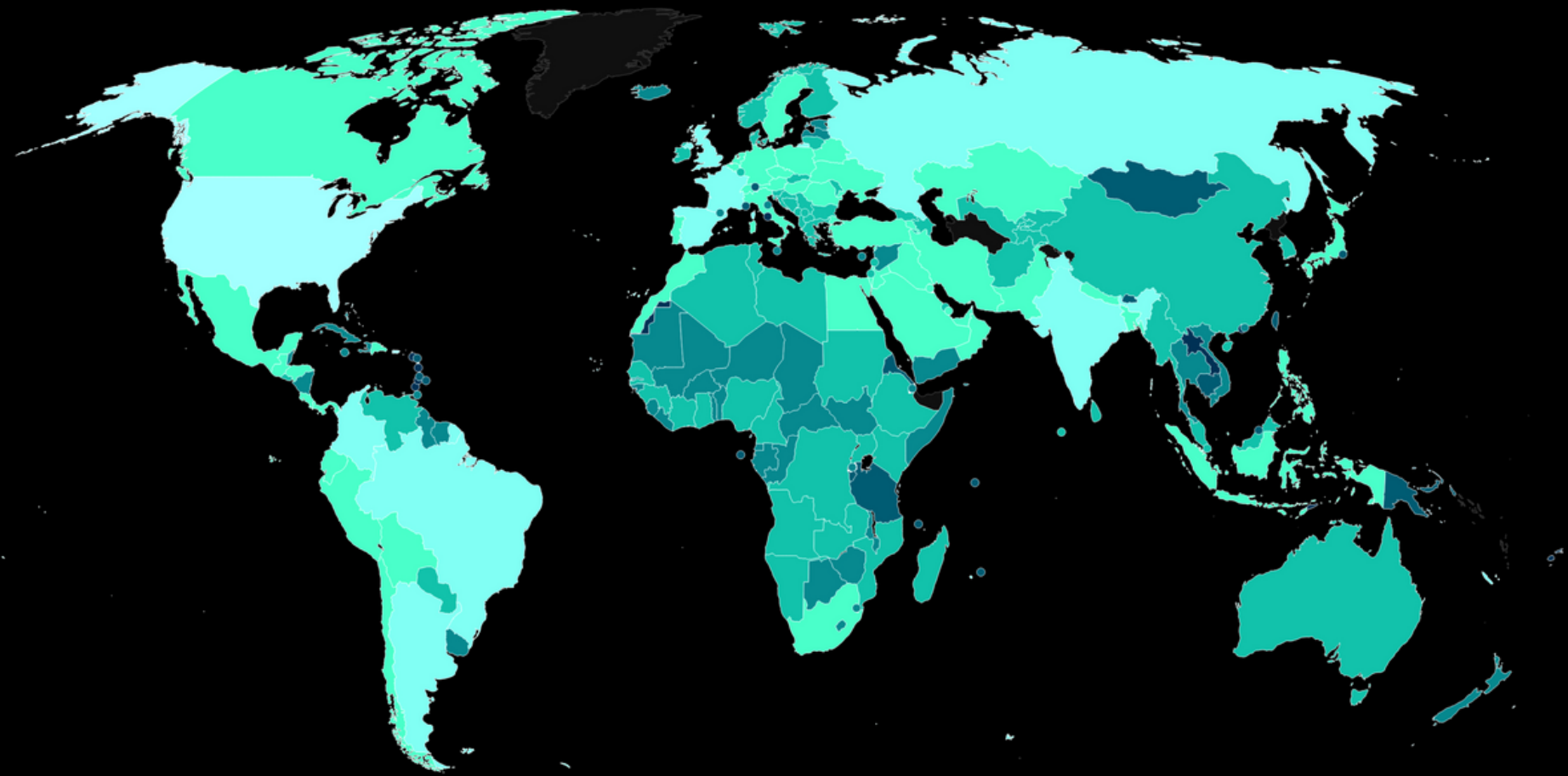
WIKIPEDIA  
*La enciclopedia libre*



# What is **Web Scrapping**



# What can I do with it ?



**Covid Cases Map**



**Stocks visualisation**

How do I read this **HTML** nonsense



# Some **tips** and **resources**





[illegible]

# ●● Medium

