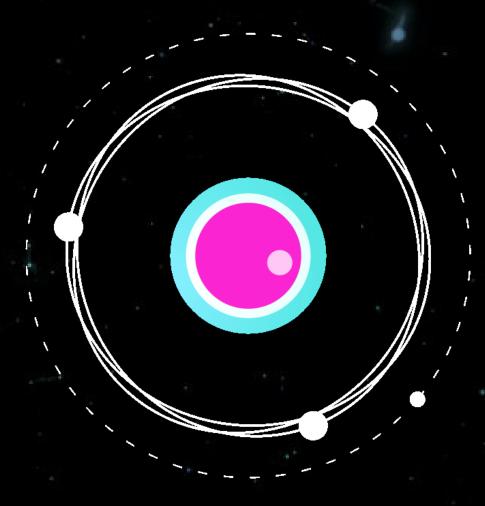
CIT

DATA CELL TRAINGING ROADMAP

INPT



MADE BY:

TAHA ELGHABI

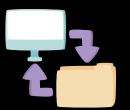




OVERVIEW, DATA ROADMAP



DATA COLLECTION



EXPLORATORY DATA ANALYSIS



CLEANING, FEATURE ENGINEERING



VISUALIZATION





Data Science buzzwords





Business Intelligence

Data Science

Analysis

Big Data

Machine Learning

Business Analytics

Deep Learning

Statistics



What the heck is Data Science





What is Data?



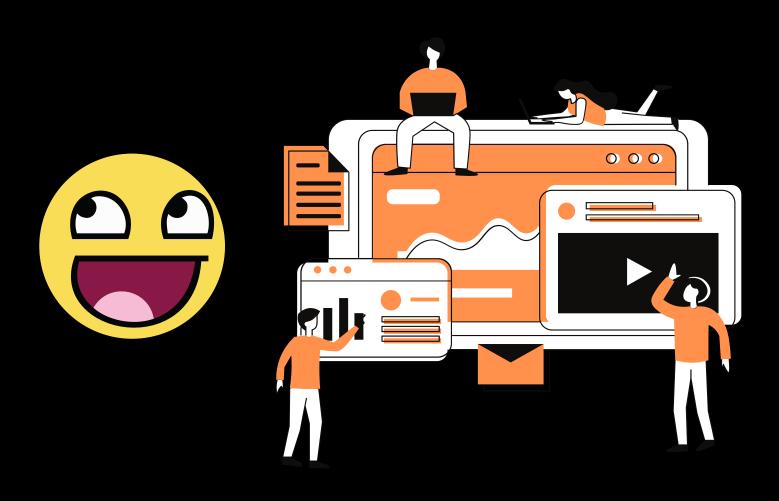
Data is a bunch of facts, such as numbers, words, measurements... or could be a descriptions of things used to better understand a subject

Types of Data

Structured



Unstructured



"It all starts with a Dataset"

- Unknown Data Scientist

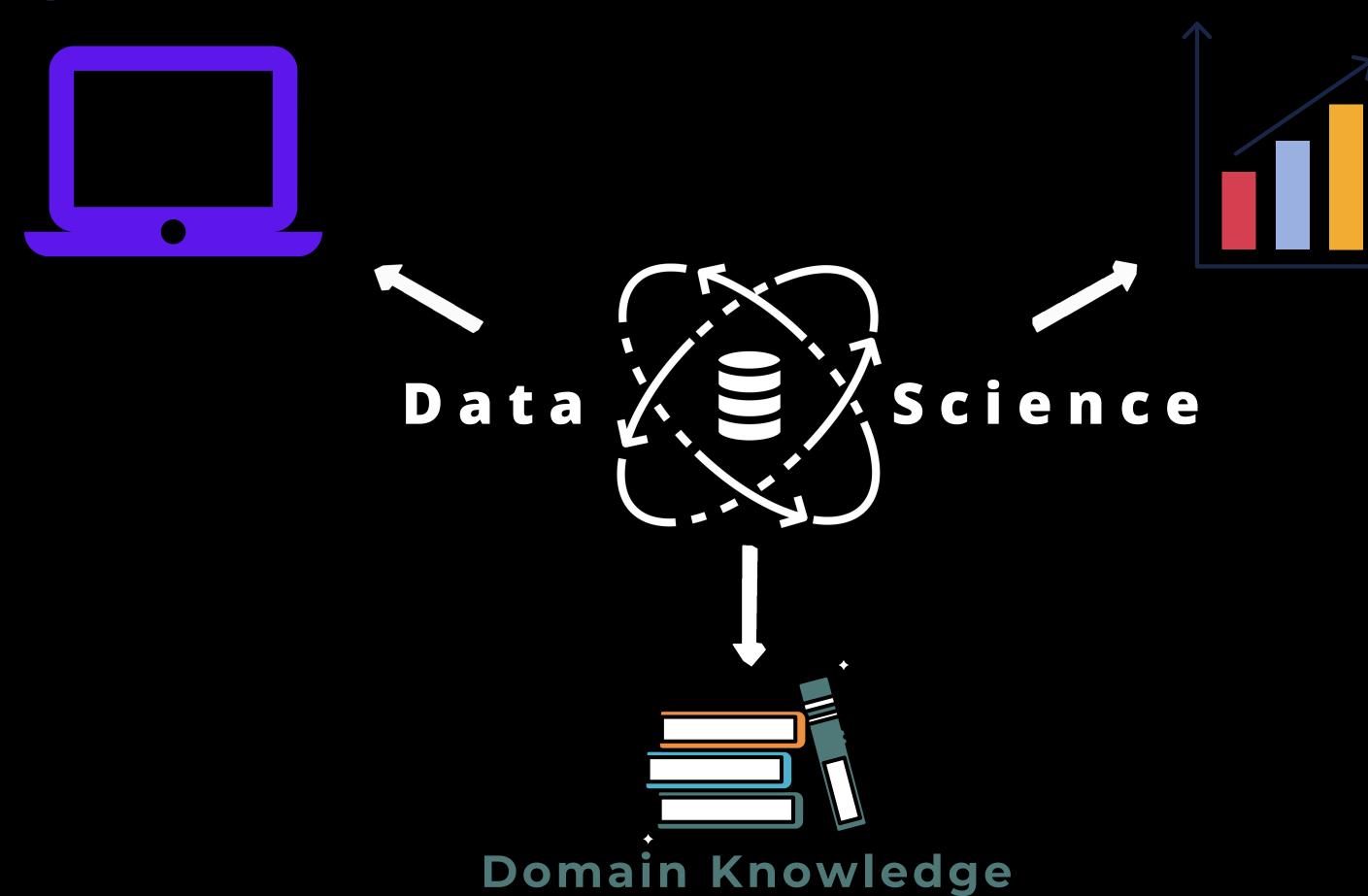
	Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/02. 3101282	7.9250	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

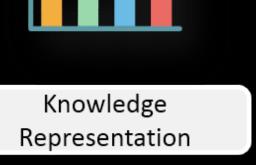
Math & Statistics



Data Science Lifecycle



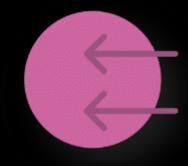
Pattern Evaluation







••••



Data Pre-processing

Data Acquisition

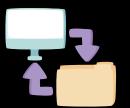


OVERVIEW, DATA ROADMAP





DATA COLLECTION



EXPLORATORY DATA ANALYSIS



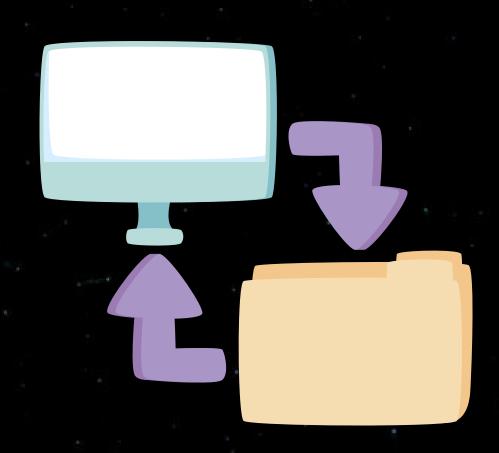
CLEANING, FEATURE ENGINEERING



VISUALIZATION











• 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







Campany Data



Sales reports

There is a lot of the Data here!



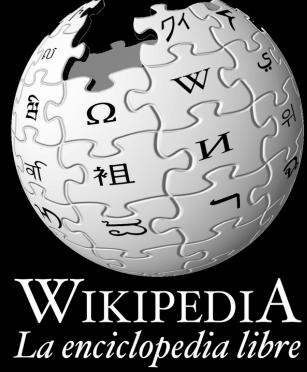


INTERNET ARCHIVE
UMBACKARINE

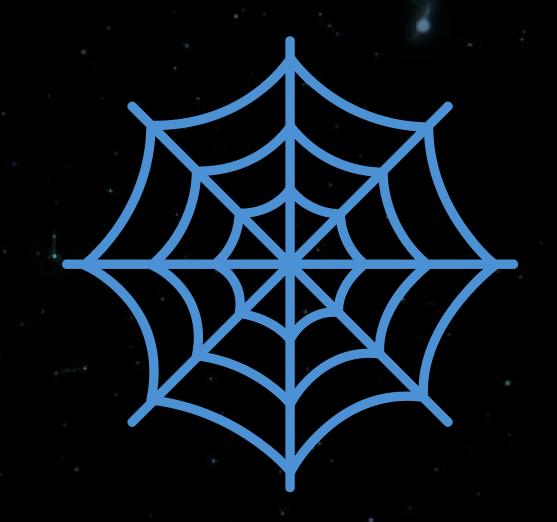
MALLER MET ARCHIVE

MALLER MET ARCHIVE

MALLER MET ARCHIVE

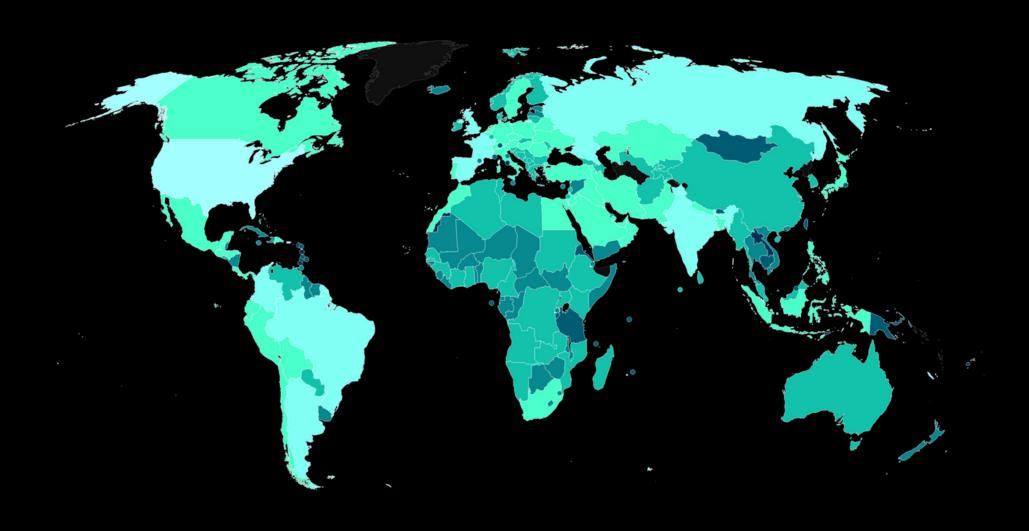


What is Web Scraping

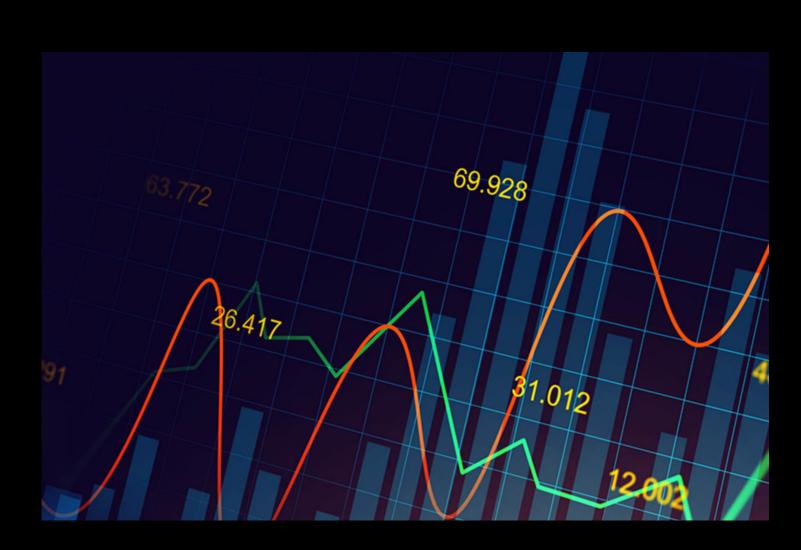




What can I do with it?



Covid Cases Map



Stocks visualisation

How do I read this HTML nonsense





Some tips and resources





vopdown-menu)"),d=b.data("target");if(d||(d=b.attr("href"),d=d&&d.replace(/.*(?=#[^\s]*\$)/,"")),! st a"),f=a.Event("hide.bs.tab",{relatedTarget:b[0]}),g=a.Event("show.bs.tab",{relatedTarget:e[0]} FaultPrevented()){var h=a(d);this.activate(b.closest("li"),c),this.activate(h,h.parent(),functio $rigger(\{type: "shown.bs.tab", relatedTarget: e[0]\})\}\}\}, c.prototype.activate=function(b,d,e)\{function(b,d,e), function(b,d,e), function(b,d,$ u > .active").removeClass("active").end().find('[data-toggle="tab"]').attr("aria-expanded",!1), 'ia-expanded",!0),h?(b[0].offsetWidth,b.addClass("in")):b.removeClass("fade"),b.parent(".dropdov ().find('[data-toggle="tab"]').attr("aria-expanded",!0),e&&e()}var g=d.find("> .active"),h=e&& fe")||!!d.find("> .fade").length);g.length&&h?g.one("bsTransitionEnd",f).emulateTransitionEnd ;var d=a.fn.tab;a.fn.tab=b,a.fn.tab.Constructor=c,a.fn.tab.noConflict=function(){return a.fn.t "show")};a(document).on("click.bs.tab.data-api",'[data-toggle="tab"]',e).on("click.bs.tab.data se strict"; function b(b){return this.each(function(){var d=a(this),e=d.data("bs.affix"),f="ob" *typeof b&&e[b]()})}var c=function(b,d){this.options=a.extend({}},c.DEFAULTS,d),this.\$target=a ",a.proxy(this.checkPosition,this)).on("click.bs.affix.data-api",a.proxy(this.checkPositionWi null, this.pinnedOffset=null, this.checkPosition()};c.VERSION="3.3.7",c.RESET="affix affix-top State=function(a,b,c,d){var e=this.\$target.scrollTop(),f=this.\$element.offset(),g=this.\$target "bottom"==this.affixed)return null!=c?!(e+this.unpin<=f.top)&&"bottom":!(e+g<=a-d)&&"bottom" !!=c&&e<=c?"top":null!=d&&i+j>=a-d&&"bottom"},c.prototype.getPinnedOffset=function(){if(this .RESET).addClass("affix"); var a=this.\$target.scrollTop(),b=this.\$element.offset(); return WithEventLoop=function(){setTimeout(a.proxy(this.checkPosition,this) 1)} int.height(), d=this.options.offset, e=d.top, f=d.bottom

Take a Look at source code

Kaggle Datasets

OMedium



Data Science Essentials: Scraping Data From the Web

Web scraping is an important skill for any data scientist to have in their toolbox

👊 Medium / Eijaz Allibhai / May 17, 2020