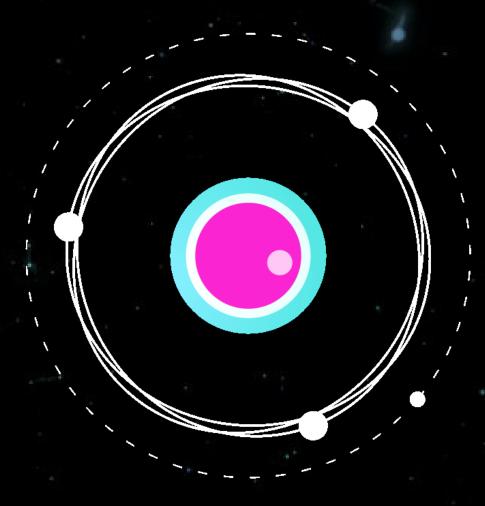
CIT

DATA CELL TRAINGING ROADMAP

INPT

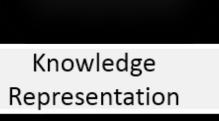


MADE BY:

TAHA ELGHABI

Data Science Lifecycle





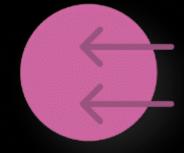


Model Building

Pattern Evaluation



Data Pre-processing



Data Acquisition



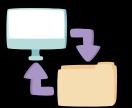


OVERVIEW, DATA ROADMAP





DATA COLLECTION



EXPLORATORY DATA ANALYSIS



CLEANING, FEATURE ENGINEERING

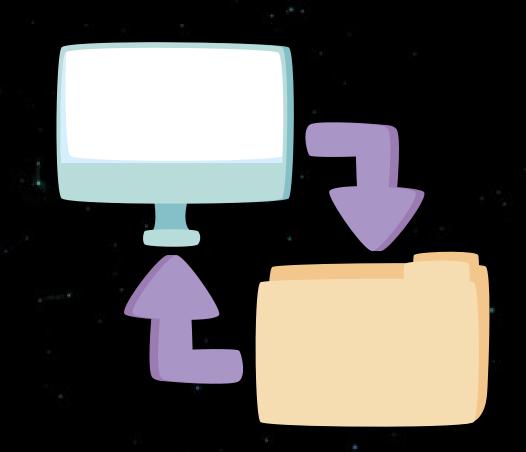


VISUALIZATION





DATA CELL TRAINING S2: DATA COLLECTION





• 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











Sales reports

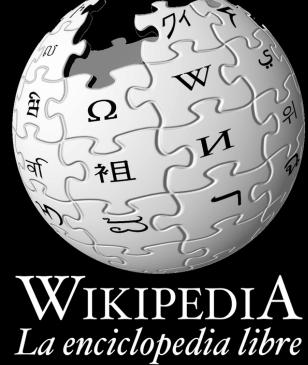


There is a lot of the Data here!



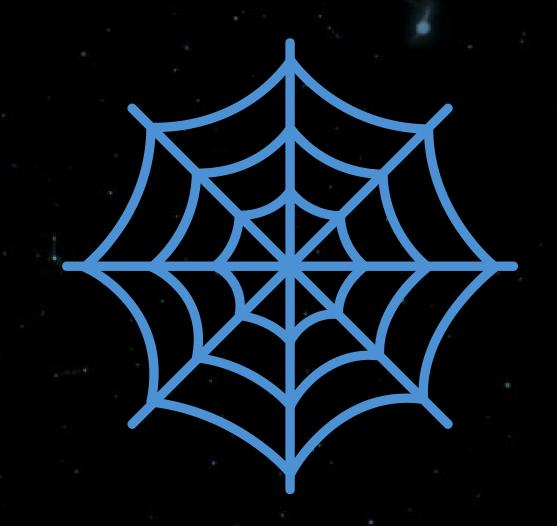


INTERNET ARCHIVE
UdyBack Alachine



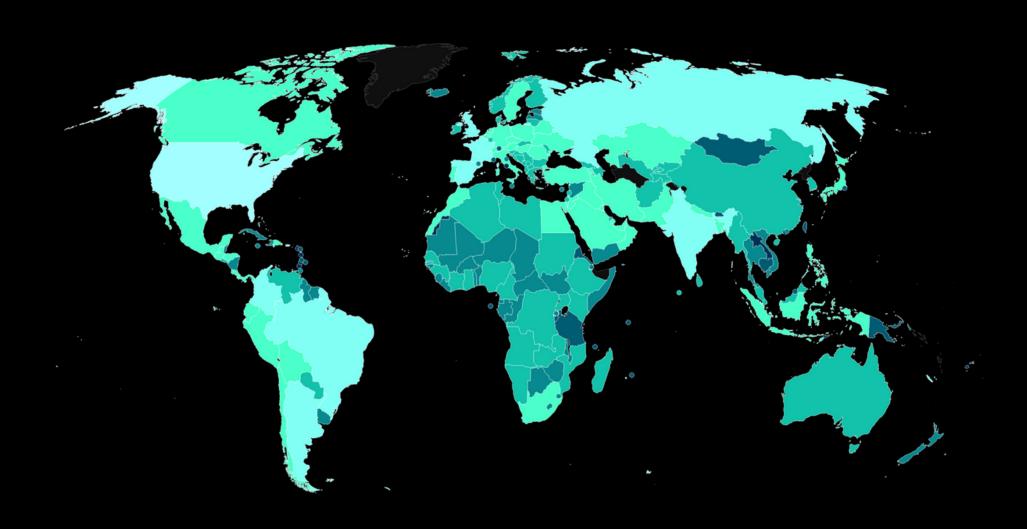


What is Web Scraping





What can I do with it?



Covid Cases Map



Stocks visualization



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 $\label{linear_rigger} \emph{rigger}(\{\textit{type:"shown.bs.tab"}, \textit{relatedTarget:e[0]}\})\}\}\}, \textit{c.prototype.activate=function(b,d,e)}\{\textit{function(b,d,e)}\}\}, \textit{c.prototype.activate=function(b,d,e)}\{\textit{function(b,d,e)}\}\}\}.$ 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

