

The AI Journey

14 JUL, 10:00 AM

Data Science
experiments with
ML.NET & Jupyter

ANTIMO MUSONE (EY)



Deltatre
Innovation
Lab

About Me



- Senior Manager & Digital Innovator of **EY Italy**
- Co-Founder of **ItaliaDotNet** Community
- Microsoft **MVP** AI
- **Intel Innovator** Program
- Co-Founder of **Fifth Ingenium**

Contacts

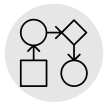
<https://www.linkedin.com/in/antimo-musone/>

<https://twitter.com/AntimoMusone>

antimo.musone@hotmail.com/antimo.musone@it.ey.com

ML : Starting with Jupyter

Project Jupyter exists to develop **open-source software**, **open-standards**, and services for **interactive computing** across dozens of programming languages. Allows you to create and share documents that **contain live code, equations, visualizations** and **narrative** text and it is powered by **Microsoft Azure**.



Do quick experiments with **Julia**, **Python**, **R** (from here the name Jupyter), C++, Ruby, Go, Java, C# ,F# and more



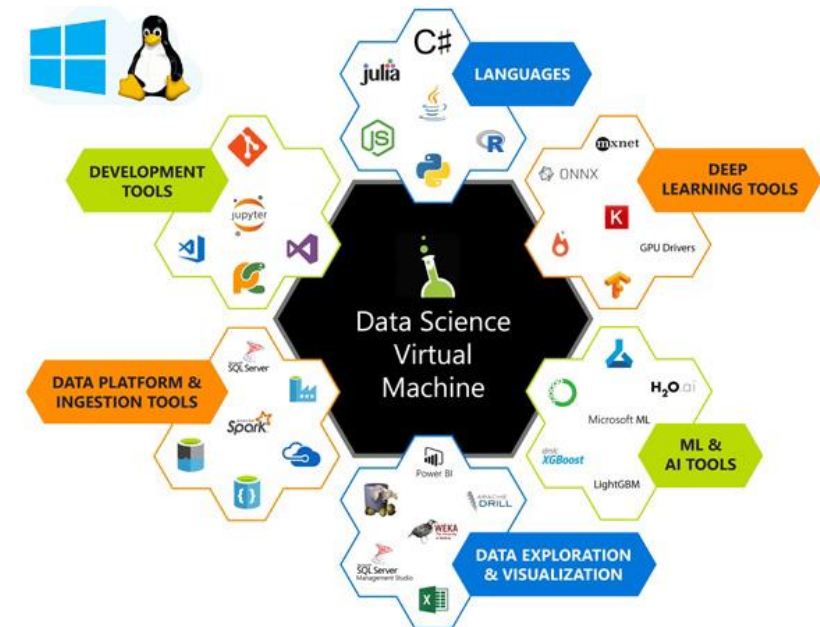
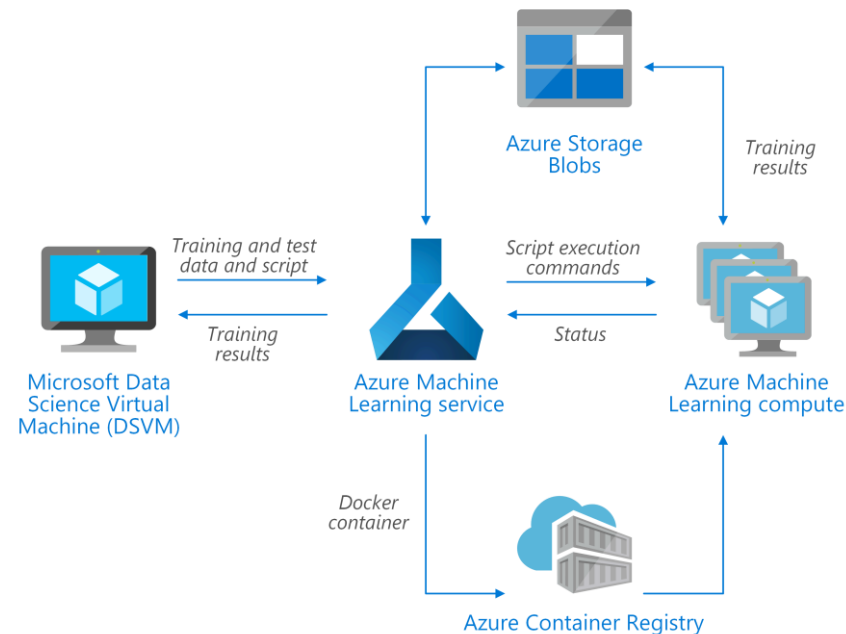
Mix live code and documentation, both formatted with Markdown and syntax highlighting



Visualize data inside the notebook or can be exported/shared as results



Scale up your Jupyter Notebook on Azure Machine Learning Service or DVSM



Interactive Computing

“The immediate exchange of input and output between a computer and the user”.

IP[y]: IPython
Interactive Computing

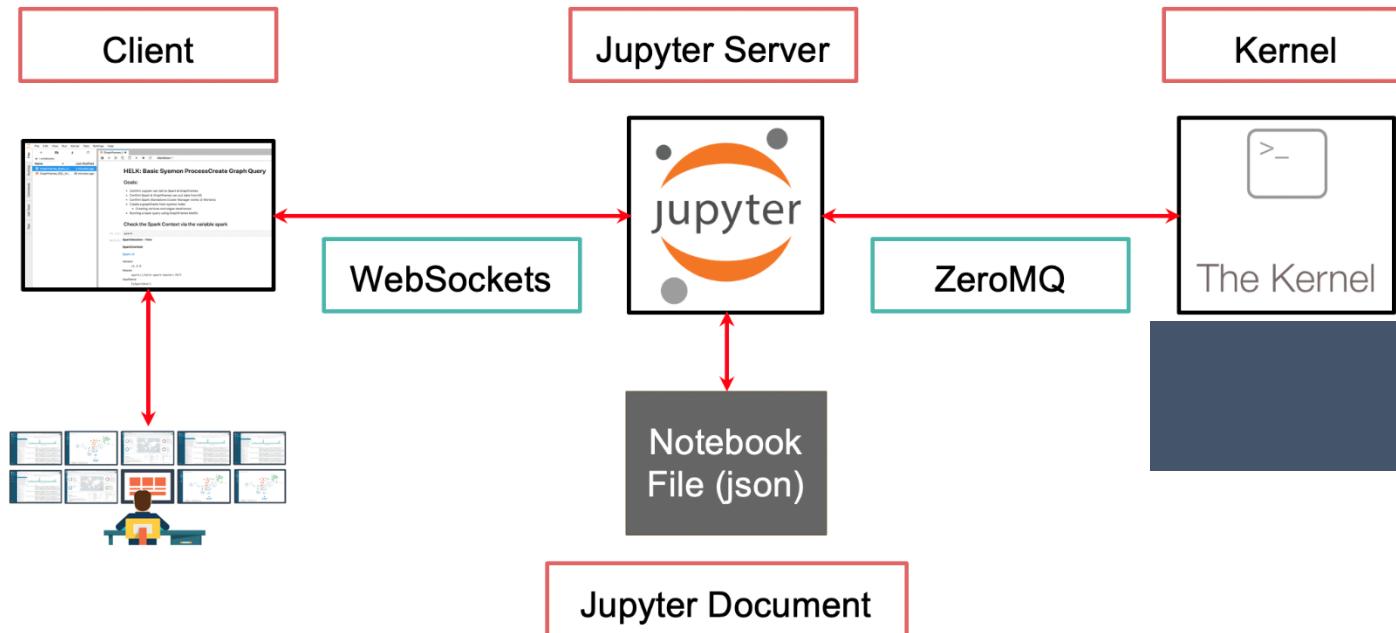
Provides a rich architecture for interactive computing with:

- A Powerful interactive shell.
- A Kernel for Jupyter.
- Support for interactive data visualization and use of GUI toolkits.
- Flexible, embeddable interpreters to load into your own projects.
- Easy to use, high performance tools for parallel computing.

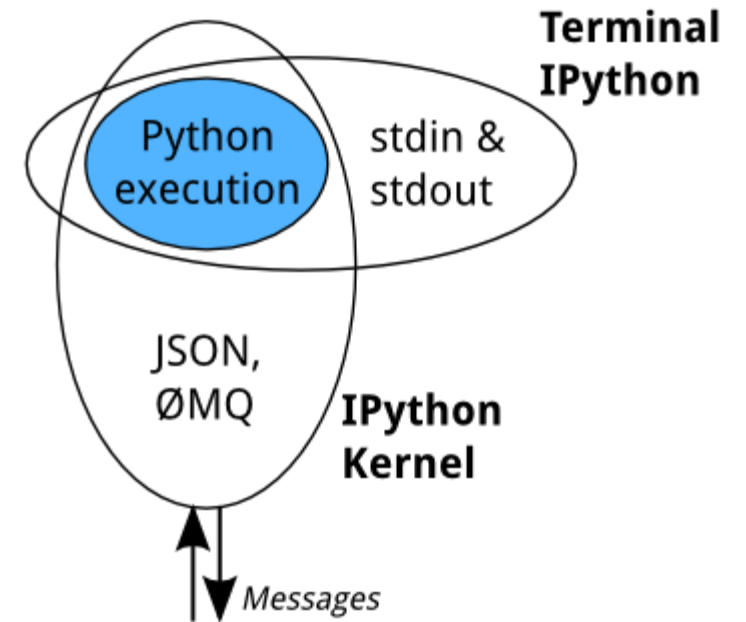
- **IPython 3.x** was the monolithic release of IPython, containing the notebook server, qtconsole, etc.
- **IPython 4.0** is the language-agnostic parts of the project, the notebook format, message protocol, qtconsole, notebook web application, etc. have moved to new projects under the name Jupyter

Jupyter Architecture

Highline Architecture



Kernel



<https://github.com/jupyter/jupyter/wiki/Jupyter-kernels>

ML.NET

An **open source** and cross-platform machine learning framework

Built for .NET developers, you can create custom **ML models** using C# or F# without having to leave the .NET ecosystem.

ML.NET offers AutoML and productive tools to help you easily build, train, and deploy high-quality custom ML models.

ML.NET has been designed as an extensible platform so that you can consume other popular ML frameworks (**TensorFlow, ONNX, Infer.NET, and more**) and have access to even more machine learning scenarios, like image classification, object detection, and more.

C#

F#

```
//Step 1. Create a ML Context
var ctx = new MLContext();

//Step 2. Read in the input data for model training
IDataView dataReader = ctx.Data
    .LoadFromTextFile<MyInput>(dataPath, hasHeader: true);

//Step 3. Build your estimator
IEstimator<ITransformer> est = ctx.Transforms.Text
    .FeaturizeText("Features", nameof(SentimentIssue.Text))
    .Append(ctx.BinaryClassification.Trainers
        .LbfgsLogisticRegression("Label", "Features"));

//Step 4. Train your Model
ITransformer trainedModel = est.Fit(dataReader);

//Step 5. Make predictions using your model
var predictionEngine = ctx.Model
    .CreatePredictionEngine<MyInput, MyOutput>(trainedModel);
```


ML.NET



Sentiment analysis

Analyze the sentiment of customer reviews using a binary classification algorithm.



Product recommendation

Recommend products based on purchase history using a matrix factorization algorithm.



Price prediction

Predict taxi fares based on parameters such as distance traveled using a regression algorithm.



Customer segmentation

Identify groups of customers with similar profiles using a clustering algorithm.



Object detection

Recognize objects in an image using an ONNX deep learning model.



Fraud detection

Detect fraudulent credit card transactions using a binary classification algorithm.



Sales spike detection

Detect spikes and changes in product sales using an anomaly detection model.



Image classification

Classify images (for example, broccoli vs. pizza) using a TensorFlow deep learning model.



Sales forecasting

Forecast future sales for products using a regression algorithm.

C# - .Net Notebook- Preview 1 – November 2019

FilesRunningClusters

Select items to perform actions on them.

0 /

3D Objects

Contacts

Desktop

Documents

Downloads

Evernote

Favorites

Links

Music

OneDrive

Pictures

Projects

Saved Games

Searches

source

Videos

UploadNew

Name

.NET (C#)

.NET (F#)

Other:

Text File

jupyter Notebook-Copy1

Last Checkpoint: a minute ago (unsaved changes)

Logout

FileEditViewInsertCellKernelWidgetsHelp

Trusted | .NET (C#)

In [1]:

```
#r "nuget:Microsoft.ML"

Installing package Microsoft.ML.....done!

Successfully added reference to package Microsoft.ML, version 1.4.0
```

In [2]:

```
using System;
using System.IO;
using System.Linq;
using Microsoft.ML;
using Microsoft.ML.Data;
using XPlot.Plotly;
```

In [3]:

```
/// <summary>
/// The HouseBlockData class holds one single housing block data record.
/// </summary>
public class HouseBlockData
{
    [LoadColumn(0)] public float Longitude { get; set; }
    [LoadColumn(1)] public float Latitude { get; set; }
    [LoadColumn(2)] public float HousingMedianAge { get; set; }
    [LoadColumn(3)] public float TotalRooms { get; set; }
    [LoadColumn(4)] public float TotalBedrooms { get; set; }
    [LoadColumn(5)] public float Population { get; set; }
    [LoadColumn(6)] public float Households { get; set; }
    [LoadColumn(7)] public float MedianIncome { get; set; }
    [LoadColumn(8)] public float MedianHouseValue { get; set; }
}
```

In [4]:

```
// filename for data set
var dataPath = Path.Combine(Environment.CurrentDirectory, "california_housing.csv");

// create the pipeline (loading content)
```

> Install .Net Core 3.0 SDK

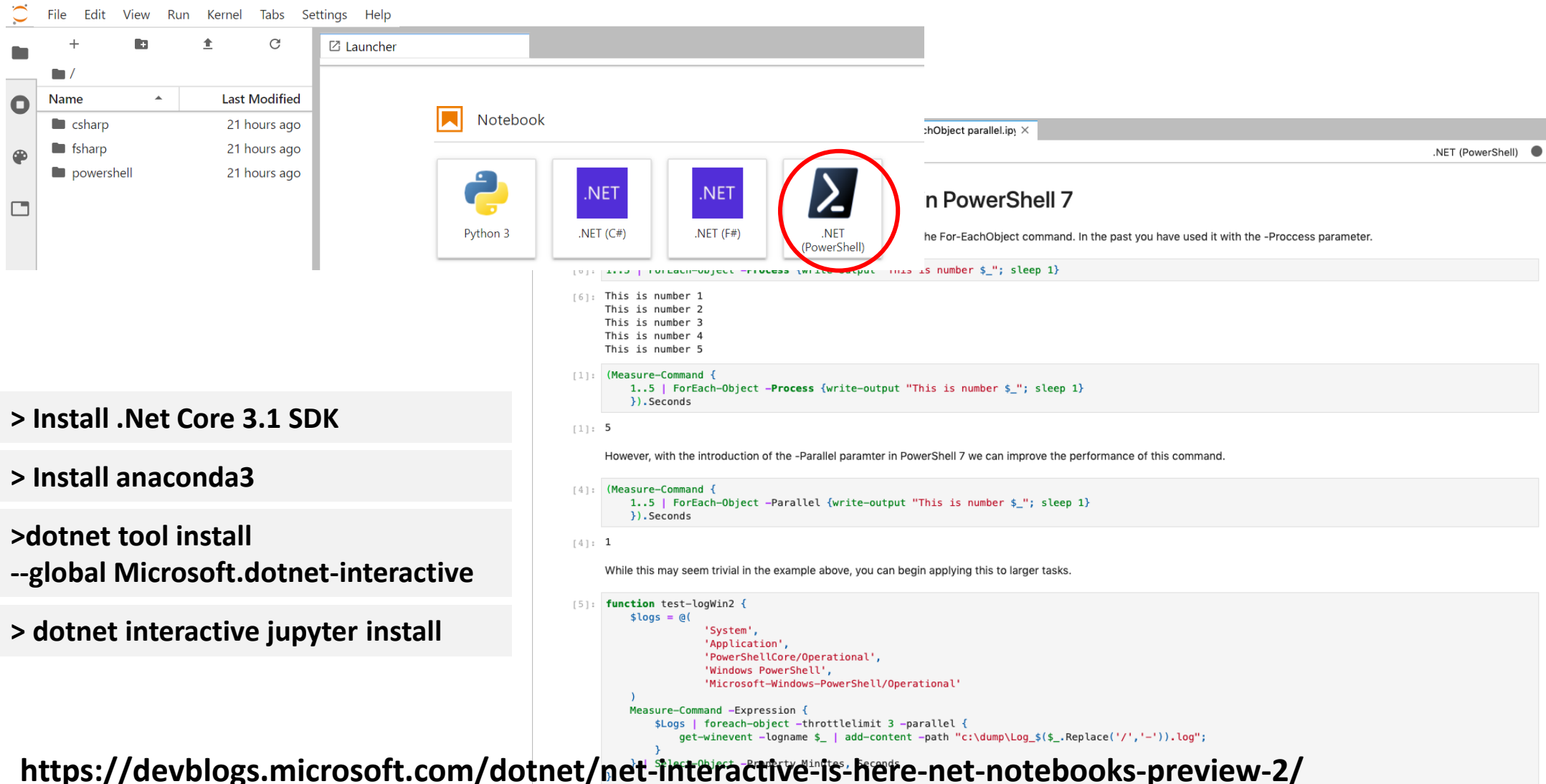
> Install anaconda3

> dotnet tool install -g dotnet-try

> dotnet try jupyter install

<https://devblogs.microsoft.com/dotnet/net-core-with-jupyter-notebooks-is-here-preview-1/>

C# - .Net Notebook – Preview 2 – February 2020



The screenshot displays the .NET Notebook application interface. On the left, a file explorer shows a directory structure with folders for 'csharp', 'fsharp', and 'powershell'. The main area features a 'Notebook' tab with a grid of icons for Python 3, .NET (C#), .NET (F#), and .NET (PowerShell). The .NET (PowerShell) icon is highlighted with a red circle. Below the icons, a PowerShell 7 notebook is open, showing a code cell with the following content:

```
[0]: 1..5 | ForEach-Object -Process {write-output "This is number $_"; sleep 1}
```

The output of the command is displayed as:

```
[6]: This is number 1  
This is number 2  
This is number 3  
This is number 4  
This is number 5
```

The notebook also includes a section titled 'n PowerShell 7' with the text: 'he For-EachObject command. In the past you have used it with the -Process parameter.'

Below this, a code cell shows a command to measure the execution time of the ForEach-Object command:

```
[1]: (Measure-Command {  
1..5 | ForEach-Object -Process {write-output "This is number $_"; sleep 1  
}).Seconds
```

The output of the command is:

```
[1]: 5
```

However, with the introduction of the -Parallel parameter in PowerShell 7 we can improve the performance of this command.

A code cell shows the command with the -Parallel parameter:

```
[4]: (Measure-Command {  
1..5 | ForEach-Object -Parallel {write-output "This is number $_"; sleep 1  
}).Seconds
```

The output of the command is:

```
[4]: 1
```

While this may seem trivial in the example above, you can begin applying this to larger tasks.

A code cell shows a function to test the performance of the ForEach-Object command:

```
[5]: function test-logWin2 {  
$logs = @(  
'System',  
'Application',  
'PowerShellCore/Operational',  
'Windows PowerShell',  
'Microsoft-Windows-PowerShell/Operational'  
)  
Measure-Command -Expression {  
$logs | foreach-object -throttlelimit 3 -parallel {  
get-winevent -logname $_ | add-content -path "c:\dump\Log_$($_.Replace('/', '-')).log";  
}  
}  
Select-Object -Property Minutes, Seconds
```

> Install .Net Core 3.1 SDK

> Install anaconda3

> dotnet tool install

--global Microsoft.dotnet-interactive

> dotnet interactive jupyter install

<https://devblogs.microsoft.com/dotnet/net-interactive-is-here-net-notebooks-preview-2/>

Try Jupyter for Free

<https://mybinder.org/>

<https://notebooks.azure.com/>

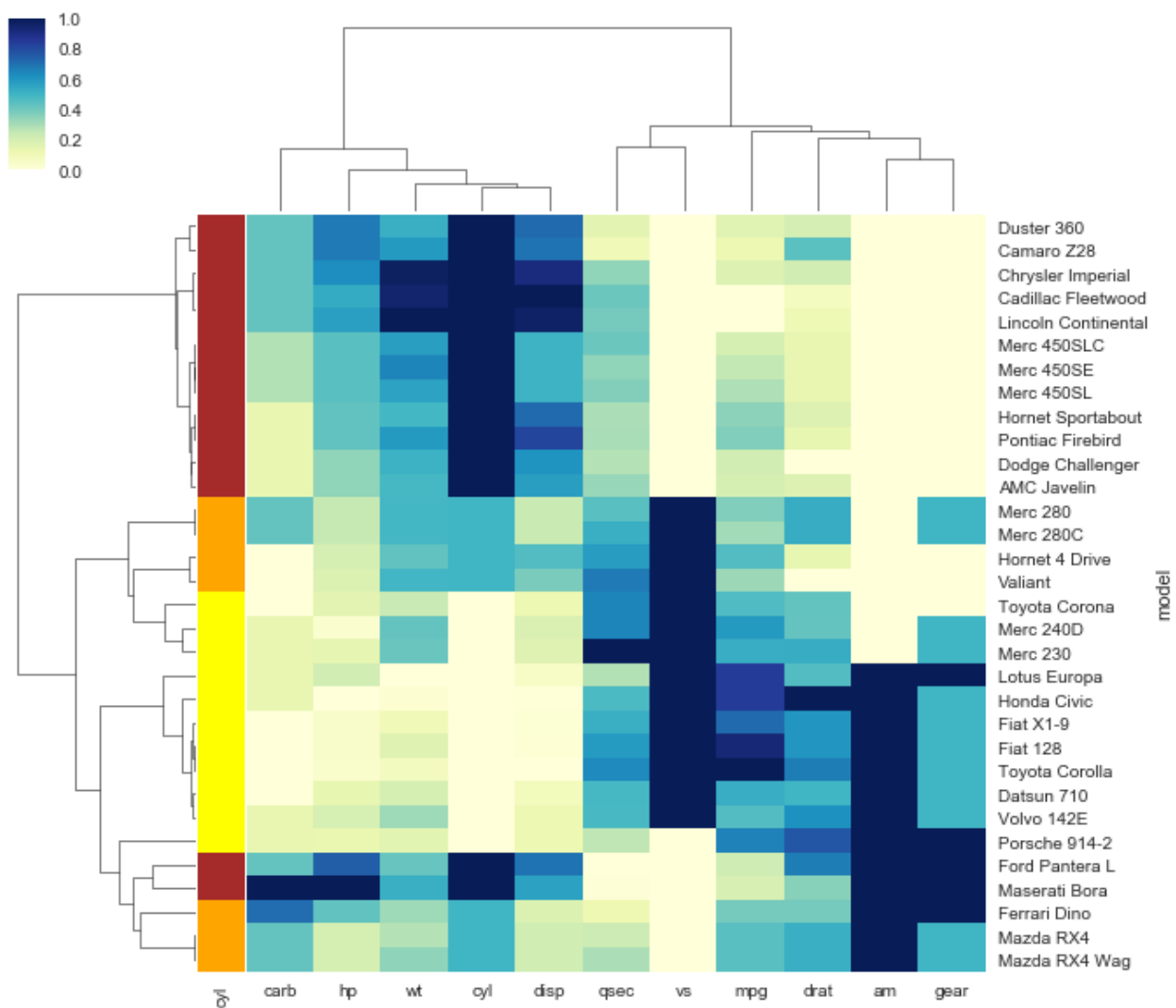
<https://colab.research.google.com/>

<https://www.kaggle.com/notebooks>

A black and white close-up photograph of a woman with short, wavy blonde hair. She has a very wide-eyed, open-mouthed expression, looking directly at the camera with a look of surprise or shock. The background is a plain, dark grey.

**Data Science
Experimentation?**

Heatmap



Folium

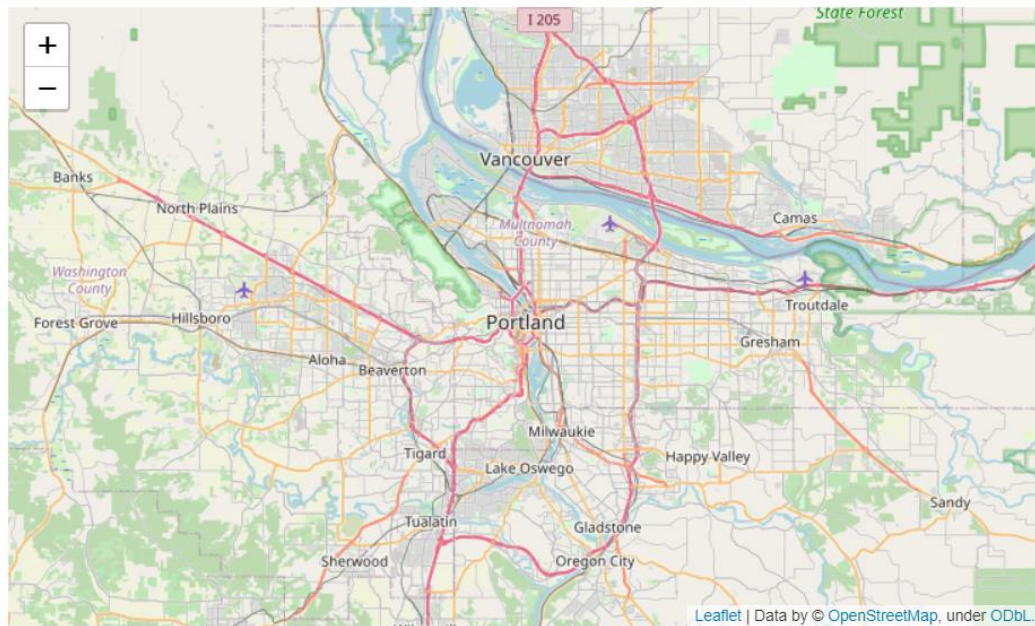
To create a base map, simply pass your starting coordinates to Folium:

```
import folium

m = folium.Map(location=[45.5236, -122.6750])
```

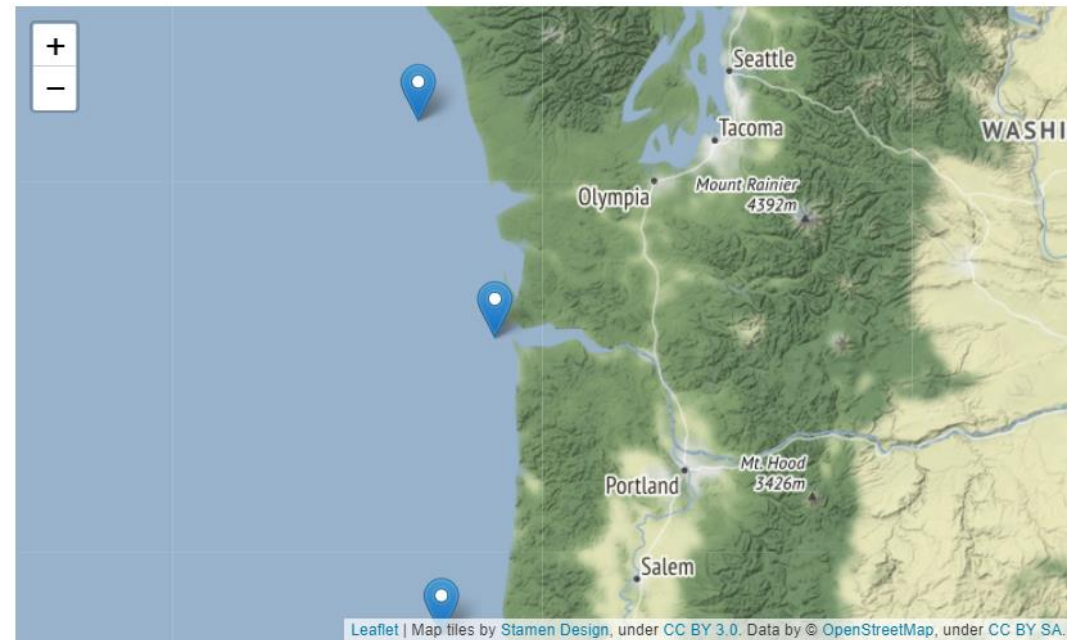
By default, if you display it in a Jupyter notebook, simply ask for the object representation:

m



```
folium.Marker(
    location=[46.216, -124.1280],
    popup=folium.Popup(max_width=450).add_child(
        folium.Vega(vis3, width=450, height=250))
).add_to(m)
```

m



<https://python-visualization.github.io/folium/>

ipywidgets

Widgets are eventful python objects that have a representation in the browser, often as a control like a slider, textbox, etc.

```
[1]: import ipywidgets as widgets
```

```
[2]: widgets.IntSlider()
```

[2]:  62

```
[12]: a = widgets.FloatText()  
      b = widgets.FloatSlider()  
      display(a,b)  
  
      mylink = widgets.jslink((a, 'value'), (b, 'value'))
```

50,7

 50.70

<https://ipywidgets.readthedocs.io/en/latest/index.html>

Voilà

Voilà turns Jupyter notebooks into standalone web applications. Unlike the usual HTML-converted notebooks, each user connecting to the Voilà tornado application gets a dedicated Jupyter kernel which can execute the callbacks to changes in Jupyter interactive widgets.

```
pip install voilà
```

```
cd notebooks/  
voilà bqplot.ipynb
```

The logo for Voilà, featuring the word "voilà" in a white, cursive script font on a teal background. A small yellow banana is positioned above the final 'à'.

<https://voila.readthedocs.io/en/stable/>

<https://voila-gallery.org/>

Dataset COVID 19

Data repository is Novel Coronavirus (COVID-19) operated by the **Johns Hopkins University** Center for Systems Science and Engineering, open and available on Github.

<https://github.com/CSSEGISandData/COVID-19>

Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE <https://systems.jhu.edu/research/publ...>

johns-hopkins-university systems-science engineering covid-19 2019-ncov coronavirus csse jhu

592 commits

3 branches

0 packages

0 releases

4 contributors

Branch: master

New pull request

Find file

Clone or download

Ryan Lau updated data

Latest commit 22753a5 1 hour ago

archived_data

archived_0325

21 days ago

csse_covid_19_data

updated data

1 hour ago

who_covid_19_situation_reports

Add files via upload

11 days ago

.gitignore

update

2 months ago

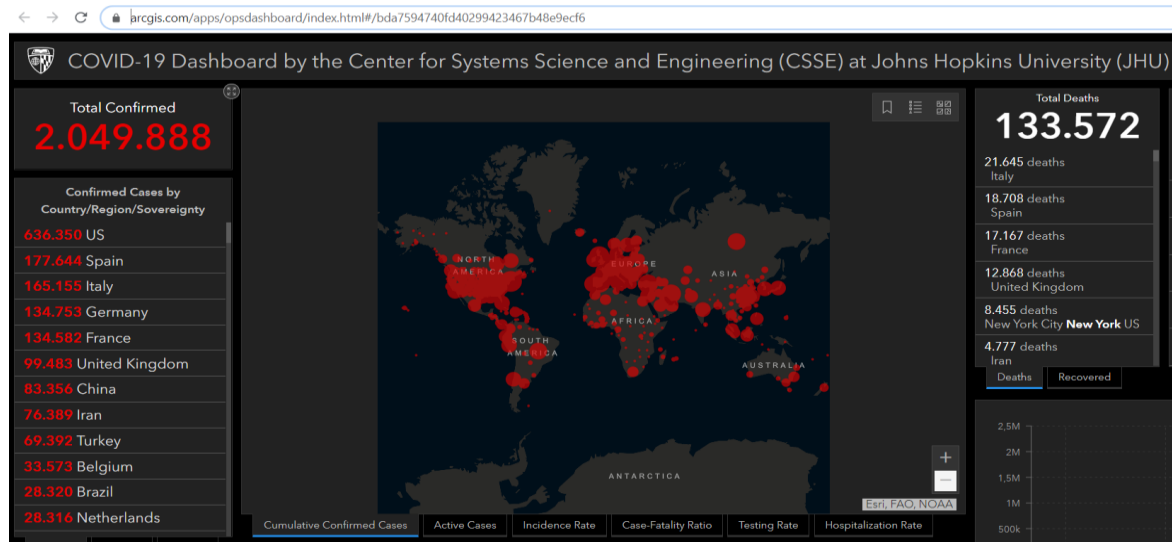
README.md

add covid tracking prj

4 days ago

README.md

2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE



Sample Dashboard based on Data repository

<https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

Metrics to Manage COVID-19

R_t: the effective reproduction number

*R*0 is the [basic reproduction number](#) of an epidemic. It's defined as the number of secondary infections produced by a single infection. If *R*0 is greater than one, the epidemic spreads quickly, If *R*0 is less than one, epidemic don't spreads and disappears before everyone becomes infected. While *R*0 is a useful measure but it's static so because the Epidemy changes so rapidly the only true way to combat COVID19 is to understand and manage by *R*_t.

Bettencourt & Ribeiro's Approach

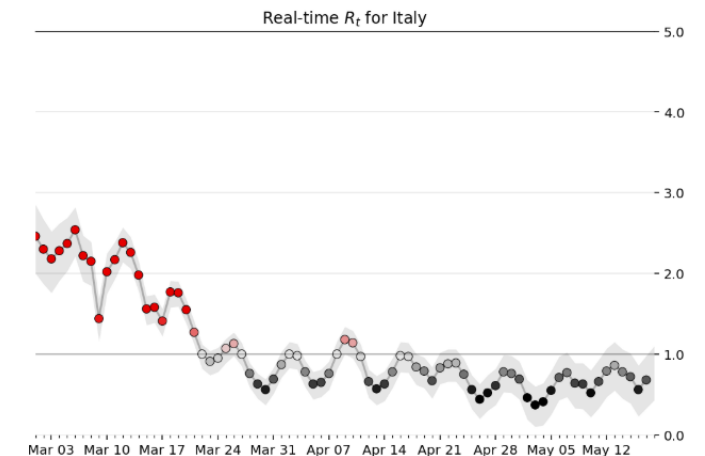
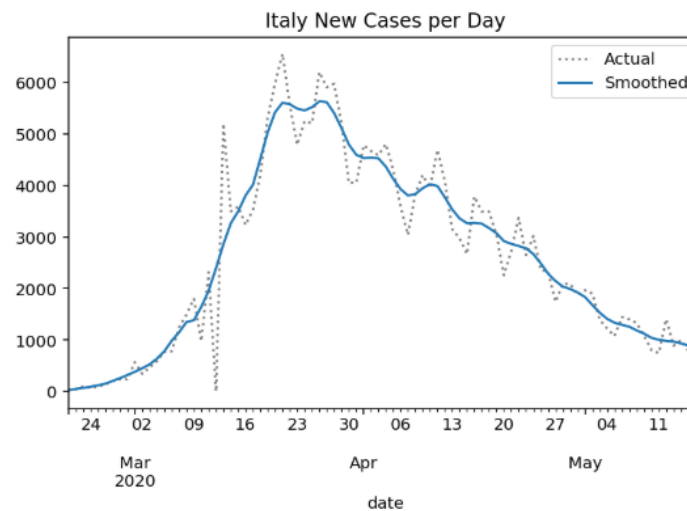
Described in the paper, “Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases.”

This is **Baves' Theorem** :

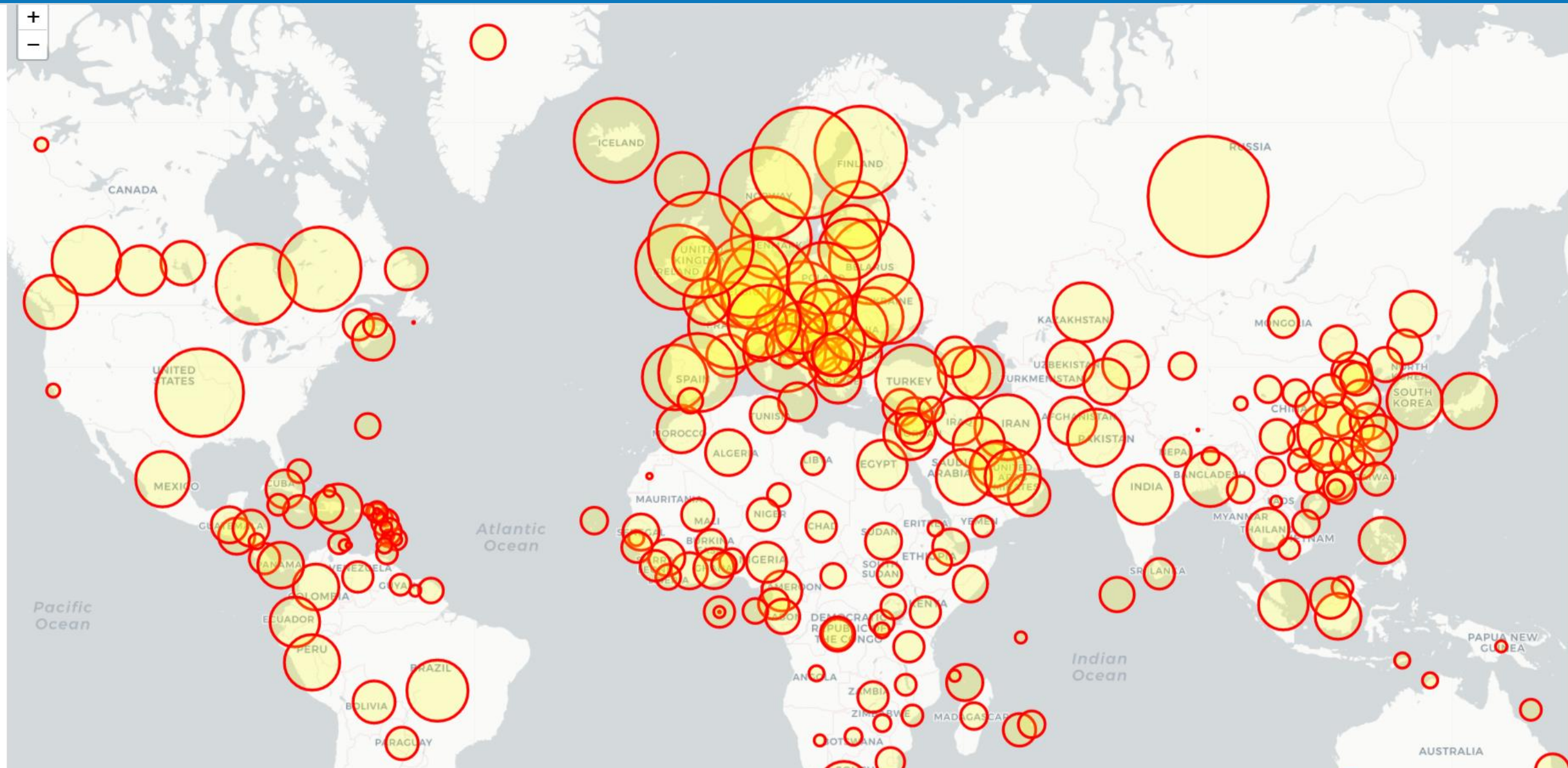
$$P(R_t|k) = \frac{P(k|R_t) \cdot P(R_t)}{P(k)}$$

This is **Poisson Distribution**

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$



<https://bit.ly/build2020-covid>



DEMO – DASHBOARD COVID




```
from ipywidgets import interact, interactive, fixed, interact_manual
from IPython.core.display import display, HTML
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
import folium
import plotly.graph_objects as go
import seaborn as sns
import ipywidgets as widgets
import seaborn as sn
import matplotlib.pyplot as plt
```

```
In [59]: death_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/covid19_deaths_global.csv')
confirmed_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/covid19_confirmed_global.csv')
recovered_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/covid19_recovered_global.csv')
country_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/web-data/data/cases_country.csv')
country_df.to_csv('cases_country.csv')
```

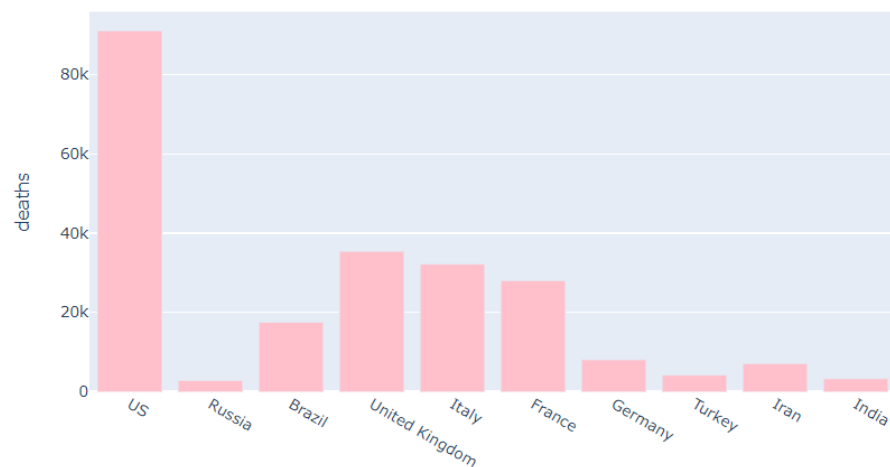
Descrizione COVID-19 Data Repository by Johns Hopkins University

```
In [60]: death_df.head()
```

```
Out[60]:
```

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	5/9/20	5/10/20	5/11/20	5/12/20	5/13/20	5/14/20
0	NaN	Afghanistan	33.0000	65.0000	0	0	0	0	0	0	...	115	120	122	127	132	137
1	NaN	Albania	41.1533	20.1683	0	0	0	0	0	0	...	31	31	31	31	31	31
2	NaN	Algeria	28.0339	1.6596	0	0	0	0	0	0	...	494	502	507	515	522	529
3	NaN	Andorra	42.5063	1.5218	0	0	0	0	0	0	...	48	48	48	48	49	49
4	NaN	Angola	-11.2027	17.8739	0	0	0	0	0	0	...	2	2	2	2	2	2

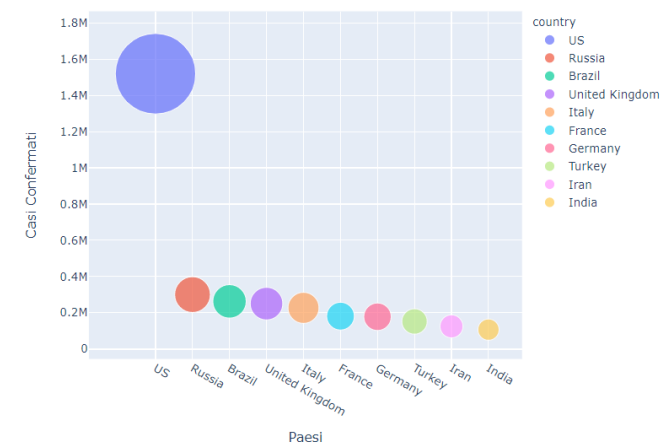
Top 10 Paesi per i Casi di Decesso



```
ipynLayout = widgets.Layout(border='solid 2px green')
ipynLayout.display='none'
widgets.VBox([fig, layout=ipynLayout])
```

n 10

10 Paesi + Colpiti

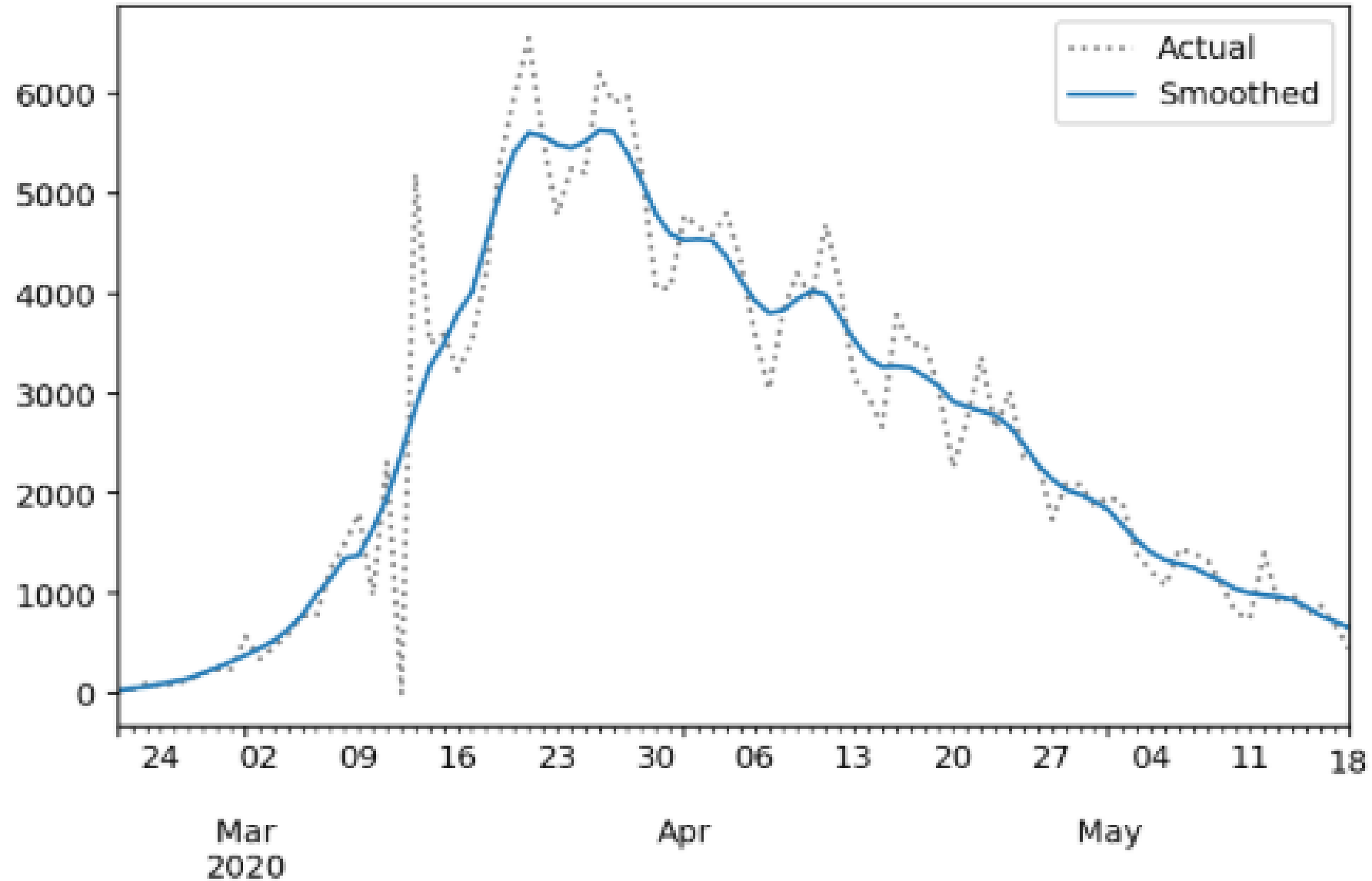


```
In [46]: def plot_cases_of_a_country(country):
labels = ['confirmed', 'deaths']
colors = ['blue', 'red']
```

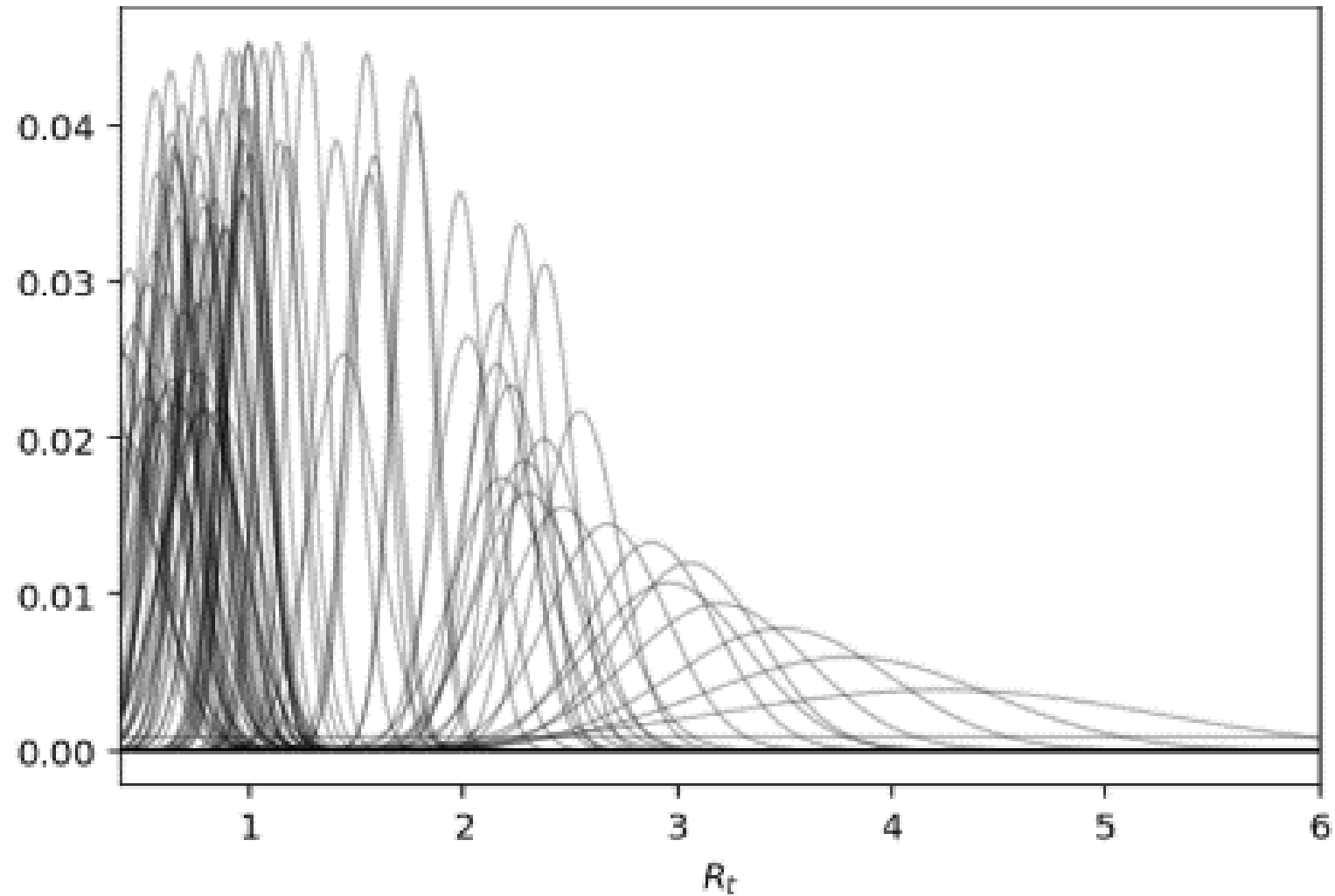
n 10

	country	last_update	confirmed	deaths	recovered	active	incident_rate	mortality_rate
17	US	2020-05-19 18:06:47	1519317	90995	283178	1143038	461.144912	5.989204
13	Russia	2020-05-19 18:06:47	299941	2837	76130	220974	205.531305	0.945853
40	Brazil	2020-05-19 18:06:47	262545	17509	100459	144577	123.516057	6.668952
16	United Kingdom	2020-05-19 18:06:47	250121	35421	1094	213606	368.442662	14.161546
10	Italy	2020-05-19 18:06:47	226699	32169	129401	65129	374.945660	14.190182
6	France	2020-05-19 18:06:47	180933	28025	62678	90230	277.192071	15.489159
7	Germany	2020-05-19 18:06:47	177574	8081	155681	13812	211.942753	4.550779
175	Turkey	2020-05-19 18:06:47	151615	4199	112895	34521	179.768410	2.769515
91	Iran	2020-05-19 18:06:47	124603	7119	97173	20311	148.349350	5.713346
89	India	2020-05-19 18:06:47	106446	3301	40865	62280	7.713454	3.101103

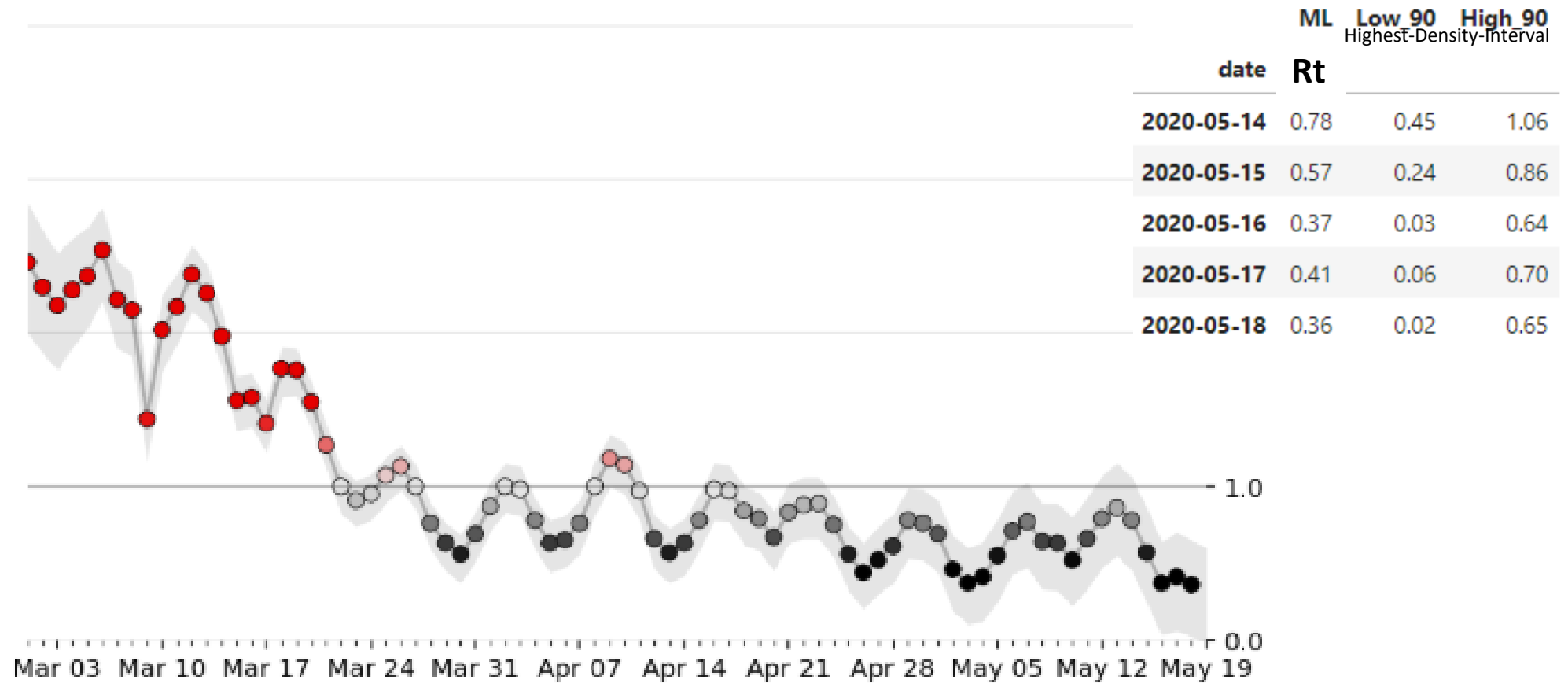
Contagion Trend- Italy



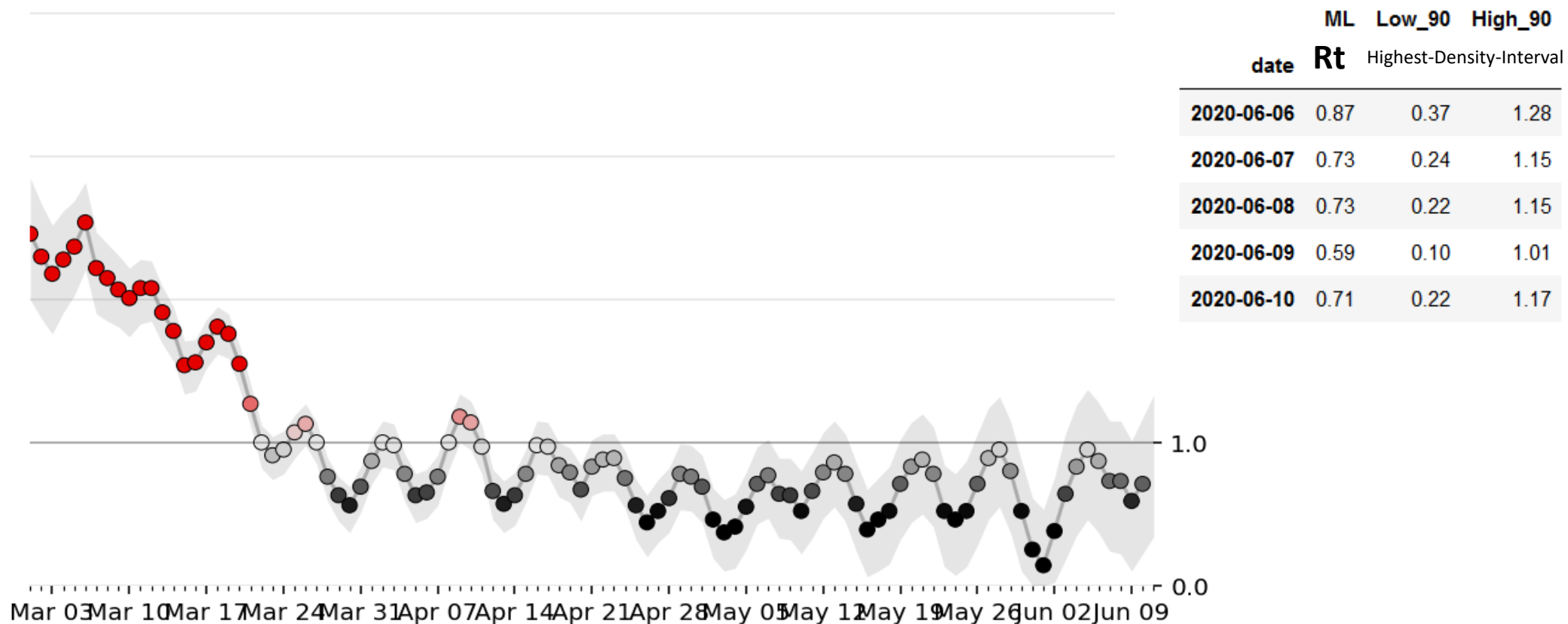
Probability Trend of R_t value – 7 Days



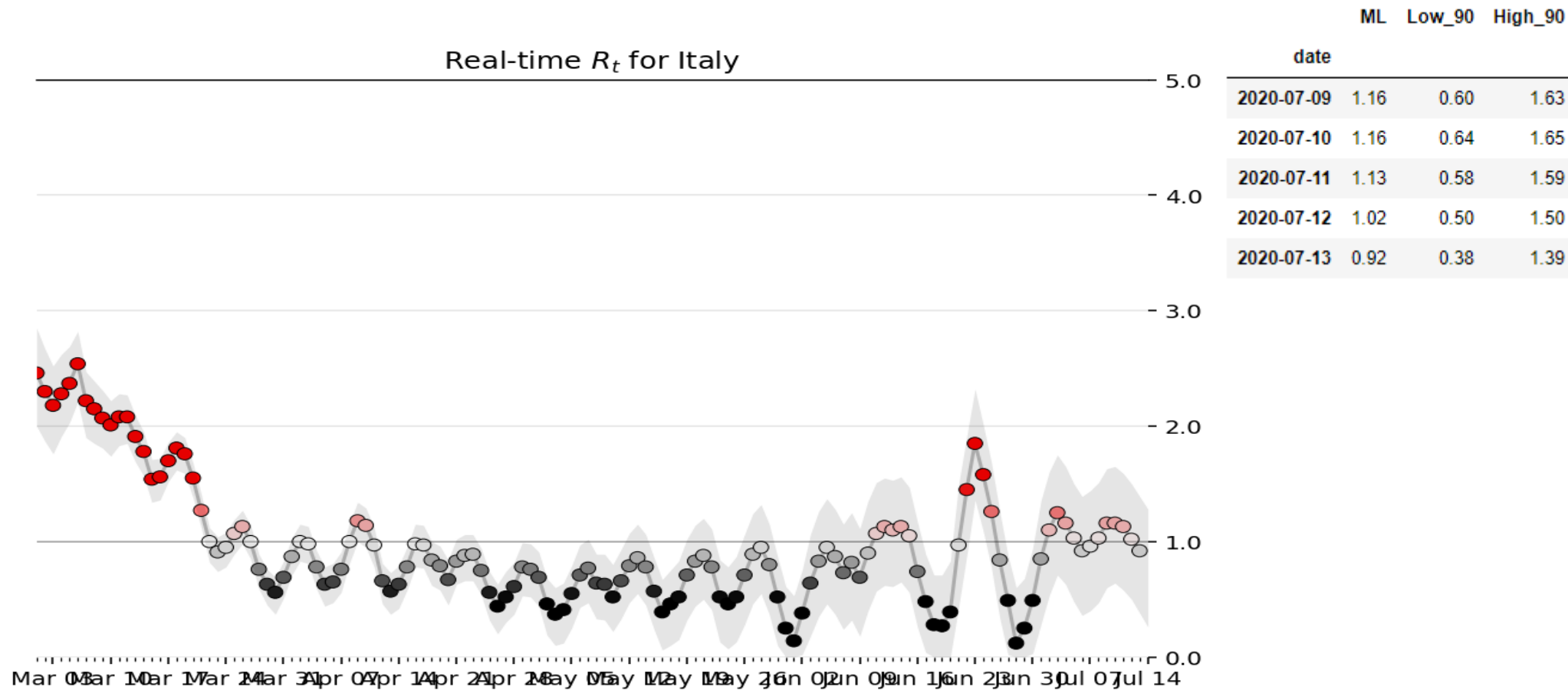
Rt Trend with Highest-Density-Interval – 19 May



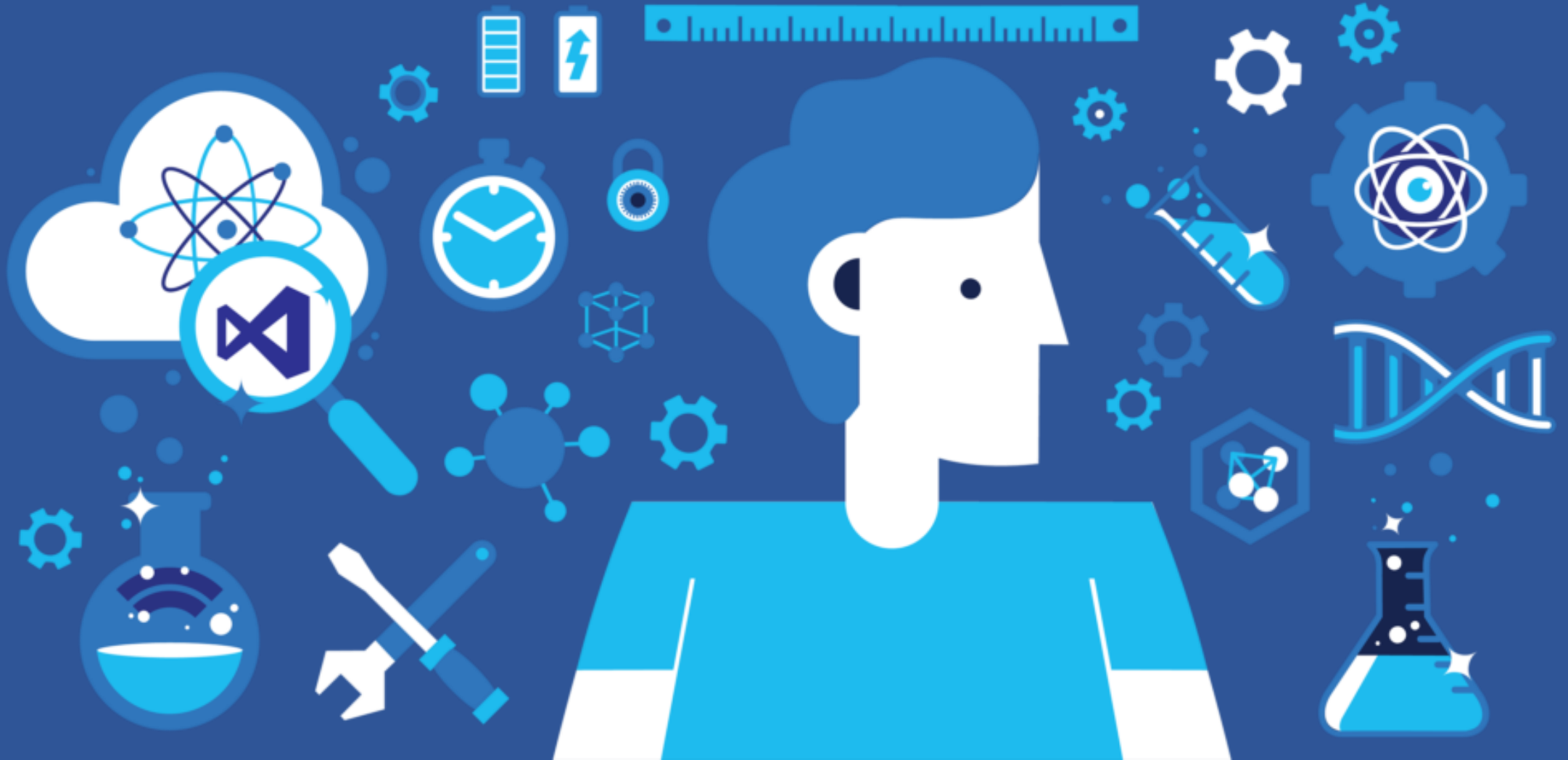
Rt Trend with Highest-Density-Interval – 11 June



Rt Trend with Highest-Density-Interval – 14 July



DEMO – .NET Nootebook & ML.Net



Part - 1

DataFrame is a new type introduced in .Net. It is similar to DataFrame in Python is used to manipulate data in notebooks. It's a collection of columns containing data similar to a table

- Load a CSV
- Metadata
 - Description
 - Info
- Display records
 - Head
 - Sample
- Filtering, Grouping, Aggregate
- Reports

Part - 2

Applying machine learning for making a prediction using time-series, building a model on the number of confirmed cases and predicting for the next 7 days

- Dataset and Transformations
- Data Classes Input/Output
 - ConfirmedData : Provides a map between columns in a dataset
 - ConfirmedForecast :
- Data Analysis
- ML Pipeline
- Train Model
- Prediction/Forecasting
- Prediction Visualization



Question Time

About Me



- Senior Manager & Digital Innovator of **EY Italy**
- Co-Founder of **Fifth Ingenium**
- Co-Founder of **ItaliaDotNet** Community
- Microsoft **MVP** AI
- **Intel Innovator** Program

Contacts

<https://www.linkedin.com/in/antimo-musone/>

<https://twitter.com/AntimoMusone>

antimo.musone@hotmail.com / antimo.musone@it.ey.com

+39 345 3636642

Thank You!!!