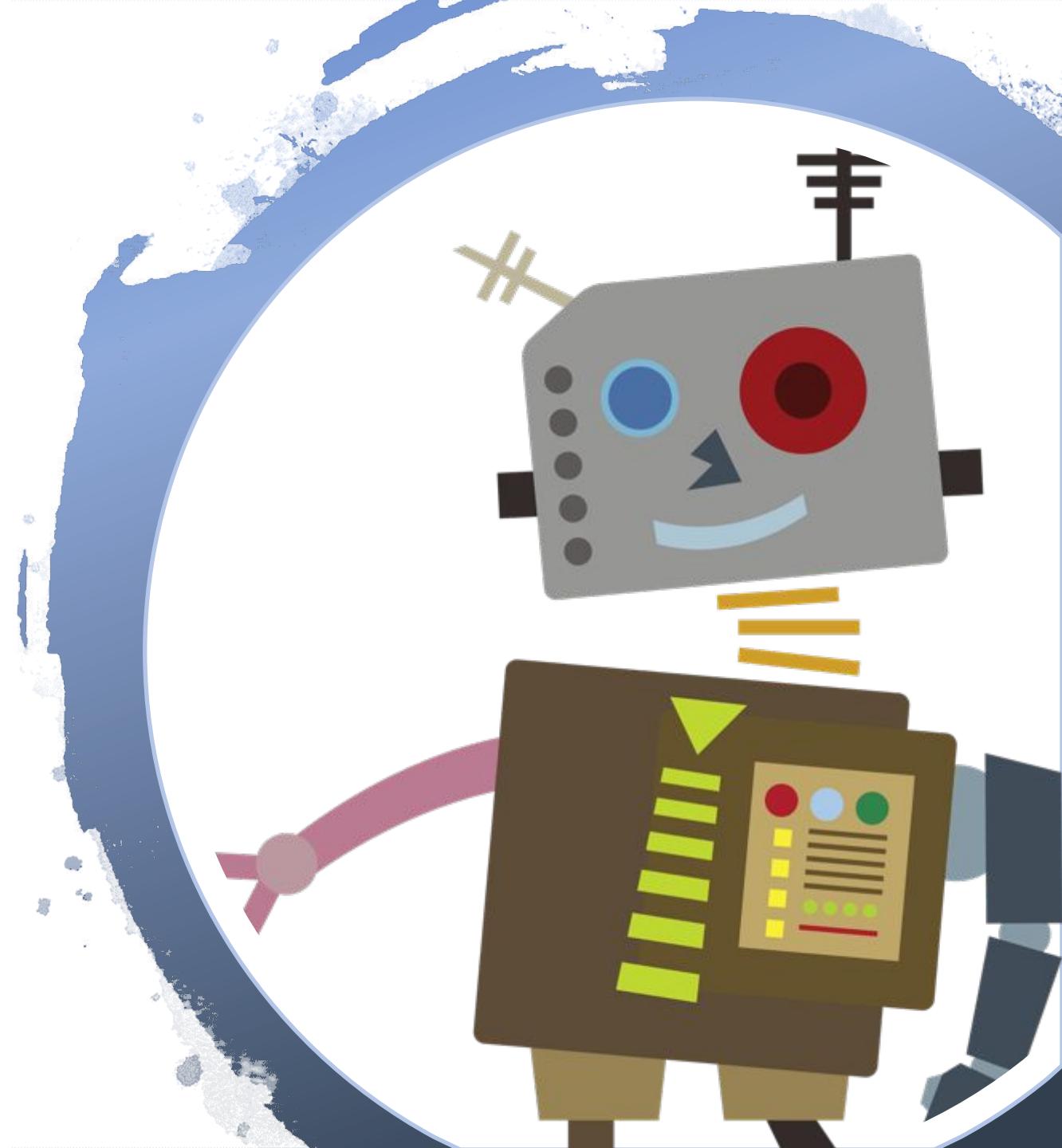
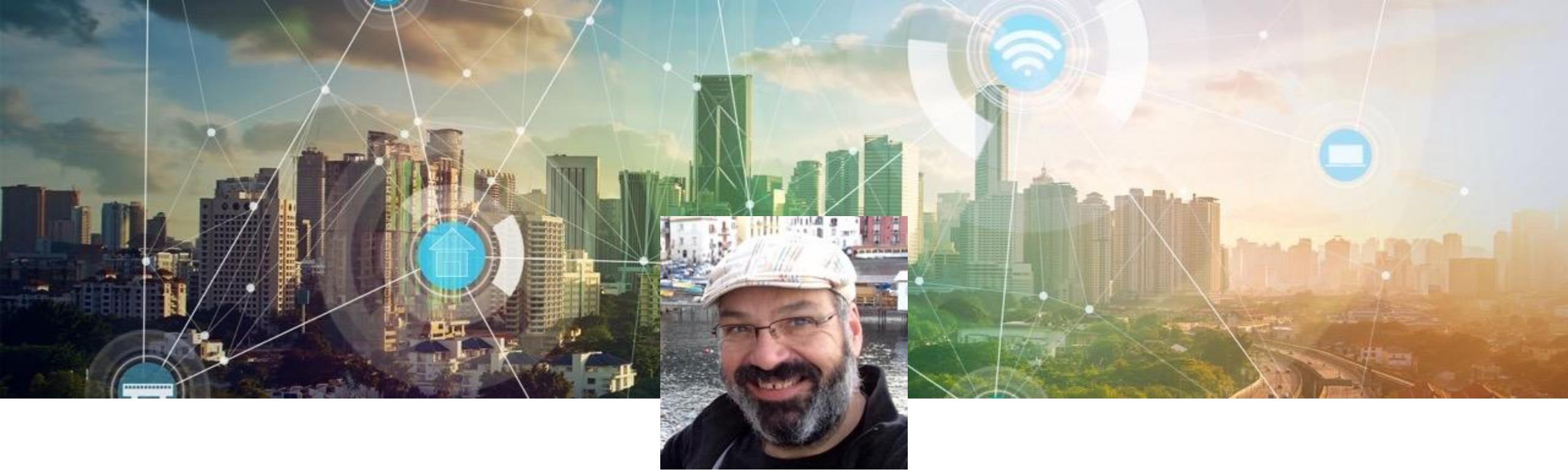


# Data Science Applied to Business

By Marcelo Rovai

May 20<sup>th</sup>, 2020





## Marcelo Rovai

Brazilian from São Paulo, Master's degree in Data Science by UDD, Chile, and MBA by IBMEC, Brazil. Graduated in 1982 as an Engineer from UNIFEI with pos graduation by Poli/USP, both in Brazil. Marcelo worked as a teacher, engineer, and executive in several companies in the technology area such as AVIBRAS Aeroespacial, SID Informática, ATT-GIS, NCR, DELL, COMPAQ (HP) and more recently at IGT where he continues as Senior Advisor.

In 2016, Marcelo began writing about electronics, publishing his works in sites of the area as MJRoBot.org (Editor/Writer), Hackster.io (#1 Contributor), Instructables.com, and Medium.com (TDS – Towards Data Science). Besides winning several Instructables competitions in the areas of electronics, robotics, and IoT.

Marcelo lives with his wife Ilza in Santiago, Chile, where he divides his time between his consultant work and sharing ideas in the field of Data Science, Electronics, IoT, Physical Computing and Robotics.

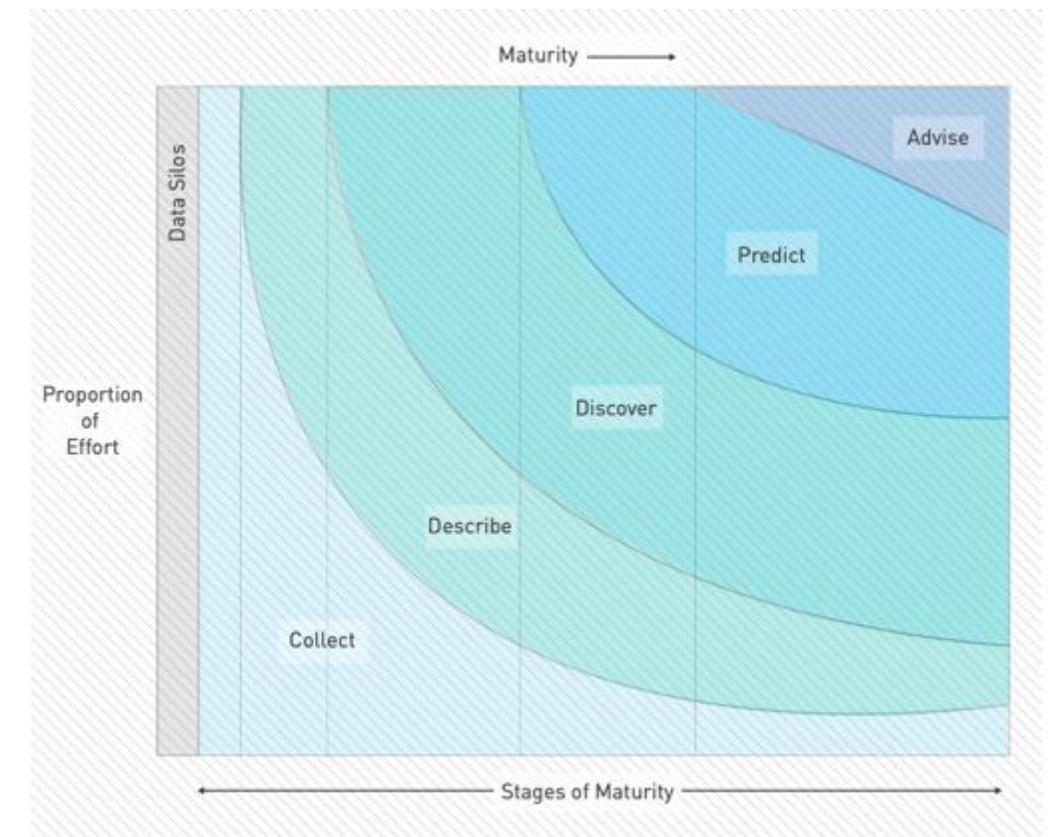
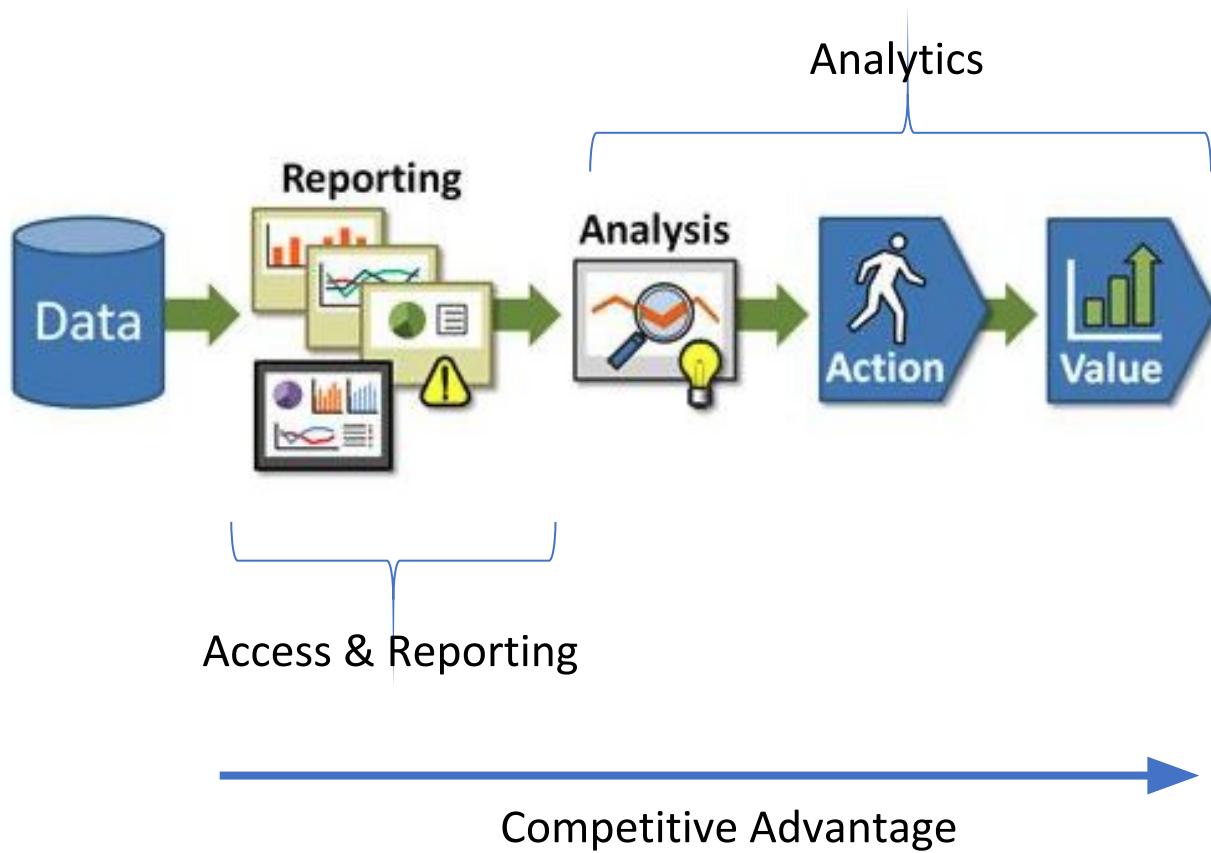
# What is Data Science?

“Data science is a broad field of study aimed at maintaining data sets and deriving meaning out of them.

The unique combination of statistics, mathematics, and business domain knowledge offered by Data Science has the potential to help organizations find ways to reduce costs, get into new markets, tap on different demographics, gauge the effectiveness of marketing campaigns or launch a new products and services.”

McCombs School of Business at The University of Texas at Austin.

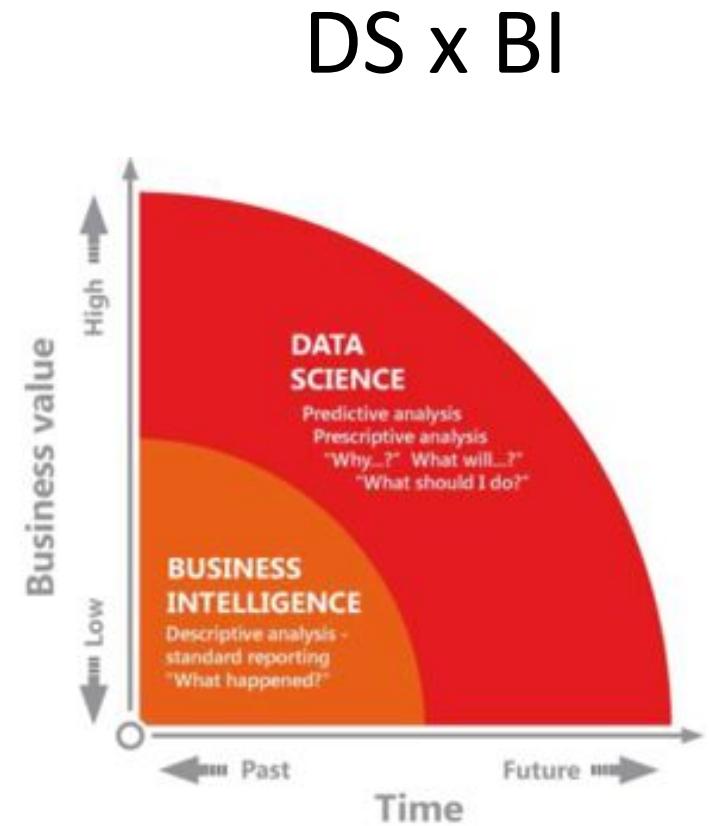
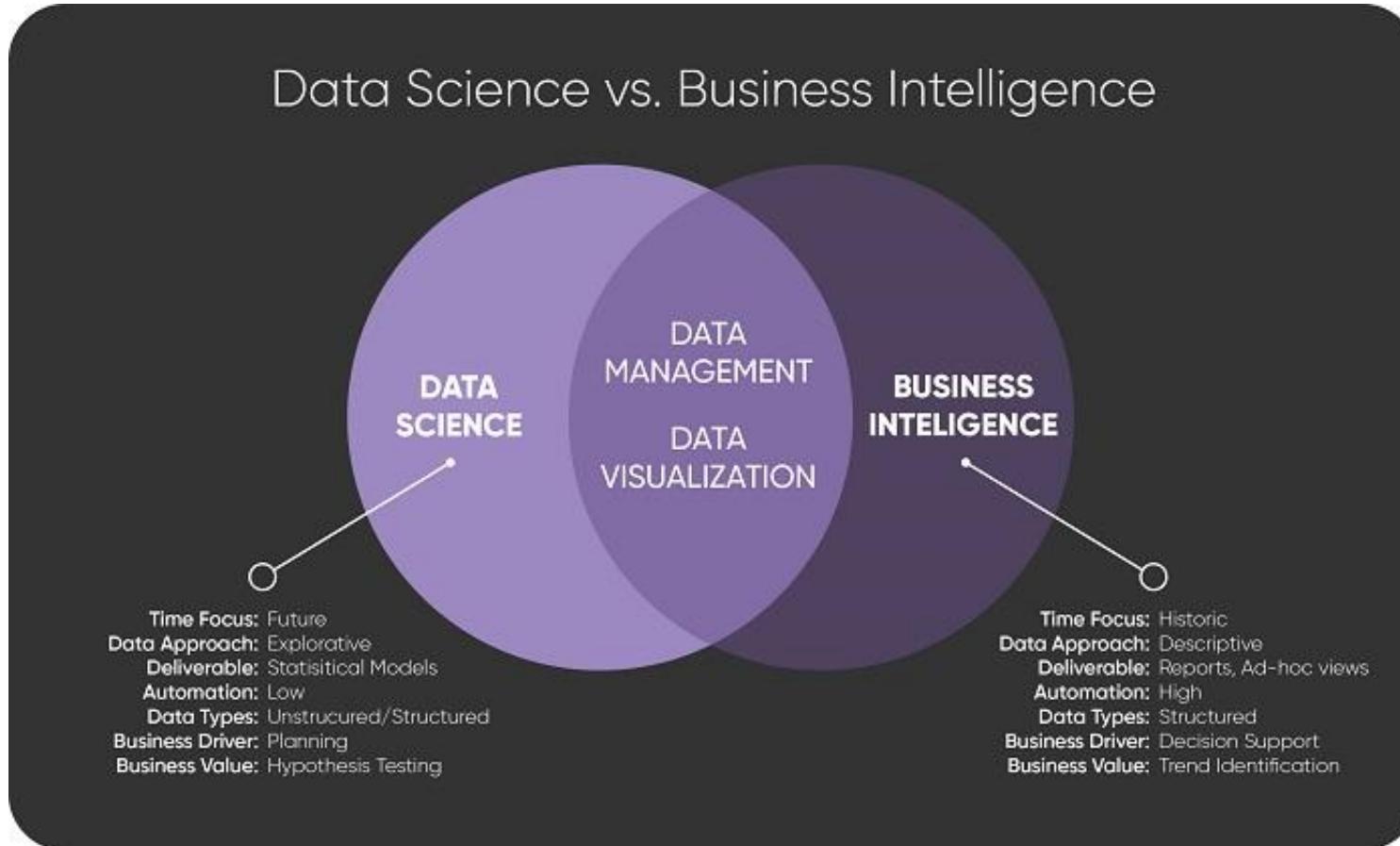
# What to do with Data?



Source: Booz Allen Hamilton  
www.BoozAllen.com

Adapted from Creating Data Driven Organization, Carl Anderson

# Data Science versus Business Intelligence

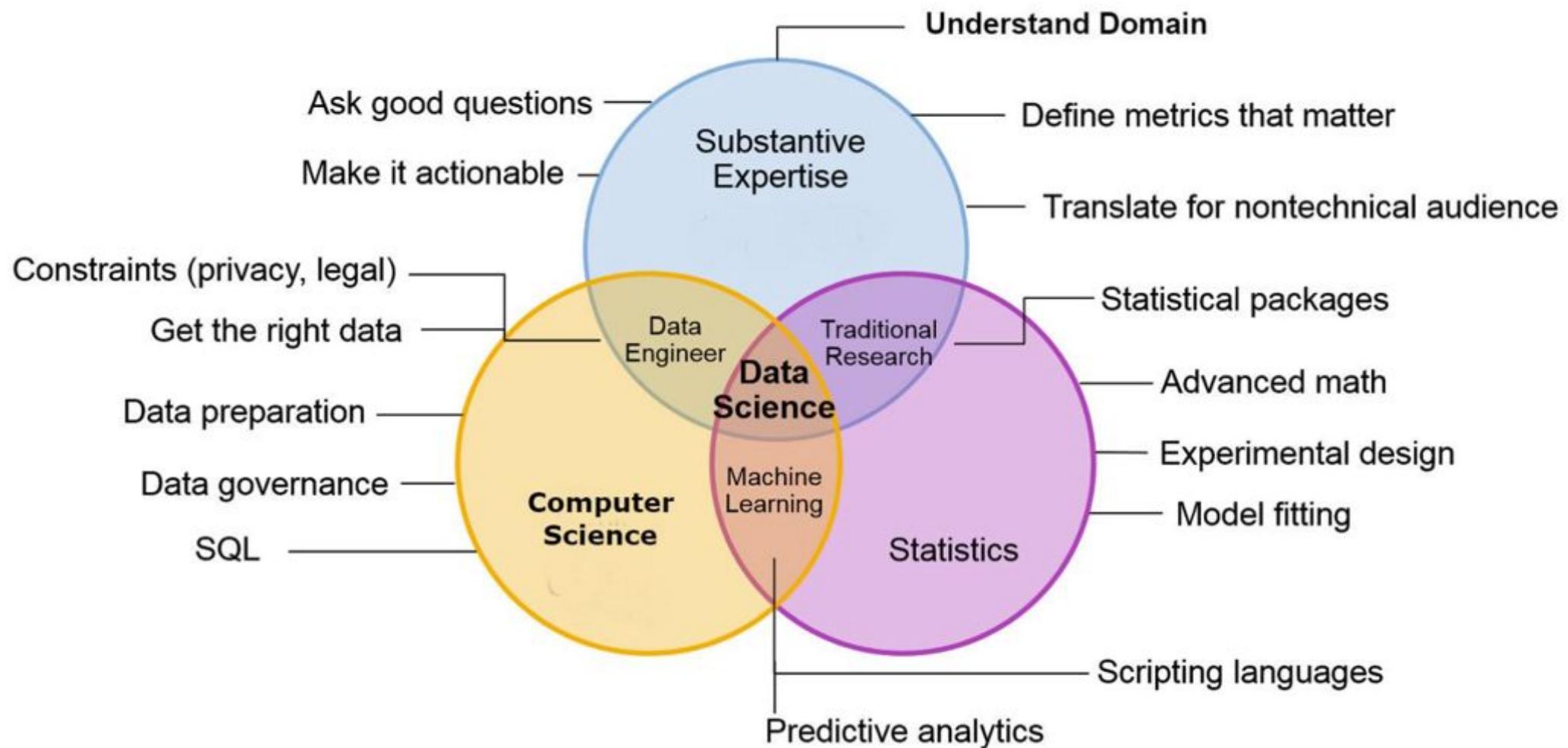


# Difference Between Data Science and Business Analytics

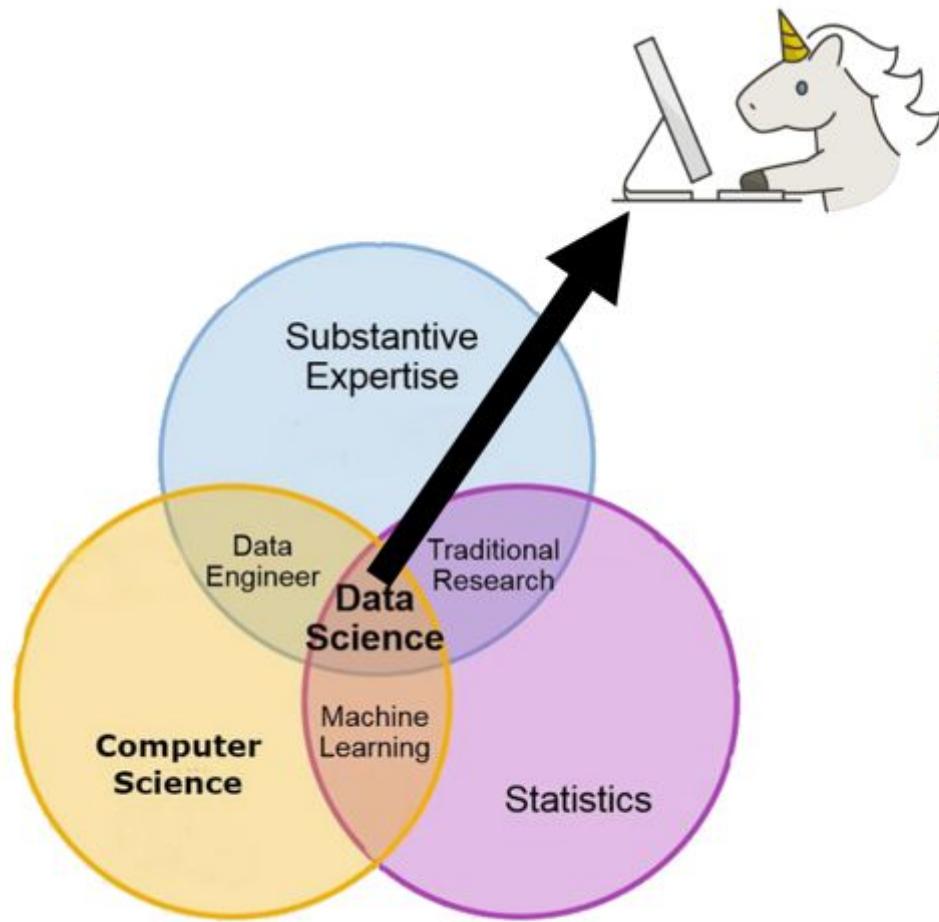
Data Science	Business Analytics
It is the science of Data study using statistics, algorithms and technology	It is the statistical study of business data
Uses both structured and unstructured data	Uses mostly structured data
It is a combination of traditional analytics practices with sound computer science knowledge including coding	It is oriented more towards statistics and does not involve much coding
Top industries where Data Science finds its applications are: a. Technology b. Finance c. E-commerce d. Academic	Top Industries where Business analytics finds its applications are a. Finance b. Technology c. Marketing d. Retail
The future applications of Data Science would be witnessed in Artificial Intelligence and Machine Learning	The future applications of Business Analytics would be witnessed in Cognitive Analytics and Tax analytics
Data Science results give insights but usually not used for making Business Decisions	Business Analytics results are vital to the key decision makers
Studies trends and patterns	Works on specific business problems



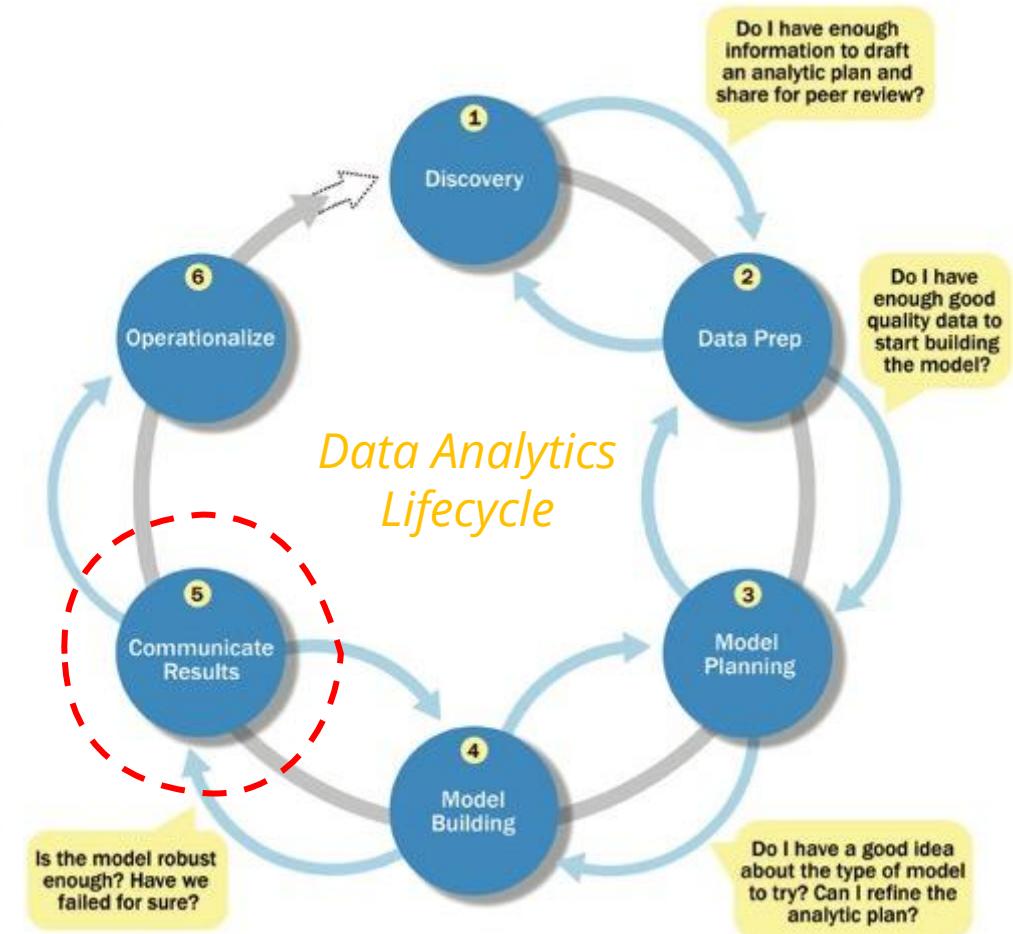
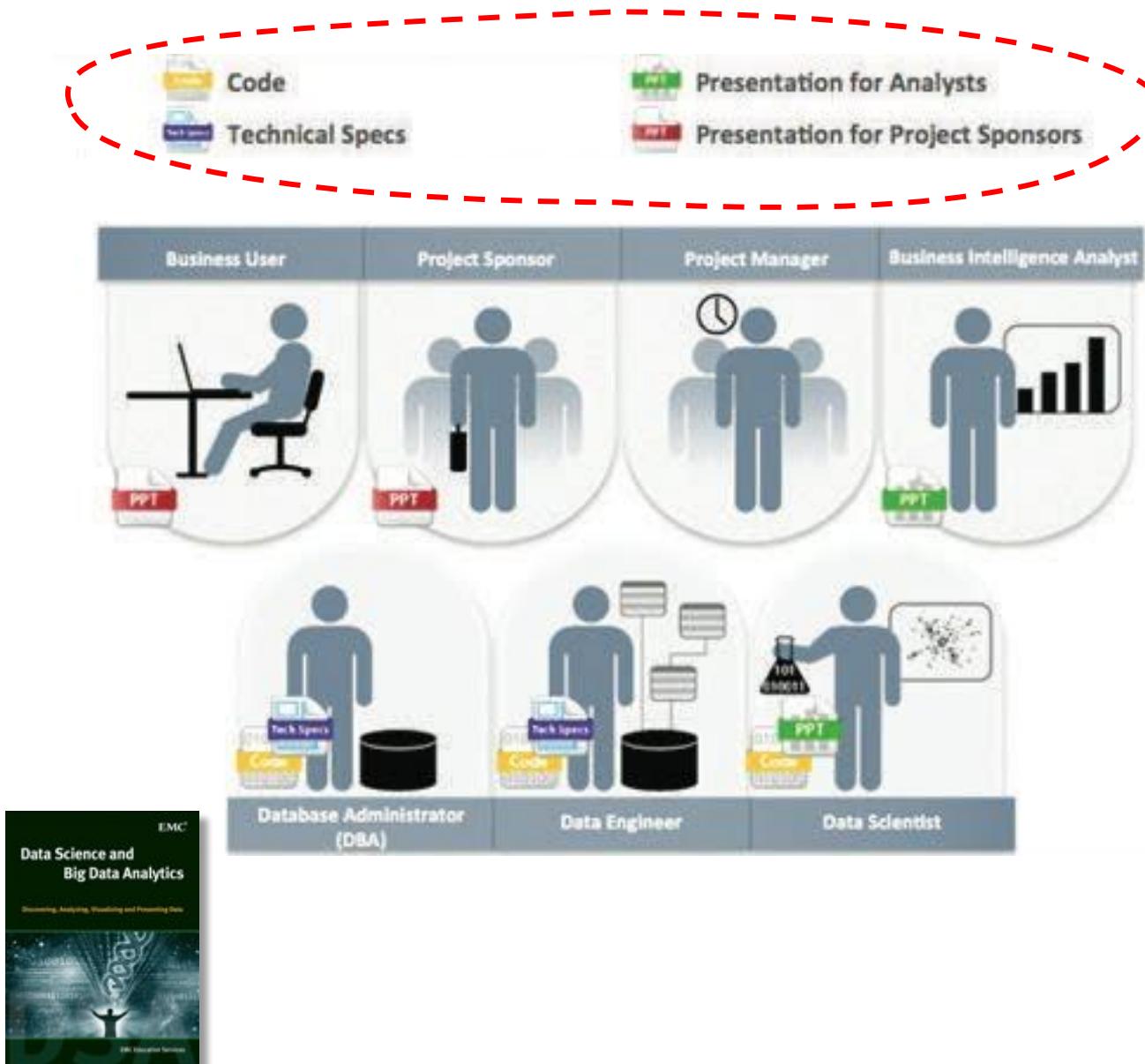
# What is a Data Scientist?



# Data Scientist is not a person, is a team!



# The fourth leg: Visualization!



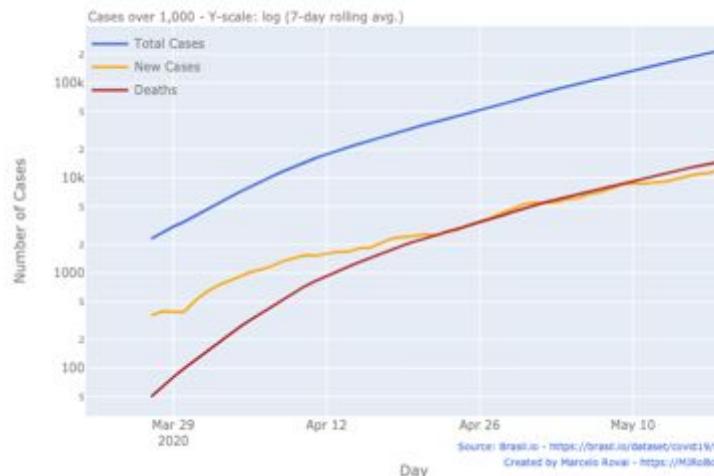
# Data Science Project Examples

# Mapping in the Time of Pandemic

---

Tracking the pandemic outbreak  
in Brazil

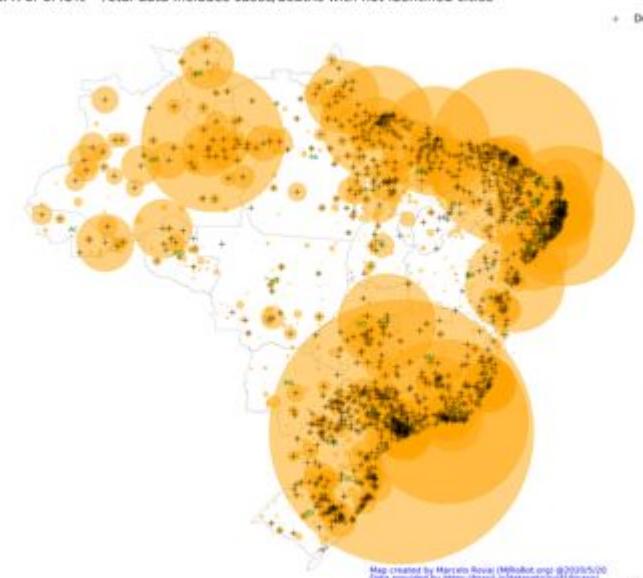




Brazil: Covid19 total cases at 2020/5/20: 288,308 (18,692 fatal in red) in 3,486 identified cities  
CFR: 6.48% - Total data includes cases/deaths with not identified cities



Brazil: Covid19 total cases at 2020/5/20: 288,308 (18,692 fatal) in 3,486 identified cities  
CFR of 6.48% - Total data includes cases/deaths with not identified cities



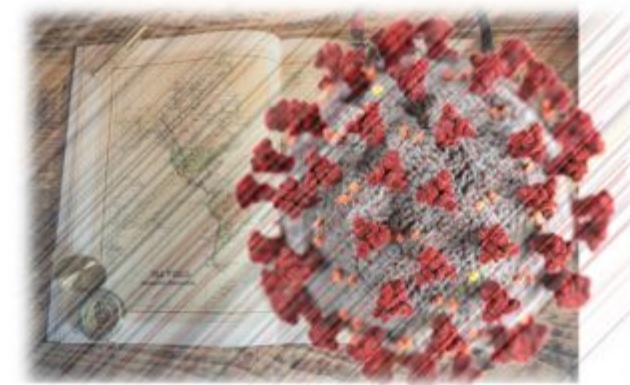
[https://mjrovai.github.io/Mapping\\_Covid-19\\_Brazil/](https://mjrovai.github.io/Mapping_Covid-19_Brazil/)

# Mapping\_Covid-19\_Brazil

[View on GitHub](#)

## Mapping in the Time of Pandemic

Tracking the pandemic outbreak in Brazil



The primary purpose of this repository is to help people to develop their graphs, maps, and videos using Python, GeoPandas, and Plotly. I will try to keep it updated in terms of Covid-19 Pandemic information in Brazil, but I strongly suggest you visit the below sites for more info.

### More information:

- <https://wcota.me/covid19br>
- <https://brasilio/covid19/>

### Content of subfolders:

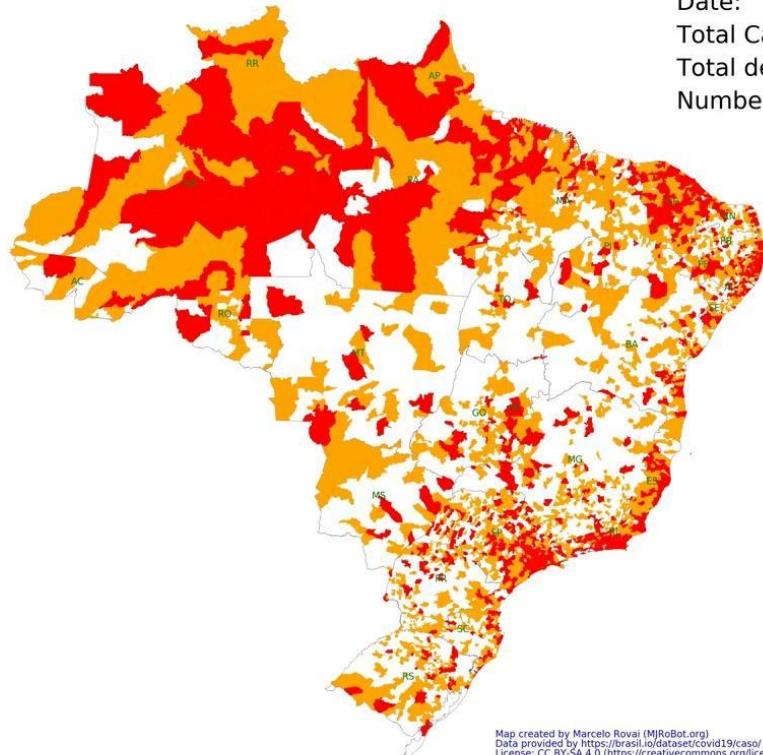
- /videos - Short videos showing the development of Covid-19 in Brazil and main States.
- /graphs - Timeline graphs of main brazilian cities
- /images - Most updated maps
- /notebooks - Code used to get Graphs, Maps, Gifs, and Videos

### Datasets:

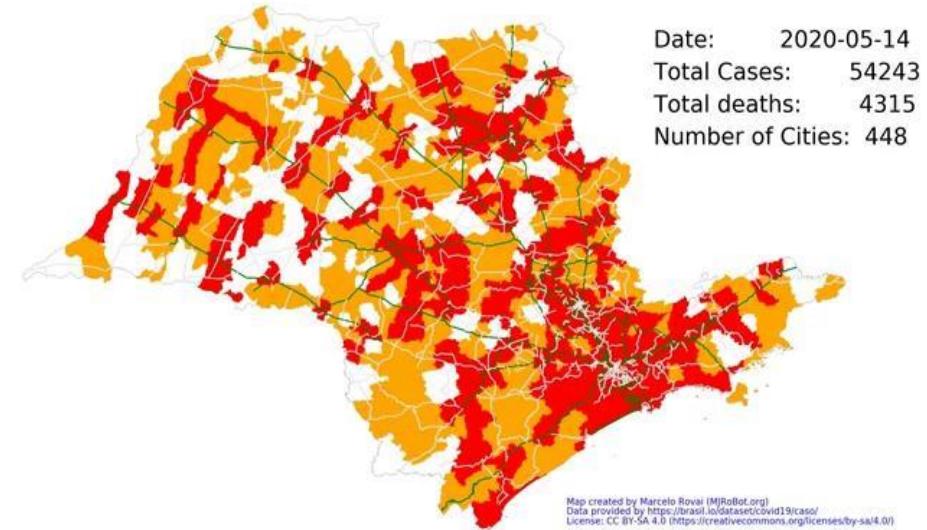
- Worldometers Daily Data
- Confirmed cases by day, using information from the news
- Raw data by city compiled from original dataset provided by Brasil.IO

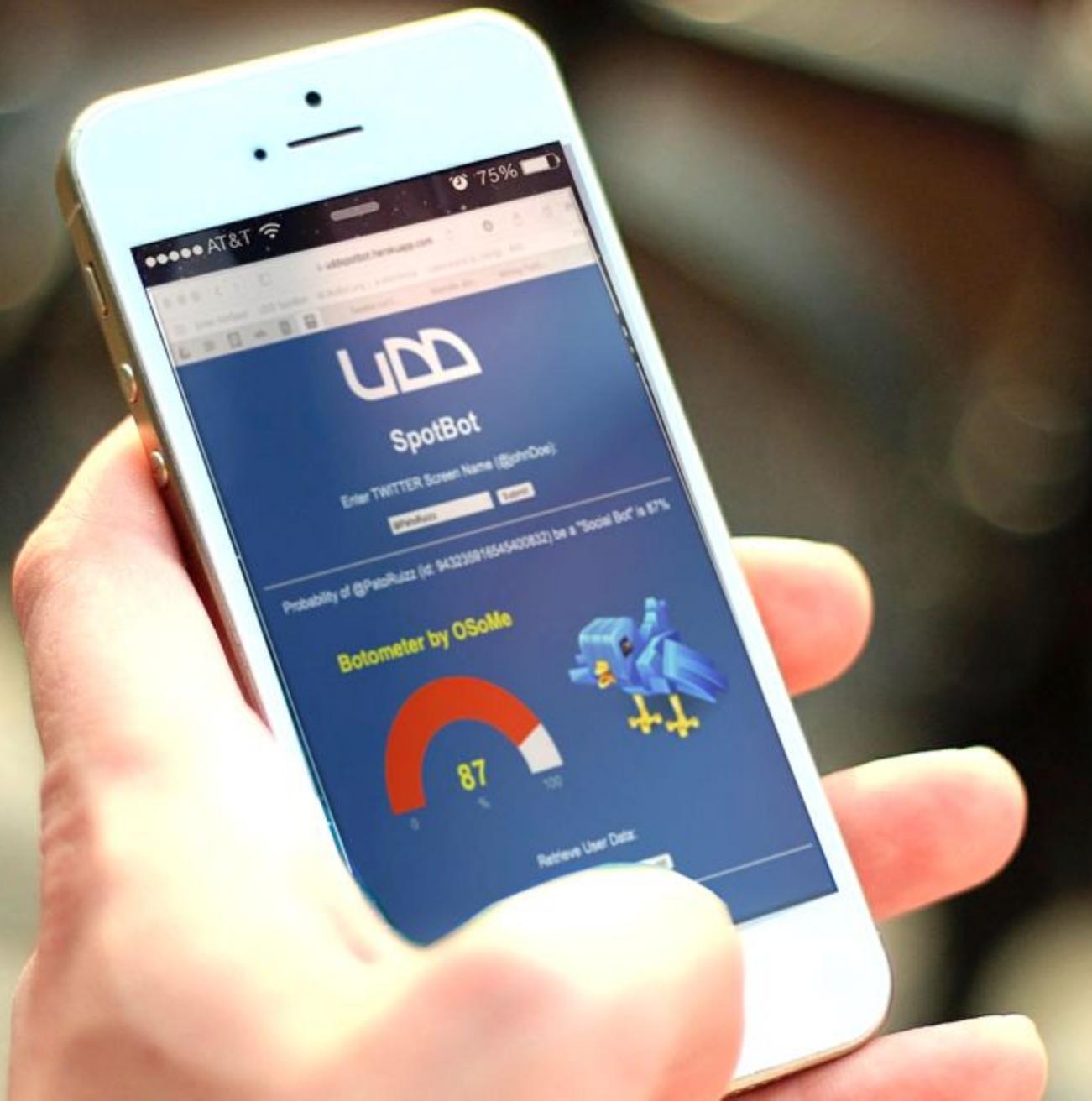
# Outbreak Spreading – Timeline

Brazilian cities reported with Covid19 cases (orange) and deaths (red)



SP state cities reported with Covid19 cases (orange) and deaths (red)



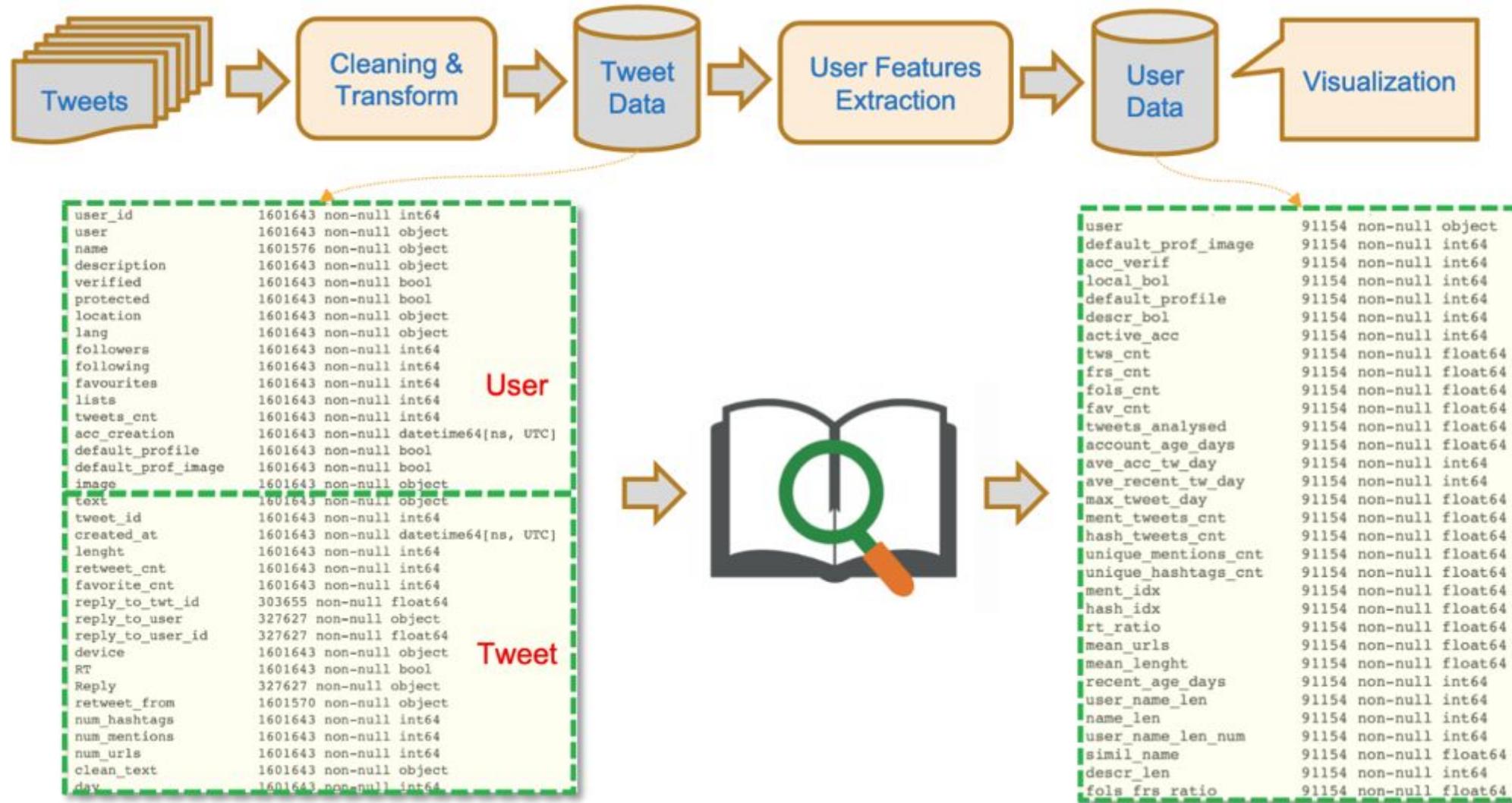


# SpotBot

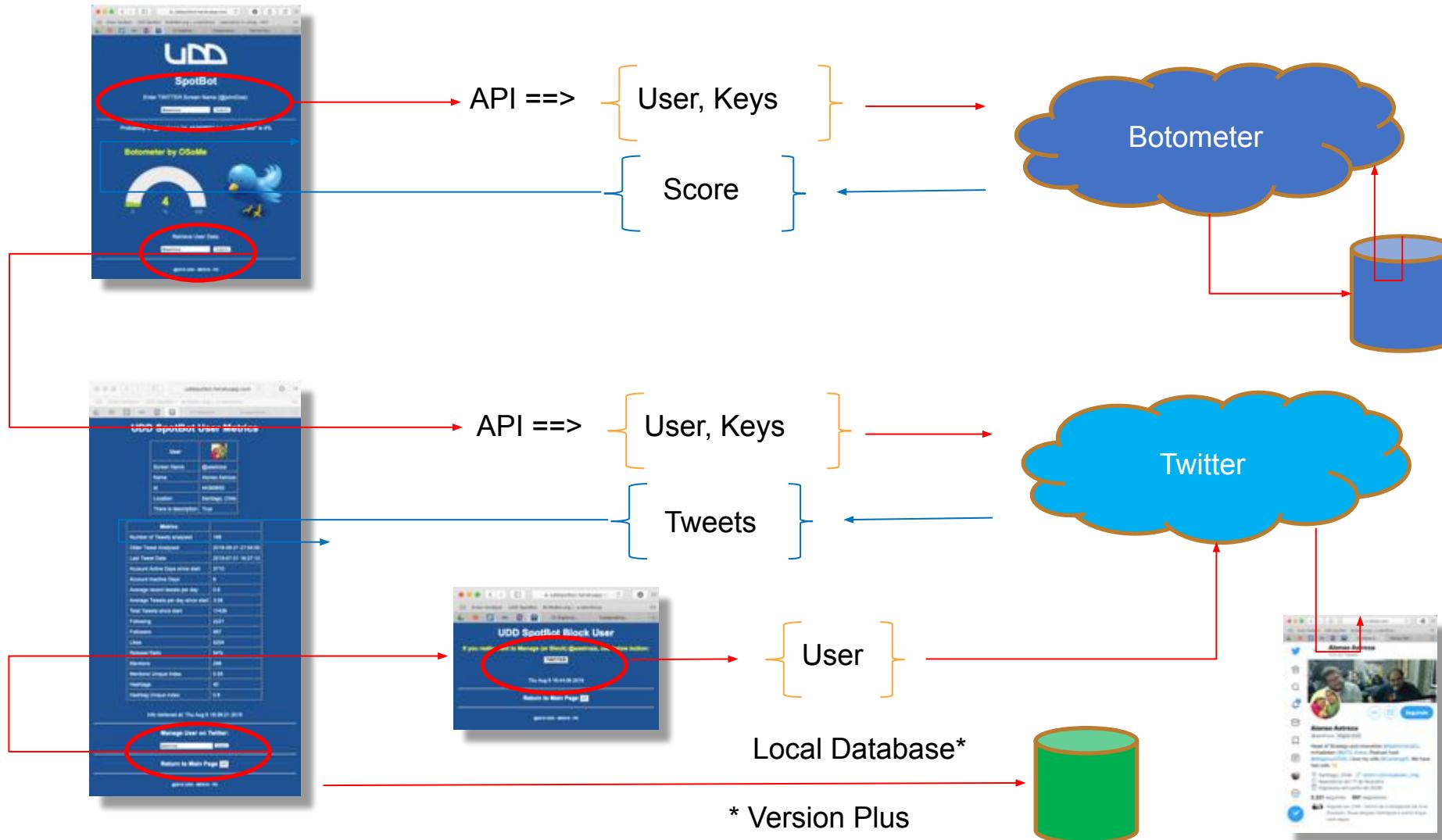
Detecting Social Bots  
in Twitter



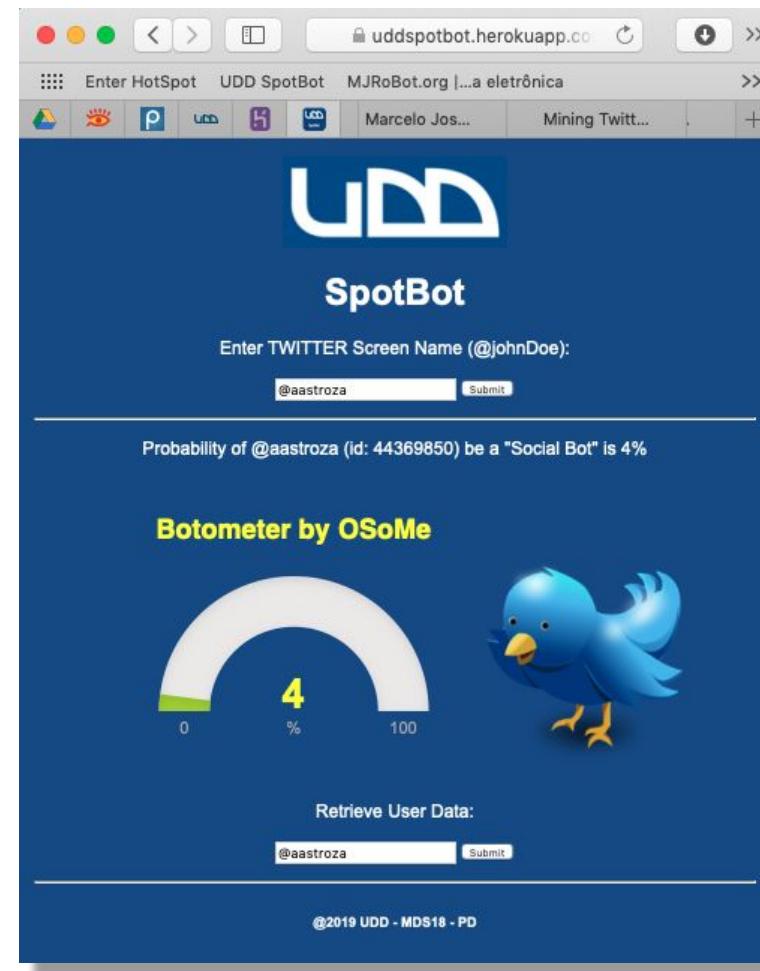
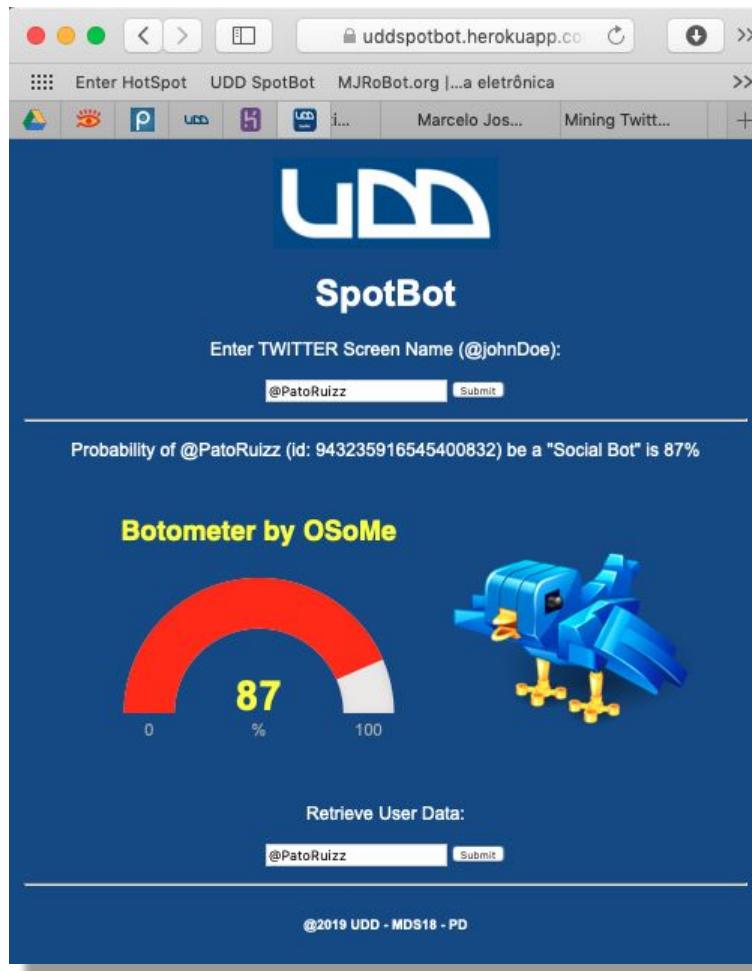
# Extracting Twitter User Features



# How it works?



# Screen #1



# Screen #2

The screenshot shows a web application titled "UDD SpotBot User Metrics". The interface includes:

- A header bar with tabs: Enter HotSpot, UDD SpotBot, MJRoBot.org |... a eletrônica, and a search bar.
- A top navigation bar with icons for file operations, a user profile, and other functions.
- A main section titled "UDD SpotBot User Metrics" containing two tables:
  - The first table, "User", displays the following information:

User	
Screen Name	@aastroza
Name	Alonso Astroza
Id	44369850
Location	Santiago, Chile
There is description	True
  - The second table, "Metrics", displays various statistics:

Metrics	
Number of Tweets analysed	199
Older Tweet Analysed	2018-08-31 21:54:00
Last Tweet Date	2019-07-31 16:27:12
Account Active Days since start	3710
Account Inactive Days	8
Average recent tweets per day	0.6
Average Tweets per day since start	3.08
Total Tweets since start	11438
Following	2221
Followers	997
Likes	5254
Retweet Ratio	54%
Mentions	298
Mentions Unique Index	0.59
Hashtags	40
Hashtag Unique Index	0.8
- An info message at the bottom: "Info retrieved at: Thu Aug 8 18:51:07 2019".
- Action buttons:
  - "Manage User on Twitter" with a form field containing "aastroza" and a "Submit" button.
  - "Return to Main Page" with a back arrow icon.
- Page footer: "@2019 UDD - MDS18 - PD".

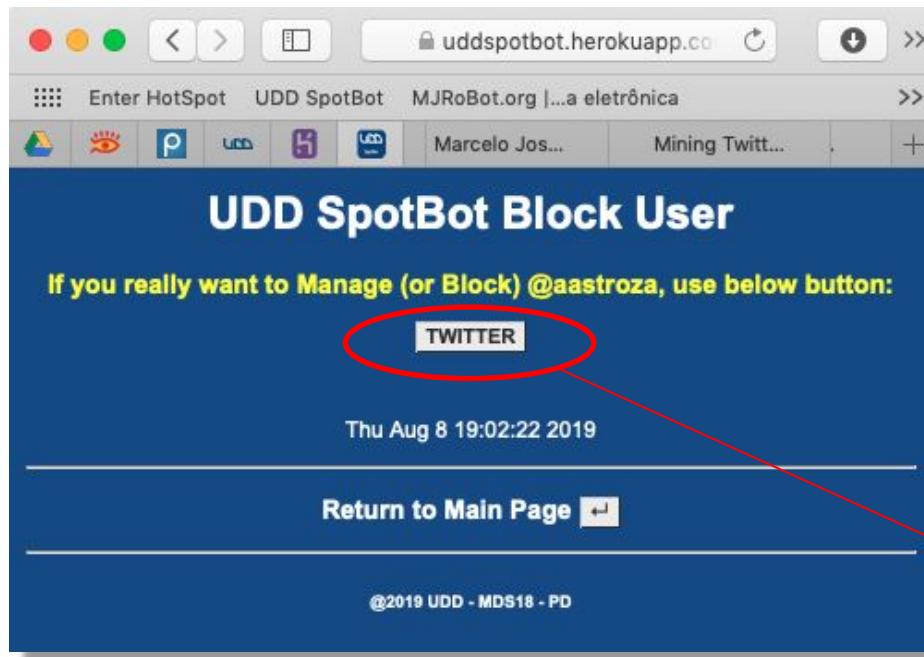
Twitter User Data

Features generated from real tweets (last 200)

Possible Actions from User

- Manage / Block user on Twitter
- Return to Main Screen #1

# Screen #3



uddspotbot.herokuapp.com

Enter HotSpot UDD SpotBot MJRoBot.org |...a eletrônica

Marcelo Jos... Mining Twitt...

## UDD SpotBot Block User

If you really want to Manage (or Block) @aastroza, use below button:

**TWITTER**

Thu Aug 8 19:02:22 2019

[Return to Main Page](#)

©2019 UDD - MDS18 - PD



twitter.com

Enter HotSpot UDD SpotBot MJRoBot.org |...a eletrônica

(1) Alonso A...

### Alonso Astroza

11,4 mil Tweets



Segundo

- Desativar Retweets
- Adicionar/remover das listas
- Ver Listas
- Visualizar Moments**
- Compartilhar perfil por...
- Copiar link para o perfil
- Silenciar @aastroza
- Bloquear @aastroza
- Denunciar @aastroza

@GeoVictoriaCL.  
ast host  
@CsolangeS. We have

/mrhadoken\_mtg

le Investigacion del Arvo  
enriquez e outros 9 que

as Mídia C >

# Screen #3

UDD SpotBot Block User

If you really want to Manage (or Block) @patoruizz, use below button:

**TWITTER**

Thu Aug 8 19:27:19 2019

[Return to Main Page](#)

@2019 UDD - MDS18 - PD

Patricio Ruiz  
13 Tweets

Seguir

Adicionar/remover das listas

Ver Listas

Visualizar Moments

Compartilhar perfil por...

Copiar link para o perfil

Silenciar @PatoRuizz

Bloquear @PatoRuizz

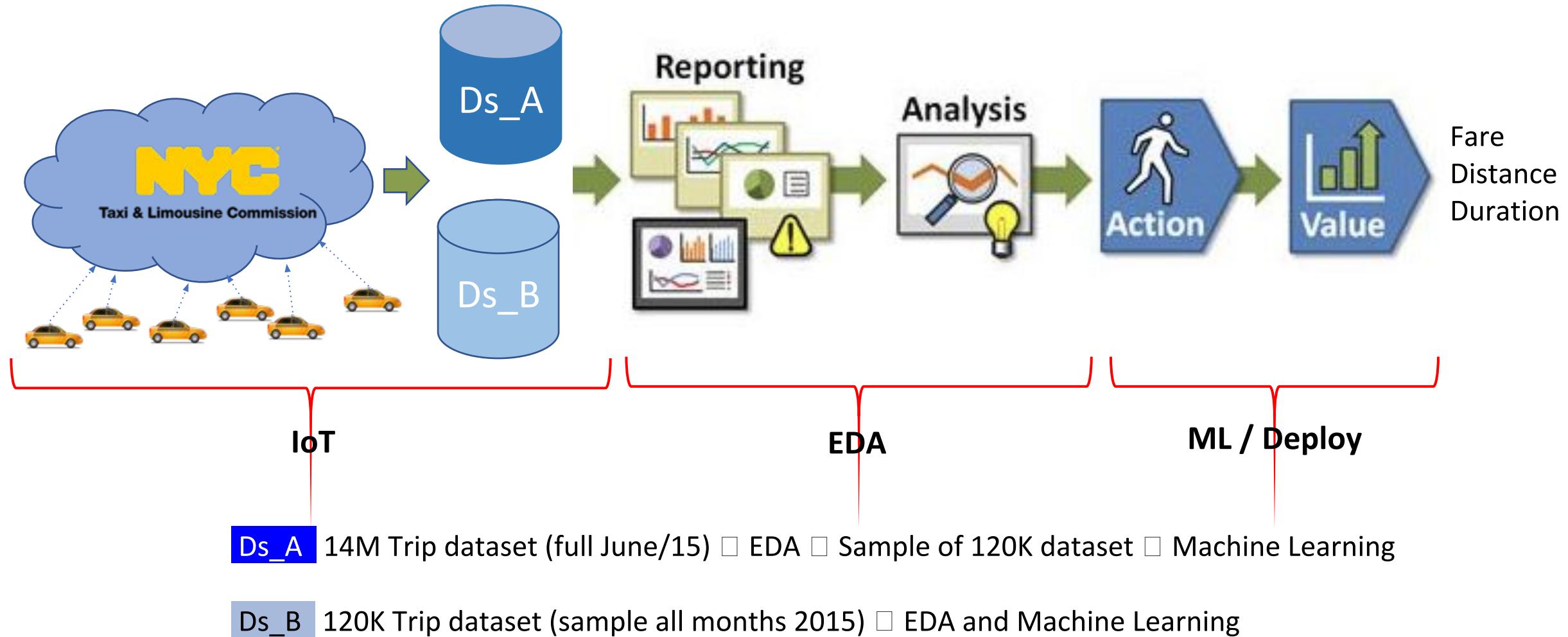
Denunciar @PatoRuizz

[https://github.com/Mjrovai/UDD\\_Master\\_Data\\_Science/  
tree/master/PD\\_SpotBot\\_Final\\_Project](https://github.com/Mjrovai/UDD_Master_Data_Science/tree/master/PD_SpotBot_Final_Project)

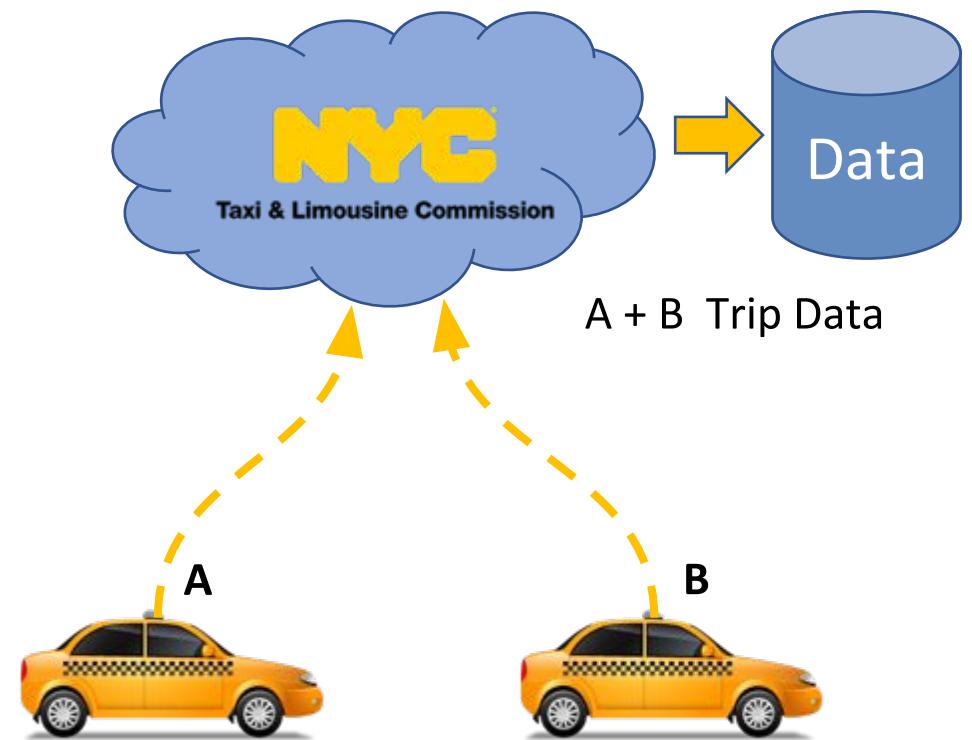
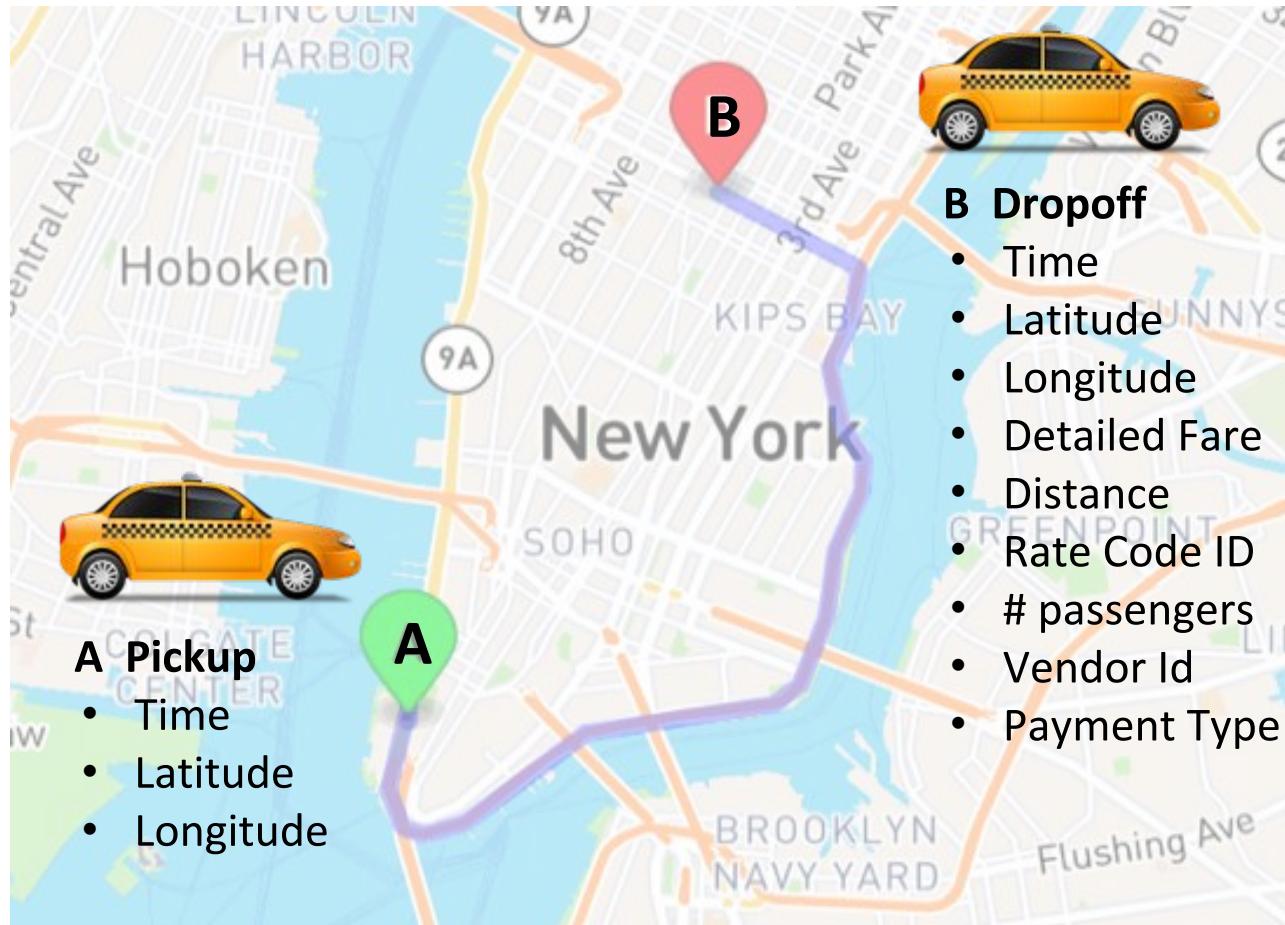


# NYC Taxi – Trip Prediction Project

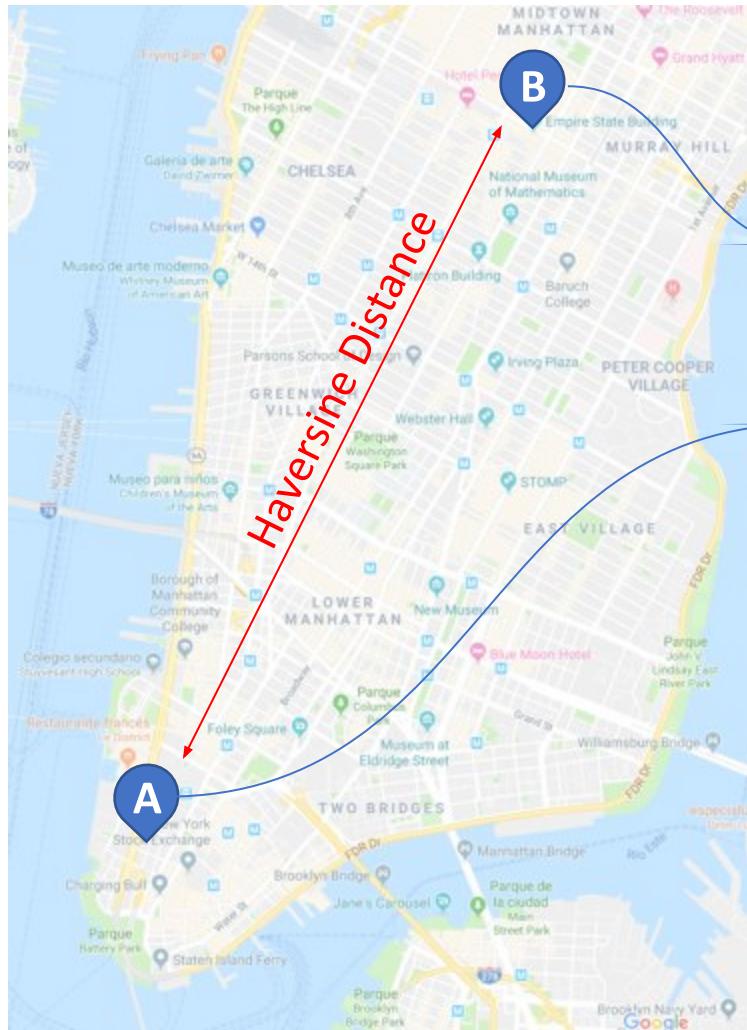
# NYC Taxi Trip Prediction Project



# NYC Taxi – IoT – Getting data



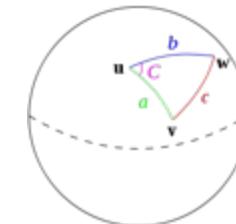
# NYC Taxi Trip Prediction – ML



Dropoff Latitude  
Dropoff Longitude  
Pickup Latitude  
Pickup Longitude  
Pickup Time

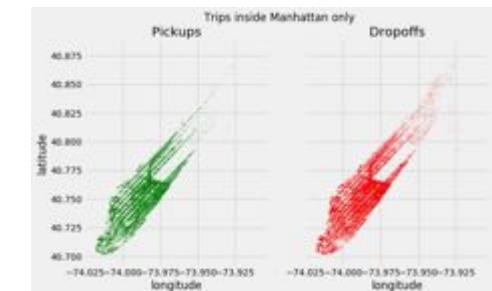


Fare  
Distance  
Duration



$$2r \arcsin \left( \sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)} \right)$$

Hav. Dist.  
Rate Code  
In/Out Manh.



# NYC Taxi Trip Prediction – ML

	Fare DS_A	Fare DS_B		Fare DS_A	Fare DS_B
	RMSE	RMSE	Comparison A-B	Base line comparison	Base line comparison
Base Line	10,25	9,88	3,6%	0,0%	0,0%
Decisión Tree Regresor	3,25	3,21	1,2%	68,3%	67,5%
Random Forest Regressor	3,01	2,92	3,0%	70,6%	70,4%
Random Forest Regressor (Reducing features)	3,00	2,92	2,7%	70,7%	70,4%
Random Forest Regressor (Changing Hyper_param)	2,77	2,72	1,8%	73,0%	72,5%

## ML Model: Random Forest Regressor

Fare: < \$3.00

	Time DS_A	Time DS_B		Time DS_A	Time DS_B
	RMSE	RMSE	Comparison A-B	Base line comparison	Base line comparison
Base Line	14,52	13,94	4,0%	0,0%	0,0%
Decisión Tree Regresor	11,47	10,75	6,3%	21,0%	22,9%
Random Forest Regressor	6,03	6,26	-3,8%	58,5%	55,1%
Random Forest Regressor (Reducing features)	5,60	5,73	-2,3%	61,4%	58,9%
Random Forest Regressor (Changing Hyper_param)	5,26	5,38	-2,3%	63,8%	61,4%

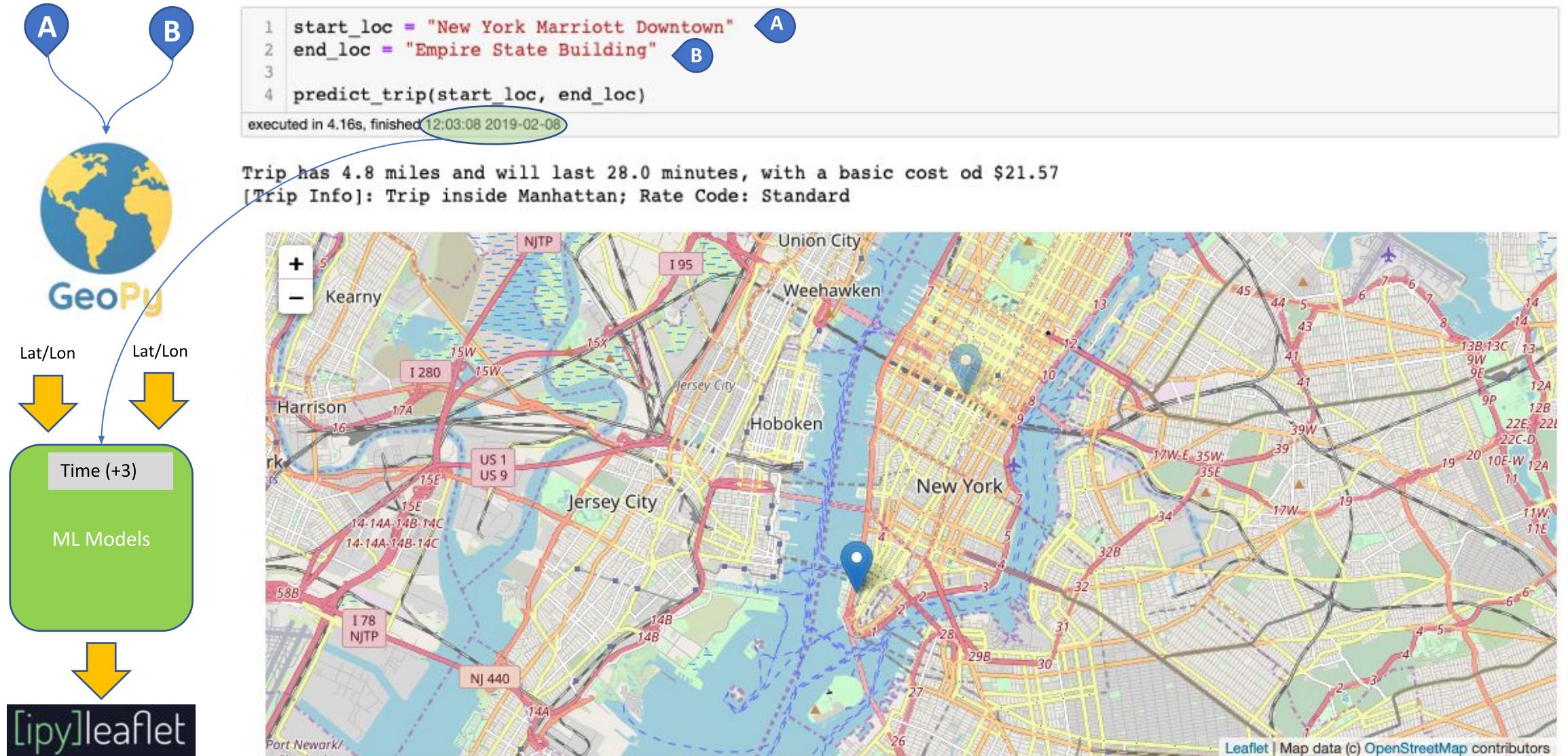
Errors

Time: > 5 min

	Dist DS_A	Dist DS_B		Dist DS_A	Dist DS_B
	RMSE	RMSE	Comparison A-B	Base line comparison	Base line comparison
Base Line	3,61	3,46	4,2%	0,0%	0,0%
Decisión Tree Regresor	0,90	0,82	8,9%	75,1%	76,3%
Random Forest Regressor	0,83	0,75	9,6%	77,0%	78,3%
Random Forest Regressor (Changing Hyper_param)	0,78	0,71	9,0%	78,4%	79,5%

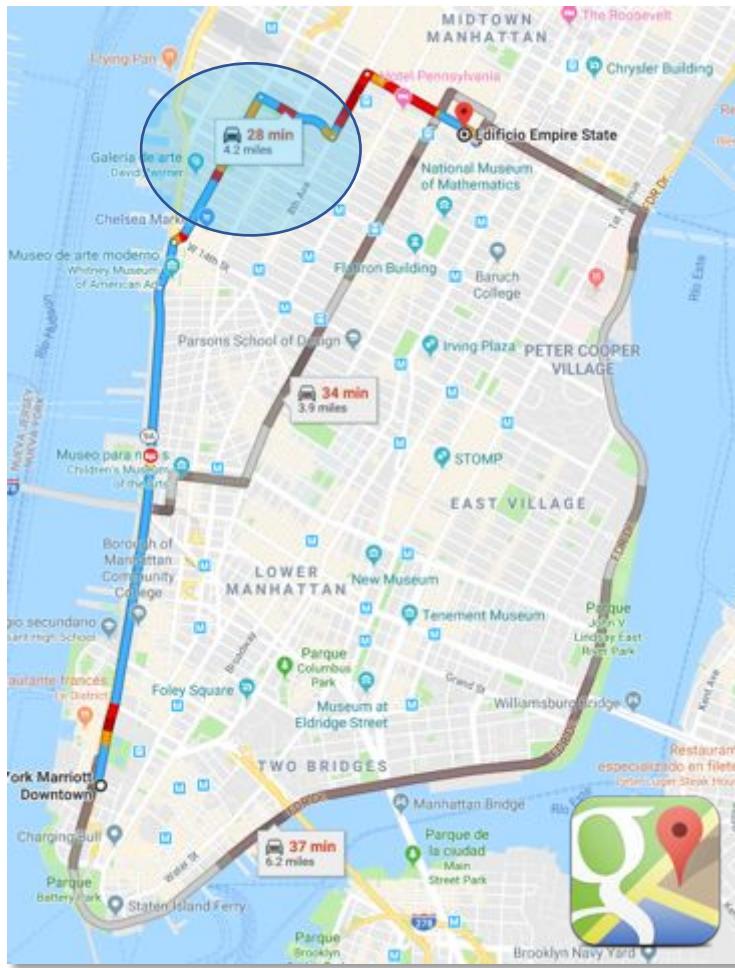
Dist: > 0.7 mile

# NYC Taxi Trip Prediction - Deploy



# NYC Taxi Trip Prediction - Benchmark

Trip has 4.8 miles and will last 28.0 minutes, with a basic cost od \$21.57  
[Trip Info]: Trip inside Manhattan; Rate Code: Standard



All last **28.0 minutes**, with a basic cost od **\$21.57**

**Manhattan; Rate Code: Standard**



**TAXI  
FARE FINDER**

[Sign up for our newsletter!](#)   [Like 103K](#) [New York, NY](#)  [English](#)

### 1 | Taxis To and From Airport

Cheap and reliable Transfer to/from the Airport.  
Book your Taxi Today! [Taxi2Airport.com](#)

### 2 | Private Airport Transfer

Best price. 24/7. Everywhere. Discounts & special offers. Book online! [gettransfer.com](#)

[NEW YORK, NY](#)
[TAXI FARE CALCULATOR](#)
[FARE COMPARISON](#)
[RATE CHART](#)
[NEWSROOM](#)
[LOCAL TAXI INFO](#)

Please take our quick (2 question) survey on tipping your ride-share drivers and automatically be entered to win one of 5 RideGuru Swag Bags!

**Estimated Fare**

**\$24.51**

For New York, NY rates  
\$29.41 incl. 20% tip

[See Uber, Lyft, and more...](#)

**Other possible fares**

8:00PM to 6:00AM: **\$30.01**

**Cost with Traffic**

\$19.30	\$24.51	\$47.32
---------	---------	---------

**How much does a taxi cab cost from Marriott Downtown Grand Ballroom B, New York, New York, United States, 10006 to Empire State Building, 350 5th Ave, New York, New York, United States, 10001 in New York, NY?**



Suggested routes: 1 of 1 (\$24.51)  Fastest  Cheapest  Shortest

Marriott Downtown Grand Ballroom B, New York, New York  or select a point of interest. [Get Fare!](#)

Empire State Building, 350 5th Ave, New York, New York  or select a point of interest. [Get Fare!](#)

**Fare Information**

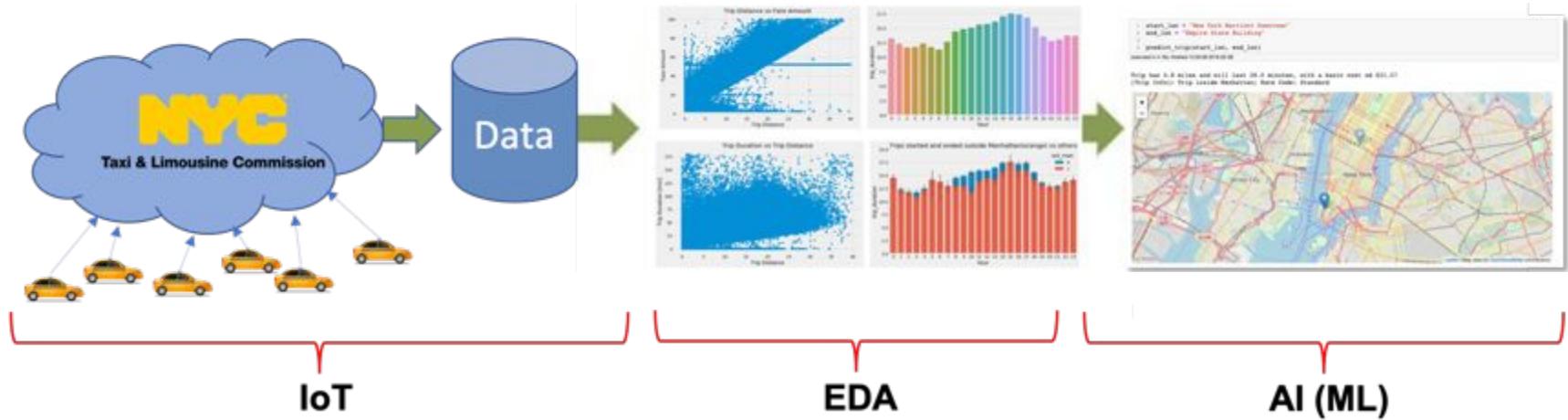
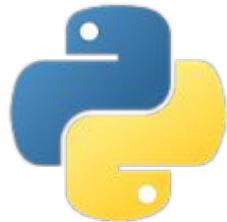
**Trip Information**

Trip is 6.1 mi, 17 mins

**Notes**

- Add \$1.00 Peak Hour surcharge (4:00PM to 8:00PM, M-F)
- \$52+ tolls flatrate JFK --> Manhattan. Add \$17.50 to Newark Airport (EWR).
- Trips beyond city limits may be under flat rate or have the portion of the trip beyond the city limit doubled (Rate 4).
- Click here for the NY City Taxi rates information.
- NYC Yellow Taxis can only be street-hailed. They cannot be pre-booked over the phone.
- Tolls & surcharges may apply

# Data Science Environment



Pandas



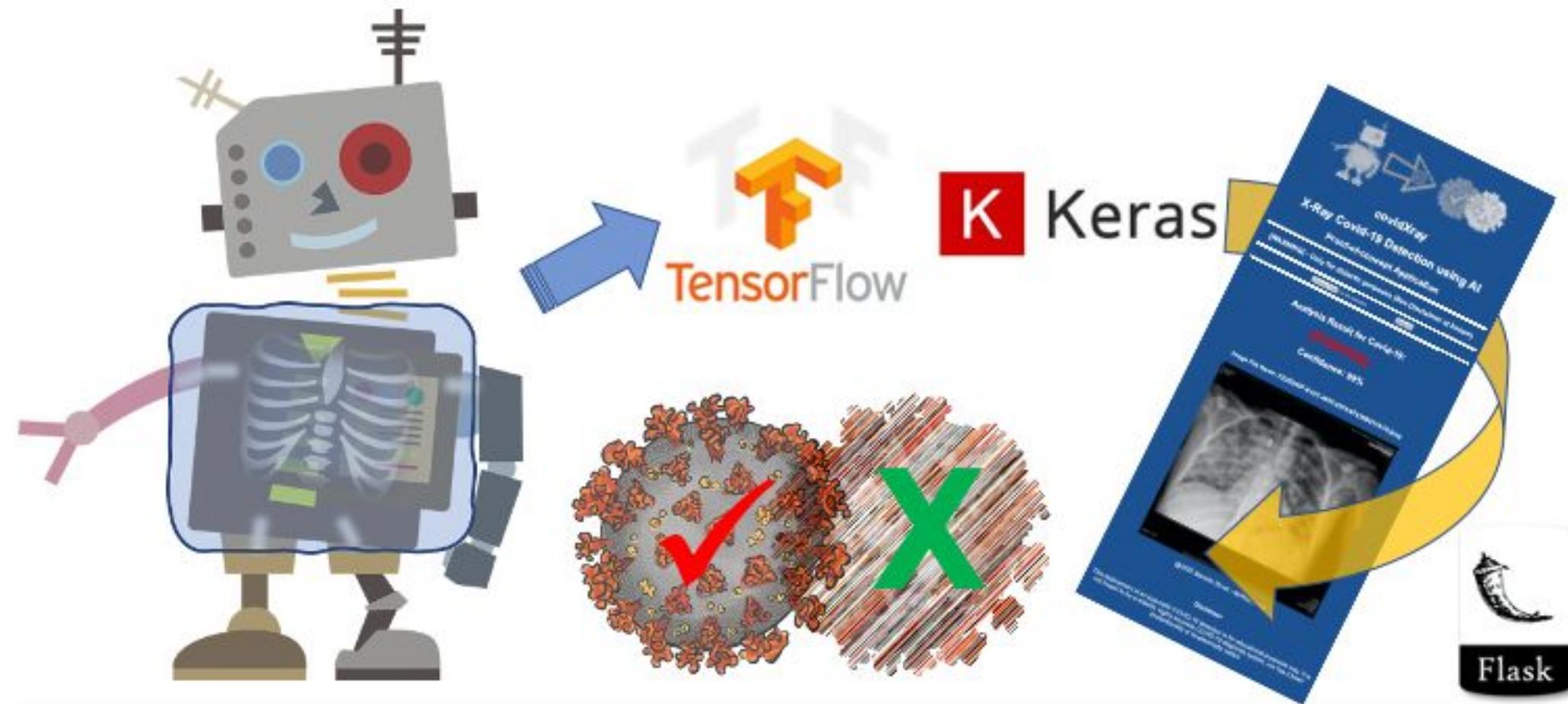
matplotlib



[ipy]leaflet

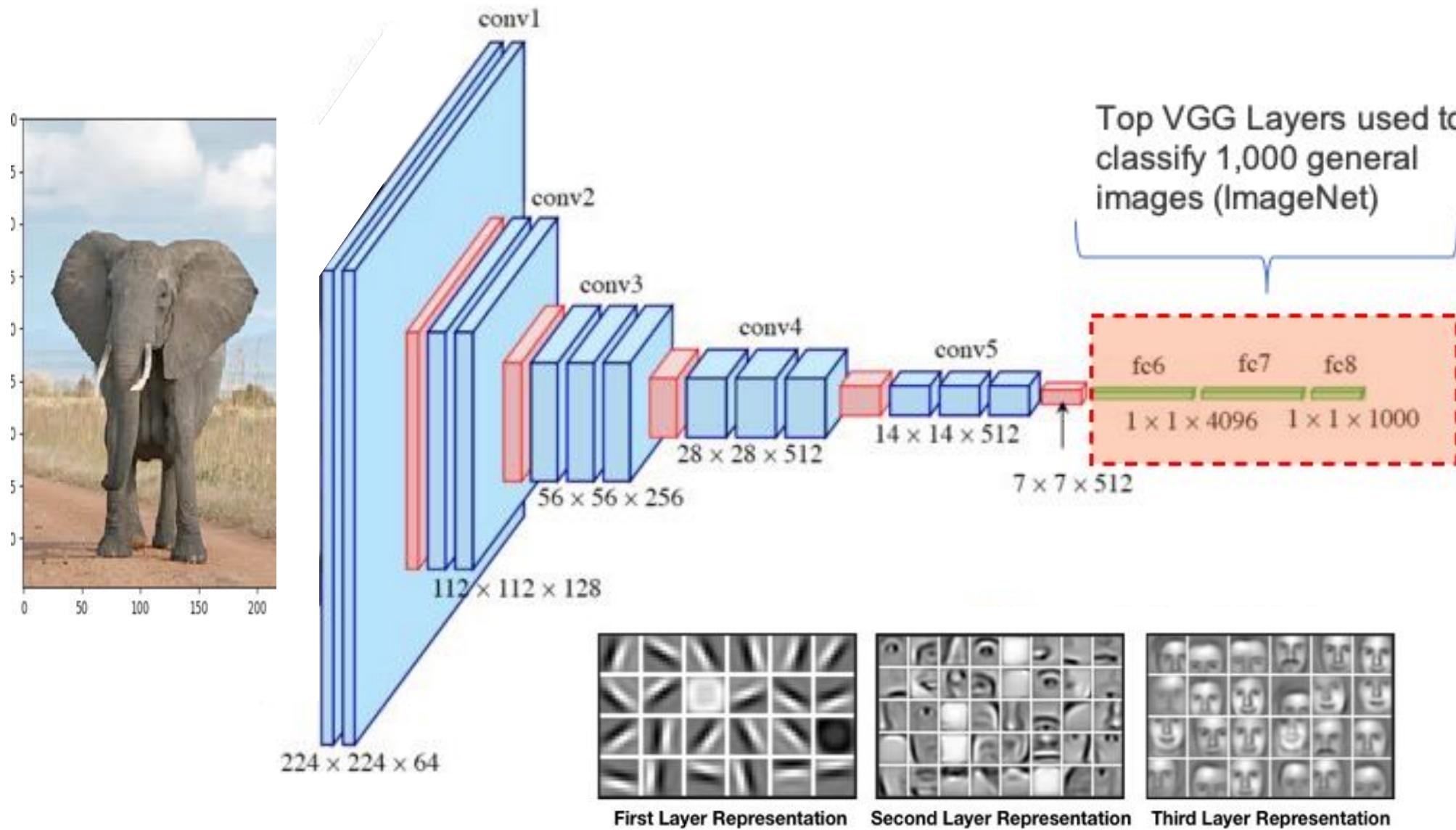


[https://github.com/Mirovai/UDD\\_Master\\_Data\\_Science/tree/master/AML-NYC\\_TAXI\\_TRIP\\_PREDICTION](https://github.com/Mirovai/UDD_Master_Data_Science/tree/master/AML-NYC_TAXI_TRIP_PREDICTION)

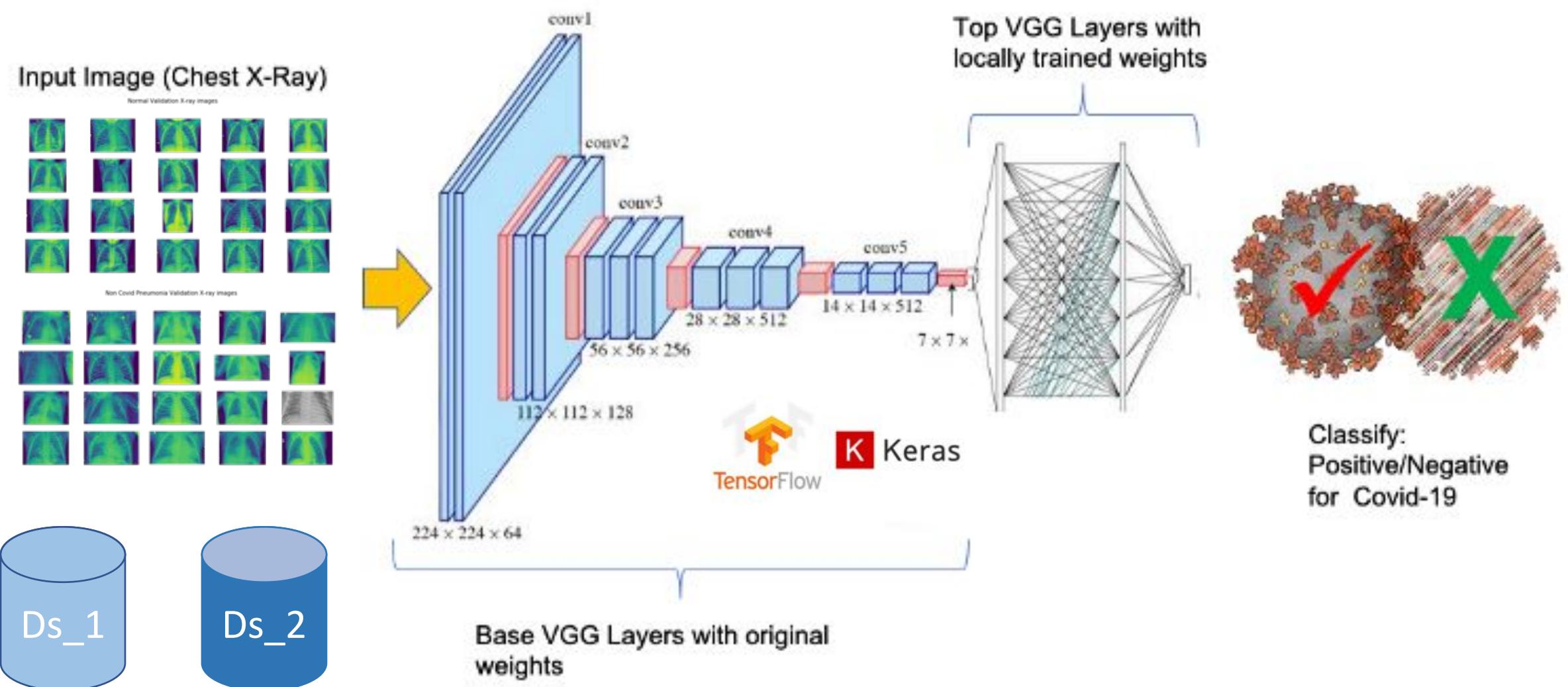


Applying Artificial Intelligence techniques in the development of a web-app for  
the detection of Covid-19 in X-ray images

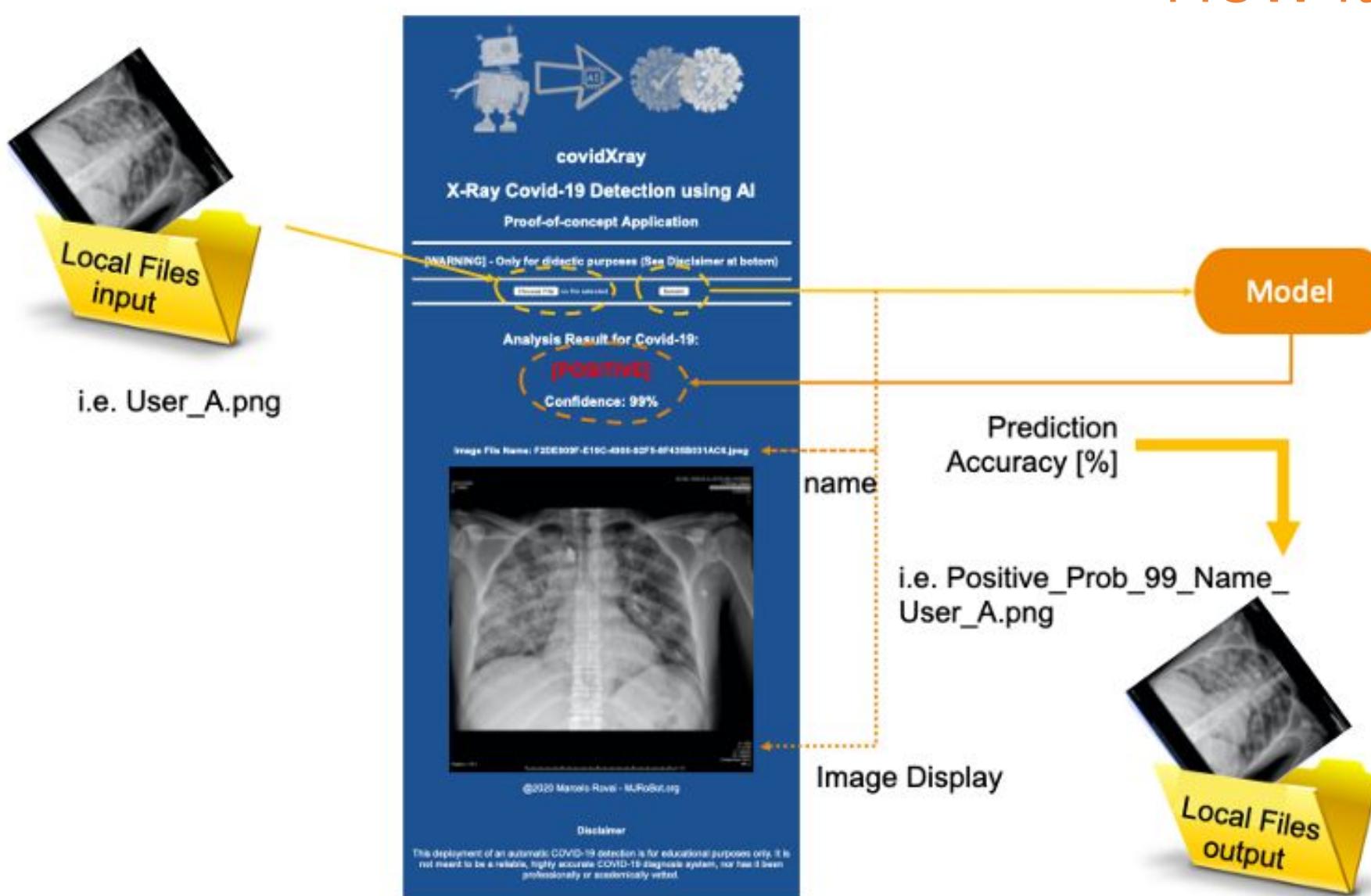
# VGG-16 Convolutional Neural Network Model



# Training the model (Transfer Learning)



# How it works

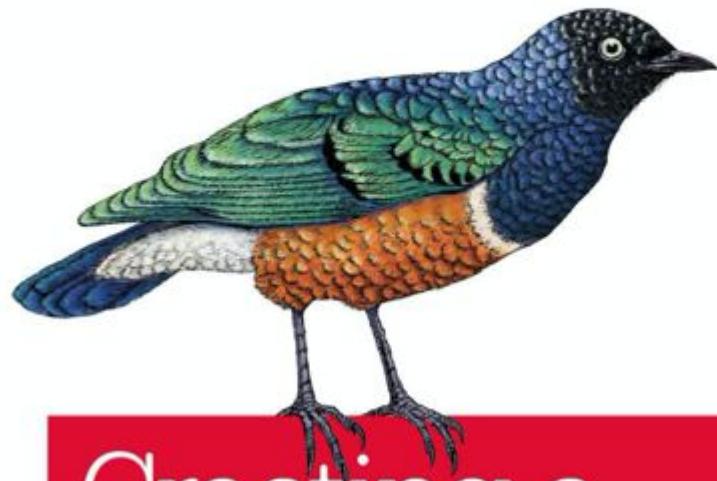


<https://github.com/Mjrovai/covid19Xray>



# Conclusion

O'REILLY®



# Creating a Data-Driven Organization

PRACTICAL ADVICE FROM THE TRENCHES

Carl Anderson



<http://shop.oreilly.com/product/0636920035848.do>



Quarta-feira  
20/05  
às 19:00



EXCLUSIVO PARA MEMBROS

## Data Science aplicado a negócios

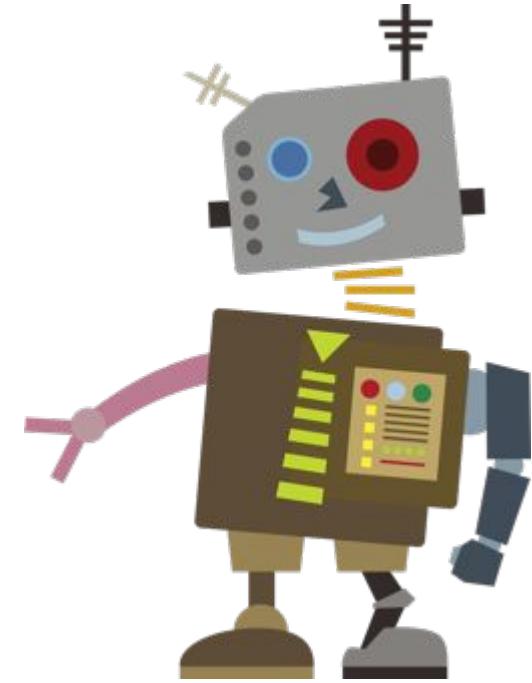
TUDO  
SOBRE **iot**



**Marcelo Rovai**

Mestre Cientista de Dados

# Thanks



[MJRoBot.org](http://MJRoBot.org)

[github.com/Mjrovai](https://github.com/Mjrovai)

[hackster.io/mjrobot](https://hackster.io/mjrobot)

[medium.com/@rovai](https://medium.com/@rovai)

[instructables.com/member/mjrovai](https://instructables.com/member/mjrovai)