COVID_19

GLOBAL EPIDEMIC ANALYSIS

Cathy Li

1/07/2020

BACKGROUND

- Dec 2019, a pneumonia of unknown cause was detected in the city of Wuhan in Hubei province, China.
- Jan 2020, the virus broke out in Wuhan which named Covid-19 by WHO.
- Mar 2020, Covid-19 outbreaks in Italy and Spain.
- Apr 2020, U.S. confirmed cases surged to 1 million.
- Until Now, Covid-19 swept the world, total confirmed cases exceeds 10 million.

DATA ACQUISITION

Many ways to get relevant Data

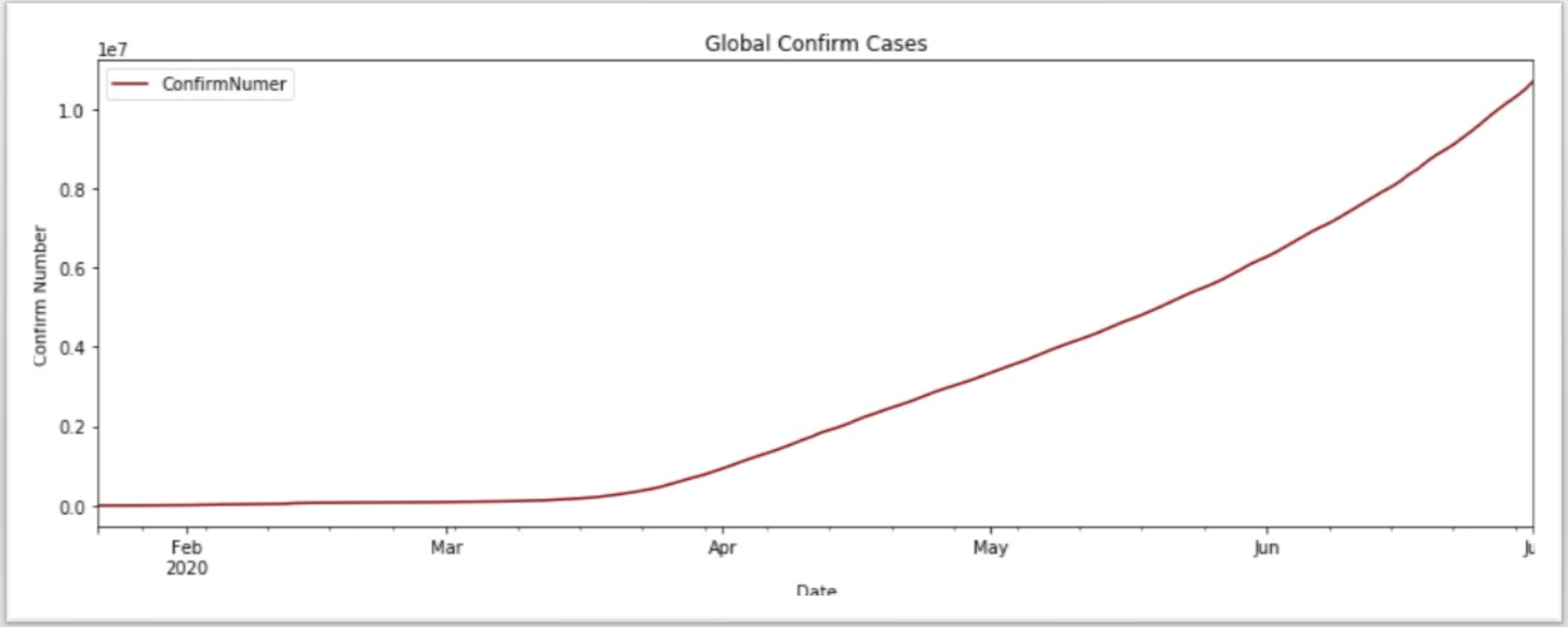
- WHO
- JOHNS HOPKINS
- Other Open Source

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

GLOBAL CONFIRM TREND MAP

- Global Confirm Number 11,088,671, 4th July 2020.

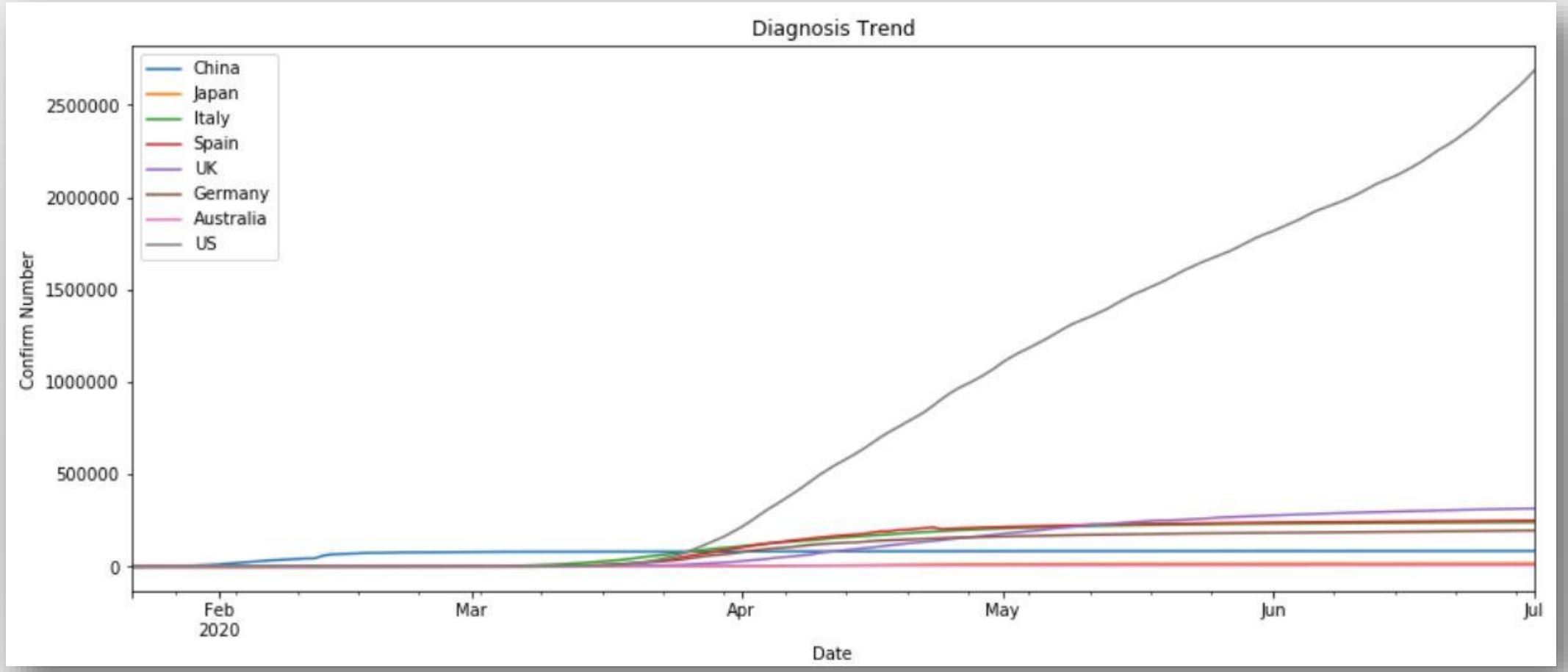
From 22/1/2020 to 1/7/2020



SEVERAL TYPICAL COUNTRIES TREND MAP

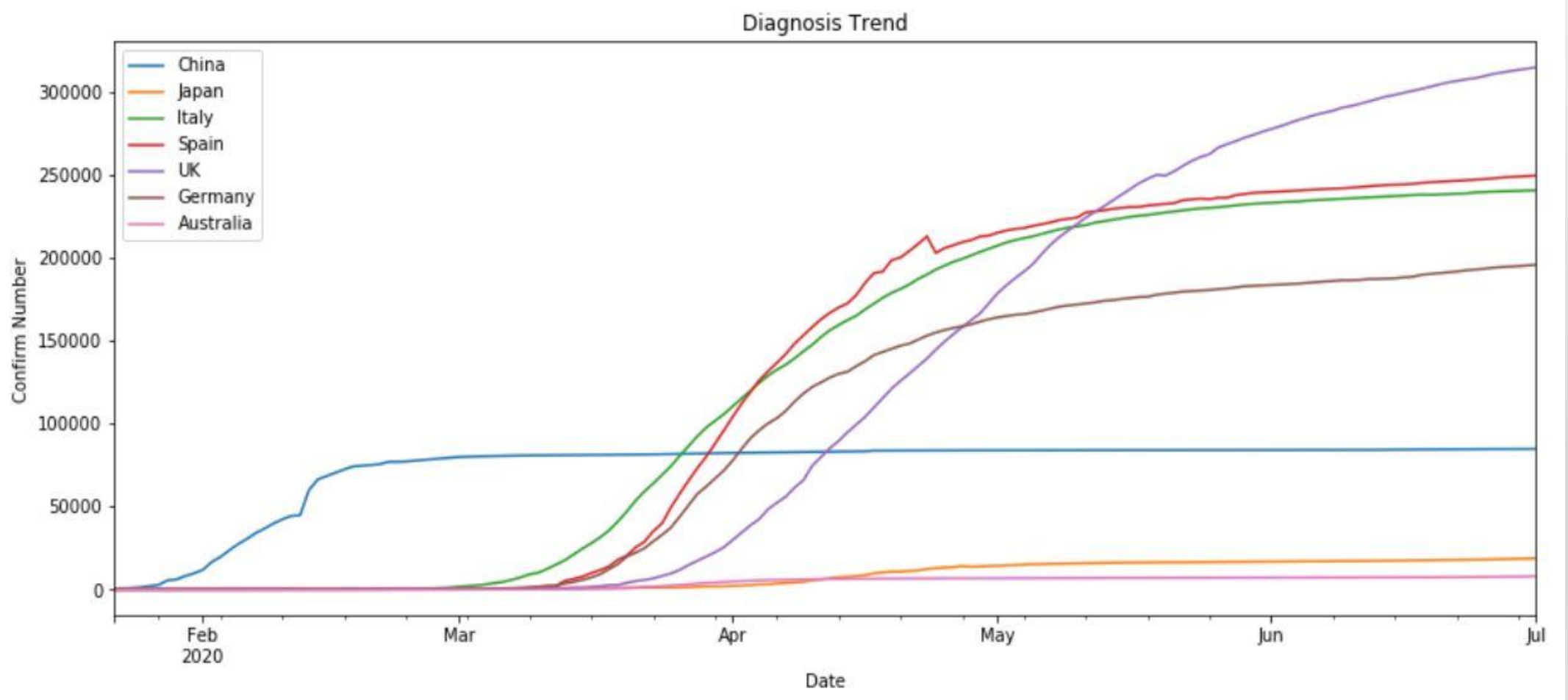
- Eight Countries from different Continents

From 22/1/2020 to 1/7/2020

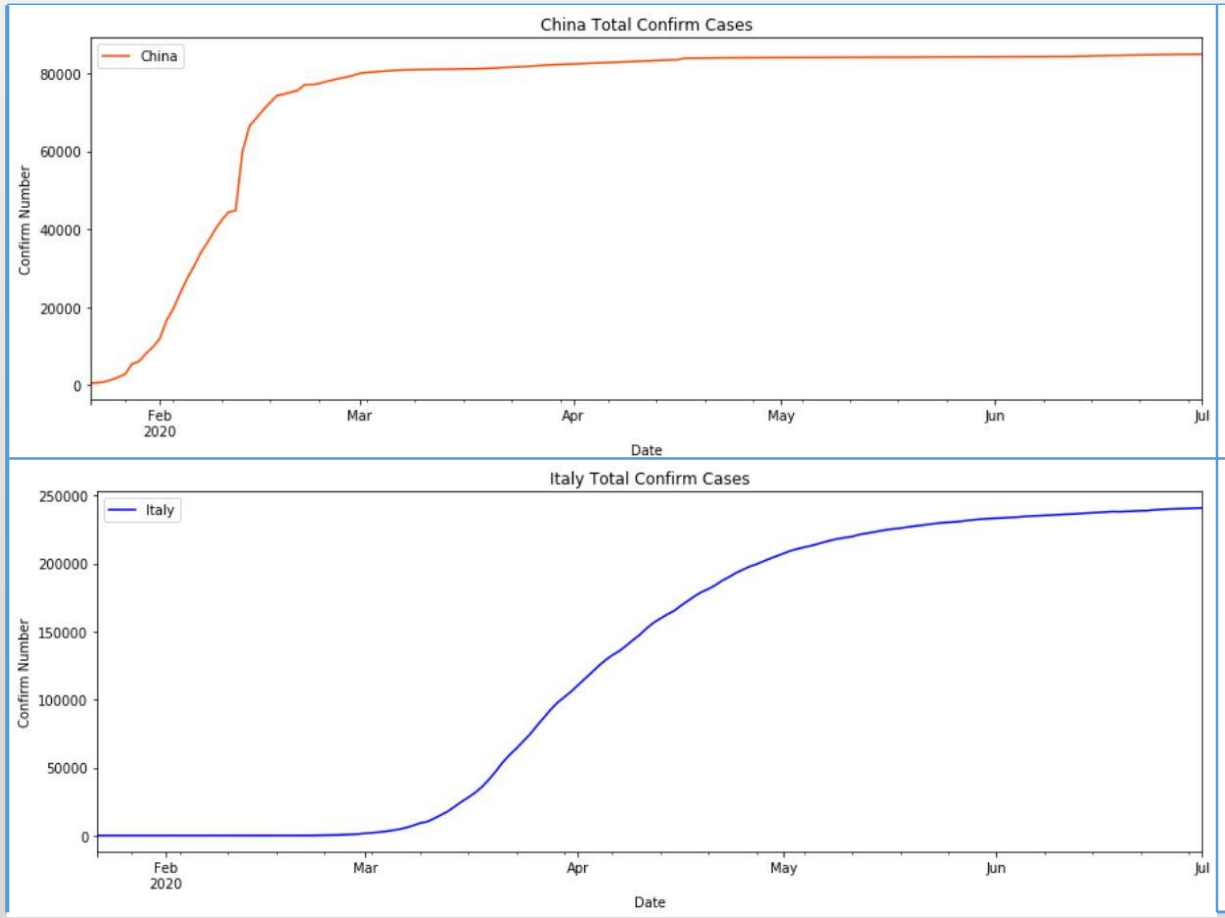


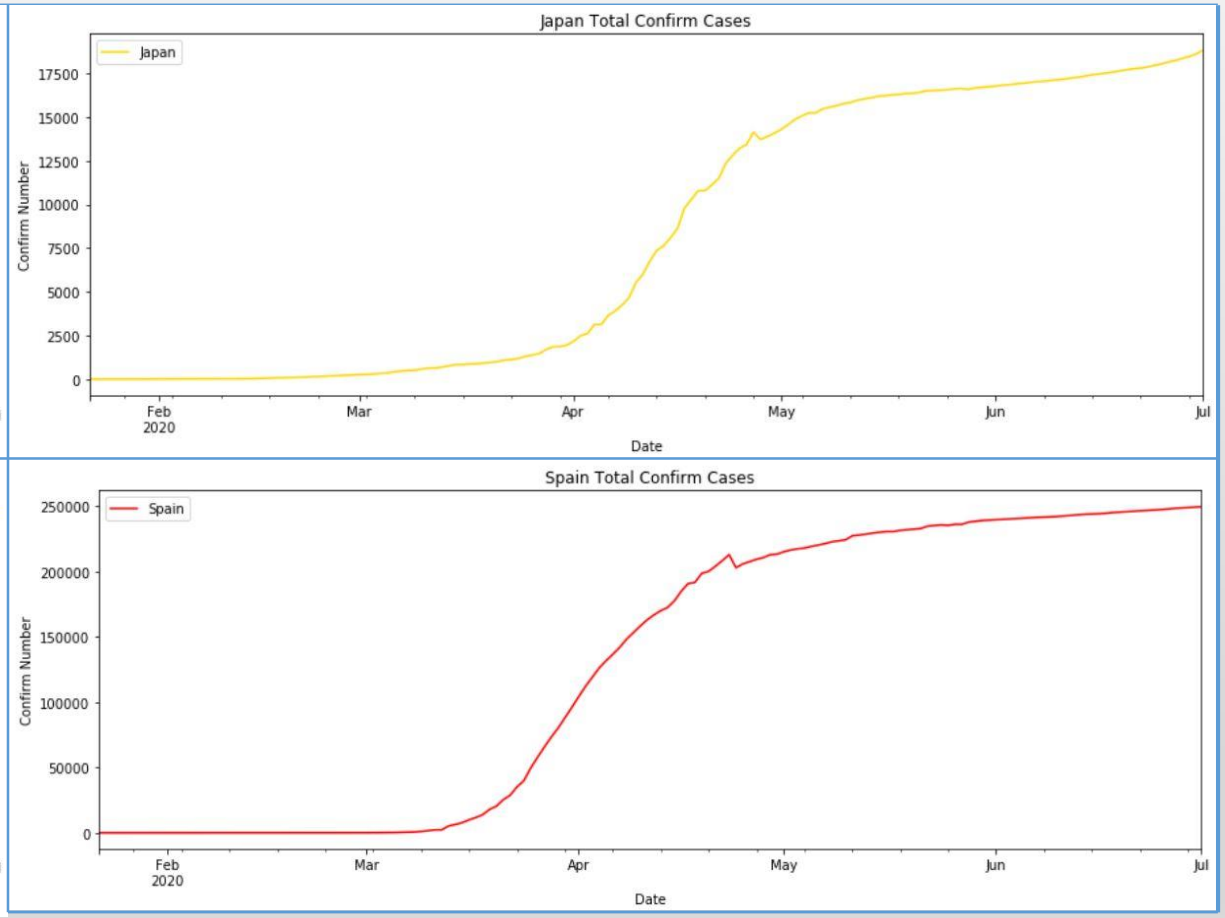
SEVERAL TYPICAL COUNTRIES TREND MAP

- Remove US in order to get other countries clearly trend



TREND MAP DETAILS

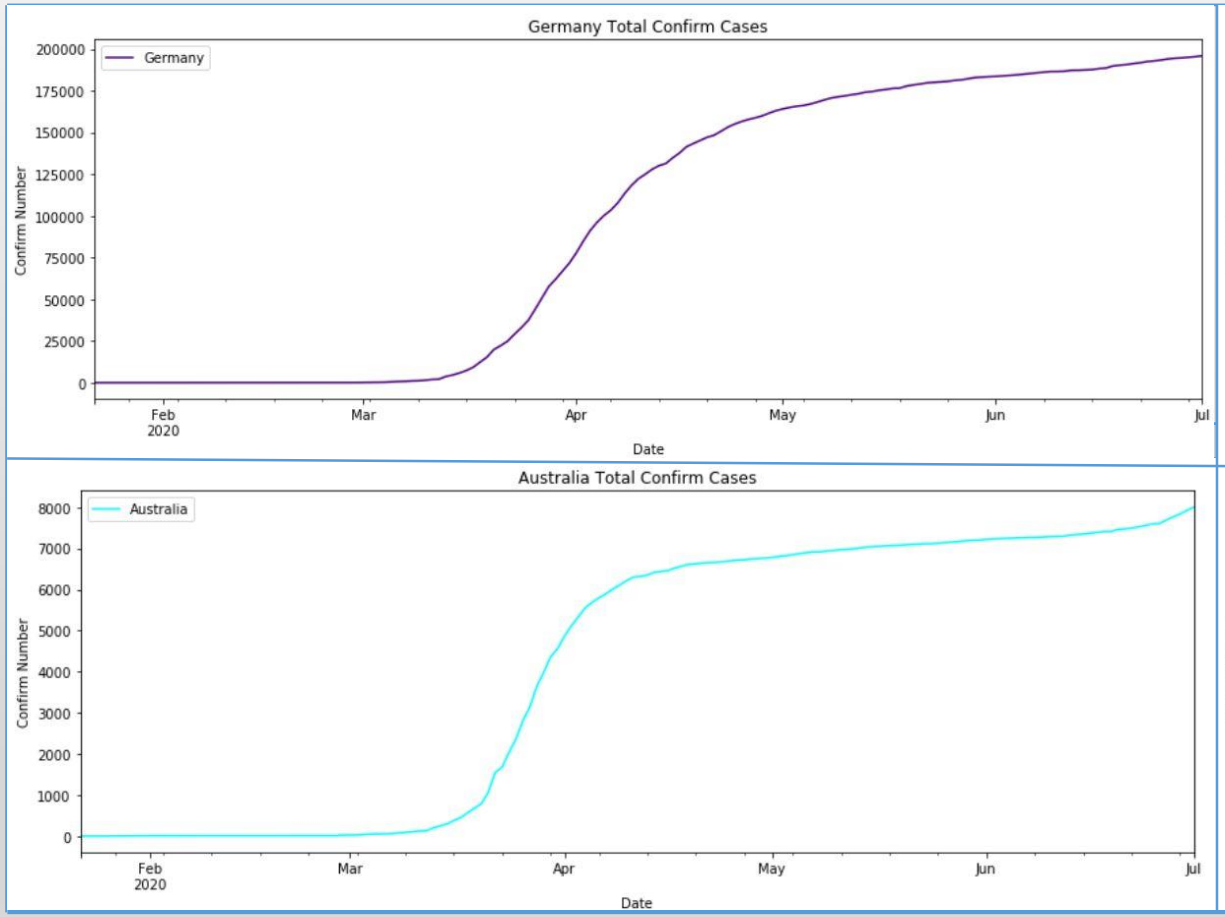


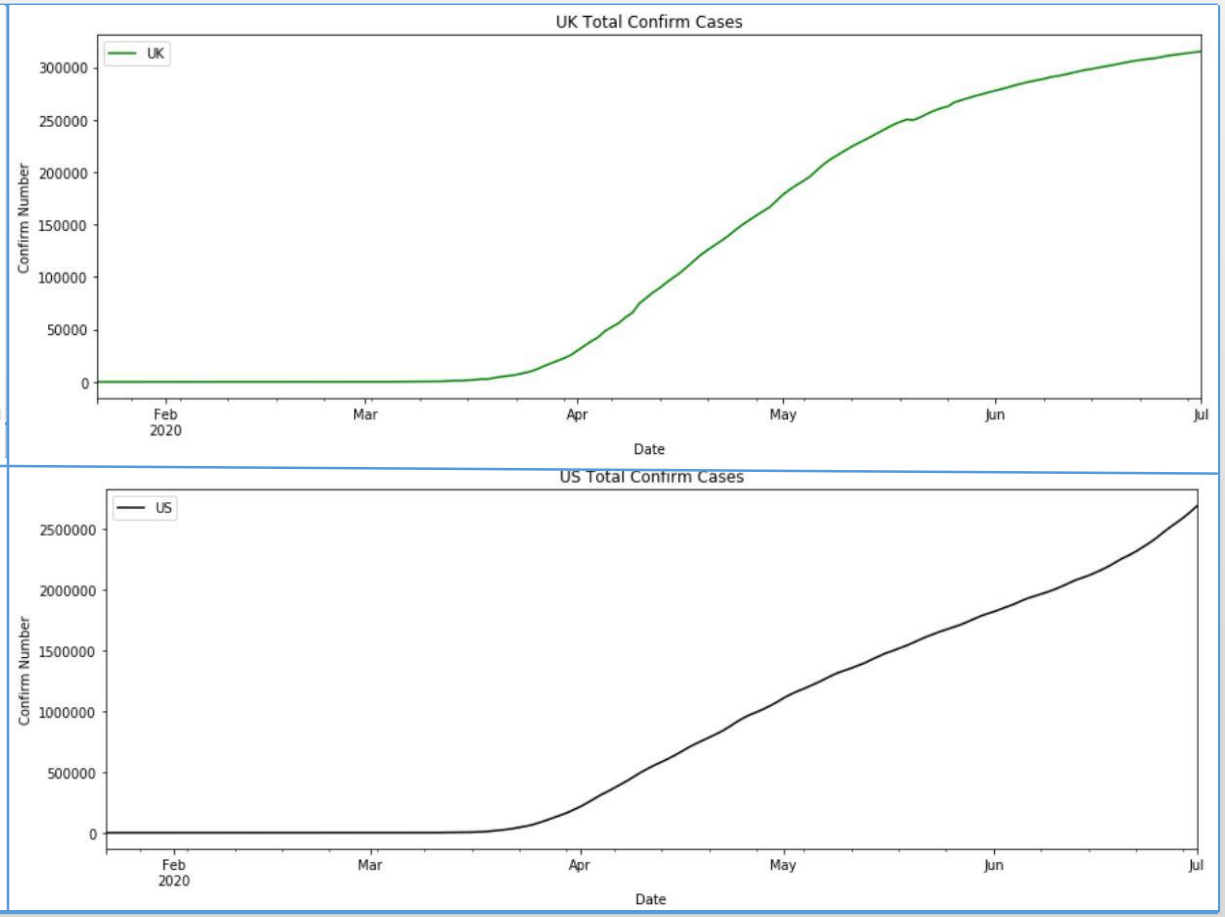


China, Japan, Italy, Spain;

From 22/1/2020 to 1/7/2020

TREND MAP DETAILS





Germany, UK, Australia, US;

From 22/1/2020 to 1/7/2020

DIAGNOSIS RATE ANALYSIS

	Country/Region	Total Confirm Number	Population	Diagnosis Rate
0	China	84785	1439323776	0.00589061%
1	Japan	18615	126476461	0.01471815%
2	Italy	240578	60461826	0.39790065%
3	Spain	249271	46754778	0.53314551%
4	United Kingdom	314160	67886011	0.46277576%
5	Germany	195418	83783942	0.23324040%
6	Australia	7920	25499884	0.03105896%
7	US	2635603	331002651	0.79624831%

From 22/1/2020 to 30/6/2020

- Replace China with Hubei.
- Replace U.S. with New York.

Because there's population are too large to do comparative analysis.

DIAGNOSIS RATE ANALYSIS

	Country/Region	Total Confirm Number	Population	Diagnosis Rate
0	Japan	18615	126476461	0.01471815%
1	Italy	240578	60461826	0.39790065%
2	Spain	249271	46754778	0.53314551%
3	United Kingdom	314160	67886011	0.46277576%
4	Germany	195418	83783942	0.23324040%
5	Australia	7920	25499884	0.03105896%
6	China/Hubei	68135	59270000	0.11495698%
7	US/New York	393454	19450000	2.02289974%

From 22/1/2020 to 30/6/2020

- Japan has the lowest Diagnosis Rate with largest population.
- Australia has the lowest Confirm Number.
- NY ranks first in Confirm Number and Diagnosis Rate.

MORTALITY RATE ANALYSIS

From 22/1/2020 to 30/6/2020

	Country/Region	Total Deaths Number	Mortality Rate by Confirmed Number	Mortality Rate by Population
0	China/Hubei	4512	6.62214721%	0.00761262%
1	Japan	972	5.22159549%	0.00076852%
2	Italy	34767	14.45144610%	0.05750240%
3	Spain	28355	11.37517000%	0.06064621%
4	United Kingdom	43815	13.94671505%	0.06454202%
5	Germany	8990	4.60039505%	0.01072998%
6	Australia	104	1.31313131%	0.00040784%
7	US/New York	32032	8.14123125%	0.16468895%

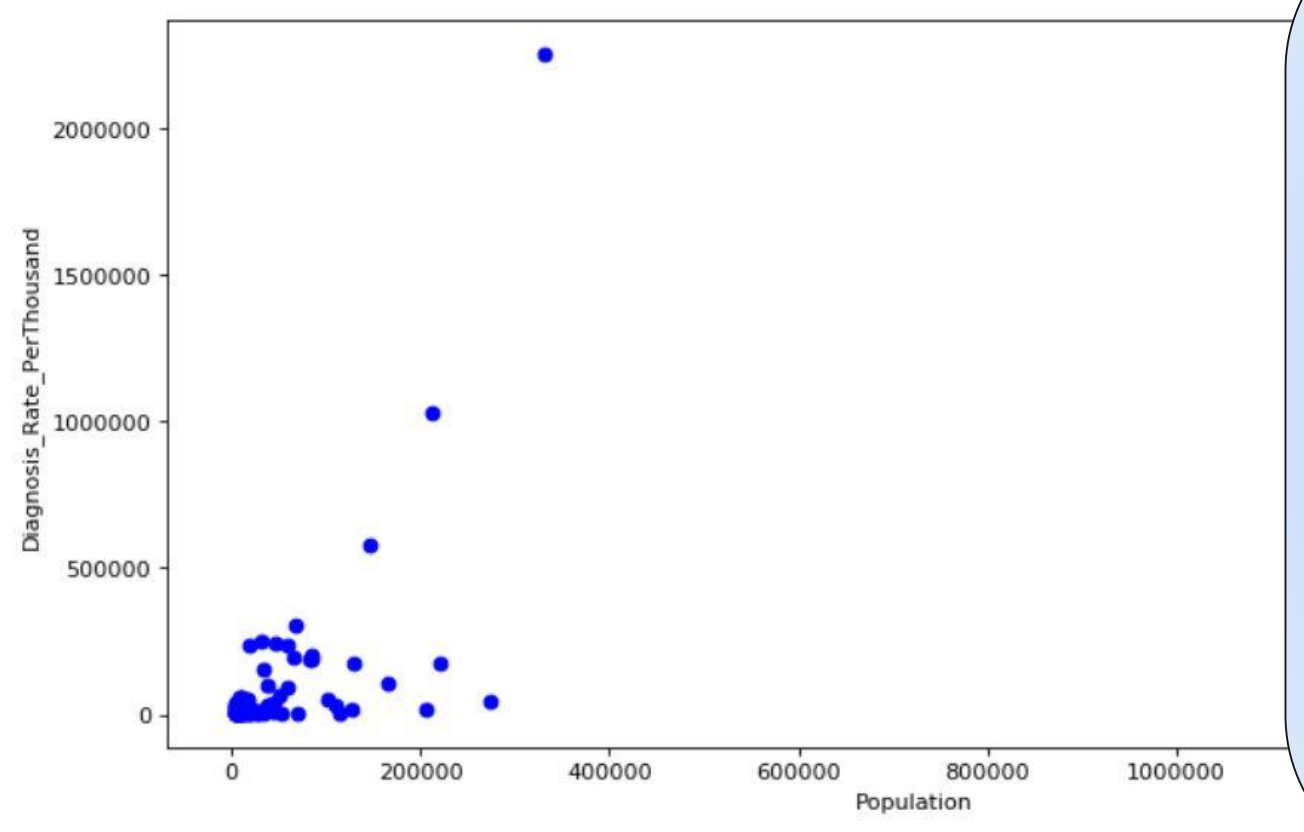
MORTALITY RATE ANALYSIS

- Rank-1 (MR/C<=5%)
  - Australia and Germany
- Rank-2 (5%< MR/C <=10%)
  - Hubei/China, Japan and New York/US
- Rank-3 (MR/C >10%)
  - Italy, Spain and UK

Conclusion:

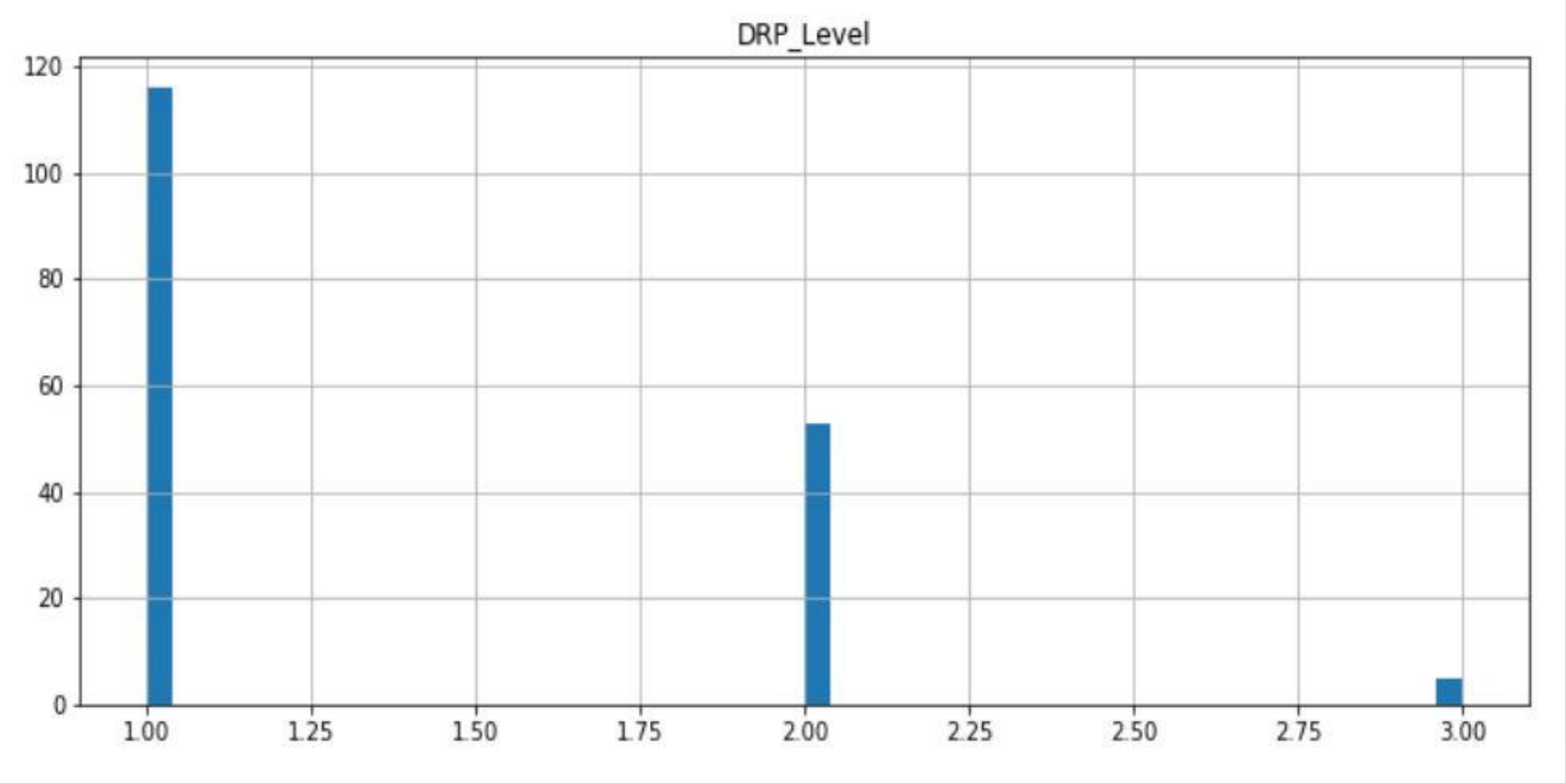
1. Japan and Australia did the best job at controlling coronavirus spreading.
2. Germany and New York, US also in good performance even they has a big amount confirmed cases, it shows their adequate medical resources.

PREDICT MODEL -KNN



K-Nearest Neighbors is an algorithm for supervised learning. Where the data is 'trained' with data points corresponding to their classification. Once a point is to be predicted, it takes into account the 'K' nearest points to it to determine it's classification.

PREDICT MODEL -KNN

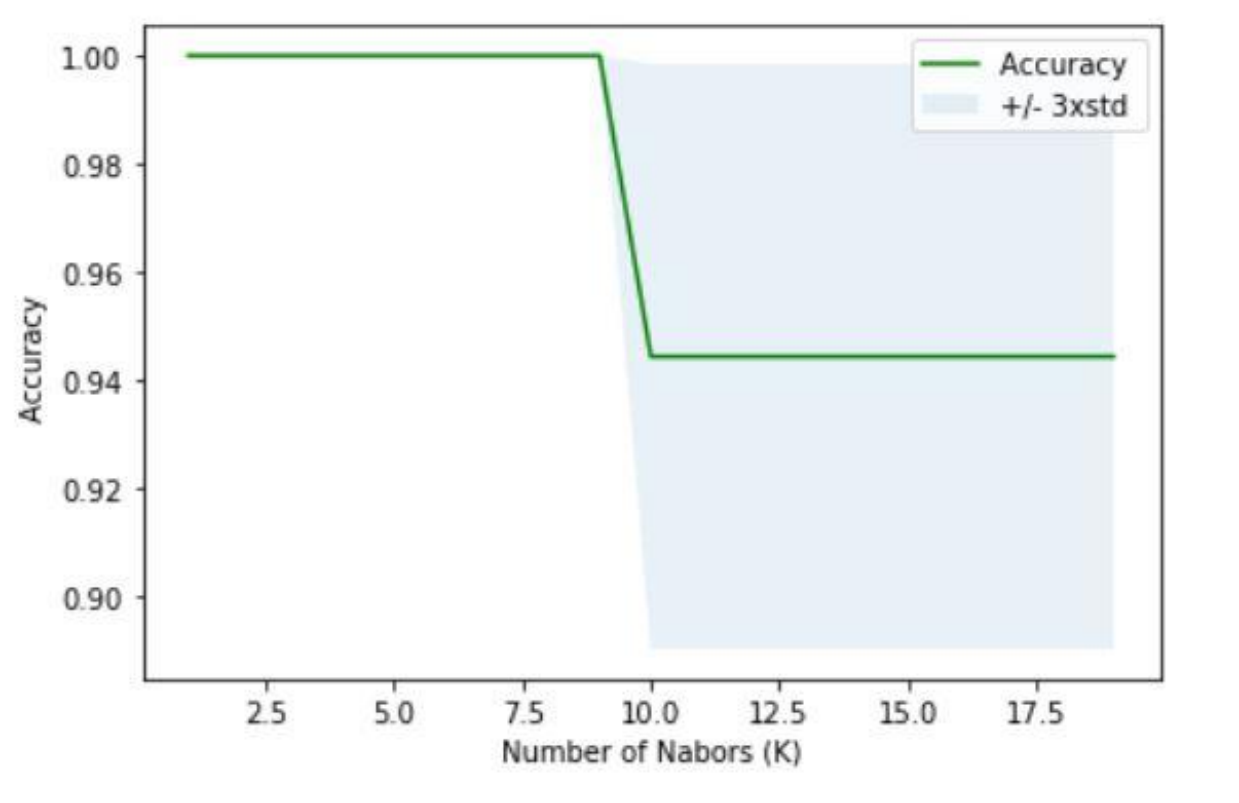


Import Feature – DRP Level

*Diagnosis Rate Per Thousand Level

- Level-1, DRP <=1
- Level-2, 1<DRP <=10
- Level-3, DRP>10

PREDICT MODEL -KNN



Calculate Best K

It's obvious that the best K's range are in 0 to 10 from the figure, here I'll take the median k = 5.

KNN models accuracy while

k = 5:

- Train set Accuracy: 0.9615384615384616
- Test set Accuracy: 1.0

MODEL PREDICTION - US

Use this Model to predict U.S.

- Sample: 52 states of U.S.
- Not Matching Data: 12 states
- Test set Accuracy: 0.7692307692307693

✓ Not Matching Data Real DRP as below:

MODEL PREDICTION - US

	Province_State	Total_Confirm_Number	Population	Density	Diagnosis_Rate_PerThousand	DRP_Level
0	Connecticut	46059	3563077	735.868900	12.926748	3.0
1	Delaware	11017	982895	504.307300	11.208725	3.0
2	District of Columbia	10185	720687	11814.541000	14.132349	3.0
3	Hawaii	866	1412687	219.941900	0.613016	1.0
4	Illinois	140291	12659682	228.024300	11.081716	3.0
5	Louisiana	54769	4645184	107.517500	11.790491	3.0
6	Maryland	66115	6083116	626.673100	10.868607	3.0
7	Massachusetts	108070	6976597	894.435500	15.490360	3.0
8	Montana	829	1086759	7.466800	0.762819	1.0
9	New Jersey	170584	8936574	1215.199100	19.088299	3.0
10	New York	391220	19440469	412.521100	20.124000	3.0
11	Rhode Island	16661	1056161	1021.432300	15.775057	3.0

MODEL PREDICTION - CHINA

Use this Model to predict China

- Sample: 33 Provinces
- Only 1 Province meet predict result – Hubei (*first outbreak region)
- Others Province/Region all lower than predict DRP Level

✓Partial of Not Matching Data Real DRP as below:

MODEL PREDICTION - CHINA

	Province/State	Total_Confirm_Number	Population	Density	Diagnosis_Rate_PerThousand	DRP_Level
0	Hubei	68135	59270.000000	325.000000	1.149570	2.0
1	Hong Kong	1196	7500.700000	6544.000000	0.159452	1.0
2	Macau	46	679.600000	20778.000000	0.067687	1.0
3	Beijing	891	21536.000000	1322.740000	0.041373	1.0
4	Shanghai	706	24281.400000	3814.000000	0.029076	1.0
5	Heilongjiang	947	37513.000000	81.000000	0.025245	1.0
6	Zhejiang	1269	58500.000000	460.000000	0.021692	1.0
7	Jiangxi	932	46661.000000	247.000000	0.019974	1.0
8	Chongqing	582	31243.200000	374.000000	0.018628	1.0
9	Hainan	171	9447.200000	224.000000	0.018101	1.0
10	Anhui	991	63659.000000	429.000000	0.015567	1.0
11	Hunan	1019	69183.800000	304.000000	0.014729	1.0
12	Guangdong	1637	115210.000000	481.000000	0.014209	1.0

CONCLUSION

- Covid-19 is a global healthy crisis, must be taken seriously.
- Herd Immunity is definitely unfeasible, we can't afford the price at all.
- Lock down city is a good way to slow down it's spreading.
- Wearing masks and Social distancing are the simplest and most effective way that we can do to saving lives.
- At last, hope everyone is safe.

Thanks