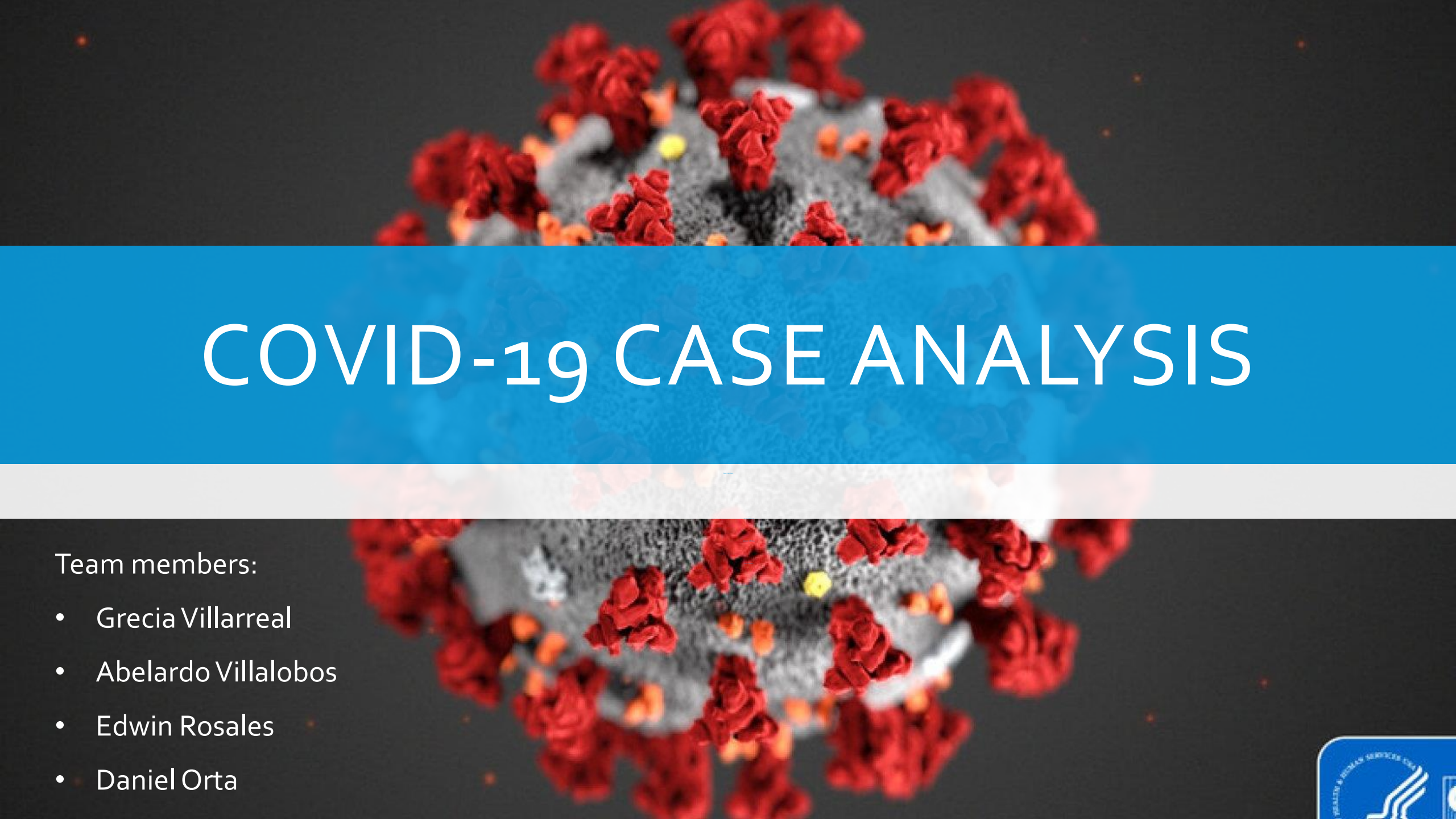


COVID-19 CASE ANALYSIS

Team members:

- Grecia Villarreal
- Abelardo Villalobos
- Edwin Rosales
- Daniel Orta

MOTIVATION & SUMMARY

- **COVID-19** is being the headline in the entire world for the last 3 months, it affects not just Health Sector, but some other like Economy, Finances, Tourism, among others.
- The scope of our project is to analyze the **evolution of the virus in humans**: confirmed cases, deaths and recoveries.
- We are seeking validation of the effectiveness of **public policy** adopted by **China** (Dictatorial Regime-One Party) versus the Rest of the World, regarding the specific case of **COVID-19** spread.

6 MARCH 2020 ANALYSIS

Coronavirus outbreak: safety measures at major international airports

By **Praveen Duddu** SHARE

Following preventive safety measures by some of the major and busiest international airports to prevent the spread of coronavirus (Covid-19) after its outbreak in Wuhan, China, many others initiated similar steps. The coronavirus has spread to 77 more countries in a month, inviting border closures and health emergencies declared by a number of countries.

QUESTIONS WE WANT TO ANSWER:

- What is the percentage of **recovered**, **death** and **active cases** worldwide? This will help us understand the overview of the current status of the virus.
- How many confirmed, death and recovered **cases** there are **per day** across the world? This will help us understand how have the cases increased worldwide in a daily basis and the cumulative comparison of the different status of the disease.
- What is the **concentrations of cases by region**? To visually represent the spread of COVID-19 cases globally.
- What is the **rate of increase** in confirmed cases in **China** compared to the **rest of the world**? This will help us identify if measures haven been taken in these two regions and identify their effectiveness in general.

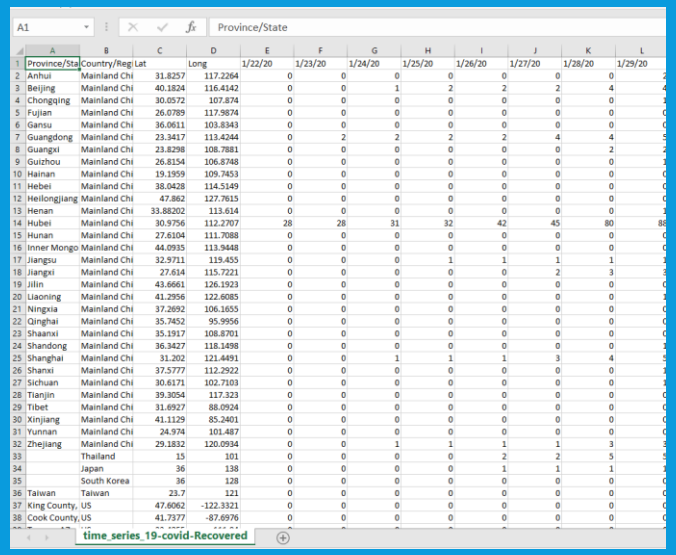
DATA CLEANUP & EXPLORATION

- We found three public csv's of data updated daily, that we consolidated, transposed and aggregated to get the data in the format we needed it to create our visualizations and answer our questions.

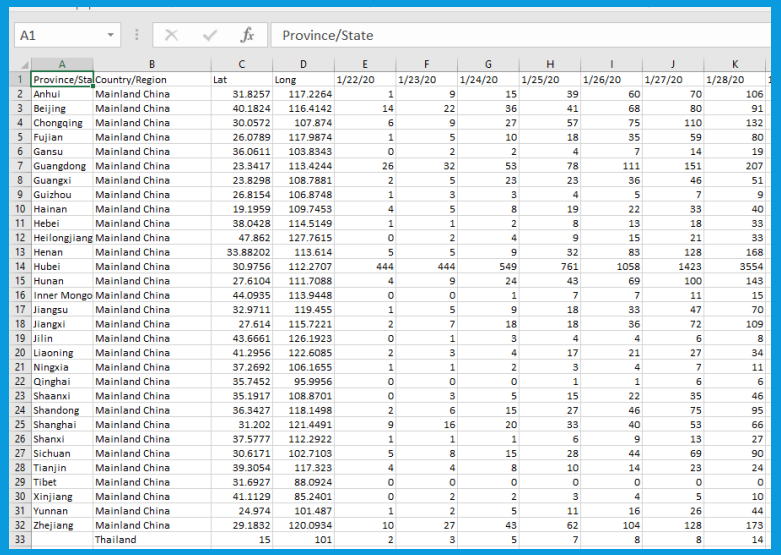
time_series_19-covid-Confirmed

time_series_19-covid-Deaths

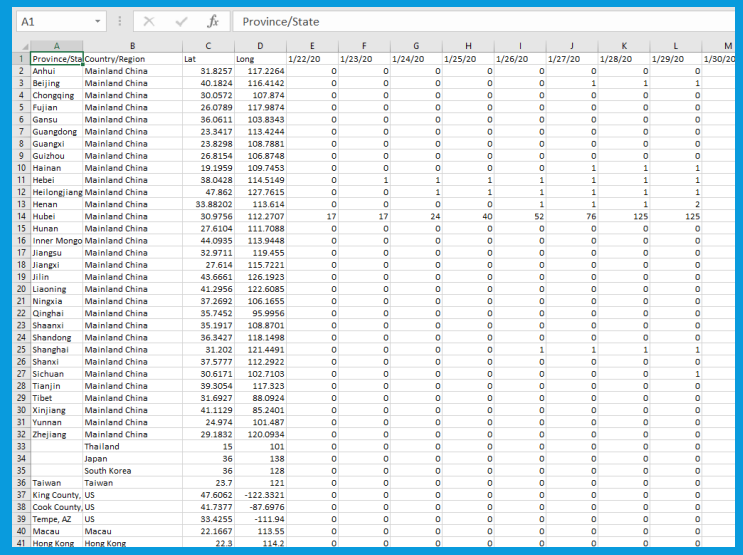
time_series_19-covid-Recovered



Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20
Anhui	Mainland Chi	31.8257	117.2264	0	0	0	0	0	0	0	2
Beijing	Mainland Chi	40.1824	116.4142	0	0	1	2	2	2	4	4
Chongqing	Mainland Chi	30.0572	107.874	0	0	0	0	0	0	0	1
Fujian	Mainland Chi	26.0789	117.9874	0	0	0	0	0	0	0	0
Gansu	Mainland Chi	36.0611	103.8343	0	0	0	0	0	0	0	0
Guangdong	Mainland Chi	23.3417	113.4244	0	2	2	2	2	4	4	5
Guangxi	Mainland Chi	23.8298	108.7881	0	0	0	0	0	0	2	2
Guizhou	Mainland Chi	26.8154	106.8748	0	0	0	0	0	0	0	1
Hainan	Mainland Chi	19.1959	109.7453	0	0	0	0	0	0	0	0
Hebei	Mainland Chi	38.0428	114.5149	0	0	0	0	0	0	0	0
Heilongjiang	Mainland Chi	47.862	127.7615	0	0	0	0	0	0	0	0
Henan	Mainland Chi	33.88202	113.614	0	0	0	0	0	0	0	1
Hubei	Mainland Chi	30.9756	112.2707	28	28	31	32	42	45	80	88
Hunan	Mainland Chi	27.6104	111.7088	0	0	0	0	0	0	0	0
Inner Mongo	Mainland Chi	44.0935	113.9448	0	0	0	0	0	0	0	0
Jiangsu	Mainland Chi	32.9711	119.455	0	0	0	1	1	1	1	1
Jiangxi	Mainland Chi	27.614	115.7221	0	0	0	0	0	2	3	3
Jilin	Mainland Chi	43.6661	126.1923	0	0	0	0	0	0	0	0
Liaoning	Mainland Chi	41.2956	122.6085	0	0	0	0	0	0	0	0
Ningxia	Mainland Chi	37.2692	106.1655	0	0	0	0	0	0	0	0
Qinghai	Mainland Chi	35.7452	95.9956	0	0	0	0	0	0	0	0
Shaanxi	Mainland Chi	35.1917	108.8701	0	0	0	0	0	0	0	0
Shandong	Mainland Chi	36.3427	118.1498	0	0	0	0	0	0	0	1
Shanghai	Mainland Chi	31.202	121.4491	0	0	1	1	1	3	4	5
Shanxi	Mainland Chi	37.5777	112.2922	0	0	0	0	0	0	0	1
Sichuan	Mainland Chi	30.6171	102.7103	0	0	0	0	0	0	0	0
Tianjin	Mainland Chi	39.3054	117.323	0	0	0	0	0	0	0	0
Tibet	Mainland Chi	31.6927	88.0924	0	0	0	0	0	0	0	0
Xinjiang	Mainland Chi	41.1129	85.2401	0	0	0	0	0	0	0	0
Yunnan	Mainland Chi	24.974	101.487	0	0	0	0	0	0	0	0
Zhejiang	Mainland Chi	29.1832	120.0934	0	0	0	1	1	1	1	3
	Thailand	15	101	0	0	0	0	2	2	5	5
	Japan	36	138	0	0	0	0	0	1	1	1
	South Korea	36	128	0	0	0	0	0	0	0	0
Taiwan	Taiwan	23.7	121	0	0	0	0	0	0	0	0
King County,	US	47.6062	-122.3321	0	0	0	0	0	0	0	0
Cook County,	US	41.7377	-87.6976	0	0	0	0	0	0	0	0



Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20
Anhui	Mainland China	31.8257	117.2264	1	9	15	39	60	70	106
Beijing	Mainland China	40.1824	116.4142	14	22	36	41	68	80	91
Chongqing	Mainland China	30.0572	107.874	6	9	27	57	75	110	132
Fujian	Mainland China	26.0789	117.9874	1	5	10	18	35	59	80
Gansu	Mainland China	36.0611	103.8343	0	2	2	4	7	14	19
Guangdong	Mainland China	23.3417	113.4244	26	32	53	78	111	151	207
Guangxi	Mainland China	23.8298	108.7881	2	5	23	23	36	46	51
Guizhou	Mainland China	26.8154	106.8748	1	3	3	4	5	7	9
Hainan	Mainland China	19.1959	109.7453	4	5	8	19	22	33	40
Hebei	Mainland China	38.0428	114.5149	1	1	2	8	13	18	33
Heilongjiang	Mainland China	47.862	127.7615	0	2	4	9	15	21	33
Henan	Mainland China	33.88202	113.614	5	5	9	32	83	128	168
Hubei	Mainland China	30.9756	112.2707	444	444	549	761	1058	1423	3554
Hunan	Mainland China	27.6104	111.7088	4	9	24	43	69	100	143
Inner Mongo	Mainland China	44.0935	113.9448	0	0	1	7	7	11	15
Jiangsu	Mainland China	32.9711	119.455	1	5	9	18	33	47	70
Jiangxi	Mainland China	27.614	115.7221	2	7	18	18	36	72	109
Jilin	Mainland China	43.6661	126.1923	0	1	3	4	4	6	8
Liaoning	Mainland China	41.2956	122.6085	2	3	4	17	21	27	34
Ningxia	Mainland China	37.2692	106.1655	1	1	2	3	4	7	11
Qinghai	Mainland China	35.7452	95.9956	0	0	0	1	1	6	6
Shaanxi	Mainland China	35.1917	108.8701	0	3	5	15	22	35	46
Shandong	Mainland China	36.3427	118.1498	2	6	15	27	46	75	95
Shanghai	Mainland China	31.202	121.4491	9	16	20	33	40	53	66
Shanxi	Mainland China	37.5777	112.2922	1	1	1	6	9	13	27
Sichuan	Mainland China	30.6171	102.7103	5	8	15	28	44	69	90
Tianjin	Mainland China	39.3054	117.323	4	4	8	10	14	23	24
Tibet	Mainland China	31.6927	88.0924	0	0	0	0	0	0	0
Xinjiang	Mainland China	41.1129	85.2401	0	2	2	3	4	5	10
Yunnan	Mainland China	24.974	101.487	1	2	5	11	16	26	44
Zhejiang	Mainland China	29.1832	120.0934	10	27	43	62	104	128	173
	Thailand	15	101	2	3	5	7	8	8	14



Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20
Anhui	Mainland China	31.8257	117.2264	0	0	0	0	0	0	0	0	
Beijing	Mainland China	40.1824	116.4142	0	0	0	0	0	1	1	1	
Chongqing	Mainland China	30.0572	107.874	0	0	0	0	0	0	0	0	
Fujian	Mainland China	26.0789	117.9874	0	0	0	0	0	0	0	0	
Gansu	Mainland China	36.0611	103.8343	0	0	0	0	0	0	0	0	
Guangdong	Mainland China	23.3417	113.4244	0	0	0	0	0	0	0	0	
Guangxi	Mainland China	23.8298	108.7881	0	0	0	0	0	0	0	0	
Guizhou	Mainland China	26.8154	106.8748	0	0	0	0	0	0	0	0	
Hainan	Mainland China	19.1959	109.7453	0	0	0	0	0	1	1	1	
Hebei	Mainland China	38.0428	114.5149	0	1	1	1	1	1	1	1	
Heilongjiang	Mainland China	47.862	127.7615	0	0	1	1	1	1	1	1	
Henan	Mainland China	33.88202	113.614	0	0	0	0	1	1	1	2	
Hubei	Mainland China	30.9756	112.2707	17	17	24	40	52	76	125	125	
Hunan	Mainland China	27.6104	111.7088	0	0	0	0	0	0	0	0	
Inner Mongo	Mainland China	44.0935	113.9448	0	0	0	0	0	0	0	0	
Jiangsu	Mainland China	32.9711	119.455	0	0	0	0	0	0	0	0	
Jiangxi	Mainland China	27.614	115.7221	0	0	0	0	0	0	0	0	
Jilin	Mainland China	43.6661	126.1923	0	0	0	0	0	0	0	0	
Liaoning	Mainland China	41.2956	122.6085	0	0	0	0	0	0	0	0	
Ningxia	Mainland China	37.2692	106.1655	0	0	0	0	0	0	0	0	
Qinghai	Mainland China	35.7452	95.9956	0	0	0	0	0	0	0	0	
Shaanxi	Mainland China	35.1917	108.8701	0	0	0	0	0	0	0	0	
Shandong	Mainland China	36.3427	118.1498	0	0	0	0	0	0	0	0	
Shanghai	Mainland China	31.202	121.4491	0	0	0	0	1	1	1	1	
Shanxi	Mainland China	37.5777	112.2922	0	0	0	0	0	0	0	0	
Sichuan	Mainland China	30.6171	102.7103	0	0	0	0	0	0	0	1	
Tianjin	Mainland China	39.3054	117.323	0	0	0	0	0	0	0	0	
Tibet	Mainland China	31.6927	88.0924	0	0	0	0	0	0	0	0	
Xinjiang	Mainland China	41.1129	85.2401	0	0	0	0	0	0	0	0	
Yunnan	Mainland China	24.974	101.487	0	0	0	0	0	0	0	0	
Zhejiang	Mainland China	29.1832	120.0934	0	0	0	0	0	0	0	0	
	Thailand	15	101	0	0	0	0	0	0	0	0	
	Japan	36	138	0	0	0	0	0	0	0	0	
	South Korea	36	128	0	0	0	0	0	0	0	0	
Taiwan	Taiwan	23.7	121	0	0	0	0	0	0	0	0	
King County,	US	47.6062	-122.3321	0	0	0	0	0	0	0	0	
Cook County,	US	41.7377	-87.6976	0	0	0	0	0	0	0	0	
Tempe, AZ	US	33.4255	-111.94	0	0	0	0	0	0	0	0	
Macau	Macau	22.1667	113.55	0	0	0	0	0	0	0	0	
Hong Kong	Hong Kong	22.3	114.2	0	0	0	0	0	0	0	0	

DATA CLEANUP & EXPLORATION

- We had cumulative numbers, while for some analysis we needed the daily cases. We applied a loop calculation to have both options as separated data frames.

```
#Confirmed cases df. Confirmed cases include deaths and recoveries.
confirmed_df = pd.read_csv("Resources/time_series_19-covid-Confirmed.csv")
#Create Status field
confirmed_df["Status"] = confirmed_df.loc[:,'Status'] = 'Confirmed'

#Create a table of daily confirmed cases
confirmedByday=confirmed_df.copy()
i=47
j=0
while (j<=172):
    while (i >=5):
        confirmedByday.iloc[j,i]=confirmedByday.iloc[j,i]-confirmedByday.iloc[j,i-1]
        i = i - 1
    j=j+1
    i=47

deaths_df = pd.read_csv("Resources/time_series_19-covid-Deaths.csv")
#Create Status field
deaths_df["Status"] = deaths_df.loc[:,'Status'] = 'Deaths'

##Create a table of daily death cases
deathsByday=deaths_df.copy()
i=47
j=0
while (j<=172):
    while (i >=5):
        deathsByday.iloc[j,i]=deathsByday.iloc[j,i]-deathsByday.iloc[j,i-1]
        i = i - 1
    j=j+1
    i=47
```

Before

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20
0	Anhui	Mainland China	31.8257	117.2264	1	9	15	39	60	70
1	Beijing	Mainland China	40.1824	116.4142	14	22	36	41	68	80
2	Chongqing	Mainland China	30.0572	107.8740	6	9	27	57	75	110
3	Fujian	Mainland China	26.0789	117.9874	1	5	10	18	35	59
4	Gansu	Mainland China	36.0611	103.8343	0	2	2	4	7	14

After

]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20
0	Anhui	Mainland China	31.8257	117.2264	1	8	6	24	21	10
1	Beijing	Mainland China	40.1824	116.4142	14	8	14	5	27	12
2	Chongqing	Mainland China	30.0572	107.8740	6	3	18	30	18	35
3	Fujian	Mainland China	26.0789	117.9874	1	4	5	8	17	24
4	Gansu	Mainland China	36.0611	103.8343	0	2	0	2	3	7

DATA CLEANUP & EXPLORATION

- The days in our data source were added in columns, but we needed a different structure to be able to analyze it easily. We transposed the data to have days and cases into rows.

```
: #Transpose the data to put days and cases into rows rather than having them by columns.
transposed_df = combined_data.melt(['Province/State', 'Country/Region', 'Lat', 'Long', 'Status'],
                                   var_name='Original Date', value_name='Cases')
transposed_df.head()
```

:

	Province/State	Country/Region	Lat	Long	Status	Original Date	Cases
0	Anhui	Mainland China	31.8257	117.2264	Confirmed	1/22/20	1
1	Beijing	Mainland China	40.1824	116.4142	Confirmed	1/22/20	14
2	Chongqing	Mainland China	30.0572	107.8740	Confirmed	1/22/20	6
3	Fujian	Mainland China	26.0789	117.9874	Confirmed	1/22/20	1
4	Gansu	Mainland China	36.0611	103.8343	Confirmed	1/22/20	0

DATA CLEANUP & EXPLORATION

- The day format was improtant for the charts to show the information chronologically. We applied a timeseries function to create a new date field with correct date format.

```
[7]: #Create a new field to have a Date format out of the Original Date that is a string.
transposed_df['Date']  = pd.to_datetime(transposed_df['Original Date'])
transposed_df
```

[7]:

	Province/State	Country/Region	Lat	Long	Status	Original Date	Cases	Date
0	Anhui	Mainland China	31.8257	117.2264	Confirmed	1/22/20	1	2020-01-22
1	Beijing	Mainland China	40.1824	116.4142	Confirmed	1/22/20	14	2020-01-22
2	Chongqing	Mainland China	30.0572	107.8740	Confirmed	1/22/20	6	2020-01-22
3	Fujian	Mainland China	26.0789	117.9874	Confirmed	1/22/20	1	2020-01-22
4	Gansu	Mainland China	36.0611	103.8343	Confirmed	1/22/20	0	2020-01-22
...	...	...	...	...	...	...	...	...

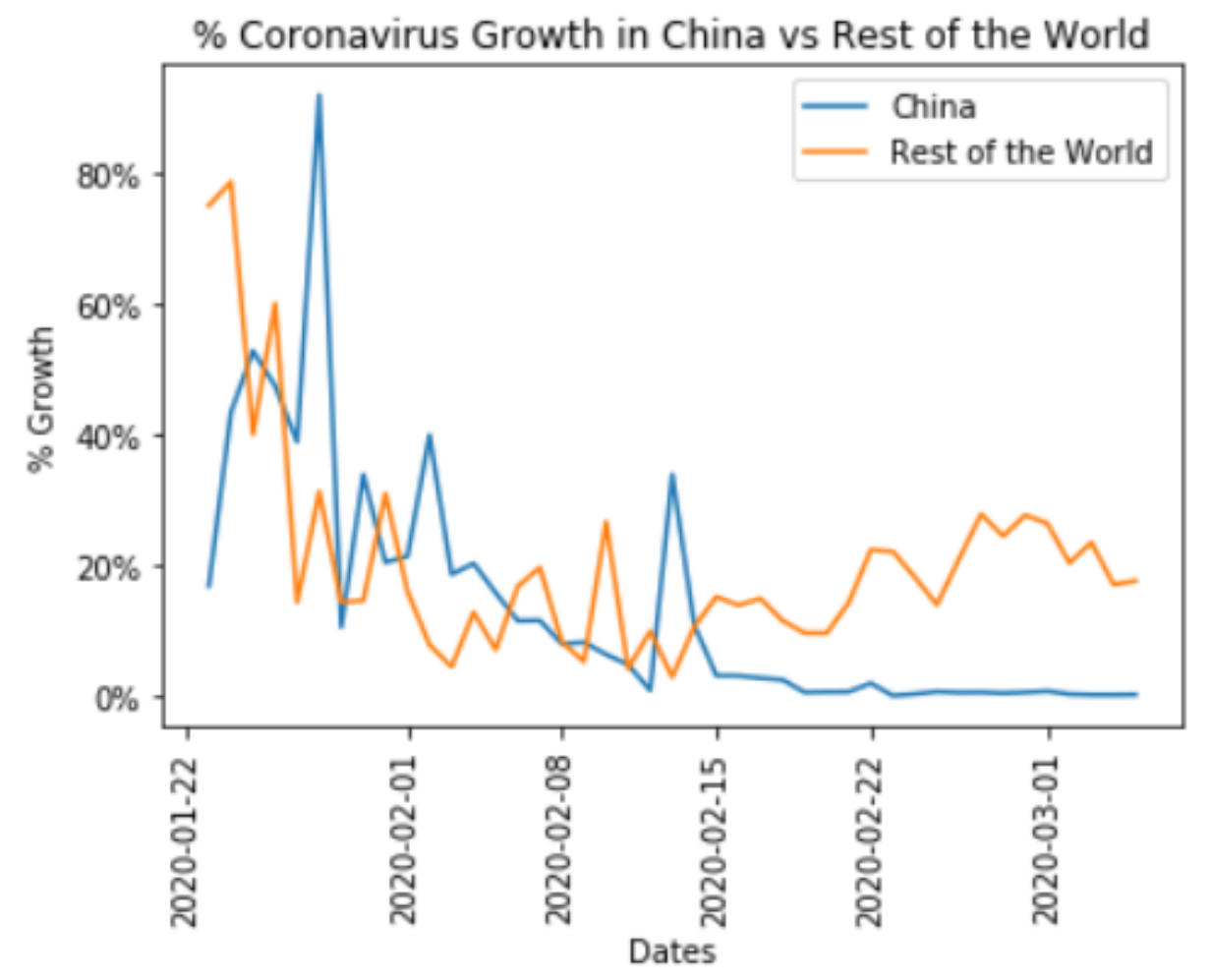
DATA CLEANUP & EXPLORATION

- Also, the calculation and addition of new fields was necessary to create some of the charts:

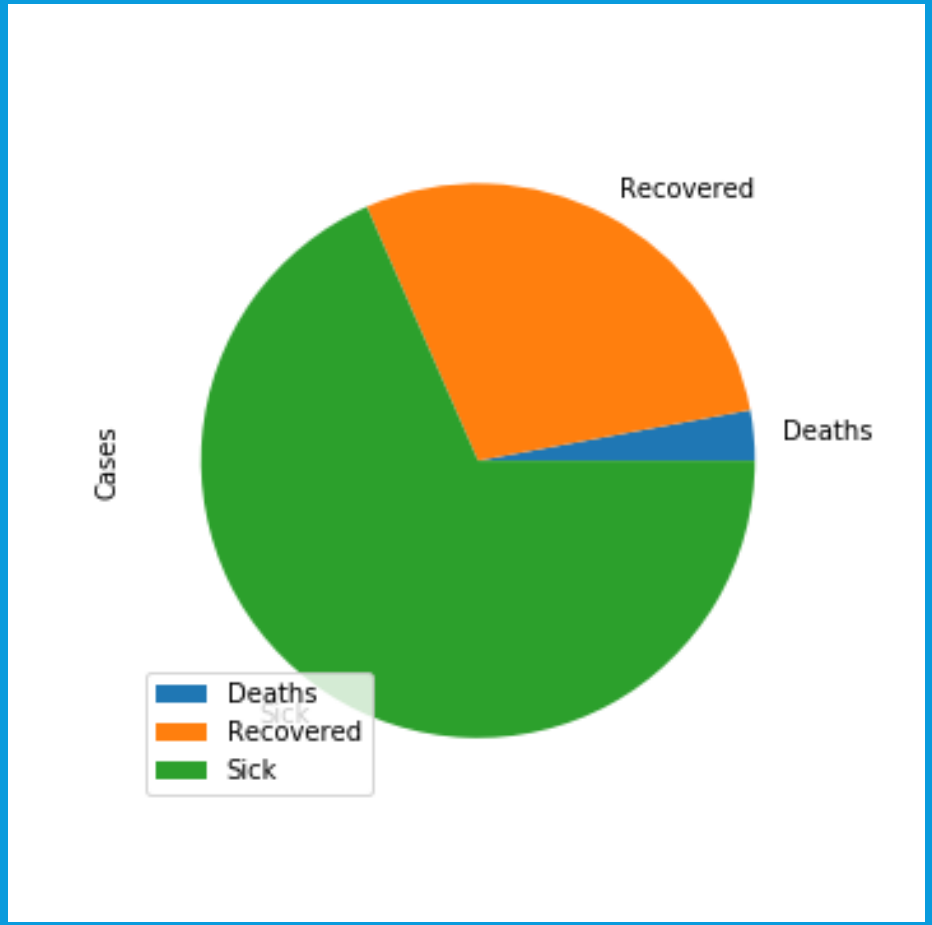
```
#Add percent change variable to the 'rest of the world' df
restofworld_confirmed['% Growth'] = restofworld_df['Cases'].pct_change() #* 100
```

	Status	Date	Cases	% Growth
0	Confirmed	2020-01-22	8	NaN
1	Confirmed	2020-01-23	14	0.750000
2	Confirmed	2020-01-24	25	0.785714
3	Confirmed	2020-01-25	35	0.400000
4	Confirmed	2020-01-26	56	0.600000
5	Confirmed	2020-01-27	64	0.142857
6	Confirmed	2020-01-28	84	0.312500
7	Confirmed	2020-01-29	96	0.142857
8	Confirmed	2020-01-30	110	0.145833
9	Confirmed	2020-01-31	144	0.309091

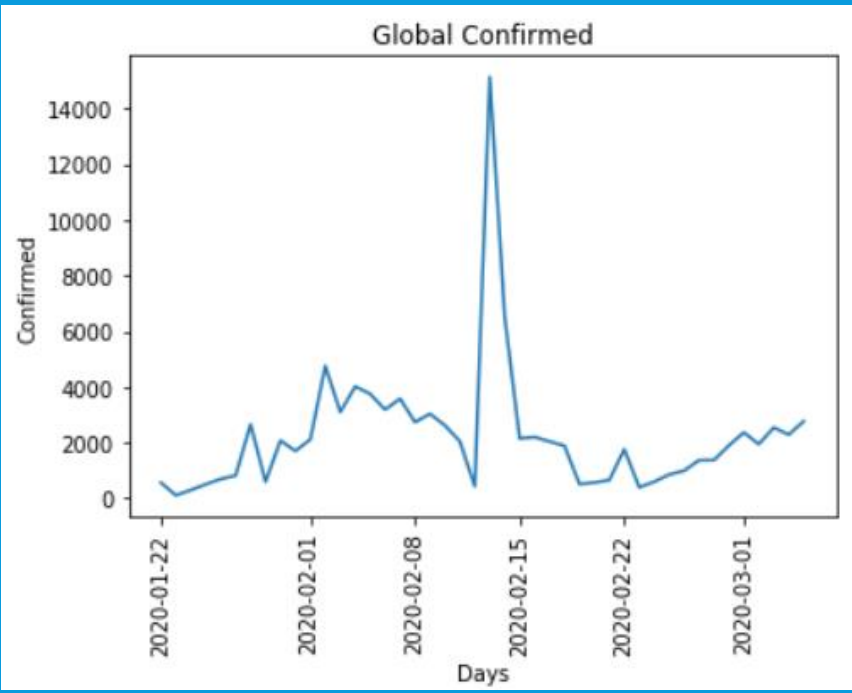
VISUALIZATION

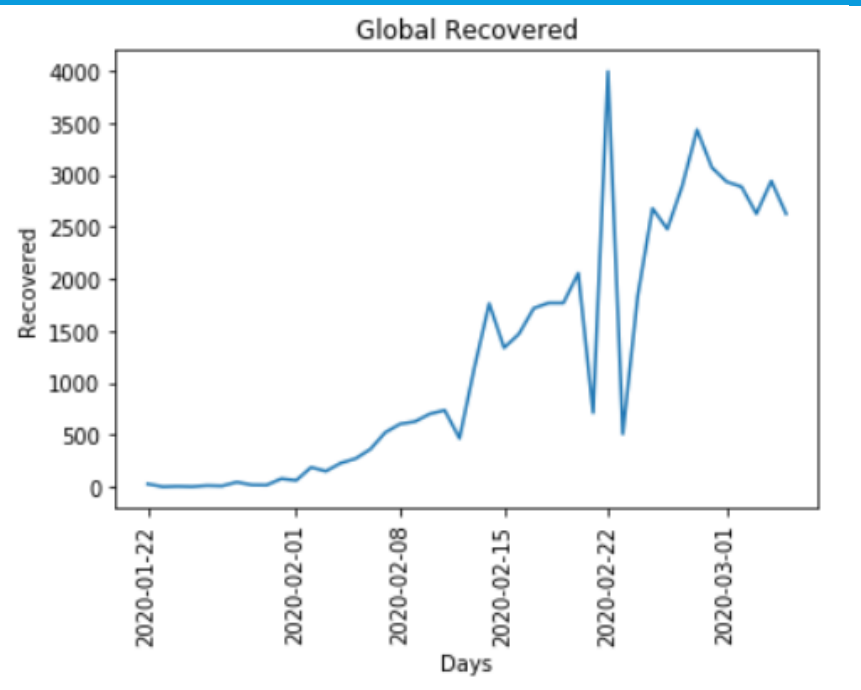


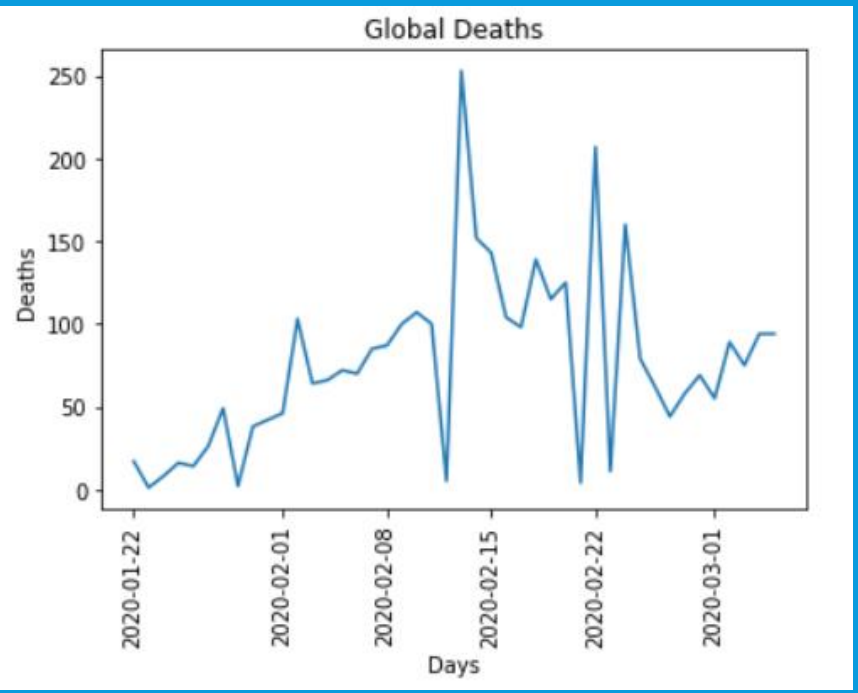
VISUALIZATION



VISUALIZATION

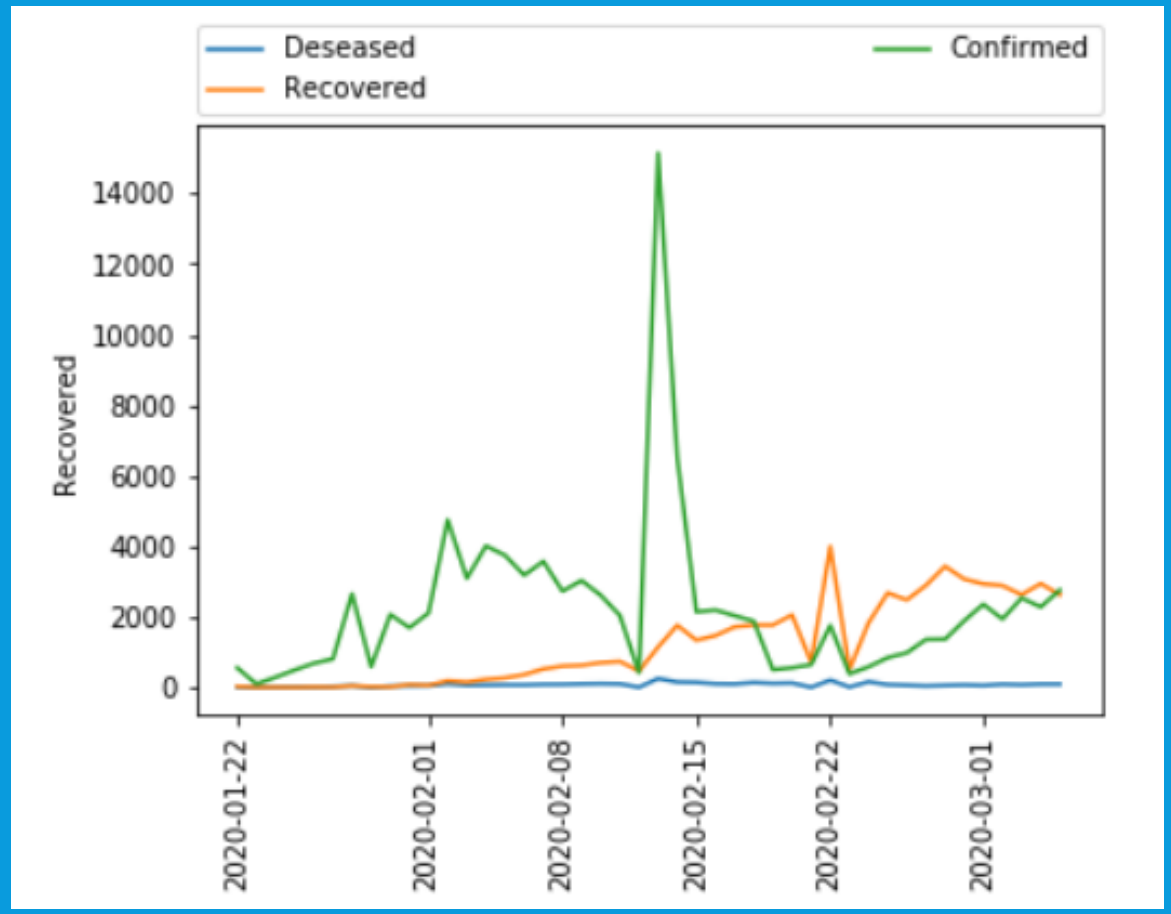




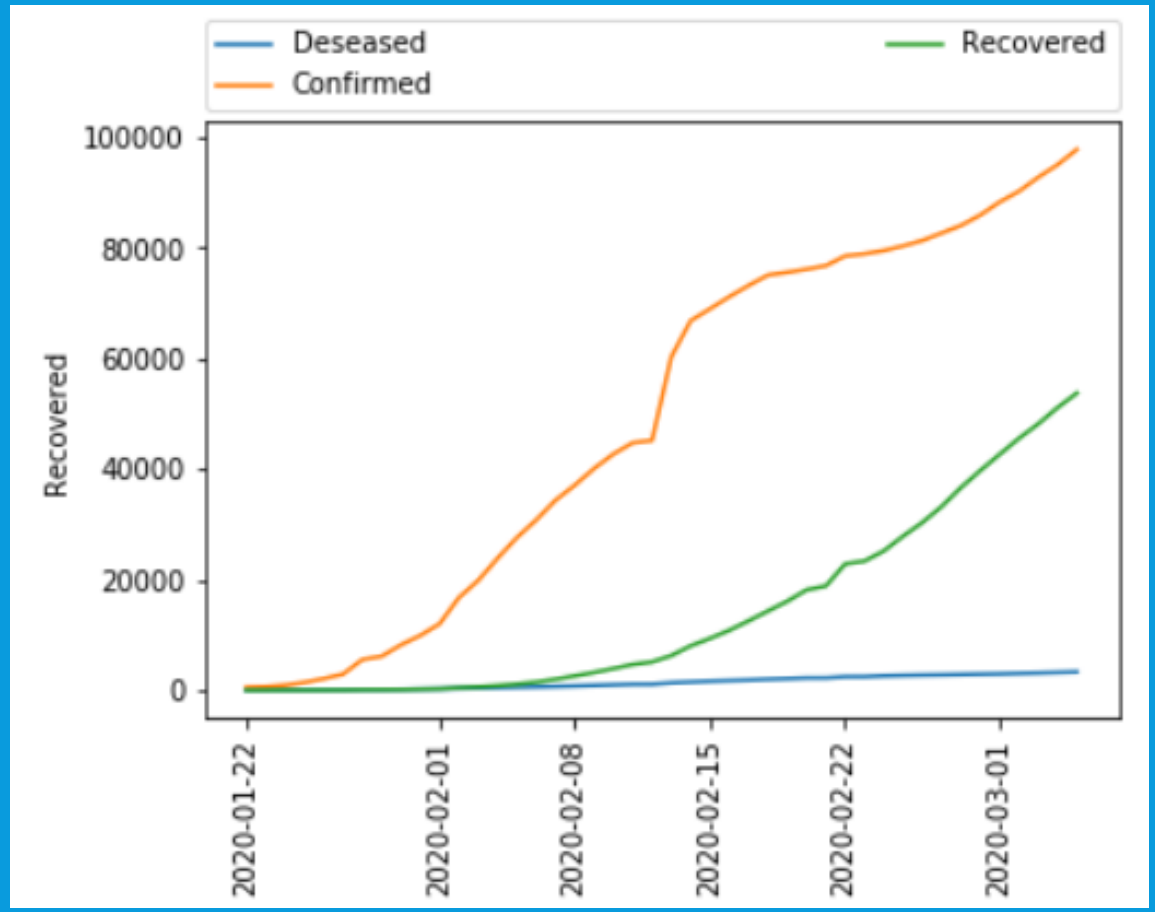


VISUALIZATION

Global Day by Day



Global Cumulative



VISUALIZATION

QUESTIONS & DATA

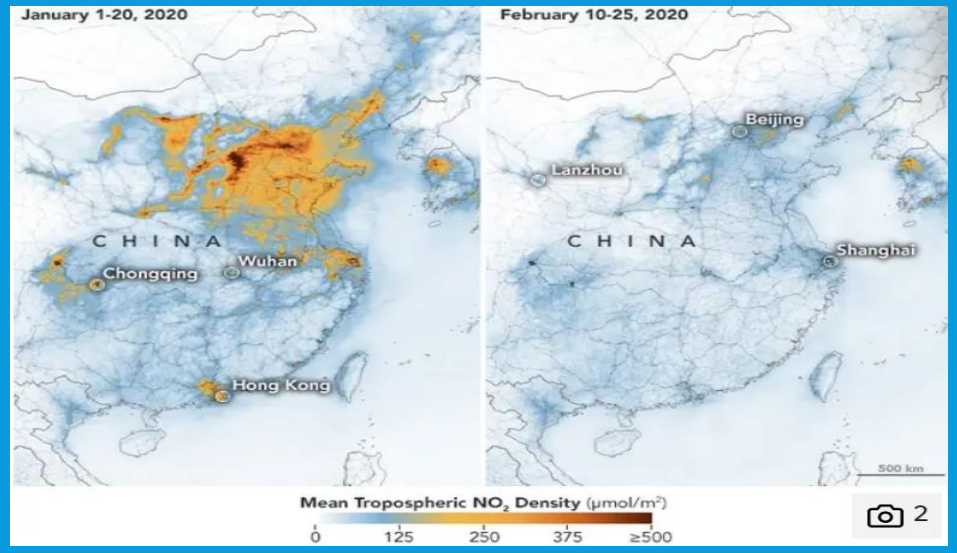
- We were able to successfully answer the questions posed in our scope.
- From our results, we conclude that **COVID-19 death rate** represent on average **3.5%** of the total confirmed cases worldwide. Additionally we found that, even though the confirmed cases have steadily increased around the world, the **increase in deaths** remain relatively **stable**.
- The number of confirmed cases in China remains predominant compared to the cases reported in the rest of the world.

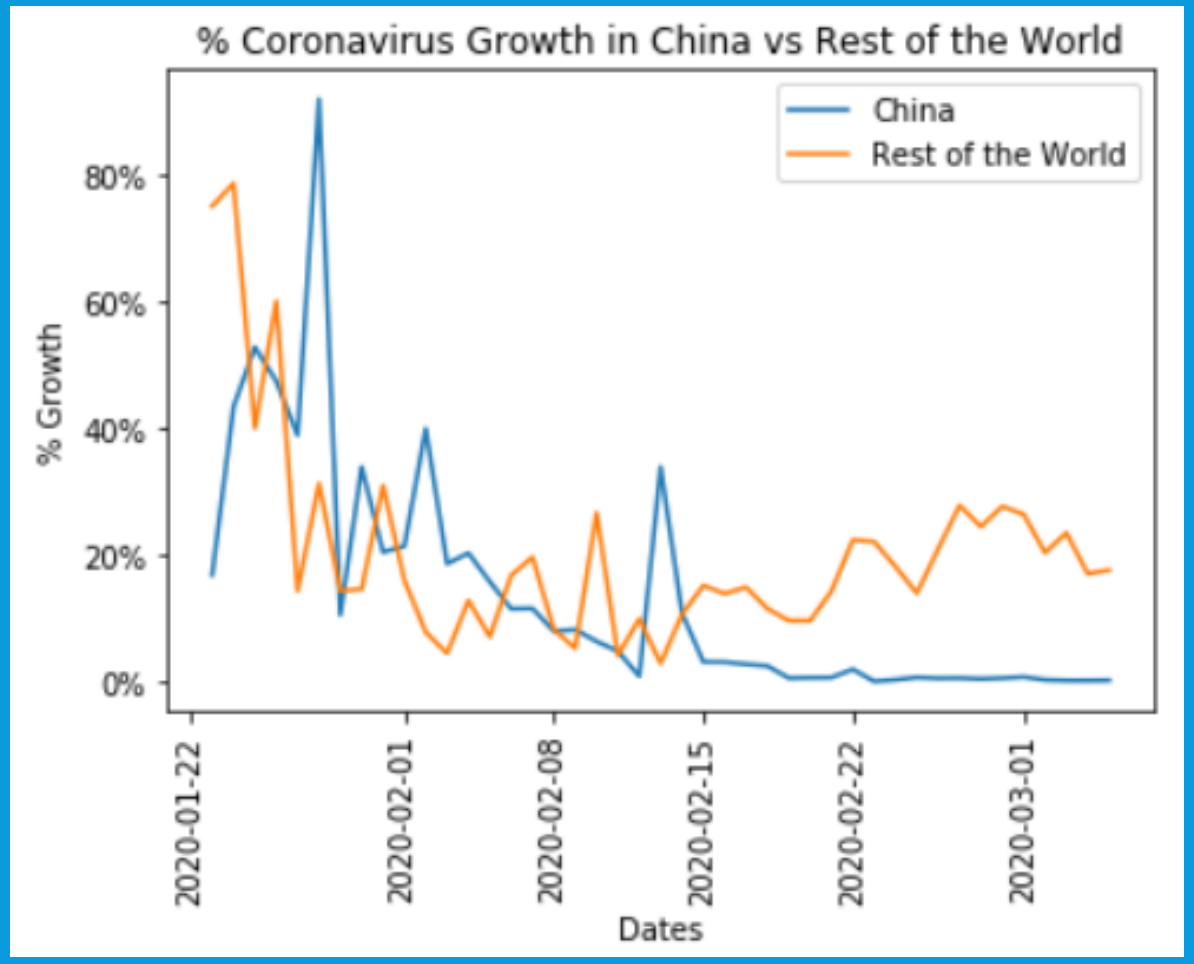
DATA ANALYSIS

- We downloaded the **data** from the reports' GitHub page, processed it, removed data that we did not need, transformed the tables to the convenience of each chart, we also **added the fields and variables needed for analysis**, investigated libraries, and used this information to plot. Once we had the charts, we were able to answer our questions.

DISCUSSION

- As per the data shown in plot "**% Coronavirus growth vs the rest of the world**", the trend of cases in China shows a clear **decreasing tendency** that starts in mid February and continues until march, while the cases confirmed in rest of the world, show an increase of approximately 25% in the same time frame (2 weeks).





POSTMORTEM

- The data **updates everyday.**
- **ApiKey** for GoogleMaps Api.
- Join the notebooks into one **functional Notebook**.
- **Cross-regional data**in order to analyze the spreading of the virus in small environments.

Q&A

Thank you!