NYU TANDON SCHOOL OF ENGINEERING

Covid-19 Economic Impact Analysis

Presented:
Tirupal Rao Ravilla, Yufei Lin, Tinghao Ruan, Huiyi Teng, Kaihan Xia

NYU TANDON SCHOOL OF ENGINEERING

Agenda

01. **Introduction**: comparing between the UK, USA and China
02. **Data Source**: the data we used to analyze
03. **Sanity checks** and **compression** (data preparation)
04. Methods of **Clustering** and Classification
05. Analysis and Plots: **Implication on Stock Market**
06. **Predication** : Stock Value
07. **Predication** : Unemployee Rate

Introduction

Motivation:

- Investigate the effect of the COVID-19 pandemic
- Analyze its impact on different Economic aspects
- Predict the unemployment rate under the influence of this pandemic

Task Definition:

- Comparison among three countries:
  - China
  - United Kingdom
  - United States
- COVID-19 related data
- Stock market indicator
- Unemployment rate
- Analysis of Economic Impact
- Prediction

Data Source

Covid-19 Data:

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

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Center for Systems Science and Engineering

Economic Data:

https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge (Historical Unemployment Rates)

https://tradingeconomics.com (Historical Stock Value)

kaggle

TRADING ECONOMICS

Sanity Check

- Filtering
- Handling missing values
- Preprocessing data: outlier and normalization

Country/Region	Afghanistan	Algeria	Argentina	Armenia	Australia
0	0.0	0.0	0.0	0.0	0.000000
1	0.0	0.0	0.0	0.0	0.000000
2	0.0	0.0	0.0	0.0	0.000000
3	0.0	0.0	0.0	0.0	0.000000
4	0.0	0.0	0.0	0.0	0.089712

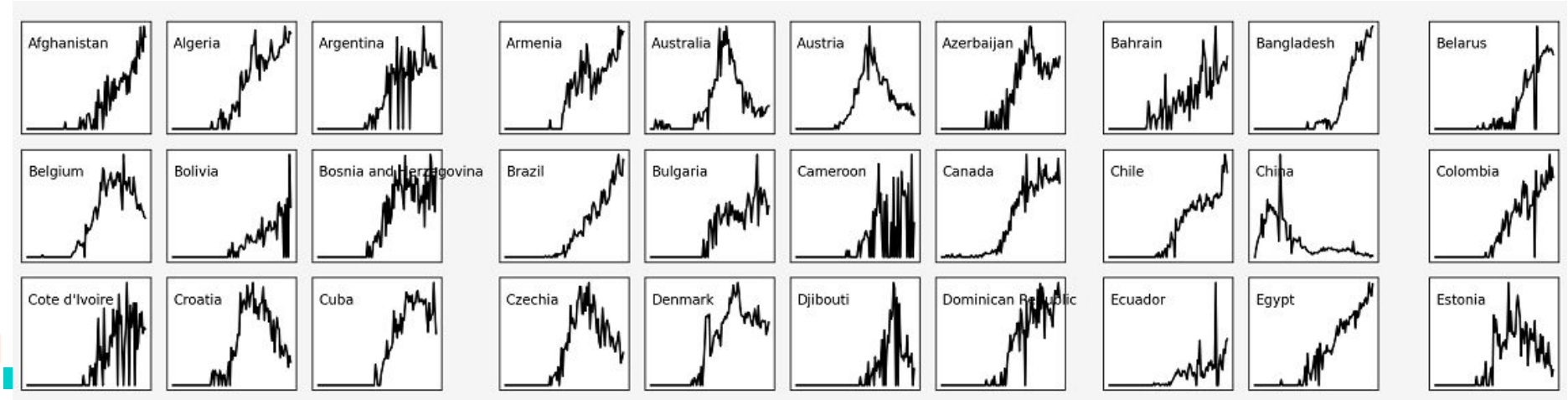
5 rows × 89 columns

```
Province/State     object
Country/Region     object
Lat               float64
Long              float64
1/22/20             int64
                    ...
4/30/20             int64
5/1/20              int64
5/2/20              int64
5/3/20              int64
5/4/20              int64
Length: 108, dtype: object
```

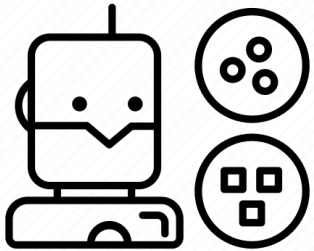
Data Compression

We can get the trend of every country like the graph.

- Feature are the difference of pattern (distance)
- Using Bray-curtis distance to define the feature
- Using PCA to compress 89 features into 2 features
- PCA Variance:
  [1.20720431 0.37731189]



Clustering



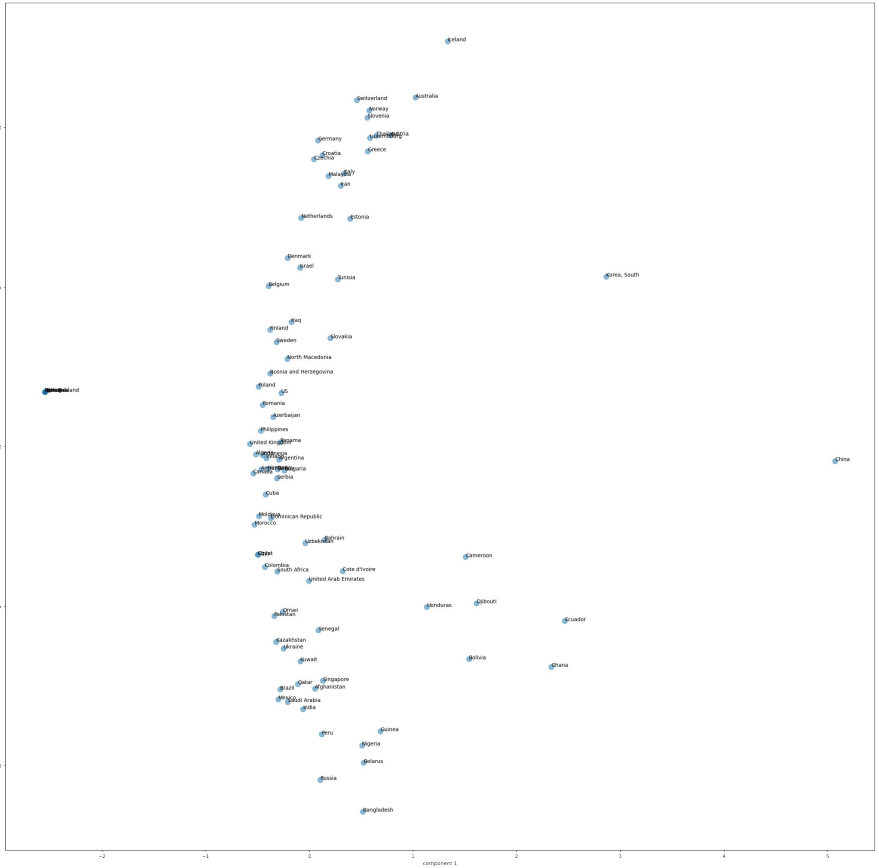
- Clustering **89** countries into **5** clusters by using **K-means++**
- After getting the new dataframe using cluster and two features, We can see the specific cluster of each country

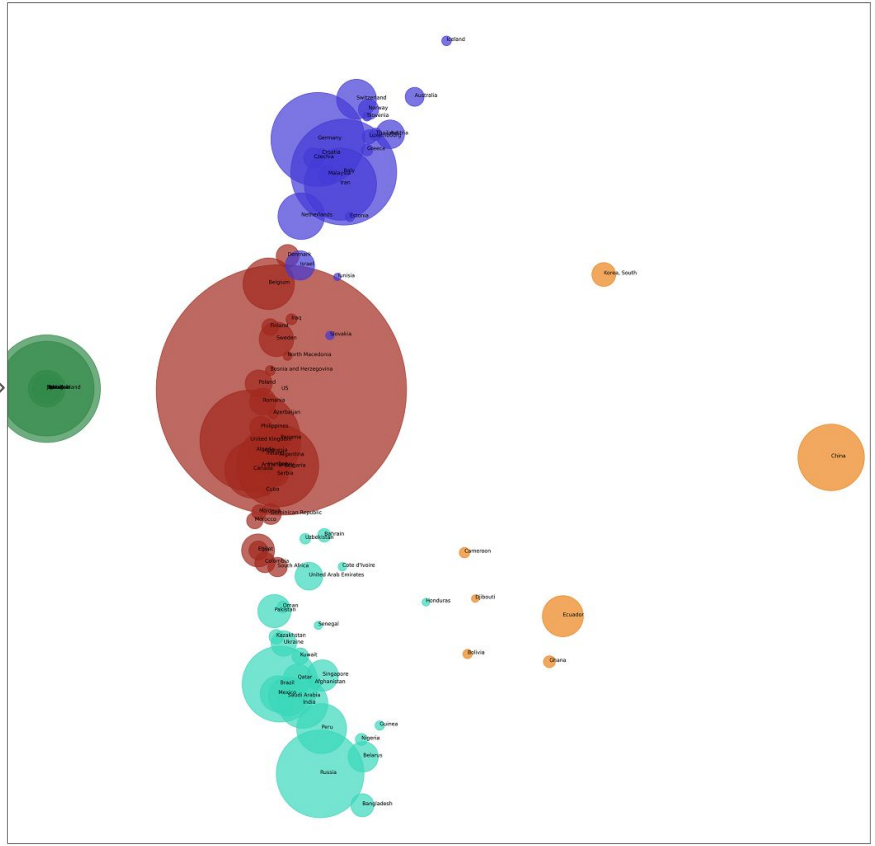
	Cluster	Country	Color	Cases	PC1	PC2
0	4	Afghanistan	#3bd9bc	2894.0	0.053722	-0.757719
1	0	Algeria	#a32a1f	4648.0	-0.515906	-0.023914
2	0	Argentina	#a32a1f	4887.0	-0.292753	-0.040125
3	0	Armenia	#a32a1f	2507.0	-0.464397	-0.071726
4	2	Australia	#453bd9	6847.0	1.024037	1.093765
...	...	...	...	...	...	...
84	0	US	#a32a1f	1180374.0	-0.271424	0.168282
85	4	Ukraine	#3bd9bc	12331.0	-0.248670	-0.632788
86	4	United Arab Emirates	#3bd9bc	14730.0	-0.003275	-0.420277
87	0	United Kingdom	#a32a1f	191832.0	-0.573301	0.008152
88	4	Uzbekistan	#3bd9bc	2189.0	-0.039467	-0.301999

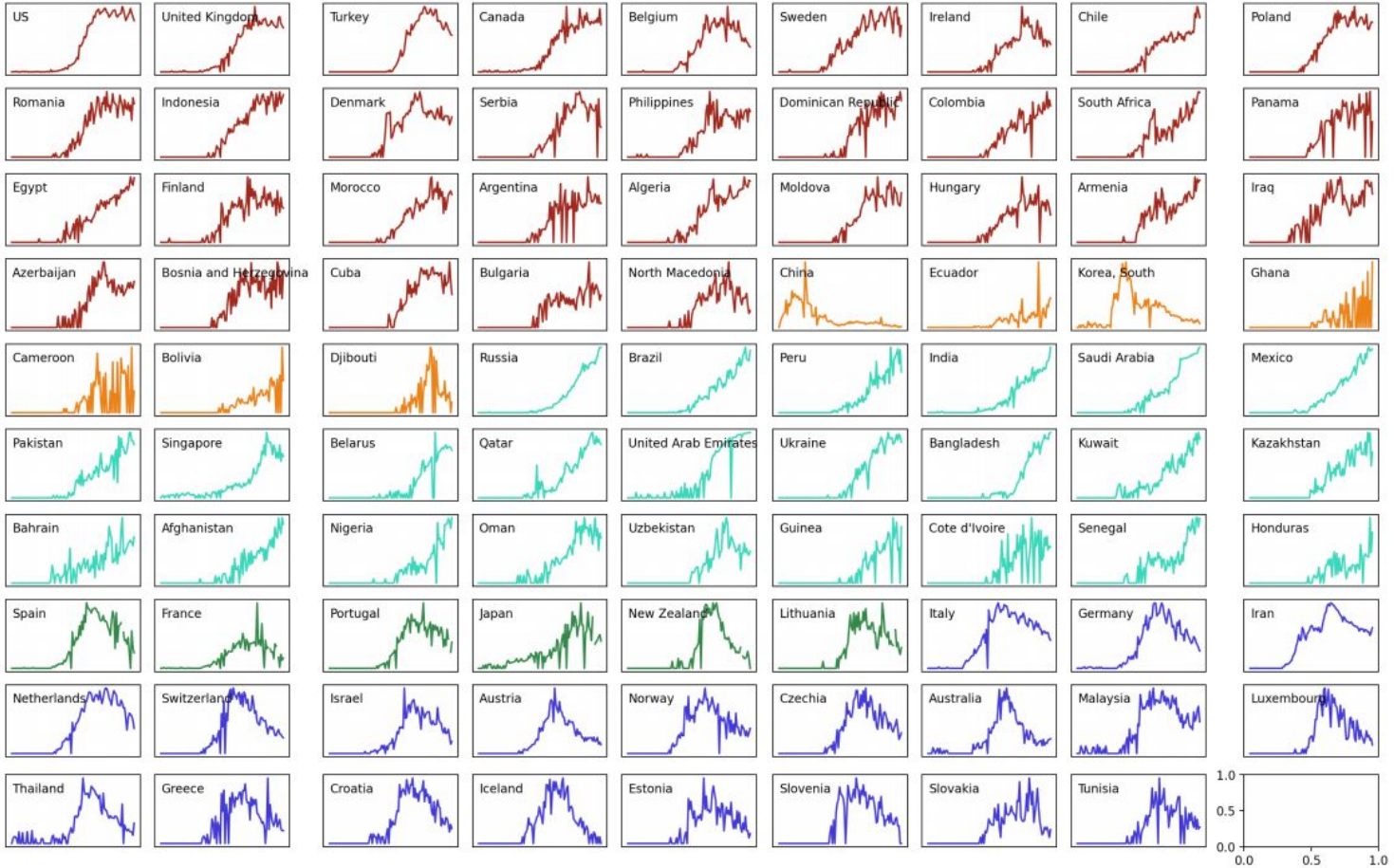
89 rows × 6 columns

NYU TANDON SCHOOL OF ENGINEERING

PCA Plots for Clustering

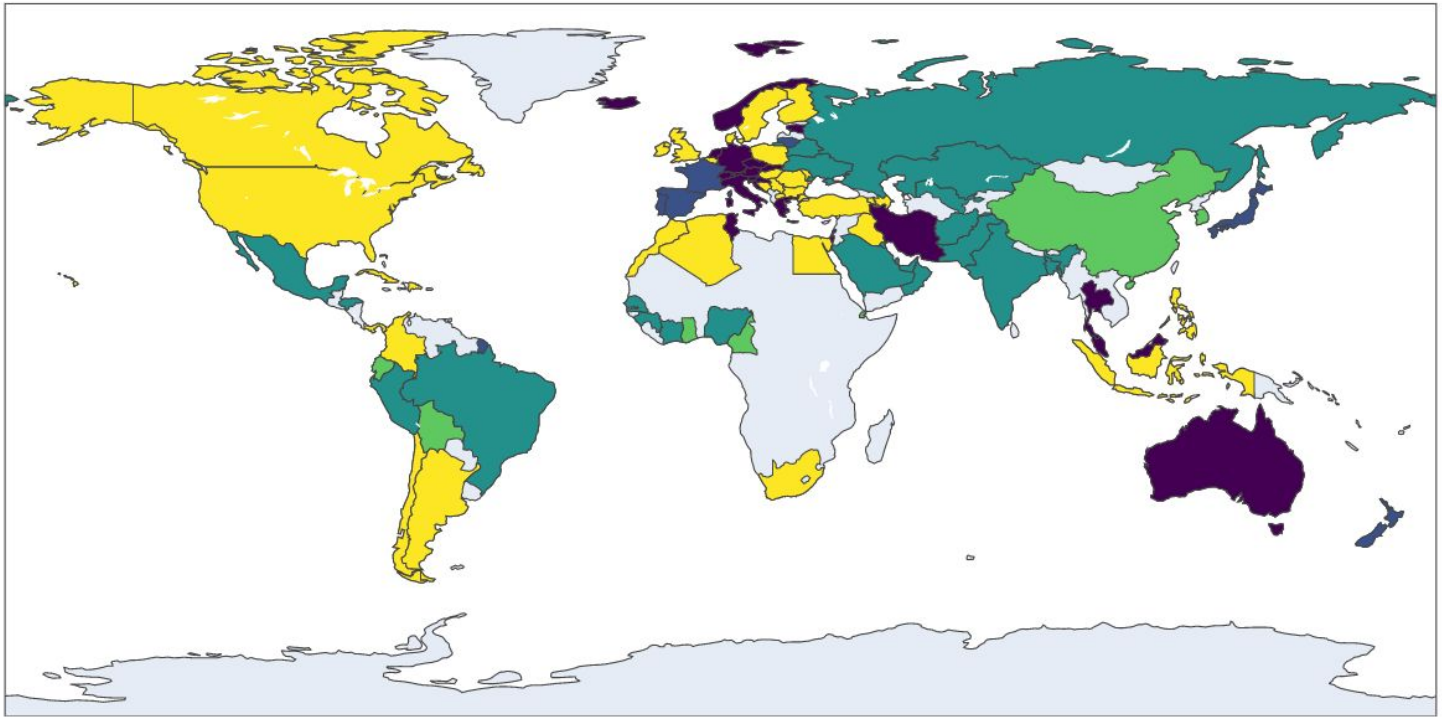




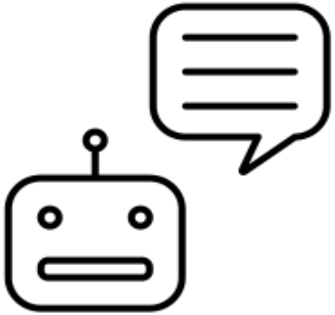

US
United Kingdom
Turkey
Canada
Belgium
Sweden
Ireland
Chile
Poland
Romania
Indonesia
Denmark
Serbia
Philippines
Dominican Republic
Colombia
South Africa
Panama
Egypt
Finland
Morocco
Argentina
Algeria
Moldova
Hungary
Armenia
Iraq
Azerbaijan
Bosnia and Herzegovina
Cuba
Bulgaria
North Macedonia
China
Ecuador
Korea, South
Ghana
Cameroon
Bolivia
Djibouti
Russia
Brazil
Peru
India
Saudi Arabia
Mexico
Pakistan
Singapore
Belarus
Qatar
United Arab Emirates
Ukraine
Bangladesh
Kuwait
Kazakhstan
Bahrain
Afghanistan
Nigeria
Oman
Uzbekistan
Guinea
Cote d'Ivoire
Senegal
Honduras
Spain
France
Portugal
Japan
New Zealand
Lithuania
Italy
Germany
Iran
Netherlands
Switzerland
Israel
Austria
Norway
Czechia
Australia
Malaysia
Luxembourg
Thailand
Greece
Croatia
Iceland
Estonia
Slovenia
Slovakia
Tunisia
1.0
0.5
0.0
0.0
0.5
1.0

NYU TANDON SCHOOL OF ENGINEERING

Covid-19 Cluster World Map



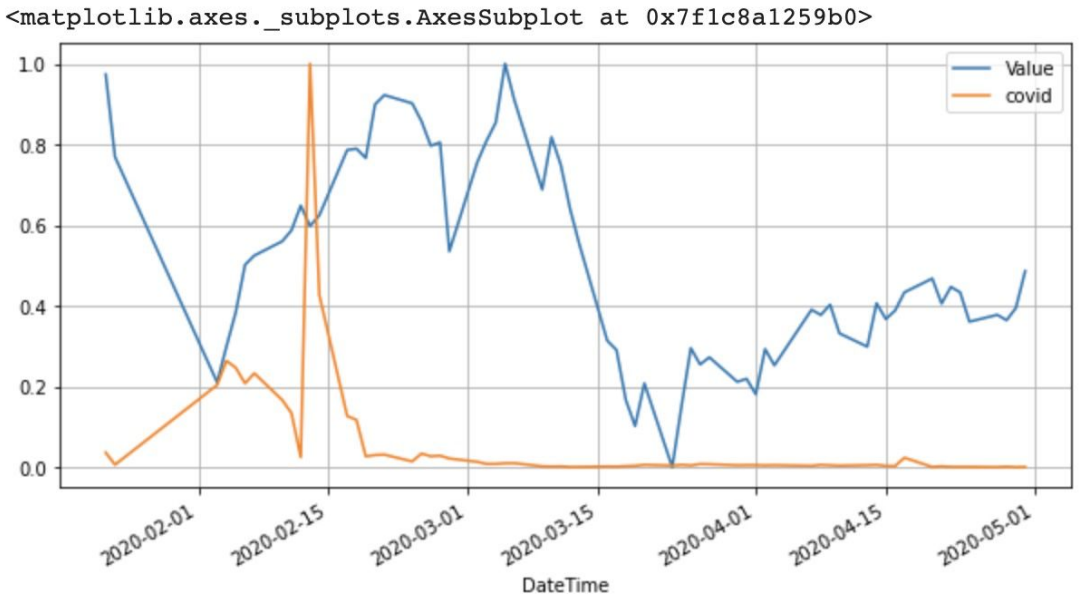
Classification

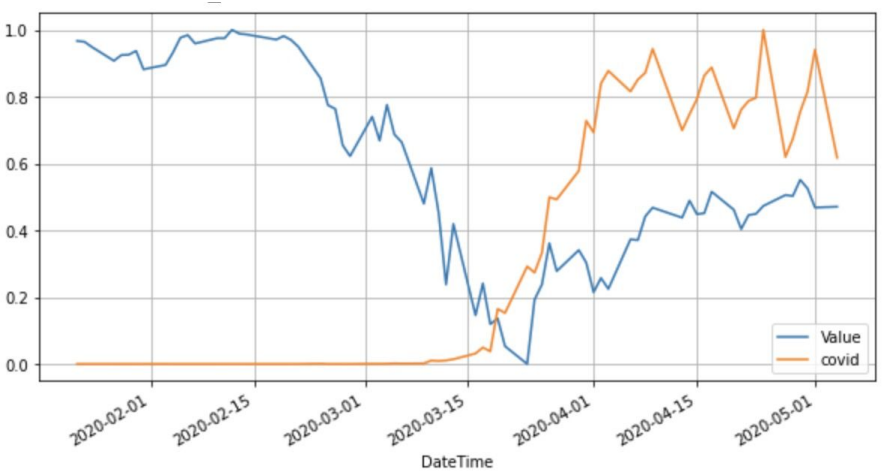


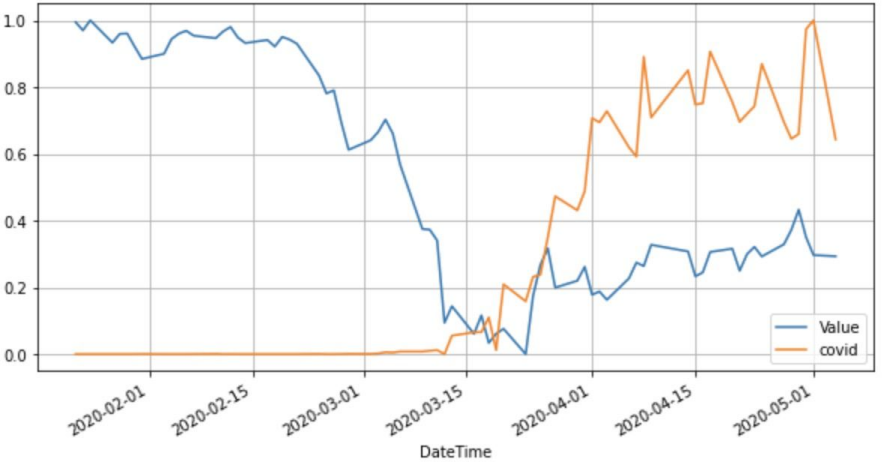
- Using **SVC** to do classification and Defining two PCA features as x and clusters as y
- The accuracy is : **97.22%**
- Then using **KNN** to do Classification and selecting 3 neighbors
- The accuracy is : **94.44%**
- In this case, **SVM** becomes a more accurate model.

Implication on Stock Market

We combined the graph of Stock Price with the graph of Covid-19 Cases so that we can see the correlation between them.







- The stock market keeps decreasing when the number of Covid-19 cases increases, which shows a “bear market”.
- Then, when Covid-19 curve starts decreasing or being flattened, the stock market curve will start increasing, getting the stock market back to a “bull market”.
- So there is a negative correlation between the number of cases and the economy, which means that the pandemic can “hurt” the economy.

NYU TANDON SCHOOL OF ENGINEERING

Prediction of Stock Value

We used **Stock Value** and **Covid-19 case numbers** as two features to predict stock price in the following 10 days after 05/04/2020

```
Including the COVID cases count
Linear Regression
0.6859399986851307
[5803.86036544 5914.07859486 5968.33139149 5980.12794226 5957.92318297
 6015.20675171 6140.56058926 6147.58385183 6057.96915731 5860.46176818]
Ridge Regression:
r2 score is = 0.6849809686497884
0.6849809686497884
[5801.07719691 5909.68016869 5962.98713917 5972.90808734 5953.35694937
 6010.6567771  6134.35399743 6136.98528732 6048.05150787 5857.74039798]
Lasso Regression:
r2 score is = 0.685891796370531
0.685891796370531
[5803.22611145 5913.24948028 5967.38234567 5978.88678652 5957.09218203
 6014.41080308 6139.57176843 6145.89934299 6056.34233844 5859.86943461]
Support Vector Regression:
r2 score is = 0.743338416577092
0.743338416577092
[5801.8701086  5830.01462487 5847.03001871 5860.43840365 5839.42235766
 5855.32695997 5910.36251842 5902.2940079  5883.8126865  5804.52295907]
```

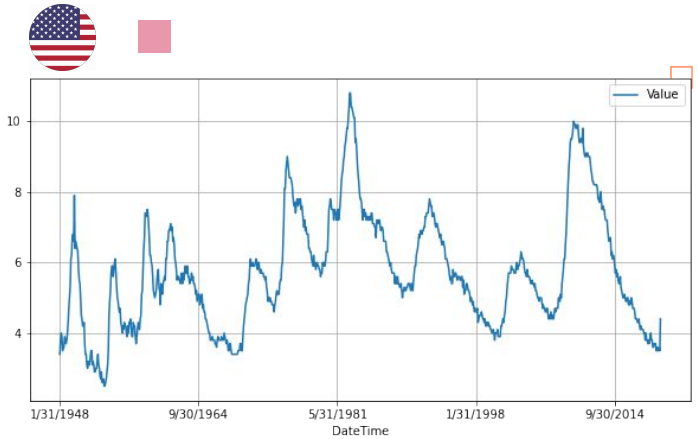
```
Not inclusing covid count
Linear Regression:
0.6711618482987114
[5635.45256212 5717.69646279 5753.22125878 5706.01985961 5766.0274464
 5836.91838888 5936.39162523 5800.56226951 5712.89255594 5707.00348651]
Ridge
r2 score is = 0.6696723100209541
0.6696723100209541
[5638.43259625 5720.19139822 5755.50665852 5708.58366708 5768.23731148
 5838.71011834 5937.59663229 5802.56843805 5715.41582623 5709.56149226]
Lasso:
r2 score is = 0.6710299481435071
0.6710299481435071
[5635.71927562 5717.91975989 5753.42580241 5706.24932077 5766.22522966
 5837.07874894 5936.49947354 5800.7418219  5713.11838902 5707.23242841]
SVR
r2 score is = 0.6957319816690711
0.6957319816690711
[5721.44619647 5744.2184805  5756.98981401 5740.41084738 5762.03451376
 5794.22317921 5851.6212633  5776.81094948 5742.62870456 5740.72419253]
```

- Using SVR method gives us the highest R^2 compared to other 4 methods, which is 0.7433.
- After comparing the R^2 of predictions **with** Covid-19 cases and **without** the cases, we can see that R^2 of prediction with the cases is higher and we can conclude that there is a influence of Covid-19 on the stock market.

Prediction of Unemployment Rate

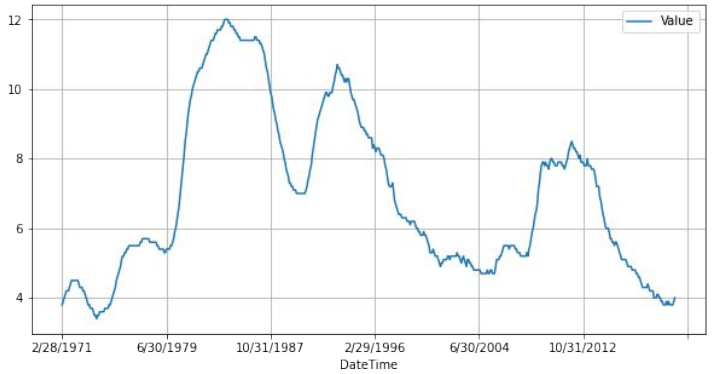
Using historical data, we predict unemployment rate in the following 10 months for United States and United Kingdom, 5 months for China.

```
         Country           Category  ... HistoricalDataSymbol           LastUpdate
0  United States  Unemployment Rate  ...               USURTOT  2012-02-23T11:41:00
1  United States  Unemployment Rate  ...               USURTOT  2012-02-23T11:41:00
2  United States  Unemployment Rate  ...               USURTOT  2012-02-23T11:41:00
3  United States  Unemployment Rate  ...               USURTOT  2012-02-23T11:41:00
4  United States  Unemployment Rate  ...               USURTOT  2012-02-23T11:41:00

[5 rows x 7 columns]
0.7205453277782565
[4.1804584  4.1804584  4.1804584  4.02528328 4.10287084 4.02528328
 4.02528328 4.10287084 4.02528328 4.72357132]
Ridge
r2 score is = 0.7204817440835063
0.7204817440835063
[4.18115035 4.18115035 4.18115035 4.02604363 4.10359699 4.02604363
 4.02604363 4.10359699 4.02604363 4.72402386]
Lasso:
r2 score is = 0.6542115008470094
0.6542115008470094
[4.55182531 4.55182531 4.55182531 4.43336083 4.49259307 4.43336083
 4.43336083 4.49259307 4.43336083 4.96645102]
SVR
r2 score is = 0.7235918186238486
0.7235918186238486
[3.72449464 3.72449464 3.72449464 3.62888601 3.6709389  3.62888601
 3.62888601 3.6709389  3.62888601 4.3056083 ]
```



- 4 Algorithms are used: Linear Regression, Ridge, Lasso, and SVR
- SVR performs better over all others with 72.4 r^2 score.

```
Regression Analysis forecast of Unemployment rate  for United Kingdom
          Country           Category  ... HistoricalDataSymbol           LastUpdate
0  United Kingdom  Unemployment Rate  ...              UKUEILOR  2015-12-16T10:10:00
1  United Kingdom  Unemployment Rate  ...              UKUEILOR  2015-12-16T10:10:00
2  United Kingdom  Unemployment Rate  ...              UKUEILOR  2015-12-16T10:10:00
3  United Kingdom  Unemployment Rate  ...              UKUEILOR  2015-12-16T10:10:00
4  United Kingdom  Unemployment Rate  ...              UKUEILOR  2015-12-16T10:10:00

[5 rows x 7 columns]
0.9350197861315745
[3.99966332 4.09419263 3.99966332 4.09419263 3.99966332 3.99966332
 3.99966332 3.99966332 4.09419263 4.18872194]
Ridge
r2 score is = 0.9350036716616507
0.9350036716616507
[4.00170083 4.09616583 4.00170083 4.09616583 4.00170083 4.00170083
 4.00170083 4.00170083 4.09616583 4.19063082]
Lasso:
r2 score is = 0.9140801471292809
0.9140801471292809
[4.41696871 4.49832564 4.41696871 4.49832564 4.41696871 4.41696871
 4.41696871 4.41696871 4.49832564 4.57968257]
SVR
r2 score is = 0.9411990963170547
0.9411990963170547
[3.84648533 3.92190864 3.84648533 3.92190864 3.84648533 3.84648533
 3.84648533 3.84648533 3.92190864 4.00054389]
```



- 4 Algorithms are used: Linear Regression, Ridge, Lasso, and SVR
- SVR performs better over all others with 94.1 r^2 score.

```
Regression Analysis forecast of Unemployment rate  for China
  Country           Category  ... HistoricalDataSymbol  LastUpdate
0   China  Unemployment Rate  ...             CNUERATE   6/27/2011
1   China  Unemployment Rate  ...             CNUERATE   6/27/2011
2   China  Unemployment Rate  ...             CNUERATE   6/27/2011
3   China  Unemployment Rate  ...             CNUERATE   6/27/2011
4   China  Unemployment Rate  ...             CNUERATE   6/27/2011

[5 rows x 7 columns]
0.844989778548143
[5.18481317 5.28368413 5.3825551  6.27239378 5.97578088]
Ridge
r2 score is = 0.8457921306151513
0.8457921306151512
[5.18013374 5.27836979 5.37660584 6.26073028 5.96602213]
Lasso:
r2 score is = 0.8479539266937678
0.8479539266937678
[5.16365009 5.2596496  5.3556491  6.21964466 5.93164614]
SVR
r2 score is = 0.9337285549821069
0.9337285549821069
[5.17292196 5.24218446 5.32085733 8.14023506 6.69738112]
```

- 4 Algorithms are used: Linear Regression, Ridge, Lasso, and SVR
- SVR performs better over all others with 93.4 r^2 score.

Questions?

NYU TANDON SCHOOL OF ENGINEERING

Thank You for Listening!

Group 6:
Tirupal Rao Ravilla
Yufei Lin
Tinghao Ruan
Huiyi Teng
Kaihan Xia