COVID-19 Data Analysis
for
Capstone Project/ IBM Data Scientist

Pradeep Sharma (夏普迪), PhD (Nanotechnology)

11th April, Year 2020

Introduction

- As of 11th April Year 2020, COVID-19 pandemic is affecting many countries around the world. It is interesting to investigate different factors driving COVID-19, and predict number of COVID cases in near future. Therefore, this topic was selected for Capstone project for IBM Data Science Certification.

- **Process**:

  - Collect information on known/potential factors influencing COVID-19. Collect data from different

  - resources. Prepare clean database for COVID-19 analysis. Install/deploy needed software/scripts.

  - Perform data analysis. Conduct discussion and review to make project report.

- **Goals of Capstone Proj**ect:

  1. Descriptive, Predictive and Prescriptive analysis for COVID-19 cases.

  2. Apply different machine learning algorithms to identify best R-sq, and predictors.

  3. Identify key factors that drive COVID-19.

Executive Summary (Data Analysis)

- Starting from 15 March 2020, most countries are showing exponential rise in number of COVID-19 reported cases.

- **Key drivers of COVID-19 spread/elimination**:

(a) Tourism/Travel score, (b) Diagnostic Tests/Million of Population,

(c) Health Care and Awareness Score, and (d) Human Freedom Score.

- **Temperature** and **Humidity** can possibly help improve immune system of people.

But, as per analysis, it is estimated to reduce # of COVID cases about 5-10% only.

Executive Summary (Recommendations)

- Key **recommendations for eliminating, or reducing** COVID cases.

(a) Minimum 2000 diagnostic tests/million population in a country.

(b) Improved social distancing, and reduce travels.

(c) Improved health care and awareness for people.

Rich Economic Score ≠ Best Health Care and Awareness.

Note: Above recommendation are based on data analysis in this project report.

On top of that all citizens must follow guidance of CDC, and WHO.

Weblinks for IBM Watson Studio & Github

IBM Webpage

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/8e308421-4c8c-4dc7-8691-29630823f91d/view?access_token=f24a085479eb3b5fda37a3eb9dffb8cfb08cd908cb247610154596c1bcfd3bf0

Github Webpage

https://github.com/PradeepTaiwan1980/Coursera-capstone-project

- PDF format project report uploaded to Github page

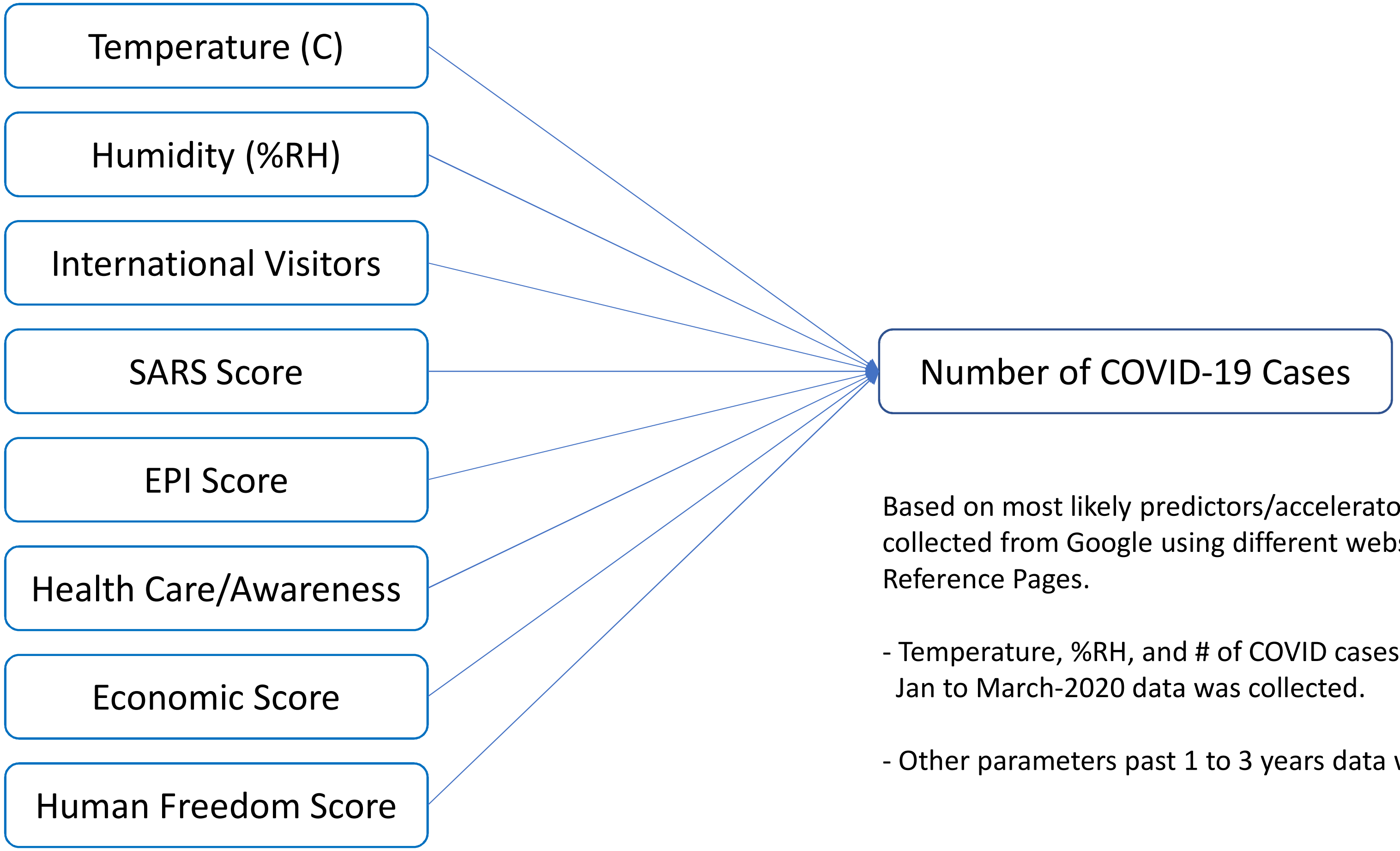
Table of Contents

- Introduction
- Executive Summary (Analysis and Recommendation)
- Metadata for Analysis
- Data Preparation
- COVID-19 Scenario in Different Countries
- Correlation and Regression Analysis using Different Algorithms
- Predictive and Prescriptive Analysis using Machine Learning Algorithms
- Conclusions
- References
- Disclaimer for Project Report

Metadata for ‘COVID-19 Data’ Data Frame

- Countries = List of countries in world
- COVID-19 cases = Total number of COVID-19 cases by 11 April 2020
- Tests per Million = # of COVID-19 diagnosis tests per million of population
- Temperature = Maximum temperature in Celsius units of a country in month of March/April-2020
- Humidity (%RH) = Average humidity of a country in month of March/April-2020
- Tourism score = # of million passengers arriving at airports in different countries, or visitors
- SARS score = Measure of impact of SARS on individual country
- EPI score = Environmental performance index of individual country
- Healthcare score (HCS) = Global rating of healthcare facilities
- Human freedom score (HFS) = Global rating of people making own choices, social interactions, and people friendly governance.
- Economic Score = GDP per capita/Cost of living in individual country
- Tourism_X = Multiplication of Tourism score index with X score of a country.
  X = Economic, Temperature, HFS, and HCS

Data Preparation for COVID-19 Analysis



Temperature (C)

Humidity (%RH)

International Visitors

SARS Score

EPI Score

Health Care/Awareness

Economic Score

Human Freedom Score

Number of COVID-19 Cases

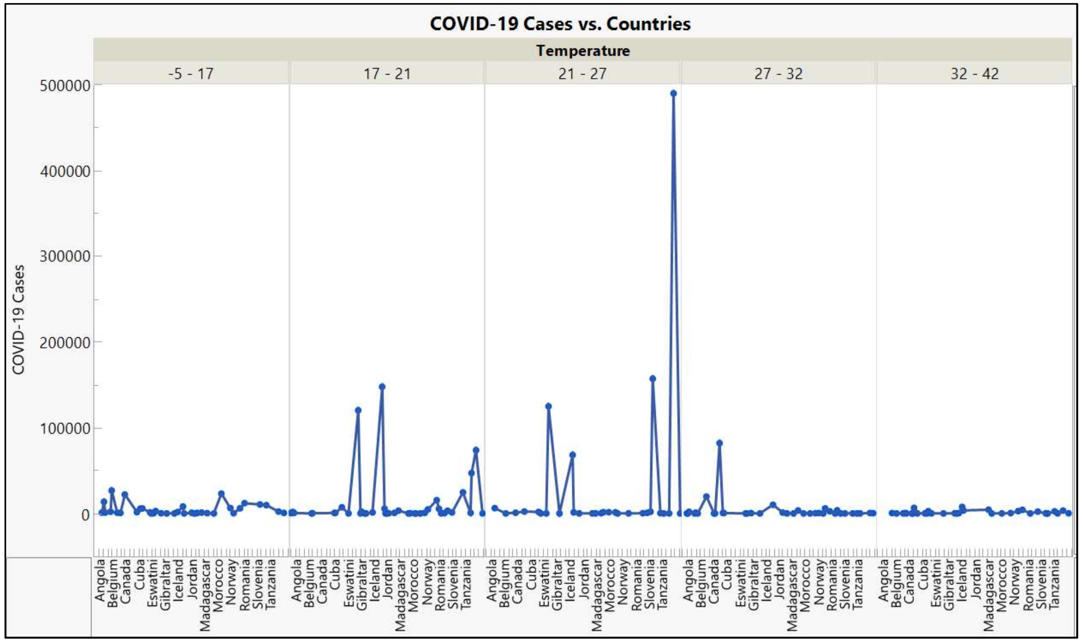
Based on most likely predictors/accelerators, data was collected from Google using different websites listed in Reference Pages.

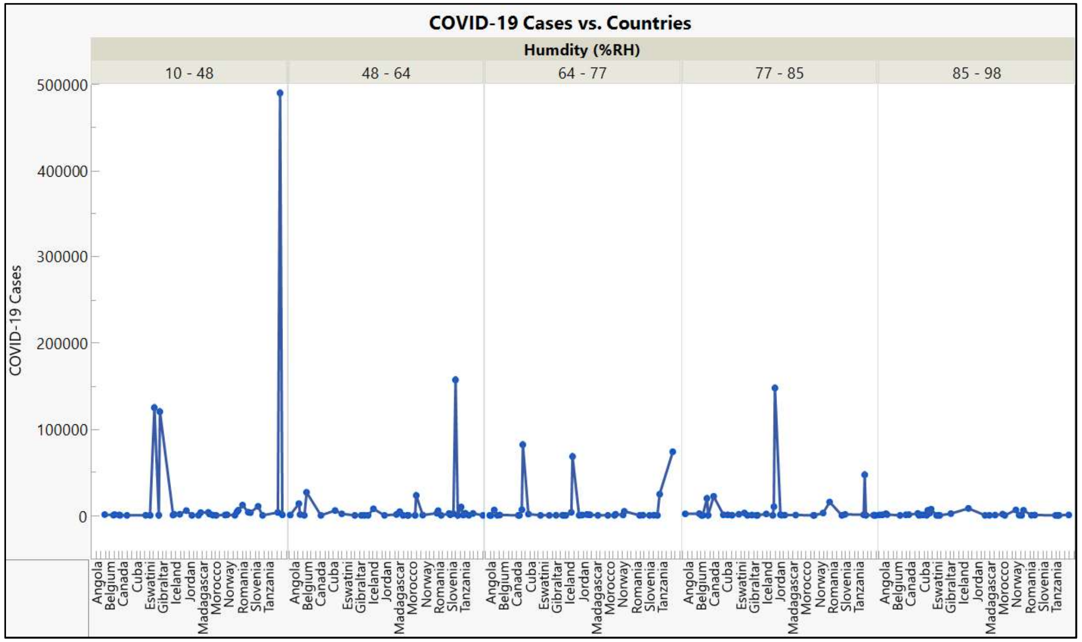
- Temperature, %RH, and # of COVID cases:
  Jan to March-2020 data was collected.

- Other parameters past 1 to 3 years data was collected.

Data Analysis using JMP Pro 14 Software

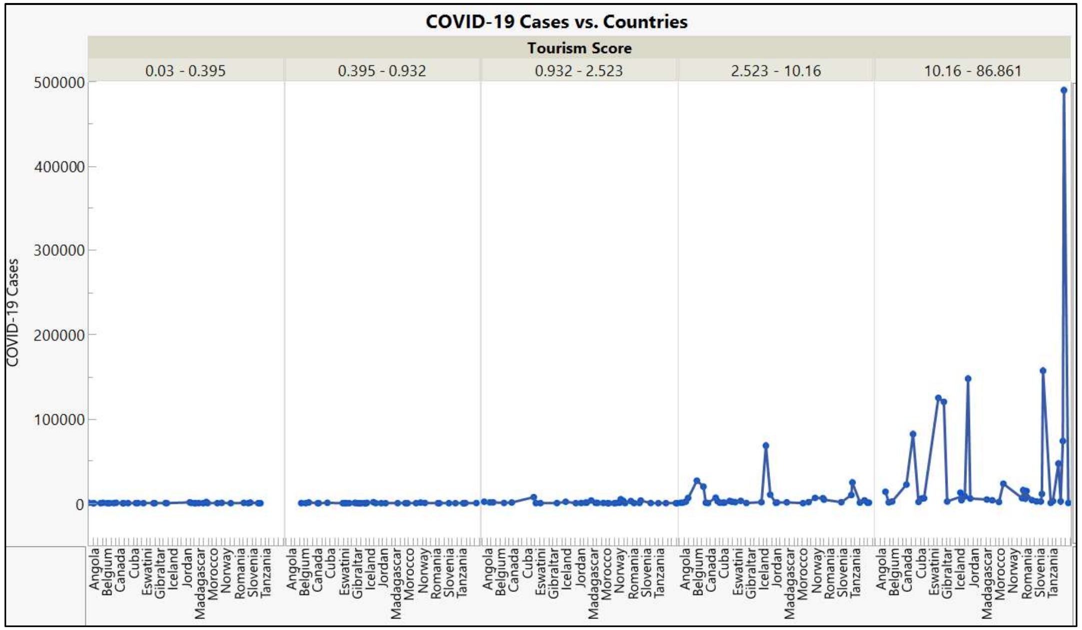
COVID-19 Cases in Different Countries

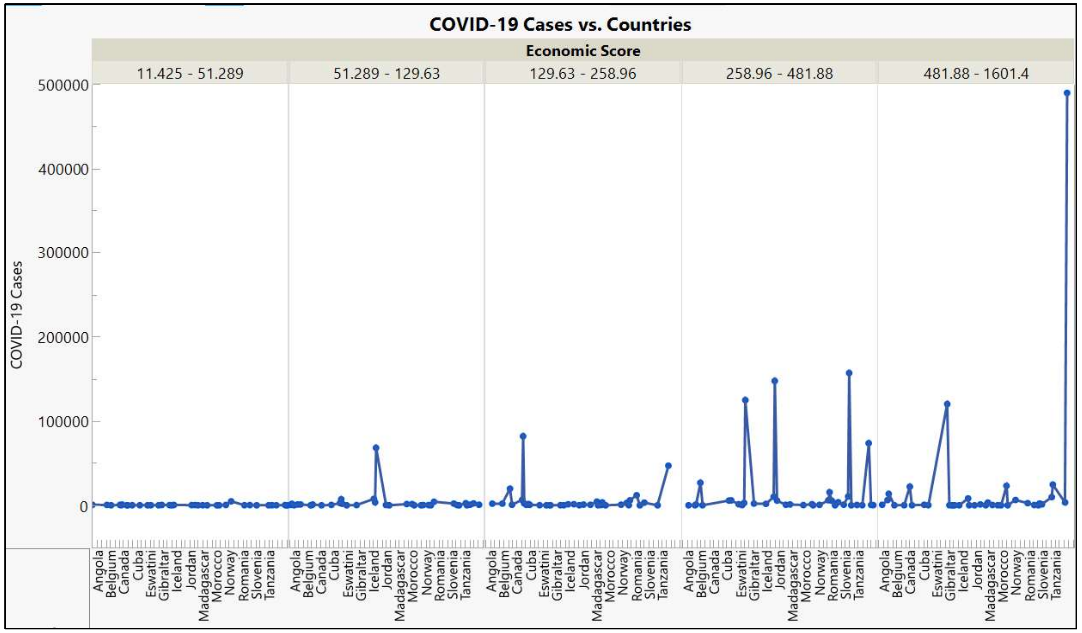




- # of COVID-19 cases have no obvious association with Temperature/Humidity of a country

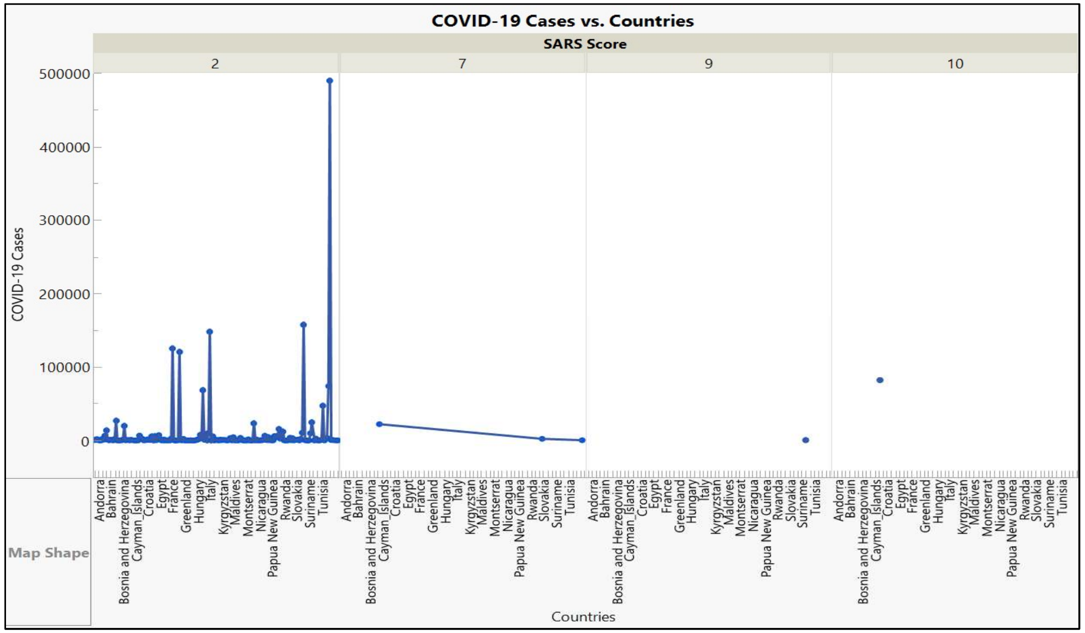
COVID-19 Cases in Different Countries

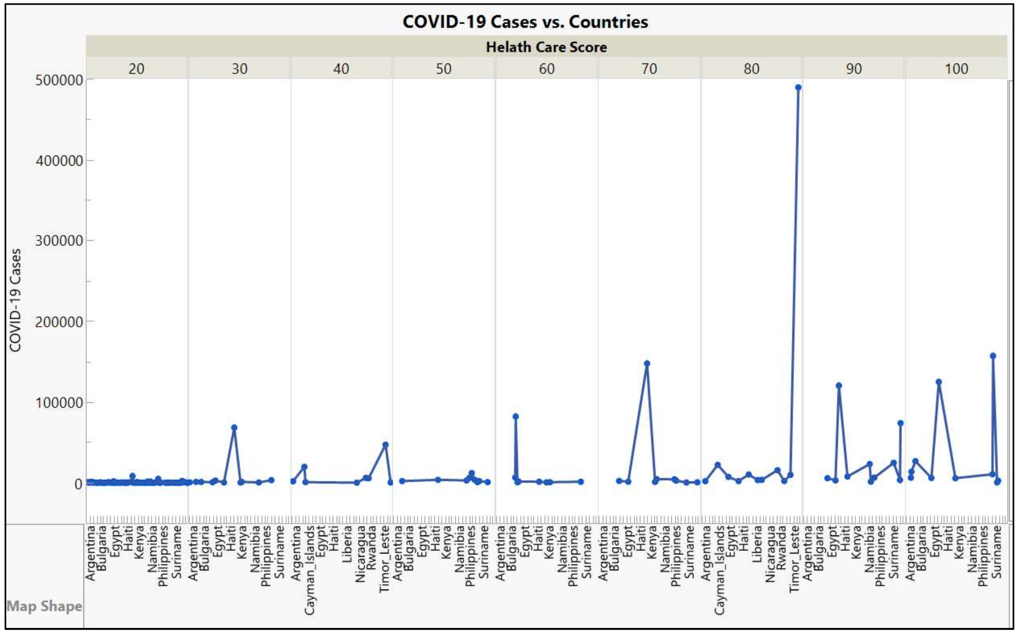




- # of COVID-19 cases seems to have association with following parameters
  - Number of tourists/visitors, and
  - Economic score facilitating business and tourism

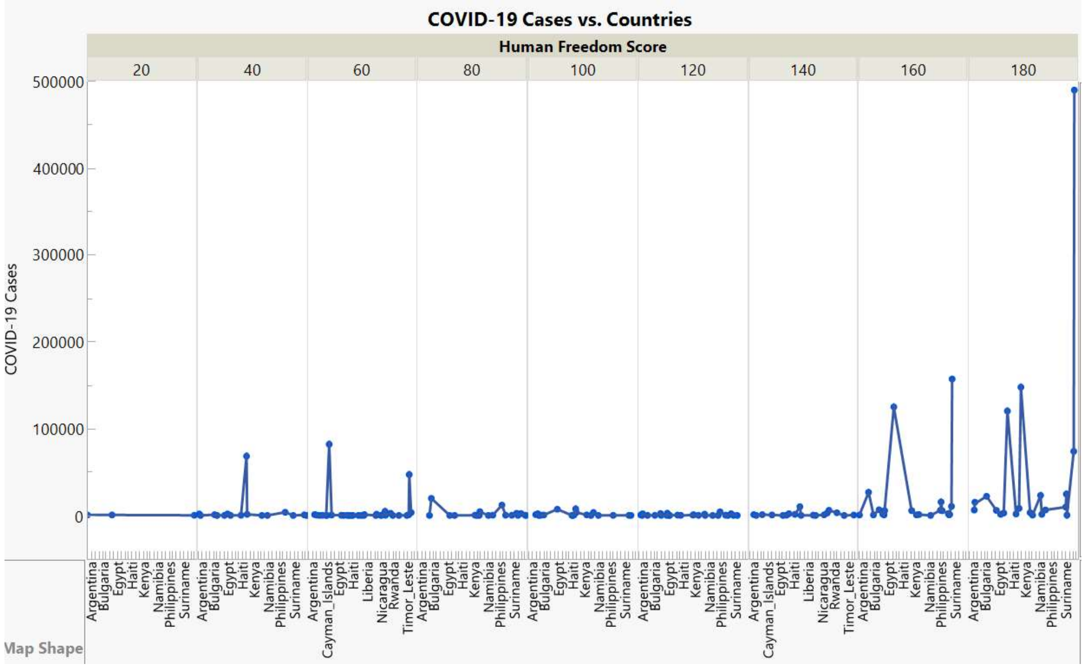
COVID-19 Cases in Different Countries





- # of COVID-19 cases more in countries with <u>no prior high exposure</u> to SARS etc (score = 2).
- No obvious association with Heath care system.

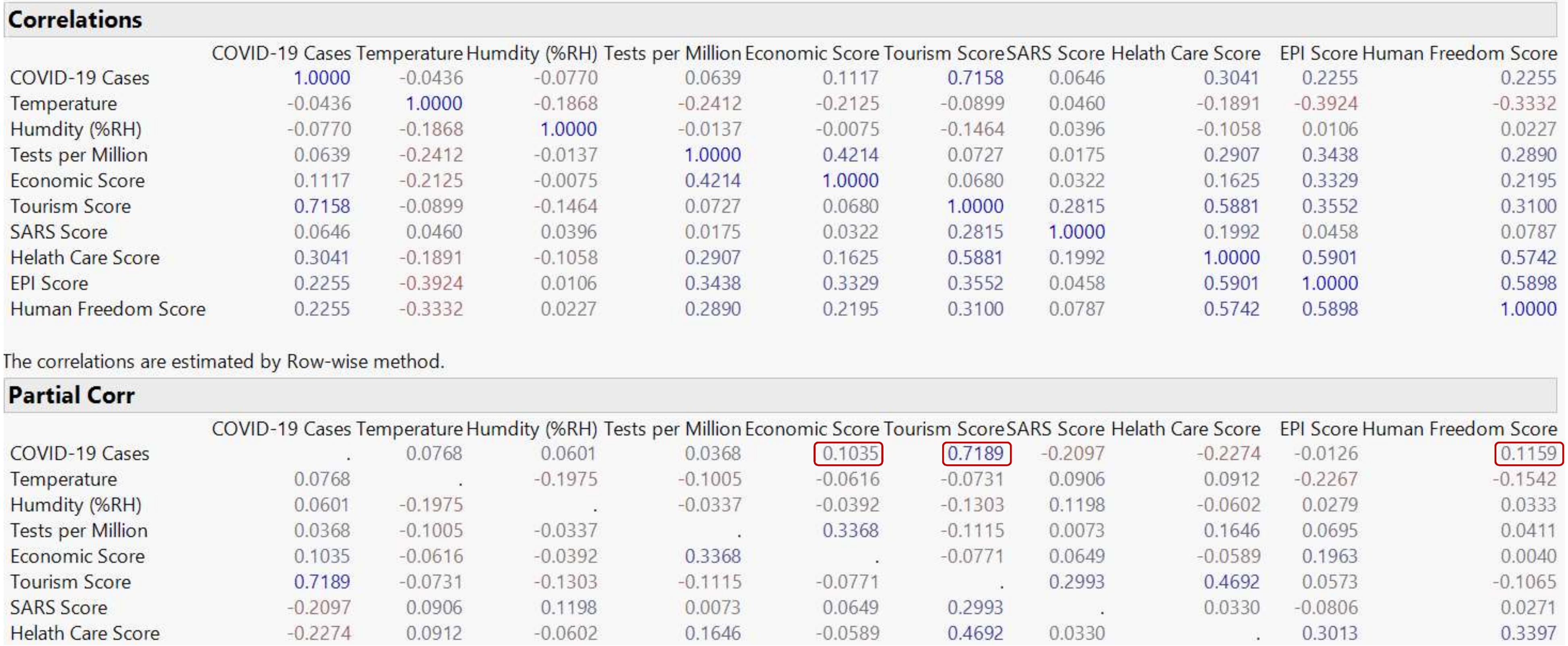
COVID-19 Cases in Different Countries



- # of COVID-19 cases more in countries with higher human freedom score, which represent following.

  - Freedom of people to make choices
  - Good social interactions
  - People friendly governance

Pairwise Correlation Analysis



Correlations

	COVID-19 Cases	Temperature	Humdity (%RH)	Tests per Million	Economic Score	Tourism Score	SARS Score	Helath Care Score	EPI Score	Human Freedom Score
COVID-19 Cases	1.0000	-0.0436	-0.0770	0.0639	0.1117	0.7158	0.0646	0.3041	0.2255	0.2255
Temperature	-0.0436	1.0000	-0.1868	-0.2412	-0.2125	-0.0899	0.0460	-0.1891	-0.3924	-0.3332
Humdity (%RH)	-0.0770	-0.1868	1.0000	-0.0137	-0.0075	-0.1464	0.0396	-0.1058	0.0106	0.0227
Tests per Million	0.0639	-0.2412	-0.0137	1.0000	0.4214	0.0727	0.0175	0.2907	0.3438	0.2890
Economic Score	0.1117	-0.2125	-0.0075	0.4214	1.0000	0.0680	0.0322	0.1625	0.3329	0.2195
Tourism Score	0.7158	-0.0899	-0.1464	0.0727	0.0680	1.0000	0.2815	0.5881	0.3552	0.3100
SARS Score	0.0646	0.0460	0.0396	0.0175	0.0322	0.2815	1.0000	0.1992	0.0458	0.0787
Helath Care Score	0.3041	-0.1891	-0.1058	0.2907	0.1625	0.5881	0.1992	1.0000	0.5901	0.5742
EPI Score	0.2255	-0.3924	0.0106	0.3438	0.3329	0.3552	0.0458	0.5901	1.0000	0.5898
Human Freedom Score	0.2255	-0.3332	0.0227	0.2890	0.2195	0.3100	0.0787	0.5742	0.5898	1.0000

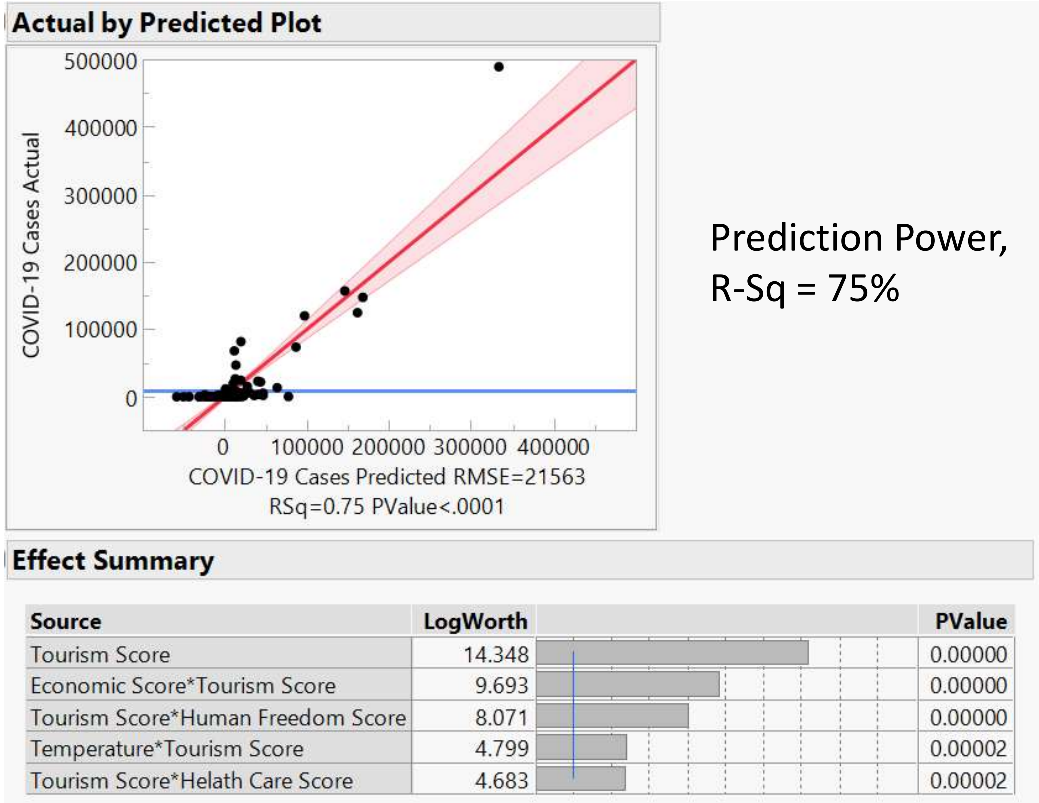
The correlations are estimated by Row-wise method.

Partial Corr

	COVID-19 Cases	Temperature	Humdity (%RH)	Tests per Million	Economic Score	Tourism Score	SARS Score	Helath Care Score	EPI Score	Human Freedom Score
COVID-19 Cases	.	0.0768	0.0601	0.0368	0.1035	0.7189	-0.2097	-0.2274	-0.0126	0.1159
Temperature	0.0768	.	-0.1975	-0.1005	-0.0616	-0.0731	0.0906	0.0912	-0.2267	-0.1542
Humdity (%RH)	0.0601	-0.1975	.	-0.0337	-0.0392	-0.1303	0.1198	-0.0602	0.0279	0.0333
Tests per Million	0.0368	-0.1005	-0.0337	.	0.3368	-0.1115	0.0073	0.1646	0.0695	0.0411
Economic Score	0.1035	-0.0616	-0.0392	0.3368	.	-0.0771	0.0649	-0.0589	0.1963	0.0040
Tourism Score	0.7189	-0.0731	-0.1303	-0.1115	-0.0771	.	0.2993	0.4692	0.0573	-0.1065
SARS Score	-0.2097	0.0906	0.1198	0.0073	0.0649	0.2993	.	0.0330	-0.0806	0.0271
Helath Care Score	-0.2274	0.0912	-0.0602	0.1646	-0.0589	0.4692	0.0330	.	0.3013	0.3397

- **Key predictors (90%) of COVID-19** = Tourism Score, Human Freedom Score, and Economic Score.
- **Minor predictors** = Temperature, Humidity, Tests/Million

Least Square Fit Analysis of Predictors



Actual by Predicted Plot

Prediction Power, R-Sq = 75%

Effect Summary

Source	LogWorth		PValue
Tourism Score	14.348		0.00000
Economic Score*Tourism Score	9.693		0.00000
Tourism Score*Human Freedom Score	8.071		0.00000
Temperature*Tourism Score	4.799		0.00002
Tourism Score*Helath Care Score	4.683		0.00002

- **Key predictors of COVID-19 with low p-values** = Tourism Score, Human Freedom Score, and Economic Score

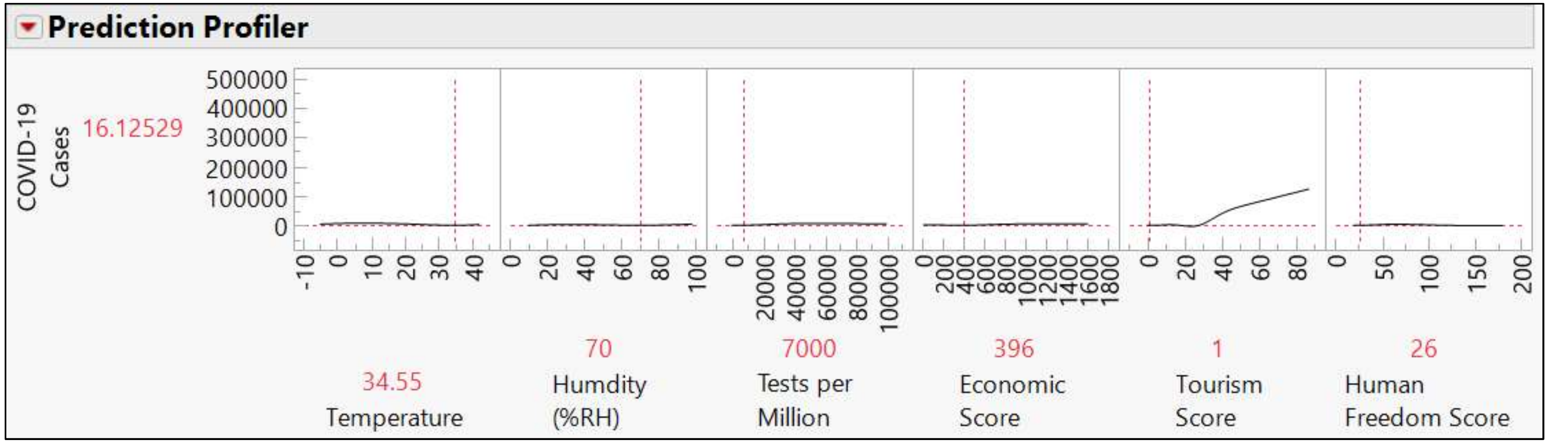
Discussion on Correlation Analysis

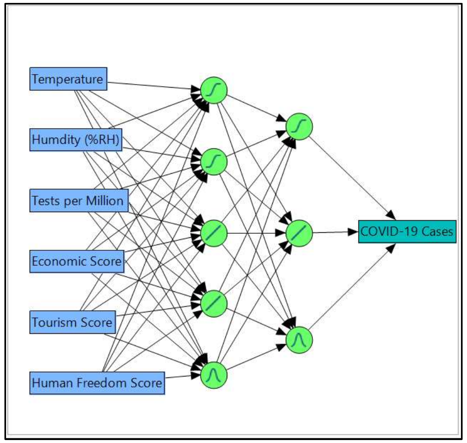
- Key predictors (90%) for # of COVID cases are Tourism Score, Human Freedom Score, and Economic Score. Higher international visitors in a country coupled with economic prosperity has highest potential for a pandemic to spread globally. Situation can gets possibly get worse in democratic countries, wherein violate people government advised rules, and regulations for social distancing, and health care. Violations can be in ‘ignorant category’, and ‘on-purpose’ category (religious, business, political, personal etc).

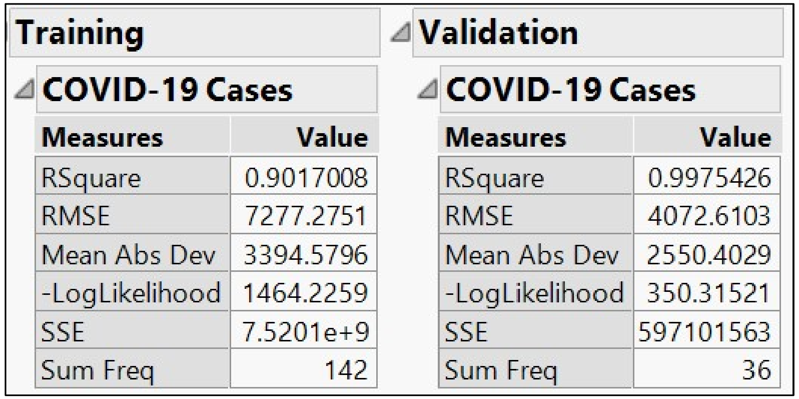
- As per available data, Temperature, and Humidity not expected to have more than 5 to 10% influence on # of COVID cases in any country.

- Rising temperatures, UV light exposures, and good food and exercise do help to enhance immune system of people to effectively fight COVID-19 infection. Yet, it is not expected to reduce # of COVID cases. Social distance, personal hygiene, and good health care for people are needed.

Neural Network Analysis of COVID-19 Cases







Training

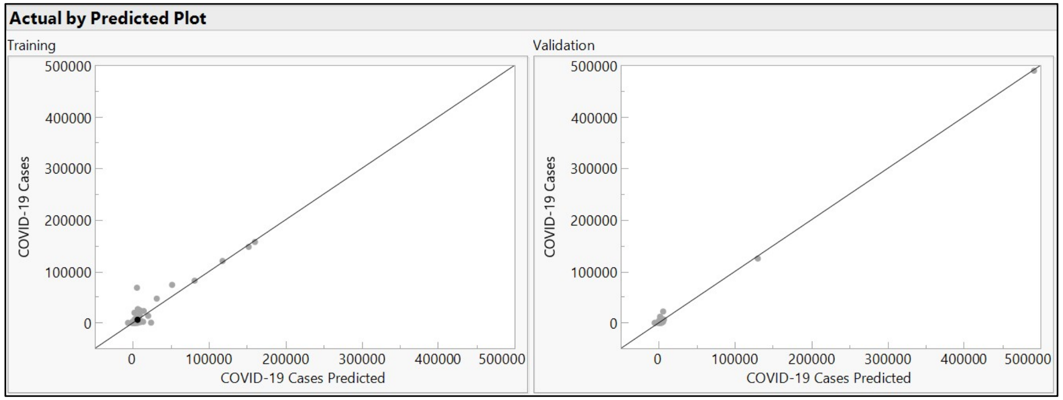
COVID-19 Cases

Measures	Value
RSquare	0.9017008
RMSE	7277.2751
Mean Abs Dev	3394.5796
-LogLikelihood	1464.2259
SSE	7.5201e+9
Sum Freq	142

Validation

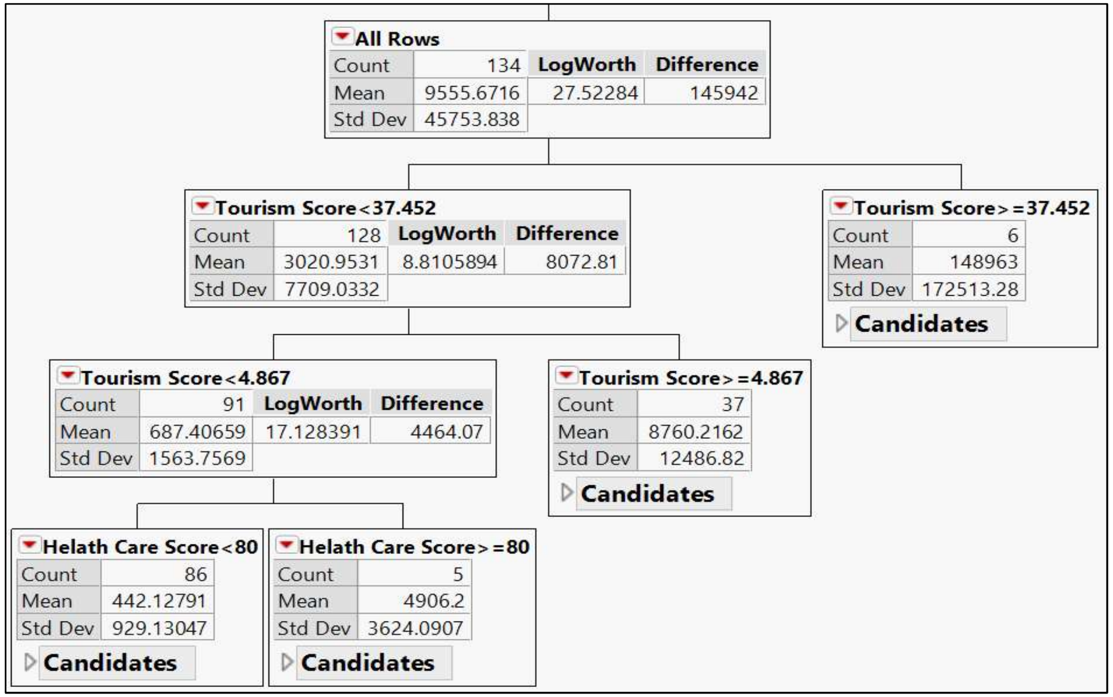
COVID-19 Cases

Measures	Value
RSquare	0.9975426
RMSE	4072.6103
Mean Abs Dev	2550.4029
-LogLikelihood	350.31521
SSE	597101563
Sum Freq	36



- **Recommendation** = Health Care + Minimum 2000 tests/million population, Social Distancing and Reduced Travels.

Decision Tree Analysis



All Rows

Count	134	**LogWorth**	**Difference**
Mean	9555.6716	27.52284	145942
Std Dev	45753.838		

Tourism Score<37.452

Count	128	**LogWorth**	**Difference**
Mean	3020.9531	8.8105894	8072.81
Std Dev	7709.0332		

Tourism Score>=37.452

Count	6
Mean	148963
Std Dev	172513.28

Candidates

Tourism Score<4.867

Count	91	**LogWorth**	**Difference**
Mean	687.40659	17.128391	4464.07
Std Dev	1563.7569		

Tourism Score>=4.867

Count	37
Mean	8760.2162
Std Dev	12486.82

Candidates

Helath Care Score<80

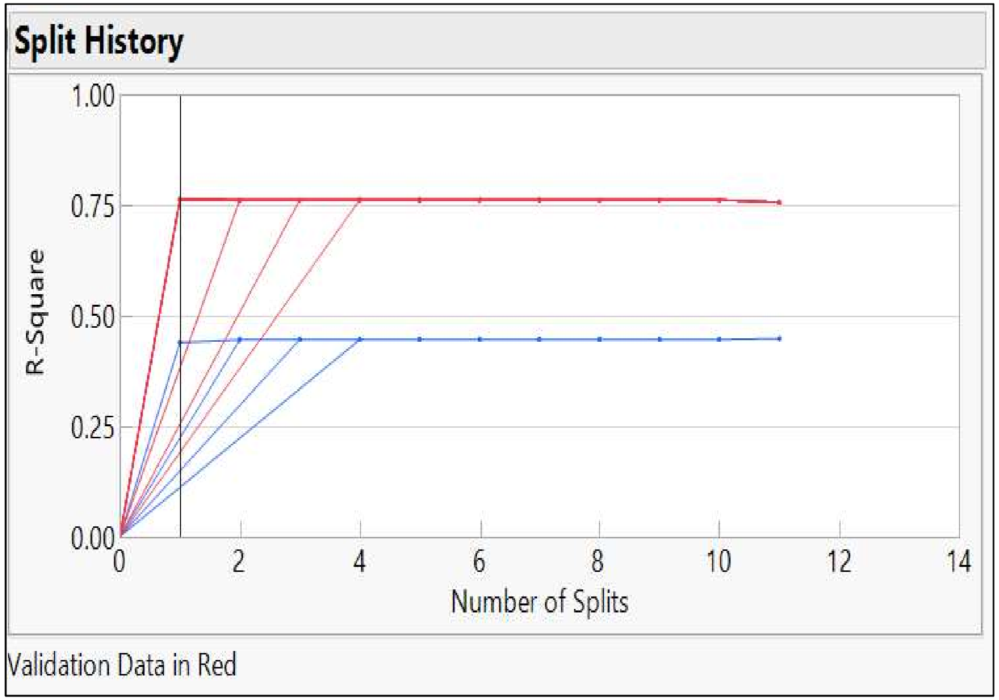
Count	86
Mean	442.12791
Std Dev	929.13047

Candidates

Helath Care Score>=80

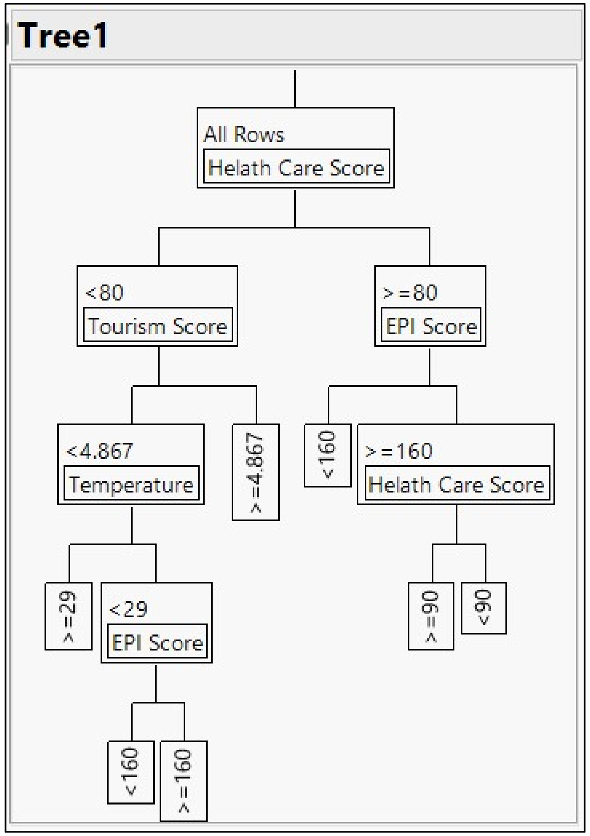
Count	5
Mean	4906.2
Std Dev	3624.0907

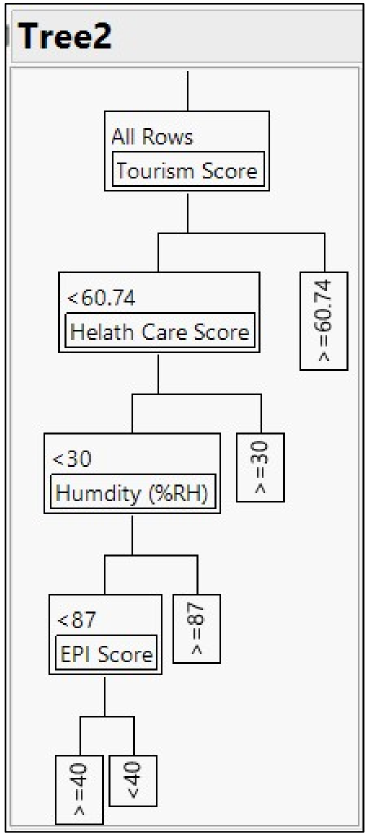
Candidates

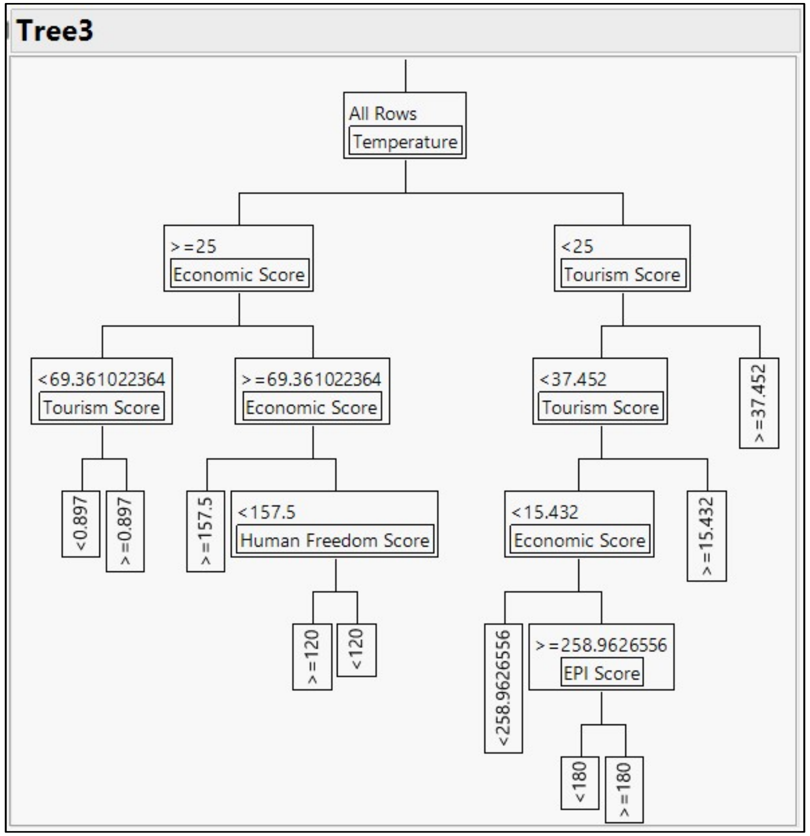


- Important factor driving # of COVID-19 cases: (a) # of Visitors, and (b) Health Care and Awareness Score.
- **Recommendation** = Improve Social Distancing and Health Care Services and Awareness.

Bootstrap Analysis

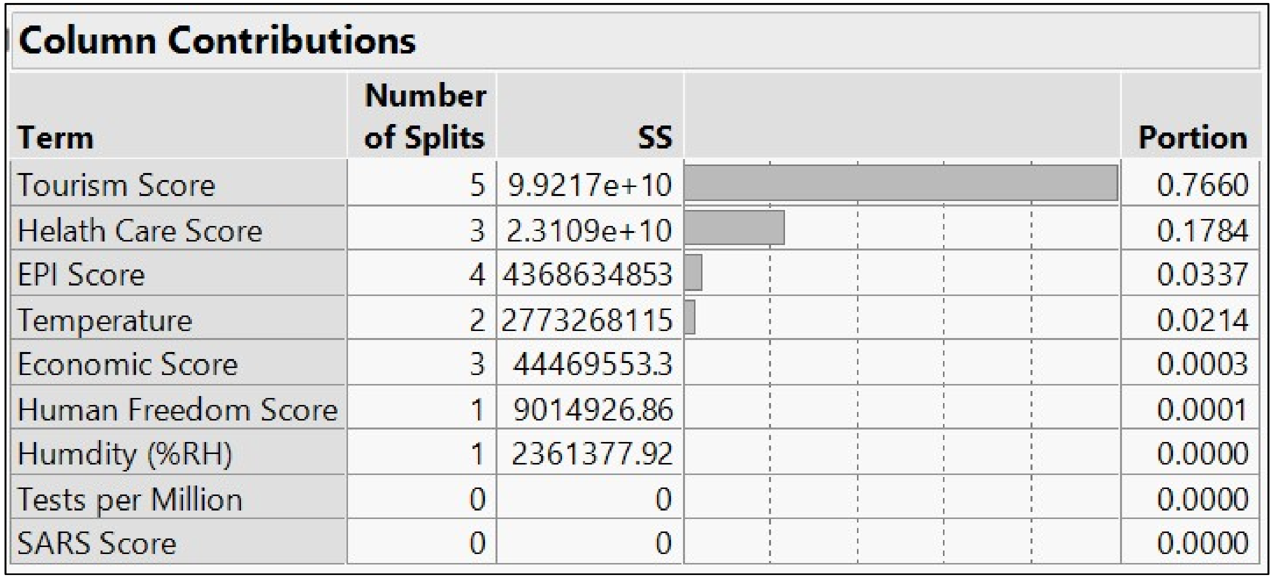






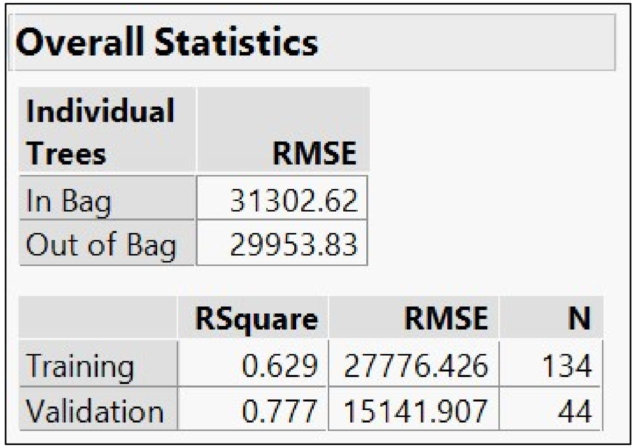
- Important factor driving # of COVID-19 cases: (a) # of Visitors/Tourism Score , and (b) Health Care and Awareness Score

Bootstrap Analysis



Column Contributions

Term	Number of Splits	SS		Portion
Tourism Score	5	9.9217e+10		0.7660
Helath Care Score	3	2.3109e+10		0.1784
EPI Score	4	4368634853		0.0337
Temperature	2	2773268115		0.0214
Economic Score	3	44469553.3		0.0003
Human Freedom Score	1	9014926.86		0.0001
Humdity (%RH)	1	2361377.92		0.0000
Tests per Million	0	0		0.0000
SARS Score	0	0		0.0000



Overall Statistics

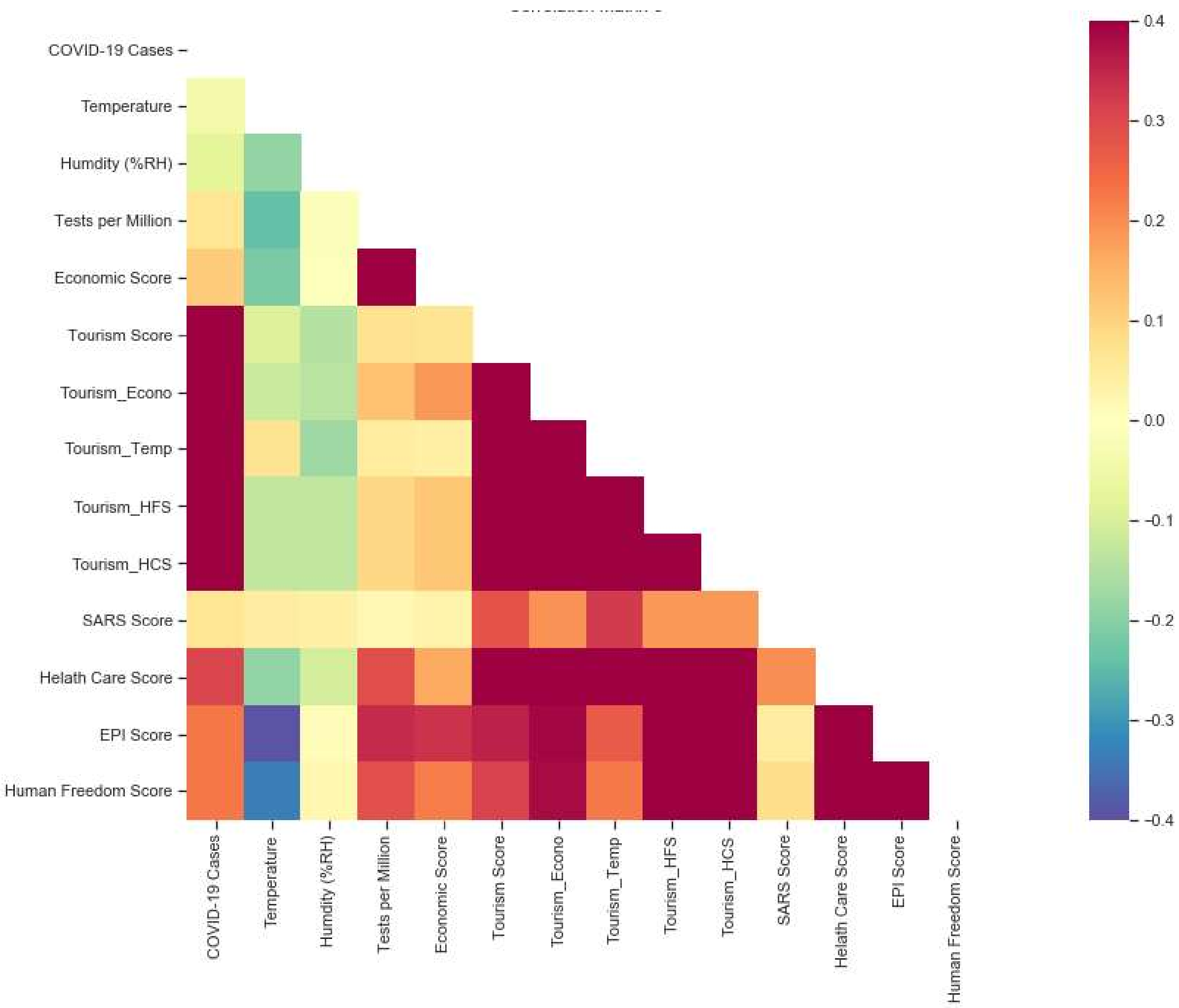
Individual Trees	RMSE
In Bag	31302.62
Out of Bag	29953.83

	RSquare	RMSE	N
Training	0.629	27776.426	134
Validation	0.777	15141.907	44

- Important factor driving # of COVID-19 cases: (a) # of Visitors/Tourism Score , and (b) Health Care and Awareness Score
- **Recommendation** = Reduce International/National Travels, Improve Social Distancing and Health Care Services and Awareness.

Data Analysis using Jupyter Notebook

Correlation Matrix of Predictors of COVID-19 Cases

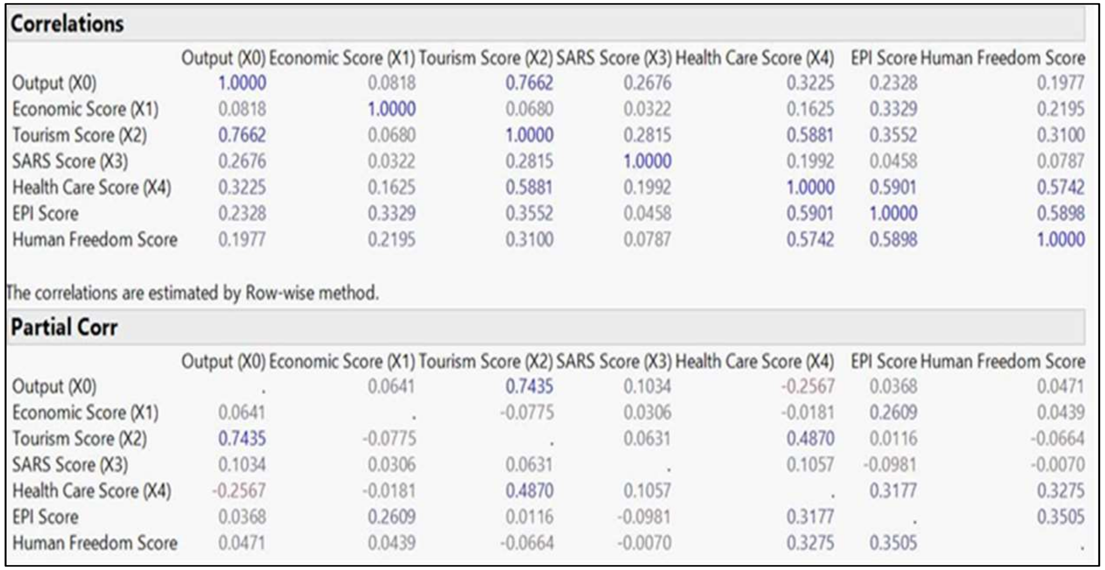


Hot colors = More correlation

Cool colors = Lower correlation

- **Key predictors (80%) of COVID-19** = Tourism Score, Human Freedom Score, and Economic Score.

Partial Correlation Matrixes of Predictors



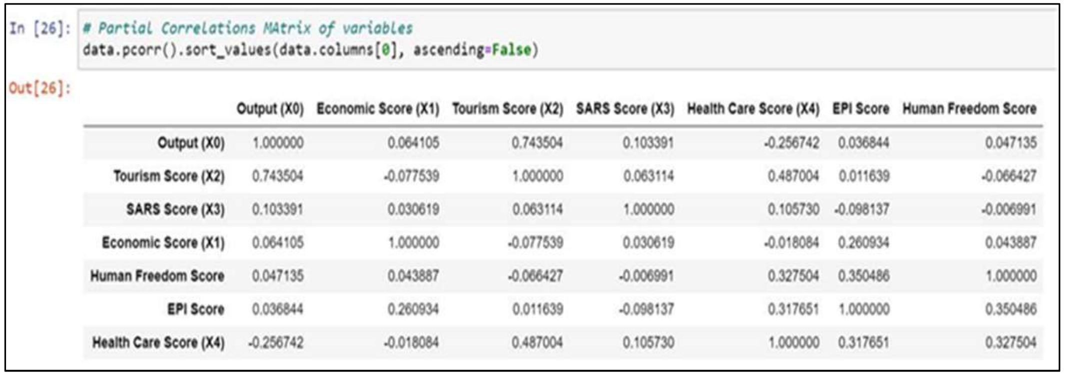
Correlations

	Output (X0)	Economic Score (X1)	Tourism Score (X2)	SARS Score (X3)	Health Care Score (X4)	EPI Score	Human Freedom Score
Output (X0)	1.0000	0.0818	0.7662	0.2676	0.3225	0.2328	0.1977
Economic Score (X1)	0.0818	1.0000	0.0680	0.0322	0.1625	0.3329	0.2195
Tourism Score (X2)	0.7662	0.0680	1.0000	0.2815	0.5881	0.3552	0.3100
SARS Score (X3)	0.2676	0.0322	0.2815	1.0000	0.1992	0.0458	0.0787
Health Care Score (X4)	0.3225	0.1625	0.5881	0.1992	1.0000	0.5901	0.5742
EPI Score	0.2328	0.3329	0.3552	0.0458	0.5901	1.0000	0.5898
Human Freedom Score	0.1977	0.2195	0.3100	0.0787	0.5742	0.5898	1.0000

The correlations are estimated by Row-wise method.

Partial Corr

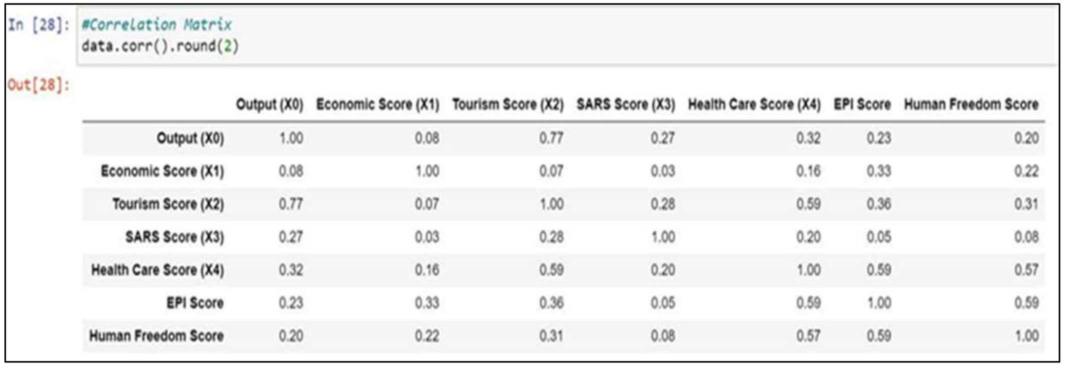
	Output (X0)	Economic Score (X1)	Tourism Score (X2)	SARS Score (X3)	Health Care Score (X4)	EPI Score	Human Freedom Score
Output (X0)	.	0.0641	0.7435	0.1034	-0.2567	0.0368	0.0471
Economic Score (X1)	0.0641	.	-0.0775	0.0306	-0.0181	0.2609	0.0439
Tourism Score (X2)	0.7435	-0.0775	.	0.0631	0.4870	0.0116	-0.0664
SARS Score (X3)	0.1034	0.0306	0.0631	.	0.1057	-0.0981	-0.0070
Health Care Score (X4)	-0.2567	-0.0181	0.4870	0.1057	.	0.3177	0.3275
EPI Score	0.0368	0.2609	0.0116	-0.0981	0.3177	.	0.3505
Human Freedom Score	0.0471	0.0439	-0.0664	-0.0070	0.3275	0.3505	.



```
In [26]: # Partial Correlations MAtrix of variables
         data.pcorr().sort_values(data.columns[0], ascending=False)
```

Out[26]:

	Output (X0)	Economic Score (X1)	Tourism Score (X2)	SARS Score (X3)	Health Care Score (X4)	EPI Score	Human Freedom Score
Output (X0)	1.000000	0.064105	0.743504	0.103391	-0.256742	0.036844	0.047135
Tourism Score (X2)	0.743504	-0.077539	1.000000	0.063114	0.487004	0.011639	-0.066427
SARS Score (X3)	0.103391	0.030619	0.063114	1.000000	0.105730	-0.098137	-0.006991
Economic Score (X1)	0.064105	1.000000	-0.077539	0.030619	-0.018084	0.260934	0.043887
Human Freedom Score	0.047135	0.043887	-0.066427	-0.006991	0.327504	0.350486	1.000000
EPI Score	0.036844	0.260934	0.011639	-0.098137	0.317651	1.000000	0.350486
Health Care Score (X4)	-0.256742	-0.018084	0.487004	0.105730	1.000000	0.317651	0.327504



```
In [28]: #Correlation Matrix
         data.corr().round(2)
```

Out[28]:

	Output (X0)	Economic Score (X1)	Tourism Score (X2)	SARS Score (X3)	Health Care Score (X4)	EPI Score	Human Freedom Score
Output (X0)	1.00	0.08	0.77	0.27	0.32	0.23	0.20
Economic Score (X1)	0.08	1.00	0.07	0.03	0.16	0.33	0.22
Tourism Score (X2)	0.77	0.07	1.00	0.28	0.59	0.36	0.31
SARS Score (X3)	0.27	0.03	0.28	1.00	0.20	0.05	0.08
Health Care Score (X4)	0.32	0.16	0.59	0.20	1.00	0.59	0.57
EPI Score	0.23	0.33	0.36	0.05	0.59	1.00	0.59
Human Freedom Score	0.20	0.22	0.31	0.08	0.57	0.59	1.00

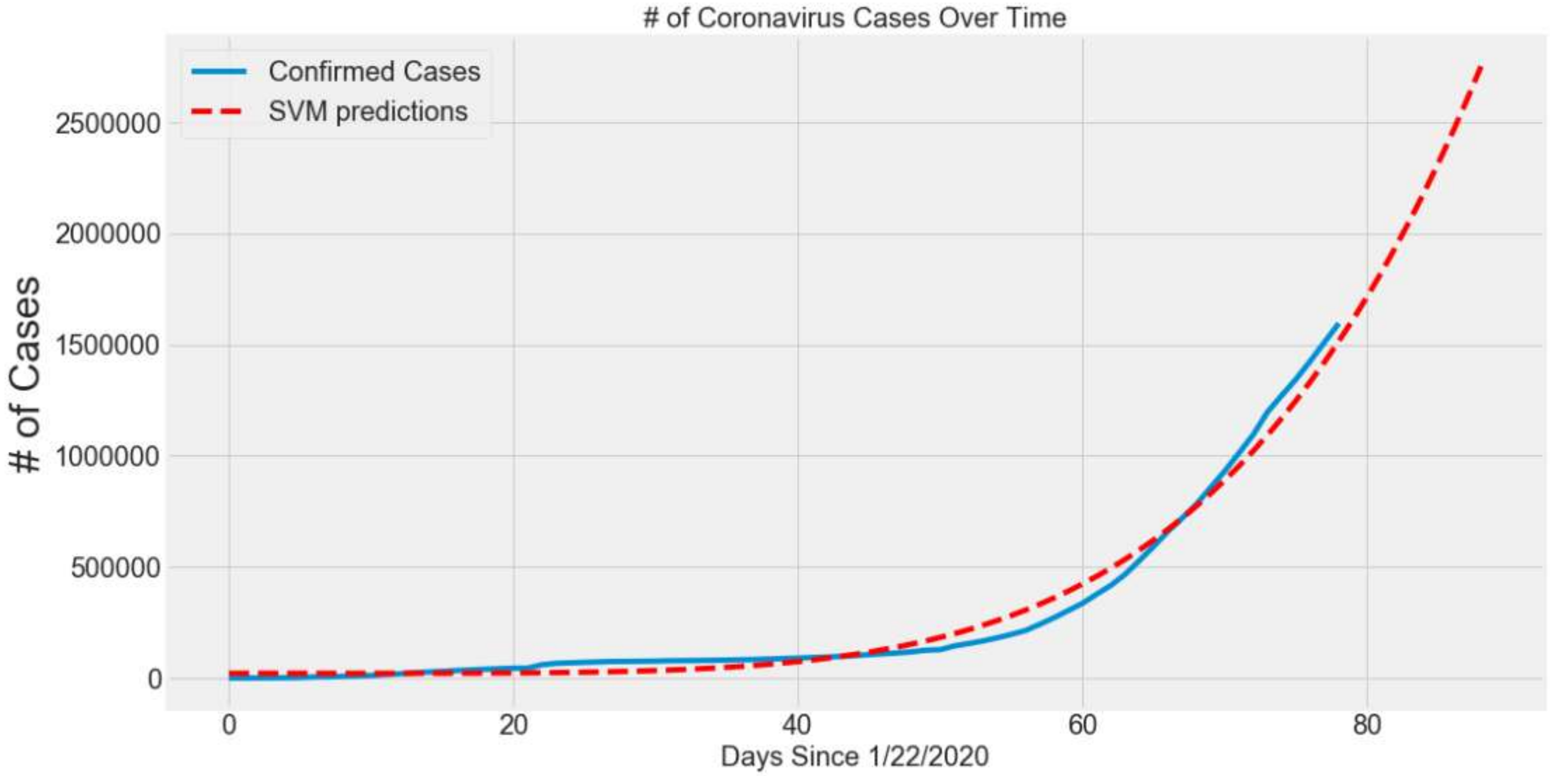
- Key predictors (80%) = Tourism Score, Human Freedom Score, and Economic Score.
- Analysis results from JMP Pro 14, and Jupyter Notebook Script analysis are same.

p-value Correlation Matrix of COVID-19 Predictors

	COVID-19 Cases	Temperature	Humdity (%RH)	Tests per Million	Economic Score	Tourism Score	Tourism_Econo	Tourism_Temp	Tourism_HFS	Tourism_HCS	SARS Score	Helath Care Score	EPI Score	Human Freedom Score
COVID-19 Cases	-	0.563732	0.307096	0.396616	0.137577	0.000000	0.000000	0.000000	0.000000	0.000000	0.391382	0.000037	0.002478	0.002477
Temperature	-0.043557	-	0.012512	0.00118	0.004398	0.232602	0.119316	0.343591	0.087782	0.087782	0.541973	0.011451	0.000000	0.000006
Humdity (%RH)	-0.076981	-0.18685	-	0.856434	0.921193	0.051111	0.063194	0.021042	0.086543	0.086543	0.600154	0.159833	0.888247	0.763758
Tests per Million	0.063922	-0.241225	-0.013656	-	0.000000	0.33476	0.093063	0.520447	0.224141	0.224141	0.81663	0.000083	0.000003	0.000091
Economic Score	0.111734	-0.21253	-0.007468	0.421405	-	0.367137	0.013617	0.61517	0.112877	0.112877	0.669634	0.030181	0.000006	0.003238
Tourism Score	0.715817	-0.089922	-0.146442	0.072714	0.067997	-	0.000000	0.000000	0.000000	0.000000	0.000141	0.000000	0.000001	0.000025
Tourism_Econo	0.868365	-0.117174	-0.139551	0.126266	0.184634	0.913859	-	0.000000	0.000000	0.000000	0.010267	0.000000	0.000000	0.000000
Tourism_Temp	0.6879	0.0714	-0.172849	0.048482	0.037933	0.960874	0.849054	-	0.000000	0.000000	0.000010	0.000000	0.000245	0.002727
Tourism_HFS	0.775237	-0.128336	-0.128836	0.091566	0.119246	0.950854	0.963825	0.877335	-	0.	0.013697	0.000000	0.000000	0.000000
Tourism_HCS	0.775237	-0.128336	-0.128836	0.091566	0.119246	0.950854	0.963825	0.877335	1.0	-	0.013697	0.000000	0.000000	0.000000
SARS Score	0.064632	0.046009	0.039552	0.017502	0.032197	0.281549	0.191944	0.32371	0.184479	0.184479	-	0.00767	0.544167	0.296532
Helath Care Score	0.304068	-0.189147	-0.105806	0.290731	0.16254	0.588126	0.550765	0.554376	0.593797	0.593797	0.199241	-	0.000000	0.000000
EPI Score	0.225467	-0.392383	0.010607	0.34383	0.332943	0.35521	0.391386	0.271617	0.39751	0.39751	0.045759	0.590114	-	0.000000
Human Freedom Score	0.225476	-0.333178	0.022684	0.288996	0.219526	0.309962	0.385515	0.223359	0.423413	0.423413	0.078676	0.574167	0.589843	-

- **Key predictors (80%)** = Tourism Score, Human Freedom Score, and Economic Score.

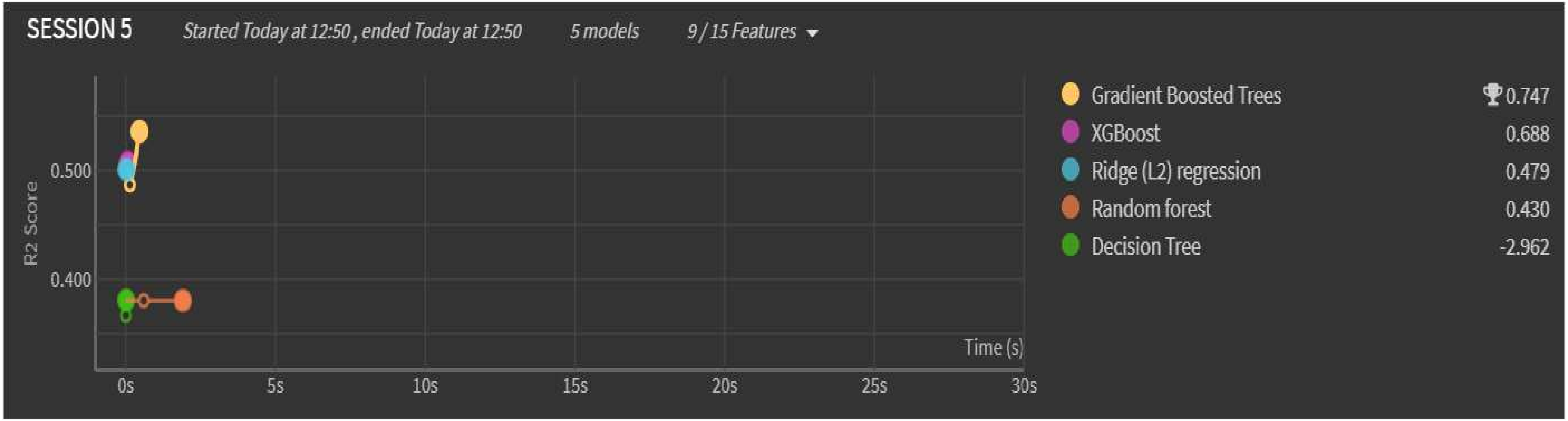
Support Vector Machine Predictions



Starting from 15 March 2020, most countries are showing <u>exponential rise </u>in number of COVID-19 reported cases.

Data Analysis using Dataiku DSS

COVID predictors using different algorithms

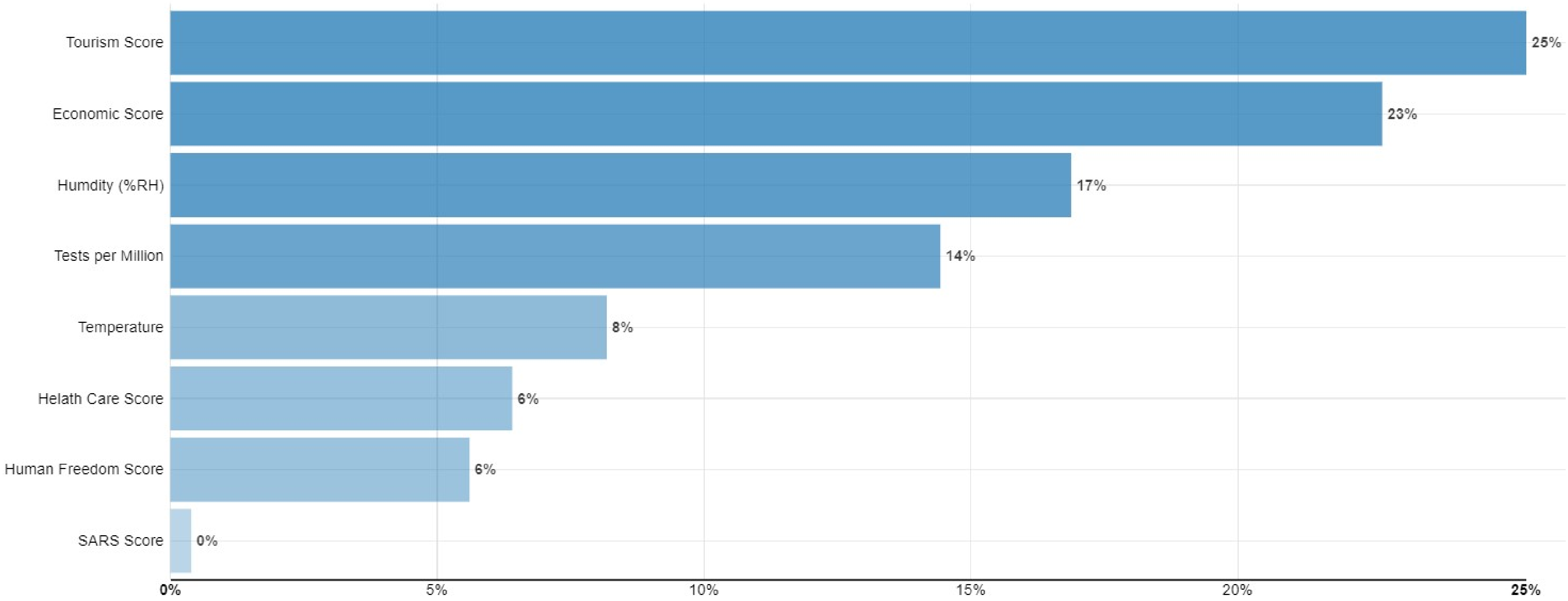


SESSION 5 *Started Today at 12:50 , ended Today at 12:50* *5 models* *9 / 15 Features* ▾

● Gradient Boosted Trees	🏆 0.747
● XGBoost	0.688
● Ridge (L2) regression	0.479
● Random forest	0.430
● Decision Tree	-2.962

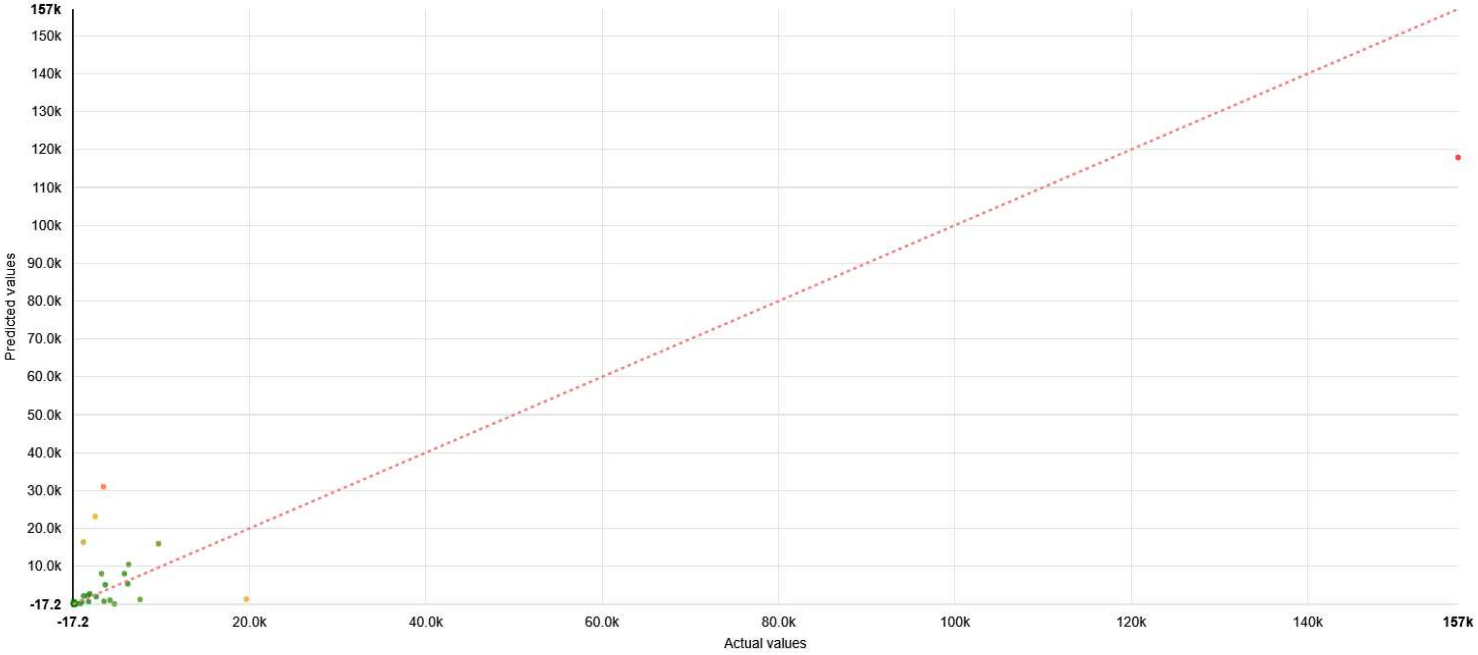
- Number of COVID-19 reported cases are best predicted by Gradient Boost Trees Network. R-Sq = 0.74

Gradient Boost Trees, Variable Importance



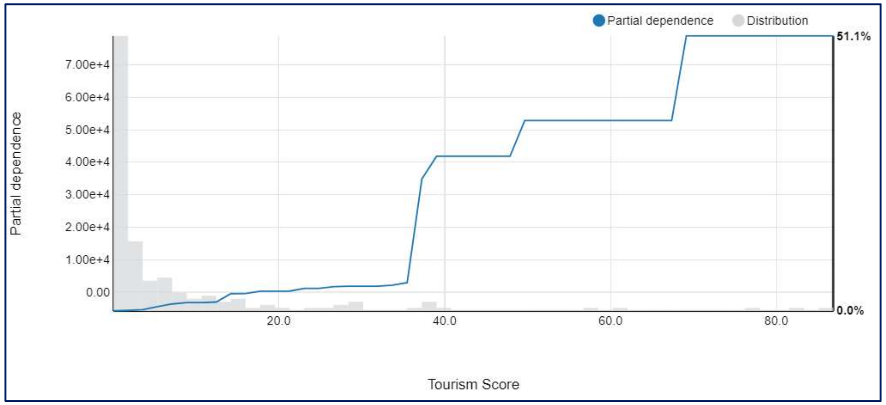
- Key predictors of number of COVID-19 cases = Tourism score, Economics Score, Humidity, and Tests/Million

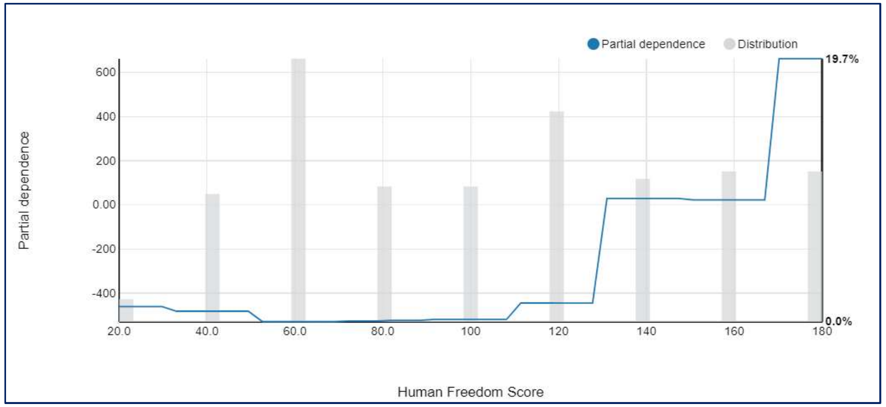
Gradient Boost Trees, Predicted vs Actual

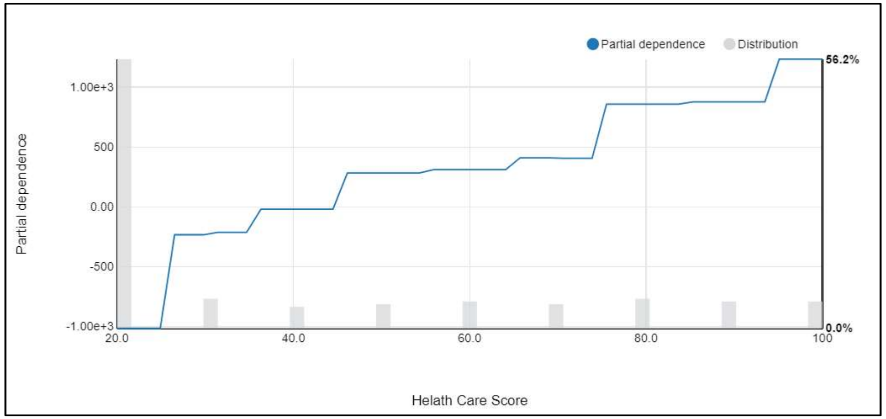


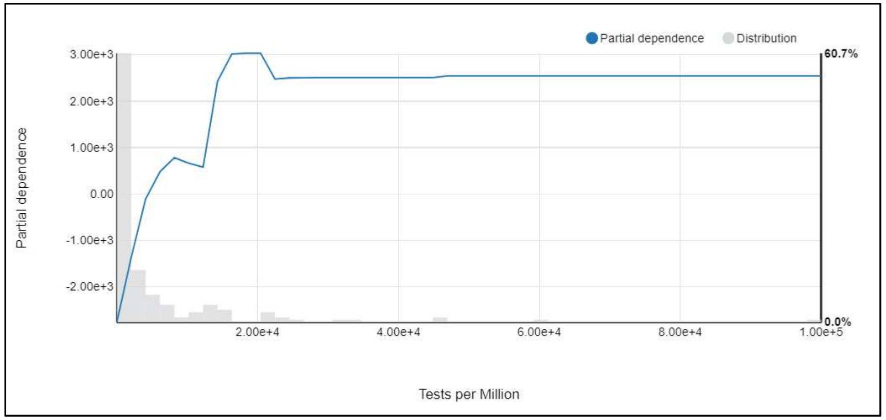
- Number of COVID-19 reported cases are best predicted by Gradient Boost Trees Network. R-sq = 0.747

Gradient Boost Trees, Partial Dependencies



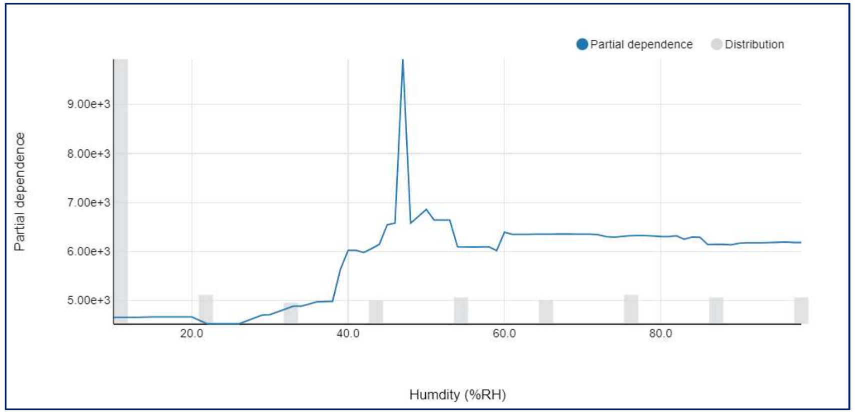


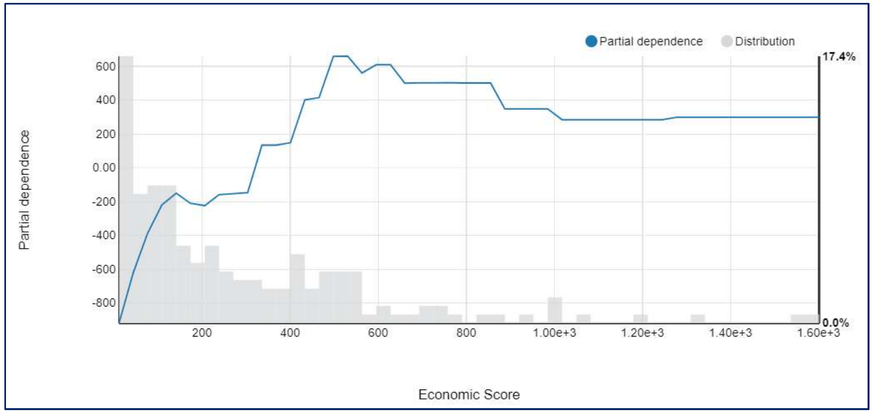


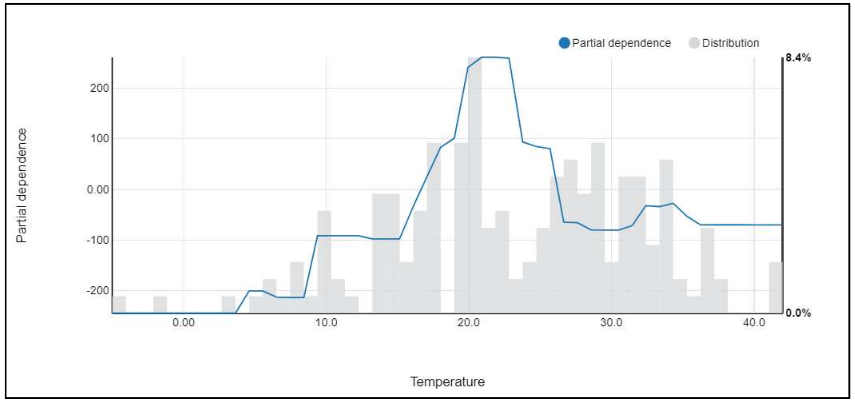


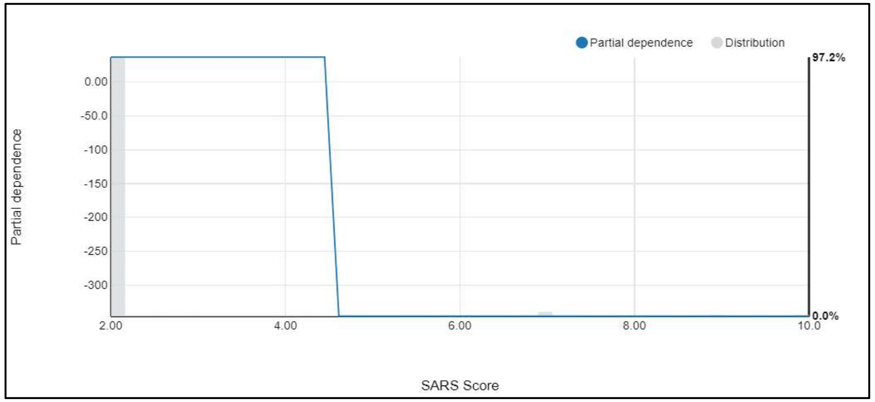
- Number of COVID-19 reported cases have strong partial dependencies on Tourism Scores, Health Care and Awareness, and Human Freedom Scores, and testing/million population.

Gradient Boost Trees, Partial Dependencies



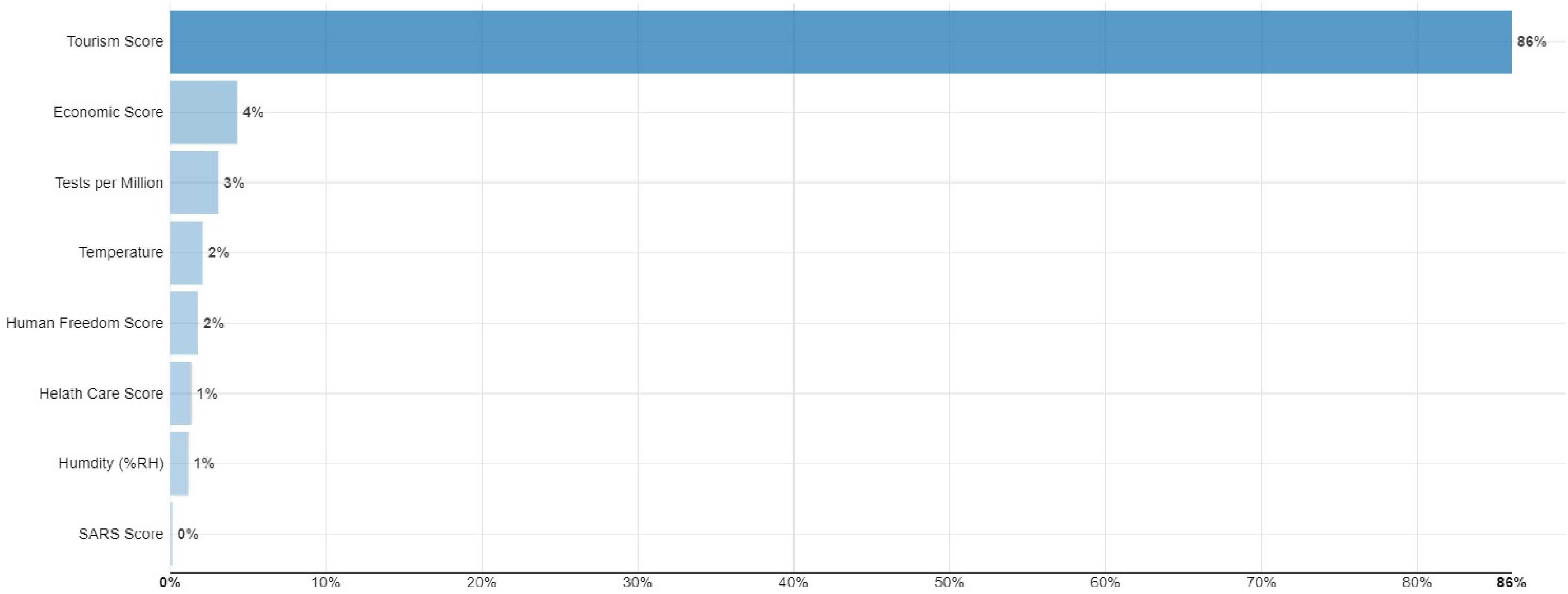






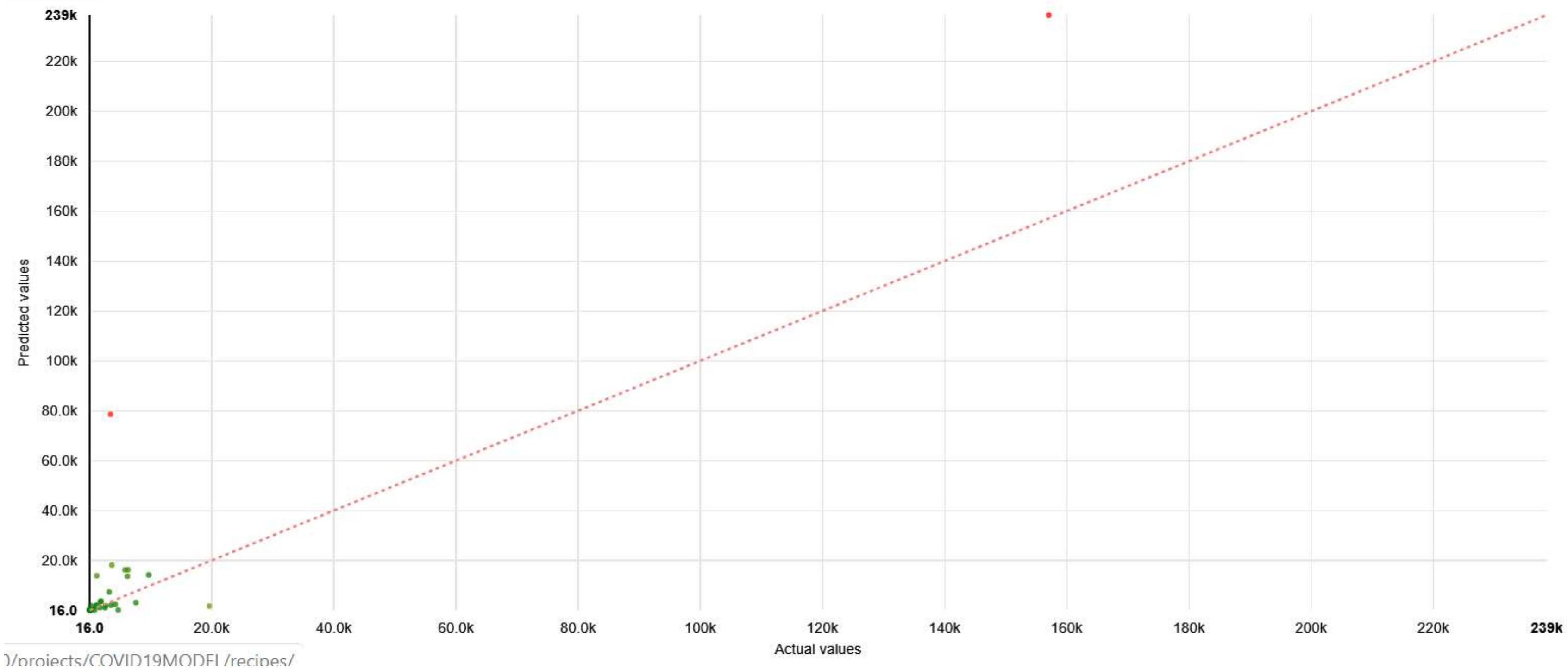
- Number of COVID-19 reported cases have <u>weak</u> partial dependencies on temperature, humidity, and economic score. But, <u>strong</u> partial dependency observed on SARS Score.

Random Forest, Variable Importance



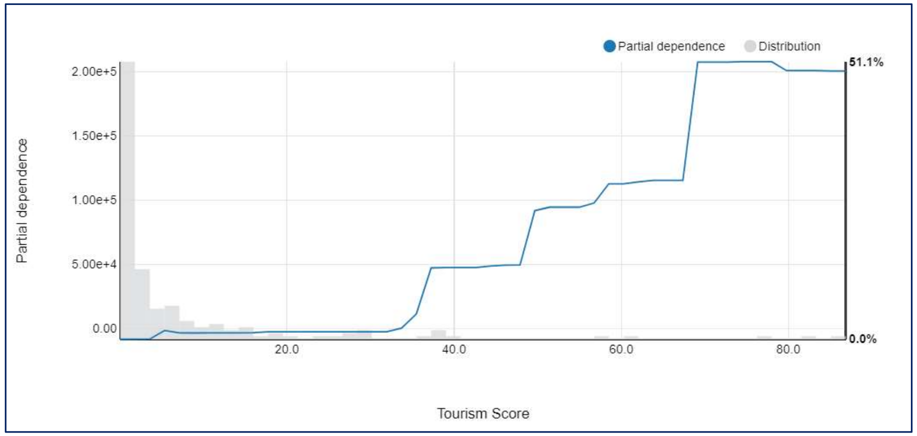
- Key predictors of number of COVID-19 cases = Tourism score, Economics Score, Humidity, and Tests/Million

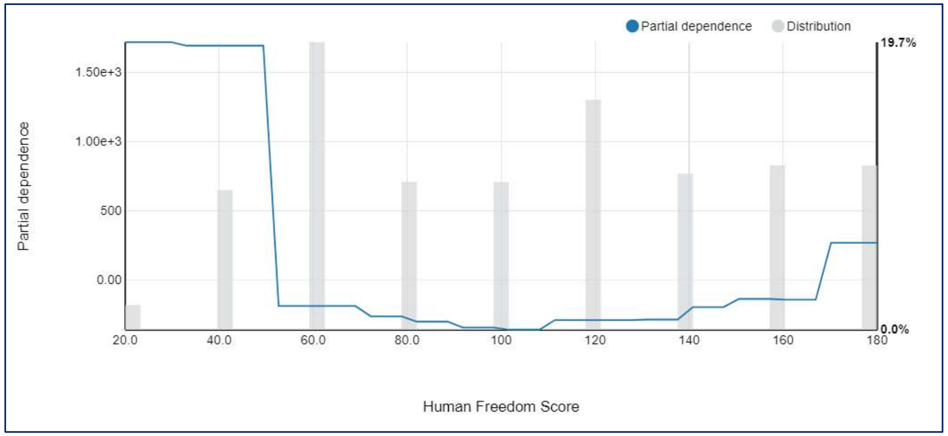
Random Forest, Predicted vs Actual

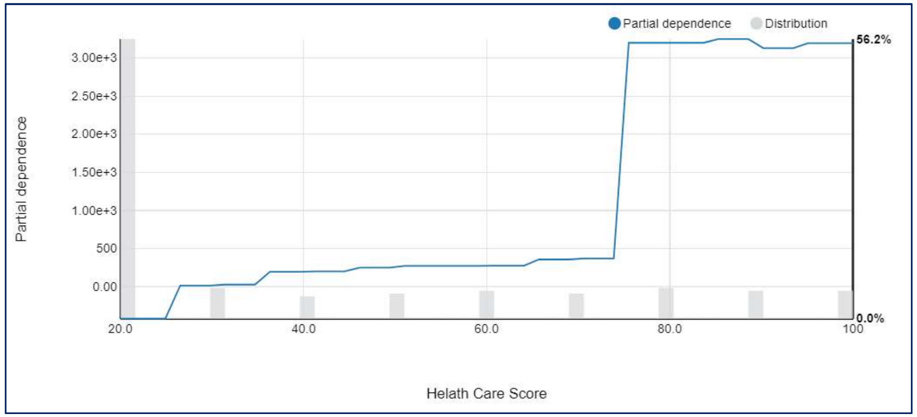


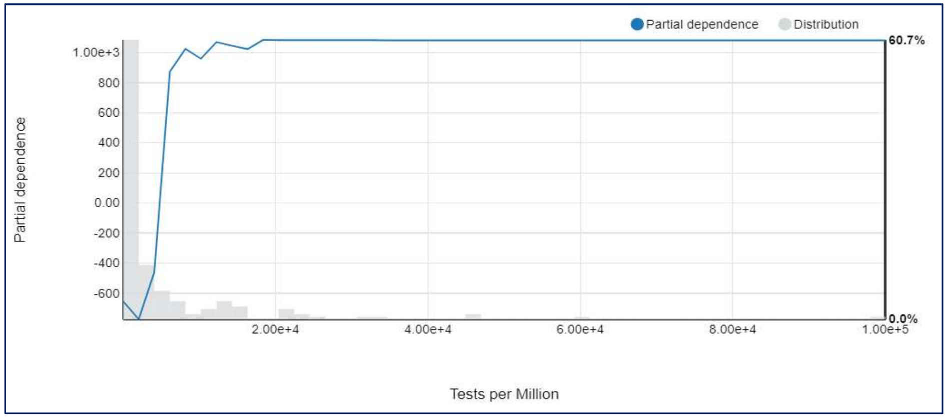
- Number of COVID-19 reported cases are best predicted by Gradient Boost Trees Network. R-sq = 0.43

Random Forest, Partial Dependencies



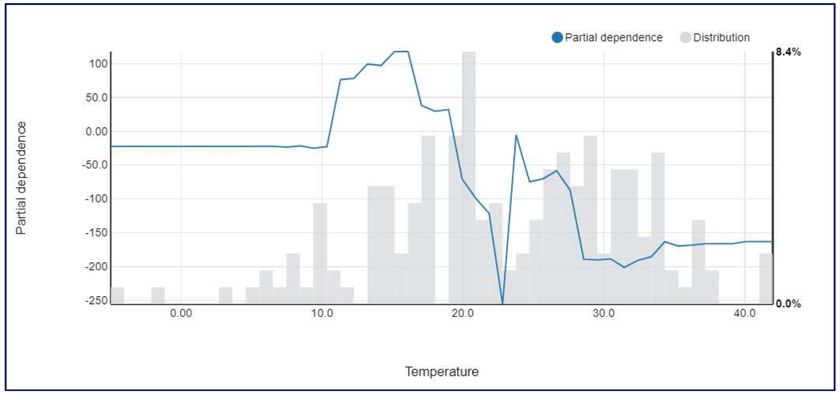


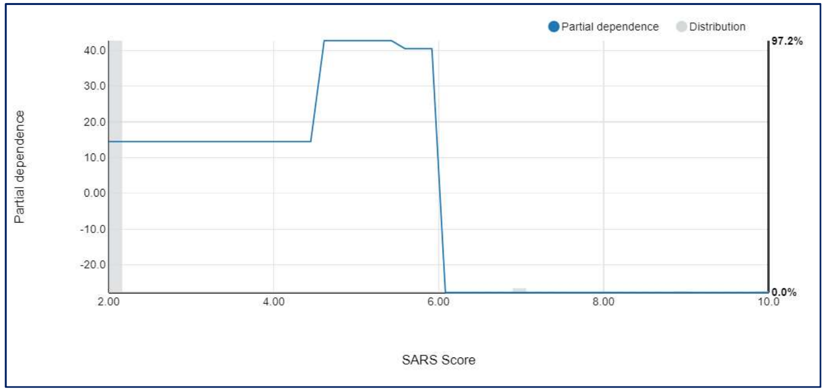


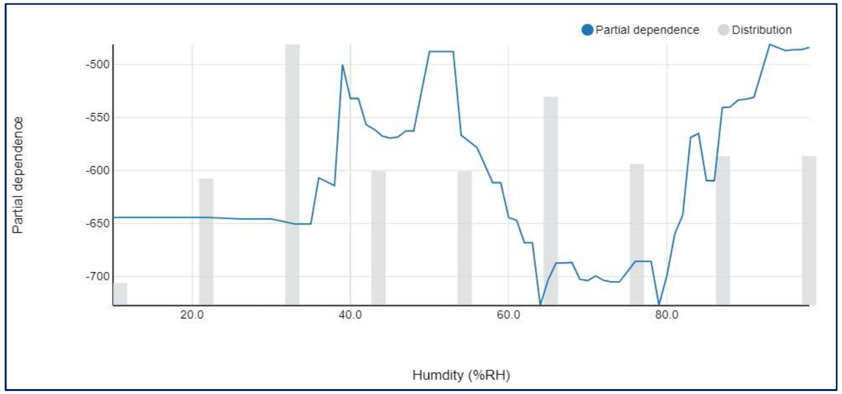


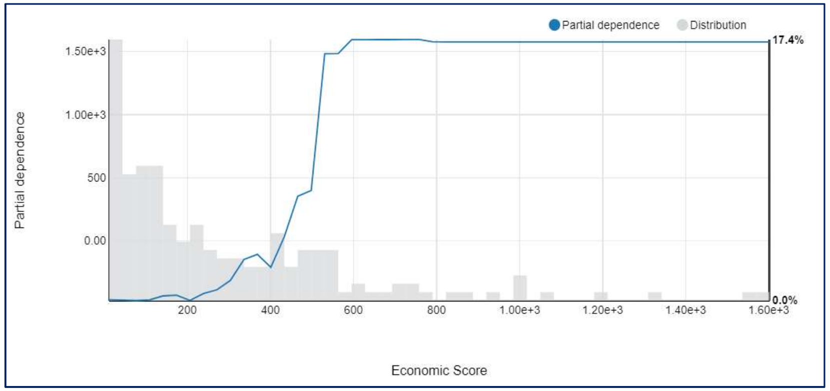
- Number of COVID-19 reported cases have strong partial dependencies on Tourism Scores, Health Care and Awareness, and Human Freedom Scores, and testing.

Random Forest, Partial Dependencies



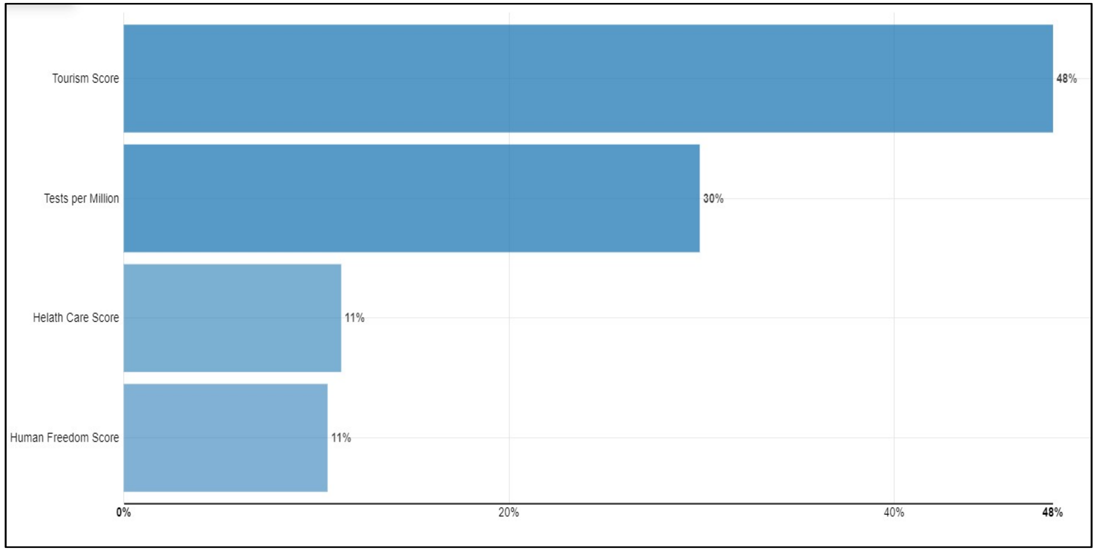


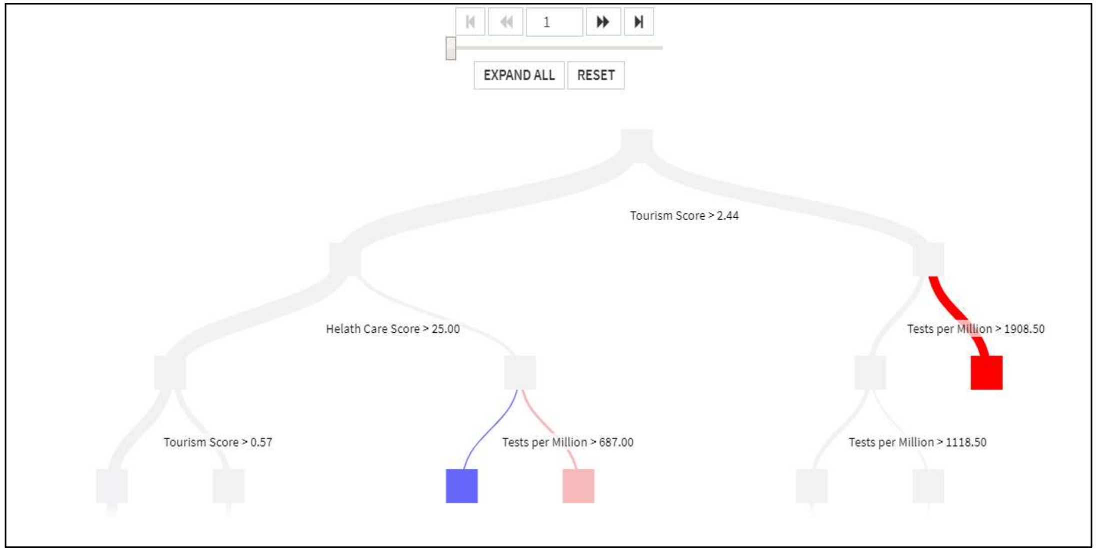




- Number of COVID-19 reported cases have weak partial dependencies on temperature, and humidity.
- Number of COVID-19 reported cases have strong partial dependencies on SARS score, and Economics.

Revised Gradient Boosted Trees Analysis




1
EXPAND ALL
RESET
Tourism Score > 2.44
Helath Care Score > 25.00
Tests per Million > 1908.50
Tourism Score > 0.57
Tests per Million > 687.00
Tests per Million > 1118.50

- Number of COVID-19 reported cases are best predicted by Gradient Boost Trees Network. R-Sq = 0.75
- Most suitable predictors = Tourism score, Tests/Million, Health Care and Awareness Score, and Human Freedom Score.

Discussion on results of Machine Learning Algorithms for COVID-19 Predictions

- Analysis of COVID-19 data using different algorithms, and software show different R-sq, MAE, and Pearson Coefficient.

- Key to identify best predictors is optimized R-sq, MAE, and Pearson Coefficient, and Partial dependencies. Due to nature of data, certain algorithm may show enhanced percentage contribution of a predictor. This can be diagnosed by doing examining partial dependencies.

- Number of COVID-19 reported cases are best predicted by Gradient Boost Trees Network.

  And key predictors are Tourism Scores, Health Care and Awareness, Human Freedom Scores,

  COVID-19 testing/million population, and SARS score.

Conclusions

- For capstone project COVID-19 data was analyzed. Descriptive, Predictive and Prescriptive Analysis were performed.

- Number of COVID cases were predicted using support vector machine algorithm.

  As of 11th April, there is exponential trend for increase of number of COVID-19 cases.

- COVID-19 spread/containment can best be predicted by Tourism score, Tests/Million, Health Care and Awareness Score, and Human Freedom Score.

References

- JMP Pro 14. https://www.jmp.com/en_ch/software/new-release/preview-jmp14.html
- Dataiku DSS https://www.dataiku.com/product/
- Jupyter Notebook https://jupyter.org/
- Anconda https://www.anaconda.com/distribution/
- Github https://github.com/topics/machine-learning

- EPI Results https://epi.envirocenter.yale.edu/epi-topline
- Visitors by Country https://www.indexmundi.com/facts/indicators/ST.INT.ARVL/rankings
- Health Care Systems in World https://ceoworld.biz/2019/08/05/revealed-countries-with-the-best-health-care-systems-2019/
- Cost of Living Comparison https://www.worlddata.info/cost-of-living.php
- Social Interaction and Prosperity Index https://www.prosperity.com/rankings
- Human Freedom Index https://www.cato.org/human-freedom-index-new
- GDP per Capita https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita
- Temperature and Humidity Data https://www.accuweather.com/

Disclaimer for Project Report

- This report is an academic study for capstone project/IBM Data Scientist Course.
- Multiple instructions and script codes from Github, Dataiku, JMP, SKLearn, IBM data science courses were adapted to analyze COVID-19 data. No original source were written for this project.
- ‘COVID-19 data’ file was created by using multiple resources from Google survey. No verification on authenticity of data retrieved from Google was performed.
- Results of analysis are inferences based on available/used data and its analysis using machine learning algorithms. Prior to any usage, or making any conclusions, these results need field, or experimental validation, and modification of results based on new evidences, or corrected data sets.