COVID-19: 100 Cases

Predicting the number of days until 100 cases of COVID-19 occur using disease, nutrition, and social factors.

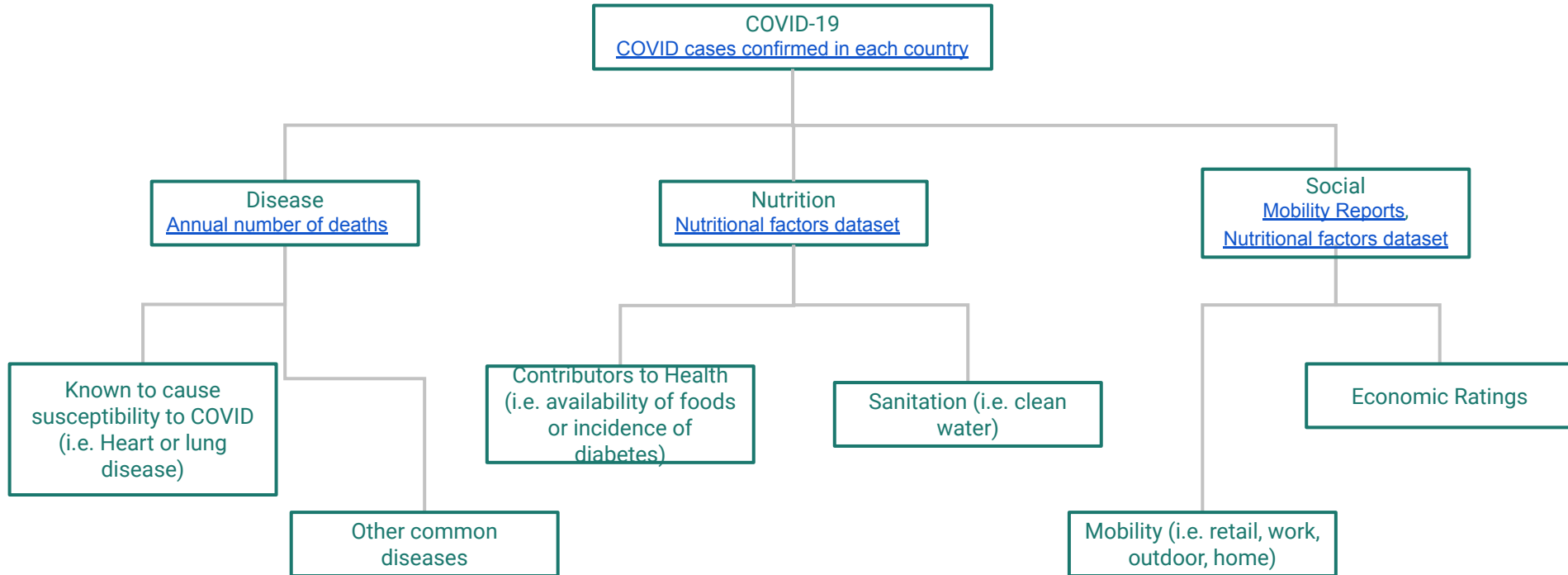# Can we predict when 100 cases of COVID will occur?

The COVID-19 virus has spread worldwide and changed the way everyone lives.

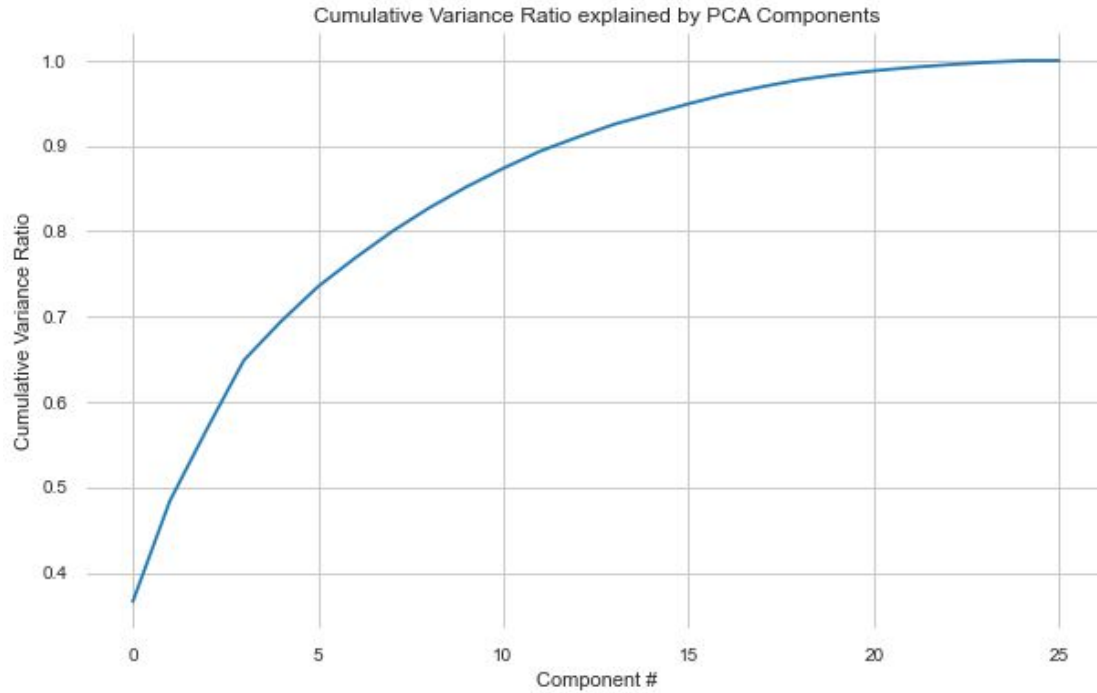The occurrence of 100 days of COVID correlates to rates of the disease later on.

What factors affect when a country reaches 100 cases?

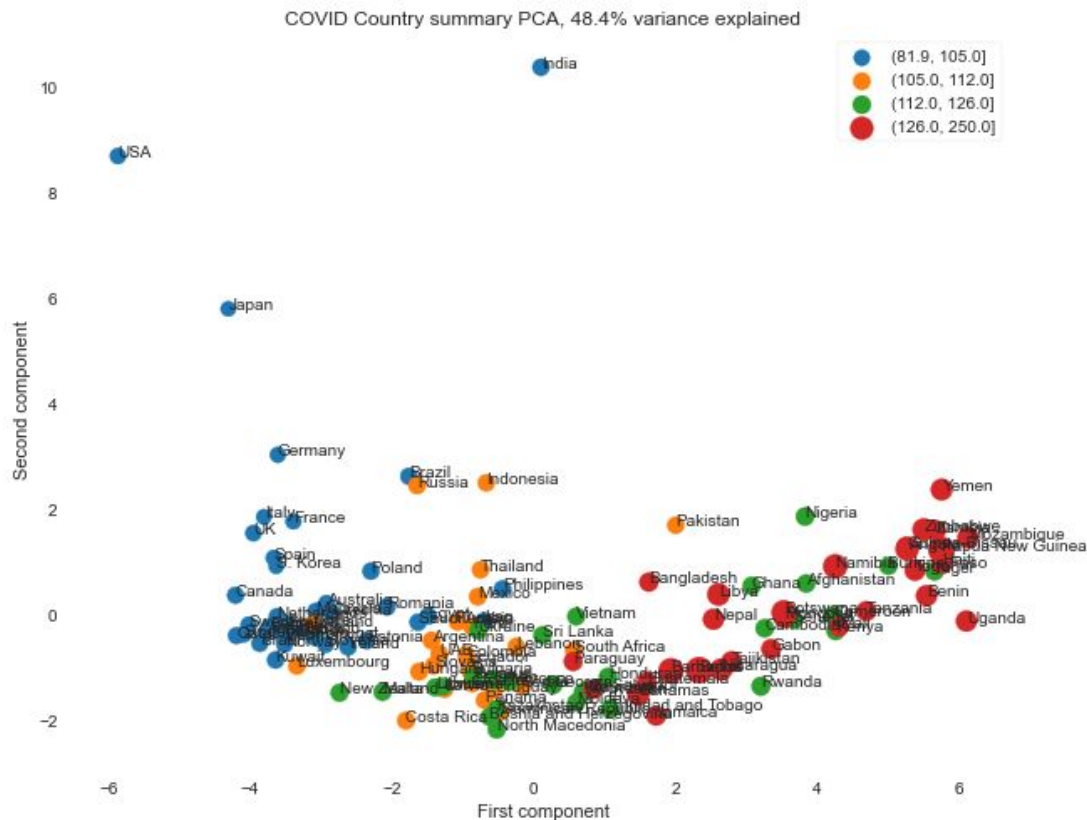Can we predict when a country will reach 100 cases?

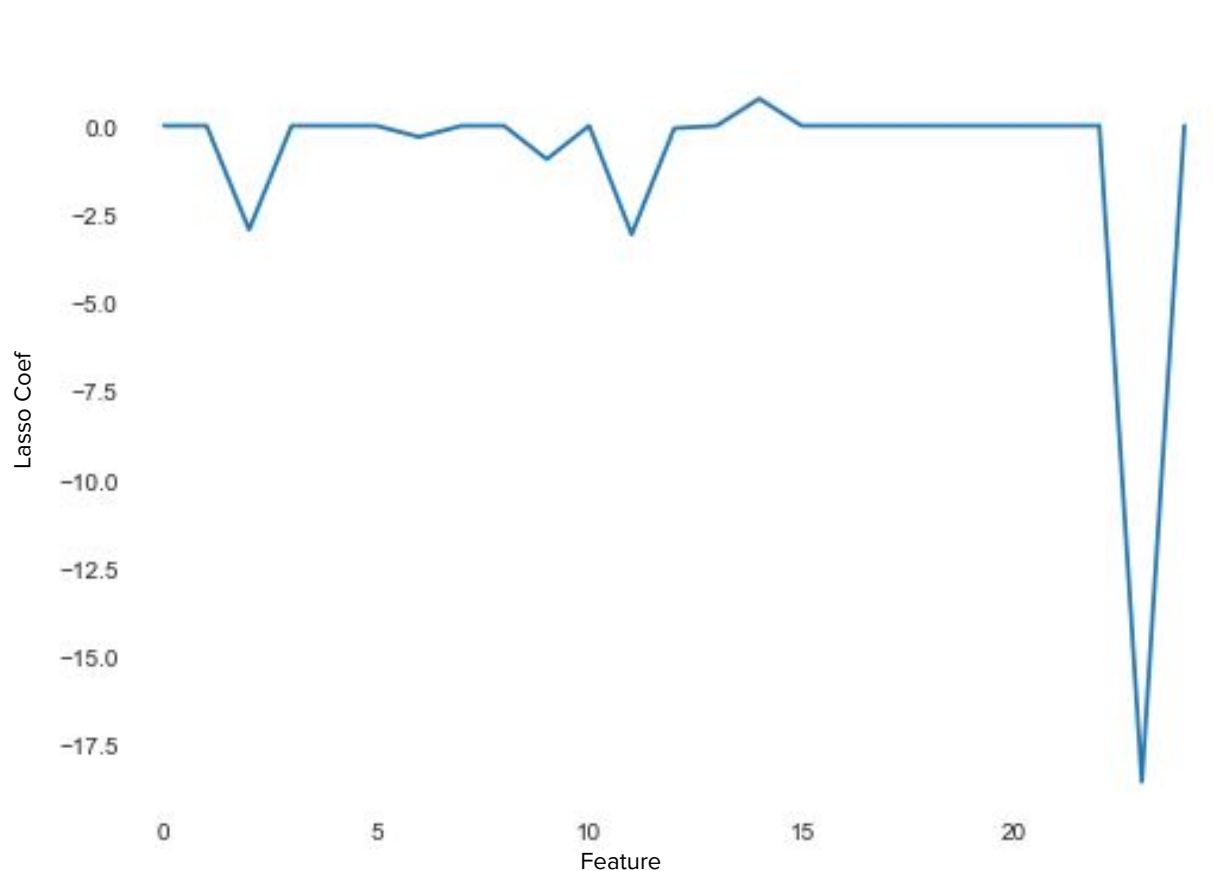# What factors might result in more COVID infection in a country?

COVID-19
COVID cases confirmed in each country

Disease
Annual number of deaths

Nutrition
Nutritional factors dataset

Social
Mobility Reports,
Nutritional factors dataset

Known to cause susceptibility to COVID (i.e. Heart or lung disease)

Other common diseases

Contributors to Health (i.e. availability of foods or incidence of diabetes)

Sanitation (i.e. clean water)

Economic Ratings

Mobility (i.e. retail, work, outdoor, home)

What data is actually important?

Cumulative Variance Ratio explained by PCA Components

What data is actually important?

COVID Country summary PCA, 48.4% variance explained

What data is actually important?

	feature	lasso_coef
0	adult_anemia_2016	-0.000000
1	adult_blood_pressure_2015	0.000000
2	adult_sodium_2017	-2.939390
3	basic_sanitation_2017	0.000000
4	basic_water_2017	0.000000
5	fruit_veg_availability_2013	-0.000000
6	government_revenue_2020	-0.315790
7	limited_sanitation_2017	0.000000
8	limited_water_2017	0.000000
9	nurses_and_midwives_2016	-0.947067
10	open_defecation_2017	0.000000
11	physicians_2016	-3.070958
12	safely_managed_sanitation_2017	-0.071508
13	safely_managed_water_2017	-0.000000
14	surface_water_2017	0.766957
15	undernourishment_prev_2017	0.000000
16	unimproved_sanitation_2017	0.000000
17	unimproved_water_2017	0.000000
18	workplaces_percent_change_from_baseline	0.000000
19	parkinson_disease_deaths	-0.000000
20	dementia_deaths	-0.000000
21	cancers_deaths	-0.000002
22	economic_rating	-0.000000
23	macroeconomic_rating	-18.564933
24	cases_per_cap	-0.000000

Model summary

Model	Score	Explained Variance Score	Mean Absolute Error	Root Mean Square
Linear Regression	0.78	0.80	9.11	20.60
Gradient Boosting	0.85	1.0	0.0	18.51
Random Forest	0.85	0.99	0.13	30.93

Linear Regression is the worst model and Gradient Boosting is the best model.