Could COVID Kill You?

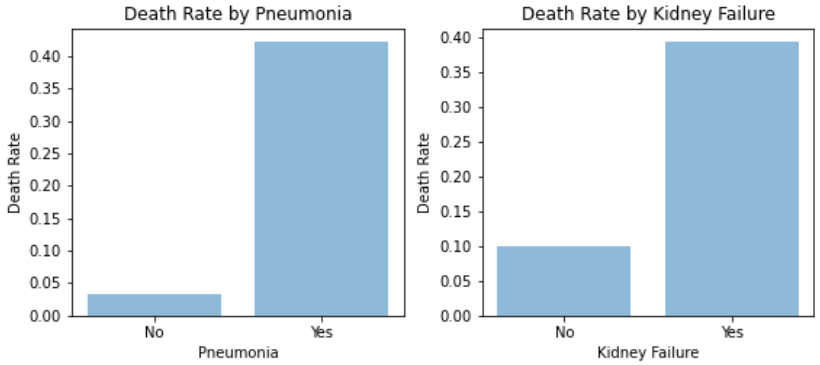
By: David Harper

Project Objectives

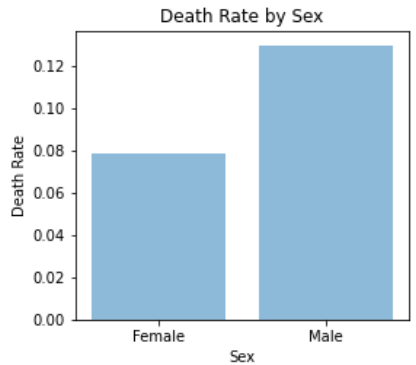
- Apply machine learning classification techniques toward predicting the binary outcome of life or death for somebody infected with COVID.

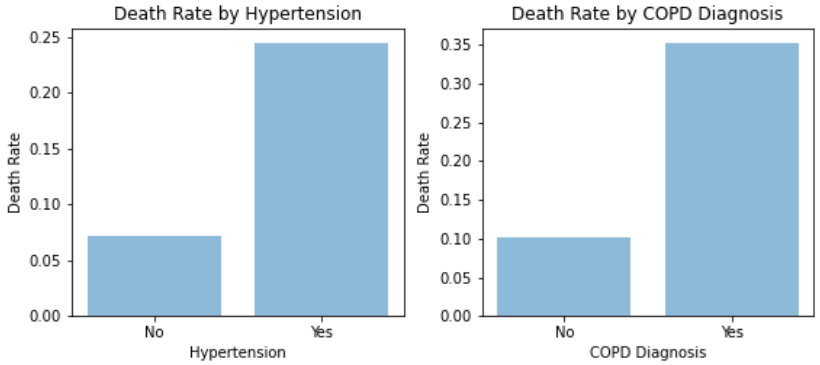
Data

- The data used in this project comes from the Mexico Government.
- Includes Risk Factors
- Includes Residency information

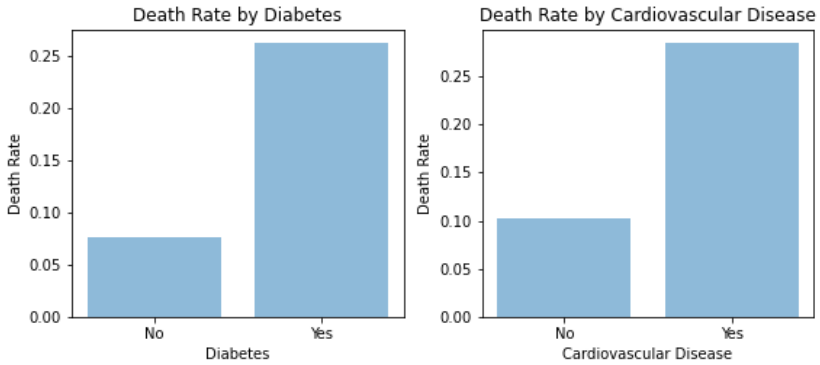

Death Rate by Pneumonia
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
No
Yes
Pneumonia

Death Rate by Kidney Failure
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
No
Yes
Kidney Failure

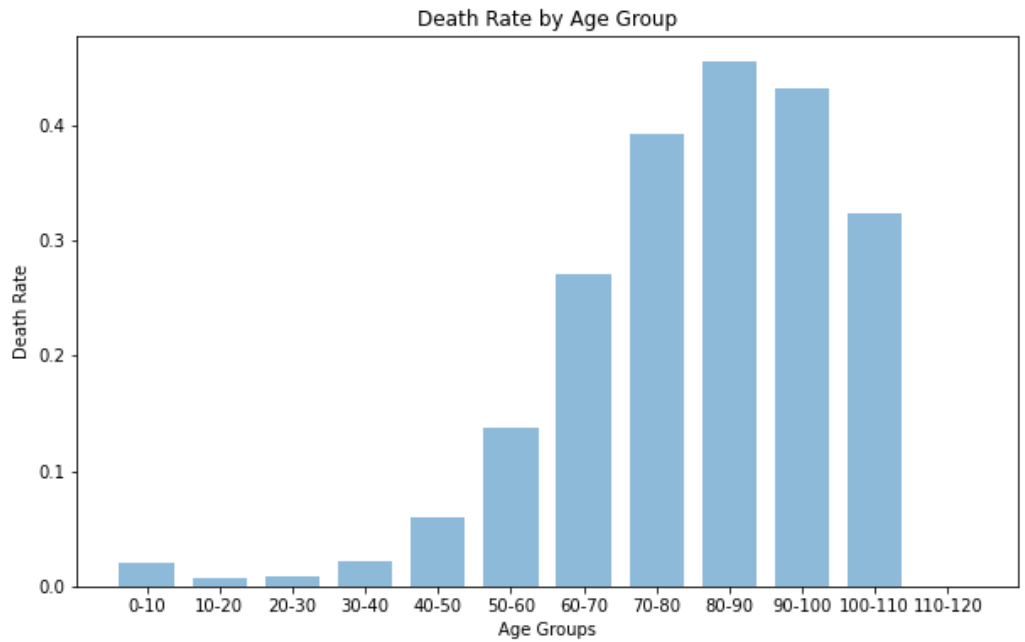

Death Rate by Sex
Death Rate
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Female
Male
Sex

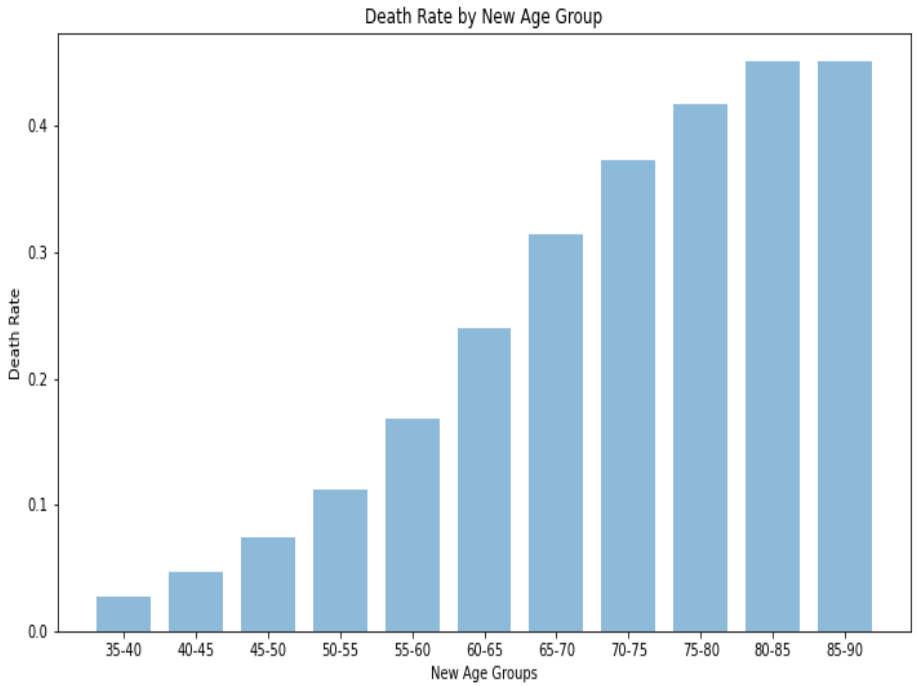

Death Rate by Hypertension
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
No
Yes
Hypertension

Death Rate by COPD Diagnosis
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
No
Yes
COPD Diagnosis


Death Rate by Diabetes
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
No
Yes
Diabetes

Death Rate by Cardiovascular Disease
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
No
Yes
Cardiovascular Disease


Death Rate by Age Group
Death Rate
0.0
0.1
0.2
0.3
0.4
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
100-110
110-120
Age Groups



	New Age	Pneumonia	Diabetes	Hypertension	Cardiovascular Disease	Obesity	Kidney Failure
New Age	1.000000	0.347417	0.344362	0.416245	0.142152	0.047360	0.096741
Pneumonia	0.347417	1.000000	0.218480	0.203332	0.074159	0.075989	0.099744
Diabetes	0.344362	0.218480	1.000000	0.375544	0.099436	0.103694	0.152789
Hypertension	0.416245	0.203332	0.375544	1.000000	0.151755	0.153008	0.171496
Cardiovascular Disease	0.142152	0.074159	0.099436	0.151755	1.000000	0.049884	0.093428
Obesity	0.047360	0.075989	0.103694	0.153008	0.049884	1.000000	0.009059
Kidney Failure	0.096741	0.099744	0.152789	0.171496	0.093428	0.009059	1.000000

All Considered Models

Model Type	Accuracy	Precision	Recall	Accuracy Std	AUCs
Forward Selection	0.907242	0.597784	0.361092	5.453823e-04	0.912943
Backward Selection	0.907738	0.595481	0.383066	6.230909e-04	0.910598
L2 Regularization	0.907362	0.593598	0.378197	6.415157e-04	0.914667
Three Leaved Tree	0.904084	0.545436	0.528804	3.178006e-07	0.821973
Pruned Tree	0.904084	0.545436	0.528804	6.002595e-04	0.874612
Random Forest	0.898631	0.638752	0.083476	4.469809e-04	0.909334

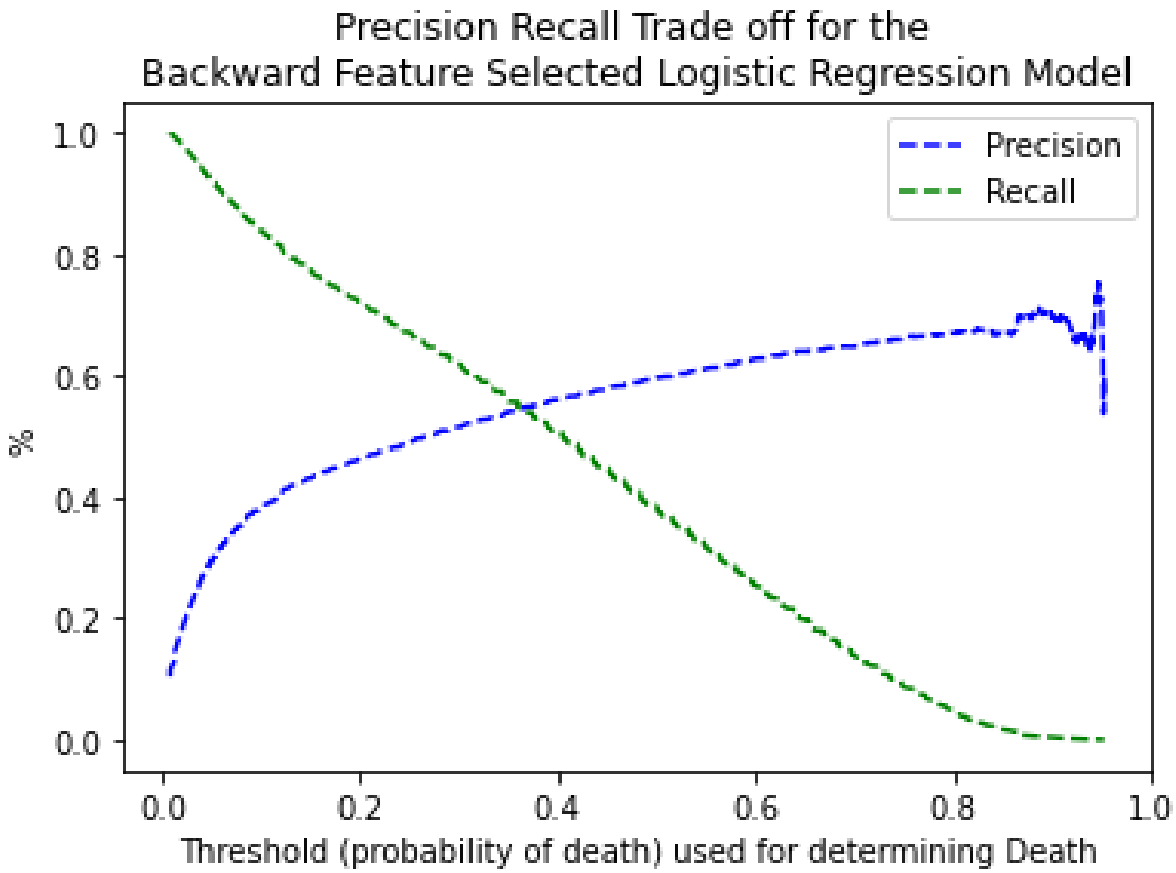
Backward Selection Logistic Regression Model

Features	Coefficients
Pneumonia	2.498118
Kidney Failure	1.132058
New Age	0.941551
Sex_1	-1.439251
Sex_2	-1.008705

Backward Selection provides the simplest model and yields top accuracy.

Test Accuracy of 90.5%
Test Precision of 59.7%
Test Recall of 36.2%

Precision Recall Trade Off



Interpretation of Model Strength

Future Research

More data features

Different groups of interest