Could COVID Kill You?

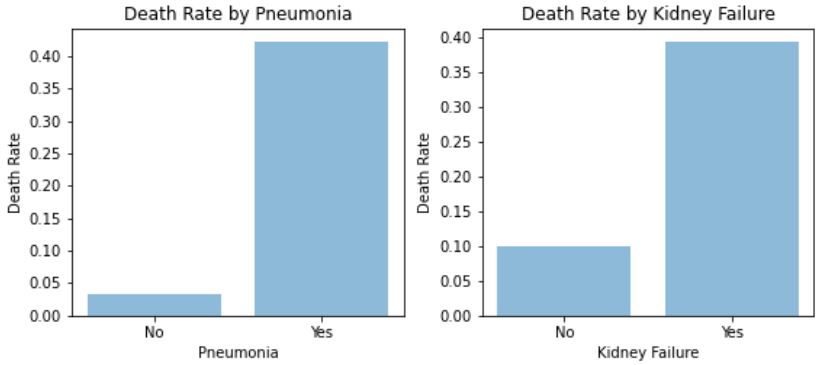
By: David Harper

Project Objectives

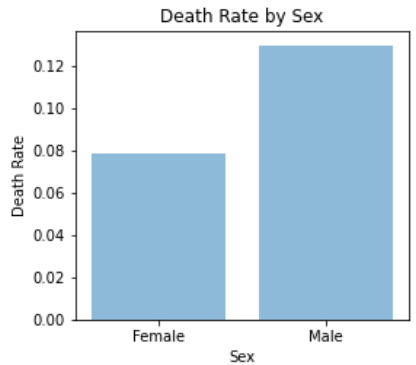
- Apply machine learning classification techniques toward predicting the binary outcome of life or death for somebody infected with COVID.

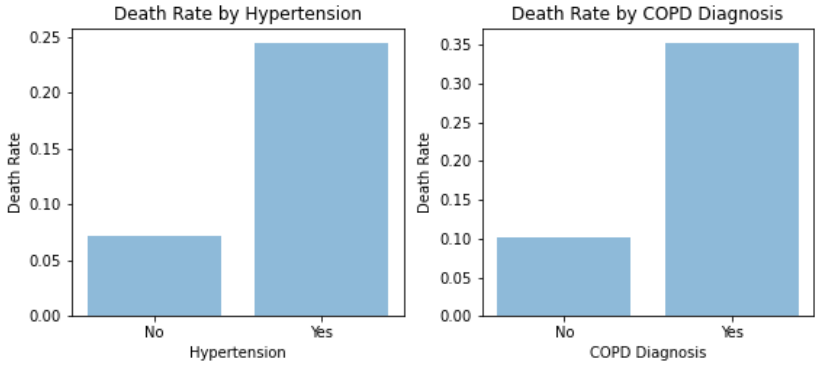
Data

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

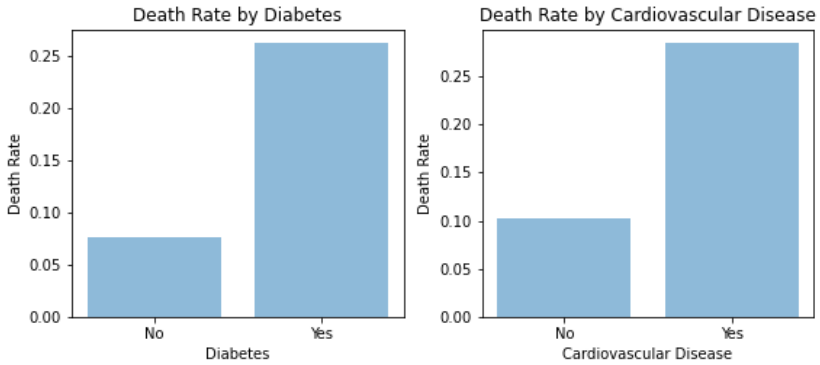

Death Rate by Pneumonia
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
No
Yes
Pneumonia

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

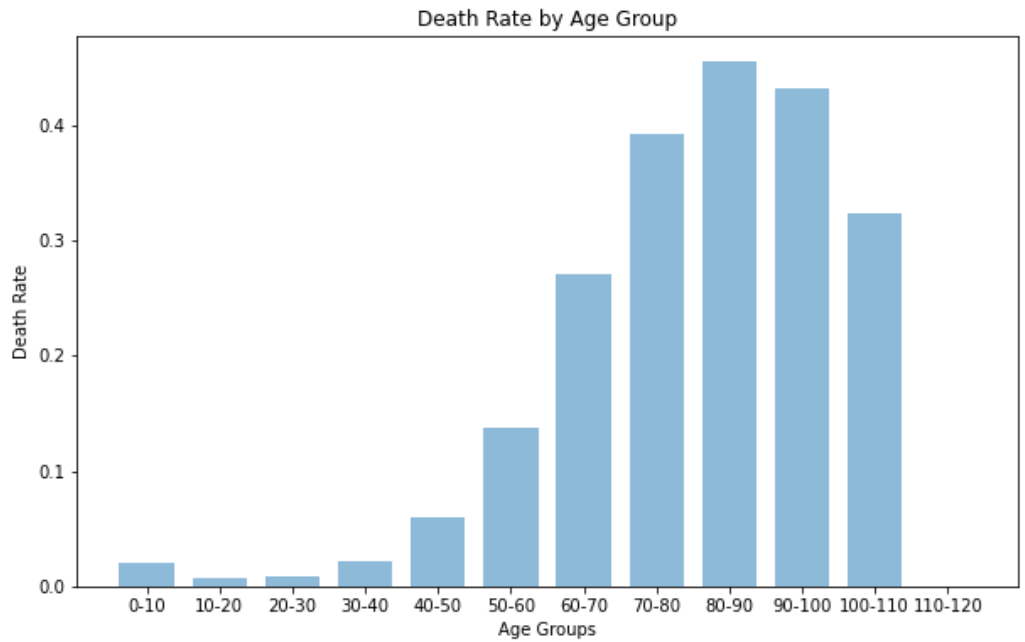

Death Rate by Sex
Death Rate
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Female
Male
Sex


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

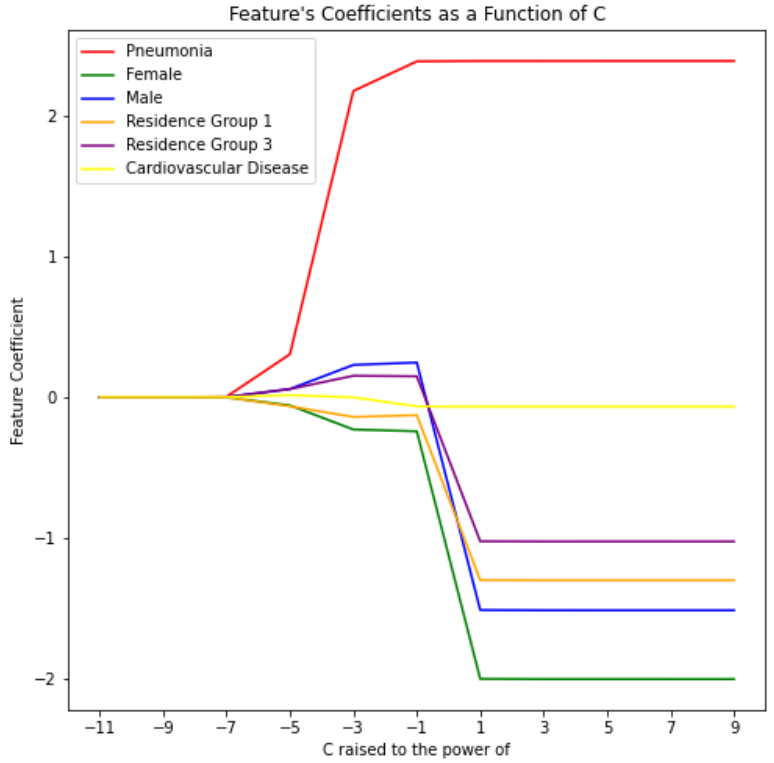
Death Rate by COPD Diagnosis
Death Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
No
Yes
COPD Diagnosis


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

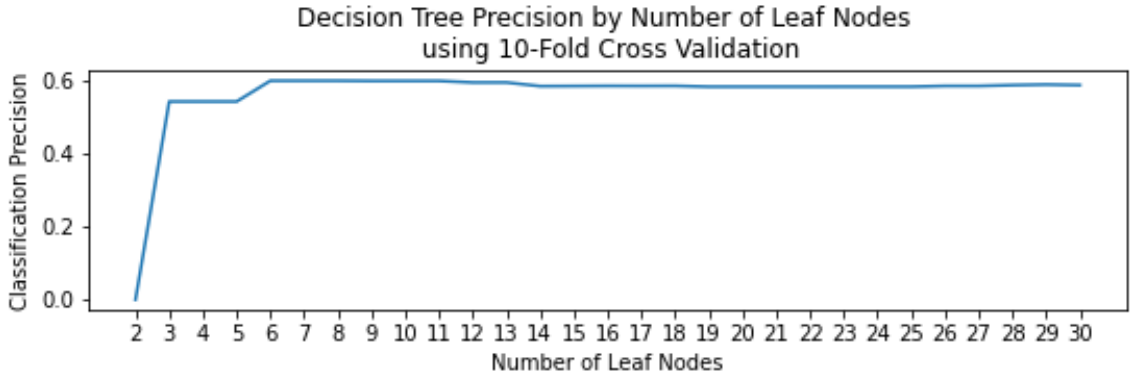

Death Rate by Age Group
Death Rate
0.0
0.1
0.2
0.3
0.4
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
100-110
110-120
Age Groups

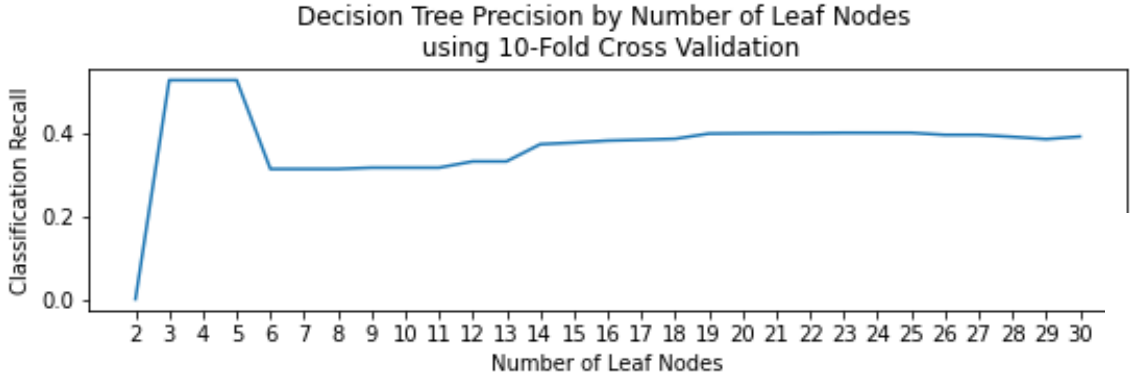
L2 Regularization Logistic Regression Model

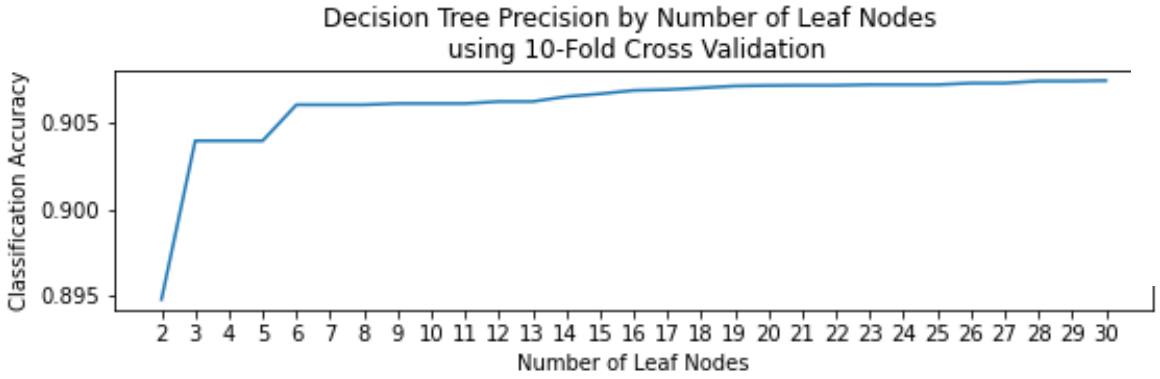


Regularization had the greatest impact on the sizing and interpretation of the coefficients.

Accuracy of 90.6%
Precision of 60.2%
Recall of 35.5%

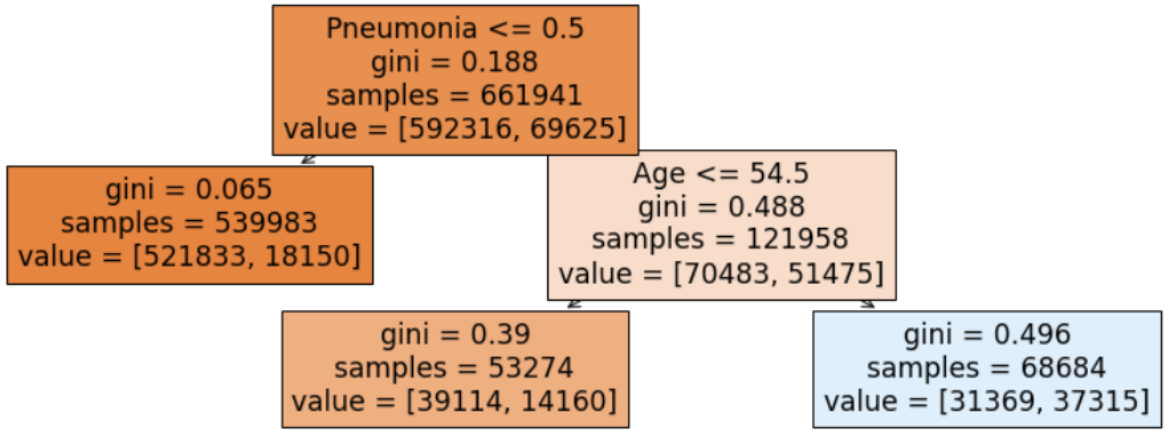






Decision Tree

Simplest tree with maximized test metrics is a tree with only 3 leaf nodes



Accuracy of 90.3%
Precision of 54.6%
Recall of 52.9%

Interpretation of Model Strength

Which Model to Use?

- **L2 Regularization Model**
- **Decision Tree Model**

Future Research

More data features

Different groups of interest