Using Machine learning for insight on vaccine distribution

Covid-19 Who should get vaccines and why?

Jephte Estiverne

Introduction

In commencement, the analysist of this paper aims to address the problem of vaccine distribution in the age of pandemic. The hypothesis emphasizes the need to distribute vaccines not by states but places of accumulative deaths. To address this, data was collected from the CDC (The Centers for Disease Control and Prevention). The first column in our database is the state. Listing state name in alphabetic order. Our Second column vector is the Place of Death. The categorical order are :(Healthcare Setting, Inpatient), (Healthcare Setting, Outpatient or Emergency Room), (Healthcare Setting, Dead on Arrival), (Decedent's Home), (Hospice Facility), (Nursing Home or Long Term Care Facility), (Other), (Place of Death Unknown), s.t. we have a total of 12 categorical variables. For the analysis, I only used the total death column. I then convert the column into a binary function. For total death equal to one or greater is 1, and otherwise, the value is zero. In commencement, this paper aims to address the problem of vaccines distribution in the age of pandemic. To address this, data was collected from the CDC (The Centers for Disease Control and Prevention).

Methods

In our categorial column listed above, we can observe the term "Healthcare Setting" as a column vector value mode. The column variables are then converted to a unique number ranging 1:8. Next, we obtain a correlation heatmap. As we can observe from the heatmap, we notice a .33 value for the column Places/Place of Death. Our value indicates an increase in additional death is positively related to the setting of why it occurs. In other words, we obtain cross-validation on our dataset.

Next, we split our dataset to train, test, and validate proportionately in (249, 3), (83, 3), and (83, 3). We then the death binary column through all three variables. Resulting in x_train, y_train, x_test, y_test, x-validate, and y_validate. Next, we use the logistic model and test X_train and y_train. We end up with a value of 1, which makes sense because the variable of interest is quite definitive. Remember that our inquiry is to measure the importance of types of facilities.

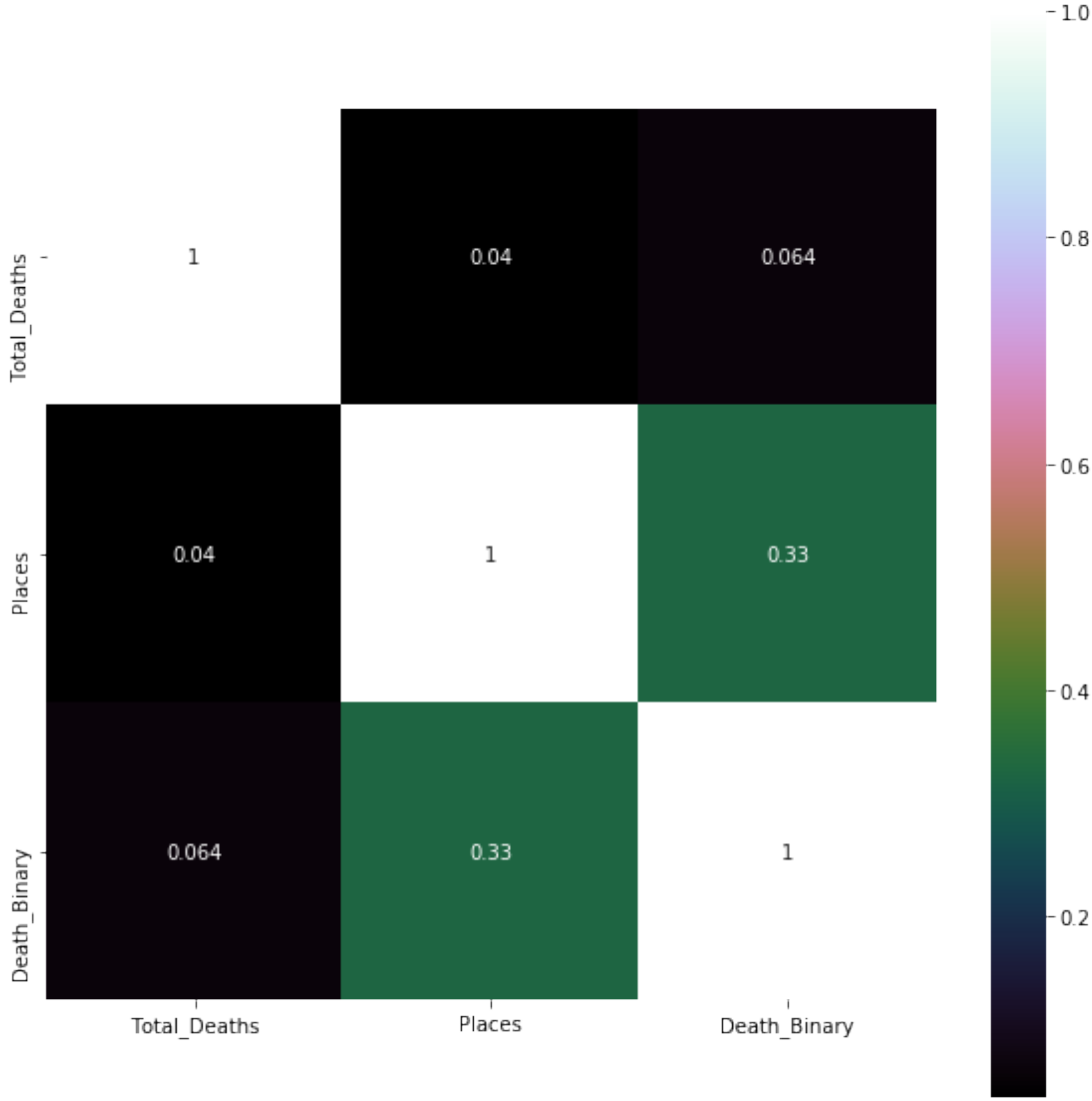
To further scrutinize the categorical, we use the kMeans pack from sklearn. We create 3 clusters then split the sample dataset into 10 sam[ples. In like manners, we obtain heatmap cross-validation and correlation. We see that total death has a positive correlation with states. A string quality of this plot is, for instance: holding everything constant, a vaccine that is distributed accordingly with the place of death we can decrease total death at a much rapid pace than focusing on states alone.

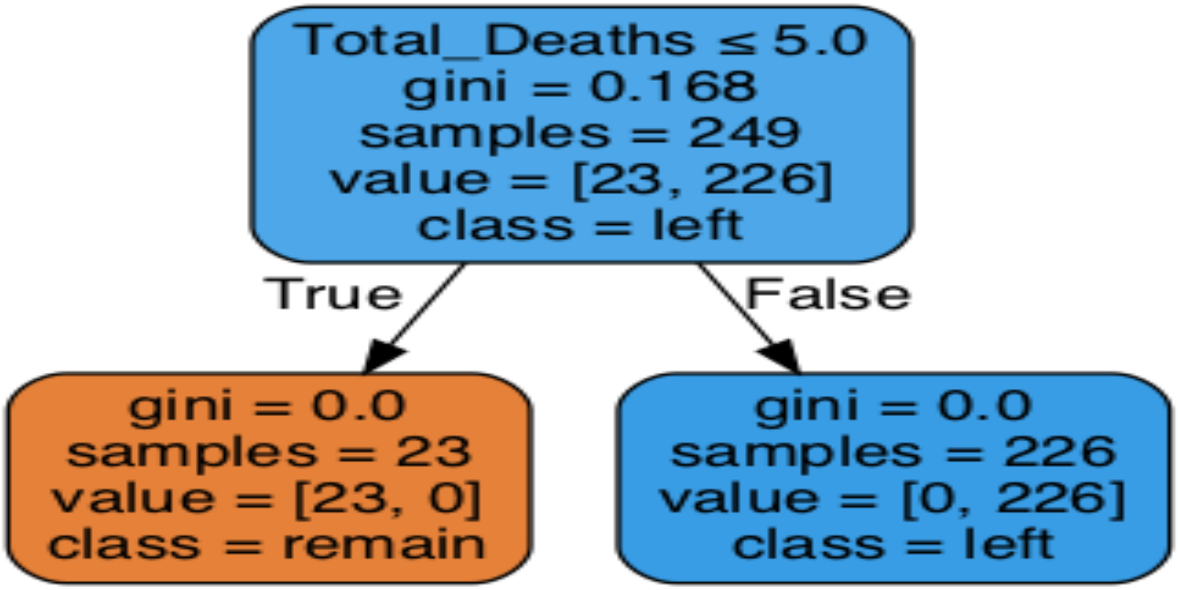
Summary

Our machine learning models reject the null hypothesis. Our findings strongly emphasize that policymakers should be focused on a strict strategic distribution process. Focusing on places of death is the best route for the given dataset.

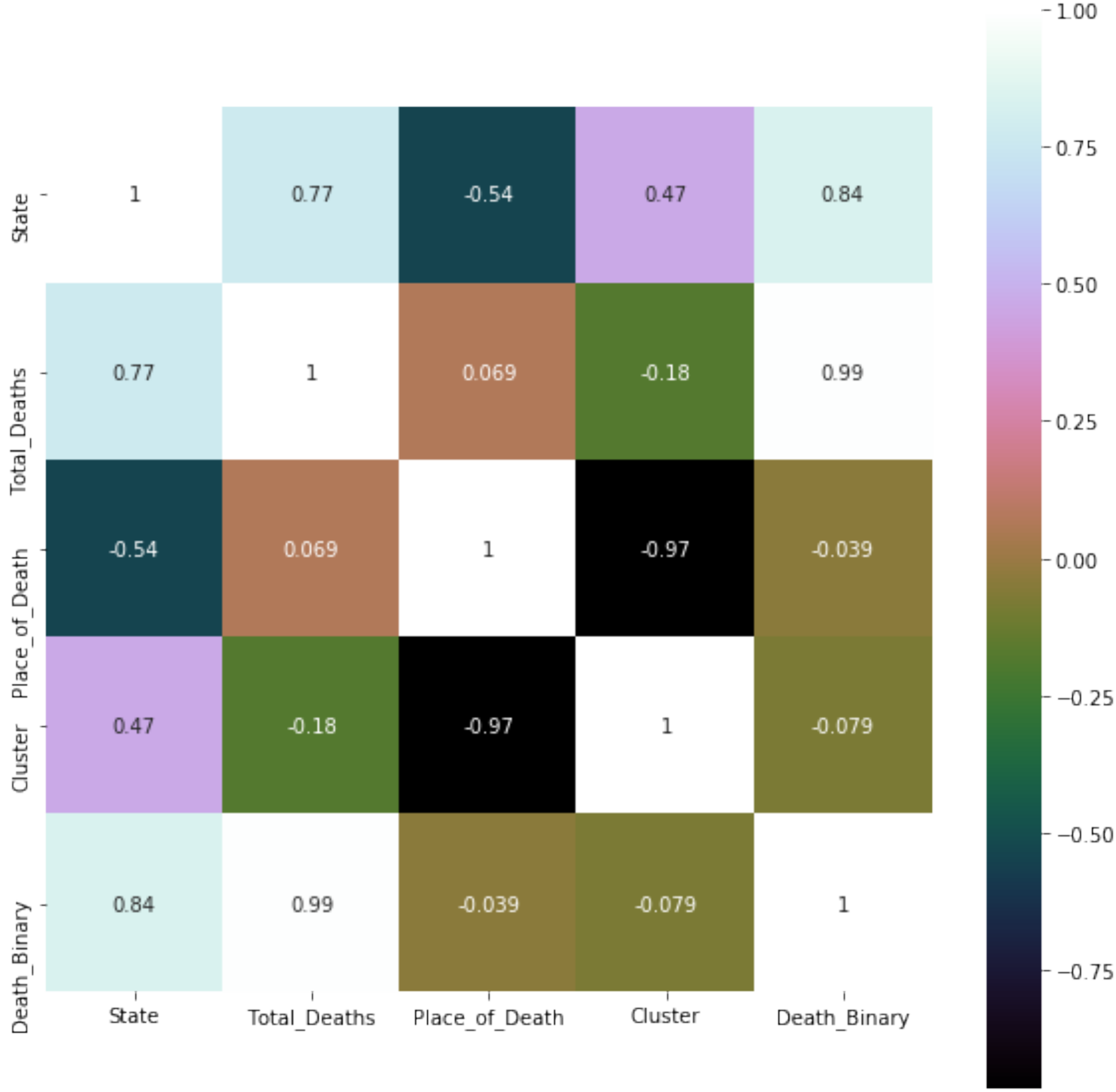
Tables Plots and Graphs

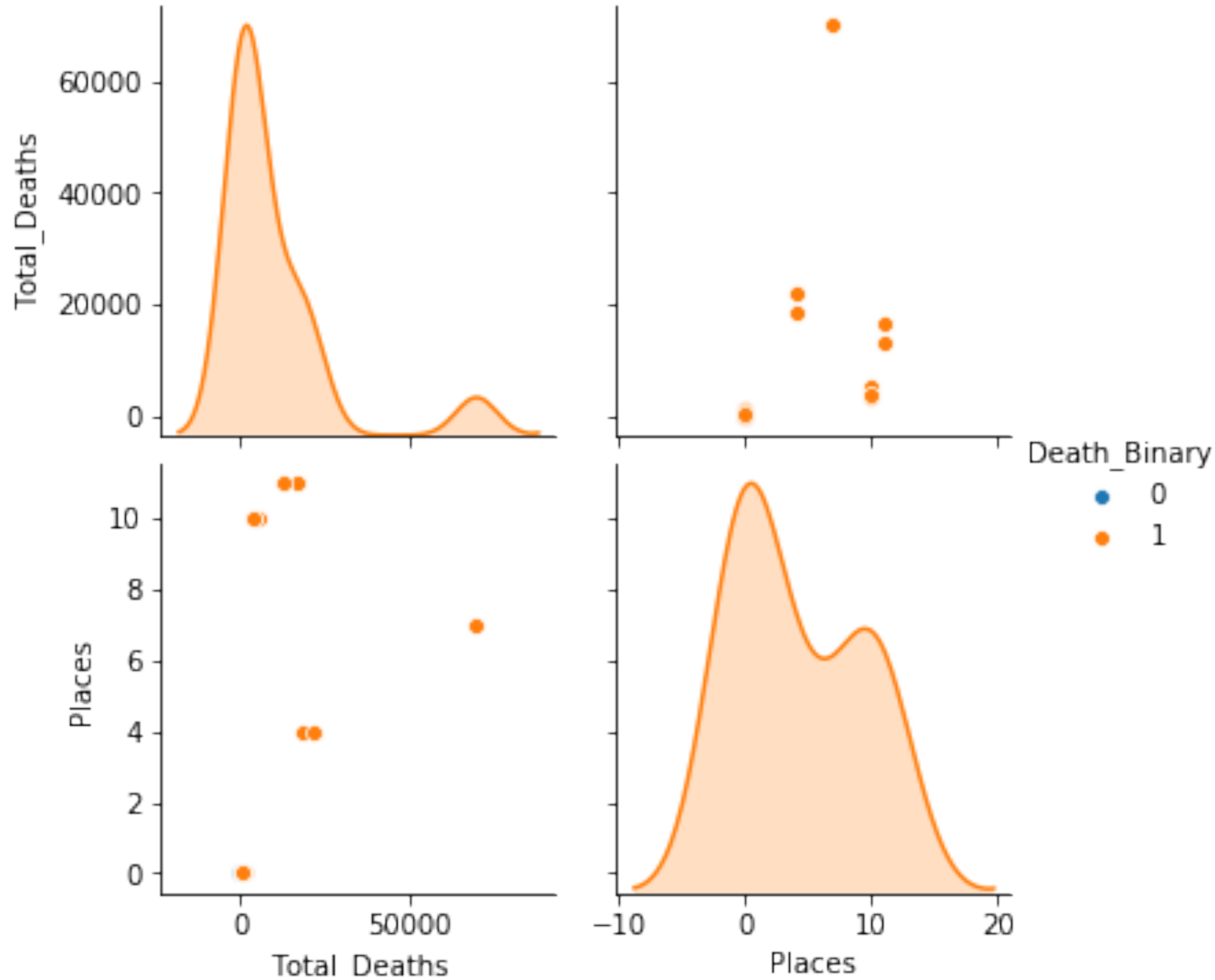
Total_Deaths	**Places**	**Death_Binary**	
0	716707	1	1
1	145939	5	1
2	7013	8	1
3	833215	6	1
4	153348	5	1
5	424540	2	1

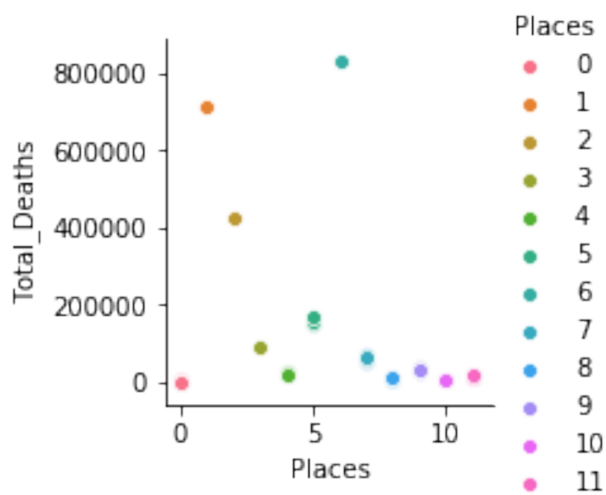

Total_Deaths
Places
Death_Binary
1
0.04
0.064
0.04
1
0.33
0.064
0.33
1
Total_Deaths
Places
Death_Binary
1.0
0.8
0.6
0.4
0.2



	State	Total_Deaths	Place_of_Death	Cluster	Death_Binary
0	-7.602652	-6.231181	1.559039	1	0
1	-9.170452	-6.598017	-0.143467	1	0
2	0.766870	4.338255	3.862665	0	1
3	9.557854	4.674793	-6.761902	2	1
4	9.031327	4.284545	-5.876104	2	1
5	-8.380741	-7.964337	-0.970486	1	0


State
Total_Deaths
Place_of_Death
Cluster
Death_Binary
1
0.77
-0.54
0.47
0.84
0.069
-0.18
0.99
-0.97
-0.039
-0.079
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Total_Deaths
Places
Death_Binary
0
1
60000
40000
20000
0
10
8
6
4
2
0
50000
−10
20


Places
Total_Deaths
800000
600000
400000
200000
0
5
10
0
1
2
3
4
5
6
7
8
9
10
11

```
Call:
lm(formula = Total_Deaths ~ Place_of_Death, data = Sample_Data)

Residuals:
   Min     1Q Median     3Q    Max 
-30860 -10145  -3997   -150 802355 

Coefficients:
                                                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                                       30860       8014   3.851 0.000137 ***
Place_of_DeathHealthcare Setting, Dead on Arrival                -30574      11618  -2.632 0.008824 ** 
Place_of_DeathHealthcare Setting, Inpatient                       -4315      11333  -0.381 0.703581    
Place_of_DeathHealthcare Setting, Outpatient or Emergency Room   -25455      11333  -2.246 0.025235 *  
Place_of_DeathHospice Facility                                   -25180      11333  -2.222 0.026842 *  
Place_of_DeathNursing Home or Long Term Care Facility            -15136      11333  -1.336 0.182431    
Place_of_DeathOther                                              -24629      11333  -2.173 0.030343 *  
Place_of_DeathPlace of Death Unknown                             -30709      12115  -2.535 0.011627 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 58890 on 407 degrees of freedom
Multiple R-squared:  0.03441,	Adjusted R-squared:  0.0178 
F-statistic: 2.072 on 7 and 407 DF,  p-value: 0.0455
```

=