Saudi Containment and Closure Policies effects on COVID-19 Pandemic

1. Introduction

Health care is one of the important sectors in life, and Saudi Arabia had given the health sector a priority before even covid started. In the Saudi 2030 vision they developed a Transformation Program, that aims to restructure the health sector in the Kingdom to be a comprehensive, effective, and integrated health system that is based on the health of the individual and society (including the citizen, the resident, and the visitor).

It also aims to improve access to health services through optimal coverage and a comprehensive and equitable geographical distribution, expanding the provision of e-health services and digital solutions, as well as improving the quality of health services.

The spread of coronavirus has become a concern to everyone over the past couple years, the many ICU cases and deaths because of this virus! The ministry of health has set different policies to contain the spread of this virus and we are here to see is it really affecting the number of deaths? Which policy is more effective?

2. Exploratory Data Analysis (EDA)

Plotting our data visually to summarize how the Saudi containment policies effects on the number of deaths and number of cases.

2.1


New Deaths Vs Total Vaccination
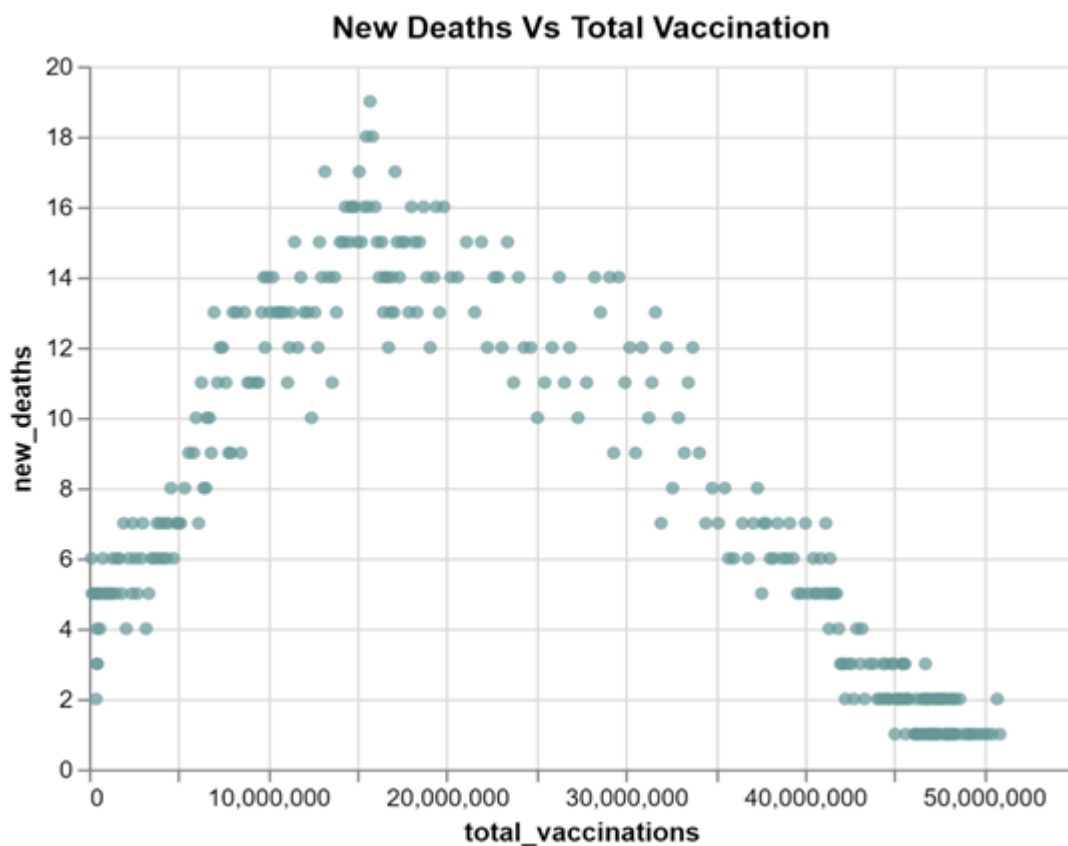
When the number of vaccinations is low, the number of new deaths rises steadily to about 20. This in turn leads to more vaccinations for the people leading to a drop in the number of new deaths. As the total vaccinations approach 30,000,000 the number of deaths drops steadily from 20 and at the total vaccinations reach 50,000,000, the number of deaths has dropped to about 1.

2.2.
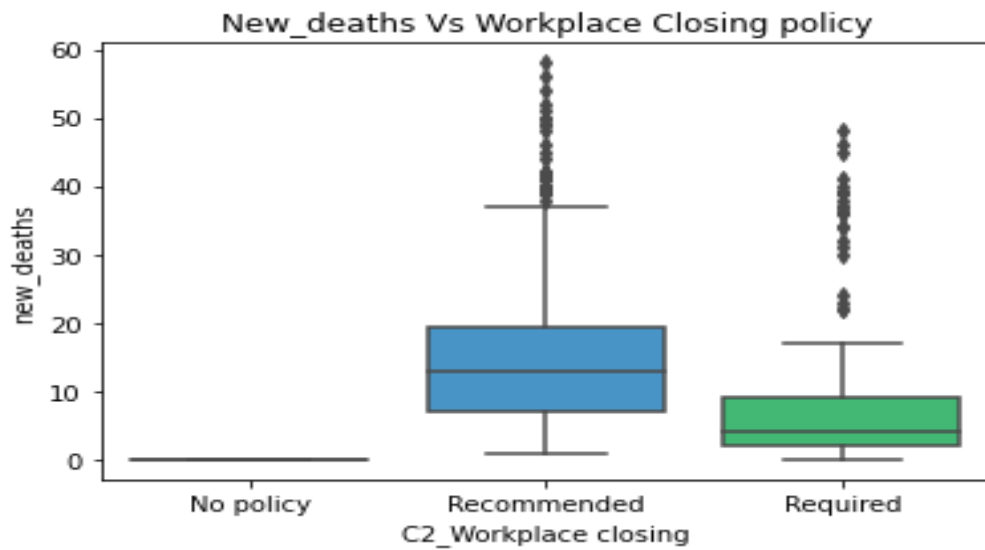
The number of new deaths is low while there is no policy implemented for the workplaces. Afterwards the number of new deaths rises averaging at about 15 new deaths, this leads to recommendations for closing the workplaces. The number of deaths drops when the workplace closing policy is required.

2.3.
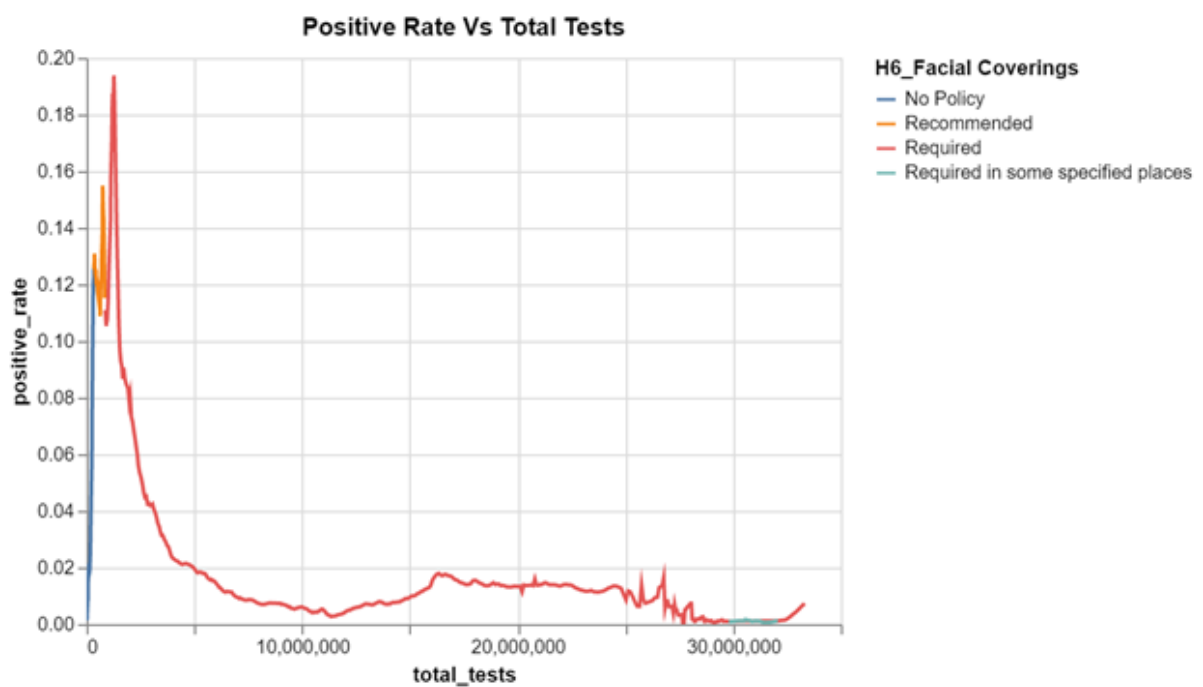
The positivity rate for covid-19 steadily increases when there is no facial coverings policy. The rate goes as high as 0.12 afterwards, facial covering is recommended though not enforced still the positivity rate goes up. This could be because many people ignored the recommendations. When the positivity rate reaches a high of 0.19, the facial coverings policy becomes required, and the positivity rate has steadily declined. When the facial coverings policy becomes required only in some places the positivity rate rises.

In this part of EDA we will take the three main policies that affect on the number of cases and number of deaths.
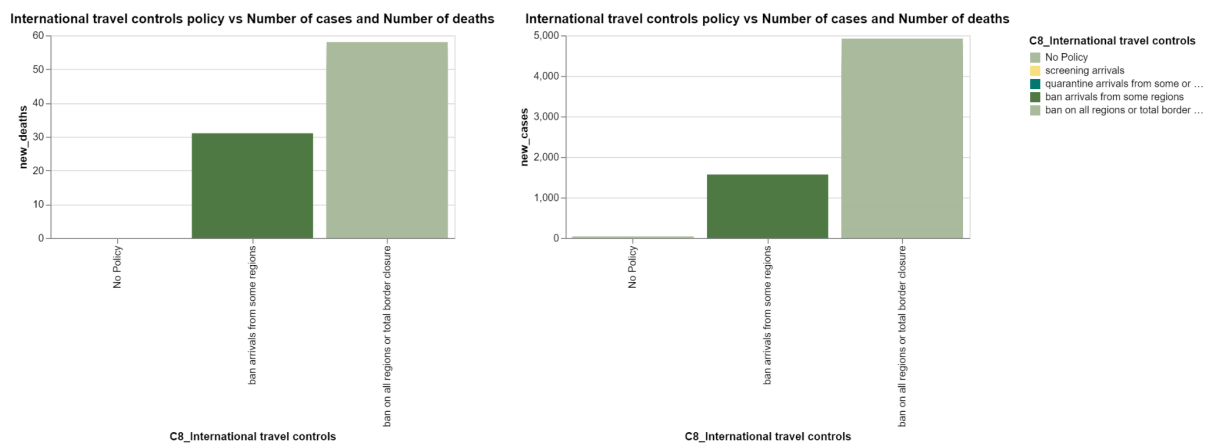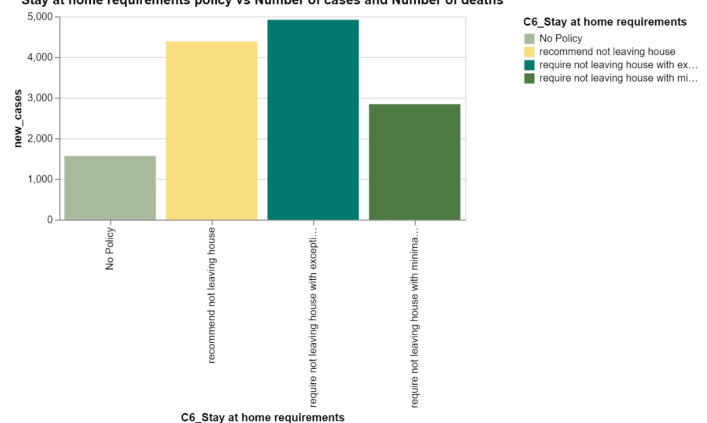

First, When the international travel controls policy is applied and it allows arrivals in all regions the number of cases reaches to 5000 and number of deaths is less than 40. On other hand , When the Saudi government allows travel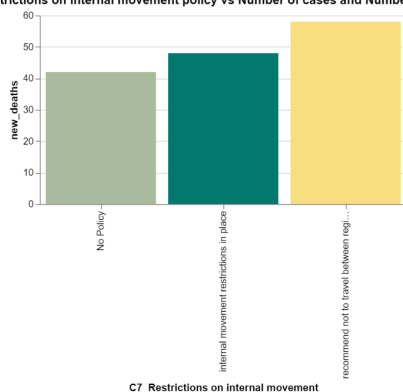 to some regions the number of cases and deaths clearly goes down to less than 2000 in number of cases and 35 in number of deaths.
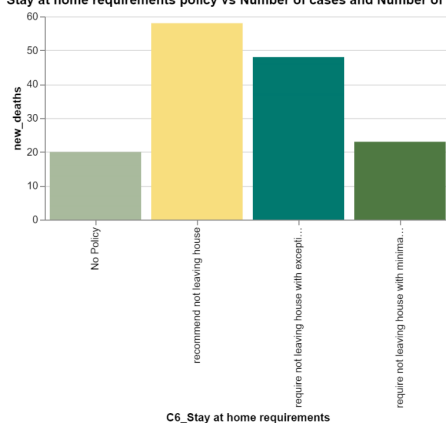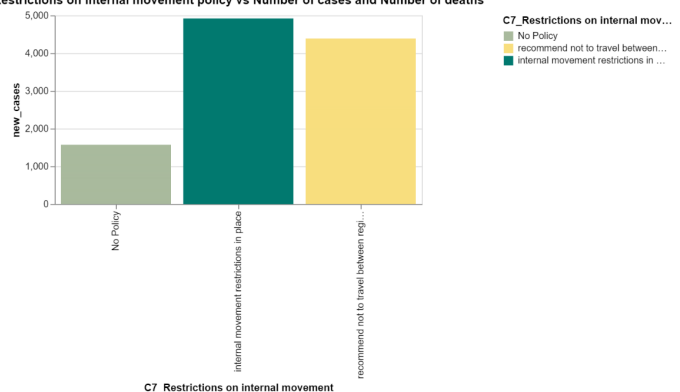
Second, Stay at home requirements policy. It's considered 4 main categories, no policy, recommend not leaving house, require not leaving house with exceptions, and require not leaving house with minimal exceptions. As shown above when recommend not leaving house policy applied the number of deaths was in height level compared with other policies, and the number of cases was high when recommend not leaving house policy applied.

Finally, Restrictions on internal movement policy. When the policy applied to be internal movement restrictions in place the number of cases reached to 5000. On other hand the number of deaths reached the maximum point to be 58.

3.Finding and Results

As our data consists of numerical and categorical features, we have performed two methods to find the correlation between the features and the target. Pearson coefficient was performed for the numerical features where all the features with r coefficient of 0.4 to 1 or -0.4 to -1 would be selected. However, for the

categorical features we have performed ANOVA test to check if the feature category will affect the target. By taking the p-value for each feature, when p-value < 0.05 then the feature will be selected as this indicates the mean of each group in the feature has different effect on the target. However, when p-value > 0.05 we will accept the null hypothesis and remove the feature from the training set.

In order to build the model, the dataset will be splitted 75% for training and 25%. Then we will set a base model and evaluate the model by calculating r2 score and the mean squared error. After that we will use grid search to find the best parameters for the model to gain the highest possible accuracy.

Since our target is a continuous value, regression analysis was used to predict the target. In this section, we applied four models to study which model can record the highest accuracy score and lowest error score. The machine learning algorithms were used in this Project :

- Linear Regression
- Lasso
- Lasso with Grid Search
- Decision Tree
- Decision Tree with Grid Search
- Random Forest

In this process, we have used GridSearchCV to tune parameters for two algorithms. Then all models have been tested. The table below illustrates the results of each model.

Model	r2 Score	MSE
Linear Regression	0.896	3.15
Lasso Lasso with Grid Search	0.890 0.910	3.17
Decision Tree Decision Tree with Grid Search	0.9541 0.9587	2.54
Random Forest	0.9661	2.33

Thus, for model accuracy and sensitivities, we can say that most of the models give a good score. However, The Random Forest and Decision Tree models performed effectively for predicting the number of deaths.