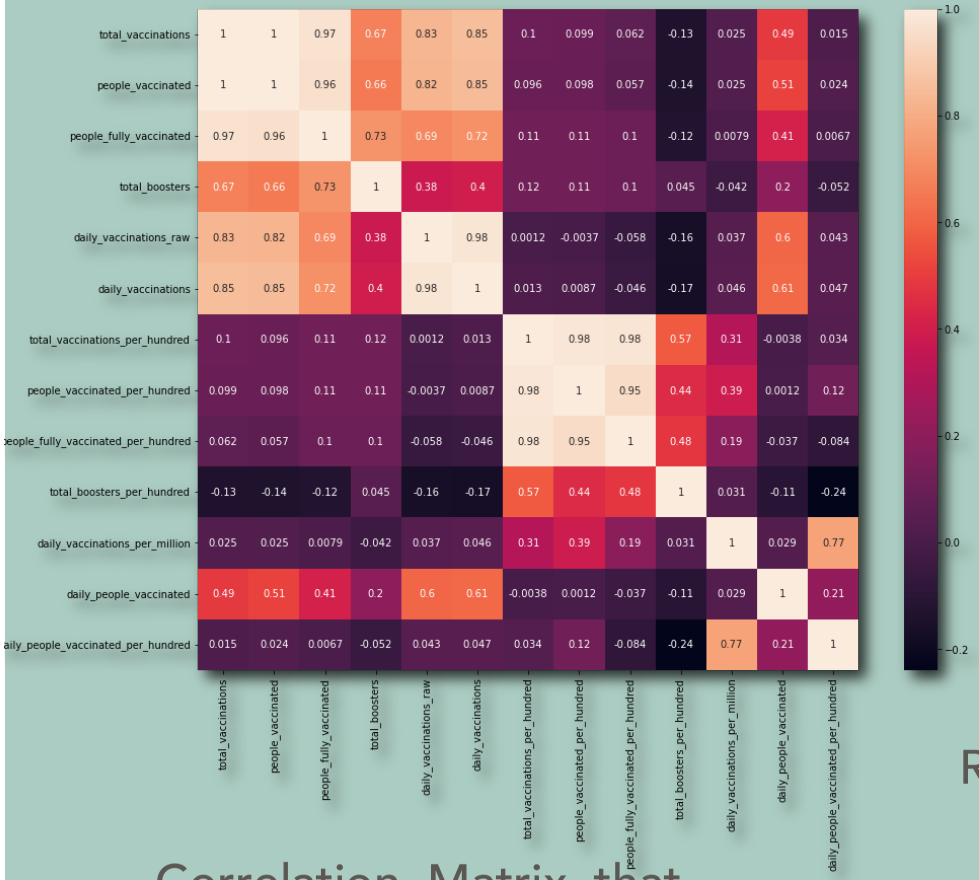
# Data Cleaning

1. Explore the Dataset and check the sum of the Nans
2. Drop the Nans  for the Total Vaccination Feature
3. Check the Correlation Matrix to see which features are correlated
4. Run normality tests to know the Distribution of each correlated feature ignorer to know which Statistical Test to perform
5. Run Statistical Tests on the Correlated Feature
6. Fill out the Nans according to the result of the Statistical Test.

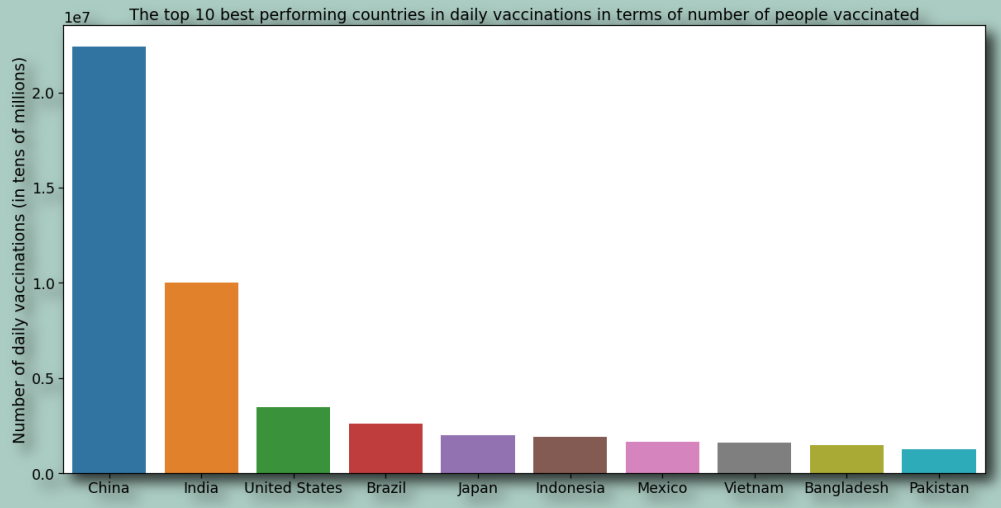


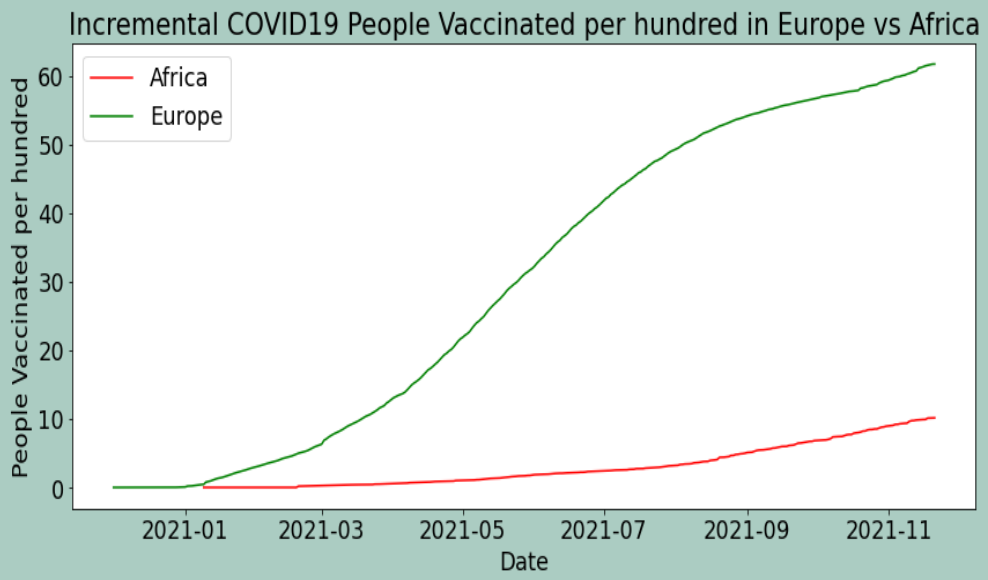

Correlation  Matrix  that shows which features are correlated

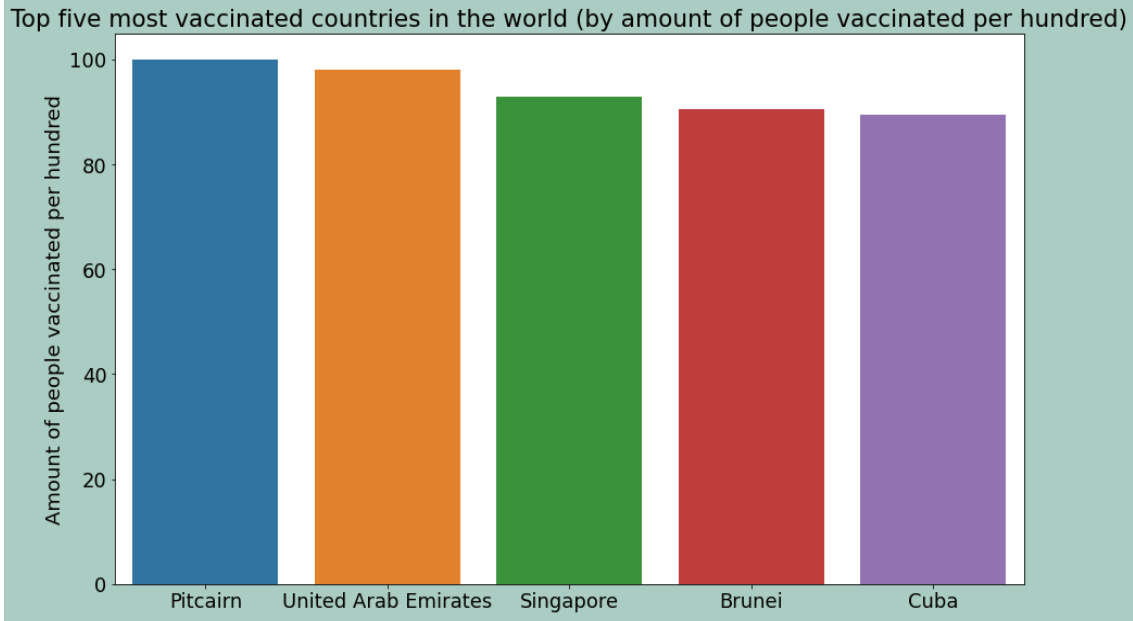
```
location                                0
iso_code                                0
date                                    0
total_vaccinations                      0
people_vaccinated                       0
people_fully_vaccinated                 0
total_boosters                          0
daily_vaccinations_raw                  0
daily_vaccinations                      0
total_vaccinations_per_hundred          0
people_vaccinated_per_hundred           0
people_fully_vaccinated_per_hundred     0
total_boosters_per_hundred              0
daily_vaccinations_per_million          0
daily_people_vaccinated                 0
daily_people_vaccinated_per_hundred     0
dtype: int64
```

Result After Carrying Out Data Cleaning Process

# Exploratory Data Analysis

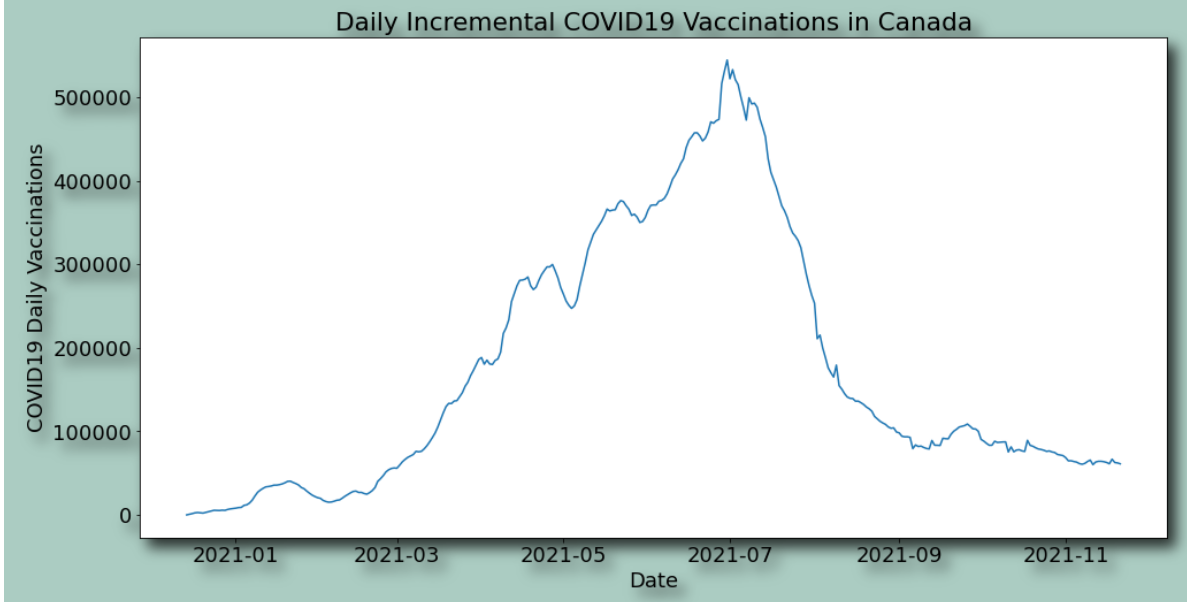


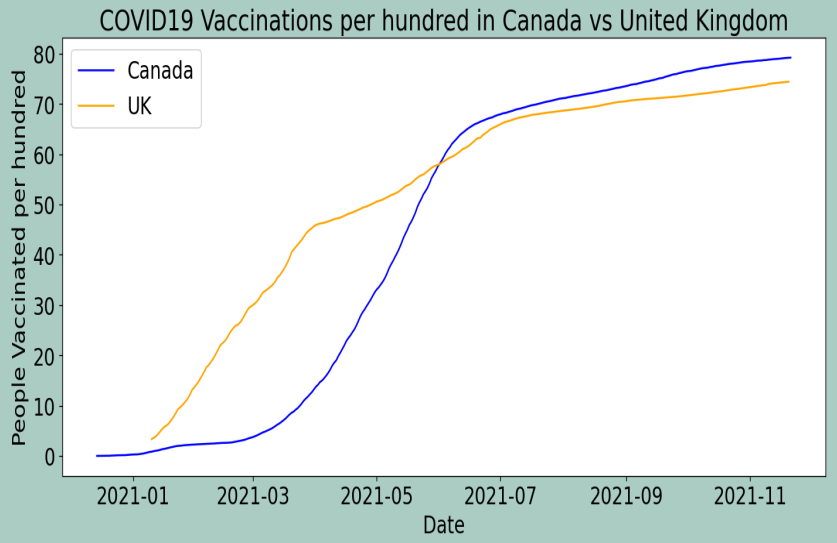



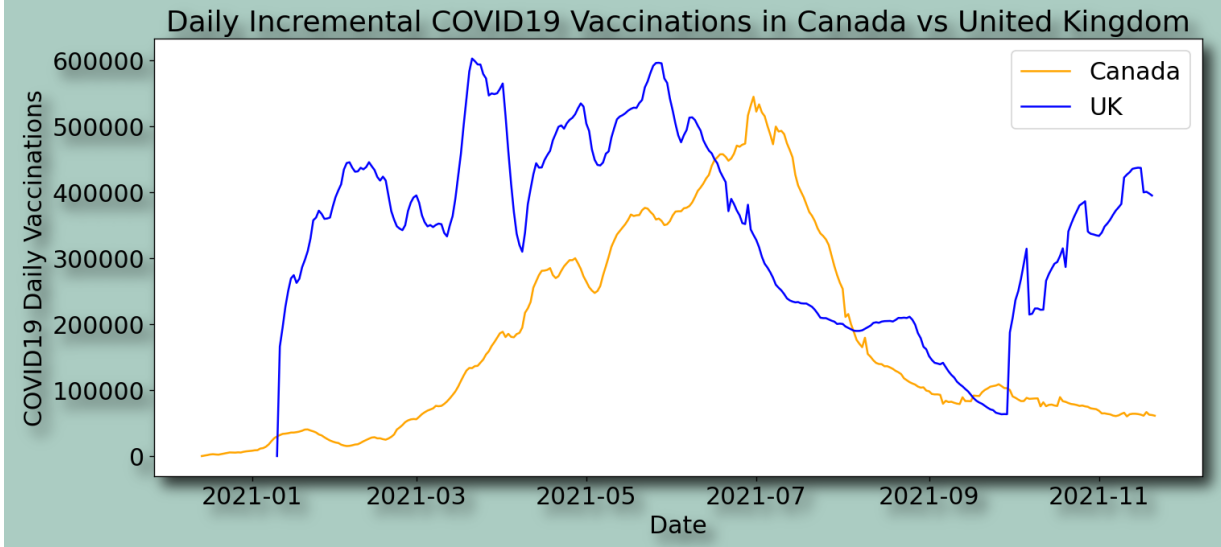




The Initial Exploratory Data Analysis Carried out to view the top ten performing Countries in terms of who has vaccinated the most people per hundred and to also see the countries who are racking up the most daily Vaccination Numbers. We see that the Countries that have the most people vaccinated per hundred do not appear in countries with the highest daily vaccination numbers

# Exploratory Data Analysis



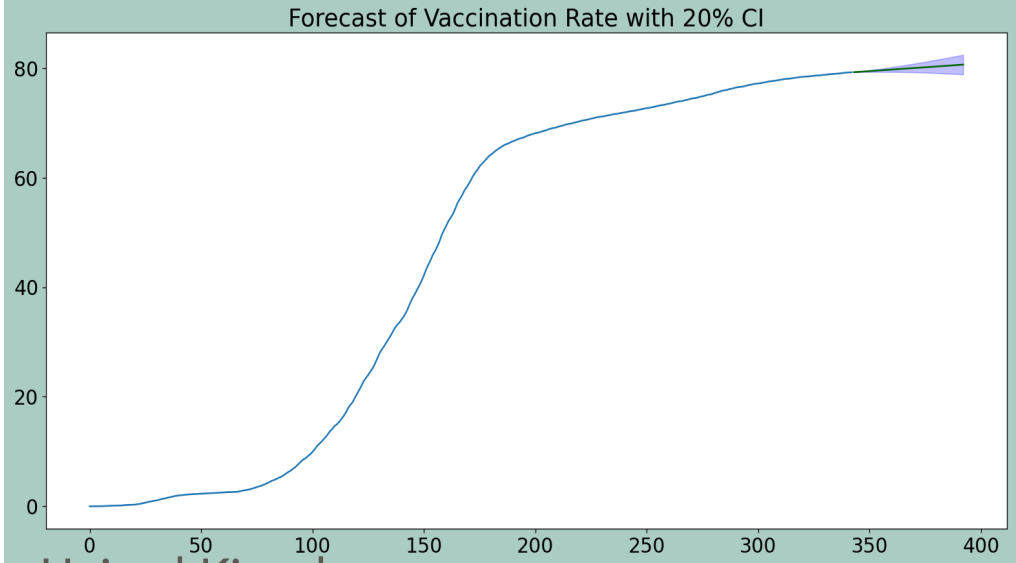



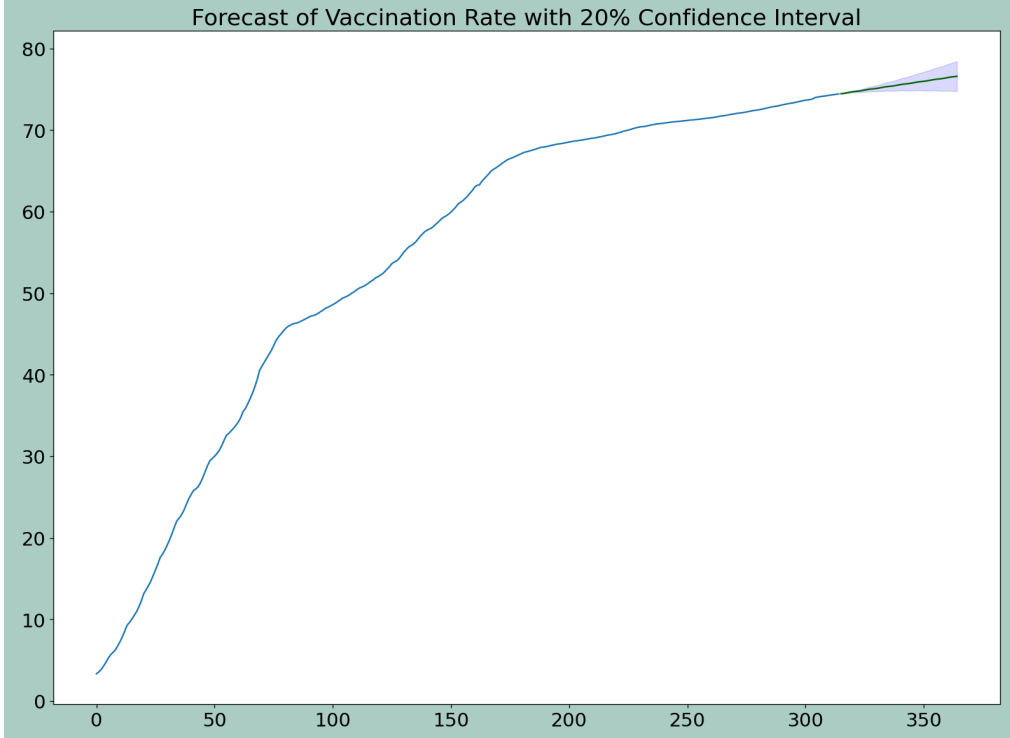

These Images Analyze the Vaccine Rates between the United Kingdom and Canada, the two countries chosen for further analysis, Here we see the trends in daily vaccinations ad the people vaccinated per hundred. We can see that the united Kingdom had a quicker covid 19 vaccine campaign but Canada has overtaken it in recent times in terms of people vaccinated by hundred

# ARIMA MODEL ANALYSIS

Canada




United Kingdom



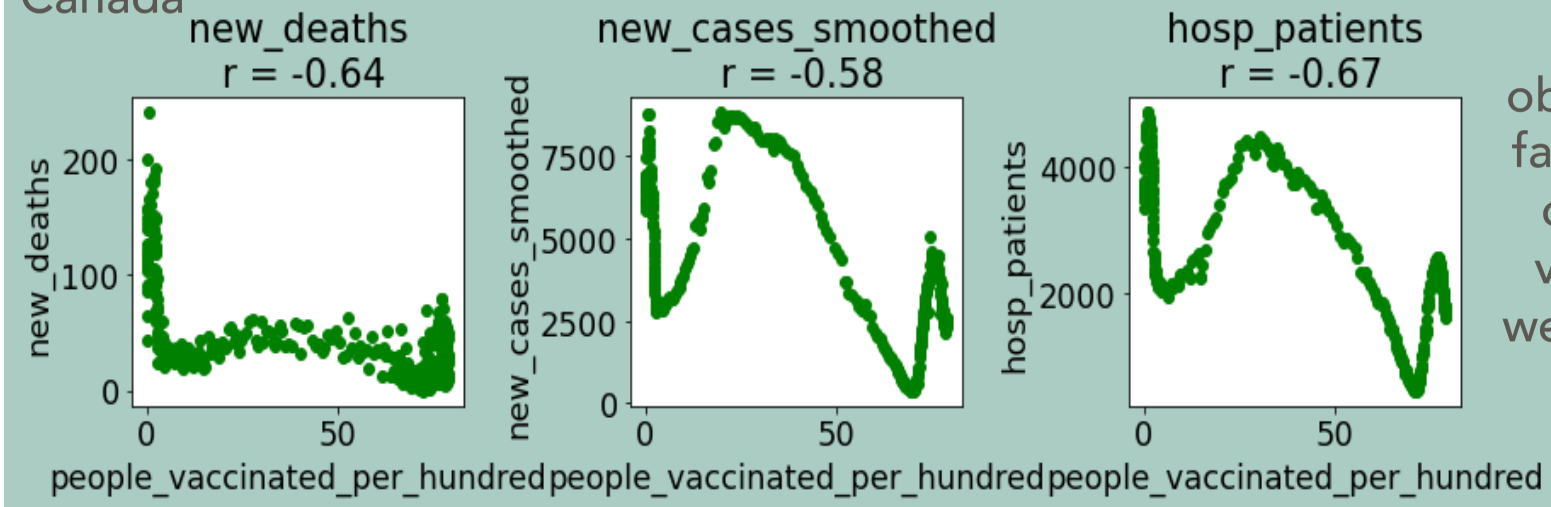

Using the Auto Arima Module in Python We predicted the vaccination rate for Canada and the United Kingdom for the Next 50 Days. The images here show our the current trend and our forecasted prediction. In the images we have included a best case, worse case and base model with a 20% confidence interval

For the Canada Projections our model predicts that in the best case with an alpha of 0.8 the highest vaccination rate possible in the next 50 days is 82.39 people per hundred and worst-case is 78.81 and the base prediction the mean between the two. While for the United Kingdom our model predicts that in the best case w the highest vaccination rate possible in the next 50 days is 78.40 people per hundred and worst-case is 74.75 and the base prediction the mean between the two.

# RELATING VACCINATION RATES TO SECOND DATASET

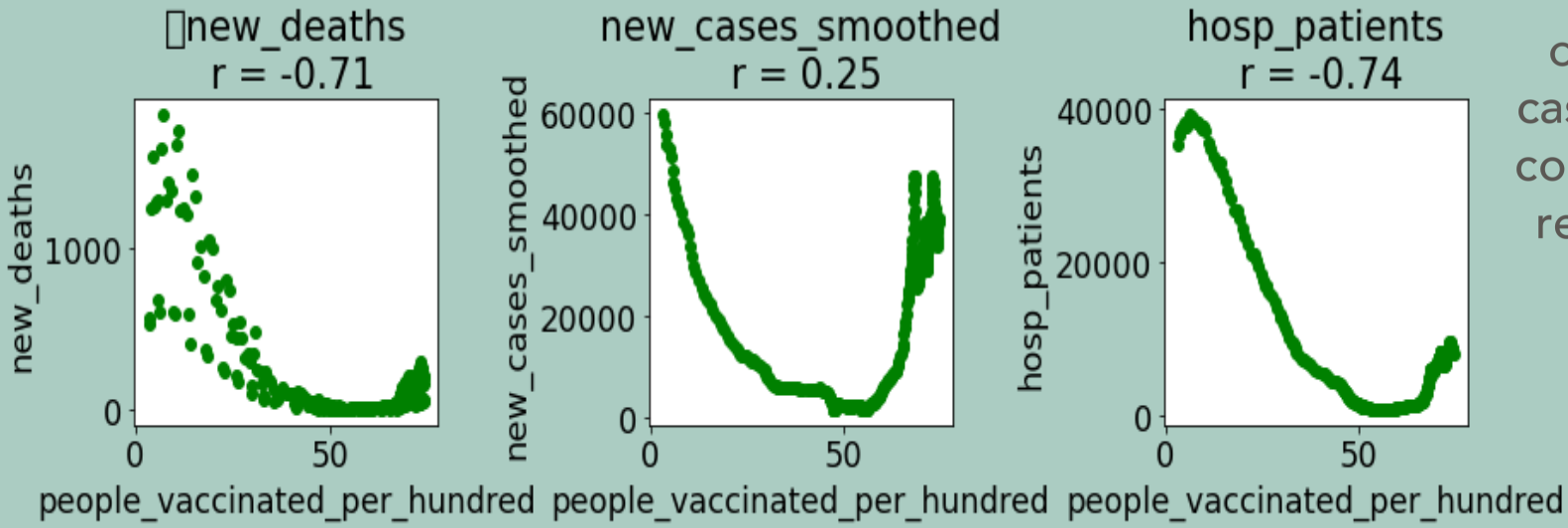
Here we Related the Number of People Vaccinated to three other factors and the relationship is shown below

Canada




For Canada We observed that all three factors had a negative correlation with the vaccination rate and we see the trend in the figure

United Kingdom




Whilst For United Kingdom We observed the new cases have a positive correlation whilst the rest had a negative correlation