Covid-19 Report and Analysis

Introduction

This report analyzes data collected on 204 countries, from approximately February 24, 2020 through approximately October 03, 2021.

Body

Data

This report was written using MyJupyterNotebooks. First I imported Pandas, MatPlotLib, Seaborn, and Numpy to analyze and write reports on the data. Next I imported the Covid-19 data from a .csv file. Then I requested the number of rows and columns (121034 and 65 respectively), the data types of each column, and the column names. Next I addressed the null values by dropping rows that had no date, and by replacing null values in the other columns with 0 so that I could perform math operations with those columns.

Method

First, I calculated the percentages of my needed columns when compared to the population. I calculated the percentage of total cases, people vaccinated, tests administered, positive rate, and hospital patients by dividing each column by the population and then multiplying by 100. After calculating my needed values, I split the dataset into two smaller dataframes to make them more manageable:
2020: All data from the year 2020
2021: All data from the year 2021
With these two datasets, I calculated the average percentage of Covid-19 cases for the year 2020 and then the year 2021, separated by country. I subtracted the year 2020 value from the year 2021 value to get percentage of change between the two years.

Results

The client wanted to know which countries had the best response to Covid-19 over the course of two years, and which countries had the worst response. To answer this question, I used the percentage of change between the years 2021 and 2020 that I had calculated. Countries with slow growth of Covid-19 cases over the timeframe would have a smaller quantity, because they had better success flattening the curve of Covid-19 cases. I found that more successful countries typically started out with a relatively low percentage of the population suffering from Covid-19 and managed to maintain that percentage; less successful countries started with a low percentage, and then total cases climbed rapidly in 2021. There were no countries that managed to decline their percentage of total cases in that time.

I found that several factors showed a slight correlation with each country's average (over the course of the two years); percentage of people vaccinated, percentage of tests administered, rate of positive cases, and percentage of hospital patients showed a direct correlation with total overall cases. These factors may serve as valuable when predicting the percentage of Covid-19 cases. Population density showed no correlation.
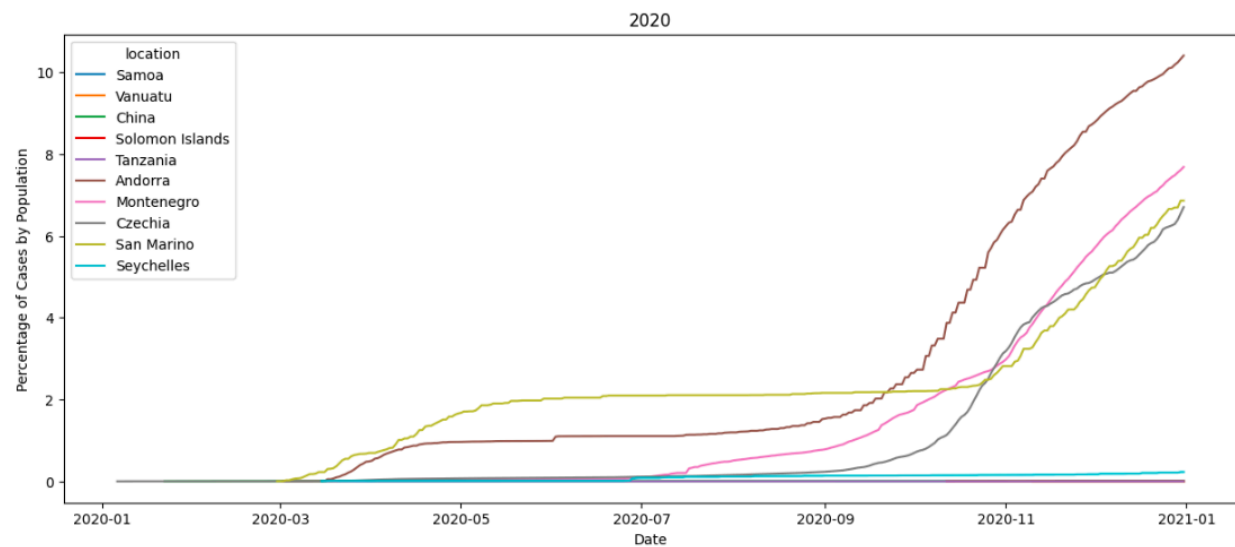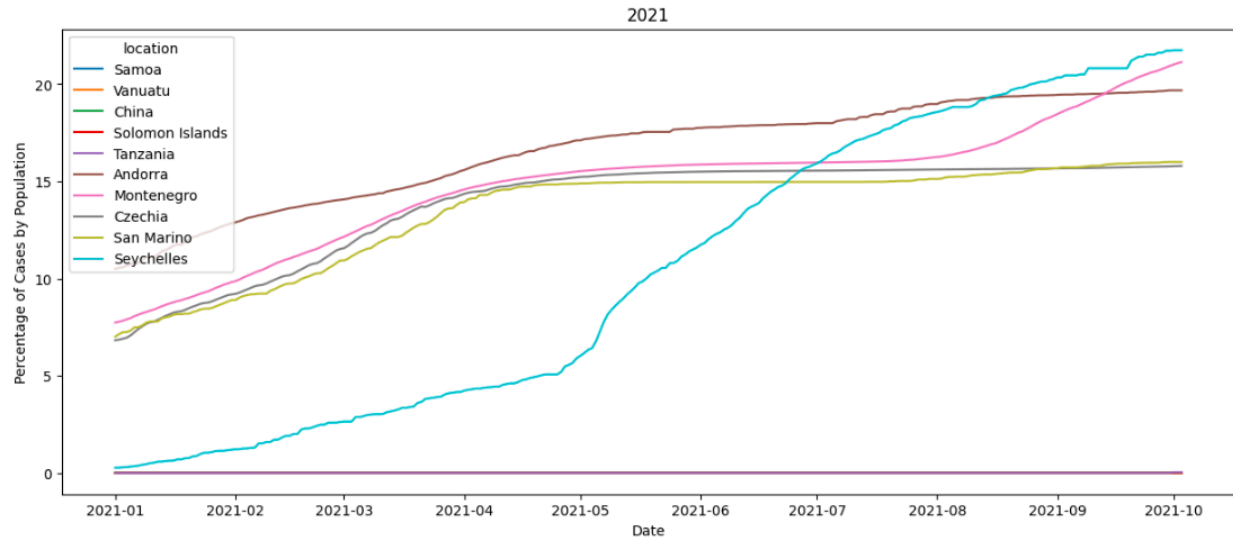
Analysis

Heatmap and plots showing the correlation of different factors compared to overall cases (grouped by country):
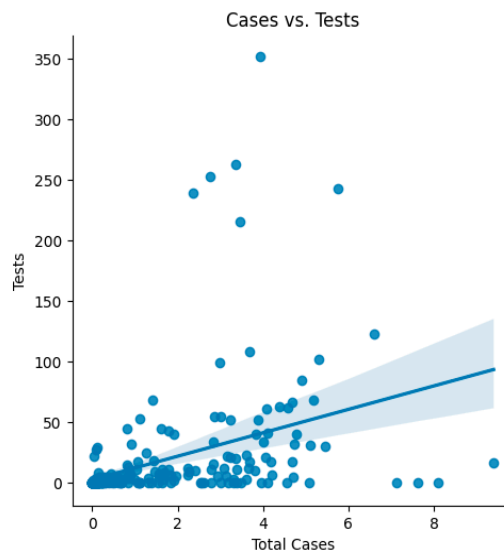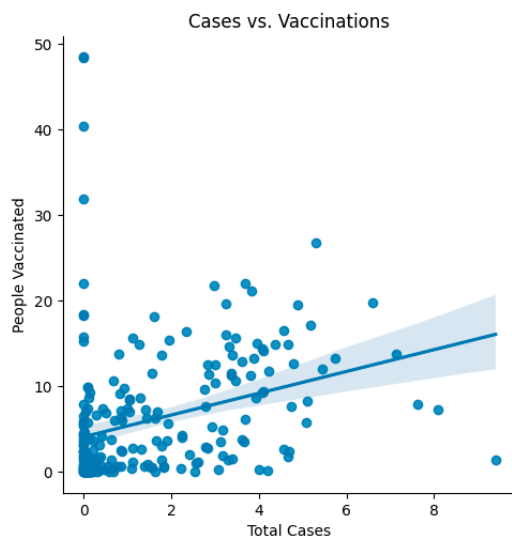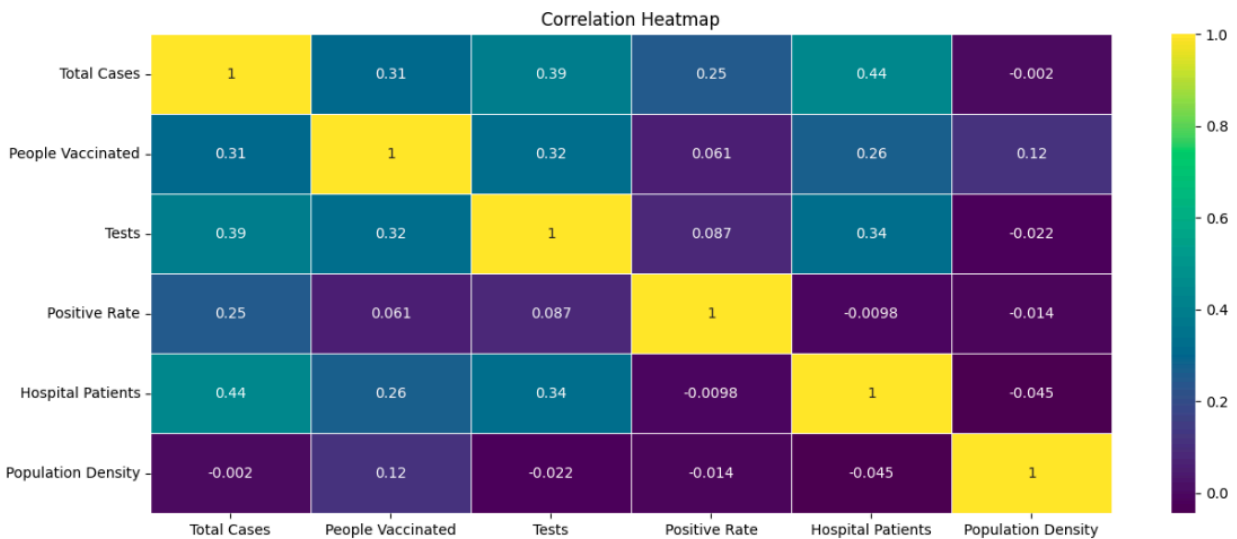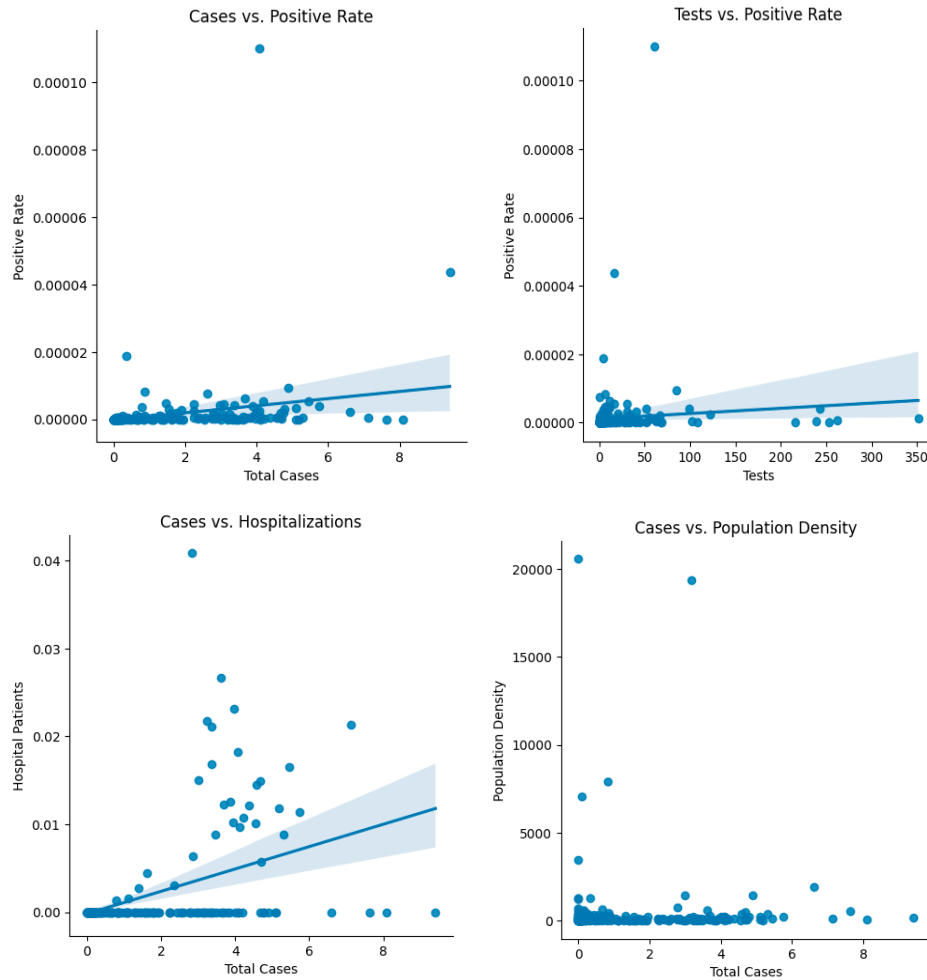
Line graphs showing the percentage of the population suffering from Covid-19 for the top 5 and bottom 5 countries. Note: the 10 countries included were calculated by calculating the degree of change between the years 2020 and 2021.

Conclusion

Over the course of approximately two years, the percentage of total cases by population has risen and continues to rise. The percentage of cases rises abruptly in the autumn of 2020, and then plateaus slightly in the spring of 2021.

The five countries with the best response to Covid-19 were Samoa, Vanuatu, China, Solomon Islands, and Tanzania. Over the course of the two years, these countries had a low case count of Covid-19 from the beginning of the dataset.

The five countries with the worst response to Covid-19 were Andorra, Montenegro, Czechia, San Marino, and Seychelles. These countries unfortunately suffered from an explosive increase in cases of Covid-19 over the course of the two years.