Covid 19 Analysis: Causes?

Introduction:

This report will analyze the Covid 19 dataset and attempt to answer the following questions;

- Do vaccination numbers affect total number of cases?
- Does GDP or poverty levels affect total number of cases?
- Does age play a roll?
- What if you are a Smoker?
- Does Human Development Index affect total cases?

Because of the size of the dataset I focused my analysis on just the continents of Asia and South America. I felt that focusing on two continents that were as far apart to see if there was any substantial difference between the results. In this report I will take each question and relay the results before moving on to the next.

Body:

Data:

In the original dataset there were 121,034 rows of data and 65 columns. The date was collected from Feb 2020 until October 2021. The size of the data gave me an idea to find a way to look at how two different geographic locations were affected (or not) by Covid. I will give a detailed looked at which data I chose to use to answer each question.

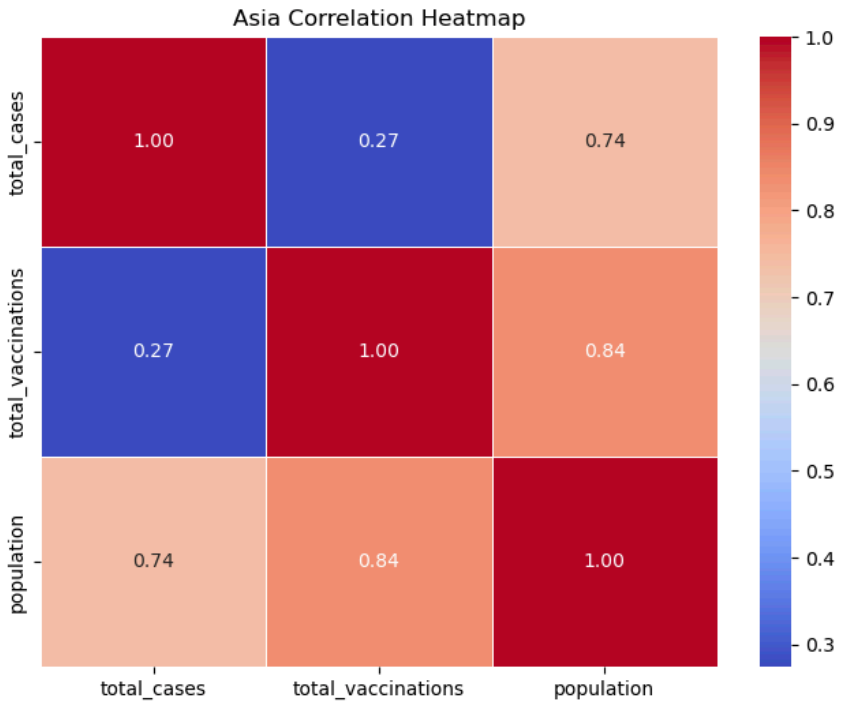
I did decide that for my final question about the Human Development Index I would use the whole dataset as I didn't think this measure would be greatly affected by geographic location.

Process:

1.) "Do vaccination numbers affect total number of cases?"

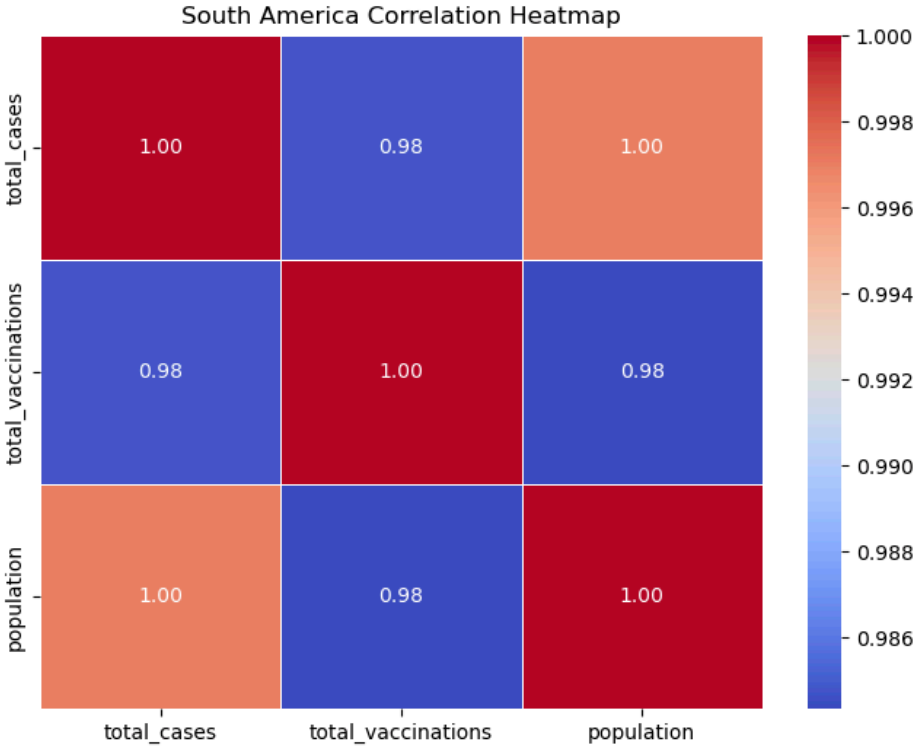
I began with looking at just the data from Asia. I used only the data contained within the Continent, Location, Total Cases, Total Vaccinations and Population columns as I felt they were the most relevant. After cleaning, I aggregated the data by Location which gave me the information for each country. Below is the results of my analysis, followed by a heat map for that correlation analysis.

	total_cases	total_vaccinations	population
total_cases	1.000000	0.274179	0.738645
total_vaccinations	0.274179	1.000000	0.837451
population	0.738645	0.837451	1.000000



Now, lets look at South America.

	total_cases	total_vaccinations	population
total_cases	1.000000	0.984755	0.996982
total_vaccinations	0.984755	1.000000	0.984352
population	0.996982	0.984352	1.000000



Results:

According to my analysis, there is not a correlation between the total number of vaccinations and total number of cases in Asia. The result of 0.27 is low. If there were to be a strong correlation that number would be close to 1.0.

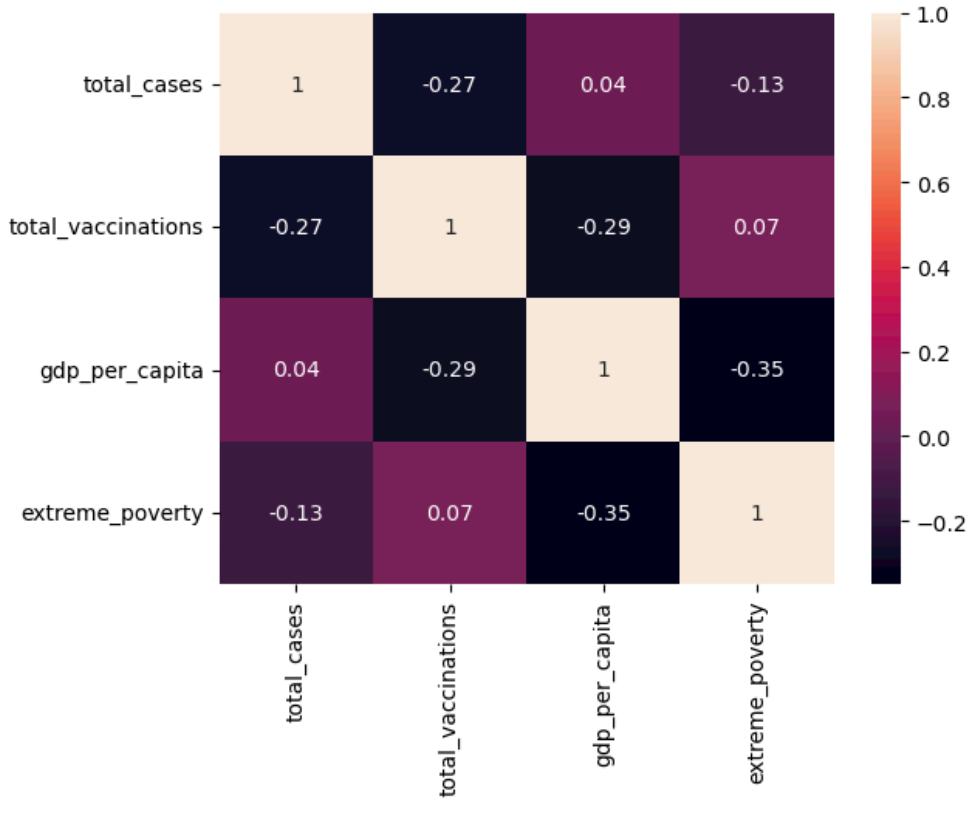
In looking at the South America data, however, there appears to be a strong correlation between the total cases and total vaccinations. The resulting matrix has many values that are close enough to 1.0 to represent a strong correlation.

2.) Does GDP or poverty levels affect total number of cases?

For this question I focused the analysis on the Total Cases, Total Vaccinations, GDP Per Capita and Extreme Poverty columns. I decided to simplify a bit here and look at the Highest GDP per capita countries as well as the lowest GDP per capita countries within Asia.

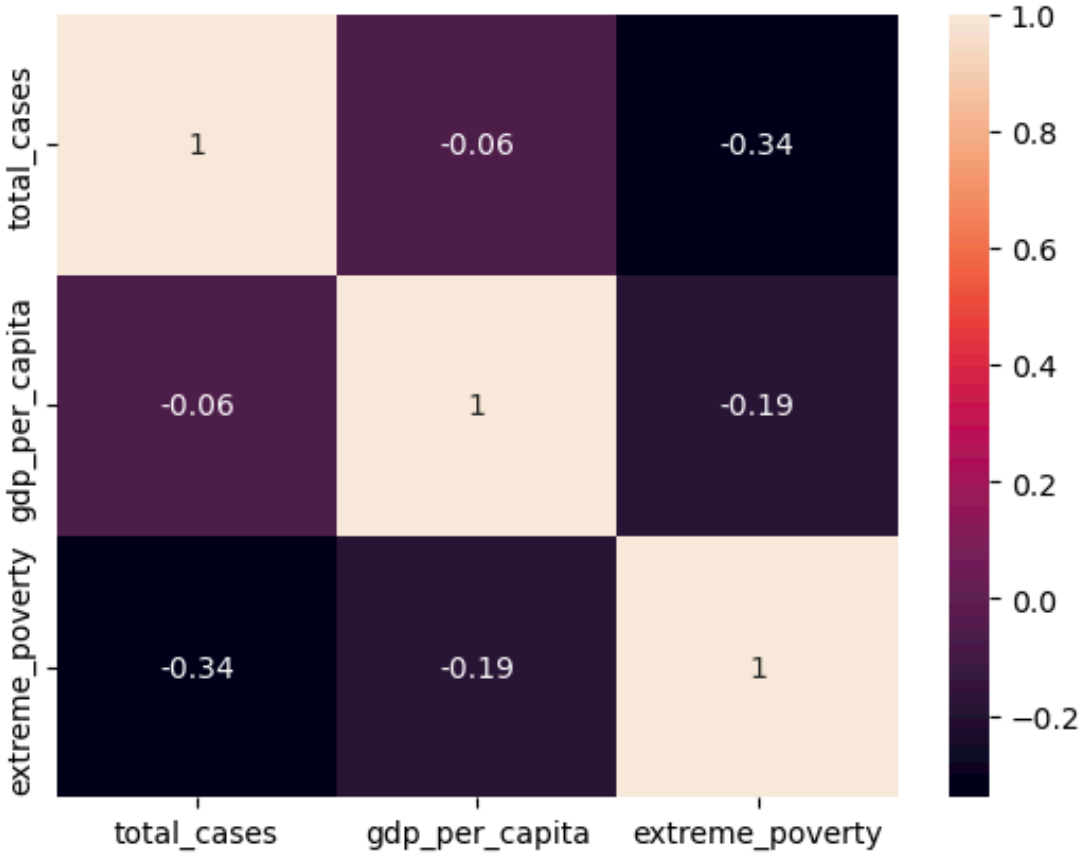
First, the highest data.

	total_cases	gdp_per_capita	extreme_poverty
total_cases	1.000000	0.028691	-0.132939
gdp_per_capita	0.028691	1.000000	-0.354536
extreme_poverty	-0.132939	-0.354536	1.000000



Second, the lowest.

	total_cases	gdp_per_capita	extreme_poverty
total_cases	1.000000	-0.058821	-0.339603
gdp_per_capita	-0.058821	1.000000	-0.193039
extreme_poverty	-0.339603	-0.193039	1.000000



Now I looked at the the Extreme Poverty column and focused the analysis on only the South American data. After a bit of workin with the data, I decided to use only the top 5 and bottom 5 in the column do to a lack of data in many of the countries.

Top 5:

	total_cases	gdp_per_capita	extreme_poverty
total_cases	1.000000	0.652642	-0.418367
gdp_per_capita	0.652642	1.000000	-0.831895
extreme_poverty	-0.418367	-0.831895	1.000000

Bottom 5:

	total_cases	gdp_per_capita	extreme_poverty
total_cases	1.000000	-0.229417	0.838898
gdp_per_capita	-0.229417	1.000000	-0.529234
extreme_poverty	0.838898	-0.529234	1.000000

Results:

For the Extreme Poverty column in the top 5, there was negative correlation between total cases and extreme poverty which indicates there was not a strong relationship. However when looking at the bottom 5 in the Extreme Poverty column, there was a 0.84 indicating a fairly strong relationship.

3. Does age play a roll?

Data:

For this question, I chose the Continent, Location, Total Cases, Aged 65 or Older and Median Age columns as I felt they were the most relevant. For this question I split it into two part; firstly to look at what relationship there was in South America using the Highest Median Age. I found the ten largest values in the Median Age and using the correlation. I then used the Lowest Median Age from Asia.

South American 10 Highest for Median Age

	total_cases	aged_65_older	median_age
total_cases	1.000000	0.015561	0.330814
aged_65_older	0.015561	1.000000	0.829356
median_age	0.330814	0.829356	1.000000

Asia 10 Lowest for Median Age.

	total_cases	aged_65_older	median_age
total_cases	1.000000	0.469674	0.725111
aged_65_older	0.469674	1.000000	0.713593
median_age	0.725111	0.713593	1.000000

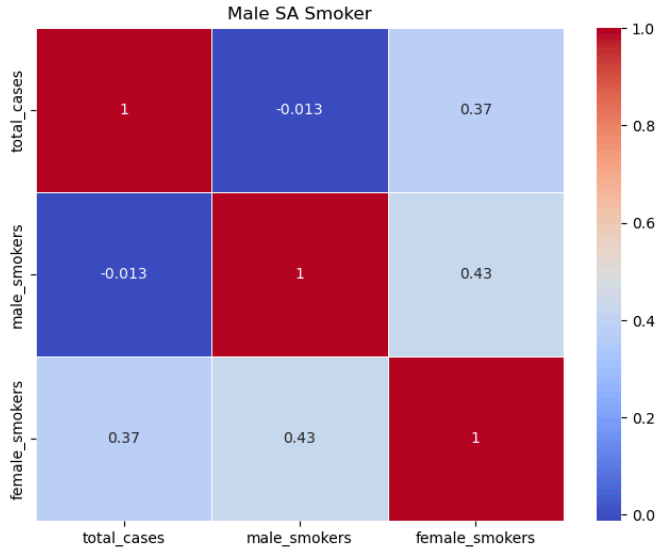
Results:

According to my analysis, the countries with the ten highest median age had no relationship between the total number of case and median age. The result of 0.33 signifies this. When looking at the South America data(lowest median age) there is a stronger relationship between to the two but it is not strong enough to clearly state that there is a relationship.

4.) What if you are a Smoker?

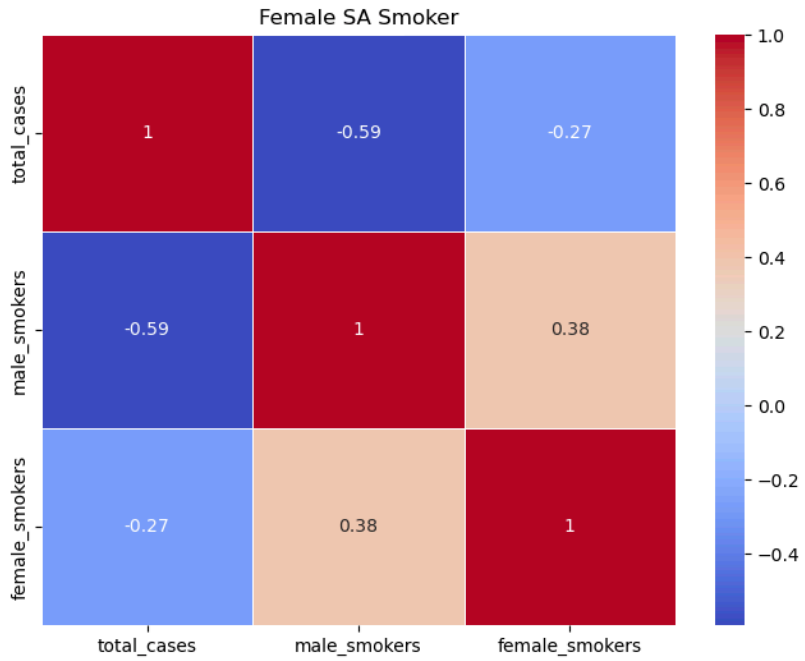
The dataset included a column that had data for males and females. I decided to keep the columns separate as I felt it would be an interesting way to look at the differences. Firstly, we will look at the Males:

South America:

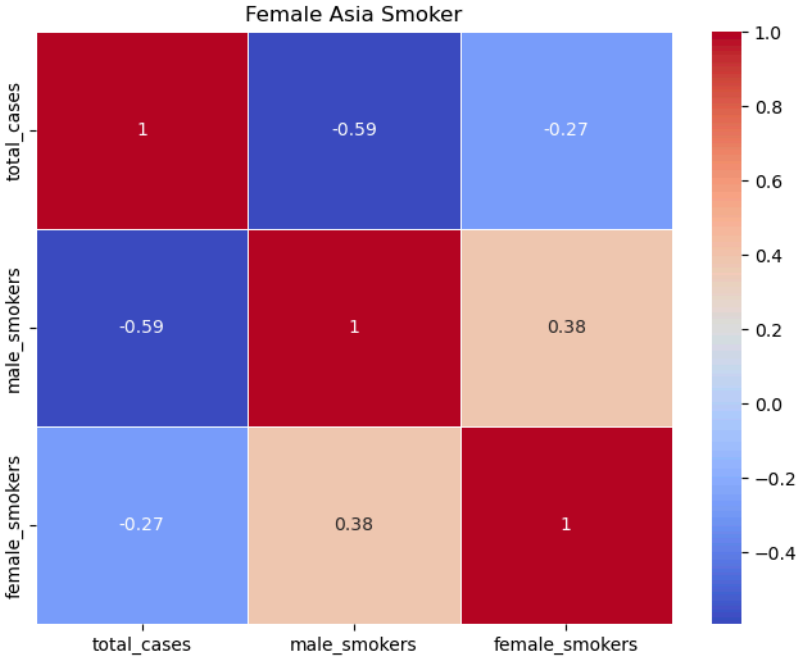


Asia:

Now Female, beginning with South America;



Asia;



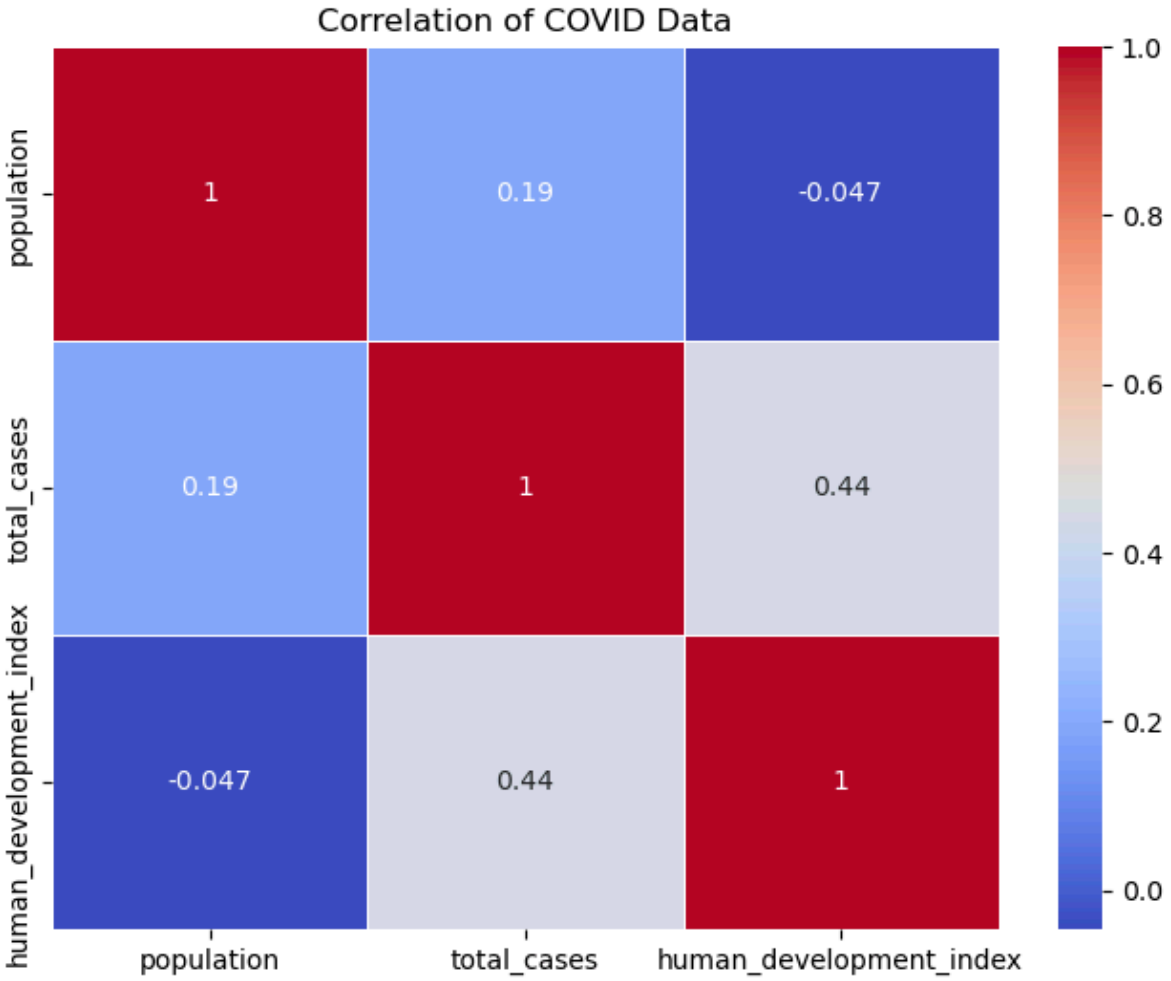
Results:

After reviewing the analysis, I did not find any relationship between a smoker and the total number of cases. This was true both for the South America and Asia data.

5.) Does Human Development Index affect total cases?

Process:

My final analysis was to use the Human Development Index to see if there was any relationship between that and the total cases. I decided it would be worth using the full dataset as I wanted to give a bit broader scope for the final question.



Results:

As with the rest of my analysis, there was not a strong relationship between the total number of cases and the Human Development Index. It was demonstrated by the result being 0.44 which looking at the correlation between the two variables.

Final Results

Through various questions I have determined that in answer to all of the questions I did not find any relationship that could be confidently pointed to a factor in the total number of cases.