**COVID 19 VACCINE ANALYSIS**

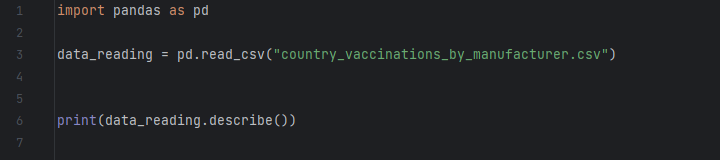
**DEVELOPMENT PART 2**

**INTRODUCTION:**

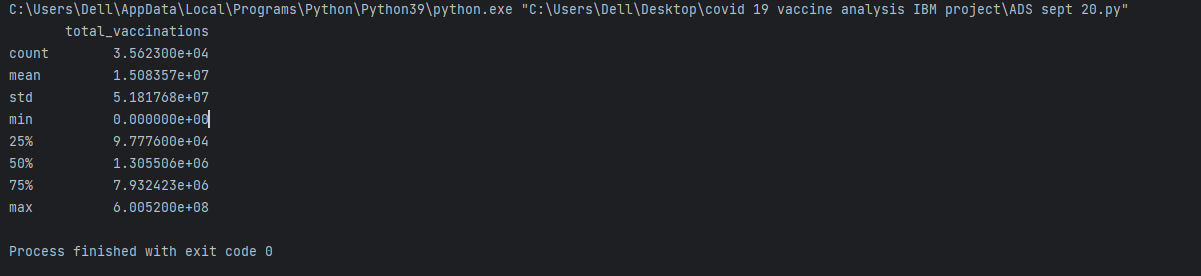
The COVID 19 pandemic caused due to the virus named corona. This virus was originated in Wuhan, China in 2019 and was later spread throughout the world. The disease showed symptoms such as mild fever and cold but also caused life threatening symptoms like breathing problems caused by damage to the lungs. As this virus was new to the world and there was no vaccine or cure to it at the earlier period there were several deaths around the world. So the countries around the world announced lockdowns to prevent the spread of the disease. An effort was made by several health organizations to discover the cure or a vaccine to fight against the virus.

In later stages of 2020 several experimental vaccines were developed and was tested to humans. The efforts were successful as the vaccines were helpful in reducing the affects the virus and even if people were infected, they were not in any life threatening situation and escaped the illness having only minor symptoms. Many countries later developed their own vaccines and also helped other countries without the resources by providing them with vaccines developed.

**EXPLORATORY DATA ANALYSIS:**

Describe():

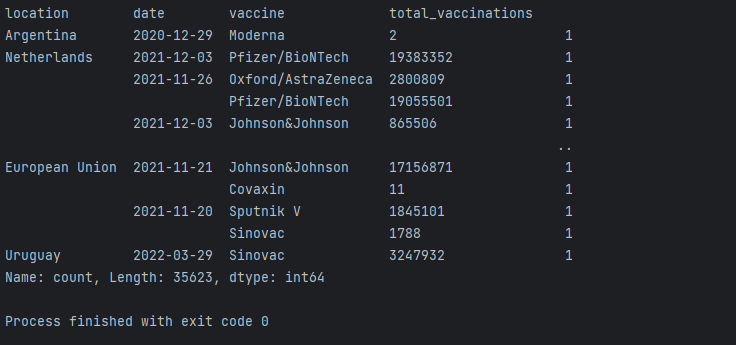
Output:



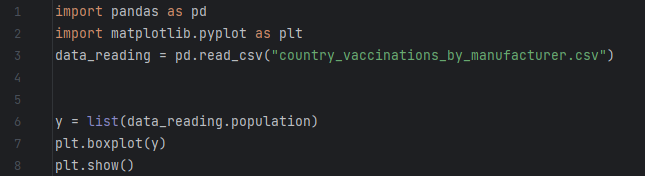
Value\_Counts():

count.PNG

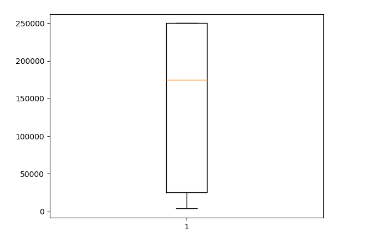
Output:



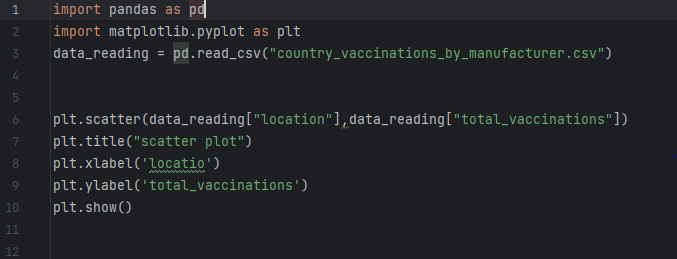
Boxplot:



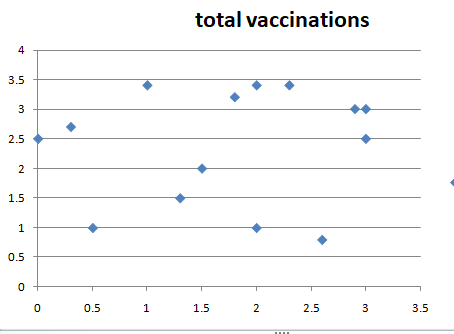
Output:

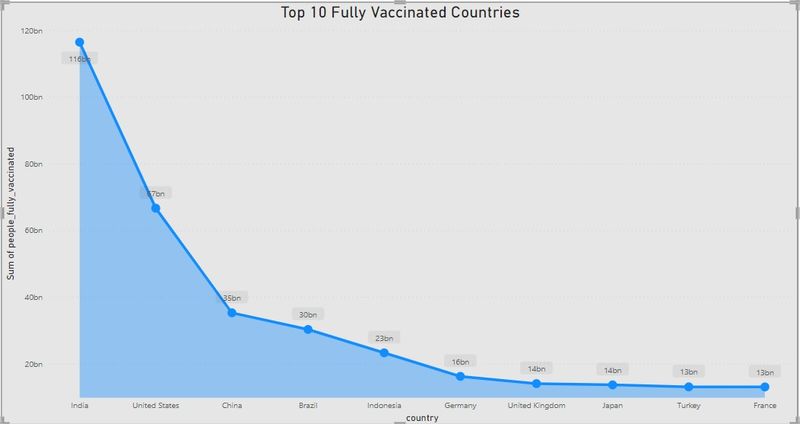


**VISUALIZATION:**



Output:





**EXPLANATION:**

EDA is a phenomenon under data analysis used for gaining a better understanding of data aspects.

Descriptive statistics is a helpful way to understand characteristics of your data and to get a quick summary of it. Pandas in python provide an interesting method **describe()**. The describe function applies basic statistical computations on the dataset like extreme values, count of data points standard deviation etc. Any missing value or NaN value is automatically skipped. describe() function gives a good picture of distribution of data. Another useful method if value\_counts() which can get count of each category in a categorical attributed series of values. For an instance suppose you are dealing with a dataset of customers who are divided as youth, medium and old categories under column name age and your dataframe is “DF”. You can run this statement to know how many people fall in respective categories. In our data set example education column can be used One more useful tool is boxplot which you can use through matplotlib module. Boxplot is a pictorial representation of distribution of data which shows extreme values, median and quartiles. We can easily figure out outliers by using boxplots. Now consider the dataset we’ve been dealing with again and lets draw a boxplot on attribute *population* Scatter plots are used to observe relationships between variables and uses dots to represent the relationship between them. The [**scatter()**](https://www.geeksforgeeks.org/matplotlib-pyplot-scatter-in-python/) method in the matplotlib library is used to draw a scatter plot. Bar Chart can be of two types horizontal bars and vertical bars. Each can be created using the hbar() and vbar() functions of the plotting interface respectively This is how we can effectively use available libraries to make various visualizations using python which will be very helpful in analyzing a dataset effectively.. 