COVID VACCINE ANALYSIS

NAME: BHARATHI.S

NAN MUDHALVAN ID:au920821104011

PHASE 4 PROJECT

Exploratory data analysis, Statistical analysis, Visualization

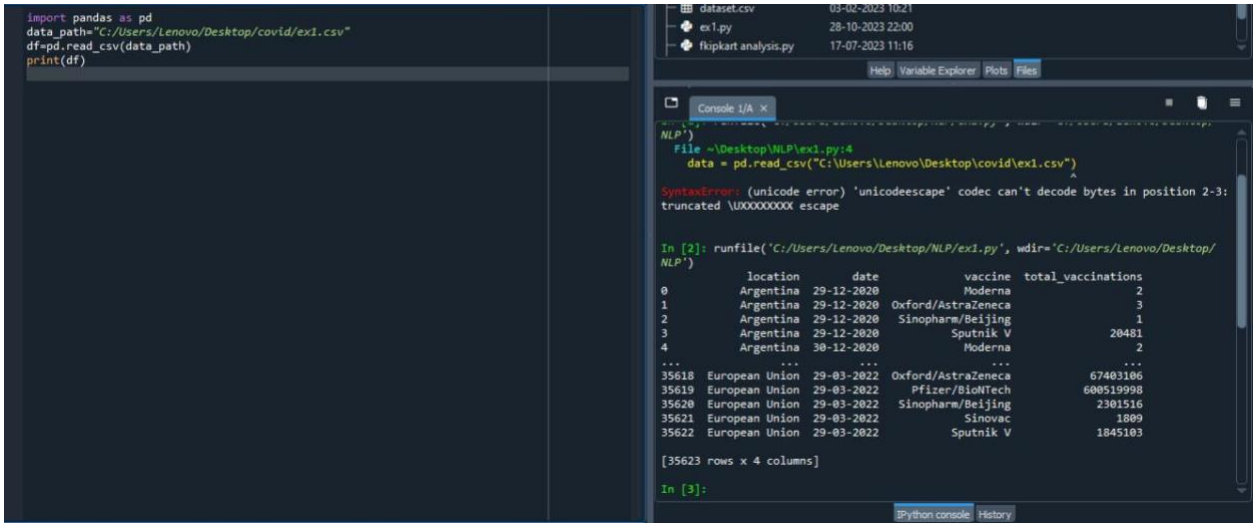
Abstraction:

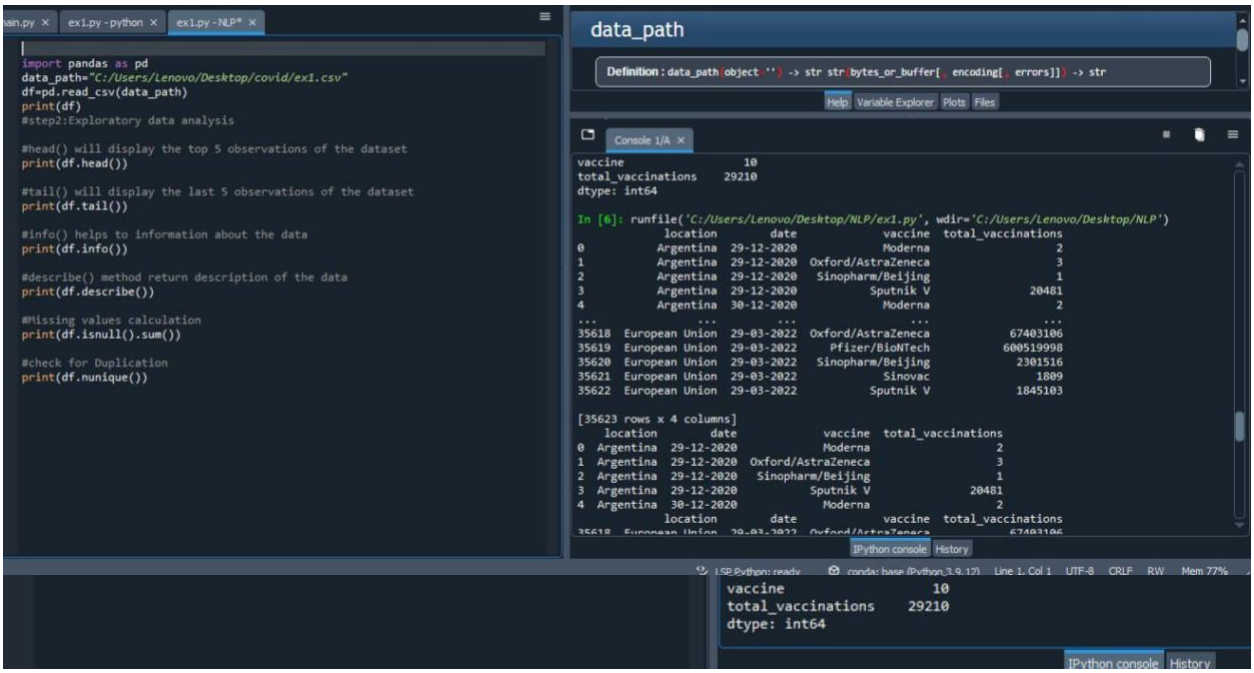
- The goal of this research is to analyze data on vaccinations, vaccination administration, and forecasting vaccination rates on a country-by-country basis for the general public, policymakers, vaccine manufacturers, national governments, and international governments to better understand the current state of COVID-19 vaccination.
- Summarize your findings in a clear and concise manner.
- Highlight key insights and trends discovered during EDA and statistical analysis.
- Provide actionable recommendations based on your analysis, such as targeting vaccination campaigns in regions with low vaccination rates.

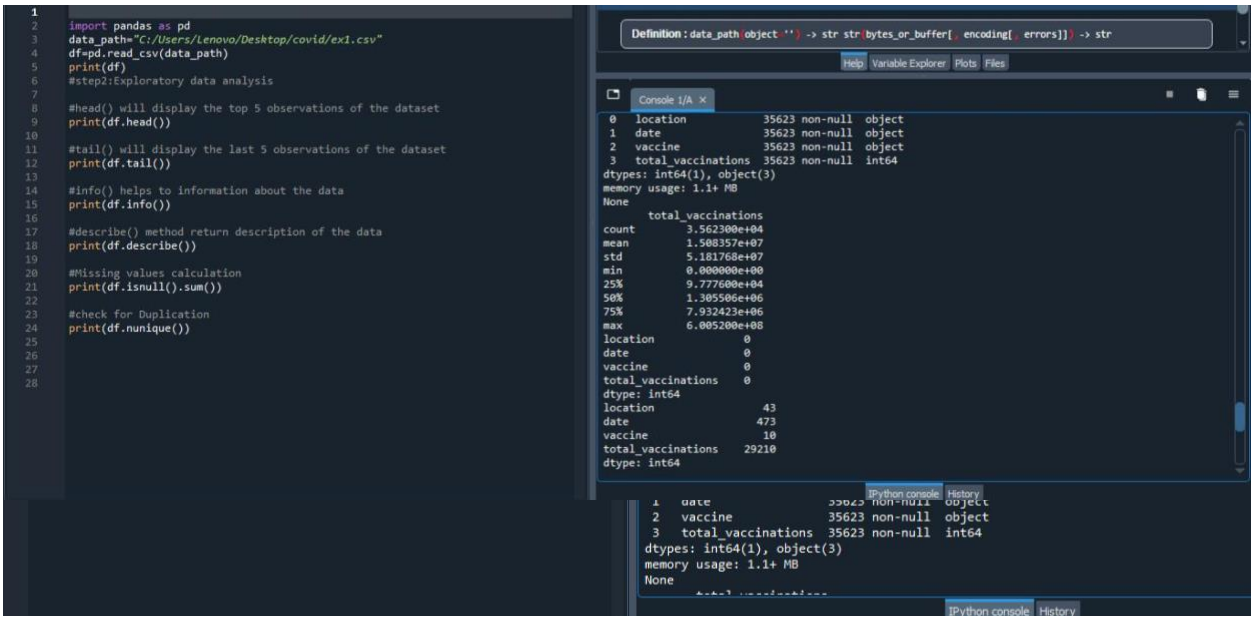
❖ **Exploratory data analysis :**

- Conduct descriptive statistics like mean, median, and standard deviation.
- Explore the distribution of vaccination rates across different regions or demographics.

- Identify trends and patterns in the data through plots, such as histograms, box plots, and scatter plots.
- Perform correlation analysis to understand relationships between variables (e.g., vaccination rates and infection rates).

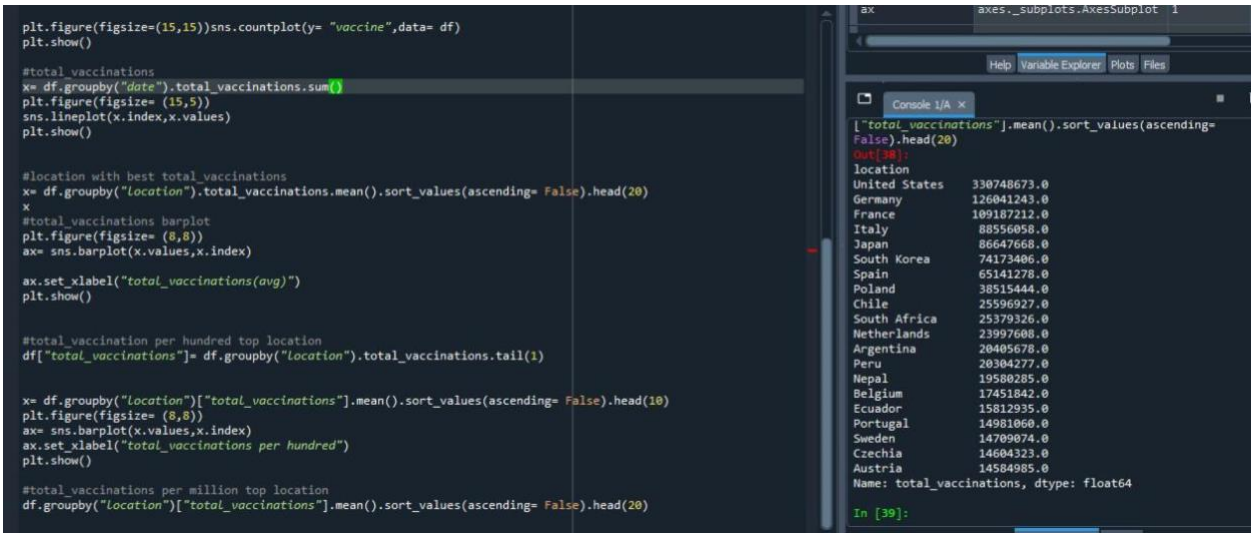






❖ Statistical Analysis:

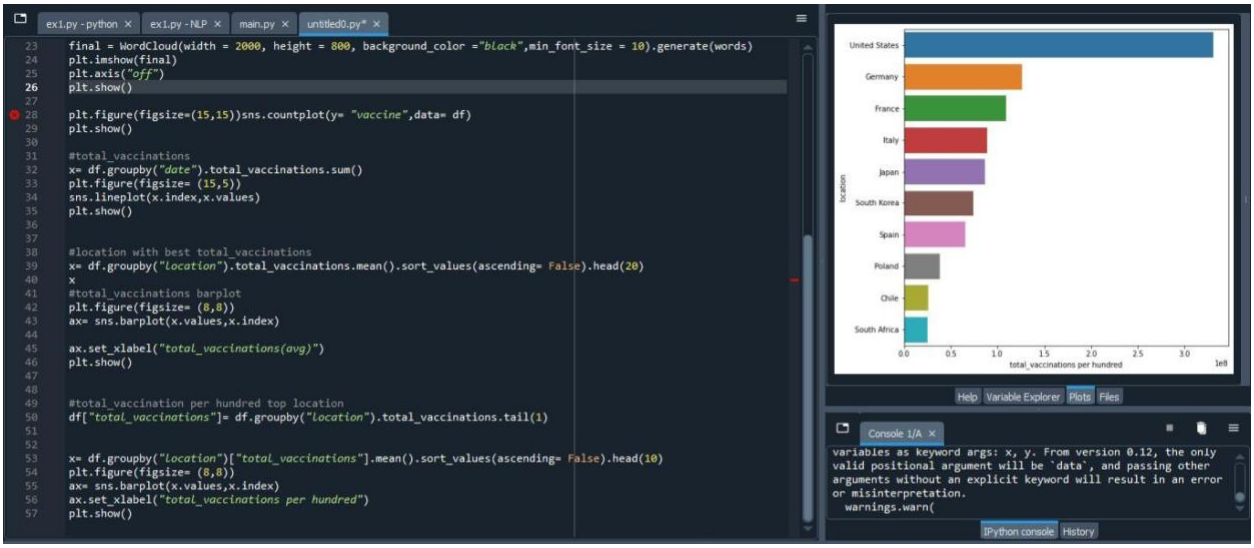
- Conduct hypothesis tests to assess the significance of differences in vaccination rates between groups (e.g., age groups or geographical regions).
- Apply regression analysis to understand the impact of various factors on vaccination rates.
- Calculate confidence intervals to estimate the precision of your findings.



```
plt.figure(figsize=(15,15))sns.countplot(y= "vaccine",data= df)
plt.show()

#total_vaccinations
x= df.groupby("date").total_vaccinations.sum()
plt.figure(figsize= (15,5))
sns.lineplot(x.index,x.values)
plt.show()


#location with best total_vaccinations
x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
x
#total_vaccinations barplot
plt.figure(figsize= (8,8))
ax= sns.barplot(x.values,x.index)

ax.set_xlabel("total_vaccinations(avg)")
plt.show()


#total_vaccination per hundred top location
df["total_vaccinations"]= df.groupby("location").total_vaccinations.tail(1)


x= df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(10)
plt.figure(figsize= (8,8))
ax= sns.barplot(x.values,x.index)
ax.set_xlabel("total_vaccinations per hundred")
plt.show()

#total_vaccinations per million top location
df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(20)
```

```
ax                axes._subplots.AxesSubplot  1

Help  Variable Explorer  Plots  Files

Console 1/A
["total_vaccinations"].mean().sort_values(ascending=
False).head(20)
Out[38]:
location
United States     330748673.0
Germany           126041243.0
France            109187212.0
Italy              88556058.0
Japan              86647668.0
South Korea        74173406.0
Spain              65141278.0
Poland             38515444.0
Chile              25596927.0
South Africa       25379326.0
Netherlands        23997608.0
Argentina          20405678.0
Peru               20304277.0
Nepal              19580285.0
Belgium            17451842.0
Ecuador            15812935.0
Portugal           14981060.0
Sweden             14709074.0
Czechia            14604323.0
Austria            14584985.0
Name: total_vaccinations, dtype: float64

In [39]:
```

- ❖ **Visualization** :
  - Create visually engaging plots to communicate your findings effectively:
  - Bar charts or stacked bar charts to compare vaccination rates by region or age group.
  - Time series plots to track vaccination progress over time.
  - Heatmaps to visualize correlations between variables.
  - Utilize interactive dashboards for dynamic exploration of data.

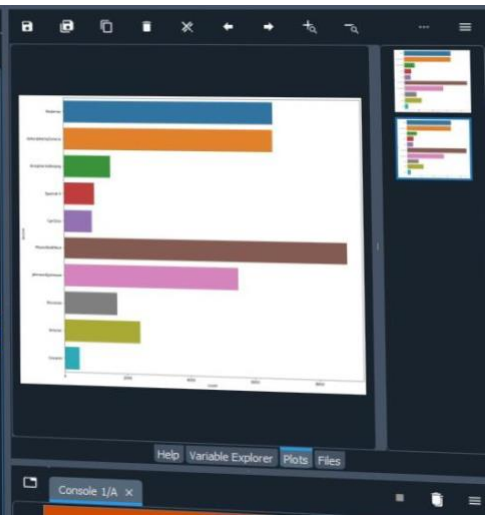


```
ex1.py - python   ex1.py - NLP   main.py   untitled0.py*

    final = WordCloud(width = 2000, height = 800, background_color ="black",min_font_size = 10).generate(words)
    plt.imshow(final)
    plt.axis("off")
    plt.show()

    plt.figure(figsize=(15,15))sns.countplot(y= "vaccine",data= df)
    plt.show()

    #total_vaccinations
    x= df.groupby("date").total_vaccinations.sum()
    plt.figure(figsize= (15,5))
    sns.lineplot(x.index,x.values)
    plt.show()


    #location with best total_vaccinations
    x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
    x
    #total_vaccinations barplot
    plt.figure(figsize= (8,8))
    ax= sns.barplot(x.values,x.index)

    ax.set_xlabel("total_vaccinations(avg)")
    plt.show()


    #total_vaccination per hundred top location
    df["total_vaccinations"]= df.groupby("location").total_vaccinations.tail(1)


    x= df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(10)
    plt.figure(figsize= (8,8))
    ax= sns.barplot(x.values,x.index)
    ax.set_xlabel("total_vaccinations per hundred")
    plt.show()
```

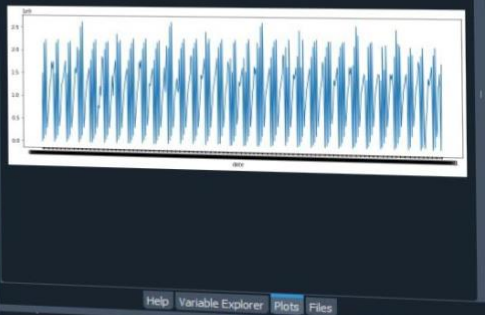


```
Help  Variable Explorer  Plots  Files

Console 1/A
variables as keyword args: x, y. From version 0.12, the only
valid positional argument will be `data`, and passing other
arguments without an explicit keyword will result in an error
or misinterpretation.
  warnings.warn(

IPython console  History
```

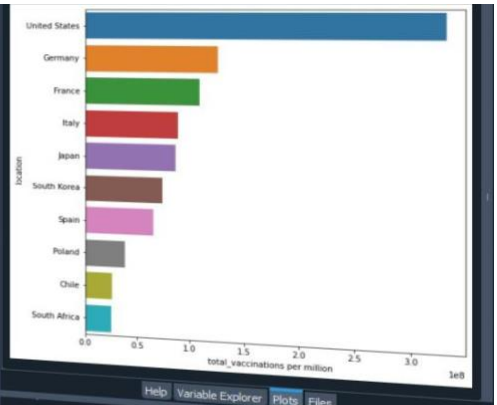
```
# -*- coding: utf-8 -*-
"""Created on Tue Oct 31 18:49:26 2023

@author: Lenovo
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


data_path="C:/Users/Lenovo/Desktop/covid/ex1.csv"
df=pd.read_csv(data_path)
print(df)

import numpy as np
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
plt.figure(figsize= (20,20))
words= "".join(df["vaccines"])
final = WordCloud(width = 2000, height = 800, background_color ="black",min_font_size = 10).generate(words)
plt.imshow(final)
plt.axis("off")
plt.show()

plt.figure(figsize=(15,15))
sns.countplot(y= "vaccine",data= df)
plt.show()
```



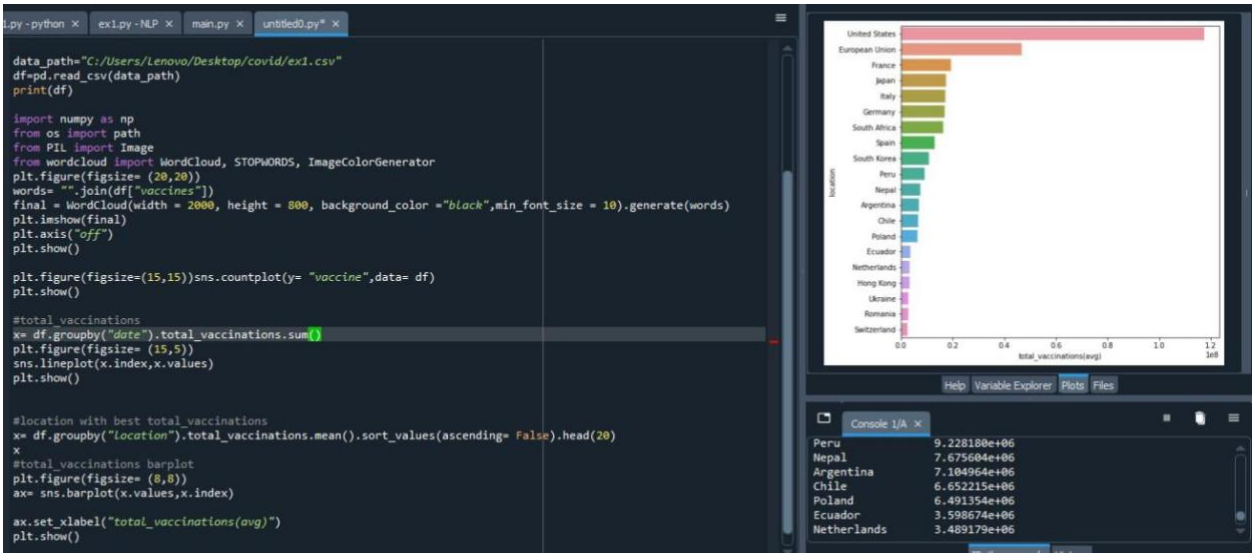

```
@author: Lenovo
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


data_path="C:/Users/Lenovo/Desktop/covid/ex1.csv"
df=pd.read_csv(data_path)
print(df)

import numpy as np
from os import path
from PIL import Image
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
plt.figure(figsize= (20,20))
words= "".join(df["vaccines"])
final = WordCloud(width = 2000, height = 800, background_color ="black",min_font_size = 10).generate(words)
plt.imshow(final)
plt.axis("off")
plt.show()

plt.figure(figsize=(15,15))
sns.countplot(y= "vaccine",data= df)
plt.show()

#total_vaccinations
x= df.groupby("date").total_vaccinations.sum()
plt.figure(figsize= (15,5))
sns.lineplot(x.index,x.values)
plt.show()
```




```
#total_vaccinations
x= df.groupby("date").total_vaccinations.sum()
plt.figure(figsize= (15,5))
sns.lineplot(x.index,x.values)
plt.show()


#location with best total_vaccinations
x= df.groupby("location").total_vaccinations.mean().sort_values(ascending= False).head(20)
x
#total_vaccinations barplot
plt.figure(figsize= (8,8))
ax= sns.barplot(x.values,x.index)

ax.set_xlabel("total_vaccinations(avg)")
plt.show()


#total_vaccination per hundred top location
df["total_vaccinations"]= df.groupby("location").total_vaccinations.tail(1)

#total_vaccination per million
x= df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(10)
plt.figure(figsize= (8,8))
ax= sns.barplot(x.values,x.index)
ax.set_xlabel("total_vaccinations per million")
plt.show()

#total_vaccinations per million top location
df.groupby("location")["total_vaccinations"].mean().sort_values(ascending= False).head(20)
#total_vaccination per million
plt.figure(figsize= (15,5))
sns.lineplot(x= "date",y= "total_vaccinations per million",data= df)
plt.show()
```




```
#total_vaccinations
x= df.groupby("date").total_vaccinations.sum()
plt.figure(figsize= (15,5))
sns.lineplot(x.index,x.values)
plt.show()
```



Conclusion

- Based on the data analysis, addressing questions or hypotheses posed earlier. Discuss the practical implications of your findings in the context of COVID-19 vaccination efforts.