```
from google.colab import drive
drive.mount('/content/drive')
```

```
    Drive already mounted at /content/drive; to attempt to forcibly remount, call dr
```

```
%cd /content/drive/Shareddrives/CSE544_Project/covid_dataset
!ls
```

```
    /content/drive/Shareddrives/CSE544_Project/covid_dataset
    backup
    colab_pdf.py
    COVID-19_Vaccinations_in_the_United_States_Jurisdiction.csv
    covid_la_cleaned.csv
    covid_la_cleaned_removed_outliers.csv
    covid_md_cleaned.csv
    covid_md_cleaned_removed_outliers.csv
    __pycache__
    United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv
    vacc_la_clean.csv
    vacc_la_clean_removed_outliers.csv
    vacc_md_clean.csv
    vacc_md_clean_removed_outliers.csv
```

```
import numpy as np
import pandas as pd
import math
```

```
df_md=pd.read_csv("vacc_md_clean_removed_outliers.csv")
df_la=pd.read_csv("vacc_la_clean_removed_outliers.csv")
```

```
df_md["Date"] = pd.to_datetime(df_md["Date"], format="%Y-%m-%d")
df_la["Date"] = pd.to_datetime(df_md["Date"], format="%Y-%m-%d")
```

```
df_la.dtypes
```

```
    Date                    datetime64[ns]
    Administered                     int64
    Administered_daily             float64
    dtype: object
```

```
#Selecting month and year
df_md_septnov= df_md [df_md["Date"].dt.month.isin([9,11]) & df_md["Date"].dt.year.isir
df_la_septnov= df_la [df_la["Date"].dt.month.isin([9,11]) & df_la["Date"].dt.year.isir
```

```
df_md_septnov.reset_index(inplace=True)
df_la_septnov.reset_index(inplace=True)
```

We assume the test is valid

▾ T-test:

Null Hpothesis H0: There is no difference between the mean of administered doses for the two states for September 2021 and Novemeber 2021.

Alternative Hypothesis H1: There is a difference between the mean of administered doses for the two states for September 2021 and Novemeber 2021.

T = d/(SD/√n)

After applying T-test, if |T| > criticial_value, reject *H0*

Here critical_value = t(n-1,alp/2)

```
#Creating difference column
df_md_septnov ["Difference"] = " "
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWa
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stab
```

```
df_md_septnov["Difference"] = df_md_septnov['Administered_daily'] - df_la_septnov['Adm
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWa
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stab
  """Entry point for launching an IPython kernel.
```

```
d = df_md_septnov['Difference'].mean()
n=len(df_md_septnov)
```

```
stdd = df_md_septnov['Difference'].std()
print(stdd)
```

```
13106.258123316033
```

```
denominator = stdd / math.sqrt(n)
```

```
T = d / denominator
print(T)
```

    4.969030904078582

```
t_crit=2.003241
if (abs(T)>t_crit):
  print("We reject null hypothesis")
else:
  print("Accept null hypothesis")
```

    We reject null hypothesis

Our critical value, t(56,0.025) = 2.003241

Calculated T = 4.9690

As |T| > critical_value, we reject H0

The difference in mean of the administered vaccination is significant between the two states as the population of Maryland is 6.038 million and of Louisiana is 4.665 million. The diference of 1.5 million people significantly increases the mean administered vaccine of Maryland state. Also Maryland made a vaccination mandate starting month of September, this could also be a reason for higher vaccination observed in Maryland.