Covid Analysis - Visualisation. Answer the following questions:

Import library

In [1]:
```
import pandas as pd
from matplotlib import pyplot as plt
```

Loading Datasets

In [2]:
```
df = pd.read_csv('country_vaccination_preprocessed.csv')
df.head()
```

Out[2]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	d
0	Afghanistan	AFG	2021-05-27	593313.0	479574.0	113739.0	
1	Afghanistan	AFG	2021-06-03	630305.0	481800.0	148505.0	
2	Afghanistan	AFG	2022-01-27	5081064.0	4517380.0	3868832.0	
3	Albania	ALB	2021-02-18	3049.0	2438.0	611.0	
4	Albania	ALB	2021-05-11	622507.0	440921.0	181586.0	

Preprocessing

In [3]:
```
df.isnull().sum()  # Checking null values
```

Out[3]:
```
country                                0
iso_code                               0
date                                   0
total_vaccinations                     0
people_vaccinated                      0
people_fully_vaccinated                0
daily_vaccinations_raw                 0
daily_vaccinations                     0
total_vaccinations_per_hundred         0
people_vaccinated_per_hundred          0
people_fully_vaccinated_per_hundred    0
daily_vaccinations_per_million         0
vaccines                               0
source_name                            0
source_website                         0
year                                   0
month                                  0
day                                    0
month_name                             0
dtype: int64
```

In [4]:
```
df = df.fillna(0) # replace null with
```

In [8]:
```
def name(data):

    name = ['Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov',

    return name[data-1]
```

In [9]:
```
# Adding Year and Months columns

df['year']               = df[            'date' ].apply( lambda x : x.split(

df['month']              = df[            'date' ].apply( lambda x : int( x.s

df['month_name']         = df[            'month'].apply( name )

df['total_vaccinations'] = df['total_vaccinations'].apply( lambda x : int( x )
```

In [10]:
```
df.head()
```

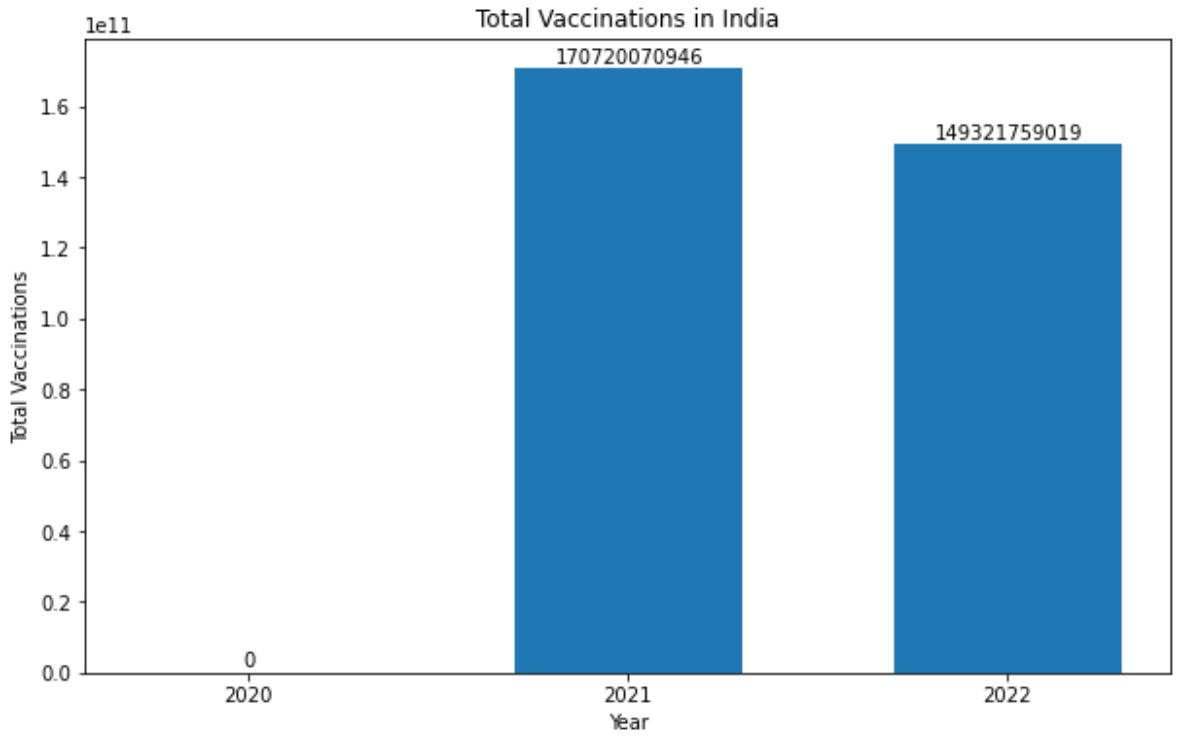
Out[10]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	d
0	Afghanistan	AFG	2021-05-27	593313	479574.0	113739.0	
1	Afghanistan	AFG	2021-06-03	630305	481800.0	148505.0	
2	Afghanistan	AFG	2022-01-27	5081064	4517380.0	3868832.0	
3	Albania	ALB	2021-02-18	3049	2438.0	611.0	
4	Albania	ALB	2021-05-11	622507	440921.0	181586.0	

1. Find the number of Total Vaccinations in India in Year 2020, 2021 and 2022

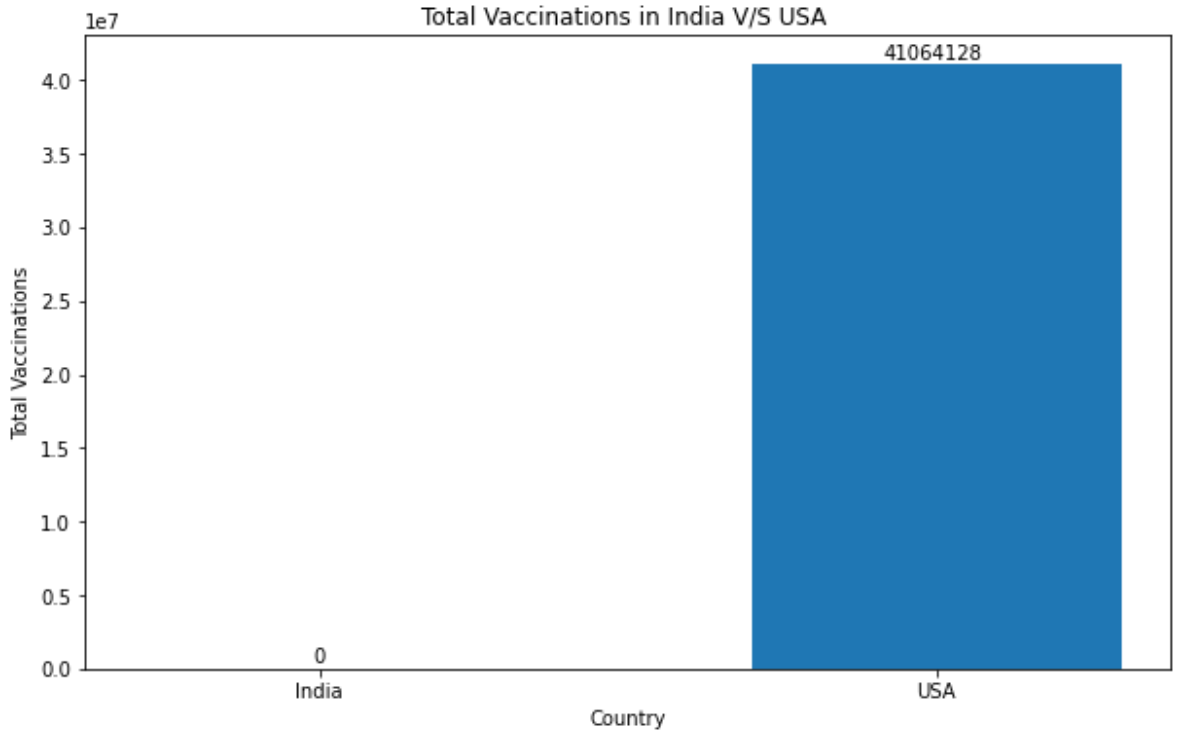
In [11]:
```
data  = df[df['country'] == 'India']
data  = data.groupby('year').sum()
data  = data['total_vaccinations']

x     = list(data.index)
y     = list(data.values)

x     = ['2020'] + x
y     = [0]      + y
print(x,y)

plt.figure(figsize=(10,6))
ax = plt.bar(x,y,width = 0.6)

for k,i in enumerate(ax):

    bar_width = i.get_width()
    x_cord    = i.xy[0] + bar_width / 2
    y_cord    = i.get_height()
    txt       = str(y[k])

    test      = ( bar_width + len(txt) ) / 2 + bar_width / 2

    offset    = -3 if y[k]==0 else -37

    plt.annotate(
        text   = txt,
        xy     = (x_cord , y_cord),
        xytext = ( offset , 2.9),
        textcoords="offset points"
                )

plt.xlabel('Year')
plt.ylabel('Total Vaccinations')

#plt.bar_label(ax.containers[0])

plt.title('Total Vaccinations in India')


plt.show()
```

```
['2020', '2021', '2022'] [0, 170720070946, 149321759019]
```


Total Vaccinations in India
1e11
170720070946
149321759019
0
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Total Vaccinations
2020
2021
2022
Year

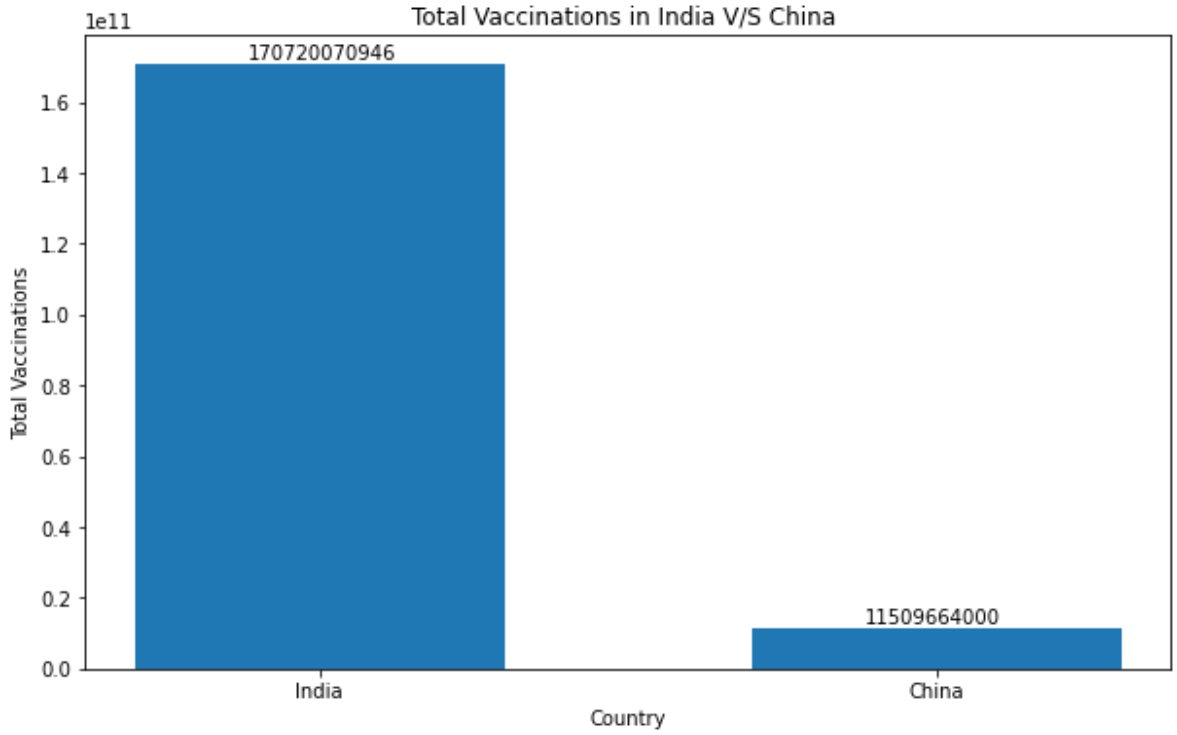
2. Compare number of total vaccinations in year 2020 of India and USA

In [12]:
```
data = df[  df[      'year'] ==  '2020']

IND  = data[data[ 'country'] == 'India'].total_vaccinations.sum()

USA  = data[data['iso_code'] ==   'USA'].total_vaccinations.sum()

x    = [ 'India', 'USA' ]
y    = [  IND   ,  USA  ]


plt.figure(figsize=(10,6))
ax = plt.bar(x,y,width = 0.6)

for k,i in enumerate(ax):

    bar_width = i.get_width()
    x_cord    = i.xy[0] + bar_width / 2
    y_cord    = i.get_height()
    txt       = str(y[k])

    test      = ( bar_width + len(txt) ) / 2 + bar_width / 2

    offset    = -3 if y[k]==0 else -27

    plt.annotate(
        text   = txt,
        xy     = (x_cord , y_cord),
        xytext = ( offset , 2.9),
        textcoords="offset points"
                )

plt.xlabel('Country')
plt.ylabel('Total Vaccinations')


plt.title('Total Vaccinations in India V/S USA ')


plt.show()
```


Total Vaccinations in India V/S USA
1e7
41064128
0
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Total Vaccinations
India
USA
Country

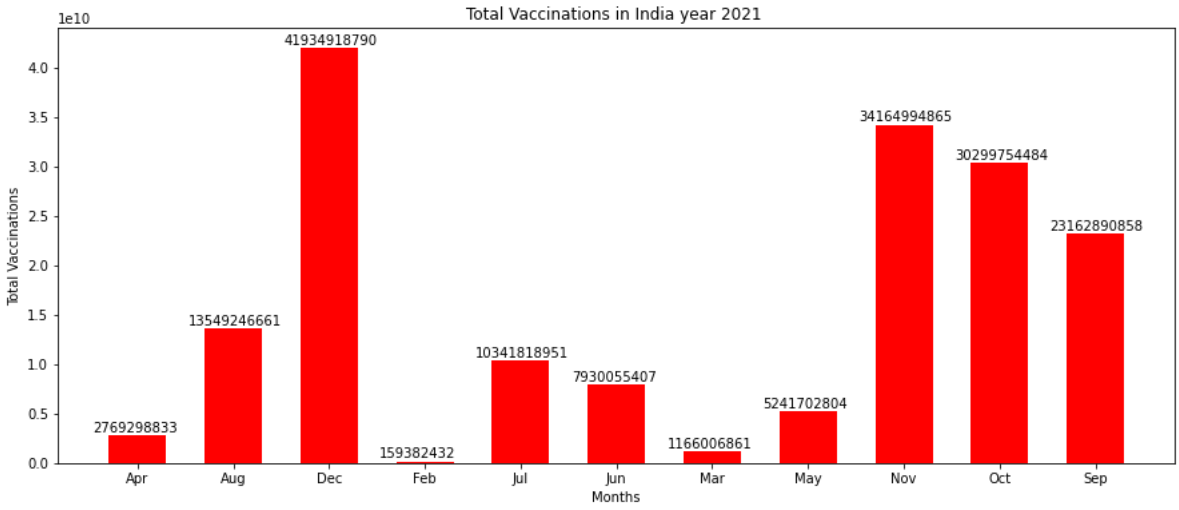
3. Compare number of total vaccinations in year 2021 of India and China

In [13]:
```
data = df[  df[      'year'] ==  '2021']

IND  = data[data['country'] == 'India'].total_vaccinations.sum()

CHI  = data[data['country'] == 'China'].total_vaccinations.sum()

x    = [ 'India', 'China' ]
y    = [  IND   ,  CHI  ]


plt.figure(figsize=(10,6))
ax = plt.bar(x,y,width = 0.6)

for k,i in enumerate(ax):

    bar_width = i.get_width()
    x_cord    = i.xy[0] + bar_width / 2
    y_cord    = i.get_height()
    txt       = str(y[k])

    test      = ( bar_width + len(txt) ) / 2 + bar_width / 2

    offset    = -3 if y[k]==0 else -37

    plt.annotate(
        text   = txt,
        xy     = (x_cord , y_cord),
        xytext = ( offset , 2.9),
        textcoords="offset points"
                )

plt.xlabel('Country')
plt.ylabel('Total Vaccinations')


plt.title('Total Vaccinations in India V/S China ')


plt.show()
```


Total Vaccinations in India V/S China
1e11
170720070946
11509664000
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Total Vaccinations
India
China
Country

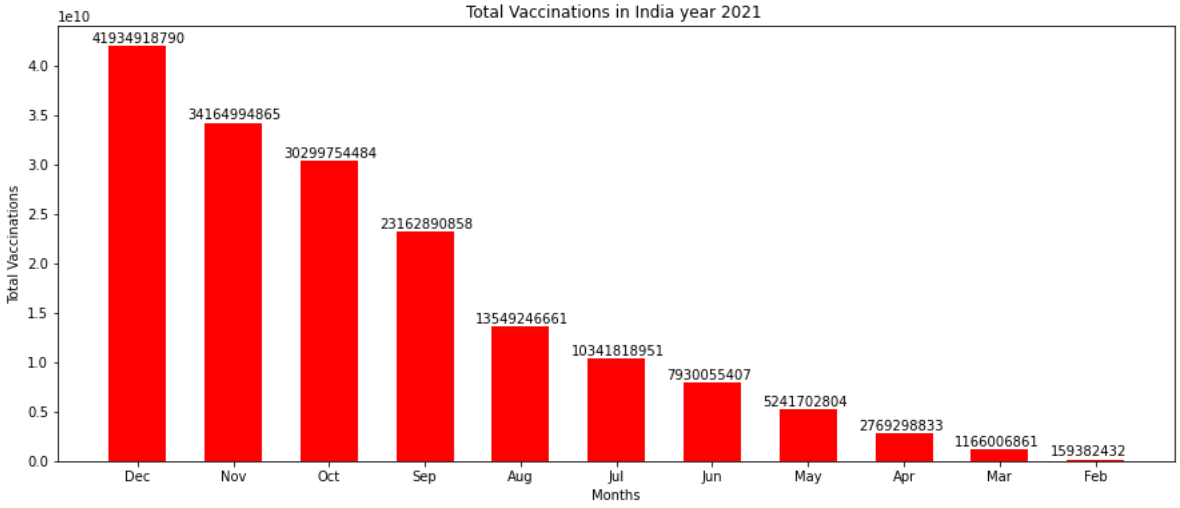
4. Find the number of Vaccinations in each month in India in the year 2021

In [14]:
```
data = df[   df[      'year'] ==  '2021']
data = data[ data[ 'country'] == 'India']

data = data.groupby('month_name').sum().total_vaccinations

x    = data.index
y    = data.values


plt.figure(figsize=(15,6))
ax = plt.bar(x,y,width = 0.6, color = 'red')

for k,i in enumerate(ax):

    bar_width = i.get_width()
    x_cord    = i.xy[0] + bar_width / 2
    y_cord    = i.get_height()
    txt       = str(y[k])

    test      = ( bar_width + len(txt) ) / 2 + bar_width / 2

    plt.annotate(
        text   = txt,
        xy     = (x_cord , y_cord),

        xytext = ( -34 , 2.9 ),
        textcoords="offset points"

                )

plt.xlabel('Months')
plt.ylabel('Total Vaccinations')


plt.title('Total Vaccinations in India year 2021 ')


plt.show()
```


Total Vaccinations in India year 2021
1e10
Total Vaccinations
Months
Apr
Aug
Dec
Feb
Jul
Jun
Mar
May
Nov
Oct
Sep
2769298833
13549246661
41934918790
159382432
10341818951
7930055407
1166006861
5241702804
34164994865
30299754484
23162890858

5. Which month has the most number of total vaccinations in India in 2021?

In [15]:
```
data = df[   df[      'year'] ==  '2021']
data = data[ data[ 'country'] == 'India']

data = data.groupby('month_name').sum().total_vaccinations
data = data.sort_values(ascending=False)

x    = data.index
y    = data.values


plt.figure(figsize=(15,6))
ax = plt.bar(x,y,width = 0.6, color = 'red')

for k,i in enumerate(ax):

    bar_width = i.get_width()
    x_cord    = i.xy[0] + bar_width / 2
    y_cord    = i.get_height()
    txt       = str(y[k])

    test      = ( bar_width + len(txt) ) / 2 + bar_width / 2

    plt.annotate(
        text   = txt,
        xy     = (x_cord , y_cord),

        xytext = ( -34 , 2.9 ),
        textcoords="offset points"

               )

plt.xlabel('Months')
plt.ylabel('Total Vaccinations')


plt.title('Total Vaccinations in India year 2021 ')


plt.show()
```


Total Vaccinations in India year 2021
1e10
41934918790
34164994865
30299754484
23162890858
13549246661
10341818951
7930055407
5241702804
2769298833
1166006861
159382432
Total Vaccinations
Dec
Nov
Oct
Sep
Aug
Jul
Jun
May
Apr
Mar
Feb
Months