BDP - Assignment 4 - BhoiteShivani - CodeOnly

In [1]:

```
import pandas as pd
import requests
import matplotlib
import matplotlib.pyplot as plt
import numpy as np
import us
```

Year 1

url = "https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2016-01-01&endtime=2017-01-01&minmagnitude=4 (https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2016-01-01&endtime=2017-01-01&minmagnitude=4)"

Year1 = requests.get(url).text f = open("Year1.csv", "w") f.write(Year1) f.close()

Year 2

url = "https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2017-01-01&endtime=2018-01-01&minmagnitude=4 (https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2017-01-01&endtime=2018-01-01&minmagnitude=4)"

Year2 = requests.get(url).text f = open("Year2.csv", "w") f.write(Year2) f.close()

Year 3

url = "https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2018-01-01&endtime=2019-01-01&minmagnitude=4 (https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2018-01-01&endtime=2019-01-01&minmagnitude=4)"

Year3 = requests.get(url).text f = open("Year3.csv", "w") f.write(Year3) f.close()

Year 4

url = "https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2019-01-01&endtime=2019-10-02&minmagnitude=4 (https://earthquake.usgs.gov/fdsnws/event/1/query?format=csv&starttime=2019-01-01&endtime=2019-10-02&minmagnitude=4)"

Year4 = requests.get(url).text f = open("Year4.csv", "w") f.write(Year4) f.close()

In [65]:

```
df_year1 = pd.read_csv('Year1.csv')
df_year2 = pd.read_csv('Year2.csv')
df_year3 = pd.read_csv('Year3.csv')
df_year4 = pd.read_csv('Year4.csv')


frames=[df_year1,df_year2,df_year3,df_year4]
all_data_df=pd.concat(frames,sort=False)

all_data_df.to_csv('All_Data.csv')

#All the data is combined and Stored into All_data.csv
```

In [ ]:

```
#Reading from the csv which contains the entire data
final_df = pd.read_csv('All_Data.csv')

final_df = final_df.loc[:, ~final_df.columns.str.contains('^Unnamed')]

final_df.count()
```

In [ ]:

```
final_df=final_df[final_df['type']=='earthquake']
final_df.count()
```

1) Use describe to get the basic statistics of all the columns (5 points)

In [ ]:

```
final_df.describe()
```

2) Get the top 10 earthquakes by magnitude (5 points)

In [ ]:

```
largest_mag_df = final_df.nlargest(10, 'mag')

largest_mag_df[['place','mag']]
```

3) Handle all Null/empty data by filling it with zeros (10 points)

In [ ]:

```
final_df=final_df.fillna(0)
final_df
```

4) Find the top 10 places where the strongest earthquakes occurred (15 points)

```
In [ ]: newdf=final_df.nlargest(10, 'mag')
        newdf['place'].str.split('of').str[-1]
```

5) Find the top 10 places where the weakest earthquakes occurred (15 points)

```
In [ ]: newdf=final_df.nsmallest(10, 'mag')
        newdf['place'].str.split('of').str[-1]
```

6) On a per-year basis, use a bar chart to plot the number of earthquakes for each of the following magnitude groups ranges: Group 1: [4,4.5), Group 2: [4.5,5), Group 3: [5,6), Group 4: [6,7), Group 5: (7,MAX]. Pay close attention to the group ranges. (20 points) Please add labels and colors to the plot.

In [ ]:

```
#Dividing the entire year into 4 groups
maxmag= final_df['mag'].max()
bins = 4,4.5,5,6,7,maxmag #for naming the xticks


df_year1_group1 = df_year1[(df_year1['mag']>=4) & (df_year1['mag']<4.5)]
df_year1_group2 = df_year1[(df_year1['mag']>=4.5) & (df_year1['mag']<5)]
df_year1_group3 = df_year1[(df_year1['mag']>=5) & (df_year1['mag']<6)]
df_year1_group4 = df_year1[(df_year1['mag']>=6) & (df_year1['mag']<7)]
df_year1_group5 = df_year1[(df_year1['mag']>7) & (df_year1['mag']<=maxma
g)]


df_year2_group1 = df_year2[(df_year2['mag']>=4) & (df_year2['mag']<4.5)]
df_year2_group2 = df_year2[(df_year2['mag']>=4.5) & (df_year2['mag']<5)]
df_year2_group3 = df_year2[(df_year2['mag']>=5) & (df_year2['mag']<6)]
df_year2_group4 = df_year2[(df_year2['mag']>=6) & (df_year2['mag']<7)]
df_year2_group5 = df_year2[(df_year2['mag']>7) & (df_year2['mag']<=maxma
g)]


df_year3_group1 = df_year3[(df_year3['mag']>=4) & (df_year3['mag']<4.5)]
df_year3_group2 = df_year3[(df_year3['mag']>=4.5) & (df_year3['mag']<5)]
df_year3_group3 = df_year3[(df_year3['mag']>=5) & (df_year3['mag']<6)]
df_year3_group4 = df_year3[(df_year3['mag']>=6) & (df_year3['mag']<7)]
df_year3_group5 = df_year3[(df_year3['mag']>7) & (df_year3['mag']<=maxma
g)]


df_year4_group1 = df_year4[(df_year4['mag']>=4) & (df_year4['mag']<4.5)]
df_year4_group2 = df_year4[(df_year4['mag']>=4.5) & (df_year4['mag']<5)]
df_year4_group3 = df_year4[(df_year4['mag']>=5) & (df_year4['mag']<6)]
df_year4_group4 = df_year4[(df_year4['mag']>=6) & (df_year4['mag']<7)]
df_year4_group5 = df_year4[(df_year4['mag']>7) & (df_year4['mag']<=maxma
g)]
```

In [ ]:

```
#Year 2016
plt.hist((df_year1_group1['mag'],df_year1_group2['mag'],df_year1_group3[
'mag'],df_year1_group4['mag'],df_year1_group5['mag']),
         bins=None, histtype='bar',align='left', rwidth=10)
plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'),rotation=
90)
plt.xlabel('Maginitude')
plt.ylabel('Frequency of Earthquake')
plt.title("Year 2016 - Earthquake count vs Magnitude")
plt.legend(('Group1','Group2','Group3','Group4','Group5'))
plt.show()

#Year 2017
plt.hist((df_year2_group1['mag'],df_year2_group2['mag'],df_year2_group3[
'mag'],df_year2_group4['mag'],df_year2_group5['mag']),
         bins=None, histtype='bar',align='left', rwidth=10)
plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'),rotation=
90)
plt.xlabel('Maginitude')
plt.ylabel('Frequency of Earthquake')
plt.title("Year 2017 - Earthquake count vs Magnitude")
plt.legend(('Group1','Group2','Group3','Group4','Group5'))
plt.show()


#Year 2018
plt.hist((df_year3_group1['mag'],df_year3_group2['mag'],df_year3_group3[
'mag'],df_year3_group4['mag'],df_year3_group5['mag']),
         bins=None, histtype='bar',align='left', rwidth=10)
plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'),rotation=
90)
plt.xlabel('Maginitude')
plt.ylabel('Frequency of Earthquake')
plt.title("Year 2018 - Earthquake count vs Magnitude")
plt.legend(('Group1','Group2','Group3','Group4','Group5'))
plt.show()


#Year 2019
plt.hist((df_year4_group1['mag'],df_year4_group2['mag'],df_year4_group3[
'mag'],df_year4_group4['mag'],df_year4_group5['mag']),
         bins=None, histtype='bar',align='left', rwidth=10)
plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'),rotation=
90)
plt.xlabel('Maginitude')
plt.ylabel('Frequency of Earthquake')
plt.title("Year 2019 - Earthquake count vs Magnitude")
plt.legend(('Group1','Group2','Group3','Group4','Group5'))
plt.show()
```

In [ ]:

```
#Trying to plot the entire data in one Map

bins = 4,4.5,5,6,7,maxmag

plt.hist((df_year1['mag'],df_year2['mag'],df_year3['mag'],df_year4['mag'
]), bins, histtype='bar', align='left', rwidth=0.60)

plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'))
plt.xlabel('Maginitude')
plt.ylabel('Earth quake count')
plt.title("Earthquake count vs Magnitude")
plt.legend(('2016','2017','2018','2019'))
plt.show()
```

7) Find the 10 countries with the highest number of earthquakes (30 points) (Note: Yes, this is only countries, not full place)

In [ ]:

```
newdf=final_df
newdf['place']=newdf['place'].str.split('of').str[-1]
newdf['place']=newdf['place'].str.split(',').str[-1]

newdf['place'].value_counts().head(10)
```

8) Analyze the distribution of the Earthquake magnitudes. This is, make a histogram of the Earthquake count versus magnitude. Make sure to use a Logarithmic scale. What sort of relationship do you see? (20 points) P lease add labels and colors to the plot.

In [ ]:

```
plt.hist(x=((df_year1['mag'],df_year2['mag'],df_year3['mag'],df_year4['m
ag'])), align='left', log=True)
plt.xlabel('Maginitude')
plt.ylabel('Earth quake count')
plt.title("Earthquake count vs Magnitude")
plt.legend(('2016','2017','2018','2019'))
plt.show()
```

Observation - The bar graph here is a UniModal bar graph. It has 1 peak. This shows that there are more number of 4 magnitude earth quakes. And as the magnitude of earth quake increases the frequency of that earth quake reduces. Showing a inverse relationship between the magnitude and the frequency.

9) Analyze the distribution of the Earthquake depths. This is, make a histogram of the Earthquake count versus depth. Make sure to use a Logarithmic scale. What sort of relationship do you see? (20 points) Please add labels and colors to the plot.

In [ ]:

```
plt.hist(x=((df_year1['depth'],df_year2['depth'],df_year3['depth'],df_year4['depth'])), align='left',log=True)
plt.xlabel('Depth')
plt.ylabel('Earth quake count')
plt.title("Earthquake count vs Magnitude")
plt.legend(('2016','2017','2018','2019'))
plt.show()
```

Observation - The bar graph here is a Bi Modal bar graph. Here we can observe that there are 2 peaks. The Highest Peak is observed for the lowest depth and then 2nd highest peak is observed at depth between 500-600.

10) Visualize the locations of earthquakes by making a scatterplot of their latitude and longitude. (20 points) Please add labels and colors to the plot.

In [ ]:

```
plt.scatter(x=final_df['latitude'],y=final_df['longitude'], alpha=0.4,edgecolors="blue")
plt.xlabel('Latitude')
plt.ylabel('Longitude')
plt.title("Latitude v/s Longitude")
plt.show()
```

In [ ]:

```
import matplotlib.image as mpimg
import matplotlib.image as mpimg
world_img=mpimg.imread('world.png')

plt.scatter(x=final_df['latitude'],y=final_df['longitude'], alpha=0.01,edgecolors="blue")
plt.imshow(world_img, cmap='viridis',extent=[120,-120,-150, 160], alpha=1)

plt.ylabel("Longitude", fontsize=14)
plt.xlabel("Latitude", fontsize=14)
plt.title("Longitude v/s Latitude",fontsize=16)
plt.show()
```

11) Using the US package (https://pypi.org/project/us/ (https://pypi.org/project/us/)), clean the dataset you used previously to only have data from the USA . You need to create a function that accommodates this. (20 points)

In [20]:

```
import us
```

In [ ]:

```
newdf=final_df
#Taking only the last variable, i.e country from the place
newdf['place']=newdf['place'].str.split('of').str[-1]
newdf['place']=newdf['place'].str.split(',').str[-1]
newdf['place']=newdf['place'].str.lstrip()

#making a new dataframe that will store only the US data
US_df = pd.DataFrame(columns=['time', 'latitude', 'longitude', 'depth', 'mag', 'magType', 'nst',
        'gap', 'dmin', 'rms', 'net', 'id', 'updated', 'place', 'type',
        'horizontalError', 'depthError', 'magError', 'magNst', 'status',
        'locationSource', 'magSource'])
US_df
```

In [ ]:

```
for index, row in newdf.iterrows():
    if((us.states.lookup(row['place']))!=None):
        US_df=US_df.append(row)
```

In [ ]:

```
US_df.describe()
```

In [ ]:

```
US_df.count()
```

12) Using the cleaned US-only version of your dataset answer the following questions (30 points total):

Note: Question 12 has to be correct and match my own solutions, otherwise you will receive no marks for these responses.

a) Find the top 10 US states where the strongest earthquakes occurred (10 points)

In [ ]:

```
temp_df=US_df.nlargest(10, 'mag')
temp_df[['place','mag']]
```

b) On a per-year basis, use a bar chart to plot the number of earthquakes for each of the following magnitude groups ranges: Group 1: [4,4.5), Group 2: [4.5,5), Group 3: [5,6), Group 4: [6,7), Group 5: (7,MAX]. Pay close attention to the group ranges. (10 points) Please add labels and colors to the plot.

In [ ]:

```
maxmag=US_df['mag'].max()

US_df_group1 = US_df[(US_df['mag']>=4) & (US_df['mag']<4.5)]
US_df_group2 = US_df[(US_df['mag']>=4.5) & (US_df['mag']<5)]
US_df_group3 = US_df[(US_df['mag']>=5) & (US_df['mag']<6)]
US_df_group4 = US_df[(US_df['mag']>=6) & (US_df['mag']<7)]
US_df_group5 = US_df[(US_df['mag']>7) & (US_df['mag']<=maxmag)]


plt.hist((US_df_group1['mag'],US_df_group2['mag'],US_df_group3['mag'],US
_df_group4['mag'],US_df_group5['mag']),
         bins=None, histtype='bar',align='left', rwidth=10)
plt.xticks(bins,('Group1','Group2','Group3','Group4','Group5'),rotation=
90)
plt.xlabel('Maginitude')
plt.ylabel('Frequency of Earthquake')
plt.title("US Earthquake count vs Magnitude")
plt.legend(('Group1','Group2','Group3','Group4','Group5'))
plt.show()
```

c) Visualize the locations of earthquakes by making a scatterplot of their latitude and longitude. Overlay a US map on top of this plot to match the locations. (20 points) Please add labels and colors to the plot.

In [41]:

```
import matplotlib.image as mpimg
import matplotlib.image as mpimg
us_img=mpimg.imread('us map.png')
```

In [ ]:

```
plt.scatter(x=US_df['latitude'],y=US_df['longitude'], alpha=0.4,edgecolo
rs="blue")

plt.imshow(us_img, cmap='viridis',extent=[-16, 200, -270, 200], alpha=1)

plt.ylabel("Longitude", fontsize=14)
plt.xlabel("Latitude", fontsize=14)
plt.title("Mapping of US Earthquake",fontsize=19)
plt.show()
```

In [ ]: