**t**

**Table of content**

**1.READING THE DATA SET**

1. **DATA PREPROCESSING**
2. **VISUALIZTION OF DATA**
3. **Clustering techniques used**

**erarD**

**Earthquakes, 1965-2016**

**Significant Earthquakes from 1965 to 2016**

Date time and location of all the earthquakes with magnitude of 5.5 or higherThis dataset includes a record of the date, time, location, depth, magnitude, and source of every earthquake with a reported magnitude 5.5 or higher since 1965.

The data set was downloaded from

<https://www.kaggle.com/usgs/earthquake-database>

Jupyter notebook was used to run the program. Different necessaries libraries were installed to read and show the output. Each line were run to see the output.

Plotly was also installed in the anaconda file in this project

Pip install plotly

conda install -c plotly plotly

1. **READING THE DATA SET**

This is a large containing attribute like date, time, latitude, longitude and many different attributes. Most of the attributes here are not necessary we take Latitude, longitude, depth and magnitude into consideration while performing the clustering.

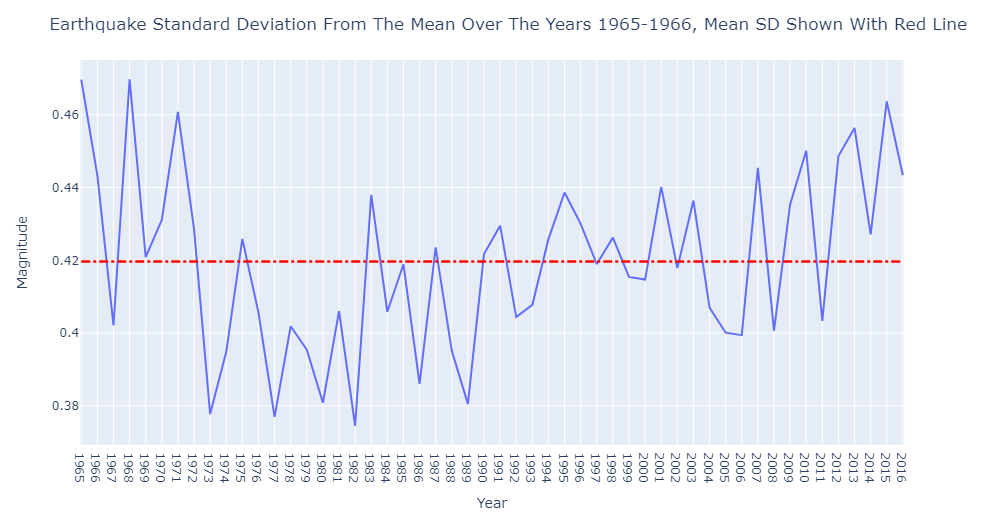
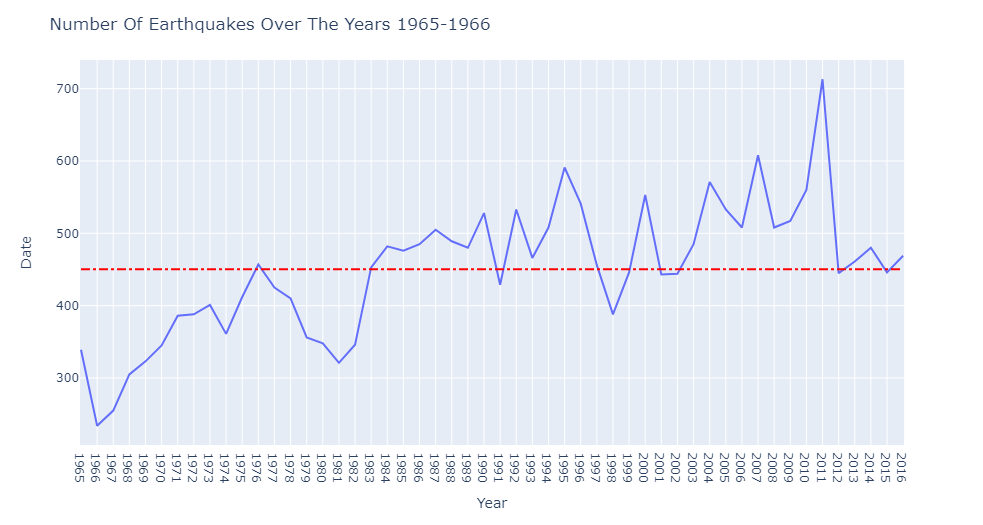
**2.DATA PREPROCESSING**

### There were missing data in the data set which was processed in the jupyter notebook. we then use the mode of the magnitude type feature to replace the small number of missing values, as for the other features which can be done by regression and nearest neighbor approach.

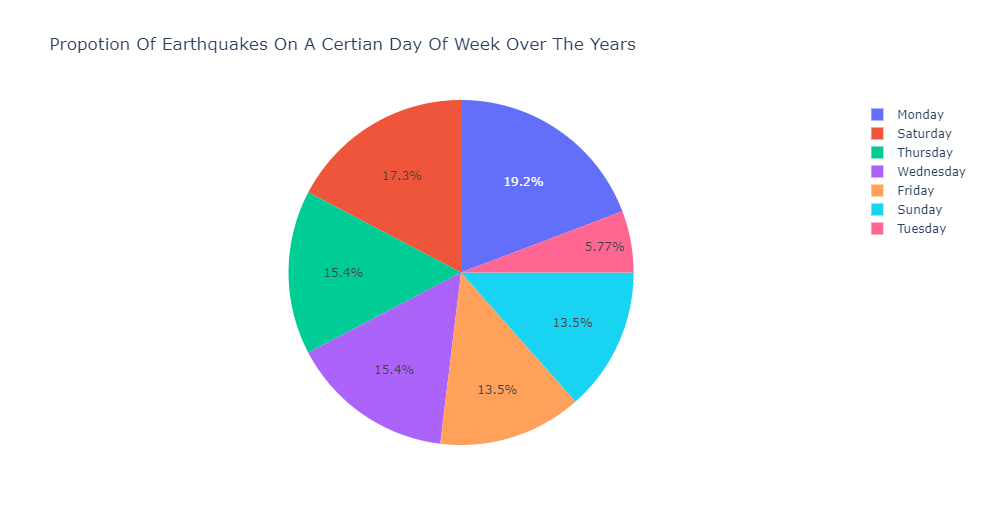
### We then perform the skewness of the data to see the positive skewness and negative skewness of the data. It helps us to see the direction and relative magnitude of a distribution's deviation from the normal distribution.

### 3.VISUALIZTION OF DATA

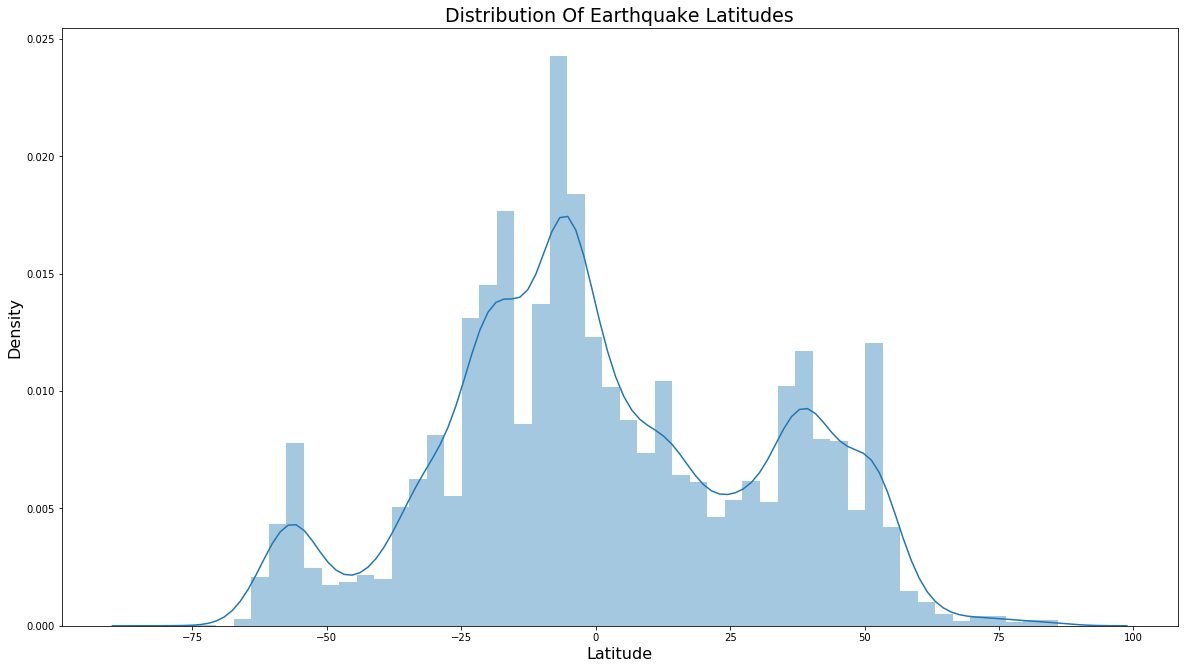
### Distibution Of Yearly Magnitude/Deapth Over Our Time Line For Each Partition

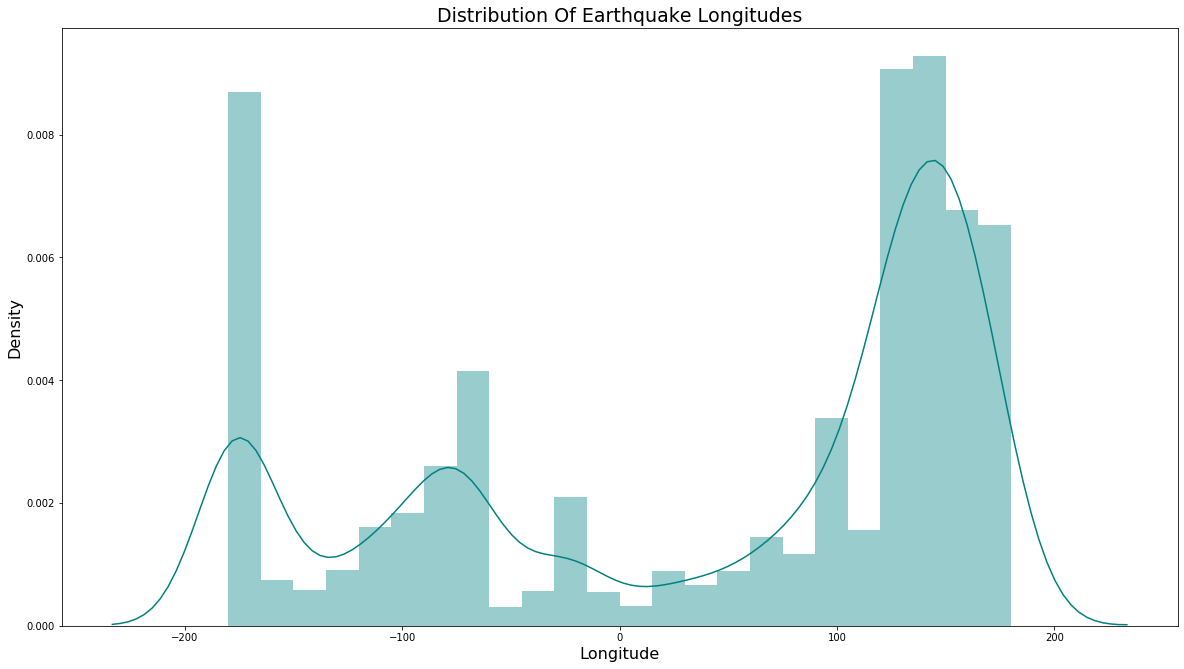


Plotting number of earth quake over the years . we see the level of magnitude have increased and decreased over the years and it shows growing trend from yeas of lowest magnitude record in 1982 and highest in 1965.



The pie chard shows the amount of earthquake among the days of the week which indicates equal amount distributed in the weeks.





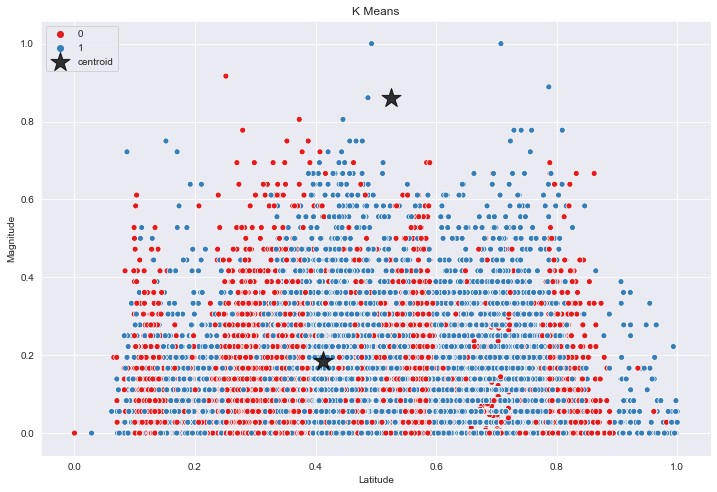
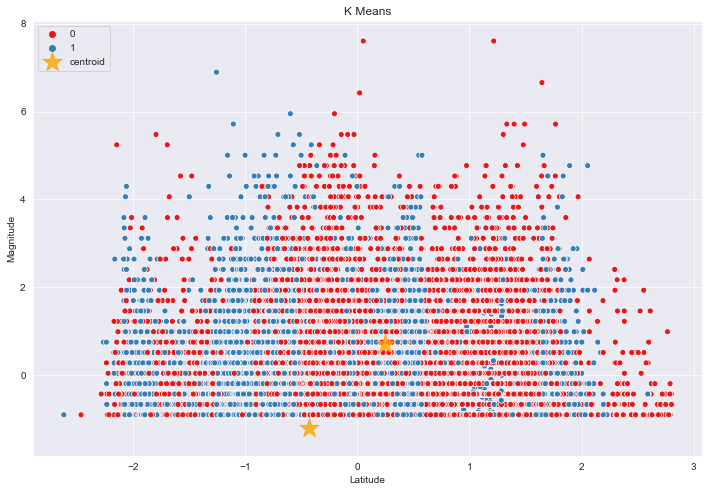
Distribution of earthquake in latitude and longitude.

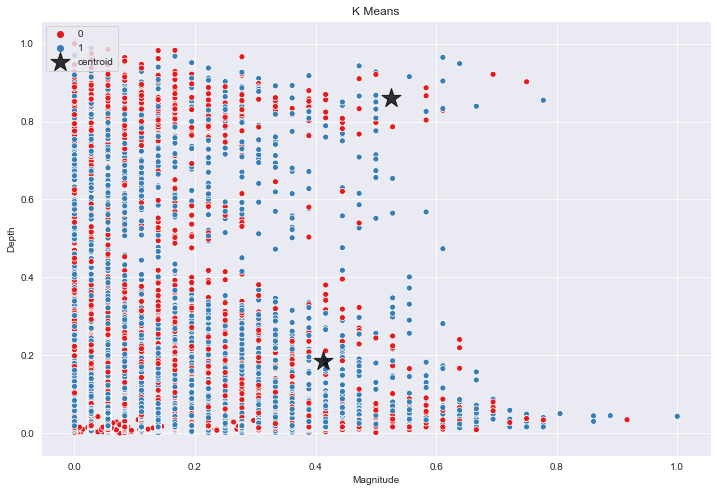
**4. Clustering techniques used**

K mean clustering and DB Scan clustering algorithm were used by the mean of sklearn libraries in jupyter notebook.

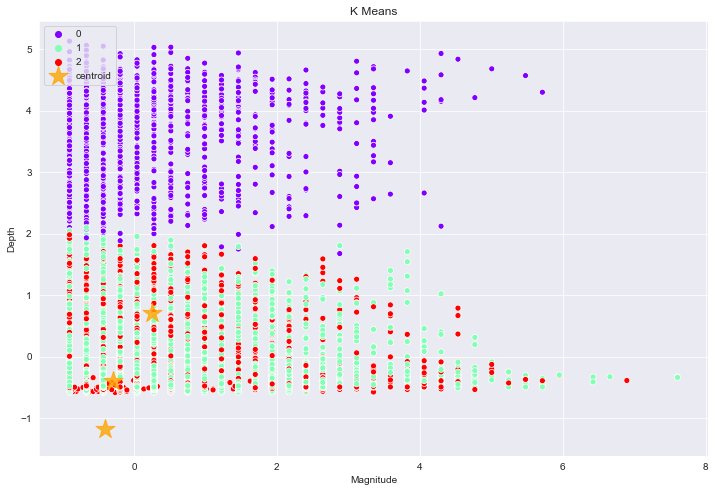
4 attributes, 'Latitude','Longitude','Depth','Magnitude' were chosen which was normalized .

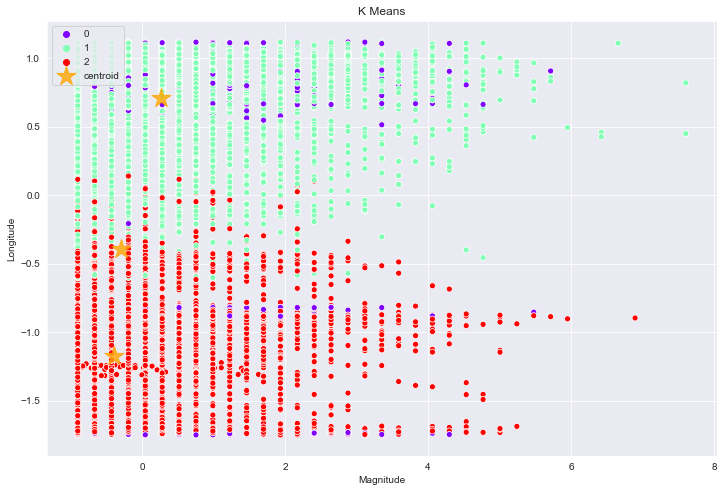
Using k as 2

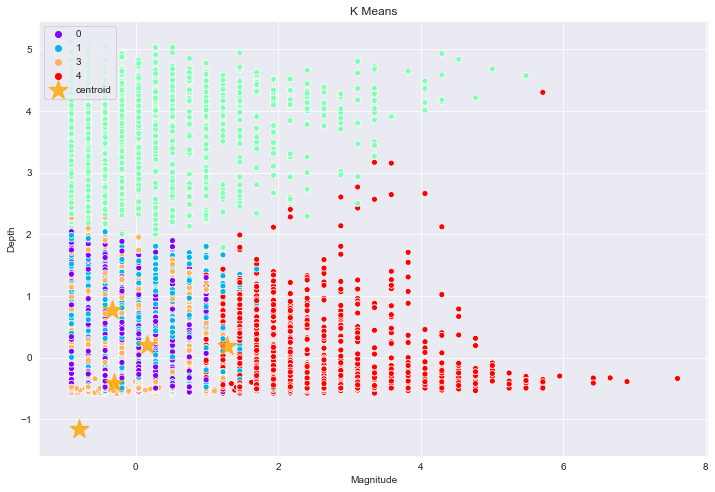
 



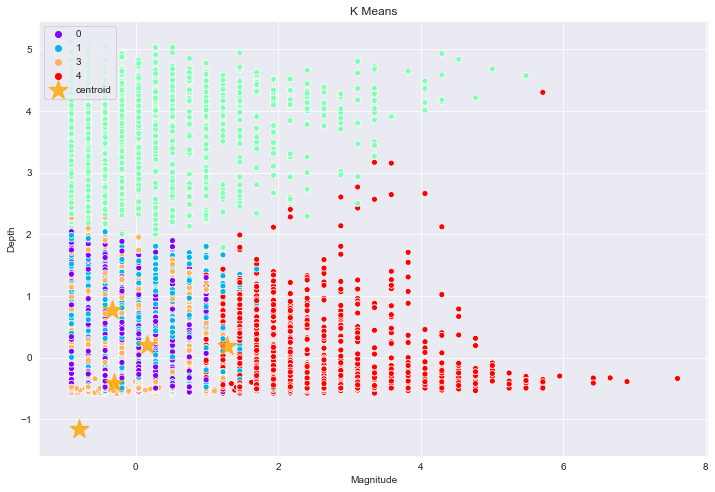
Using K as 3

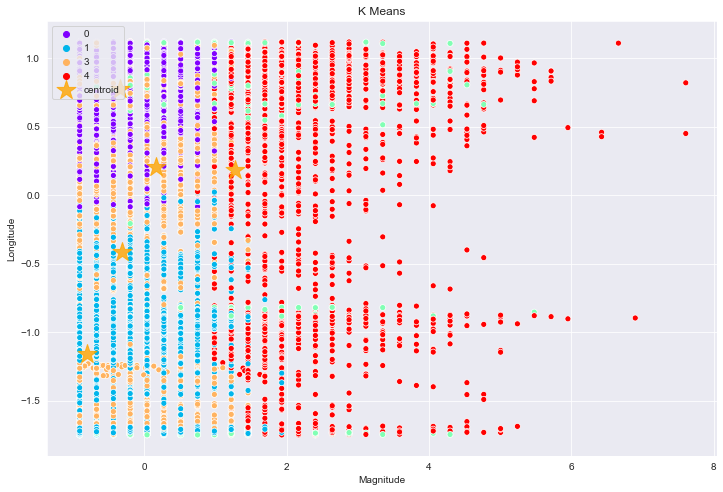






Using K as 5





DBSCAN ALGORITHM

Using EPS as 0.5 and Min sample as 15

