**EarthQuake Analysis**

**Introduction**

Creating a big data analytics pipeline, using big data technologies like PySpark, MLlib, Power BI and MongoDB.

**Problem Statement**

Create a model to predict the possibility of earth quakes based on sampledata from 1965-2016.

Working with earthquake data that will be transformed into summary tables.Then use these tables to train predictive models and predict future earthquakes. Write the final datasets to Mongodb.After that analyze the data by building reports and dashboards in Power BI Desktop

**Dataset Details**

Csv files.

Dabase.csv : This file will bethe source file containing earth quake details

**Steps Performed**

1.Load the dataset containing earthquake details

2.Drop the columns which are not required

3.Created a year column from date

4.Build the quakes frequency dataframe using the year field and counts for each year

5.Check the schema and convert some fields from string to numeric

6.Create avg magnitude and max magnitude fields and add to df\_quake\_freq

7.Write the dataframes to mongodb

**Details**

Dataset containing earth quake details are uploaded into Filestore of databricks

/dbfs/FileStore/ProjectPyspark

1.Load the dataset containing earthquake details

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Sample data:

Date='01/02/1965', Time='13:44:18', Latitude='19.246', Longitude='145.616', Type='Earthquake', Depth='131.6', Depth Error=None, Depth Seismic Stations=None, Magnitude='6', Magnitude Type='MW', Magnitude Error=None, Magnitude Seismic Stations=None, Azimuthal Gap=None, Horizontal Distance=None, Horizontal Error=None, Root Mean Square=None, ID='ISCGEM860706', Source='ISCGEM', Location Source='ISCGEM', Magnitude Source='ISCGEM', Status='Automatic'

2.Drop the columns which are not required

A screenshot of a computer

Description automatically generated

3.Create a year column from date

A screenshot of a computer

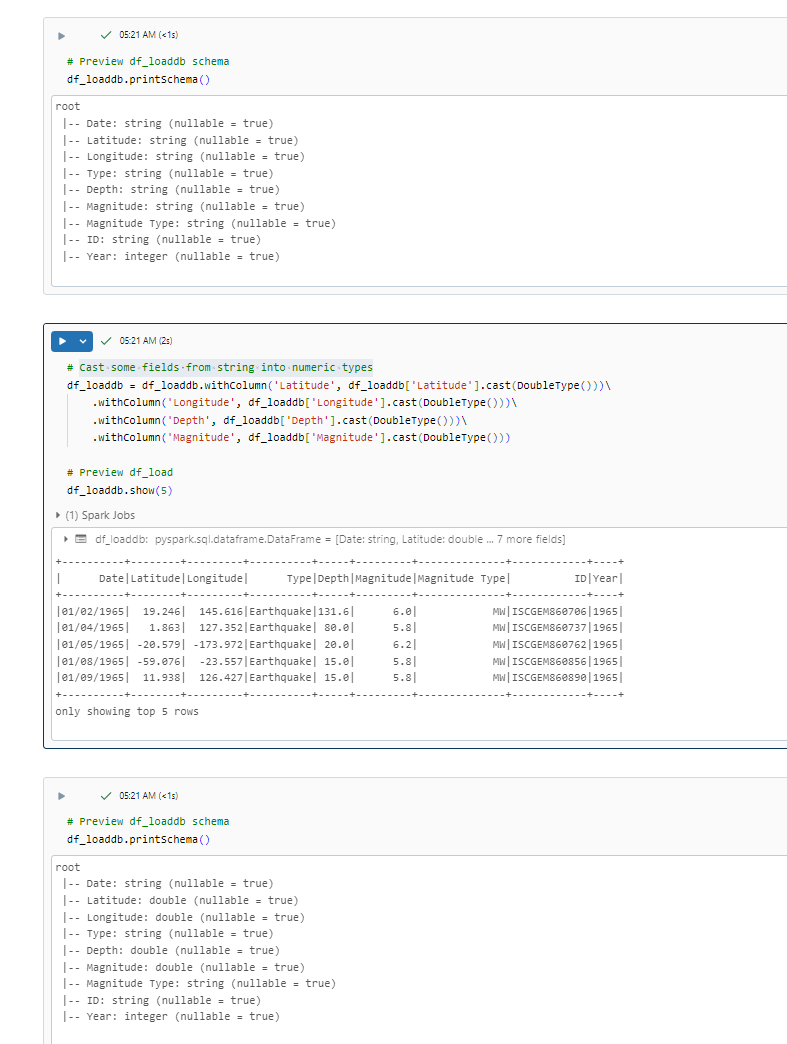
Description automatically generated

4.Build the quakes frequency data frame using the year field and counts for each year

A screenshot of a computer

Description automatically generated

5.Check the schema and convert some fields from string to numeric



6.Create avg magnitude and max magnitude fields and join to df\_quake\_freq

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A white rectangular object with black text

Description automatically generated

7.Write the data frames to mongo Db

To establish a connection from databricks to mongodb we need to install mongo-spark connector library.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Write from databricks to mongodb

A screenshot of a computer

Description automatically generated

Mongodb:

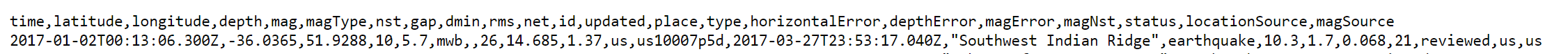
A screenshot of a computer

Description automatically generated

**Machine Learning**

8.Load the Query.csv file which was stored in the HDFS which contains the earthquake test data

Sample



A screenshot of a computer

Description automatically generated

9.Load the training data from mongodb and Perform the data cleansing activity such as selecting only required fields,Renaming fields,casting specific fields to the desired datatype



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

10.Create testing and training dataframes

A screen shot of a computer

Description automatically generated

11. Import the relevant modules for machine learning and create mode land evaluate it

RandomForestRegressor

VectorAssembler

RegressionEvaluator

Select features to parse into model and then create the feature vector. After that create and train the model and make the predictions.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Evaluate the model.rmse should be less than 0.5 for the model to be useful

A screenshot of a computer

Description automatically generated

12.Create the prediction Dataset

A screenshot of a computer

Description automatically generated

13.Load the Prediction dataset to Mongodb

A black and red line

Description automatically generated

A screenshot of a computer

Description automatically generated

14.Connect the tables from MongoDB using Power BI

<https://www.mongodb.com/docs/atlas/data-federation/query/sql/powerbi/connect/#std-label-sql-connect-powerbi>

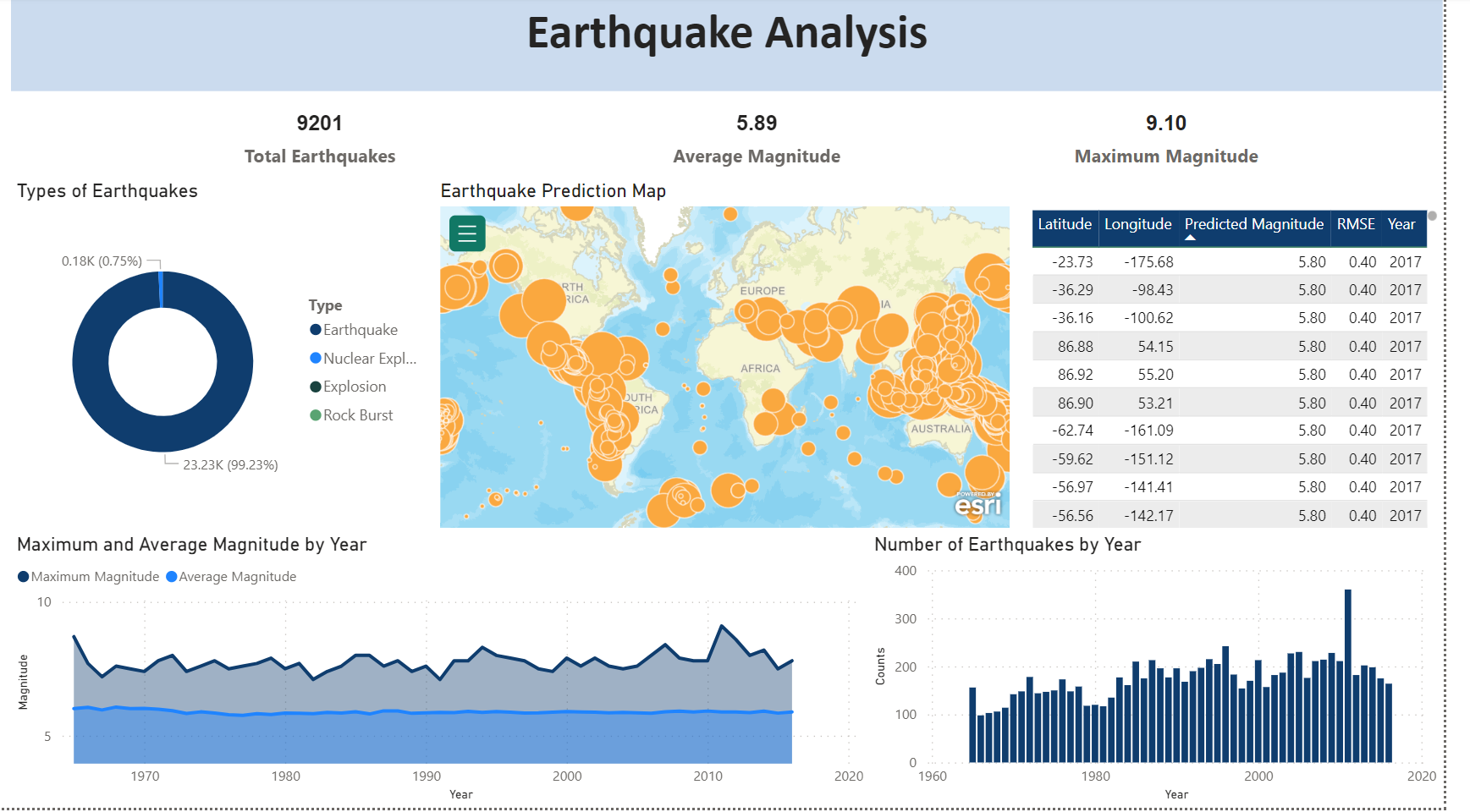
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

15.Create a report in Power BI which will showcase the different metrics and predictions.



A screenshot of a computer screen

Description automatically generated