

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from datetime import datetime
```

```
df = pd.read_csv('database.csv')
```

```
df.head()
```

	Date	Time	Latitude	Longitude	Type	Depth	Depth
Error \							
0	01/02/1965	13:44:18	19.246	145.616	Earthquake	131.6	
NaN							
1	01/04/1965	11:29:49	1.863	127.352	Earthquake	80.0	
NaN							
2	01/05/1965	18:05:58	-20.579	-173.972	Earthquake	20.0	
NaN							
3	01/08/1965	18:49:43	-59.076	-23.557	Earthquake	15.0	
NaN							
4	01/09/1965	13:32:50	11.938	126.427	Earthquake	15.0	
NaN							

	Depth	Seismic Stations	Magnitude	Magnitude	Type	...	\
0		NaN	6.0		MW	...	
1		NaN	5.8		MW	...	
2		NaN	6.2		MW	...	
3		NaN	5.8		MW	...	
4		NaN	5.8		MW	...	

	Magnitude	Seismic Stations	Azimuthal	Gap	Horizontal	Distance	\
0		NaN		NaN		NaN	
1		NaN		NaN		NaN	
2		NaN		NaN		NaN	
3		NaN		NaN		NaN	
4		NaN		NaN		NaN	

	Horizontal Error	Root Mean Square	ID	Source	Location
Source \					
0	NaN	NaN	ISCGEM860706	ISCGEM	
ISCGEM					
1	NaN	NaN	ISCGEM860737	ISCGEM	
ISCGEM					
2	NaN	NaN	ISCGEM860762	ISCGEM	
ISCGEM					
3	NaN	NaN	ISCGEM860856	ISCGEM	
ISCGEM					
4	NaN	NaN	ISCGEM860890	ISCGEM	
ISCGEM					

	Magnitude	Source	Status
0		ISCGEM	Automatic
1		ISCGEM	Automatic
2		ISCGEM	Automatic
3		ISCGEM	Automatic
4		ISCGEM	Automatic

[5 rows x 21 columns]

df.tail()

	Date	Time	Latitude	Longitude	Type	Depth	\
23407	12/28/2016	08:22:12	38.3917	-118.8941	Earthquake	12.30	
23408	12/28/2016	09:13:47	38.3777	-118.8957	Earthquake	8.80	
23409	12/28/2016	12:38:51	36.9179	140.4262	Earthquake	10.00	
23410	12/29/2016	22:30:19	-9.0283	118.6639	Earthquake	79.00	
23411	12/30/2016	20:08:28	37.3973	141.4103	Earthquake	11.94	

	Depth	Error	Depth	Seismic Stations	Magnitude	Magnitude
Type ... \						
23407		1.2		40.0		5.6
ML ...						
23408		2.0		33.0		5.5
ML ...						
23409		1.8		NaN		5.9
MWW ...						
23410		1.8		NaN		6.3
MWW ...						
23411		2.2		NaN		5.5
MB ...						

	Magnitude	Seismic Stations	Azimuthal Gap	Horizontal Distance
\				
23407		18.0	42.47	0.120
23408		18.0	48.58	0.129
23409		NaN	91.00	0.992
23410		NaN	26.00	3.553
23411		428.0	97.00	0.681

	Horizontal Error	Root Mean Square	ID	Source	Location
Source \					
23407	NaN	0.1898	NN00570710	NN	
NN					
23408	NaN	0.2187	NN00570744	NN	
NN					
23409	4.8	1.5200	US10007NAF	US	

US				
23410	6.0	1.4300	US10007NL0	US
US				
23411	4.5	0.9100	US10007NTD	US
US				

	Magnitude	Source	Status
23407	NN		Reviewed
23408	NN		Reviewed
23409	US		Reviewed
23410	US		Reviewed
23411	US		Reviewed

[5 rows x 21 columns]

`df.shape` # *representing the dimensions of the DataFrame*

(23412, 21)

`df.info()` # *provides a concise summary of the DataFrame.*

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 23412 entries, 0 to 23411

Data columns (total 21 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Date	23412 non-null	object
1	Time	23412 non-null	object
2	Latitude	23412 non-null	float64
3	Longitude	23412 non-null	float64
4	Type	23412 non-null	object
5	Depth	23412 non-null	float64
6	Depth Error	4461 non-null	float64
7	Depth Seismic Stations	7097 non-null	float64
8	Magnitude	23412 non-null	float64
9	Magnitude Type	23409 non-null	object
10	Magnitude Error	327 non-null	float64
11	Magnitude Seismic Stations	2564 non-null	float64
12	Azimuthal Gap	7299 non-null	float64
13	Horizontal Distance	1604 non-null	float64
14	Horizontal Error	1156 non-null	float64
15	Root Mean Square	17352 non-null	float64
16	ID	23412 non-null	object
17	Source	23412 non-null	object
18	Location Source	23412 non-null	object
19	Magnitude Source	23412 non-null	object
20	Status	23412 non-null	object

dtypes: float64(12), object(9)

memory usage: 3.8+ MB

`df.isnull()` #returns a DataFrame of the same shape as the input df, with True indicating missing (NaN) values and False indicating non-missing values in each cell.

	Date	Time	Latitude	Longitude	Type	Depth	Depth	Error	\
0	False	False	False	False	False	False		True	
1	False	False	False	False	False	False		True	
2	False	False	False	False	False	False		True	
3	False	False	False	False	False	False		True	
4	False	False	False	False	False	False		True	
...	
23407	False	False	False	False	False	False		False	
23408	False	False	False	False	False	False		False	
23409	False	False	False	False	False	False		False	
23410	False	False	False	False	False	False		False	
23411	False	False	False	False	False	False		False	

	Depth	Seismic	Stations	Magnitude	Magnitude	Type	...	\
0			True	False		False	...	
1			True	False		False	...	
2			True	False		False	...	
3			True	False		False	...	
4			True	False		False	...	
...			
23407			False	False		False	...	
23408			False	False		False	...	
23409			True	False		False	...	
23410			True	False		False	...	
23411			True	False		False	...	

	Magnitude	Seismic	Stations	Azimuthal	Gap	Horizontal	Distance
\							
0			True		True		True
1			True		True		True
2			True		True		True
3			True		True		True
4			True		True		True
...		
23407			False		False		False
23408			False		False		False
23409			True		False		False
23410			True		False		False

23411	False	False	False
-------	-------	-------	-------

Source \	Horizontal Error	Root Mean Square	ID	Source	Location
0	True	True	False	False	
False					
1	True	True	False	False	
False					
2	True	True	False	False	
False					
3	True	True	False	False	
False					
4	True	True	False	False	
False					
...
...					
23407	True	False	False	False	
False					
23408	True	False	False	False	
False					
23409	False	False	False	False	
False					
23410	False	False	False	False	
False					
23411	False	False	False	False	
False					

	Magnitude	Source	Status
0		False	False
1		False	False
2		False	False
3		False	False
4		False	False
...	
23407		False	False
23408		False	False
23409		False	False
23410		False	False
23411		False	False

[23412 rows x 21 columns]

`df.isnull().sum()` #returns the count of missing (NaN) values in each column of the DataFrame.

Date	0
Time	0
Latitude	0

```

Longitude          0
Type               0
Depth             0
Depth Error       18951
Depth Seismic Stations 16315
Magnitude          0
Magnitude Type     3
Magnitude Error   23085
Magnitude Seismic Stations 20848
Azimuthal Gap     16113
Horizontal Distance 21808
Horizontal Error   22256
Root Mean Square   6060
ID                0
Source            0
Location Source    0
Magnitude Source   0
Status            0
dtype: int64

```

```

df.dropna(how="all")#removes rows from the DataFrame df where all
values in a row are missing (NaN).

```

	Date	Time	Latitude	Longitude	Type	
Depth \						
0	01/02/1965	13:44:18	19.2460	145.6160	Earthquake	131.60
1	01/04/1965	11:29:49	1.8630	127.3520	Earthquake	80.00
2	01/05/1965	18:05:58	-20.5790	-173.9720	Earthquake	20.00
3	01/08/1965	18:49:43	-59.0760	-23.5570	Earthquake	15.00
4	01/09/1965	13:32:50	11.9380	126.4270	Earthquake	15.00
...
23407	12/28/2016	08:22:12	38.3917	-118.8941	Earthquake	12.30
23408	12/28/2016	09:13:47	38.3777	-118.8957	Earthquake	8.80
23409	12/28/2016	12:38:51	36.9179	140.4262	Earthquake	10.00
23410	12/29/2016	22:30:19	-9.0283	118.6639	Earthquake	79.00
23411	12/30/2016	20:08:28	37.3973	141.4103	Earthquake	11.94
	Depth Error	Depth Seismic Stations	Magnitude	Magnitude		
Type ... \						
0	NaN	NaN	6.0			

MW	...				
1		NaN		NaN	5.8
MW	...				
2		NaN		NaN	6.2
MW	...				
3		NaN		NaN	5.8
MW	...				
4		NaN		NaN	5.8
MW	...				
...	
...					
23407		1.2		40.0	5.6
ML	...				
23408		2.0		33.0	5.5
ML	...				
23409		1.8		NaN	5.9
MWW	...				
23410		1.8		NaN	6.3
MWW	...				
23411		2.2		NaN	5.5
MB	...				
	Magnitude	Seismic Stations	Azimuthal Gap	Horizontal Distance	
\					
0		NaN	NaN		NaN
1		NaN	NaN		NaN
2		NaN	NaN		NaN
3		NaN	NaN		NaN
4		NaN	NaN		NaN
...	
23407		18.0	42.47		0.120
23408		18.0	48.58		0.129
23409		NaN	91.00		0.992
23410		NaN	26.00		3.553
23411		428.0	97.00		0.681
	Horizontal Error	Root Mean Square	ID	Source	\
0	NaN	NaN	ISCGEM860706	ISCGEM	
1	NaN	NaN	ISCGEM860737	ISCGEM	

2	NaN	NaN	ISCGEM860762	ISCGEM
3	NaN	NaN	ISCGEM860856	ISCGEM
4	NaN	NaN	ISCGEM860890	ISCGEM
...
23407	NaN	0.1898	NN00570710	NN
23408	NaN	0.2187	NN00570744	NN
23409	4.8	1.5200	US10007NAF	US
23410	6.0	1.4300	US10007NL0	US
23411	4.5	0.9100	US10007NTD	US

	Location	Source	Magnitude	Source	Status
0		ISCGEM		ISCGEM	Automatic
1		ISCGEM		ISCGEM	Automatic
2		ISCGEM		ISCGEM	Automatic
3		ISCGEM		ISCGEM	Automatic
4		ISCGEM		ISCGEM	Automatic
...	
23407		NN		NN	Reviewed
23408		NN		NN	Reviewed
23409		US		US	Reviewed
23410		US		US	Reviewed
23411		US		US	Reviewed

[23412 rows x 21 columns]

```
# Create a dictionary to specify filling methods for each column
fill_methods = {
    'Depth Error': df['Depth Error'].mean(),
    'Depth Seismic Stations': df['Depth Seismic Stations'].mean(),
    'Magnitude Seismic Stations': df['Magnitude Seismic
Stations'].mean(),
    'Azimuthal Gap':df['Azimuthal Gap'].median(),
    'Horizontal Distance':df['Horizontal Distance'].mean(),
    'Horizontal Error':df['Horizontal Error'].mean(),
    'Root Mean Square':df['Root Mean Square'].mean(),
    'Magnitude Error': df['Magnitude Error'].mean(),
    'Magnitude': df['Magnitude'].mean(),
    'Magnitude Type': 'Unknown',
    'Date': df['Date'].mode()[0],
    'Latitude': df['Latitude'].mode()[0],
    'Longitude': df['Longitude'].mode()[0],
    'Type': df['Type'].mode()[0],
    'Depth': df['Depth'].mode()[0],
    'ID': df['ID'].mode()[0],
    'Source': df['Source'].mode()[0],
    'Location Source': df['Location Source'].mode()[0],
    'Magnitude Source': df['Magnitude Source'].mode()[0],
    'Status': df['Status'].mode()[0],
}
```



```
# Apply the filling methods using fillna()
df.fillna(fill_methods, inplace=True)

df.isnull().sum()# returns the count of missing (NaN) values in each
column of the DataFrame.
```

```
Date      0
Time      0
Latitude  0
Longitude 0
Type      0
Depth     0
Depth Error      0
Depth Seismic Stations  0
Magnitude      0
Magnitude Type  0
Magnitude Error  0
Magnitude Seismic Stations  0
Azimuthal Gap  0
Horizontal Distance  0
Horizontal Error  0
Root Mean Square  0
ID            0
Source       0
Location Source      0
Magnitude Source     0
Status        0
dtype: int64
```

```
df.head()
```

	Date	Time	Latitude	Longitude	Type	Depth	Depth Error
0	01/02/1965	13:44:18	19.246	145.616	Earthquake	131.6	4.993115
1	01/04/1965	11:29:49	1.863	127.352	Earthquake	80.0	4.993115
2	01/05/1965	18:05:58	-20.579	-173.972	Earthquake	20.0	4.993115
3	01/08/1965	18:49:43	-59.076	-23.557	Earthquake	15.0	4.993115
4	01/09/1965	13:32:50	11.938	126.427	Earthquake	15.0	4.993115

	Depth	Seismic Stations	Magnitude	Magnitude Type	...	\
0		275.364098	6.0	MW	...	
1		275.364098	5.8	MW	...	
2		275.364098	6.2	MW	...	
3		275.364098	5.8	MW	...	
4		275.364098	5.8	MW	...	

	Magnitude	Seismic Stations	Azimuthal Gap	Horizontal Distance \
0		48.944618	36.0	3.99266
1		48.944618	36.0	3.99266
2		48.944618	36.0	3.99266
3		48.944618	36.0	3.99266
4		48.944618	36.0	3.99266

	Horizontal Error	Root Mean Square	ID	Source Location
Source \				
0	7.662759	1.022784	ISCGEM860706	ISCGEM
ISCGEM				
1	7.662759	1.022784	ISCGEM860737	ISCGEM
ISCGEM				
2	7.662759	1.022784	ISCGEM860762	ISCGEM
ISCGEM				
3	7.662759	1.022784	ISCGEM860856	ISCGEM
ISCGEM				
4	7.662759	1.022784	ISCGEM860890	ISCGEM
ISCGEM				

	Magnitude	Source	Status
0		ISCGEM	Automatic
1		ISCGEM	Automatic
2		ISCGEM	Automatic
3		ISCGEM	Automatic
4		ISCGEM	Automatic

[5 rows x 21 columns]

`df.describe()` *#provides summary statistics (count, mean, std, min, 25%, 50%, 75%, and max) for each numerical column in the DataFrame*

	Latitude	Longitude	Depth	Depth Error \
count	23412.000000	23412.000000	23412.000000	23412.000000
mean	1.679033	39.639961	70.767911	4.993115
std	30.113183	125.511959	122.651898	2.127886
min	-77.080000	-179.997000	-1.100000	0.000000
25%	-18.653000	-76.349750	14.522500	4.993115
50%	-3.568500	103.982000	33.000000	4.993115
75%	26.190750	145.026250	54.000000	4.993115
max	86.005000	179.998000	700.000000	91.295000

	Depth	Seismic Stations	Magnitude	Magnitude Error \
count		23412.000000	23412.000000	23412.000000
mean		275.364098	5.882531	0.071820
std		89.267086	0.423066	0.006073
min		0.000000	5.500000	0.000000
25%		275.364098	5.600000	0.071820
50%		275.364098	5.700000	0.071820

75%	275.364098	6.000000	0.071820				
max	934.000000	9.100000	0.410000				
	Magnitude	Seismic Stations	Azimuthal Gap				
\	Horizontal Distance						
count	23412.000000	23412.000000	23412.000000				
mean	48.944618	38.545089	3.992660				
std	20.826318	18.339697	1.407077				
min	0.000000	0.000000	0.004505				
25%	48.944618	36.000000	3.992660				
50%	48.944618	36.000000	3.992660				
75%	48.944618	36.000000	3.992660				
max	821.000000	360.000000	37.874000				
	Horizontal Error	Root Mean Square					
count	23412.000000	23412.000000					
mean	7.662759	1.022784					
std	2.316764	0.162319					
min	0.085000	0.000000					
25%	7.662759	0.940000					
50%	7.662759	1.022784					
75%	7.662759	1.100000					
max	99.000000	3.440000					
# Replace the original dataset file							
df.to_csv('database.csv', index=False)							
df.head()							
	Date	Time	Latitude	Longitude	Type	Depth	Depth
Error \							
0	01/02/1965	13:44:18	19.246	145.616	Earthquake	131.6	
4.993115							
1	01/04/1965	11:29:49	1.863	127.352	Earthquake	80.0	
4.993115							
2	01/05/1965	18:05:58	-20.579	-173.972	Earthquake	20.0	
4.993115							
3	01/08/1965	18:49:43	-59.076	-23.557	Earthquake	15.0	
4.993115							
4	01/09/1965	13:32:50	11.938	126.427	Earthquake	15.0	
4.993115							
	Depth	Seismic Stations	Magnitude	Magnitude	Type	...	\

0	275.364098	6.0	MW	...
1	275.364098	5.8	MW	...
2	275.364098	6.2	MW	...
3	275.364098	5.8	MW	...
4	275.364098	5.8	MW	...

	Magnitude	Seismic Stations	Azimuthal Gap	Horizontal Distance \
0		48.944618	36.0	3.99266
1		48.944618	36.0	3.99266
2		48.944618	36.0	3.99266
3		48.944618	36.0	3.99266
4		48.944618	36.0	3.99266

	Horizontal Error	Root Mean Square	ID	Source Location
Source \				
0	7.662759	1.022784	ISCGEM860706	ISCGEM
ISCGEM				
1	7.662759	1.022784	ISCGEM860737	ISCGEM
ISCGEM				
2	7.662759	1.022784	ISCGEM860762	ISCGEM
ISCGEM				
3	7.662759	1.022784	ISCGEM860856	ISCGEM
ISCGEM				
4	7.662759	1.022784	ISCGEM860890	ISCGEM
ISCGEM				

	Magnitude	Source	Status
0		ISCGEM	Automatic
1		ISCGEM	Automatic
2		ISCGEM	Automatic
3		ISCGEM	Automatic
4		ISCGEM	Automatic

[5 rows x 21 columns]

--Feature engineering--

Create a new feature "MagnitudeSquared" by squaring the "Magnitude" column

`df['MagnitudeSquared'] = df['Magnitude'] ** 2`

`df.head()`

	Date	Time	Latitude	Longitude	Type	Depth	Depth Error \
0	01/02/1965	13:44:18	19.246	145.616	Earthquake	131.6	4.993115
1	01/04/1965	11:29:49	1.863	127.352	Earthquake	80.0	4.993115
2	01/05/1965	18:05:58	-20.579	-173.972	Earthquake	20.0	4.993115

```

3 01/08/1965 18:49:43 -59.076 -23.557 Earthquake 15.0
4.993115
4 01/09/1965 13:32:50 11.938 126.427 Earthquake 15.0
4.993115

```

```

      Depth Seismic Stations Magnitude Magnitude Type ... Azimuthal
Gap \
0      275.364098      6.0      MW ...
36.0
1      275.364098      5.8      MW ...
36.0
2      275.364098      6.2      MW ...
36.0
3      275.364098      5.8      MW ...
36.0
4      275.364098      5.8      MW ...
36.0

```

```

      Horizontal Distance Horizontal Error Root Mean Square
ID \
0      3.99266      7.662759      1.022784
ISCGEM860706
1      3.99266      7.662759      1.022784
ISCGEM860737
2      3.99266      7.662759      1.022784
ISCGEM860762
3      3.99266      7.662759      1.022784
ISCGEM860856
4      3.99266      7.662759      1.022784
ISCGEM860890

```

```

      Source Location Source Magnitude Source Status MagnitudeSquared
0  ISCGEM      ISCGEM      ISCGEM Automatic      36.00
1  ISCGEM      ISCGEM      ISCGEM Automatic      33.64
2  ISCGEM      ISCGEM      ISCGEM Automatic      38.44
3  ISCGEM      ISCGEM      ISCGEM Automatic      33.64
4  ISCGEM      ISCGEM      ISCGEM Automatic      33.64

```

```
[5 rows x 22 columns]
```

```

# Save the updated DataFrame to a new or the same file
df.to_csv('updated_database.csv', index=False)

```

```
#----model development ---
```

```

# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Load the updated earthquake dataset with the added
"MagnitudeSquared" feature
df = pd.read_csv('updated_database.csv')

# Define your feature matrix X (including "Magnitude" and
"MagnitudeSquared") and target variable y
X = df[['Magnitude', 'MagnitudeSquared']]
y = df['Magnitude'] # Replace 'YourTargetVariable' with the actual
target variable name

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Instantiate the linear regression model
model = LinearRegression()

# Train the model on the training data
model.fit(X_train, y_train)

LinearRegression()

# Make predictions on the testing data
y_pred = model.predict(X_test)

# Calculate the mean squared error to evaluate the model's performance
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

Mean Squared Error: 2.4762449939673018e-31

#---model evaluation---

from sklearn.metrics import mean_absolute_error, r2_score

# Calculate and print the mean absolute error (MAE)
mae = mean_absolute_error(y_test, y_pred)
print(f"Mean Absolute Error: {mae}")

Mean Absolute Error: 2.788004008027299e-16

# Calculate and print the R-squared (R2) score
r2 = r2_score(y_test, y_pred)
print(f"R-squared (R2) Score: {r2}")

R-squared (R2) Score: 1.0

```

#--Visualization--

```
pip install folium
```

```
Requirement already satisfied: folium in
/usr/local/lib/python3.10/dist-packages (0.14.0)
Requirement already satisfied: branca>=0.6.0 in
/usr/local/lib/python3.10/dist-packages (from folium) (0.6.0)
Requirement already satisfied: jinja2>=2.9 in
/usr/local/lib/python3.10/dist-packages (from folium) (3.1.2)
Requirement already satisfied: numpy in
/usr/local/lib/python3.10/dist-packages (from folium) (1.23.5)
Requirement already satisfied: requests in
/usr/local/lib/python3.10/dist-packages (from folium) (2.31.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2>=2.9->folium)
(2.1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(3.3.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests->folium) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests->folium)
(2023.7.22)
```

```
import folium
```

```
import pandas as pd
```

```
# Replace 'your_data.csv' with the path to your data file
```

```
df = pd.read_csv('updated_database.csv')
```

```
# Calculate the center of the map based on latitude and longitude
```

```
center_lat = df['Latitude'].mean()
```

```
center_lon = df['Longitude'].mean()
```

```
# Create a map centered at the calculated location
```

```
m = folium.Map(location=[center_lat, center_lon], zoom_start=5)
```

```
# Iterate through your data and add markers for each earthquake
```

```
for index, row in df.iterrows():
```

```
    folium.CircleMarker(
```

```
        location=[row['Latitude'], row['Longitude']],
```

```
        radius=5,
```

```
        color='blue',
```

```
        fill=True,
```

```
        fill_color='blue',
```

```
        fill_opacity=0.6,
```

```
        popup=f"Magnitude: {row['Magnitude']}, Date: {row['Date']}"  
    ).add_to(m)
```

```
# Display the map  
m
```

```
<folium.folium.Map at 0x7f75e63b6bf0>
```