

Earthquake Severity Prediction System

Enter the parameters below to predict the magnitude class:

Magnitude (Richter Scale)

5.5



Depth (km)

50



0  10

0  700

Soil Type (Encoded)

1




Region Cluster (Encoded)

0



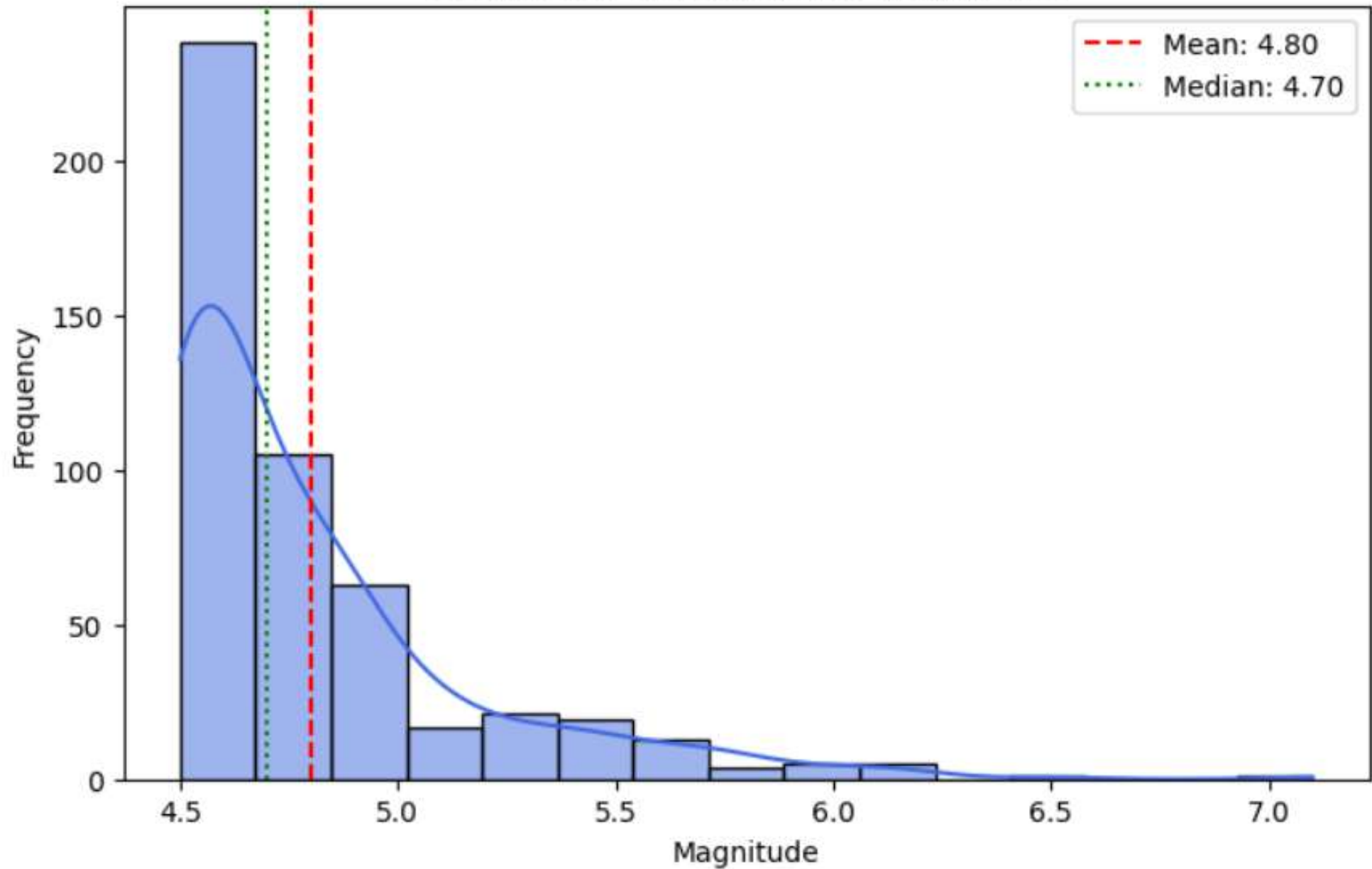
 Predict Severity

Prediction Result

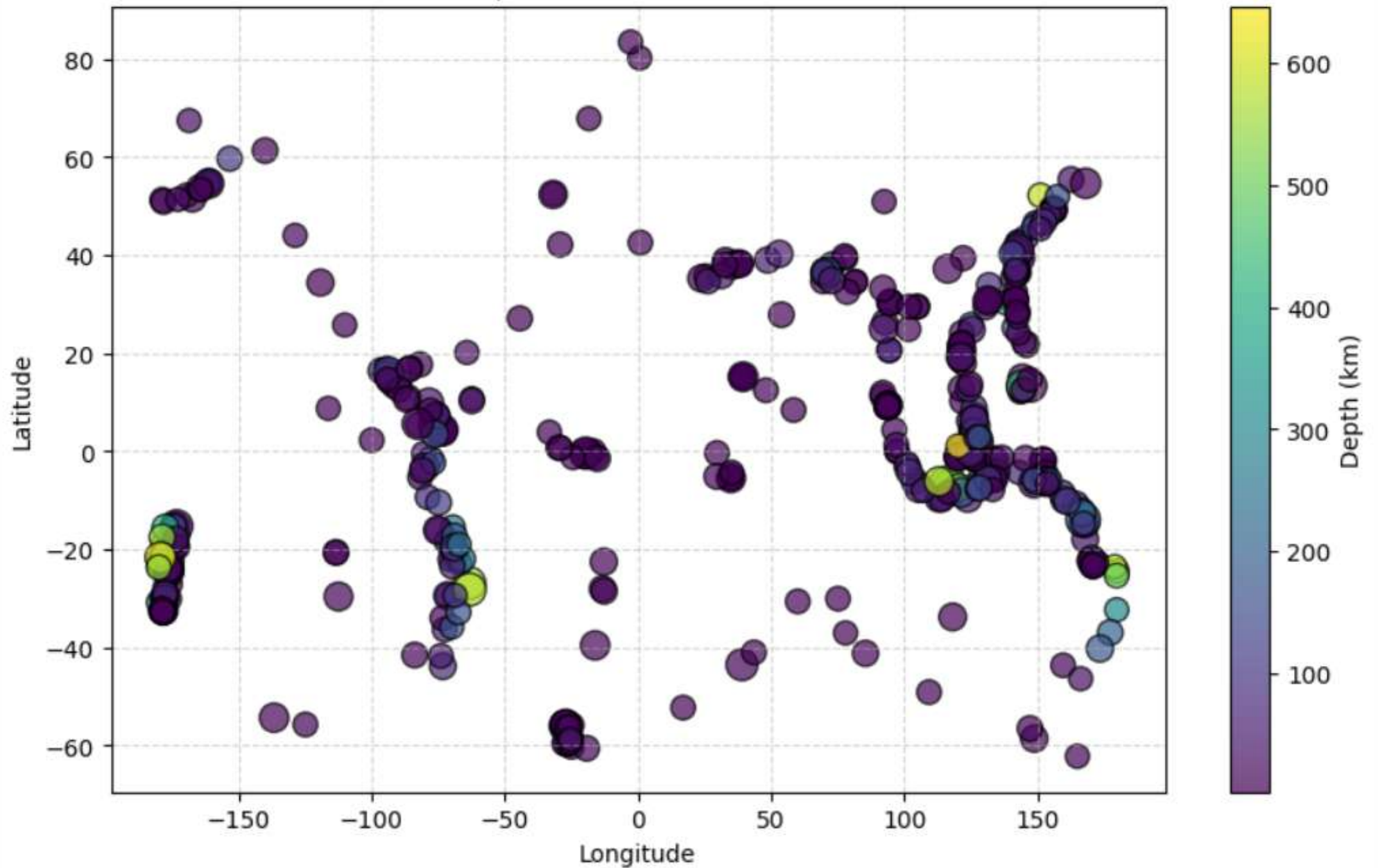
 Predicted Earthquake Severity: **Moderate**

Based on trained machine learning model (Logistic Regression / Decision Tree / XGBoost)

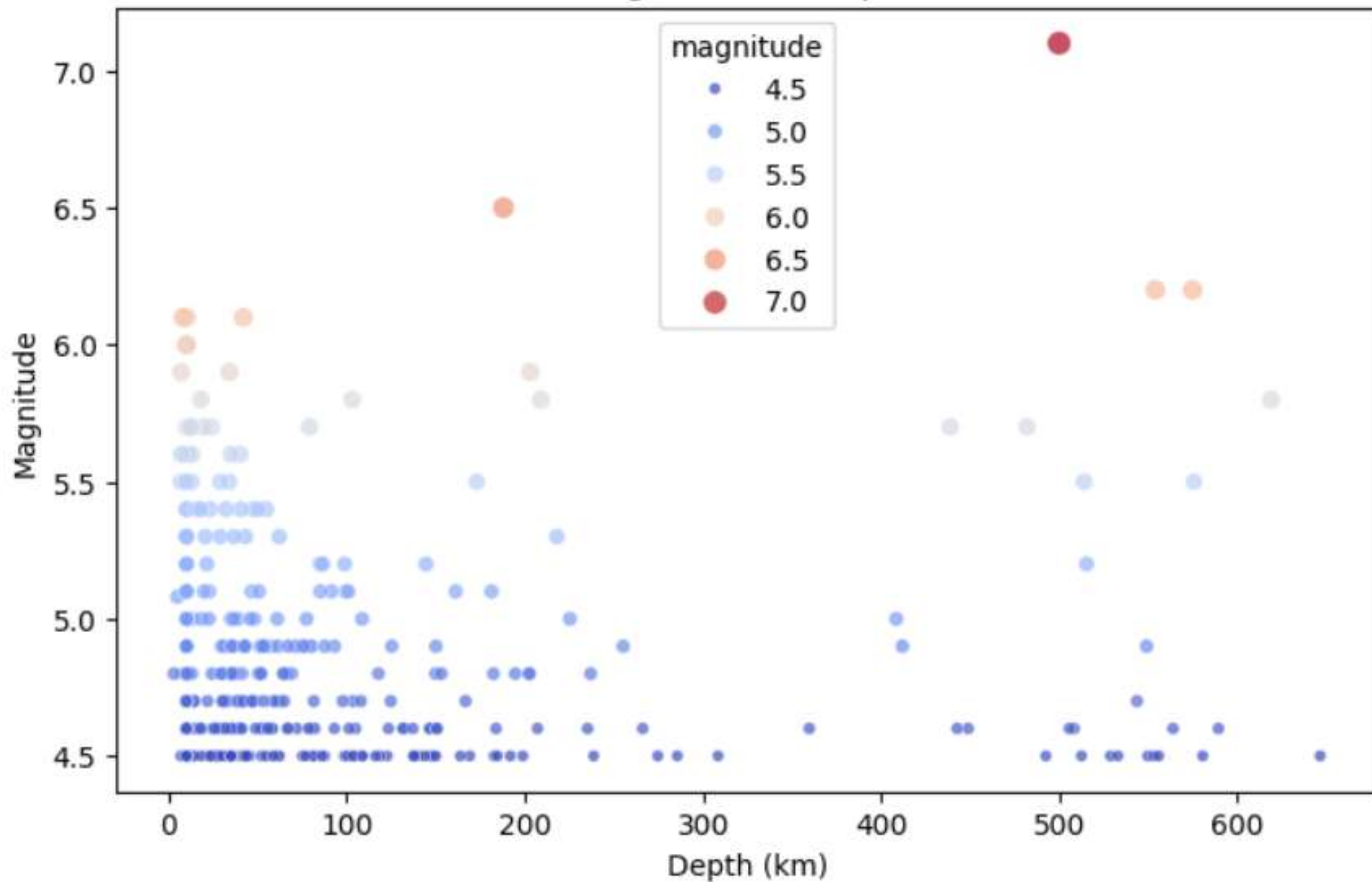
Distribution of Earthquake Magnitudes



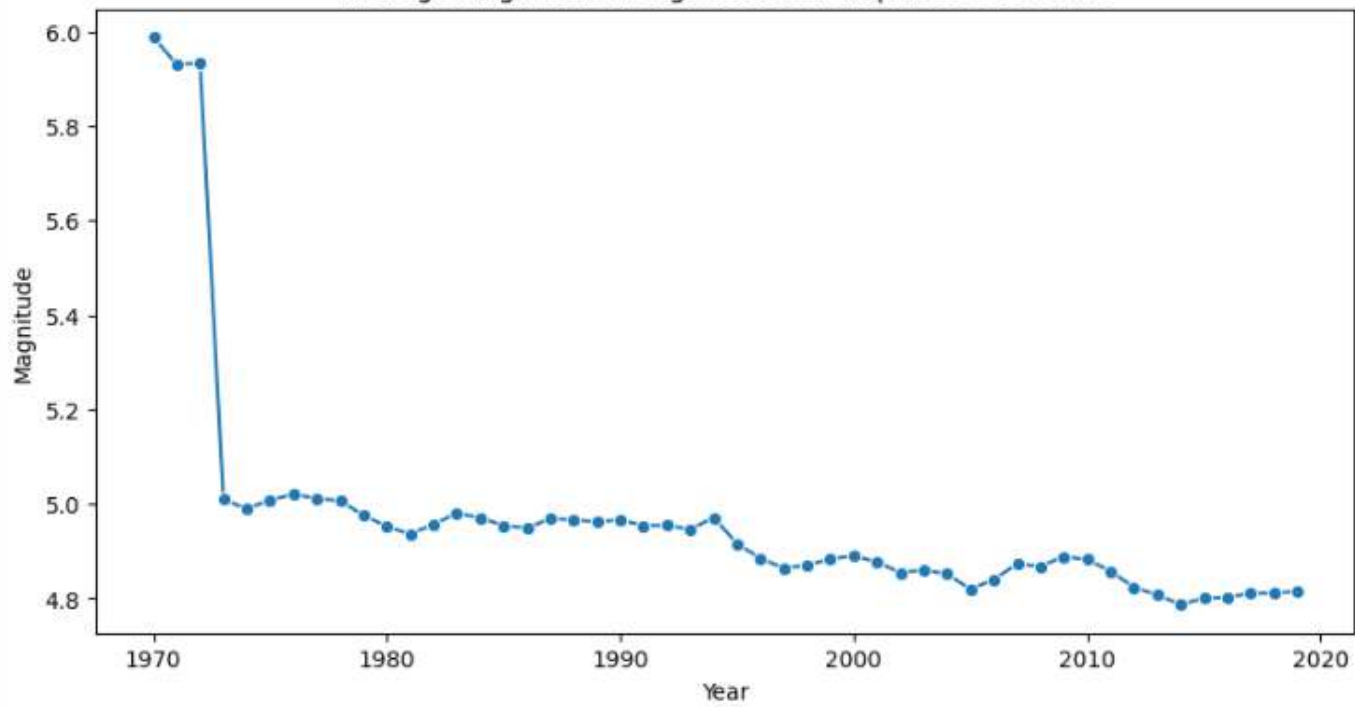
Earthquake Locations (API Data)



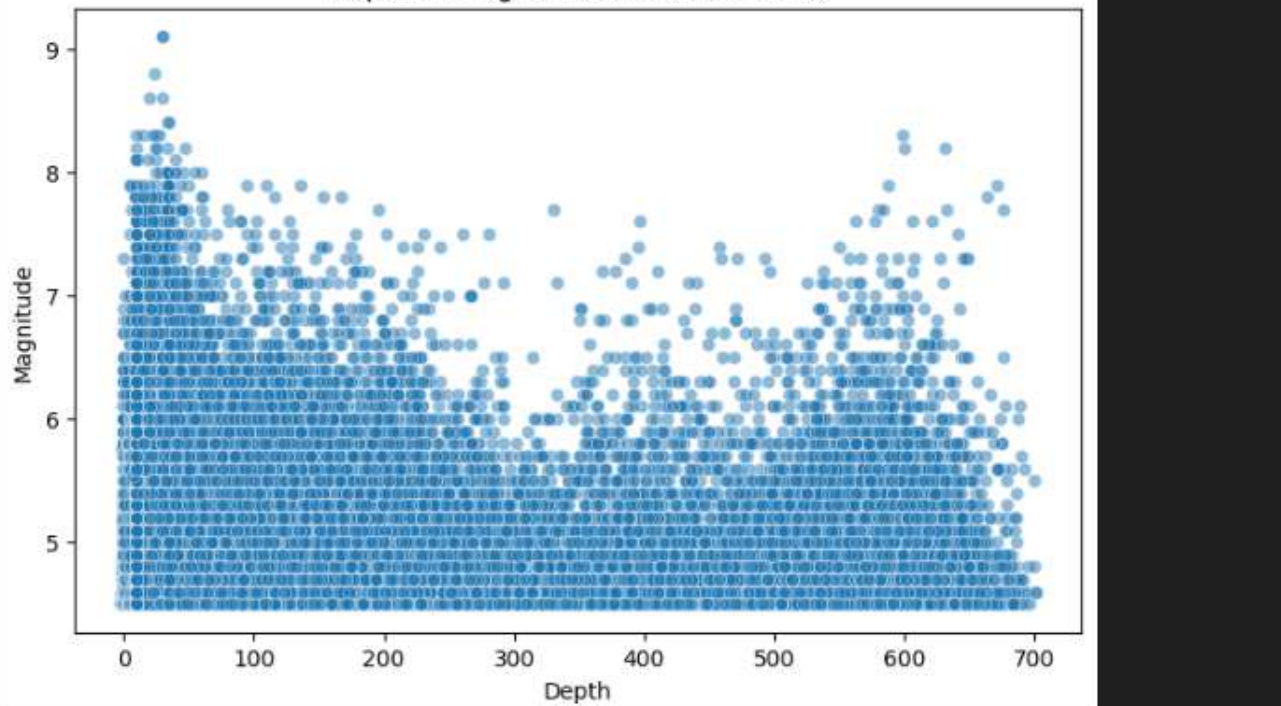
Magnitude vs Depth



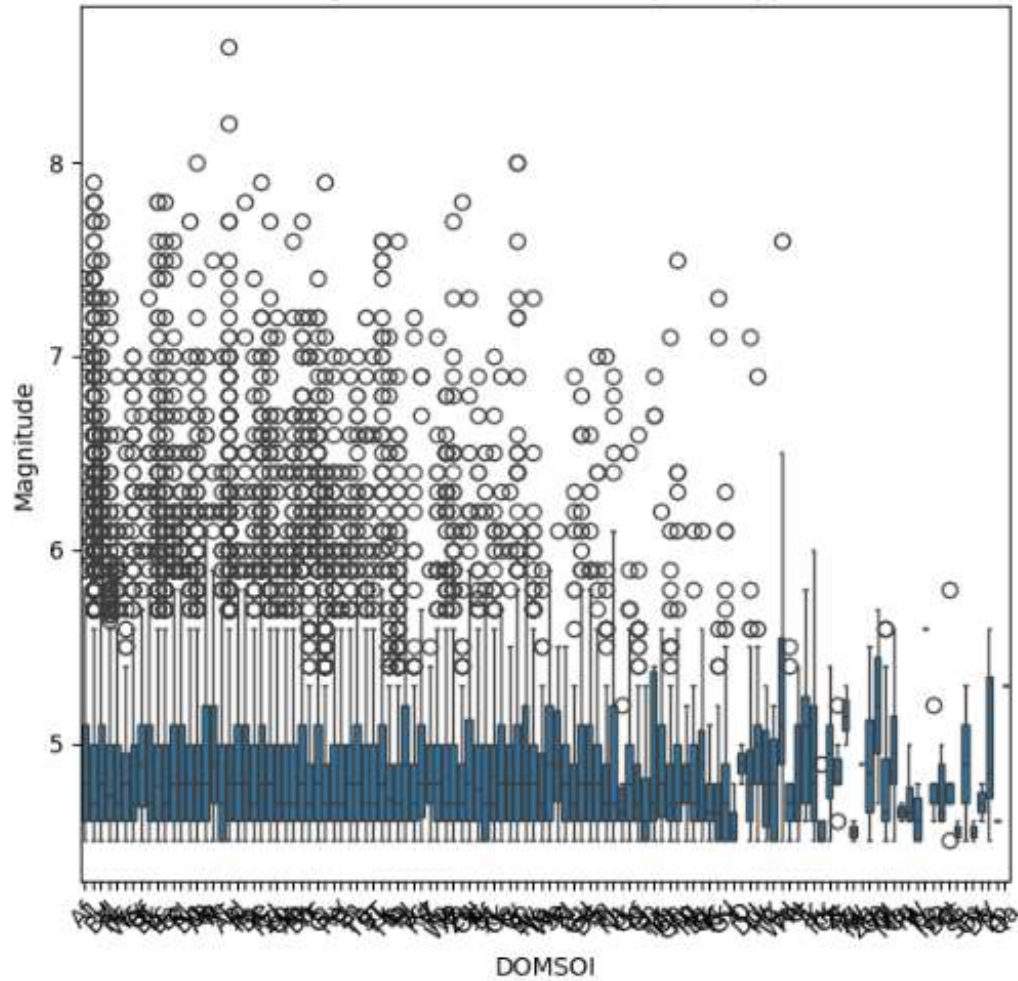
Average Magnitude of Significant Earthquakes Over Time



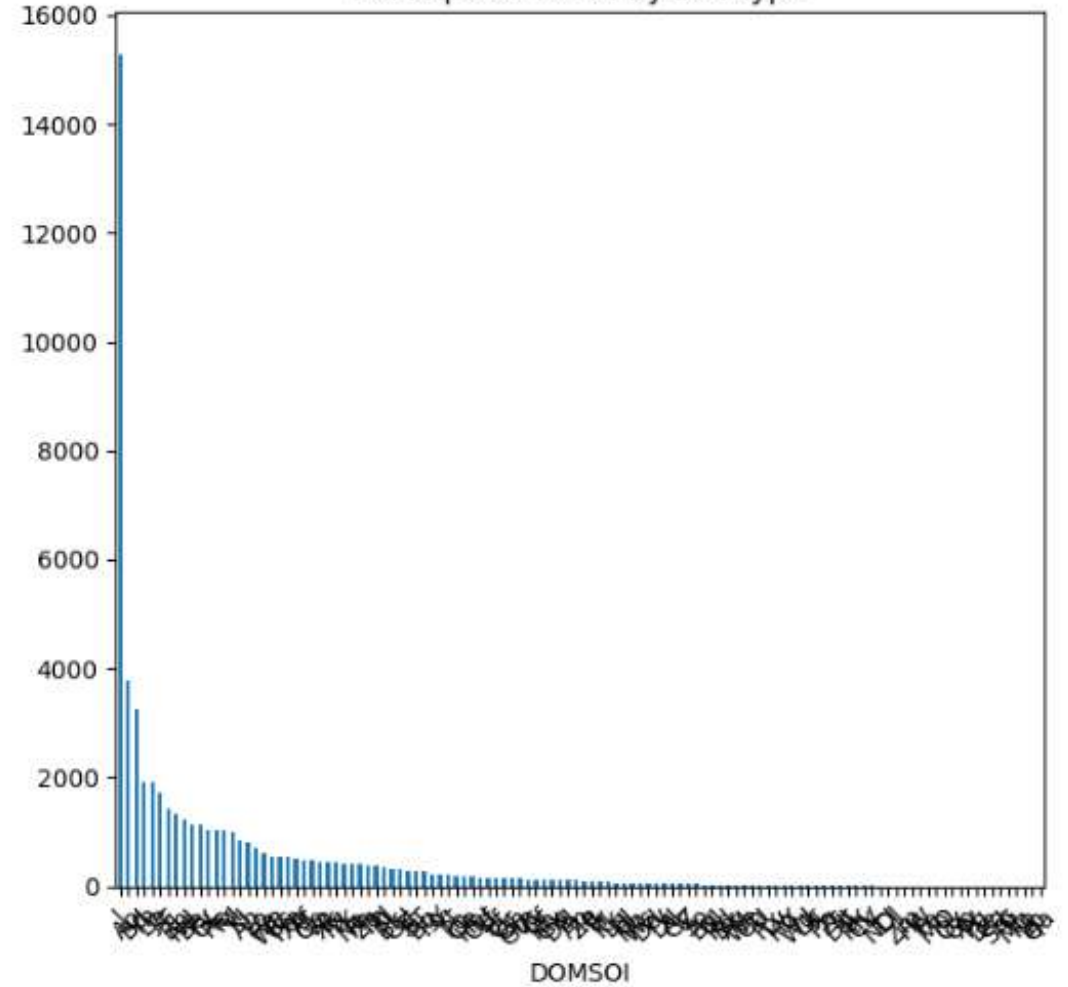
Depth vs Magnitude (Historical Data)



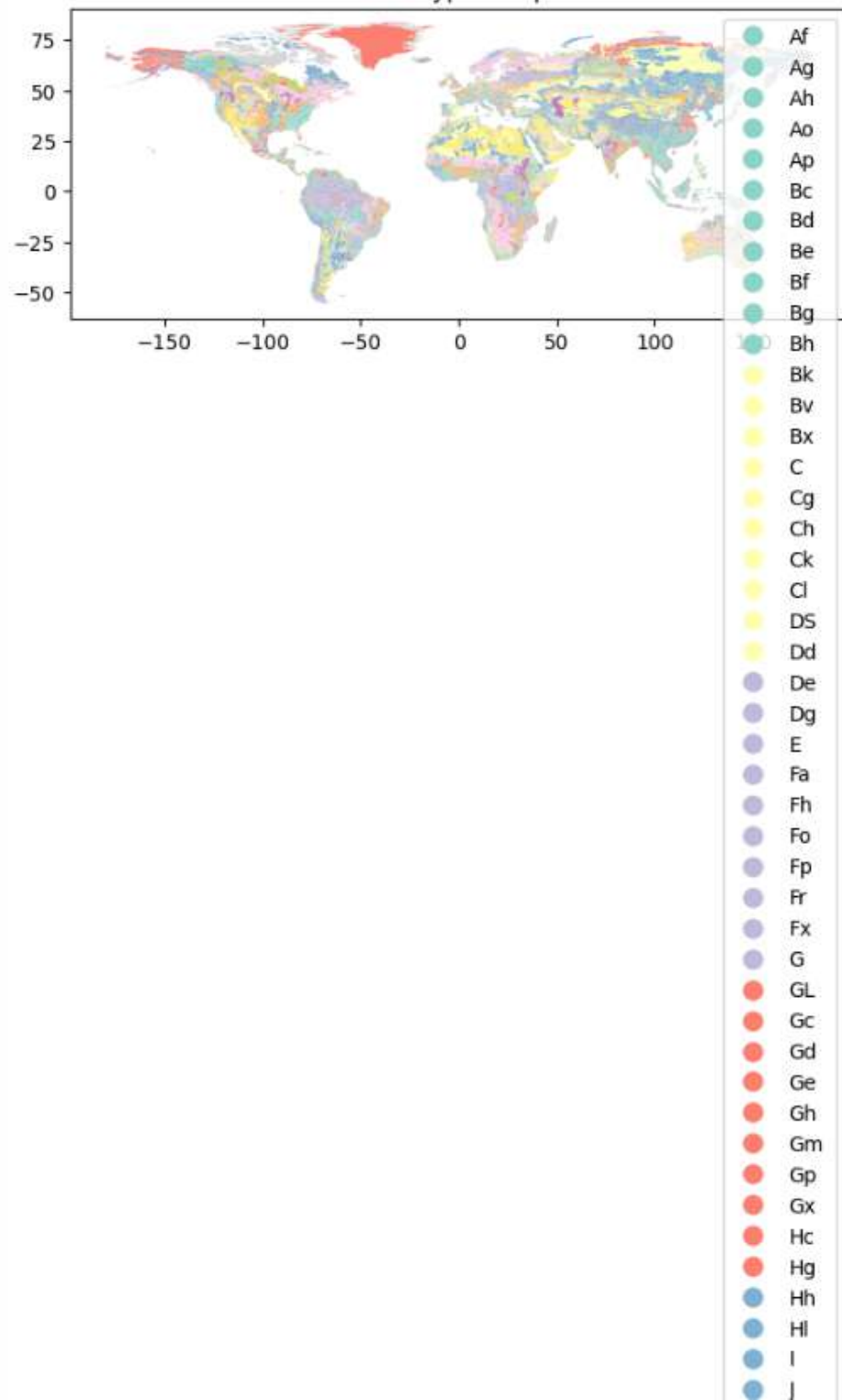
Magnitude Distribution by Soil Type



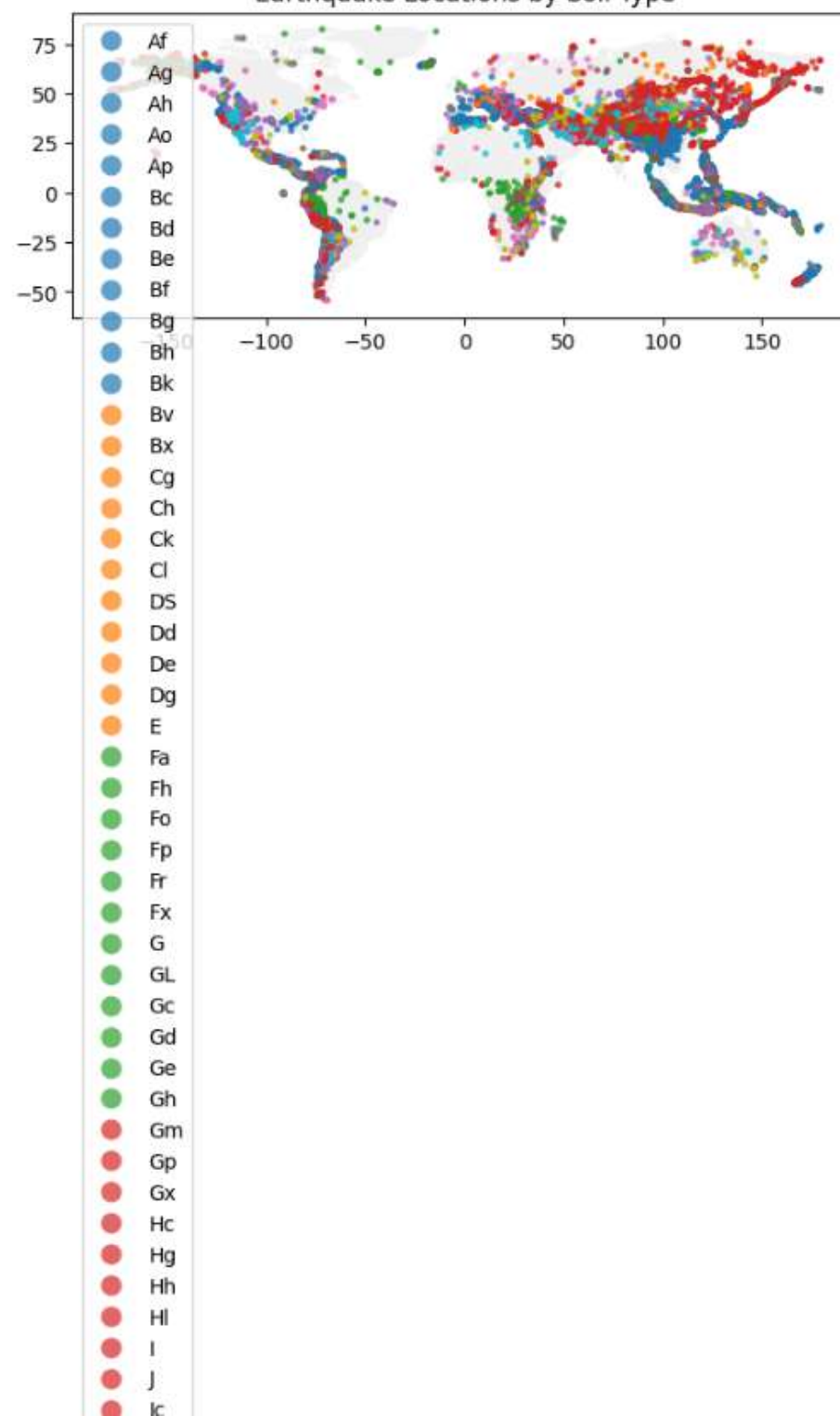
Earthquake Count by Soil Type



Soil Types Map



Earthquake Locations by Soil Type

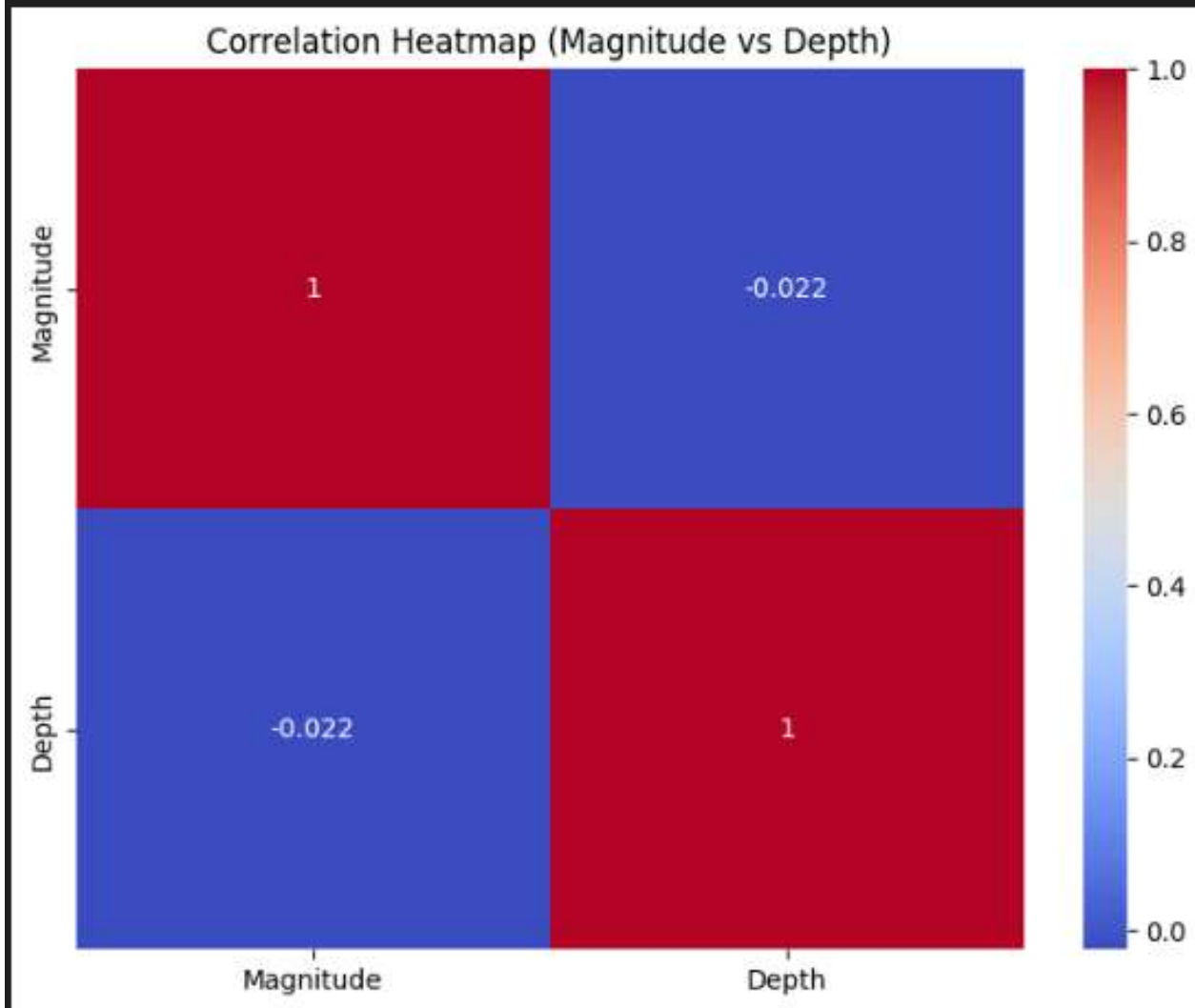


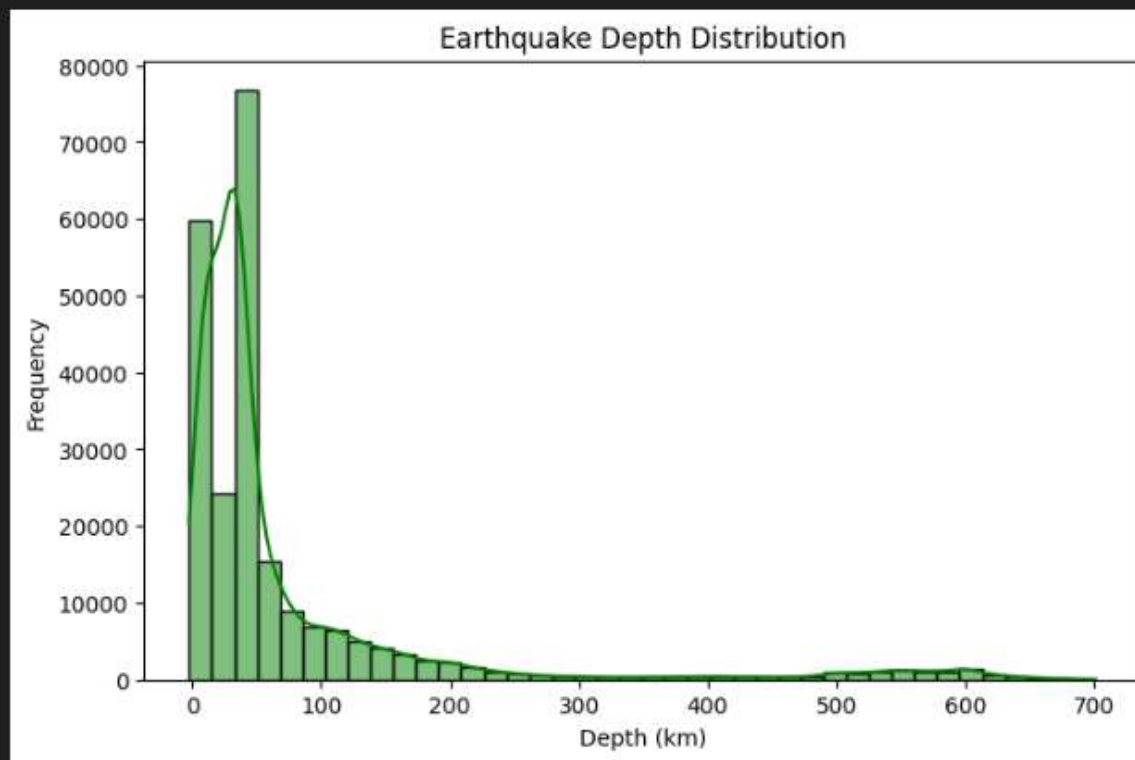
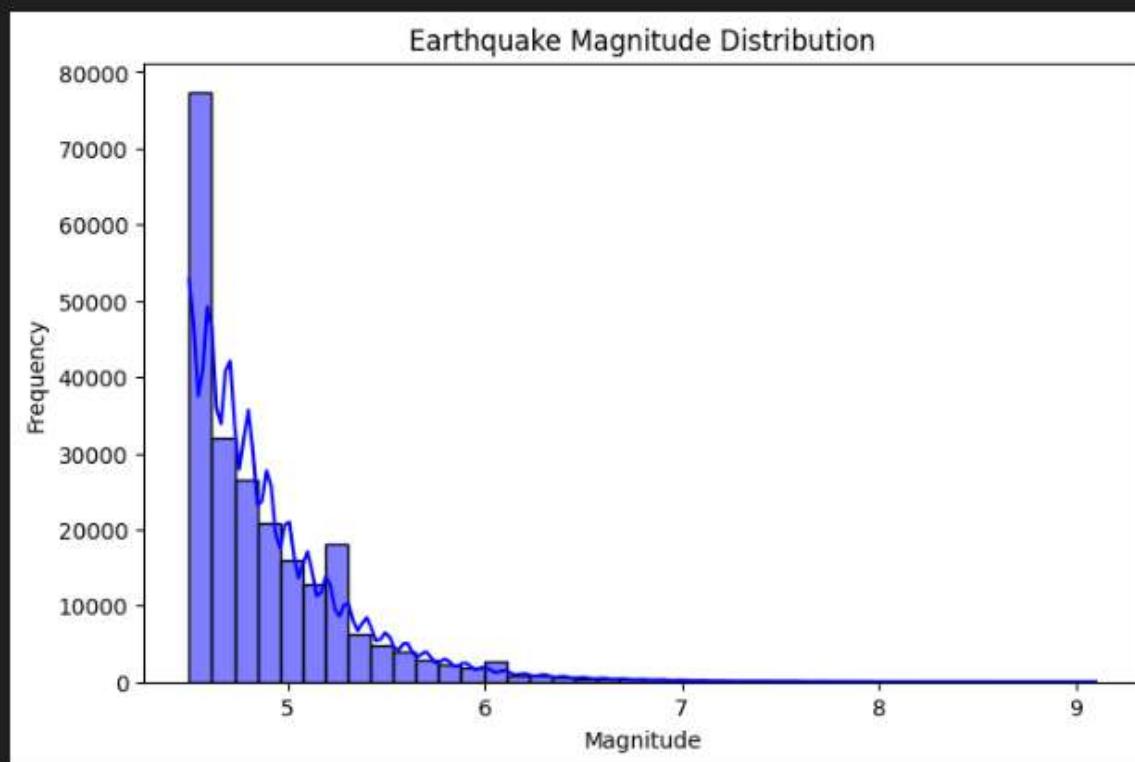
Statistical summary by soil type:

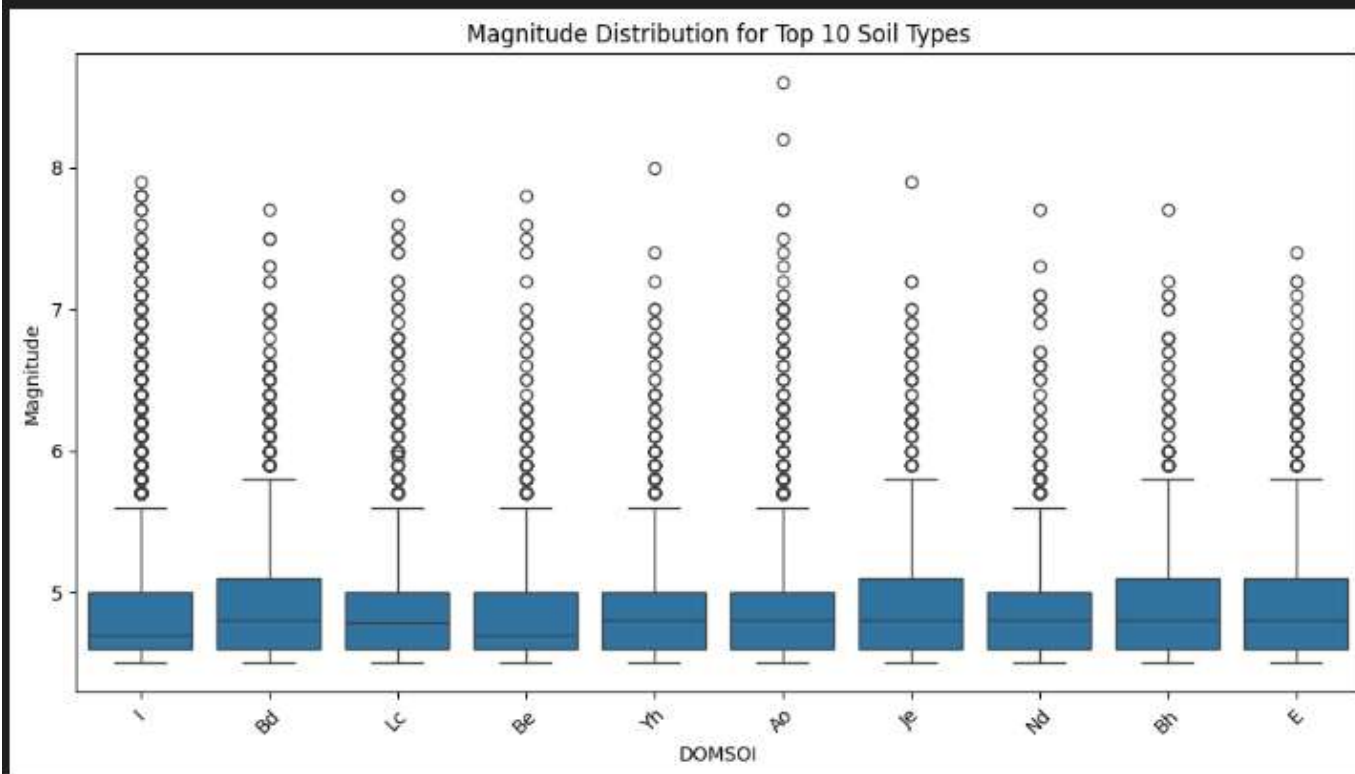
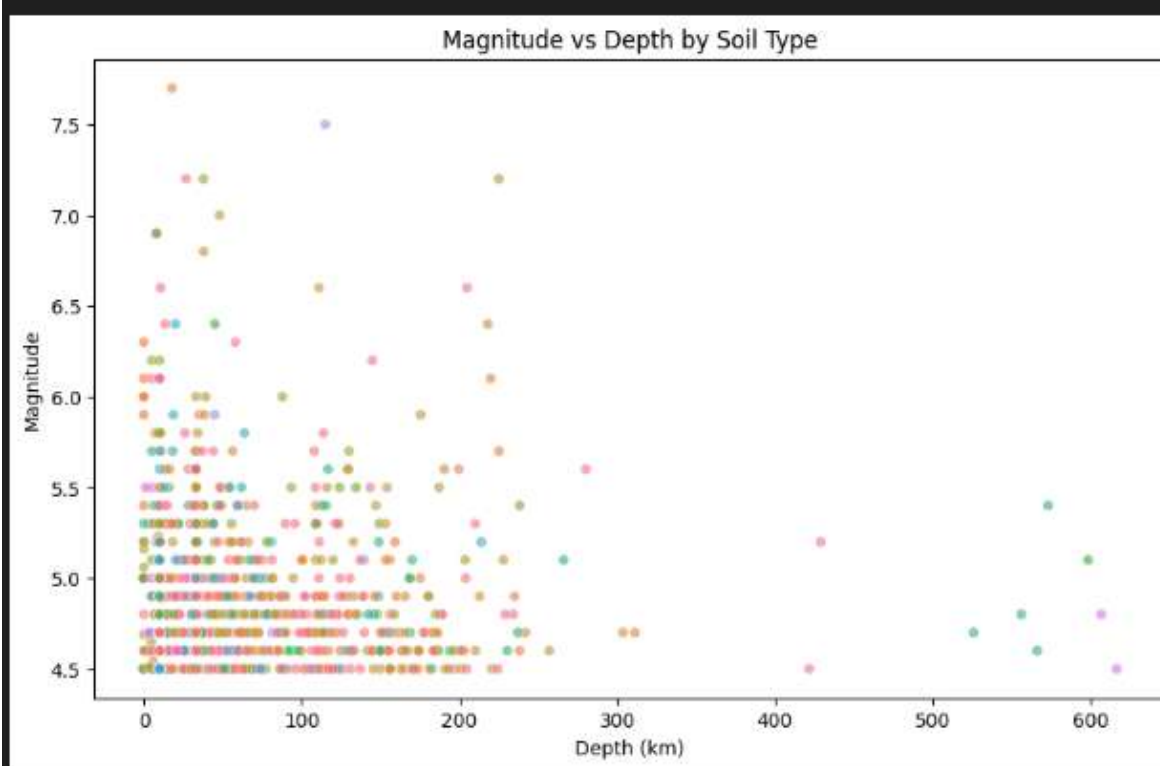
	count	mean	std	min	max
DOMSOI					
Af	488	4.898566	0.461426	4.5	7.4
Ag	22	4.813636	0.225294	4.5	5.2
Ah	800	4.875212	0.411510	4.5	7.6
Ao	3789	4.877590	0.418964	4.5	8.6
Ap	112	5.006250	0.538145	4.5	7.0
...
Z	42	4.888095	0.293176	4.5	5.9
Zg	87	4.849425	0.439046	4.5	7.8
Zm	4	4.925000	0.434933	4.5	5.5
Zo	387	4.846253	0.384418	4.5	7.2
Zt	8	4.862500	0.176777	4.6	5.2

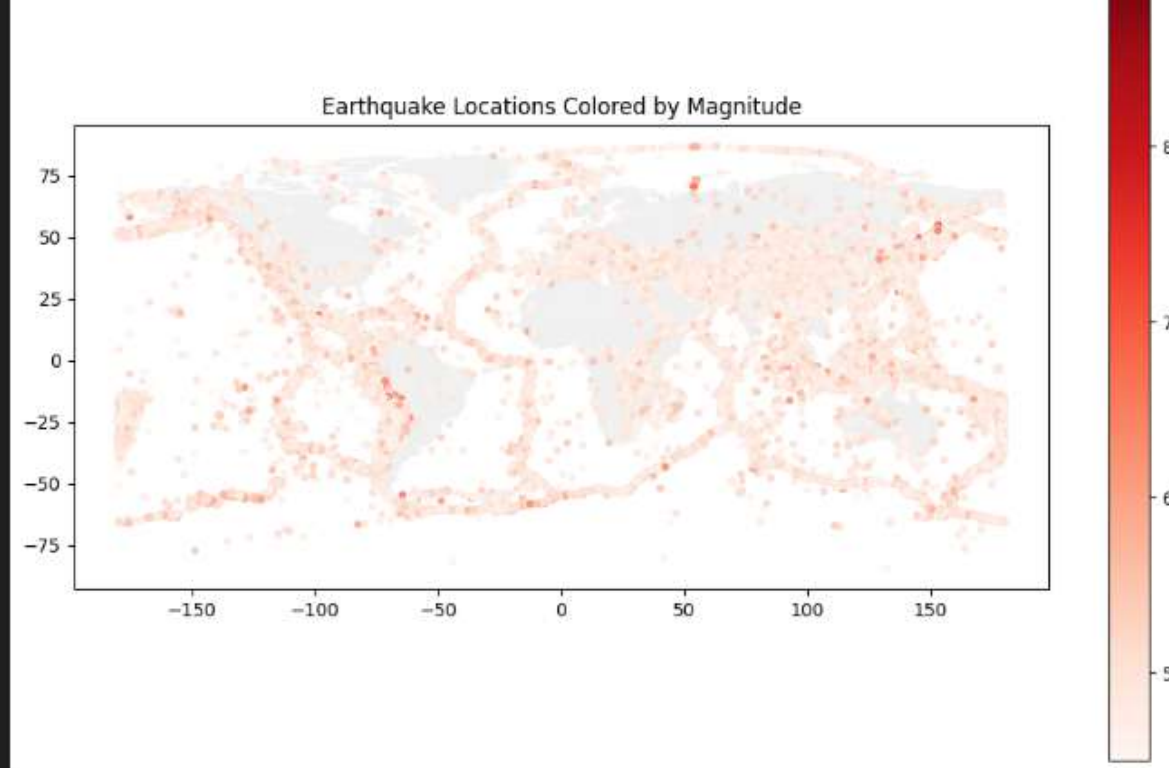
[116 rows x 5 columns]


```
Missing values per column:
Date          0
Latitude      0
Longitude     0
Depth         0
Magnitude     0
Location      0
Year          0
geometry      0
index_right   178912
DOMSOI        178912
dtype: int64
```

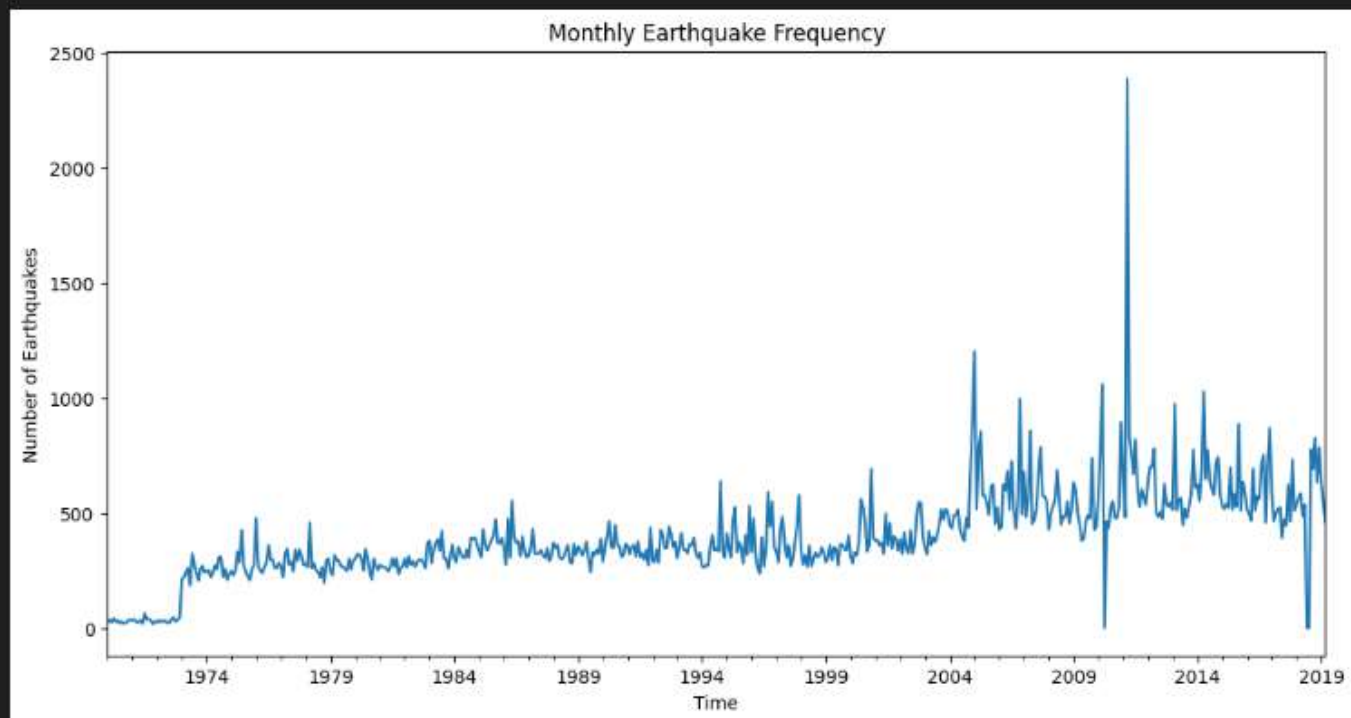


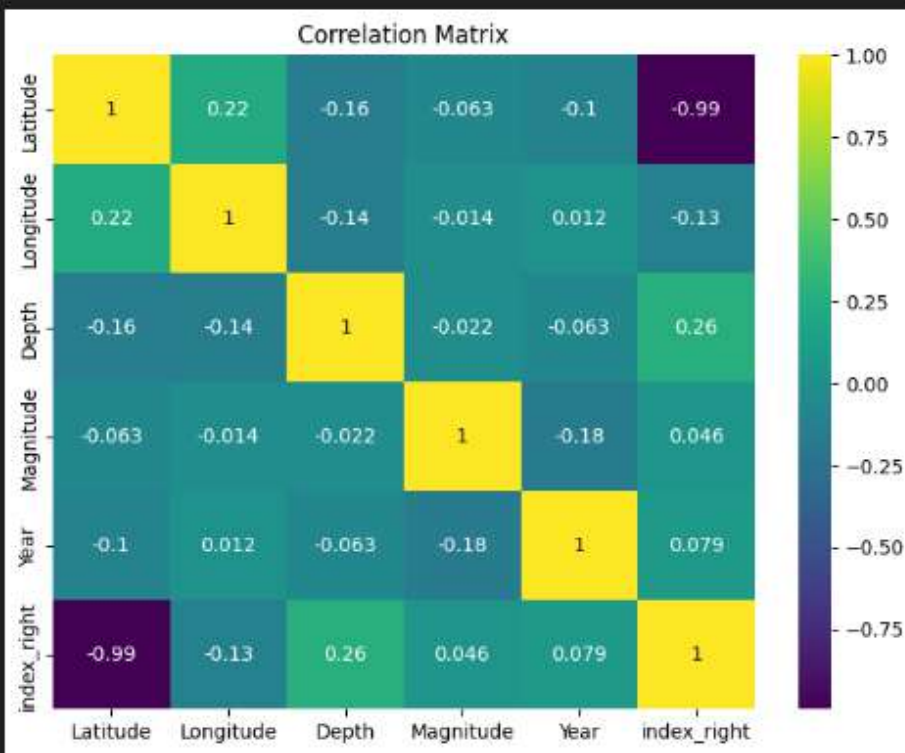
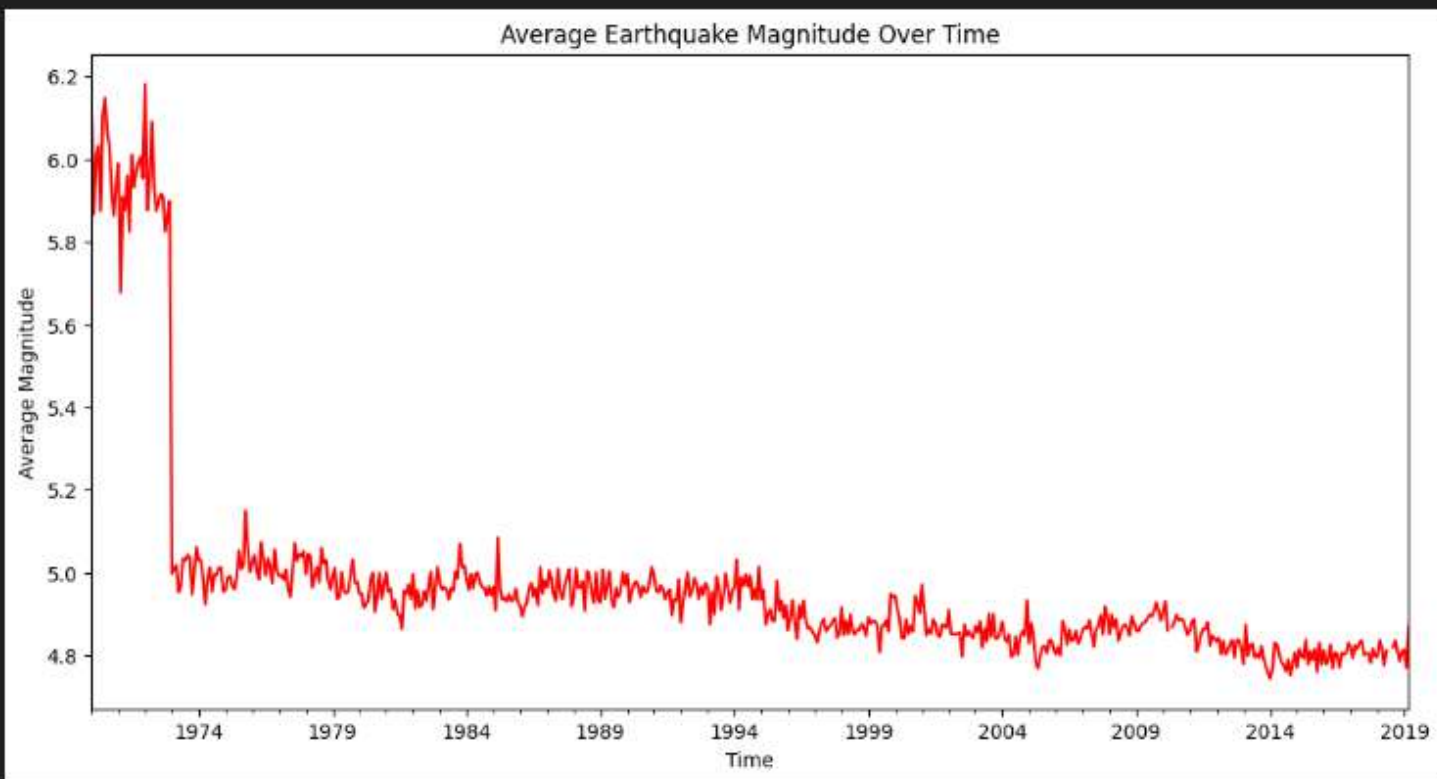




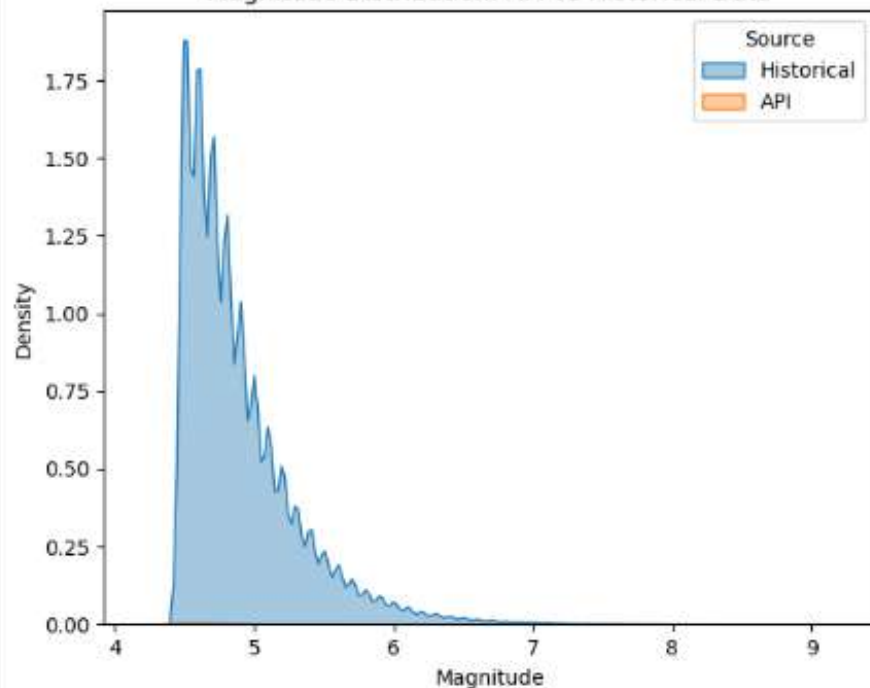


```
/tmp/ipython-input-2643866979.py:70: FutureWarning: 'M' is deprecated and will be removed in a future version, please use 'ME' instead.  
eq_with_soil.set_index('Date').resample('M')['Magnitude'].count().plot()
```

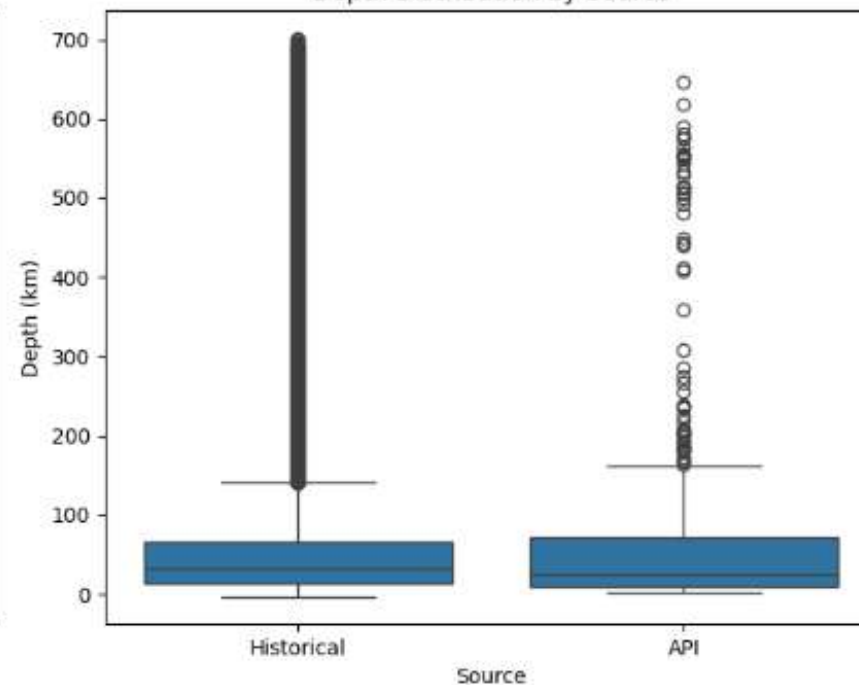




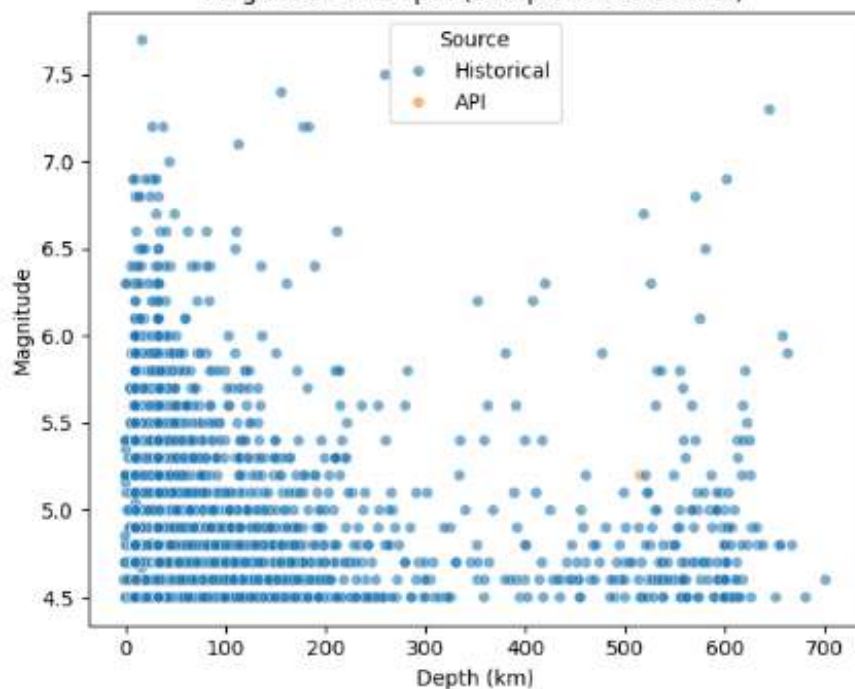
Magnitude Distribution: API vs Historical Data



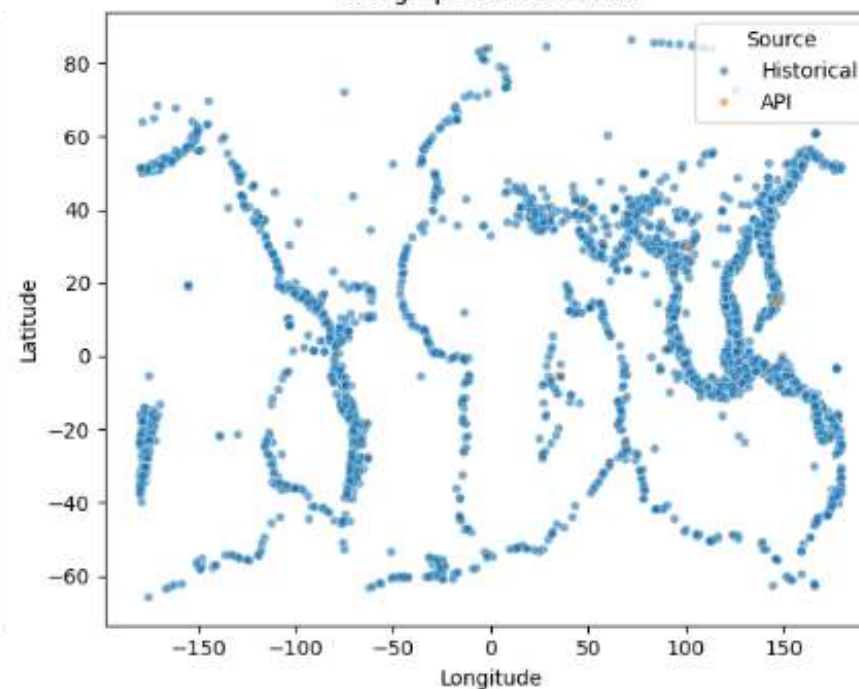
Depth Distribution by Source



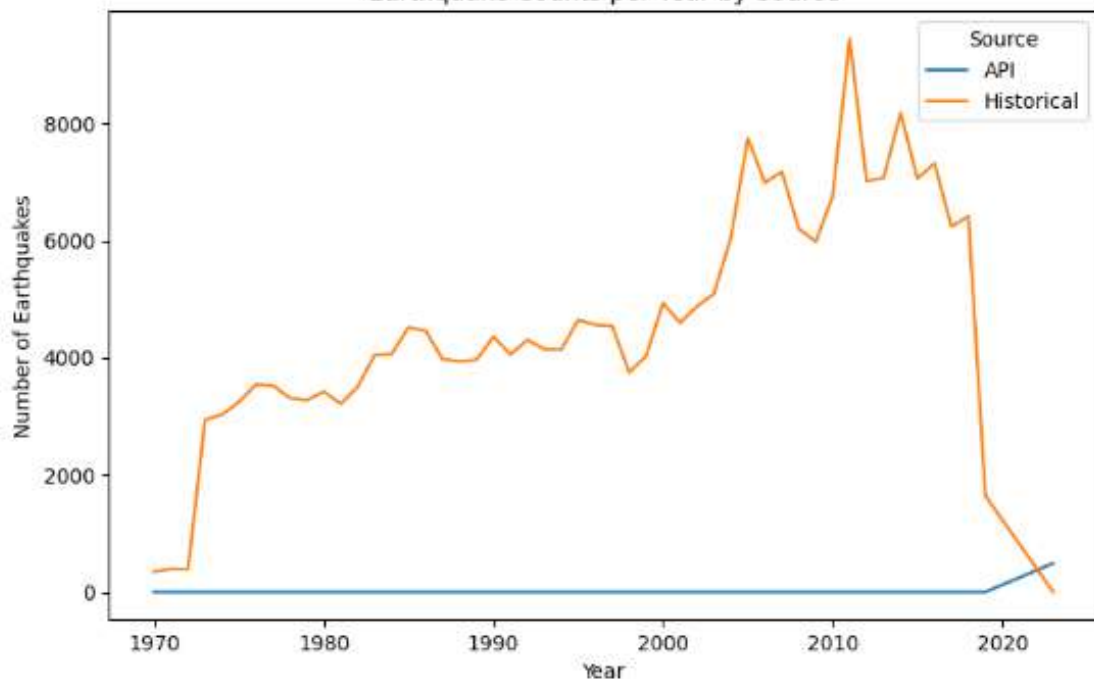
Magnitude vs Depth (Sample: 5000 events)



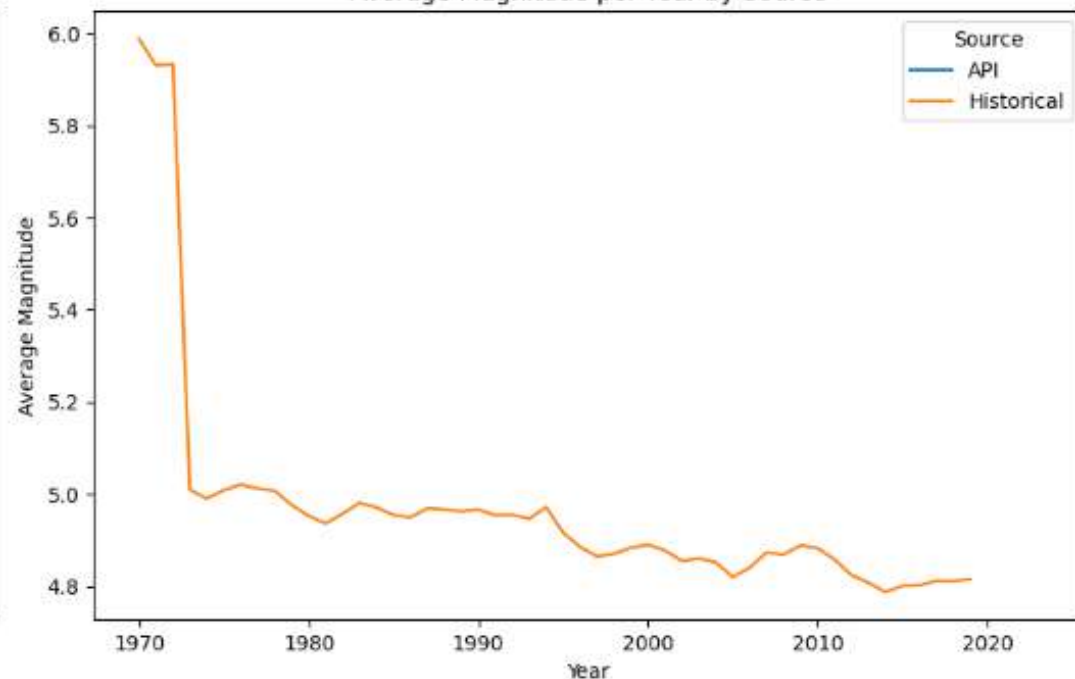
Geographic Distribution



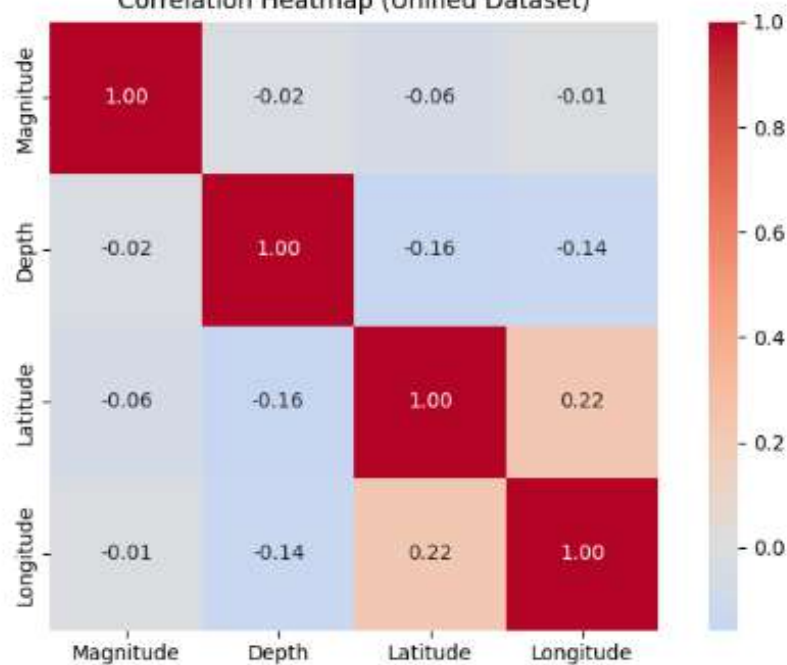
Earthquake Counts per Year by Source



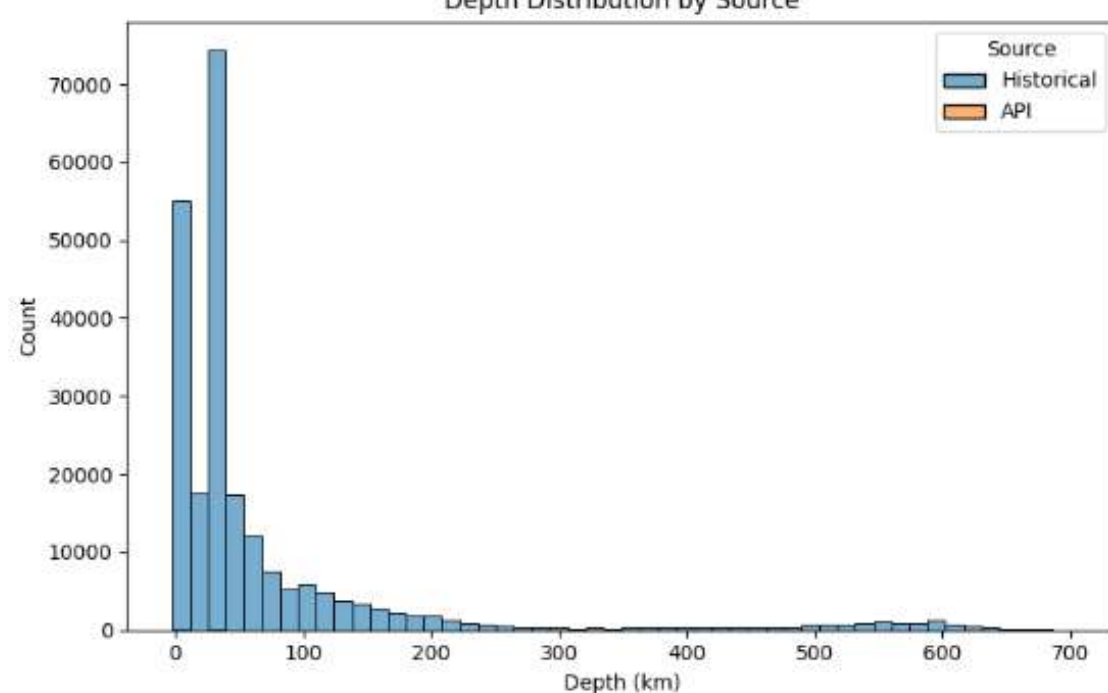
Average Magnitude per Year by Source



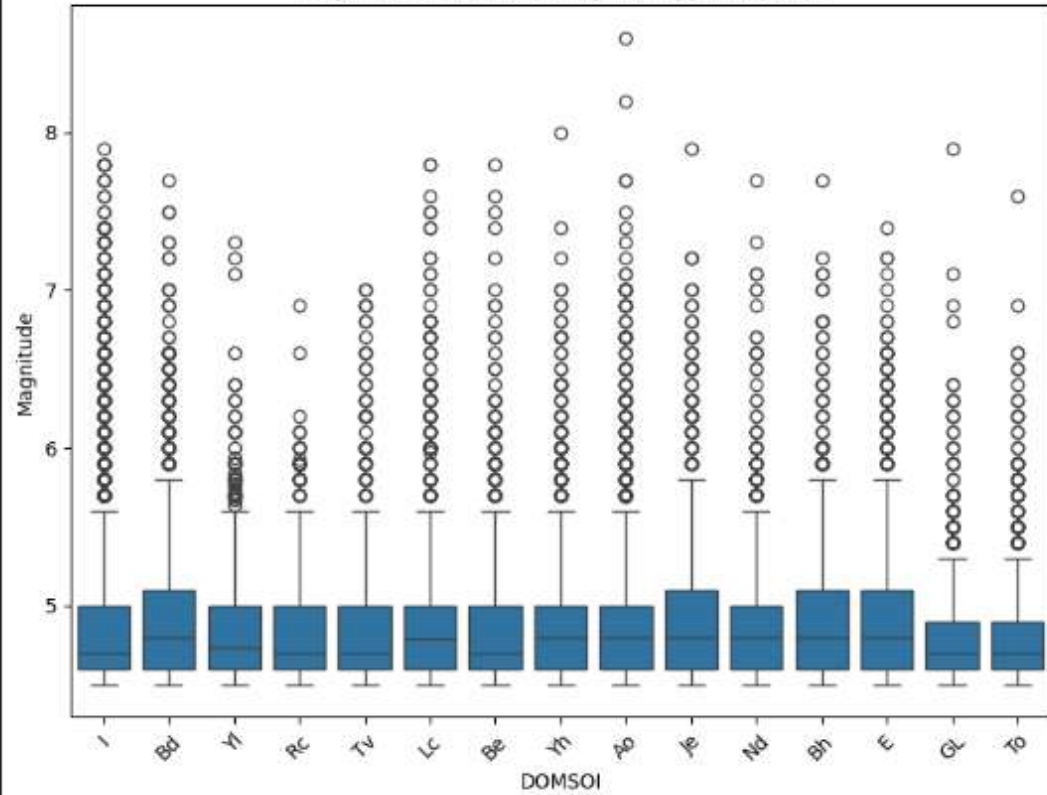
Correlation Heatmap (Unified Dataset)



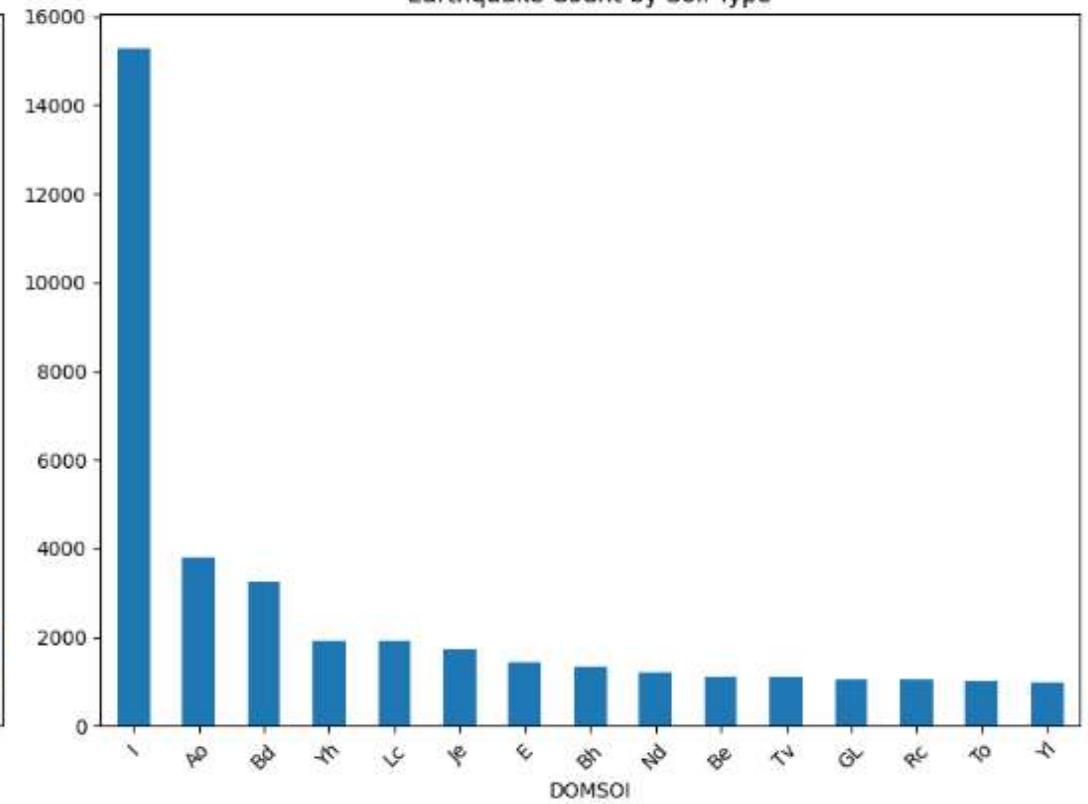
Depth Distribution by Source



Magnitude Distribution by Soil Type (Top 15)

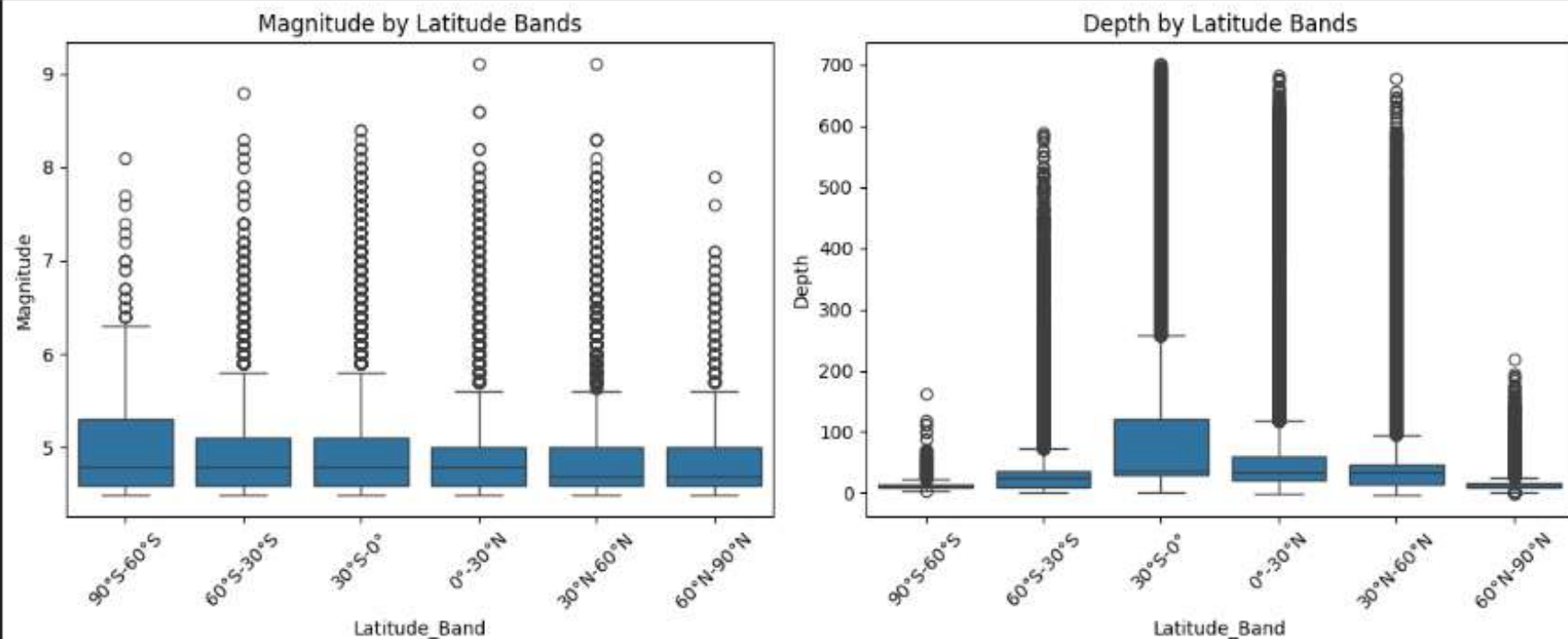


Earthquake Count by Soil Type



=== SOIL TYPE ANALYSIS ===

DOMSOI	Magnitude			Depth	
	count	mean	std	mean	std
Af	488	4.899	0.461	33.461	42.803
Ag	22	4.814	0.225	37.786	35.943
Ah	800	4.875	0.412	53.914	55.742
Ao	3789	4.878	0.419	62.065	63.428
Ap	112	5.006	0.538	239.019	246.852
Bc	536	4.842	0.390	51.995	63.716
Bd	3248	4.914	0.430	57.220	51.892
Be	1121	4.886	0.446	37.259	44.585
Bf	159	4.817	0.385	70.035	37.554
Bg	268	4.878	0.436	65.671	72.325



=== UNIFIED DATASET SUMMARY ===

Total earthquakes: 233,085

Time range: 1970-01-01 17:11:00 to 2023-08-31 21:43:25.888000

Geographic range: Lat(-84.1 to 87.2)

Lon(-180.0 to 180.0)

Magnitude range: 4.5 to 9.1

Depth range: -2.5 to 700.9 km

=== DATA QUALITY CHECK ===

Missing values per column:

Date	0
Latitude	0
Longitude	0
Depth	0
Magnitude	0
Location	0
Source	0
merge_id	0
DONSOI	179404
Year	0
Latitude_Band	0
dtype:	int64

Soil data coverage: 23.0%

Unified dataset saved to: unified_earthquake_data.csv

Missing values:

Date	0
Latitude	0
Longitude	0
Depth	0
Magnitude	0
Location	0
Source	0
merge_id	0
DOMSOI	179404
Year	0
Latitude_Band	0
dtype: int64	

Missing values after imputation:

Date	0
Latitude	0
Longitude	0
Depth	0
Magnitude	0
Location	0
Source	0
merge_id	0
DOMSOI	0
Year	0
Latitude_Band	0
dtype: int64	

Magnitude

Depth

count	2.330850e+05	2.330850e+05
mean	2.347045e-15	-6.133436e-17
std	1.000002e+00	1.000002e+00
min	-9.430561e-01	-6.419712e-01
25%	-7.049620e-01	-5.013006e-01
50%	-2.287737e-01	-3.433193e-01
75%	4.855086e-01	-7.105370e-02
max	1.000927e+01	5.269218e+00

	DOMSOI	SoilType_Encoded
--	--------	------------------

0	I	42
---	---	----

1	Af	0
---	----	---

2	Af	0
---	----	---

3	I	42
---	---	----

4	I	42
---	---	----

	Date	Latitude	Longitude	Depth	Magnitude	\
0	1970-01-01 17:11:00	-29.400	-177.169	-0.326513	1.675979	
1	1970-01-04 17:00:41	24.185	102.543	-0.525670	5.247391	
2	1970-01-05 11:49:10	23.984	102.732	-0.494578	2.390262	
3	1970-01-06 05:35:54	-9.583	151.493	-0.494578	3.342638	
4	1970-01-07 07:56:14	15.785	-59.808	-0.312227	2.628356	

	Location	Source	merge_id	DOMS01	\
0	Kermadec Islands, New Zealand	Historical	-29.4_-177.169	I	
1	Yunnan, China	Historical	24.185_102.543	Af	
2	Yunnan, China	Historical	23.984_102.732	Af	
3	D'Entrecasteaux Islands region	Historical	-9.583_151.493	I	
4	east of Guadeloupe, Leeward Islands	Historical	15.785_-59.808	I	

	Year	Latitude_Band	SoilType_Encoded	Location_Encoded	Month	Day	Hour
0	1970	30°S-0°	42	31746	1	1	17
1	1970	0°-30°N	0	32243	1	4	17
2	1970	0°-30°N	0	32243	1	5	11
3	1970	30°S-0°	42	31570	1	6	5
4	1970	0°-30°N	42	32264	1	7	7

	Magnitude	Depth	Risk_Score	Region_Cluster
0	1.675979	-0.326513	1.054600	0
1	5.247391	-0.525670	3.449418	4
2	2.390262	-0.494578	1.468475	4
3	3.342638	-0.494578	2.135138	1
4	2.628356	-0.312227	1.727560	2

```

... Sample Magnitude Classes:
      Magnitude Magnitude_Class
0    1.675979             Low
1    5.247391          Moderate
2    2.390262             Low
3    3.342638             Low
4    2.628356             Low

Class Distribution (Magnitude_Class):
Magnitude_Class
Low            0.982223
Moderate       0.015615
Strong         0.001573
Major          0.000590
Name: proportion, dtype: float64

Magnitude Statistics:
count      2.330850e+05
mean       2.349484e-15
std        1.000002e+00
min        -9.430561e-01
25%        -7.049620e-01
50%        -2.287737e-01
75%         4.855086e-01
max         1.000927e+01
Name: Magnitude, dtype: float64

```

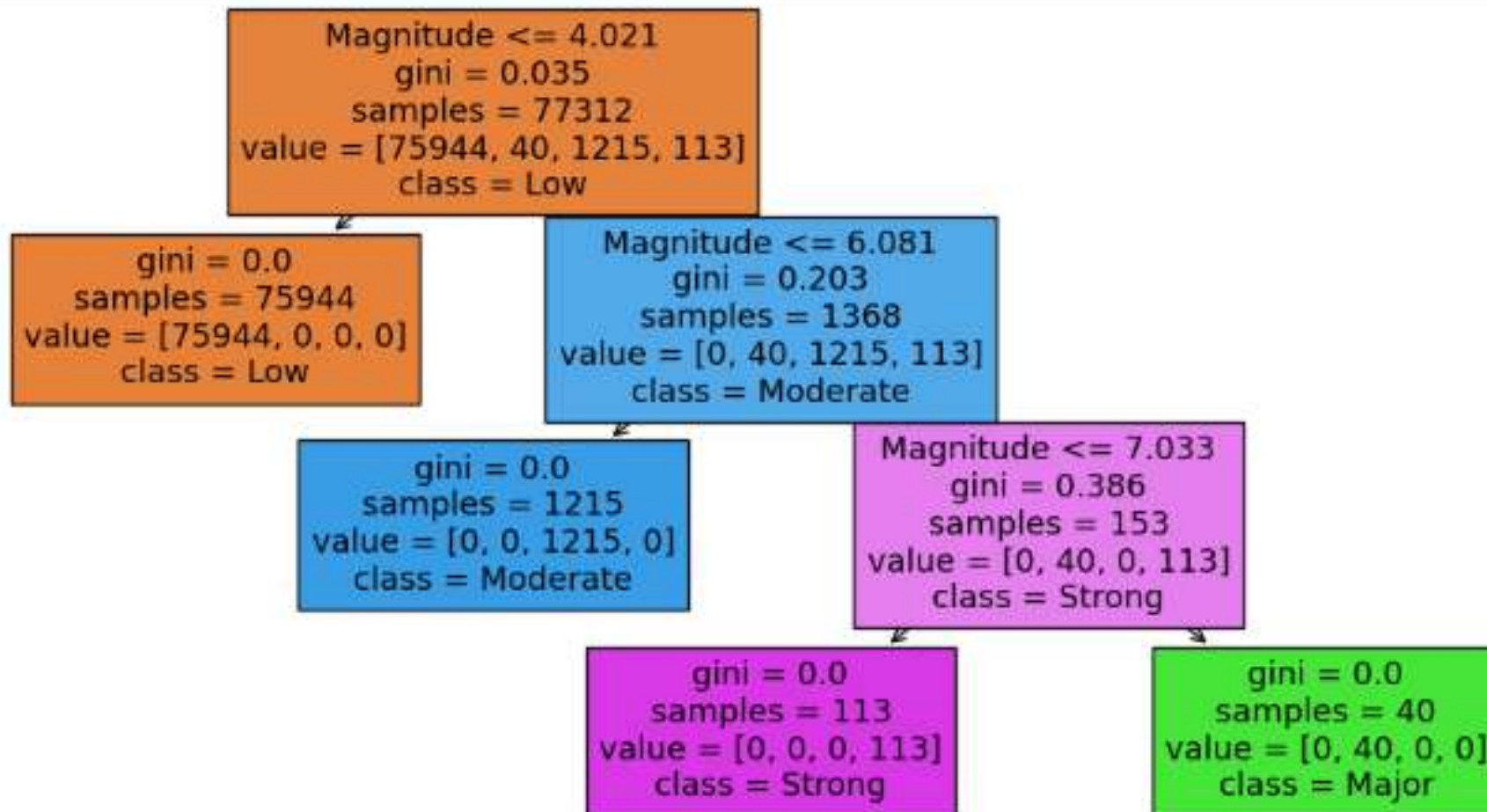
```
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:1247: FutureWarning: 'multi_class' was deprecated in version 1.5 and will be removed in 1.7. From then on, it will always use 'multinomial'. Leave it to its default value
warnings.warn(
```

```
=== Logistic Regression Report ===
```

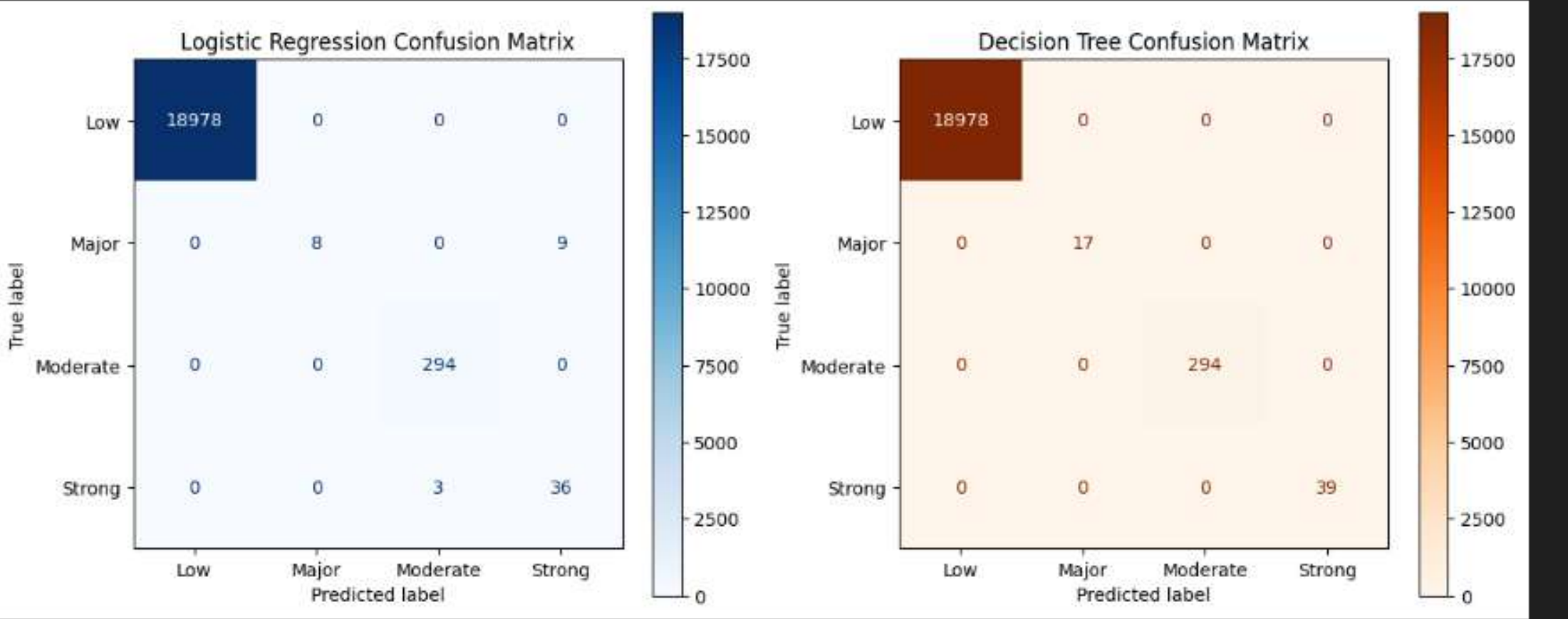
	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	0.47	0.64	17
Moderate	0.99	1.00	0.99	294
Strong	0.80	0.92	0.86	39
accuracy			1.00	19328
macro avg	0.95	0.85	0.87	19328
weighted avg	1.00	1.00	1.00	19328

=== Decision Tree Report ===

	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	1.00	1.00	17
Moderate	1.00	1.00	1.00	294
Strong	1.00	1.00	1.00	39
accuracy			1.00	19328
macro avg	1.00	1.00	1.00	19328
weighted avg	1.00	1.00	1.00	19328



Logistic Regression Accuracy: 0.999
Decision Tree Accuracy: 1.000

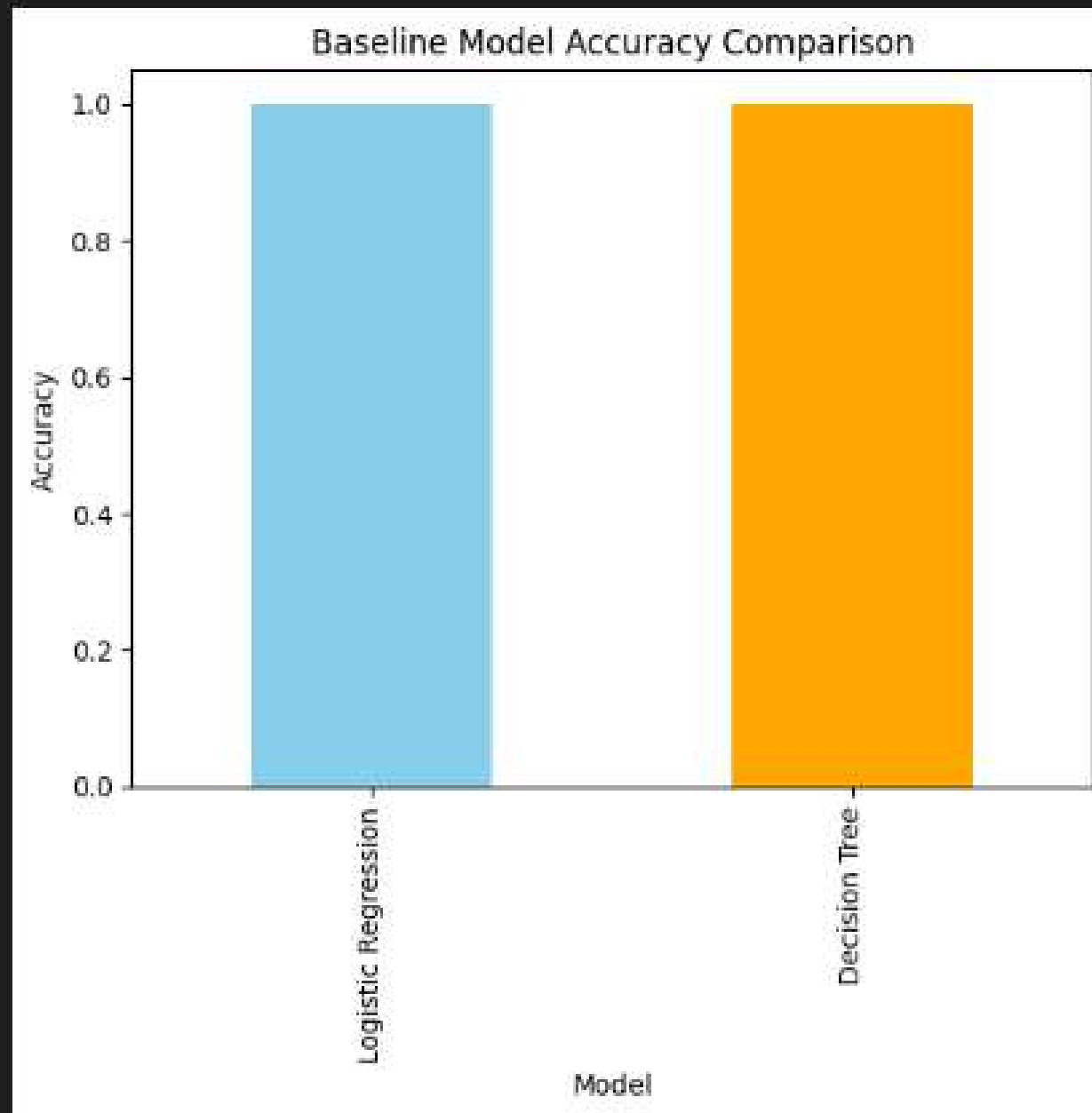


=== Baseline Model Comparison ===

	Model	Accuracy
0	Logistic Regression	0.999379
1	Decision Tree	1.000000

Updated Model Comparison (with F1-score)

	Model	Accuracy	F1-score
0	Logistic Regression	0.999379	0.873017
1	Decision Tree	1.000000	1.000000



Original Classes: ['Low' 'Major' 'Moderate' 'Strong']
Encoded Mapping: {'Low': np.int64(0), 'Major': np.int64(1), 'Moderate': np.int64(2), 'Strong': np.int64(3)}
Unique Encoded Classes in Train: [0 1 2 3]
Unique Encoded Classes in Test: [0 1 2 3]
Computed class weights: {np.int64(0): 0.2545033182344886, np.int64(2): 15.907818930041152, np.int64(3): 171.04424778761063, np.int64(1): 483.2}
Using cv=3 (min class count = 40)

Starting XGBoost fine-tuning...

Fitting 3 folds for each of 64 candidates, totalling 192 fits

/usr/local/lib/python3.12/dist-packages/xgboost/training.py:199: UserWarning: [18:10:21] WARNING: /workspace/src/learner.cc:790:

Parameters: { "use_label_encoder" } are not used.

```
bst.update(dtrain, iteration=i, fobj=obj)
```

Best Parameters Found: {'colsample_bytree': 0.8, 'gamma': 0, 'learning_rate': 0.01, 'max_depth': 4, 'n_estimators': 100, 'subsample': 0.8}

XGBoost Accuracy (Tuned): 1.0000

Classification Report (XGBoost):

	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	1.00	1.00	17
Moderate	1.00	1.00	1.00	294
Strong	1.00	1.00	1.00	39
accuracy			1.00	19328
macro avg	1.00	1.00	1.00	19328
weighted avg	1.00	1.00	1.00	19328

Saved tuned model as 'xgboost_tuned_weighted.pkl'

Saved LabelEncoder as 'label_encoder.pkl'

✦ Loading tuned models...

✓ Loaded XGBoost model

✓ Loaded label encoder

🔍 Evaluating models...

=== Logistic Regression ===

Accuracy: 0.9994

	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	0.47	0.64	17
Moderate	0.99	1.00	0.99	294
Strong	0.88	0.92	0.86	39
accuracy			1.00	19328
macro avg	0.95	0.85	0.87	19328
weighted avg	1.00	1.00	1.00	19328

=== XGBoost (Tuned) ===

Accuracy: 1.0000

	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	1.00	1.00	17

...

0 Logistic Regression 0.999379

1 XGBoost (Tuned) 1.000000

📄 Saved summary as: model_comparison_summary.csv

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#)...

☒ Evaluation Metrics

• XGBoost Model

Accuracy: 1.0000

Precision (macro): 1.0000

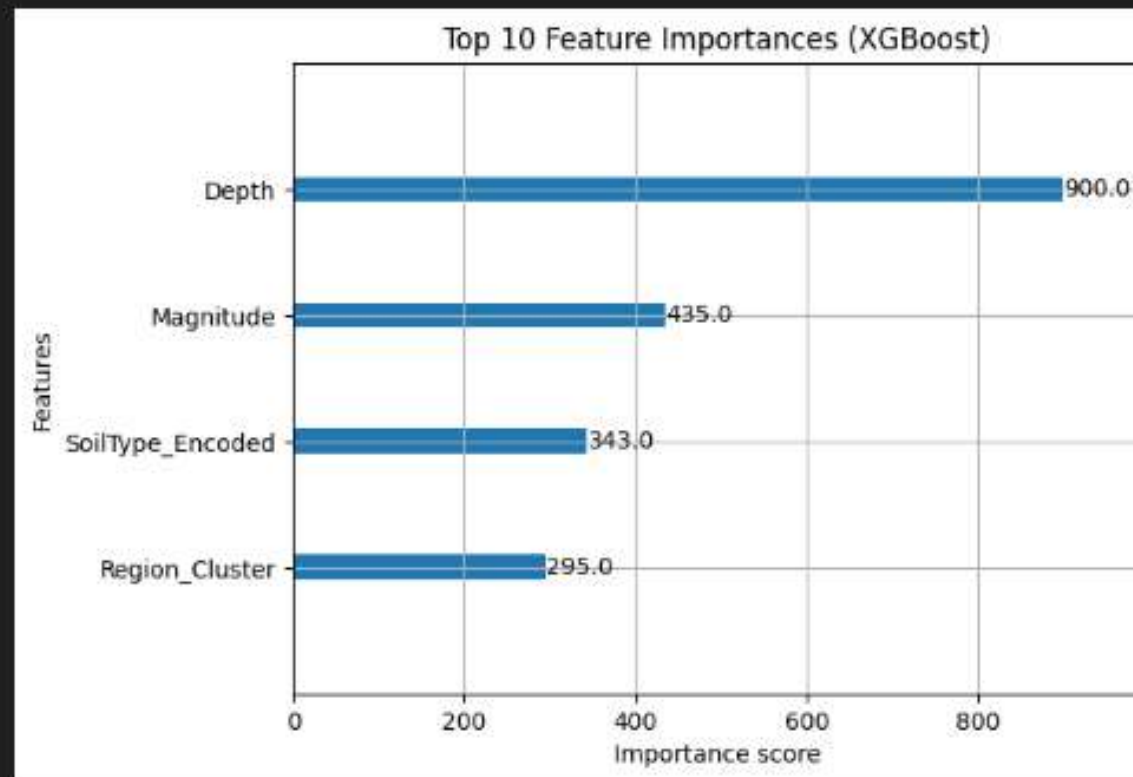
Recall (macro): 1.0000

F1-score (macro): 1.0000

Classification Report:

	precision	recall	f1-score	support
Low	1.00	1.00	1.00	18978
Major	1.00	1.00	1.00	17
Moderate	1.00	1.00	1.00	294
Strong	1.00	1.00	1.00	39
accuracy			1.00	19328
macro avg	1.00	1.00	1.00	19328
weighted avg	1.00	1.00	1.00	19328

<Figure size 1000x600 with 0 Axes>



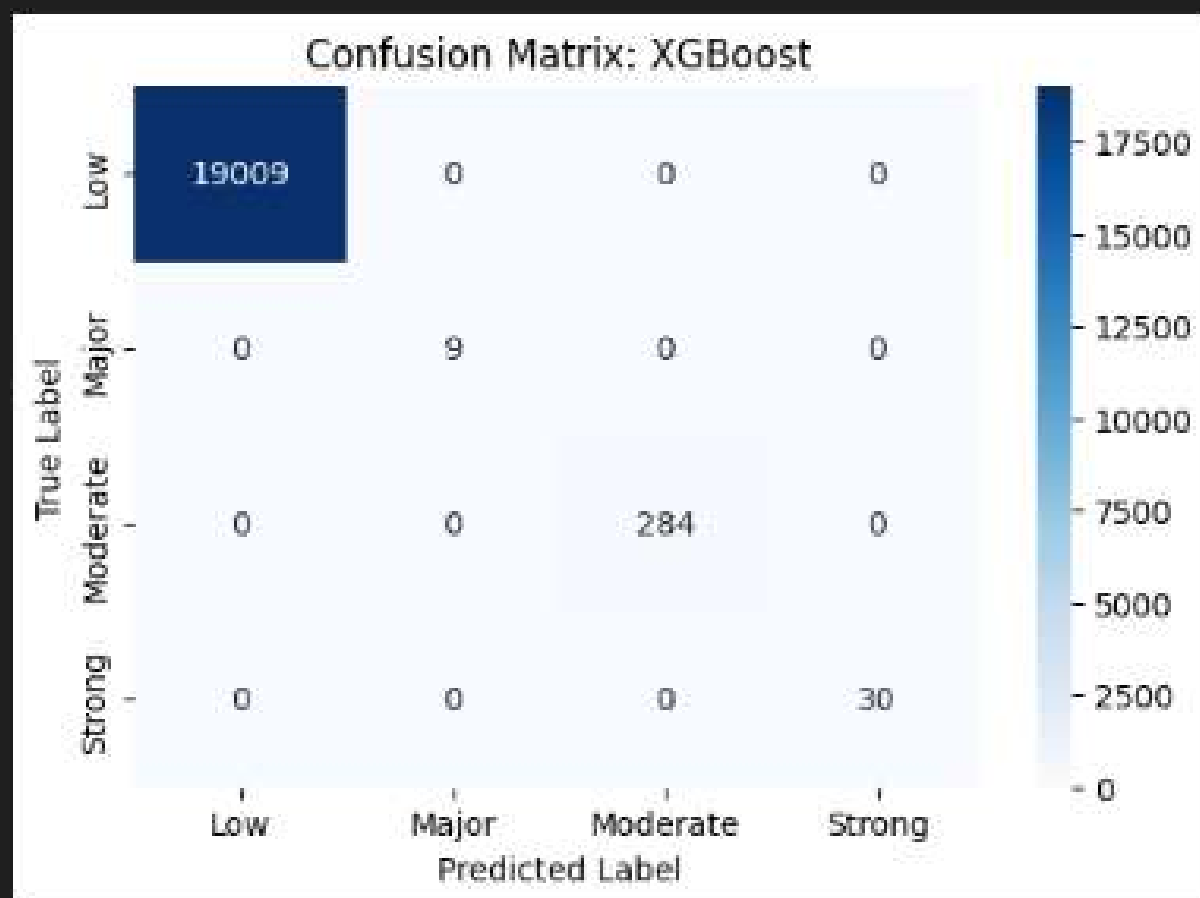
🟢 New Test Distribution: target

Low	19009
Moderate	284
Strong	30
Major	9

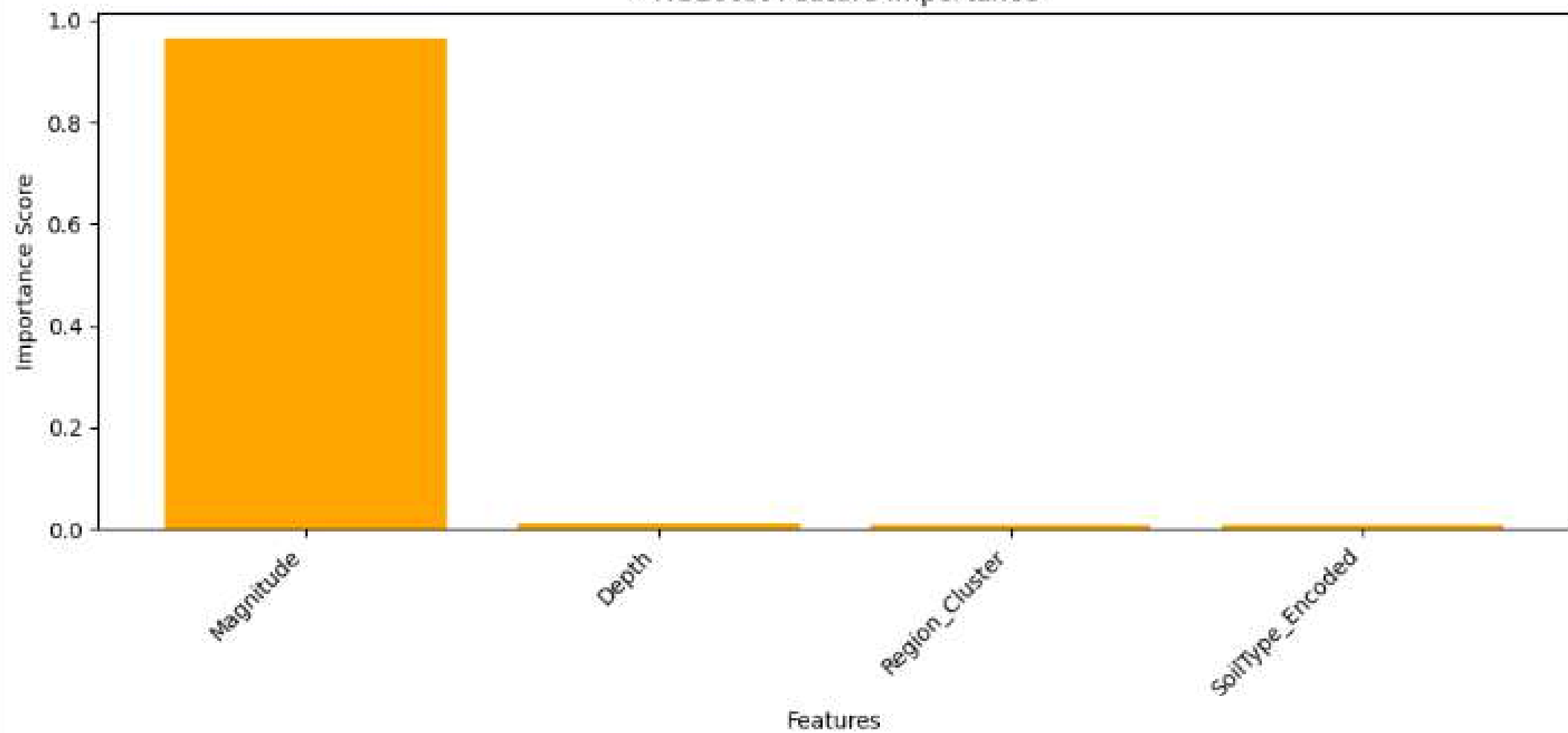
Name: count, dtype: int64

/tmp/ipython-input-1182867885.py:9: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass `include_groups=False` to exclude the groupings or explicitly select the grouping columns.
test_samples = df.groupby("target", group_keys=False).apply(lambda x: x.sample(n=min(1, len(x)), random_state=42))

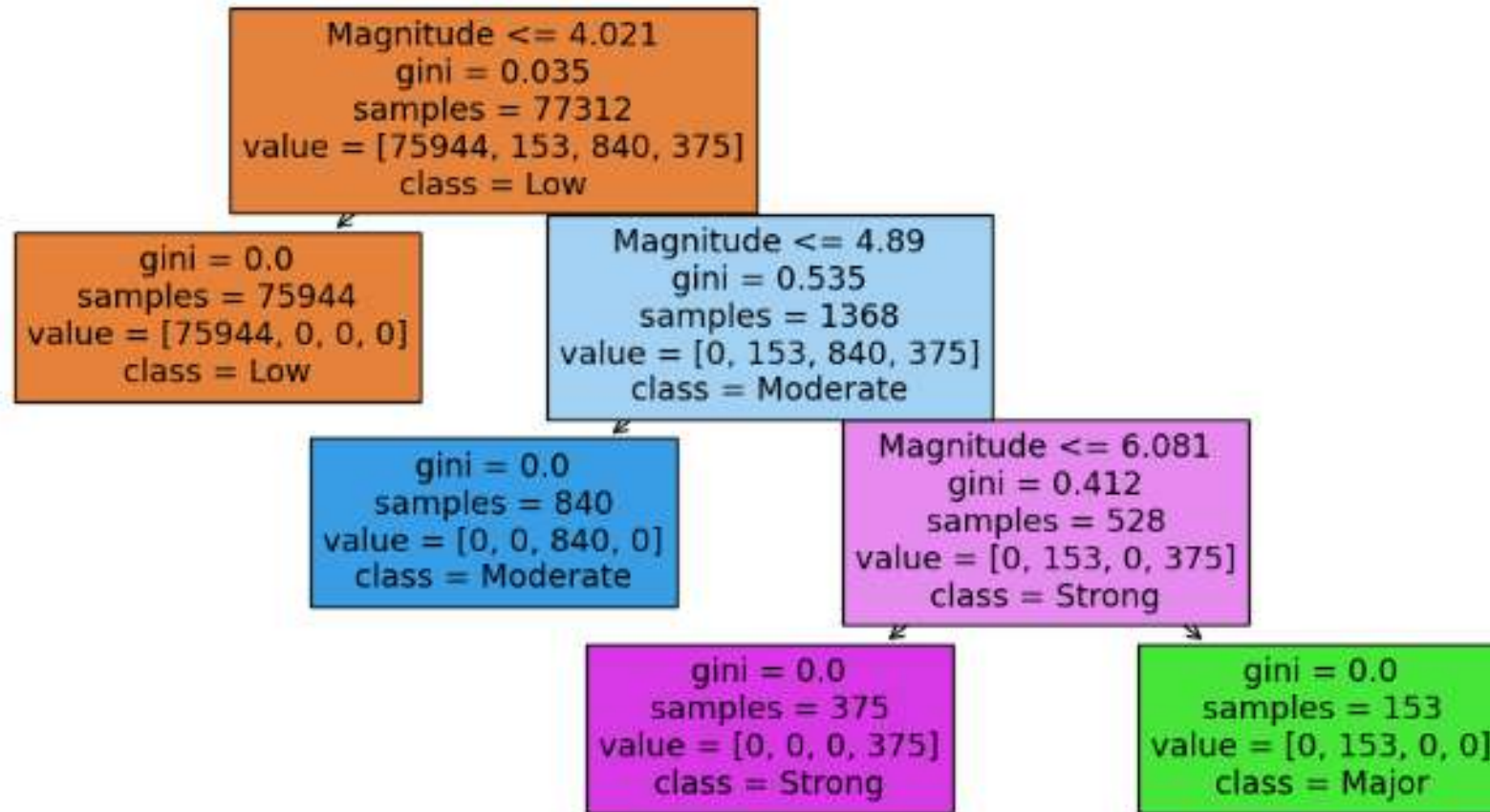
Classification Report for XGBoost Model				
	precision	recall	f1-score	support
Low	1.00	1.00	1.00	19009
Major	1.00	1.00	1.00	9
Moderate	1.00	1.00	1.00	284
Strong	1.00	1.00	1.00	30
accuracy			1.00	19332
macro avg	1.00	1.00	1.00	19332
weighted avg	1.00	1.00	1.00	19332



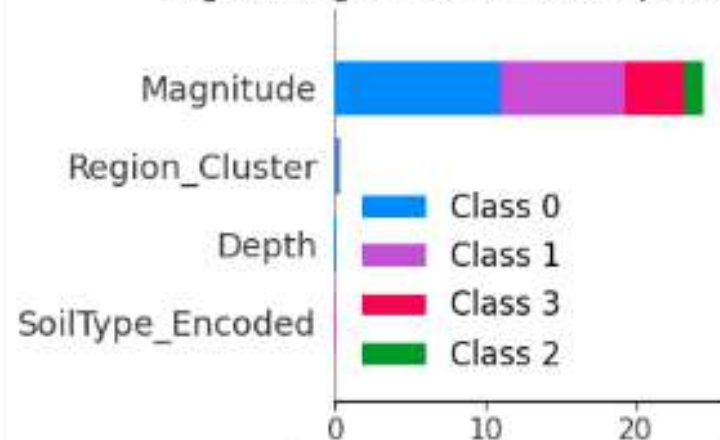
✂ XGBoost Feature Importance



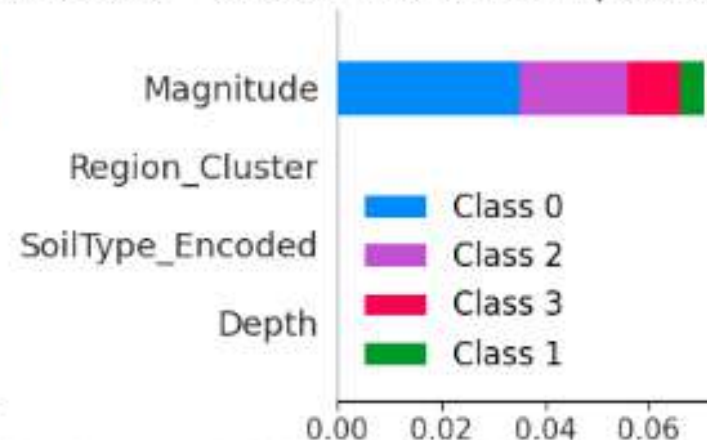
Decision Tree Structure



Logistic Regression Feature Importance (SHAP)



Decision Tree Feature Importance (SHAP)



ean(|SHAP value|) (average impact on model output) (SHAP value) (average impact on model output)