In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```
data = pd.read_csv('earthquake_data_tsunami.csv')
data
```

Out[4]:

	magnitude	cdi	mmi	sig	nst	dmin	gap	depth	latitude	longitude	Year
0	7.0	8	7	768	117	0.509	17.0	14.000	-9.7963	159.596	2022
1	6.9	4	4	735	99	2.229	34.0	25.000	-4.9559	100.738	2022
2	7.0	3	3	755	147	3.125	18.0	579.000	-20.0508	-178.346	2022
3	7.3	5	5	833	149	1.865	21.0	37.000	-19.2918	-172.129	2022
4	6.6	0	2	670	131	4.998	27.0	624.464	-25.5948	178.278	2022
...	...	...	...	...	...	...	...	...	...	...	...
777	7.7	0	8	912	427	0.000	0.0	60.000	13.0490	-88.660	2001
778	6.9	5	7	745	0	0.000	0.0	36.400	56.7744	-153.281	2001
779	7.1	0	7	776	372	0.000	0.0	103.000	-14.9280	167.170	2001
780	6.8	0	5	711	64	0.000	0.0	33.000	6.6310	126.899	2001
781	7.5	0	7	865	324	0.000	0.0	33.000	6.8980	126.579	2001

782 rows × 13 columns

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 782 entries, 0 to 781
Data columns (total 13 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   magnitude  782 non-null    float64
 1   cdi        782 non-null    int64  
 2   mmi        782 non-null    int64  
 3   sig        782 non-null    int64  
 4   nst        782 non-null    int64  
 5   dmin       782 non-null    float64
 6   gap        782 non-null    float64
 7   depth      782 non-null    float64
 8   latitude   782 non-null    float64
 9   longitude  782 non-null    float64
 10  Year       782 non-null    int64  
 11  Month      782 non-null    int64  
 12  tsunami    782 non-null    int64  
dtypes: float64(6), int64(7)
memory usage: 79.6 KB
```

In [6]:
```
data.describe()
```

Out[6]:

	magnitude	cdi	mmi	sig	nst	dmin	
count	782.000000	782.000000	782.000000	782.000000	782.000000	782.000000	782.000
mean	6.941125	4.333760	5.964194	870.108696	230.250639	1.325757	25.038
std	0.445514	3.169939	1.462724	322.465367	250.188177	2.218805	24.225
min	6.500000	0.000000	1.000000	650.000000	0.000000	0.000000	0.000
25%	6.600000	0.000000	5.000000	691.000000	0.000000	0.000000	14.625
50%	6.800000	5.000000	6.000000	754.000000	140.000000	0.000000	20.000
75%	7.100000	7.000000	7.000000	909.750000	445.000000	1.863000	30.000
max	9.100000	9.000000	9.000000	2910.000000	934.000000	17.654000	239.000

In [7]:
```
data.isnull().sum()
```

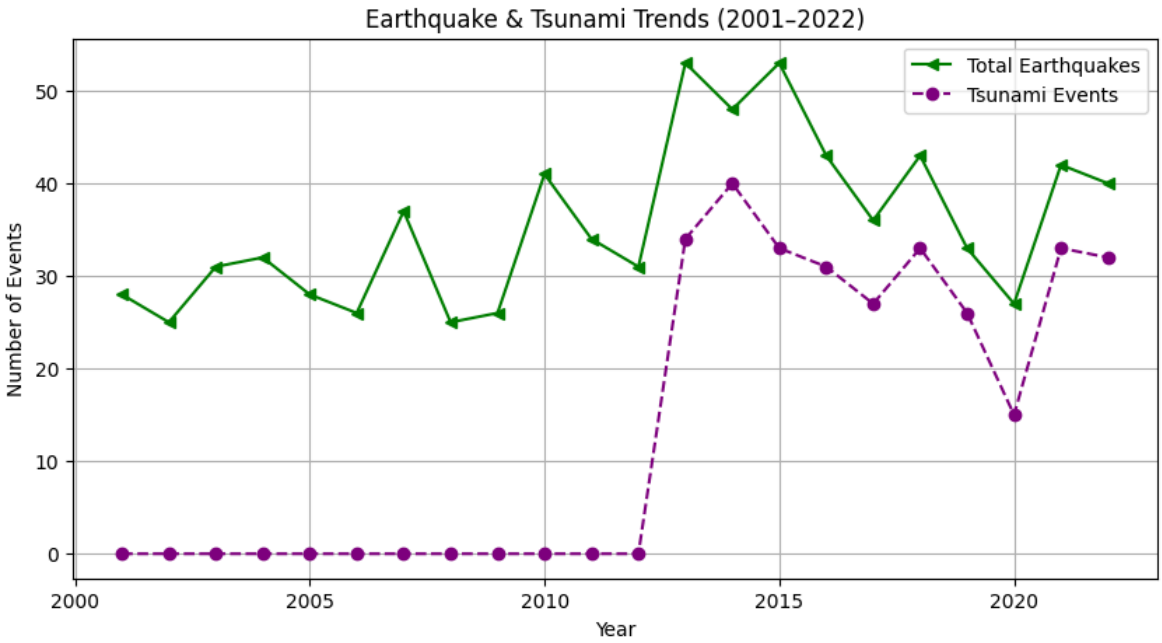
Out[7]:
```
magnitude    0
cdi          0
mmi          0
sig          0
nst          0
dmin         0
gap          0
depth        0
latitude     0
longitude    0
Year         0
Month        0
tsunami      0
dtype: int64
```

1. Time-Based Analysis:

o Explore how earthquake occurrences and tsunami events have changed over the 22-year period (2001–2022).

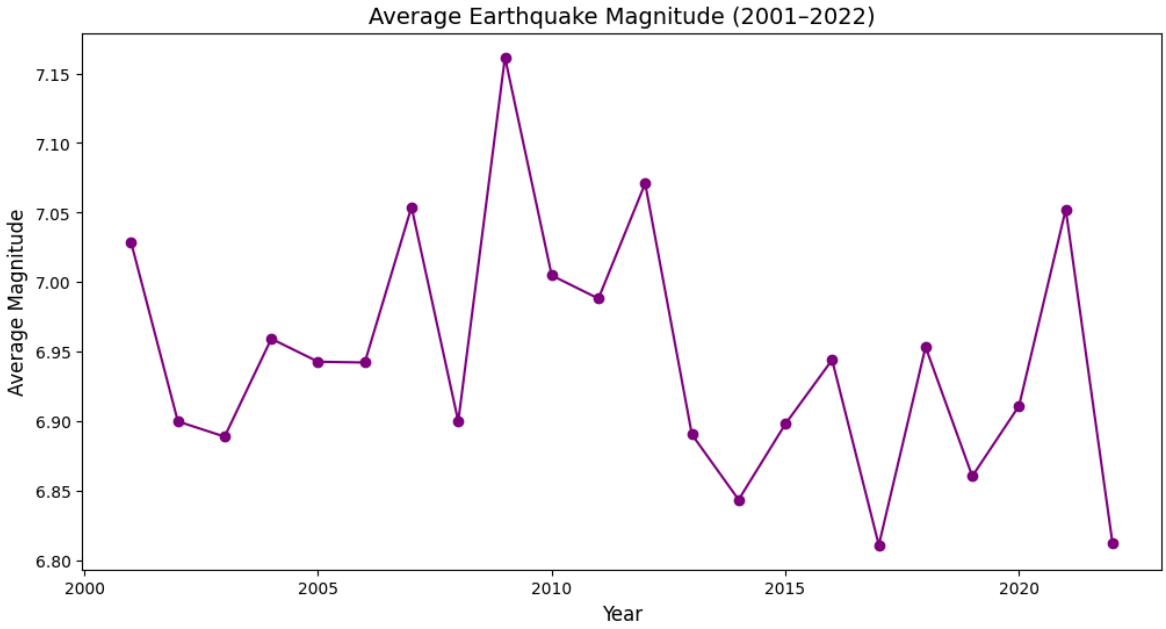
In [8]:
```
# Group earthquake counts per year
earthquake_count = data.groupby('Year').size()

# Group tsunami events per year (sum of 1s and 0s)
tsunami_count = data.groupby('Year')['tsunami'].sum()

# Plot directly using those grouped results
plt.figure(figsize=(10,5))
plt.plot(earthquake_count.index, earthquake_count.values,
         marker='<', linestyle='-', color='green', label='Total Earthquakes')

plt.plot(tsunami_count.index, tsunami_count.values,
         marker='o', linestyle='--', color='purple', label='Tsunami Events')

plt.title("Earthquake & Tsunami Trends (2001-2022)")
plt.xlabel("Year")
plt.ylabel("Number of Events")
```

```
plt.legend()
plt.grid(True)
plt.show()
```



o Identify any trends in the frequency or magnitude of earthquakes over time.

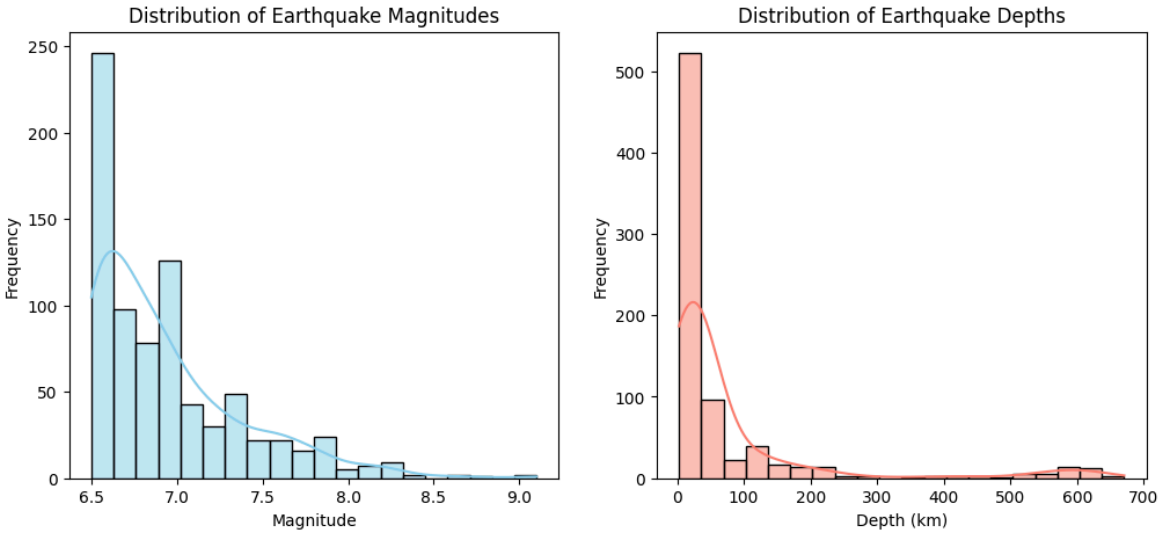
In [11]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Group by Year to calculate average magnitude
yearly_magnitude = data.groupby("Year")["magnitude"].mean().reset_index()

# Plot the average magnitude trend
plt.figure(figsize=(12,6))
plt.plot(yearly_magnitude["Year"], yearly_magnitude["magnitude"],
         marker='o', color='purple')

# Add title and labels
plt.title("Average Earthquake Magnitude (2001-2022)", fontsize=14)
plt.xlabel("Year", fontsize=12)
plt.ylabel("Average Magnitude", fontsize=12)

plt.show()
```



2. Magnitude and Depth Analysis

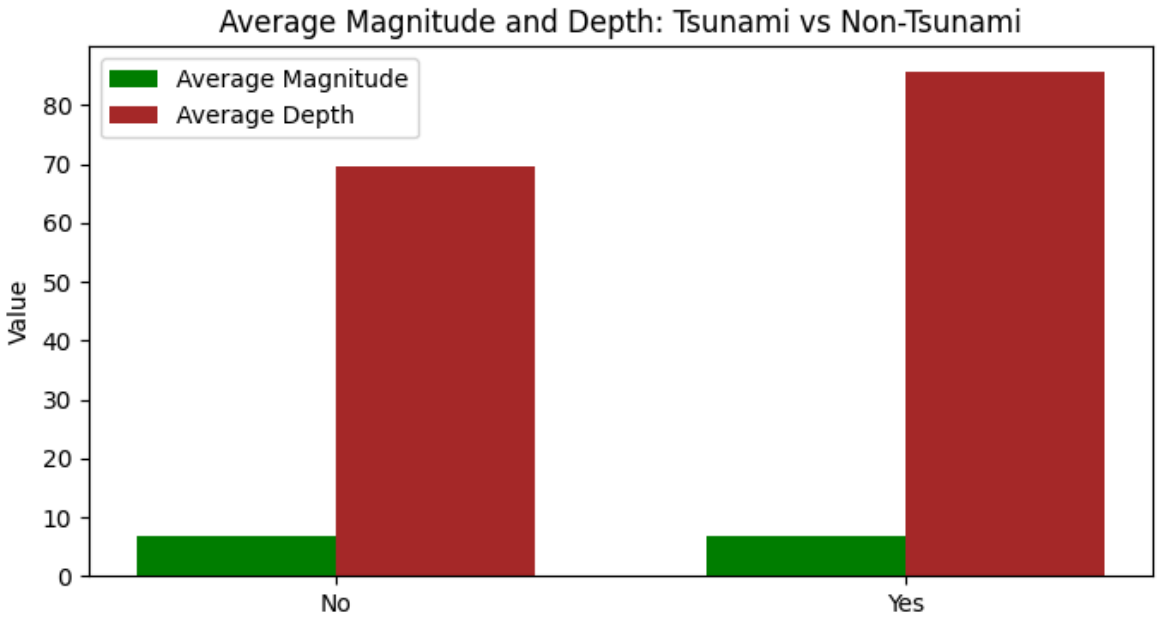
o Analyze the distribution of earthquake magnitudes and depths.

In [5]:
```
# Plot the distribution of Magnitude and Depth
plt.figure(figsize=(12,5))

# Magnitude distribution
plt.subplot(1,2,1)
sns.histplot(data['magnitude'], bins=20, kde=True, color='skyblue')
plt.title('Distribution of Earthquake Magnitudes')
plt.xlabel('Magnitude')
plt.ylabel('Frequency')

# Depth distribution
plt.subplot(1,2,2)
sns.histplot(data['depth'], bins=20, kde=True, color='salmon')
plt.title('Distribution of Earthquake Depths')
plt.xlabel('Depth (km)')
plt.ylabel('Frequency')

plt.show()
```



o Compare the average magnitude and depth of tsunami vs. non-tsunami events.

In [6]:
```
import matplotlib.pyplot as plt
import pandas as pd

# Group by tsunami (0 = No, 1 = Yes) and calculate mean
avg_stats = data.groupby('tsunami')[['magnitude','depth']].mean().reset_index()
avg_stats['Tsunami'] = avg_stats['tsunami'].map({0:'No', 1:'Yes'})

# Plotting
plt.figure(figsize=(8,4))
bar_width = 0.35
x = range(len(avg_stats))

# Average Magnitude
plt.bar(x, avg_stats['magnitude'], width=bar_width, label='Average Magnitude', c

# Average Depth
plt.bar([i + bar_width for i in x], avg_stats['depth'], width=bar_width, label='

# X-axis labels
plt.xticks([i + bar_width/2 for i in x], avg_stats['Tsunami'])
plt.ylabel('Value')
plt.title('Average Magnitude and Depth: Tsunami vs Non-Tsunami')
plt.legend()
plt.show()
```



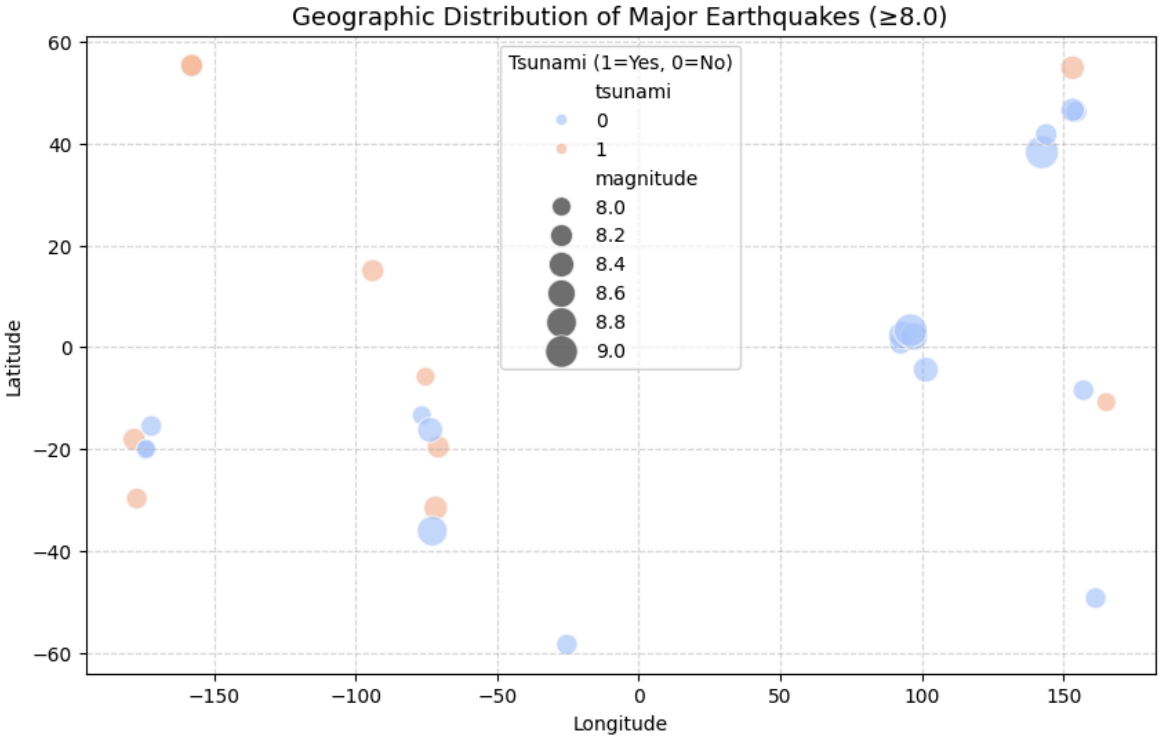
o Highlight major earthquakes (≥8.0) and their characteristics.

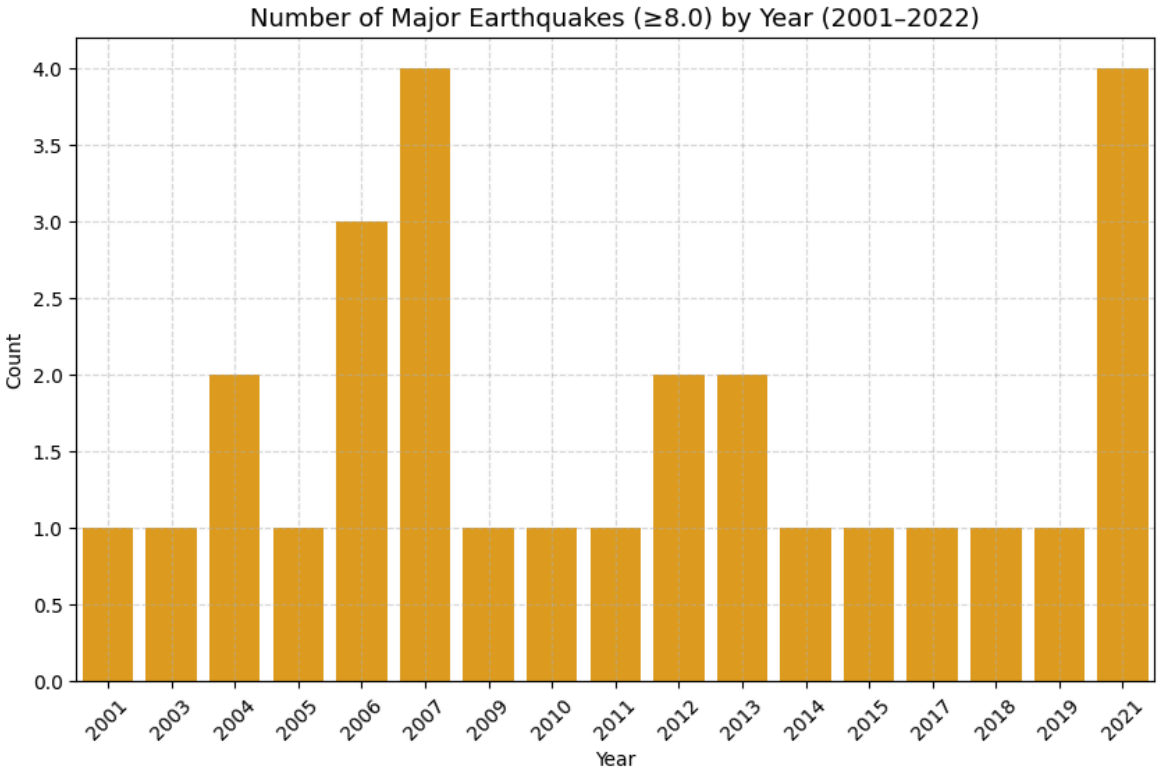
In [41]:
```
# Convert time column to datetime and extract year if not already present
if "Year" not in df.columns:
    df["time"] = pd.to_datetime(df["time"], errors="coerce")
    df["Year"] = df["time"].dt.year

# Filter major earthquakes (magnitude ≥ 8.0)
major_eq = df[df["magnitude"] >= 8.0]
print(f"Total Major Earthquakes (≥8.0): {len(major_eq)}\n")

#  Summary Statistics
print("Summary Statistics for Major Earthquakes:\n")
summary = major_eq[["magnitude", "depth", "latitude", "longitude", "Year", "tsun
print(summary)

# Geographic Distribution Plot
plt.figure(figsize=(10,6))
sns.scatterplot(
    data=major_eq,
    x="longitude", y="latitude",
    size="magnitude", hue="tsunami",
    sizes=(100,300), palette="coolwarm", alpha=0.7
)
plt.title("Geographic Distribution of Major Earthquakes (≥8.0)", fontsize=13)
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.legend(title="Tsunami (1=Yes, 0=No)")
plt.grid(True, linestyle="--", alpha=0.5)
plt.show()

# Yearly Frequency Trend
plt.figure(figsize=(10,6))
sns.countplot(x="Year", data=major_eq, color="orange")
plt.title("Number of Major Earthquakes (≥8.0) by Year (2001–2022)", fontsize=13)
plt.xlabel("Year")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.grid(True, linestyle="--", alpha=0.5)
```

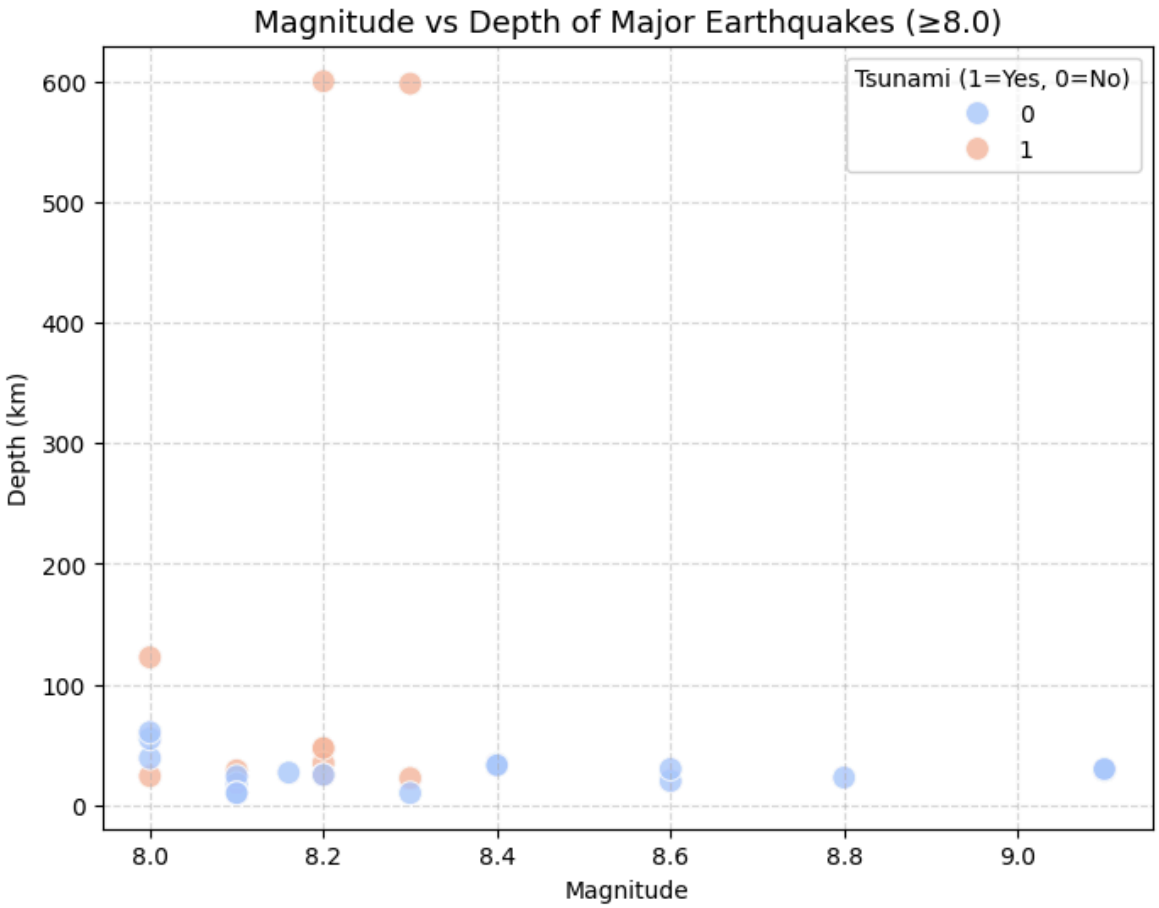
```
plt.show()

# Magnitude vs Depth Visualization
plt.figure(figsize=(8,6))
sns.scatterplot(
    data=major_eq,
    x="magnitude", y="depth",
    hue="tsunami", palette="coolwarm", s=100, alpha=0.8
)
plt.title("Magnitude vs Depth of Major Earthquakes (≥8.0)", fontsize=13)
plt.xlabel("Magnitude")
plt.ylabel("Depth (km)")
plt.legend(title="Tsunami (1=Yes, 0=No)")
plt.grid(True, linestyle="--", alpha=0.5)
plt.show()
```

```
Total Major Earthquakes (≥8.0): 28

Summary Statistics for Major Earthquakes:
```

	magnitude	depth	latitude	longitude	Year	tsunami
count	28.000000	28.000000	28.000000	28.000000	28.000000	28.000000
mean	8.280714	73.227857	0.158446	-1.448004	2011.071429	0.357143
std	0.303631	150.047721	32.430338	133.458224	6.163985	0.487950
min	8.000000	10.000000	-58.415700	-178.153000	2001.000000	0.000000
25%	8.100000	22.872500	-19.704775	-109.881725	2006.000000	0.000000
50%	8.200000	28.965000	-7.138950	-48.044850	2010.500000	0.000000
75%	8.325000	40.915000	20.840900	142.757250	2015.500000	1.000000
max	9.100000	600.000000	55.474200	165.114000	2021.000000	1.000000



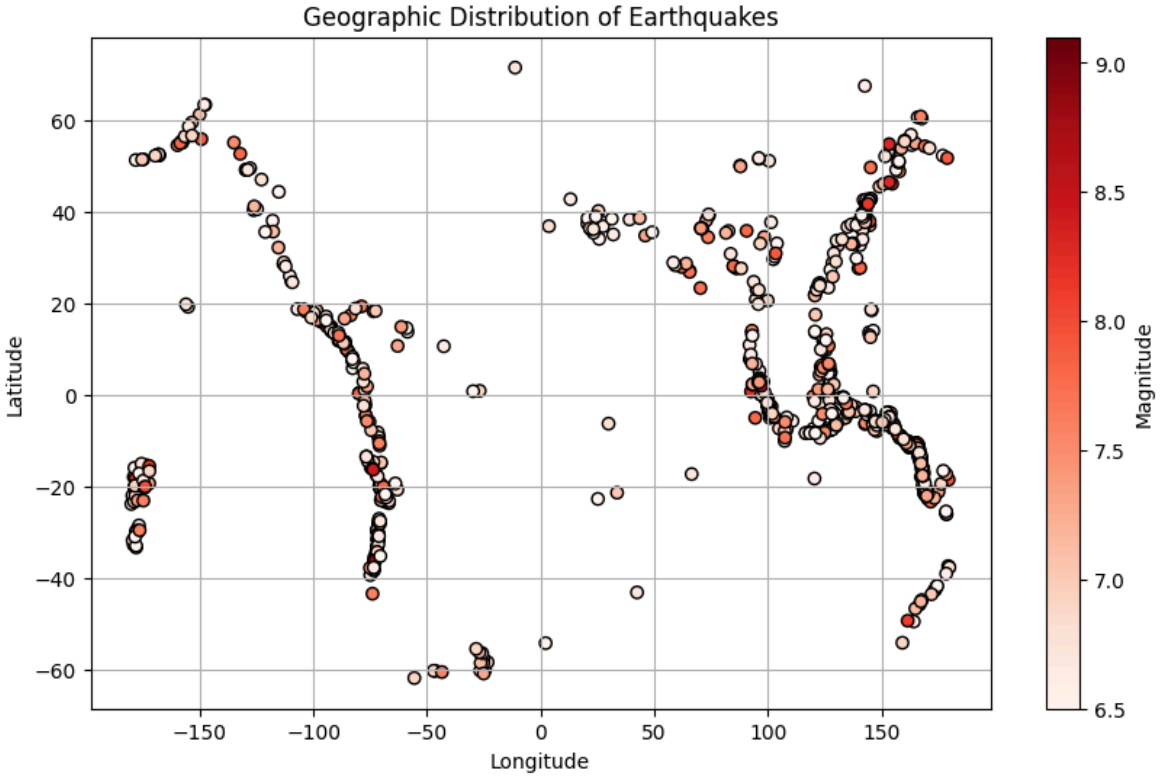




3.Geographic Distribution Using 2D Plotting

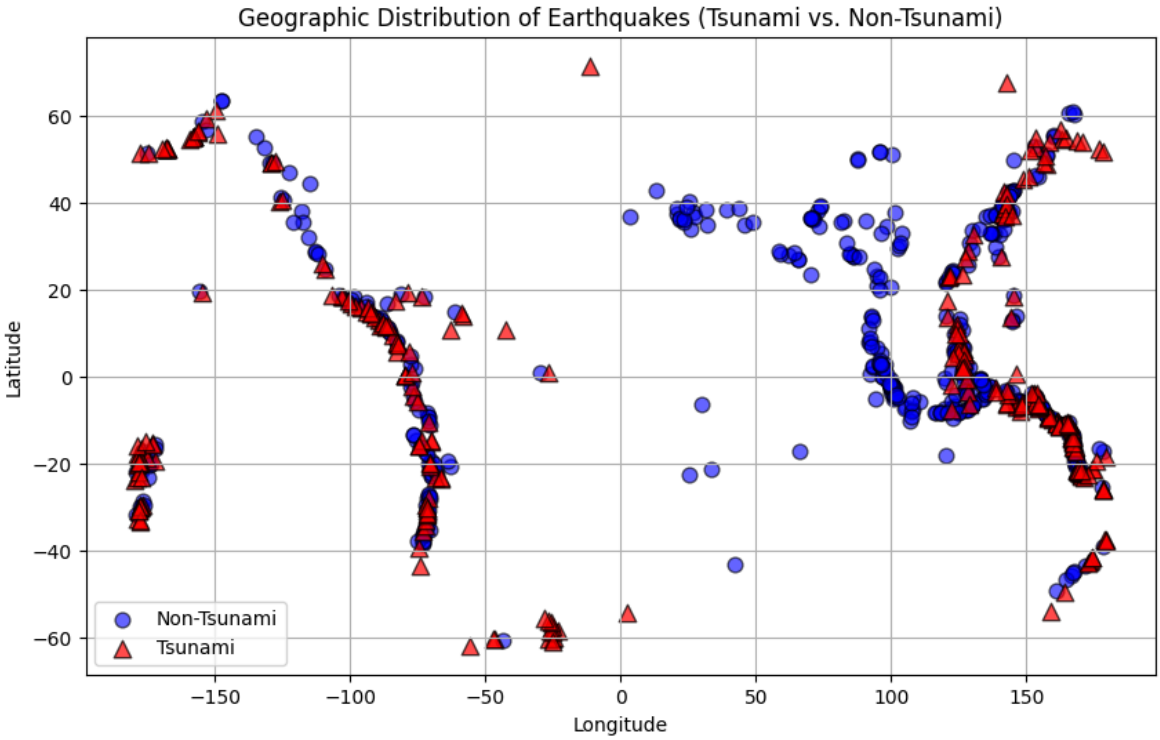
o Plot earthquake locations using latitude and longitude on a 2D scatter plot.

In [58]:
```
plt.figure(figsize=(10, 6))
plt.scatter(data['longitude'], data['latitude'], c=df['magnitude'],cmap='Reds',e

plt.title('Geographic Distribution of Earthquakes')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.colorbar(label='Magnitude')
plt.grid(True)
plt.show()

#Longitude (X-axis) → East-West position.
#Latitude (Y-axis) → North-South position.
```



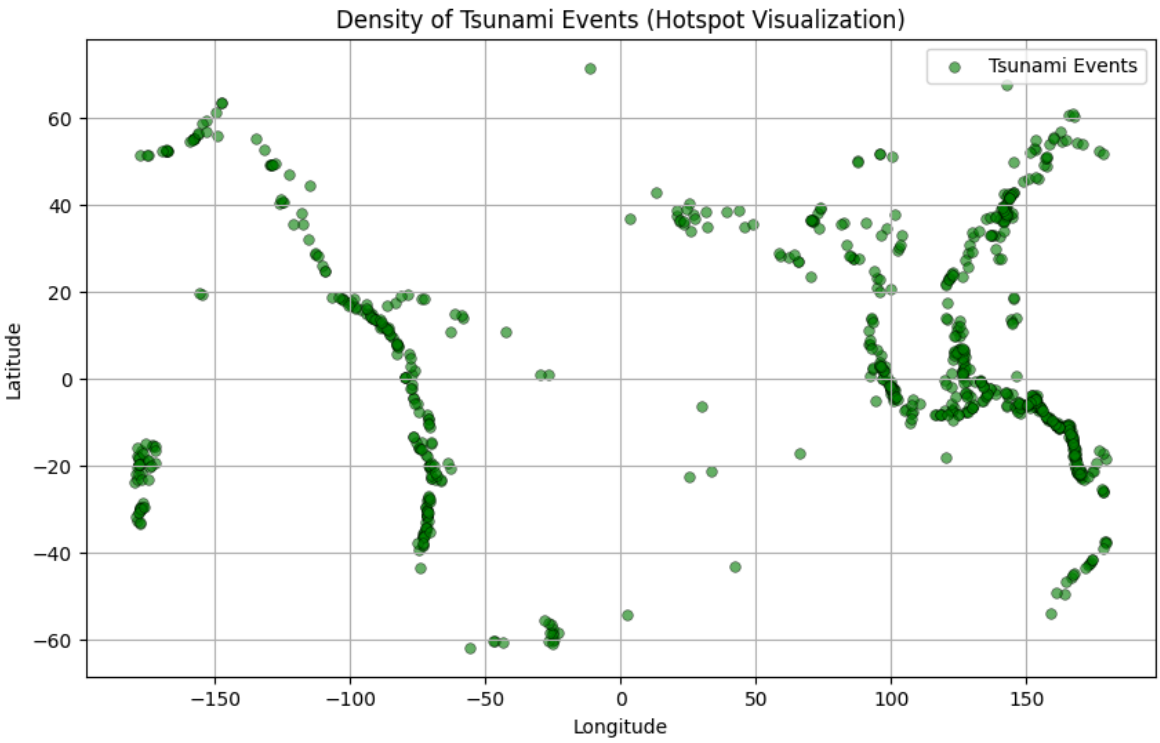
o Visually distinguish between tsunami and nontsunami events.

In [27]:
```
df = pd.DataFrame(data)
plt.figure(figsize=(10, 6))

# Non-Tsunami events
plt.scatter(df[data['tsunami'] == 0]['longitude'],
            df[data['tsunami'] == 0]['latitude'],
            c='blue', label='Non-Tsunami', alpha=0.6, edgecolors='k', s=60)

# Tsunami events
plt.scatter(df[data['tsunami'] == 1]['longitude'],
            df[data['tsunami'] == 1]['latitude'],
            c='red', label='Tsunami', alpha=0.7, edgecolors='k', s=80, marker='^

plt.title('Geographic Distribution of Earthquakes (Tsunami vs. Non-Tsunami)')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend()
plt.grid(True)
plt.show()
```



o Identify clusters or regions with higher concentration of tsunami events (without using map tiles or interactive maps).

In [60]:
```
plt.figure(figsize=(10,6))
sns.scatterplot(
    x=data['longitude'],
    y=data['latitude'],
    color='green', s=30, alpha=0.6, label='Tsunami Events' ,edgecolor='k')

plt.title('Density of Tsunami Events (Hotspot Visualization)')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend()
plt.grid(True)
plt.show()
```

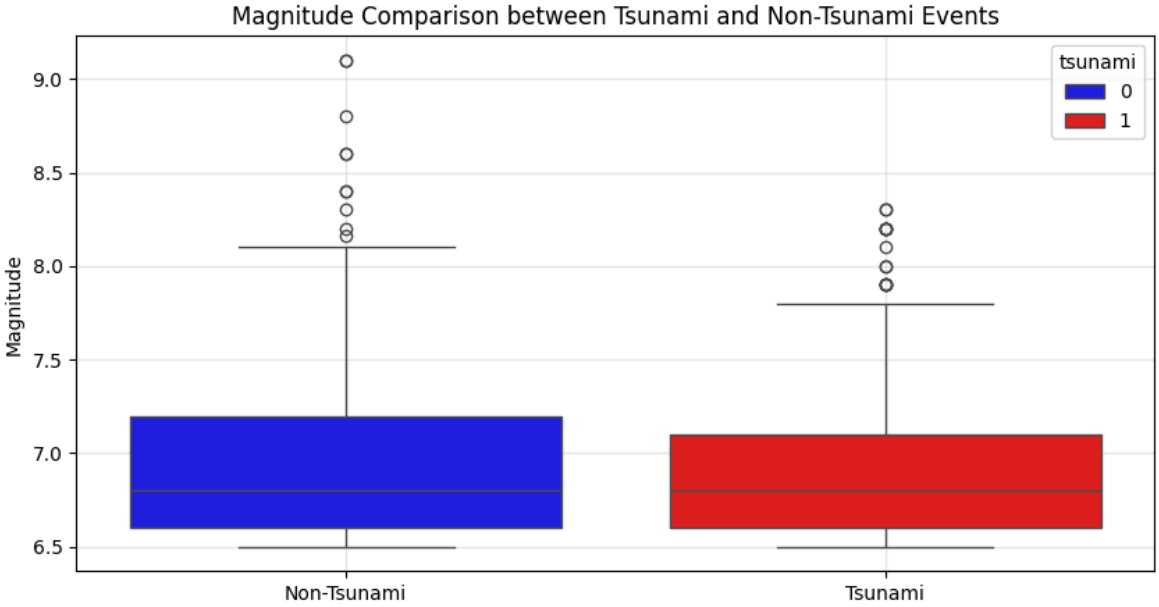


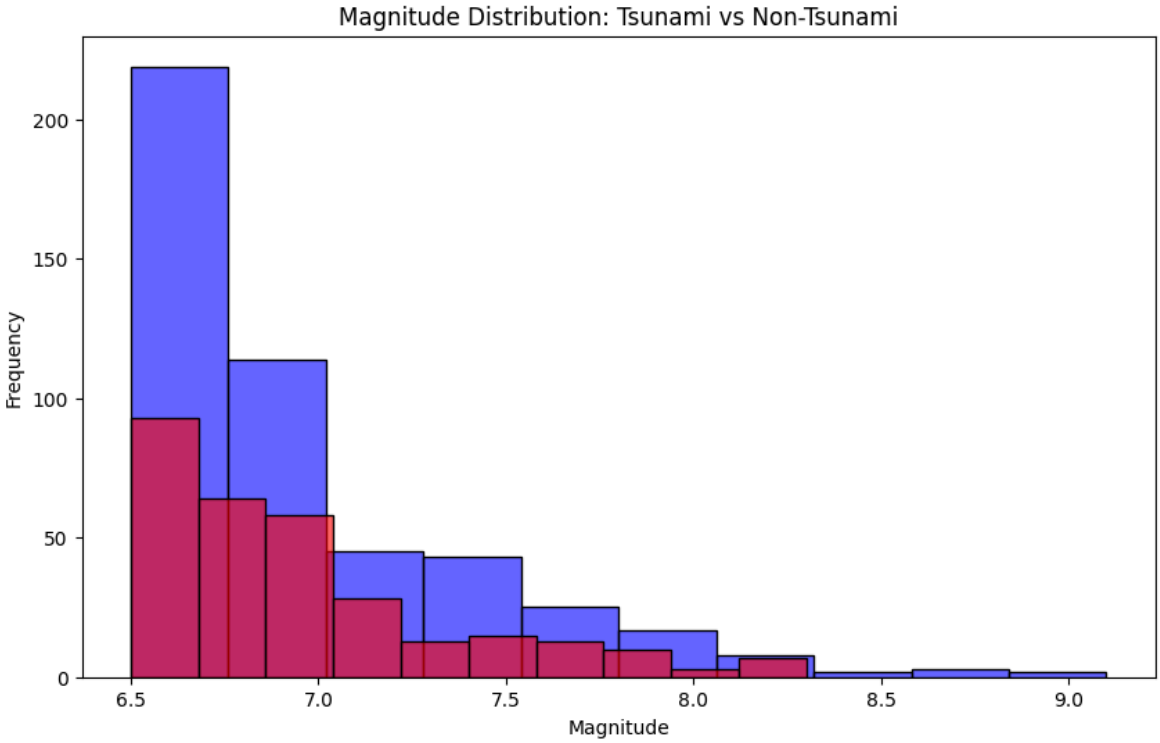
In [ ]:

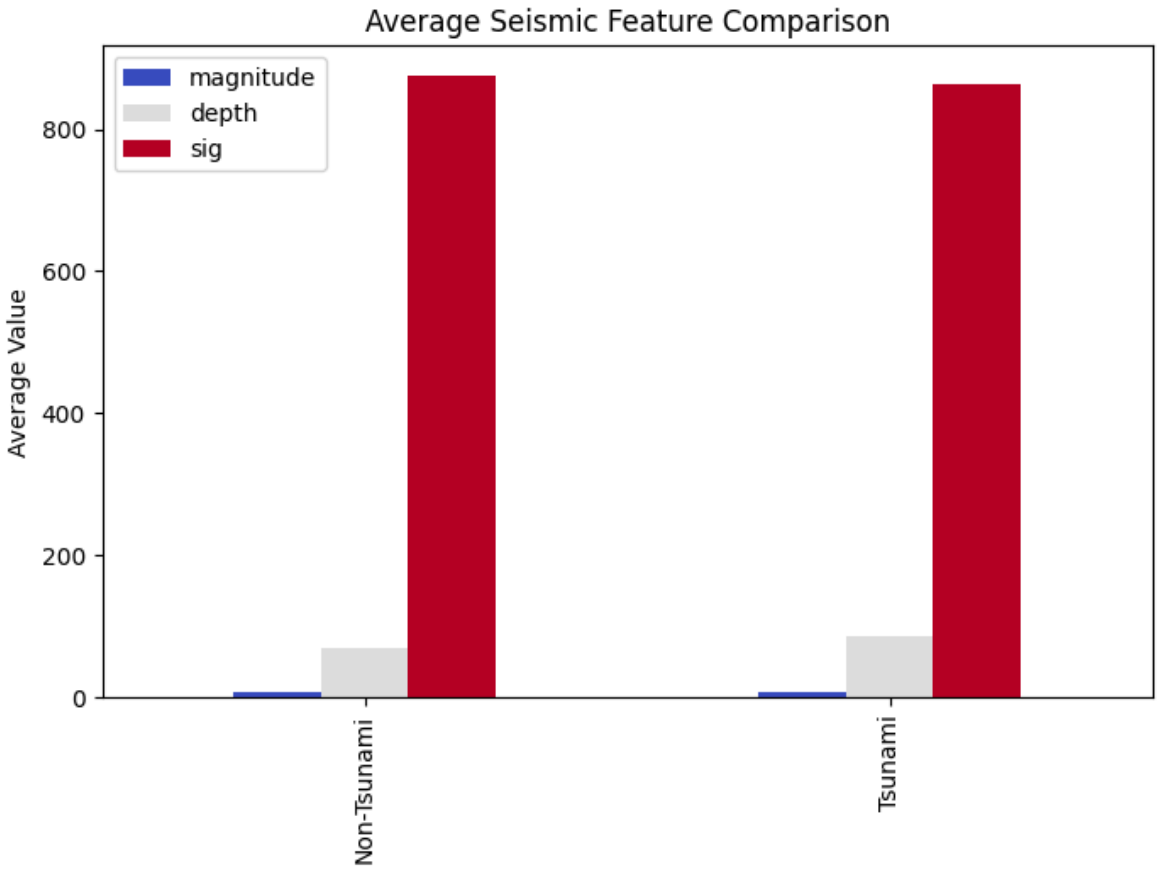
4.Statistical and Comparative Analysis

o Use box plots, histograms, and bar charts to compare seismic features between tsunami and non tsunami events.

In [93]:
```
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='tsunami', y='magnitude',  hue='tsunami',  palette=['blue
plt.xticks([0, 1], ['Non-Tsunami', 'Tsunami'])
plt.title('Magnitude Comparison between Tsunami and Non-Tsunami Events')
plt.xlabel('')
plt.ylabel('Magnitude')
plt.grid(True, alpha=0.3)
plt.show()

plt.figure(figsize=(10,6))
sns.histplot(df[df['tsunami']==0]['magnitude'], bins=10, color='blue', label='No
sns.histplot(df[df['tsunami']==1]['magnitude'], bins=10, color='red', label='Tsu
plt.title('Magnitude Distribution: Tsunami vs Non-Tsunami')
plt.xlabel('Magnitude')
plt.ylabel('Frequency')
plt.show()

avg_values = df.groupby('tsunami')[['magnitude','depth','sig']].mean().reset_ind
avg_values['tsunami'] = avg_values['tsunami'].map({0: 'Non-Tsunami', 1: 'Tsunami

avg_values.plot(x='tsunami', kind='bar', figsize=(8,5), colormap='coolwarm')
plt.title('Average Seismic Feature Comparison')
plt.ylabel('Average Value')
plt.xlabel('')
plt.show()
```

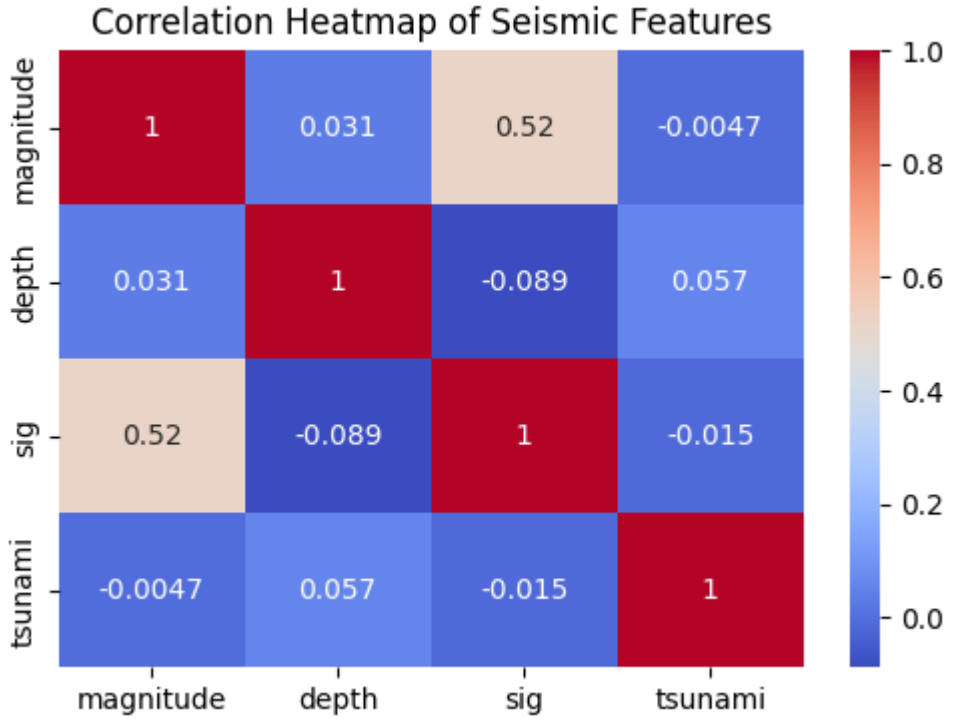

Magnitude Comparison between Tsunami and Non-Tsunami Events
tsunami
0
1
Magnitude
9.0
8.5
8.0
7.5
7.0
6.5
Non-Tsunami
Tsunami


Magnitude Distribution: Tsunami vs Non-Tsunami
Frequency
200
150
100
50
0
6.5
7.0
7.5
8.0
8.5
9.0
Magnitude



o Analyze correlations between variables using heatmaps.

In [12]:
```
plt.figure(figsize=(6,4))
corr = data[['magnitude', 'depth', 'sig', 'tsunami']].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap of Seismic Features')
plt.show()
```



5. Insight and observations

In [ ]:

```
🌍 Earthquake & Tsunami Analysis (2001–2022)
🔎 Project Summary

Analyzed global earthquake data (2001–2022) to explore trends in magnitude, dept
The project highlights major earthquakes (≥8.0) and examines their relationship

📊 Key Insights

Stable Magnitude Trend: Average earthquake magnitudes stayed between 6.8–7.1 ove

Shallow Earthquakes Dominate: Most events occurred at depths <300 km, causing st

Tsunami Triggers: High-magnitude, shallow, and coastal earthquakes have a higher

Major Quakes (≥8.0): Rare but catastrophic — mainly in the Pacific Ring of Fire

No Clear Frequency Rise: The number of earthquakes fluctuated yearly without a s

🧠 Tools & Libraries

Python, Pandas, Matplotlib, Seaborn, Plotly

📁 Includes

Earthquake frequency & magnitude trend

Magnitude vs. depth distribution

Major earthquake highlights

Tsunami vs. non-tsunami comparison
```