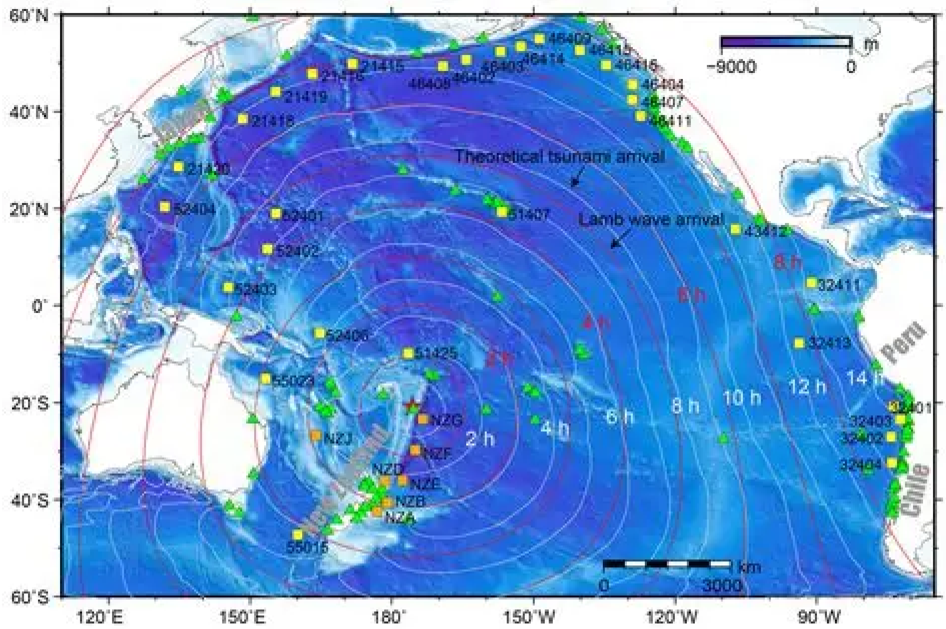
Exploratory Analysis of Global Earthquake-Tsunami Events (2001–2022)



In [1]:

```
# import libraries.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [37]:

```
# Load the dataset.
data=pd.read_csv('earthquake_data_tsunami.csv')
data
```

Out[37]:

	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 [60]:
```
data.head()
```

Out[60]:

	magnitude	cdi	mmi	sig	nst	dmin	gap	depth	latitude	longitude	Year	Mo
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	

In [61]:
```
data.tail()
```

Out[61]:

	magnitude	cdi	mmi	sig	nst	dmin	gap	depth	latitude	longitude	Year	M
777	7.7	0	8	912	427	0.0	0.0	60.0	13.0490	-88.660	2001	
778	6.9	5	7	745	0	0.0	0.0	36.4	56.7744	-153.281	2001	
779	7.1	0	7	776	372	0.0	0.0	103.0	-14.9280	167.170	2001	
780	6.8	0	5	711	64	0.0	0.0	33.0	6.6310	126.899	2001	
781	7.5	0	7	865	324	0.0	0.0	33.0	6.8980	126.579	2001	

In [62]:
```
data.sample(5)
```

Out[62]:

	magnitude	cdi	mmi	sig	nst	dmin	gap	depth	latitude	longitude	Year
141	6.8	3	3	715	0	1.181	16.0	570.41	-8.1440	-71.5870	2019
694	6.7	0	7	691	256	0.000	33.3	25.70	-3.6650	135.3390	2004
425	6.5	6	6	745	691	0.000	34.4	17.00	10.0860	-85.2980	2012
212	6.5	7	8	1297	0	2.360	15.0	29.00	-22.6784	25.1558	2017
554	6.6	0	6	670	463	0.000	14.1	96.00	1.8850	127.3630	2008

In [63]:

```
data.columns
```

Out[63]:

```
Index(['magnitude', 'cdi', 'mmi', 'sig', 'nst', 'dmin', 'gap', 'depth',
       'latitude', 'longitude', 'Year', 'Month', 'tsunami'],
      dtype='object')
```

In [64]:

```
print(type(data))
```

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

In [65]:

```
data.shape
```

Out[65]:

```
(782, 13)
```

In [66]:

```
data.info
```

Out[66]:

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

     Year  Month  tsunami
0    2022     11        1
1    2022     11        0
2    2022     11        1
3    2022     11        1
4    2022     11        1
..    ...    ...      ...
777  2001      1        0
778  2001      1        0
779  2001      1        0
780  2001      1        0
781  2001      1        0

[782 rows x 13 columns]>
```

Dealing With Duplicate Values.

In [67]:
```
data.duplicated()
```

Out[67]:
```
0      False
1      False
2      False
3      False
4      False
       ...
777    False
778    False
779    False
780    False
781    False
Length: 782, dtype: bool
```

In [68]:
```
data.duplicated().sum()
```

Out[68]:
```
np.int64(0)
```

In [69]:
```
# particular columns duplicted data
data['magnitude'].duplicated()
```

Out[69]:
```
0      False
1      False
2       True
3      False
4      False
       ...
777     True
778     True
779     True
780     True
781     True
Name: magnitude, Length: 782, dtype: bool
```

Dealing with Null Values.

In [70]:
```
data.isnull()
```

Out[70]:

	magnitude	cdi	mmi	sig	nst	dmin	gap	depth	latitude	longitude	Yea
0	False	False	False	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	...	...	..
777	False	False	False	False	False	False	False	False	False	False	Fals
778	False	False	False	False	False	False	False	False	False	False	False
779	False	False	False	False	False	False	False	False	False	False	Fals
780	False	False	False	False	False	False	False	False	False	False	Fals
781	False	False	False	False	False	False	False	False	False	False	Fals

782 rows × 13 columns

In [14]:

```
df.isnull().sum()
```

Out[14]:

```
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
```

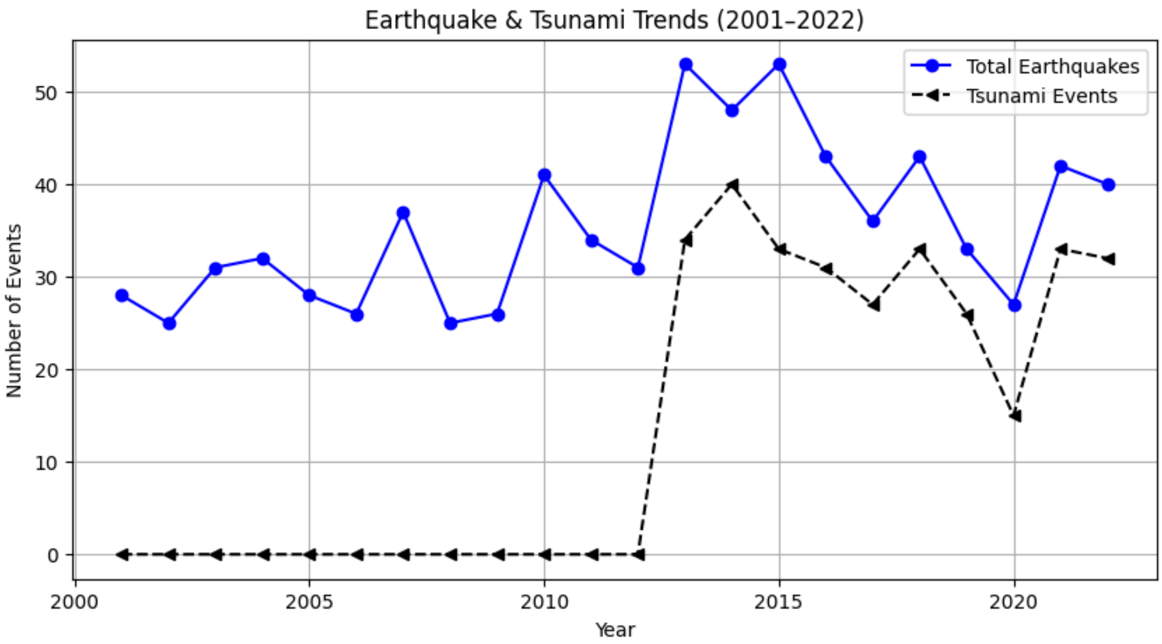
1. Time-Based Analysis:

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

In [46]:

```
# 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,
```

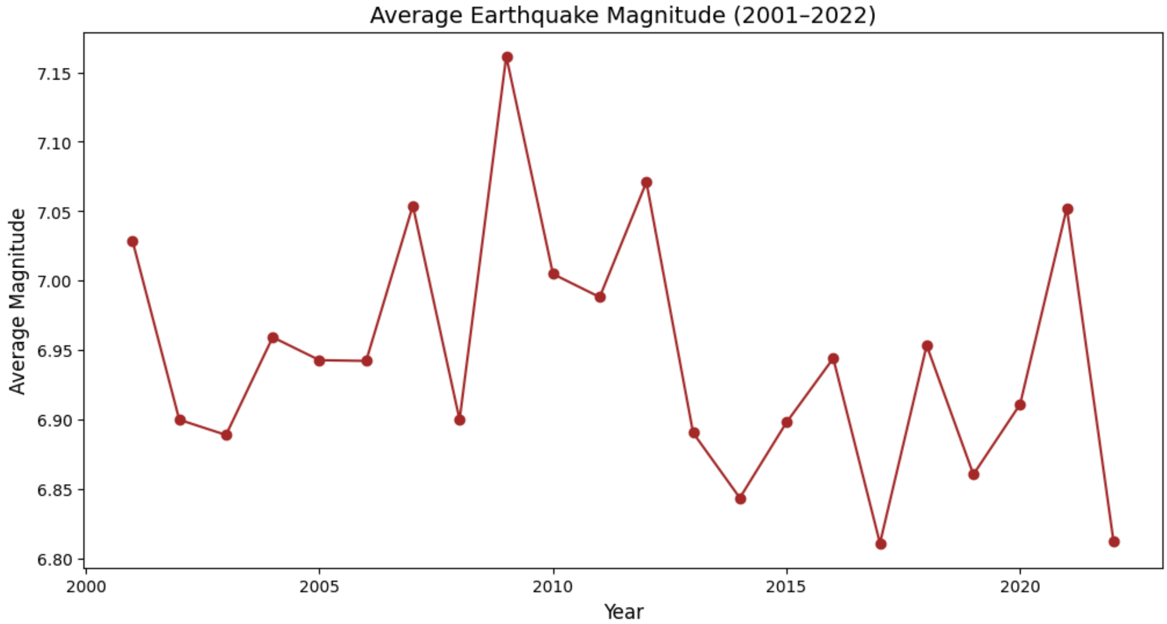
```
        marker='o', linestyle='-', color='blue', label='Total Earthquakes')

plt.plot(tsunami_count.index, tsunami_count.values,
         marker='<', linestyle='--', color='black', 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()
```



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

In [48]:

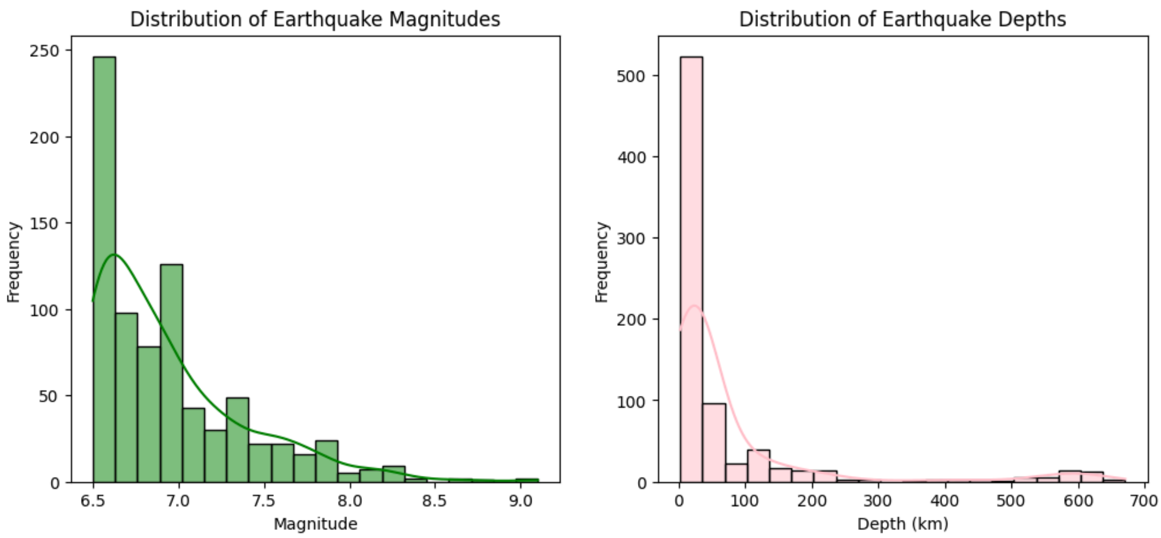
```
# Group by Year to calculate average magnitude
yearly_magnitude = data.groupby("Year")["magnitude"].mean().reset_index()

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

# 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:

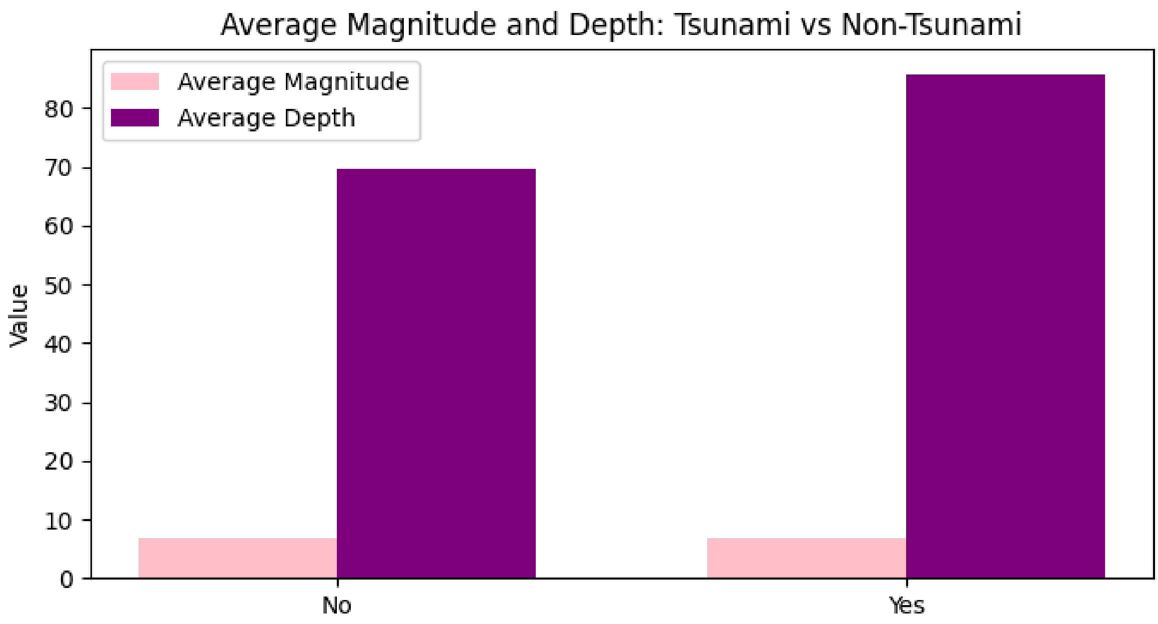
1) Analyze the distribution of earthquake magnitudes and depths.

In [51]:
```python
# 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='green')
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='pink')
plt.title('Distribution of Earthquake Depths')
plt.xlabel('Depth (km)')
plt.ylabel('Frequency')

plt.show()
```



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

In [58]:
```
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()
```



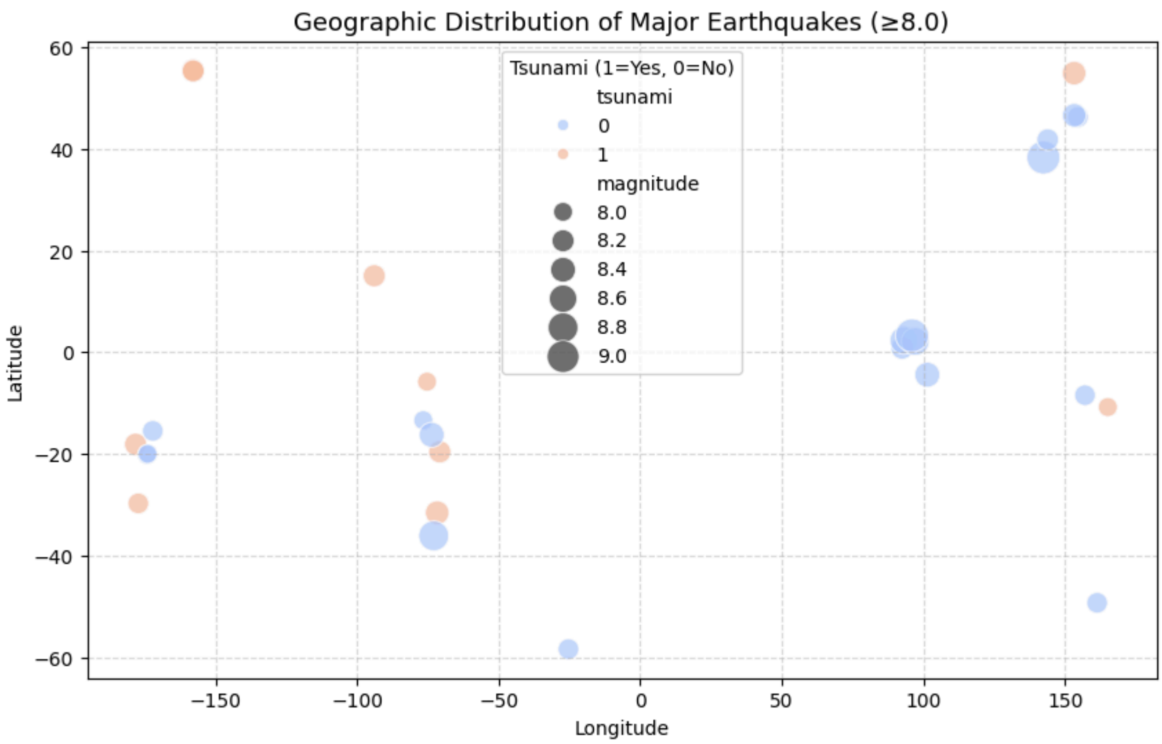
3) Highlight major earthquakes (≥8.0) and their characteristics.

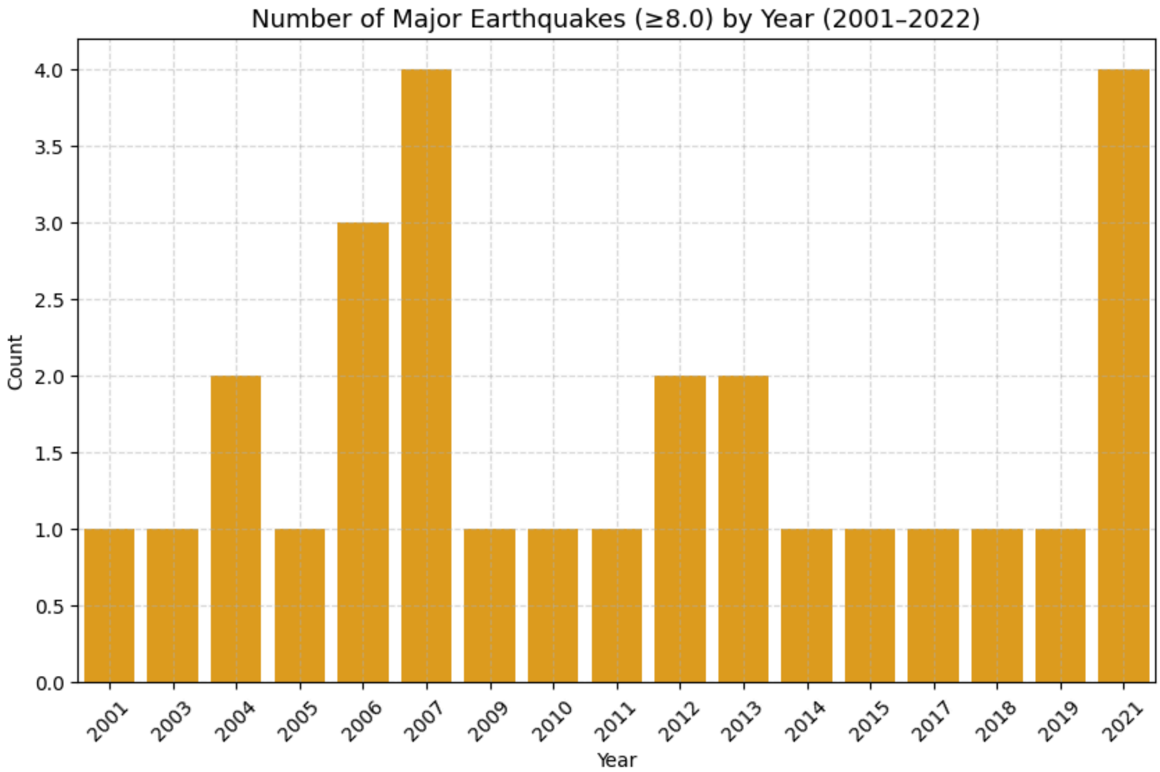
In [73]:

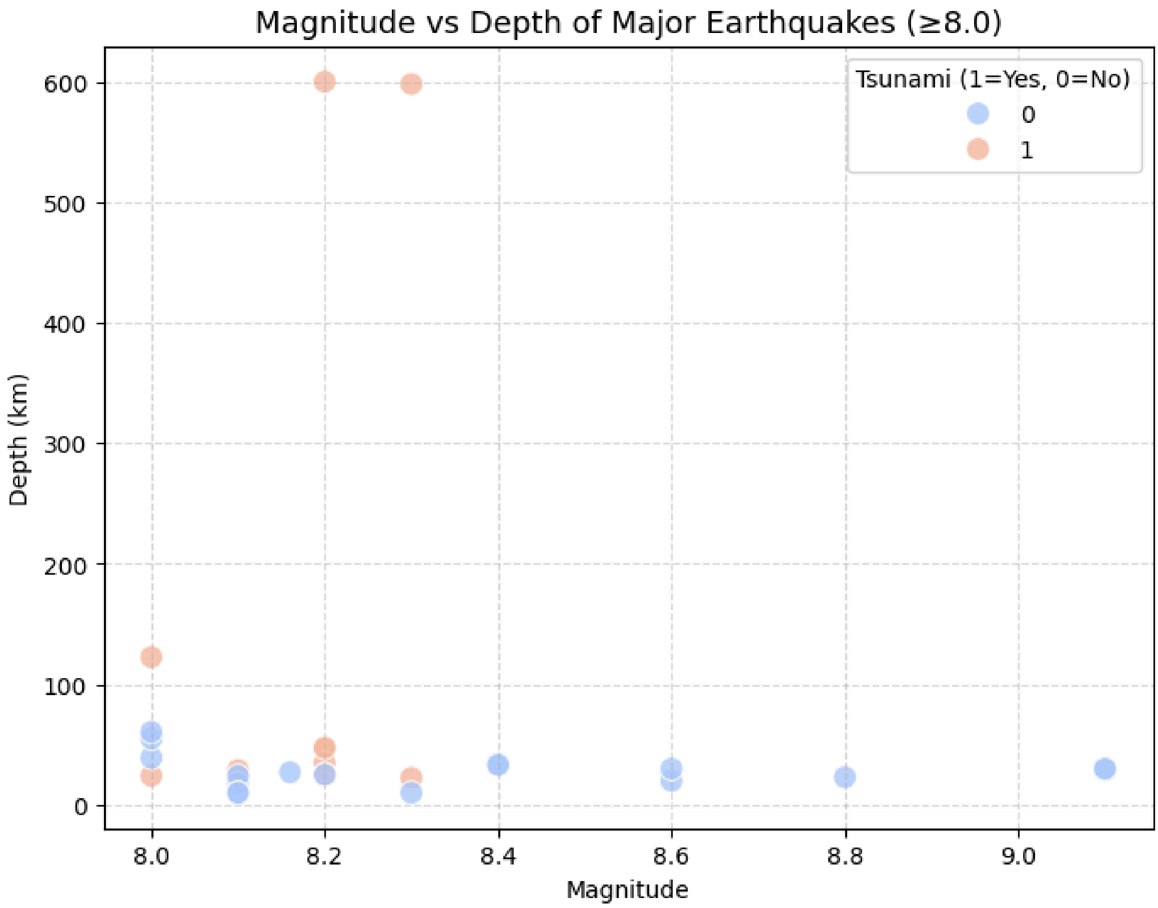
```
# Convert time column to datetime and extract year if not already present
if "Year" not in df.columns:
    df["time"] = pd.to_datetime(df["time"], errors="coerce")
    df["Year"] = df["time"].dt.year

# 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)
```

```python
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
```



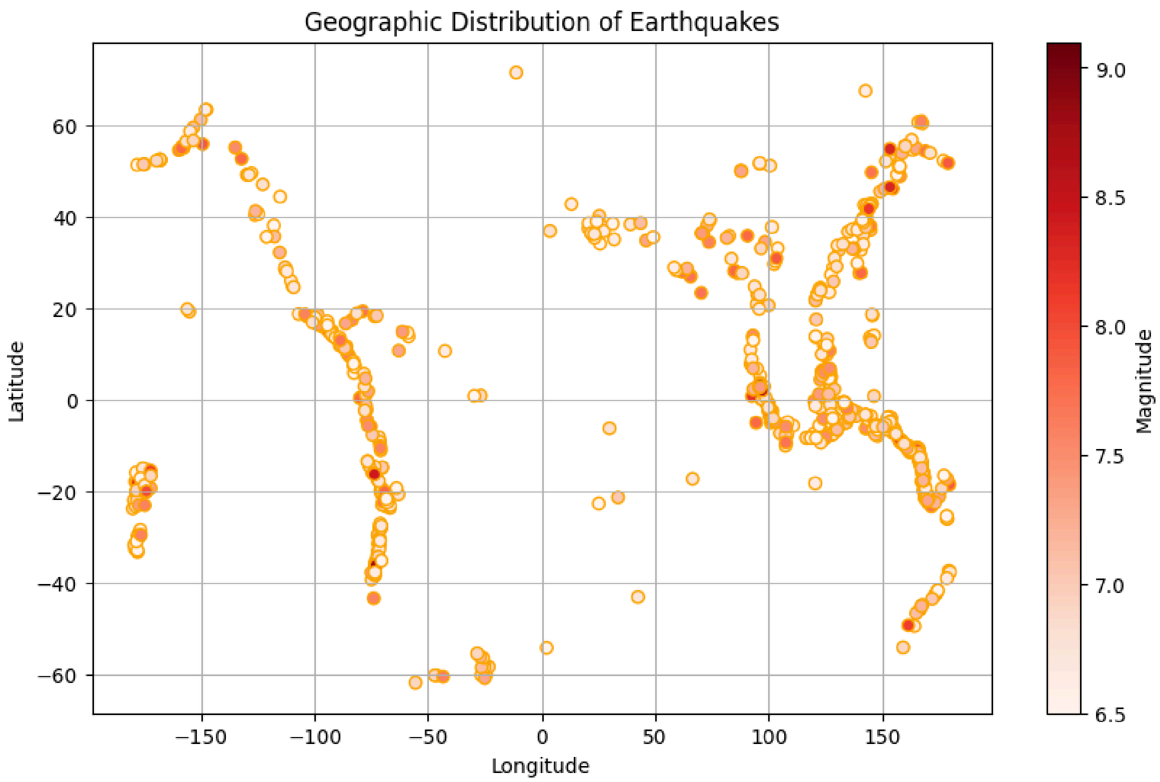




3. Geographic Distribution Using 2D Plotting:

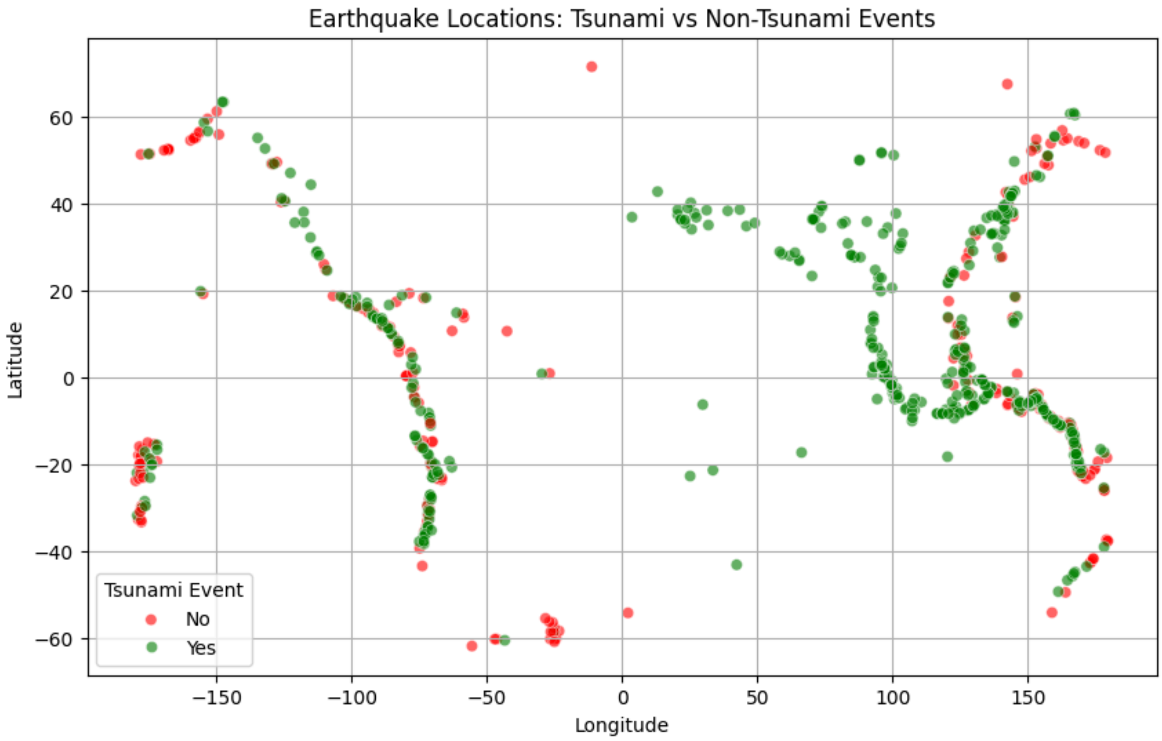
1) Plot earthquake locations using latitude and longitude on a 2D scatter plot.

In [80]:
```
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()
```



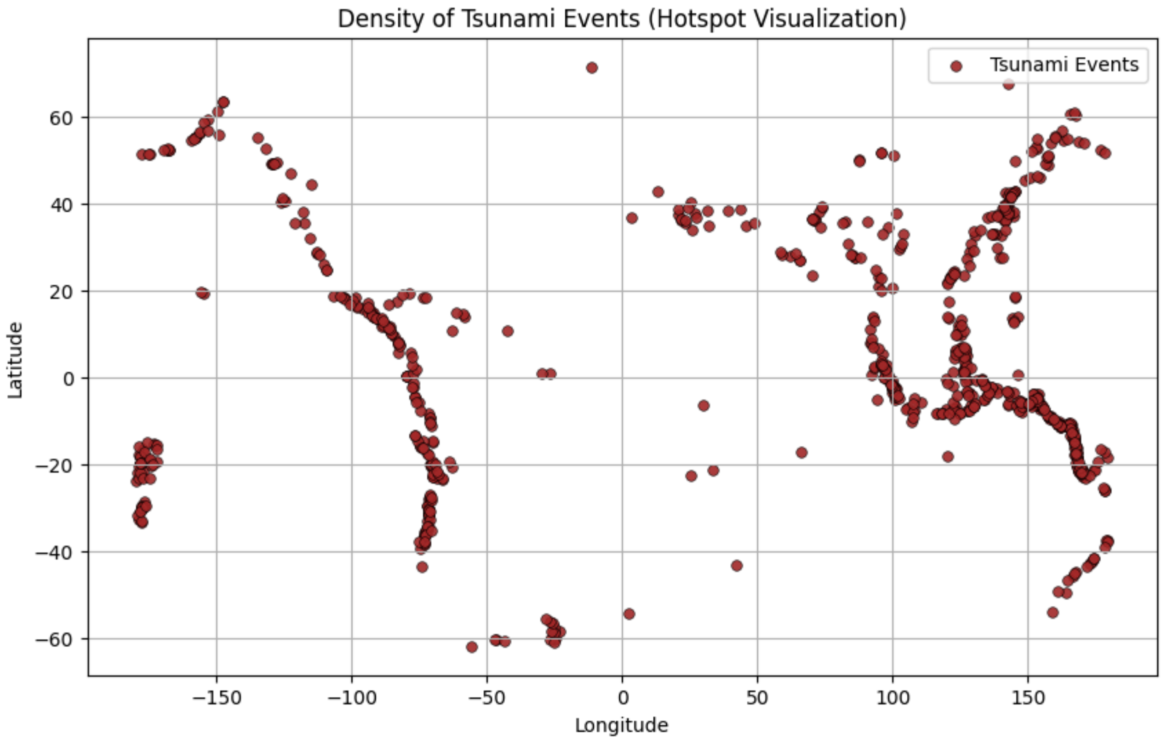
2) Compare the average magnitude and depth of tsunami vs. non-tsunami events.

In [82]:
```
plt.figure(figsize=(10,6))
sns.scatterplot(x='longitude',y='latitude',hue='tsunami',data=df,palette=['green
plt.title('Earthquake Locations: Tsunami vs Non-Tsunami Events')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.legend(title='Tsunami Event', labels=['No', 'Yes'])
plt.grid(True)
plt.show()
```



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

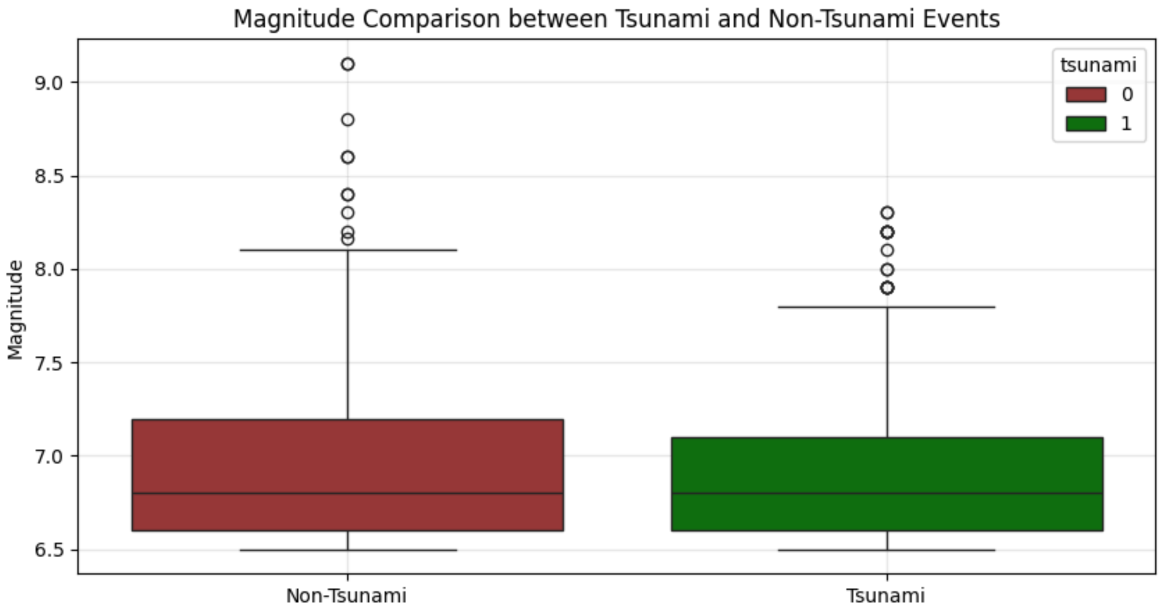
In [34]:
```
plt.figure(figsize=(10,6))
sns.scatterplot(
    x=df['longitude'],
    y=df['latitude'],
    color='brown',s=30, alpha=0.9, label='Tsunami Events' ,edgecolor='k')

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

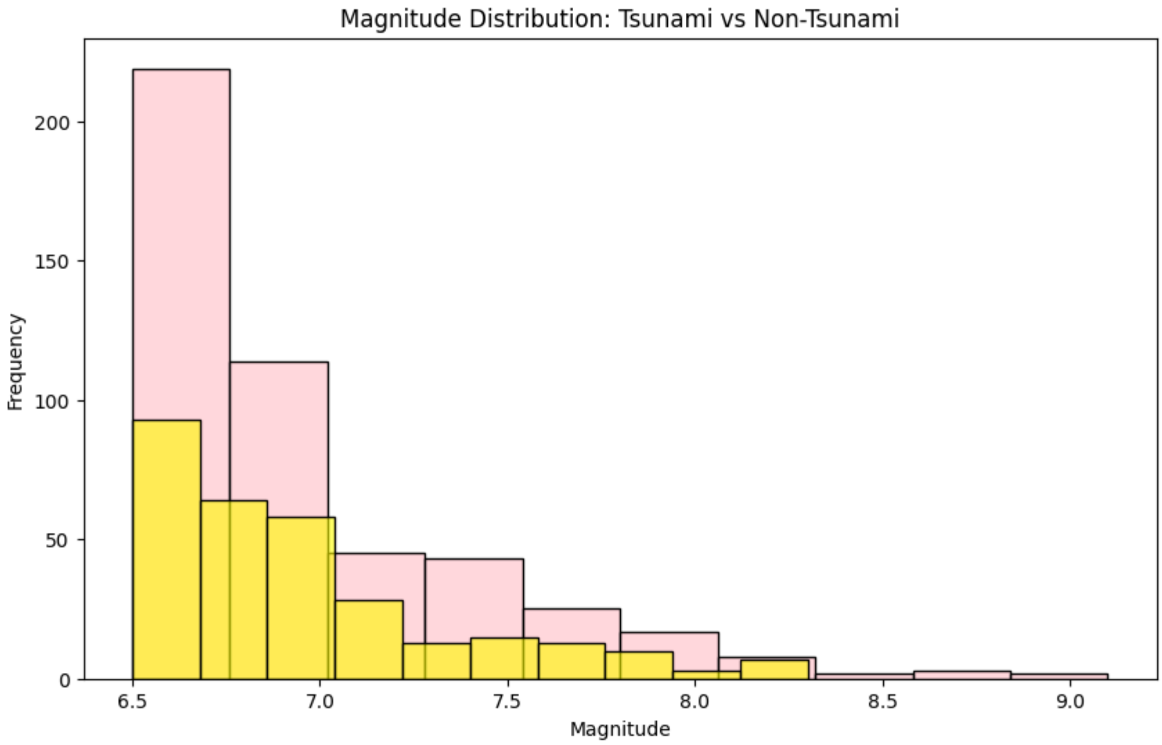


4.Statistical and Comparative Analysis:

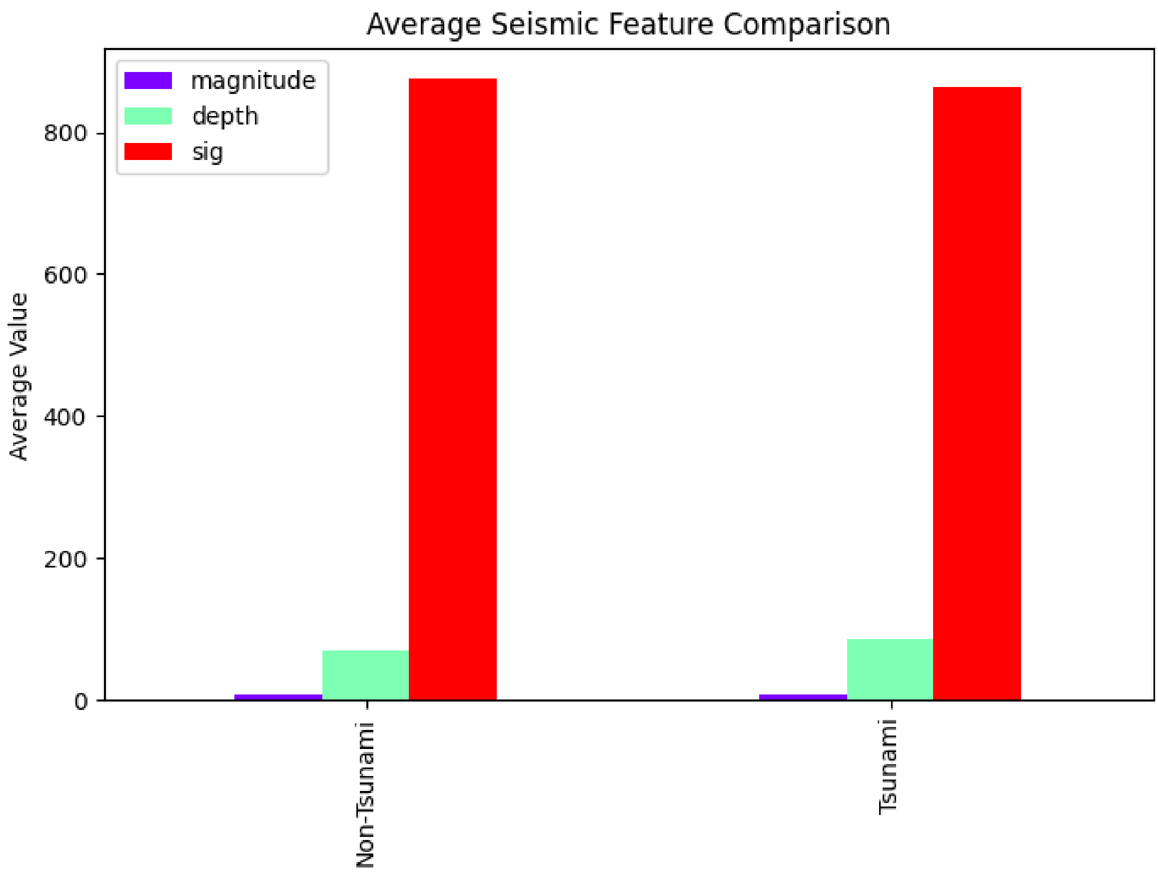
1) Use box plots, histograms, and bar charts to compare seismic features between tsunami and non-tsunami events.

In [43]:
```
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='tsunami',y='magnitude',hue='tsunami',palette=['brown', '
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='pink', label='No
sns.histplot(df[df['tsunami']==1]['magnitude'], bins=10, color='yellow', label='
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='rainbow')
plt.title('Average Seismic Feature Comparison')
plt.ylabel('Average Value')
plt.xlabel('')
```



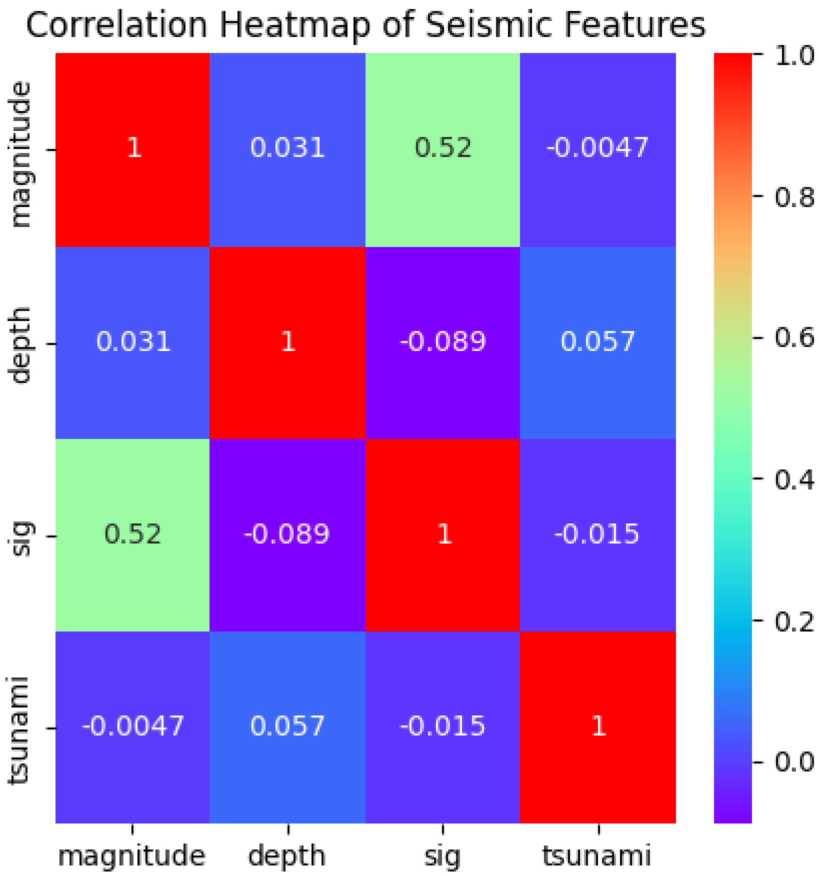



```
Out[43]:  Text(0.5, 0, '')
```



2)Analyze correlations between variables using heatmaps.

In [86]:
```
plt.figure(figsize=(5,5))
corr = data[['magnitude','depth','sig','tsunami']].corr()
sns.heatmap(corr, annot=True, cmap='rainbow')
plt.title('Correlation Heatmap of Seismic Features')
plt.show()
```



5.Insights and Observations:

In [ ]:
```
Earthquake & Tsunami Analysis (2001–2022)
🔎 Project Summary:
This project focuses on analyzing global earthquake records from 2001 to 2022 to
Using Python, statistical summaries and visualizations were created to observe h
The study also draws attention to large earthquakes (magnitude ≥ 8.0) and their

📊 Key Insights (Simplified Explanations)
1.Stable Magnitude Trend:Over the 22-year period, the overall strength of earthq
The average earthquake size stayed roughly between 6.8 and 7.1, showing that glo

2.Shallow Earthquakes Are Common Most earthquakes occur at shallow depths (below
Shallow quakes are generally more damaging because they happen close to the Eart

3.Tsunami-Generating Earthquakes that produce tsunamis are usually stronger in m
Their shallow position under the seabed helps displace large volumes of water, w

4.Major Earthquakes(≥8.0):Very powerful earthquakes (8.0 or more) are rare but h

5. No Increase in Earthquake Frequency

🧠 Tools & Libraries Used:Python,Matplotlib,Seaborn,Pandas

📁 Project Includes
Yearly earthquake and tsunami trend analysis

Magnitude and depth comparisons

Major earthquake (≥8.0) summary
```

```
Tsunami vs non-tsunami characteristics

Geographic earthquake distribution
```