

**VIETNAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF INFORMATION TECHNOLOGY**



Subject: Decision support and business intelligence applications

Topic: Analysis the Global Historical Tsunami Database

Lecturer: Do Phuc

Lecturer: Nguyen Thi Kim Phung

Group: 2

Members:

- Pham Thuy Dung - 20521214
- Dau Dinh Quang Anh - 20521059

COMMENT

Table of Contents

1	TOPIC OVERVIEW	5
2	DICTIONARY OF DATA.....	6
3	DATA WAREHOUSE DESIGN.....	10
4	BUID DATA WAREHOUSE (SSIS).....	11
5	DATA WAREHOUSE ANALYSIS (SSAS).....	14
5.1	CREATE DATA WAREHOUSE.....	16
5.2	CREATE DATA SOURCE VIEW.....	17
5.3	CREATE DATA CUBE.....	18
5.4	EDIT THE DIMENSION.....	19
5.4.1	<i>DimDate</i>	19
5.4.2	<i>DimTime</i>	20
5.4.3	<i>DimLocation</i>	21
5.4.4	<i>DimDescription</i>	22
5.4.5	<i>DimDetailDamage</i>	23
5.4.6	<i>DimTotalDamage</i>	24
5.4.7	<i>Run the Process for Deployment</i>	25
6	QUERY MDX.....	26
6.1	QUERY TSUNAMI YEAR AND DESCRIPTION	26
6.1.1	<i>SQL Query</i>	27
6.1.2	<i>Get SQL result</i>	27
6.2	QUERY NUMBER OF DEATHS AND THEIR DESCRIPTION	27
6.2.1	<i>SQL Query</i>	27
6.2.2	<i>Get SQL Result</i>	28
6.3	QUERY TSUNAMI INTENSITY OVER THE YEAR	28
6.3.1	<i>SQL query</i>	28
6.3.2	<i>Get SQL Result</i>	29
6.4	QUERY TOTAL OF DEATHS OVER THE YEAR WITH COUNTRY	29
6.4.1	<i>SQL Query</i>	29
6.4.2	<i>Get SQL Result</i>	29
6.5	QUERY NUMBER OF HOUSES DAMAGED AND THEIR DESCRIPTIONS	30
6.5.1	<i>SQL Query</i>	30
6.5.2	<i>Get SQL Result</i>	30
6.6	QUERY FOR THE TOTAL NUMBER OF TSUNAMIS BY TIME OF DAY WITH DESCRIPTIONS	31
6.6.1	<i>SQL Query</i>	31
6.6.2	<i>Get SQL Result</i>	31
6.7	QUERY COUNTRIES AND LOCATION HAS TSUNAMI OVER THE MONTHS	32
6.7.1	<i>SQL Query</i>	32
6.7.2	<i>Get SQL Result</i>	32
6.8	TOP 10 COUNTRIES BY EVENT COUNT IN 2020	33
6.8.1	<i>SQL Query</i>	33
6.8.2	<i>Get SQL Result</i>	33
6.9	TOP 3 YEARS BY EVENT COUNT	34
6.9.1	<i>SQL Query</i>	34
6.9.2	<i>Get SQL Result</i>	34
6.10	QUERY FOR THE TOTAL DAMAGE (\$) BY COUNTRY AND YEAR	34
6.10.1	<i>SQL Query</i>	34
6.10.2	<i>Get SQL Result</i>	35

7 REPORT DATA (BY USING POWERBI)	35
7.1 SUM OF DEATHS BY YEAR.....	35
7.2 SUM OF TSUNAMI INTENSITY BY YEAR	36
7.3 COUNT OF TOTAL HOUSES DAMAGED BY YEAR.....	36
7.4 SUM OF MAXIMUM WATER HEIGHT (M) BY YEAR.....	37
7.5 SUM OF VOL BY YEAR.....	37
7.6 SUM OF EARTHQUAKE MAGNITUDE BY YEAR.....	38
7.7 SUM OF NUMBER OF RUNUPS AND SUM OF EARTHQUAKE MAGNITUDE BY YEAR.....	38
7.8 COUNT OF TOTAL DAMAGE (\$MIL) BY COUNTRY	39
7.9 SUM OF TSUNAMI INTENSITY IN MONTH BY COUNTRY	39
7.10 COUNT OF TSUNAMI DAYS IN MONTHS.....	40
8 PIVOT TABLE.....	40
8.1 DATASET.....	40
8.2 STATISTICS ON THE NUMBER OF DEATHS AND CAUSES OVER THE YEARS.....	41
8.2.1 <i>Top 10 Number of Tsunami Event Validity per Country.</i>	41
8.2.2 <i>Number of Earthquake magnitude over the year</i>	41
8.2.3 <i>Number of Tsunami Intensity over the month</i>	42
8.2.4 <i>Top 10 of Country has Damage (\$Mil)</i>	43
8.2.5 <i>Sum of Total House damage description.....</i>	43
8.2.6 <i>Count of Country per Tsunami Cause.....</i>	43
8.2.7 <i>Top 10 of Country has Volcano.....</i>	44
8.2.8 <i>Top 10 of Country has total runup highest.....</i>	45
8.2.9 <i>Top 10 of Year has number of deaths highest.....</i>	45
8.2.10 <i>Top 10 of Country has Maximum Water Height.....</i>	46
9 DATA MINING	46
9.1 LOAD DATASET	46
9.2 VISUALIZATION OF DATASET.....	47
9.2.1 <i>Visualization Number of Tsunami per Country</i>	48
9.2.2 <i>Visualization number of Tsunami per Cause</i>	48
9.2.3 <i>Number of Tsunamis per Month (Sorted from High to Low)</i>	49
9.2.4 <i>Number of Earthquakes by Magnitude Range</i>	50
9.2.5 <i>Number of Tsunamis by Maximum Water Height (m) Range</i>	50
9.2.6 <i>Number of Tsunamis by Number of Runups Range</i>	51
9.2.7 <i>Number of Tsunamis by Event Validity</i>	52
9.2.8 <i>Number of Tsunamis by Total Death Description</i>	52
9.2.9 <i>Number of Tsunamis by Total Damage Description</i>	53
9.3 CLUSTERING DATASET BY USING SOME ALGORITHMS.....	53
9.3.1 <i>Clustering by using K-means</i>	53
9.3.2 <i>Clustering by using Agglomerative.....</i>	55
9.3.3 <i>Clustering by using Birch</i>	55
9.3.4 <i>Comparison between alogirthm</i>	56
10 CONCLUSION	56

1 Topic Overview

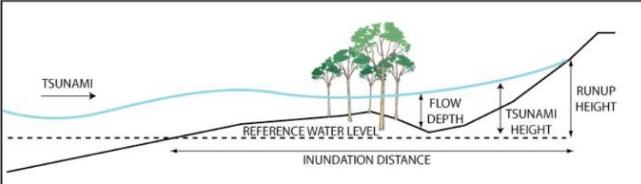
This report conducts research on the causes, consequences, and damage of tsunamis based on the Global Historical Tsunami Database.

This dataset represents historical tsunamis and related information, providing data on over 2,400 tsunamis from 2100 BC to the present in the Atlantic, Indian, and Pacific Oceans, as well as the Mediterranean and Caribbean Seas. This dataset download from [NCEI Global Historical Hazard Database \(noaa.gov\)](#)

2 Dictionary of Data

Attribute	Description
Year	The year in which the event occurred.
Mo (Month)	The month in which the event occurred.
Dy (Day)	The day on which the event occurred.
Hr (Hour)	The hour at which the event occurred.
Mn (Minute)	The minute at which the event occurred.
Sec (Second)	The second at which the event occurred.
Tsunami Event Validity	<p>Valid values: -1 to 4</p> <p>Validity of the actual tsunami occurrence is indicated by a numerical rating of the reports of that event:</p> <ul style="list-style-type: none"> -1: erroneous entry 0: event that only caused a seiche or disturbance in an inland river/lake 1: very doubtful tsunami 2: questionable tsunami 3: probable tsunami 4: definite tsunami
Tsunami Cause Code	<p>Valid values: 0 to 11</p> <p>The source of the tsunami:</p> <ul style="list-style-type: none"> 0: Unknown 1: Earthquake 2: Questionable Earthquake 3: Earthquake and Landslide 4: Volcano and Earthquake 5: Volcano, Earthquake, and Landslide 6: Volcano 7: Volcano and Landslide 8: Landslide 9: Meteorological 10: Explosion 11: Astronomical Tide
Earthquake Magnitude	Valid values: 0.0 to 9.9

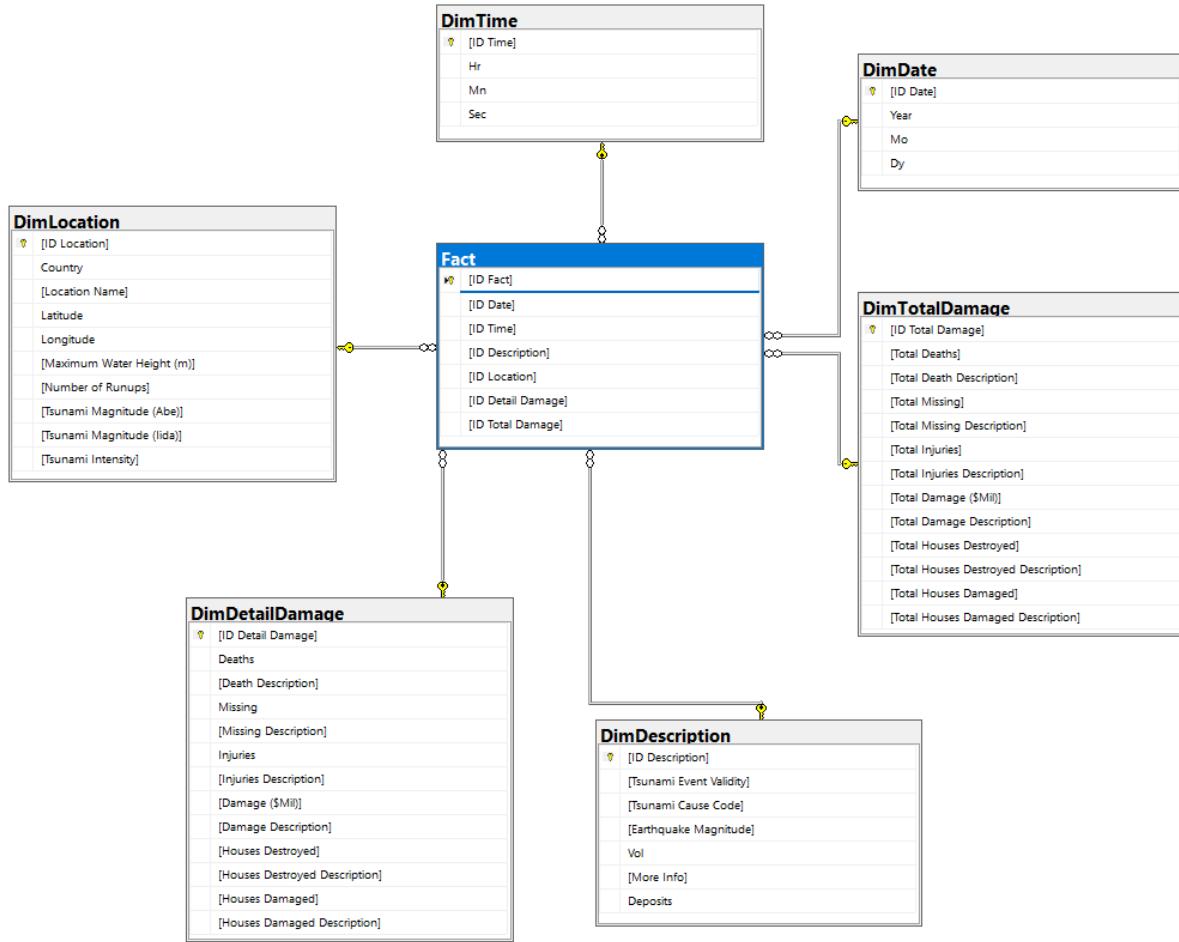
	<p>The value in this column contains the primary earthquake magnitude. Magnitude measures the energy released at the source of the earthquake. Magnitude is determined from measurements on seismographs. For pre-instrumental events, the magnitudes are derived from intensities. There are several different scales for measuring earthquake magnitudes. The primary magnitude is chosen from the available magnitude scales in this order:</p> <ul style="list-style-type: none"> • Mw Magnitude • Ms Magnitude • Mb Magnitude • Ml Magnitude • Mfa Magnitude • Unknown Magnitude
Vol	Volcano identification number if a volcanic eruption caused the tsunami.
More Info	Additional information or notes about the event.
Deposits	Information about deposits left by the tsunami.
Country	The Country where the tsunami source occurred (For example enter: Japan or Russia).
Location Name	The Country, State, Province or Island where the tsunami source occurred (For example enter: Japan or Honshu)
Latitude (kinh độ)	Valid values: -90 to 90 Latitude: 0 to 90 (Northern Hemisphere) -90 to 0 (Southern Hemisphere)
Longitude (vĩ độ)	Valid values: -180 to 180 Longitude: 0 to 180 (Eastern Hemisphere) -180 to 0 (Western Hemisphere)
Maximum Water Height (m)	<p>Maximum Water Height (m)</p> <ul style="list-style-type: none"> • The maximum water height above sea level in meters for this event. If the type of measurement of the runup was a: • Tide Gauge - half of the maximum height (minus the normal tide) of a tsunami wave recorded at the coast by a tide gauge.

	<ul style="list-style-type: none"> • Runup Height - the maximum elevation the wave reaches at the maximum inundation.
Number of Runups	The total number of runups link will display the runup locations associated with a particular tsunami event. 
Tsunami Magnitude (Abe)	Tsunami magnitude scale by Abe.
Tsunami Magnitude (Iida)	Tsunami magnitude scale by Iida.
Tsunami Intensity	Intensity of the tsunami.
Deaths	Whenever possible, numbers of deaths are listed.
Death Description	<p>Valid values: 0 to 4</p> <p>When a description was found in the historical literature instead of an actual number of deaths, this value was coded and listed in the Deaths column. If the actual number of deaths was listed, a descriptor was also added for search purposes.</p> <p>0: None 1: Few (~1 to 50 deaths) 2: Some (~51 to 100 deaths) 3: Many (~101 to 1000 deaths) 4: Very many (over 1000 deaths)</p>
Missing	Number of people missing after the tsunami.
Missing Description	Description of missing people.
Injuries	Whenever possible, numbers of injuries from the tsunami are listed.
Injuries Description	<p>Valid values: 0 to 4</p> <p>When a description was found in the historical literature instead of an actual number of injuries, this value was coded and listed in the Injuries column. If the actual number of injuries was listed, a descriptor was also added for search purposes.</p> <p>0: None 1: Few (~1 to 50 injuries) 2: Some (~51 to 100 injuries) 3: Many (~101 to 1000 injuries) 4: Very many (over 1000 injuries)</p>

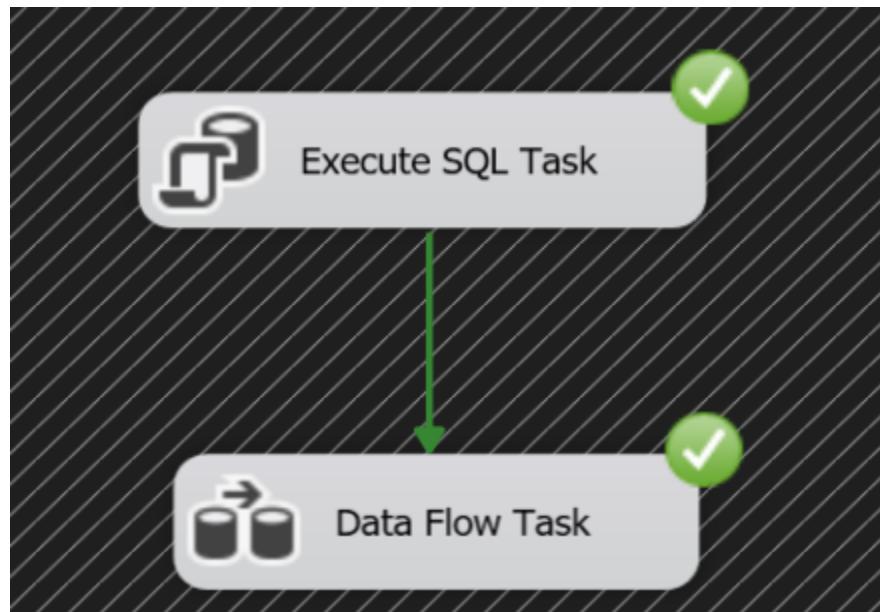
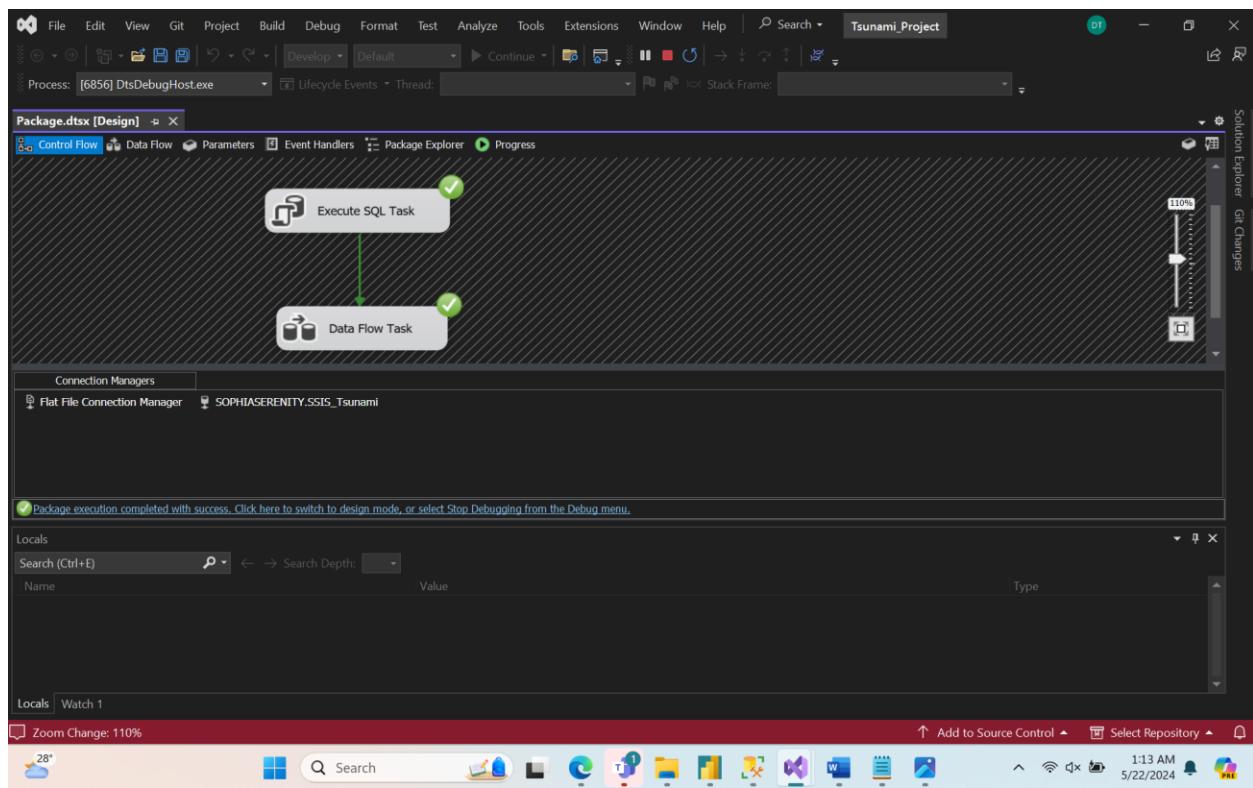
Damage (\$Mil)	The value in the Damage column should be multiplied by 1,000,000 to obtain the actual dollar amount.
Damage Description	<p>Valid values: 0 to 4</p> <p>For those events not offering a monetary evaluation of damage, the following five-level scale was used to classify damage (1990 dollars) and was listed in the Damage column. If the actual dollar amount of damage was listed, a descriptor was also added for search purposes.</p> <p>0: NONE 1: LIMITED (roughly corresponding to less than \$1 million) 2: MODERATE (~\$1 to \$5 million) 3: SEVERE (~\$5 to \$25 million) 4: EXTREME (~\$25 million or more)</p>
Houses Destroyed	Whenever possible, numbers of houses destroyed are listed.
Houses Destroyed Description	<p>Valid values: 0 to 4</p> <p>For those events not offering an exact number of houses damaged, the following four-level scale was used to classify the damage and was listed in the Houses Destroyed column. If the actual number of houses destroyed was listed, a descriptor was also added for search purposes.</p> <p>0: None 1: Few (~1 to 50 houses) 2: Some (~51 to 100 houses) 3: Many (101 to 1000 houses) 4: Very Many (~over 1000 houses)</p>
Houses Damaged	Number of houses damaged.
Houses Damaged Description	Description of the damage to houses.
Total Deaths	Whenever possible, total number of deaths from the tsunami and secondary effects are listed.
Total Death Description	<p>Valid values: 0 to 4</p> <p>When a description was found in the historical literature instead of an actual number of deaths, this value was coded and listed in the Deaths column. If the actual number of deaths was listed, a descriptor was also added for search purposes.</p>

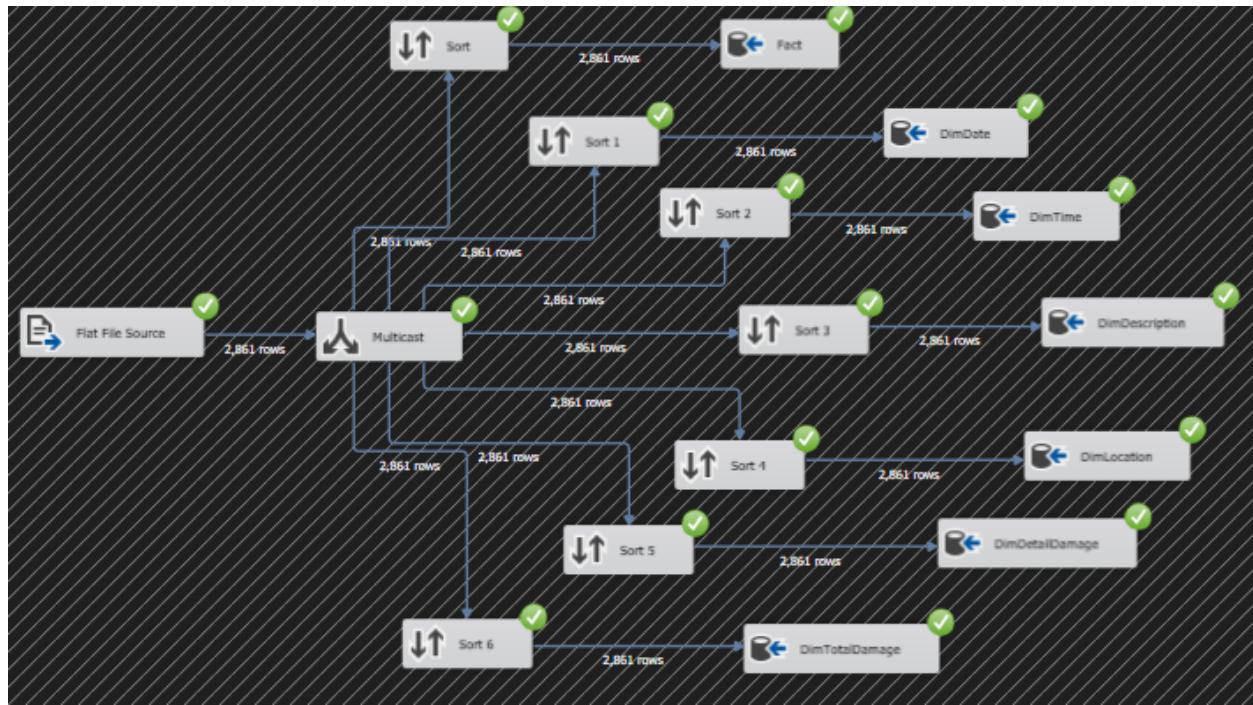
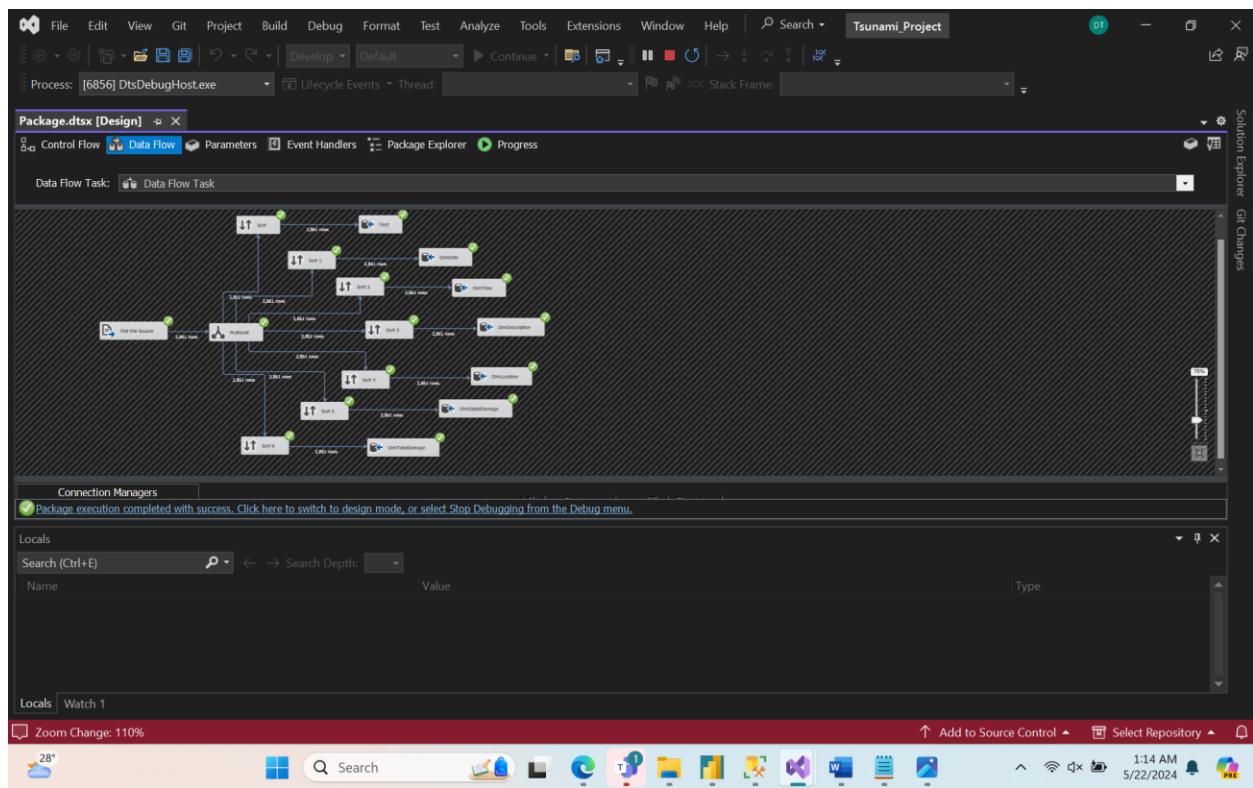
	0: None 1: Few (~1 to 50 deaths) 2: Some (~51 to 100 deaths) 3: Many (~101 to 1000 deaths) 4: Very many (over 1000 deaths)
Total Missing	Total number of missing including secondary effects.
Total Missing Description	Description of total missing.
Total Injuries	Total number of injuries including secondary effects.
Total Injuries Description	Description of total injuries.
Total Damage (\$Mil)	Total Damage in Millions of Dollars from the Tsunami and the Source Event
Total Damage Description	<p>Valid values: 0 to 4</p> <p>For those events not offering a monetary evaluation of damage, the following five-level scale was used to classify damage (1990 dollars) and was listed in the Damage column. If the actual dollar amount of damage was listed, a descriptor was also added for search purposes.</p> <p>0: NONE 1: LIMITED (roughly corresponding to less than \$1 million) 2: MODERATE (~\$1 to \$5 million) 3: SEVERE (~\$5 to \$25 million) 4: EXTREME (~\$25 million or more)</p>
Total Houses Destroyed	Total number of houses destroyed.
Total Houses Destroyed Description	Description of total houses destroyed.
Total Houses Damaged	Total number of houses damaged.
Total Houses Damaged Description	Description of total houses damaged.

3 Data Warehouse Design



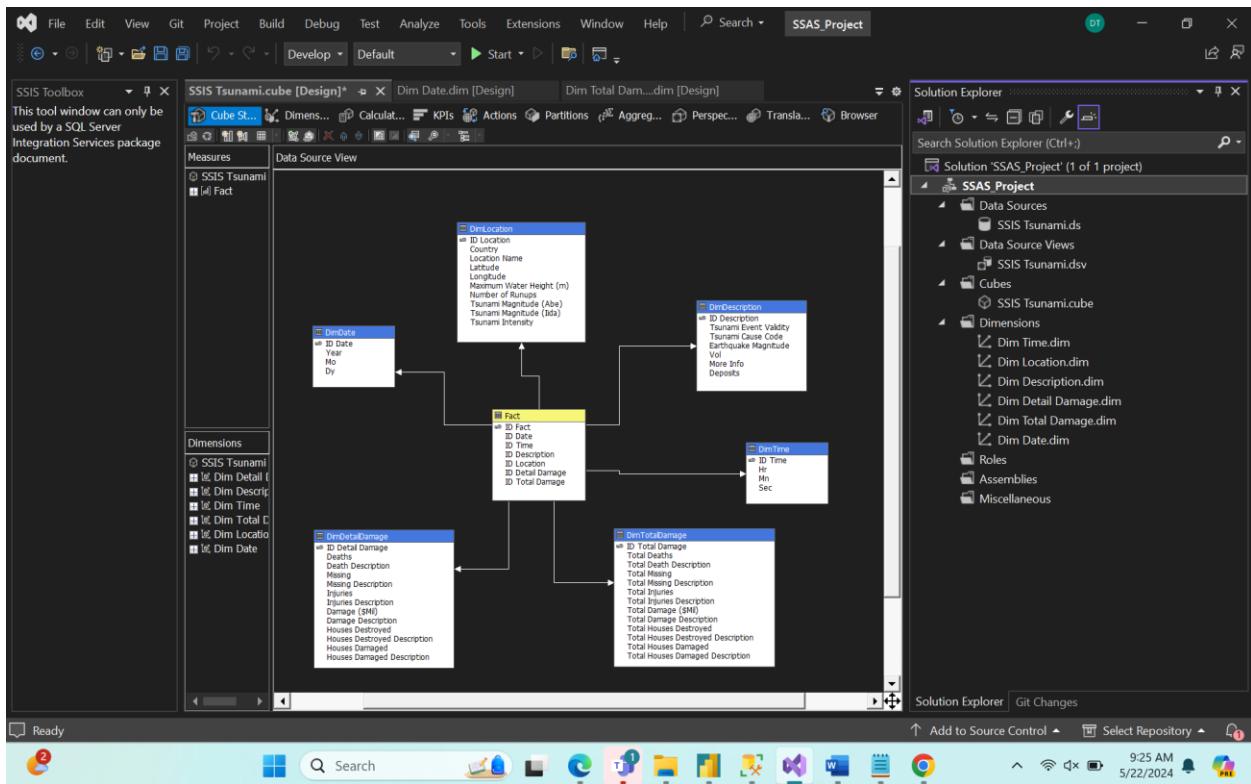
4 Build Data Warehouse (SSIS)

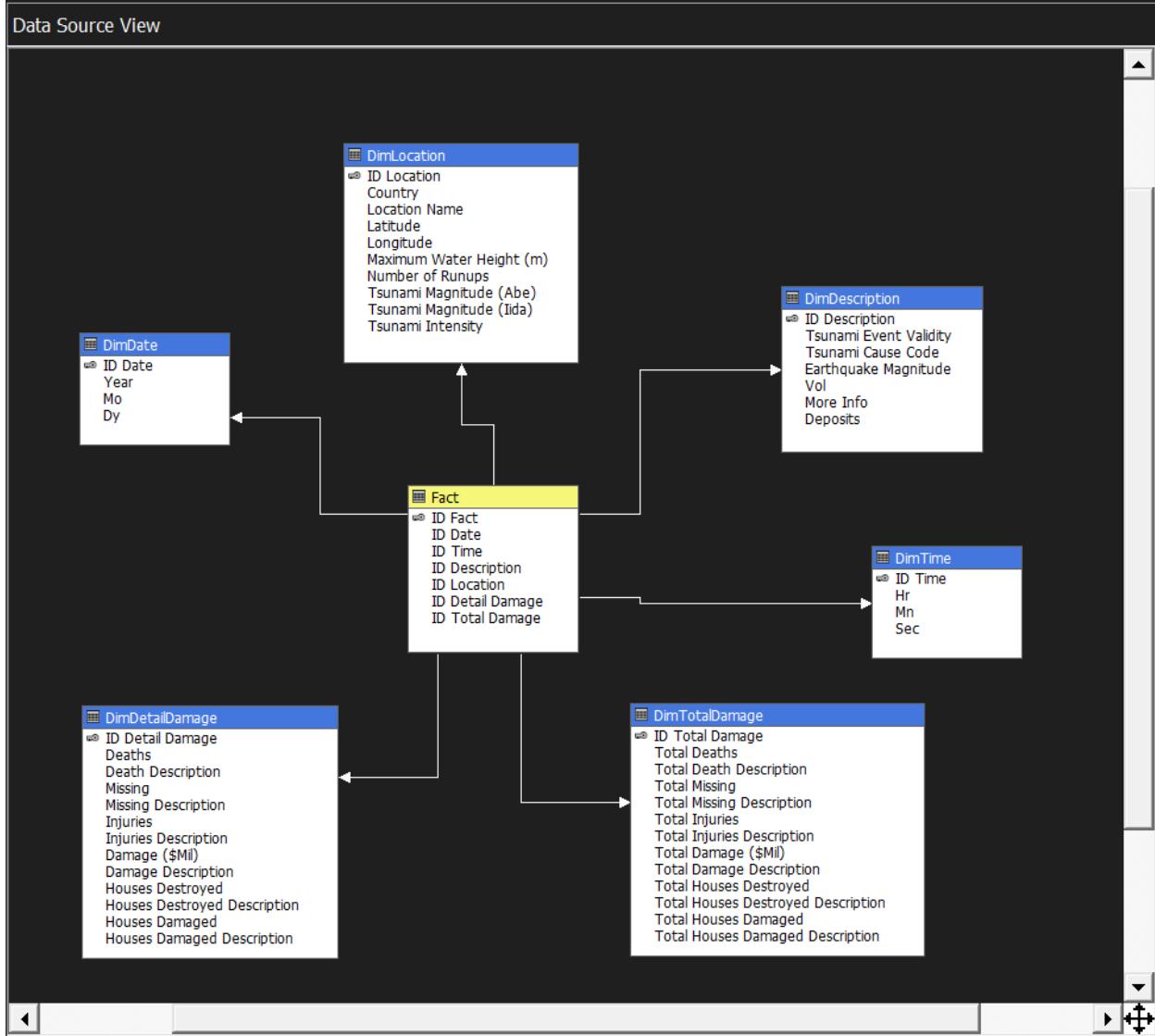


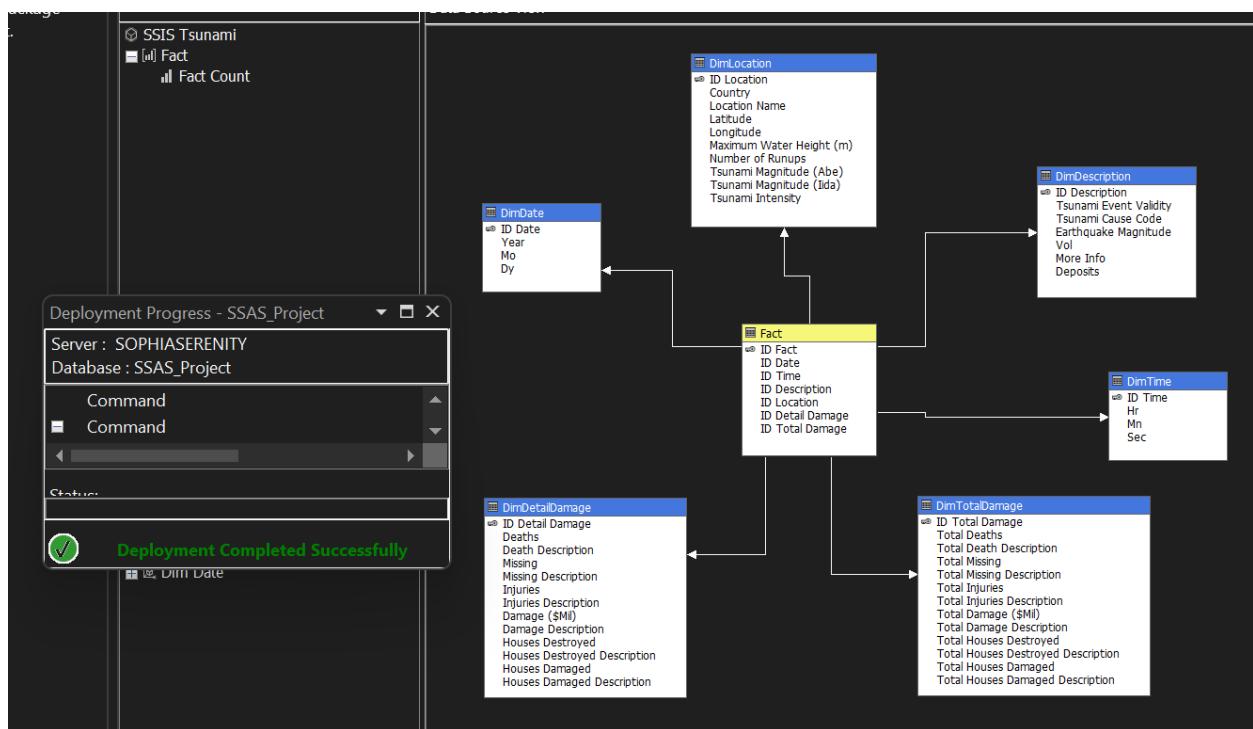
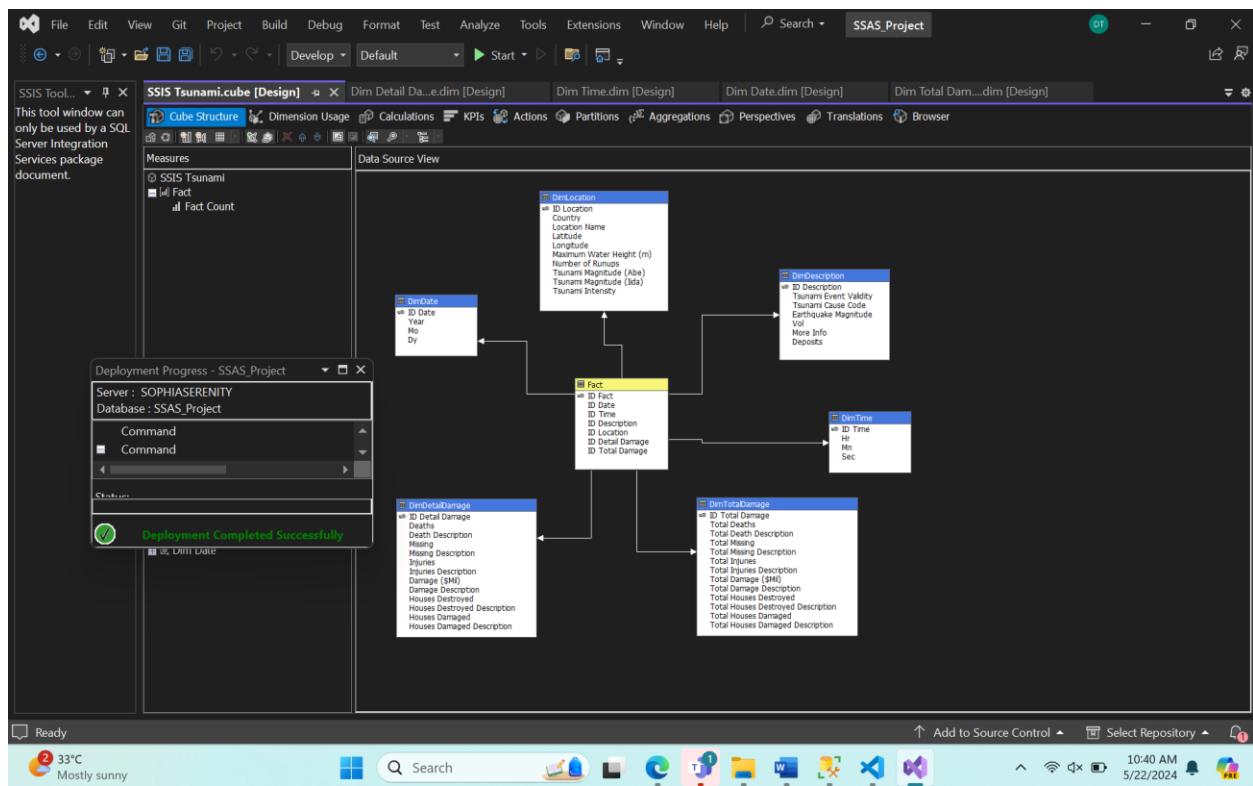


The source data used for analysis is a csv file, so we use flat file sources to integrate the data, use sort to remove and order the lines.

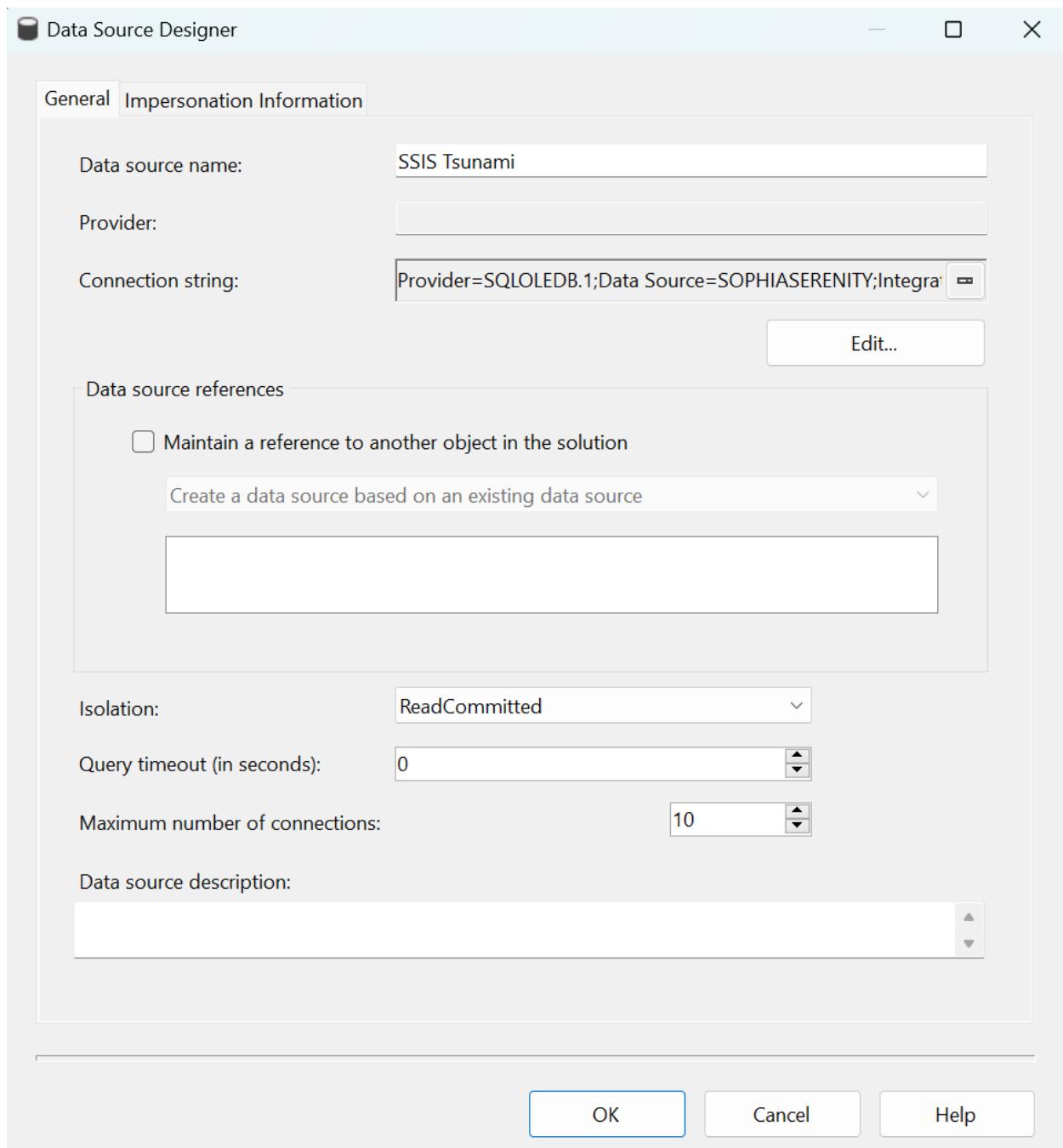
5 Data Warehouse Analysis (SSAS)



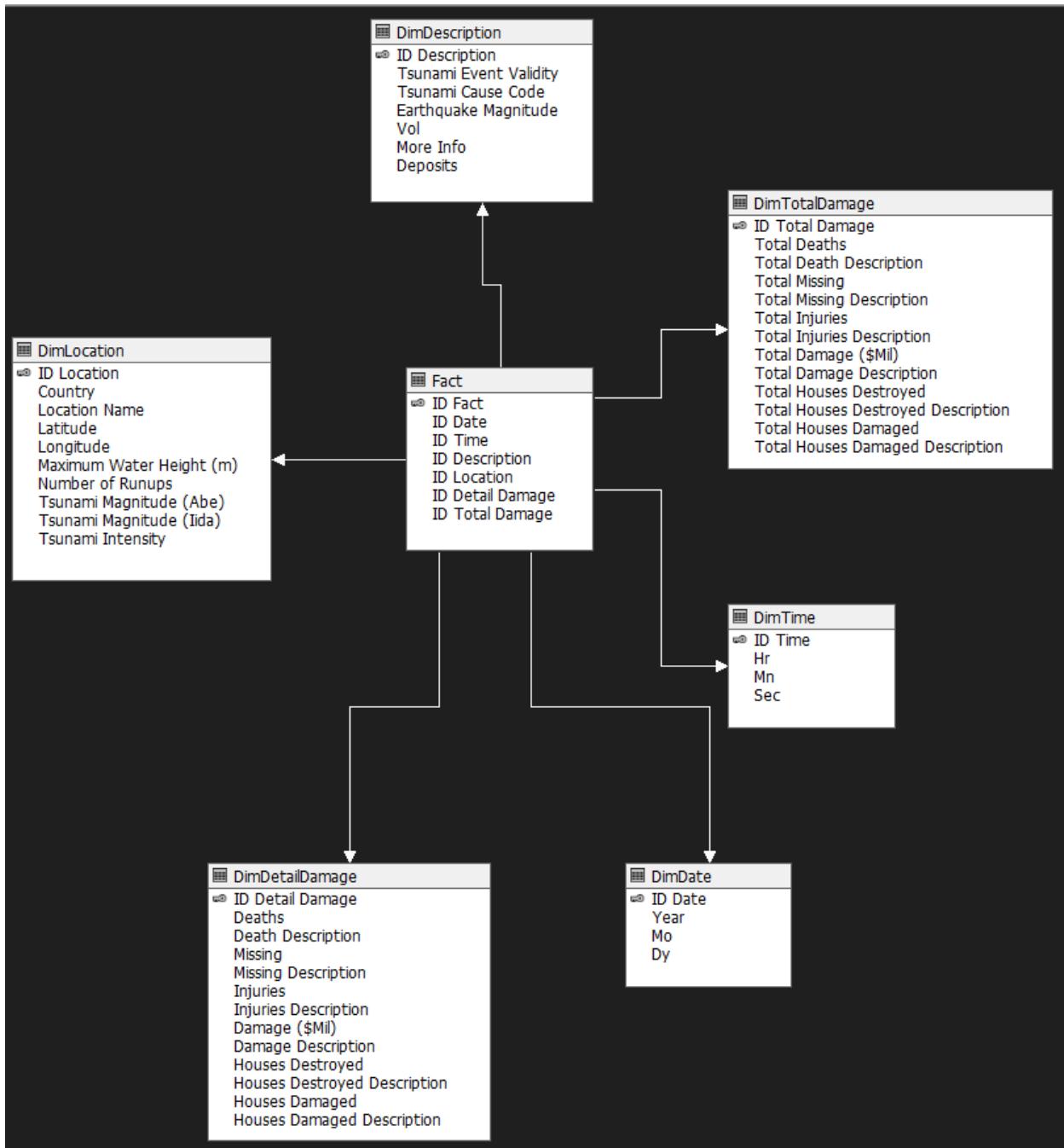




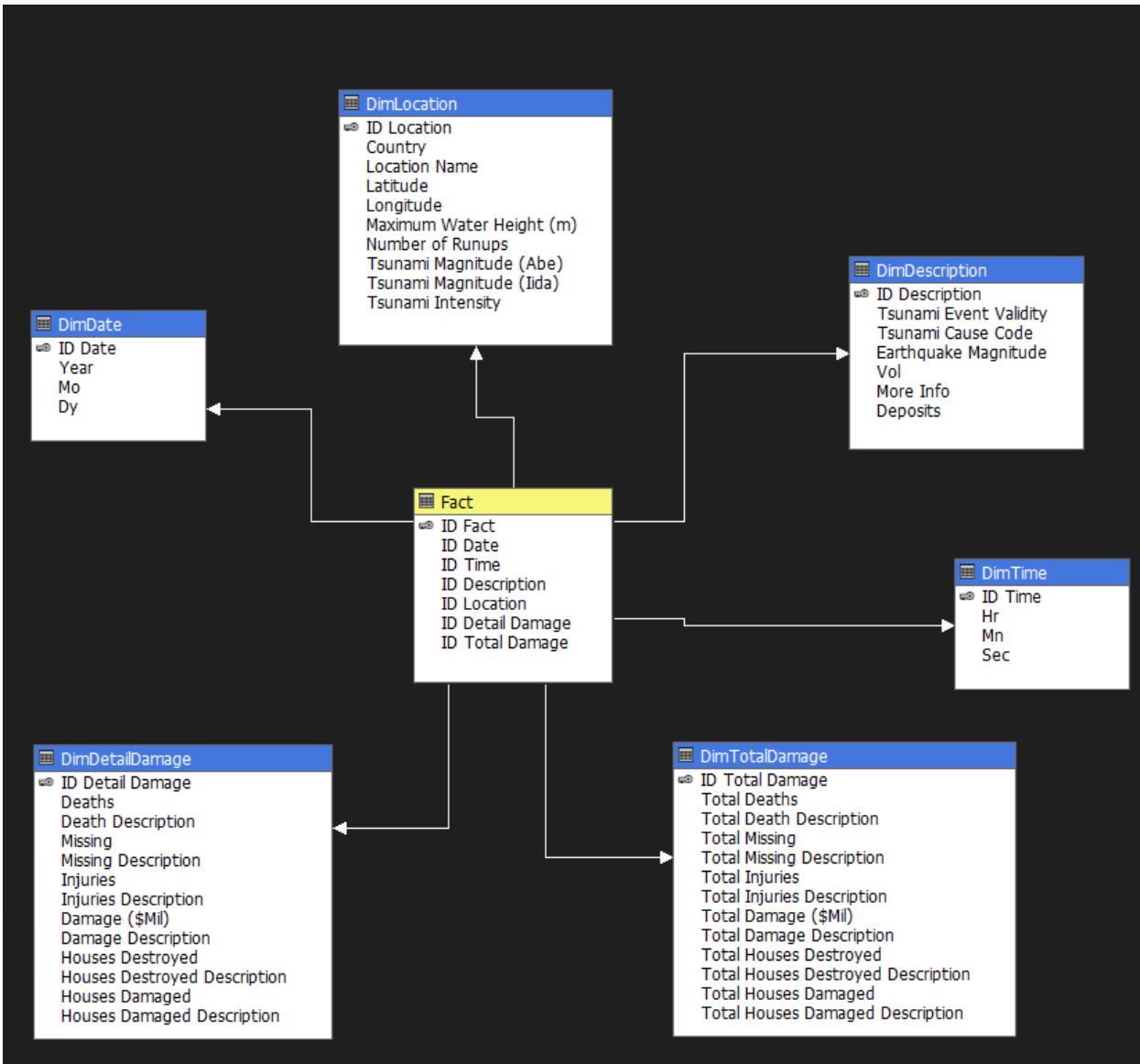
5.1 Create Data Warehouse



5.2 Create Data Source View



5.3 Create Data Cube



5.4 Edit the Dimension

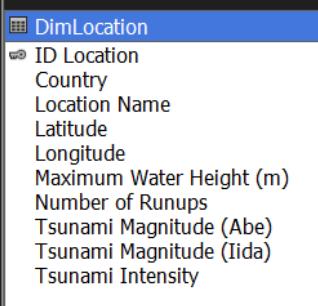
5.4.1 DimDate

Attributes	Hierarchies	Data Source View
<ul style="list-style-type: none">↳ Dim Date<ul style="list-style-type: none">■ Dy■ ID Date■ Mo■ Year	To create a new hierarchy, drag an attribute here.	 <ul style="list-style-type: none">DimDate<ul style="list-style-type: none">ID DateYearMoDy

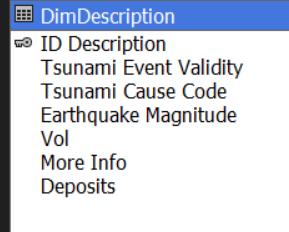
5.4.2 DimTime

Attributes	Hierarchies	Data Source View
<ul style="list-style-type: none">↳ Dim Time<ul style="list-style-type: none">■ Hr■ ID Time■ Mn■ Sec	To create a new hierarchy, drag an attribute here.	<ul style="list-style-type: none">DimTime<ul style="list-style-type: none">ID TimeHrMnSec

5.4.3 DimLocation

Attributes	Hierarchies	Data Source View
<p>↳ Dim Location</p> <ul style="list-style-type: none"> ■ Country ■ ID Location ■ Latitude ■ Location Name ■ Longitude ■ Maximum Water Height m ■ Number Of Runups ■ Tsunami Intensity ■ Tsunami Magnitude Abe ■ Tsunami Magnitude Iida 	<p>To create a new hierarchy, drag an attribute here.</p>	 <p>DimLocation</p> <ul style="list-style-type: none"> ↳ ID Location Country Location Name Latitude Longitude Maximum Water Height (m) Number of Runups Tsunami Magnitude (Abe) Tsunami Magnitude (Iida) Tsunami Intensity

5.4.4 DimDescription

Attributes	Hierarchies	Data Source View
<p>↳ Dim Description</p> <ul style="list-style-type: none"> ■ Deposits ■ Earthquake Magnitude ■ ID Description ■ More Info ■ Tsunami Cause Code ■ Tsunami Event Validity ■ Vol 	<p>To create a new hierarchy, drag an attribute here.</p>	 <pre> DimDescription ↳ ID Description Tsunami Event Validity Tsunami Cause Code Earthquake Magnitude Vol More Info Deposits </pre>

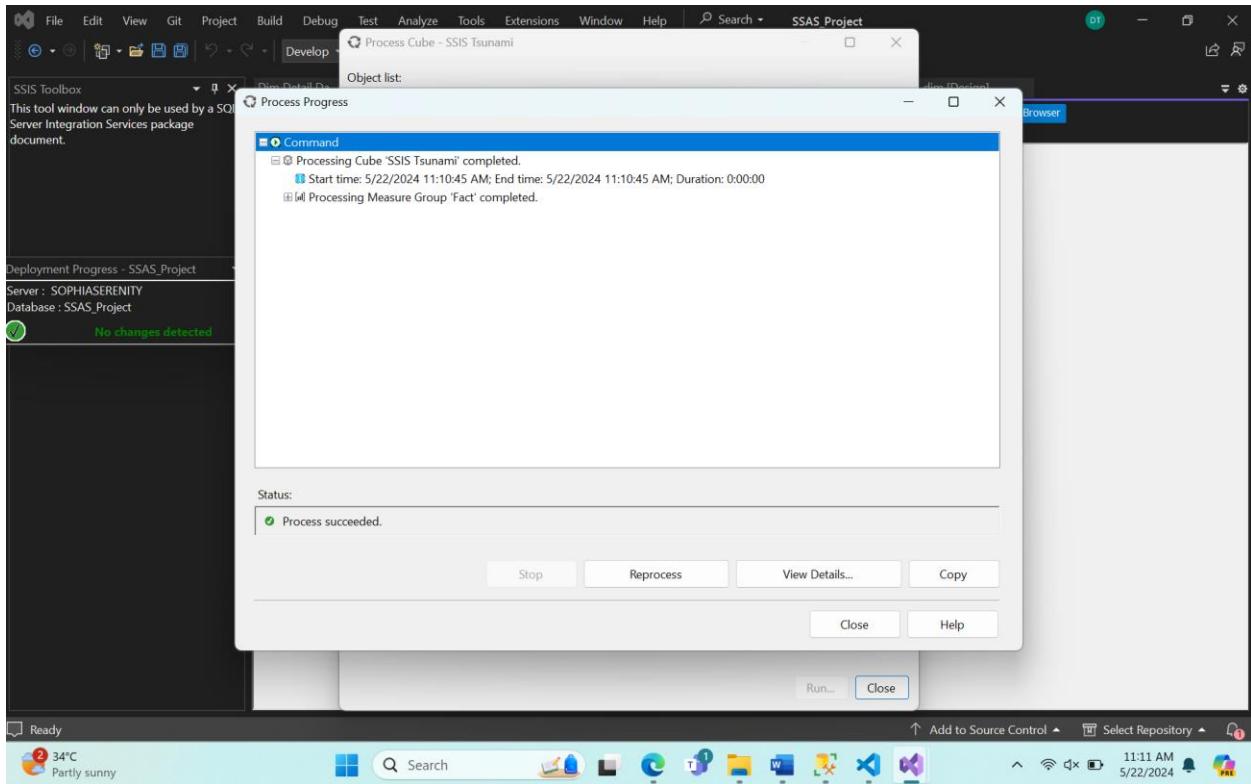
5.4.5 DimDetailDamage

Attributes	Hierarchies	Data Source View
<p>Dim Detail Damage</p> <ul style="list-style-type: none"> Damage Mil Damage Description Death Description Deaths Houses Damaged Houses Damaged Description Houses Destroyed Houses Destroyed Description ID Detail Damage Injuries Injuries Description Missing Missing Description 	To create a new hierarchy, drag an attribute here.	<p>DimDetailDamage</p> <ul style="list-style-type: none"> ID Detail Damage Deaths Death Description Missing Missing Description Injuries Injuries Description Damage (\$M\$) Damage Description Houses Destroyed Houses Destroyed Description Houses Damaged Houses Damaged Description

5.4.6 DimTotalDamage

Attributes	Hierarchies	Data Source View														
<ul style="list-style-type: none"> Dim Total Damage <ul style="list-style-type: none"> ID Total Damage Total Damage Mil Total Damage Description Total Death Description Total Deaths Total Houses Damaged Total Houses Damaged Description Total Houses Destroyed Total Houses Destroyed Description Total Injuries Total Injuries Description Total Missing Total Missing Description 	To create a new hierarchy, drag an attribute here.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #0070C0; color: white; padding: 2px;">DimTotalDamage</td> </tr> <tr> <td>ID Total Damage</td> </tr> <tr> <td>Total Deaths</td> </tr> <tr> <td>Total Death Description</td> </tr> <tr> <td>Total Missing</td> </tr> <tr> <td>Total Missing Description</td> </tr> <tr> <td>Total Injuries</td> </tr> <tr> <td>Total Injuries Description</td> </tr> <tr> <td>Total Damage (\$Mil)</td> </tr> <tr> <td>Total Damage Description</td> </tr> <tr> <td>Total Houses Destroyed</td> </tr> <tr> <td>Total Houses Destroyed Description</td> </tr> <tr> <td>Total Houses Damaged</td> </tr> <tr> <td>Total Houses Damaged Description</td> </tr> </table>	DimTotalDamage	ID Total Damage	Total Deaths	Total Death Description	Total Missing	Total Missing Description	Total Injuries	Total Injuries Description	Total Damage (\$Mil)	Total Damage Description	Total Houses Destroyed	Total Houses Destroyed Description	Total Houses Damaged	Total Houses Damaged Description
DimTotalDamage																
ID Total Damage																
Total Deaths																
Total Death Description																
Total Missing																
Total Missing Description																
Total Injuries																
Total Injuries Description																
Total Damage (\$Mil)																
Total Damage Description																
Total Houses Destroyed																
Total Houses Destroyed Description																
Total Houses Damaged																
Total Houses Damaged Description																

5.4.7 Run the Process for Deployment



SSIS Toolbox

This tool window can only be used by a SQL Server Integration Services package

Deployment Progress - SSAS_Project

Server : SOPHIASERENITY
Database : SSAS_Project

No changes detected

Dim Detail Damage [Design]

Dim Date.dim [Design]

SSIS Tsunami.cube [Design]

Dim Total Dam...dim [Design]

Cube Structure Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

MDX

Dimension Hierarchy Operator Filter Expression Parameters

<Select dimension>

Country	Location Name	Year	Longitude	Latitude	Fact Count
ALBAN...	ALBANIAN C...	1893	19.7	40.2	1
ALBAN...	ALBANIAN C...	1920	19.5	40.5	1
ALBAN...	BAY OF VLORE	1866	19.4	40.4	1
ALBAN...	OCHRIDA LAKE	1911	20.8	40.9	1
ALGER...	"ALGIERS	1989	TIPAZA"	CHER...	1
ALGER...	ALGERIA	1884	4.53	36.85	1
ALGER...	ALGIERS	1365	3	36.8	1
ALGER...	ALGIERS	2003	3.578	36.939	1
ALGER...	EL ASNAM	1980	1.374	36.199	1
ALGER...	HARBOUR OF...	1801	2.9	36.5	1
ALGER...	N. ALGERIA	2003	3.634	36.964	1
ALGER...	N. ALGERIA	2021	5.198	36.915	1
ALGER...	ORAN	1790	-0.6	35.7	1
ALGER...	SEA OF ALBO...	1856	5.7	37.1	2
ALGER...	SEA OF ALBO...	1857	1.3	36.5	1
ALGER...	SEA OF ALBO...	1891	1.8	36.5	1
ALGER...	SEA OF ALBO...	1954	1.594	36.274	1
ANTA...	"TELEPHONE ...	1967	-63	DECE...	1
ANTA...	BALLENY ISL...	1998	149.527	-62.877	1
ANTA...	SCOTIA SEA	2003	-43.411	-60.532	1
ANTA...	SCOTIA SEA	2013	-46.401	-60.274	1
ANTA...	SCOTIA SEA	2021	-55.494	-61.825	1
ANTIG...	ANTIGUA ISL...	1802	-62.4	17.2	1
ANTIG...	ANTIGUA; S...	1690	-61.5	17.5	1
ARGE...	MAR DEL PLA...	1954	-57	-38	1
ATLA...	ATLANTIC O...	1894	-34.5	49	1

6 Query MDX

6.1 Query Tsunami year and description

6.1.1 SQL Query

```
-- 1. Query Tsunami date and description

SELECT
{
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY
(
    [Dim Date].[Year].[Year].Members *
    [Dim Description].[Tsunami Event Validity].[Tsunami Event Validity].Members *
    [Dim Description].[Tsunami Cause Code].[Tsunami Cause Code].Members *
    [Dim Description].[Earthquake Magnitude].[Earthquake Magnitude].Members *
    [Dim Description].[Vol].[Vol].Members
) ON ROWS
FROM [SSIS Tsunami]
```

6.1.2 Get SQL result

				Fact Count
1008	2	1	6.5	1
1026	3	1	7.5	1
103	3	1	7	1
1033	4	1		1
1039	1	1	6.8	1
1050	3	6		4845
1065	3	1		1
1068	3	1		1
1076	1	0		1
1088	3	1		1
1088	4	1	7	1
1092	-1	9		1
1096	4	1	8.4	1
1099	1	0		1
1000	4	1	9.4	1

6.2 Query Number of Deaths and their description

6.2.1 SQL Query

```
-- 2. Query number of deaths and their description
```

```
SELECT
{
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY
(
    [Dim Detail Damage].[Deaths].[Deaths].Members *
    [Dim Detail Damage].[Death Description].[Death Description].Members
) ON ROWS
FROM [SSIS Tsunami]
```

6.2.2 Get SQL Result

		Fact Count
11	1	2
110	3	1
1105	4	1
1169	4	1
12	1	5
1200	4	1
1223	4	1
1239	4	1
128	3	1
13	1	2
1300	4	1
13486	4	1
1362	4	1
14	1	2

6.3 Query Tsunami intensity over the year

6.3.1 SQL query

```
-- 4. Query total of deaths over the year with country
```

```
SELECT
NON EMPTY {
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY {
    [Dim Location].[Country].[Country].Members *
    [Dim Date].[Year].[Year].Members *
    [Dim Total Damage].[Total Deaths].[Total Deaths].Members
} ON ROWS
FROM [SSIS Tsunami]
```

6.3.2 Get SQL Result

			Fact Count
"MICRONESIA		1837	1
"MICRONESIA		1912	2
"MICRONESIA		1925	1
ALBANIA		1273	1
ALBANIA		1833	1
ALBANIA		1851	2000
ALBANIA		1866	3
ALBANIA		1866	60
ALBANIA		1893	1
ALBANIA		1911	1
ALBANIA		1920	1
ALBANIA		346	1
ALBANIA		-58	1
ALGERIA		1365	1
ALGERIA		1790	3000
ALGERIA		1802	1
ALGERIA		1856	5
ALGERIA		1856	8
ALGERIA		1857	1
ALGERIA		1885	1
ALGERIA		1901	20

6.4 Query total of deaths over the year with country

6.4.1 SQL Query

```
-- 4. Query total of deaths over the year with country
```

```
SELECT
    NON EMPTY {
        [Measures].[Fact Count]
    } ON COLUMNS,
    NON EMPTY {
        [Dim Location].[Country].[Country].Members *
        [Dim Date].[Year].[Year].Members *
        [Dim Total Damage].[Total Deaths].[Total Deaths].Members
    } ON ROWS
FROM [SSIS Tsunami]
```

6.4.2 Get SQL Result

Messages Results

			Fact Count
"MICRONESIA	1837		1
"MICRONESIA	1912		2
"MICRONESIA	1925		1
ALBANIA	1273		1
ALBANIA	1833		1
ALBANIA	1851	2000	1
ALBANIA	1866		3
ALBANIA	1866	60	2
ALBANIA	1893		1
ALBANIA	1911		1
ALBANIA	1920		1
ALBANIA	346		1
ALBANIA	-58		1
ALGERIA	1365		1
ALGERIA	1790	3000	1
ALGERIA	1802		1
ALGERIA	1856	5	1
ALGERIA	1856	8	1
ALGERIA	1857		1
ALGERIA	1885		1
ALGERIA	1891	36	1
ALGERIA	1054	1242	1

6.5 Query number of houses damaged and their descriptions

6.5.1 SQL Query

```
-- 5. Query for number of houses damaged and their descriptions

SELECT
{
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY
(
    [Dim Detail Damage].[Houses Damaged].[Houses Damaged].Members *
    [Dim Detail Damage].[Houses Damaged Description].[Houses Damaged Description].Members
) ON ROWS
FROM [SSIS Tsunami]
```

6.5.2 Get SQL Result

		Fact Count
182	3	1
19	1	1
2		6
22	1	1
238	3	1
280920	4	1
2844	4	1
3		9
3	1	1
300	3	1
39191	4	1
4		5
478	3	1
500	2	1

6.6 Query for the total number of tsunamis by time of day with descriptions

6.6.1 SQL Query

```
-- 6. Query for the total number of tsunamis by time of day with descriptions

SELECT
{
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY
(
    [Dim Time].[Hr].[Hr].Members *
    [Dim Description].[Tsunami Event Validity].[Tsunami Event Validity].Members *
    [Dim Description].[Tsunami Cause Code].[Tsunami Cause Code].Members
) ON ROWS
FROM [SSIS Tsunami]
```

6.6.2 Get SQL Result

Messages Results

Fact Count			
1	1	9	1
1	-1	1	1
1	2	1	8
1	2	3	1
1	3	1	11
1	3	3	1
1	3	9	1
1	4	1	27
1	4	3	2
1	4	6	1
1	4	9	1
10	0	1	2
10	0	2	1
10	1	0	1

6.7 Query Countries and location has tsunami over the months

6.7.1 SQL Query

```
-- 7. Countries and location has tsunami over the dim date

SELECT
{
    [Measures].[Fact Count]
} ON COLUMNS,
NON EMPTY
(
    [Dim Location].[Country].[Country].Members *
    [Dim Location].[Location Name].[Location Name].Members *
    [Dim Date].[Mo].[Mo].Members
)
ON ROWS
FROM [SSIS Tsunami]
```

6.7.2 Get SQL Result

Messages Results

			Fact Count
"MICRONESIA	FED. STATES OF"	10	2
"MICRONESIA	FED. STATES OF"	12	2
ALBANIA	ADRIATIC SEA		1
ALBANIA	ALBANIA		1
ALBANIA	ALBANIA	1	1
ALBANIA	ALBANIAN COASTS	1	2
ALBANIA	ALBANIAN COASTS	10	1
ALBANIA	ALBANIAN COASTS	12	1
ALBANIA	ALBANIAN COASTS	3	2
ALBANIA	ALBANIAN COASTS	6	1
ALBANIA	ALBANIAN COASTS	9	1
ALBANIA	BAY OF VLORE	3	1
ALBANIA	OCHRIDA LAKE	2	1
ALGERIA	"ALGIERS	10	1
ALGERIA	ALGERIA	1	1
ALGERIA	ALGIERS	1	1
ALGERIA	ALGIERS	5	1
ALGERIA	EL ASNAM	10	1
ALGERIA	HARBOUR OF ALGIERS	11	1
ALGERIA	N. ALGERIA	3	1
ALGERIA	N. ALGERIA	5	1

6.8 Top 10 Countries by Event Count in 2020

6.8.1 SQL Query

```
--8. Top 10 Countries by Event Count in 2020
SELECT
    NON EMPTY {
        [Measures].[Fact Count]
    } ON COLUMNS,
    NON EMPTY {
        TOPCOUNT(
            [Dim Location].[Country].[Country].Members,
            10,
            [Measures].[Fact Count]
        )
    } ON ROWS
FROM [SSIS Tsunami]
WHERE ([Dim Date].[Year].&[2020])
```

6.8.2 Get SQL Result

	Fact Count
USA	4
CANADA	1
CHILE	1
CUBA	1
GREECE	1
INDONESIA	1
MEXICO	1
NEW ZEALAND	1
RUSSIA	1
SPAIN	1

6.9 Top 3 Years by Event Count

6.9.1 SQL Query

```
--9. Top 3 Years by Event Count
SELECT
    NON EMPTY {
        [Measures].[Fact Count]
    } ON COLUMNS,
    NON EMPTY {
        TOPCOUNT(
            [Dim Date].[Year].[Year].Members,
            3,
            [Measures].[Fact Count]
        )
    } ON ROWS
FROM [SSIS Tsunami]
```

6.9.2 Get SQL Result

	Fact Count
1928	25
1938	24
1868	23

6.10 Query for the total damage (\$) by country and year

6.10.1 SQL Query

```
--10. Top 3 Countries by Event Count for Each Year

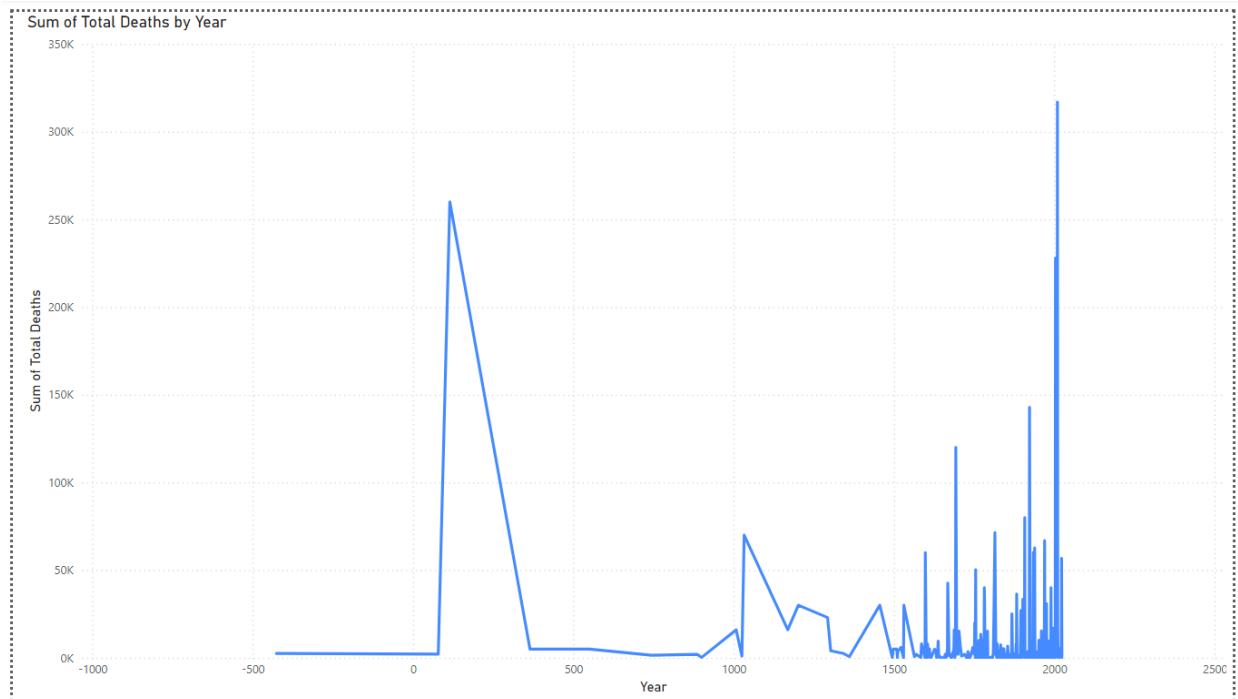
SELECT
    NON EMPTY {
        [Measures].[Fact Count]  -- Replace with the correct measure name for deaths
    } ON COLUMNS,
    NON EMPTY {
        GENERATE(
            [Dim Date].[Year].[Year].Members,
            TOPCOUNT(
                [Dim Location].[Country].[Country].Members,
                3,
                [Measures].[Fact Count]  -- Replace with the correct measure name for deaths
            )
        )
    } ON ROWS
FROM [SSIS Tsunami]
```

6.10.2 Get SQL Result

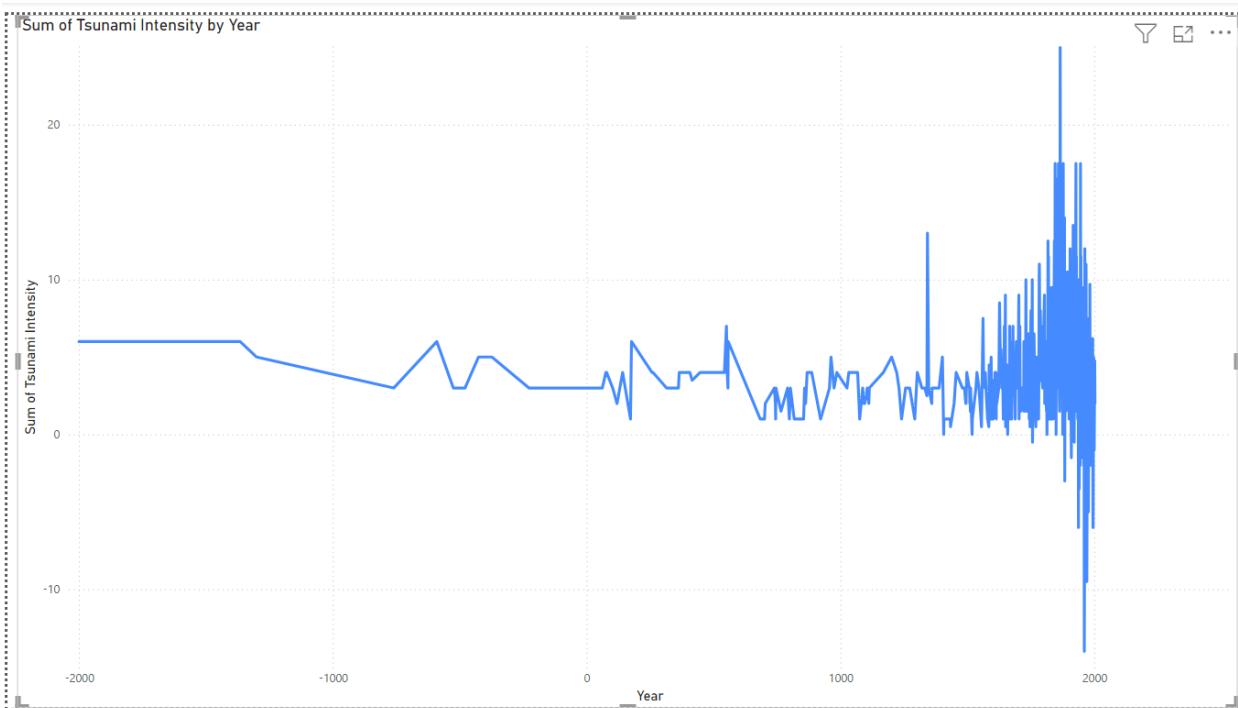
	Fact Count
IRAN	6
	1
"MARSHALL ISLANDS	1
JAPAN	371
UKRAINE	12
SYRIA	8
TURKEY	67
GREECE	179
ISRAEL	9
CHINA	43
UK	23
ITALY	160
NORTH SEA	1
CYPRUS	5
SOUTH KOREA	14
ALBANIA	13
ALGERIA	14

7 Report Data (by using PowerBI)

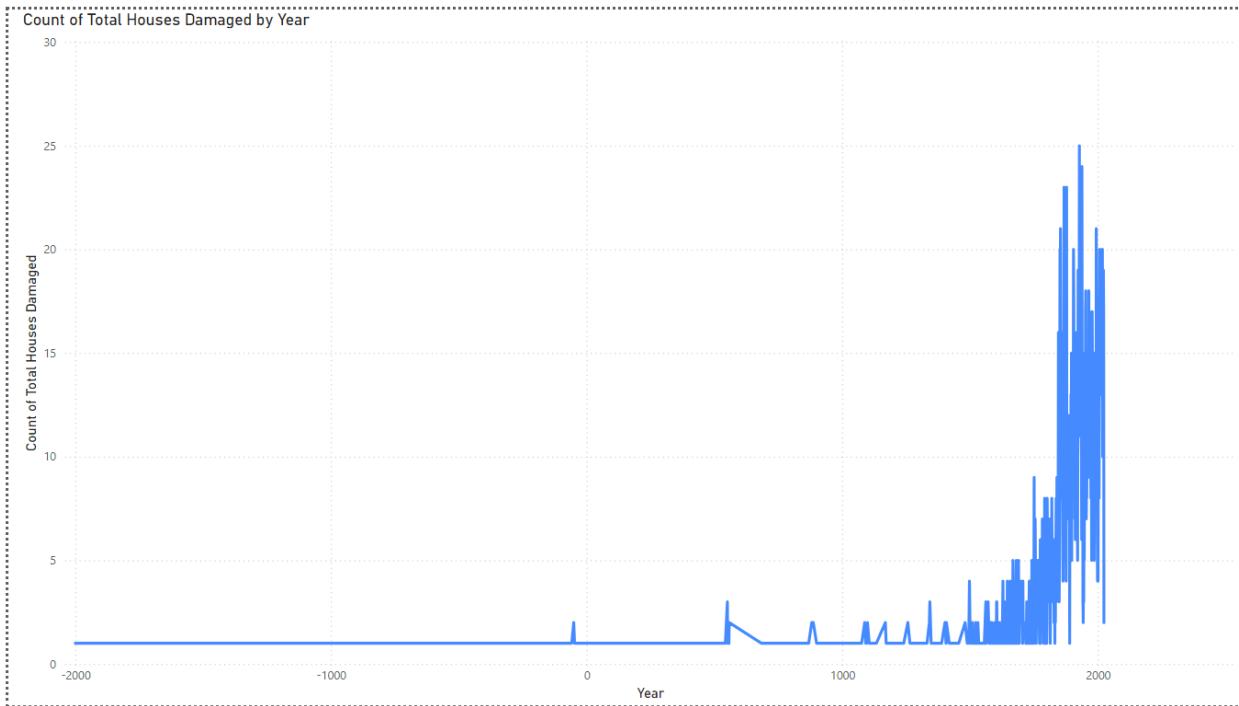
7.1 Sum of Deaths by Year



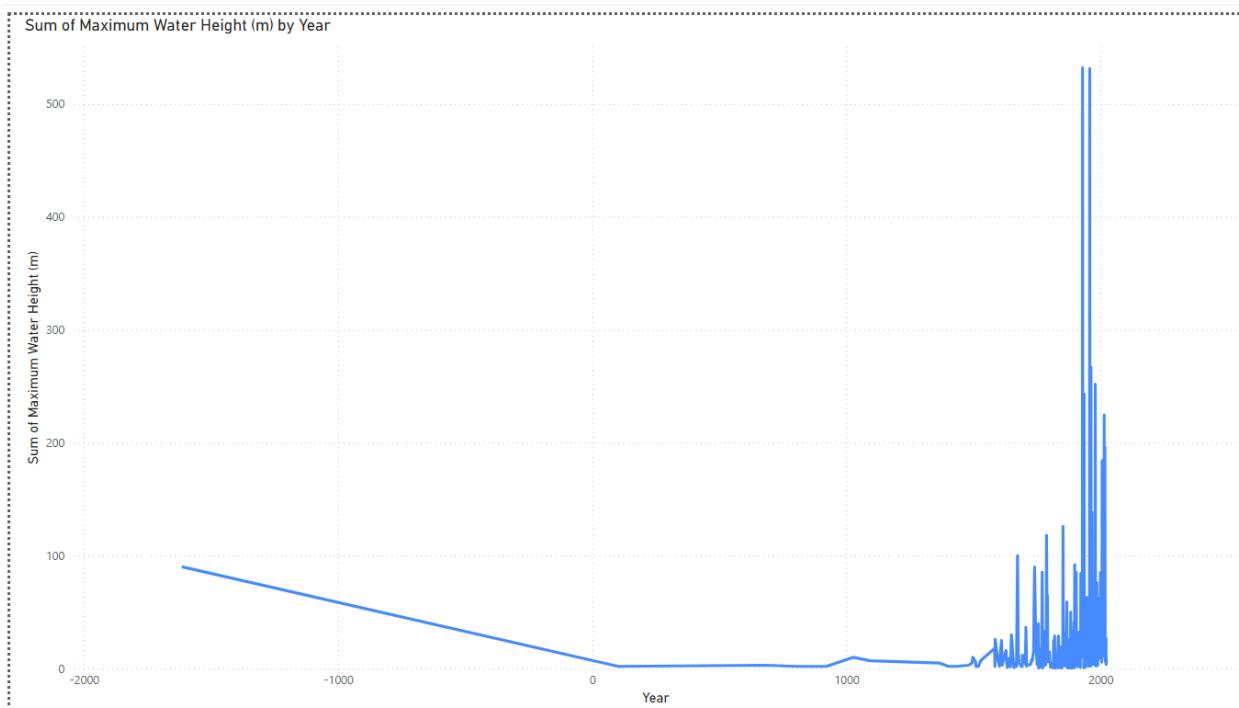
7.2 Sum of Tsunami Intensity by Year



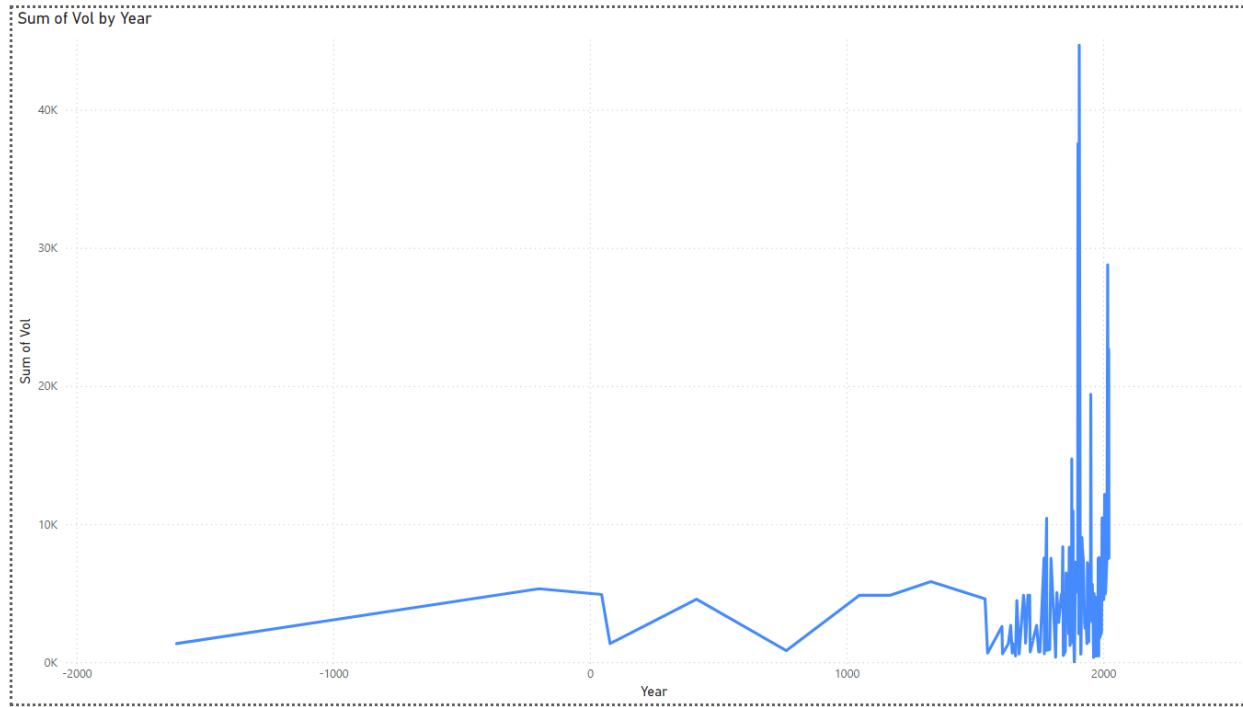
7.3 Count of Total Houses Damaged by Year



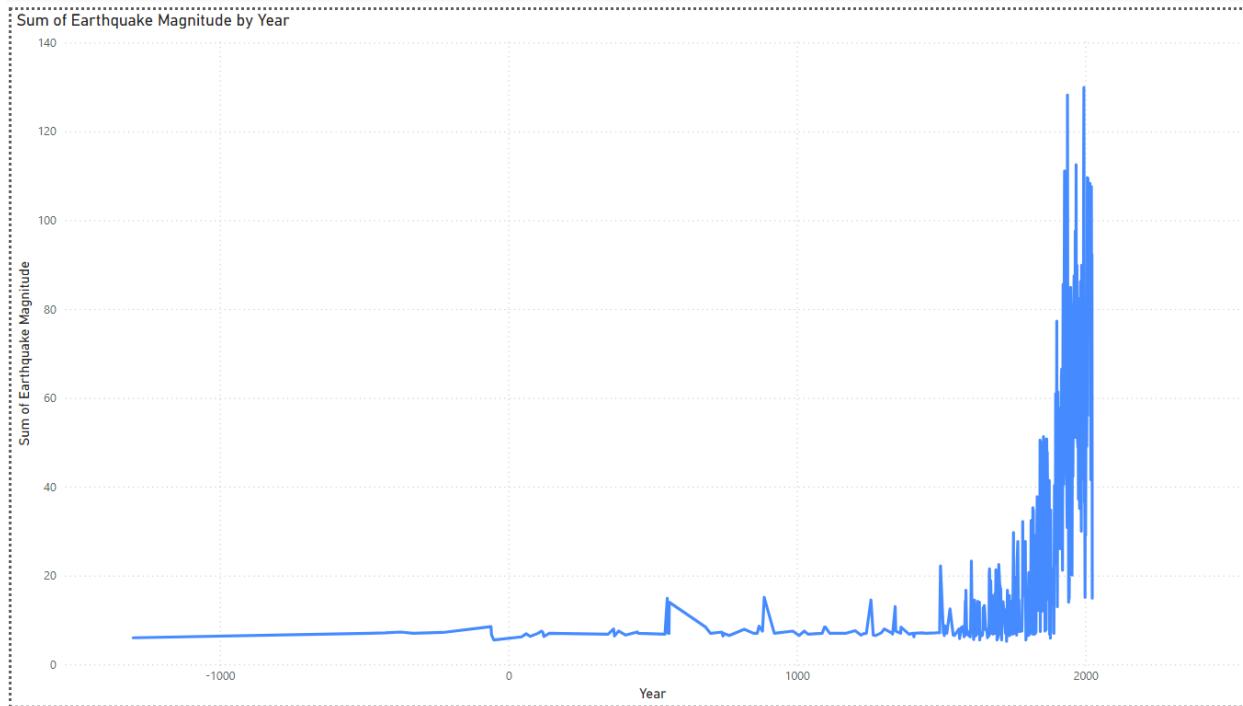
7.4 Sum of Maximum Water Height (m) by Year



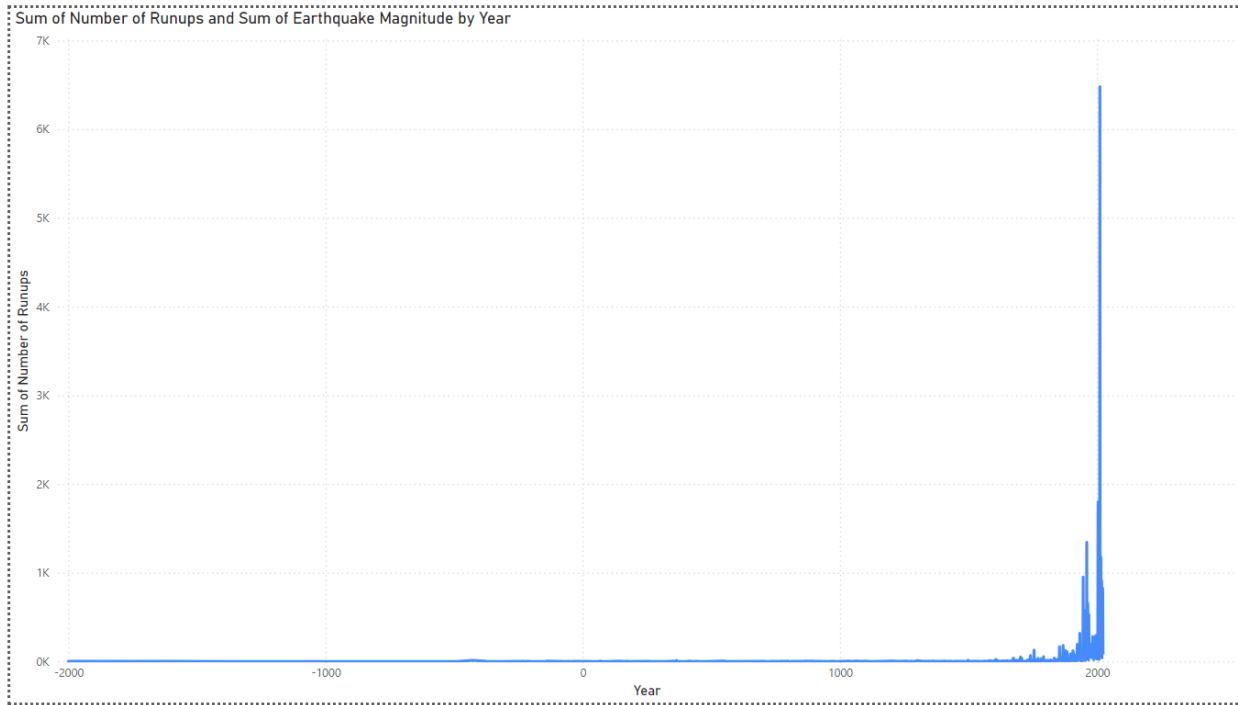
7.5 Sum of Vol by Year



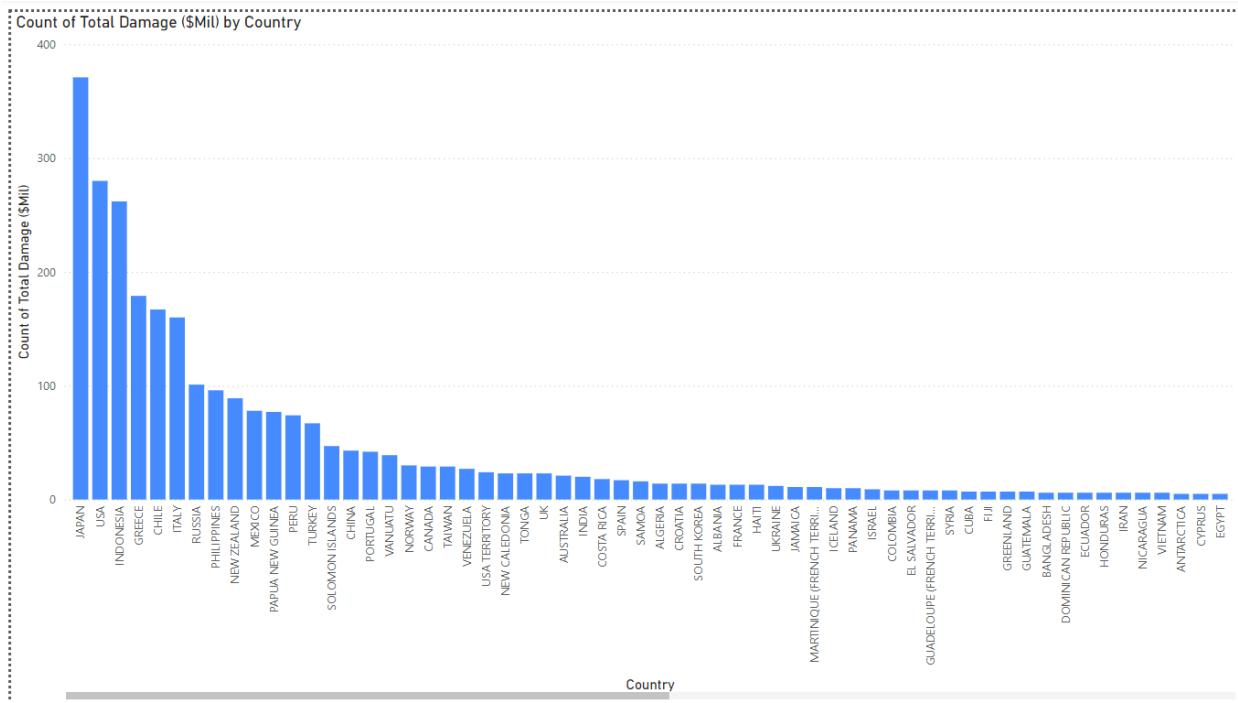
7.6 Sum of Earthquake Magnitude by Year



7.7 Sum of Number of Runups and Sum of Earthquake Magnitude by Year



7.8 Count of Total Damage (\$Mil) by Country



7.9 Sum of Tsunami Intensity in Month by Country

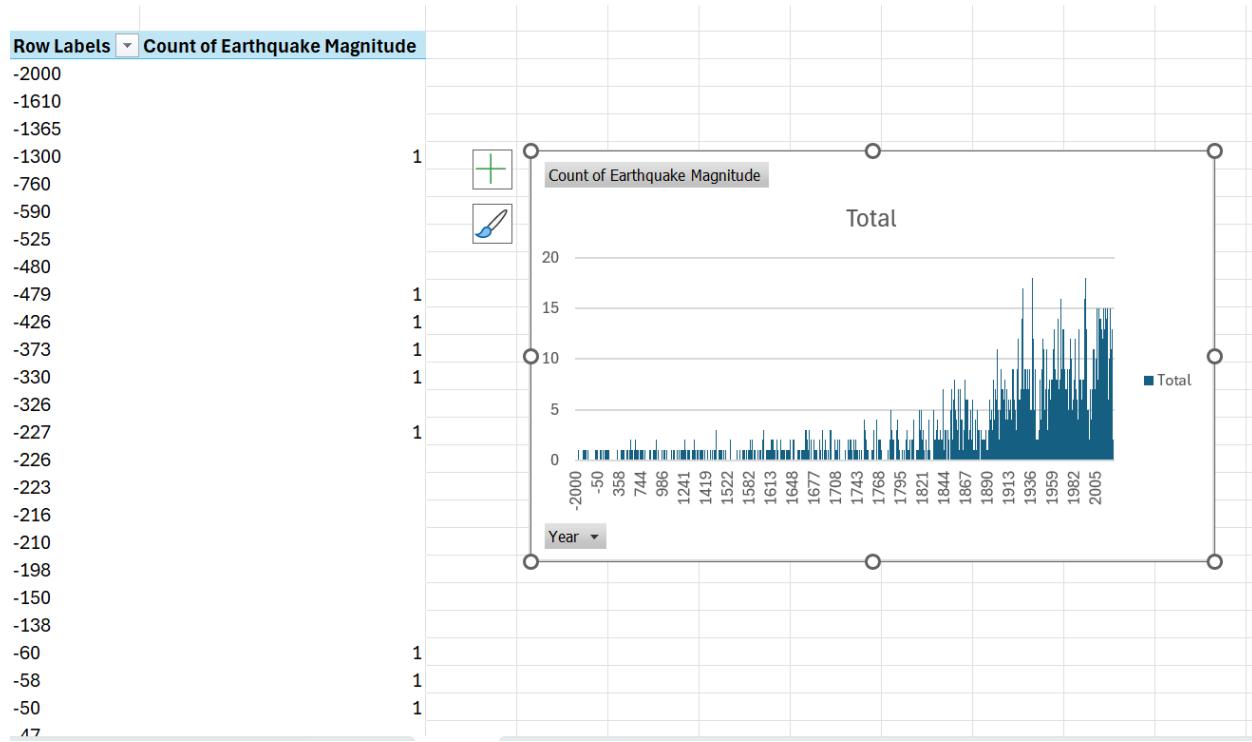
Country	1	2	3	4	5	6	7	8	9	10	11	12	Total
ALBANIA	3.00	11.00		10.00		3.00		3.00	3.00		5.00	38.00	
ALGERIA					1.50		8.00		6.00	2.00		17.50	
ANTARCTICA												0.00	
ATLANTIC OCEAN												2.00	
AUSTRALIA												-5.00	
BULGARIA				4.00								4.00	
CANADA			0.00		0.00	2.00						2.00	
CHILE	2.00	1.50	7.50	9.00	4.50	13.50	0.00	7.50	24.00	0.50	6.20	17.00	7.50
CHINA		1.00	-1.00			2.00	1.00		1.50	2.00	1.00	2.00	9.50
COLOMBIA						1.00						2.50	3.50
COOK ISLANDS				2.50									2.50
COSTA RICA							-3.00		3.00		-2.00		-7.00
CROATIA			3.00		5.00	4.00		6.00	10.00	4.00		3.00	35.00
CYPRUS				2.00			4.00	4.00		2.00			12.00
DEAD SEA		3.00											3.00
EAST CHINA SEA								0.00					0.00
ECUADOR			5.00	0.50							1.50		7.00
EGYPT		3.00					3.00						6.00
EL SALVADOR				2.00					1.50	0.50		1.50	5.50
FJJI								1.00		1.00			2.00
FRANCE			4.00			3.00		4.00					11.00
FRENCH POLYNESIA								0.50					0.50
GREECE	34.00	17.00	36.00	13.00	28.00	12.00	24.00	22.00	26.00	16.00	14.00	24.00	16.00
GUATEMALA											-1.00		-1.00
INDONESIA	5.00	14.80	16.30	18.50	9.50	10.00	9.50	4.00	20.00	9.00	9.82	15.50	15.20
													157.12

7.10 Count of Tsunami Days in Months

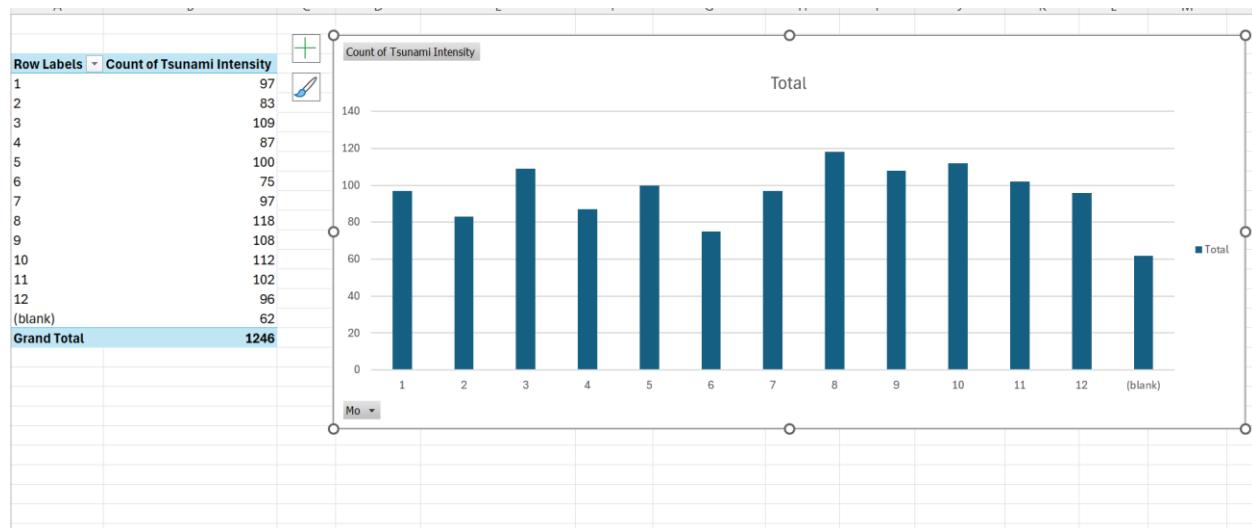
Mo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Total		
	151																														151			
1	13	14	5	7	8	4	5	4	11	10	6	4	5	7	8	10	10	4	7	8	9	7	7	8	4	9	5	7	8	6	7	11	238	
2	4	2	12	8	7	5	8	10	8	8	6	11	3	10	7	3	8	8	3	6	7	8	5	10	5	5	6	5	4	4		196		
3	13	7	7	4	10	5	5	5	4	11	2	10	9	4	8	5	8	8	11	9	12	5	8	6	7	9	6	3	8	3	7	12	231	
4	7	8	6	8	7	7	6	5	6	7	9	7	7	11	7	4	8	6	7	9	6	7	8	7	9	3	3	5	5	4	7		206	
5	13	6	6	5	11	8	9	16	8	1	6	6	5	7	8	6	6	7	3	10	10	6	8	10	6	9	9	6	4	6	5	8	234	
6	4	4	4	8	2	5	2	4	6	5	7	4	10	11	4	13	8	5	12	4	6	5	6	7	6	4	9	7	5	8	3		188	
7	8	6	4	6	5	5	4	3	8	13	12	6	9	6	3	5	5	8	11	7	10	6	3	7	6	11	6	3	6	8	11	5		216
8	14	7	11	4	7	6	10	5	11	12	9	11	11	6	9	11	11	9	8	11	11	10	5	9	6	5	5	9	3	9	4	6		265
9	10	12	8	4	6	12	7	7	14	7	6	7	7	7	4	13	3	6	6	9	10	4	6	10	9	6	6	6	6	8		233		
10	12	1	7	7	10	8	10	5	7	12	9	8	10	5	5	2	7	4	14	9	13	5	10	3	4	7	10	3	7	5	4	11		234
11	7	13	8	1	10	10	7	6	5	4	7	12	5	11	10	7	11	10	11	9	5	11	12	6	11	9	5	4	11	9	9		256	
12	8	7	5	6	15	4	4	6	4	11	3	4	6	5	5	4	9	11	10	3	6	9	5	5	4	9	11	4	11	10	4	5		213
Total	264	87	83	68	98	79	77	76	92	101	82	90	87	90	81	74	104	83	103	91	104	89	81	84	78	89	81	62	78	78	69	58	2861	

8 Pivot Table

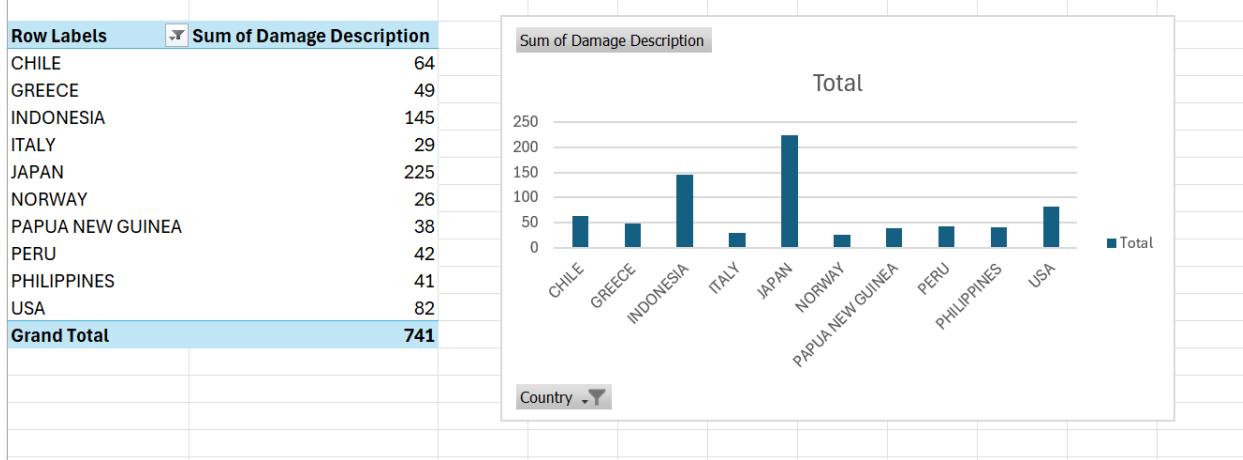
8.1 Dataset



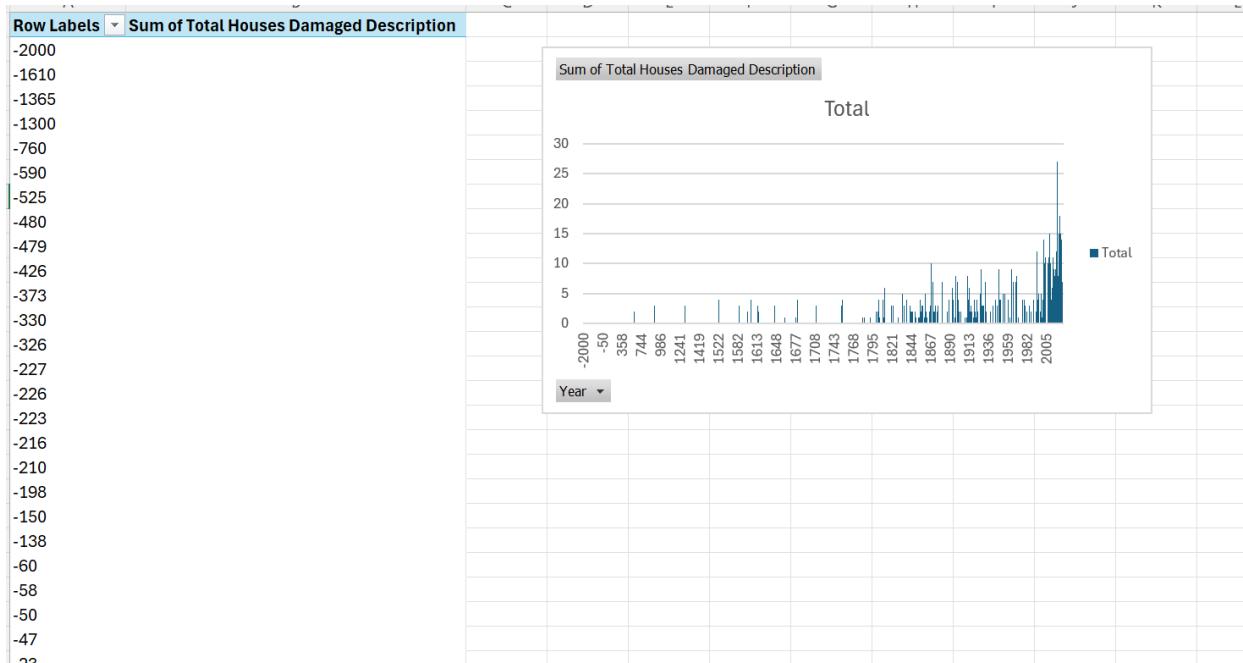
8.2.3 Number of Tsunami Intensity over the month



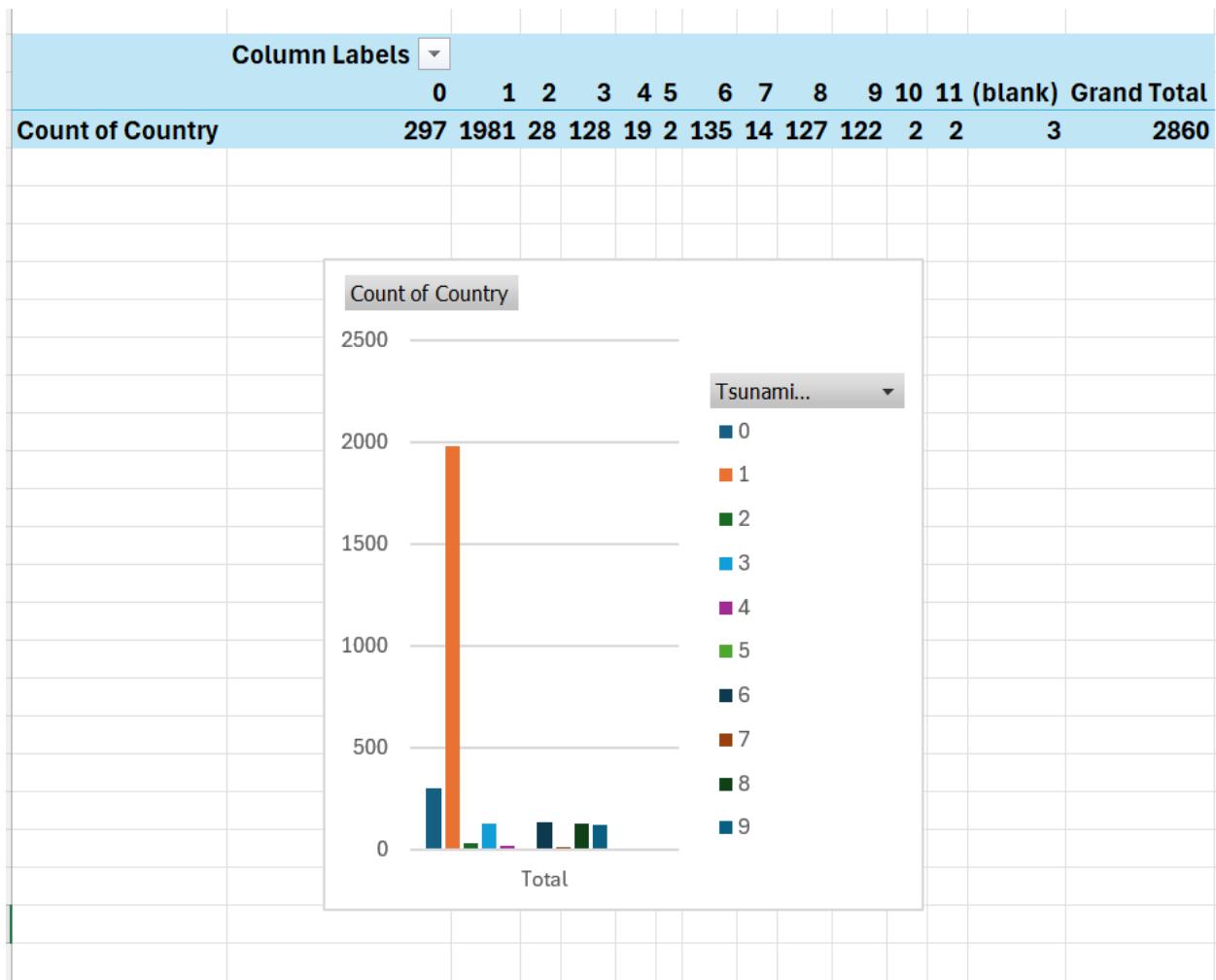
8.2.4 Top 10 of Country has Damage (\$Mil)



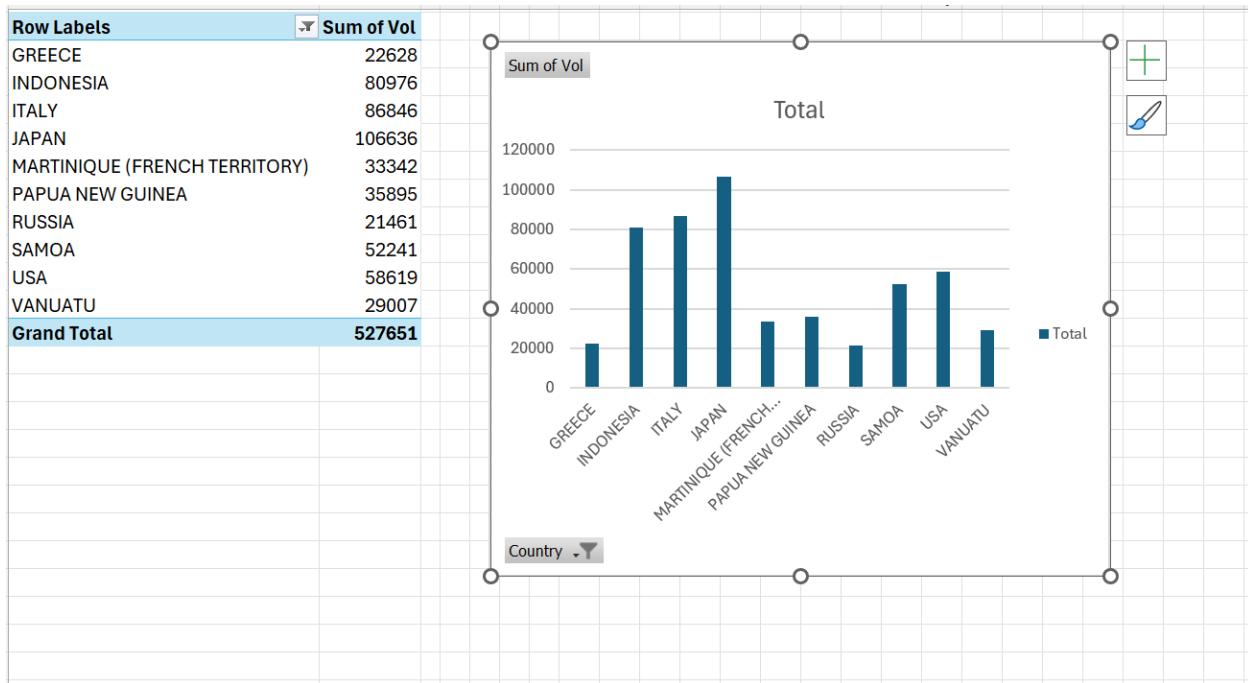
8.2.5 Sum of Total House damage description



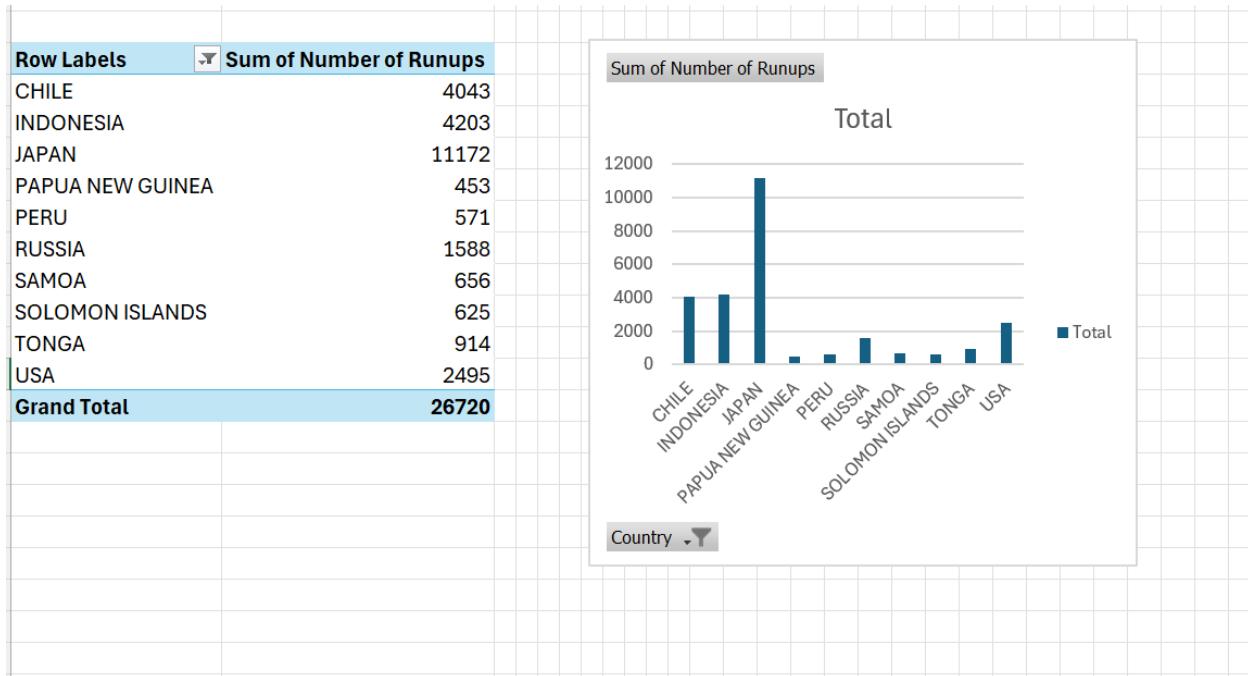
8.2.6 Count of Country per Tsunami Cause



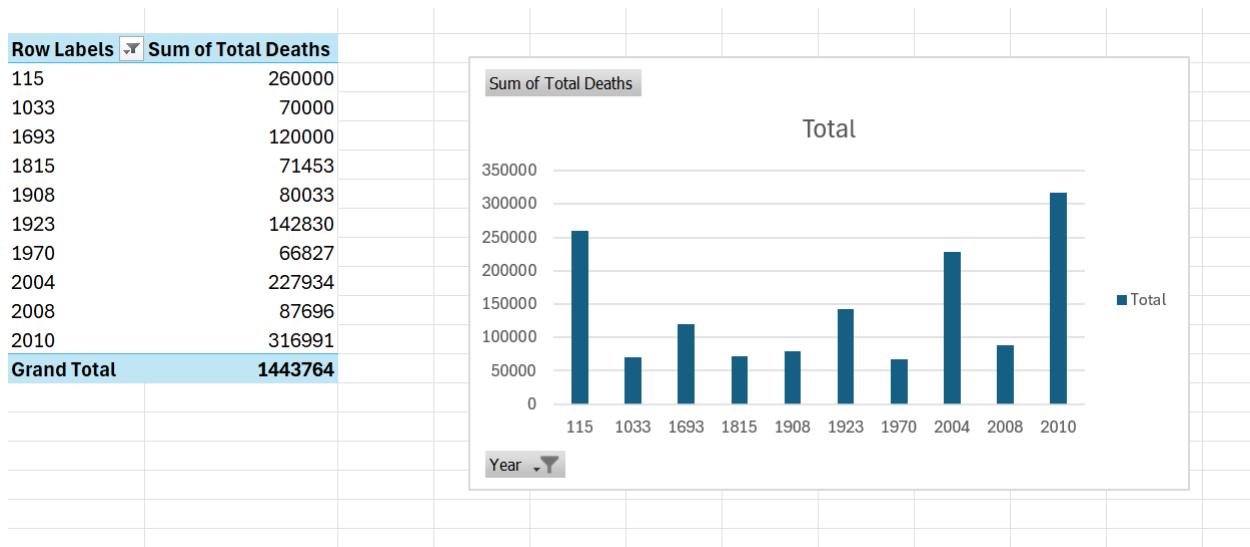
8.2.7 Top 10 of Country has Volcano



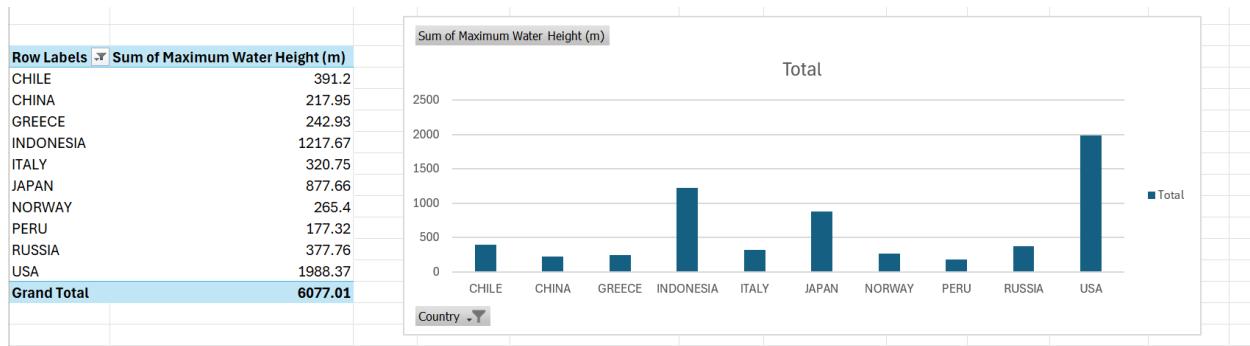
8.2.8 Top 10 of Country has total runup highest



8.2.9 Top 10 of Year has number of deaths highest



8.2.10 Top 10 of Country has Maximum Water Height



9 Data Mining

9.1 Load Dataset

```
In [2]: # Load the dataset
tsunami = pd.read_csv(r'C:\Users\ungdu\Downloads\Smart-Decision\Tsunami_Cut.csv')
tsunami.head()
```

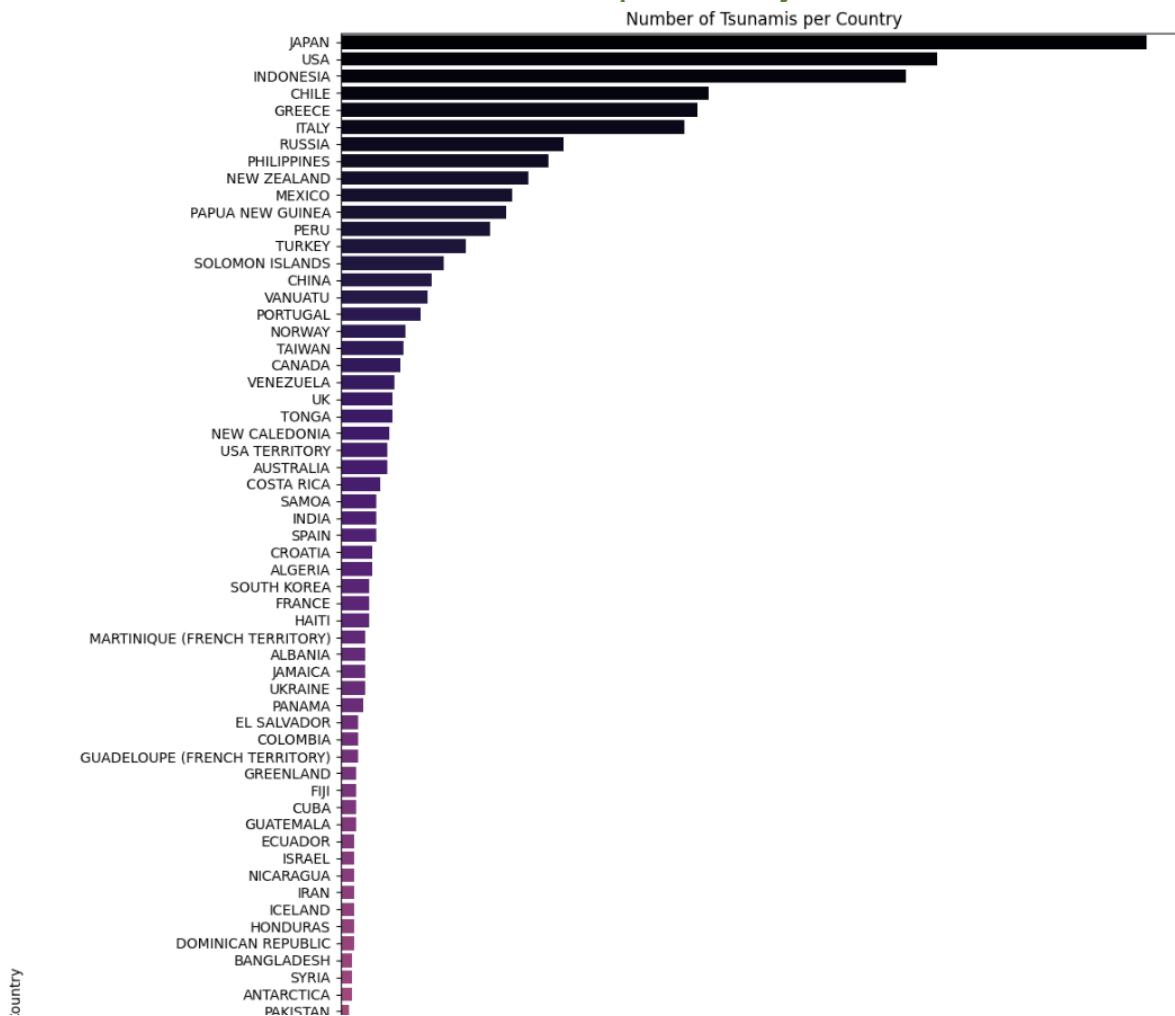
Out[2]:

	Year	Mo	Tsunami Event Validity	Tsunami Cause Code	Earthquake Magnitude	Vol	Deposits	Country	Location Name	Latitude	...	Damage (\$Mil)	Damage Description	House Destroye
0	-2000	NaN	1	1.0	NaN	NaN	0	SYRIA	SYRIAN COASTS	35.683	...	NaN	4.0	Na
1	-1610	NaN	4	6.0	NaN	1351.0	28	GREECE	TERA ISLAND (SANTORINI)	36.400	...	NaN	3.0	Na
2	-1365	NaN	1	1.0	NaN	NaN	0	SYRIA	SYRIAN COASTS	35.683	...	NaN	NaN	Na
3	-1300	NaN	2	0.0	6.0	NaN	0	TURKEY	IONIAN COASTS, TROAD	39.960	...	NaN	NaN	Na
4	-760	NaN	2	0.0	NaN	NaN	0	ISRAEL	ISRAEL AND LEBANON COASTS	NaN	...	NaN	NaN	Na

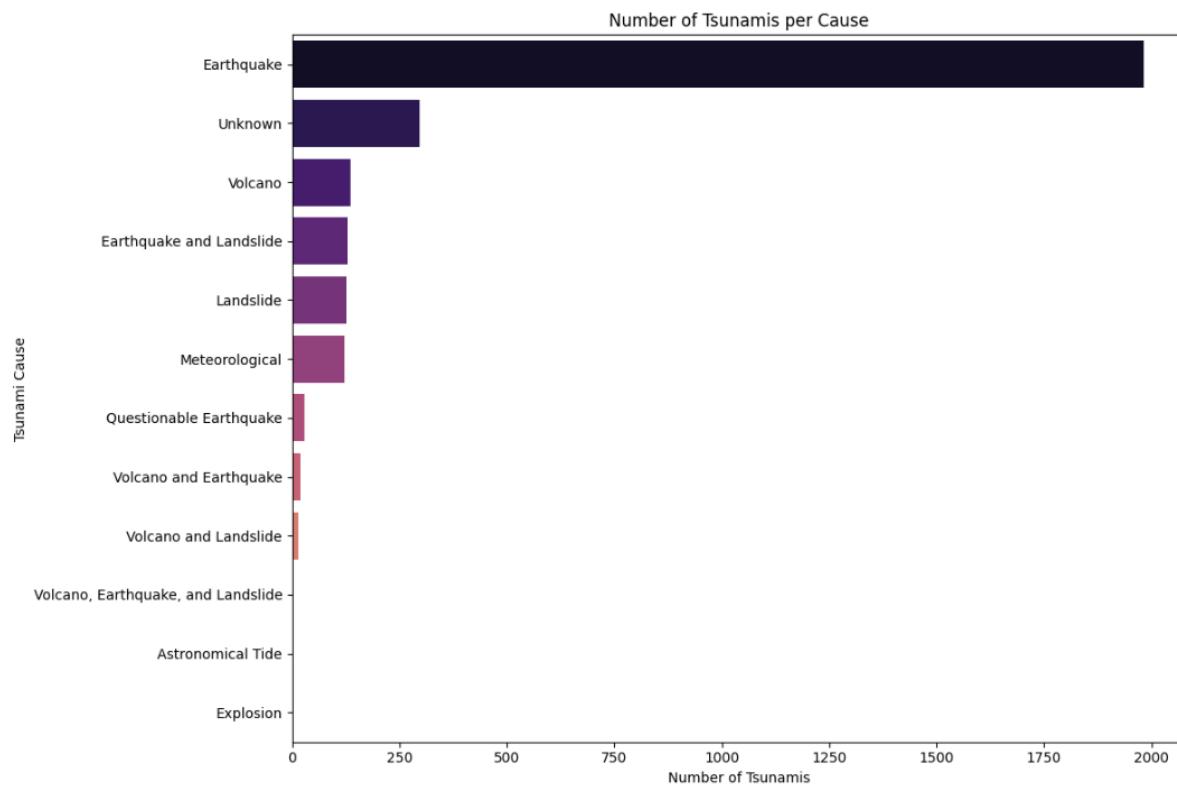
5 rows × 26 columns

9.2 Visualization of Dataset

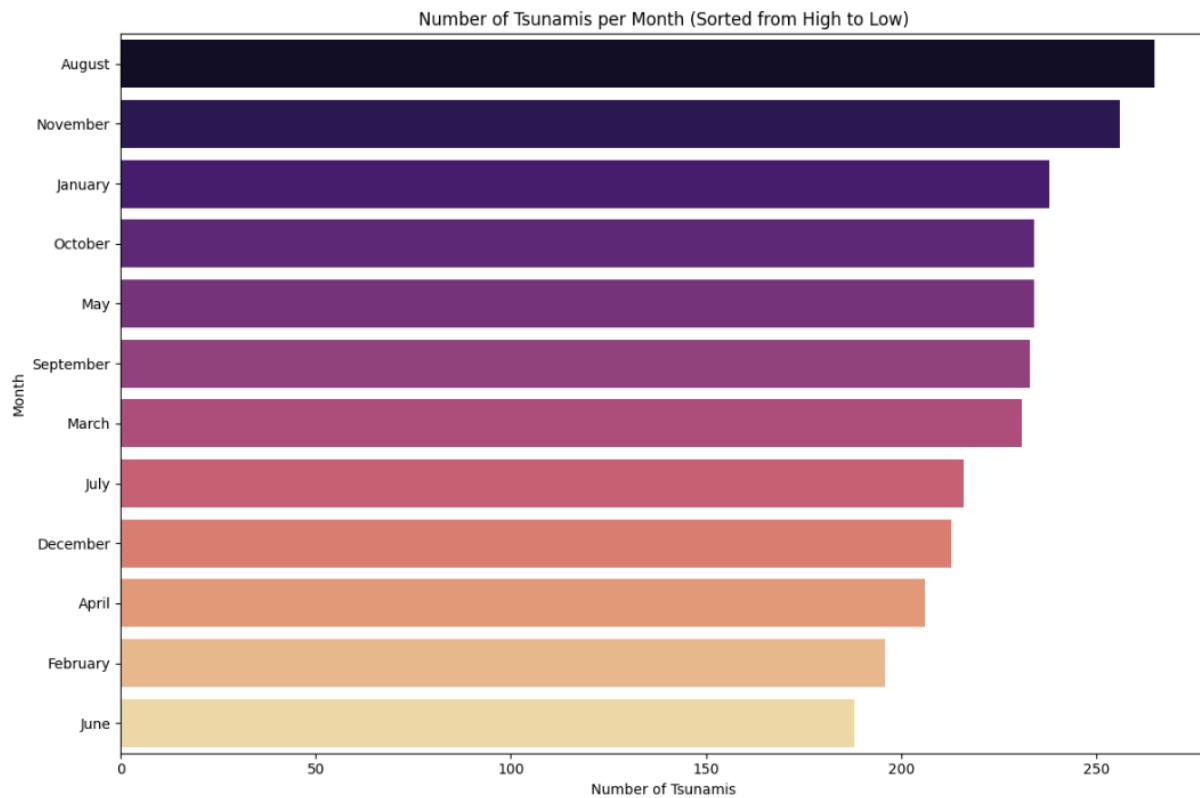
9.2.1 Visualization Number of Tsunami per Country



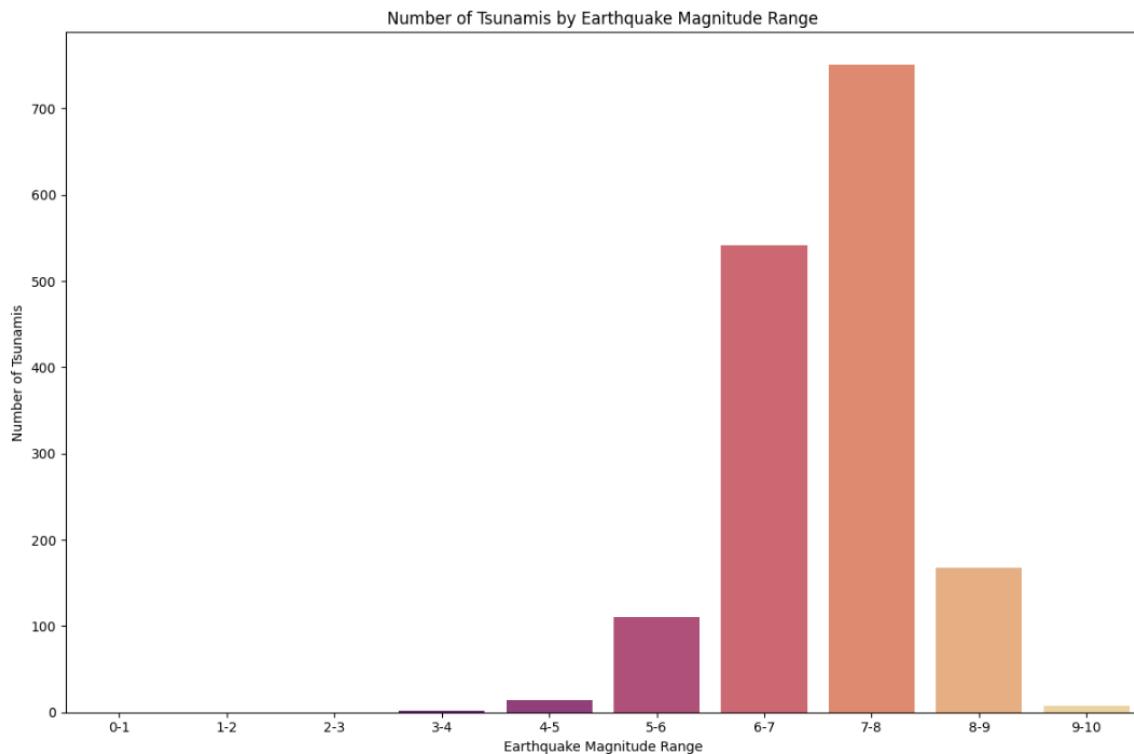
9.2.2 Visualization number of Tsunami per Cause



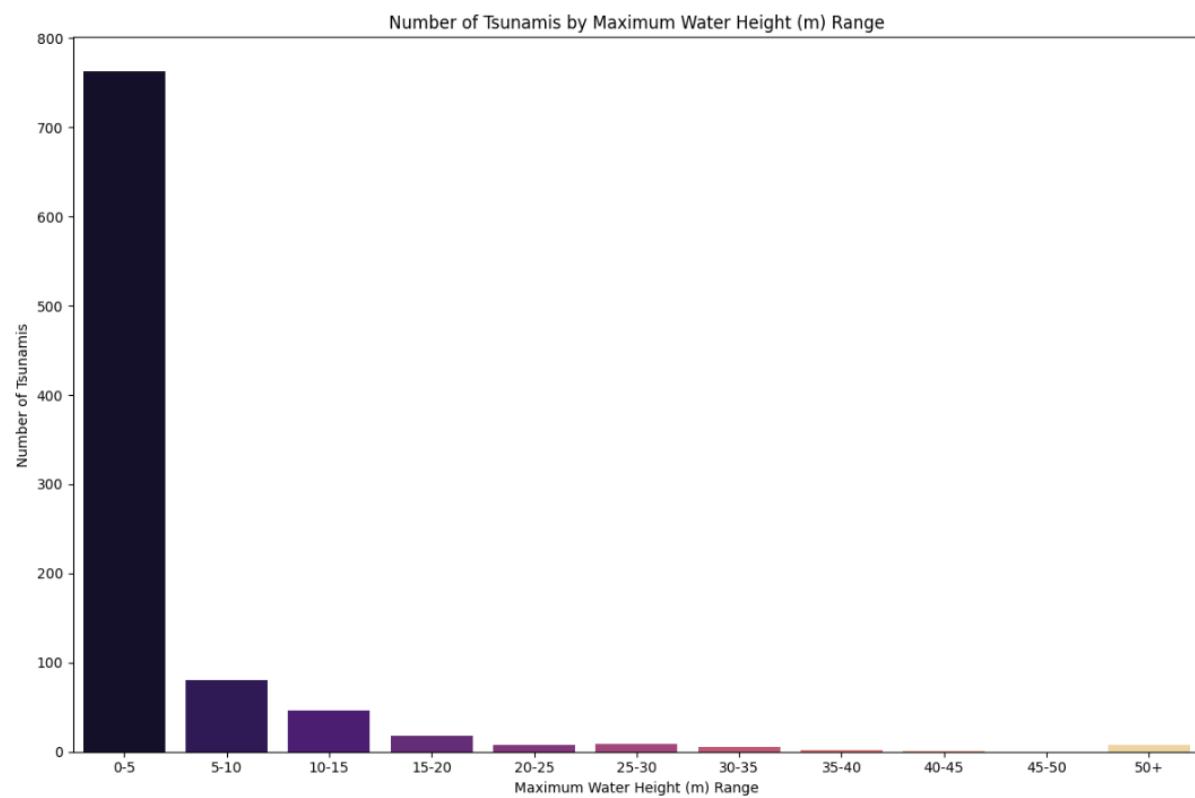
9.2.3 Number of Tsunamis per Month (Sorted from High to Low)



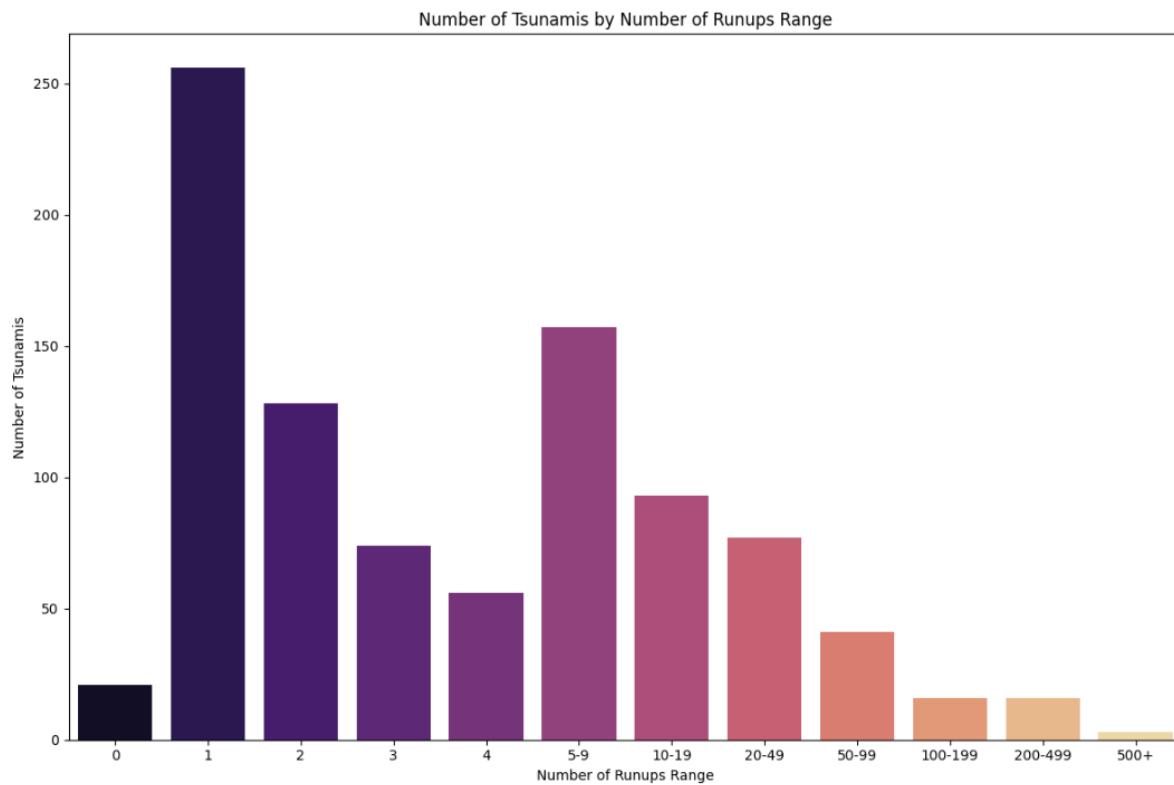
9.2.4 Number of Earthquakes by Magnitude Range



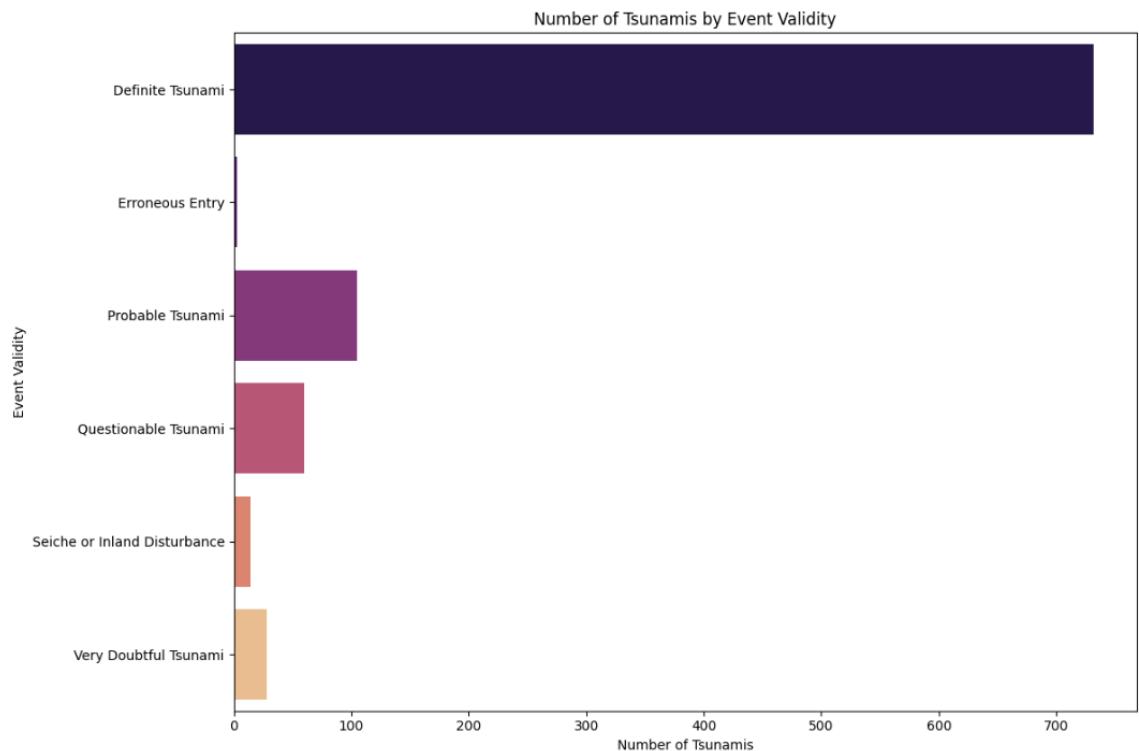
9.2.5 Number of Tsunamis by Maximum Water Height (m) Range



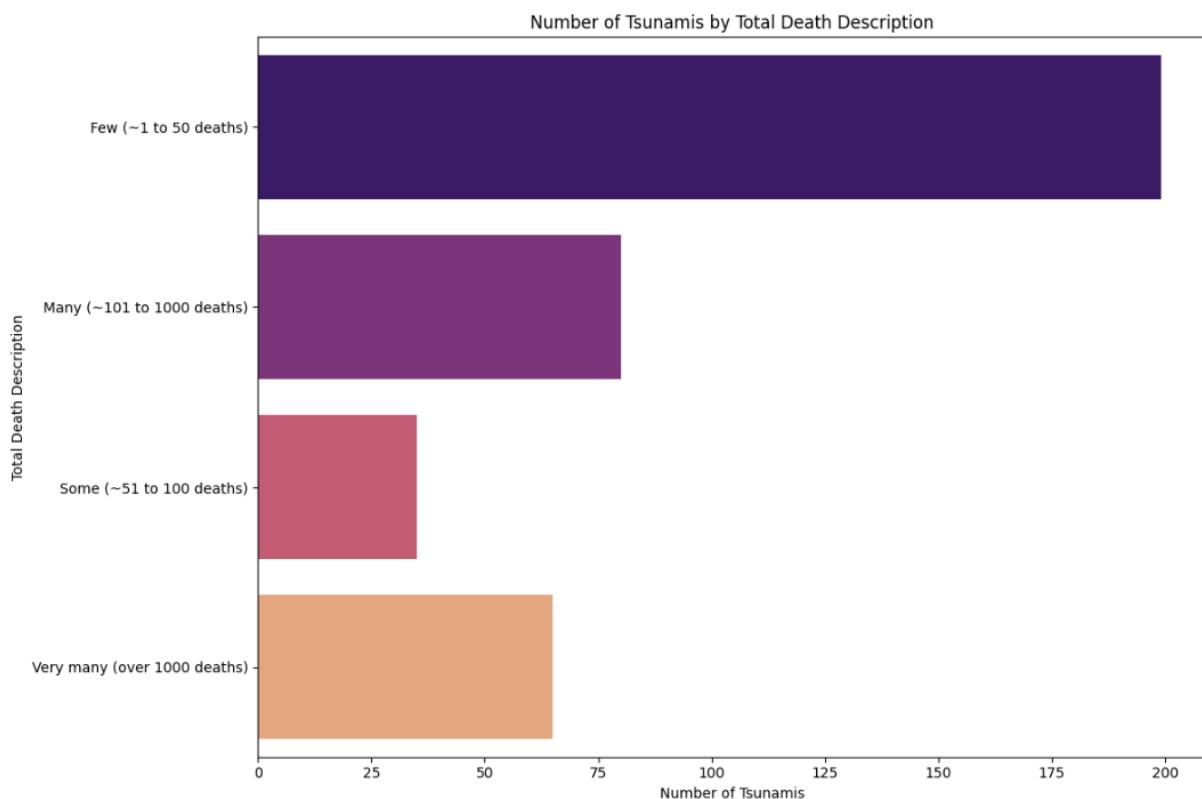
9.2.6 Number of Tsunamis by Number of Runups Range



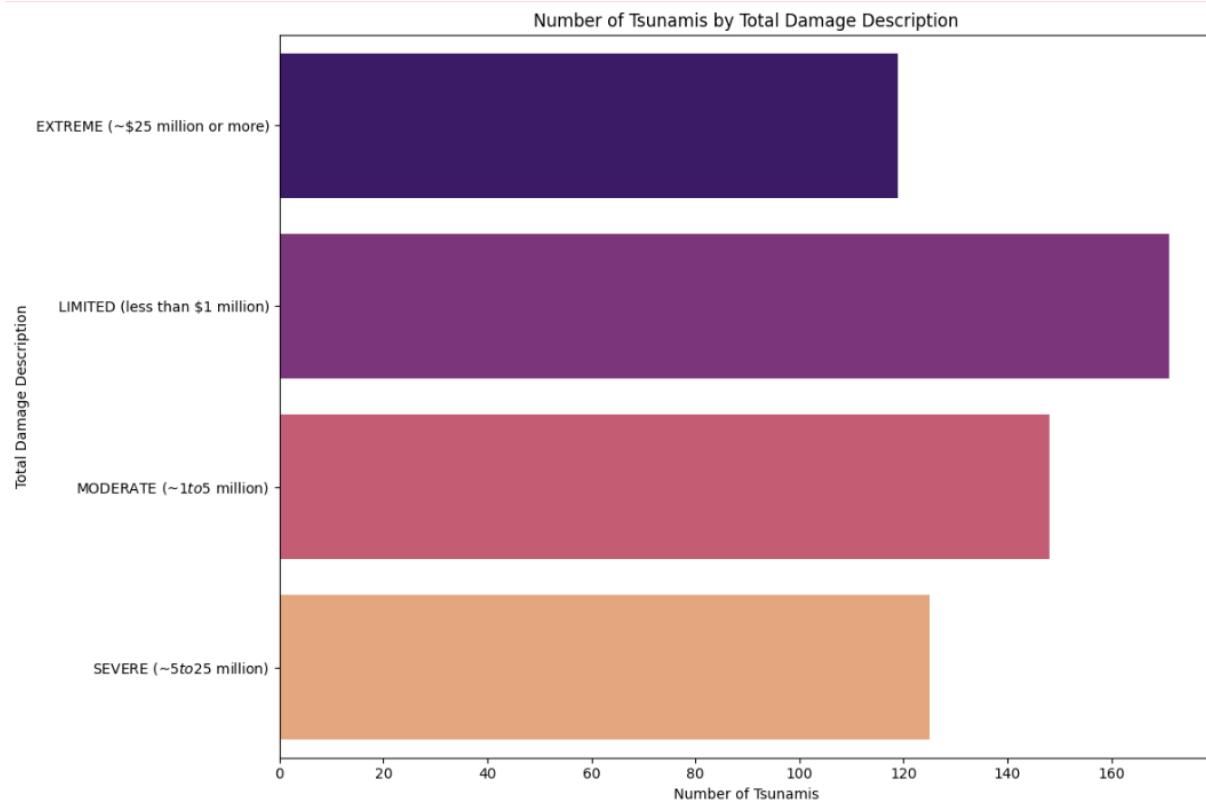
9.2.7 Number of Tsunamis by Event Validity



9.2.8 Number of Tsunamis by Total Death Description

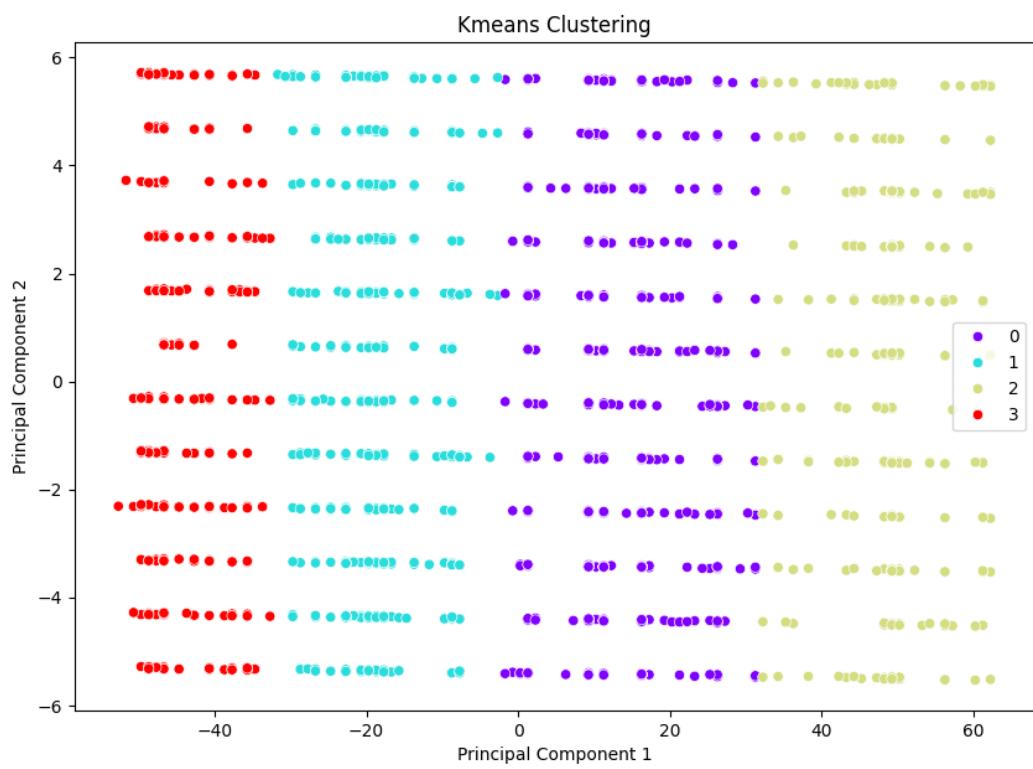


9.2.9 Number of Tsunamis by Total Damage Description

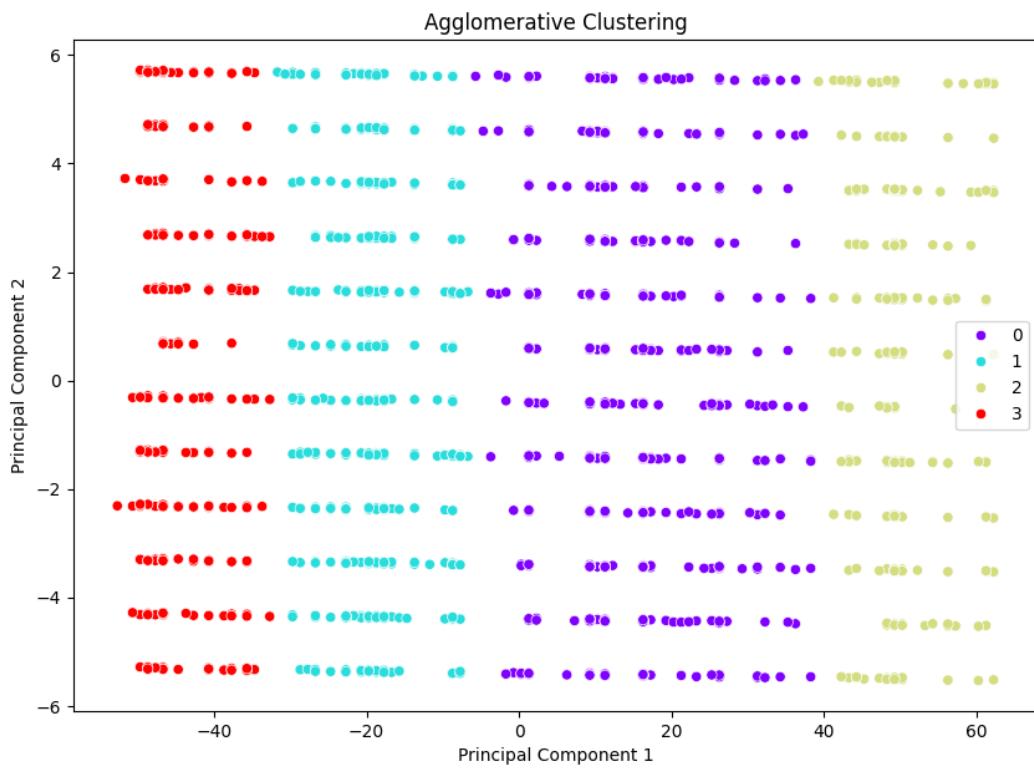


9.3 Clustering Dataset by using some algorithms

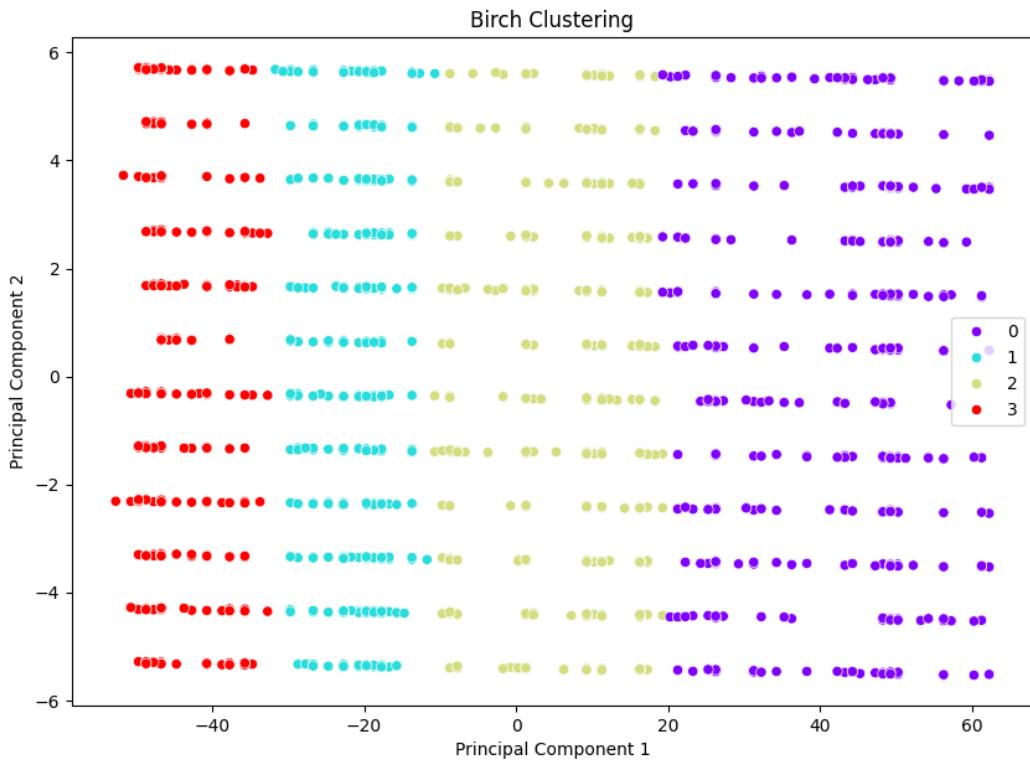
9.3.1 Clustering by using K-means



9.3.2 Clustering by using Agglomerative



9.3.3 Clustering by using Birch



9.3.4 Comparison between algorithm

Algorithm	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
KMeans	0.652535	0.420917	14819.58
Agglomerative	0.64618	0.417298	13931.45
Birch	0.582364	0.512778	10347.99

Overall Conclusion:

- Kmeans performs the best in terms of the Silhouette Score and Calinski-Harabasz Index, suggesting it forms the most well-defined and separated clusters.
- Agglomerative clustering performs the best in terms of the Davies-Bouldin Index, indicating the most optimal clustering with respect to intra-cluster similarity.
- Birch consistently shows lower performance across all three metrics compared to kmeans and agglomerative clustering.

10 Conclusion

After analysis, we have some conclusions. Firstly, the main cause of tsunamis is earthquakes, the year with the highest number of deaths by tsunami is the 2000s. Secondly, Japan has the highest number of Tsunami. Moreover, Most of Tsunami has Earthquake magnitude range in 7-8. Then,

Tsunamis occur most frequently in August each year. Lastly, in data mining, the data used for analysis gives quite good results, highlighting visualization of data on the causes, deaths, and damages of Tsunamis. In Comparison, Kmeans appears to be the most balanced and effective clustering method overall, with agglomerative clustering also being a strong contender, especially in terms of intra-cluster similarity. Birch is less effective in this comparison.