EXTRACT

Source

Volcano_data_2010.csv

Raw Data
Year
Month
Day
TSU
EQ
Name
Location
Country
Latitude
Longitude
Elevation
Type
Status
Time
VEI
Agent
DEATHS
DEATHS_DESCRIPTION
MISSING
MISSING_DESCRIPTION
INJURIES
INJURIES_DESCRIPTION
DAMAGE_MILLIONS_DOLLARS
DAMAGE_DESCRIPTION
HOUSES_DESTROYED
HOUSES_DESTROYED_DESCRIPTION
TOTAL_DEATHS
TOTAL_DEATHS_DESCRIPTION
TOTAL_MISSING
TOTAL_MISSING_DESCRIPTION
TOTAL_INJURIES
TOTAL_INJURIES_DESCRIPTION
TOTAL_DAMAGE_MILLIONS_DOLLARS
TOTAL_DAMAGE_DESCRIPTION
TOTAL_HOUSES_DESTROYED
TOTAL_HOUSES_DESTROYED_DESCRIPTION

TRANSFORM

From

Dataframe: volcano_df

Action	
Keep - Not Null	as Year
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Elevation
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Total_Dealths
Remove	
Keep	as Total_Missing_Persons
Remove	
Keep	as Total_Inquries
Remove	
Remove	
Remove	
Remove	
Remove	
Merge df (country_df)	Country_ID

LOAD

Into

Table: Volcano

Table: Volcano	
Year	Integer
Country_ID	Foreign Key
Elevation	Integer
Total_Deaths	Integer
Total_Missing_Persons	Integer
Total_Injuries	Integer

EXTRACT

Source

Worldwide-Earthquake-database.csv

Raw Data
I_D
FLAG_TSUNAMI
YEAR
MONTH
DAY
HOUR
MINUTE
SECOND
FOCAL_DEPTH
EQ_PRIMARY
EQ_MAG_MW
EQ_MAG_MS
EQ_MAG_MB
EQ_MAG_ML
EQ_MAG_MFA
EQ_MAG_UNK
INTENSITY
COUNTRY
STATE
LOCATION_NAME
LATITUDE
LONGITUDE
REGION_CODE
DEATHS
DEATHS_DESCRIPTION
MISSING
MISSING_DESCRIPTION
INJURIES
INJURIES_DESCRIPTION
DAMAGE_MILLIONS_DOLLARS
DAMAGE_DESCRIPTION
HOUSES_DESTROYED
HOUSES_DESTROYED_DESCRIPTION
HOUSES_DAMAGED
HOUSES_DAMAGED_DESCRIPTION
TOTAL_DEATHS
TOTAL_DEATHS_DESCRIPTION
TOTAL_MISSING
TOTAL_MISSING_DESCRIPTION
TOTAL_INJURIES
TOTAL_INJURIES_DESCRIPTION
TOTAL_DAMAGE_MILLIONS_DOLLARS
TOTAL_DAMAGE_DESCRIPTION
TOTAL_HOUSES_DESTROYED
TOTAL_HOUSES_DESTROYED_DESCRIPTION
TOTAL_HOUSES_DAMAGED
TOTAL_HOUSES_DAMAGED_DESCRIPTION

TRANSFORM

From

Dataframe: earthquake_df	

Action	
Remove	
Remove	
Keep - Not null	as Year
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Magnitude
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	Total_Deaths
Remove	
Keep	
Remove	
Keep	Total_Missing_Persons
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Merge df (country_df)	Country_ID

LOAD

Into

Table: World_Earthquake

Table: World_Earthquake	
Year	Integer
Magnitude	Integer
Country_ID	Foreign Key
Total_Dealths	Integer
Total_Missing_Persons	Integer
Total_Injuries	Integer

EXTRACT

Source

tsunami_dataset.csv

Raw Data
ID
YEAR
MONTH
DAY
HOUR
MINUTE
LATITUDE
LONGITUDE
LOCATION_NAME
COUNTRY
REGION
CAUSE
EVENT_VALIDITY
EQ_MAGNITUDE
EQ_DEPTH
TS_INTENSITY
DAMAGE_TOTAL_DESCRIPTION
HOUSES_TOTAL_DESCRIPTION
DEATHS_TOTAL_DESCRIPTION
URL
COMMENTS

TRANSFORM

From

Dataframe: tsumani_df

Action	
Remove	
Keep	as Year
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Cause
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Total_Deaths
Remove	
Remove	
Merge df (country_df)	Country_ID

LOAD

Into

Table: Tsunami

Table: Tsunami	
Year	Integer
Country_ID	Foreign Key
Cause	Varchar
Total_Deaths	Varchar

EXTRACT

Source

WH_2015.csv , WH_2016.csv , WH_2017.csv , WH_2018.csv

Raw Data
Country
Region
Happiness Rank
Happiness Score
Standard Error
Economy (GDP per Capita)
Family
Health (Life Expectancy)
Freedom
Trust (Government Corruption)
Generosity
Dystopia Residual

TRANSFORM

From

The following csv files were merged WH_2015,2016,2017,2018 into: Dataframe: WH_df

Action	
Remove	
Remove	
Keep (for yr 2015-2018)	as Rank (2015, 2016, 2017, 2018)
Keep (for yr 2015-2018)	as Score (2015, 2016, 2017, 2018)
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Merge df (country_df)	Country_ID

LOAD

Into

Table: World_Happiness

Table: World_Happiness	
Country_ID	Foreign Key
Rank_2015	Varchar
Score_2015	Varchar
Rank_2016	Varchar
Score_2016	Varchar
Rank_2017	Varchar
Score_2017	Varchar
Rank_2018	Varchar
Score_2018	Varchar

EXTRACT

Source

UNSD - Methodology.csv

Raw Data
Global Code
Global Name
Region Code
Region Name
Sub-region Code
Sub-region Name
Intermediate Region Code
Intermediate Region Name
Country or Area
M49 Code
ISO-alpha2 Code
ISO-alpha3 Code
Least Developed Countries (LDC)
Land Locked Developing Countries (LLDC)
Small Island Developing States (SIDS)
Developed / Developing Countries

TRANSFORM

From

country_df

Action	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Remove	
Keep	as Country_Name
Keep	as Country_ID
Keep	as Country_Abbreviation
Remove	
Remove	
Remove	
Remove	
Keep	as Developed_Or_Developing

LOAD

Into

Table: Country

Table: Country	
Country_Abbreviation	Varchar
Country_ID	Primary Key
Developed_Or_Developing	Varchar
Country_Name	Varchar