INTRODUCTION TO DATASCIENCE

UE18CS203

PROJECT FINAL PRESENTATION

TSUNAMI WAVES ANALYSIS

PRESENTED BY:

UTKARSH VERMA PES1201800330

NAYAN VARMA PES1201800208

SARAN SAI PES1201800845

DESCRIPTION OF DATASET

NAME OF DATASET-->waves.csv

SOURCE:https://www.kaggle.com/noaa/seismic-waves#sources.csv

Dataset conatins 30 rows and 26,204 columns

The column headings are as follows:

SOURCE_ID	WAVE_ID	YEAR	MONTH	DAY	REGION_ CODE	COUNTRY	STATE/ PROVINCE	LOCATION	LATITUDE
LONGITUDE	DISTANCE_ FROM_ SOURCE	TRAVEL_ TIME_ HOURS	TRAVEL_ TIME_ MINUTES	VALIDITY	MEASUREM ENT_ TYPE	PERIOD	FIRST_ MOTION	MAXIMUM_ HEIGHT	HORIZONTAL_ INUNDATION
INJURIES	INJURY_ ESTIMATE	FATALITIES	FATALITY_ ESTIMATE	DAMAGE_ MILLIONS_ DOLLARS	DAMAGE_ ESTIMATE	HOUSES_ DAMAGED	HOUSE_ DAMAGE_ ESTIMATE	HOUSES_ DESTROYE D	HOUSE_ DESTRUCTION_ ESTIMATE

DATASET BEFORE CLEANING

SOURCE_ID	WAVE_ID	YEAR	MONTH	DAY	REGION_CODE	COUNTRY	STATE/PROVINCE	LOCATION	LATITUDE	LONGITUDE	DISTANCE_FROM_SOURCE	TRAVEL_TIME_HOUR
1	11014	-2000			50	SYRIA		UGARIT	35.583	35.75	12	
3	17601	-1610			50	SYRIA		UGARIT	35.583	35.75	935	
3	1	-1610			50	GREECE		N. AND E. COAST CRETE	35.5	25	106	
9	2	-479			50	GREECE		POTIDAEA, MACEDONIA	40.3	23.33	67	
10	19364	-426	6		50	GREECE		TARFI				
10	19365	-426	6		50	GREECE		THERMOPLYLES	38.8	22.55	17	
10	5	-426	6		50	GREECE		DAPHNUS	38	23.68	132	
10	6	-426	6		50	GREECE		SKOPELOS	39.12	23.72	91	
10	3	-426	6		50	GREECE		ATALANTI	38.651	22.999	38	
10	17612	-426	6		50	GREECE		THRONIUM	38.817	22.733	10	
10	17613	-426	6		50	GREECE		LICHADES ISLAND	38.815	22.816	14	
10	7	-426	6		50	GREECE		OPOUS	38.633	23.0833	45	
10	4	-426	6		50	GREECE		CENEUM	40.173	22.487	143	
10	17611	-426	6		50	GREECE		OROBIES	38.817	23.233	47	
11	8	-373			50	GREECE		HELICE, PELOPONNESUS	38.21	22.13	11	
3092	17621	-227			50	GREECE		ISLAND OF RHODES	36.167	28	31	
5382	17622	-218			73	SPAIN		CADIZ	36.533	-6.3	86	
4281	17623	-210			73	SPAIN		CADIZ	36.533	-6.3	381	
15	19366	-138			50	LEBANON		SIDON	33.5631	35.3689		
15	19367	-138			50	ISRAEL		AKKO (ACRE)	32.924	35.073		
15	19368	-138			50	LEBANON		SUR [TYRE]	33.2733	35.1939		
3714	17631	-58			50	ALBANIA		DURAZZO	41.323	19.441	6	
19	16093	-23			50	EGYPT		PELUSIUM	31.133	32.667		
19	16092	-23			50	EGYPT		ALEXANDRIA	31.2	29.919		
23	17632	62			50	GREECE		LEBENA	34.931	24.929	34	
4396	16533	115	12	13	50	ISRAEL		CAESAREA	32.483	34.883	418	
3583	25334	123			84	SOUTH KOREA		KYONGJU	35.8428	129.2117	5	
27	10	142			50	GREECE		ISLAND OF SYMI	36.58	27.83	20	
27	12	142			50	GREECE		ISLAND OF SERIFOS	37.15	24.5	315	
27	11	142			50	GREECE		ISLAND OF KOS	36.8933	27.2889	67	
27	9	142			50	GREECE		ISLAND OF RHODES	36.167	28	59	
29	25331	173	6	28	84	CHINA		YEHSIAN, SHANDONG	37.1717	119.9214	37	
29	25330	173	6	28	84	CHINA		HUANGXIAN, SHANDONG	37.6481	120.5261	49	
29	25332	173	6	28	84	CHINA		CHANYI, SHANDONG	36.8536	119.3908	90	
3427	10864	258			50	ITALY		ROME	41.9	12.45	0	
34	10865	342	7		50	GREECE		THRACE	41.3	19.5		
3474	25838	346			50	ALBANIA		DURRES (DYRRACHIUM)	41.3231	19.4414	9	
36	10866	348			50	SYRIA		ARWAD ISLAND	34.85	35.85	110	
36	10868	348			50	SYRIA		SYRIAN COAST				
36	10867	348			50	LEBANON		BEIRUT	33.53	35.3	45	
2314	25832	358	8	24	50	TURKEY		IZMIT (NICOMEDEA)	40.77	29.94	3	
38	19371	365	7	21	50	CROATIA		EPIDAURUS	42.5811	18.2181	939	
38	10857	365	7	21	50	ITALY	REGIONE SICILIANA	ISLAND OF SICILY	38.18	15.56	752	
38	19370	365	7	21	50	SPAIN		ADRA	36.7333	-3.0167	2346	
38	19369	365	7	21	73	PORTUGAL		CAPE SOUTH VICENTE	37.025	-8.9944	2874	
38	17	365	7	21	50	GREECE		ACHAEA, PELOPONNESUS	38.25	21.75	378	
38	16	365	7	21	50	GREECE		KNOSSOS, CRETE	35.3	25.167	200	

CLEANING:

- The orignal data had nearly 50% of its data missing.
- The first step involved pairing up columns with high correlation,linear regression was then applied on those columns which helped to reduce the number of missing values appreciably.
- The INJURY_ESTIAMTE is a catogorical data which is used to estimate INJURY as follows:
  - 0 = None
  - 1 = Few (~1 to 50 injuries)
  - 2 = Some(~51 to 100 injuries)
  - 3 = Many (~101 to 1000 injuries)
  - 4 = Very Many (~1001 or more
    - injuries)


```
for i in data['INJURIES'].index:
        if(np.isnan(data.loc[i,'INJURIES']) and not(np.isnan(data.loc[i,'INJURY_ESTIMATE']))):
                if(data.loc[i,'INJURY_ESTIMATE']==0):
                        data.loc[i,'INJURIES']=0
                elif(data.loc[i,'INJURY_ESTIMATE']==1):
                        data.loc[i,'INJURIES']=25
                elif(data.loc[i,'INJURY_ESTIMATE']==2):
                        data.loc[i,'INJURIES']=76
                elif(data.loc[i,'INJURY_ESTIMATE']==3):
                        data.loc[i,'INJURIES']=551
                else:
                        data.loc[i,'INJURIES']=1001
```

CLEANING

- The FATALITY_ESTIAMTE is a catogorical data which is used to estimate FATALITY as follows:
    - 0 = None
    - 1 = Few (~1 to 50 deaths)
    - 2 = Some (~51 to 100 deaths)
    - 3 = Many (~101 to 1000 deaths)
    - 4 = Very Many (~1001 or more
    - deaths)
- THE DAMAGE_ESTIMATE was similarly
- used to estimate
- DAMAGE_MILLIONS_DOLLARS
    - 0 = NONE
    - 1 = LIMITED (roughly corresponding to
    - less than $1 million)
    - 2 = MODERATE (~$1 to $5 million)
    - 3 = SEVERE (~>$5 to $24 million)
    - 4 = EXTREME (~$25 million or more)

```python
for i in data['HOUSES_DAMAGED'].index:
    if(np.isnan(data.loc[i,'HOUSES_DAMAGED']) and not(np.isnan(data.loc[i,'HOUSE_DAMAGE_ESTIMATE']))):
        if(data.loc[i,'HOUSE_DAMAGE_ESTIMATE']==0):
            data.loc[i,'HOUSES_DAMAGED']=0
        elif(data.loc[i,'HOUSE_DAMAGE_ESTIMATE']==1):
            data.loc[i,'HOUSES_DAMAGED']=25
        elif(data.loc[i,'HOUSE_DAMAGE_ESTIMATE']==2):
            data.loc[i,'HOUSES_DAMAGED']=76
        elif(data.loc[i,'HOUSE_DAMAGE_ESTIMATE']==3):
            data.loc[i,'HOUSES_DAMAGED']=551
        else:
            data.loc[i,'HOUSES_DAMAGED']=1001

for i in data['HOUSES_DESTROYED'].index:
    if(np.isnan(data.loc[i,'HOUSES_DESTROYED']) and not(np.isnan(data.loc[i,'HOUSE_DESTRUCTION_ESTIMATE']))):
        if(data.loc[i,'HOUSE_DESTRUCTION_ESTIMATE']==0):
            data.loc[i,'HOUSES_DESTROYED']=0
        elif(data.loc[i,'HOUSE_DESTRUCTION_ESTIMATE']==1):
            data.loc[i,'HOUSES_DESTROYED']=25
        elif(data.loc[i,'HOUSE_DESTRUCTION_ESTIMATE']==2):
            data.loc[i,'HOUSES_DESTROYED']=76
        elif(data.loc[i,'HOUSE_DESTRUCTION_ESTIMATE']==3):
            data.loc[i,'HOUSES_DESTROYED']=551
        else:
            data.loc[i,'HOUSES_DESTROYED']=1001
```

CLEANING

- HOUSES_DESTROYED was also estimated via same procedure(using HOUSE_DESTRUCTION_ESTIMATE.
  - 0 = None
  - 1 = Few (~1 to 50 houses)
  - 2 = Some (~51 to 100 houses)
  - 3 = Many (~101 to 1000 houses)
  - 4 = Very Many (~1001 or more houses)
- The following columns were found to be either irrelevant or having to many missiing values and were dropped.
- An python module called reverse_geocoder provides with the functionality to estimate STATE/PROVINCE based on latitude and longitude

```
for i in data.index:
        if(pd.isnull(data.loc[i,'STATE/PROVINCE'])):
                if(not(pd.isnull(data.loc[i,'LATITUDE']) and pd.isnull(data.loc[i,'LONGITUDE']))):
                        coordinates=((data.loc[i,'LATITUDE']),(data.loc[i,'LONGITUDE']))
                        result = rg.search(coordinates)
                        data.loc[i,'STATE/PROVINCE']=result[0]['admin2']
                        print(result[0]['admin2'])

print(data.describe())
```

CLEANING:

- A dictionary was created which had the the country with the STATE/PROVINCE with highest frequency.This was used to replace the missing values corresponding to a particular COUNTRY.
- Finally the rest of numeriacal values were dealt by replacing with mean and categorical values were replace by 'ffill' method.

YEAR	MONTH	DAY	REGION_CODE	COUNTRY	STATE/PROVINCE	LOCATION	LATITUDE	LONGITUDE	DISTANCE_FROM_SOURCE	TRAVEL_TIME_HOURS	MEASUR
1802	3	19	74	ANTIGUA AND BARBUDA	ANTIGUA	ANTIGUA ISLAND	17.05	-61.8	66	0.220827945323144	
1802	3	19	74	SAINT KITTS AND NEVIS	SAINT KITTS	SAINT CHRISTOPHER (SAINT KITTS)	17.333	-62.75	40	0.183287870315442	
1802	8	15	74	VENEZUELA	Municipio Pedernales	ORINOCO RIVER	9.8	-62.3	247	0.482164621338293	
1802	8	15	83	INDONESIA	MALUKU	AMBON ISLAND	-3.683	128.187	2	0.128421606842649	
1802	11	7	50	ALGERIA	MALUKU	ALGIERS	36.763	3.051	32	0.171737078005381	
1802	12	9	85	JAPAN	NIIGATA	OGI	37.28	137.25	104	0.275694208795937	
1804	1	13	50	SPAIN	Provincia de Almeria	ALMERIA	36.833	-2.433	36	0.177512474160412	
1804	7	10	85	JAPAN	YAMAGATA	SAKATA	38.917	139.85	17	0.150079342424015	
1804	7	10	85	JAPAN	AKITA	KISAGATA	39.2167	139.9	19	0.15296704050153	
1805	5	8	81	AUSTRALIA	North Sydney	NORFOLK ISLAND	-29.058	167.955	1	0.126977757803891	
1805	7	26	50	ITALY	CAMPANIA	BAY OF NAPLES	40.83	14.25	77	0.236710284749479	
1806	3	25	88	USA	CA	SANTA BARBARA, CA	34.42	-119.68	3	0.129865455881406	
1806	12	1	89	PERU	Callao	CALLAO	-12.05	-77.15	8	0.137084701075195	
1808	4	2	50	FRANCE	Departement des Bouches-du-Rhone	MARSEILLES	43.3	5.366	227	0.453287640563138	
1808	8	8	85	JAPAN	TOKUSHIMA	NAKA RIVER	33.94	134.67	13	0.144303946268984	
1809	12	4	70	SOUTH AFRICA	City of Cape Town	TABLE BAY	-33.88	18.45	0	0.125533908765133	
1810	2	16	50	EGYPT	City of Cape Town	RASHID (ROSETTA)	31.4	30.417	679	1.10590740608163	
1810	2	16	50	EGYPT	City of Cape Town	ALEXANDRIA	31.2	29.919	661	1.07991812338399	
1810	9	25	85	JAPAN	AKITA	OGA	39.9	139.9	0	0.125533908765133	
1811	11	19	89	CHILE	VALPARAISO	VALPARAISO	-33.033	-71.633	22	0.157298587617803	
1811	12	16	75	USA	MI	ORCHARD LAKE, MI	46.257	-89.571	1187	1.83938271777056	
1811	12	16	75	USA	AR	NEW MADRID, AR	36.5837	-89.5266	135	0.320453528997427	
1811	12	16	75	USA	LA	LAKE BISTINEAU, TX-LA BORDER	32.3167	-93.4167	459	0.788260617554932	
1812	3	26	74	VENEZUELA	Municipio Vargas	LA GUAIRA	10.6	-66.933	4	0.131309304920164	
1812	6	23	50	FRANCE	Departement des Bouches-du-Rhone	MARSEILLES	43.3	5.366	0	0.125533908765133	
1812	11	11	74	JAMAICA	Departement des Bouches-du-Rhone	ANNOTTO BAY	18.267	-76.767	30	0.168849379927865	
1812	12	21	80	USA	HI	HOOKENA, HAWAII, HI	19.38	-155.9	3909	5.76953980126911	
1812	12	21	88	USA	CA	SANTA BARBARA, CA	34.42	-119.68	32	0.171737078005381	
1812	12	21	88	USA	CA	VENTURA, CA	34.27	-119.28	58	0.209277153013082	
1812	12	21	88	USA	CA	EL REFUGIO (GAVIOTA), CA	34.47	-120.2	41	0.1847317193542	
1813	5	17	50	ITALY	NAPLES	TORRE DEL GRECO	40.7853	14.363	7	0.135640852036437	
1815	4	10	60	INDONESIA	WEST NUSA TENGGARA	SUMBAWA ISLAND	-8.84	118.08	72	0.22949103955569	
1815	4	10	83	INDONESIA	WEST SULAWESI	SULAWESI I	-2.823	118.806	605	0.999062577213561	
1815	4	10	60	INDONESIA	EAST JAVA	SUMENEP, JAVA	-7.017	113.867	474	0.809918353136298	
1815	4	10	60	INDONESIA	WEST NUSA TENGGARA	BIMA, SUMBAWA ISLAND	-8.467	118.717	84	0.246817228020783	
1815	4	11	83	INDONESIA	MALUKU	AMBON ISLAND	-3.683	128.187	24	0.160186285695319	
1815	11	22	60	INDONESIA	BALI	BALI ISLAND	-8.5	115	56	0.206389454935566	
1816	5	1	60	MALAYSIA	Langkawi	PENANG ISLAND	5.411	100.335	427	0.742057448314685	
1817	1	8	75	USA	PA	PHILADELPHIA, PENNSYLVANIA	39.95	-75.15	4	0.131309304920164	
1818	2	23	50	ITALY	LIGURIA	GENOA	44.398	8.95	91	0.256924171292087	
1818	2	23	50	FRANCE	Departement des Alpes-Maritimes	ANTIBES	43.5	7.1	88	0.252592624175814	
1818	3	18	60	INDONESIA	BENGKULU	FORT MALBORO, BENGKULU, SUMATRA	-3.767	102.267	89	0.254036473214571	
1818	11	8	60	INDONESIA	WEST NUSA TENGGARA	BIMA, SUMBAWA ISLAND	-8.467	118.717	250	0.486496168454566	
1818	12	9	50	ITALY	LIGURIA	HARBOUR OF GENOA	44.398	8.95	121	0.300239642454819	
1819	1	8	50	ITALY	LIGURIA	GENOA	44.398	8.95	109	0.282913453989726	
1819	4	12	80	USA	HI	HAKAAANO, MOLOKAI, HI	21.176	-156.788	10621	15.460654549411	
1819	4	12	89	CHILE	COPIAPO	COPIAPO	-27.366389	-70.333056	122	0.301683491493577	

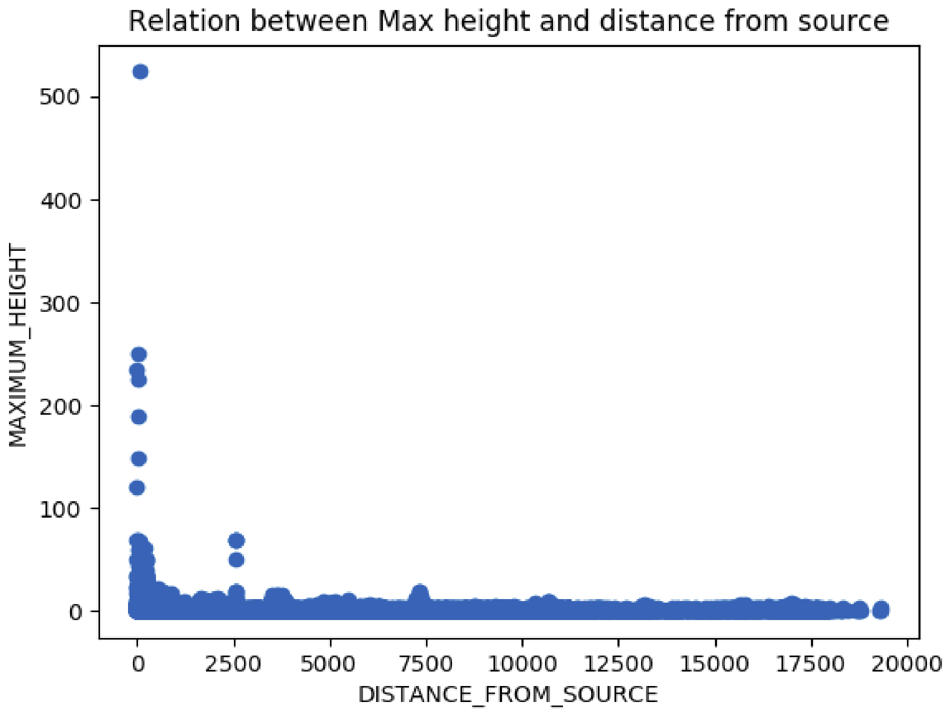
NORMALIZATION:

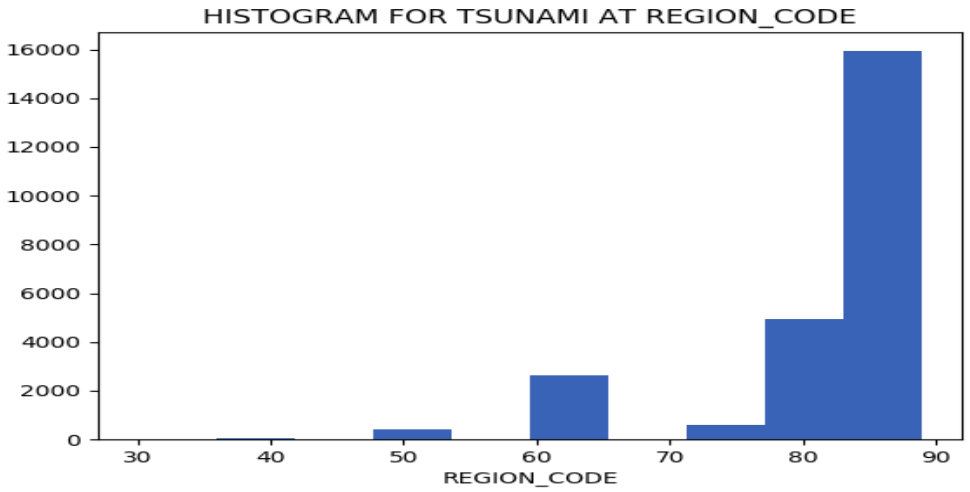
All numerical data were normalized using the preprocessing.normalize() function.

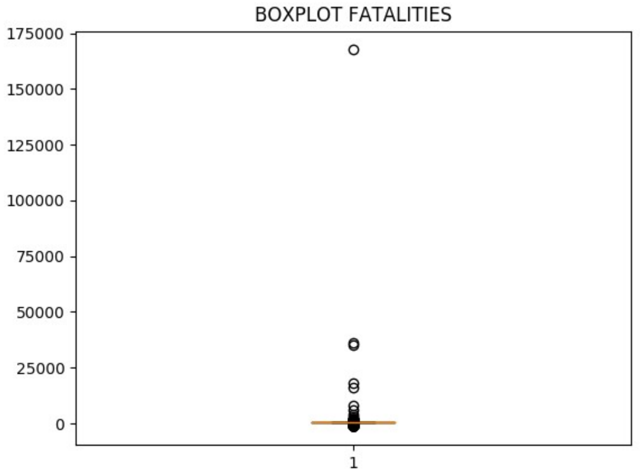
A1 YEAR

	G	H	I	J	K	L	M	N	O	P	Q
1	LOCATION	LATITUDE	LONGITUDE	DISTANCE_FROM_SOURCE	TRAVEL_TIME_HOURS	MEASUREMENT_TYPE	PERIOD	MAXIMUM_HEIGHT	HORIZONTAL_INUNDATION	FATALITIES	FATALITY_EST
2	ANTIGUA ISLAND	0.027289765084692	-0.098915394852433	0.105637800327841	0.000353451187801	1	0.038026356390124	0.006822262606343	0.25773782018258	0.577806756338646	
3	SAINT CHRISTOPHER (SAINT KITTS)	0.027836588948709	-0.100775743179572	0.064239517564667	0.000294358109113	1	0.038155010703965	0.006845344321179	0.258609822802714	0.579761646021121	
4	ORINOCO RIVER	0.014660115333231	-0.093196447475539	0.369494743602859	0.000721284587594	1	0.035540260896688	0.006376235220846	0.240887380275316	0.540030779111872	
5	AMBON ISLAND	-0.008566407277731	0.298154235598835	0.00465186384889	0.000298699915144	1	0.055259417155216	0.00991402519449	0.374541320131423	0.839661424724655	
6	ALGIERS	0.059305625500877	0.004921836177765	0.051622011697306	0.000277044170302	1	0.038326066328412	0.006876033203866	0.259769216127774	0.582360819460234	
7	OGI	0.058015147075237	0.213588490774578	0.161844830896584	0.000429035409632	1	0.036972157477942	0.006633130054582	0.250592593675018	0.561788307246796	
8	ALMERIA	0.059397393931237	-0.003923488704007	0.058054086865705	0.000286259016518	1	0.038312421141362	0.006873585137362	0.259676730780759	0.582153482181098	
9	SAKATA	0.074333788236681	0.267121830688385	0.032471012668592	0.000286660484067	1	0.045379134871085	0.001910059568741	0.307574020067887	0.689531504315389	
10	KISAGATA	0.090005147050121	0.321080561911427	0.043606366521209	0.000351070359672	1	0.054526246196399	0.002295071922169	0.369571980371179	0.828520963902967	
11	NORFOLK ISLAND	-0.066964382044391	0.387053575134754	0.002304507607006	0.00029262120878	1	0.054750418898574	0.011522538035032	0.371091394511477	0.831927246129297	
12	BAY OF NAPLES	0.06540727252196	0.022827666750868	0.123349497531005	0.000379196034861	1	0.038058876157966	0.006828096938558	0.25795823504406	0.578300891021986	
13	SANTA BARBARA, CA	0.054600775900287	-0.189849531079209	0.004758928753657	0.000206006817367	1	0.037687492978112	0.006761467531456	0.255441046958878	0.57265776002335	
14	CALLAO	-0.029279050826232	-0.187458819190355	0.019438373992519	0.000333087961019	1	0.057727034376124	0.002429796749065	0.391266516651358	0.877156626412417	
15	MARSEILLES	0.065629097481026	0.008133157900305	0.344060164623394	0.00068704061777	1	0.036009561752817	0.000227352531689	0.244068241949979	0.547161759599318	
16	NAKA RIVER	0.053575489056789	0.212581352718848	0.020520959273372	0.000227788877259	1	0.03750279245376	0.002367802993081	0.254189170233044	0.569851253668257	
17	TABLE BAY	-0.054718678255903	0.029798099581506	0	0.000202746445216	1	0.038370856810664	0.006884069010112	0.260072800335055	0.583041406445723	
18	RASHID (ROSETTA)	0.039680499885244	0.038438272771002	0.858059217263708	0.001397546455418	1	0.030023186696253	0.005386423611553	0.203493351154633	0.456199377661559	
19	ALEXANDRIA	0.040211201181711	0.03856022205627	0.851910384009966	0.001391820670486	1	0.030619757913212	0.005493453732023	0.207536835191474	0.465264218801207	
20	OGA	0.062850127496647	0.220369244029597	0	0.000197739904036	1	0.037423341925591	0.00157519116533	0.253650664630967	0.56864401068395	
21	VALPARAISO	-0.05299143213232	-0.114913427721807	0.035292329092454	0.000252337887272	1	0.038112456329333	0.006416787107719	0.258321394649241	0.579115036471631	
22	ORCHARD LAKE, MI	0.034468219585776	-0.066743474425871	0.884488329297527	0.001370608716916	7	0.017703155665732	0.003176101745682	0.119989743556486	0.268997714301944	
23	NEW MADRID, AR	0.083749661087924	-0.20494981120975	0.309050321505746	0.000733601971587	7	0.054388205714861	0.009757722203723	0.368636359496974	0.826423452322772	
24	LAKE BISTINEAU, TX-LA BORDER	0.041639832651317	-0.120366737780721	0.591418158009768	0.001015668066375	7	0.030611969297561	0.00549205638591	0.207484044942532	0.465145871550166	
25	LA GUAIRA	0.025873410570729	-0.163375942427412	0.009763551158765	0.000320511279052	1	0.057990534968821	0.010404011738089	0.39305249024509	0.881160492078581	
26	MARSEILLES	0.069894807670141	0.008661790714965	0	0.000202636683816	1	0.03835008387441	0.006880342162744	0.259932004008055	0.582725763716415	
27	ANNOTTO BAY	0.029287942413813	-0.123082469769592	0.048099757618349	0.000270720474951	1	0.038091750713803	0.006833994922859	0.258181054613571	0.578800416674127	
28	HOOKENA, HAWAII, HI	0.004893118042971	-0.03936207961296	0.986955543342261	0.001456709974264	1	0.005998480073742	0.000757448613463	0.040656937066323	0.091146316486712	
29	SANTA BARBARA, CA	0.054532149374433	-0.189610913339108	0.050698105170884	0.000272085763827	1	0.037640124389138	0.00316863157318	0.255119988671254	0.571937998959039	
30	VENTURA, CA	0.080332974256463	-0.279606570449691	0.135958929293109	0.000490570649124	1	0.055691516301297	0.004688238941142	0.377470033334214	0.846227128876076	
31	EL REFUGIO (GAVIOTA), CA	0.081123696111142	-0.282885647593828	0.096491776633502	0.000434758336611	1	0.055913379996518	0.008001757086681	0.378973797318255	0.849598325968152	
32	TORRE DEL GRECO	0.065834290699463	0.023184282506599	0.011299169918972	0.000218947005016	1	0.038349331692937	0.006880207214776	0.259926905817439	0.582714334392691	
33	SUMBAWA ISLAND	-0.021028092416053	0.280882030824375	0.171269530990473	0.000545900315434	1	0.007136230457936	0.008325602200926	0.383045486468869	0.858726398438342	
34	SULAWESI I	-0.003235069783351	0.136147963400903	0.693311094200178	0.001144894493525	1	0.003437906252232	0.004884563880641	0.184533624661204	0.413694719018619	
35	SUMENEP, JAVA	-0.008914797493622	0.144663138977667	0.602196666948407	0.001028966524808	1	0.030183479709524	0.0019056856549	0.204579796866317	0.458635014279272	
36	BIMA, SUMBAWA ISLAND	-0.019715320694663	0.276431289347855	0.195593118973861	0.000574711326727	1	0.055320180229864	0.000698546853478	0.374953164543937	0.840584713685282	
37	AMBON ISLAND	-0.005831350628007	0.20296045152112	0.037999569663904	0.000253625413437	1	0.037616357299404	0.006748705165379	0.25495889835176	0.571576860361225	
38	BALI ISLAND	-0.006503932650903	0.087994382923983	0.042849438641244	0.000157922719383	1	0.018178850162149	0.003261445519922	0.123213940514028	0.918202256598086	
39	PENANG ISLAND	0.007140129137264	0.132397866750581	0.563451329072587	0.000979187951993	1	0.031350020771706	0.005624469308199	0.212486464216604	0.476360491323663	
40	PHILADELPHIA, PENNSYLVANIA	0.064043412730288	-0.12047215185685	0.006412356718927	0.000210500525916	1	0.038086142070109	0.00048092675392	0.258143039938039	0.578715193883205	
41	GENOA	0.070892949970053	0.014291001897202	0.14530515895479	0.000410246236801	1	0.037935773345179	0.006806011211811	0.25712385979389	0.576430355853618	
42	ANTIBES	0.06951629909176	0.011346338472448	0.14063067402471	0.000403662170358	1	0.037967035331464	0.006811619886959	0.257335749571521	0.576905378669548	

VISUALIZATIONS:







INSIGHTS FROM THE GRAPHS

- The histogram shows that the maximum number tsunamis occur at region code 80-90.
- The scatter plot has a slightly decreasing tendency which shows that as the height increases the distance decreases.

HYPOTHESIS TESTING

1)

NULL HYPOTESIS:mean latitude=23.141

p-value=0.201297892931

null hypothesis is plausible.

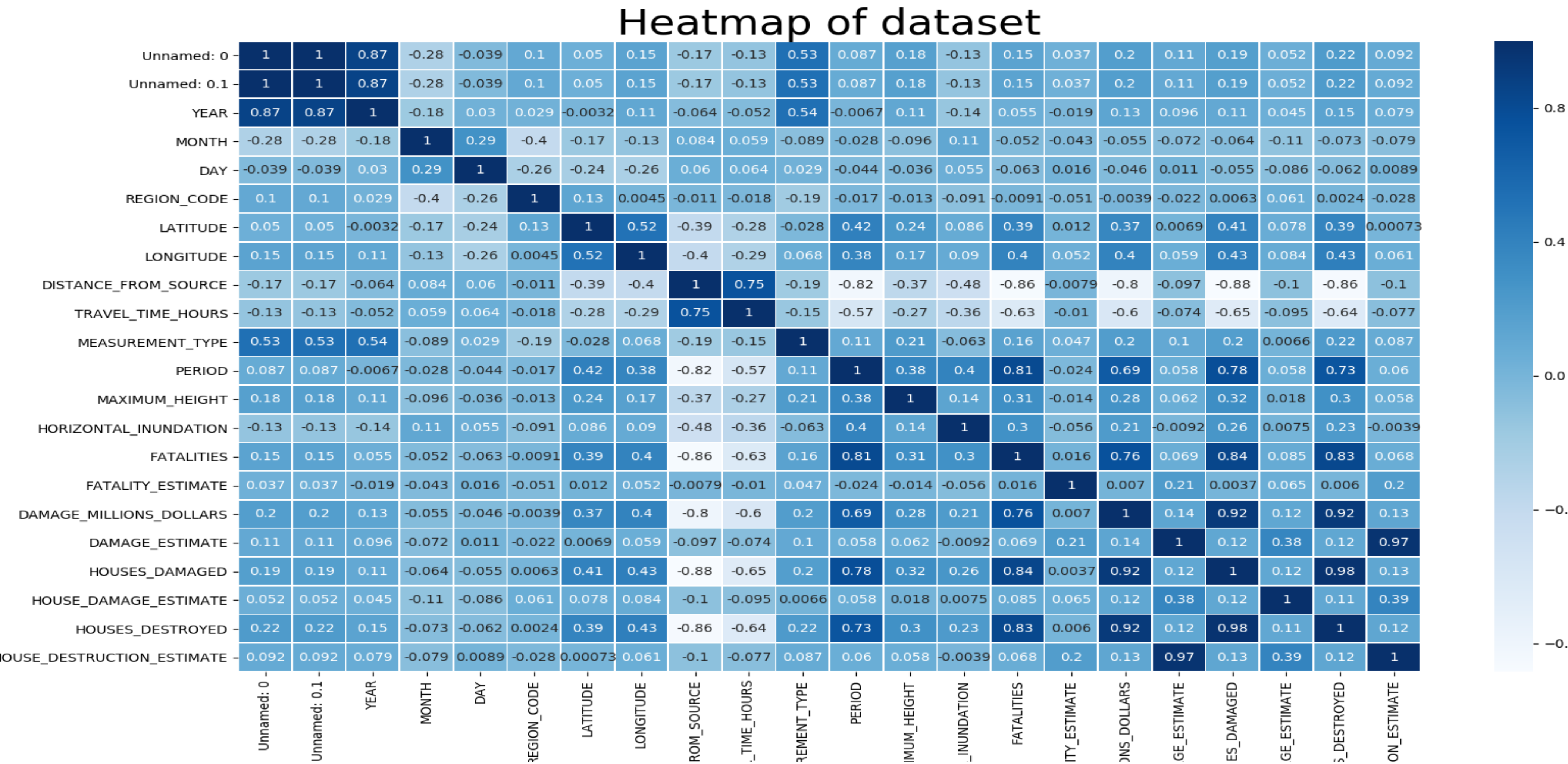
2)

NULL HYPOTESIS:mean longitude=52

p-value=0.290841542469

null hypothesis is plausible.

CORRELATATION



CORRELATION

- DISTANCE_FROM_SOURCE AND TRAVEL_TIME_HOURS are higly correlated with an corr. value of 0.75.This is obvious because the larger the distance from source the more the time it takes.
- Their is a high correlation of PERIOD with fatalities,houses_damaged and houses destroyed.It can be inferred that a larger lasting tsunami is going to have a larger impact on the number of fatalities and economic losses.
- FATALITIES and HOUSES_DESTROYED also shares an high corr. 0.83 this implies that the more the number of houses are destroyed the more are the cases of fatalities.
- HOUSES_ DESTROYED and DAMAGE_MILLIONS of dollars has a high corr. Of 0.92 which is self explanatory because the houses_destroyed will lead to more economic loss.
- MONTH and DISTANCE_FROM SOURCE share a low corr. value because an tsunami is a random event based solely on tectonic activities and the month on which a tsuanmi ocuurs has nothing to do with the distance_from_source.

THANK YOU