SHARKS

¿Hay una correlación entre mes del año, Nº muertes y hemisferio?

Oriol Bàguena - Project II - IRONHACK

Primera Limpieza

df.isna().sum()

del df['Unnamed: 22']
del df['Unnamed: 23']

df.drop_duplicates()

df.drop_duplicates(subset=['Case Number'],
inplace=True)

Nombres y Países

```python
df2 = df[['Date', 'Country', 'Sex ', 'Fatal (Y/N)',
'Activity']]

df2["Fatal"].fillna("U", inplace = True)

df2.dropna(subset=['Country'], inplace=True)

countries = df2["Country"].unique()

sum(df2['Country'].value_counts()==1)
```

Limpieza de únicos

```python
num_cases = df2['Country'].value_counts()

one_case = num_cases[num_cases == 1].index

df3 = df2[~df2.Country.isin(one_case)]

countries = df3["Country"].unique()
```

203 a 121

Op. 1 Listas

```
1   north = ['USA', 'BAHAMAS', 'SPAIN', 'CHINA', 'JAPAN', 'C(
2            'INDONESIA', 'CAPE VERDE', 'DOMINICAN REPUBLIC
3            'THAILAND', 'ITALY', 'MEXICO', 'GREECE', 'FRAN(
4            'ISRAEL', 'TAIWAN', 'JAMAICA', 'GUAM', 'BELIZE
5            'NIGERIA', 'TONGA', 'SCOTLAND', 'CANADA', 'CRO/
6            'CHILE', 'KENYA', 'RUSSIA', 'TURKS & CAICOS',
7            'MALAYSIA', 'UNITED ARAB EMIRATES (UAE)', 'AZO|
8            'MALTA', 'VIETNAM', 'PANAMA', 'SOMALIA', 'CUBA
9            'SENEGAL', 'YEMEN', 'LIBERIA', 'VENEZUELA', 'S|
10           'CARIBBEAN SEA', 'OKINAWA', 'MARSHALL ISLANDS'
11           'BERMUDA', 'MONTENEGRO', 'IRAN', 'TUNISIA', 'N(
12           'PORTUGAL', 'PALAU', 'GRENADA', 'IRAQ', 'TURKE
13           'JOHNSTON ISLAND', 'NORTH PACIFIC OCEAN', 'PER!
14           'SIERRA LEONE', 'NICARAGUA', 'BARBADOS', 'HOND(
15           'CRETE', 'EGYPT ', 'BURMA', 'LEBANON', 'GUINEA

1   south = ['AUSTRALIA', 'NEW CALEDONIA', 'REUNION', 'SOUTH
2            'NEW ZEALAND', 'FRENCH POLYNESIA', 'Fiji', 'BRA!
3            'FIJI', 'MAURITIUS', 'ECUADOR', 'PAPUA NEW GUIN|
4            'SEYCHELLES', 'SOLOMON ISLANDS', 'MADAGASCAR',
5            'URUGUAY', 'TANZANIA', 'NAMIBIA', 'NEW BRITAIN'
6            'NEW GUINEA', 'AMERICAN SAMOA', 'CENTRAL PACIFI(

1   other = ['ATLANTIC OCEAN', 'PACIFIC OCEAN ', 'MID ATLANT
```

df5["Hemisphere"] =
np.where(df5["Country"].isin(north), "N", "S")

df5 = df4[~df4.Country.isin(other)]

df5['Country'] = df5['Country'].str.upper()   (FIJI)

Op. 2 Función

```python
def country_hem(col):
    hem_lst = []
    for row in col:
        try:
            country = CountryInfo(row)
        except AttributeError:
            row = 'x'

        try:
            pos = country.latlng()
        except KeyError:
            pos = (0,0)

        if pos[0] > 0:
            temp_row = 1
        elif pos[0] < 0:
            temp_row = 0
        else:
            temp_row = np.nan

        hem_lst.append(temp_row)

    return hem_lst
```

Fechas

```
df5['Date'] = pd.to_datetime(df['Date'], errors='coerce')
```

```
df5.dropna(subset=['Date'], inplace=True)
```

```
df5['Month'] = pd.DatetimeIndex(df5['Date']).month
```

Limpieza Fatal

df_6['Fatal'].unique()

eliminate = ['U', 'UNKNOWN', 'N ', '#VALUE!', 'n']

df_def = df_6[~df_6.Fatal.isin(eliminate)]

Tabla Final

	Hemisphere	Month	Fatal
0	N	9	N
1	N	9	N
2	N	9	N
3	S	9	N
4	S	9	N
...	...	...	...
5843	N	1	N
5844	N	1	N
5846	N	1	Y
5850	N	6	Y
5851	N	3	N

Hemisphere	Month	
N	1	209
	2	85
	3	130
	4	177
	5	201
	6	294
	7	427
	8	411
	9	340
	10	220
	11	148
	12	85
S	1	455
	2	226
	3	222
	4	175
	5	121
	6	110
	7	103
	8	96
	9	92
	10	133
	11	183
	12	269

Gráfico