

US_accident project

In [1]:

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

%matplotlib inline

import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
data = pd.read_csv("D:\My project\Datasets\US_Accidents_Dec21_updated.csv")
```

In [3]:

```
data.shape
```

Out[3]:

```
(2845342, 47)
```

In [4]:

```
data.head()
```

Out[4]:

	ID	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Description	...	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal	Turning_Loop	Sunrise
0	A-1	3	2016-02-08 06:37:08	2016-02-08 06:37:08	40.108910	-83.092860	40.112060	-83.031870	3.230	Between Sawmill Rd/Exit 20 and OH-315/Olentang...	...	False	False	False	False	False	False	False
1	A-2	2	2016-02-08 05:56:20	2016-02-08 11:56:20	39.865420	-84.062800	39.865010	-84.048730	0.747	At OH-410H-235/Exit 41 - Accident.	...	False	False	False	False	False	False	False
2	A-3	3	2016-02-08 06:15:39	2016-02-08 12:15:39	39.102660	-84.524680	39.102090	-84.523960	0.055	At I-71/US-50/Exit 1 - Accident.	...	False	False	False	False	False	False	False
3	A-4	2	2016-02-08 06:51:45	2016-02-08 12:51:45	41.062130	-81.537840	41.062170	-81.535470	0.123	At Dart Ave/Exit 21 - Accident.	...	False	False	False	False	False	False	False
4	A-5	3	2016-02-08 07:53:43	2016-02-08 13:53:43	39.172393	-84.492792	39.170478	-84.501798	0.500	At Mitchell Ave/Exit 6 - Accident.	...	False	False	False	False	False	False	False

5 rows × 47 columns

In [5]:

```
data.tail(10)
```

Out[5]:

	ID	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Description	...	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal	Turning_L
2845332	A-2845333	2	2019-08-23 17:42:27	2019-08-23 18:11:10	34.064460	-118.003880	34.065330	-117.997150	0.390	At I-605 - Accident.	...	False	False	False	False	False	False

2845333	A-2845334	2	2019-08-23 17:40:12	2019-08-23 18:08:35	33.943599	-117.077880	33.943599	-117.077880	0.000	At Jack Rabbit Trl - Accident.	...	False	False	False	False	False	False
2845334	A-2845335	2	2019-08-23 17:40:12	2019-08-23 18:08:35	34.261030	-119.228000	34.262390	-119.230870	0.189	At Telephone Rd/Exit 65 - Accident.	...	False	False	False	False	False	False
2845335	A-2845336	2	2019-08-23 17:43:56	2019-08-23 18:12:27	33.741700	-117.837090	33.739170	-117.830010	0.443	At CA-55 - Accident.	...	False	False	False	False	False	False
2845336	A-2845337	2	2019-08-23 18:30:23	2019-08-23 18:58:54	34.239104	-118.416176	34.239104	-118.416176	0.000	At Osborne St/Exit 154 - Accident.	...	False	False	False	False	False	False
2845337	A-2845338	2	2019-08-23 18:03:25	2019-08-23 18:32:01	34.002480	-117.379360	33.998680	-117.370940	0.543	At Market St - Accident.	...	False	False	False	False	False	False
2845338	A-2845339	2	2019-08-23 19:11:30	2019-08-23 19:36:23	32.766960	-117.148060	32.765550	-117.153630	0.338	At Camino Del Rio/Mission Center Rd - Accident.	...	False	False	False	False	False	False
2845339	A-2845340	2	2019-08-23 19:00:21	2019-08-23 19:28:49	33.775450	-117.847790	33.777400	-117.857270	0.561	At Glassell St/Grand Ave - Accident in the n...	...	False	False	False	False	False	False
2845340	A-2845341	2	2019-08-23 19:00:21	2019-08-23 19:29:42	33.992460	-118.403020	33.983110	-118.395650	0.772	At CA-90/Marina Fwy/Jefferson Blvd - Accident.	...	False	False	False	False	False	False
2845341	A-2845342	2	2019-08-23 18:52:06	2019-08-23 19:21:31	34.133930	-117.230920	34.137360	-117.239340	0.537	At Highland Ave/Arden Ave - Accident.	...	False	False	False	False	False	False

10 rows × 47 columns

In [6]: data.describe().transpose()

Out[6]:

	count	mean	std	min	25%	50%	75%	max
Severity	2845342.0	2.137572	0.478722	1.000000	2.000000	2.000000	2.000000	4.000000e+00
Start_Lat	2845342.0	36.245201	5.363797	24.566027	33.445174	36.098669	40.160243	4.900058e+01
Start_Lng	2845342.0	-97.114633	8.317819	-124.548074	-118.033113	-92.418076	-80.372431	-6.711317e+01
End_Lat	2845342.0	36.245321	5.363873	24.566013	33.446278	36.097987	40.161049	4.907500e+01
End_Lng	2845342.0	-97.114367	8.317632	-124.545748	-118.033331	-92.417718	-80.372383	-6.710924e+01

	Distance(mi)	2845342.0	0.702678	1.560361	0.000000	0.052000	0.244000	0.764000	1.551860e+02
	Number	1101431.0	8099.408114	18360.093995	0.000000	1270.000000	4007.000000	9567.000000	9.999997e+06
	Temperature(F)	2776068.0	61.793556	18.622629	-89.000000	50.000000	64.000000	76.000000	1.960000e+02
	Wind_Chill(F)	2375699.0	59.658231	21.160967	-89.000000	46.000000	63.000000	76.000000	1.960000e+02
	Humidity(%)	2772250.0	64.365452	22.874568	1.000000	48.000000	67.000000	83.000000	1.000000e+02
	Pressure(in)	2786142.0	29.472344	1.045286	0.000000	29.310000	29.820000	30.010000	5.890000e+01
	Visibility(mi)	2774796.0	9.099391	2.717546	0.000000	10.000000	10.000000	10.000000	1.400000e+02
	Wind_Speed(mph)	2687398.0	7.395044	5.527454	0.000000	3.500000	7.000000	10.000000	1.087000e+03
	Precipitation(in)	2295884.0	0.007017	0.093488	0.000000	0.000000	0.000000	0.000000	2.400000e+01

Find the average temprature ?

In [7]: data['Temperature(F)'].mean()

Out[7]: 61.79355592154877

In [7]:	<pre>data['Temperature(F)'].mean()</pre>																	
Out [7]:	61.79355592154077																	

End_Lat	0
End_Lng	0
Distance(mi)	0
Description	0
Number	1743911
Street	2
Side	0
City	137
County	0
State	0
Zipcode	1319
Country	0
Timezone	3659
Airport_Code	9549
Weather_Timestamp	56736
Temperature(F)	69274
Wind_Chill(F)	469643
Humidity(%)	73992
Pressure(in)	59288
Visibility(mi)	79546
Wind_Direction	73775
Wind_Speed(mph)	157944
Precipitation(in)	549458
Weather_Condition	78636
Amenity	0
Bump	0
Crossing	0
Give_Way	0
Junction	0
No_Exit	0
Railway	0
Roundabout	0
Station	0
Stop	0
Traffic_Calming	0
Traffic_Signal	0
Turning_Loop	0
Sunrise_Sunset	2867
Civil_Twilight	2867
Nautical_Twilight	2867
Astronomical_Twilight	2867
dtype:	int64

```
In [9]: df=data.dropna()
```

```
In [10]: df.isnull().sum()
```

ID	0
Severity	0

```
Out[10]:
```

In [9]:	<pre>df=data.dropna()</pre>																	
In [10]:	<pre>df.isnull().sum()</pre>																	

Street	0
Side	0
City	0
County	0
State	0
Zipcode	0
Country	0
Timezone	0
Airport_Code	0
Weather_Timestamp	0
Temperature(F)	0
Wind_Chill(F)	0
Humidity(%)	0
Pressure(in)	0
Visibility(mi)	0
Wind_Direction	0
Wind_Speed(mph)	0
Precipitation(in)	0
Weather_Condition	0
Amenity	0
Bump	0
Crossing	0
Give_Way	0
Junction	0
No_Exit	0
Railway	0
Roundabout	0
Station	0
Stop	0
Traffic_Calming	0
Traffic_Signal	0
Turning_Loop	0
Sunrise_Sunset	0
Civil_Twilight	0
Nautical_Twilight	0
Astronomical_Twilight	0
dtype:	int64

```
In [11]: df.columns
```

```
Out[11]: Index(['ID', 'Severity', 'Start_Time', 'End_Time', 'Start_Lat', 'Start_Lng',  
            'End_Lat', 'End_Lng', 'Distance(mi)', 'Description', 'Number', 'Street',  
            'Side', 'City', 'County', 'State', 'Zipcode', 'Country', 'Timezone',  
            'Airport_Code', 'Weather_Timestamp', 'Temperature(F)', 'Wind_Chill(F)',  
            'Humidity(%)', 'Pressure(in)', 'Visibility(mi)', 'Wind_Direction',  
            'Wind_Speed(mph)', 'Precipitation(in)', 'Weather_Condition', 'Amenity',  
            'Bump', 'Crossing', 'Give_Way', 'Junction', 'No_Exit', 'Railway',  
            'Roundabout', 'Station', 'Stop', 'Traffic_Calming', 'Traffic_Signal',  
            'Turning_Loop', 'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight',  
            'Astronomical_Twilight'],  
            dtype='object', length=40)
```

In [11]:	<pre>df.columns</pre>																	
Out [11]:	<pre>Index(['ID', 'Severity', 'Start_Time', 'End_Time', 'Start_Lat', 'Start_Lng', 'End_Lat', 'End_Lng', 'Distance(mi)', 'Description', 'Number', 'Street', 'Side', 'City', 'County', 'State', 'Zipcode', 'Country', 'Timezone', 'Airport_Code', 'Weather_Timestamp', 'Temperature(F)', 'Wind_Chill(F)', 'Humidity(%)', 'Pressure(in)', 'Visibility(mi)', 'Wind_Direction', 'Wind_Speed(mph)', 'Precipitation(in)', 'Weather_Condition', 'Amenity', 'Bump', 'Crossing', 'Give_Way', 'Junction', 'No_Exit', 'Railway', 'Roundabout', 'Station', 'Stop', 'Traffic_Calming', 'Traffic_Signal', 'Turning_Loop', 'Sunrise_Sunset', 'Civil_Twilight', 'Nautical_Twilight', 'Astronomical_Twilight'], dtype='object')</pre>																	

In [12]:	<pre>from sklearn.preprocessing import LabelEncoder</pre>																	
In [13]:	<pre>le=LabelEncoder() for col in df.columns: df[col]=le.fit_transform(df[col])</pre>																	
In [14]:	<pre>df.head()</pre>																	

47	710188	3	1	1	397222	275471	385999	272661	1206	84002 ...	0	0	0	0	0	0	0
50	729636	1	2	2	414705	397437	403140	386812	824	71607 ...	0	0	0	0	0	0	0
66	795624	1	3	3	410276	331204	398451	326671	462	50250 ...	0	0	0	0	0	0	0

5 rows x 17 columns

```
[15]: X=df.drop('Severity', axis=1)
      y=df['Severity']

[16]: corr=data.corr()
      fig=plt.figure(figsize=(10,10))
      r=ms.heatmap(corr,cmap='Greens_r')
      r.set_title('Correlation')
```