

108年臺北市車禍統計人數

觀看108年臺北市各區死傷人數分別為多少

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn; seaborn.set()
```

```
In [2]: df = pd.read_csv('http://163.29.157.32:8080/es/dataset/2f238b4f-1b27-4085-93e9-d684ef0e2735/resource/f47c1289-9e2f-49aa-a328-37708402a9ac/download/20201223105235')
df
```

發生年	發生月	發生日	發生時	發生分	處理別	區序	肇事地點	死亡人數	受傷人數	當事人序	車種	性別	年齡	受傷程度	天候	速限	道路型態	事故位置
0	108	1	2	8	37	2	01大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	1	C03	2	37.0	2.0	6.0	30	4.0	1.0
1	108	1	2	8	37	2	01大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	2	C03	2	59.0	2.0	6.0	30	4.0	1.0
2	108	1	9	10	59	2	01大同區 大同區重慶北路1段與天水路口	0	1	1	H01	4	0.0	4.0	8.0	0	4.0	2.0
3	108	1	9	10	59	2	01大同區 大同區重慶北路1段與天水路口	0	1	2	C03	1	24.0	2.0	8.0	0	4.0	2.0
4	108	1	12	10	10	2	01大同區 大同區環河北路1段273號對面	0	1	1	C03	1	24.0	2.0	7.0	60	14.0	7.0
...
52428	108	12	22	17	31	2	07信義區 信義區松壽路與松高路20巷口	0	1	1	H01	2	27.0	2.0	8.0	0	4.0	2.0
52429	108	12	22	17	31	2	07信義區 信義區松壽路與松高路20巷口	0	1	3	H01	1	28.0	4.0	8.0	0	4.0	2.0
52430	108	12	22	17	31	2	07信義區 信義區松壽路與松高路20巷口	0	1	2	B03	1	35.0	3.0	8.0	0	4.0	2.0
52431	108	12	26	7	20	2	07信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	1	C03	2	59.0	3.0	8.0	30	4.0	2.0
52432	108	12	26	7	20	2	07信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	2	C03	2	42.0	2.0	8.0	30	4.0	2.0

52433 rows × 19 columns

```
In [3]: df['區序'] = df['區序'].apply(lambda x: ''.join([i for i in x if not i.isdigit()]))
df
```

發生年	發生月	發生日	發生時	發生分	處理別	區序	肇事地點	死亡人數	受傷人數	當事人序	車種	性別	年齡	受傷程度	天候	速限	道路型態	事故位置
0	108	1	2	8	37	2	大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	1	C03	2	37.0	2.0	6.0	30	4.0	1.0
1	108	1	2	8	37	2	大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	2	C03	2	59.0	2.0	6.0	30	4.0	1.0
2	108	1	9	10	59	2	大同區 大同區重慶北路1段與天水路口	0	1	1	H01	4	0.0	4.0	8.0	0	4.0	2.0
3	108	1	9	10	59	2	大同區 大同區重慶北路1段與天水路口	0	1	2	C03	1	24.0	2.0	8.0	0	4.0	2.0
4	108	1	12	10	10	2	大同區 大同區環河北路1段273號對面	0	1	1	C03	1	24.0	2.0	7.0	60	14.0	7.0
...
52428	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	1	H01	2	27.0	2.0	8.0	0	4.0	2.0
52429	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	3	H01	1	28.0	4.0	8.0	0	4.0	2.0
52430	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	2	B03	1	35.0	3.0	8.0	0	4.0	2.0
52431	108	12	26	7	20	2	信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	1	C03	2	59.0	3.0	8.0	30	4.0	2.0
52432	108	12	26	7	20	2	信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	2	C03	2	42.0	2.0	8.0	30	4.0	2.0

52433 rows × 19 columns

```
In [4]: df['性別'] = df['性別'].apply(lambda x: 'Male' if x == 1 else 'Female')
df
```

發生年	發生月	發生日	發生時	發生分	處理別	區序	肇事地點	死亡人數	受傷人數	當事人序	車種	性別	年齡	受傷程度	天候	速限	道路型態	事故位置
0	108	1	2	8	37	2	大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	1	C03	Female	37.0	2.0	6.0	30	4.0	1.0
1	108	1	2	8	37	2	大同區 大同區重慶北路3段137巷與民族西路182巷口	0	2	2	C03	Female	59.0	2.0	6.0	30	4.0	1.0
2	108	1	9	10	59	2	大同區 大同區重慶北路1段與天水路口	0	1	1	H01	Female	0.0	4.0	8.0	0	4.0	2.0
3	108	1	9	10	59	2	大同區 大同區重慶北路1段與天水路口	0	1	2	C03	Male	24.0	2.0	8.0	0	4.0	2.0
4	108	1	12	10	10	2	大同區 大同區環河北路1段273號對面	0	1	1	C03	Male	24.0	2.0	7.0	60	14.0	7.0
...
52428	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	1	H01	Female	27.0	2.0	8.0	0	4.0	2.0
52429	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	3	H01	Male	28.0	4.0	8.0	0	4.0	2.0
52430	108	12	22	17	31	2	信義區 信義區松壽路與松高路20巷口	0	1	2	B03	Male	35.0	3.0	8.0	0	4.0	2.0
52431	108	12	26	7	20	2	信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	1	C03	Female	59.0	3.0	8.0	30	4.0	2.0
52432	108	12	26	7	20	2	信義區 信義區永吉路127巷與基隆路1段83巷口	0	1	2	C03	Female	42.0	2.0	8.0	30	4.0	2.0

52433 rows × 19 columns

```
In [5]: df = df.groupby(['區序', '性別'])[['死亡人數', '受傷人數']].sum().reset_index().sort_values('受傷人數')
df = df.reset_index(drop=True)
df
```

Out[5]:

	區序	性別	死亡人數	受傷人數
0	南港區	Female	1	838
1	松山區	Female	4	1280
2	萬華區	Female	2	1285
3	南港區	Male	1	1809
4	大同區	Female	2	1877
5	士林區	Female	18	2196
6	信義區	Female	3	2230
7	大安區	Female	12	2290
8	內湖區	Female	9	2310
9	文山區	Female	6	2398
10	中正區	Female	3	2410
11	萬華區	Male	8	2544
12	松山區	Male	4	2654
13	北投區	Female	11	2880
14	中山區	Female	4	2937
15	大同區	Male	13	3372
16	士林區	Male	20	3752
17	內湖區	Male	15	4006
18	文山區	Male	9	4141
19	信義區	Male	10	4381
20	大安區	Male	28	4460
21	北投區	Male	20	4567
22	中正區	Male	3	4864
23	中山區	Male	18	5970

In [6]:

```
df.columns = ['District', 'Sex', 'Death', 'Injury']
df
```

Out[6]:

	District	Sex	Death	Injury
0	南港區	Female	1	838
1	松山區	Female	4	1280
2	萬華區	Female	2	1285
3	南港區	Male	1	1809
4	大同區	Female	2	1877
5	士林區	Female	18	2196
6	信義區	Female	3	2230
7	大安區	Female	12	2290
8	內湖區	Female	9	2310
9	文山區	Female	6	2398
10	中正區	Female	3	2410
11	萬華區	Male	8	2544
12	松山區	Male	4	2654
13	北投區	Female	11	2880
14	中山區	Female	4	2937
15	大同區	Male	13	3372
16	士林區	Male	20	3752
17	內湖區	Male	15	4006
18	文山區	Male	9	4141
19	信義區	Male	10	4381
20	大安區	Male	28	4460
21	北投區	Male	20	4567
22	中正區	Male	3	4864
23	中山區	Male	18	5970

中文名稱改為英文名稱

```
In [7]: df['District'] = df['District'].str.replace('南港區', 'Nangang District')
df['District'] = df['District'].str.replace('松山區', 'Songshan District')
df['District'] = df['District'].str.replace('萬華區', 'Wanhua District')
df['District'] = df['District'].str.replace('大同區', 'Datong District')
df['District'] = df['District'].str.replace('士林區', 'Shilin District')
df['District'] = df['District'].str.replace('中正區', 'Zhongzheng District')
df['District'] = df['District'].str.replace('信義區', 'Xinyi District')
df['District'] = df['District'].str.replace('文山區', 'Wenshan District')
df['District'] = df['District'].str.replace('北投區', 'Beitou District')
df['District'] = df['District'].str.replace('中山區', 'Zhongshan District')
df['District'] = df['District'].str.replace('內湖區', 'Neihu District')
df['District'] = df['District'].str.replace('大安區', 'Daan District')
df
```

Out[7]:

	District	Sex	Death	Injury
0	Nangang District	Female	1	838
1	Songshan District	Female	4	1280
2	Wanhua District	Female	2	1285
3	Nangang District	Male	1	1809
4	Datong District	Female	2	1877
5	Shilin District	Female	18	2196
6	Xinyi District	Female	3	2230
7	Daan District	Female	12	2290
8	Neihu District	Female	9	2310
9	Wenshan District	Female	6	2398
10	Zhongzheng District	Female	3	2410
11	Wanhua District	Male	8	2544
12	Songshan District	Male	4	2654
13	Beitou District	Female	11	2880
14	Zhongshan District	Female	4	2937
15	Datong District	Male	13	3372
16	Shilin District	Male	20	3752
17	Neihu District	Male	15	4006
18	Wenshan District	Male	9	4141
19	Xinyi District	Male	10	4381
20	Daan District	Male	28	4460
21	Beitou District	Male	20	4567
22	Zhongzheng District	Male	3	4864
23	Zhongshan District	Male	18	5970

```
In [8]: df = df.sort_values(['District', 'Sex'], ascending=False)
df = df.reset_index(drop=True)
df
```

Out[8]:

	District	Sex	Death	Injury
0	Zhongzheng District	Male	3	4864
1	Zhongzheng District	Female	3	2410
2	Zhongshan District	Male	18	5970
3	Zhongshan District	Female	4	2937
4	Xinyi District	Male	10	4381
5	Xinyi District	Female	3	2230
6	Wenshan District	Male	9	4141
7	Wenshan District	Female	6	2398
8	Wanhua District	Male	8	2544
9	Wanhua District	Female	2	1285
10	Songshan District	Male	4	2654
11	Songshan District	Female	4	1280
12	Shilin District	Male	20	3752
13	Shilin District	Female	18	2196
14	Neihu District	Male	15	4006
15	Neihu District	Female	9	2310
16	Nangang District	Male	1	1809
17	Nangang District	Female	1	838
18	Datong District	Male	13	3372
19	Datong District	Female	2	1877
20	Daan District	Male	28	4460
21	Daan District	Female	12	2290
22	Beitou District	Male	20	4567
23	Beitou District	Female	11	2880

```
In [9]: n_cols = len(df['District'])
index = np.arange(n_cols)

x = df['Death']
y = df['Injury']

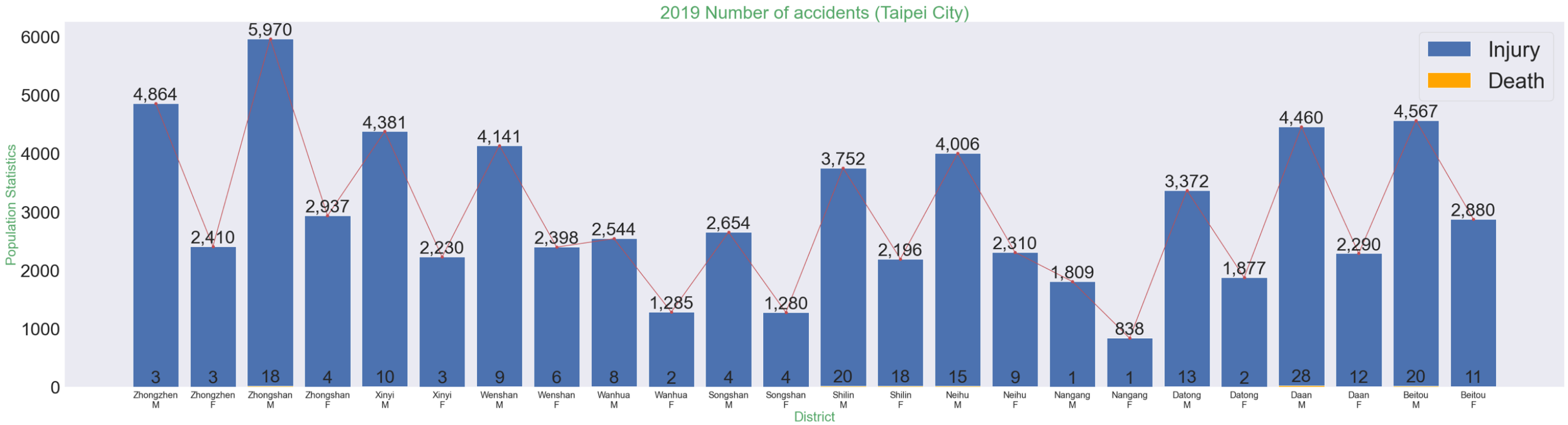
plt.subplots(figsize=(60, 15))
plt.title('2019 Number of accidents (Taipei City)', fontsize=40, color='g')

plt.xlabel('District', fontsize=30, color='g')
plt.xticks(index, ('Zhongzhen\nM', 'Zhongzhen\nF', 'Zhongshan\nM', 'Zhongshan\nF', 'Xinyi\nM', 'Xinyi\nF', 'Wenshan\nM', 'Wenshan\nF',
                  'Wanhua\nM', 'Wanhua\nF', 'Songshan\nM', 'Songshan\nF', 'Shilin\nM', 'Shilin\nF', 'Neihu\nM', 'Neihu\nF', 'Nangang\nM', 'Nangang\nF',
                  'Datong\nM', 'Datong\nF', 'Daan\nM', 'Daan\nF', 'Beitou\nM', 'Beitou\nF'), fontsize=20)

plt.ylabel('Population Statistics', fontsize=30, color='g')
plt.yticks(fontsize=40)

plt.bar(index, df['Injury'], label='Injury')
plt.bar(index, df['Death'], color='orange', label='Death')
plt.plot(index, df['Injury'], '-o', color='r')
plt.grid()
plt.legend(fontsize=48)
for xx, yy in enumerate(y): #為了更容易觀看長條圖，使用enumerate日數字顯示在各個長條圖上方
    plt.text(xx, yy, '{:,}'.format(int(yy)), ha='center', va='bottom', fontsize=40)
for xx, yy in enumerate(x):
    plt.text(xx, yy, '{:,}'.format(int(yy)), ha='center', va='bottom', fontsize=40)

plt.show()
```



整理另一個觀測年齡死亡和受傷數

```
In [31]: df2 = pd.read_csv('http://163.29.157.32:8080/es/dataset/2f238b4f-1b27-4085-93e9-d684ef0e2735/resource/f47c1289-9e2f-49aa-a328-37708402a9ac/download/2020122310523')
df2['性別'] = df2['性別'].apply(lambda x: 'Male' if x == 1 else 'Female')
df2['區序'] = df2['區序'].apply(lambda x: ''.join([i for i in x if not i.isdigit()]))
df2 = df2.groupby(['性別', '年齡'])['死亡人數', '受傷人數'].sum().reset_index().sort_values('受傷人數')
df2 = df2.reset_index(drop=True)
df2.columns = ['Sex', 'Age', 'Death', 'Injury']
df2
```

<ipython-input-31-f9533c5bcdfa>:4: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.

```
df2 = df2.groupby(['性別', '年齡'])['死亡人數', '受傷人數'].sum().reset_index().sort_values('受傷人數')
```

Out[31]:

	Sex	Age	Death	Injury
0	Male	103.0	0	1
1	Female	101.0	0	1
2	Male	99.0	0	1
3	Male	98.0	0	1
4	Male	96.0	0	1
...
197	Male	21.0	2	1402
198	Male	20.0	5	1482
199	Male	18.0	2	1603
200	Male	19.0	4	1713
201	Female	0.0	24	2312

202 rows × 4 columns

```
In [39]: df_Male = df2[df2['Sex'].str.contains('Male')].reset_index(drop=True).sort_values('Age', ascending=True)
df_Male = df_Male.reset_index(drop=True)
df_Male
```

Out[39]:

	Sex	Age	Death	Injury
0	Male	0.0	1	11
1	Male	1.0	0	32
2	Male	2.0	1	24
3	Male	3.0	0	39
4	Male	4.0	0	57
...
98	Male	99.0	0	1
99	Male	101.0	0	4
100	Male	103.0	0	1
101	Male	107.0	0	50
102	Male	108.0	0	5

103 rows × 4 columns

In [40]:

```
df_Female = df2[df2['Sex'].str.contains('Female')].reset_index(drop=True).sort_values('Age', ascending=True)
df_Female = df_Male.reset_index(drop=True)
df_Female
```

Out[40]:

	Sex	Age	Death	Injury
0	Male	0.0	1	11
1	Male	1.0	0	32
2	Male	2.0	1	24
3	Male	3.0	0	39
4	Male	4.0	0	57
...
98	Male	99.0	0	1
99	Male	101.0	0	4
100	Male	103.0	0	1
101	Male	107.0	0	50
102	Male	108.0	0	5

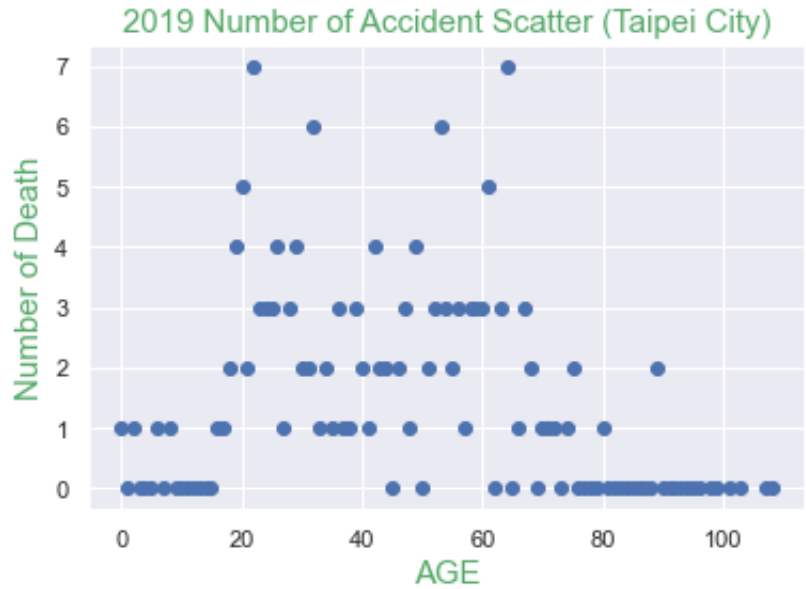
103 rows × 4 columns

製作散狀圖

依照散狀圖來分析，死亡人數座落於20-60歲，最低死亡年齡在20歲。

In [45]:

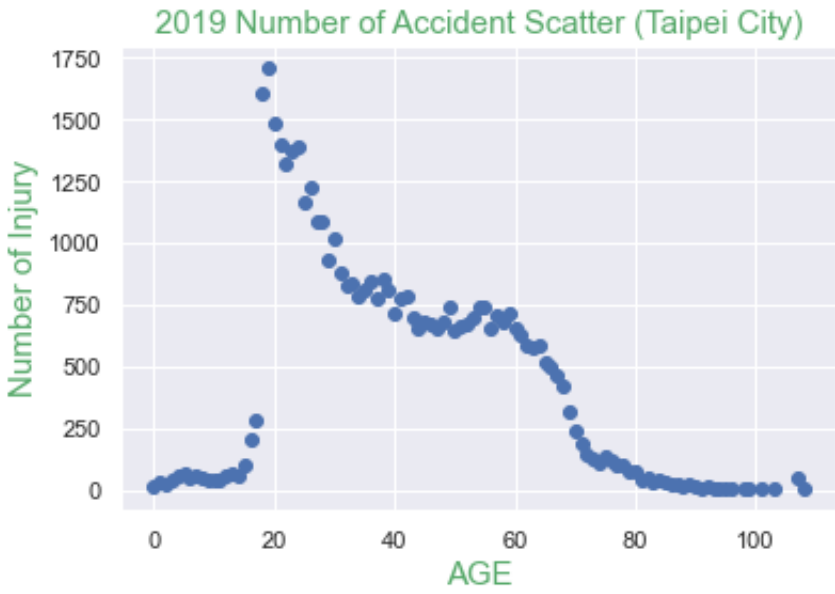
```
plt.scatter(df_Male['Age'], df_Male['Death'])
plt.title('2019 Number of Accident Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Death', fontsize=15, color='g')
plt.show()
```



依照散狀圖來分析，受傷人數於20歲最多。

In [46]:

```
plt.scatter(df_Male['Age'], df_Male['Injury'])
plt.title('2019 Number of Accident Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Injury', fontsize=15, color='g')
plt.show()
```

```
In [48]: df3 = pd.read_csv('http://163.29.157.32:8080/es/dataset/2f238b4f-1b27-4085-93e9-d684ef0e2735/resource/f47c1289-9e2f-49aa-a328-37708402a9ac/download/2020122310523')
df3 = df3.groupby(['發生月'])['死亡人數', '受傷人數'].sum().reset_index().sort_values('發生月')
df3 = df3.reset_index(drop=True)
df3.columns = ['Month', 'Death', 'Injury']
df3
```

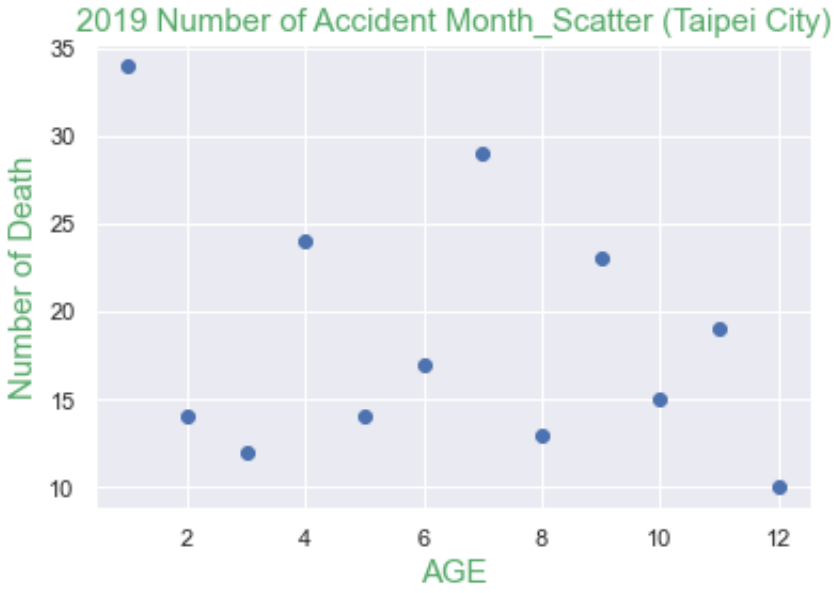
```
<ipython-input-48-cb5fcdacc618>:2: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
df3 = df3.groupby(['發生月'])['死亡人數', '受傷人數'].sum().reset_index().sort_values('發生月')
```

Out[48]:

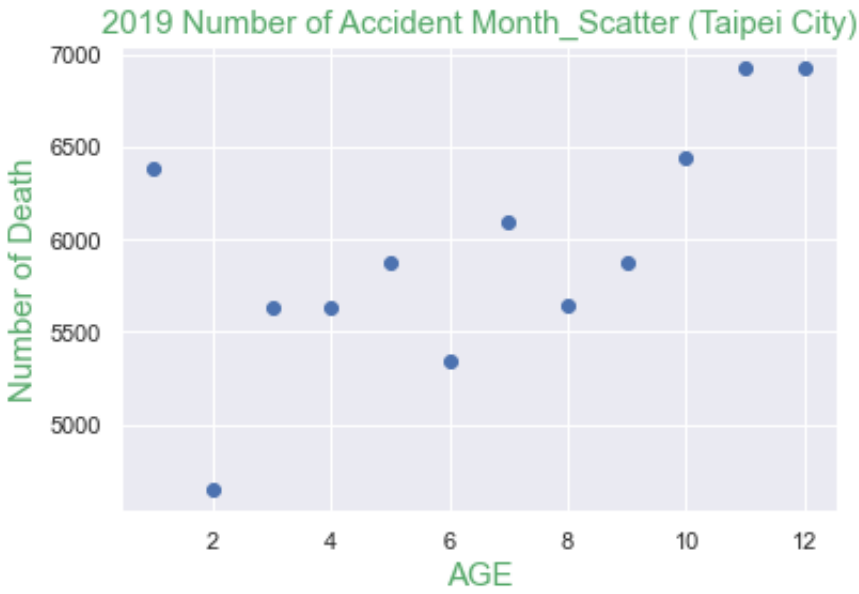
	Month	Death	Injury
0	1	34	6386
1	2	14	4644
2	3	12	5634
3	4	24	5631
4	5	14	5882
5	6	17	5345
6	7	29	6100
7	8	13	5645
8	9	23	5878
9	10	15	6447
10	11	19	6926
11	12	10	6933

依照散狀圖來分析每月的死亡人數於哪幾個月為較多

```
In [50]: plt.scatter(df3['Month'], df3['Death'])
plt.title('2019 Number of Accident Month_Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Death', fontsize=15, color='g')
plt.show()
```

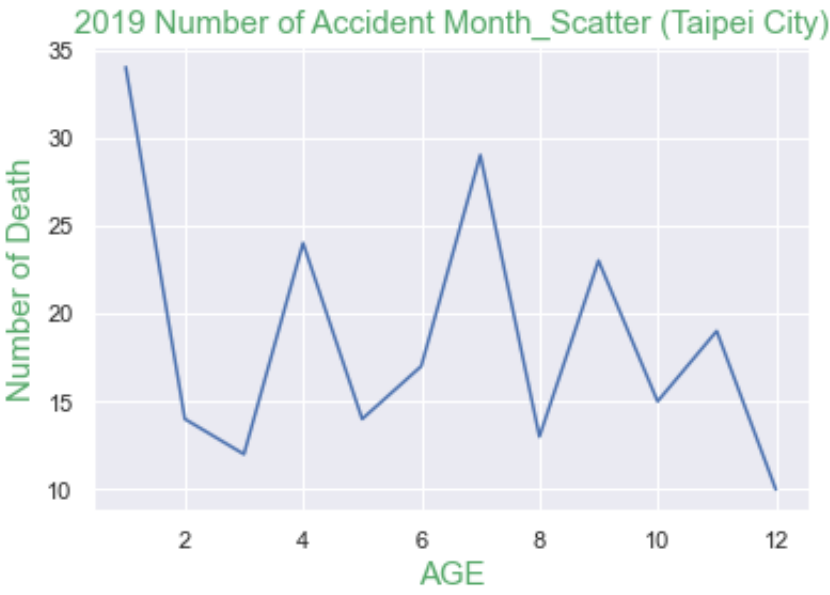


```
In [52]: plt.scatter(df3['Month'], df3['Injury'])
plt.title('2019 Number of Accident Month_Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Death', fontsize=15, color='g')
plt.show()
```



依照曲線圖來分析每月的死亡人數

```
In [53]: plt.plot(df3['Month'], df3['Death'])
plt.title('2019 Number of Accident Month_Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Death', fontsize=15, color='g')
plt.show()
```



依照曲線圖來分析每月的受傷人數

```
In [55]: plt.plot(df3['Month'], df3['Injury'])
plt.title('2019 Number of Accident Month_Scatter (Taipei City)', fontsize=15, color='g')
plt.xlabel('AGE', fontsize=15, color='g')
plt.ylabel('Number of Death', fontsize=15, color='g')
plt.show()
```

