In [17]:

```
print("Name : ")
print("In the following notebook we will how to clean and manipulate the data, and lastly c
print("Using the above skills set we will cleaned and manipulated the Austin weather data,
print("Also we have plotted a point scatter graph for average temperature and humidity as p
```

```
Name :
In the following notebook we will how to clean and manipulate the data, and
lastly come to a consulsion.
Using the above skills set we will cleaned and manipulated the Austin weathe
r data, to plot a point scatter graph for the average high and low temperatu
re as per months
Also we have plotted a point scatter graph for average temperature and humid
ity as per months
```

Activity - 1 Find the average high and low temperature and plot a point grahp as per months

In [18]:

```
#Activity 1
import pandas as pd
import matplotlib.pyplot as plt

#read the csv
df = pd.read_csv('austin_weather.csv')
df
```

Out[18]:

	Date	TempHighF	TempAvgF	TempLowF	DewPointHighF	DewPointAvgF	DewPointLowF
0	2013-12-21	74	60	45	67	49	43
1	2013-12-22	56	48	39	43	36	28
2	2013-12-23	58	45	32	31	27	23
3	2013-12-24	61	46	31	36	28	21
4	2013-12-25	58	50	41	44	40	36
...	...	...	...	...	...	...	...
1314	2017-07-27	103	89	75	71	67	61
1315	2017-07-28	105	91	76	71	64	55
1316	2017-07-29	107	92	77	72	64	55
1317	2017-07-30	106	93	79	70	68	63
1318	2017-07-31	99	88	77	66	61	54

1319 rows × 21 columns

In [19]:

```
#Fetech the months from Data column and add a new column
df['month'] = pd.DatetimeIndex(df['Date']).month
df

#Group by month column and find the average high and low temperature and make a new datafra
group_by_month = df.groupby('month')[['TempHighF','TempLowF']].mean().reset_index()
group_by_month
```

Out[19]:

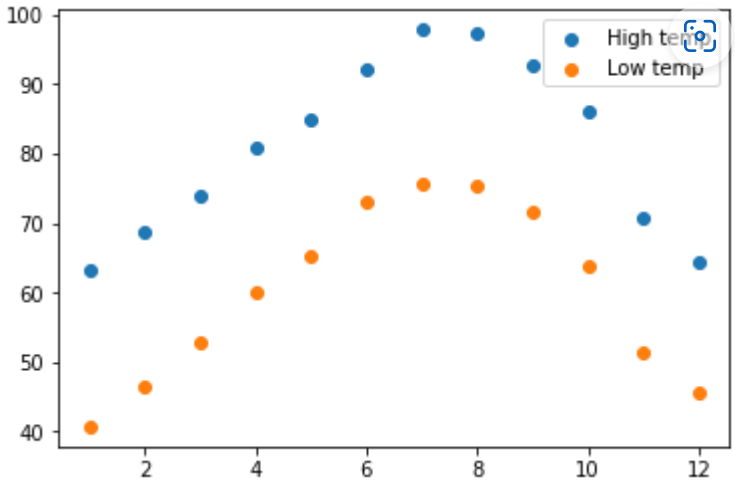
	month	TempHighF	TempLowF
0	1	63.137097	40.725806
1	2	68.849558	46.407080
2	3	73.943548	52.943548
3	4	80.933333	60.125000
4	5	85.024194	65.104839
5	6	92.116667	72.950000
6	7	97.806452	75.653226
7	8	97.322581	75.247312
8	9	92.677778	71.677778
9	10	86.118280	63.688172
10	11	70.611111	51.244444
11	12	64.298077	45.634615

In [20]:

```
#Plot a point scatter graph for the average high and low temperature as per months
month = group_by_month['month']
TempHighF = group_by_month['TempHighF']

plt.scatter(month, TempHighF, label = "High temp")

TempLowF = group_by_month['TempLowF']

plt.scatter(month, TempLowF , label = "Low temp")
plt.legend()
```

Out[20]:

```
<matplotlib.legend.Legend at 0x1dd751fefd0>
```



Conclusion - By logic the low temperature is always lesser then high temperature hence we have plot the grahp correctly

Activity - 2 Plot a point graph for showing the correlation between humidity and temperature as per month

In [21]:

```
#Activity 1
import pandas as pd
import matplotlib.pyplot as plt

#read the csv
dataframe = pd.read_csv('austin_weather.csv')
dataframe
dataframe['month']=pd.DatetimeIndex(dataframe['Date']).month
dataframe
```

Out[21]:

	Date	TempHighF	TempAvgF	TempLowF	DewPointHighF	DewPointAvgF	DewPointLowF
0	2013-12-21	74	60	45	67	49	43
1	2013-12-22	56	48	39	43	36	28
2	2013-12-23	58	45	32	31	27	23
3	2013-12-24	61	46	31	36	28	21
4	2013-12-25	58	50	41	44	40	36
...	...	...	...	...	...	...	...
1314	2017-07-27	103	89	75	71	67	61
1315	2017-07-28	105	91	76	71	64	55
1316	2017-07-29	107	92	77	72	64	55
1317	2017-07-30	106	93	79	70	68	63
1318	2017-07-31	99	88	77	66	61	54

1319 rows × 22 columns

In [22]:

```
#cleaning the data
dataframe.replace('-',float('NaN'),inplace=True)
df=dataframe.dropna()
df
#Converting datatype of HumidityHighPercent column from object to int
df['HumidityHighPercent']=df['HumidityHighPercent'].astype(int)
df
group_by_month = df.groupby('month')[['TempHighF','HumidityHighPercent']].mean().reset_inde
group_by_month
```

```
C:\Users\soumy\AppData\Local\Temp\ipykernel_16956\1651924433.py:6: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/
stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pand
as.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-v
ersus-a-copy)
  df['HumidityHighPercent']=df['HumidityHighPercent'].astype(int)
```

Out[22]:

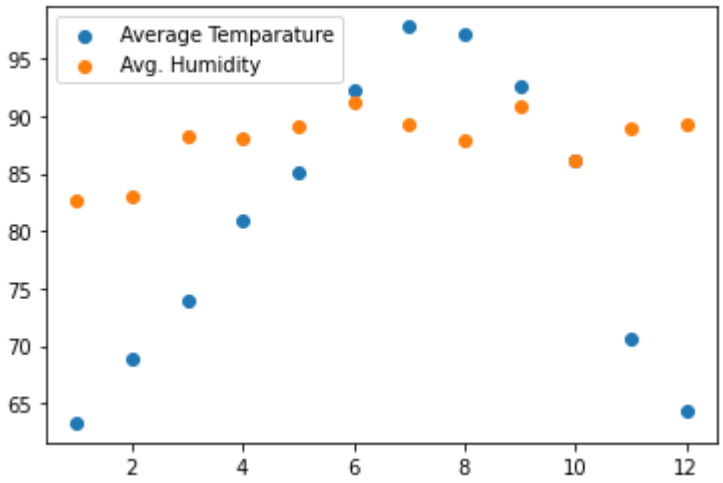
	month	TempHighF	HumidityHighPercent
0	1	63.314050	82.586777
1	2	68.849558	82.964602
2	3	73.943548	88.290323
3	4	80.933333	88.033333
4	5	85.024194	89.193548
5	6	92.198276	91.215517
6	7	97.804878	89.325203
7	8	97.088889	87.877778
8	9	92.586207	90.885057
9	10	86.118280	86.150538
10	11	70.611111	88.900000
11	12	64.298077	89.288462

In [23]:

```
#Group by month column and out the find the average temperature and humidity
month=group_by_month['month']
TempHighF=group_by_month['TempHighF']
plt.scatter(month,TempHighF,label="Average Temparature")
HumidityHighPercent=group_by_month['HumidityHighPercent']
plt.scatter(month,HumidityHighPercent,label='Avg. Humidity')
plt.legend()
df.groupby('month')['Events'].max().reset_index()
```

Out[23]:

	month	Events
0	1	Thunderstorm
1	2	Rain , Thunderstorm
2	3	Thunderstorm
3	4	Thunderstorm
4	5	Thunderstorm
5	6	Thunderstorm
6	7	Thunderstorm
7	8	Thunderstorm
8	9	Rain , Thunderstorm
9	10	Thunderstorm
10	11	Rain , Thunderstorm
11	12	Rain , Thunderstorm



In [24]:

```
#Plot a point scatter graph for the average temperature and humidity as per months
```

Conclusion - Austin Has Humid Weather

In [25]:

```
#Print the Events column(which has weather information) and cross check the observation
```

Conclusion -

In [ ]:

```

```

In [26]:

```
lol
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [26], in <cell line: 1>()
----> 1 lol

NameError: name 'lol' is not defined
```

In [ ]:

```

```