In [1]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot
import matplotlib.pyplot as plt

# 设置中文字体
plt.rcParams['font.family'] = 'SimHei'
import seaborn as sns
```

In [24]:
```
df1=pd.read_csv('./东莞.csv')
```

In [25]:
```
df2=pd.read_csv('./中山.csv')
```

In [26]:
```
df3=pd.read_csv('./云浮.csv')
```

In [27]:
```
df4=pd.read_csv('./佛山.csv')
```

In [28]:
```
df5=pd.read_csv('./广州.csv')
```

In [29]:
```
df6=pd.read_csv('./惠州.csv')
```

In [30]:
```
df7=pd.read_csv('./揭阳.csv')
```

In [31]:
```
df8=pd.read_csv('./梅州.csv')
```

In [32]:
```
df9=pd.read_csv('./汕头.csv')
```

In [33]:
```
df10=pd.read_csv('./汕尾.csv')
```

In [34]:
```
df11=pd.read_csv('./江门.csv')
```

In [35]:
```
df12=pd.read_csv('./河源.csv')
```

In [36]:
```
df13=pd.read_csv('./深圳.csv')
```

In [37]:
```
df14=pd.read_csv('./清远.csv')
```

In [38]:
```
df15=pd.read_csv('./湛江.csv')
```

In [39]:
```
df16=pd.read_csv('./潮州.csv')
```

In [40]:
```
df17=pd.read_csv('./珠海.csv')
```

In [41]:
```
df18=pd.read_csv('./肇庆.csv')
```

In [42]:
```
df19=pd.read_csv('./茂名.csv')
```

In [43]:
```
df20=pd.read_csv('./阳江.csv')
```

In [44]:
```
df21=pd.read_csv('./韶关.csv')
```

In [45]:
```
data=pd.concat([df1,df2,df3,df4,df5,df6,df7,df8,df9,df10],axis=0)
```

In [46]:
```
data1=pd.concat([df11,df12,df13,df14,df15,df16,df17,df18,df19,df20,df21],axis=0)
```

In [47]:
```
data=pd.concat([data,data1],axis=0)
```

In [49]:
```
data.to_csv('./综合.csv')
```

In [53]:
```
df=pd.read_csv('./new.csv')
```

In [54]:
```
df
```

Out[54]:

	城市	日期	累计确诊
0	东莞	2022/12/17	940
1	东莞	2022/12/16	940
2	东莞	2022/12/15	940
3	东莞	2022/12/14	939
4	东莞	2022/12/13	911
...	...	...	...
21618	韶关	2020/1/27	4
21619	韶关	2020/1/26	4
21620	韶关	2020/1/25	3
21621	韶关	2020/1/24	3
21622	韶关	2020/1/23	3

21623 rows × 3 columns

In [55]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21623 entries, 0 to 21622
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   城市      21623 non-null  object
 1   日期      21623 non-null  object
 2   累计确诊    21623 non-null  int64 
dtypes: int64(1), object(2)
memory usage: 506.9+ KB
```

提取年月日

In [58]:
```
import numpy as np
import datetime
df['日期']= pd.to_datetime(df['日期'])
df['Year'] = df['日期'].dt.year
df['Month'] = df['日期'].dt.month
```

In [59]:
```
df
```

Out[59]:

	城市	日期	累计确诊	Year	Month
0	东莞	2022-12-17	940	2022	12
1	东莞	2022-12-16	940	2022	12
2	东莞	2022-12-15	940	2022	12
3	东莞	2022-12-14	939	2022	12
4	东莞	2022-12-13	911	2022	12
...	...	...	...	...	...
21618	韶关	2020-01-27	4	2020	1
21619	韶关	2020-01-26	4	2020	1
21620	韶关	2020-01-25	3	2020	1
21621	韶关	2020-01-24	3	2020	1
21622	韶关	2020-01-23	3	2020	1

21623 rows × 5 columns

In [60]:
```
df1=df[df['Year']==2020]
```

In [61]:
```
df1
```

Out[61]:

	城市	日期	累计确诊	Year	Month
716	东莞	2020-12-31	101	2020	12
717	东莞	2020-12-30	101	2020	12
718	东莞	2020-12-29	101	2020	12
719	东莞	2020-12-28	101	2020	12
720	东莞	2020-12-27	101	2020	12
...	...	...	...	...	...
21618	韶关	2020-01-27	4	2020	1
21619	韶关	2020-01-26	4	2020	1
21620	韶关	2020-01-25	3	2020	1
21621	韶关	2020-01-24	3	2020	1
21622	韶关	2020-01-23	3	2020	1

6880 rows × 5 columns

In [62]:
```
df1_1=df[df['Month']==1]
```

In [63]:
```
df1_2=df[df['Month']==2]
```

In [64]:
```
df1_3=df[df['Month']==3]
```

In [65]:
```
df1_4=df[df['Month']==4]
```

In [66]:
```
df1_5=df[df['Month']==5]
```

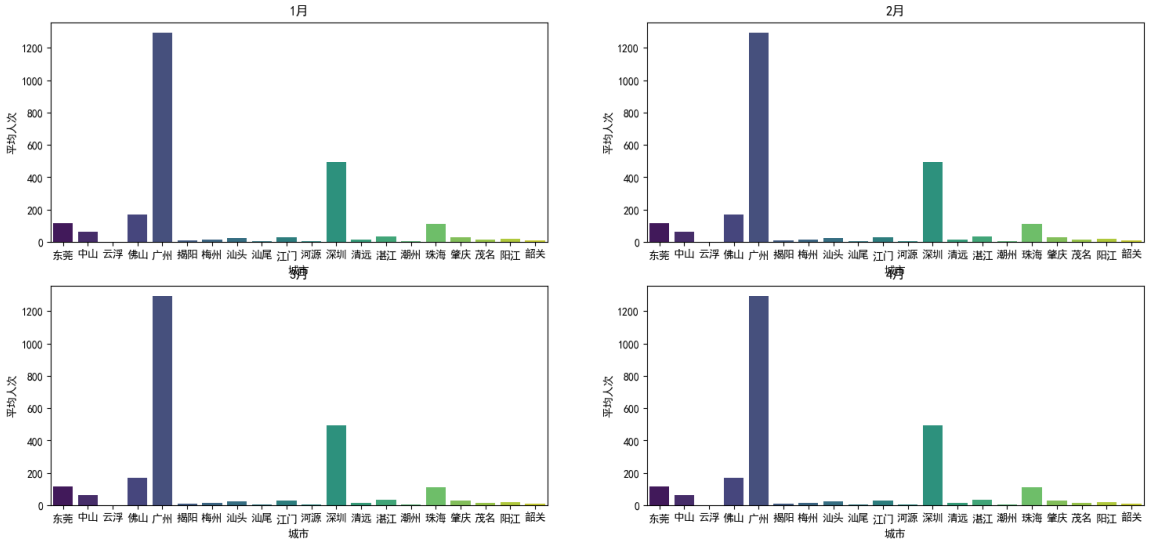
In [67]:
```
df1_6=df[df['Month']==6]
```

In [68]:
```
df1_7=df[df['Month']==7]
```

In [69]:
```
df1_8=df[df['Month']==8]
```

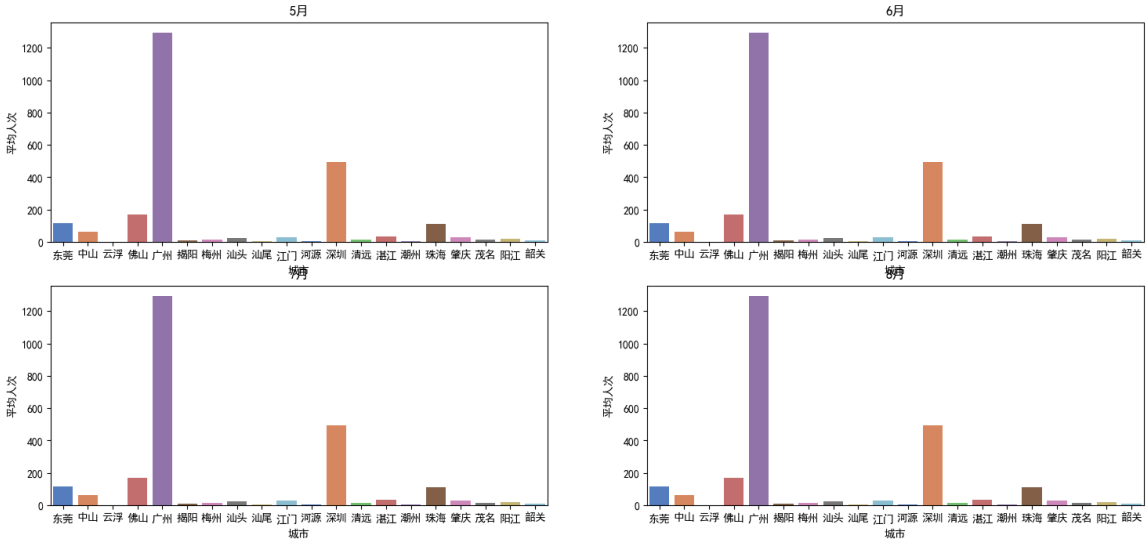
In [74]:
```
fig, axes = plt.subplots(2, 2, figsize=(18, 8))

month1 = df1_1.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='viridis', ax=axes[0, 0])
axes[0, 0].set_title('1月')
axes[0, 0].set_xlabel('城市')
axes[0, 0].set_ylabel('平均人次')

month2 = df1_2.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='viridis', ax=axes[0, 1])
axes[0, 1].set_title('2月')
axes[0, 1].set_xlabel('城市')
axes[0, 1].set_ylabel('平均人次')

month3 = df1_3.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='viridis', ax=axes[1, 0])
axes[1, 0].set_title('3月')
axes[1, 0].set_xlabel('城市')
axes[1, 0].set_ylabel('平均人次')

month4 = df1_4.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='viridis', ax=axes[1, 1])
axes[1, 1].set_title('4月')
axes[1, 1].set_xlabel('城市')
axes[1, 1].set_ylabel('平均人次')
```

Out[74]: Text(0, 0.5, '平均人次')



In [75]:
```
fig, axes = plt.subplots(2, 2, figsize=(18, 8))

month5 = df1_5.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='muted', ax=axes[0, 0])
axes[0, 0].set_title('5月')
axes[0, 0].set_xlabel('城市')
axes[0, 0].set_ylabel('平均人次')

month6 = df1_6.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='muted', ax=axes[0, 1])
axes[0, 1].set_title('6月')
axes[0, 1].set_xlabel('城市')
axes[0, 1].set_ylabel('平均人次')

month7 = df1_7.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='muted', ax=axes[1, 0])
axes[1, 0].set_title('7月')
axes[1, 0].set_xlabel('城市')
axes[1, 0].set_ylabel('平均人次')

month8 = df1_8.groupby('城市')['累计确诊'].mean().reset_index()
sns.barplot(x='城市', y='累计确诊', data=month1, palette='muted', ax=axes[1, 1])
axes[1, 1].set_title('8月')
axes[1, 1].set_xlabel('城市')
axes[1, 1].set_ylabel('平均人次')
```

Out[75]: Text(0, 0.5, '平均人次')



In [ ]:
```
```