

```
In [1]: import geopandas as gpd
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
import pylab as mpl #导入中文字体，避免显示乱码
mpl.rcParams['font.sans-serif']=['SimHei'] #设置为黑体字

## 读取数据
selected_area = gpd.read_file("outputs/selected_area_wgs84.geojson", driver="GeoJSON")
```

西建大周边建筑物类型分析

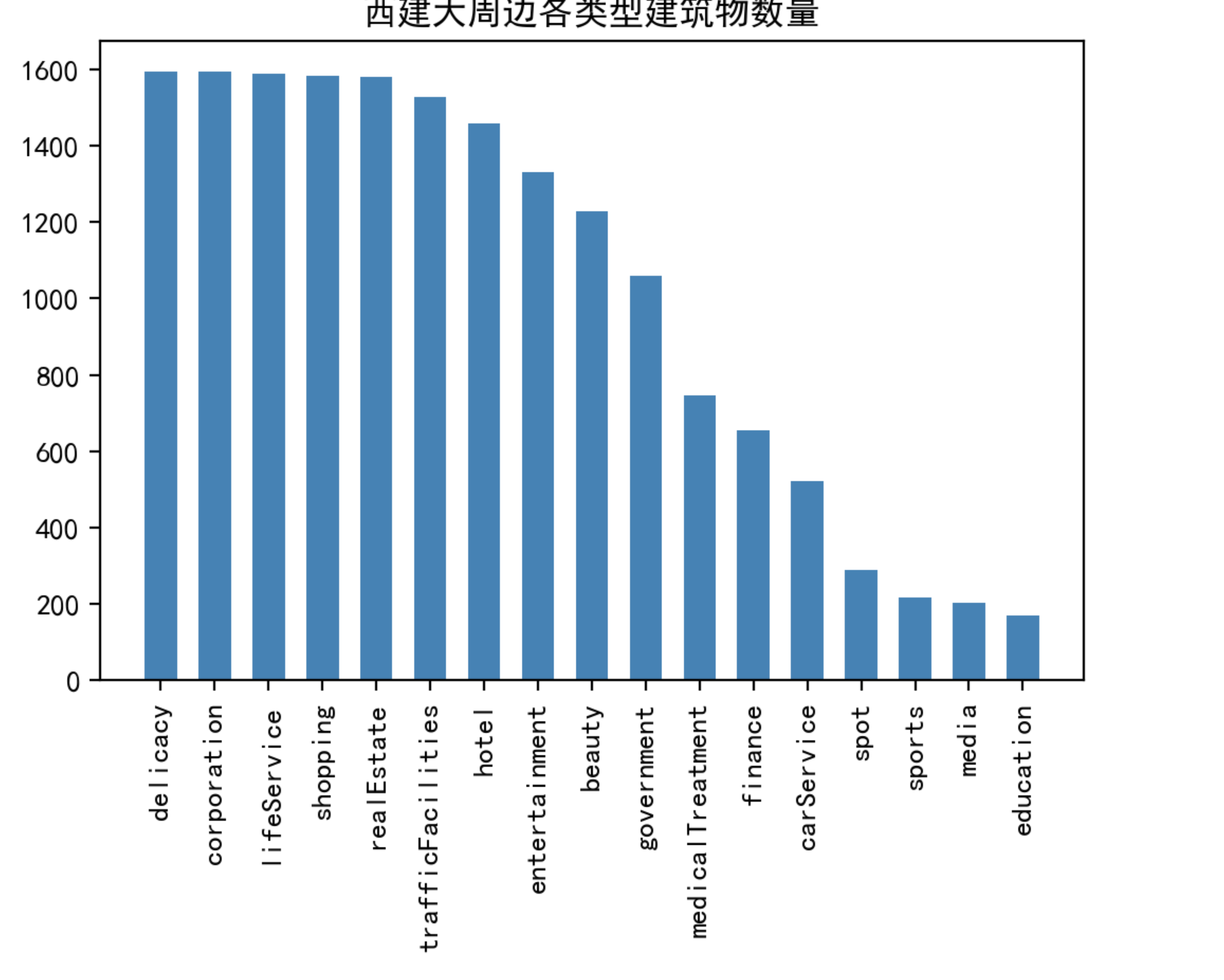
```
In [3]: ## 获取不同类型建筑的数量
tags = np.unique(selected_area.level_0)

tags_list = []
tags_num = []
for i in range(len(tags)):
    tags_list.append(tags[i][tags[i].index('_', 4)+1:])
    tags_num.append(len(selected_area[selected_area['level_0'] == tags[i]]))

df_tags = pd.DataFrame()
df_tags['tag'] = tags_list
df_tags['num'] = tags_num
df_tags = df_tags.sort_values('num',ascending=0)

In [4]: ## 柱状图
#生成figure对象
fig = plt.figure(figsize = (6,4), dpi = 200)
#生成axis对象
ax = fig.add_subplot(111) #本案例的figure中只包含一个图表
x = np.arange(len(df_tags))
x_ticks = df_tags['tag']
plt.xticks(x,x_ticks,rotation=90)
ax.bar(x,df_tags['num'],color='steelblue',width = 0.6)
plt.title('西建大周边各类型建筑物数量')

plt.show()
```



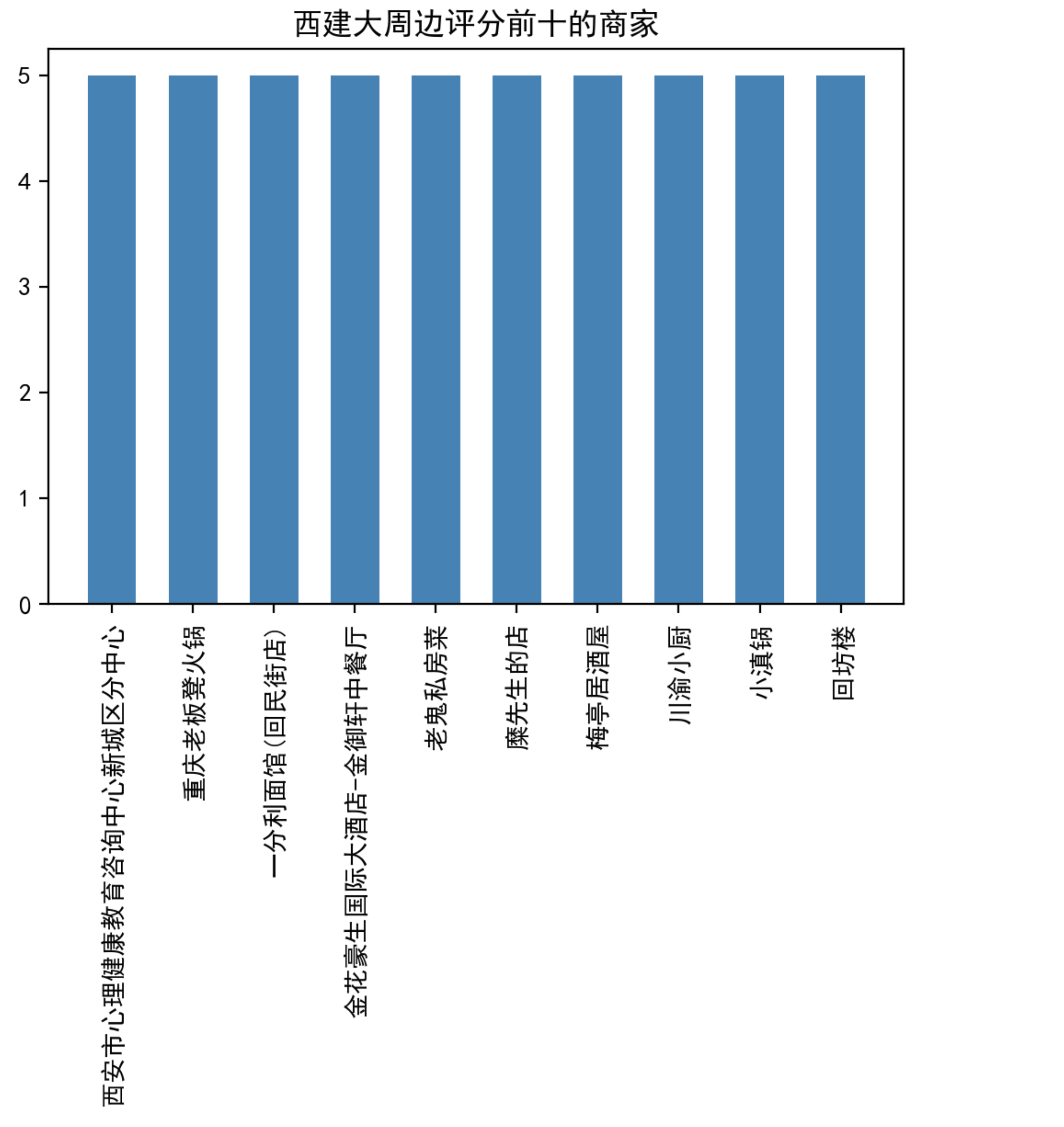
小结：
通过上面的分析可以看出，西建大周边建筑以金融和政府机构为主，健身娱乐等建筑较少。

商家评分情况统计

```
In [6]: df_rating = selected_area.sort_values('detail_info_overall_rating',ascending=0)

## 柱状图
#生成figure对象
fig = plt.figure(figsize = (6,4), dpi = 200)
#生成axis对象
ax = fig.add_subplot(111) #本案例的figure中只包含一个图表
x = np.arange(10)
x_ticks = df_rating['name'][:10]
plt.xticks(x,x_ticks,rotation=90)
ax.bar(x,df_rating['detail_info_overall_rating'][:10],color='steelblue',width = 0.6)
plt.title('西建大周边评分前十的商家')

plt.show()
```



```
In [7]: def frequency_bins(df,bins):
import pandas as pd
'''function-频数分布计算'''

#A-组织数据
column_name=df.columns[0]
column_bins_name=df.columns[0]+'_bins'
df[column_bins_name]=pd.cut(x=df[column_name],bins=bins,right=False) #参数right=False指定为包含左边值，不包括右边值。
df_bins=df.sort_values(by=column_name) #按照分割区间排序
df_bins.set_index([column_bins_name,df_bins.index],drop=False,inplace=True) #以price_bins和原索引值设置多重索引，同时配置drop=False参数保留原列。
#print(df_bins.head(10))

#B-频数计算
dfBins_frequency=df_bins[column_bins_name].value_counts() #dropna=False
dfBins_relativeFrequency=df_bins[column_bins_name].value_counts(normalize=True) #参数normalize=True将计算相对频数(次数) dividing all values by the
dfBins_freqANDrelFreq=pd.DataFrame({'fre':dfBins_frequency,'relFre':dfBins_relativeFrequency})
#print(dfBins_freqANDrelFreq)

#C-组中值计算
df_bins['rating']=df_bins['rating'].astype(float)
dfBins_median=df_bins.median(level=0)
dfBins_median.rename(columns={column_name:'median'},inplace=True)
#print(dfBins_median)

#D-合并分割区间、频数计算和组中值的DataFrame格式数据。
df_fre=dfBins_freqANDrelFreq.join(dfBins_median).sort_index().reset_index() #在合并时会自动匹配index
#print(ranmen_fre)

#E-计算频数比例
df_fre['fre_percent%']=df_fre.apply(lambda row:row['fre']/df_fre.fre.sum()*100,axis=1)

return df_fre
bins=np.arange(0,5.5,1) #配置分割区间 (组距)
rating_df = pd.DataFrame()
rating_df['rating'] = df_rating['detail_info_overall_rating']
Rating_fre=frequency_bins(rating_df,bins)
print(Rating_fre)

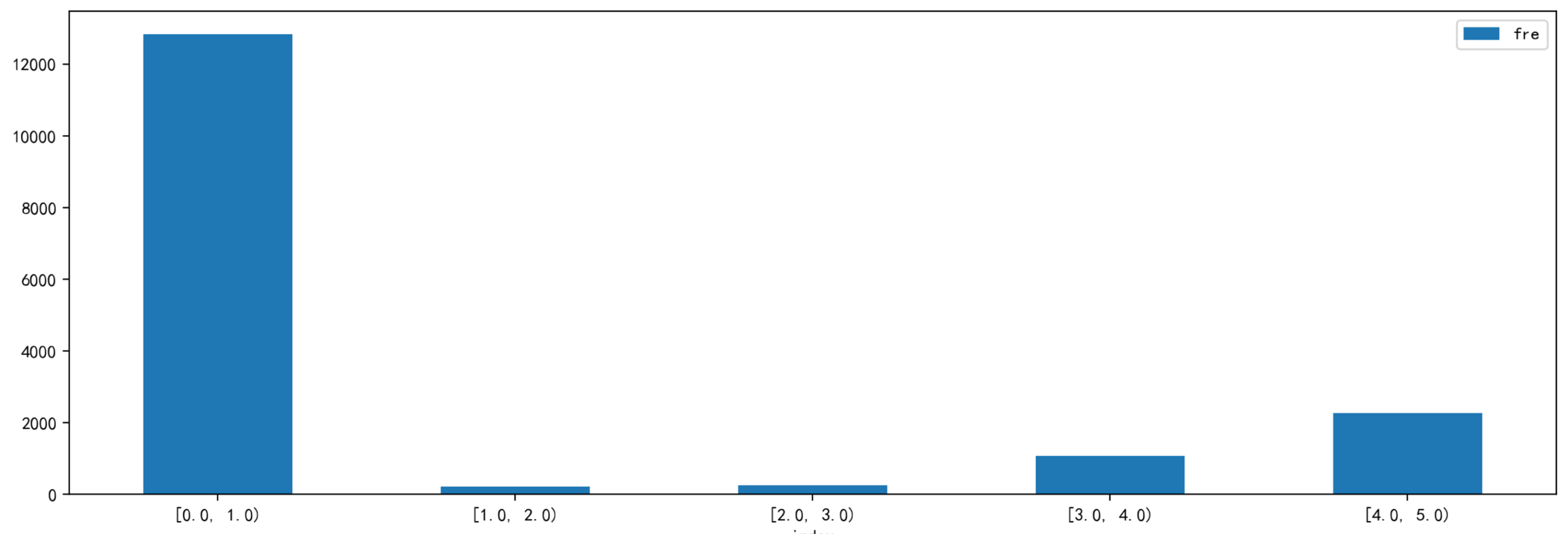
index    fre    relFre    median    fre_percent%
0 [0.0, 1.0)  12827  0.770159    0.0      77.015911
1 [1.0, 2.0)    215  0.012909    1.0      1.290904
2 [2.0, 3.0)   264  0.015851    2.3      1.585110
3 [3.0, 4.0)  1083  0.065026    3.5      6.502552
4 [4.0, 5.0)   2266  0.136055    4.4     13.605524

/var/folders/8n/dl2lhpvxini5_23k0mlvf19000000gn/T/ipykernel_15012/667557660.py:21: FutureWarning: Using the level keyword in DataFrame and Series aggregations is deprecated and will be removed in a future version. Use groupby instead. df.median(level=1) should use df.groupby(level=1).median().
dfBins_median=df_bins.median(level=0)
/Users/gdmcsie/opt/anaconda3/lib/python3.9/site-packages/pandas/core/generic.py:10351: FutureWarning: Dropping invalid columns in DataFrameGroupBy .median is deprecated. In a future version, a TypeError will be raised. Before calling .median, select only columns which should be valid for the function.
return self._agg_by_level(

In [8]: #生成figure对象
fig = plt.figure(figsize = (6,4), dpi = 200)
#生成axis对象
ax = fig.add_subplot(111) #本案例的figure中只包含一个图表

Rating_fre.loc[:,['fre','index']].plot.bar(x='index',rot=0,figsize=(15,5),ax = ax)

Out[8]: <AxesSubplot: xlabel='index'>
```



小结：
通过上面的分析可以看出，西建大周边评分前十的商家主要为汽车服务和酒店。西建大周边商家评分主要集中在0-1之间。

```
In [ ]:
```