In [ ]:
```
import json
import os
import pandas as pd
import os,pathlib
import util_poi
import geopandas as gpd
from shapely.geometry import Point
from pyproj import CRS
from tqdm import tqdm
```

In [ ]:
```
path='data/xianPOI_36/'
path
```

Out[ ]:
```
'data/xianPOI_36/'
```

1.选择区域

In [ ]:
```
loc1 = (108.93087,34.281836)
loc2 = (108.977654,34.258477)
```

2. 数据及预处理

POI（西安）数据

In [ ]:
```
fls = os.listdir(path)
fls
```

Out[ ]:
```
['poi_0_delicacy.json',
 'poi_10_medicalTreatment.json',
 'poi_11_carService.json',
 'poi_12_trafficFacilities.json',
 'poi_13_finance.json',
 'poi_14_realEstate.json',
```

```
'poi_15_corporation.json',
'poi_16_government.json',
'poi_1_hotel.json',
'poi_2_shopping.json',
'poi_3_lifeService.json',
'poi_4_beauty.json',
'poi_5_spot.json',
'poi_6_entertainment.json',
'poi_7_sports.json',
'poi_8_education.json',
'poi_9_media.json']
```

In [ ]:

```
df1 = []
for i in fls:
    # 读取json文件内容,返回字典格式
    with open(os.path.join(path,i) ,'r',encoding='utf8')as fp:
        json_data = json.load(fp)
    tmp = [str(i) for i in json_data]
    df1 += tmp
```

In [ ]:

```
pd.DataFrame(df1).to_csv('data/all.csv',index=False)
```

In [ ]:

```
fields_extraction=['name','location_lat', 'location_lng','detail_info_tag','detail_info_overall_rating', 'detail_info_price'] #配置需
save_path={'geojson':'./data/poiAll_gpd.geojson','shp':'./data/poiAll_gpd.shp','pkl':'./data/poiAll_gpd.pkl'} #分别存储为GeoJSON、Sha
poi_df_dic={}
i=0
poi_df=util_poi.csv2df('data/all.csv') #注释掉了了csv2df()函数内部的print("%s data type is not converted..."%(col))语句，以pass替代,
```

```
incorrect format of data_row number:0
__________________________________________________
.csv to DataFrame is completed!
```

In [ ]:

```
poi_fieldsExtraction=poi_df.loc[:,fields_extraction]
poi_fieldsExtraction.head()
```

Out[ ]:

	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price
1	户县兴伦美食广场	34.113672	108.614730	美食;中餐厅	4.5	10.0
2	户县机场老马家烧烤	34.157196	108.597683	美食;其他	3.5	NaN

	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price
3	户县印象	34.106033	108.630919	美食;中餐厅	3.0	31.0
4	长虹饭店(渼陂东路)	34.122130	108.615176	美食;中餐厅	3.7	138.0
5	聚鑫缘食府	34.115677	108.659891	美食;中餐厅	4.7	20.0

In [ ]:

```
poi_geoDF=poi_fieldsExtraction.copy(deep=True)
poi_geoDF['geometry']=poi_geoDF.apply(lambda row:Point(row.location_lng,row.location_lat),axis=1)
crs_32749=CRS('epsg:32749') #配置坐标系统，参考：https://spatialreference.org/
poiAll_gpd=gpd.GeoDataFrame(poi_geoDF,crs=crs_32749)
poiAll_gpd.to_pickle(save_path['pkl'])
poiAll_gpd.to_file(save_path['geojson'],driver='GeoJSON',encoding='utf-8')

poi_gpd = poiAll_gpd.copy()
idx = (poi_gpd['location_lng'].between(loc1[0],loc2[0]).values == True) *  (poi_gpd['location_lat'].between(loc2[1],loc1[1]).values
# poi_gpd[idx].plot(column='detail_info_overall_rating') #提取index为'poi_0_delicacy'的行查看结果

poiMy_gpd2shp=poiAll_gpd[idx].reset_index() #不指定level参数，例如Level=0，会把多重索引中的所有索引转换为列
poiMy_gpd2shp.rename(columns={
    'location_lat':'lat', 'location_lng':'lng',
    'detail_info_tag':'tag','detail_info_overall_rating':'rating', 'detail_info_price':'price'},inplace=True)
poiMy_gpd2shp.to_file(save_path['shp'],encoding='utf-8')
```

In [ ]:

```
poiAll_gpd.head(-5)
```

Out[ ]:

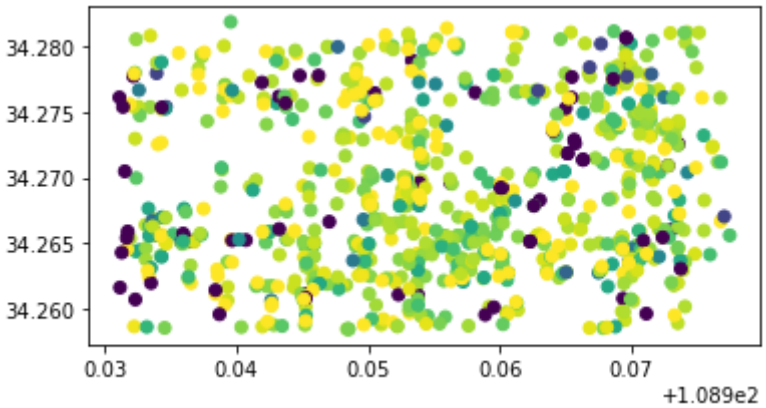
	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price	geometry
1	户县兴伦美食广场	34.113672	108.614730	美食;中餐厅	4.5	10.0	POINT (108.615 34.114)
2	户县机场老马家烧烤	34.157196	108.597683	美食;其他	3.5	NaN	POINT (108.598 34.157)
3	户县印象	34.106033	108.630919	美食;中餐厅	3.0	31.0	POINT (108.631 34.106)
4	长虹饭店(渼陂东路)	34.122130	108.615176	美食;中餐厅	3.7	138.0	POINT (108.615 34.122)
5	聚鑫缘食府	34.115677	108.659891	美食;中餐厅	4.7	20.0	POINT (108.660 34.116)
...	...	...	...	...	...	...	...
88805	西安秦影文化艺术传播有限公司	34.383195	109.202724	文化传媒;其他	NaN	NaN	POINT (109.203 34.383)

	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price	geometry
88806	西安市古德馨驿站	34.394470	109.293366	文化传媒;美术馆	NaN	NaN	POINT (109.293 34.394)
88807	西安工程大学校史馆	34.371336	109.194185	文化传媒;展览馆	NaN	NaN	POINT (109.194 34.371)
88808	西安科技大学临潼校区-与后勤管理处	34.370985	109.200650	文化传媒;文化宫	0.0	NaN	POINT (109.201 34.371)
88809	西安市临潼区东方婚庆歌舞艺术团	34.380975	109.221616	文化传媒;艺术团体	3.2	NaN	POINT (109.222 34.381)

88809 rows × 7 columns

In [ ]:
```
poi_gpd = poiAll_gpd.copy()
idx = (poi_gpd['location_lng'].between(loc1[0],loc2[0]).values == True) *  (poi_gpd['location_lat'].between(loc2[1],loc1[1]).values
poi_gpd[idx].plot(column='detail_info_overall_rating') #提取index为'poi_0_delicacy'的行查看结果
# poi_gpd.loc[:,:].plot(column='detail_info_overall_rating') #提取index为'poi_0_delicacy'的行查看结果
```

Out[ ]: <AxesSubplot:>



建筑高度数据（西安）

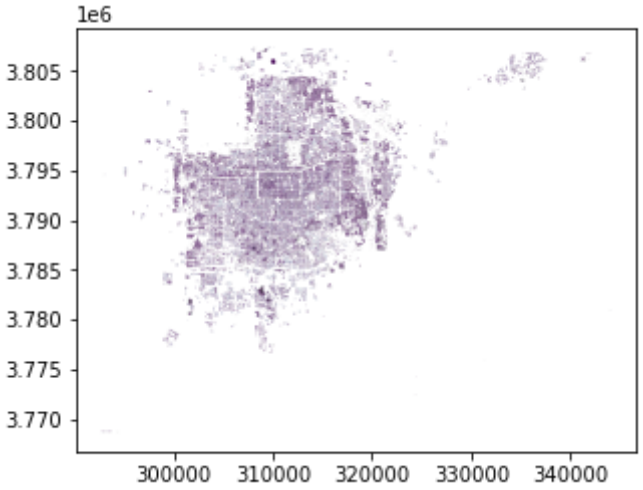
In [ ]:
```
import plotly.express as px
```

In [ ]:

```
poi_gpd=gpd.read_file('data/xianBuildingHeight/xian_Project.shp',crs='EPSG:32749') #读取存储的.shp格式文件
```

In [ ]:
```
poi_gpd.loc[:,:].plot(column='Floor',figsize=(5,5)) #提取index为'poi_0_delicacy'的行查看结果
```

Out[ ]: <AxesSubplot:>



In [ ]:
```
poi_gpd.head()
tmp = poi_gpd.copy()
```

In [ ]:
```
poi_gpd.dropna(axis=0,inplace=True)
poi_gpd.reset_index(drop=True, inplace=True)
poi_gpd.isna().sum()
```

Out[ ]:
```
Id          0
Floor       0
geometry    0
dtype: int64
```

In [ ]:
```
my_poi = []
for i in tqdm(range(poi_gpd.shape[0])):
    try:
        if poi_gpd['geometry'][i].exterior.coords.xy[0][0]<312800 and poi_gpd['geometry'][i].exterior.coords.xy[0][0]>308450 and p
            my_poi.append(list(poi_gpd.values[i]))
```

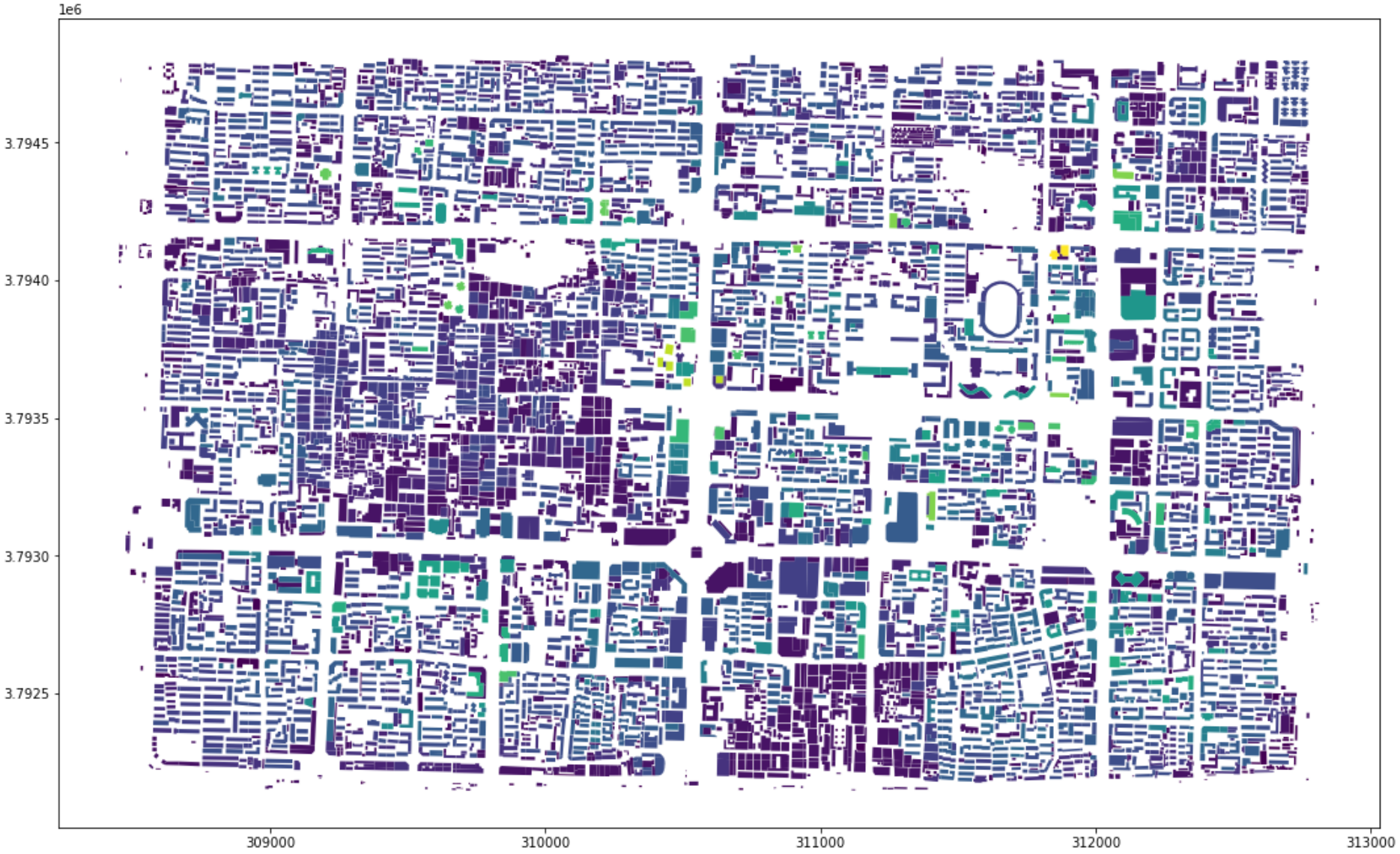
```
    except:
        continue
# my_poi_gpd = pd.DataFrame(my_poi,columns=poi_gpd.columns)
```

```
100%|██████████| 144481/144481 [01:11<00:00, 2021.40it/s]
```

In [ ]:

```
my_poi_gpd1=gpd.GeoDataFrame(my_poi,crs=crs_32749,columns=poi_gpd.columns)
my_poi_gpd1.plot(column='Floor',figsize=(18,18),aspect=1) #提取index为'poi_0_delicacy'的行查看结果
```

Out[ ]: <AxesSubplot:>

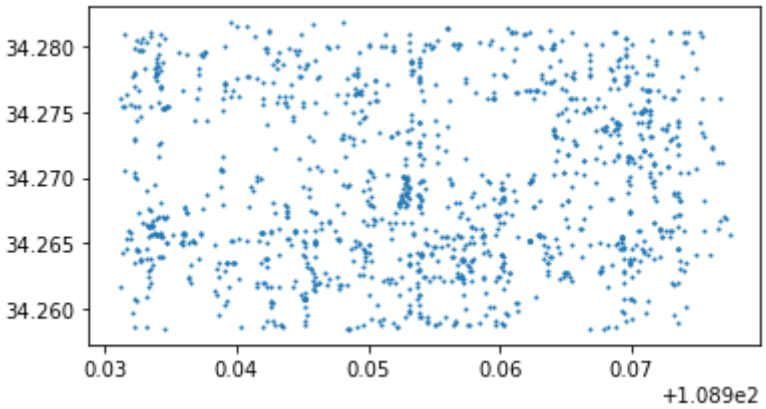


3. 分析

计算POI的核密度

In [ ]:
```
import pandas as pd
poi_gpd=pd.read_pickle('data/poiAll_gpd.pkl')[idx].reset_index()  #读取已经存储为.pkl格式的POI数据，其中包括geometry字段，为GeoDataF
poi_gpd.plot(marker=".",markersize=5) #查看POI数据是否读取正常
```

Out[ ]:
```
<AxesSubplot:>
```



In [ ]:
```
from scipy import stats
poi_coordinates=poi_gpd[['location_lng','location_lat']].to_numpy().T  #根据stats.gaussian_kde()输入参数要求确定数组结构
poi_coordi_kernel=stats.gaussian_kde(poi_coordinates) #核密度估计
poi_gpd['poi_kde']=poi_coordi_kernel(poi_coordinates)

import plotly.express as px
poi_gpd.detail_info_price=poi_gpd.detail_info_price.fillna(0)
mapbox_token='pk.eyJ1IjoicmljaGllYmFvIiwiYSI6ImNrYjB3N2NyMzBlMG8yc254dTRzNnMyeHMifQ.QT7MdjQKs9Y6OtaJaJAnOA'
px.set_mapbox_access_token(mapbox_token)
fig=px.scatter_mapbox(poi_gpd,lat=poi_gpd.location_lat, lon=poi_gpd.location_lng,color='poi_kde',color_continuous_scale=px.colors.s
fig.show()

poi_gpd.head()
```

Out[ ]:

	index	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price	geometry	poi_kde
0	3231	东湖黄鹤楼1958(南马道店)	34.258683	108.932237	美食;中餐厅	4.9	78.0	POINT (108.932 34.259)	152.783562
1	3291	老墙根	34.277957	108.932298	美食;中餐厅	3.9	78.0	POINT (108.932 34.278)	690.976069

	index	name	location_lat	location_lng	detail_info_tag	detail_info_overall_rating	detail_info_price	geometry	poi_kde
2	3361	大魏道火锅(西大街店)	34.266400	108.933109	美食;中餐厅	4.5	100.0	POINT (108.933 34.266)	986.459421
3	3364	鑫泰熙乐(西大街店)	34.265669	108.933951	美食;外国餐厅	4.6	34.0	POINT (108.934 34.266)	1126.397635
4	3390	宝华酒家	34.275345	108.933457	美食;中餐厅	4.2	38.0	POINT (108.933 34.275)	622.804393

使用K-means进行聚类

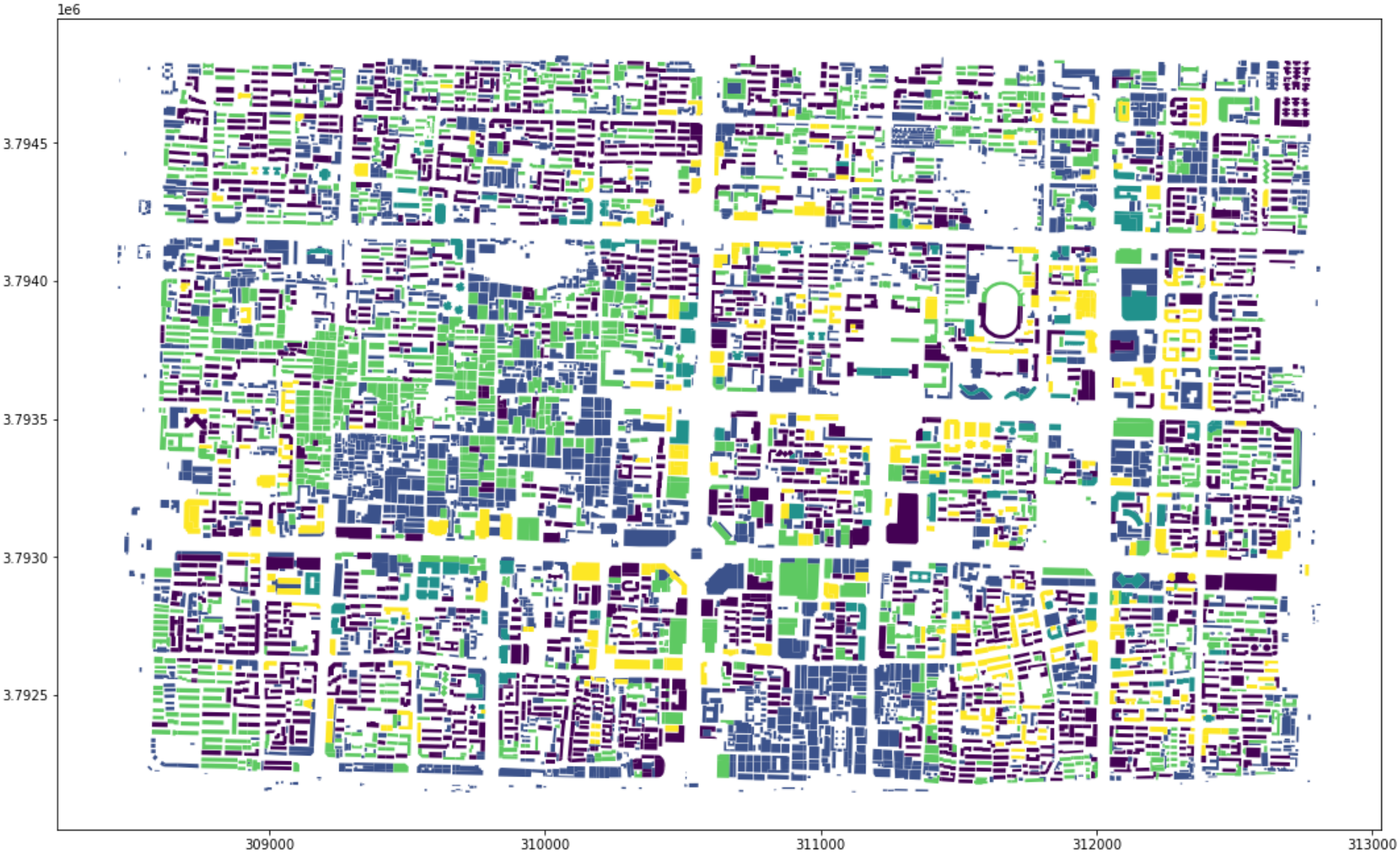
In [ ]:
```
from sklearn.cluster import KMeans
```

In [ ]:
```
estimator = KMeans(n_clusters=5)
estimator.fit(my_poi_gpd1['Floor'].values.reshape(-1, 1))
label_pred = estimator.labels_
label_pred
```

Out[ ]:
```
array([1, 1, 1, ..., 1, 1, 1])
```

In [ ]:
```
my_poi_gpd1['lable']=label_pred
my_poi_gpd1.plot(column='lable',figsize=(18,18),aspect=1)
```

Out[ ]:
```
<AxesSubplot:>
```



In [ ]:
```
my_poi_gpdl.head()
```

Out[ ]:

	Id	Floor	geometry	lable
0	0	1	POLYGON ((308455.70663 3794732.85541, 308457.6...	1

	Id	Floor	geometry	lable
1	0	2	POLYGON ((308618.39953 3794584.77384, 308692.2...	1
2	0	2	POLYGON ((308621.59962 3794790.22151, 308649.2...	1
3	0	3	POLYGON ((308617.28897 3794627.08183, 308626.1...	1
4	0	3	POLYGON ((308704.37140 3794786.45367, 308717.1...	1