**爬取去哪儿网热门景点地理位置、销量信息，利用百度地图API生成热力图：**

百度地图API Key申请方法：

访问百度地图开放平台：<http://lbsyun.baidu.com/>

点击页面下方的“申请密钥”，需要登录你的百度账号。



进入之后会看到百度地图LBS开放平台。

1. 选择“创建应用”

2.“应用名称”可以自己随意编辑，如“热力图”

3. “启用服务”勾选“Javascript API”和“Geocoding API v2”。

4.“Referer白名单”处建议就填\*，方便在不同的电脑上访问。

5. 点击“提交”。



进入应用列表页面，可以看到新创建的应用“热力图”的API Key（AK）：



Geocoding API v2 用于将景点地址信息转换为经纬度信息：



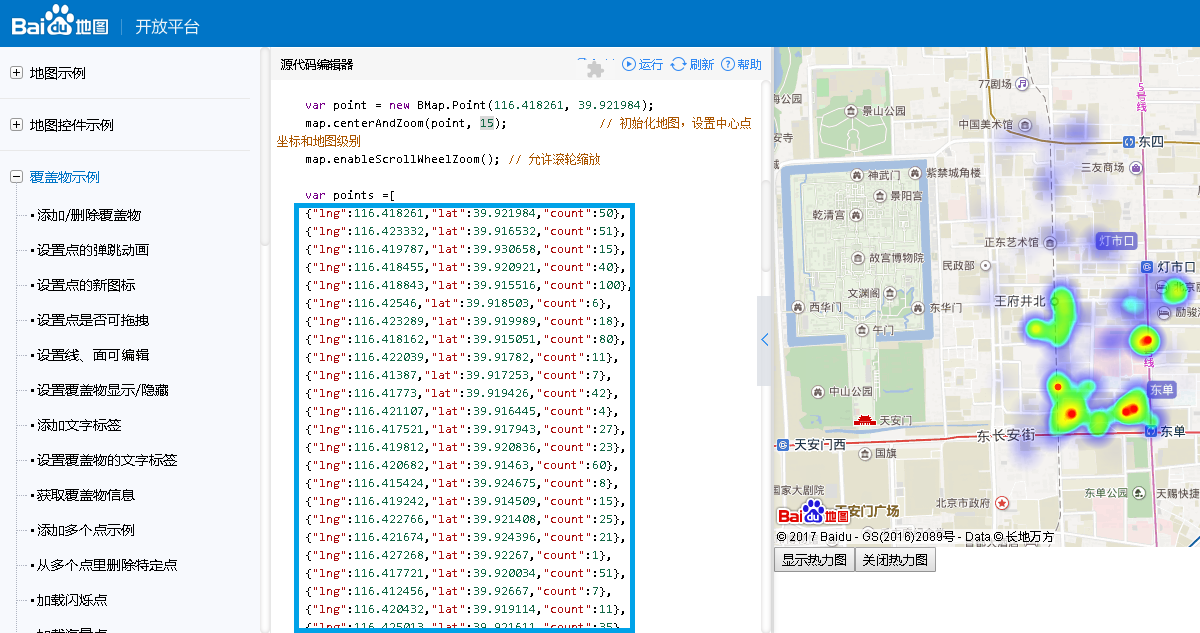
Javascript API 用于在网页中调用百度地图Heatmap\_min.js将JSON数据渲染成热力图：



可访问下面网址查看热力图示例HTML代码：

<http://developer.baidu.com/map/jsdemo.htm#c1_15>

蓝色方框内的是热力地图JSON源数据：

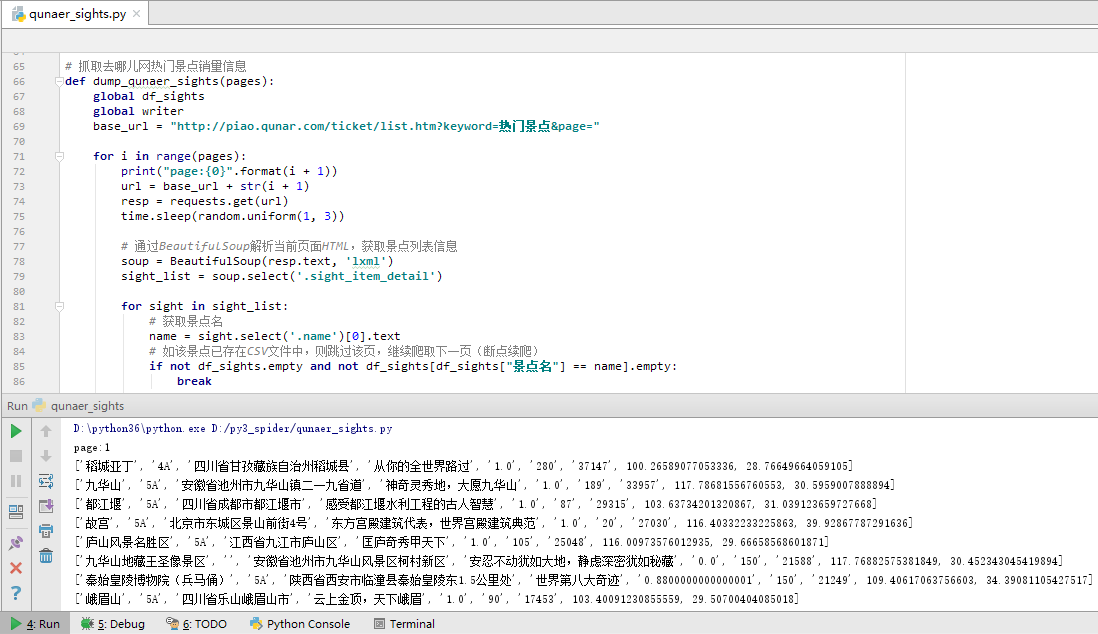


**1）去哪儿网热门景点爬虫代码(qunaer\_sights.py)：**

爬取时间2017年9月15日，获取当月热门景点门票销量数据：

**import** requests  
**from** bs4 **import** BeautifulSoup  
**import** pandas **as** pd  
**import** time  
**import** random  
**import** urllib.parse  
**import** json  
**import** csv  
  
  
*# 读取上次已爬取的CSV文件，以便断点续爬***def** init\_df():  
 **global** df\_sights  
 **try**:  
 df\_sights = pd.read\_csv(**"D:/qunaer\_sights.csv"**)  
 **except**:  
 df\_sights = pd.DataFrame()  
  
  
*# 初始化CSV writer，第一次需写入列名***def** init\_csv():  
 **global** f  
 **global** writer  
 **global** df\_sights  
 csvFile = **"D:/qunaer\_sights.csv"** f = open(csvFile, **"a+"**, newline=**""**, encoding=**'utf-8'**)  
 writer = csv.writer(f)  
 **if** df\_sights.columns.empty:  
 writer.writerow([**"景点名"**, **"等级"**, **"地址"**, **"介绍"**, **"热度"**, **"价格"**, **"月销量"**, **"经度"**, **"纬度"**])  
  
  
*# 程序结束前关闭CSV文件***def** close\_csv():  
 **global** f  
 f.close()  
  
  
*# 调用百度地图API获取景点地址对应的经纬度***def** get\_geo\_info(address):  
 geo\_url = **"http://api.map.baidu.com/geocoder/v2/?"** geo\_params = {  
 **"output"**: **"json"**,  
 **"ak"**: **"tAwAffII9G0F8Gs4VRuXnuFmIdbOKCEu"** *# 替换成你的百度地图API Key* }  
  
 *# 更新url中的地址参数* geo\_params.update({**"address"**: address})  
 data = urllib.parse.urlencode(geo\_params)  
 cur\_geo\_url = geo\_url + data  
 geo\_resp = requests.get(cur\_geo\_url)  
  
 json\_data = json.loads(geo\_resp.text)  
  
 *# 调用成功， 获取JSON data中的经纬度信息* **if** json\_data[**"status"**] == 0:  
 longitude = json\_data[**"result"**][**"location"**][**"lng"**]  
 latitude = json\_data[**"result"**][**"location"**][**"lat"**]  
 **else**:  
 longitude = **""** latitude = **""  
  
 return** longitude, latitude  
  
  
*# 抓取去哪儿网热门景点销量信息***def** dump\_qunaer\_sights(pages):  
 **global** df\_sights  
 **global** writer  
 base\_url = **"http://piao.qunar.com/ticket/list.htm?keyword=热门景点&page="  
  
 for** i **in** range(pages):  
 print(**"page:{0}"**.format(i + 1))  
 url = base\_url + str(i + 1)  
 resp = requests.get(url)  
 time.sleep(random.uniform(1, 3))  
  
 *# 通过BeautifulSoup解析当前页面HTML，获取景点列表信息* soup = BeautifulSoup(resp.text, **'lxml'**)  
 sight\_list = soup.select(**'.sight\_item\_detail'**)  
  
 **for** sight **in** sight\_list:  
 *# 获取景点名* name = sight.select(**'.name'**)[0].text  
 *# 如该景点已存在CSV文件中，则跳过该页，继续爬取下一页（断点续爬）* **if not** df\_sights.empty **and not** df\_sights[df\_sights[**"景点名"**] == name].empty:  
 **break** *# 获取景点等级* **try**:  
 level = sight.select(**'.level'**)[0].text.replace(**"景区"**, **""**)  
 **except**:  
 level = **""** *# 获取景点地址* address = sight.select(**'.address.color999 span'**)[0].text.replace(**"地址："**, **""**)  
  
 *# 获取景点介绍* intro = sight.select(**'.intro.color999'**)[0].text  
  
 *# 获取景点热度* star = sight.select(**'.product\_star\_level em span'**)[0].text.replace(**"热度 "**, **""**)  
  
 *# 获取门票价格，月销量* **try**:  
 price = sight.select(**'.sight\_item\_price em'**)[0].text  
 sales = sight.select(**'.hot\_num'**)[0].text  
 **except**:  
 **continue** *# 将景点地址转换为经纬度* longitude, latitude = get\_geo\_info(address)  
  
 *# 向CSV文件中插入一条景点信息* sight\_item = [name, level, address, intro, star, price, sales, longitude, latitude]  
 print(sight\_item)  
 writer.writerow(sight\_item)  
  
  
**if** \_\_name\_\_ == **"\_\_main\_\_"**:  
 init\_df()  
 init\_csv()  
 dump\_qunaer\_sights(pages=400)  
 close\_csv()

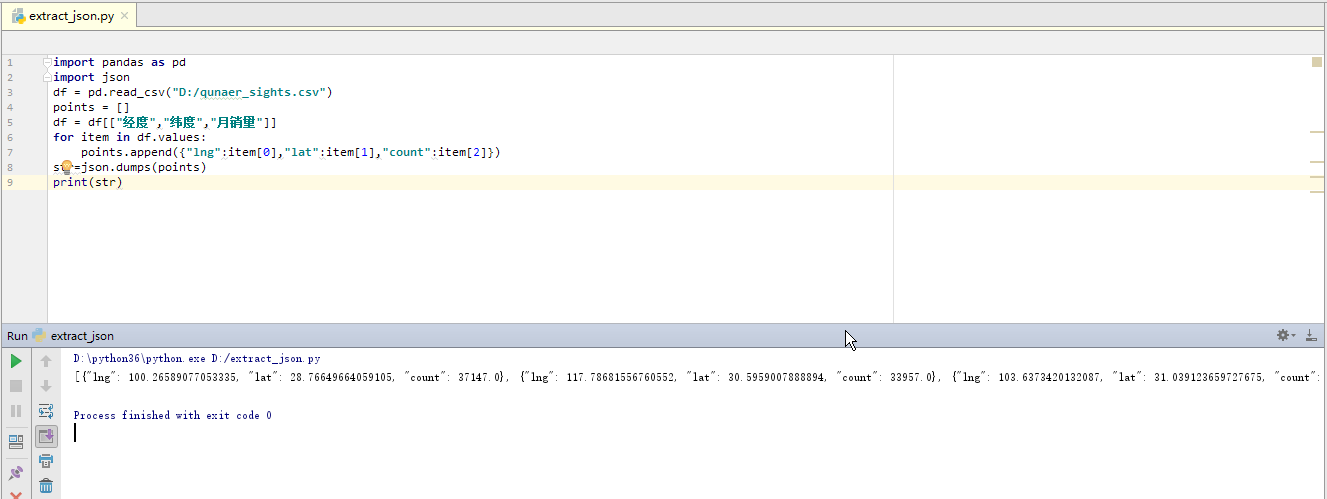
PyCharm运行结果：



**2）提取出CSV文件中的经纬度和销量信息，转化为JSON数据并打印(extract\_json.py)：**

**import** pandas **as** pd  
**import** json  
df = pd.read\_csv(**"D:/qunaer\_sights.csv"**)  
points = []  
df = df[[**"经度"**,**"纬度"**,**"月销量"**]]  
**for** item **in** df.values:  
 points.append({**"lng"**:item[0],**"lat"**:item[1],**"count"**:item[2]})  
str=json.dumps(points)  
print(str)

PyCharm运行结果：



1. **创建景点门票销量热力地图HTML文件(heat.html)：**

红色字体的API key替换成你的百度地图API Key：

<!DOCTYPE html>

<html>

<head>

<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<meta name="viewport" content="initial-scale=1.0, user-scalable=no" />

<script type="text/javascript" src="http://api.map.baidu.com/api?v=2.0&ak=tAwAffII9G0F8Gs4VRuXnuFmIdbOKCEu"></script>

<script type="text/javascript" src="http://api.map.baidu.com/library/Heatmap/2.0/src/Heatmap\_min.js"></script>

<title>热力图功能示例</title>

<style type="text/css">

ul,li{list-style: none;margin:0;padding:0;float:left;}

html{height:100%}

body{height:100%;margin:0px;padding:0px;font-family:"微软雅黑";}

#container{height:100%;width:100%;}

</style>

</head>

<body>

<div id="container"></div>

</body>

</html>

<script type="text/javascript">

var map = new BMap.Map("container"); // 创建地图实例

var point = new BMap.Point(110, 37.5);

map.centerAndZoom(point, 5); // 初始化地图，设置中心点坐标和地图级别

map.enableScrollWheelZoom(); // 允许滚轮缩放

var points = [{"lng": 100.26589077053335, "lat": 28.76649664059105, "count": 37147.0}, {"lng": 117.78681556760552, "lat": 30.5959007888894, "count": 33957.0}]; **//这里需填写上步中打印出的JSON数据（约6千条）**

if(!isSupportCanvas()){

alert('热力图目前只支持有canvas支持的浏览器,您所使用的浏览器不能使用热力图功能~')

}

//详细的参数,可以查看heatmap.js的文档 https://github.com/pa7/heatmap.js/blob/master/README.md

//参数说明如下:

/\* visible 热力图是否显示,默认为true

\* opacity 热力的透明度,1-100

\* radius 势力图的每个点的半径大小

\* gradient {JSON} 热力图的渐变区间 . gradient如下所示

\* {

.2:'rgb(0, 255, 255)',

.5:'rgb(0, 110, 255)',

.8:'rgb(100, 0, 255)'

}

其中 key 表示插值的位置, 0~1.

value 为颜色值.

\*/

heatmapOverlay = new BMapLib.HeatmapOverlay({"radius":8});

map.addOverlay(heatmapOverlay);

heatmapOverlay.setDataSet({data:points,max:100});

heatmapOverlay.show();

//是否显示热力图

function openHeatmap(){

heatmapOverlay.show();

}

function closeHeatmap(){

heatmapOverlay.hide();

}

//closeHeatmap();

function setGradient(){

/\*格式如下所示:

{

0:'rgb(102, 255, 0)',

.5:'rgb(255, 170, 0)',

1:'rgb(255, 0, 0)'

}\*/

var gradient = {};

var colors = document.querySelectorAll("input[type='color']");

colors = [].slice.call(colors,0);

colors.forEach(function(ele){

gradient[ele.getAttribute("data-key")] = ele.value;

});

heatmapOverlay.setOptions({"gradient":gradient});

}

//判断浏览区是否支持canvas

function isSupportCanvas(){

var elem = document.createElement('canvas');

return !!(elem.getContext && elem.getContext('2d'));

}

</script>

用浏览器打开heat.html，查看景点人气（门票月销量）热力图：

可发现北京、江浙沪、广州、青岛、四川等地是游客扎堆的地方，国庆期间可避开这些过于热门的景点，而选择一些相对小众的去处。