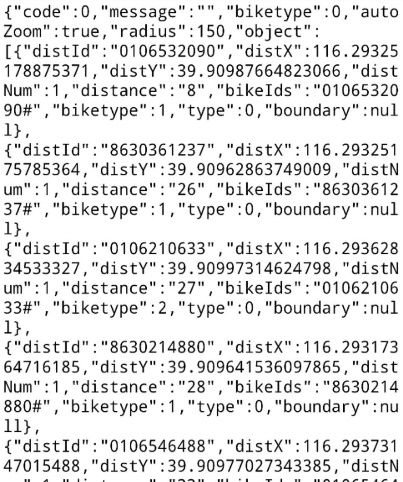
基于网络爬虫的摩拜单车移动趋势大数据分析

引言：在大数据的背景下，摩拜单车的骑行数据自然是分析城市人群出行爱好的最佳数据，故利用python网络爬虫对摩拜公司的单车位置信息进行实时抓取，通过GIS系统对骑行数据进行分析，从而了得出0时-9时、9时-16时、16时-22时出行喜好的结论。

# 实现方法

## 1.数据获取：

在手机上对微信摩拜小程序进行抓包，得到以下JSON数据。

 于是，通过对本地POST发送数据分析，获得摩拜单单车GPS信息获取API接口。

此后利用Python，将API中发送数据中的latitude变量，和longitude分别写入两个嵌套循环，在北京（经度：116~116.8000 ; 纬度：39.6000~40.3000）范围内，以0.008作为步长对北京市六环以内进行快速扫描，从而得到摩拜单车GPS当前数据。

经过与手机APP（摩拜单车）的对比，可以确定抓取数据为有效数据，非反爬虫机制的模拟数据。

对于车辆类型，JSON数据中的biketype有1和2两种可能，经过与手机客户端的对比，1代表lite型车，2为老款车。

至此，数据获取部分结束。

## 2.数据处理：

得到抓取的原始数据后，还不能将其输入GIS系统进行分析。由于其中某些数据由于GPS波动造成车辆频繁移动或大幅度一定，因此需要对数据进行进一步筛选。

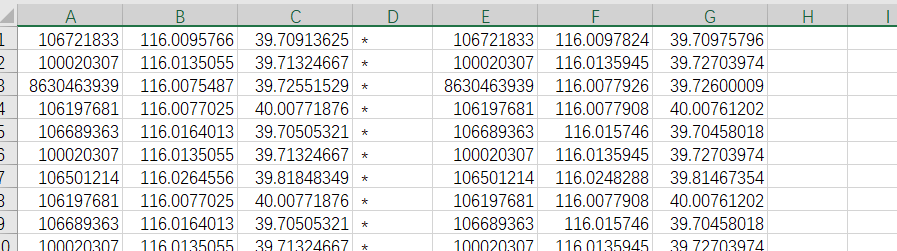
根据研究目的，选择以下筛选方法：

保留：车辆 ID相同且位置变化适中 的车辆

去除：车辆 ID不同或位置变化极大（小） 的车辆

根据所选研究的时间区段，移动经纬度超过0.4可能性不大，于是选择0.4作为数据的上限，由于车辆的小规模移动（GPS波动）同样可能，于是选择经纬度变化0.0002作为下限。

处理后，分别将同一ID自行车的前后位置存在excel表格中的同一列，如下图：



自左向右，分别为ID号，起点经度，起点纬度，ID号（同上），终点经度，终点纬度。

核心代码如下:

for temline1 in range(1,line1):

for temline2 in range(1,line2):

if file1[temline1][id] == file2[temline2][id]:

if (file1[temline1][x] != file2[temline2][x] ) and (file1[temline1][y] != file2[temline2][y] ):

#print(file1[temline1][x])

#print(file2[temline2][x])

subx = abs(str2float(str(file1[temline1][x])) - str2float(str(file2[temline2][x])))

suby = abs(str2float(str(file1[temline1][y])) - str2float(str(file2[temline2][y])))

#print(str(subx) + str(suby))

#print(temline2)

if (subx > 0.0002 ) and (suby > 0.0002):

if (subx < 0.4) and (suby < 0.4 ):

print('subX : ' + str(subx))

print('subY : ' + str(suby))

print(file1[temline1])

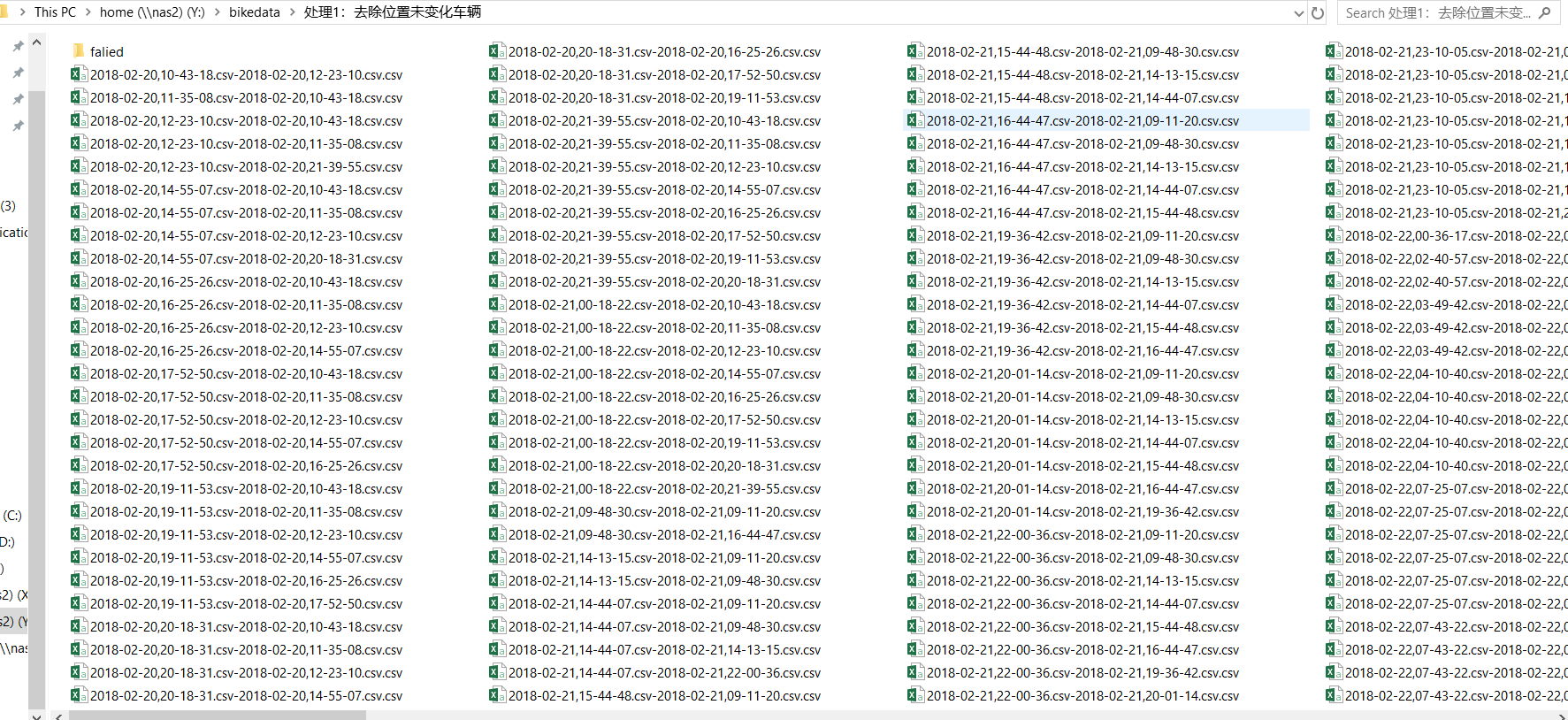
print(file2[temline2])

csv\_writer = csv.writer(out)

csv\_writer.writerow([file1[temline1][0],file1[temline1][1],file1[temline1][2] , " \* " , file2[temline2][0],file2[temline2][1],file2[temline2][2]] )

break

以下为处理后文件：

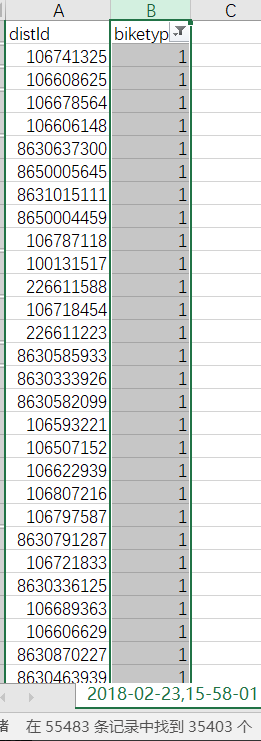
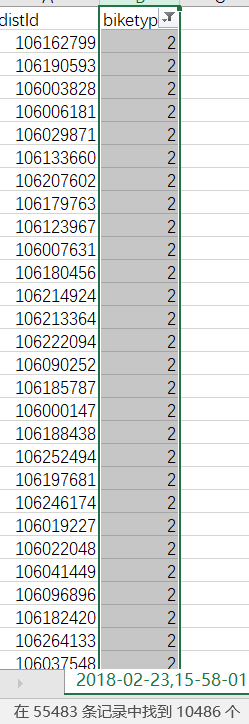
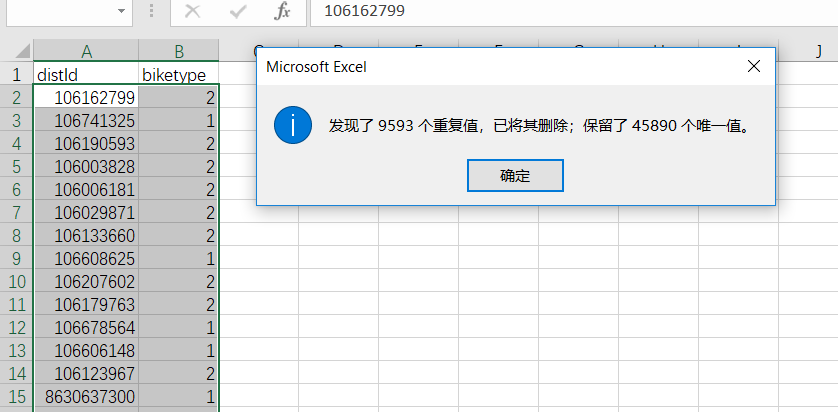


至此，数据处理部分结束。

# 数据分析

## 车辆类型分析

首先对ID进行去重处理：



除去重复项后，分别统计1和2类型的数量：

2类型包含10486个

1类型包含35403个

摩拜北京官方总投放量：37万辆

对北京市的总车辆类型估计如下：  
2类型约为: (69985辆

1类型约为: (236284辆

可以看出摩拜公司在车辆改进后，大量投放新车的现象，原旧版本车辆单车成本高达5000元而新车仅3000元，数据符合客观规律。

与时隔一年前的网上非官方数据分析产生了较大变化：

原文地址：https://www.jianshu.com/p/2a20d2a97ac0

## 车辆移动分析

根据数据统计，有近4成车辆在研究时间内从未发生移动。对于此现象分析有两种可能：

* 1. 由于时间是冬天，骑行人较少
  2. 车辆处在小区内、偏僻位置或无法开始骑行（损坏，上锁）

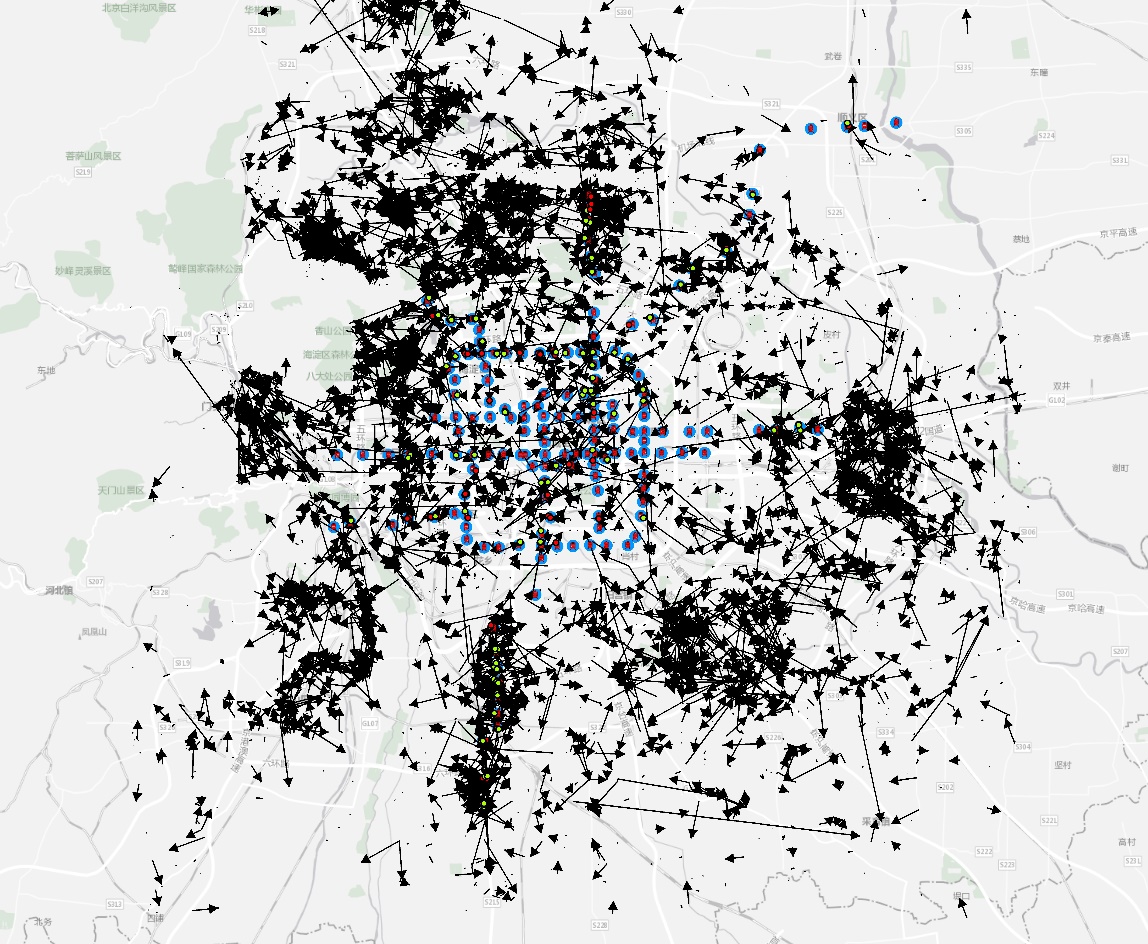
3.移动趋势分析：  
 首先，通过百度地图开放平台获取北京市内主要地铁站的坐标，并将其输入GIS系统，并在其周围200米建立缓冲区，便于统计车辆移动信息。

之后，将处理后文件输入GIS系统，由起点向终点连线。同时将起点和终点坐标输入GIS系统，与缓冲区做相交处理，将起点设为红色，终点绿色。

以下为数据分析：

### 宏观分析：

根据总体数据，用车集中区不在北京市市中心附近，而是呈现围绕地铁的条带状，或是在郊区的聚集形态



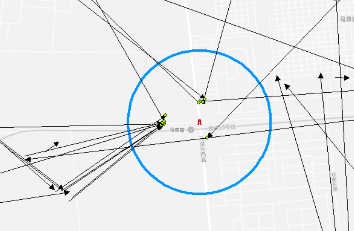
例如在地图的左下方的聚集区均围绕地铁，而位于地图左上的聚集区经百度地图查询，为三个产业园聚集处，青年人居多，故出现了聚集现象。

同时由图可见，中心城区骑行次数并不多

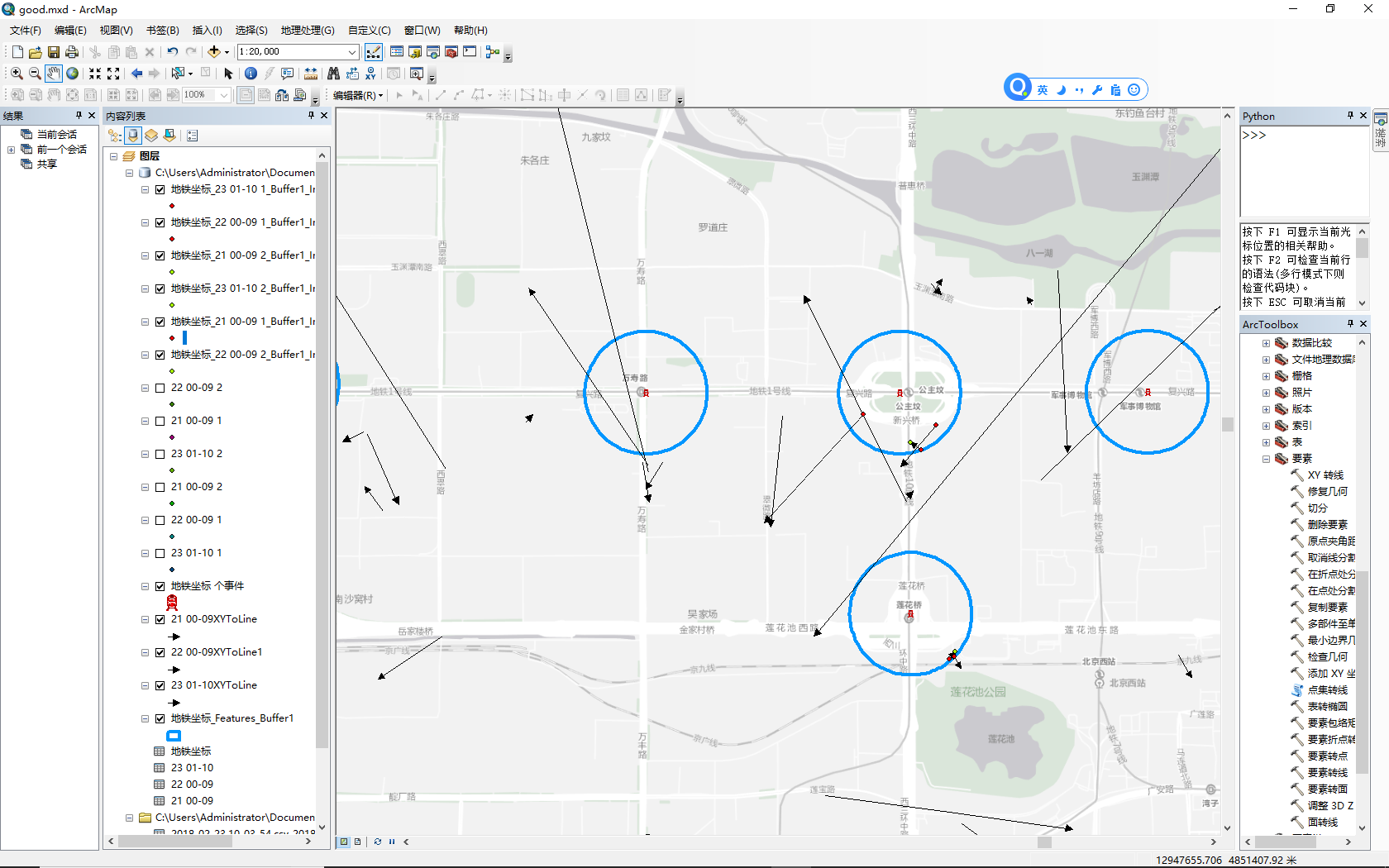
可能原因：

1. 城市生活水平较高，大部分居民在冬天选择乘私家车出行。
2. 城市公共交通便利，路网发达，冬天骑行相比较于公交车舒适性较差。

### 0时-9时：

 此时间段内，在郊区的地铁多呈现为绿色点聚集，也就是大量居民选择地铁站作为终点。

此为北京市东北五环外的马泉营地铁站流向图，在上班高峰期，居民多乘坐地铁向城市中心移动，而几乎没有人从地铁站出发向外骑行。



在市中心的情况就与之相反，多呈现红色聚集，多数地铁乘客下车后，乘摩拜自行车前往上班的地点。例如：处于西三环的公主坟地铁站，大多数骑行都呈现向外的发散状。

### C:\Users\ADMINI~1\AppData\Local\Temp\WeChat Files\271521428608022795.png9时-16时：

此区间内城市和郊区地铁区域无较大差别，而相比于早上，产业园附近车辆移动更为密集，且出发点与终点大多重合，经手机百度地图确认，这些地点附近大多有麦当劳，肯德基之类的餐饮企业。