

# OpenStreetMap 案例研究

## 一、地图区域

美国纽约

下载地址: <https://www.openstreetmap.org/export#map=11/40.7241/-73.9689>

选择纽约是想了解下纽约这个城市

## 二、地图中发现的问题

- 1.街道名存在缩写、写错、首字母大小写不统一等情况
- 2.带有数字的街道名写的方式不统一, 有的带“#”有的不带

对数据进行修改

```
def update_value(value, mapping):  
    if value in mapping:  
        value=mapping[value]  
    else:  
        last=value.split()[-1]  
        if last in mapping:  
            value=value.replace(last,mapping[last])  
    return value
```

## 三、数据统计信息

### 1.文件大小

```
map.osm ..... 53.9 MB  
osm.db ..... 31.1 MB  
nodes.csv ..... 15.9 MB  
nodes_tags.csv ..... 2.31 MB  
ways.csv ..... 2.05 MB  
ways_tags.csv ..... 6.32 MB  
ways_nodes.cv ..... 7.04 MB
```

### 2.nodes 数量

```
sqlite> SELECT COUNT(*) FROM nodes;  
185622
```

### 3.ways 数量

```
sqlite> SELECT COUNT(*) FROM ways;  
33390
```

### 4.unique users 数量

```
sqlite> SELECT COUNT(DISTINCT(e.uid))  
FROM (SELECT uid FROM nodes UNION ALL SELECT uid FROM ways) e;  
1051
```

#### 5.前十位贡献者数量

```
SELECT e.user, COUNT(*) as num  
FROM (SELECT user FROM nodes UNION ALL SELECT user FROM ways) e  
GROUP BY e.user  
ORDER BY num DESC  
LIMIT 10;
```

```
Rub21_nycbuildings|88077  
robgeb|42555  
lxbarth_nycbuildings|33760  
ALE!|6709  
Korzun|6100  
celosia_nycbuildings|2511  
tre1994|2105  
zingbot_nycbuildings|1794  
lxbarth|1396  
cityracks|1267
```

#### 6.highway 的数量

```
sqlite> SELECT COUNT(*) FROM ways_tags WHERE key="highway";  
4469
```

#### 7.bicycle 的数量

```
sqlite> SELECT COUNT(*) FROM ways_tags WHERE key="bicycle";  
636
```

#### 8.餐厅的数量

```
sqlite> SELECT COUNT(*) FROM nodes_tags WHERE value="restaurant";  
792
```

### 四、改进建议

在数据清洗过程中，节点的 k 值命名比较混乱，有的大写有的小写有的简写有的写错有的有符号有的没符号，没有统一的标准，这给数据清洗增加了很大的工作量，特别是数据比较大的时候，建议地图贡献者严格按照统一的标准去编辑地图，同时从地图前十位贡献者贡献的数量可以看到，他们的工作量是非常大的，这就更容易出错了，建议激励更多的人去完成这个工作，大家每个人的工作量减小，每个人都按照要求按照标准去编辑地图，那么错误会大大减少，更利于使用地图者进行数据清洗工作，减少很多不必要的麻烦。不过这也可能会带来一些问题，新参加的贡献者对地图不熟悉，可能会出现很多错误，并且贡献者太多可能会导致贡献风格很难统一，即使按照了一定的规范来贡献可能也很难统一起来。