

# 摩拜单车的 Tableau 故事

第一版[工作簿链接](#)

第二版[工作簿链接](#)

## 总结：

对摩拜单车 2016 年 8 月份的骑行信息进行可视化分析。经过分析可以得出以下结论：摩拜单车日订单数不断攀升，骑行高峰对应上下班高峰，在上海的分布主要是在市中心地区。

## 设计：

1. 首先对数据进行清洗，包括把 id 改成文字类型，把坐标改成地理-经纬度等。
2. 工作表：
  - a) 每日订单总数：统计这个月每天的订单数，做线形图。线形图可以很好的展示变量随时间变化的趋势。
  - b) 不同时段的订单数：统计不同时段的订单数并将周末与非周末分开，做线形图。周末与非周末的使用时段应该是有所不同的，可以通过聚合来对比。时间粒度为小时，且修改成离散变量，这样就可以将不同日期的相同时段聚合，有助于观察整体的分布情况。
  - c) 骑行路径图：把 track 这一列拆分之后再用 pandas 进行处理画出 track 线路图。由于路径图很多，加入了日期时间的筛选器。路径图展示了骑行的路线分布情况。
  - d) 每日骑行密度图：展示了 8 月 1 日至 31 日的订单起始位置分布情况。用日期（天）分页，可以动态显示密度变化。从密度图中可以很清楚的观察到摩拜单车在上海的主要分布范围，也可以看到摩拜单车分布范围在扩大的趋势。
  - e) 每时骑行密度图：展示了 24 小时的订单起始位置分布情况。用日期（小时）分页，可以动态显示密度变化。从密度图中可以清楚的观察到摩拜单车在不同时段的使用分布情况。
  - f) 骑行时间统计（分钟）：用订单结束时间减去订单开始时间，得到骑行时间。对不同骑行时间计数做数据桶。数据桶很好的展示了骑行时间的分布情况。
  - g) 周末与非周末骑行时间对比：周末和非周末的骑行情况是很有对比价值的。将周末与非周末的骑行时间聚合取平均值，就可以清楚的看到两者的不同。
  - h) 不同时段的平均骑行时间：统计不同时段的平均骑行时间并区分周末与非周末，做线形图。周末与非周末的不同时段骑行时间可能有差异，可以通过聚合来对比。时间粒度为小时，且修改成离散变量，这样就可以将不同日期的相同时段聚合，有助于观察整体的分布情况。
3. 仪表板：
  - a) 骑行订单数及其分布：将骑行订单数的日变化和时变化放在一起，清晰明了地展示

订单数的变化情况。以“每日订单总数”作为筛选器控制其它三个工作表，以“不同时间段的订单数”作为筛选器可以控制下方两个工作表。

- b) 骑行时间：骑行时间的分布情况以及周末/非周末的骑行时间平均值可以很好的反映 8 月份骑行时间的特征。
- 4. 将仪表盘整理成故事（“汇总”），加上注释。

## 反馈:

1. 给朋友讲解 Tableau 故事并获取反馈：
  - a) 希望知道骑行时间相关的特征。修改方案：计算骑行时间，分析骑行时间的分布以及周末与非周末骑行时间的不同。
  - b) 希望知道天气情况是否与订单数有关。不修改理由：查询 2016 年 8 月份天气，下雨天气多是阵雨，无法确定下雨时段。而且，摩拜的日订单是逐渐增加的，会掩盖掉天气的影响。
  - c) 骑行起始位置在地图中以点来展示不够清晰。修改方案：用密度代替点。
2. 审阅老师给出修改建议：
  - a) “不同时间段的订单数”用聚合方式展示更好。修改方案：原来的方案展示的信息比较混乱，所以讲日期修改成“离散”，就能够展现周末与非周末整体的特征。
  - b) “骑行路径图”比较混乱。修改方案：加入筛选器，可以选择某天某时的骑行路线图，从图中可以观察骑行的距离，一般是短途骑行。最终没有把这个工作表放到故事中，因为其展示的信息与数据桶展示的信息相似。
  - c) 密度图的强度不合适。修改方案：降低“每日骑行密度图”的强度，使其能够展示 8 月 1 日至 31 的密度变化。
  - d) “骑行时间统计”的气泡图不能够很好的展示骑行时间的分布情况。修改方案：用数据桶代替气泡图。
  - e) “周末与非周末的骑行时间对比”分日期展示有点多余。修改方案：取周末与非周末骑行时间的平均值做条形图，用均值代替原来的中位数。（ps.采用中位数有防止异常骑行时间干扰）
  - f) 其它反馈，如标题，图例等都做了相应修改。

## 资源:

1. 为了做 track 图，把摩拜单车数据进行了转换，转换的代码在文件夹中（[把摩拜单车数据进行转换.ipynb](#)）