

2016年8月摩拜单车数据分析

6 - 通过数据讲故事

Tableau可视化地址：https://public.tableau.com/profile/muming#!/vizhome/Mobikedraft/20168_1

总结：

摩拜单车于 2016 年 4 月 22 日开始在上海正式推出共享单车服务^[1]，当年8月末的订单量约为8月初的3倍，覆盖地点也明显增加，扩张迅速。

8月份单个用户最骑行次数最多的次数为5次，已经养成了一定的用户习惯。

但是当月79%的单车的使用率只有1次，至少使用过3次的单车不足6%，需要通过在周一、周二、周日进行营销、在重复路径的起止点增加车辆等方法进一步提高单车使用率。

设计：

该摩拜单车数据集包含了上海摩拜单车在2016年8月的102361条订单数据，共10个变量，无缺失值和重复数据，是个比较干净的数据集。

但每个订单的路径坐标都存在同一个单元格内，无法对骑行路径进行可视化，因此将其拆分为每一个路径坐标后，与起、止点的坐标存入同一列^[2]，并增加坐标类型列以说明该坐标为起、止点坐标还是路径坐标。

首先查看摩拜单车用量和使用范围的情况。采用折线图的形式比较直观且能反映订单增长的趋势，可以看出8月末的订单数量约为8月初的3倍。查看8月初与8月末的用车地点情况，为避免星期中不同日子的影响，选取周一作比较，可以看出，摩拜单车主要活跃在江以西，8月末比8月初的覆盖范围明显增大，特别是在原本活跃度很低的江东地区。

考虑到用户在工作日和周末的行为应有不同，制作星期中每天的订单数量柱状图。图中显示，周五、周六的订单数最高，周日的订单数其次，周一、周二订单数最低。结合订单折线图，可以看出周日的订单量不稳定，2次都明显低于周一的量，为避免天气的影响，查看上海市2016年8月的天气后报^[3]，未发现周日订单量低的原因与天气相关，可能是用户在结束一周的工作后，更愿意在紧接着的周五晚上和周六进行出行活动，而周日在家中休息。所以应该考虑与活动方合作，在周日举办更多的活动以吸引用户出行，并佐之以相应的营销来提高单车的周日用量。周一、周二一向为用户的消费不活跃时间，中国的银行一般在周三开展信用卡活动以提高商家的销量，可以考虑与银行合作，像某些国家（如新西兰）一样，与商家（特别是餐饮业）在周一、周二开展活动以提高用户的出行率，从而提高单车的使用率。

为了解每天摩拜单车的使用高峰，制作星期中每天与每小时订单数的热力图：工作日的高峰使用段为7-8时和17-19时，主要用于通勤；周末使用时段比较平均，上午的出行时间晚于工作日，下午的单车使用量相对更高。可以考虑在周末凉爽的早晨增加全民骑行健身活动来提高单车使用率。

为了解单次的骑行时长的分布，制作骑行时长的直方图，数据桶大小为10，图中剔除了单次骑行时长超过120分钟。可以看出单车主要为短途使用，出骑行次数随骑行时间的增大而降低。为了解工作日与周末单词骑行时长的分布，制作骑行时长箱线图，可以看出工作日的骑行时长中位数为18分钟，周末的骑行时长为20分钟，可见用户在周末愿意用单车去比上班/学更远的地方。

为了解每个用户的活跃程度，制作每个用户骑行次数的直方图，以每个用户的骑行次数为颜色梯度。骑行次数在1-10次之间的分布基本呈正态分布，最多的次数为5次，证明已经养成了一定的用户习惯。

为了解摩拜单车的使用率，制作每辆单车的使用次数直方图，以每辆单车的使用次数为颜色梯度。可以看出8月中79%的单车的使用率只有1次，至少使用过3次的单车不足6%，单车的使用率还比较低。

为了确认单车的投入地点，在地图上绘制重复出现的路径，并列出现该路径的骑行时间中位数来查看该路径的长短。可以看出重复出现次数最多的路径有1条，该路径重复出现了14次，骑行时长多为8分钟。重复出现13次的路径有1条，骑行时长多为3分钟。重复出现次数多于5次的路径的骑行时间均不超过10分钟，其中37%的路径为超短途路径，骑行时间不超过37%。

通过百度地图拾取坐标系统^[4]查询至少重复出现了10次的路径的起止点，发现这些地点分布于虹口区的华严路和奎照路、杨浦区的世界路和上海财经大学、普陀区的真如科技创业园和李子园工业园区，在应当在这些重复路径的起止点加大单车的投放量，并向目标人群投放广告。

反馈：

第一份草图地址：<https://public.tableau.com/profile/muming#!/vizhome/Mobikedraft/20168>

第二份草图地址：<https://public.tableau.com/profile/muming#!/vizhome/MobikeAnalysis/20168>

反馈意见及修改结果：

- 修改坐标轴
修改结果：删除了坐标轴多余的部分和部分多余的筛选器，提高data/ink ratio，并对部分坐标轴的说明做了修改，使之更准确地描述坐标轴的含义。
- 图表色系要统一
修改结果：统一以摩拜橙为主色调，使风格统一。
- 图例颜色需要统一
修改结果：把图例颜色的顺序进行了统一，低值为浅色，高值为深色。
- “摩拜在2016-08的扩展迅速，养成了一定的客户习惯，但需要增加单车频率”这三点都属于探索性分析。核心问题可以是：如何增加单车的使用频率，并具体加以解释。如何分时段刺激用户，哪里追加投放，增加多少投放等作为你解释性分析的主要内容
修改结果：核心问题定为如何增加单车的使用频率，并增加了相应的可视化。
- 故事2月首末的地点比较，虽然结论正确，但并不能得出车辆使用数量增加的结论。该结论需要比较月首末的自行车数量，而不是行程重点的数量（订单量）
修改结果：已修改。
- 故事1，周日的订单数并不比周五多，和你的文字说明不符
修改结果：已将描述修改为与图相符。
- 故事4、5下方的图的结论缺失、故事5下方图表标题缺失外、故事3散点图不适合描述趋势，直接使用折线图描述平均骑行时间的趋势，添加趋势线来表现增减、故事3散点图的结论不正确，单次（平均）骑行时间随着时间推移并没有明显增长
修改结果：这3个散点图不适合描述趋势，已删除
- 暂未使用任何可视化交互或动画，可通过仪表盘-操作添加
修改结果：故事3已增加可视化交互
- 文字报告设计部分记录得已经比较完整，但请相应解释，为何这样的可视化是最佳的，而不仅仅是做了什么。例如故事5下方的散点图，增加大小和颜色后显得比较杂乱，这样的设计的目的/结论是什么？故事3上方柱状图，4-14分钟是一个相当的区间，你不能说4-14分钟的比0-4分钟的订单多，因为范围的大小不一；建议使用骑行时间数据桶。同此栏第一点，故事5

的柱状图1-9次的范围也过大

修改结果：已增加相应的解释，骑行时间数据桶大小设为10；骑行次数数据桶大小设为1，因为值比较小，不太适合分组。

- 反馈的内容并不只是修改的方案，例如：为何要修改坐标轴？改的好处是什么，或不修改的原因是什么？

修改结果：已增加改/不改的原因。

资源：

[1] <https://mobike.com/cn/about/>

[2] <https://zhuanlan.zhihu.com/p/28337202>

[3] <http://www.tianqihoubao.com/lishi/shanghai/month/201608.html>

[4] <http://api.map.baidu.com/lbsapi/getpoint/index.html>