

# Tableau 故事

## 总结

我选择创建有关摩拜上海城区用户使用数据的可视化，主要展示上海地区骑行时长分布，骑行时间段分布，以及骑行地点分布。从图中我们可以发现大多数人骑行时长不超过 20min，而且骑行时长为 7min 的最多；工作日比周末用车多，周三骑行数最多，而且在 8 点和 18 点时间用车最多；上海市区内用车比其他地区用车多等规律。

## 作品链接

版本 1:

[https://public.tableau.com/profile/wei.chai#!/vizhome/mobike\\_0/sheet10?publish=yes](https://public.tableau.com/profile/wei.chai#!/vizhome/mobike_0/sheet10?publish=yes)

版本 2:

[https://public.tableau.com/profile/wei.chai#!/vizhome/mobike\\_modify1/sheet12?publish=yes](https://public.tableau.com/profile/wei.chai#!/vizhome/mobike_modify1/sheet12?publish=yes)

版本 3:

[https://public.tableau.com/profile/wei.chai#!/vizhome/mobike\\_modify2/sheet0?publish=yes](https://public.tableau.com/profile/wei.chai#!/vizhome/mobike_modify2/sheet0?publish=yes)

## 设计

我创建的故事主要由三个仪表板组成，分别为骑行时长分布，骑行时间分布，骑行地点分布，下面我将依此为依据来讲述我的设计：

### 1. 骑行时长分布

最初的想法就是想展现各订单的骑行时长分布，由于提供的原始数据中，并没有骑行时长这一参数，所以我就利用 `end_time - start_time`，创建了一个计算字段，并命名为 `cycling_time`，表示骑行时长。一开始并没有注意到这样计算会有问题，直到我将图表给其他人看，给我反馈的结果是骑行时长为 0.2，0.3 分钟的数据很多，我越发觉得这个计算可能有错误，然后我又去观察了一下原始数据，发现就是有问题，然后我就使用 `DATEDIFF` 函数去计算，得到正确的结果。

然后感觉只有这一个图的话，好像并不能体现出说明，而且感觉图标的也不是很符合。于是我将图标题改为“各订单骑行时长”。同时我又增加两个图，分别为“骑行时长分布图”和“骑行时长特征统计”。在一开始画“骑行时长分布图”时发现横坐标太大，于是用将横坐标用对数表示，同时在图中添加了筛选器，用于选择周一到周天各天的骑行时长分布情况。在“骑行时长特征统计”图中用箱线图去表示相关统计特性，同时我在图中，添加了平均值。

在第一次提交之后，根据反馈我将第一个图换成“各时长段内骑行数”，这样比较有意义，可以很清楚地看出各时间段内总的骑行数是多少。在“骑行时长分布图”中我添加了图例，并修改线条的颜色，避免红色和绿色同时出现。

结论：从“各时长段内骑行数”中我们可以看出骑行时长位于 0-20min 的最多。从“骑行时长特征统计”图中可以看出骑行时长呈偏态分布，平均骑行时长为 16.79min。从“骑行

时长分布图”中可以很清晰地看出骑行时长为 7min 左右的人数最多。

## 2. 骑行时间分布

这里我放置了三个图，图标题分别为“周内每天的每小时骑行分布”，“每小时骑行总数”以及“各天的骑行总数”。经别人反馈后，说第一个图和第二个图的标题不太好。于是这里我将图一和图二的标题分别修改为“周内每天的各时段骑行分布”和“各时段骑行总数”。这里的三个图，我想展示的是到底一周内哪天的骑行数比较多，还有每天说明时间段骑行数比较多，以及一周内每天是不是都是有相同的规律。同时，为了更好地显示三个图之间的关联，我添加了突出显示的功能，这样可以帮助别人更好地理解。

结论：从第一个图中可以看出，工作日的用车量要大于周末，工作日每天的骑行时间分布规律几乎一样，周末两天的分布规律也很相似。从第二个图中，可以看出在 8 点与 18 这两个时间点骑行人数最多，这也正是大部分人上下班的时间。最后一幅图可以很清楚地看出周三这一天的用车量要大于周内其他天。

## 3. 骑行地点分布

这里我放置三个图，分别为“骑行出发点分布”，“骑行终点分布”以及“骑行起始-终点路线”。一开始我在“骑行出发点分布”，“骑行终点分布”图中，将各个点的颜色标记为黄色，同时还给各个点使用了红色的边框颜色。别人看过之后感觉不好，于是我将点的颜色调成蓝色，边框设置成无色。同时为了将强对比，我在这里统一将地图的背景修改成黑色背景。在“骑行起始-终点路线”中，其实我一开始的想法是将每个订单的路线图都给画出来，所以我将原始数据进行清理，然后再 tableau 中尝试画轨迹图，虽然每个订单的各个点都可以画出，但是因为顺序的缘故，因为有的轨迹有好几百个点，因此在此放弃了，从而改用起始-终点这样的设计。

在这里，我一开始在图中标注各订单 Origin 与 Destination 的标签，同时还用颜色去区分，别人给我反馈的情况是不应该显示文字，颜色种类使用过多，同种颜色有关联吗。于是我将 Origin 与 Destination 的标签去除，同时也将颜色给去除，改用统一的颜色。

结论：从图中可以看出，用户多数集中在上海市区内。

最后，我将各个仪表板的标题进行修改，并将整个故事的工作簿的主题修改成现代主题，并且添加了海报部分。

# 反馈

## 版本 1:

### 反馈 1:

1. 骑行时长图，从图中看骑行时长为 0.2，0.3 分钟的数据很多，感觉数据有误。
2. 骑行时间分布仪表板中，第一个图和第二个图的标题不太好。
3. 骑行起始-终点路线图中不应该显示文字，还有图中的点不应该有边框，图中颜色不好看。

### 反馈 2:

你在此可视化中注意到了什么？

1. 上下班时间用车最多，周末总体用量大体比工作日更少；

你注意到什么关系？

1. 上下班时间用车最多，周末总体用量大体比工作日更少。

你认为在此可视化中的主要收获是什么？

1. 上下班高峰期用车巨多，可以早起一点或者加会儿班。

反馈 3:

Origin 与 Destination 标注过多，颜色种类使用过多，同种颜色有关联吗？

## 版本 2:

反馈 (第一次提交修改意见)

1. 根据你在故事中的文字叙述，你仅仅提出了问题，并使用了可视化图表，但并没有对可视化做任何的解释。
2. 此部分中，你展示了所有订单的时长，以条形图进行展示。但实际上订单数量过多，而订单号本身没有意义，因此可视化效果并不好，至多只是能够直观看出时长前几位的订单。因此建议只显示时长前几名的订单，或者添加交互式筛选器（使观众能够筛选时长范围），以更好符合条形图的效果；
3. 从你的条形图和箱线图中，可以看出，有少量订单的时长在 500 甚至 1000 分钟以上，这意味着这个订单进行了近 10 个、20 个小时，这显然不符合正常的订单情况。你可以思考这些订单产生的原因（在未关锁的情况下连续被多人骑行、系统故障等），并考虑剔除这些异常值，因为其不具备代表性，且对整体情况的分析会产生负面影响（比如严重拉高均值等）；
4. 你的骑行时长分布图从思路上看没有问题，但你并未对颜色增加图例，使得观众无法了解颜色代表的日期，请在下次提交时添加。同时请注意，不建议同时在图表中添加红色和绿色，这对红绿色盲并不友好；

## 参考资源

[http://onlinehelp.tableau.com/current/pro/desktop/zh-cn/maps\\_howto\\_origin\\_destination.html](http://onlinehelp.tableau.com/current/pro/desktop/zh-cn/maps_howto_origin_destination.html)