

Tableau 故事-泰坦尼克号

一、总结

本项目根据提供的数据集探讨了泰坦尼克号上乘客的有关信息，以及这些信息与船沉没后乘客是否幸存之间的关系。最终得到了，妇女、儿童以及地位较高的乘客在船沉没后会有较大的生存概率。

二、设计

在用 Tableau 可视化探索之前，预览数据发现年龄、客舱以及登船港口的数据缺失，需进行数据清理。根据数据特点，分别用年龄均值、登船港口的众数以及“None”值对相应列的空值进行填充（使用的最终数据为 `titanic_cleaned.csv`）。

第一个故事的链接如下：

https://public.tableau.com/profile/.10707992#!/vizhome/TableauStory_Titanic/sheet21_1?publish=yes

第一页，我想先描述泰坦尼克号整体的一个情况，利用饼图展现该船沉没后存活以及遇难的人数比例，饼图能够更好的展现某个部分或多个部分占用整体的比重；同时我使用了灰色代表遇难信息，绿色代表生还信息，并且这两种特定颜色的选择贯穿了整个故事，体现出一致性。

第二页，我想初步探索并展示乘客的信息，主要是性别，年龄，登船港口以及乘客等级的分布情况。在此我建立了年龄的集，因为原始的离散数据并不能很好的展现存活信息的特点。采用条形图便于进行各个分类的度量值对比，同时使用蓝色和橙色代表女性和男性来体现比重上的对比情况。

第三页，我想展示乘客等级与生存率之间的关系，在此我创建了生存率的变量。、小多组图和条形图来能够更好的展示各个分类的对比情况，同时采用色彩展现分类的差异。能够看出乘客等级越高，生存的概率越大。

第四页，我想展示乘客年龄与生存率之间的关系，首先绘制反应年龄分布的直方图，直方图能够很好地展示连续数据的分布情况，同时利用色彩区分我之前所建立的关于年龄的集。存活人数以及存活率的条形图，能够更好的展示各个分类的对比情况，并且使用了年龄的集进行绘制，能够很好的展现数据信息。能够看到，儿童得到生存的概率最大。

第五页，我想展示乘客性别与生存率之间的关系，同之前一样，我利用条形图及色彩区分来建立生还人数及生还率的对比图。能够看到，相对于男性，女性得到生存的概率会大很多。

第六页，我想展示来自不同登船港口的乘客与生存率之间的差异关系。建立生还人数及生还率的对比图，采用条形图及色彩区分便于进行不同分类之间的对比。能够看到，在 **Cherbourg** 港上船的乘客得到生存的概率较大。同时，我建立了关于乘客等级和登船港口人数的小多组图，小多组图能够很好地对比几组不同的数据，在这里对各个港口登船乘客分成了共三个组（一等、二等和三等）。能够看出，在生还率最高的 **Cherbourg** 港上船的乘客中，有相对于其它两个港口最多的一等乘客数比例和最少的三等乘客比例。说明港口的因素并不会对乘客生还率产生影响，起到影响的还是乘客本身的属性。

三、反馈

1，我把故事发布在 **Tableau** 网站上，收到了反馈：

建议将图中代表乘客等级的阿拉伯数字“1,2,3”以及代表生还与否的阿拉伯数字“0,1”，代表登船港口的“C,Q,S”更改为相对详细的信息，以便有更好的阅读体验。

2，提交项目后，审阅老师对图面也提出了意见：

在故事中，你需要用文字说明从故事中得出的发现和结论（在故事的标题处或使用单独的文本框）；

3，反馈修改：

- 1) 将图中代表乘客等级的阿拉伯数字“1,2,3”更改为“一等，二等，三等”；
- 2) 将图中代表乘客生还与否的阿拉伯数字“0,1”更改为“殉难，生还”；
- 3) 将代表登船港口的“C,Q,S”更改为“Cherbourg, Queenstown, Southampton”；
- 4) 在标题处增加每页故事的发现和结论。

最后的故事链接可以在下面找到：

https://public.tableau.com/profile/.10707992#!/vizhome/TableauStory_Titanic-V2/sheet21?publish=yes