泰坦尼克的故事

故事地址

泰坦尼克的故事v1：<https://public.tableau.com/workbooks/v1_15633608878540.twb>

泰坦尼克的故事final：<https://public.tableau.com/workbooks/8766.twb>

总结

1. 在泰坦尼克乘客撤离过程中，优先让女士以及6岁以下的婴幼儿逃生。66岁以上老人由于人数较少，生还率不具备统计意义；
2. 高级船票的乘客生还率最高，其次是中等，最后是低等船票乘客。由于瑟堡乘客中高等船票占比最高，整体的生还率最高。
3. 各个等级船票乘客中，女性均比男性生还率高；虽然来自皇后镇的低等船票的乘客占比高达90%，但由于皇后镇的女性乘客占比高，整体皇后镇乘客的生还率要高于南安普顿。
4. 7-65岁有家人陪伴的乘客，生存率高于独自乘船的乘客。家人的数量与生存率有一些相关性，但不是很明显。6岁以下（无家人陪同）”和“66岁以上”乘客由于人数较少，生还率不具备统计意义。

设计

故事按照介绍泰坦尼克事件、描述乘客基本情况、分析逃生结果的顺序进行讲述。

* 为了凸现这场悲剧，在介绍事件时展示的生存和死亡人数占比。根据朋友的反馈，使用饼图没有冲击力，无法体现事件的严重性。因此改为使用较大的字体直接显示生存和死亡人数。
* 描述乘客基本情况使用了年龄、性别、经济实力、是否携带家属、出发城市等维度。这些也是该数据集仅有的人口及社会属性。
* 分析逃生结果前期先提出了几个假设：女性幸存率高、孩子和老人幸存率高、社会地位和经济实力对幸存率没有显著影响，并对假设一一进行了可视化。此外对乘客基本信息中的其他维度进行单维度或双维度的生存率交叉分析。

反馈

第一次提交，来自朋友反馈：

1. 首页故事用饼图没有冲击力，无法体现事件的严重性。修改点如下：
   1. 可视化方式由饼图改为了表格；
   2. 字体大小由默认大小（9号）改为因此改为使用较大的15号字体直接显示生存和死亡人数。
2. 图表的颜色比较淡，色弱和色盲可能无法很好识别颜色。修改点如下：
   1. 将各个工作表的饼图和堆积柱状图的调色板，由自动改为了色盲；
   2. 年龄分布的柱状图，由于是单色无法修改调色板，使用取色器将色盲调色板的蓝色用在了柱状图上。
3. 柱状图上有很多多余的线条。将图表信息直接作为标签并隐藏y轴后，界面上的刻度线其实就没有意义了。修改点如下：
   1. 设置线格式，将工作表的轴标尺，和行的网格线设置为无。

第二次提交，来自审阅意见：

1. 既然你呈现的最终结果为一个故事，其他部分已经包含在故事中，那么可以将其隐藏，使得工作簿更加简洁。需要注意，如果只保留故事部分，你需要先在仪表板选项卡中隐藏工作表，再在故事选项卡中隐藏仪表板。对于在故事中没有用到的仪表板或工作表，请作删除处理。修改方式如下：
   1. 隐藏了“泰坦尼克的故事”以外所有的工作表和仪表盘。
2. 你的部分仪表板以及故事的尺寸较小，导致可视化内容比较拥挤，请在故事和各仪表板（二者均需要选择）的大小中选择“自动”以解决此问题（对于故事及所有故事点中的仪表板，都应当选择此选项）。修改方式如下：
   1. 将“泰坦尼克乘客基本情况”、“年龄性别-幸存”、“经济实力对幸存的影响”这三个仪表盘的大小，由“台式机浏览器”改为了“自动”；
   2. 将“泰坦尼克的故事”故事的大小，由“固定大小改为了“自动”。
3. 故事点2展示各变量分布的思路没问题，但部分变量并未用于后续的生还分析中。如果不考虑变量对生还的影响，那么展示本身没有太大意义，无法突出主题，观众对这部分内容也不会感兴趣。请注意变量分布的展示和后续分析的结合，并调整可视化的展示内容。修改方式如下：
   1. 重新设计了故事3、4、5中，现在整个故事已经包含了所有故事2（乘客的基本信息）中展示的所有维度在生还率上的分析。
4. 目前，针对生还率的分析仅仅停留在单变量的因素分析，而未结合至少2个变量综合分析，分析角度较为缺乏。请补充多变量因素分析图表（例如，各舱位分性别的生还率、各年龄段人群在有无携带亲属状况下的生还率比较等），充实故事的内容。修改方式如下：
   1. 故事4中增加了出发城市和船票等级、性别和船票等级、出发城市和性别的交叉分析；
   2. 故事5中增加了年龄段和家人是否同性的交叉分析。
5. 生还分析使用的变量不足。在数据集简单并存在明确分析目的情况下，你至少还需要分析sibsp/parch（兄弟姐妹数量/父母子女数量）等因素，对生还状况的影响，使得可视化和故事整体更充实、完善。请补充此部分。修改方式如下：
   1. 增加了出发城市的分析。结合船票等级分布解释了瑟堡生存率高的原因，又结合年龄分布解释了皇后镇生存率高于南安普顿的原因。
   2. 使用sibsp+parch计算了“同行加人人数”，并分析了同行家人人数对生还率的影响。
6. Pclass变量虽然一定程度可以反映乘客经济实力，但二者显然不等同，直接替换是不合适的。展示变量时应严格遵循其在数据中的本身含义，只有在分析说明时可以适当推论、延展。请修改相应图表。修改方式如下：
   1. 修改“泰坦尼克乘客基本情况”仪表盘中“经济实力分布”图表标题，改为“船票等级分布”。
7. “老人小孩并不优先”不是一个准确的结论，从你的图表中无法得出，原因在于分组存在问题。目前的分组将18岁以下和51岁以上都分为一组，却在中间划分大量组别，不符合数据背景下的年龄特点。举例而言，15岁的青少年和1岁的婴幼儿逃生能力完全不同，52岁的中年人和80岁老人同样存在明显区别，反而是20-50岁人群在体力区别上没有很大，没有必要细分为多组。同时是否能将18岁以下的人群称为“小孩”，51岁以上即为“老人”，也有待商榷。因此在这方面，请重新对年龄进行分组，得出更加合理、准确的结论。修改方式如下：
   1. 根据中国的年龄分段方式（见资源1链接），将乘客年龄划分改为：0-6岁（婴幼儿）、7-17岁（儿童）、18-40岁（青年）、41-65岁（中年）、65岁以上（老年）。
8. 整体生存分析的思路正确，但目前全部使用堆叠式条形图，图表显得较为臃肿。关键信息实际上只有“生还率”一项（死亡率就是1-生还率），因此完全可以通过不堆叠的条形图或者折线图，直接展示生还率，使图表更简洁。修改方式：
   1. 将隐藏的Survived字段显示，增加度量生还率，计算方式=sum([Survived])/sum([记录数])。应用在图表后格式改为百分数，保留两位小数。

第三次提交，来自审阅意见：

1. 除了在故事界面调整尺寸，你还需要在每个故事点对应仪表板的界面调整尺寸（到“自动”），否则可视化效果仍然不好。你的项目中，部分故事点对应的仪表板尺寸未进行相应调整。请注意这点，并检查和修改尺寸。修改方式：
   1. 将“出发城市生还率分析”、“家人生还”这两个仪表盘的大小由“台式机浏览器”改为“自动”。
   2. 确认了故事使用的4个仪表盘，和所有五个故事点大小均已改为“自动”。
2. 故事点3和故事点5中有关年龄分组展示的图表，66岁以上的人群均未列出，也并未说明相关理由。请完整呈现每个分组的内容，使得图表内容和故事结论准确、全面。修改方式：
   1. 故事3的年龄分组取消了隐藏“66岁以上”人群的显示。由于改人群的总人数仅7人，生还率并不具备统计意义，因此还在工具提示上增加了各年龄段的“人数”以做提示。
   2. 故事3取消了隐藏样本量大于等于10的限制。由于“6岁以下（无家人陪同）”和“66岁以上”的样本量不到10，生还率不具备统计意义，因此在工具提示上增加“人数”和“生还人数”，以及将人数应用于散大小以做提示。

资源

1. [少年、青年、中年、老年的年龄段的划分](https://wenku.baidu.com/view/0456a159580216fc710afd3e.html)