

# Tableau 数据可视化报告

数据可视化第一版链接:

[https://public.tableau.com/profile/sz.ydp#!/vizhome/-\\_1183/sheet14](https://public.tableau.com/profile/sz.ydp#!/vizhome/-_1183/sheet14)

数据可视化第二版链接:

[https://public.tableau.com/profile/sz.ydp#!/vizhome/-\\_1186/sheet16](https://public.tableau.com/profile/sz.ydp#!/vizhome/-_1186/sheet16)

数据可视化第三版链接:

[https://public.tableau.com/profile/sz.ydp#!/vizhome/-\\_1187/sheet22](https://public.tableau.com/profile/sz.ydp#!/vizhome/-_1187/sheet22)

数据可视化第四版链接:

[https://public.tableau.com/profile/sz.ydp#!/vizhome/-\\_1190/sheet20](https://public.tableau.com/profile/sz.ydp#!/vizhome/-_1190/sheet20)

## 总结

在这份可视化数据中,主要可视化泰坦尼克号的生还与哪些因素有关。主要从性别、年龄、舱别这几个因素进行可视化数据,最后综合可视化性别、年龄、舱别对生存率的影响。

通过可视化数据得出,性别、年龄、舱别对生存率均有影响。其中女性乘客生还率高于男性乘客生还率。年龄对生还率的影响是,年幼的生存率高于年长的生存率。舱别对生存率的影响,1 舱别生还率高于 2 舱别,2 舱别生还率高于 3 舱别生还率。其中,对生还率影响最大的是性别(女性生还率 74.2%),其次是舱别(1 舱别生还率 62.96%),最后是年龄(0-10 岁生还率 59.38%)。

## 设计

首先,拿到原始数据,发现年龄舱别数据有缺失。使用 jupyter notebook 对其进行清洗。然后使用清洗后 titanic-data-clean.csv 的数据进行可视化。

在这份可视化报告中,主要采取蓝黄配色,保持一致颜色的编码,以便更容易读取图表。

使用了 6 个故事点讲述泰坦尼克号生还与性别、年龄、舱别的关系。首先对性别做总体分析,分析所有乘客的性别比例分布情况,采用的是条形图,性别以不同颜色进行区分;其次,分生存与遇难进行性别比例的分析,采用的是条形图;最终得出相关的可视化数据结果。

首先对年龄设置以 10 为单位的年龄组,得到 age group 项,同样方式,对年龄组、舱别与生还率的关系进行一一可视化分析,主要采用的是条形图分析。

然后进行生还与年龄、性别两者之间的关系可视化,采用是方形图,并且可以对不同年龄进行筛选。这样,使得生存与遇难之间不同性别,不同年龄组之间的比例更加直观的显示。

生还率与舱别、性别之间的可视化，使用的是条形图，性别用不同颜色区分，便于区分不同性别的数据。

最后，综合可视化了性别、年龄、舱别与生还率之间的关系，采用了气泡图，颜色区分性别，气泡大小表示不同的比例。

根据第一版的反馈进行相应的更改：

问题 1：横坐标的 0 10 20 30 40 是分别指哪个年龄段，建议应该换成具体年龄段分析；

更改：对原始数据进行年龄分组列的增加，按照 10 岁对年龄 0-90 岁进行分组。

问题 2：还可以增加舱别、年龄对生还率影响的可视化

更改：增加了加舱别、年龄对生还率影响的可视化，采用了条形图，并增加了筛选器。

问题 3：建议将生存者数据体现在显眼的地方，易于读取。

更改：对 `survived` 进行排序，将生存者数据放在显眼处。

问题 4：对于最后一个展示性别、年龄、舱别对生还率的影响的图不是很明白。

更改：采用气泡图，将性别用颜色和标签表示，年龄分组和舱别使用标签显示，这样能够一目了然的看懂可视化数据。

根据第二版的反馈进行相应的更改：

问题：占比并不能反映各个因素的对生还率的影响

更改：增加了性别、年龄、舱位中的生还率计算的可视化，同时增加两两因素对生存率的可视化图形。

## 反馈

### 第一版：

反馈：

1、你在此可视化中注意到了什么？

生还率与年龄关系：生存者中年龄主要分布在 0-40 岁，40 岁区间也占了 10%左右

2、你对此数据有何问题？

横坐标的 0 10 20 30 40 是分别指哪个年龄段，建议应该换成具体年龄段分析；

还可以增加舱别、年龄对生还率影响的可视化；

建议将生存者数据体现在显眼的地方，易于读取。

3、你注意到什么关系？

发现泰坦尼克中的生还率与性别、年龄、舱别有一定的关系。

4、你认为在此可视化中的主要收获是什么？

主要收获是了解了泰坦尼克号的生还率与性别的关系很大。

5、此图形中是否有你不理解的部分？

对于最后一个展示性别、年龄、舱别对生还率的影响的图不是很明白。

## 第二版:

反馈:

1、你在此可视化中注意到了什么?

生还率与年龄: 怎么得出年轻的生还率高, 老的低。是通过对比生还人数中年轻年老的人数占比吗? 可能年老人数本来就少, 所以占比也少, 并不能说年轻生还率高。注意到女性生还率占比都比较高, 可能是因为 女士小孩优先, 可以再看看是不是小孩的生还率也比较高。

2、你对此数据有何问题?

比较生还率高低时, 应该这样比较: 比较女性与男性生还率, 女性总共 100、男性总共 200, 女性生还 70 男性生还 70; 女性生还率应该是  $70/100 = 70\%$  男性应该是  $70/200 = 35\%$ 。主要是分析生还率时, 需要看此类因素 本来的人数有多少, 可能人数多, 生还的也多, 生还占总生还的比率就大, 但不能说明他的生还率高。

## 第三版:

反馈:

这次的生还率清晰明了, 改进了很多之前模糊的地方。能够看出性别、年龄、舱别与生还率的关系。

## 资源

1、数据来源: [优达数据集](#)

2、参考资料来源: [泰坦尼克号相应数据分析](#)

3、tableau 教程: [tableau 教程](#)