

Ali's Titanic Dataset Analysis

Ali Alkhalaf

1/28/2022

This is an R HTML document. This page shows few analysis of the Taitanic dataset. For any question or comment, please contact me on ali2alkhalaf@gmail.com

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(readr)
```

```
##1. Read a .csv file containing data elements about Titanic travelers
```

```
Titanicsurvival <- read_csv("titanic.csv")
```

```
## Rows: 891 Columns: 12
```

```
## -- Column specification -----
## Delimiter: ","
## chr (5): Name, Sex, Ticket, Cabin, Embarked
## dbl (7): PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
Titanicsurvival <- as_tibble(Titanicsurvival)
Titanicsurvival
```

```
## # A tibble: 891 x 12
##   PassengerId Survived Pclass Name    Sex    Age SibSp Parch Ticket   Fare Cabin
##         <dbl>   <dbl> <dbl> <chr>  <chr> <dbl> <dbl> <dbl> <chr>   <dbl> <chr>
## 1             1       0     3 Braun~ male    22     1     0 A/5 2~   7.25 <NA>
## 2             2       1     1 Cumin~ fema~   38     1     0 PC 17~  71.3  C85
## 3             3       1     3 Heikk~ fema~   26     0     0 STON/~   7.92 <NA>
## 4             4       1     1 Futre~ fema~   35     1     0 113803  53.1  C123
## 5             5       0     3 Allen~ male    35     0     0 373450   8.05 <NA>
## 6             6       0     3 Moran~ male    NA     0     0 330877   8.46 <NA>
## 7             7       0     1 McCar~ male    54     0     0 17463   51.9  E46
## 8             8       0     3 Palss~ male     2     3     1 349909  21.1  <NA>
## 9             9       1     3 Johns~ fema~   27     0     2 347742  11.1  <NA>
## 10            10       1     2 Nasse~ fema~   14     1     0 237736  30.1  <NA>
## # ... with 881 more rows, and 1 more variable: Embarked <chr>
```

##2. Calculate the total number of passengers in the dataset.

```
TitanicPassengers <- nrow(Titanicsurvival)
TitanicPassengers
```

```
## [1] 891
```

```
Firstclass<- nrow(filter(Titanicsurvival, Pclass == 1))
Firstclass
```

```
## [1] 216
```

```
Secondclass <- nrow(filter (Titanicsurvival, Pclass == 2))
Secondclass
```

```
## [1] 184
```

```
Thirdclass <- nrow(filter(Titanicsurvival, Pclass == 3))
Thirdclass
```

```
## [1] 491
```

```
Male<-nrow(filter(Titanicsurvival, Sex == "male"))
Male
```

```
## [1] 577
```

```
Female<-nrow(filter(Titanicsurvival, Sex == "female"))
Female
```

```
## [1] 314
```

##3. Calculate the total proportion of passengers surviving.

```
TitanicSurvived <- filter(Titanicsurvival, Survived == 1)
TitanicSurvived1<- nrow(TitanicSurvived)
TitanicSurvived1
```

```
## [1] 342
```

```
TitanicSurvived2<-((TitanicSurvived1/TitanicPassengers)*100)
TitanicSurvived2
```

```
## [1] 38.38384
```

##4. Calculate the proportion of passengers surviving for each class of passenger.

###1st Class

```
TitanicSurvivor_C1 <- filter(Titanicsurvival, Pclass == 1,Survived==1 )
TitanicSurvivor_C1
```

```
## # A tibble: 136 x 12
```

```
##   PassengerId Survived Pclass Name    Sex    Age SibSp Parch Ticket  Fare Cabin
##   <dbl>      <dbl>  <dbl> <chr>  <chr> <dbl> <dbl> <dbl> <chr>  <dbl> <chr>
## 1         2         1      1 Cumin~ fema~  38     1     0 PC 17~  71.3 C85
## 2         4         1      1 Futre~ fema~  35     1     0 113803  53.1 C123
## 3        12         1      1 Bonne~ fema~  58     0     0 113783  26.6 C103
## 4        24         1      1 Slope~ male   28     0     0 113788  35.5 A6
## 5        32         1      1 Spenc~ fema~  NA     1     0 PC 17~ 147. B78
## 6        53         1      1 Harpe~ fema~  49     1     0 PC 17~  76.7 D33
## 7        56         1      1 Wooln~ male   NA     0     0 19947   35.5 C52
## 8        62         1      1 Icard~ fema~  38     0     0 113572   80 B28
## 9        89         1      1 Fortu~ fema~  23     3     2 19950  263 C23 ~
## 10       98         1      1 Green~ male   23     0     1 PC 17~  63.4 D10 ~
## # ... with 126 more rows, and 1 more variable: Embarked <chr>
```

```
FirstClassSurvivor<-nrow(TitanicSurvivor_C1)
FirstClassSurvivor
```

```
## [1] 136
```

```
PropFirstClassSurvivor<-((FirstClassSurvivor/Firstclass)*100)
PropFirstClassSurvivor
```

```
## [1] 62.96296
```

###2nd Class

```
TitanicSurvivor_C2 <- filter(Titanicsurvival, Pclass == 2,Survived==1 )
TitanicSurvivor_C2
```

```
## # A tibble: 87 x 12
```

```
##   PassengerId Survived Pclass Name    Sex    Age SibSp Parch Ticket  Fare Cabin
##   <dbl>      <dbl>  <dbl> <chr>  <chr> <dbl> <dbl> <dbl> <chr>  <dbl> <chr>
## 1         10         1      2 Nasse~ fema~  14     1     0 237736  30.1 <NA>
```

```
## 2      16      1      2 Hewle~ fema~ 55      0      0 248706 16 <NA>
## 3      18      1      2 Willi~ male  NA      0      0 244373 13 <NA>
## 4      22      1      2 Beesl~ male  34      0      0 248698 13 D56
## 5      44      1      2 Laroc~ fema~  3      1      2 SC/Pa~ 41.6 <NA>
## 6      54      1      2 Faunt~ fema~ 29      1      0 2926   26 <NA>
## 7      57      1      2 Rugg,~ fema~ 21      0      0 C.A. ~ 10.5 <NA>
## 8      59      1      2 West,~ fema~  5      1      2 C.A. ~ 27.8 <NA>
## 9      67      1      2 Nye, ~ fema~ 29      0      0 C.A. ~ 10.5 F33
## 10     79      1      2 Cald~ male  0.83    0      2 248738 29 <NA>
## # ... with 77 more rows, and 1 more variable: Embarked <chr>
```

```
SecondClassSurvivor<-nrow(TitanicSurvivor_C2)
SecondClassSurvivor
```

```
## [1] 87
```

```
PropSecondClassSurvivor<-((SecondClassSurvivor/Secondclass)*100)
PropSecondClassSurvivor
```

```
## [1] 47.28261
```

```
###3rd Class
```

```
TitanicSurvivor_C3 <- filter(Titanicsurvival, Pclass == 3, Survived==1 )
TitanicSurvivor_C3
```

```
## # A tibble: 119 x 12
##   PassengerId Survived Pclass Name    Sex    Age SibSp Parch Ticket  Fare Cabin
##         <dbl>   <dbl> <dbl> <chr>  <chr> <dbl> <dbl> <dbl> <chr>  <dbl> <chr>
## 1           3       1      3 "Heik~ fema~  26     0     0 STON/~  7.92 <NA>
## 2           9       1      3 "John~ fema~  27     0     2 347742 11.1 <NA>
## 3          11       1      3 "Sand~ fema~   4     1     1 PP 95~ 16.7 G6
## 4          20       1      3 "Mass~ fema~  NA     0     0 2649   7.22 <NA>
## 5          23       1      3 "McGo~ fema~  15     0     0 330923 8.03 <NA>
## 6          26       1      3 "Aspl~ fema~  38     1     5 347077 31.4 <NA>
## 7          29       1      3 "O'Dw~ fema~  NA     0     0 330959 7.88 <NA>
## 8          33       1      3 "Glyn~ fema~  NA     0     0 335677 7.75 <NA>
## 9          37       1      3 "Mame~ male   NA     0     0 2677   7.23 <NA>
## 10         40       1      3 "Nico~ fema~  14     1     0 2651  11.2 <NA>
## # ... with 109 more rows, and 1 more variable: Embarked <chr>
```

```
ThirdClassSurvivor<-nrow(TitanicSurvivor_C3)
ThirdClassSurvivor
```

```
## [1] 119
```

```
PropThirdClassSurvivor<-((ThirdClassSurvivor/Thirdclass)*100)
PropThirdClassSurvivor
```

```
## [1] 24.23625
```

```
##5 Plot of comparsion the number of survivors among the three passengers' classes
```

```
number_survivor <- data.frame(Firstclass, Secondclass, Thirdclass)
number_survivor
```

```
##   Firstclass Secondclass Thirdclass
## 1         216         184         491
```

```
barplot(as.matrix(number_survivor), main = "Number of Survivors Comparison", col = rainbow(20), ylim = c(0, 500))
```

