

Exercise Problems

We will use the `titanic.csv` file to answer the following using Rattle:

1. Read the `titanic.csv` in rattle.
 - a. Rename the variable “Embarked” to “Onboarded”.
 - b. Rename the variable “Sex” to “Gender”
2. Typecast Survived column to categorical variable. Perform cross-tabulation of categorical variables with Survived column.
3. Create a Bar plot of Survived with Gender.
4. After having solved question 1 to 3, use the code generated in rattle – log files and clean-up the code for further use.

Using `r_titanic_exploration.RMD`, `titanic.csv` and `dplyr` package and `magrittr` package (optional):

5. Create a new variable named ‘`pclass_gender_age_df`’ and calculate the average age of passengers based on Pclass and Gender. Use `group_by()` and `summarise()` from `dplyr` package.
6. Create a new variable named ‘`pclass_gender_sur_df`’ and calculate how many passengers survived in each of the pclass and for each of the Gender type. Use `group_by()` and `summarise()` from `dplyr` package.
7. Merge `pclass_gender_age_df` with `pclass_gender_sur_df` based on pclass and Gender. Use `inner_join()` from `dplyr` package.

You may write the above code in Groupby and Aggregate section of `r_titanic_exploration.RMD` file.