

For this assignment, first we need to import the required order to proceed with our analysis. After that, we have to read our CSV file and do a quick glance that everything is right. It is and we continue.

Next step, we check the data types, information and check the summary statistics table in order to insure there is no existing incorrect data (missing data will not be checked here) i.e. negative data. We continue.

Next step, we check for the missing data. As we can see, there are a lot of missing data in some columns and we have to clean it according to our problem (assignment). I chose to delete rows of data regarding our Age column ONLY for this assignment (if the question was different maybe additional cleaning is needed).

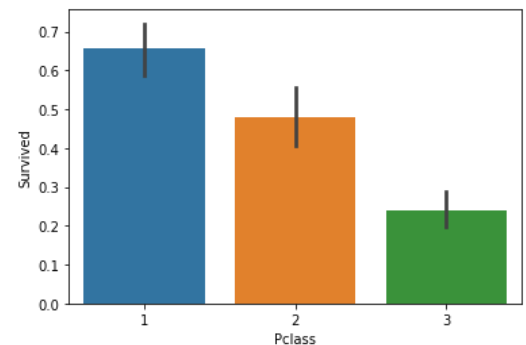
First, I chose deleting over mean imputing because it will create bias in our analysis. Finally, I chose delete of age because any other deleting will reduce our data significantly and this is loss of information. We can do our analysis for this assignment now.

Three hypotheses are in question here. I answer each one with a table, which shows the survival rate depending on the classification asked about in the question. Followed by a bar plot in each one to visualize also.

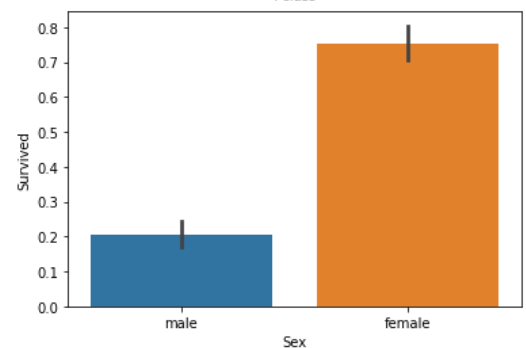
- Survival rate associated with class of passenger
- Survival rate associated with gender of passenger
- Survival rate associated with age of passenger

In all three we can safely say that the tables and bar plots indicate an association between survival rate and the respective variable of the passenger.

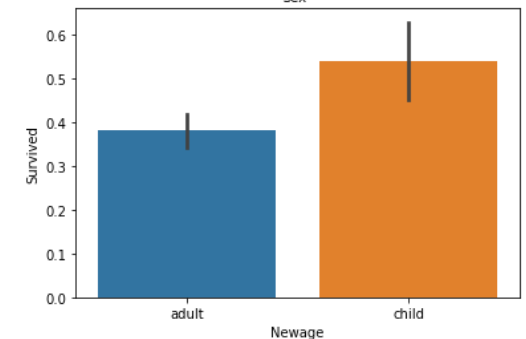
Pclass	Mean(Survival Rate)
1	0.655914
2	0.479769
3	0.239437



Sex	Mean(Survival Rate)
female	0.754789
male	0.205298



Newage	Mean(Survival Rate)
adult	0.381032
child	0.539823



link for GitHub repo:

https://github.com/Darariahi/CPSC4800-Assignment-3_2