

Download the “train_titanic.csv” from the channel and apply the following steps. The submission file should be downloaded as notebook (.ipynb)

Basic (7 out of 9 is a full credit. You can answer more for bonus):

1. Create a directory in Desktop named “python_files”, then a subdirectory as “week5”. Then put the “train_titanic.csv” there. [You can do all this not using python or linux]. Then change path to read the file from the proper directory.
2. Show info(), and describe the data frame. Write as output text your 2-line interpretation of the loaded file.
3. Drop the row with index 1
4. Drop the column called “Embarked”
5. Drop rows with null values
6. Plot fare histogram
7. Plot survived histogram
8. Draw correlation matrix
9. Interpret your results (the way you think might be interesting to explore in more details) and print your text as output

Advanced:

1. Create a new column called “doubledFare” and double the fares. Append the new column to your dataframe
2. Fill null age values by mean(age)
3. Scatter plot age to fare and interpret
4. Present using groupby the count, mean, and standard deviation of survived column
5. Find top 3 correlated attributes to survive