ML Challenge – Intuit

This is a challenge to test your understanding of how you will approach and solve a Machine Learning problem.

The data set provided to you is a data set of passengers who travelled in a ship. The ship met with an accident and few of the passengers survived.
Data set is present in: ship_data.csv

Now, here is what you need to do:
1) Count the number of passengers on the ship based on Gender.
2) Count the number of passengers on the ship based on Class.
3) Count the number of passengers on the ship who are children (Age less than 16 years).
4) Find the number of children in the various classes.
Bonus: Can you plot the above results ?
5) Replace all NULL and NAN values in the dataset programatically.
Note: You may replace the missing values in numeric feature by mean/median and in categorical features by a suitable dummy value.
6) Find the number of passengers who were alone and who were with family.
7) What percentage of passengers survived in each Class?
8) Find the percentage of passengers who survived, based on Gender.
9) Did having a family member increased the chances of survival?
Hint : Find the percentage of passengers who survived and were with Family members.
10) Fit a classification model which can predict whether a person survived or not. Try to maximize the accuracy.
Hint: Possible features can be: Class, Gender, Age, Siblings Count, Parents Count, Fare, Embarked.

Rules:
1. There will be 3 winners.
2. The evaluation criteria is AUC for the model. The higher the AUC, the better is the model and the students are ranked accordingly.
3. Code submission should be in the form of jupyter notebook and the code should run with no error.
4. Last date for submission is Monday, 5th Feb, midnight 11:59 PM