For reviewer1:
First of all, the features shown in the report are only part of the features we use. We have clearly explained the reason we choose them, and we believe that using useful variables should be a priority rather than using the more variables the better.

Secondly, the missing values of most variables we deleted are more than half of the data. In this case, artificially replacing missing values will make it hard to have high predictive accuracy. In addition, since the sample is large enough, deleting variables containing missing values is a reasonable strategy.

Finally, we have compared the Lasso model with the Ridge model, and too many models can make the report jumbled. And since this is a warm-up project, we only use the model taught by the professor in class for fitting. We have read the reference book and had a deeper understanding of regularization. As we state in the report, the result of logistic models we used to fit the data may not be as good as other models like Random Forest, AdaBoost, Gradient Boosting, and so on. We will try more technical models in the next project.

For reviewer2:
Thanks for the recognition.

For reviewer3:
Few grammatical errors should not be a reason for 3 points. It makes no sense. The structure and other aspects of our report are very good.

Furthermore, we have compared the Lasso and Ridge methods and have shown the results of fitted coefficients. Lasso is proved to be a variable selector.

Besides, I think the reason why different models get the same results has been explained by the professor in class. So, we do not repeat it too much.