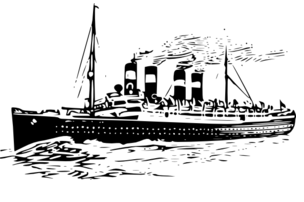
# Lab 2 Regression

Please submit your answers in this word document, including detailed discussion for each question with supportive screenshots of your analysis in python. Be sure your answers are numbered to correspond to the directions so that your responses are clear.

## Linear Regression

Welcome to CastAway Cruise Lines ™!

CastAway Cruise Lines™ is a short to midterm cruise operator based out of the port of Miami. We maintain a fleet of 3 elegant vessels that each have approximately 750 cabins per ship. Each ship has 30 suites (S), 120 Balcony (B) cabins, 150 cabins equipped with a window (W) and 450 Interior (I) cabins. In addition to room type, we track total cost, room costs, ship board expenses, casino expenses, and excursion expenses for all of our customers. Our primary concern is maintaining our customer satisfaction amongst our patrons. For all of analysis, make sure that Customer Satisfaction Average is ‘Target’ while Overall\_1\_5\_Satisfaction is ‘Rejected’.

As our newest (and only) data analyst you are tasked with the following projects:

1. Is there a relationship between how much a patron’s entire trip costs and their satisfaction? Provide the visual and numerical values that you referenced and justify your conclusion. (Hint: Calculate the correlation between the variables).
2. Build the best possible model you can, without any categorical variables, to predict customer satisfaction (Make sure the role of Customer Satisfaction is your dependent variable). Show the screenshot of the output of the parameter estimates and significance. How predictive is this model? Is there any potential for multicollinearity? How did you come to this conclusion?
3. Does the choice of the type of room have significant impact on the customer satisfaction if it is the only predictor? If so which room type(s) is different from the reference type (cabins with a window)? How did you come to this conclusion using the software output?
4. Considering all predictor candidates including the categorical variable(s), create the best model you can and show the screenshot of the output of the parameter estimates and significance. Also, without real implementations, please discuss a general idea of how to improve this model by creating or adding other categorical variables.
5. How can you use this model to increase customer satisfaction? Give specific examples and tie it back to the individual coefficients.

## Logistic Regression

1. Using the same dataset as you did for the Linear Regression portion of the assignment, please estimate a logistic regression to predict whether a customer booked a suite (Room Type = S) as opposed to any other type of room. (HINT: You need to create a new dependent variable which =1 if Room Type = S and 0 otherwise) Select 2 independent variables which you believe to be most important and explain why you chose those variables in your model.
2. Add 2 additional variables. How did your misclassification rate change? Which model do you believe is better and why? Please be sure your results screenshot includes anything you reference in your answer.