1. Prepare the Environment (do not copy and paste...manually type so that you remember what you need and why you are importing the things that you are. You will do this often!)
   1. Read the information into a dataframe.
   2. How many rows are in your dataframe?
   3. Print to the notebook the last 10 rows of your dataframe.
   4. What is the inner quartile range of tenure?
   5. How many missing values are in each variable?
   6. Fill any missing values with 0.
   7. Split your data into a test and train dataset. Set the random\_state. You should end up with 4 dataframes: X\_train, X\_test, y\_train, y\_test.
2. Using the in-sample data (X\_train and y\_train)
   1. Create a scatterplot for each combination of variables.
   2. Create a heatmap containing the 3 variables.
   3. Compute Pearson’s correlation coefficient and print it's value in the sentence "Pearson's R is \_\_\_\_ with a significance p-value of \_\_\_\_"
   4. Train (aka *fit*) a linear regression model, modeling total charges as a linear function.
   5. What is the y-intercept of the regression line? Translate the intercept in plain english, i.e. what is means in the context of the data.
   6. What the slope of the regression line? Translate the slope in plain english, i.e. what is means in the context of the data.
   7. Write the linear function in the form of y=mx+by=mx+b using the parameters that were estimated from the algorithm and the variable names for y and x specific to your data.
   8. Test your model on the training sample.
   9. Evaluate the model's performance using r-squared, mean squared error, and median absolute error.
3. Using the out-of-sample data (X\_test, y\_test)
   1. Test your model on the test sample.
   2. Evaluate your model's performance on test sample. How do each of the metrics compare to that of the training sample? How can you explain the reason for the difference?
   3. Calculate the p-value of the regressor for the train and test sample. Hint: f\_regression
   4. Create a scatter plot of the predicted values versus the actual values in the test sample.
   5. Create a scatter plot of the actual values of x and y in the test sample with the regression line layered over.
   6. Create a scatter plot of the residuals (x axis: predicted values, y axis: the residual values (remember definition of residual from previous lesson). Layer over it a horizonal line that represents no residual.