ASSIGNMENT 1

THE FTIC.csv file contains data for 9218 first-time college freshman. The goal of this assignment is to explore the effect of increased admissions standards on enrollment and retention.

1. Import the FTIC.csv file as a pandas data frame
2. Find the dimension of the data frame, and print the first few rows of the data frame to begin inspecting the data.
3. The column "2^{nd}_FALL" shows whether each student was retained until their second fall semester. Store that column of data in a vector called retention. Check the length of retention, view the first few elements of the vector, and table the vector.
4. Store the PERCENTILE and SAT columns in vectors called rank and sat, create histograms, summaries, and find the mean of these vectors.
5. Replace the 'Y'/'N' values of retention with 1's and 0's, and find the total number of students retained, and the retention rate.
6. Basic statistics:
   (a) Plot rank vs sat
   (b) Create a linear regression model for predicting sat using rank, and summarize that model
   (c) Plot retention vs rank
7. Determining how to predict retention accurately will allow a school to admit students with the highest chance of being retained, and therefore increase retention rate.
   (a) Create a logistic regression model for predicting retention using rank.
   (b) Find the p-value for rank in the model above.
   (c) Create a logistic regression model for predicting retention using sat.
   (d) Find the p-value for sat in the model above.
   (e) Create a logistic regression model for predicting retention using rank and sat.
   (f) Find the p-value for sat and rank in the model above.
8.