

Capstone 2: Statistical Analysis

A different form of this dataset was analyzed in the first project, using resampling techniques, such as bootstrapping and permutations. A parallel analysis was performed on each of the two modified datasets in the second project. Hypothesis tests were performed to determine if the mean ages and age distributions were each identical between the genders. It was found that they were not. The Pearson correlation coefficient, between gender and age, was -0.22, as was shown in the [Statistical Analysis](#) notebooks from the second project. The same analysis was done to the revenue distributions with respect to gender. It was shown that these distributions were distinct, as well. The Pearson coefficient was much weaker, at -0.02. The results of all three analyses were nearly identical, regardless of the form the target variable took. As such, the resampling processes used in this analysis was thoroughly explained in the detailed [PDF](#) from the first project.