Appointment No-Shows - Inferential Statistics

Dovid Burns

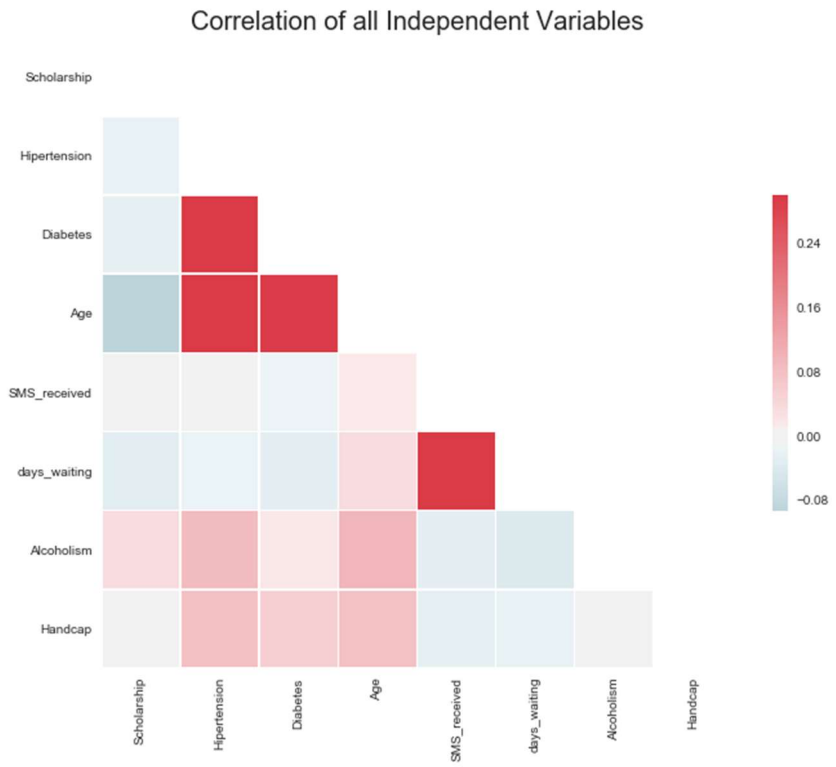
2/5/18

Springboard Capstone One

Examination of the data story telling showed that the leading indicators of missing an appointment are appointment day, patient handicap, hypertension, diabetes, scholarship receipt, text message reminder, age binned, time between appointment scheduled to appointment day, and neighborhood. Our initial survey indicated that there did not seem to be a correlation between missing an appointment and alcoholism and gender. To further verify these claims, we ran t-tests for independence and chi-squared tests between these variables. To check for correlation between the independent variables above, we created a correlation heat map of all the numerical variables.

Using a chi-squared contingency test built into python, we found the following p-values for correlation with no-show rate: appointment day = 1.8e-5, handicap = 0.11, hypertension = 2.4e-32, diabetes = 4.9e-7, scholarship = 3.2e-22, received a text reminder =0.0, age binned = 6e-174, time between appointment scheduled to appointment day binned = 0.0 and neighborhood = 1.3e-60. Given the previous analysis of mean values, the p-values that we expected to refute any correlation was p-value for gender = 0.19 and alcoholism = 0.98. Holding these test to a 99% accuracy, we would only accept variables as statistically significant in predicting no-shows with a p-value of less than 0.01. Therefore, appointment day, hypertension, diabetes, scholarship, received a text reminder, age binned, time between appointment scheduled to appointment day binned, and neighborhood are all good predictors of a patient not showing up for an appointment. The difference in no-show rates between patients with and without these features are as follows: appointment day= up

to 1.9%, hypertension= 3.6%, diabetes= 2.4%, scholarship= 3.9%, received a text reminder 10.9%, age binned= up to 9.8%, time between appointment scheduled to appointment day binned= up to 27.8%, and neighborhood = up to 14%. The strongest indicators based on smallest p-value and largest difference in no-show rates are received a text, age binned, time between appointments scheduled to appointment day binned and neighborhood.



The heat map above depicts some correlation between several independent variables. We see a positive correlation between diabetes and hypertension. Likewise, we also note a positive correlation between age and hypertension, age and diabetes and a slight negative correlation between age and scholarship. Lastly, the time that elapses between scheduling and the appointment date is correlated with receiving a reminder text. As the highest correlation found between these variables is only 0.24, we can confidently use all of the important variables without a correlation bias.