

# TD Hospital

# Modifications Made to the Data

## Kept Features:

- extraprimary\_Cancer
- primary\_Lung Cancer
- psych6
- primary\_MOSF w/Malig
- psych2
- information
- cancer\_no
- disability\_no(M2 and SIP pres)
- dnr\_no dnr
- timeknown

# Explanation

The columns which were deemed irrelevant to death logically, were removed first. Some of these features were race, income, cost, etc.

# Explanation

Using pandas, we found correlation values between -1 and 1 for each feature. Using a particular threshold, we can remove specific columns based on their correlations. For a higher threshold, more features get removed. Selecting a threshold of 0.05, we kept all of the other features, and got an accuracy of 0.85.