David Kuralt

Description of this project

This project is the final assignment that I completed for COMP 4441, one of my classes at the University of Denver, where I am working on an MS in Data Science. It is a survival analysis I conducted on two different hard drive models maintained by BackBlaze, which is a cloud storage company that serves individual and corporate clients. Their Website is https://www.backblaze.com/, and I downloaded zip files of the data they maintain on all their hard drives for the fourth quarter of 2019 and the first three quarters of 2020. For each day of the quarter, they maintain an Excel csv file with data on each hard drive.

I wrote two Python programs to scan each file in the zip folder. The first program looks for the first record of each hard drive of the two chosen models, and the second program looks for the last record. This way, we can identify the day a hard drive "died," or determine if the hard drives survived the entire time period from Q4 of 2019 through Q3 of 2020.

The two text files contain the data I obtained on each of hard drives, and I load these files into the Rmd file where I conduct the survival analysis.