Assignment 2

TPK4186

Torstein Heltne Hovde
Lars Magnus Johnsen
Simen Eger Heggelund

Intro:
In this assignment we are exploring reliability data for 10 different components. The data have been recorded in 123 different Excel spreadsheets. The program records data from all the spreadsheets to a database. It can also print all the data to one single Excel spreadsheet, in other words: combining all the data into one Excel spreadsheet.
We calculate the Kaplan-Meier estimator for the 10 different components, this can be exported in .png and .pdf files.
Finally the program generates a HTML page for reporting the Kaplan-Meier estimator for each component.

How to run:
With all files and folders in the same path as delivered in the .zip file, all that is necessary to run the program is to write in the terminal “pip install -r requirements.txt” to install the required packages, then run the **tester.py**. This will print out all the processed lists in the terminal aswell as displaying the survival analysis plots for the engineer in turn, to review. After these are crossed out(reviewed), the HTML page is generated. This displaying of the plots was a design preference from our side to gave the user a chance to individually analyze the results before it gets produced to a finalized report.

Task 1:
The function is called “listFilesInFolder”. It returns a list with all the files in the contained folder.

Task 2:
The function “extractExcel” takes an Excel workbook as input. The function returns a dictionary with row number as key, and the corresponding column data as values.

Task 3:
We have implemented a function called “dateFromString”. When given a string with date information recorded from Excel, the function returns a date object.

To calculate the difference between to dates in hours the function “diffDate” is implemented. This function takes two date objects as inputs, and returns the total of hours between them.

Task 4:
All units are loaded from the 123 different Excel workbooks with the function “createDatabase”. The database is implemented as a dictionary with the different

components names as keys, and all the corresponding unit objects as values. The data stored in a unit object is “code”, “description”, “inDate”, “outDate” and “failureDate”.

Task 5:
The function “printToExcel” prints the whole database to one Excel spreadsheet. The spreadsheet has 12453 rows and is sorted so all the data regarding the same type of unit follow each other.

Task 6:
The class “KaplanMeierEstimator” generates the Kaplan-Meier estimates with the function “generateKME”. We create a list of tuples to show them in the same way as the assignment set. E.g the Kaplan-Maier estimate for “Vibration sensor” looks like (0,1296),(2,1293),(3,1291),(4,1288),(5,1286).....(6554,785),(6570,784).

Task 7:
The “Calculator” class is implemented with methods to extract Kaplan-Meier estimates. Most of the functionality lies in the KaplanMeierEstimator class for simplicity sake, but the preparePlotValues() function lies in the Calculator class which takes in all relevant data, and by scaling, creates interpretable values ready for the matplotlib library to display the survival analysis.

Task 8:
In the Calculator class it was added the function exportKMEtofile() that has much of the same functionality as plotKME() but also saves the plots in .png and .pdf format for further use to generate a report.

Task 9:
The ReportGenerator Class is designed to be able to automatically generate a HTML report from the data in our Database class. It utilizes data processed in the KaplanMeierEstimator class and the Calculator class to present Survival Analysis plots for each consecutive component as a readible, feasible HTML report.