HW02 — STAT/CS 287  
NAME: Cecily Page  
DATE: Friday September 28, 2018

## P1.1

I unzipped the reports files so there is a folder inside the hw\_01\_crpage folder called reports and this folder contains all the .dat files

## P1.2

Load\_reports takes in a filename and read the file line by line. It strips the line of \n because it’s not useful to the data and then it splits the line at the ‘:’ character. I did this because : appears in every file separating the type of error by the number of times it occurred.

The next part is just putting the relevant information into a dictionary.

This function returns a dictionary with the following format {error name: number of times occurred, …}

## P1.3

I created a loop that went through every file in the reports folder. It uses a try/ except statement to add the new number of errors to the existing total for that error type, or if the error type doesn’t yet exist, creates a new entry in the dictionary for that error type. I know this is working well because I printed out the length of the final dictionary while adding each error type to a set, and at the end of this event there were the same number of entries in the dictionary and set.

## P1.4

I created a function called get\_id\_nums() which returns a list of datacenter numbers. it iterates through every file in the reports folder and gets the DATACENTER number from each file. It does this by finding the word ‘DATACENTER’ in the line and taking the integer that follows it and returns a list of data center numbers. To see if all these numbers were unique I took the length of the list and compared it to the length of the list turned into a set. They were the same length, so we know all the data centers have unique numbers. I also noticed that the file name was corresponding with the data center number, so to make sure no file was skipped, I ran through this loop for numbers (0, 1431) and got not print statements that said there was a file missing. Therefore all the datacenter numbers are sequential.

## P1.5

The error modes were not quite consistent. I am defining consistency as each unique string implying a different type of error. In this case the errors were not consistant. There were two different error types meaning the same error; A/C error. I cleaned this by adding the number of errors from each one up and creating a new AC error type that has the sum of both types of errors. I then deleted the other two AC error types from the dictionary.

## P1.6

Yes. The error mode counts are consistent. The way I created the dictonary didn’t allow for error in this situation. If there was an error in an individual data center it as added to the total number.

## P2.1

Yes. Some data centers experience much more errors than others. I created tables to show this.

The Centers with the most errors

000223 : 111967

000353 : 105275

000959 : 28322

001357 : 14999

000841 : 12979

001386 : 11777

001035 : 8386

001176 : 6920

001092 : 4592

000159 : 4057

The minimum number of errors at a data center is 6 while the maximum is 111967.

## P2.2

The expected number of errors is 262.

## P2.3

Total Errors in highest occurrence

Misc. elec. : 214269

Physical intrusion (water) : 50684

Fiber pipeline in : 42601

AC : 34400

Fiber pipeline out : 12237

HVAC : 8163

Power/generator reduction : 7762

Operator (non-employee) error : 4489

Physical intrusion (person) : 617

Power/generator loss : 0

The most common error is Miscellaneous electric error.

## P2.4

## P3.1

## P3.2

## P3.2 Bonus

## P3.3 Bonus