CHECKPOINTS AND REFRENCE GUIDE

Q1 Folder

1. There are a lot of files in this folder, but the naming scheme will make it easier to recognize what each file stands for. Data files are in fits and csv formats. The fits files are the original files provided. The csv files have been created from the fits files. There are two sets of csv data files: One with the original data extracted from the fits files and another with the same dataset with two extra colums for FieldID and MJD added. There's also a csv file named tabledatalocationq1.csv which lists the FieldID, Filename of obsservations, Filter and MJD for all the data files. This in addition with the data files will form our relational database. There's also another csv file which lists all original file info and was provided by the instructor.

2. The conversion from fits to csv involves two python files. They are apprpriately named: fitstocsv.py and addmjdandfieldid.py.

3. The steps and queries for 1a, 1b and some of 1c is described in the text file named: q1queriesandsolutions.txt (NOTE Jan 23rd: A very important edit has been made in this file which needs to be read to see all results in context.)

4. There are certain other csv files and another python file. Their origin and pupose will be known after going through the text file indicated above. In addition, there is also a .sql file which has the schema for creation of database test2.db.

5. Lastly, there is also an ipython notebook named q1_c_simulation.ipynb. This is for generating the 100,000 samples for the Euclid mission as part of question 1c.

Q2 Folder

1. There are two vot files are two csv files. The csv files contain data extracted from the vot files. The python script for this is given in vottocsv.py.

2. The ipython notebook named q2.ipynb gives the solution to the second question.

Added Notes:

I'm very new to programming and hence my use of coding will be very unaesthetic. I'm not proud of that fact at all. But, for now it gets the job done. I've hence added comments wherever needed to facilitate better understanding of what I want that code to do.

This project was done with the help of code in the assignments, lecture materials and more prominently sql documentation and scikit-learn documentation. So, it is implied that any code based on that has an implicit reference to these sources. Stack Overflow was also used in many cases and references for the webpages have been provided in the files containing the codes.

I also discussed the nitty gritty of this project with three friends and a lot of ideas was shared. I have made it my first priority to write the code myself, so while the ideas may seem same, the codes will differ and the explanations for those codes will vary.