File IO - Ms.Quesnelle
2023-03-10

Aditya Sen

Solar Flares and Sunspots Analysis

Description of Data:
Two datasets are used for this project, both are csv files. Solar flares and Sunspots.

RHESSI Solar Flare Data from 2002 - 2018:
This dataset tracks when solar flares episodes occur. Counting the number of flares that happen in an episode, and corresponding data that includes:

1. **Date**: YYYY-MM-DD
2. **Duration**: in seconds
3. **Number of flares**: integer or scientific notation (e.g. 2e+5) *FLOAT CAST*
4. **Energy**: range of the energy observed (e.g 60-100) in kilo electron volts

This data was uploaded by a Kaggle member, retrieved directly from NASA's RHESSI mission data. All NASA data is available to the public.

SAMAHA, K. (2021). hessi.solar.flare.UP_To_2018.csv, Version 3. Retrieved March 10, 2023 from https://www.kaggle.com/datasets/khsamaha/solar-flares-rhessi.

Sunspot Data from 1818 - 2019:
This dataset tracks the number of sunspots observed on an Earth day. This data includes:

1. **Year**: int
2. **Month**: int
3. **Day**: int
4. **Number of Sunspot**: int (if data is missing, integer is -1)

This data was uploaded by a Kaggle member, originally taken from the Solar Influences Data Analysis Center (SIDC).

ABHINAND. (2020). sunspot_data.csv, Version 2. Retrieved March 10, 2023 from https://www.kaggle.com/datasets/abhinand05/daily-sun-spot-data-1818-to-2019.

Application / Purpose:
The importance of this data is to analyze and find patterns or predictions about the activity of our mother star. Solar flares release massive amounts of electromagnetic radiation, this radiation can interfere with electronics back on Earth, from simple miscalculations on computers to mass area blackouts of the electric grid. First analysis is to graph a time series of the sunspots and peak energy of flares throughout the decades, specifically 2002 to 2018. Second analysis to graph the correlation between sunspots on an Earth day, to the solar flares observed. Sunspots is the independent variable, while solar flares are the dependent variable in the analysis.

SUPPORTING General Data Structure:
The first dictionary is used as a lookup to reference in other following active data structures. All_flares organizes all the flares in a certain day into a list of tuples, each flare episode is mapped to a certain day it happened on.

```
All_flares = { date: [((low_energy_value, high_energy_value), Duration, number_of_flares), …], … }
```

The second structure is a list of tuples. Having the date and the number of sunspots on that day, since this is a list it can be iterable for the active data structures following. We can use the date in this tuple for a lookup in the All_flares dictionary to correlate and analyze these two data structures.

```
All_sunspots_daily = [(date, num_of_sunspots), … ]
```

ACTIVE Time Series Energy & Number of Sunspots:
Each date (independent variable) maps to two dependent variables, number of sunspots on that day, and the peak flare energy on that day. To graph the time series graph.

```
Time_series_analysis = { date: (num_of_sunspots, peak_flare_energy), … }
```

ACTIVE Correlation between the Sunspots and Solar Flares:
Each is used as a lookup for two variables, number of sunspots which is the independent variable, and number of flares which is the dependent variable in the correlation application.

```
Corrleation_flare_sunspots = { date: (num_of_sunspots, num_of_flares), … }
```