Abstract

The study assessed the occurrence of sunspots and flares using sunspot and flare data available from the reference provided on the website ‘https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/documentation/miscellaneous. The solar flare data showed 9 columns and 72, 054 rows. Viewing the sunspot data using appropriate functions showed that the data has 16 columns and 271,892 rows. Similar steps were taken to load the solar flares dataset. Data exploration was done to understand better the data after the needed libraries for the analysis had been imported. The data was visualised to see the highest number of sunspots and solar flares occurrences. The correlation between sunspot and flares were then accessed. The study was restricted to the X and M sun flare in determining the occurrence and correlations between sunspots and flares. Evaluating monthly and weekly reports visually demonstrated a linear association between sunspots and solar flares. Thus, to a lesser degree, more sunspots may lead to more solar flares.

Keywords: Sunspot, Flares, Occurrence, Correlation, Visualisation

INTRODUCTION

Sunspots are a phenomenon that manifests as transient, darker-than-the-neighbouring-areas spots on the photosphere of the Sun. They are areas of lower surface temperature brought on by magnetic flux concentrations that prevent convection. Sunspots are areas of lower surface temperature linked with concentrated magnetic field flux and vigorous magnetic activity on the Sun's photosphere that are darker than the surrounding photosphere regions on the visible solar disk. They typically come in opposite magnetic polarity pairs. (Oyedokun & Cilliers, 2018). The solar cycle, periodic increase and reduction in the number of sunspots, lasts for around 11 years. A single sunspot may last for a few days to several months. Secondary solar phenomena such as prominences, reconnection events, and coronal loops are present in conjunction with sunspots. The magnetically active areas connected to sunspot groups are the origin of most solar flares and coronal mass ejections. (Ichimoto, 2019).

It is well known that the solar magnetic cycle tends to be followed by solar activity, such as solar flares and Coronal Mass Ejections (CMEs). For instance, the phase and amplitude of the solar cycle (SC) tend to correlate with the occurrence of CMEs (Webb & Howard 1994). Numerous investigations have demonstrated that the most significant phases of Sunspots do not coincide with the epochs of important flares activity (Hudson et al., 2014).

BACKGROUND

Sunspot groupings have been explained using Zurich, Mount Wilson magnetic, and McIntosh classifications. Walmeier (1947) devised the Zurich sunspot classification by updating Cortie's categorisation method (1901). (Lee et.al., 2016). The McIntosh classification considers sunspot group size, stability, and complexity, while Mount Wilson uses magnetic characteristics (McIntosh, 1990). (Hale, 1919). McIntosh classification describes active solar zones using white-light pictures (Bornmann, Kalmbach, and Kulhanek, 1994). McIntosh's categorisation uses Z, p, and c. The modified Zurich class is defined by group length, penumbra presence, and distribution. p represents the primary sunspot type's penumbra.

A parameter determines group compactness. Finding the 60 McIntosh sunspot categories begins with classifying sunspots as unipolar or bipolar (McIntosh, 1990). Morphological, statistical, and machine-learning solar flare forecasting systems exist. Sunspot grouping characteristics underlie all three. (Lee et.al., 2016). However, only the morphological traits of sunspot groups are used in the first category of schemes (morphological schemes). To estimate the flare occurrence rates and probabilities, the schemes use historical records of flare rates for each type of sunspot group (Gallagher, Moon, and Wang, 2002). The second type of scheme, called statistical scheme (Yu et al. 2009), is based on morphological schemes but relies on a statistical analysis of the flare-related sunspot properties and flare intensities by mathematical models. The third category of schemes, called "machine learning schemes," is a subset of artificial intelligence that enables computers to develop a model system based on empirical information, like the nature of sunspot groups. Recently, flare forecasting has also used machine learning techniques like neural networks (Yuan et al., 2010), but few exist to find correlations among occurrences. This study seeks to discover hidden correlations between sunspot data and solar flare occurrences to help researchers identify crucial elements and correctly predict solar flares. These disasters could harm Earth's communication and energy networks. Predicting these disasters has been difficult for experts. Space weather prediction is the trickiest. Sunspots shape the solar cycle. The study will analyse the facts and evaluate their meaning. It will also create data processing tools to read and extract the essential sunspots and flares data, developing a visualisation method to show the association between sunspot statistics and flare occurrences.

MAIN PART

- First Step

Importing of needed Libraries

```python
#import libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from IPython.core.interactiveshell import InteractiveShell

#auto print items
InteractiveShell.ast_node_interactivity = "all"

#setting params
sns.set_style("darkgrid")

from prettytable import PrettyTable
import json

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
warnings.simplefilter("ignore")

%matplotlib inline
```

Figure 1: Importing Libraries

The needed libraries important for statistical analysis, as seen in Fig. 1, were imported for the study.

- Loading and Previewing the Dataset

For this study, two datasets (Sunspot and Flare Data) were used and loaded for wrangling. The next step is specifying the columns using the reference provided on the website for the sunspots data 'https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/sunspot regions/usaf_mwl/documentation/usaf_sunspot-region-reports_format.txt'.

Viewing the sunspot data using appropriate functions showed that the data has 16 columns and 271,892 rows. Similar steps were taken to load the solar flares dataset, using the reference provided on the website ('https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/documentation/miscellaneous/software/xraydatareports.proc', viewing the solar flare data showed that the data has 9 columns and 72,054 rows.

The nature of the missing data, the magnitude of the dataset, and solutions were checked by exploration before visualising the data for better insight.

The missing data are seen in Table 1 below.

Table 1: Missing Data (SUNSPOT AND FLARE DATA)

Columns with Missing data	Sum of missing values
Date	9
Ns	36854
Location	9
Class	36948
NOAA Number	46
Date (Flare)	6
Flare type	33
Location(Flare)	37305
NOAA Number (Flare)	33299

The difference in the sizes of missing data in the sunspots data and solar flares data showed that no singular approach would apply to both datasets, so thereby, the missing data in the sunspot data constituted less than 15% of the whole data, so dropping the missing rows tend to be a better alternative to replacing the data to avoid introducing some correctional biases into the data (i.e. replacing with either the mean, modal or 0). In contrast, for the solar flare, the columns needed for the analysis had low missing data points; hence removing them would not pose any data integrity challenge. In contrast, the missing data in the other columns that would aid visualisation were filled with 0.

ANALYSIS AND DISCUSSION

From Appendix I, sunspot number 1 was the most common in the dataset. In contrast, extreme sunspot 98 has a minor occurrence across the years in the review, and there is a sharp L-curve line movement from sunspot number 1 to the highest sunspot number. Appendix II shows that plot c12 flare type, which is the C class of flare with 12 x-ray intensity level, has the most increased occurrence across the years in review, which in effect, does not pose any dangerous threat to the Earth because the only solar flare class that are of concern to researchers in terms of the danger they pose are the M and X class of solar flares.

For better cross-referencing, the two datasets were merged with date; this was done by selecting the solar flares class of importance to this analysis (M and X class) and then concatenating the M and X class flares before merging with the sunspots data. The combined dataset has 146,174 records. A sharp reduction of rows was noticed in comparison with the original sunspots data, so dates with multiple solar flares in the merged algorithm will have the rows repeated with all the flare data to reflect an actual merging of the two datasets also from observation, some dates in the solar flares data were not captured in the sunspots data (e.g. early months of 1981) so those solar flares date will not be captured in the merged datasets. To make provision for this, the visualisation techniques take the two datasets as inputs, so the combined data provides other dimensions of the analysis.

Using the merged datasets to plot the frequency of sunspots classification showed that the BXO sunspot class has the highest number of occurrences, with over 160,000 counts. This was validated by visualising only the sunspots data, as seen in APPENDIX III. Data were also resampled into daily, weekly, monthly, and yearly estimates. The data was previewed for the sunspot and flare daily, weekly, monthly, and annual analysis to ensure the data's validity. The daily analysis is in Appendix IV. Jan 14, 2005 has the highest flares (33), with M type of flare being the common flare experienced on the day. Appendix V shows the weekly sunspot and flares analysis. The highest sunspot was within the third/last week in August 1991, while 1st/2nd week of July 2012 had the highest flare occurrence.

Monthly analysis of sunspots and solar flares in Appendix VI found that the 1980s had the most solar anomalies in the years in review, with the sunspot occurring in June 1989 and the flares in February 1982.

For yearly analysis, the sunspot for monthly and weekly analysis gave a pointer to the yearly sunspot and flare analysis (Appendix VII). The class of flare was further assessed with a focus on the M and X analysis, and it was seen that the X class flare type occurred 57 times in 1989; this is a massive discovery, as this coincides with the year the first GPS satellite went up into space. It was seen that the same year had over 505 M class of solar flares.

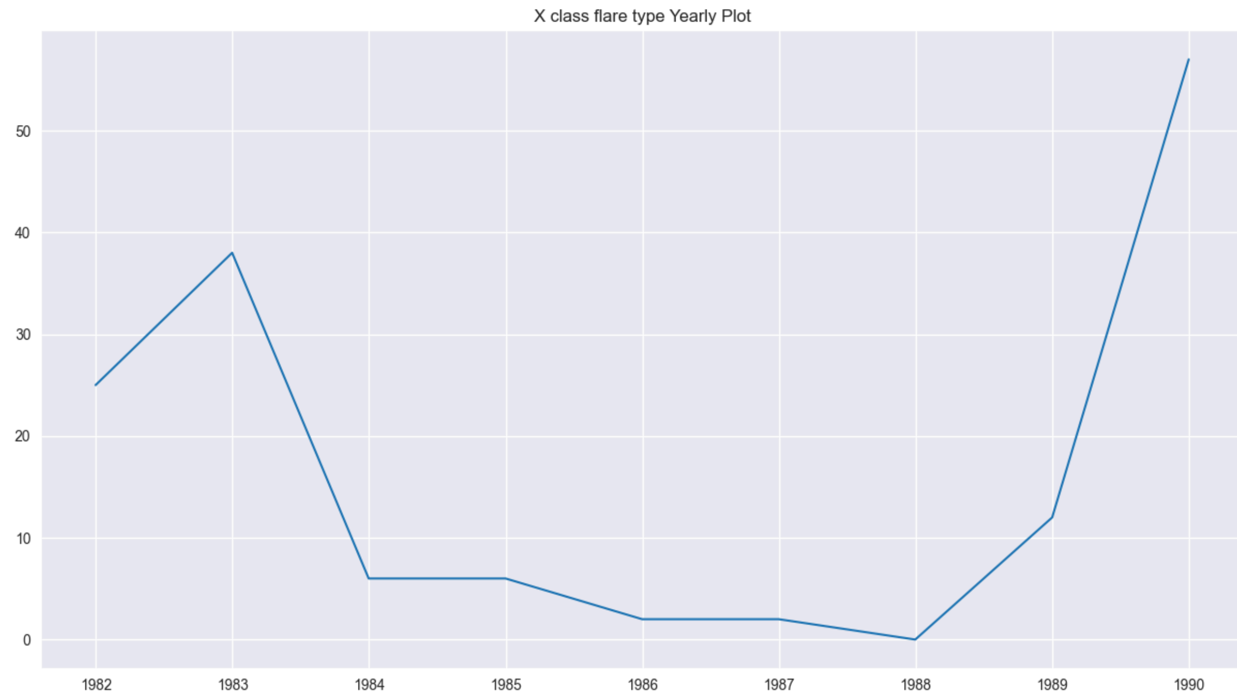


Figure 2: X class flare type

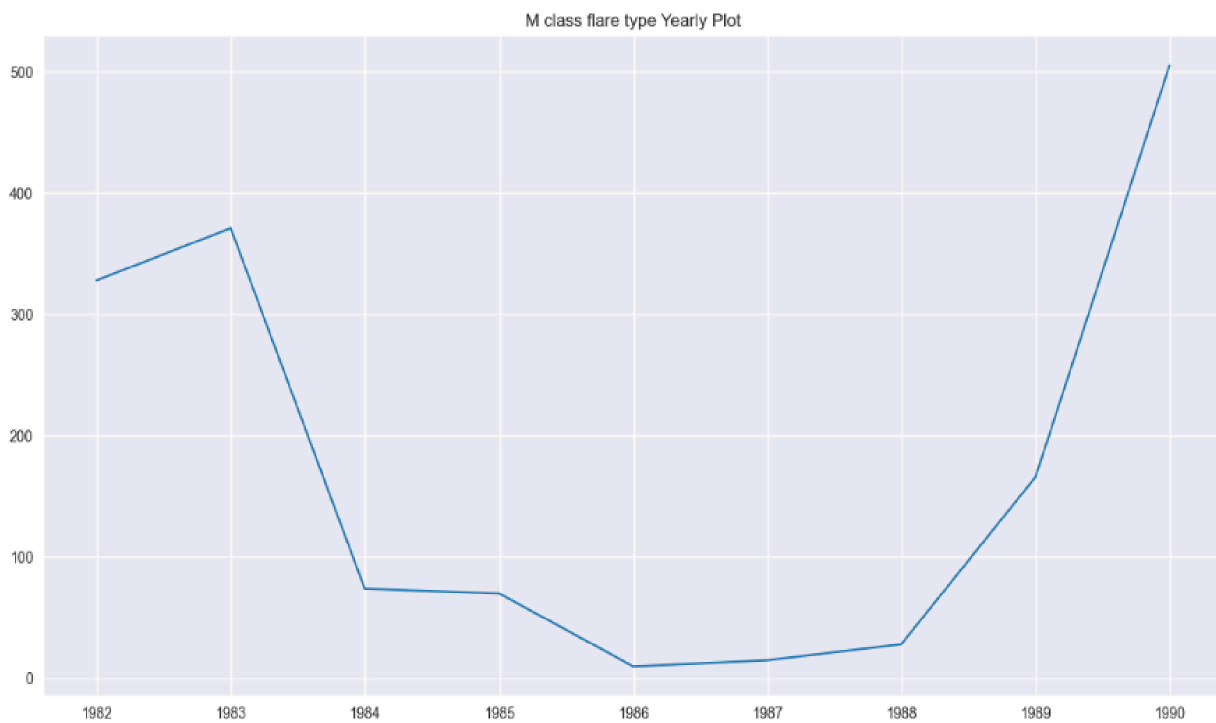


Figure 3: M type of Flare

Correlation between Sunspot Data and Solar Flares

The study showed a yearly plot linear correlation between sunspots and solar flares, as an increase in sunspots showed an increase in solar flares. (See Appendix VIII). While the M class was closely like the X and M class plot, the X class showed a disjointed pattern of X class flares, where each group starts and ends with breaks between them. The first break was from 1988; the second was from 1994 to 1997; the third was from 2008 to 2010, and the fourth was from 2016.

Apart from the first break, other breaks showed at least two years where no X class of solar flares occurred. (Appendix IX). It was seen that the monthly and weekly plots agree with the yearly analysis showing a close linear relationship between sunspots and solar flares, as seen in the Images in Appendix X. The study concluded that Sunspots and solar flares have a linear relationship, but the exact degree of linearity is unknown.

Evaluating monthly and weekly reports visually demonstrated a linear association between sunspots and solar flares. Thus, to a lesser degree, more sunspots may lead to more solar flares. The study recommends that more research be conducted to understand the degree of correlation accuracy, giving a deeper insight into the data on sunspots and flare occurrences.

BIBLIOGRAPHY

BORNMANN, P.L., KALMBACH, D., and KULHANEK, D.: 1994, Active-region class similarities and suggestions for mergers *Solar Phys*. 150, 147.

GALLAGHER, P.T., MOON, Y.-J., and WANG, H.: 2002, Active-Region Monitoring and Flare Forecasting – I. Data Processing and First Results, *Solar Phys*. 209, 171

HUDSON, H., FLETCHER, L., and MCTIERNAN, J. 2014, Cycle 23 Variation in solar flare productivity., *Solar Physics*, 289, 1341

ICHIMOTO, K., 2019. in The Sun as a Guide to Stellar Physics, 2019

Lee, Y., Moon, J., and V. M. Nakariakov, 2016. Dependence Of Occurrence Rates of Solar Flares and Coronal Mass Ejections on The Solar Cycle Phase and The Importance Of Large-Scale Connectivity, *The Astrophysical Journal*, 831:131.

MCINTOSH, P.S.1990, The classification of sunspot groups, *Solar Phys*. 125, 251.

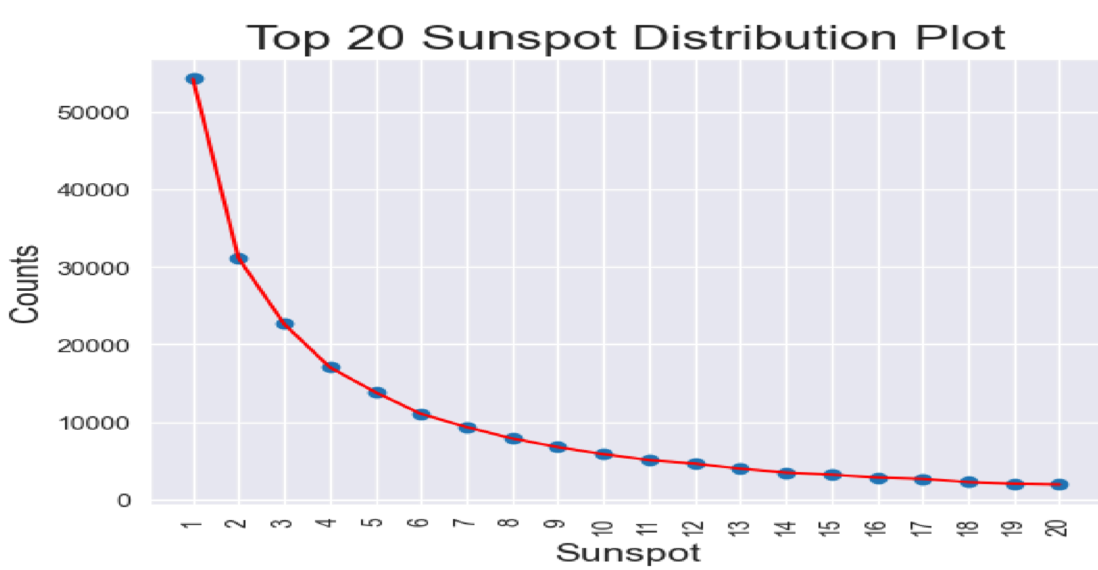
OYEDOKUN, D. and CILLIERS, P., 2018. Classical and Recent Aspects of Power System Optimization, 218

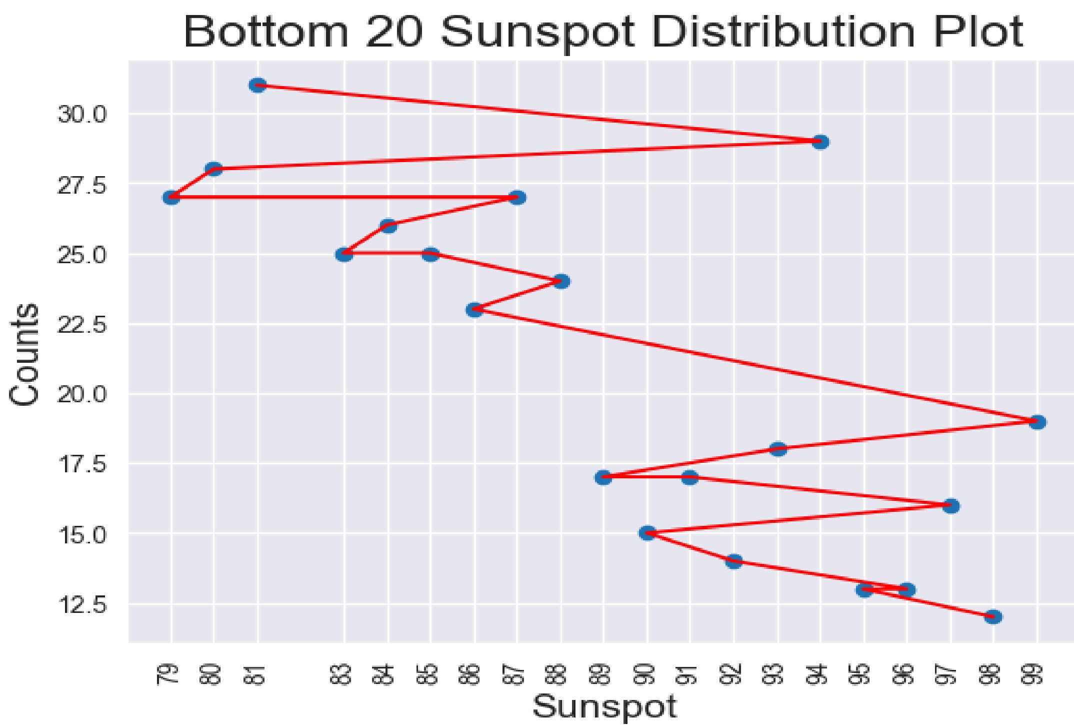
WEBB, D. F., and HOWARD, R. A. 1994, The solar cycle variation of coronal mass ejections and the solar wind mass flux, Journal *of Geographical Research*, 99, 4201

YU, D.R., HUANG, X., WANG, H.N., and CUI, Y.M. 2009, Short-Term Solar Flare Prediction Using a Sequential Supervised Learning Method, *Solar Phys.* 255, 91-105.

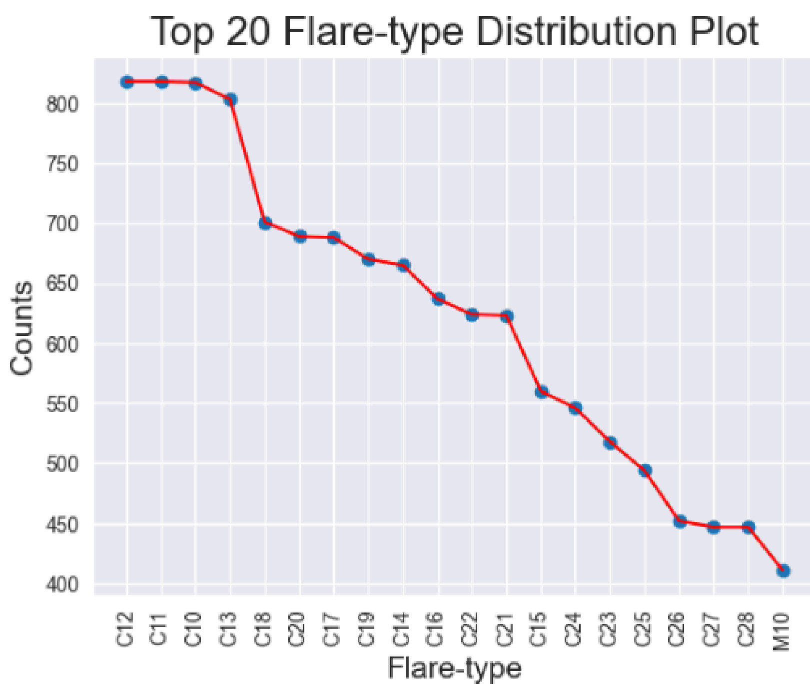
YUAN, Y., SHIH, F.Y., JING, J., and WANG, H.: 2010, Automated flare forecasting using a statistical learning technique, *Res. Astron. Astrophys. 10,* 785.

Appendix I

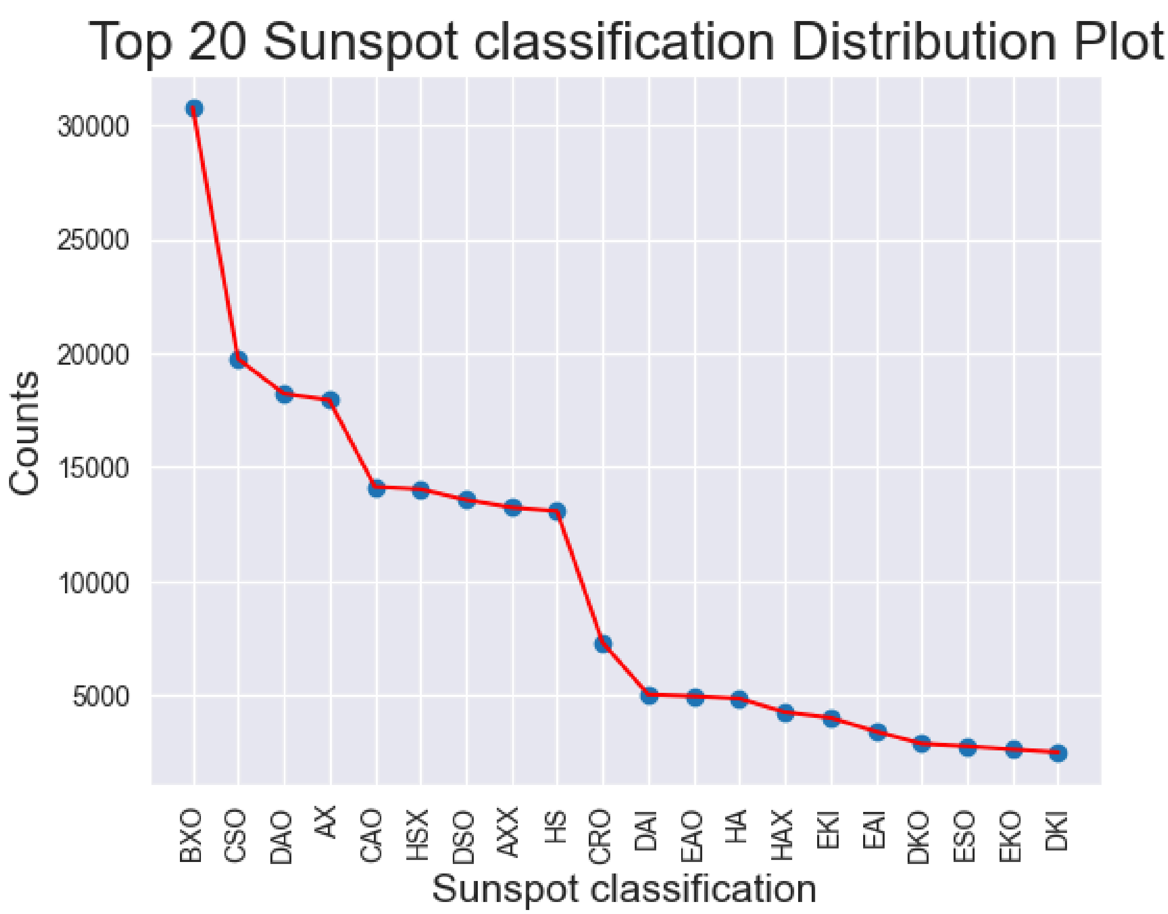



Bottom 20 Sunspot Distribution Plot
Counts
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
79
80
81
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
Sunspot

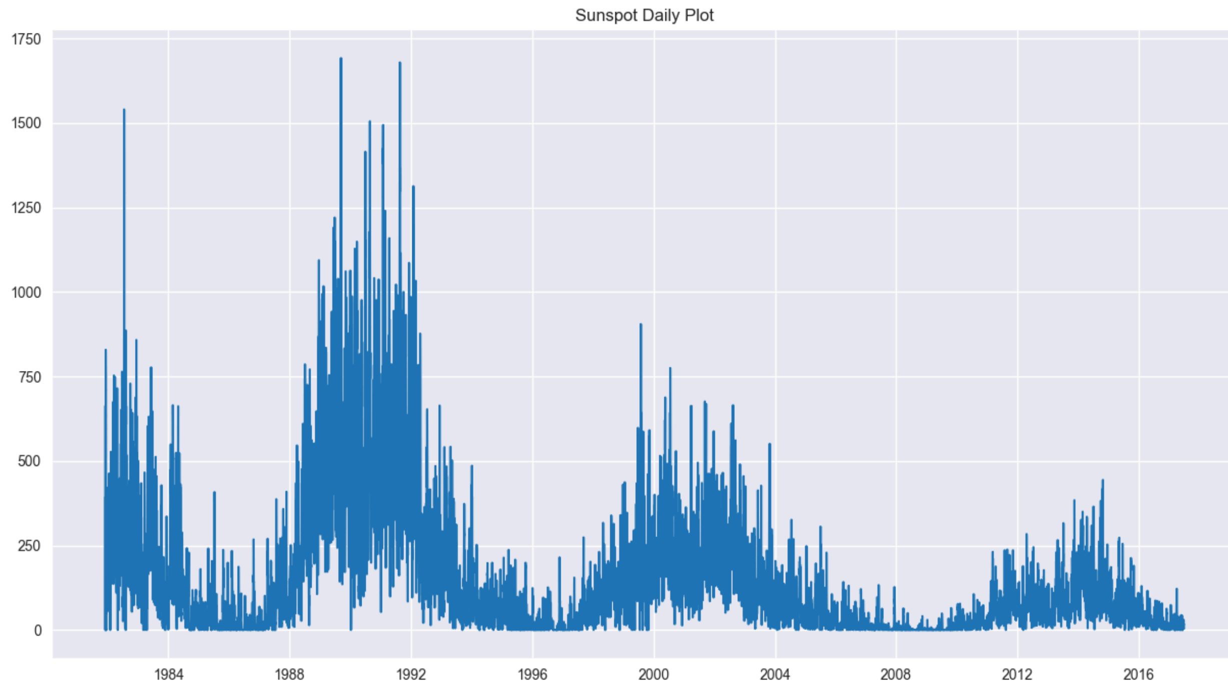
Appendix II

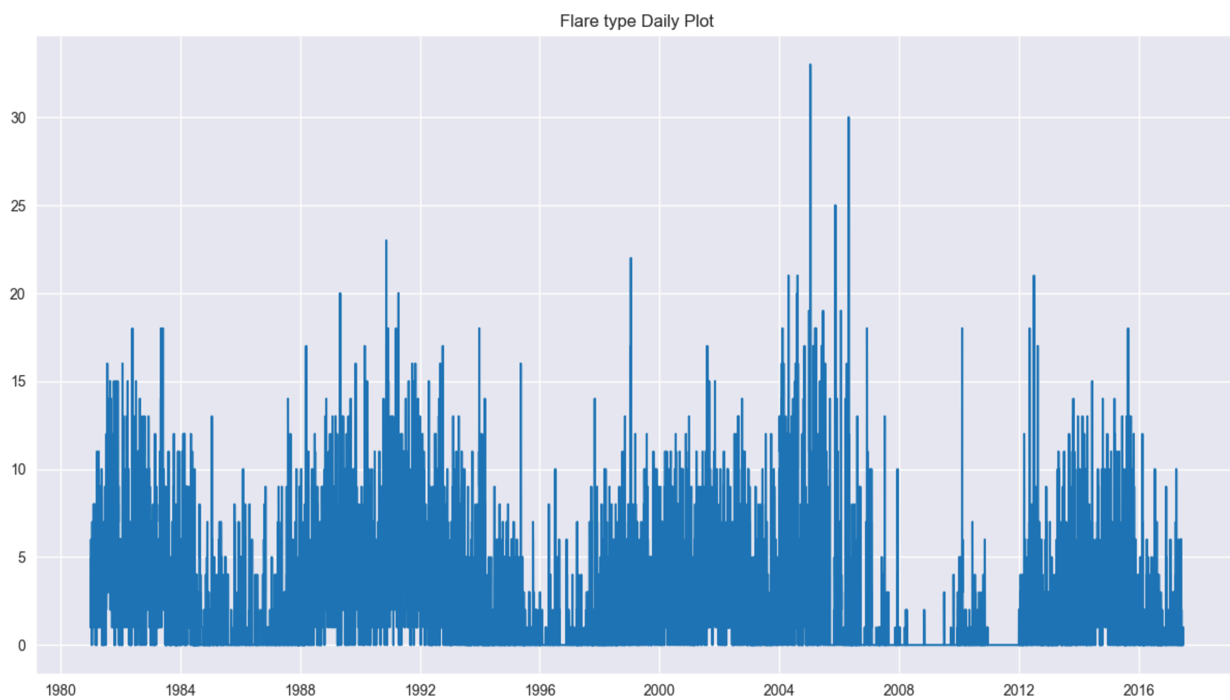


Appendix III

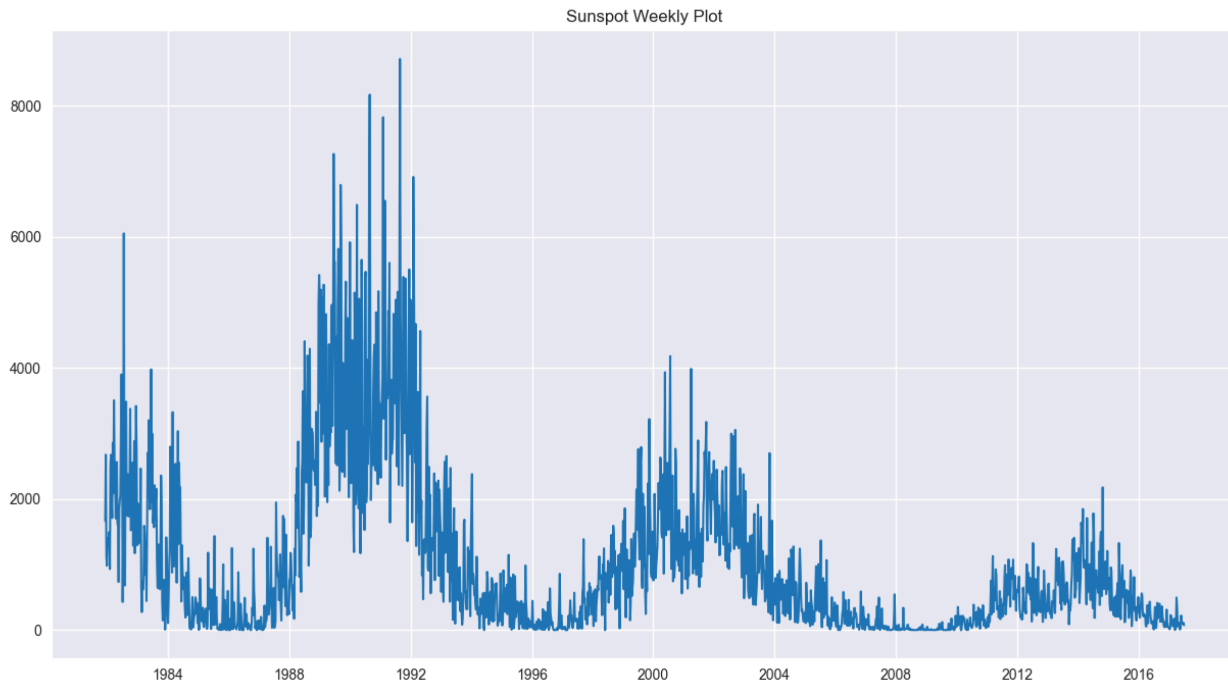


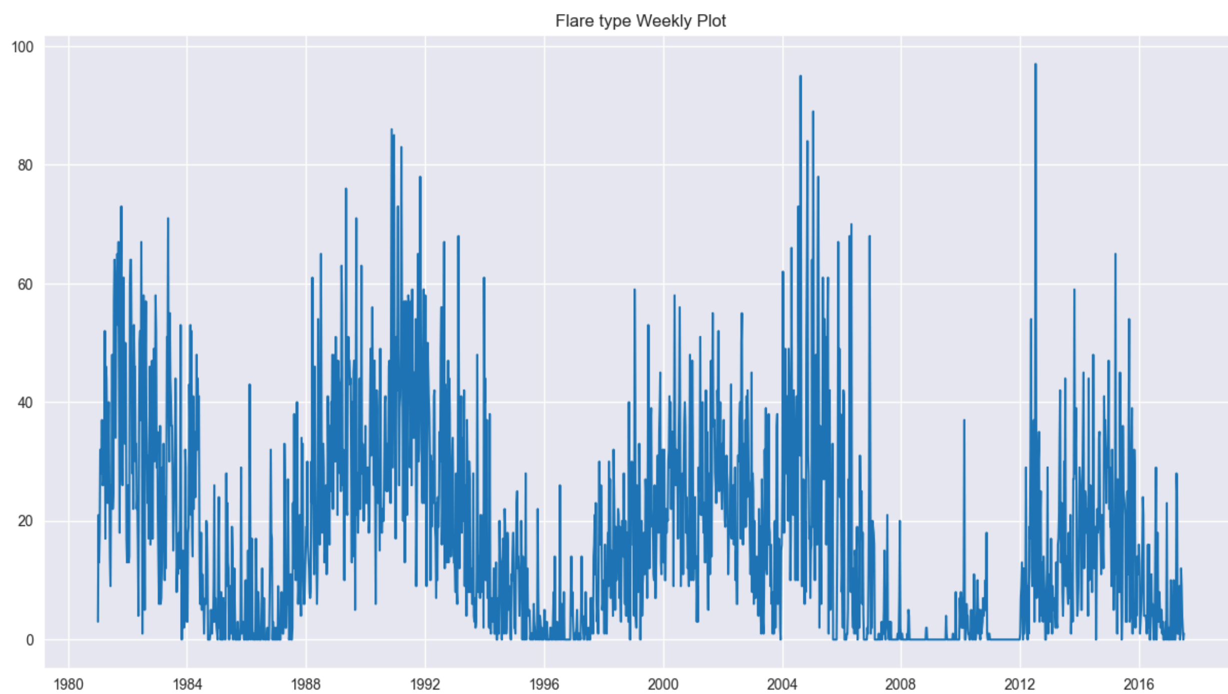
Appendix IV



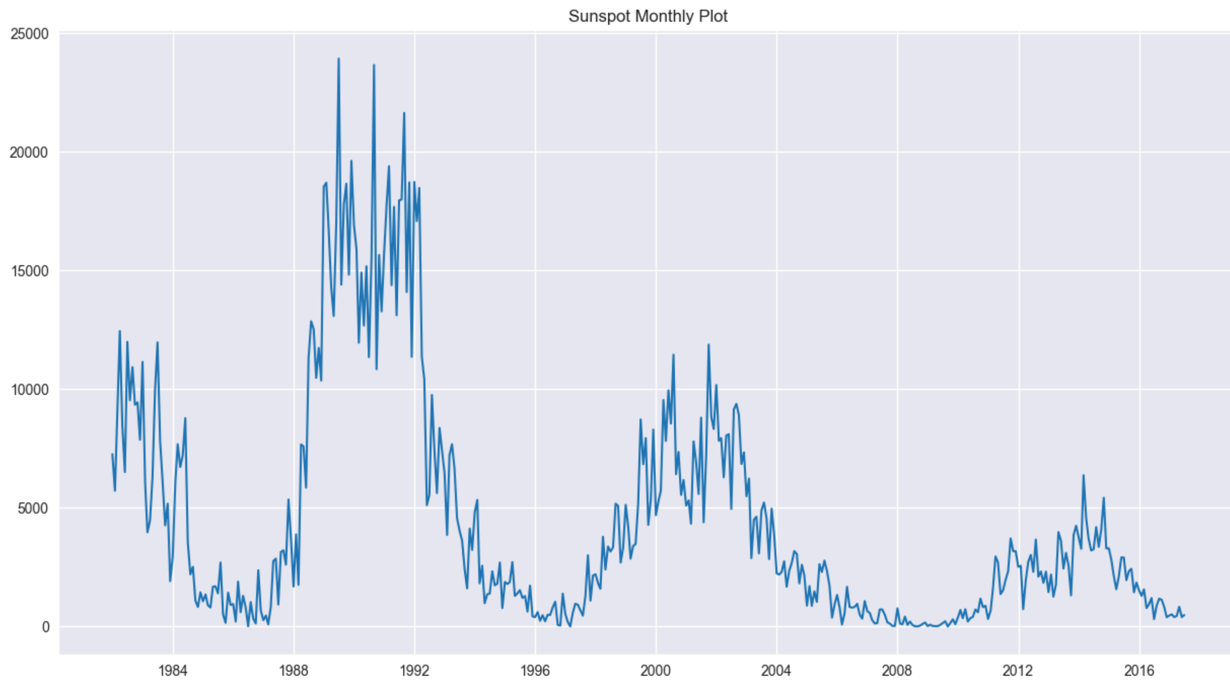


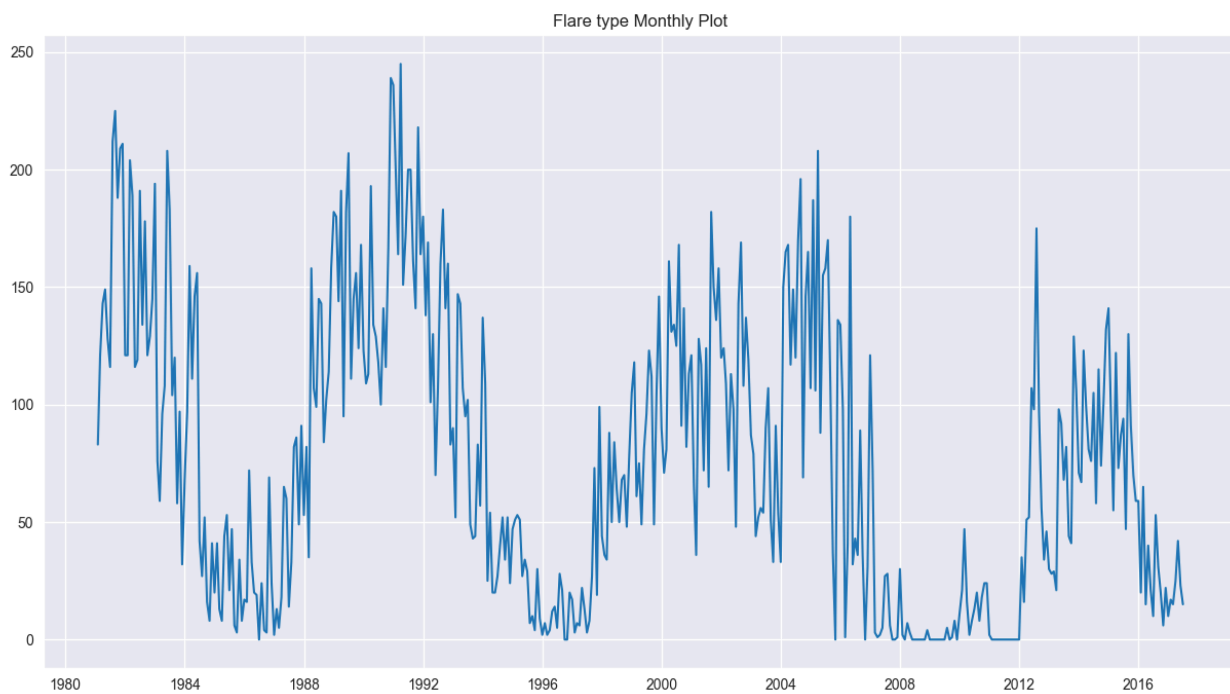
Appendix V


Sunspot Weekly Plot
8000
6000
4000
2000
0
1984
1988
1992
1996
2000
2004
2008
2012
2016

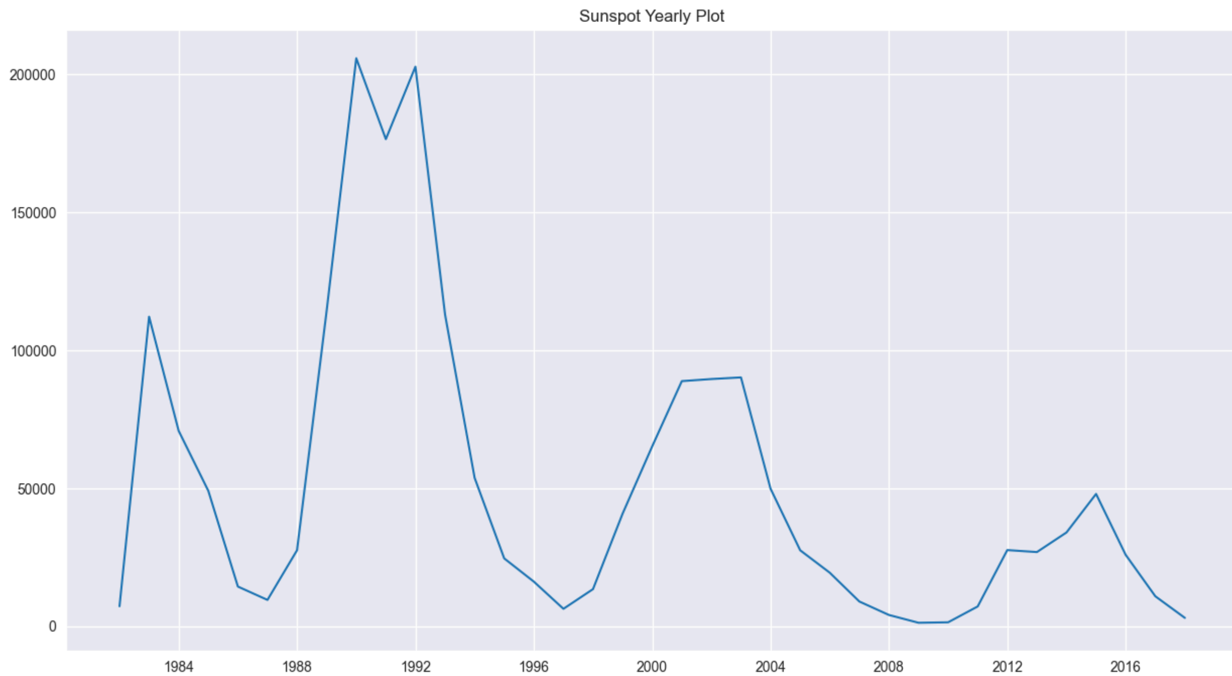

Flare type Weekly Plot
100
80
60
40
20
0
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016

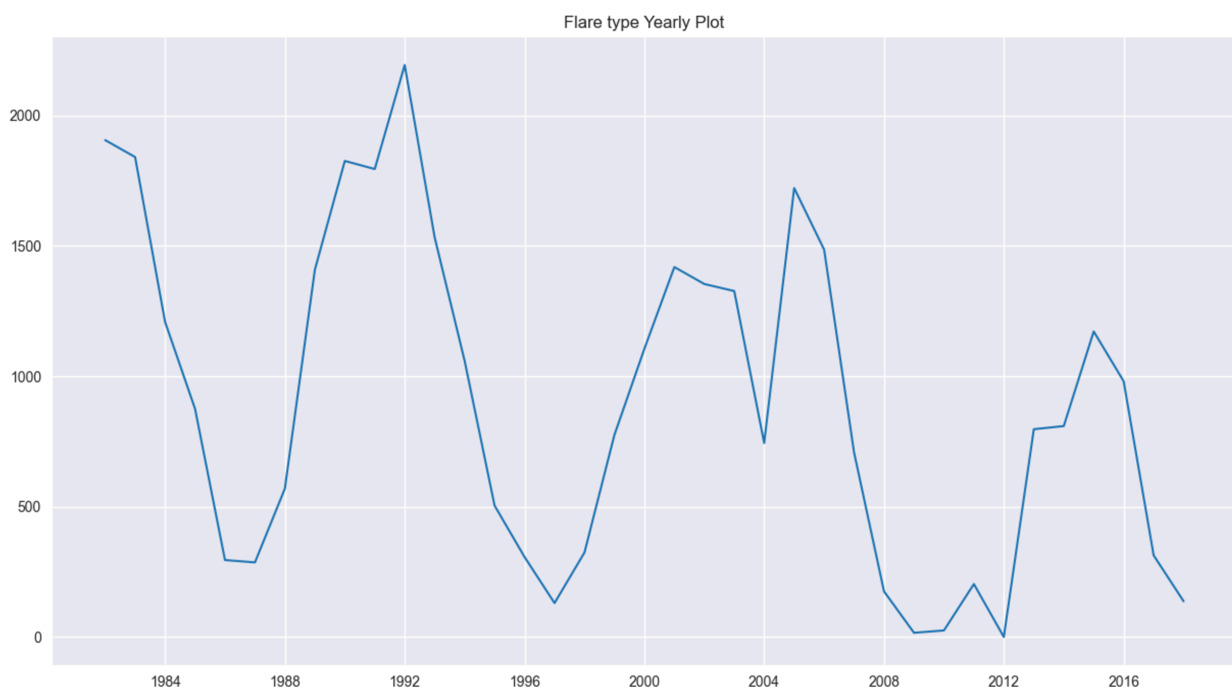
Appendix VI



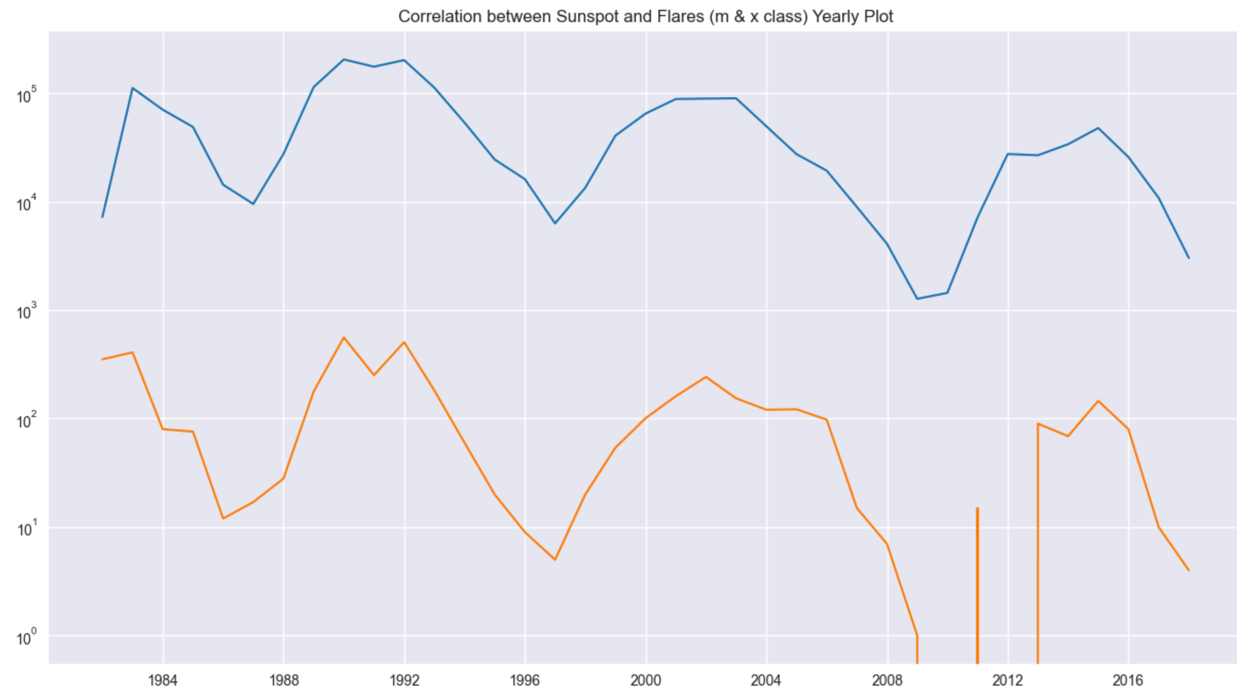


Appendix VII

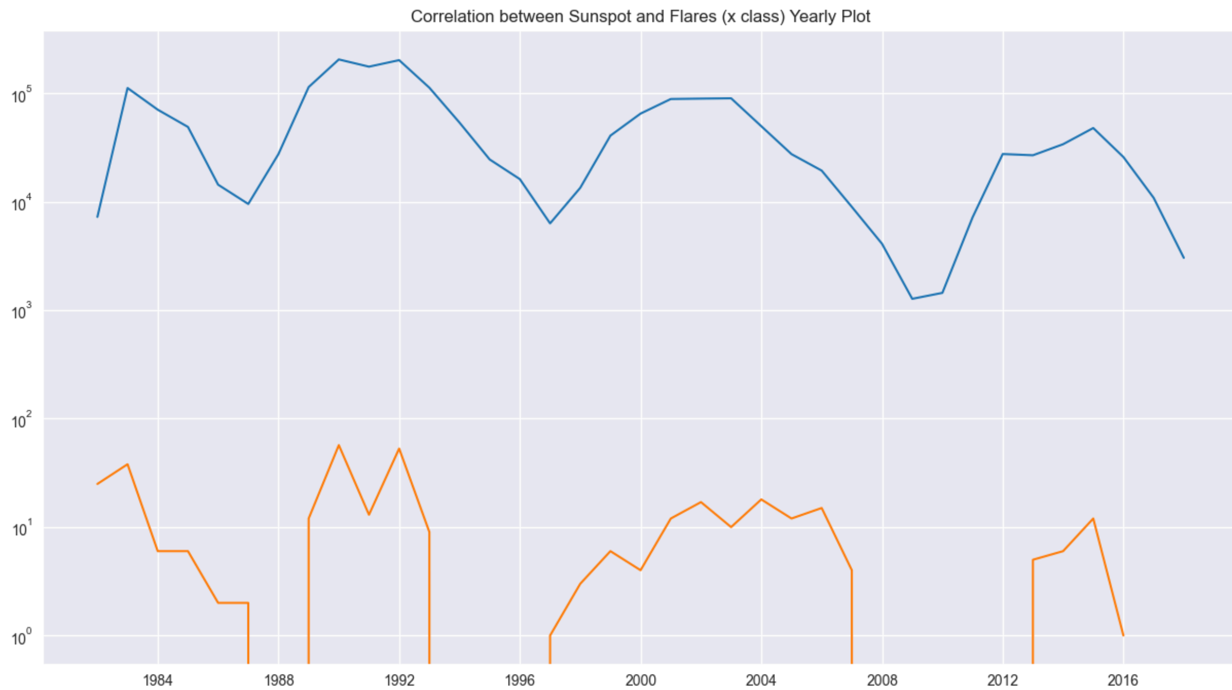




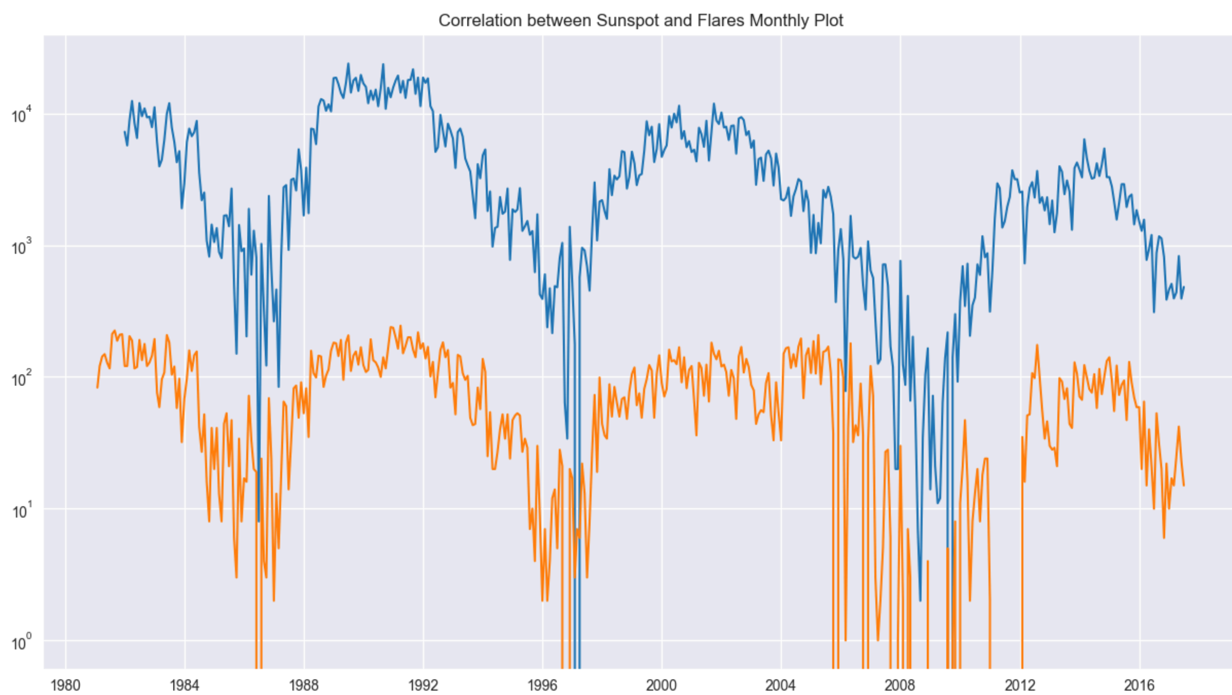
Appendix VIII


Correlation between Sunspot and Flares (m & x class) Yearly Plot
10^5
10^4
10^3
10^2
10^1
10^0
1984
1988
1992
1996
2000
2004
2008
2012
2016

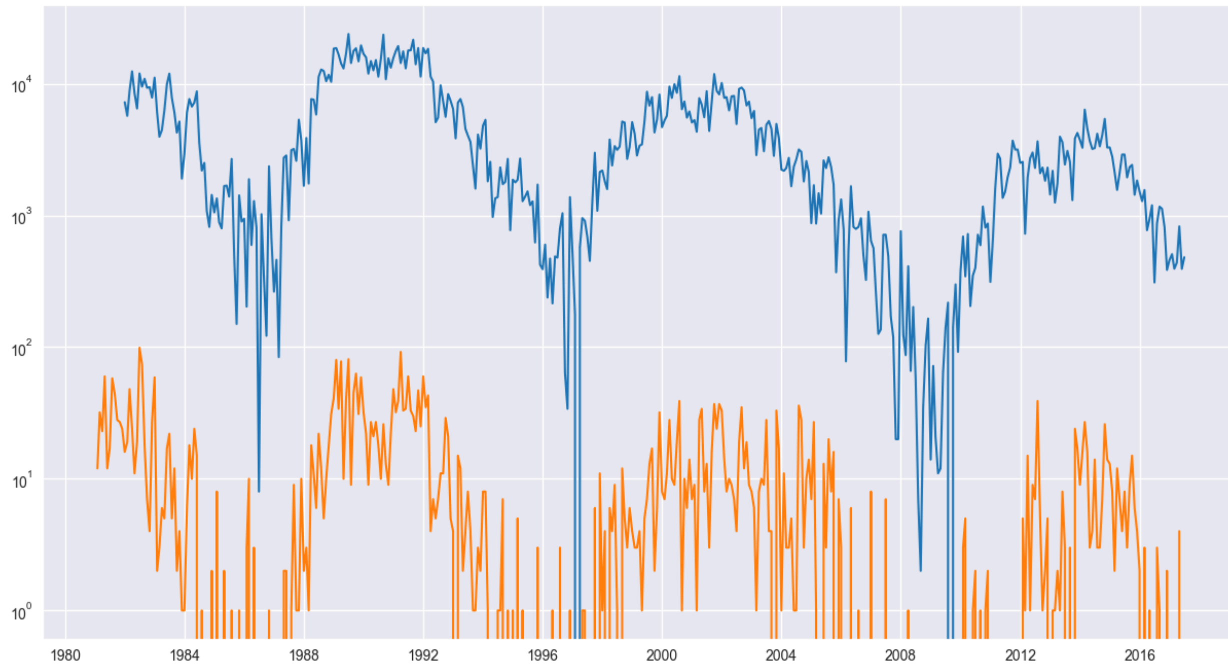
Appendix IX


Correlation between Sunspot and Flares (x class) Yearly Plot
10^5
10^4
10^3
10^2
10^1
10^0
1984
1988
1992
1996
2000
2004
2008
2012
2016

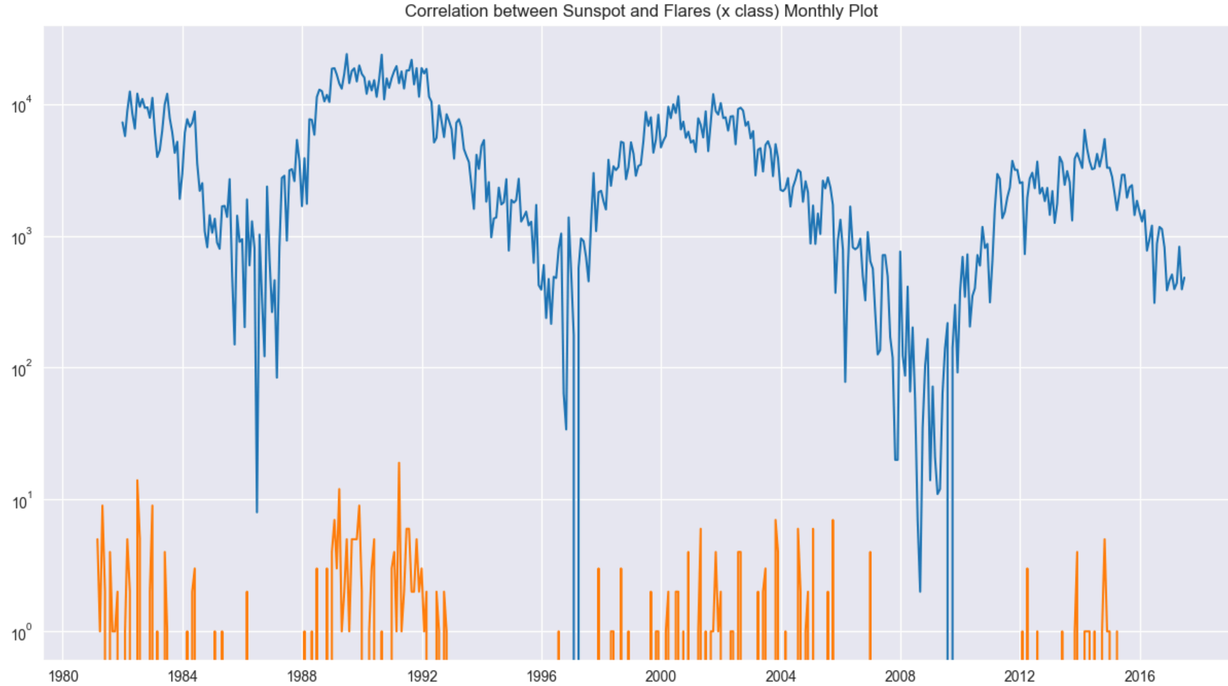
Appendix X


Correlation between Sunspot and Flares Monthly Plot
10^4
10^3
10^2
10^1
10^0
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016

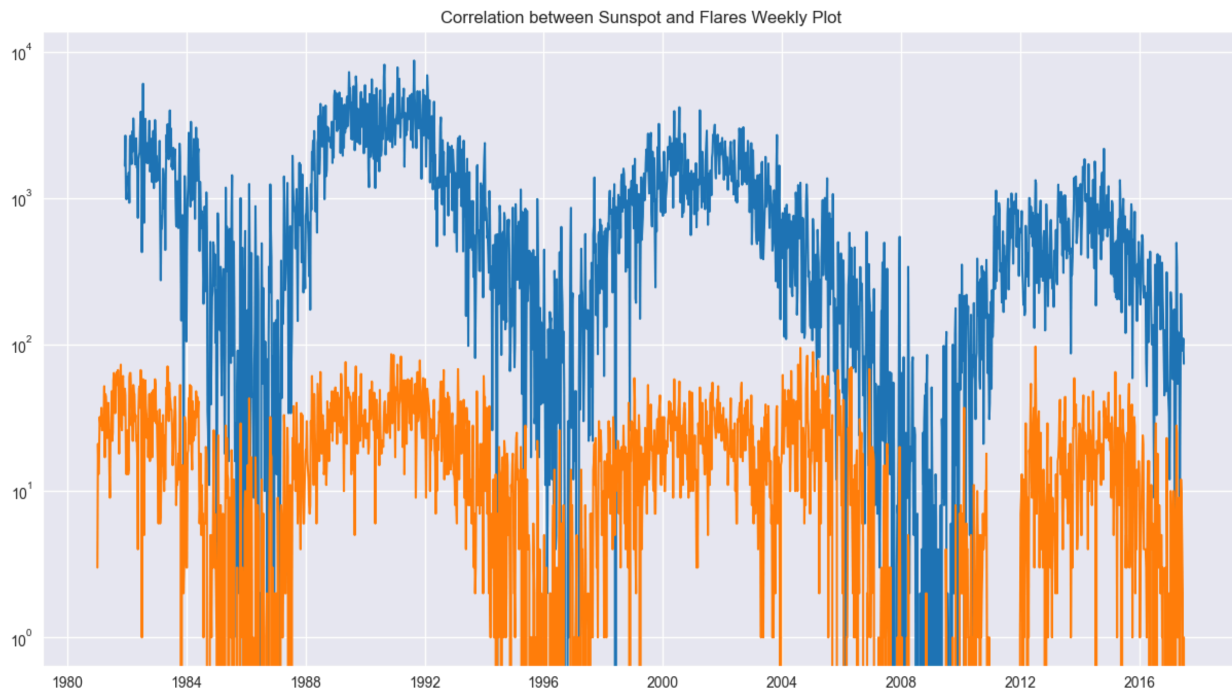
Correlation between Sunspot and Flares (m & x class) Monthly Plot

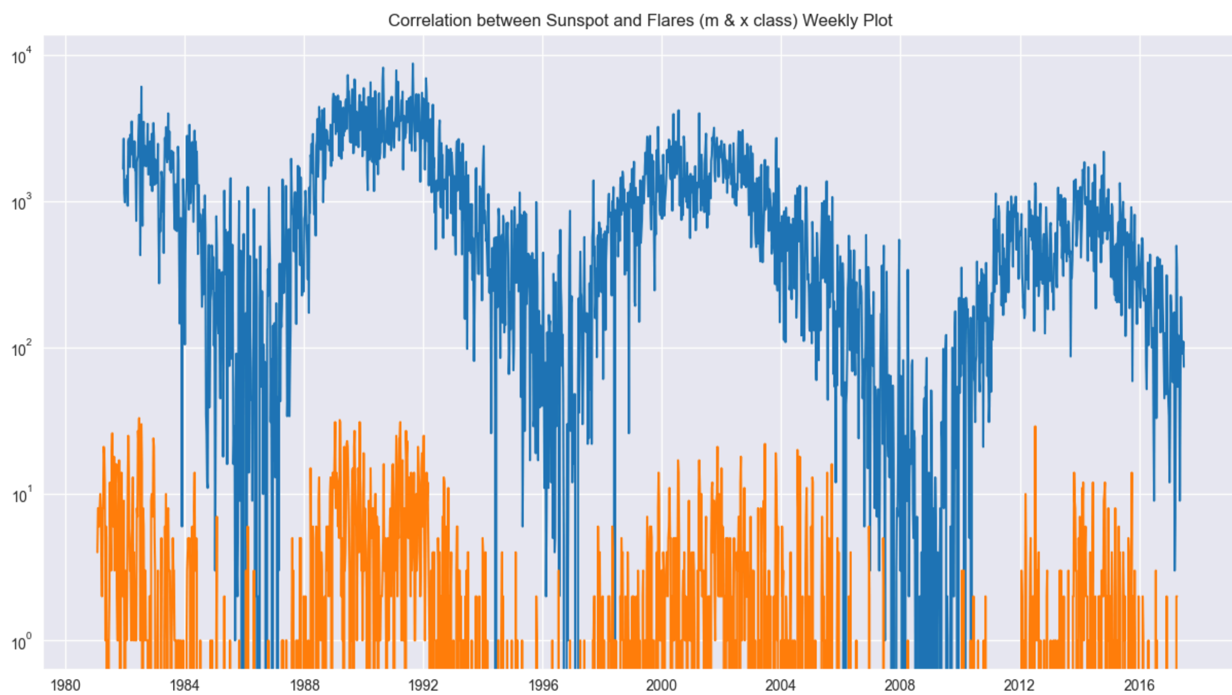


Correlation between Sunspot and Flares (x class) Monthly Plot



Appendix XI


Correlation between Sunspot and Flares Weekly Plot
10^4
10^3
10^2
10^1
10^0
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016


Correlation between Sunspot and Flares (m & x class) Weekly Plot
10^4
10^3
10^2
10^1
10^0
1980
1984
1988
1992
1996
2000
2004
2008
2012
2016