Analysis on Solar UV Radiations

Name : Edward Leen

Student ID : 23090780

GitHub : https://github.com/edwardleen95/Applied-Data-Science-1-assignment-2.git

Introduction

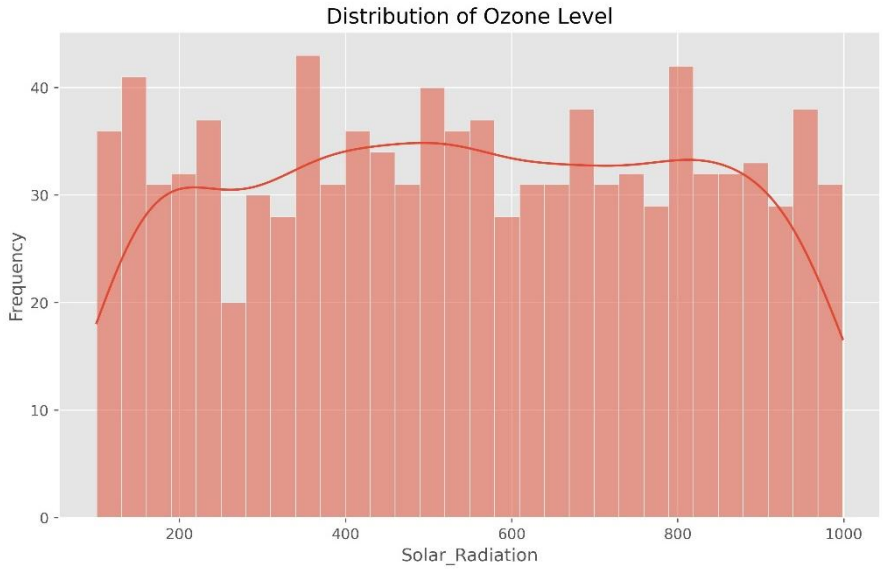
This notebook performs exploratory data analysis (EDA) on a dataset containing information about UV Radiation from the Sun. Through various visualizations such as histograms, scatter plots, correlation heatmap, cluster plots and fitted line plot, our goal is to uncover patterns, relationships, and distributions in the data. Some numeric columns (like 'Altitude' and 'Ozone Level') are excluded where appropriate.

Analysis and Results

Statistical Depth Analysis

Statistical depth provides a summary of the central tendency, dispersion, and overall range of numeric features. This summary helps to quickly understand the general properties of the data and identify any significant outliers or anomalies.

	Solar Radiation	Cloud Cover	Ozone Level	Altitude	UV Index
mean	547.959	0.49718	297.614	1.5026	1.57292
std	260.145	0.28563	57.6355	0.86333	1.32194
min	100.121	1.2E-05	200.006	0.00283	0.00104
max	998.513	0.99956	399.55	2.99805	7.15776

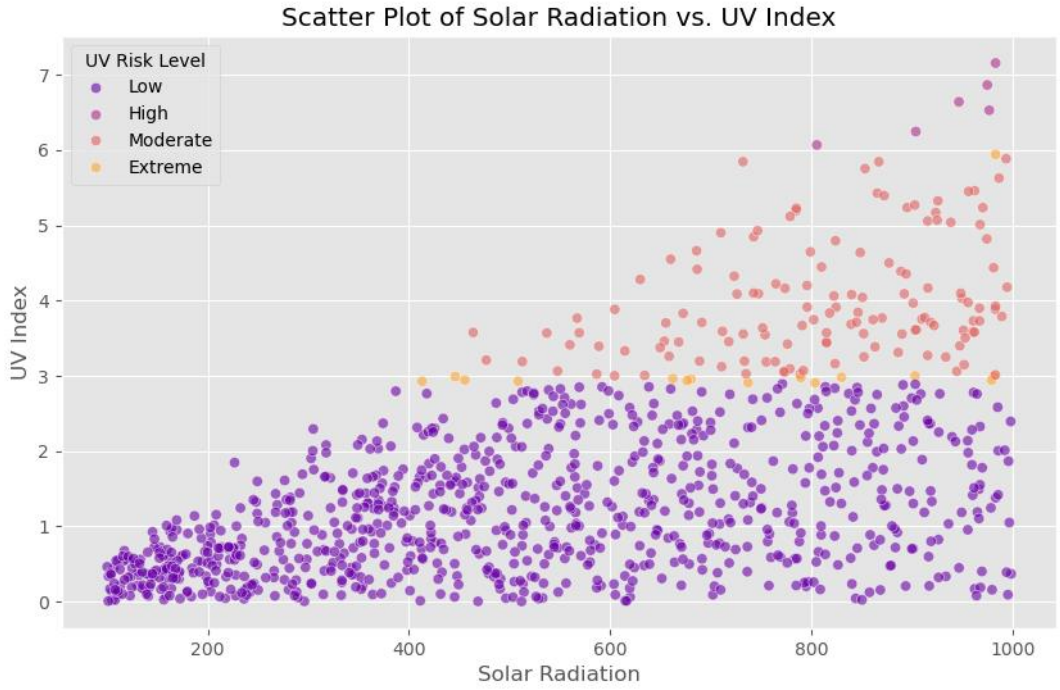


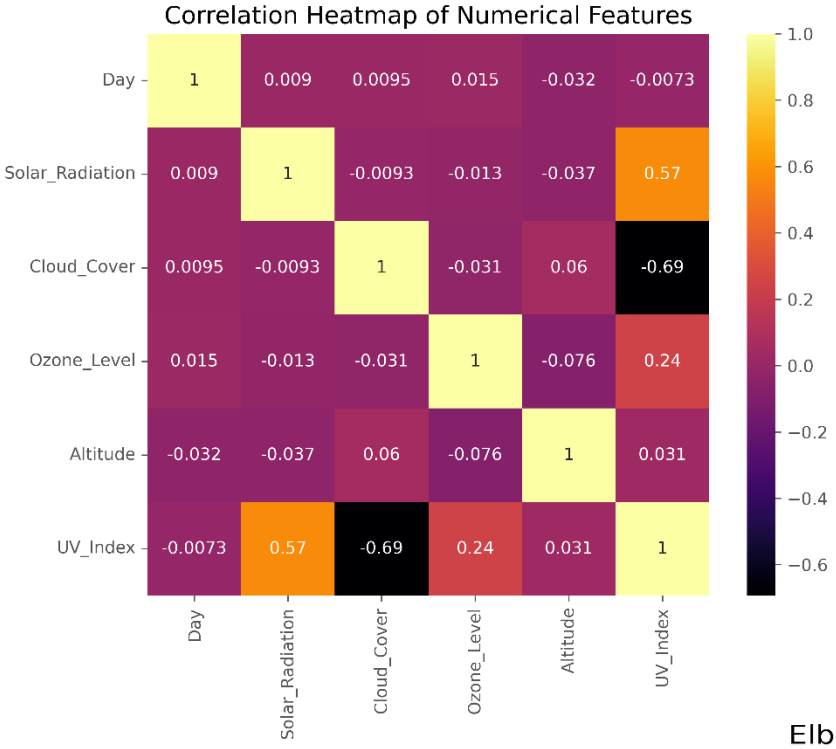
Distribution of Solar Radiation

Solar Radiation distribution reveals how UV radiation from the sun varies throughout the year. Solar radiation has a strong correlation with UV index. By observing the data we can see that the solar radiation has a frequency mixture in the high, middle and low levels. A skew of -0.00788 depicts a slight left skew since it is close to zero, which indicates it being symmetric. A kurotosis of -1.175754 depicts the distribution as less tailed and flatter than the normal distribution.

Scatter plot of Solar Radiation VS UV Index

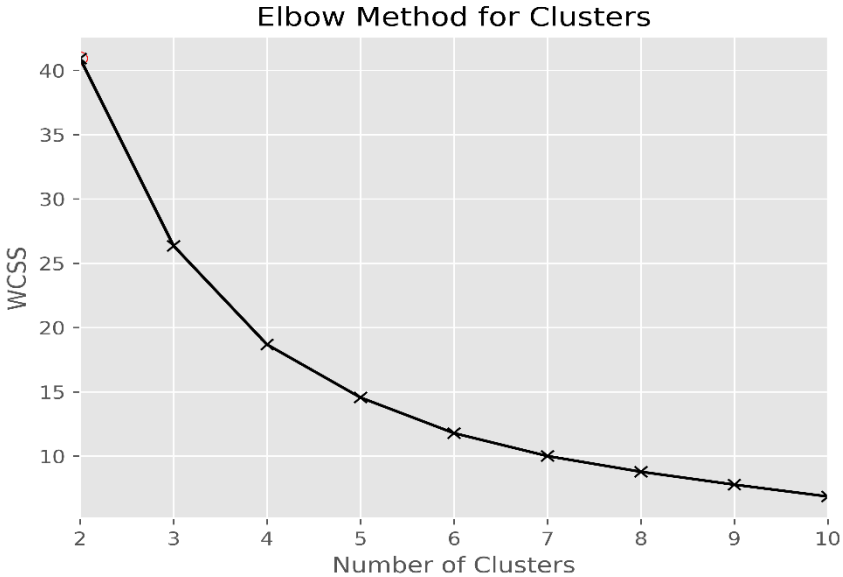
Solar Radiation distribution reveals how UV radiation from the sun varies throughout the year. Solar radiation has a strong correlation with UV index. By observing the data we can see that the solar radiation has a frequency mixture in the high, middle and low levels.





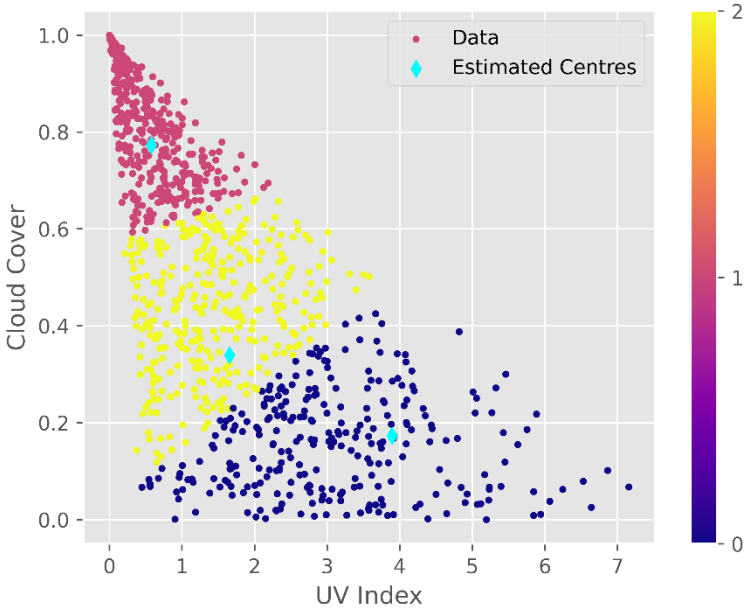
Correlation Heatmap

The correlation heatmap reveals relationships between numerical features, helping us identify pairs of variables that change together. Strong correlations, such as between Solar Radiation and UV Index, might suggest that as one variable increases, the other tends to follow.



The Elbow Method

The elbow method helps us to determine the optimum number of clusters. It involves the computation of the Sum of Squared Errors (SSE). From the graph we can see that the best number of clusters is 2. The elbow method helps us to determine the optimum number of clusters for the dataset.

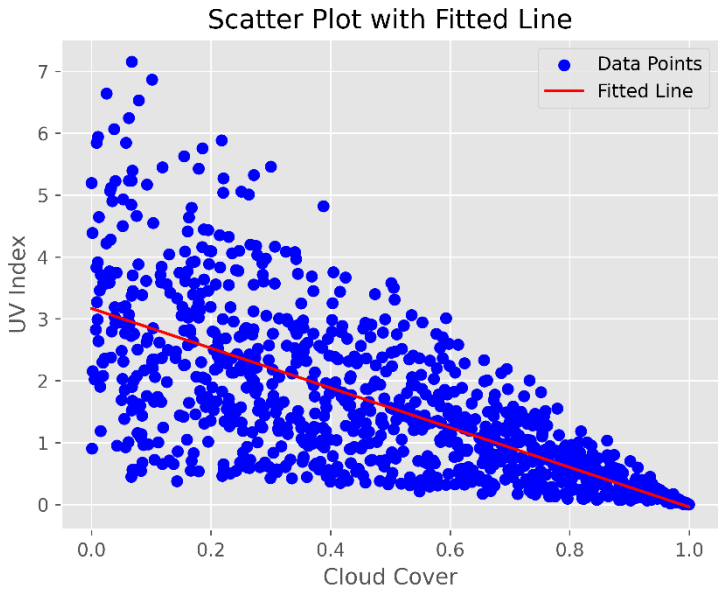


Fitted Clusters

Clustering helps us to find subsets of data that behave in a similar fashion. This means that groups behaving in a similar way also behave in a dissimilar way to other groups. Based on the elbow method I was able to conclude that the most optimum number of clusters is 3 for the data “UV Index” and “Cloud Cover”.

Least Square Fitting

It is a way of finding the best-fitted line through a set of data points by minimizing the total errors. This method is sensitive to large errors, so it prioritizes finding the best line to keep the errors as small as possible. The graph depicts the relation between “UV Index” and “Cloud Cover”. In the graph the fitted line takes the best possible route with the least number of errors to show that “Cloud Cover” changes with “UV Index”.



Conclusion

In conclusion, this analysis provides a detailed exploration of solar UV radiation and related variables through statistical and visual methods. Solar radiation shows a mostly symmetric distribution, it strongly correlates with UV Index. The clusters reveal distinct behaviour patterns. The fitted line highlighted a clear relationship between cloud cove and UV Index. Overall, the study sheds light on how environmental factors influence UV radiation.