7COM1079-0901-2024 - Team Research and Development Project

Final report title: What is the effect of sunshine duration on global radiation levels, and what correlations emerge between these variables?

Group ID: A254

Dataset number: DS312

Prepared by: Srinivasa Reddy Pulyala [23096195]

Dhanunjaya Sai Sudheer Karri [23073110]

Abhilash Komatineni [23011784]

Anudeep Reddy Yeruva [23101965]

Samuel Heath [20035949]

University of Hertfordshire

Hatfield, 2024

Contents

[1 Introduction 3](#_Toc186724988)

[1.1 Problem statement and research motivation 3](#_Toc186724989)

[1.2 The data set 3](#_Toc186724990)

[1.3 Research question. Explain how you are going to answer your RQ. 3](#_Toc186724991)

[1.4 Null hypothesis and alternative hypothesis (H0/H1) 3](#_Toc186724992)

[2 Background research 4](#_Toc186724993)

[2.1 Research papers 4](#_Toc186724994)

[2.2 Why RQ is of interest 4](#_Toc186724995)

[3 Visualisation 5](#_Toc186724996)

[3.1 Appropriate plot for the RQ output of an R script 5](#_Toc186724997)

[*3.2* Additional information relating to understanding the data 6](#_Toc186724998)

[*3.3* Useful information for the data understanding 6](#_Toc186724999)

[4 Analysis 7](#_Toc186725000)

[*4.1* Statistical test used to test the hypotheses and output 7](#_Toc186725001)

[4.2 The null hypothesis is rejected /not rejected based on the p-value 7](#_Toc186725002)

[5 Evaluation – group’s experience at 7COM1079 7](#_Toc186725003)

[5.1 What went well 7](#_Toc186725004)

[5.2 Points for improvement 8](#_Toc186725005)

[5.3 Group’s time management 8](#_Toc186725006)

[5.4 Project’s overall judgement 8](#_Toc186725007)

[5.5 Comment on the GitHub log output (50 words) 8](#_Toc186725008)

[6 Conclusions 9](#_Toc186725009)

[6.1 Results explained 9](#_Toc186725010)

[6.2 Interpretation of the results 9](#_Toc186725011)

[6.3 Reasons and/or implications for future work, limitations 9](#_Toc186725012)

[7 Reference list 10](#_Toc186725013)

[8 Appendices 11](#_Toc186725014)

[*8.1* R code used for analysis and visualisation 11](#_Toc186725015)

[8.1.1 Visualisation.R 11](#_Toc186725016)

[8.1.2 Analysis.R 11](#_Toc186725017)

[8.2 GitHub log output. 11](#_Toc186725018)

# Introduction

## Problem statement and research motivation

The relationship between sunshine duration and global radiation is understood as critical for optimising solar energy systems and planning for renewable energy. Although solar energy carries high potential, real-time correlations are needed in under-researched areas (Abdelatif et al., 2023). To that end, this research extends this gap by analysing the correlation between the duration of sunshine and global radiation using the London weather dataset. The findings hope to improve renewable energy applications, promote sustainable development, and help the move towards clean energy solutions globally through predictive modelling for those who suffer from varying climatic conditions.

## The data set

This dataset contains weather data: cloud cover, sunshine hours, global radiation and temperature metrics. Gen2 tracks the daily variation in order to investigate correlations, especially with sunshine and global radiation. Interval variables such as precipitation, pressure and snow depth span data. Useful for climate studies and predictive modelling, this London weather dataset (ID: To understand seasonal weather patterns, we utilise DS312.

## Research question. Explain how you are going to answer your RQ.

* Is there a correlation between Global radiation (global\_radiation) and Sunshine (sunshine)?

To answer the research question, a Spearman correlation test is conducted to determine the monotonic relationship between global radiation and sunshine duration. Using the London weather dataset, scatterplots and histograms validate and further illustrate this correlation.

## Null hypothesis and alternative hypothesis (H0/H1)

* The Null hypothesis (H0): There is no relation between global\_radiation and sunshine.
* Alternative hypothesis(H1) : There is a correlation between global\_radiation and sunshine.

# Background research

## Research papers

Some research studies involve the dataset which, on sunshine duration and global solar radiation, is of paramount importance. In research from Abdelatif et al. (2023), daily global solar radiation (GHI) was predicted based on sunshine duration data in Ghardaïa, Algeria, using empirical, polynomial, and Support Vector Regression (SVR) models with R² values of 0.97 with 3–4% error. In Aryani et al. (2023), NASA's 20-year climatological dataset was used to model solar irradiance for Eastern Indonesia using Artificial Neural Networks (ANN) that had high accuracy and provided spatial maps of solar energy potential for 174 districts. Lokoja City's solar energy potential was evaluated by Ikotoni et al. (2019) using meteorological data for the period 1995–2014 through the use of the Hargreaves–Samani and Angstrom–Prescott model. The study found an annual solar radiation between 3.54 and 4.88 kWh/m2/day. They point to a wide array of methodologies and applications for similar data sets, including machine learning methods like ANN and SVR, as well as many empirical approaches, as evidence of the value of the data set in renewable energy planning, climate modelling, material, and regional solar potential analysis. Taken together, the findings collectively emphasise the importance of sunshine duration and global solar radiation data for sustainable energy research.

## Why RQ is of interest

The research question encompasses the relationship between global radiation and sunshine duration to fill a gap in the knowledge of daily scale interactions of different climatic regions. Studies like Abdelatif et al. (2023) and Aryani et al. (2023) focus on advanced modelling, but few explorations in under-researched fields it is done based on real-time correlation. Future work incorporates model refinement, inclusion of more meteorological variables, and extension of the application to improve solar energy systems. Closing this gap will further the science of renewable energy planning and predictive climate modelling.

# Visualisation

## Appropriate plot for the RQ output of an R script

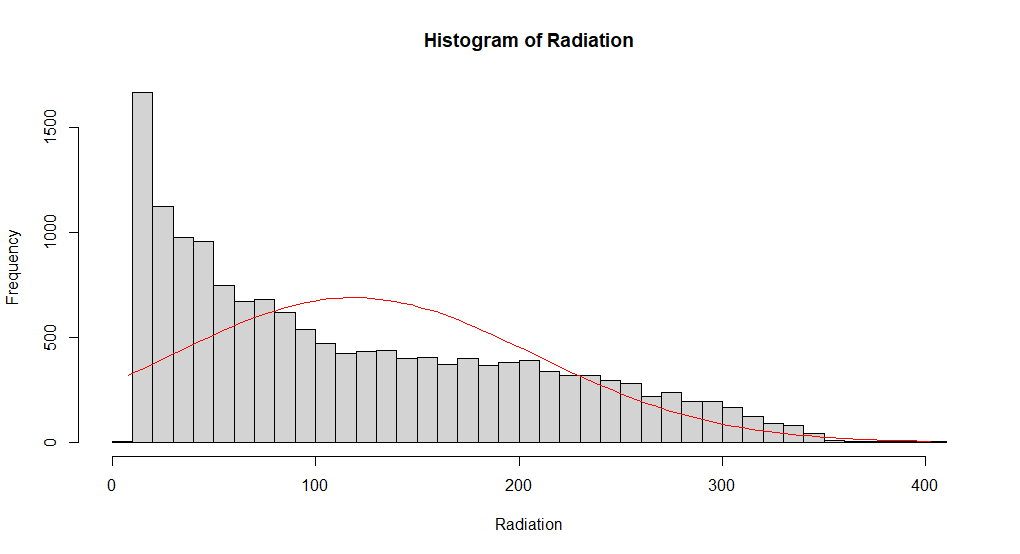


Figure 1: Histogram of Radiation

A graph of a sun shine

Description automatically generated with medium confidence

Figure 2: Scatterplot of Sunshine vs Radiation

The graph of sunshine duration and radiation provides an aggregation of radiation values to visualise frequency and skew, while the histogram of radiation shows the correlation between sunshine duration and radiation. The plots show the structure of the data and reveal how that structure relates to the research question.

## Additional information relating to understanding the data

The radiation value distribution is displayed in the histogram in terms of skewed to the right. Lower radiation levels are found to occur fairly often, and their frequency, coupled with common radiation ranges, along with data variability, provide useful perspectives on solar energy potential.

The correlation between sunshine duration and radiation is shown to be positive on the scatterplot. Their relationship is shown in the upward trend, which helps in validating the hypotheses of the influence of sunshine duration on global radiation.

## Useful information for the data understanding

As shown in the histogram, radiation is right-skewed in that it predominantly takes place at lower levels. Sunshine duration and radiation are shown on a scatterplot with a strong positive correlation confirming that greater sunshine duration tends to mean higher radiation.

# Analysis

## Statistical test used to test the hypotheses and output

A Spearman correlation test was used to determine the monotonic relationship of global radiation with duration of sunshine. The data contained ties, and normality assumptions are not required, so this test is suitable.

## The null hypothesis is rejected /not rejected based on the p-value

Spearman correlation test results show a strong positive correlation (ρ = 0.834) between global radiation and sunshine duration at a highly significant p-value (< 2.2e-16). With this, we have enough evidence to reject the null hypothesis (H0) and confirm the alternative hypothesis (H1), which is that there is a significant relationship between global radiation and sunshine. They confirm that global radiation also tends to increase as sunshine duration increases, suggesting that sunshine duration should be used as a predictor of global radiation.

# Evaluation – group’s experience at 7COM1079

## What went well

The group did an excellent job of collaborating on the technical side, taking advantage of individual coding skills and statistical analytical skills. Scatterplots and histograms, as examples, were skillfully created, creating intrigue in the data. Spearman correlation test was run without any hassle, and it yielded significant results. Everyone contributed meaningfully so that the analysis was complete. Communication and task allocation were strong, leading to an accurate and well-documented output by this group.

## Points for improvement

The task was completed successfully, but there are ways to improve. Individual responsibilities would have been clarified earlier in order to minimise redundancy in tasks. Furthermore, exploratory data analysis before statistical tests could have also revealed other insights. Had a group looked at the code before submitting it might have avoided inconsistency and failure to adhere to best practices. Discussion time with the dataset could have allowed the group more time to understand its intricacies.

## Group’s time management

Effective time management and task distribution were achieved equally among the members. Coding and analysis took place on time, but additional time would have allowed for improving final visualisations and double-checking the interpretation of results.

## Project’s overall judgement

The statistical techniques combined with strong coding skills were used on a group task, and it was a success. Teamwork and analytical rigour were very well demonstrated in the final output which adequately tackled the research objectives. Although additional work would refine the process further, the project, all in all, greatly exceeded expectations.

## Comment on the GitHub log output (50 words)

**Commit Message**: [visualization] In this commit, we used R code to filter the weather data and generate a scatterplot and histogram.

**Commit Message**: [analysis] In this commit, we performed analysis on our data using Spearman's rank correlation from the values we collected.

**Commit Message**:[dataset added] In this commit, we uploaded the dataset file that we used for the project.

# Conclusions

## Results explained

Finally, Spearman Correlation test results show a very strong positive correlation (ρ = 0.834) between global radiation and sunshine duration with a clearly significant p-value (< 2.2e-16). The fit good verifies a clear relation between the two variables, showing that the availability of sunshine duration correlates with high levels of global radiation, as has been previously indicated in studies about the potential of solar energy.

## Interpretation of the results

Results show that the research question is confirmed since sunshine duration has a significant relationship with global radiation. Predicting solar energy potential and renewable energy system planning critically depends on this relationship. This means better, cheaper renewable energy design and solar power harvesting for the population. In a general sense, the results support global efforts to achieve sustainable energy systems by emphasising the need to understand the dynamics of sunshine radiation.

## Reasons and/or implications for future work, limitations

Further work will need to refine models by incorporating other meteorological variables, such as humidity and wind speed. One limitation of the study is in its examination of a single dataset making generalisation impossible. Findings are applicable to broader datasets and can be tested in real-time.

# Reference list

Abdelatif, T., Samia, H., Redha, Y., Mohammed, L. and Amar, B., 2023, November. Prediction of Daily Global Solar Radiation in All Types of Sky Based on Sunshine Duration. In *2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM)* (Vol. 1, pp. 1-5). IEEE.

Aryani, D., Pranoto, S., Fajar, F., Intang, A.N. and Rhamadhan, F.Z., 2023, March. Artificial Neural Network Prediction to Identify Solar Energy Potential In Eastern Indonesia. In *2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA)* (pp. 252-257). IEEE.

Ikotoni, B.F., Ibrahim, H., Jatto, I.Y. and Olutade, I.O., 2019, August. Solar Radiation Potential Using Surface Meteorological Data:(A Case Study of Lokoja City, Kogi State). In *2019 IEEE PES/IAS PowerAfrica* (pp. 723-728). IEEE.

# Appendices

## R code used for analysis and visualisation

### Visualisation.R

d <- read.csv("london\_weather.csv")

colnames(d)

df <- d[!is.na(d$global\_radiation), ]

plot(df$sunshine, df$global\_radiation,

xlab = "Sunshine", ylab =

"Radiations", main

= "Scatterplot of SunShine vs Radations")

abline(lm(df$global\_radiation ~ df$sunshine), col = "red")

# Create a Histogram

h <-hist(df$global\_radiation, breaks = 30, xlab = "Radiation", ylab = "Frequency", main = "Histogram of Radiation")

# Add a normal distribution line

x <- seq(min(df$global\_radiation), max(df$global\_radiation), length = 100)

y <- dnorm(x, mean = mean(df$global\_radiation), sd = sd(df$global\_radiation)) \* length(df$global\_radiation)

box.size <- diff(h$mids[1:2])

y <- y \* box.size

lines(x, y, col = "red")

### Analysis.R

d <- read.csv("london\_weather.csv")

colnames(d)

df <- d[!is.na(d$global\_radiation), ]

# Perform Spearman correlation test

spearman\_test <- cor.test(df$global\_radiation, df$sunshine, method = "spearman")

# Print the results

print(spearman\_test)

## GitHub log output.

git log

commit 473bd001454be8ce273bf164f9d6b4d16749f164 (HEAD -> main, origin/main, origin/HEAD)

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Jan 7 12:18:32 2025 +0000

final report added

commit e69a5a77b47003e1459567d552b981cc70093436

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Sat Jan 4 20:01:37 2025 +0000

dataset added

commit c9b584a7e43aefbe9498ed4a6ea80bc9d277b1b9

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Sat Jan 4 20:00:18 2025 +0000

somechanges

commit 7cf7b43115461e2660acf68d95137dad304879b2

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 26 16:36:24 2024 +0000

ss

commit 554c34786546073a1acc3f70eae3387a3513647f

Merge: 7501bb4 1a611d7

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Mon Nov 25 00:04:18 2024 +0000

ss

commit 7501bb4669ca17edff16013bd775964075b3a2b0

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Sun Nov 24 23:53:41 2024 +0000

project

commit 1a611d793188d05088e0aa9ee7347bb258acae08

Author: Samuel Heath <sh21aac@herts.ac.uk>

Date: Sun Nov 24 15:35:56 2024 +0000

Added presentation

commit 1dda158db544f5e77ef52d466b1f54b2fa1859b8

Author: sh21aac <sh21aac@herts.ac.uk>

Date: Sun Nov 24 15:19:27 2024 +0000

Delete histogram.R

commit 84179b384c94073b4666072fdb7370d23c0bd1bd

Author: dk24aak <dk24aak@herts.ac.uk>

Date: Sun Nov 24 12:12:15 2024 +0000

analysis

commit e81c333ec4ac8309669752f92a858356a719cca2

Author: abhilash komatineni <ak23ahx@herts.ac.uk>

Date: Sun Nov 24 17:26:19 2024 +0530

visualization

commit e99c6b1bcea37e833eadcacab434ac446435c475

Merge: 4de0514 e0707c5

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Sun Nov 24 11:43:41 2024 +0000

file

commit 4de05145dc7bd1c5b0655f589986901822bab7ca

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Sun Nov 24 11:42:01 2024 +0000

file

commit e0707c5ed3a31bcd57dfa55a69f92e20269671f2

Author: Samuel Heath <sh21aac@herts.ac.uk>

Date: Thu Nov 21 18:42:57 2024 +0000

Corrected y axis title

commit 3b438b4321fa692a5c7cef4531d4290031372acd

Author: Samuel Heath <sh21aac@herts.ac.uk>

Date: Thu Nov 21 07:22:34 2024 +0000

Added R code to create histogram

commit 62189650ff0de336d6d56c36fd08d5bd05423acb

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 19 18:33:01 2024 +0000

ss

commit f8964f1afb0d41c1f694b918f07bc94cd9598c4d

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 19 18:32:03 2024 +0000

jj

commit b22ec36ead12eea23aead3079e2f1093d03f00c5

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 19 18:20:14 2024 +0000

hello

commit c5d68f04b4d75565c0beefd60f68284ccbb36048

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 19 16:05:34 2024 +0000

hello

commit 40182d7d568e977c316b48a3d9c0abf504b4c004

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Tue Nov 19 12:27:17 2024 +0000

sri

commit 17002061141e8515dc81b2fa0b3391d4f59cd26e

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Fri Nov 15 16:17:14 2024 +0000

london

commit 1e415178525896685a8c11d4f6b807b57aa3cae2

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Fri Nov 1 13:23:56 2024 +0000

sri

commit bf055ad3c711b0486736666b9057f8f7662edbbc

Author: dk24aak <dk24aak@herts.ac.uk>

Date: Fri Nov 1 13:21:46 2024 +0000

hi

commit 758d3b8f6709a62039449a78b79f040ef3b7deb5

Author: Samuel Heath <sh21aac@herts.ac.uk>

Date: Fri Nov 1 13:08:50 2024 +0000

hello

commit bedc905850e003c977519e10bb01a686e2ebdb59

Author: srinivasa reddy <sp24ach@herts.ac.uk>

Date: Fri Nov 1 12:58:31 2024 +0000

sri

(END)