7COM1079-0901-2025 - Team Research and Development Project

Final report title: Is there a difference in the median hydropower generation between the countries Brazil and India from 1965 to 2018?

Group ID: B104

Dataset number: DS335

Prepared by: Indireshvar Venkatchalam (23076055)

University of Hertfordshire
Hatfield, 2025

Table of Contents

Table of Figures

1. Introduction

1.1 Problem statement and research motivation

Hydropower is a vital component of the renewable energy systems in the world, yet each country has different geography, infrastructure, policy, and development of hydropower. According to the BP Statistical Review of World Energy (BP, 2019), trends in hydropower play an important role in defining renewable energy strategies. Brazil and India depend on hydropower to meet their increasing energy demand. Nevertheless, their comparative study of their long-term generation pattern is not available. Examining the significant difference in median hydropower generation between these two countries may provide useful information on their energy patterns and planning needs

1.2 The data set

The dataset contains annual renewable energy data on hydropower, solar, wind, geothermal, and biofuel production of each country in the year 1965 to 2018. It is obtained in the BP Statistical Review of World Energy. The data is in the form of energy production in terawatt-hours, which can be compared between the countries and over time. To analyse hydropower generation over time, the data from Brazil and India were selected in this study.

1.3 Research question

This study focuses on the following research question: Is there a difference in the median hydropower generation between the countries Brazil and India from 1965 to 2018?
This research question helps to compare hydropower trends over the period and how each country generates this type of renewable energy source.

1.4 Null hypothesis and alternative hypothesis (H0/H1)

To determine whether Brazil and India have a significant difference in the median hydropower production from 1965 to 2018. The null hypothesis (H_0) states that there is no difference in the median hydropower generation between the countries Brazil and India from 1965 to 2018. In contrast, the alternative hypothesis (H_1) proposes that there is a difference in the median hydropower generation between the countries Brazil and India from 1965 to 2018. The findings of the statistical tests help to reject or accept the null hypothesis for an evidence-based comparison of hydropower generation trends in these two countries.

2. Background research

2.1 Research papers

Hydropower is one of the important renewable energy in the world, which provides predictable and low-carbon energy, but its production is largely different. Brandao et al. (2024) examined the Brazilian hydropower infrastructure and stressed that the strategies of operating the reservoirs will have to change as variable renewable resources such as wind and solar are becoming increasingly important components of the energy infrastructure, which is different in contexts where hydropower is a significant part of the energy infrastructure.

The article by MDPI Pandit (2023) summarizes India's hydropower and its potential to reduce asymmetry in energy supply, leading to sustainable growth, although there are problems that exist, such as infrastructure and climate variability.

ScienceDirect Bortoluzzi et al. (2022) use data envelopment analysis and neural networks to evaluate the environmental impacts of Brazilian hydro projects, which provides information on the efficiency and regional effects of hydroelectric generation. IDEAS/RePEc All these studies combine to demonstrate a wide range of analyses of hydropower, with some being operational optimization and others being environmental and sustainability. However, all are lacking in comparison with median hydropower generation among countries such as Brazil and India, itself suggesting a gap in comparative statistical analysis of generation trends.

2.2 Why RQ is of interest

Although research has been conducted on optimizing operations in Brazil and exploring the potential of hydropower in India, it does not provide statistical data on the distribution of hydropower output in various decades.

This gap is filled by comparing the median hydropower generation between Brazil and India, which examines the significant difference in the dependence on and development of hydropower. The analysis is useful for policymaking and understanding the historical patterns of energy transition in various countries, which may be useful for future planning of renewable energy and mitigation of climate change.

3. Visualisation

3.1 Appropriate plot for the RQ

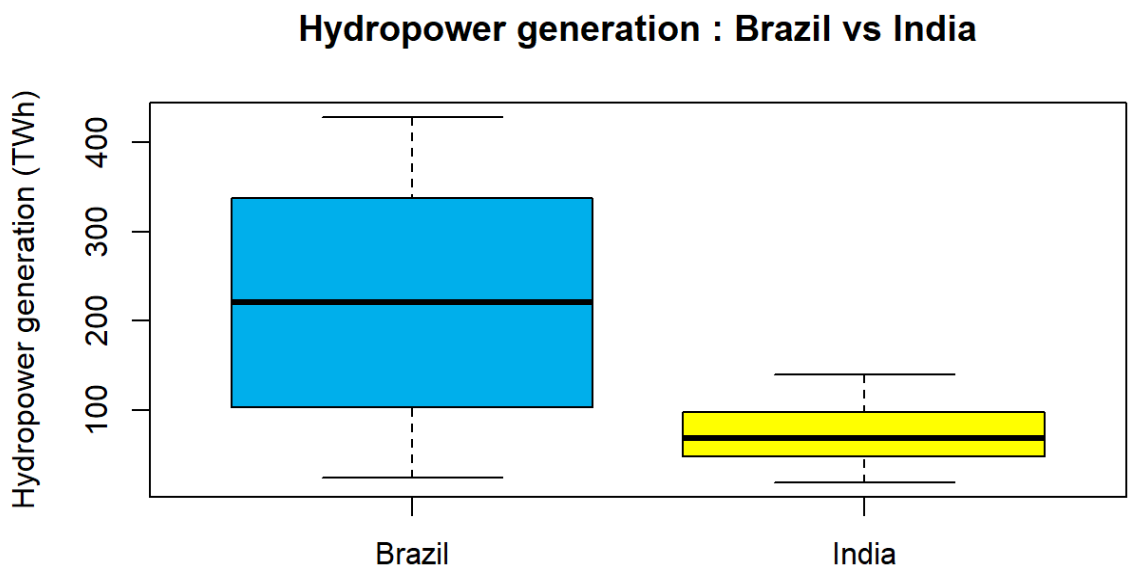

Figure 3.1. Box plot for the comparison between Brazil and India

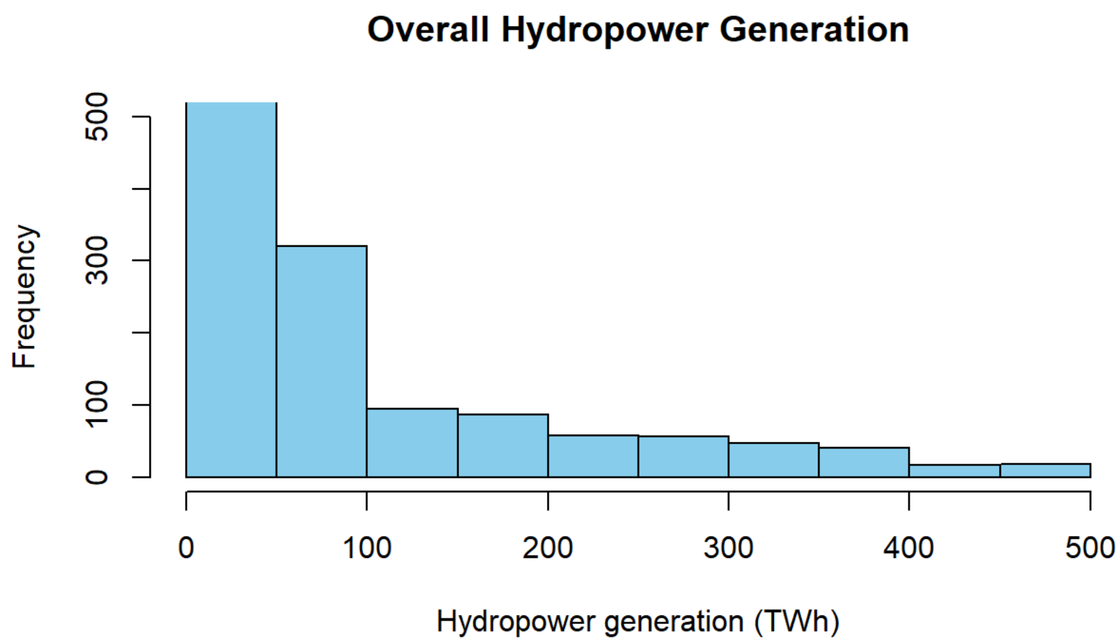

Figure 3.2. Histogram for checking the normality of overall hydropower generation

A boxplot is selected because it clearly visualizes the median, interquartile range, and spread of hydropower generation in Brazil and India, which addresses this research question. The Wilcoxon test was selected based on the shape of the histogram distribution, normality, and the non-parametric test was selected because of skewness.

3.2 Additional information relating to understanding the data

The boxplot indicates that Brazil has a higher median hydropower generation as compared to India, and a broader interquartile range with more fluctuations. The Histogram shows that the distribution is right-skewed. The results of these visualizations justify using a non-parametric test. This visually supports the use of the Wilcoxon test.

3.3 Useful information for the data understanding

The boxplot shows significant differences in central tendency, especially median and dispersion of hydropower production of Brazil and India.
Brazil's box size is large, indicating large fluctuations over the period, whereas in India, it is small, it does not fluctuate, and it is stable.

4. Analysis

4.1 Statistical test used to test the hypotheses and output

The Wilcoxon Signed-Rank Test was used to compare hydropower generation between Brazil and India because the distribution does not follow a normal distribution, as evidenced by histogram shapes and normality checks. This non-parametric test is selected for skewed and non-normally distributed values. It helps in identifying whether there is a significant difference in the median between the two countries. It provides the p-value, which is important for this analysis.

4.2 The null hypothesis is rejected based on the p-value

The null hypothesis states that there is no difference in the median hydropower generation between the countries Brazil and India from 1965 to 2018. The null hypothesis is rejected, according to the Wilcoxon Signed-Rank Test(1.67×10^{-10}), the p-value is much less than the 0.05 significance level. This shows that there is a statistically significant difference in the hydropower generation between the two countries. In particular, the findings indicate that a particular country will always produce more hydropower than the other country over the years of observation.

5. Evaluation – group's experience at 7COM1079

5.1 What went well

Although the project became an individual effort as the group was not active, communication was also broken down, leading to delays in the project's key stages. Completing those phases improved my understanding of R, data cleaning, statistical tests, hypothesis analysis, working on Git, and Visualization techniques without any obstacles. There were many feedback and support sessions available to clear the doubts and understand each concept, which were helpful during the project completion.

5.2 Points for improvement

The main issue is the workload for completing the project, as the limited participation of team members poses this problem. If there is shared responsibility among team members, it would benefit from incorporating different ideas, enabling better planning and a collaborative environment. If there is proper communication and engagement among the team members, then the time spent on the project would be reduced, thereby resolving technical challenges and finishing quickly.

5.3 Group's time management

The group time management was negatively affected by the lack of engagement of team members in the project work. This creates additional pressure, and since the tasks were not shared, I had to complete the tasks in tighter deadlines. Despite this, I organised the work effectively and finished on time.

5.4 Project's overall judgement

The group time management was negatively affected by the lack of engagement of team members in the project work. This creates additional pressure since the tasks were not shared, which puts pressure on completing the tasks in tighter deadlines. Despite this, I organised the work effectively and finished on time.

5.5 Amendments in the group arrangement

Due to non-engagement and limited participation of team members make the project functions as an individual assignment. A new GitHub repository was created to contribute to the project workflow of each process. The entire GitHub activity is under one member, and it represents the entire project work. There are no new or amended GitHub IDs done on this project work.

5.6 Comment on GitHub log output

The full GitHub log and the three commit messages are included in the appendix B.

1. Commit Message: [Change histogram with filtered data and adjusting bins structure]
   This commit improved the clarity of the visualization as the bin size is reduced to 50 TWh, thereby there is no distortion.

2. Commit Message: [Perform non-parametric statistical (Wilcoxon rank sum test) for comparison of median]
This commit introduces the main statistical test to analyse the hypothesis and to answer the research question

3. Commit Message: [Implement boxplot comparing hydropower generation between Brazil and India]
This commit shows a visual representation for comparing the median of two countries.

6. Conclusions

6.1 Results explained.

The analysis indicated that Brazil and India had significant outcomes. The patterns were not normally distributed, and the visualization showed a strong right skewness. The Wilcoxon Signed-Rank test showed a statistically significant difference with a p-value (1.67×10^{-10}) less than 0.5. The hydropower production has more fluctuation and is higher in Brazil, and India had a lower value but was constant throughout the period of time.

6.2 Interpretation of the results

The smaller p-value indicates that there is a difference in the median hydropower of Brazil and India, providing an answer to the research question. The median hydropower generation of Brazil is higher than India. The implications of these findings on the energy policy, resource management, and sustainability planning are significant in each country. The knowledge of such differences assists in the national energy strategies and demonstrates the necessity of renewable energy development in specific approaches.

6.3 Reasons and implications for future work, limitations

The shortcomings are that it considers only two countries and uses consolidated yearly data. In the future, the number of countries, capacity factors, or temporal trend modelling could be increased. Socioeconomic and environmental factors could enhance the knowledge of the hydropower dependence and performance, and provide more insights.

7. Reference list

Brandão, A., de Lima, L., Colla, V. and de Souza, A. (2024) *Hydropower reservoir operation under increasing penetration of variable renewable energy sources: A case study in Brazil*. *Energies*, 17(13), 3339.

Pandit, D. (2023). *A comprehensive overview of the hydropower potential in India: Energy matrix and sustainable growth perspectives*. *Heliyon*, 9(12), e13478.

Bortoluzzi, G., e Silva, A. and Gomes, R. (2022) *Environmental efficiency evaluation of hydroelectric power plants using data envelopment analysis and artificial neural networks*. *Renewable Energy*, 200, pp. 1316–1326.

BP (2019) *Statistical Review of World Energy 2019*. Available at: https://www.bp.com/statisticalreview (Accessed: 6 December 2025)

Gulagi, A., Bogdanov, D., and Breyer, C. (2017). *The role of Brazil and India in global renewable energy development*. *Energy Strategy Reviews*, 16, pp. 33–45.

de Faria, F.A.M., Davis, M.A.S., Seabra, J.E.A. and Macedo, I.C. (2017). *The contribution of hydropower to Brazil's energy security and sustainability*. *Renewable and Sustainable Energy Reviews*, 70, pp. 658–670.

8. Appendices

A. R code used for analysis and visualisation.

```
library(tidyverse)

#load dataset

modern_renewable_energy_consumption <- readr::read_csv(
  "modern-renewable-energy-consumption.csv"
)

# Inspect structure and missing values

str(modern_renewable_energy_consumption)

print(colSums(is.na(modern_renewable_energy_consumption)))

numeric_cols <- modern_renewable_energy_consumption %>%
  select(where(is.numeric)) %>%
  names()
numeric_cols

#Median imputation for data cleaning

data_clean <- modern_renewable_energy_consumption %>%
  group_by(Entity) %>%
  mutate(
    across(
      all_of(numeric_cols),
      ~ tidyr::replace_na(.x, median(.x, na.rm = TRUE))
    )
  ) %>%
  ungroup()

cat("\nMissing values per column:\n")
print(colSums(is.na(data_clean)))

data_clean <- data_clean %>%
  mutate(
    across(
      all_of(numeric_cols),
      ~ tidyr::replace_na(.x, median(.x, na.rm = TRUE))
    )
  )

cat("\nMissing values per column after data cleaning\n")
print(colSums(is.na(data_clean)))
```

```
# Extract hydropower data for Brazil and India

hydro_data <-
  data_clean %>%
  select(
    Country = Entity,
    Year,
    `Hydro Power Generation` = `Hydropower (terawatt-hours)`
  ) %>%
  filter(Country %in% c("Brazil", "India"))

print(hydro_data %>% head())

#Convert from long to wide format

hydro_wide <-
  hydro_data %>%
  spread(key = Country, value = `Hydro Power Generation`) %>%
  arrange(Year)

colnames(hydro_wide)[colnames(hydro_wide) == "Brazil"] <- "Brazil_Hydro"
colnames(hydro_wide)[colnames(hydro_wide) == "India"] <- "India_Hydro"

print(hydro_wide %>% head())

#Summary Statistics

cat("\nBasic Summary: Brazil\n")
print(summary(hydro_wide$Brazil_Hydro))

cat("\nBasic Summary: India\n")
print(summary(hydro_wide$India_Hydro))

summary_stats <- function(x) {
  data.frame(
    Min      = min(x, na.rm = TRUE),
    Q1       = quantile(x, 0.25, na.rm = TRUE),
    Median   = median(x, na.rm = TRUE),
    Mean     = mean(x, na.rm = TRUE),
    Q3       = quantile(x, 0.75, na.rm = TRUE),
    Max      = max(x, na.rm = TRUE),
    SD       = sd(x, na.rm = TRUE),
    Range    = max(x, na.rm = TRUE) - min(x, na.rm = TRUE)
  )
}

Brazil_summary <- summary_stats(hydro_wide$Brazil_Hydro)
India_summary  <- summary_stats(hydro_wide$India_Hydro)
```

```
cat("\nDescriptive Statistics: Brazil\n")
print(Brazil_summary)

cat("\nDescriptive Statistics: India\n")
print(India_summary)

# Overall hydropower generation histogram

x <- data_clean$`Hydropower (terawatt-hours)`
x <- x[x <= 500]
hist(x,
     freq = TRUE,
     main = "Overall Hydropower Generation",
     xlab = "Hydropower generation (TWh)",
     ylab = "Frequency",
     col = "skyblue",
     border = "black",
     ylim = c(0, 500),
     xlim = c(0, 500),
     breaks = seq(0, 500, 50),)

#Also checking the normality based on individual countries

x <- hydro_wide$Brazil_Hydro
y <- hydro_wide$India_Hydro

hist(x,
     freq = FALSE,
     main = "Brazil Hydropower generation with Normal Curve",
     xlab = "Hydropower generation (TWh)",
     col  = "skyblue")
curve(dnorm(x, mean = mean(x, na.rm = TRUE), sd = sd(x, na.rm = TRUE)),
      add = TRUE, lwd = 2)

hist(y,
     freq = FALSE,
     main = "India Hydropower generation with Normal Curve",
     xlab = "Hydropower generation (TWh)",
     col  = "skyblue")
curve(dnorm(x, mean = mean(y, na.rm = TRUE), sd = sd(y, na.rm = TRUE)),
      add = TRUE, lwd = 2)

#Wilcoxan rank sum test for hypothesis analysing

wtest <- wilcox.test(x, y, paired = TRUE)
wtest
wtest$statistic   # W value
wtest$p.value     # p-value
```

```
cat("\n Wilcoxon Signed-Rank Test: W =", wtest$statistic,
   "p-value =", wtest$p.value, "\n")

#Visualize and compare the hydropower generation using boxplot

boxplot(hydro_wide$Brazil_Hydro,
        hydro_wide$India_Hydro,
        names = c("Brazil", "India"),
        ylab  = "Hydropower generation (TWh)",
        main  = "Hydropower generation : Brazil vs India",
        col = c("deepskyblue2", "yellow")
```

B. GitHub log output.

```
commit 76368506d3ab66190dcbda2e2311dbf7fa601945
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Fri Dec 12 11:17:06 2025 +0000

    Added the previous git log

commit ae5250f9e36ff6a9dfe13ca36adb9fbd2f28bc09
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Fri Dec 12 11:14:52 2025 +0000

    Added the visualization plots as images

commit 870b596b41db03b33d7f988b3413bee2705a6d7e
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Fri Dec 12 11:10:12 2025 +0000

    Added the Rscript.log file and Rplots

commit d69b7ec0efb7ea8fc63554fdff2dc7a01086d300
Merge: fde0738 3ff0b4b
Author: Indireshvar Venkatchalam <iv24aag@herts.ac.uk>
Date:   Fri Dec 12 09:49:33 2025 +0000

    Merge pull request #4 from
indireshvar-venkatachalam/analysing-statistical-test-and-visualization

    The non parametric statistical test, wilcoxan rank sum test used to analyse the p value and
hypothesis and the box plot is used to visualize and compare hydropower generation between
Brazil and India.

commit 3ff0b4b660e0402d176d43efd5a4cc8c79527e27
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Fri Dec 12 09:45:46 2025 +0000
```

    Implement boxplot comparing hydropower generation between Brazil and India

commit bd0a7708c3dbc5cdf6da2aa7904dc9e9ff539434
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Fri Dec 12 09:38:08 2025 +0000

    Perform non-parametric statistical (Wilcoxan rank sum test) for comparison of median

commit fde07380b759c63414ceea58aa53bc7167d4ba45
Merge: d784d54 afb058e
Author: Indireshvar Venkatchalam <iv24aag@herts.ac.uk>
Date: Fri Dec 12 02:33:25 2025 +0000

    Merge pull request #3 from indireshvar-venkatachalam/hydropower-visualization

    The overall distribution of hydropower generation of brazil and India is visualized using histogram and normality is analysed .Also checked based on individual country.

commit afb058e034bab742ed9c63095f9cb3a55d164d9e
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Fri Dec 12 02:30:07 2025 +0000

    Also check the normality of each individual country based on hydropower generation

commit 4dcc7a51f77d0025fd03d9ded07c85c5c0972bfa
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 23:35:54 2025 +0000

    Change histogram with filtered data and adjusting bins structure

commit 439808a9d569543008d6f40102913f034c36920a
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 23:31:25 2025 +0000

    Histogram for overall hydropower generation

commit d784d54f1b32480efdd20fd02c883b13268342de
Merge: b2b9075 d62a615
Author: Indireshvar Venkatchalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 21:47:48 2025 +0000

    Merge pull request #2 from indireshvar-venkatachalam/summary-stats

    Added the basic summaries and descriptive statistics for the countries Brazila and India for analysing hydropower generation trends.

commit d62a6154e8d45a40c13e52b753ad53e1a5cba07e
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 21:06:08 2025 +0000

    Displaying the descriptive statistics for countries Brazil and India

commit 20875135acd6d7d18e30d142fb9e335fa22a9e65
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 21:04:56 2025 +0000

    Implement detailed descriptive statistics for countries Brazil and India

commit 14a07663a73b67bc67a6641c6775b367890ec892
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 21:01:35 2025 +0000

    Printing the basic summaries for the countries Brazil and India

commit b2b907519e832a983791f7fa20c2cf6ab5509ba2
Merge: 79df636 23ae899
Author: Indireshvar Venkatchalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 20:59:41 2025 +0000

    Merge pull request #1 from indireshvar-venkatachalam/hydro-data-processing

    Hydro data processing

commit 23ae899f9c811fcc945a560f51b9a480f8e049ec
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 19:04:12 2025 +0000

    Displaying the few rows of reshaped dataset

commit aa66a71c7165764f6039b80f538b0c14aca81356
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 19:03:03 2025 +0000

    Renamed the columns

commit 1d4619ee05453d3c69ded8f0aadfbcde56316faa
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 19:01:51 2025 +0000

    Reshaped the data from long to wide format for analaysing

commit b5bb064df2140410b5f2e1a98a3aa4eecd2fe147
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date: Thu Dec 11 18:51:48 2025 +0000

    displaying few rows

commit 3bd58f6434cbc136d2198c6f03cd437356246fe7

```
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:50:10 2025 +0000

    Select the data needed and filtering countries

commit 918078e645af414aedbe5f9903923ad6cd27b6e6
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:46:08 2025 +0000

    printing columnwise

commit baf9ae67fa64379f2413311f35473cf61caba81f
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:45:00 2025 +0000

    further data cleaning

commit 37870c70c3d4ef95915699fea4f8b62a9b3133fa
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:43:41 2025 +0000

    printing columnwise

commit 5083e4b0bd708c93c7bd41edb6fe9c33297a4456
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:40:09 2025 +0000

    Data cleaning using median imputation

commit 421798be8b49ef2275ae5509df2999da3a507b15
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:32:42 2025 +0000

    Identify the numeric columns for analaysing

commit 45feec4748c56232f5d364849b00d725f616e29d
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:29:13 2025 +0000

    Identify the missing NA values in the dataset

commit 79df636c132be30e8f8d99fb11e98422d0d1233a
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 11 18:26:44 2025 +0000

    Checking the dataset file uploaded

commit 0d19fe1305dbecac7c877c700ed7ffde0cd91a7c
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
```

```
Date:   Thu Dec 4 20:38:58 2025 +0000

    Checked the dataset structure output

commit c41cefef7e3fded8de2a85e9445ac2aec4ff8cfb
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 4 20:34:28 2025 +0000

    some files removed

commit 2b060752717767e24147edd1d415246d86f596fa
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 4 20:18:19 2025 +0000

    Removed unwanted files added to repository

commit 3528ca5ad49e5ecd2ed070a961fc27307a0c6836
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 4 20:15:18 2025 +0000

    Loaded the dataset in R script

commit fc74b47a0791e9bb6efab94dcf164f875042fe6d
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Thu Dec 4 20:06:43 2025 +0000

    Added dataset file

commit c43e1e82feba0a517714b53110ef0c5e83bf1935
Author: indireshvar-venkatachalam <iv24aag@herts.ac.uk>
Date:   Tue Dec 2 23:01:12 2025 +0530

    Initial empty branch
```