# Introduction

## Finding Suitable Data Source

The data selected for this task was that of “Western Europe Power Consumption”. The data was publically available and downloaded from Kaggle at the URL: <https://www.kaggle.com/datasets/francoisraucent/western-europe-power-consumption>. The data obtained from the link was obtained from the ENTSO E-Transparency Platform, which provides access to such data as provided by various transmission system operators (TSOs) across Europe.

The data was time-series and had a resolution of 15 minutes, 30 minutes, or 1 hour depending on the country. The data spanned from January 2015 to August 2020. The units of measurement for the power consumption were in Megawatts (MW).

Multiple time-series spanning the same time period could be obtained by considering the power consumption of different countries in Western Europe. The data was also suitable since a single time series had thousands of values. Also, the data was relevant in the current space and the insights obtained from the data could be of importance.

## Research Questions

* How did power consumption in Britain, France, and Italy vary over time?
* What is the yearly power consumption over the study time in Britain, France, and Italy?
* What is the range of daily power consumption in Britain, France, and Italy?

# Importing the Data

The dataset for this assignment was in CSV format. Each country had its CSV file with the time-series data on power consumption.

To import the data into MATLAB, the “readtable()” function was used with the file path to a CSV file as the parameter. The function parsed the CSV file into a table in MATLAB. Using this function simplified work and avoided the writing of boilerplate code to handle file manipulation and data structure creation.

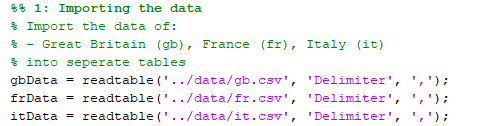


Figure : Code to import data into MATLAB

# Cleaning the Data

To clean the data, first, a feel of the data had to be obtained. To do this, the “summary()” function was used to obtain summaries of the three tables. The table of interest was passed as the parameter. A sample summary is shown in figure 2. The “head()” function was then used to get a glance of the structure and contents of each table. A sample head is shown in figure 3.

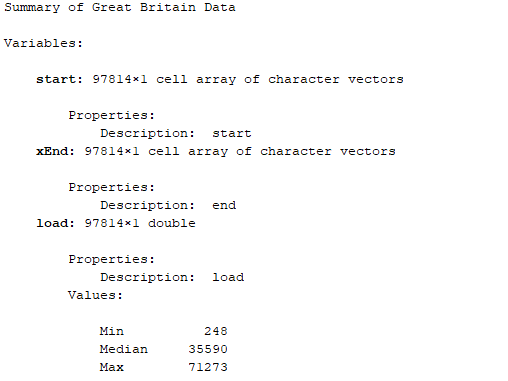


Figure : Summary of table hosting Britain data

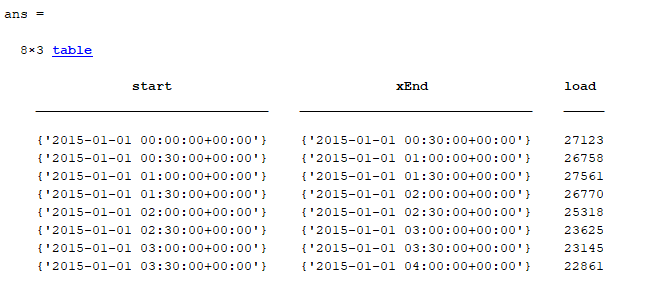


Figure : Head of the table hosting Britain data

From the summaries, it was noted that the tables didn’t have missing or NaN values. The head operation allowed the structure of the data to be seen. It was noted that the start and end timestamps within which the load consumption was recorded were contained in the table. It was also observed that the time resolution of the data in Britain was 30-minutes, while those of France and Italy were 1 hour. As such, some cleaning had to be done.

The ”start” column was removed from the tables using the “removevars()” function. In this assignment, it was desired that analysis be conducted daily. As such, the timestamp complete with time fields had to be converted to date only.

To do this, the date part from the timestamp were extracted using the “extractBefore()” function. The loads in each date were then summed using the “groupsummary()” function [1]. The dates, which were currently string objects were then converted to datetime objects using the “convertvars()” function [1]. Finally, the rows in the resulting table were sorted according to date using the “sortrows()” function.

The above operations were carried out for all three tables. At the end of that, there were three tables with dates and daily loads as fields. The three tables were joined into one using the “join()” function [2]. The date was used as the key in this operation. At the end of this, there was a single table with dates and loads of Britain, France, and Italy as shown in the table in figure 5.

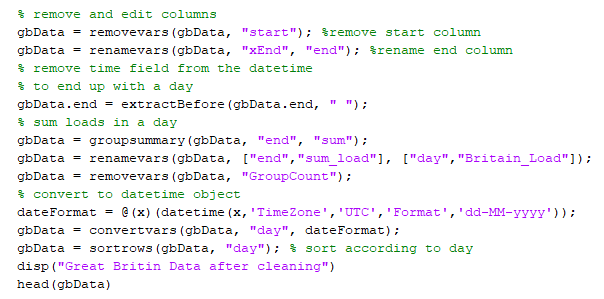


Figure : Sample code for cleaning data - Case Britain

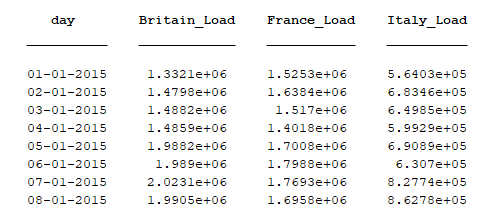


Figure : Resulting table after cleaning and joining tables

# Data Analysis

**Question: How did power consumption in Britain, France, and Italy vary over time?**

The line chart was seen to be a suitable tool to answer this question. The power consumption data had been already resolved to a daily resolution. This saw a reduction of the data points from 97814 values in Britain and 48936 values in France and Italy to 2040 values for each country. This step was necessary to reduce the computational time and resources needed to plot all those values into a single chart.

Next, a moving mean was applied to the daily data. The function “movmean()” was used to achieve this. A 20-day moving mean was found to be suitable for the data. The necessity of the function was to average the data, reducing the impact of outliers in the generation of the chart [3].

On plotting the data using, the chart shown in figure 6 was obtained.

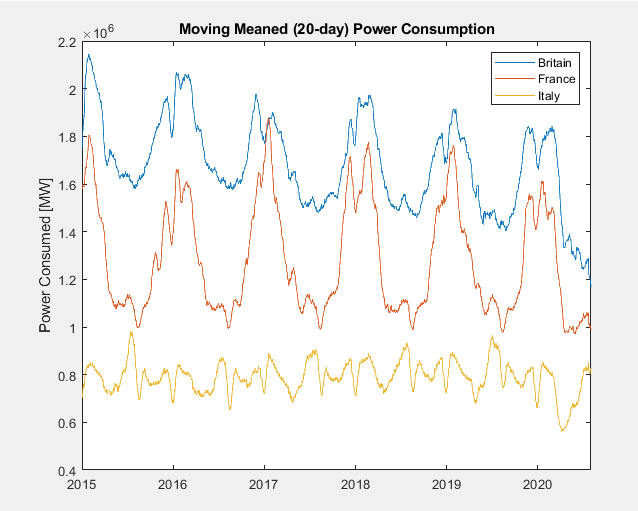


Figure : Power consumption through the years

It was observed that the power consumption in all countries was seasonal. The greatest power consumption was observed to be towards the end of the year and at the start of the succeeding year. This can be attributed to the high power consumption in the Winter season. The lowest power consumption was observed to be at the middle of the year – Summer.

From a glance, Britain was observed to be the country consuming the most power of the three. Italy was observed to be the country consuming the least amount of power.

**Question: What is the yearly power consumption over the study time in Britain, France, and Italy?**

The bar chart was deemed to be a suitable tool for answering this question. The power consumption table that had been produced from the last step in cleaning had a daily resolution. In this analysis, it was necessary to obtain the yearly power consumption.

The “day” column as shown in figure 5 was converted to string objects. The “extractAfter()” function was then used to extract the year portion from the date. The “groupsummary()” function was used to sum the loads in the same year to produce the yearly consumption. The rows for 2020 were removed since the data for this year wasn't complete. After removing unnecessary columns using the “removevars()” function and renaming the columns appropriately using the “renamevars()” function, the resulting table was displayed on a bar graph using the “bar()” function. The graph generated is shown in figure 9.

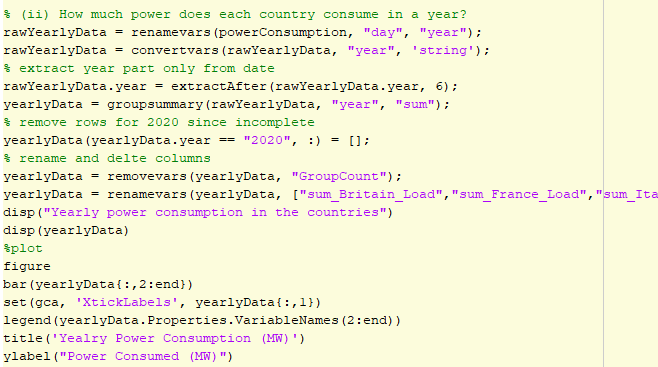


Figure : Code for plotting the bar graph

The yearly power consumption for Britain, France, and Italy is shown in the table in figure 8.

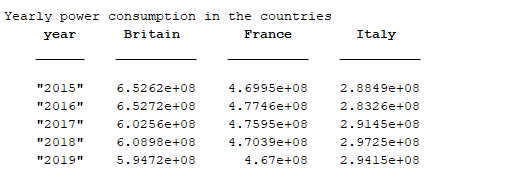


Figure : Table of yearly power consumption in Britain, France, and Italy

It was observed that Britain had the highest power consumption every year while Italy had the least power consumption.

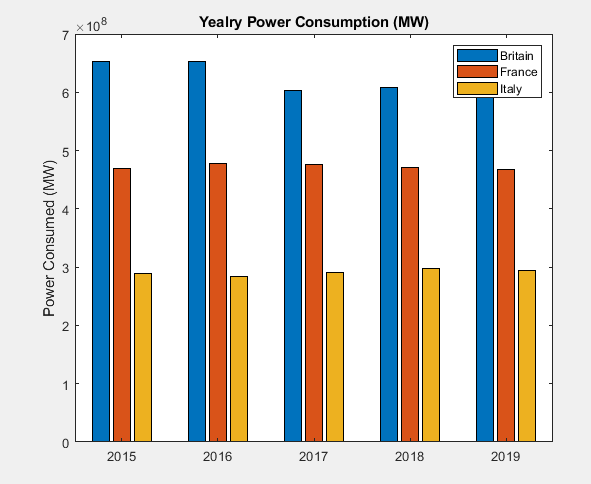


Figure : Bar graph of yearly power consumption in Britain France, and Italy

**Question: What is the range of daily power consumption in Britain, France, and Italy?**

This question was seen to be better answered empirically. The “range()” function was applied to the table containing daily power consumption. The results in figure 10 were obtained.

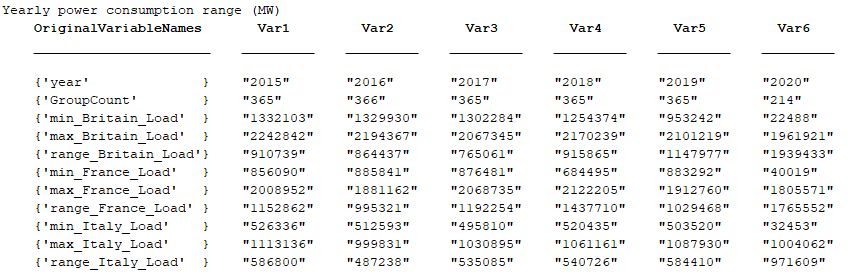


Figure : Range of power consumption in Britain, France, and Italy

# Conclusion

The purpose of this assignment was to make use of MATLAB for data analysis. Using the “readtable()” function, the importation of CSV data files into MATLAB as tables were found to be an easy task. The “summary()” function was found to be efficient in providing concise summaries of tables. It was discovered that MATLAB had useful functions for performing table manipulations – to rename columns, the “renamevars()” function was used. To drop columns, the “removevars()” function was used, and to convert datatypes, the “convertvars()” function was used. The “groupsummary()” function was useful in grouping data in tables, and the “join()” function was useful in joining multiple tables into one.

For this data analysis, it was observed that the power consumption in Britain, France, and Italy where seasonal. During the end and beginning of the year (Winter seasons), it was observed that the power consumption was highest. Likewise, the power consumption was least in the middle of the year during the Summer seasons. Britain was observed to consume the most power of the three countries, while Italy consumed the least power.

# References

|  |  |
| --- | --- |
| [1] | p. I. Kattan, MATLAB For Beginners, Petra Books, 2009. |
| [2] | A. Stormy, Matlab: A Practical Introduction to Programming and Problem Solving, Oxford: Butterworth-Heinemann, 2013. |
| [3] | R. Hyndman, "Moving averages," in *International Encyclopedia of Statistical Science*, Berlin, Springer, 200, pp. 866-869. |