7PAM2000 Applied Data Science 1

STUDENT NAME: FAISAL ZULFIQAR

Student ID: 22010044

Link: https://github.com/Faisal-Zulfiqar786/Rework-Assignment-1.git

ASSIGNMENT 1: VISUALISATION

DATASET

Dataset

Link: https://archive.ics.uci.edu/ml/machine-learning-databases/00235/household_power_consumption.zip

The dataset used is called "Household Power Consumption" and contains measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost four years. The data was collected in an individual household located in Sceaux, France, with an electrical monitoring system composed of a set of measurements taken every minute.

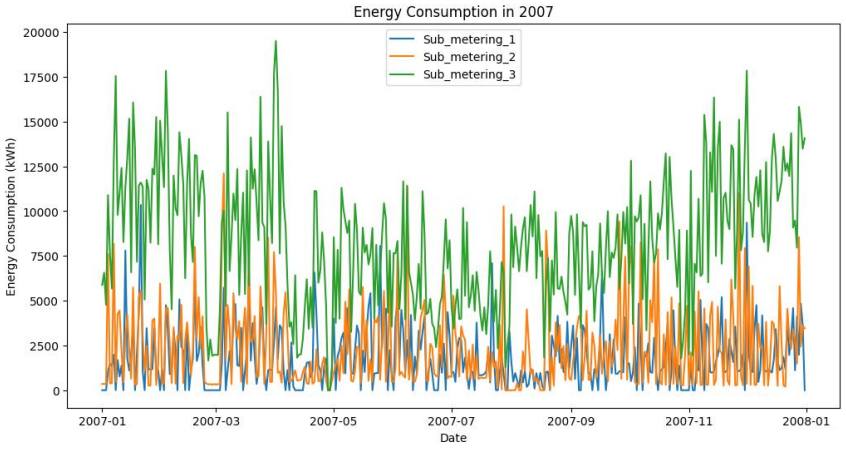
The dataset contains measurements from January 2007 to November 2010, and includes 2,075,259 instances. The data includes nine attributes:

Attribute Name	Description
Date	The date on which the power consumption was measured in the format dd/mm/yyyy
Time	The time at which the power consumption was measured in the format hh:mm:ss
Global Active Power	The total active power consumed by the household in kilowatts
Global Reactive Power	The total reactive power consumed by the household in kilowatts
Voltage	The voltage level in volts at the time of measurement
Global Intensity	The total current intensity in amperes at the time of measurement
Sub-metering 1	The active energy consumed by the kitchen in kilowatt-hours during the time interval between the current and the previous measurement
Sub-metering 2	The active energy consumed by the laundry room in kilowatt-hours during the time interval between the current and the previous measurement
Sub-metering 3	The active energy consumed by an electric water-heater and an air-conditioner in kilowatt-hours during the time interval between the current and the previous measurement

VISULAIZATION

MULTIPLE LINE PLOT

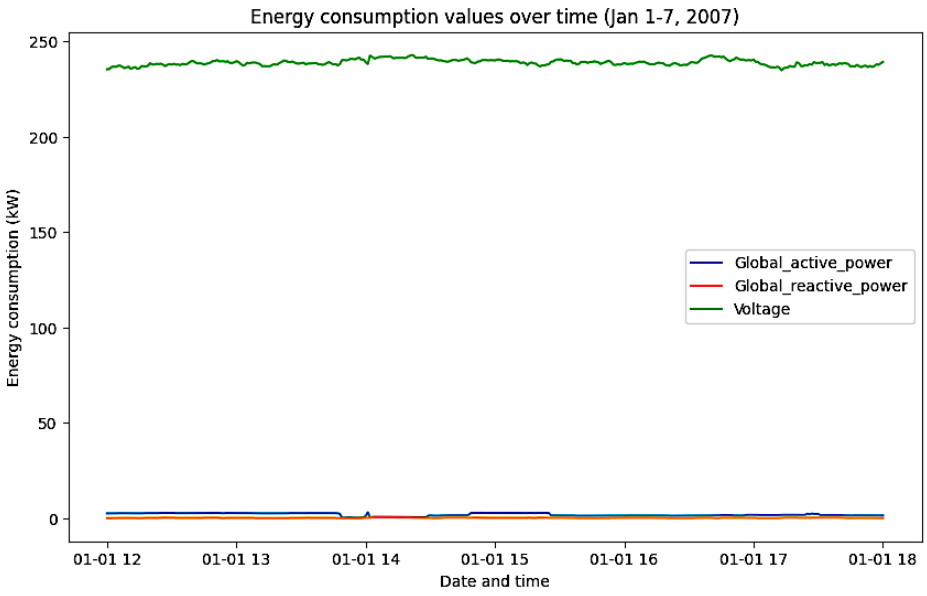
This code is reading in a dataset of household power consumption and creating a line plot that shows the energy consumption for three different sub-meters over the course of the year 2007. The code plots the sub-metering data for each day in 2007, with three different lines on the plot representing the three sub-meters.



LINE PLOT 1: ENERGY CONSUMPTION FOR THREE DIFFERENT SUB-METERS

The benefit of this plot is that it allows us to see how energy consumption varies over time for each of the three sub-meters. By visualizing the data in this way, we can identify patterns and trends in the data that might not be immediately apparent from looking at the raw data. Additionally, the plot makes it easy to compare the energy consumption for each sub-meter over the course of the year.

2nd line plot visualizes the changes in energy consumption values over time for different variables such as global active power, global reactive power, and voltage.

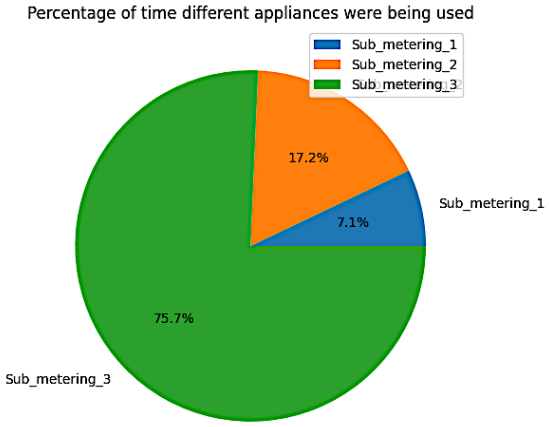


LINE PLOT 2: ENERGY CONSUMPTION VALUES OVER TIME FOR DIFFERENT VARIABLES

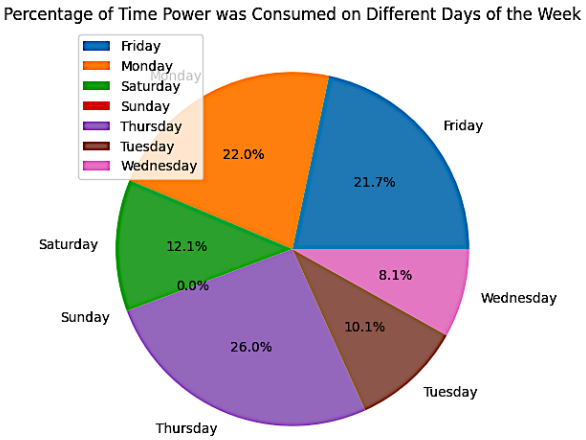
PIE CHART

7PAM2000 Applied Data Science 1

The code is using a dataset to create three pie charts using matplotlib. The first chart shows the percentage of time different appliances were being used. The chart gives us an idea that an electric water-heater and an air-conditioner consumed the most power during the first week of January 2007.



PIE CHART 1:PERCENTAGE OF TIME DIFFERENT APPLIANCES

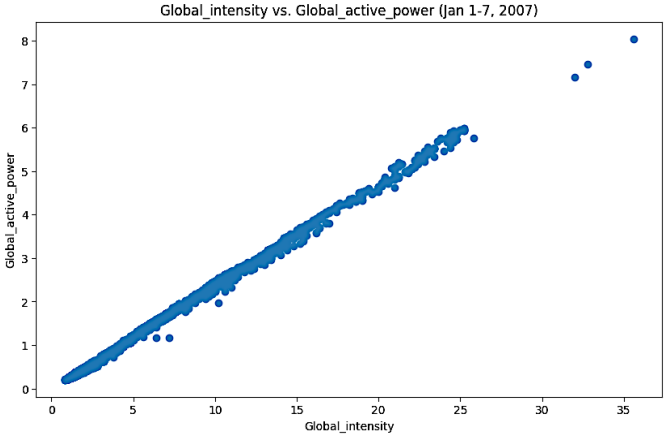


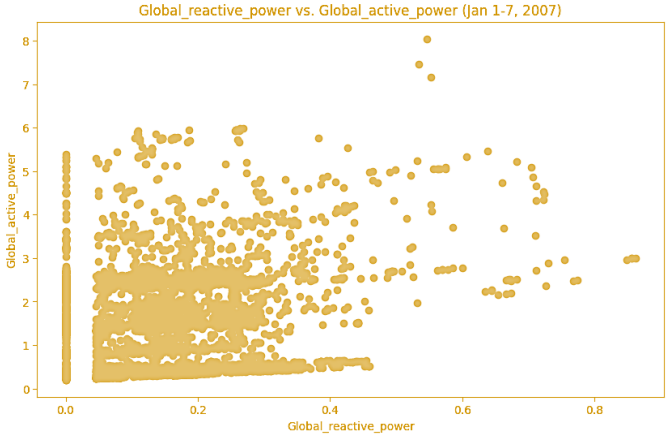
PIE CHART 2: PERCENTAGE OF TIME POWER WAS CONSUMED ON DIFFERENT DAYS OF THE WEEK

The second chart shows the percentage of time power was consumed on different days of the week. This pie chart shows that power consumption was highest on Thursday at 26% during the first week of January 2007.

SCATTER PLOT

By plotting the scatter plot of each feature variable against the output variable separately, we can visually inspect if there is any correlation or relationship between the two variables. If the points on the plot are spread out evenly and there is no obvious pattern, it suggests that the feature variable is not strongly correlated with the output variable. On the other hand, if the points are clustered or show a clear pattern, it suggests that the feature variable is related to the output variable. This information can be useful in understanding which features have a significant impact on the output variable and could be used in building a predictive model.





SCATTER PLOTS 1