

ADS506 Assignment 1.2 sp500_stocks

Gabriel E. Mancillas Gallardo

2024-11-07

Assignment: Propose a Time Series Dataset for Your Final Project ### Data Source

The data source is the S&P 500 stock data from Kaggle however because the data exceeds the 50MB capacity we will only include 2022 - 2024 dataset. The dataset contains the stock prices of the S&P 500 companies from 2010 to 2024. The dataset contains the following columns:

```
# Load necessary libraries
library(readr)
library(dplyr)
library(tsibble)

# Load AEP consumption data
aep_data <- read_csv("/Users/gabrielmancillas/Documents/GitHub/S-P-500/AEP_hourly.csv")

# Display the first few rows of the dataset
head(aep_data)
```

```
## # A tibble: 6 x 2
##   Datetime          AEP_MW
##   <dtm>             <dbl>
## 1 2004-12-31 01:00:00 13478
## 2 2004-12-31 02:00:00 12865
## 3 2004-12-31 03:00:00 12577
## 4 2004-12-31 04:00:00 12517
## 5 2004-12-31 05:00:00 12670
## 6 2004-12-31 06:00:00 13038
```

```
# Display the total number of rows and columns
dim(aep_data)
```

```
## [1] 121273      2
```

Data Source

Include public links to data if it is too large to upload (do not upload datasets larger than 50MB). **There is no need to upload the data (its under 40MB)** [github repository](<https://github.com/Gabeleo24/S-P-500/tree/main/data>)

Time Series Plot

```
# Check for duplicates
duplicates <- aep_data %>%
  group_by(Datetime) %>%
  filter(n() > 1)
```

```

# Remove duplicates if any
aep_data <- aep_data %>%
  distinct(Datetime, .keep_all = TRUE)

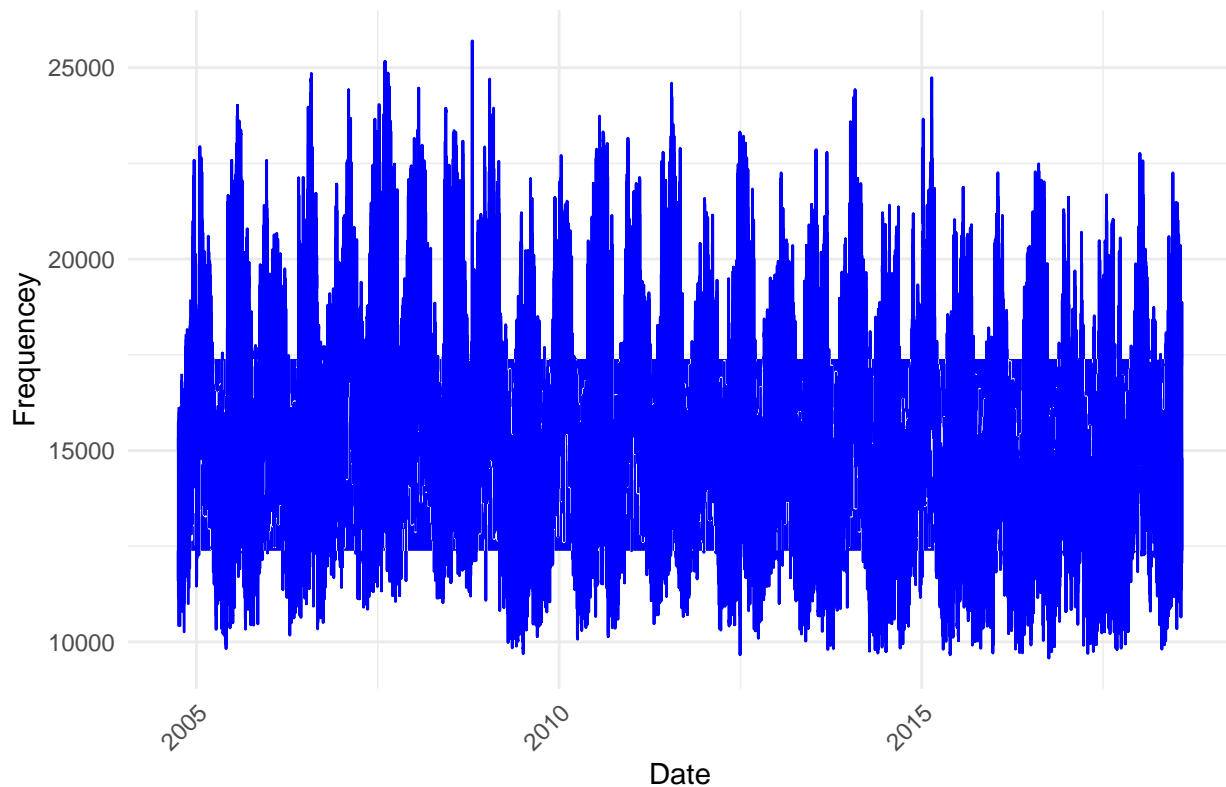
# Plot the time series
library(tsibble)
library(ggplot2)

# Convert to tsibble for time series structure
aep_data_ts <- aep_data %>%
  as_tsibble(index = Datetime)

# Time series plot for Adjusted Close prices
ggplot(aep_data_ts, aes(x = Datetime, y = `AEP_MW`)) +
  geom_line(color = "blue") +
  labs(
    title = "Hourly power consumption data from PJM",
    x = "Date",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

```

Hourly power consumption data from PJM



Discussion

The hourly power consumption data from PJM reveals notable fluctuations over time, capturing trends that reflect varying levels of energy demand influenced by seasonality, time of day, and economic factors. From 2005 onwards, we observe a periodic increase in energy demand during summer and winter months, likely driven by seasonal heating and cooling needs. These patterns highlight the strong influence of temperature and weather conditions on energy usage.

Energy demand also reflects broader economic activities. Periods of economic growth are often marked by increased industrial and residential energy consumption, whereas economic downturns can lead to reduced demand. This dataset offers insights into how external factors, such as economic slowdowns, regulatory changes, and shifts in energy efficiency practices, can directly impact consumption patterns.

Predicting future consumption trends from this dataset can provide valuable insights for energy providers, policymakers, and businesses. Accurate forecasts help ensure resource availability, support grid stability, and inform pricing strategies. However, energy demand is inherently variable, affected by unpredictable factors like extreme weather events and changes in consumer behavior. Adding contextual data, such as temperature records or economic indicators, would likely improve forecast accuracy and provide a more comprehensive understanding of the factors driving energy consumption.