Capstone

2025-02-11

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readxl)
library(lubridate)
```

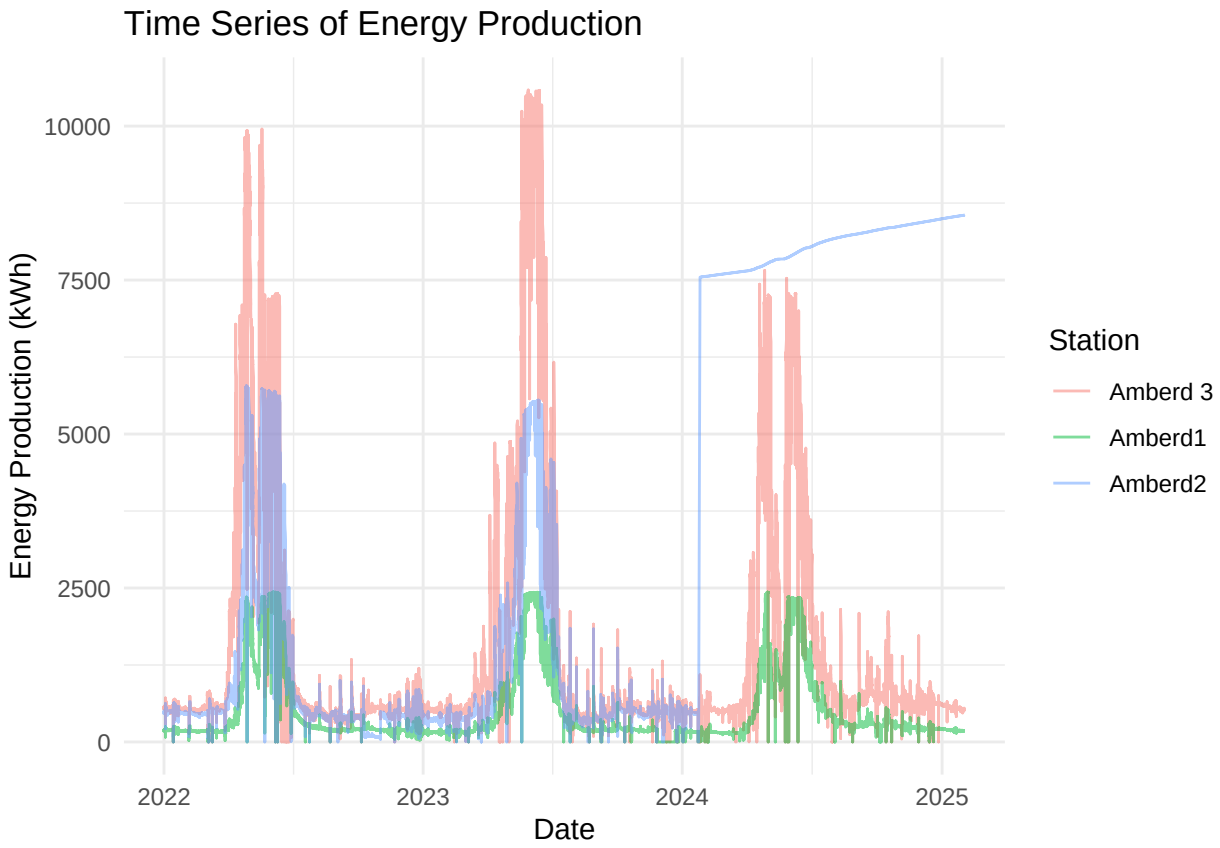
```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

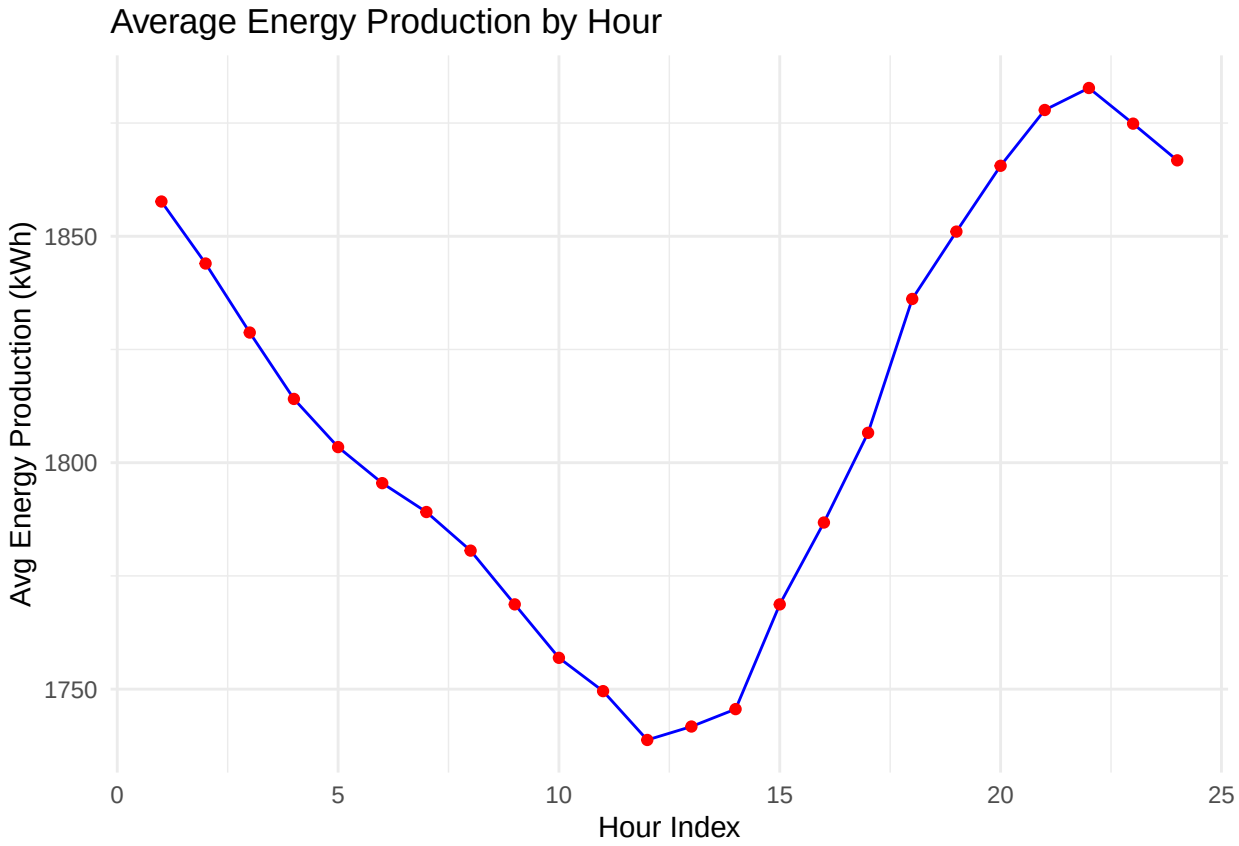
```
library(tidyr)
```

```
df <- read_excel("Amberd.xlsx")

# Convert Date column to Date type
df$Date <- as.Date(df$Date)

# Remove NA values
df <- df %>% drop_na(`Energy Production (kWh)`)
```

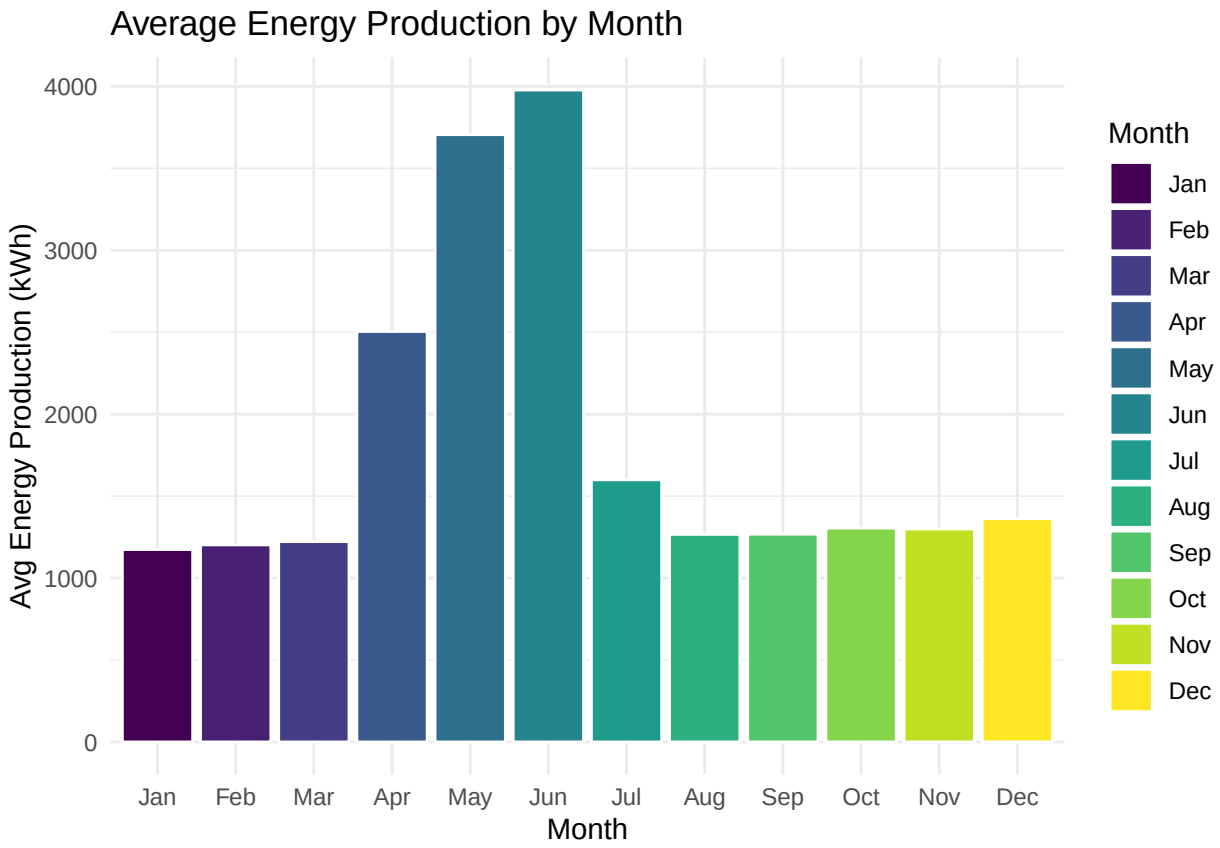
```
# Plot 1: Time Series of Energy Production
ggplot(df, aes(x = Date, y = `Energy Production (kWh)`, color = Station)) +
  geom_line(alpha = 0.5) +
  labs(title = "Time Series of Energy Production",
       x = "Date",
       y = "Energy Production (kWh)") +
  theme_minimal()
```




```
# Plot 2: Average Energy Production by Hour
df %>%
  group_by(`Hour Index`) %>%
  summarise(avg_production = mean(`Energy Production (kWh)`, na.rm = TRUE)) %>%
  ggplot(aes(x = `Hour Index`, y = avg_production)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(title = "Average Energy Production by Hour",
       x = "Hour Index",
       y = "Avg Energy Production (kWh)") +
  theme_minimal()
```



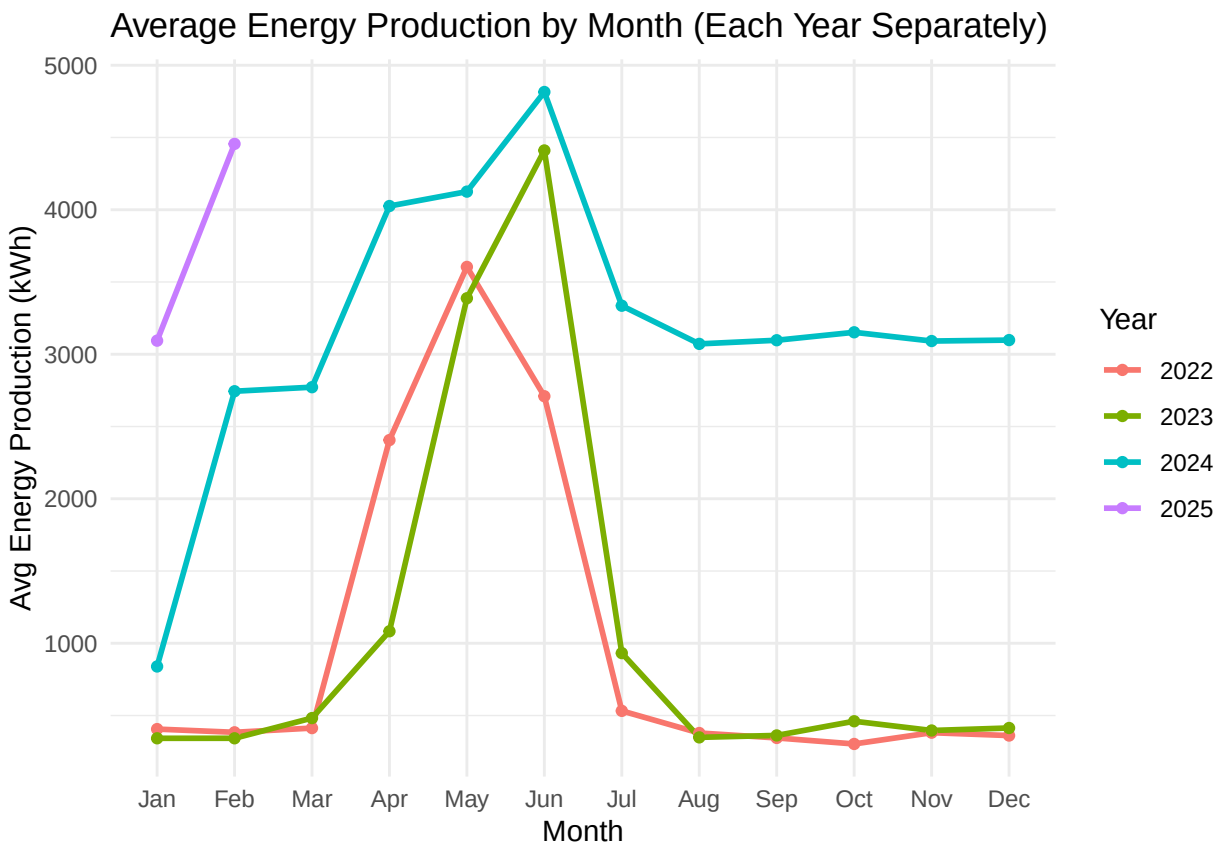

```r
# Extract Month
df$Month <- month(df$Date, label = TRUE)

# Plot 3: Average Energy Production by Month
df %>%
  group_by(Month) %>%
  summarise(avg_production = mean(`Energy Production (kWh)`, na.rm = TRUE)) %>%
  ggplot(aes(x = Month, y = avg_production, fill = Month)) +
  geom_bar(stat = "identity", color = "white") +
  labs(title = "Average Energy Production by Month",
       x = "Month",
       y = "Avg Energy Production (kWh)") +
  theme_minimal()
```




```
df$Year <- year(df$Date)  # Extracts the year
df$Month <- month(df$Date, label = TRUE)  # Extracts the month

df %>%
  group_by(Year, Month) %>%  # Groups by both Year and Month
  summarise(avg_production = mean(`Energy Production (kWh)`, na.rm = TRUE)) %>%
  ggplot(aes(x = Month, y = avg_production, group = Year, color = as.factor(Year))) +
  geom_line(size = 1) +
  geom_point() +
  labs(title = "Average Energy Production by Month (Each Year Separately)",
       x = "Month",
       y = "Avg Energy Production (kWh)",
       color = "Year") +
  theme_minimal()
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```


Average Energy Production by Month (Each Year Separately)
Avg Energy Production (kWh)
5000
4000
3000
2000
1000
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Year
2022
2023
2024
2025