District heating system (DHS) Analysis

2025-08-25

Project Goal

This analysis aims to filter a dataset of District heating system (DHS) facilities in the Ústí nad Labem Region of the Czech Republic. The project focuses on two specific objectives:

- Identify the proportion of non-renewable CZT facilities with a heat output below 5 MW.
- Calculate the average share of coal as a fuel source for these smaller, non-renewable facilities.

Key Skills Demonstrated

- Data Cleaning and Preprocessing: Handling raw data, including correcting and converting data types.
- Data Filtering and Subsetting: Selecting specific data based on regional and technical criteria.
- Data Manipulation and Aggregation: Calculating proportions and averages from filtered data.
- R Programming: Utilizing the dplyr package for efficient data wrangling.

Data Source

The data used in this analysis was transformed from a publicly available XLSX document published by the Energy Regulatory Office (ERO) of the Czech Republic. The original data is part of the "Yearly Report on the Operation of Heat Supply Systems in the CR for 2021."

Data Processing

```
# Load necessary libraries
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
```

```
# Load the dataset
# The data is assumed to be a CSV file named "Data_CZT_1.csv"
CZT <- read.csv("/cloud/project/CZT_statistika/Data_CZT_1.csv")

# Data cleaning: remove spaces and convert 'Vykon_MW' column to numeric
CZT$Vykon_MW <- gsub("[ ,]", "", CZT$Vykon_MW)
CZT$Vykon_MW <- as.numeric(CZT$Vykon_MW)
```

Filtering Data for the Ústí nad Labem Region

```
# Filter the dataset to include only facilities in the Ústí nad Labem Region ("U")
ustecky <- CZT[, c("Kraj", "Uhli", "ZP", "Biom_OZE", "Topne_oleje", "Jina_paliva", "Vykon_MW")] %>%
  filter(Kraj == "U")

# Check the dimensions of the filtered dataset
dim(ustecky)
```

```
## [1] 244   7
```

```
# There are 244 rows, indicating 244 CZT facilities in the region.
```

Calculating the Proportion of Non-Renewable Facilities

```
# Create a new dataset excluding renewable energy sources (Biom_OZE)
ustecky_bez_OZE <- ustecky %>%
  select(-Biom_OZE)

# Calculate the total number of facilities in the region
total_rows <- nrow(ustecky)

# Count the number of non-renewable facilities with a heat output less than 5 MW
rows_less_than_5k <- ustecky_bez_OZE %>%
  filter(Vykon_MW < 5.000) %>%
  nrow()

# Calculate the percentage and print the result
percentage <- (rows_less_than_5k / total_rows) * 100
cat("Proportion of non-renewable CZT facilities in Ústí nad Labem Region with output < 5 MW:", round(pe
```

```
## Proportion of non-renewable CZT facilities in Ústí nad Labem Region with output < 5 MW: 12.3 %
```

Calculating the Average Coal Share

```
# Filter for non-renewable facilities with output less than 5 MW and calculate the average coal share
average_uhli <- ustecky_bez_OZE %>%
  filter(Vykon_MW < 5.000) %>%
  summarize(average_uhli = mean(Uhli, na.rm = TRUE))

# Print the result
print(average_uhli)
```

```
##   average_uhli
## 1     83.43333
```