Final Project Manuscript

Vinayak Revisanker

2024-07-25

Libraries

library(tidyverse)

Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'tidyr' was built under R version 4.3.3

Warning: package 'dplyr' was built under R version 4.3.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.4  
✔ forcats 1.0.0 ✔ stringr 1.5.0  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.2 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(hms)

Attaching package: 'hms'  
  
The following object is masked from 'package:lubridate':  
  
 hms

library(ggplot2)

Reading and loading CSV file

methane <- read\_csv("Methane in the Furnance.csv", show\_col\_types = FALSE)  
  
methane

# A tibble: 296 × 3  
 obs co2 methane  
 <time> <dbl> <dbl>  
 1 01:01 53.8 -0.109  
 2 01:02 53.6 0   
 3 01:03 53.5 0.178  
 4 01:04 53.5 0.339  
 5 01:05 53.4 0.373  
 6 01:06 53.1 0.441  
 7 01:07 52.7 0.461  
 8 01:08 52.4 0.348  
 9 01:09 52.2 0.127  
10 01:10 52 -0.18   
# ℹ 286 more rows

checking for missing values

missing\_values <- colSums(is.na(methane))  
  
print(missing\_values)

obs co2 methane   
 0 0 0

summary statistics

co2\_summary <- summary(methane$co2)  
methane\_summary <- summary(methane$methane)  
  
  
print(co2\_summary)

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 45.60 51.20 53.50 53.51 56.00 60.50

print(methane\_summary)

Min. 1st Qu. Median Mean 3rd Qu. Max.   
-2.71600 -0.83000 0.00000 -0.05683 0.68050 2.83400

Directories for plots and results

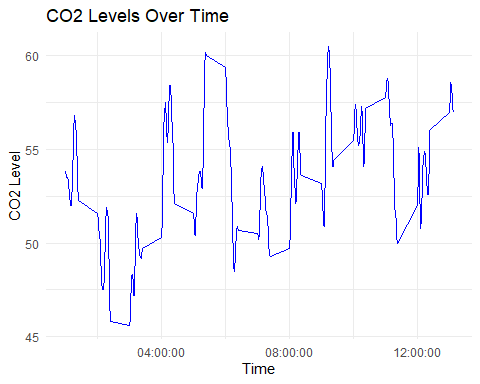
dir.create("plots", showWarnings = FALSE)  
dir.create("results", showWarnings = FALSE)

plot of CO2 levels over time

co2\_plot <- ggplot(methane, aes(x = obs, y = co2)) +  
 geom\_line(color = "blue") +  
 labs(title = "CO2 Levels Over Time", x = "Time", y = "CO2 Level") +  
 theme\_minimal()  
ggsave("plots/co2\_plot.png", plot = co2\_plot)

Saving 5 x 4 in image

print(co2\_plot)

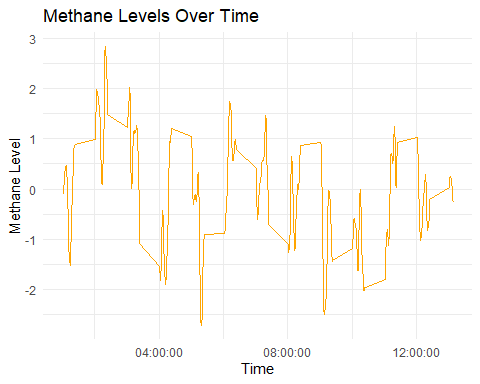


Plot of methane levels over time

methane\_plot <- ggplot(methane, aes(x = obs, y = methane)) +  
 geom\_line(color = "orange") +  
 labs(title = "Methane Levels Over Time", x = "Time", y = "Methane Level") +  
 theme\_minimal()  
ggsave("plots/methane\_plot.png", plot = methane\_plot)

Saving 5 x 4 in image

print(methane\_plot)



Correlation between CO2 and methane levels

correlation <- cor(methane$co2, methane$methane, use = "complete.obs")  
print(paste("Correlation between CO2 and methane levels:", correlation))

[1] "Correlation between CO2 and methane levels: -0.484450717109386"

Splitting data into training and testing sets

set.seed(123)  
train\_indices <- sample(1:nrow(methane), size = 0.8 \* nrow(methane))  
train\_data <- methane[train\_indices, ]  
test\_data <- methane[-train\_indices, ]

Linear model and summary

lm\_fit <- lm(co2 ~ methane, data = train\_data)  
lm\_summary <- summary(lm\_fit)  
capture.output(lm\_summary, file = "results/lm\_summary.txt")  
  
print(lm\_fit)

Call:  
lm(formula = co2 ~ methane, data = train\_data)  
  
Coefficients:  
(Intercept) methane   
 53.442 -1.376

print(lm\_summary)

Call:  
lm(formula = co2 ~ methane, data = train\_data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-6.2299 -2.0031 0.0401 2.2960 6.5693   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 53.4423 0.1828 292.378 < 2e-16 \*\*\*  
methane -1.3756 0.1701 -8.085 3.33e-14 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 2.805 on 234 degrees of freedom  
Multiple R-squared: 0.2184, Adjusted R-squared: 0.215   
F-statistic: 65.37 on 1 and 234 DF, p-value: 3.329e-14

Model evaulation

predictions <- predict(lm\_fit, newdata = test\_data)  
actuals <- test\_data$co2  
residuals <- actuals - predictions  
mse <- mean(residuals^2)  
rmse <- sqrt(mse)  
r\_squared <- lm\_summary$r.squared  
  
print(paste("RMSE:", rmse))

[1] "RMSE: 2.8156497018104"

print(paste("R-squared:", r\_squared))

[1] "R-squared: 0.218365844110571"

evaluation\_metrics <- c(paste("RMSE:", rmse), paste("R-squared:", r\_squared))  
write(evaluation\_metrics, file = "results/metrics\_summary.txt")

Manuscript directory and results

dir.create("manuscript", showWarnings = FALSE)  
  
  
  
correlation <- readLines("results/analysis\_summary.txt")  
lm\_summary <- readLines("results/lm\_summary.txt")  
metrics\_summary <- readLines("results/metrics\_summary.txt")  
  
  
print(correlation)

[1] "Correlation between CO2 and methane levels: -0.484450717109386"

print(lm\_summary)

[1] ""   
 [2] "Call:"   
 [3] "lm(formula = co2 ~ methane, data = train\_data)"   
 [4] ""   
 [5] "Residuals:"   
 [6] " Min 1Q Median 3Q Max "   
 [7] "-6.2299 -2.0031 0.0401 2.2960 6.5693 "   
 [8] ""   
 [9] "Coefficients:"   
[10] " Estimate Std. Error t value Pr(>|t|) "   
[11] "(Intercept) 53.4423 0.1828 292.378 < 2e-16 \*\*\*"   
[12] "methane -1.3756 0.1701 -8.085 3.33e-14 \*\*\*"   
[13] "---"   
[14] "Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1"  
[15] ""   
[16] "Residual standard error: 2.805 on 234 degrees of freedom"   
[17] "Multiple R-squared: 0.2184,\tAdjusted R-squared: 0.215 "   
[18] "F-statistic: 65.37 on 1 and 234 DF, p-value: 3.329e-14"   
[19] ""

print(metrics\_summary)

[1] "RMSE: 2.8156497018104" "R-squared: 0.218365844110571"

# 1. Discussion

## 1.1 Summary and Interpretation

*The analysis showed that both CO2 and methane levels fluctuate over time. The correlation between the two gases will provide insight into their relationship in the furnace environment.*

## 1.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 1.3 Conclusions

*Understanding the behavior of methane and carbon dioxide in furnaces can help in developing better control strategies to reduce emissions. Further research with more context-specific data is recommended.*

# 2. References

0a8029a263c0a128bc84961770acd801f698db05