Prova 4Intelligence Junho/2021

Henrique Jordão Figueiredo Alves

24/06/2021

1. Introdução

Este é um projeto de análise de dados e machine learning referente ao processo seletivo para o cargo de Cientista de Dados da 4Intelligence.

Se necessário, favor modificar o diretório dos arquivos na função setwd().

2. Etapa 1: Preparação

Nesta Etapa realizamos a instalação e o carregamento dos pacotes necessários para o projeto ser executado.

```
# Carregando os pacotes
library(tidyr)
library(dplyr)
library(ggplot2)
library(xlsx)
library(reshape2)
library(scales)
library(ggcorrplot)
library(caret)
library(randomForest)
library(e1071)
library(xgboost)

dados_base <- read.xlsx("Bases_Final_ADS_Jun2021.xlsx",sheetName="dados")
```

3. Etapa 2: Pré-Processamento

Agora que já temos o ambiente pronto e os dados carregados, iremos dar início ao processo de análise e manipulação dos dados.

Antes de qualquer coisa, vamos verificar o resumo do dataset para termos uma noção melhor das variáveis nas quais estamos trabalhando.

```
summary(dados_base)
```

```
##    data_tidy             com_co          com_n           com_ne
##  Min.   :2004-01-01   Min.   :283.1   Min.   :175.5   Min.   : 539.5
##  1st Qu.:2008-09-23   1st Qu.:393.4   1st Qu.:241.2   1st Qu.: 732.0
```

```
##  Median :2013-06-16   Median :520.5   Median :340.6   Median : 972.3
##  Mean   :2013-06-16   Mean   :499.4   Mean   :327.1   Mean   : 949.5
##  3rd Qu.:2018-03-08   3rd Qu.:602.6   3rd Qu.:404.0   3rd Qu.:1169.8
##  Max.   :2022-12-01   Max.   :683.1   Max.   :460.3   Max.   :1390.6
##                       NA's   :22      NA's   :22      NA's   :22
##      com_s            com_se         ind_co          ind_n
##  Min.   : 627.7   Min.   :2159   Min.   :364.3   Min.   : 810.3
##  1st Qu.: 850.1   1st Qu.:2849   1st Qu.:497.0   1st Qu.:1028.2
##  Median :1090.1   Median :3489   Median :693.9   Median :1140.1
##  Mean   :1058.5   Mean   :3399   Mean   :646.0   Mean   :1121.5
##  3rd Qu.:1217.3   3rd Qu.:3914   3rd Qu.:769.5   3rd Qu.:1228.6
##  Max.   :1552.7   Max.   :4572   Max.   :904.8   Max.   :1322.0
##  NA's   :22       NA's   :22     NA's   :22      NA's   :22
##      ind_ne          ind_s          ind_se         res_co          res_n
##  Min.   :1452   Min.   :1811   Min.   :6331   Min.   : 466.4   Min.   :320.2
##  1st Qu.:1914   1st Qu.:2279   1st Qu.:7478   1st Qu.: 585.6   1st Qu.:402.9
##  Median :2215   Median :2558   Median :7784   Median : 769.7   Median :566.8
##  Mean   :2165   Mean   :2495   Mean   :7829   Mean   : 776.7   Mean   :588.8
##  3rd Qu.:2394   3rd Qu.:2693   3rd Qu.:8272   3rd Qu.: 935.9   3rd Qu.:760.8
##  Max.   :2575   Max.   :3037   Max.   :8796   Max.   :1306.4   Max.   :962.1
##  NA's   :22     NA's   :22     NA's   :22     NA's   :22       NA's   :22
##      res_ne          res_s          res_se        renda_r      pop_ocup_br
##  Min.   : 962.9   Min.   :1057   Min.   :3433   Min.   :1401   Min.   :75778
##  1st Qu.:1301.2   1st Qu.:1276   1st Qu.:4224   1st Qu.:1611   1st Qu.:84519
##  Median :1804.6   Median :1566   Median :5049   Median :1746   Median :87672
##  Mean   :1799.7   Mean   :1540   Mean   :4904   Mean   :1787   Mean   :87295
##  3rd Qu.:2231.9   3rd Qu.:1750   3rd Qu.:5464   3rd Qu.:1928   3rd Qu.:90708
##  Max.   :2759.0   Max.   :2303   Max.   :6571   Max.   :2335   Max.   :94552
##  NA's   :22       NA's   :22     NA's   :22     NA's   :98
##     massa_r             du           pmc_a_co       temp_max_co
##  Min.   :101690   Min.   :18.00   Min.   : 39.85   Min.   :24.30
##  1st Qu.:118639   1st Qu.:20.00   1st Qu.: 68.86   1st Qu.:29.45
##  Median :139925   Median :21.00   Median : 81.59   Median :30.94
##  Mean   :142963   Mean   :20.93   Mean   : 79.01   Mean   :30.66
##  3rd Qu.:161283   3rd Qu.:22.00   3rd Qu.: 89.82   3rd Qu.:31.70
##  Max.   :201815   Max.   :23.00   Max.   :117.22   Max.   :35.42
##  NA's   :98
##   temp_min_co       pmc_r_co          pim_co          pmc_a_n
##  Min.   :14.25   Min.   : 44.37   Min.   : 59.00   Min.   : 30.02
##  1st Qu.:17.37   1st Qu.: 69.92   1st Qu.: 80.97   1st Qu.: 68.97
##  Median :20.40   Median : 81.89   Median : 91.67   Median : 86.52
##  Mean   :19.40   Mean   : 80.11   Mean   : 97.83   Mean   : 83.42
##  3rd Qu.:21.12   3rd Qu.: 89.14   3rd Qu.:117.71   3rd Qu.: 98.73
##  Max.   :22.31   Max.   :126.06   Max.   :145.30   Max.   :146.16
##
##    temp_max_n      temp_min_n       pmc_r_n          pim_n
##  Min.   :30.33   Min.   :21.56   Min.   : 35.93   Min.   : 62.18
##  1st Qu.:32.13   1st Qu.:23.11   1st Qu.: 65.09   1st Qu.: 91.02
##  Median :32.78   Median :23.47   Median : 85.83   Median : 98.93
##  Mean   :33.00   Mean   :23.44   Mean   : 82.91   Mean   : 99.17
##  3rd Qu.:33.89   3rd Qu.:23.89   3rd Qu.: 96.45   3rd Qu.:108.78
##  Max.   :35.93   Max.   :24.82   Max.   :151.52   Max.   :130.80
##
##     pmc_a_ne       temp_max_ne     temp_min_ne       pmc_r_ne
```

```
##  Min.   : 33.93   Min.   :29.12   Min.   :21.10   Min.   : 36.27
##  1st Qu.: 66.08   1st Qu.:30.25   1st Qu.:22.60   1st Qu.: 64.64
##  Median : 80.35   Median :30.92   Median :23.58   Median : 80.70
##  Mean   : 76.84   Mean   :30.86   Mean   :23.32   Mean   : 77.44
##  3rd Qu.: 87.41   3rd Qu.:31.41   3rd Qu.:24.03   3rd Qu.: 88.44
##  Max.   :122.95   Max.   :32.57   Max.   :24.73   Max.   :128.83
##
##      pim_ne          pmc_a_s         temp_max_s      temp_min_s
##  Min.   : 60.70   Min.   : 46.93   Min.   :17.71   Min.   : 8.993
##  1st Qu.: 92.28   1st Qu.: 70.59   1st Qu.:22.10   1st Qu.:12.986
##  Median : 97.90   Median : 87.10   Median :25.25   Median :16.297
##  Mean   : 97.56   Mean   : 84.45   Mean   :25.02   Mean   :16.035
##  3rd Qu.:103.25   3rd Qu.: 98.20   3rd Qu.:28.09   3rd Qu.:19.087
##  Max.   :114.00   Max.   :132.30   Max.   :31.73   Max.   :22.199
##
##     pmc_r_s           pim_s           pmc_a_se       temp_max_se
##  Min.   : 51.70   Min.   : 64.97   Min.   : 46.53   Min.   :22.31
##  1st Qu.: 71.11   1st Qu.: 90.72   1st Qu.: 74.64   1st Qu.:25.74
##  Median : 91.07   Median : 95.84   Median : 89.45   Median :27.57
##  Mean   : 87.86   Mean   : 96.30   Mean   : 85.32   Mean   :27.54
##  3rd Qu.:101.42   3rd Qu.:104.35   3rd Qu.: 96.39   3rd Qu.:29.06
##  Max.   :146.80   Max.   :116.96   Max.   :126.26   Max.   :32.02
##
##   temp_min_se       pmc_r_se          pim_se
##  Min.   :14.89   Min.   : 45.08   Min.   : 60.32
##  1st Qu.:17.27   1st Qu.: 70.20   1st Qu.: 84.20
##  Median :19.43   Median : 89.67   Median : 91.07
##  Mean   :19.04   Mean   : 84.21   Mean   : 90.90
##  3rd Qu.:20.87   3rd Qu.: 95.22   3rd Qu.: 97.19
##  Max.   :23.13   Max.   :137.33   Max.   :112.05
##
```

Como há valores 'NA' nas colunas `massa_r` e `renda_r`, teremos problemas na análise, caso não façamos algo a respeito. Como solução, optei por preencher os dados em falta com o valor da média dos valores já existentes em cada variável.

```
dados_base <- dados_base %>%
  replace_na(list(renda_r=mean(dados_base$renda_r, na.rm = TRUE),
  massa_r = mean(mean(dados_base$massa_r, na.rm = TRUE))))
```

Feito isso, já temos resolvido o problema com 'Missing Values'.

Agora, precisamos separar os dados presentes dos dados futuros, para podermos realizar a análise descritiva do negócio, e posteriormente, construirmos o modelo de aprendizado de máquina.

Como temos 206 linhas de dados preenchidas, iremos extraí-las para o nosso novo dataset.

```
dados <- dados_base[1:206,]
```

4. Etapa 3: Análise Descritiva

Agora partiremos para a nossa análise descritiva e exploratória do projeto. Nessa fase, iremos adquirir um conhecimento melhor dos nossos dados através da plotagem de gráficos.

Para responde a Questão 1 da prova, precisamos separar nossos dados em subsets e depois agrupá-los num dataframe mais organizado para realizarmos a análise dos dados de consumo nas regiões do Brasil, assim como a correlação entre as respectivas variáveis.

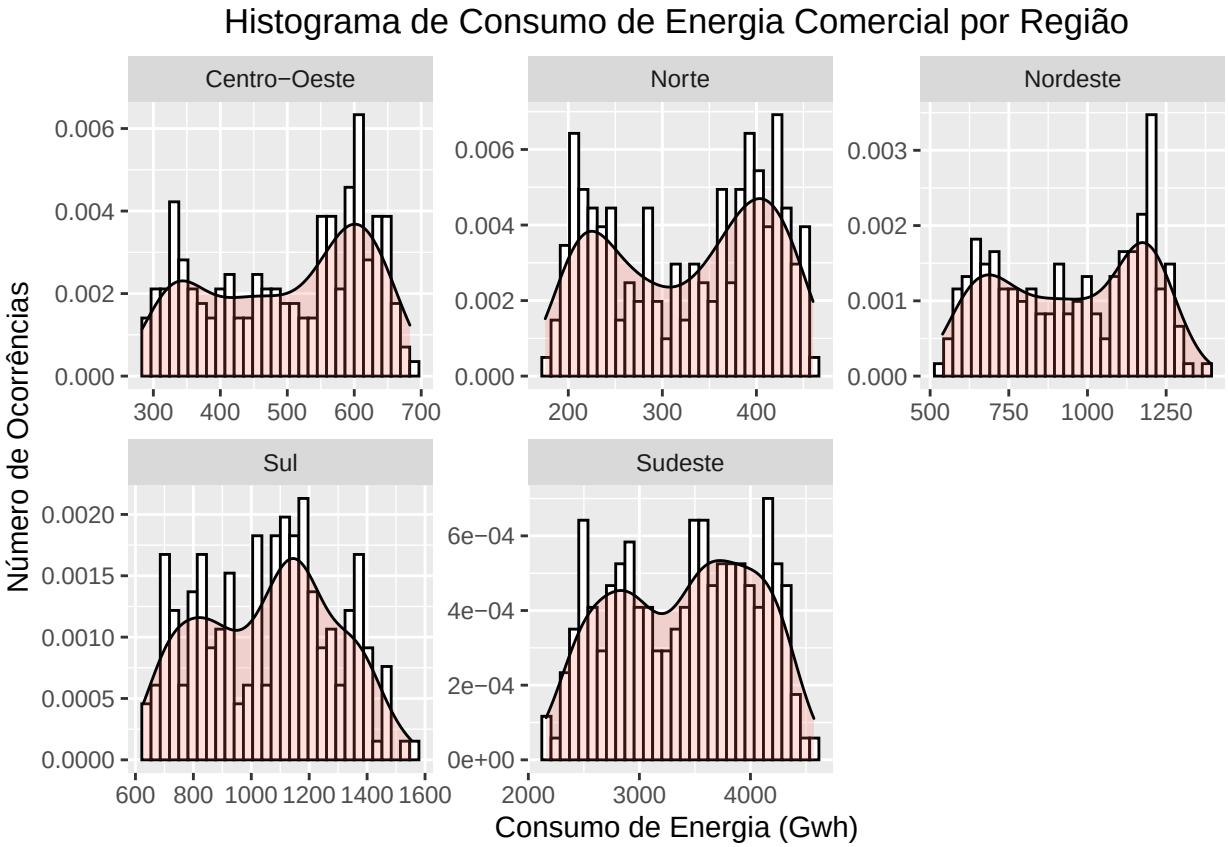
Também iremos analisar, em especial, a variável `ind_se`, correspondente ao consumo de energia industrial da região Sudeste, já que esta será nossa variável preditora do nosso modelo de machine learning.

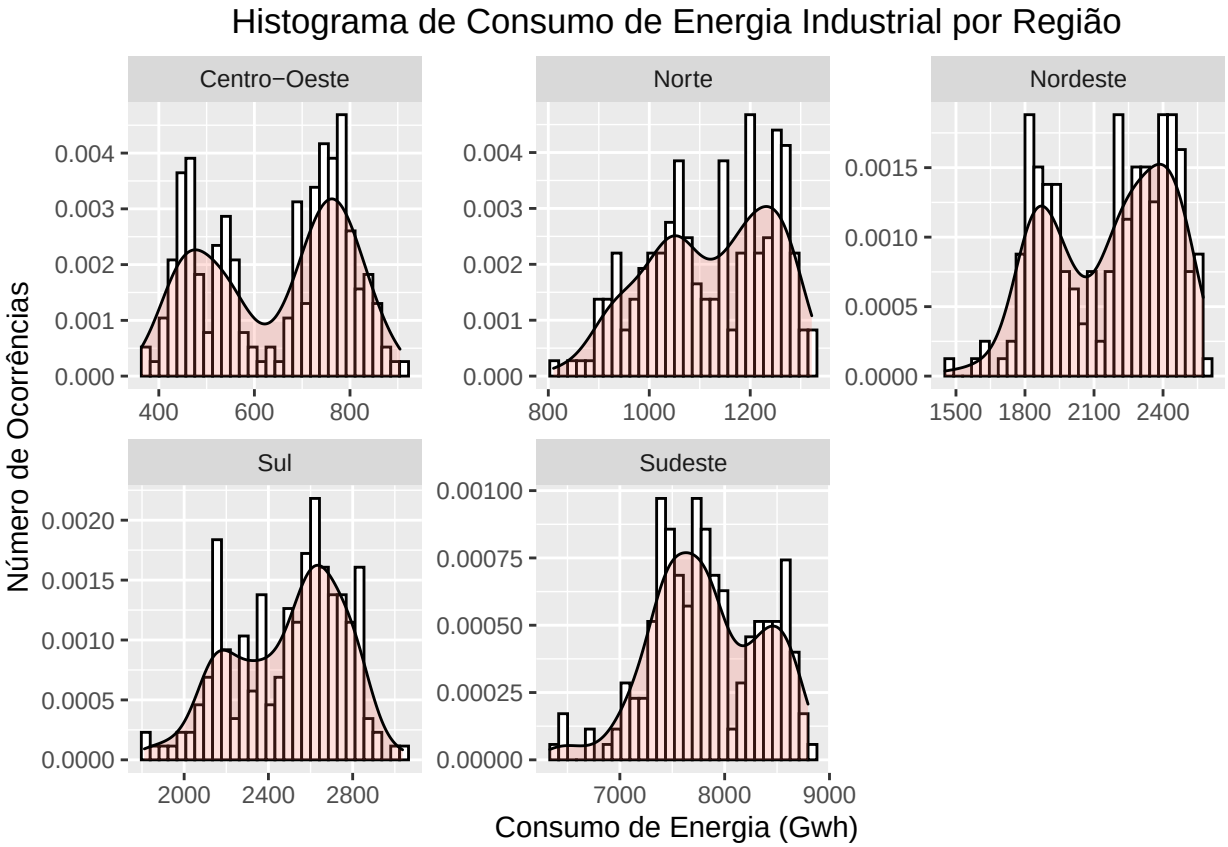
Logo:

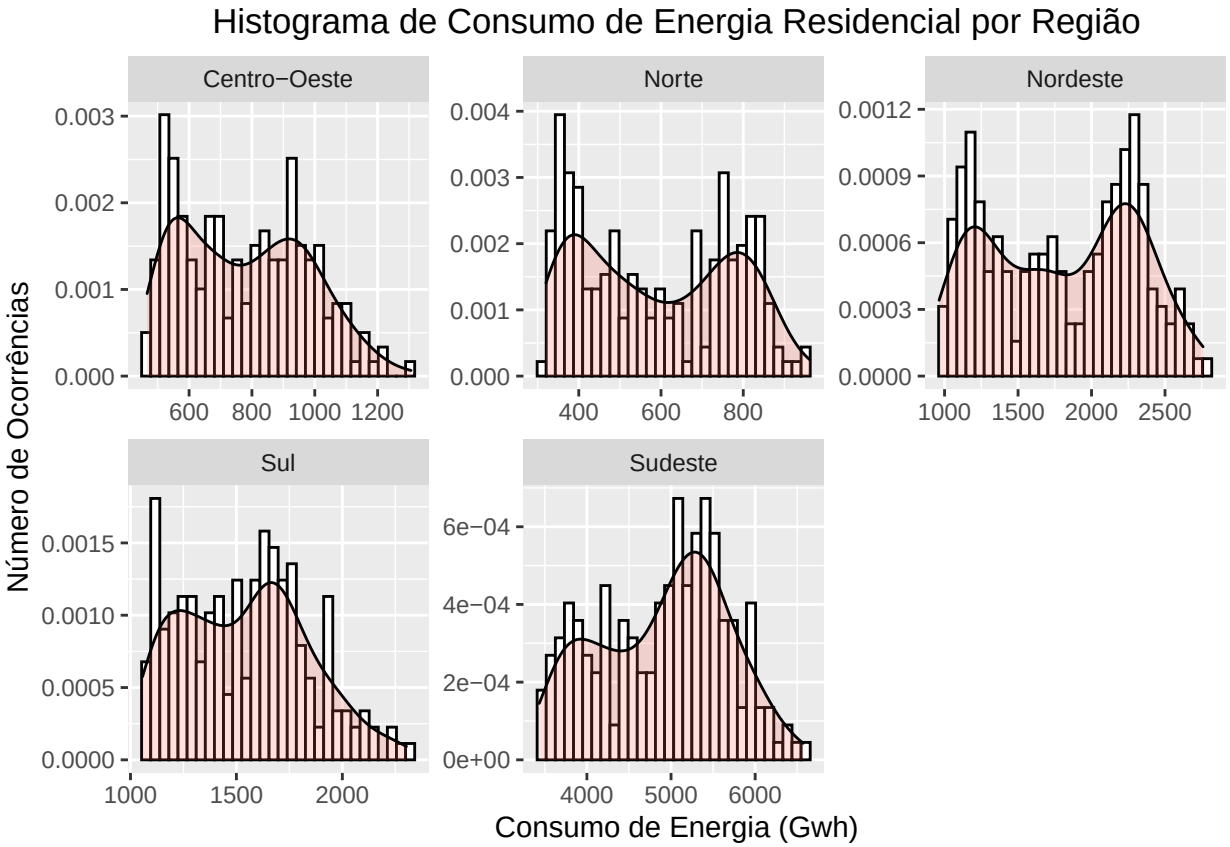
```
# Separando os dados das datas e criando um array com os nomes das regiões
regiao <- c('Centro-Oeste','Norte','Nordeste','Sul','Sudeste')
datas <- dados[,1]

# Criando subsets para análise de consumos de energia comercial, industrial e residencial

comercio <- subset(dados[,2:6])
colnames(comercio) <- regiao
comercio$categoria <- "Comércio"
comercio$data <- datas

industria <- subset(dados[,7:11])
colnames(industria) <- regiao
industria$categoria <- "Indústria"
industria$data <- datas

residencial <- subset(dados[,12:16])
colnames(residencial) <- regiao
residencial$categoria <- "Residencial"
residencial$data <- datas

dados_consumo <- do.call("rbind", list(comercio, industria, residencial))

head(dados_consumo)
```

```
##   Centro-Oeste    Norte Nordeste      Sul  Sudeste categoria       data
## 1     307.2821 193.8096 589.2903 704.0017 2450.453  Comércio 2004-01-01
## 2     290.1468 175.4953 550.7726 733.8949 2396.965  Comércio 2004-02-01
## 3     307.1001 182.7569 573.0280 738.4571 2402.521  Comércio 2004-03-01
## 4     329.1609 189.4908 603.9387 743.5705 2580.914  Comércio 2004-04-01
## 5     303.2379 191.6422 570.1765 696.5795 2344.084  Comércio 2004-05-01
## 6     284.9630 194.1630 573.7435 627.6517 2159.479  Comércio 2004-06-01
```

Agora sim podemos dar início à nossa análise descritiva. Primeiro vamos conferir os histogramas e as curvas de densidade das variáveis de consumo nos setores de comércio, indústria e residencial nas cinco regiões do país.

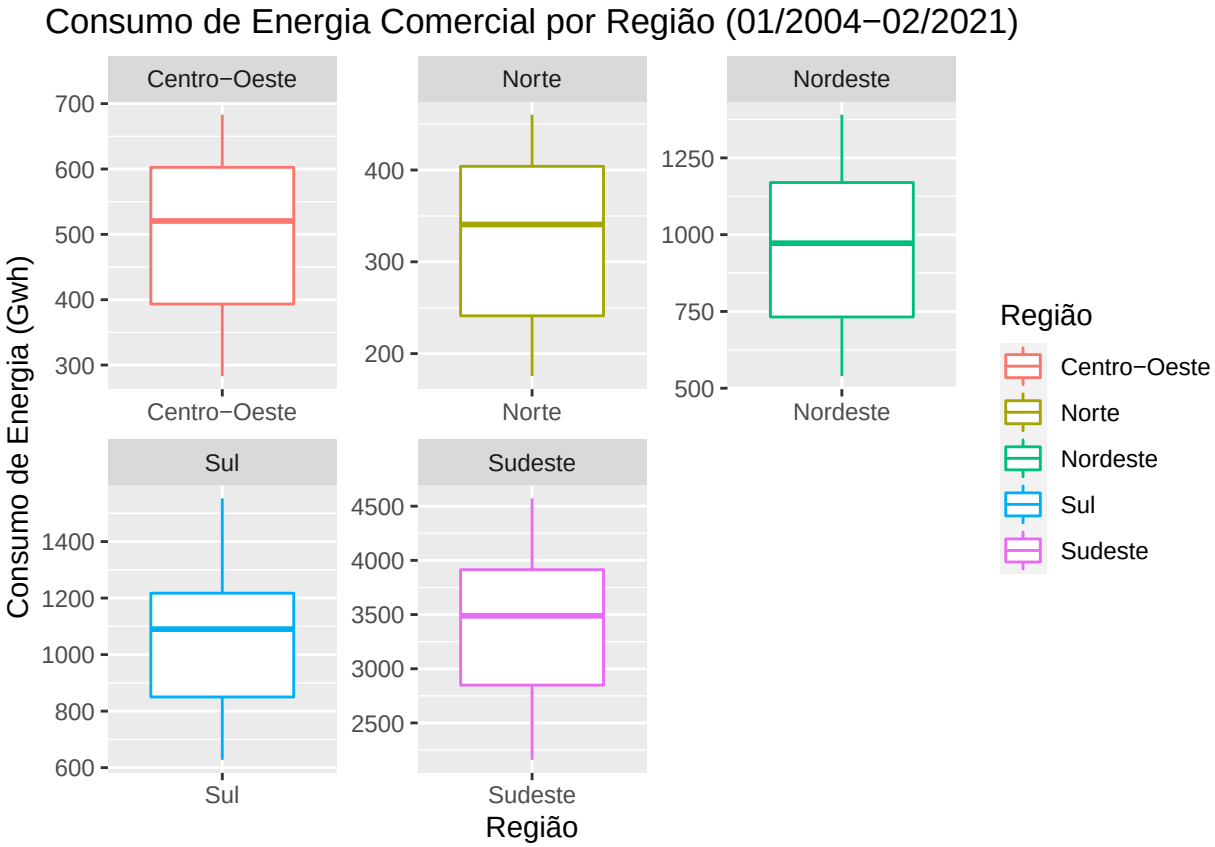

Histograma de Consumo de Energia Comercial por Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Número de Ocorrências
Consumo de Energia (Gwh)

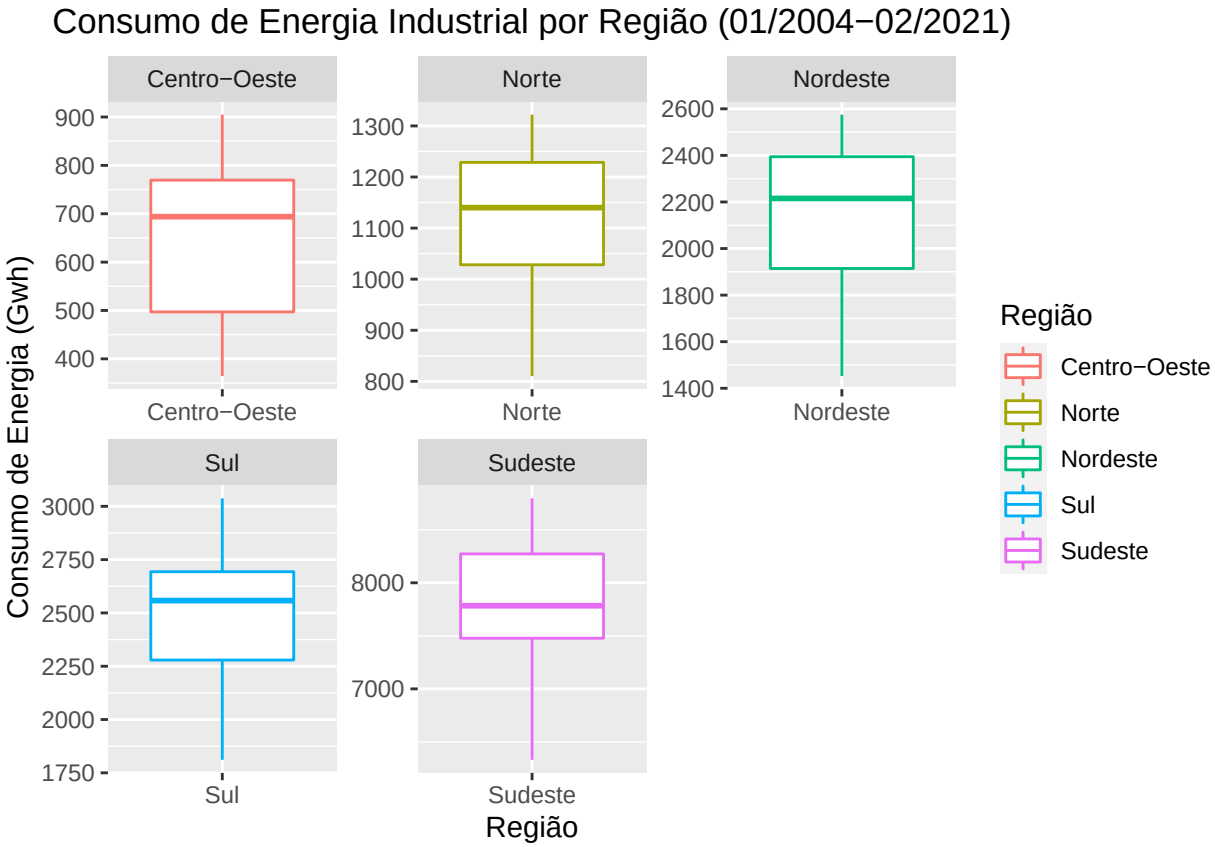

Histograma de Consumo de Energia Industrial por Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Número de Ocorrências
Consumo de Energia (Gwh)
0.004
0.003
0.002
0.001
0.000
400
600
800
1000
1200
0.0015
0.0010
0.0005
0.0000
1500
1800
2100
2400
0.0020
2000
2400
2800
0.00100
0.00075
0.00050
0.00025
0.00000
7000
8000
9000

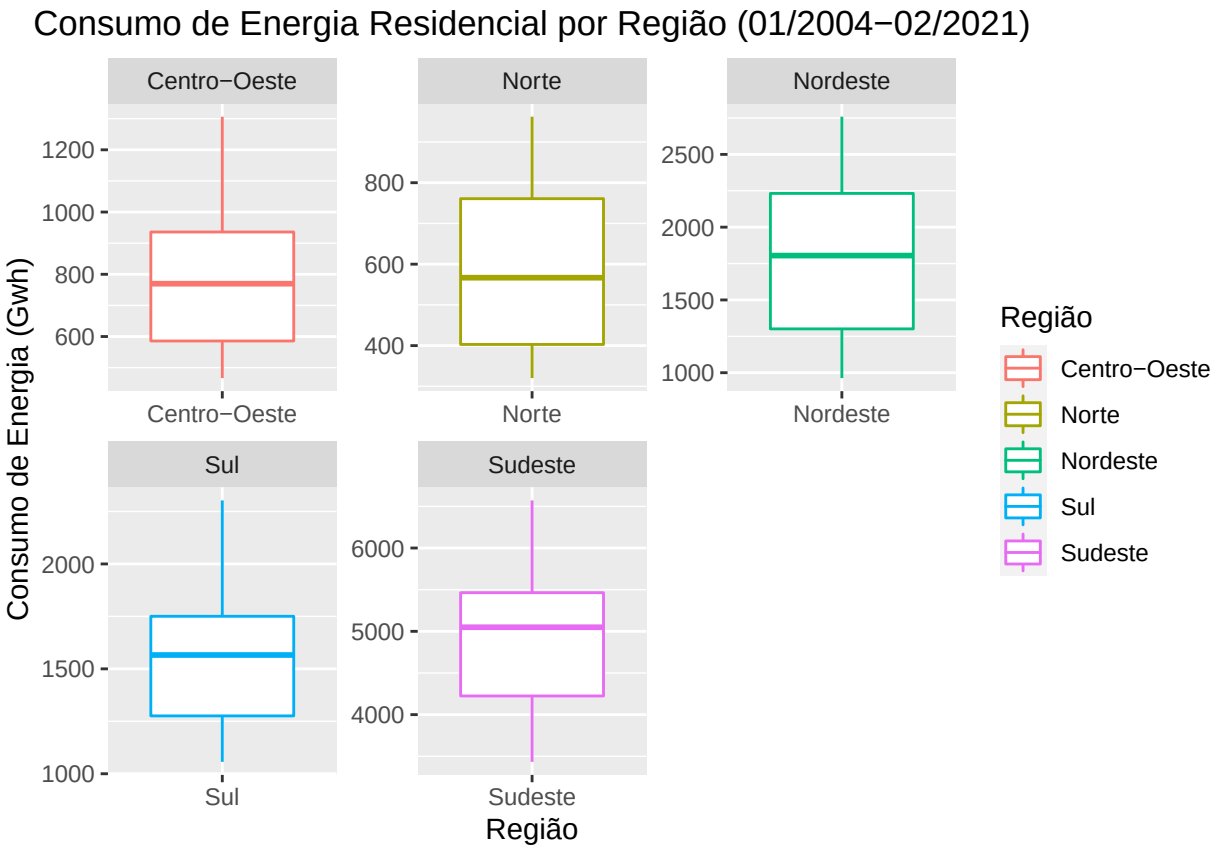


Nos histogramas acima, podemos ver que algumas variáveis não possuem distribuição normal, incluindo a variável de consumo de energia industrial da região Sudeste (`ind_se`). Nela, podemos ver que há uma assimetria mais à direita do centro de distribuição dos valores.

Em seguida, iremos analisar as variáveis de consumo por meio de boxplots. Confira abaixo:

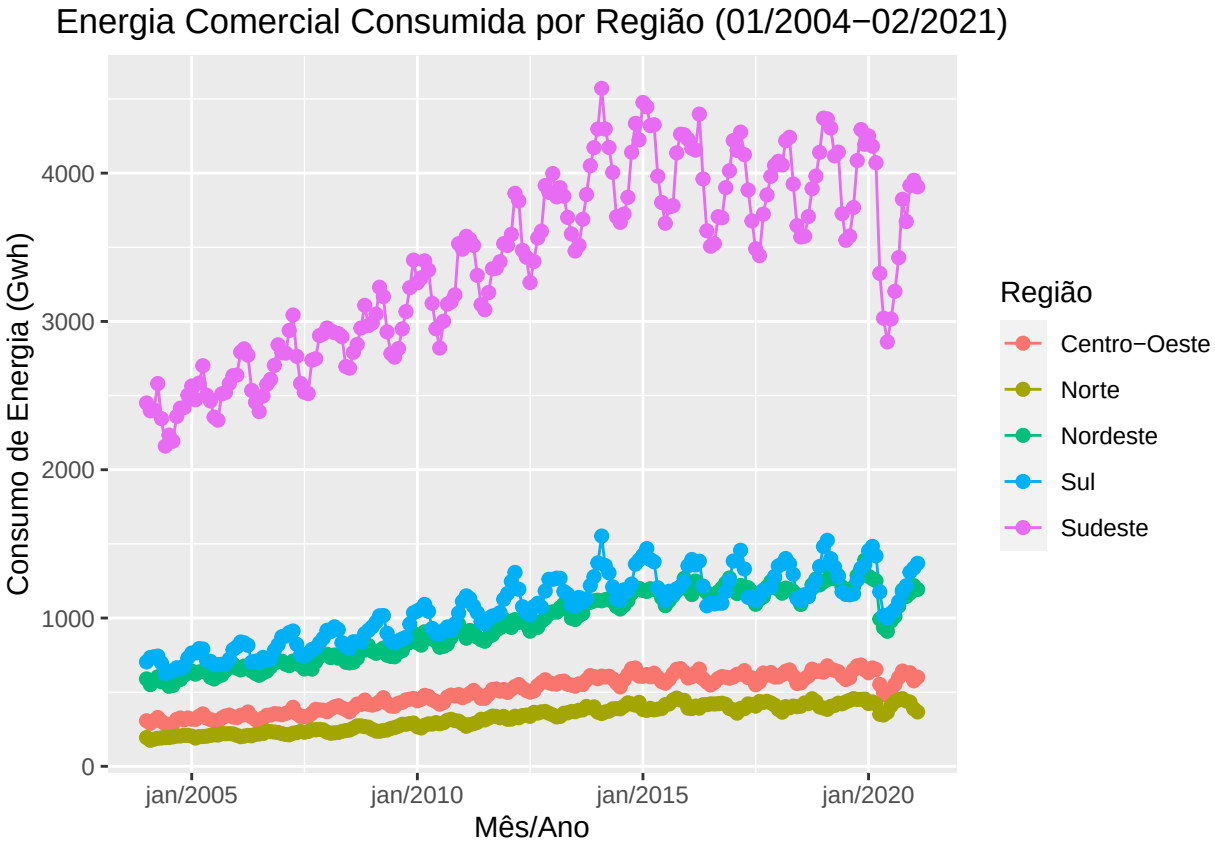

Consumo de Energia Comercial por Região (01/2004−02/2021)
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Consumo de Energia (Gwh)
700
600
500
400
300
400
300
200
1250
1000
750
500
1400
1200
1000
800
600
4500
4000
3500
3000
2500
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Região
Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste

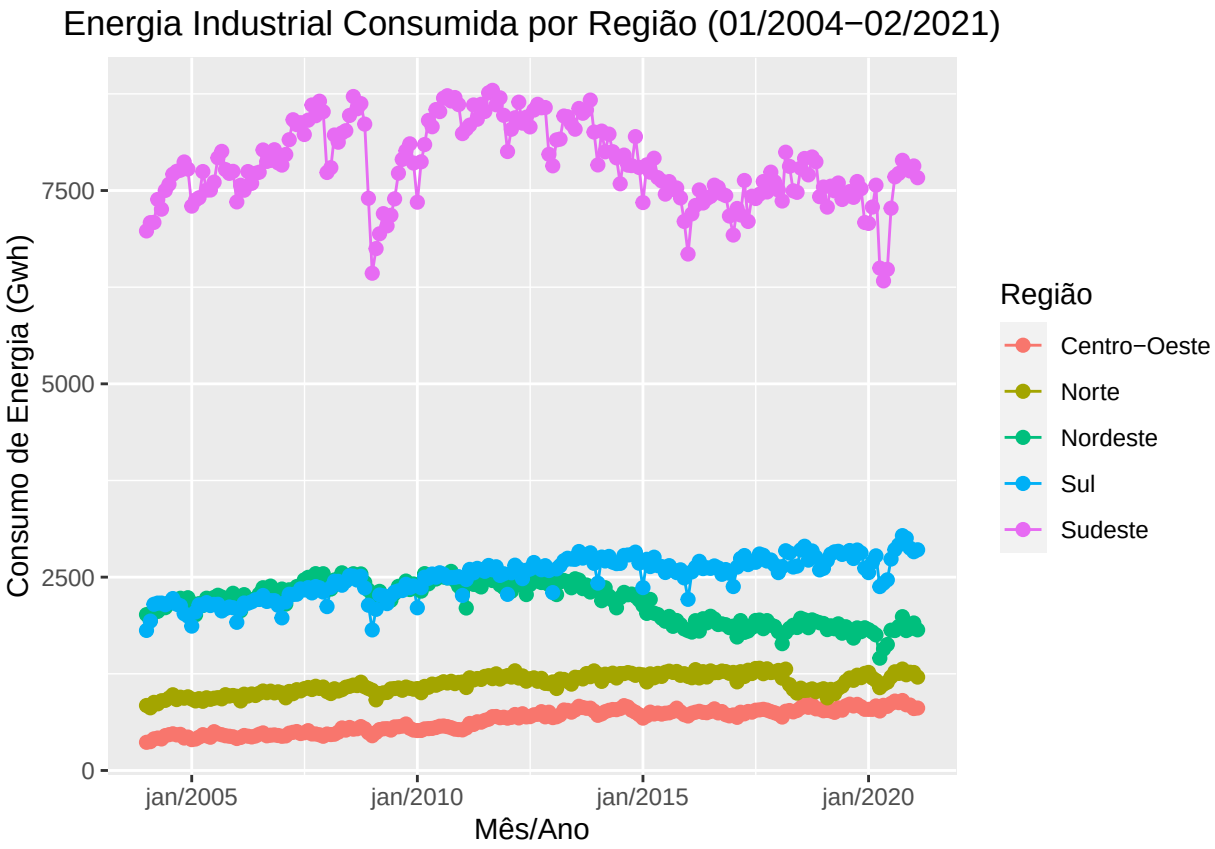

Consumo de Energia Industrial por Região (01/2004−02/2021)
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Consumo de Energia (Gwh)
900
800
700
600
500
400
1300
1200
1100
1000
900
800
2600
2400
2200
2000
1800
1600
1400
3000
2750
2500
2250
2000
1750
8000
7000
Centro−Oeste
Norte
Nordeste
Sul
Sudeste
Região
Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste

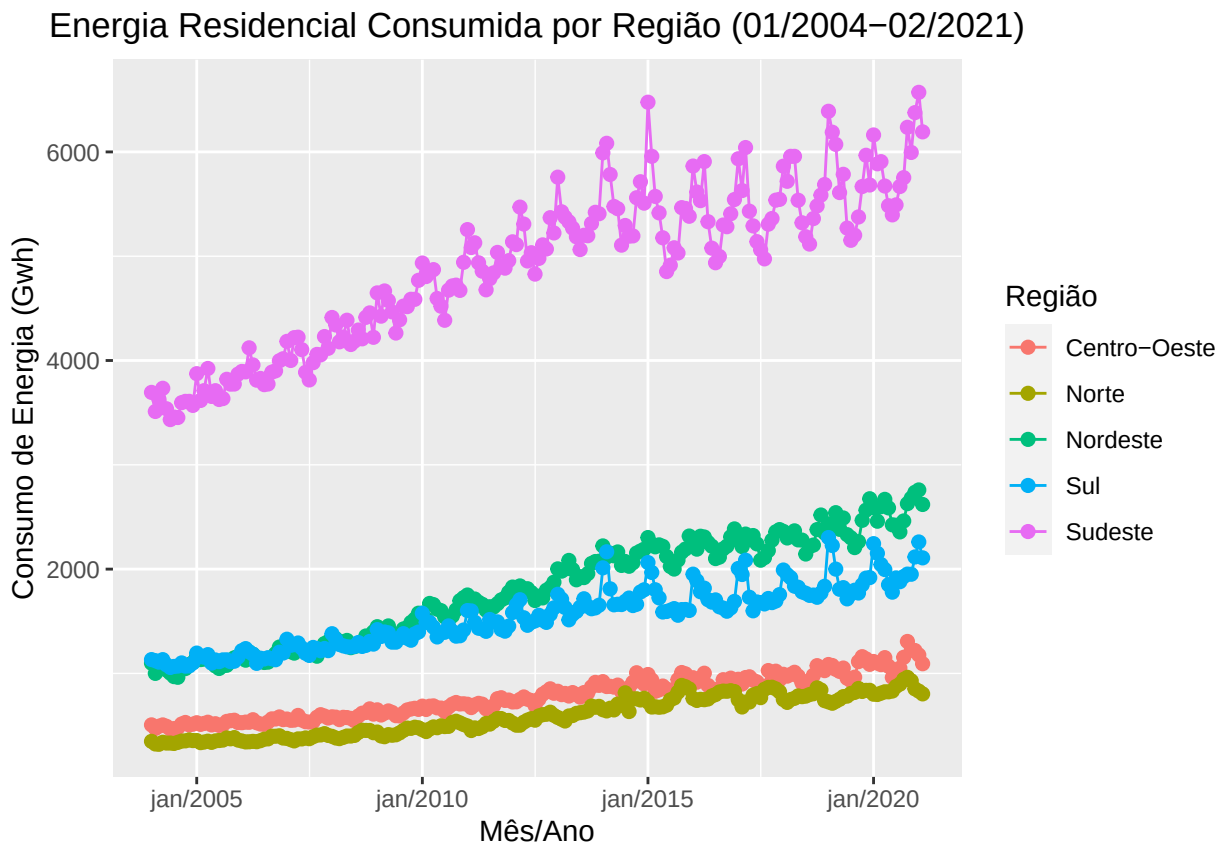


Com os boxplots acima, podemos ter uma visão de um ponto de vista diferente de como está a distribuição de valores das variáveis. Note como há uma discrepância entre o valor máximo e o valor mínimo da variável `ind_se`.

Agora, vamos analisar a evolução do consumos de energia ao longo do tempo com gráficos de série temporal


Energia Comercial Consumida por Região (01/2004−02/2021)
Consumo de Energia (Gwh)
4000
3000
2000
1000
0
jan/2005
jan/2010
jan/2015
jan/2020
Mês/Ano
Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste


Energia Industrial Consumida por Região (01/2004−02/2021)
Consumo de Energia (Gwh)
7500
5000
2500
0
jan/2005
jan/2010
jan/2015
jan/2020
Mês/Ano
Região
Centro−Oeste
Norte
Nordeste
Sul
Sudeste



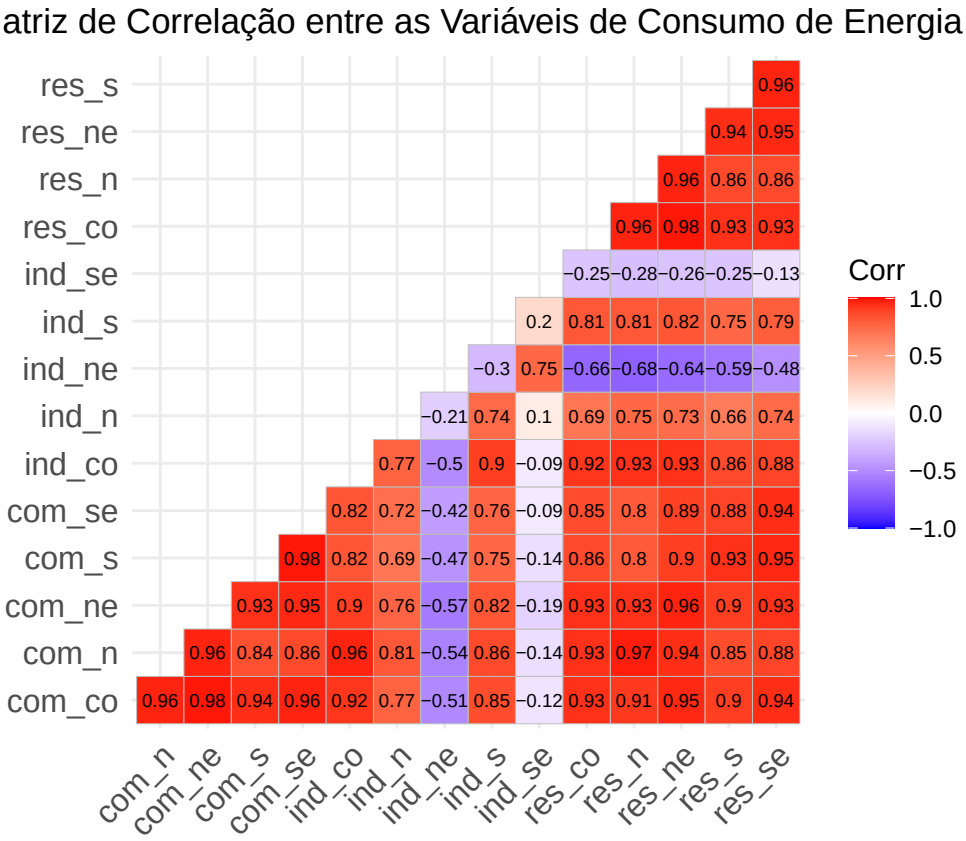
Nos gráficos, podemos ver a evolução do consumo de energia nas regiões entre os anos de 2004 e 2021.

No gráfico do setor industrial, podemos perceber que houve uma forte queda no consumo de energia em todas as regiões do país durante o biênio 2008-2009, em especial no Sudeste, onde se encontram a maior parte das indústrias em território nacional.

Esse acontecimento, pode ser explicado pela crise de 2008, que causou uma desindustrilização periódica no país e que afetou fortemente o setor de energia elétrica[1].

Outro acontecimento interessante de notar, é a queda do consumo de energia em todas as regiões no primeiro trimestre de 2020, devido à pandemia de COVID-19. Com o aumento de restrições de mobilidade, a demanda por energia elétrica diminuiu nas indústrias e no comércio, e uma leve alta no consumo residencial[2].

Por fim, iremos analisar a relação entre as variáveis de consumo de energia.



Na matriz de correlação acima, podemos perceber que quase todas as variáveis de consumo energético do país, tem forte relação entre elas, com exceção das variáveis de consumo industrial das regiões Nordeste e Sudeste, que por sua vez tem uma relação positiva forte entre elas (cor = 0.75)[3].

5. Etapa 4: Criando Modelos de Machine Learning

Nesta etapa vamos finalmente criar nossos modelos de aprendizado de máquina.

Já sabemos que, assim como a nossa variável preditora `ind_se`, todas as nossas variáveis são numéricas, com exceção da variável `data_tidy` que é uma variável do tipo 'Date'.

Logo, precisamos construir um modelo de regressão, com a melhor acurácia possível, que possa prever os dados de consumo de energia industrial na região Sudeste nos próximos 2 anos.

Antes, vamos remover a variável `data_tidy` e criar os modelos de treino e teste baseado nos nossos dados presentes.

```
dados$data_tidy <- NULL

sample1 <- sample(1:nrow(dados), 165)
sample2 <- sample(166:nrow(dados), 41)
dados_treino <- dados[sample1,]
dados_teste <- dados[sample2,]
```

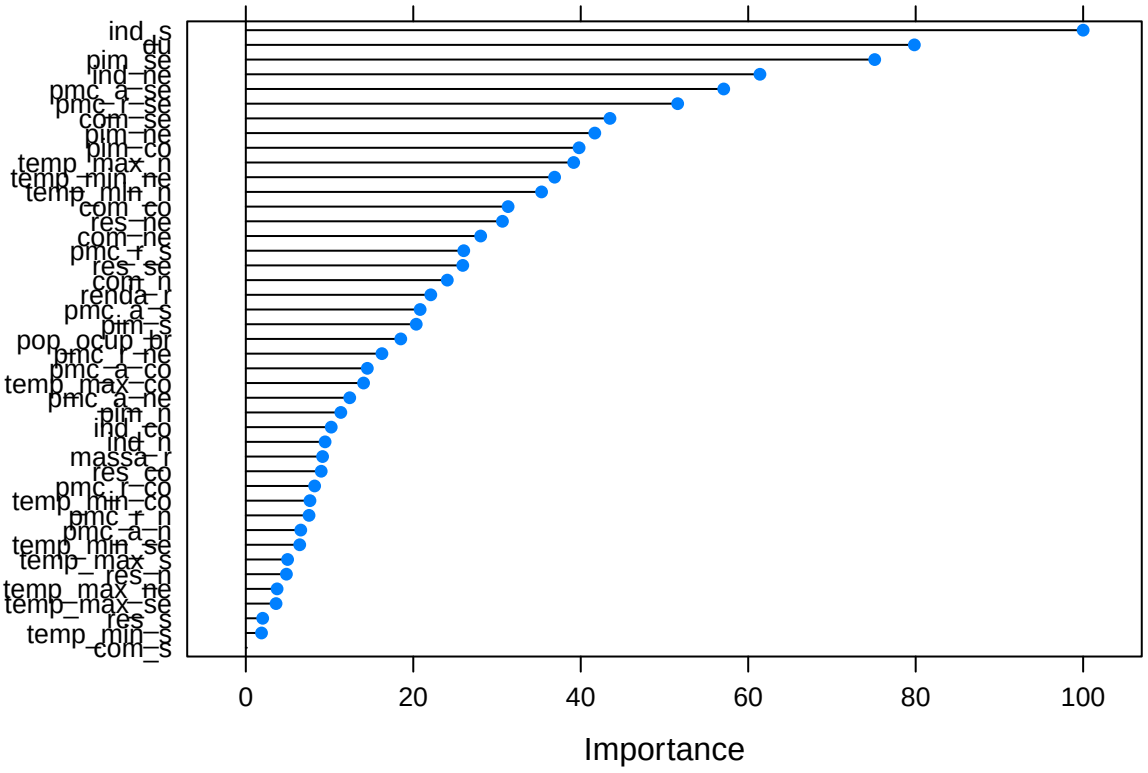
Como só temos apenas 206 observações no nosso dataset, optei por fazer uma divisão 80/20 para os dados de treino e de teste para garantir que o modelo tenha uma acurácia maior.

Vamos fazer em seguida, o nosso modelo inicial para determinarmos as variáveis mais importsntes para a criação do nosso modelo definitivo.

```
modelo_v1 <- train(ind_se ~ ., data = dados_treino, method = 'lm')
varImp(modelo_v1)
```

```
## lm variable importance
##
##   only 20 most important variables shown (out of 43)
##
##             Overall
## ind_s        100.00
## du            79.84
## pim_se        75.11
## ind_ne        61.40
## pmc_a_se      57.08
## pmc_r_se      51.58
## com_se        43.49
## pim_ne        41.68
## pim_co        39.80
## temp_max_n    39.16
## temp_min_ne   36.88
## temp_min_n    35.32
## com_co        31.33
## res_ne        30.65
## com_ne        28.05
## pmc_r_s       26.03
## res_se        25.91
## com_n         24.06
## renda_r       22.09
## pmc_a_s       20.82
```

```
plot(varImp(modelo_v1))
```



Pela lista de variáveis importantes acima e o gráfico, podemos ver que as variáveis mais importantes para a variável preditora são: `ind_s`, `pim_se`, `com_se` e `ind_ne`.

O fato do consumo industrial do Nordeste estar entre as variáveis importantes confere com o que vimos anteriormente na matriz de correlação de variáveis. As variáveis `pim_se`, que representa a produção industrial no Sudeste, e `com_se`, que representa o consumo no comercio da região, também não surpreende por estarem diretamente ligadas ao consumo industrial.

Assim vamos construir o nosso modelo de regressão linear baseado nessas informações.

Como a variável `ind_se` não possui uma distribuição normal, iremos realizar uma transformação logarítmica nela e nas demais variáveis importantes.

```
# Modelo Regressão Linear
modelo_v1 <- train(log(ind_se) ~ log(com_se) + log(ind_s) + log(pim_se) + log(ind_ne),
                   data = dados_treino, method = 'lm')
summary(modelo_v1)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.114920 -0.016526  0.001626  0.020579  0.074010
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.31782    0.36302   9.139 2.71e-16 ***
## `log(com_se)` -0.08434    0.02374  -3.553   0.0005 ***
## `log(ind_s)`   0.39012    0.05018   7.775 8.67e-13 ***
## `log(pim_se)`  0.04166    0.04701   0.886   0.3768
## `log(ind_ne)`  0.40313    0.04108   9.813  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03022 on 160 degrees of freedom
## Multiple R-squared:  0.7919, Adjusted R-squared:  0.7867
## F-statistic: 152.2 on 4 and 160 DF,  p-value: < 2.2e-16
```

Pelo resumo, obtivemos um erro residual baixo, o que é um bom sinal para o nosso modelo.

Iremos comprovar isso abaixo ao realizar a previsão dos nossos dados e conferindo com os nossos dados de teste.

```
previsao <- predict(modelo_v1, dados_teste)
previsao <- round(previsao,1)
mean(previsao==round(log(dados_teste$ind_se),1))
```

```
## [1] 0.804878
```

Conseguimos obter uma acurácia de 82.9%, o que é um bom resultado se levarmos em conta a escassez de dados usados no modelo.

Feito isso, iremos agora tentar encontrar uma alternativa ao modelo de regressão linear utilizando outros 4 modelos: randomForest, GradientBoost, regressão linear generalizada (glm) e o support vector machine (svm).

Esses modelos foram escolhidos por serem os mais eficazes para previsão de variáveis numéricas e fornecerem uma gama de opções para configurar os modelos a ponto de se obter uma melhor acurácia.

```
# Modelo randomForest
modelo_v2 <- randomForest(ind_se ~ com_se + ind_s + pim_se + ind_ne, data = dados_treino)
print(modelo_v2)
```

```
##
## Call:
##  randomForest(formula = ind_se ~ com_se + ind_s + pim_se + ind_ne,      data = dados_treino)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##           Mean of squared residuals: 62345.15
##                     % Var explained: 75.86
```

```
previsao2 <- predict(modelo_v2,dados_teste)
RMSE(previsao2,dados_teste$ind_se)
```

```
## [1] 212.7678
```

```r
# Modelo GradientBoost

set.seed(123)
modelo_v3 <- train(ind_se ~ com_se + ind_s + ind_ne + pim_se, data = dados_treino,
method = "xgbTree", trControl = trainControl("cv", number = 10),
objective = "reg:squarederror", verbose=FALSE)
```

```r
previsao3 <- predict(modelo_v3,dados_teste)
RMSE(previsao3, dados_teste$ind_se)
```

```
## [1] 197.7548
```

```r
# Modelo Regressão Linear Generalizada

modelo_v4 <- train(log(ind_se) ~ log(com_se) + log(ind_s) + log(pim_se) + log(ind_ne),
                   data = dados_treino, method = 'glm')

previsao4 <- predict(modelo_v4, dados_teste)
previsao4 <- round(previsao4,1)
mean(previsao4==round(log(dados_teste$ind_se),1))
```

```
## [1] 0.804878
```

```r
# Modelo SVM

modelo_v5 <- svm(log(ind_se) ~ log(com_se) + log(ind_s) + log(pim_se) + log(ind_ne),
                     data = dados_treino,
                     type = 'eps-regression',
                     kernel = 'linear')

pred_test <- predict(modelo_v5, dados_teste)
mean(round(pred_test,1) == round(log(dados_teste$ind_se),1))
```

```
## [1] 0.8292683
```

Acima, podemos ver que os modelos SVM e glm tiveram resultados similares aos de regressão linear. Já os modelos de randomForest e GradientBoost não obtiveram resultados satisfatórios, talvez por ser necesário uma configuração mais adequada para o dataset ou pela necessidade de mais dados para treinar o modelo.

6. Conclusão

Durante o projeto, pudemos ter uma noção melhor de como funciona o consumo de energia elétrica no país.

Na análise descritiva, vimos que a região Sudeste é dona da maior parte do consumo de energia entre todas as regiões, independente do setor. Com o passar dos anos, o consumo de energia tende a aumentar, especialmente em áreas mais industrializadas e com maior população.

Na modelagem, pudemos provar como a produção industrial e o consumo industrial do Nordeste estão diretamente ligados ao consumo industrial do Sudeste. Através de técnicas de normalização e uso de feature selection para escolha de variáveis importantes em relação à variável preditora, pudemos alcançar uma acurácia de 82.9% no nosso modelo de machine learning, mesmo com poucos dados à nossa disposição.

7. Referências Bibliográficas

[1] https://agenciabrasil.ebc.com.br/economia/noticia/2018-09/crise-de-2008-resultou-em-desindustrializacao-e-crise-fiscal-no-brasil - Acessado em 24 de junho de 2021

[2] https://www.epe.gov.br/pt/imprensa/noticias/impactado-pela-covid-19-consumo-de-energia-deve-cair-0-9-em-2020 - Acessado em 24 de junho de 2021

[3] https://g1.globo.com/jornal-nacional/noticia/2020/11/13/nordeste-produz-mais-energia-do-que-consome-e-excedente-e-distribuido-para-outras-regioes.ghtml - Acessado em 24 de junho de 2021