How do I use R to make my business to progress

AFONSO LUGO

ENERGY ENGINEER

afonso.lugo@epowerbay.com

Content

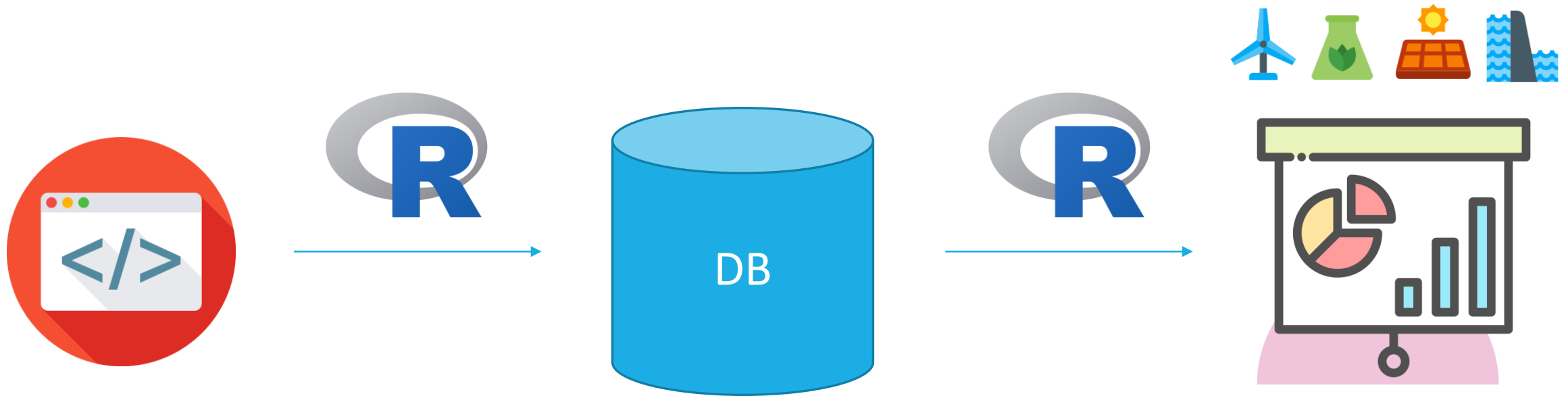**Data Collection (Web Scraping)**

Data Analysis (Reporting)

Machine Learning:

Clustering

Forecasting
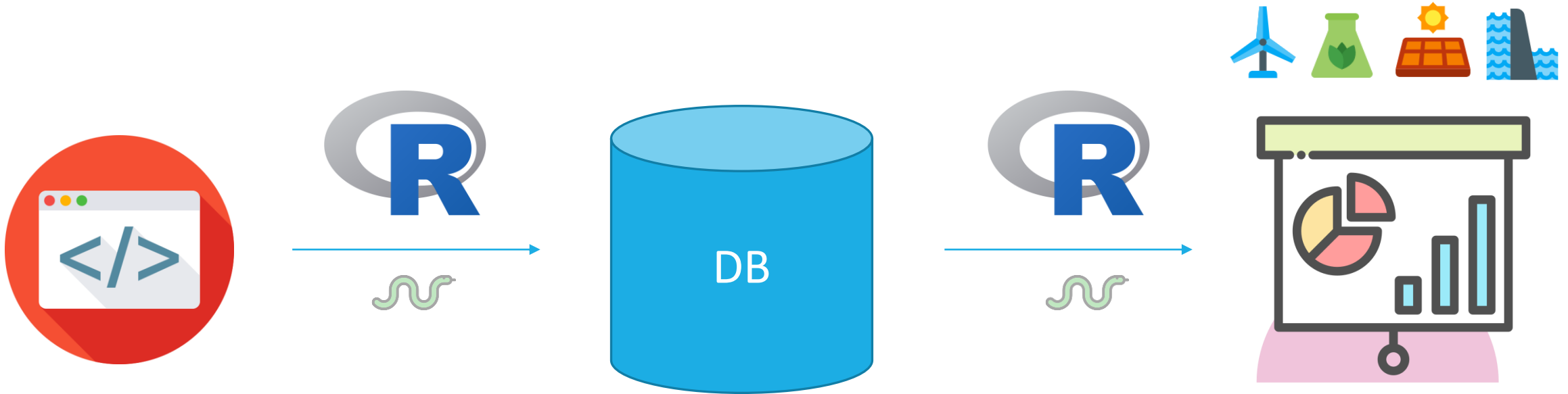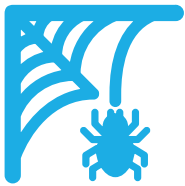
The Product
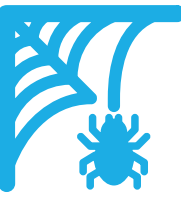
The Product

Content

Data Collection (Web Scraping)

Data Analysis (Reporting)

Machine Learning:
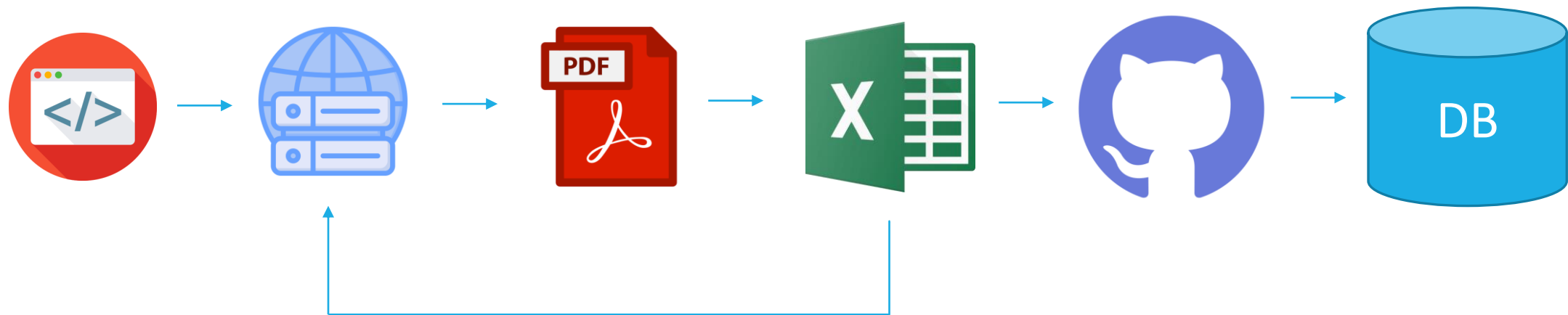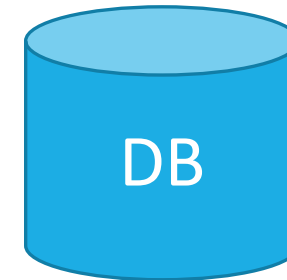
Clustering

Forecasting

Data Collection

The Old Way

Data Collection

The New Way

Data Collection

+ pdftools

The New Way

```
'http://www2.aneel.gov.br/cedoc/rea20197588sti.pdf'
```

```r
for(i in 1:quantidade) {

  n <- n+1

  if(n==ultima){
    next
  } else {

    print(n)

    # Verificar se a url possue erro #
    res <- try(pdf_text(sprintf('http://www2.aneel.gov.br/cedoc/rea%sti.pdf', n)))

    if(inherits(res, "try-error")) {
      next
    } else {⟵}
  }
}
```

```r
if(str_detect(texto2, "implantar e explorar")[1] == TRUE) {⟵}
```
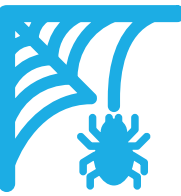
Data Collection

Web Scraping

PROS

- Time saving
- Time saving
- Time saving

CONS

- Hard to develop a generic script
- If the document changes format, you need a new script

Content

Data Collection (Web Scraping)

Data Analysis (Reporting)

Machine Learning:

Clustering

Forecasting

Data Analysis

- Statistical analysis
  - rodbc
  - dplyr
  - sf
- Graphics
  - ggplot2
  - ggspatial
  - gridExtra
  - cowplot
- Report generation (.docx)
  - ubiquity
  - officer

```r
db <- odbcDriverConnect(paste0("DRIVER={SQL Server};
                               server=",my_server,";
                               database=",my_db,";
                               uid=",my_username,";
                               pwd=",my_pwd))
```

```r
uf <- sqlQuery(db, "SELECT idUF, UF_Sigla FROM dbo.tbUF WHERE Paiz_id = 1")
```

```r
fontes_mapa <- ggplot(data = br$geometry) +
  geom_sf(aes(fill=br$Submercado)) +
  geom_point(data=df_n, aes(x=long, y=lat, color=fonte), size=4) +
  coord_sf(xlim=c(-80,-30), ylim=c(10,-35), expand = FALSE) +
  scale_color_manual(values=fontes_color) +
  scale_fill_manual(values=sub_color) +
  guides(color=guide_legend("Fonte")) +
  xlab('Longitude (º)') + ylab('Latitude (º)') +
  theme(
    axis.text=element_text(size=label_size, colour='black'),
    axis.title=element_text(size=label_size, colour='black'),
    panel.border = element_rect(colour='black',fill=NA),
    panel.grid = element_line(colour='black', linetype = 8),
    panel.background = element_rect(fill='white'),
    legend.text = element_text(size=15),
    legend.title = element_blank()
  )
```

Data Analysis

3. Emissão de Outorgas

No mês de referência, foram emitidas um total de **54 novas outorgas** com **1.940 MW de capacidade instalada**.

Resumo e Estatística Geral

Tab. 02: Estatística Geral dos projetos com Outorga emitida no mês de referência.

| Fonte	Quantidade	Capacidade Instalada (MW)	Potência Média (MW)
UFV	28	1.365,8	48,8
EOL	22	356	16,2
UTE	2	153	76,5
UHE	1	50	50,0
PCH	1	15,5	15,5
TOTAL	**54**	**1.940**	**35,9**
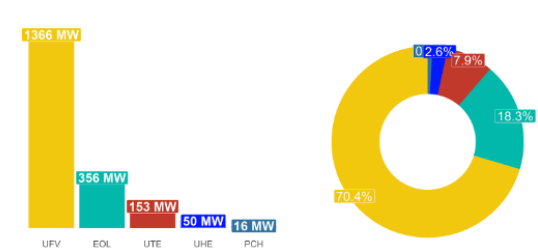
- *Fig. 9: Potência por fonte: Total (esquerda) e Parcela (direita).*

Localização

Os projetos com outorga emitida no período de referência estão localizados em **08 estados** brasileiros, sendo que **Minas Gerais recebeu a maior potência** com **668,5 MW** de capacidade instalada.

O gráfico a seguir mostra a localização dos projetos no território brasileiro.
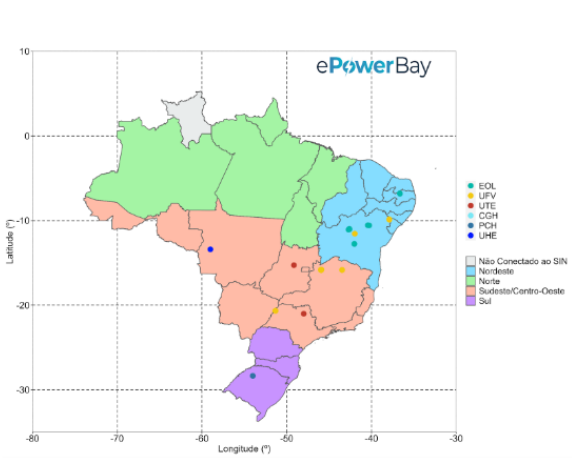
- *Fig. 10: Localização das outorgas por fonte.*

A empresa que mais adicionou capacidade instalada à matriz elétrica brasileiro foi a **Casa dos Ventos** com **174 MW** de usinas eólicas. A **Powertis** adicionou **144 MW** em usinas solares fotovoltaicas e ocupa a segunda posição.

A imagem a seguir mostra a capacidade instalada adicionada à matriz elétrica brasileira por empresa investidora e classificada por tipo de usina no último mês.
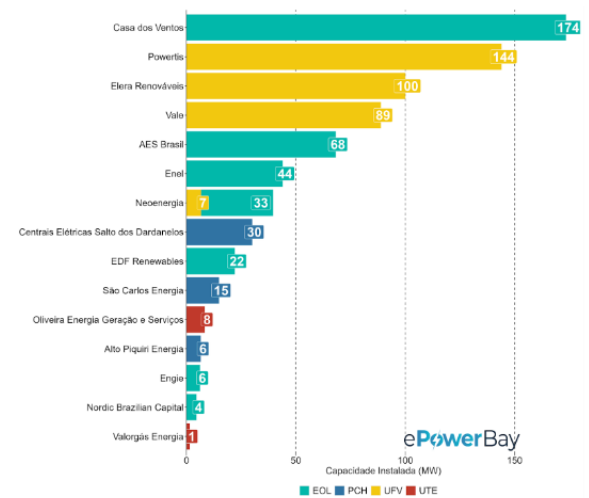
- *Fig. 20: Operação comercial por empresa investidora.*

Data Analysis

Reporting

PROS

- Time saving

- Control of graphics

- Prettier aesthetics

- Better understanding of the data

CONS

- The report might be too repetitive

- Need to pay attention to details

Content

Data Collection (Web Scraping)

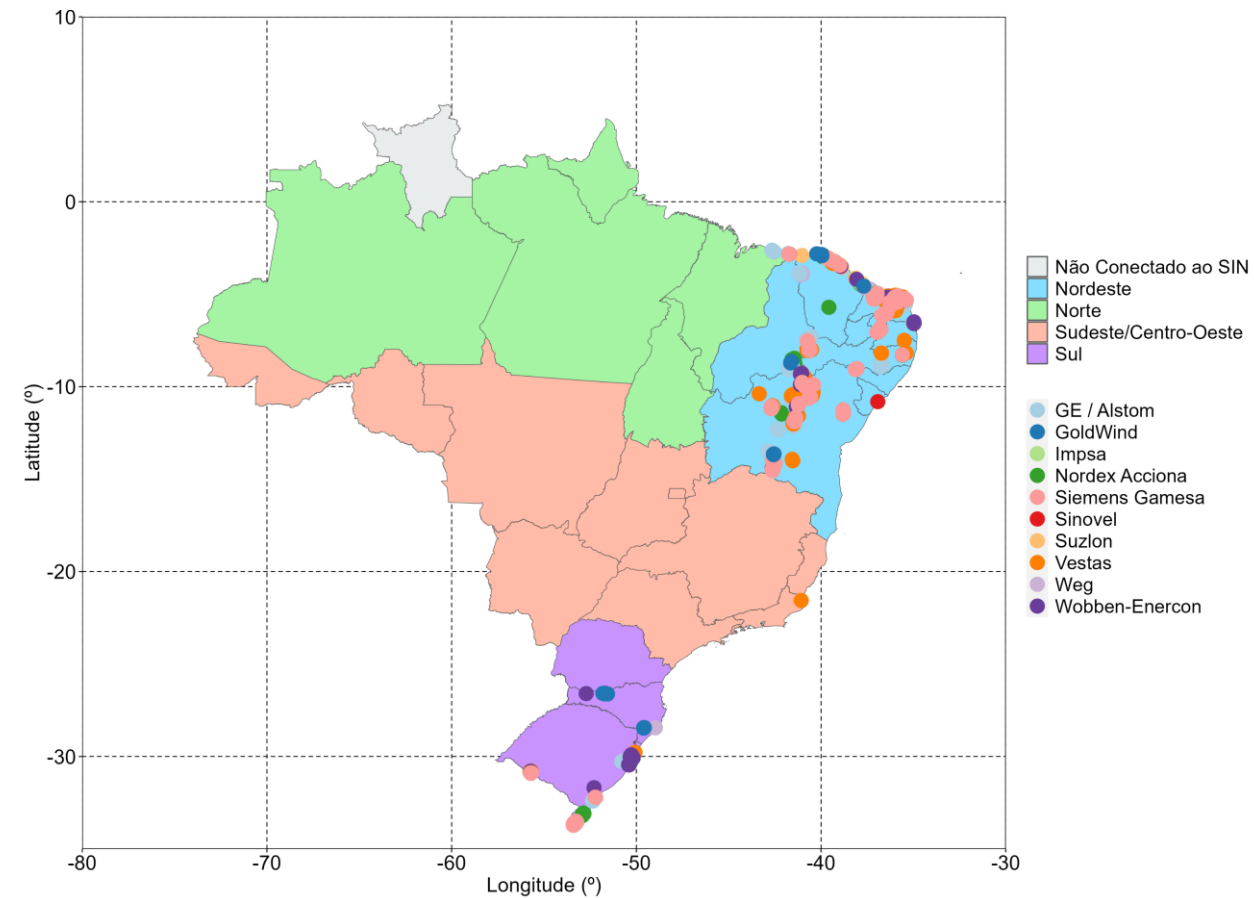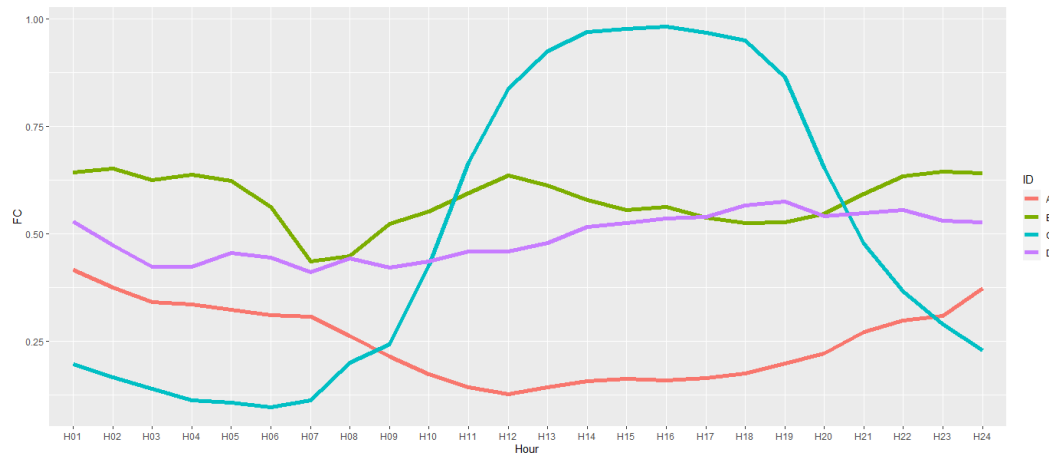Data Analysis (Reporting)

Machine Learning:

Classification

Forecasting

Machine Learning

Classification

- Which wind turbine performs better in a given wind condition?
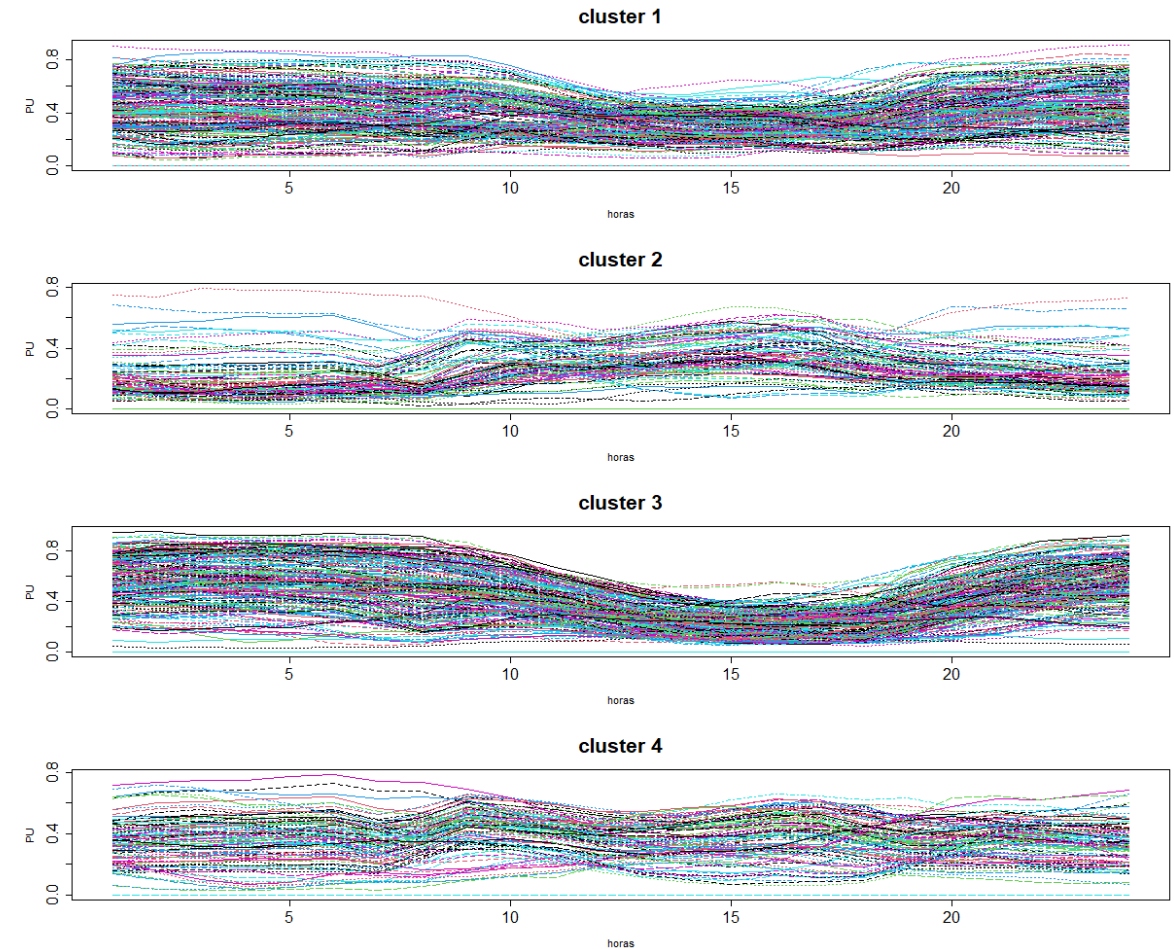   - Classify the performance curve of different turbine models

Machine Learning

Classification

```
resultado.hc = hclust(dist(dadospu),method='ward.D',members=NULL)
```

```
clusters = cutree(resultado.hc,k=4)
```

```
table(clusters)
```

Machine Learning

Classification

PROS

- Good for comparing similar curves

CONS

- Hard to understand which method is the best

- If the data is too divergent, you get too many clusters, or a too generic one

Content

Data Collection (Web Scraping)

Data Analysis (Reporting)

Machine Learning:

Classification

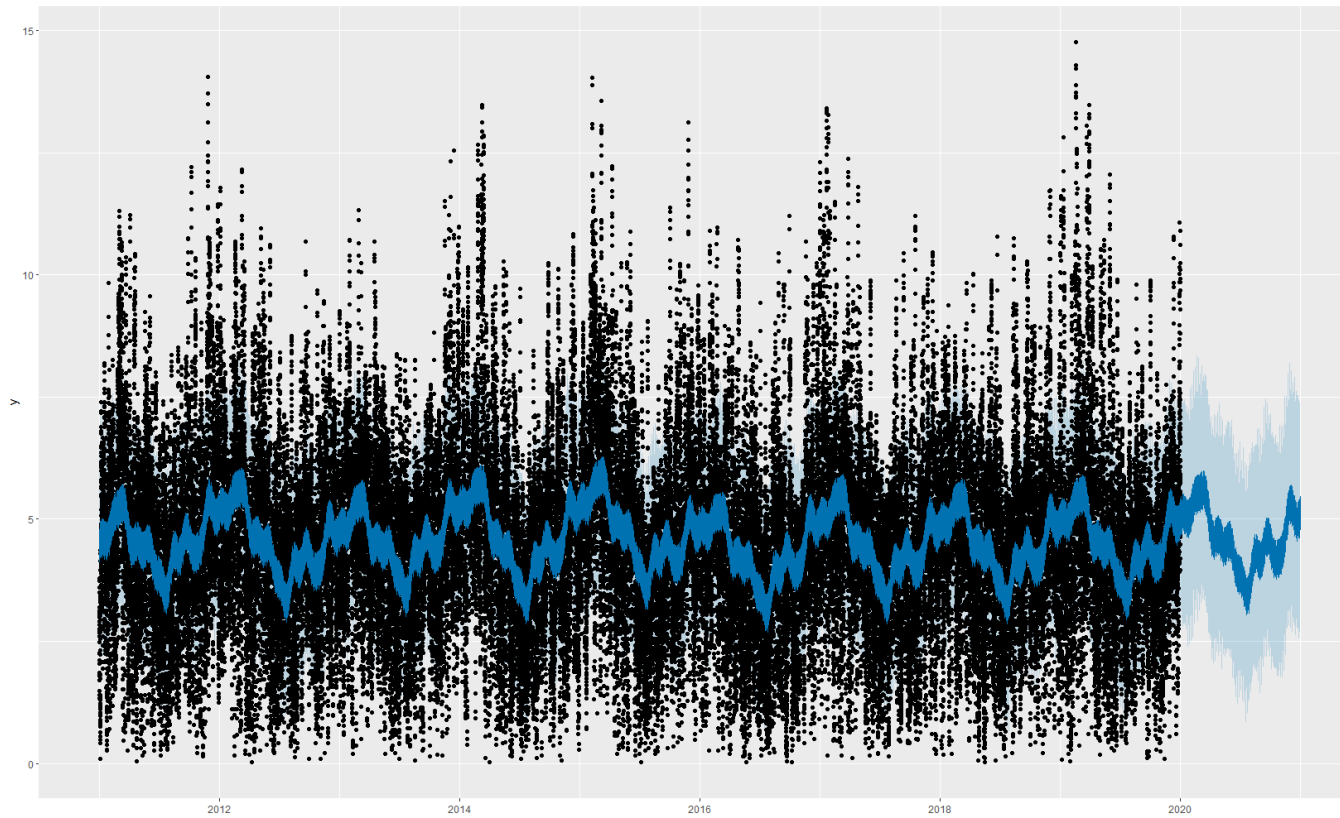Forecasting

Machine Learning

Forecasting

- When is the best time for a maintenance on my wind turbine? When is the best time for assembling a wind turbine?

    o Forecasting of wind data given a long historical period

- Answer: when the wind speed is low ☺

Machine Learning

Forecasting


```r
# function to fit the model
m <- prophet(df)

# Prediction to be made with the prophet object
future <- make_future_dataframe(m, periods = 8760, freq=3600)

# Make the predictions
forecast <- predict(m, future)

# Plot full data
plot(m, forecast)

# Plot components
prophet_plot_components(m, forecast)
```
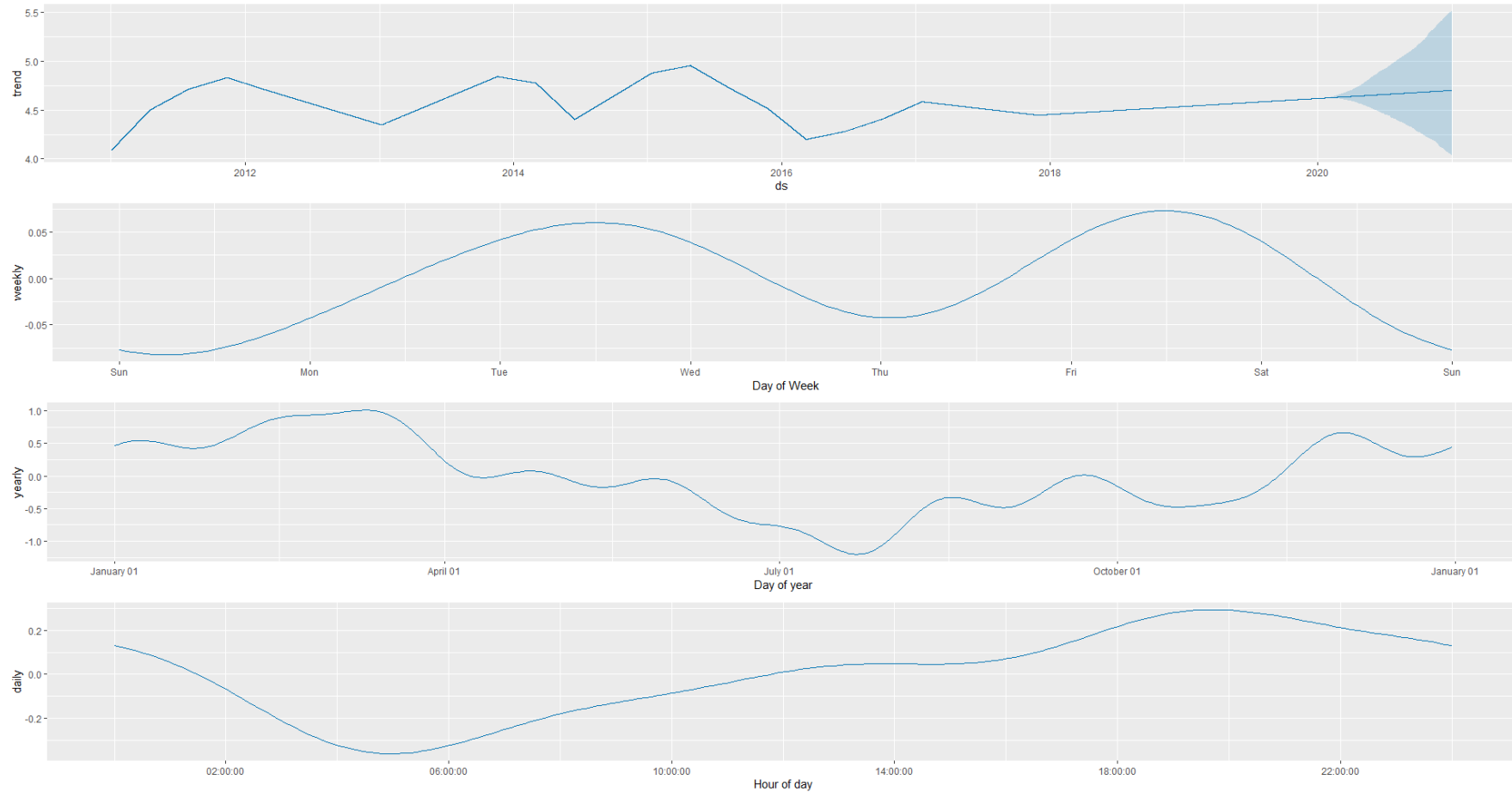
Machine Learning

Forecasting

Machine Learning

Forecasting

PROS

- Easy-to-use

- You can specify different parameters

CONS

- Computational power

- Not so good for weather data

Thank you!

afonso.lugo@epowerbay.com