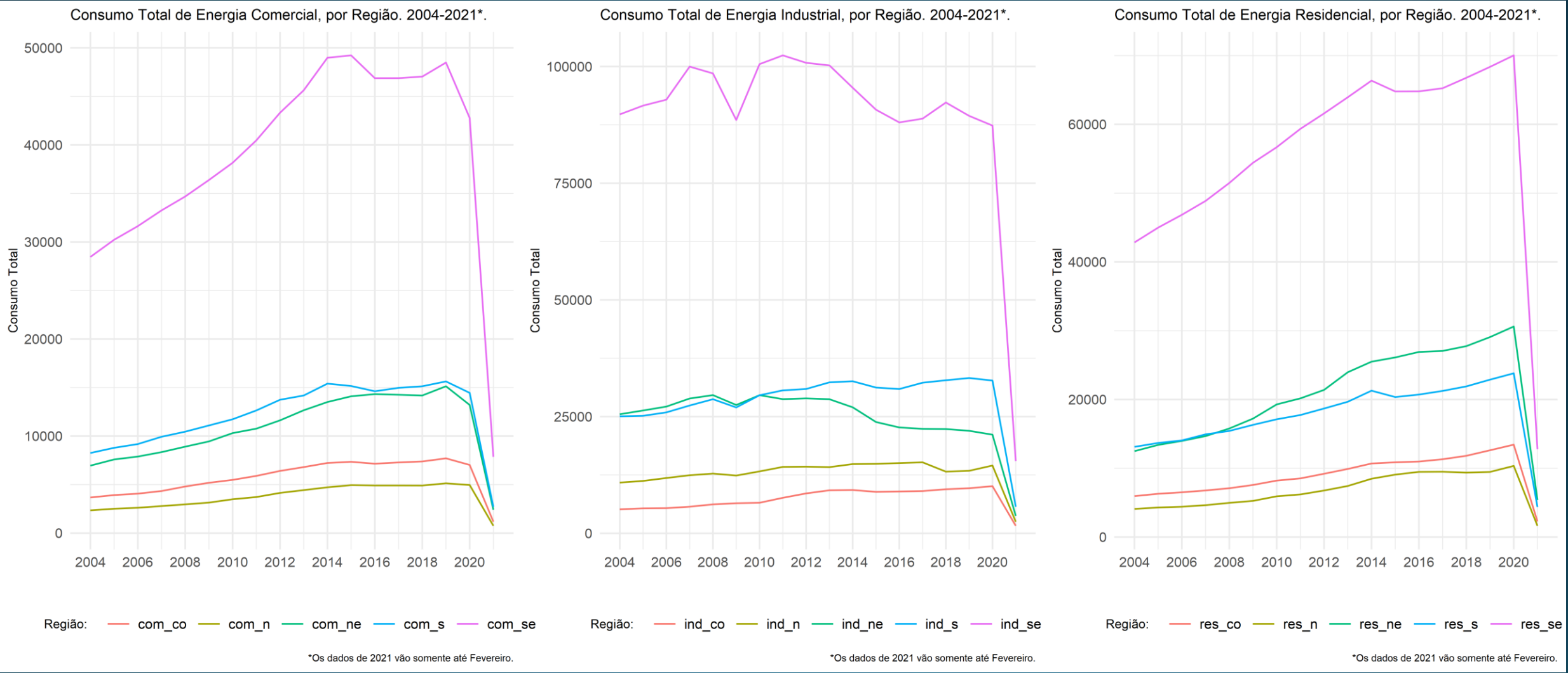


Resolução da Prova do Processo Seletivo da 4intelligence

Por
Livia Silva Sousa

1ª QUESTÃO

Gráfico 1 :

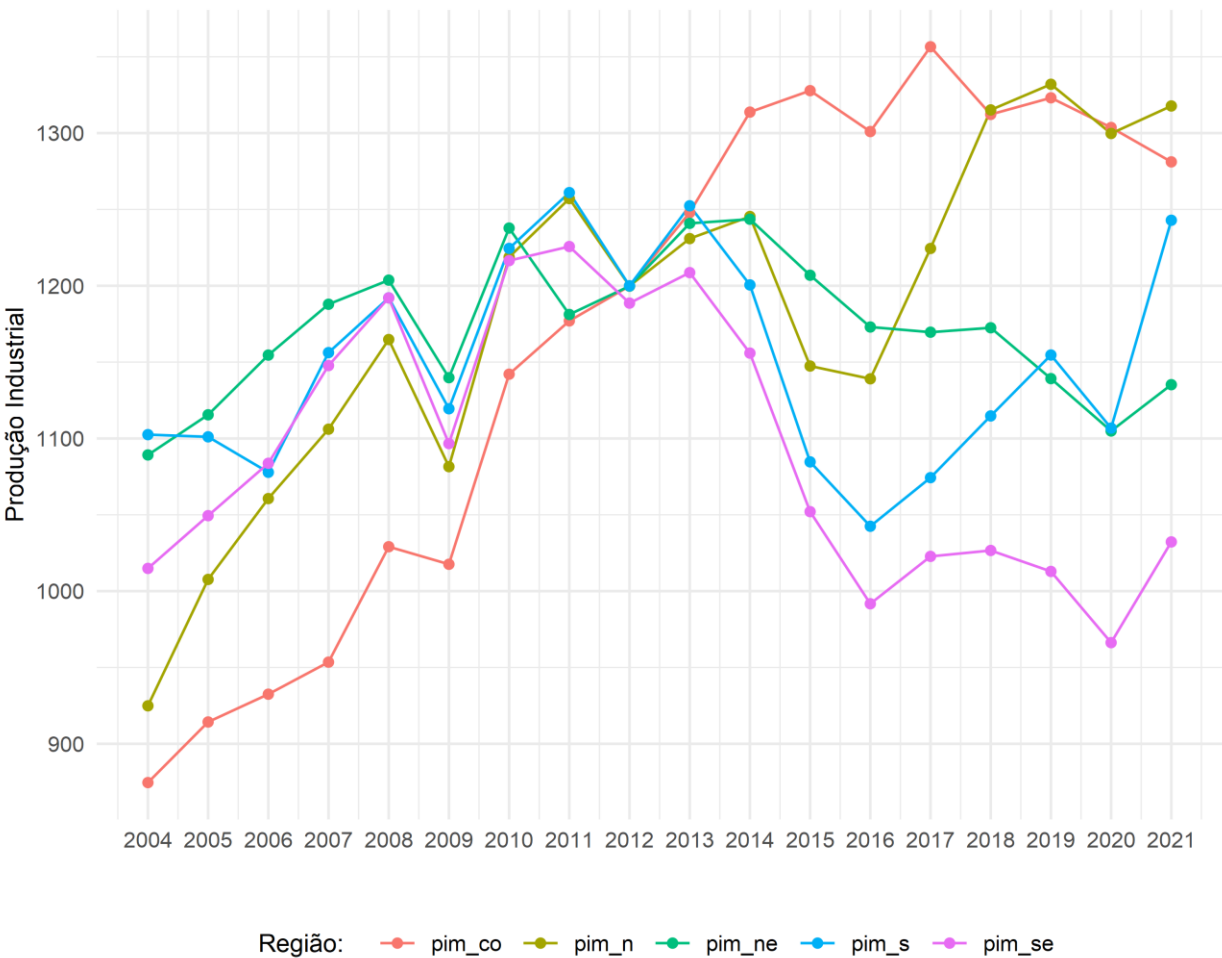


Análise do Gráfico 1

A Indústria é o setor que mais demanda consumo de energia, seguida por Residencial e Comercial. Os setores comercial e residencial possuem um padrão de crescimento similar ao longo do tempo. Nota-se um crescimento constante desses setores entre os anos de 2004 e 2014, seguida por uma queda no ano de 2015 e uma breve recuperação a posteriori. Já o consumo de energia apresenta um padrão relativamente constante ao longo do período. A região Sudeste é a região com maior consumo de energia nas três esferas. O consumo de energia residencial da Região Nordeste ultrapassa a Região Sul a partir de 2008.

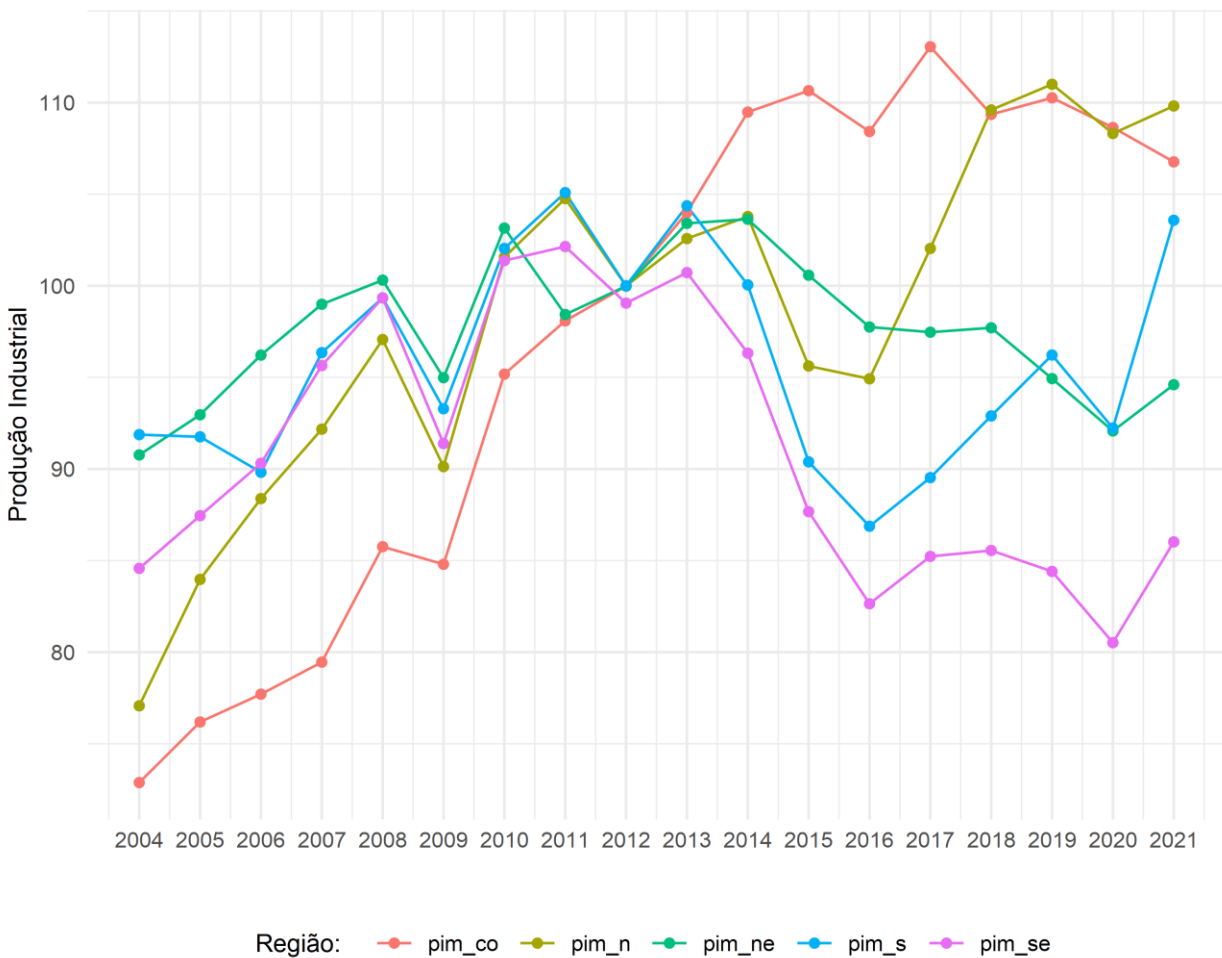
Gráfico 2 :

Produção Industrial Total, por Região. 2004-2021.



*Os dados de 2021 vão somente até Fevereiro.

Produção Média Industrial, por Região. 2004-2021.



*Os dados de 2021 vão somente até Fevereiro.

Análise do Gráfico 2

A partir do gráfico nota-se um crescimento relativamente constante da produção industrial total da região Centro – Oeste que deixa de ocupar o último lugar em 2004 para se tornar a região com maior produção industrial em 2014, todavia, a região Centro – Oeste perde o posto de primeiro lugar para região Norte, isso ocorre em 2018.

2ª QUESTÃO

- 1) O Objeto da análise caracteriza-se por ser um serie temporal. Portanto, a técnica utilizada foi uma previsão (Forecasting) utilizando o algoritmo de previsões do Facebook, o Prophet. A partir da modelagem percebe-se que a previsão a ponta um ritmo de queda no consumo de energia para os próximos 24 meses. Ao longo da serie percebe-se um tendência de crescimento entre 2004 até meados de 2011 seguindo por um período de desaceleração no final de 2011 até o momento. Nota-se também diversos pontos de mudança no comportamento da serie no período de 2010 até 2015 e 2015 - 2020.

Gráfico 4

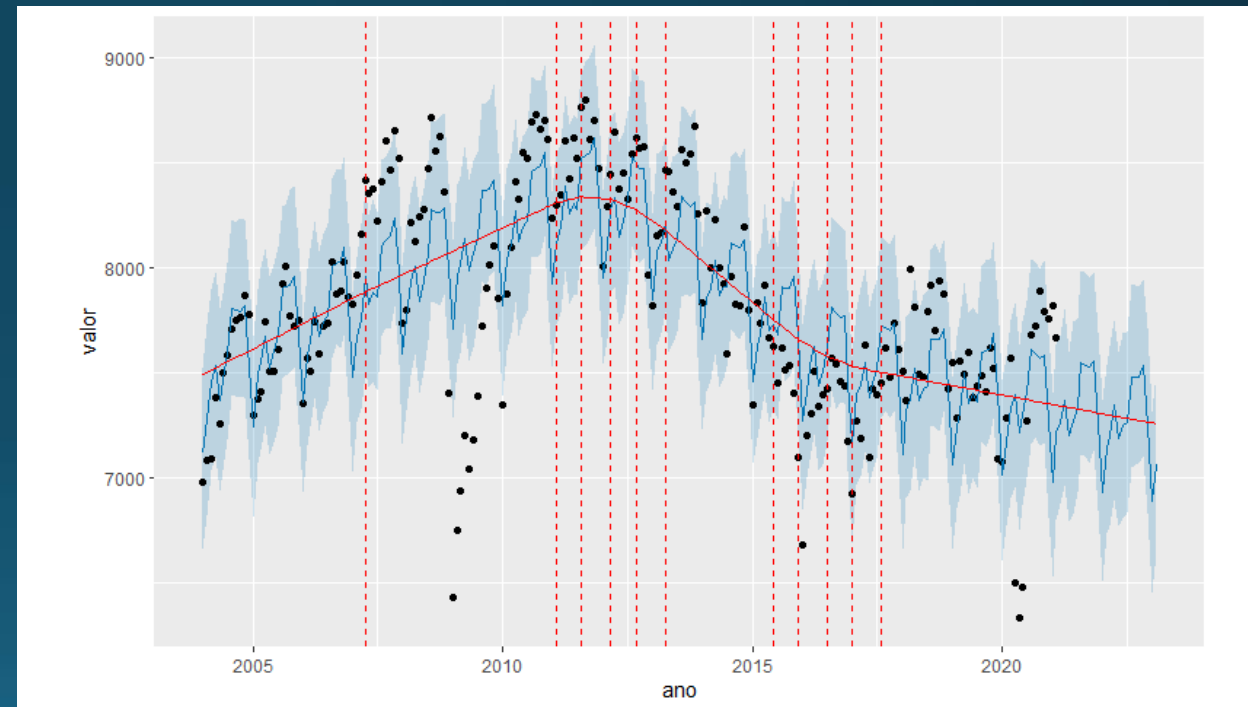
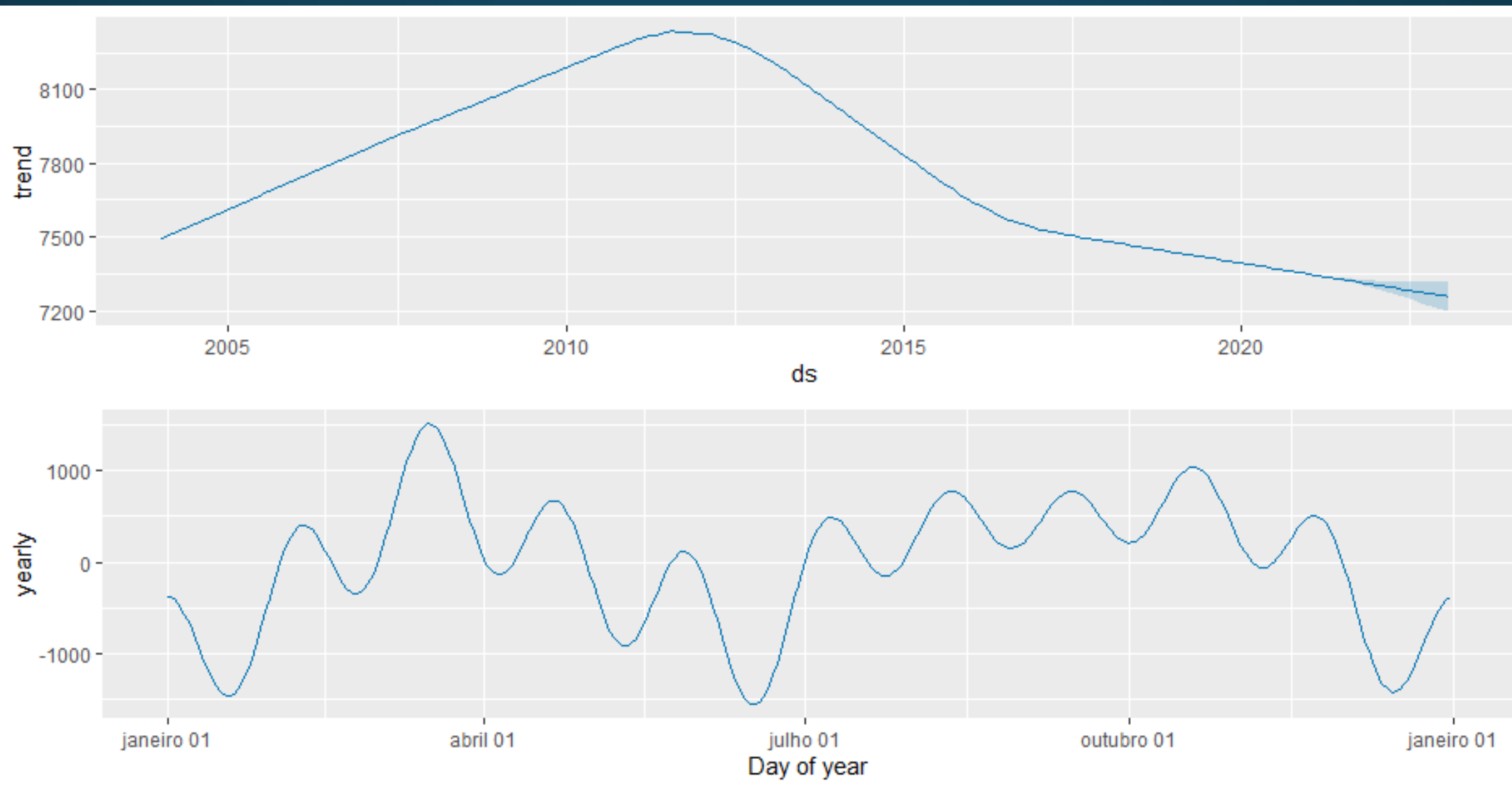


Gráfico 4: Tendência da Série



2) O modelo se mostrou estatisticamente significativo, demonstrado que os parâmetros estimados são robustos e que o modelo pode ser utilizado para previsão.

3) O modelo utilizando não fez uso de variáveis explicativas visto que a abordagem utilizada foi de um algoritmo de Machine Learning aplicado à série temporais.

```
Call:
lm(formula = dataframe_acuracia$y ~ dataframe_acuracia$modelo)

Residuals:
    Min       1Q   Median       3Q      Max
-1267.16  -146.89   41.57   205.32   907.82

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -620.53884   493.33186   -1.258    0.21
dataframe_acuracia$modelo    1.07932    0.06295   17.146 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 333.5 on 204 degrees of freedom
Multiple R-squared:  0.5904,    Adjusted R-squared:  0.5884
F-statistic: 294 on 1 and 204 DF,  p-value: < 0.00000000000000022
```

3ª QUESTÃO

Análise da 3ª Questão

A modelagem utilizando o algoritmo Prophet é voltando especificamente para uso preditivo em séries temporais, o uso dessa ferramenta não engloba outros modelos como o Random Forest, KNN, Naive Bayes , etc. O modelo adotado pelo Prophet se assemelha a vetores auto-regressivos. Para o desafio, resolvi trazer esta ferramenta apesar de ter conhecimentos dos modelos tradicionais presentes no *tidymodels*.

4^a QUESTÃO

Análise da 4ª Questão

Conclui-se que o papel do Cientista de Dados é realizar uma análise robusta dos dados, buscando extrair *insights* a partir das diversas informações na base de dados e através de ferramentas avançadas de modelagem realizar previsões que possam trazer retornos positivos para a empresa.