

Analytics para Produção Industrial

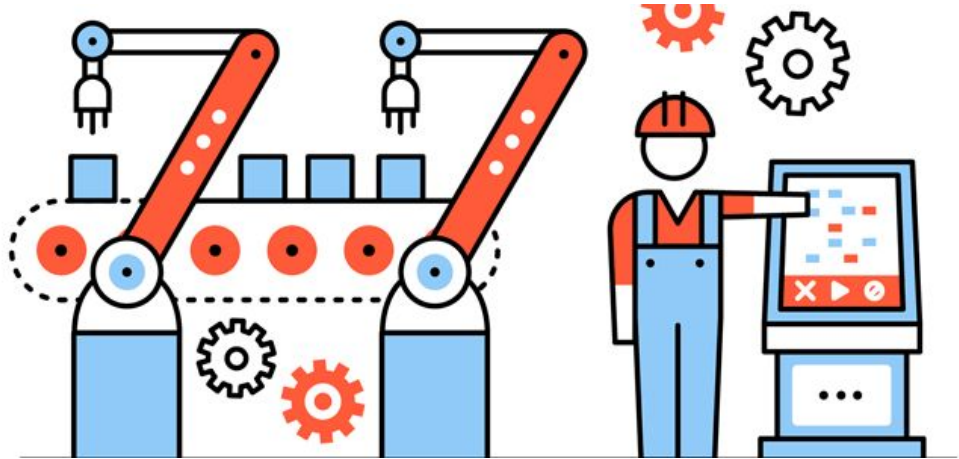
Laislla Pinheiro Brandão

Dezembro, 2022



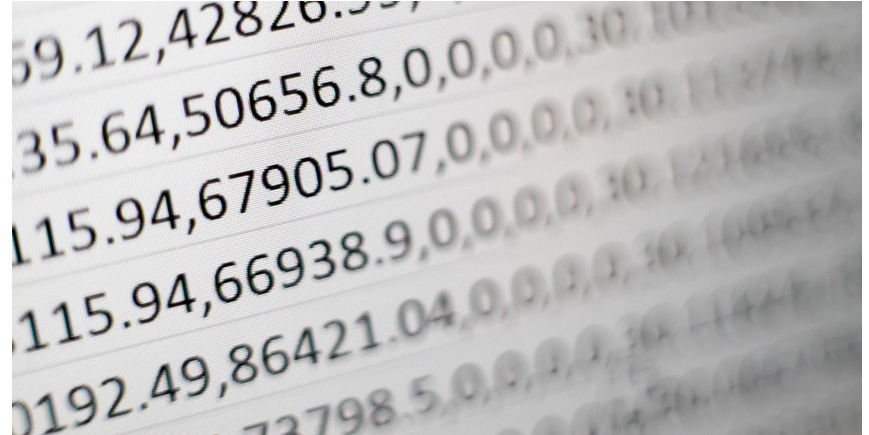
Contexto do projeto

- Uma unidade fabril que fabrica diariamente três produtos diferentes: *Produto 1*, *Produto 2* e *Produto 3*.



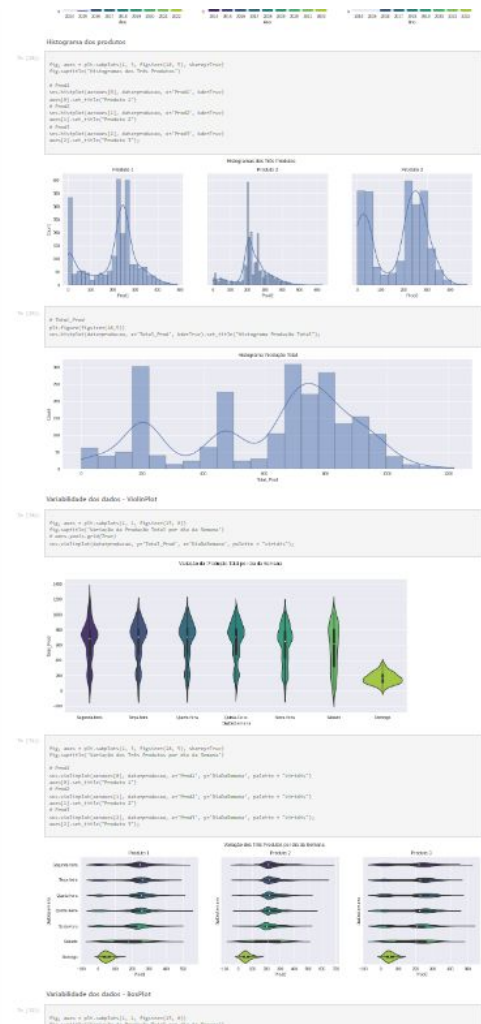
Objetivos

- Desenvolver um projeto de Análise ou Ciência de Dados sobre a base disponibilizada da produção.
- Obter *insights* explorando a base de diferentes formas.



Ferramentas

O projeto é desenvolvido utilizando o notebook do Google Colab, e todas as manipulações de dados, pré-processamentos, visualizações e modelagem são feitas utilizando a linguagem Python.



Processo



DADOS



Dados Originais

Os dados originais correspondiam às **datas de produção** e as **quantidades produzidas** para os 3 produtos.

	Data	Prod1	Prod2	Prod3
0	2014-12-01	0.0	11.0	0.0
1	2014-12-02	0.0	11.0	0.0
2	2014-12-03	0.0	11.0	0.0
3	2014-12-04	0.0	14.0	0.0
4	2014-12-05	0.0	13.0	0.0
...
2245	2022-10-27	142.0	264.0	272.0
2246	2022-10-28	155.0	302.0	285.0
2247	2022-10-29	124.0	234.0	225.0
2248	2022-10-30	46.0	77.0	70.0
2249	2022-10-31	31.0	77.0	58.0

2250 rows × 4 columns

Como as datas não estavam em ordem cronológica, foram ordenadas no momento da leitura do arquivo porque isso viria a ser útil mais à frente.

No total, a base possui **2250** registros de produção.

Transformações

A coluna de Data estava no formato de Texto e foi transformada para formato de DateTime, de forma a possibilitar a criação de **novas colunas** para agrupamento de dados posteriormente.

Além disso, a coluna com o **Total da Produção** foi criada.

	Data	Prod1	Prod2	Prod3	Total_Prod	Semana	Mes	Ano	Ano-Mes	Ano-Semana	DiaDaSemana
0	2014-12-01	0.0	11.0	0.0	11.0	49	12	2014	2014-12	2014-49	Segunda-feira
1	2014-12-02	0.0	11.0	0.0	11.0	49	12	2014	2014-12	2014-49	Terça-feira
2	2014-12-03	0.0	11.0	0.0	11.0	49	12	2014	2014-12	2014-49	Quarta-feira
3	2014-12-04	0.0	14.0	0.0	14.0	49	12	2014	2014-12	2014-49	Quinta-Feira
4	2014-12-05	0.0	13.0	0.0	13.0	49	12	2014	2014-12	2014-49	Sexta-feira
...
2245	2022-10-27	142.0	264.0	272.0	678.0	43	10	2022	2022-10	2022-43	Quinta-Feira
2246	2022-10-28	155.0	302.0	285.0	742.0	43	10	2022	2022-10	2022-43	Sexta-feira
2247	2022-10-29	124.0	234.0	225.0	583.0	43	10	2022	2022-10	2022-43	Sábado
2248	2022-10-30	46.0	77.0	70.0	193.0	43	10	2022	2022-10	2022-43	Domingo
2249	2022-10-31	31.0	77.0	58.0	166.0	44	10	2022	2022-10	2022-44	Segunda-feira

2250 rows × 11 columns

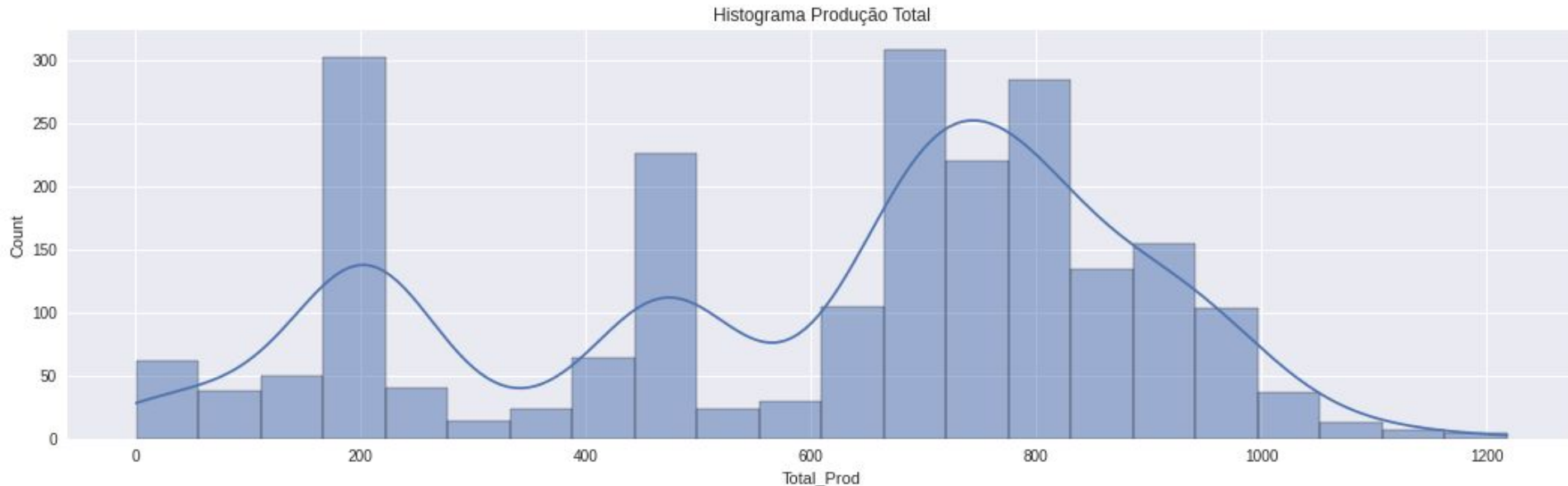
EXPLORAÇÃO DOS DADOS



Histograma da Produção Total

Os dados são bem distribuídos, não existe uma constante nos valores de produção.

Por esta imagem da distribuição, é possível ver bem como se comporta a produção em dias produtivos, semi-produtivos e não-produtivos.



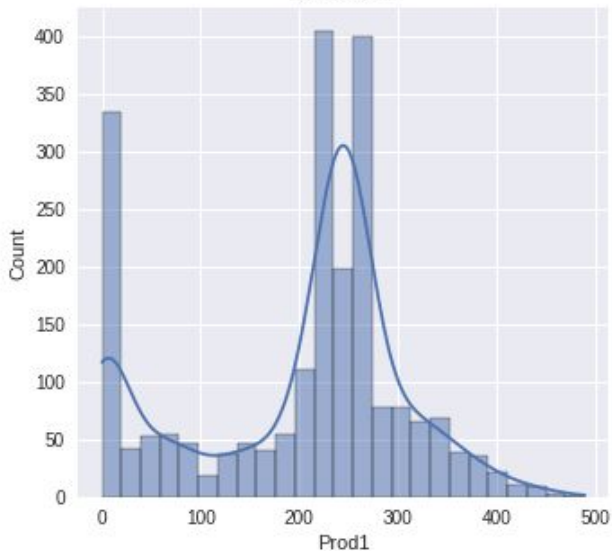
Histogramas dos Produtos

As distribuições para cada um dos produtos não é parecida, embora as médias principais estejam ao redor de um valor similar quando analisada a base inteira.

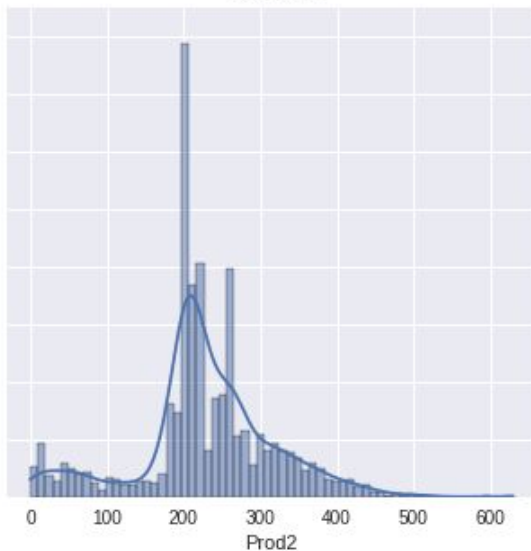
O produto 3, por exemplo, apresenta bem menos dados em distâncias grandes da média, ou seja, *outliers*.

Histogramas dos Três Produtos

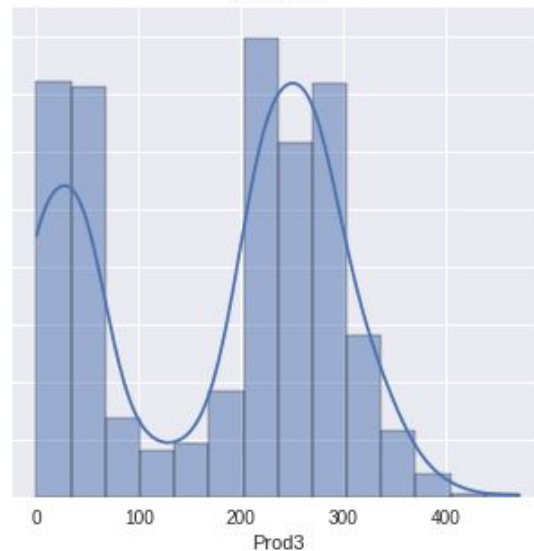
Produto 1



Produto 2



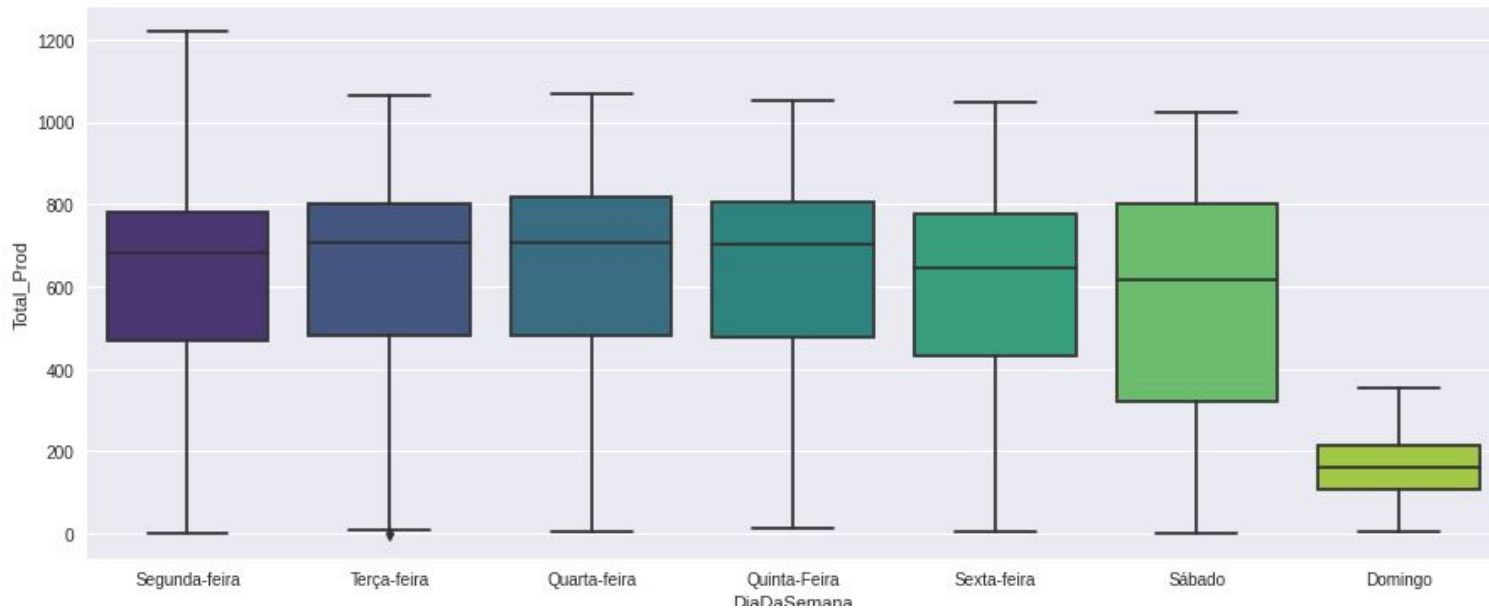
Produto 3



Variação da Produção Total

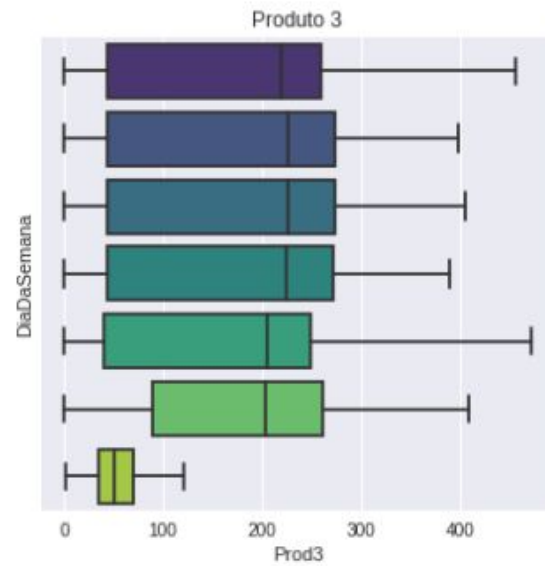
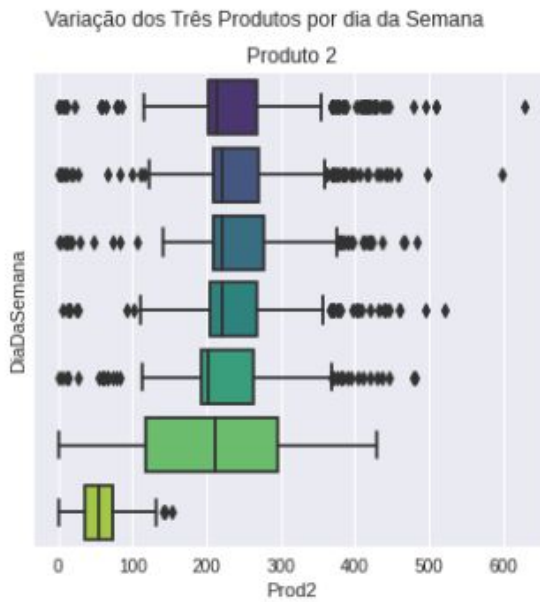
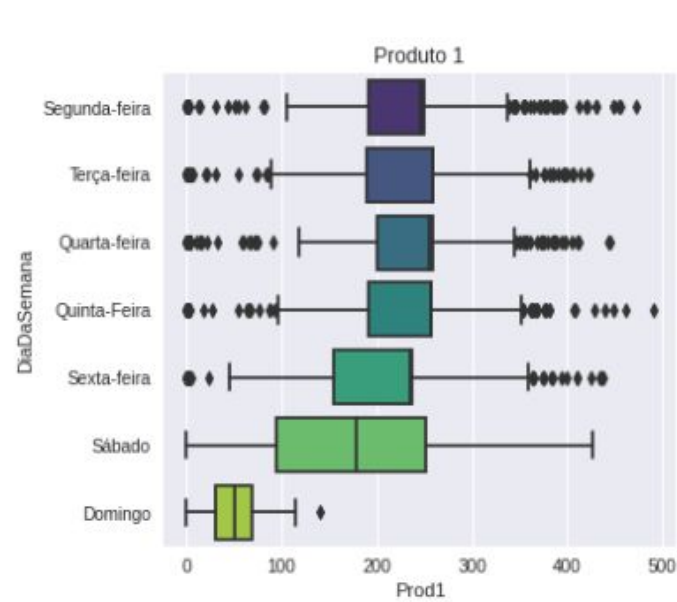
Variação da Produção Total por dia da Semana

- Notadamente, os sábados são semi-produtivos e os domingos são não-produtivos.
- Mas, existe um comportamento estranho nas sextas-feiras, por variar mais que os outros dias úteis.



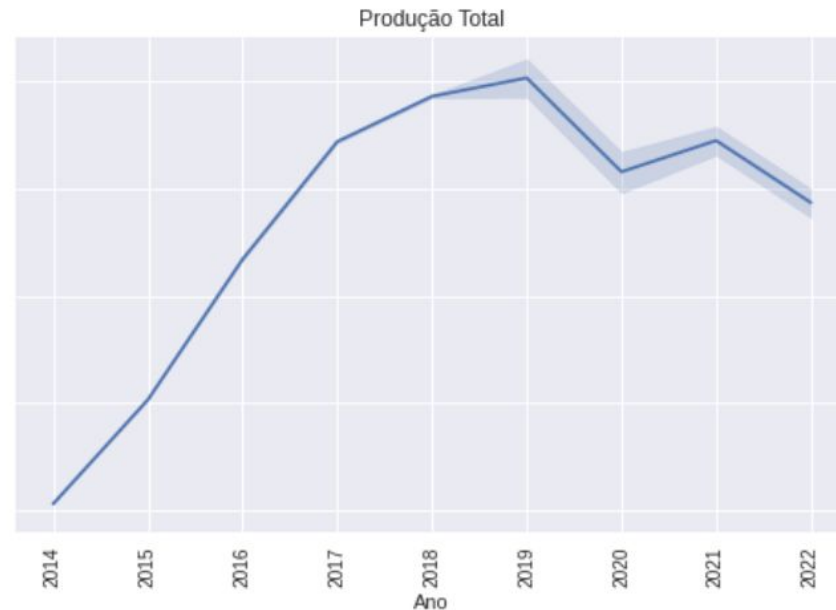
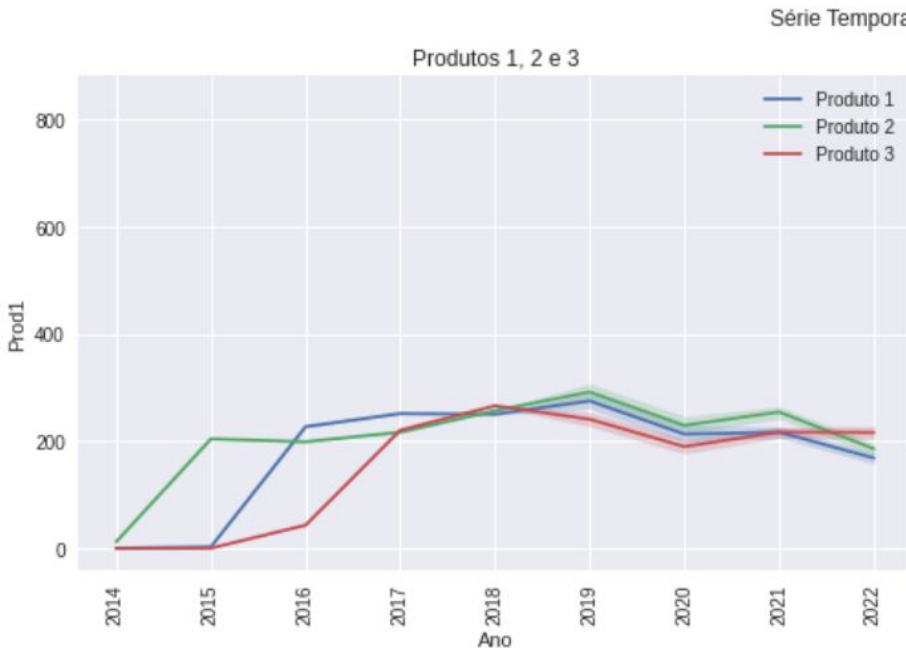
Variação dos Produtos

- Outliers não foram removidos propositalmente para utilizar os dados como séries temporais inteiras.
- Percebe-se que o produto 3 é o único cuja variação nos sábados é proporcionalmente menor que os outros.



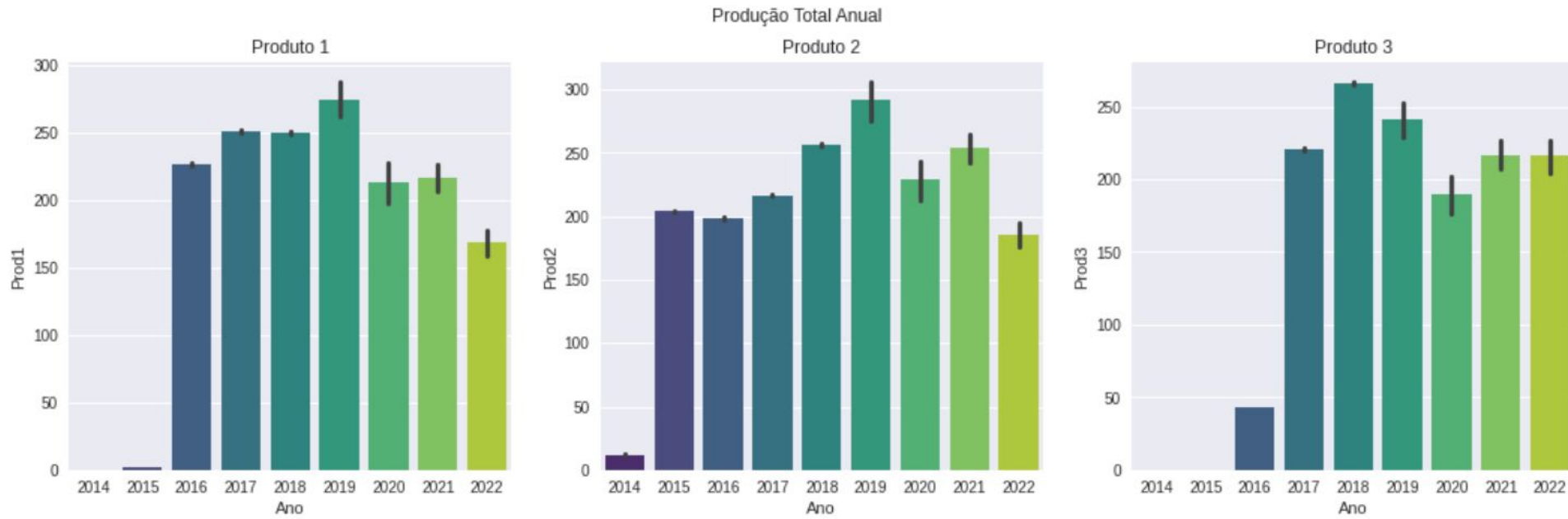
Produção Média Anual

- Os 3 produtos apresentaram queda de produção durante a pandemia, mas após ela, o Produto 3 aumentou sua produção, ultrapassando os Produtos 1 e 2 e alcançando sua produção média de 2021.
- Apesar disso, a produção total em 2022 ainda não retomou os patamares pré e logo após a pandemia.



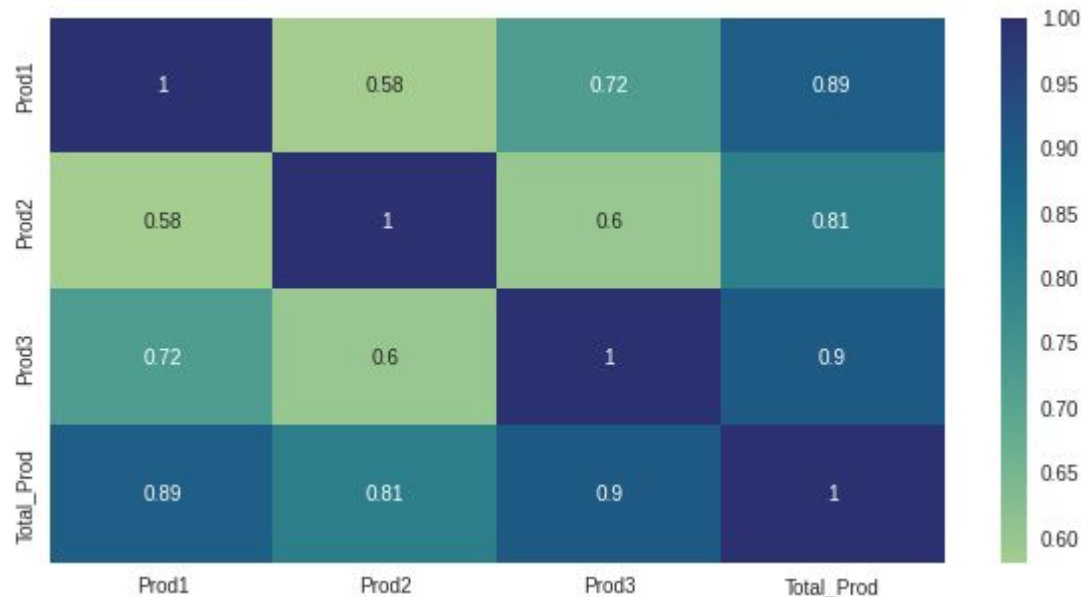
Produção Média Anual

Assim como visto anteriormente, em 2022 o Produto 3 está mantendo o aumento de produção pós-pandemia, enquanto os Produtos 1 e 2 ainda não retomaram.

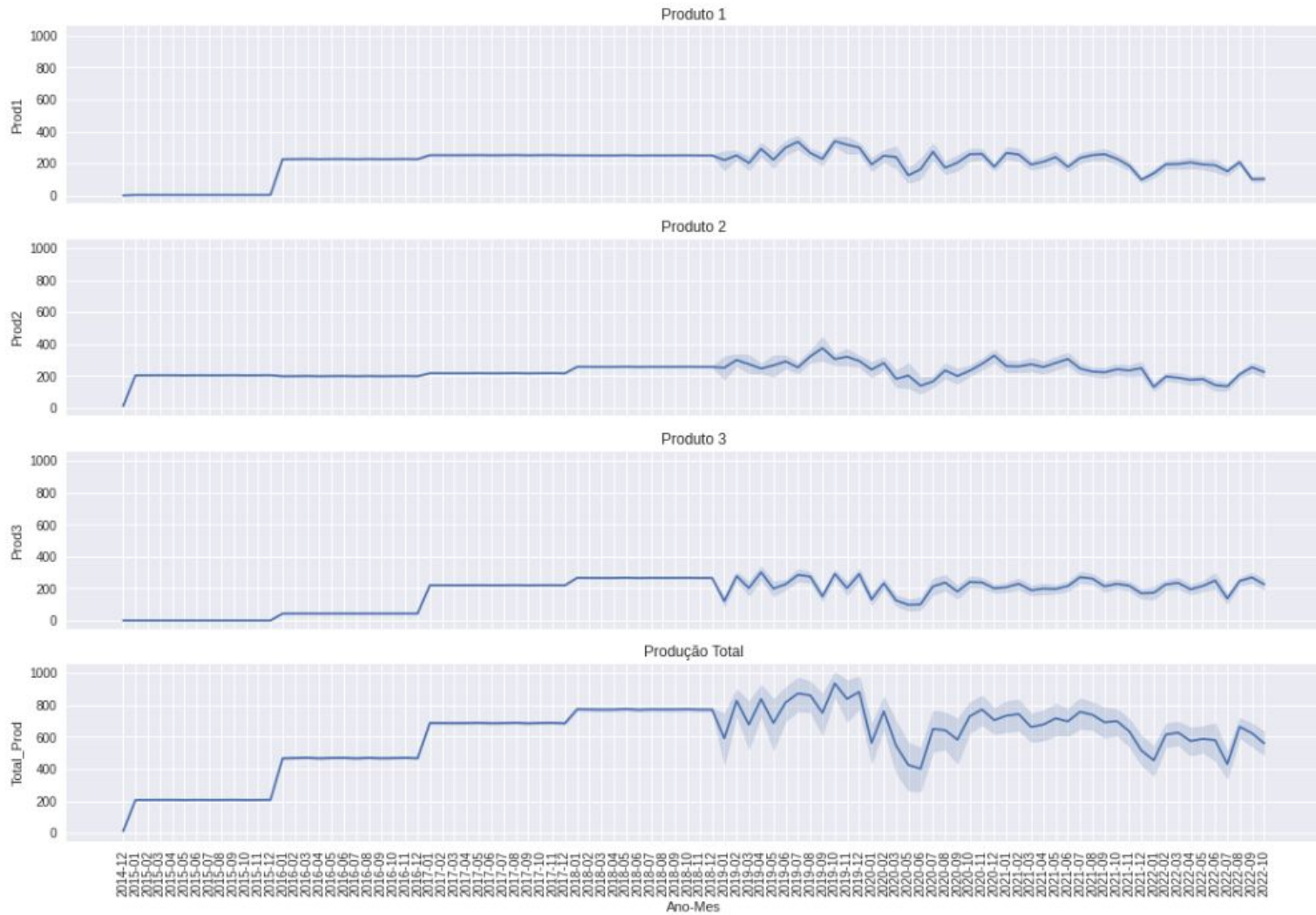


Correlação entre os Produtos

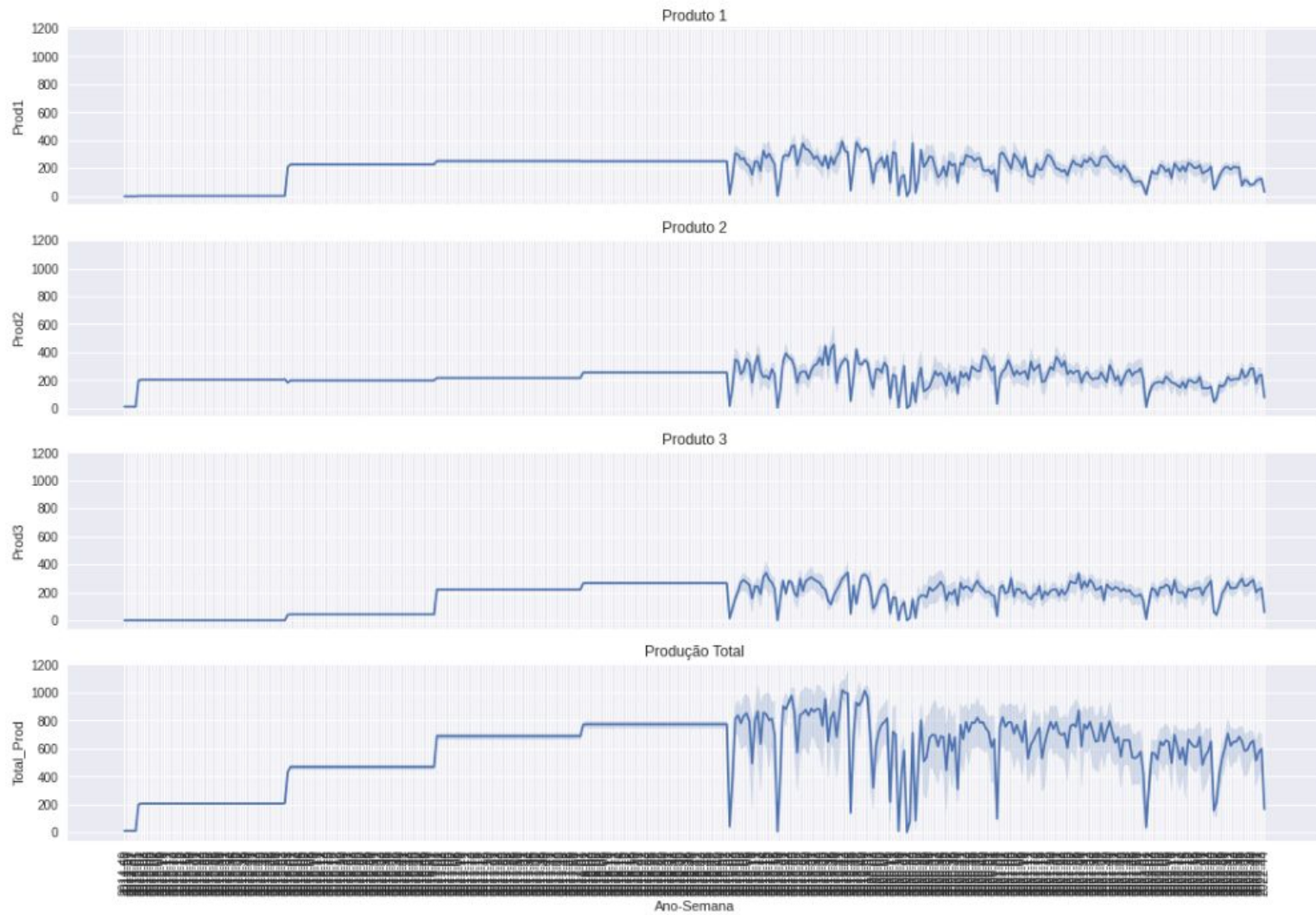
- Não é observada uma correlação significativa entre os produtos.
- De forma sutil, a produção total depende menos do Produto 2, mesmo ele tendo uma produção média ligeiramente maior que os outros dois produtos de 2019 até 2021.



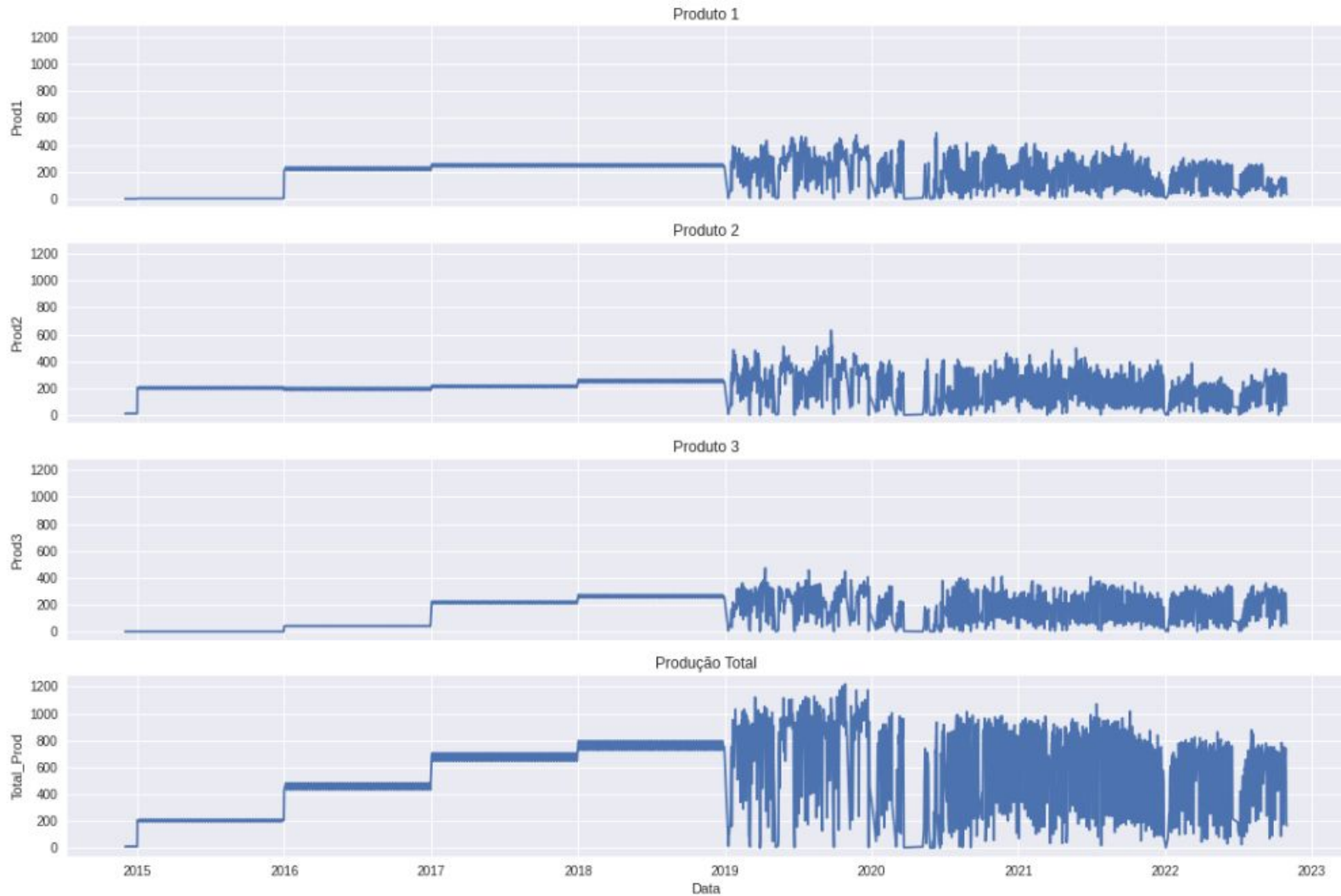
Produção Mensal



Produção Semanal



Produção Diária



MODELAGEM

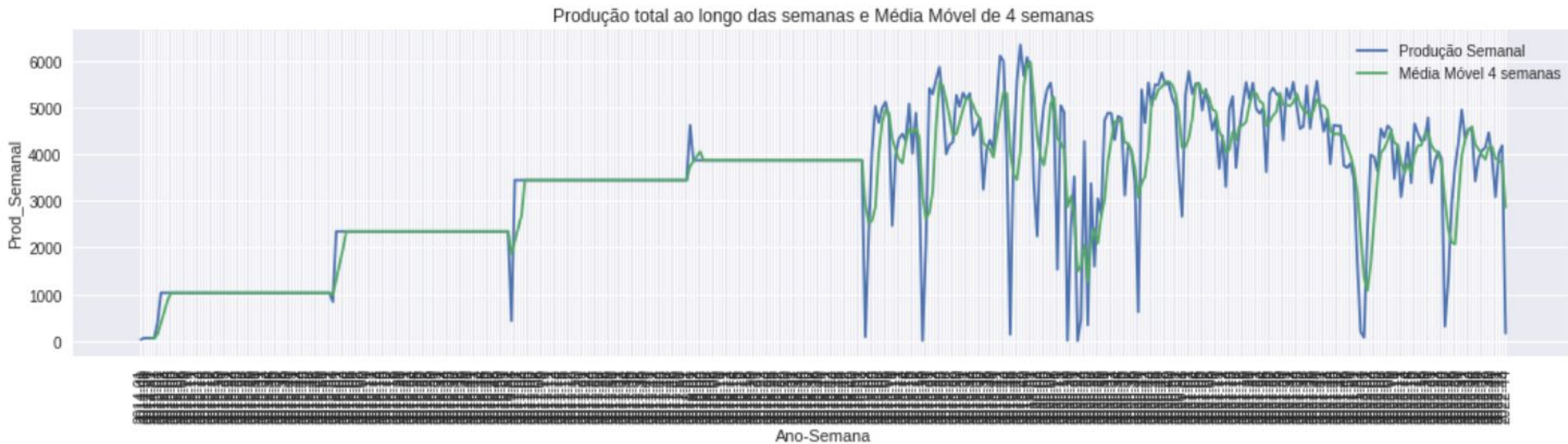


Séries Temporais

- Pelo fato de a produção ser um dado registrado em instantes de tempo, o mais intuitivo é pensar nela como uma série temporal.
- Em se tratando de séries temporais, a predição de dados em instantes futuros é interessante e será considerada como objeto de estudo neste projeto.
- Pela variabilidade da série quando considerada no grão de dia ou semana e por também considerar que seja mais útil a previsão mensal por questões de planejamento, a série de estudo será a **PRODUÇÃO TOTAL MENSAL**.

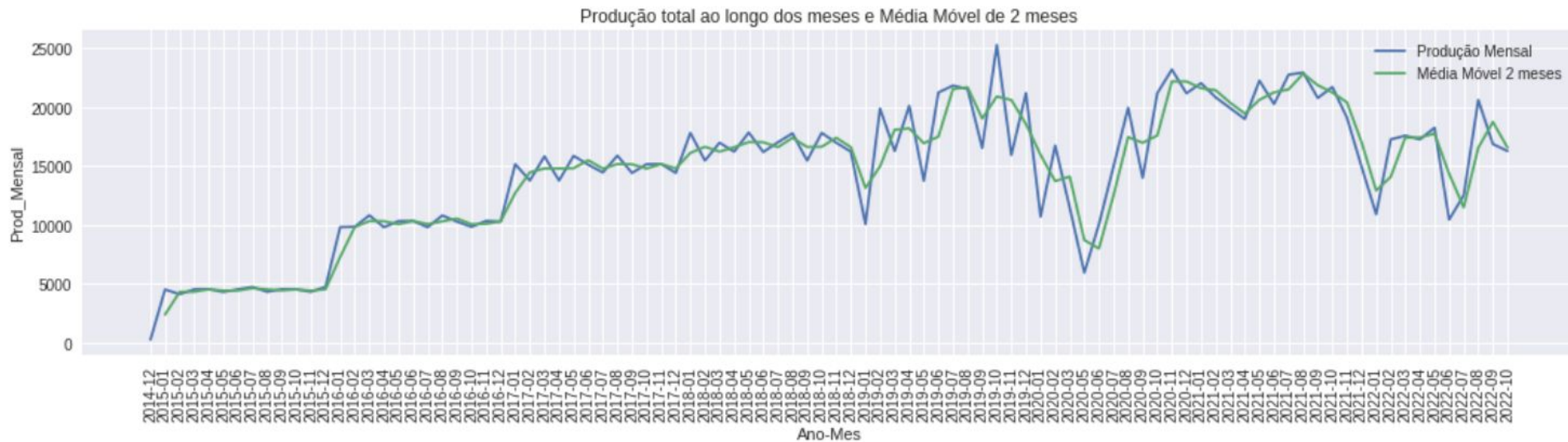
Série Semanal + MM

Para ilustração, esta é a série da Produção Total Semanal, junto com sua média móvel de 4 semanas.

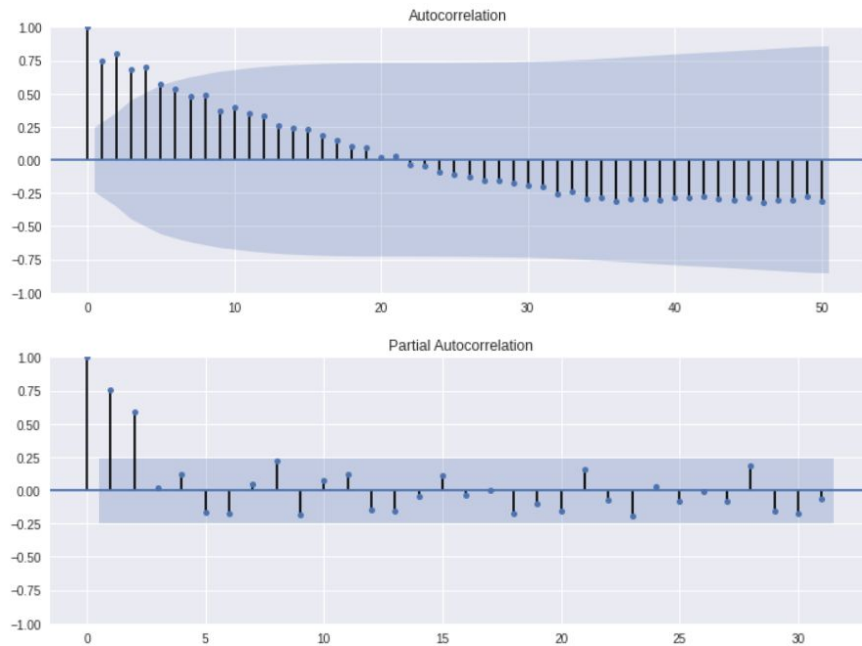


Série Mensal + MM

E esta é a série de estudo, que considera a Produção Total Mensal, aqui ilustrada junto com sua média móvel de 2 meses.



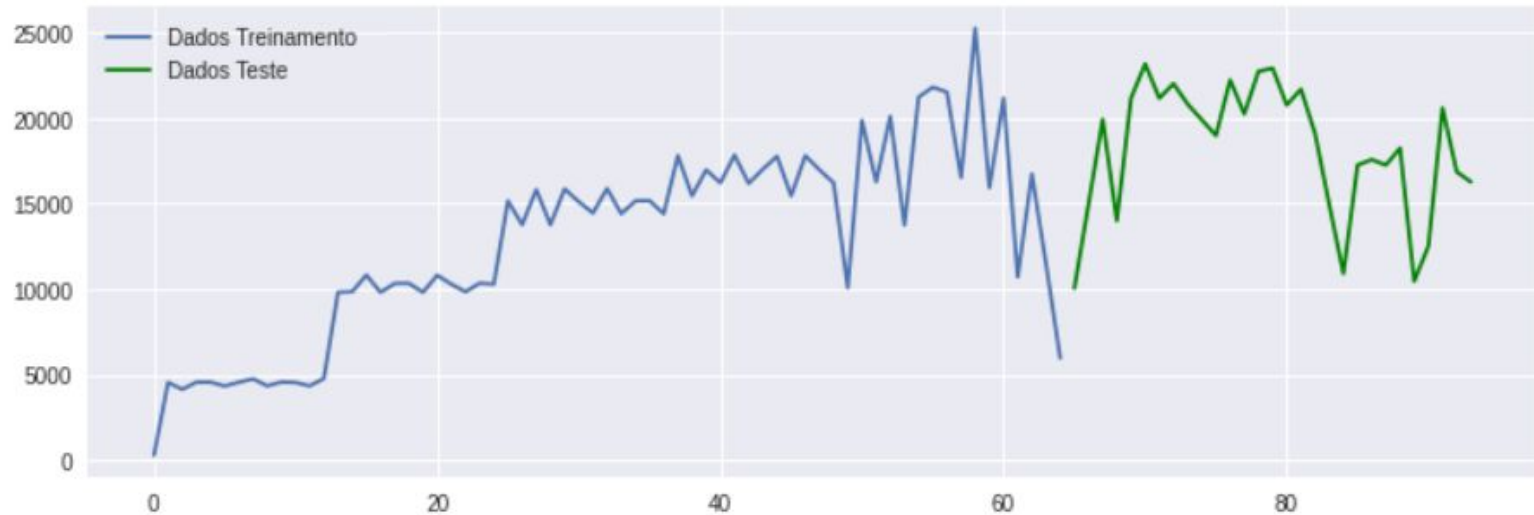
Autocorrelações e Estacionariedade



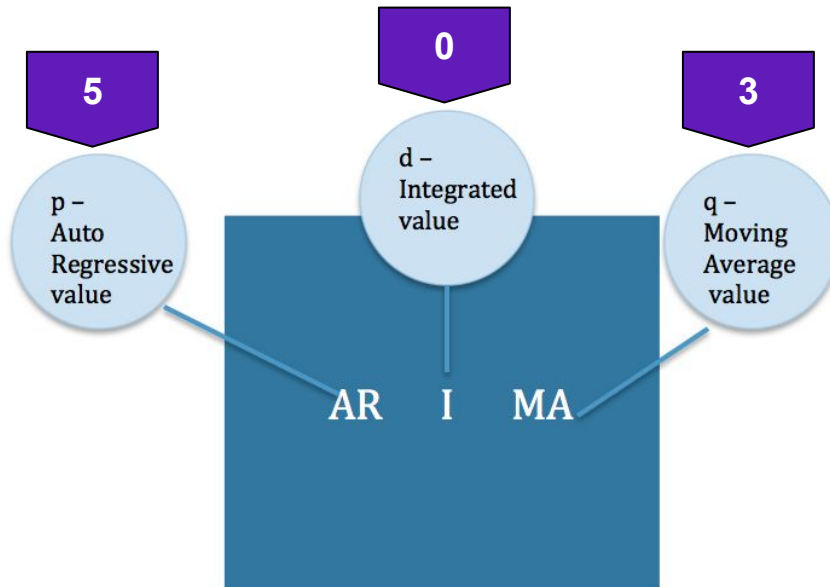
- É interessante observar as autocorrelações entre os horizontes de tempo da série (lags) para saber se ela é previsível ou se seu comportamento é fruto de aleatoriedade.
- Neste caso, **é previsível**, pois existem correlações significativas acima da zona de significância.
- Aplicando o teste KPSS (Kwiatkowski Philips Schmidt e Shin), verificou-se que **a série é estacionária**.

Divisão de Dados para o Modelo

Foram considerados os primeiros **70%** dos dados para Treinamento do modelo e **30%** dos dados seguintes para Teste.



Parâmetros ARIMA



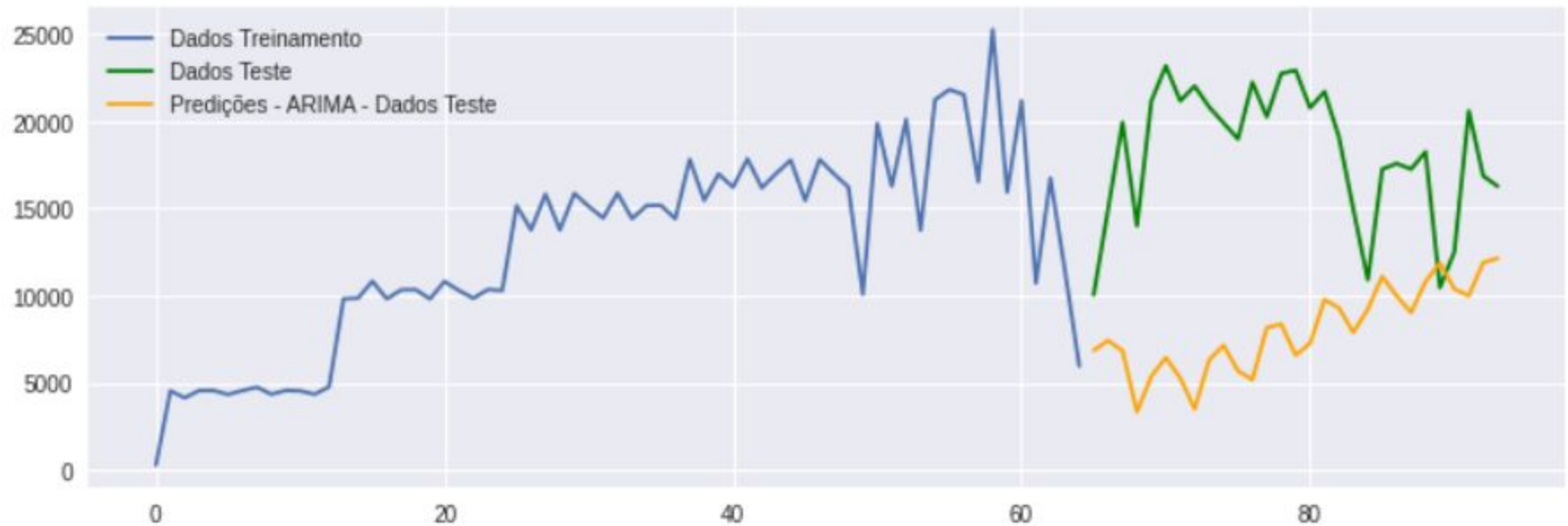
O modelo ARIMA é o mais utilizado para previsões de séries temporais e, por este motivo, será testado nos dados de produção.

Este modelo é muito bom para modelar os componentes lineares de uma série temporal e possui estes **três parâmetros**.

Como esta série é estacionária, o parâmetro "d" é zero.

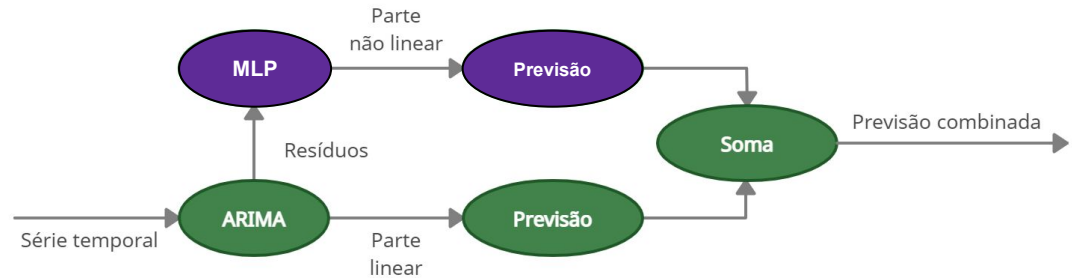
Os outros 2 parâmetros podem ser inferidos pelas autocorrelações e também junto a um grid search.

Resultado do ARIMA

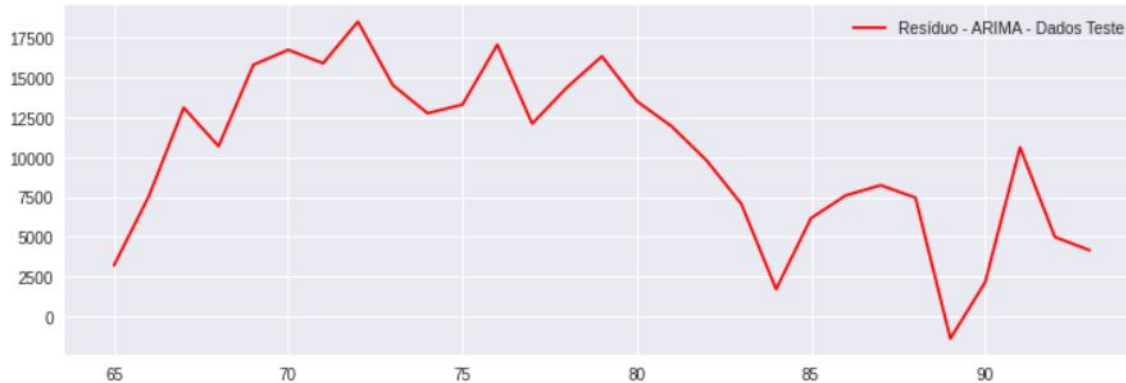


Melhoria da Performance

Como o modelo ARIMA não conseguiu modelar bem os componentes Não Lineares da série, será realizada uma **hibridização** no modelo, incluindo uma etapa de modelagem Não-Linear para **melhorar o desempenho da previsão**.



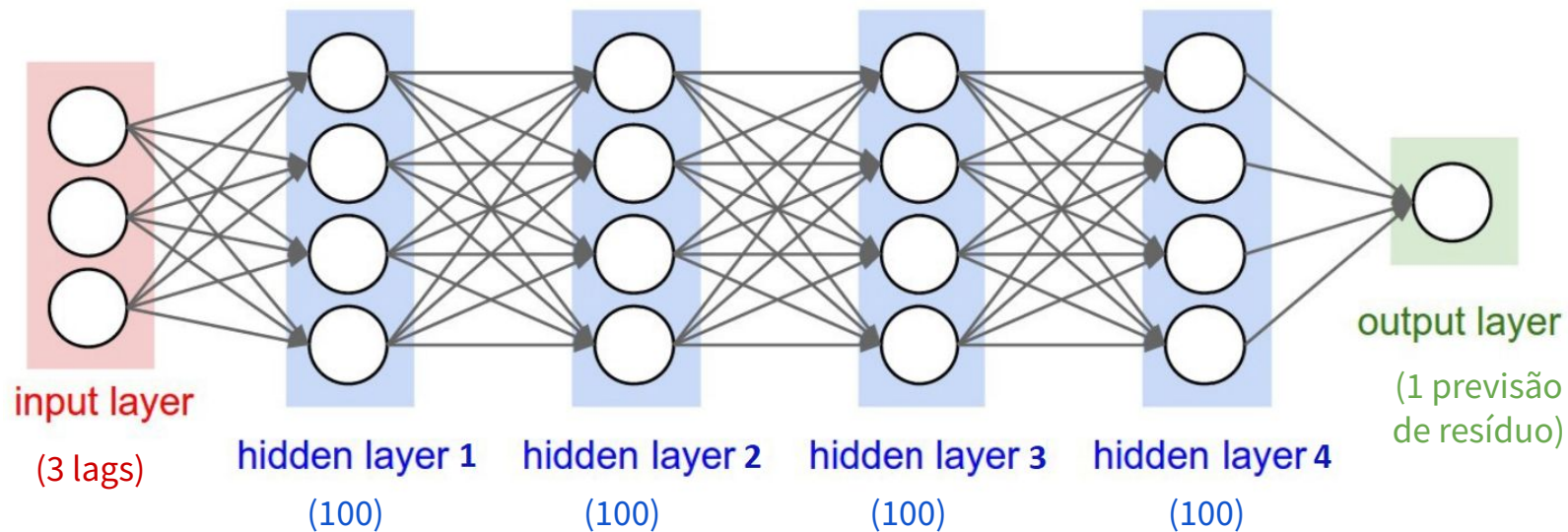
Modelagem dos Resíduos



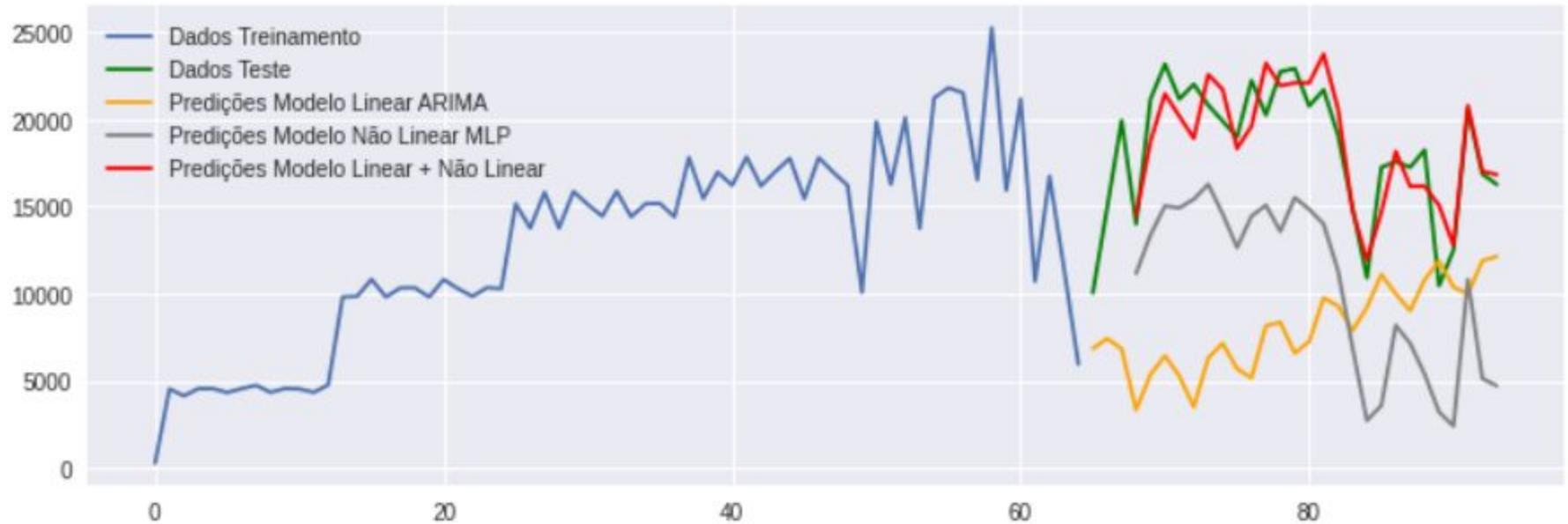
Os resíduos agora apresentam os componentes Não Lineares que não foram modelados pelo ARIMA e serão inseridos em um modelo MLP utilizando um **lag de 3** (definido pelas autocorrelações).

As previsões da MLP serão somadas às do ARIMA para obter a previsão final.

Arquitetura MLP



Resultado Final da Previsão



Comparação

Modelo	Métrica MSE (Erro Médio Quadrático)	Métrica RMSE (Raiz do Erro Médio Quadrático)
ARIMA (Modelagem dos componentes lineares)	130.926.832,96	11.442,32
Combinação ARIMA + MLP (Modelagem dos componentes lineares e não-lineares)	3.347.769,24	1.829,69

Considerações



Considerações

- O projeto mostrou uma ideia de início, meio e fim do uso de ciência de dados para tomada de decisão, juntando processamento dos dados, análise e modelos.
- Além de *insights* a respeito do comportamento dos dados, como eles variam de acordo com o dia da semana, como estão relacionados entre si e como se comportam ao longo do tempo, foi possível prever o comportamento dos dados em um instante no futuro.
- Outras técnicas poderiam ser acrescentadas, como uma classificação para indicar se a produção em um instante à frente estaria dentro de um target ou não, por exemplo.
- Além disso, outros modelos de previsão poderiam ser testados para comparar as performances com o modelo desenvolvido aqui, bem como o próprio modelo ser melhorado.

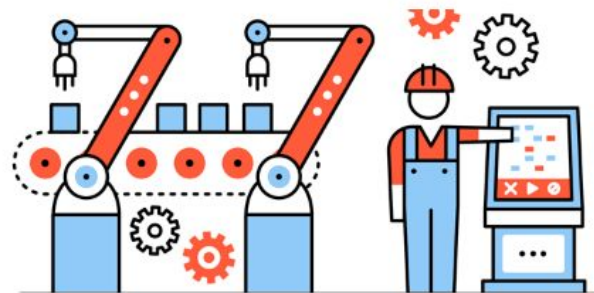
Projeto Completo

O projeto completo, notebook e dataset estão no GitHub para acesso público:

lcpbrandao/
Projeto_Analytics_Producao_Industrial

README.md

 LaislaBrandão



Projeto de Analytics para Produção Industrial

- Projeto para a etapa de seleção da Residência Tecnológica em Ciência de Dados e Analytics para o setor Automotivo.
- Parceria Universidade de Pernambuco/Poli - Fiat Chrysler Automobiles/Jeep - FACEPE
- Novembro, 2022
- DESAFIO: Desenvolver um projeto de Análise ou Ciência de Dados sobre a base disponibilizada e preparar um pitch para defendê-lo.



Principais Contribuições



Principais Contribuições



FCA Jeep



- Novos indicadores
- Novos reports
- Gestão de Microparadas
- Gestão de OEE

WCM Energy



- Co-pilar Energia Planta
- Aditoria Prata
- Extensão em Machine Learning para
- Análise de Dados C.E.S.A.R.

Spot Welding



- Projeto Pandora
- Data Analytics Center
- Mestrado em Engenharia da Computação
- Projeto Welding Wizard / Weld Ranger
- Automação OEE
- Nova gestão EWO PowerBI
- Nova gestão Cartão PM PowerBI

2018

2020

2016/2017

2019

2021/2022

Power BI



- Dashboards PM
- Nova gestão EWO PowerBI
- Nova gestão Cartão PM PowerBI
- Pilar Energia Funilaria

BigQuery e Data Studio



- Nova gestão EWO com Manusis/SQL/Data Studio
- Nova gestão EWO Planta
- Caça-vazamentos Funilaria

OBRIGADA!

