**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data**

**Frank Tetsuya Komati**

**ANÁLISE E PREVISÃO DE SÉRIES TEMPORAIS DE DADOS DE EXPORTAÇÃO DE MINÉRIO DE FERRO PELO ESTADO DO ESPÍRITO SANTO**

VITÓRIA

2023

**Frank Tetsuya Komati**

**ANÁLISE E PREVISÃO DE EXPORTAÇÃO DE MINÉRIO DE FERRO PELO ESTADO DO ESPÍRITO SANTO**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

VITÓRIA

2023**SUMÁRIO**

[1. Introdução 4](#_Toc131878634)

[1.1. Contextualização 4](#_Toc131878635)

[1.2. O problema proposto 7](#_Toc131878636)

[1.3. Objetivos 7](#_Toc131878637)

[2. Coleta de Dados 9](#_Toc131878638)

[3. Processamento/Tratamento de Dados 12](#_Toc131878639)

[4. Análise e Exploração dos Dados 17](#_Toc131878640)

[5. Criação de Modelos de Machine Learning 20](#_Toc131878641)

[6. Interpretação dos Resultados 30](#_Toc131878642)

[7. Apresentação dos Resultados 31](#_Toc131878643)

[8. Links 32](#_Toc131878644)

[REFERÊNCIAS 33](#_Toc131878645)

[APÊNDICE 34](#_Toc131878646)

# 1. Introdução

## 1.1. Contextualização

### Introdução

Antes de iniciar a Pós-Graduação em Ciência de Dados e Big Data pela PUC-MG, fiz dois MBAs sendo o meu primeiro em 2003, em Logística Empresarial. Naquela época trabalhava em um Porto Seco, ou Estação Aduaneira do Interior (EADI), ou *Dry Port*, no município de Cariacica-ES, até então entre as maiores do país.

### Portos

Relembrando meu passado e meu grande interesse no assunto visto que o estado do Espírito Santo possui uma privilegiada infraestrutura logística contando com um complexo de portos, que segundo o Governo do Espírito Santo ([Governo ES - Portos (www.es.gov.br)](https://www.es.gov.br/portos-2)) é construído dos seguintes portos:

* Terminal Norte Capixaba (Transpetro) – Localizado em Barra Nova (São Mateus), escoa o petróleo dos campos terrestres do norte do Espírito Santo por navios atracados em monoboia.
* Portocel – Localizado em Barra do Riacho (Aracruz), é especializado em celulose, sendo um dos mais eficientes do mundo, operando também com blocos de mármore e granito, além de produtos siderúrgicos.
* Terminal Aquaviário Barra do Riacho – Movimenta gás liquefeito de petróleo (GLP) e gasolina natural (C5+), resultantes do processamento de gás natural em Cacimbas.
* Complexo Portuário de Tubarão - Reúne cinco terminais, com estrutura comparável aos melhores portos do mundo:
* Terminal de Tubarão – Operado pela Vale, é o maior e mais eficiente terminal de exportação de pelotas e de minério de ferro do mundo.
* Terminal de Granéis Líquidos (TGL) – É especializado na movimentação de derivados do petróleo.
* Terminal de Praia Mole – É especializado em operações de descarga de navios com carvão (70% do volume importado pelas siderúrgicas), coque e antracito.
* Terminal de Produtos Diversos (TPD) – Tem um sistema integrado de recepção, manuseio e armazenagem de grãos (15% das exportações do Brasil) e fertilizantes (8% do volume importado).
* Terminal de Produtos Siderúrgicos (TPS) – Movimenta, além de produtos siderúrgicos, ferro gusa, mármore/granito e veículos. Operado pelo consórcio ArcelorMittal Tubarão, Usiminas e Gerdau Açominas.
* Complexo do Porto de Vitória - Administrado pela Companhia Docas do Espírito Santo (Codesa), movimenta carga geral por meio dos terminais Cais de Vitória, Capuaba, Peiú, Atalaia, Flexibrás, TVV e CPVV.
* Terminal Vila Velha (TVV) – terminal especializado em contêineres operado pela iniciativa privada (Log-In Internacional e Logística). É uma excelente alternativa para operações de importação e exportação de contêineres e carga geral, destacando-se como um dos mais produtivos terminais brasileiros nesse segmento.
* Companhia Portuária de Vila Velha - CPVV – atende às operações offshore de exploração e produção de petróleo no Espírito Santo.
* Porto de Ubu (Samarco) – Movimento de pelotas e de minério de ferro, granéis sólidos e carga geral. Utilizado em operações de supply boats para indústria de petróleo e outras.

Projetos de investimentos em Portos:

* Portocel (privado) - Terminal no município de Aracruz, especializado em celulose. Projeto de expansão para movimentação de carga geral.
* Imetame (privado) - Terminal no município de Aracruz, para apoio de atividades offshore com possibilidade de diversificação de uso.
* Itaoca Offshore (privado) – Terminal a ser instalado no município de Itapemirim, Sul do Estado, para atender às demandas de serviços logísticos offshore do mercado nacional de petróleo e gás.
* Porto Central (privado) – Parceria entre a TPK Logística e o Porto de Roterdã, maior porto marítimo da Europa. Porto multicargas: contêiner, granel sólido (soja, minério, etc), granel líquido e apoio offshore, no município de Presidente Kennedy, localizado no extremo Sul do Estado.
* Terminal Portuário de Urussuquara – Petrocity (privado) – Terminal de cargas gerais a ser instalado no município de São Mateus com projeto de atender também às demandas de serviços logísticos offshore do mercado nacional de petróleo e gás.

### Ferrovias

Além disso conta também com uma infraestrutura de transporte ferroviário, a Estrada de Ferro Vitória a Minas que é responsável por 40% da carga ferroviária do país, embora represente apenas 3% da malha nacional. Movimenta cerca de milhões de toneladas de minério de ferro por ano; e milhões de toneladas de carga geral, como carvão mineral, grãos, fertilizantes e produtos siderúrgicos.

### Minério de Ferro

Os principais produtos de exportação são: Celulose, Café, Rochas Ornamentais, petróleo, mas o principal item de exportação é o minério de ferro, tendo a Vale como principal empresa responsável por cerca de 70% das exportações de minério do Brasil ([Minério de Ferro: principais produtores no Brasil, aspectos econômicos e preços da commodity - Saga Consultoria](https://sagaconsultoria.com/minerio-de-ferro-principais-produtores-no-brasil-aspectos-economicos-e-precos-da-commodity/)). Entre os principais compradores destacam-se a China, com demanda quase metade de toda extração, seguido do Japão, Coréia do Sul, Holanda, Itália, entre outros.

Segundo a Global Iron Ore Mining, empresa que fornece estatísticas e informações globais sobre essa commodity, o Brasil possui a terceira maior reserva do mundo, atrás apenas da Austrália e da Rússia.

A indústria de mineração desempenha um papel importante na economia brasileira, respondendo por 1,61% do PIB brasileiro. No entanto, além de sua contribuição para a economia nacional, a exploração de minério de ferro também é fundamental para o sucesso do setor externo, tornando a balança comercial superavitária.

O minério de ferro é vital para a produção de aço. O aço é uma liga dura e durável elaborada principalmente de ferro, mas combinada com pequenas quantidades de outros metais, forma a base da sociedade industrial.

O aço é usado em infraestrutura de transporte, como ferrovias e automóveis, bem como em máquinas e eletrodomésticos. Além disso, é usado na construção, como na criação de pontes, arranha-céus e aeroportos. ([Minério de ferro: qual a sua importância para a economia? | Exame](https://exame.com/invest/guia/minerio-de-ferro-qual-a-sua-importancia-para-a-economia/)).

## 1.2. O problema proposto

As exportações

(Why?) Analisar os dados das exportações de minério pelo estado do Espírito Santo podem auxiliar na análise de novos projetos logísticos como Portos e Ferrovias, pois o minério é um dos principais produtos de exportação do Brasil

(Who?) Os dados objeto da análise são do Governo Federal do Brasil, disponibilizados pelo sistema Comex Stat do Ministério da Indústria, Comércio Exterior e Serviços (MDIC)

(What?): A tendência de exportação do minério de ferro pelo estado do Espírito Santo

(Where?): Trataremos especificamente das exportações referentes ao estado do Espírito Santo. Veremos mais abaixo que o estado de Minas Gerais também é um importante exportador de minério e muito se utiliza da logística ferroviária e portuária para o escoamento de sua produção, mas neste estudo nos ateremos apenas aos dados do estado do Espírito Santo.

(When?): Analisaremos dados de 5 anos, de 2018 a 2022, neste intervalo houve um impacto muito grande da pandemia de COVID

# **1.3. Objetivos**

O objetivo principal deste trabalho é, com base nas informações coletadas, prever o comportamento da exportação de minério de ferro pelos portos do Espírito Santo, visando assim auxiliar no planejamento de investimentos em infraestrutura do estado.

2. Coleta de Dados

Os dados das exportações do Brasil estão disponíveis no site do governo ([Estatísticas de Comércio Exterior em Dados Abertos — Ministério da Economia (www.gov.br)](https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas/base-de-dados-bruta)) onde são disponibilizados dados de exportações, e importações, desde 19978 até 2022. Para este projeto trabalharemos com dados de 5 anos, ou seja, de 2018 a 2022.

EXP\_2018, EXP\_2019, EXP\_2020, EXP\_2021 e EXP\_2022

|  |  |  |
| --- | --- | --- |
| Nome da coluna/campo | Descrição | Tipo |
| CO\_ANO | Ano da exportação. Ex.: 2020 | Inteiro |
| CO\_MES | Mês da exportação. Ex.: 02 | Inteiro |
| CO\_NCM | Código NCM da Mercadoria - Código de 8 dígitos adotado como convenção da categorização de mercadorias no Mercosul. Baseado no Sistema Harmonizado (código internacional de mesmo propósito com 6 dígitos) é utilizado para classificar e definir o tratamento tributário e administrativo de cada mercadoria importada ou exportada. Ex.: 0710.10.00 é a NCM de Batatas congeladas não cozidas ou cozidas em água ou vapor | Inteiro |
| CO\_UNID | Código de 2 dígitos da unidade de medida estatística da mercadoria.  Ex: 17 é Litro | Inteiro |
| CO\_PAIS | Código do país. Ex: 105 é o Brasil | Inteiro |
| SG\_UF\_NCM | Sigla do Estado da Operação.  Ex: SP é São Paulo | String |
| CO\_VIA | Código da via de transporte da  operação. Ex: 1 é Marítima. | Inteiro |
| CO\_URF | Código de 7 dígitos da  Unidade da RFB responsável pela operação. Ex: 0817600 é o Aeroporto Internacional de Guarulhos | Inteiro |
| QT\_ESTAT | Quantidade da mercadoria na  unidade de medida estatística  determinada pelo CO\_UNID. Ex.: 2000. | Inteiro |
| KG\_LIQUIDO | Peso em kg líquidos (sem considerar peso da embalagem). Ex: 50 kgs. | Inteiro |
| VL\_FOB | Valor da mercadoria em si, no local de embarque (sem considerar os custos de transporte ou seguro) em dólares. FOB vem da sigla em inglês de Free On Board, que significa “livre a bordo”, sendo toda a responsabilidade pelo transporte da mercadoria é do cliente, incluindo os riscos e os custos. Ex: 3.500,00 USD. | Inteiro |

Em relação aos datasets utilizados para enriquecimento dos dados:

NCM

|  |  |  |
| --- | --- | --- |
| Nome da coluna/campo | Descrição | Tipo |
| CO\_NCM | Código NCM da Mercadoria - Código de 8 dígitos adotado como convenção da categorização de mercadorias no Mercosul. Baseado no Sistema Harmonizado (código internacional de mesmo propósito com 6 dígitos) é utilizado para classificar e definir o tratamento tributário e administrativo de cada mercadoria importada ou exportada. Ex.: 0710.10.00 é a NCM de Batatas congeladas não cozidas ou cozidas em água ou vapor | Inteiro |
| NO\_NCM\_POR | Descrição dessa NCM em  português. Ex: a descrição do código 8517.12.11 é “Portáteis (por exemplo, walkie talkie e handle talkie)” | String |
| CO\_PPE | Código da PPE (Pauta de Produtos Exportados). Ex: 1258 é o Minérios de ferro e seus concentrados | Inteiro |
| NO\_PPE | Descrição dessa Pauta de Produtos Exportados (em maiúsculo). Ex: 1258 é “MINÉRIOS DE FERRO E SEUS CONCENTRADOS” | String |
| NO\_PPE\_MIN | Descrição dessa Pauta de Produtos Exportados (em minúsculo). Ex: 1258 é “Minérios de ferro e seus concentrados” | String |
| NO\_PPE\_ING | Descrição dessa Pauta de Produtos Exportados (em inglês e em maiúsculo). Ex: 1258 é “IRON ORE AND CONCENTRATES ITS” | String |

PAÍS

|  |  |  |
| --- | --- | --- |
| Nome da coluna/campo | Descrição | Tipo |
| CO\_PAIS | Código de 3 dígitos da  país responsável pela operação. Ex: 13 Afeganistão | Inteiro |
| CO\_PAIS\_ISON3 | Código de 3 dígitos desse País segundo a norma ISO. Ex: 276 – Alemanha | Inteiro |
| CO\_PAIS\_ISOA3 | Código de 3 caracteres desse País segundo a norma ISO. Ex: DEU - Alemanha | String |
| NO\_PAIS | Descrição desse PAÍS em português. Ex: Alemanha | String |
| NO\_PAIS\_ING | Descrição desse PAÍS em português. Ex: Germany | String |
| NO\_PAIS\_ESP | Descrição desse PAÍS em português. Ex: Alemania | String |

VIA

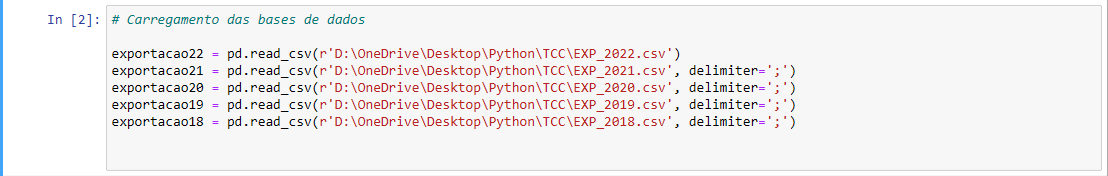
|  |  |  |
| --- | --- | --- |
| Nome da coluna/campo | Descrição | Tipo |
| CO\_VIA | Código de 2 dígitos da  Via utilizada. Ex: 1 Marítima | Inteiro |
| NO\_VIA | Descrição dessa VIA em português. Ex: Marítima | String |

# 3. Processamento/Tratamento de Dados

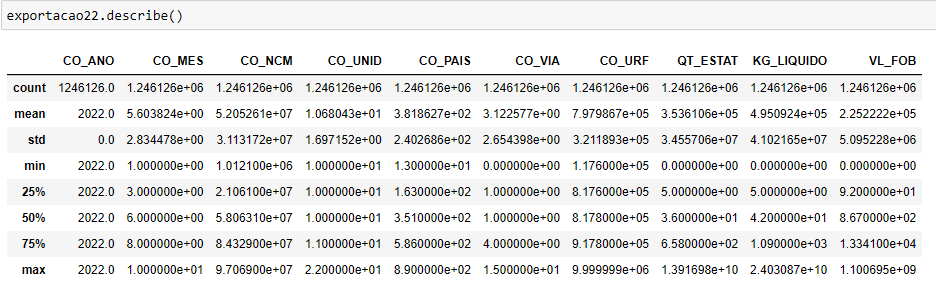
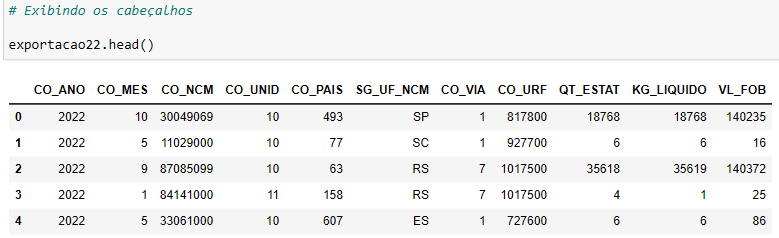
Foi feito download dos arquivos referentes aos anos de 2018 até 2022:

* EXP\_2018.csv;
* EXP\_2019.csv;
* EXP\_2020.csv;
* EXP\_2021.csv; e
* EXP\_2022.csv.

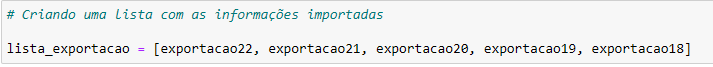
Guardados no diretório de trabalho eles foram carregados para processamento:



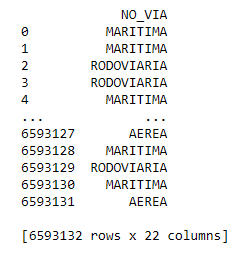
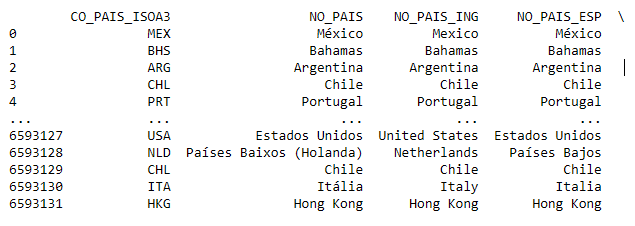
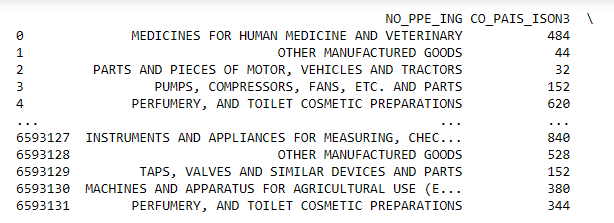
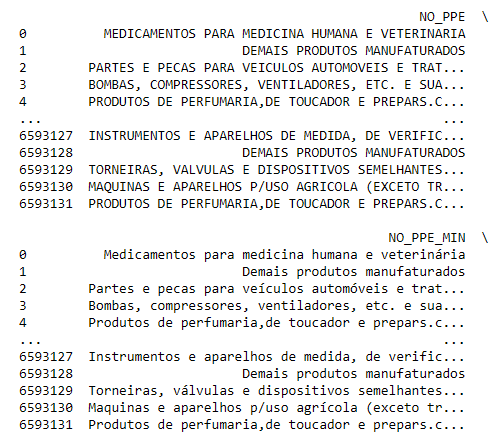
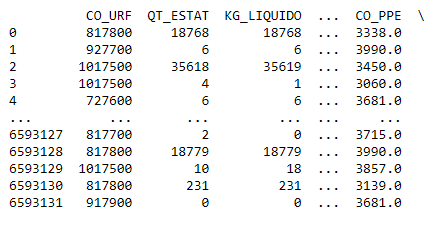
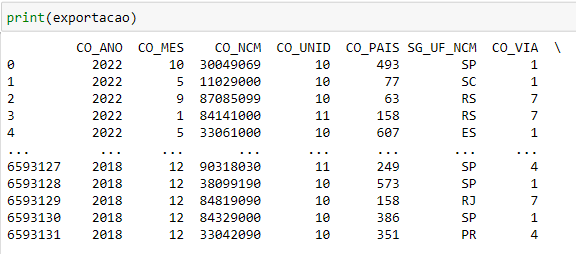
Os dados foram validados passo a passo para verificar se foram devidamente carregados na base:



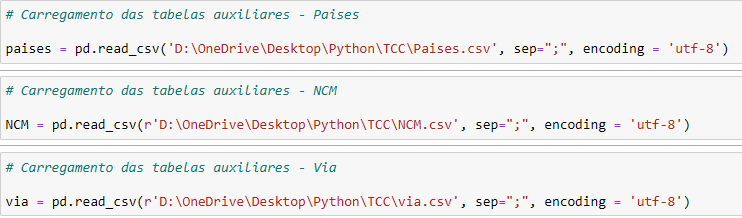
Após o processo de carragamento das bases foi necessário uní-las:



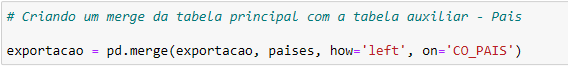
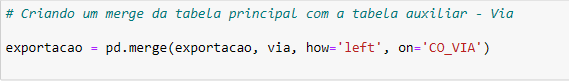
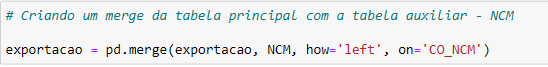
Após as junções verificar as informações unificadas:



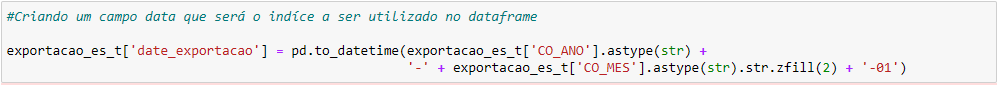
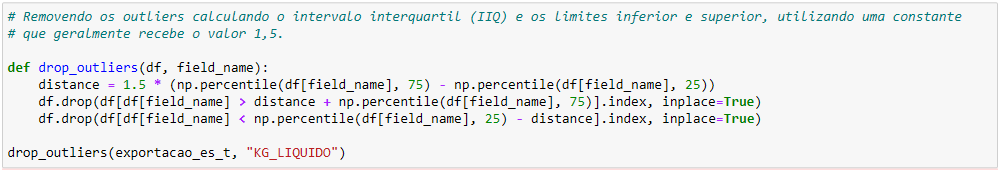
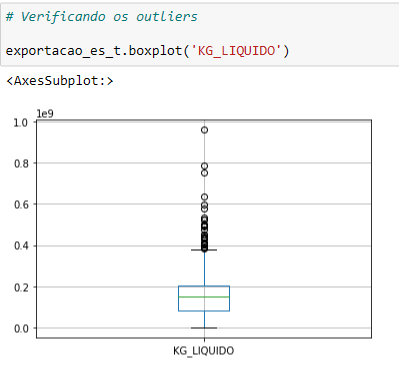
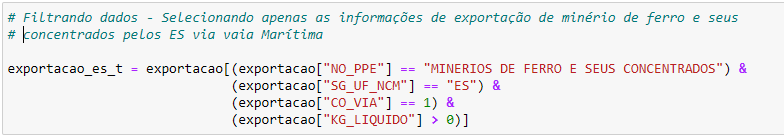
Carregamento de tabelas auxiliares (países, NCM e vias):



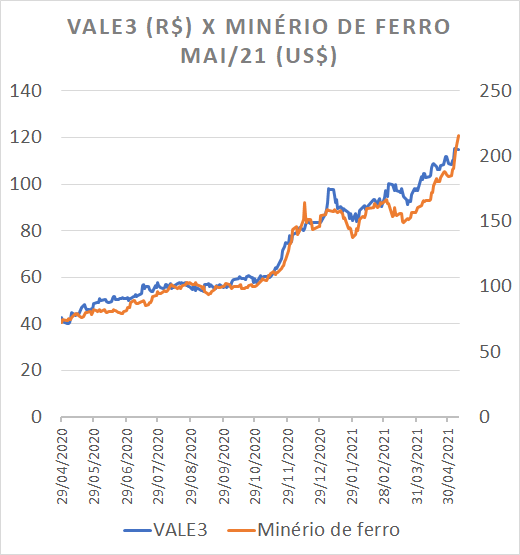
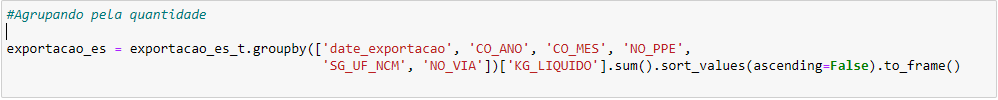
Fazendo merge com as tabelas auxiliares:



Filtrando as informações:



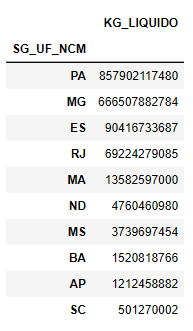
Para agrupar as informações de exportações poderíamos ter como base dois campos, ‘KG\_LIQUIDO’ ou ‘VL\_FOB’, mas se fossemos levar em consideração o valor, como o minério de ferro é uma commoditie teríamos uma variação por conta do mercado (conforme podemos ver na fig XXX), o que poderia nos levar a inferências errôneas. No caso do peso é um dado absoluto que não depende de outras variáveis.



[O Setor Mineral no Mercado de Ações - Saga Consultoria](https://sagaconsultoria.com/o-setor-mineral-no-mercado-de-acoes/)

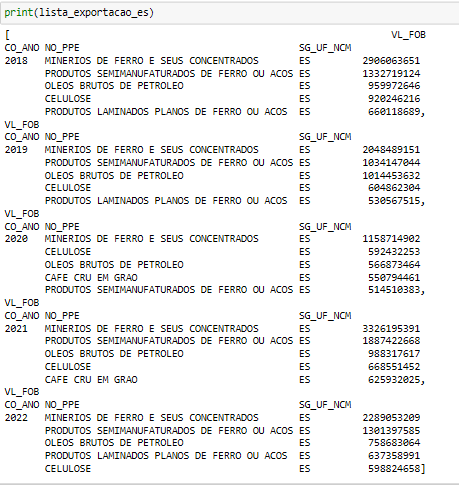
4. Análise e Exploração dos Dados

Minha intenção inicial era de analisar os dados para fundamentar sobre a exportação por meio dos Portos do Espírito Santo, mas ao carregar os arquivos baixados de dados.gov.br, e fazer alguns filtros e agrupamentos, por exemplo, ao filtrar por produto = "MINERIOS DE FERRO E SEUS CONCENTRADOS" e via = “MARÍTIMA” o resultado abaixo é mostrado:



Ora, como bem sabemos Minas Gerais não tem portos, mas aparece em 2º lugar como maior exportador de minério por via marítima, logo meu entendimento das informações ficou mais claro, o estado é somente a origem e não o estado de saída de exportação. Podemos também inferir que grande parte das exportações do estado de Minas Gerais são escoados pelos porto do estado do Espírito Santo pela proximidade e pela boa logística ferroviária disponível. Mas mesmo assim a soma dos dois estado, Mingas Gerais e Espírito Santo, não ultrapassam o montante do estado do Pará.

A seguir podemos verificar quais foram os principais itens de exportação do Espírito Santo desde 2018, onde podemos verificar que o minério de ferro está sempre colocado em primeiro lugar.



Legenda:

Minério de ferro

2º Lugar nas exportações

3º Lugar nas exportações

4º Lugar nas exportações

5º Lugar nas exportações

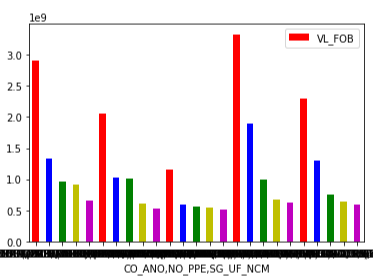
2022

2021

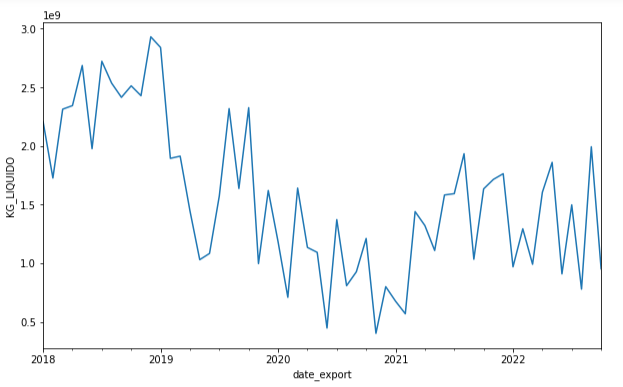
2020

2019

2018



Podemos verificar pelo gráfico abaixo que as exportações caem vertiginosamente durante o período da pandemia, resultado muito provavelmente do lockdown adotado em vários países ao redor do Globo, impactando o consumo Mundial.



Existe uma retomada a partir do início de 2021, mas que não retornaram ao mesmo patamar de antes da pandemia.

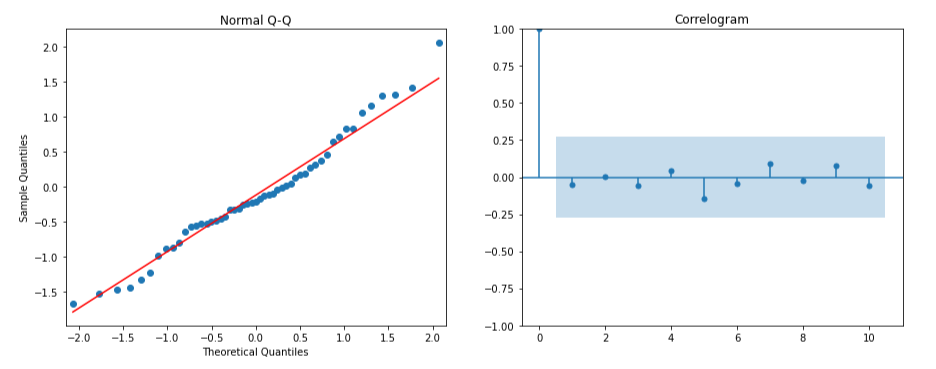
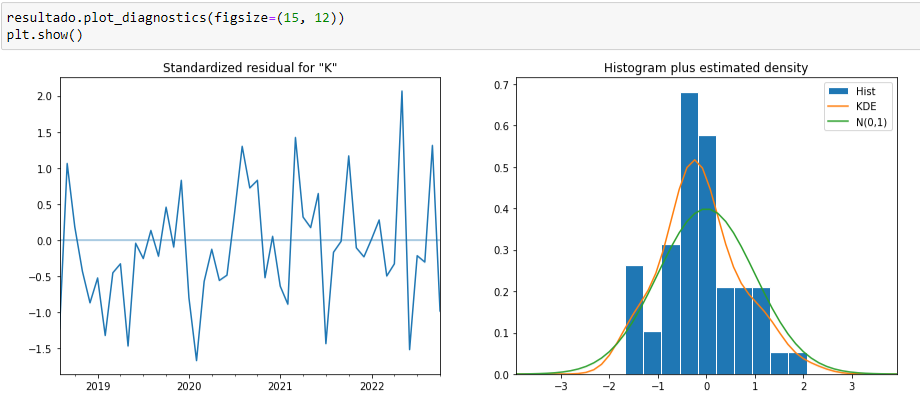
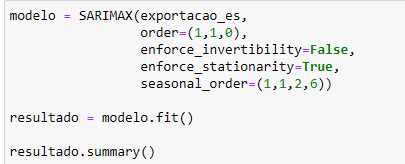
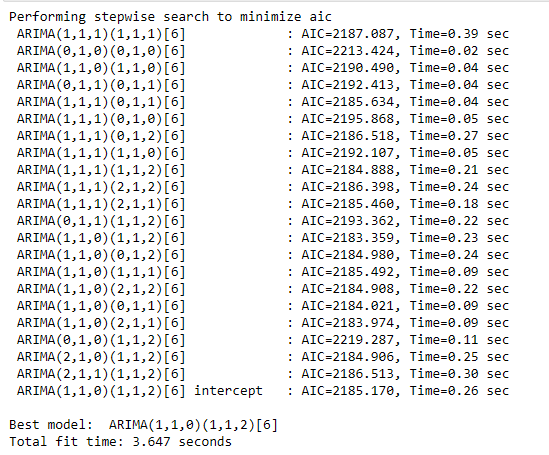
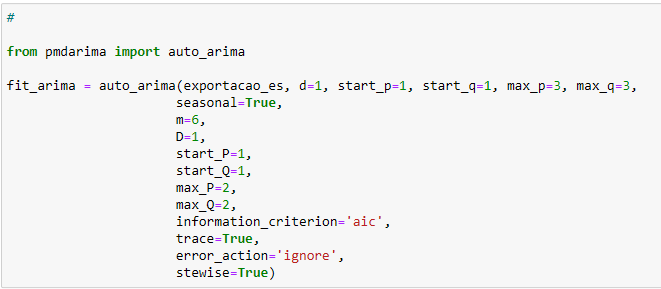
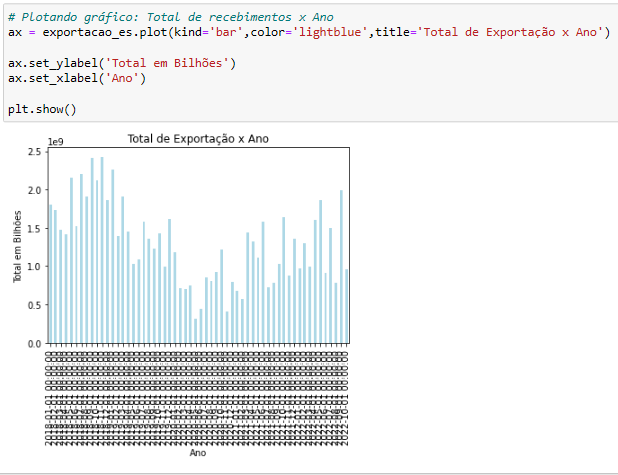
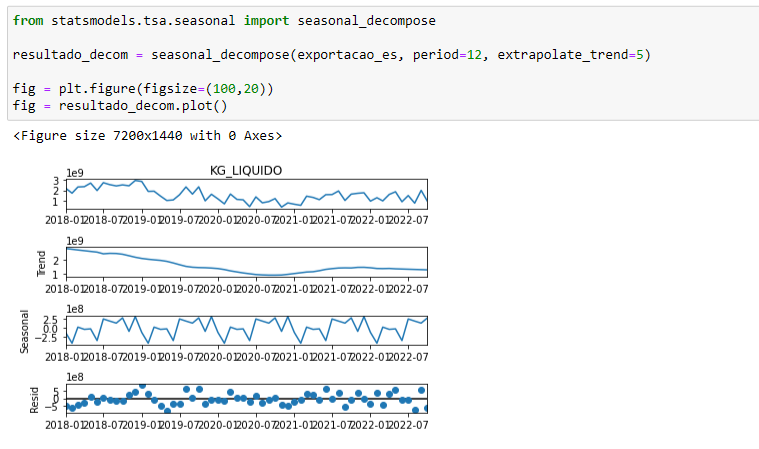
# 5. Criação de Modelos de Machine Learning

As séries temporais representam um conjunto de observações ordenadas no tempo e fundamentadas na ideia de que a história dos acontecimentos, ao longo deste, pode ser usada para prever o futuro. A previsão de uma série temporal é o estabelecimento dos valores futuros da série, sendo uma previsão a estimativa acerca da verossimilhança de eventos futuros, baseados na informação atual e histórica. Pressupõe a modelagem matemática do fenômeno, obtenção de conclusões e avaliação do modelo em termos de precisão (SOUZA; CAMARGO, 2004).

Segundo Tubino (2007), as previsões de demanda baseadas em séries temporais partem do princípio de que a demanda futura será uma projeção dos valores passados, não sofrendo influência de outras variáveis. Métodos estatísticos de previsão de séries temporais buscam identificar um padrão de comportamento da série e utilizá-lo para prever os valores futuros. Estas séries, em sua grande maioria, apresentam características repetitivas que podem ser utilizadas no momento de realizar previsões. Um modelo clássico para séries temporais supõe que a série possa ser escrita como o agrupamento dos três seguintes componentes: tendência, ciclo e sazonalidade; e o processo de construção de valores previstos para a série é realizado por meio da reunificação de cada um desses componentes (SOUZA; SAMOHYL; MIRANDA, 2008). Um modelo é uma descrição probabilística de sua série temporal e cabe ao usuário definir como aplica-lo, levando em conta seus objetivos (MORETTIN; TOLOI, 2004). Um modelo pode levar a um procedimento de previsão; e os procedimentos baseados em séries temporais, deixam “os dados falarem por si”, para a construção deste. Assim, utilizam unicamente as observações da própria série de dados, não dependendo de nenhuma variável externa para efetuar as previsões.

Ressalta-se que a previsão não constitui um fim em si, mas apenas um meio de fornecer informações para uma posterior tomada de decisões, visando a objetivos específicos. Na classe de modelos paramétricos, a análise é feita no domínio do tempo. Dentre esses modelos, os mais frequentemente usados são os modelos de erro (ou de regressão), os modelos auto-regressivos-médias móveis (ARMA) e os modelos auto-regressivos-integrados-médias móveis (ARIMA) e, para incorporar o comportamento de sazonalidade, utilizam-se os modelos ARIMA sazonais multiplicativos (SARIMA) (Morettin & Toloi, 2004).

SARIMA

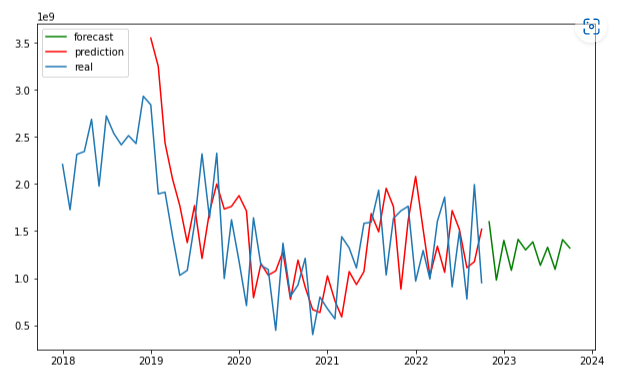
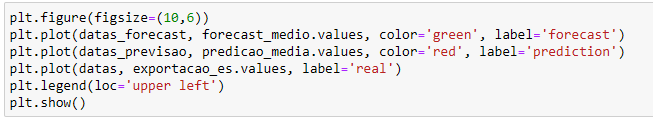
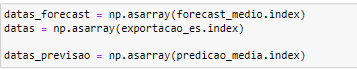
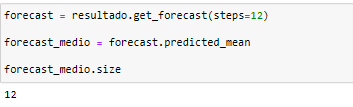
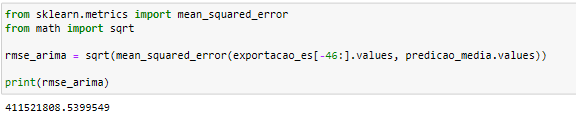
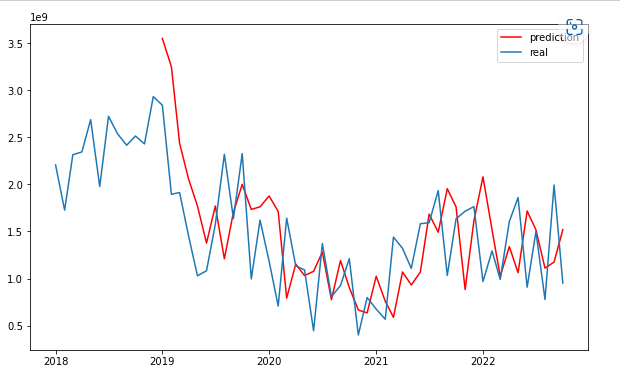
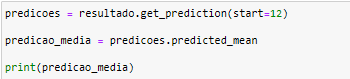
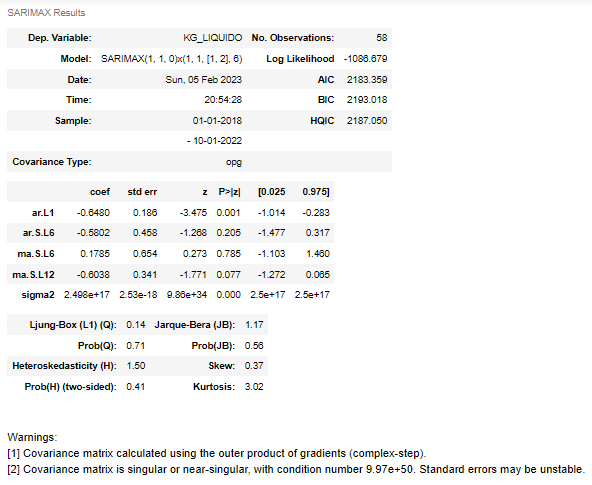


Nesse caso, nosso diagnóstico de modelo sugere que os resíduos do modelo são normalmente distribuídos com base no seguinte:

No gráfico “Histogram Plus estiimated density”, vemos que a linha KDE vermelha segue de perto a linha N(0,1) (onde N(0,1)) é a notação padrão para uma distribuição normal com média 0 e desvio padrão de 1). Esta é uma boa indicação de que os resíduos são normalmente distribuídos.

O qq-plot, gráfico “Normal Q-Q” mostra que a distribuição ordenada de resíduos (pontos azuis) segue a tendência linear das amostras retiradas de uma distribuição normal padrão com N(0, 1). Novamente, esta é uma forte indicação de que os resíduos são normalmente distribuídos.

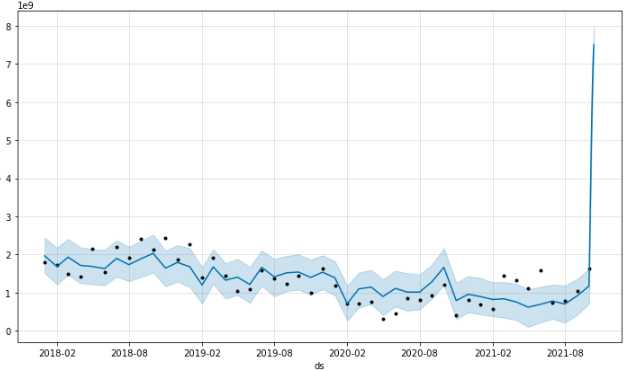
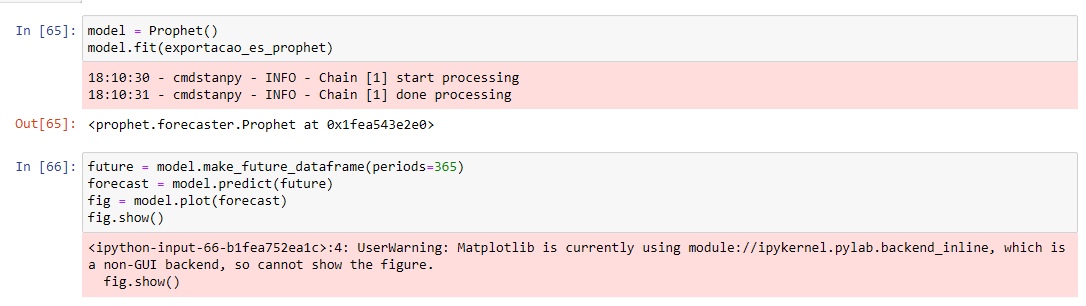
Os resíduos ao longo do tempo (Standardized Residual for “K”) não exibem nenhuma sazonalidade óbvia e parecem ser ruído branco. Isso é confirmado pelo gráfico de autocorrelação (ou seja, correlograma) no canto inferior direito, que mostra que os resíduos da série temporal têm baixa correlação com versões defasadas de si mesmos.



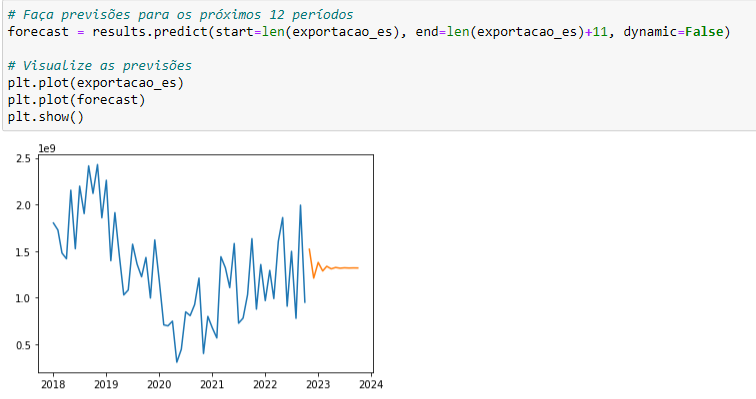
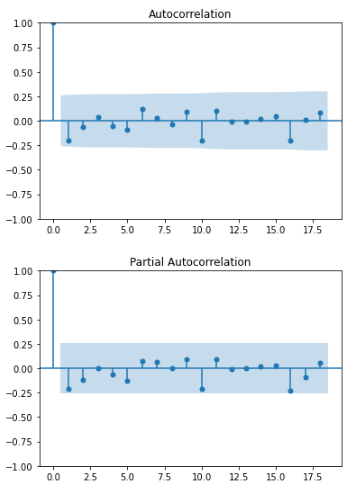
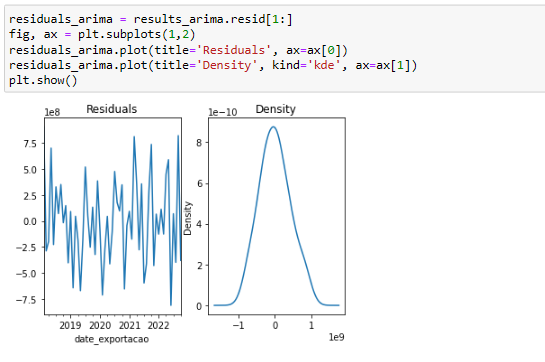
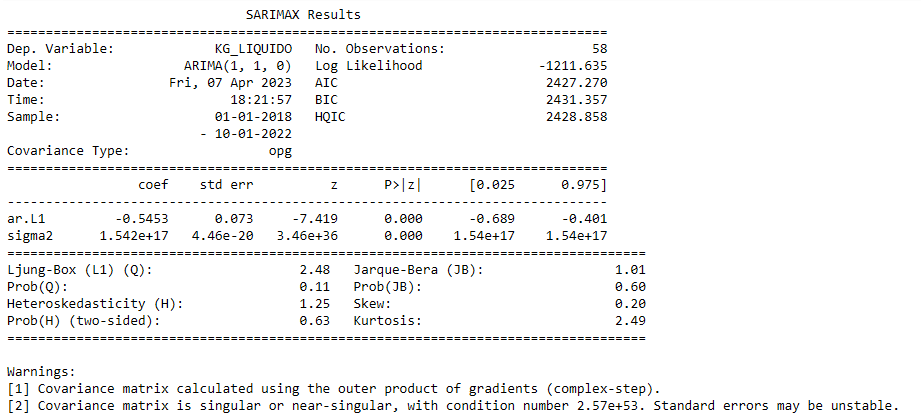
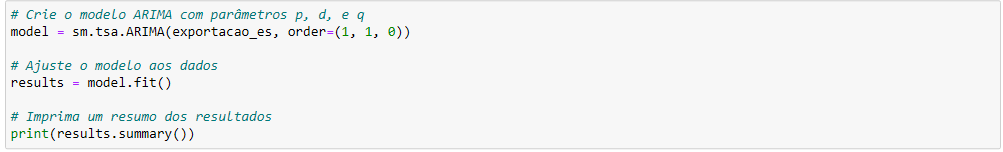
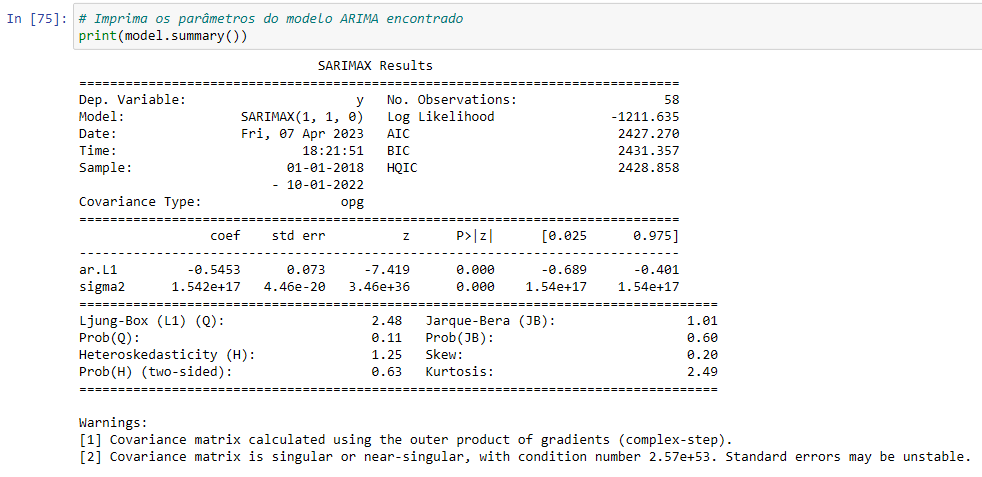
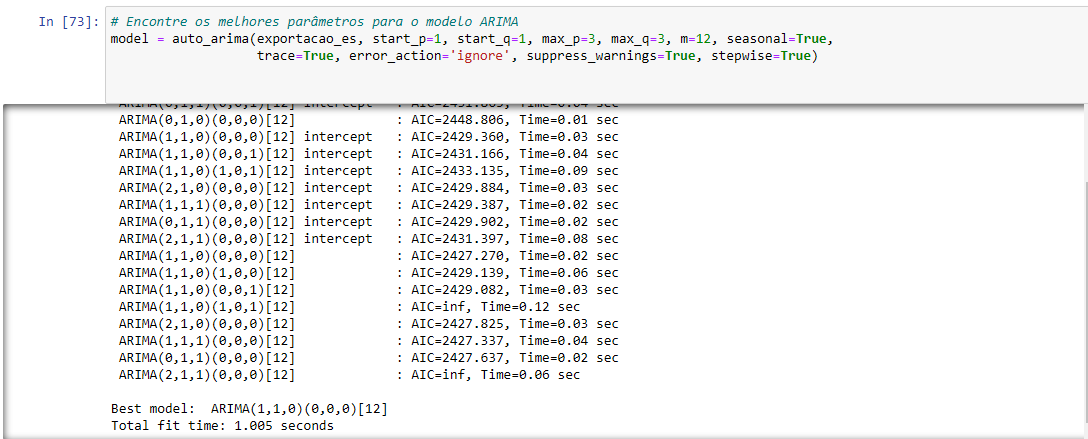
Prophet

O Prophet é um pacote para R e Python usado em produção no Facebook (daí o nome do pacote, fbprophet). Ele implementa o algoritmo de previsão de séries temporais, programado para detectar automaticamente os padrões sazonais de uma série de input. Contudo, é possível customizar alguns inputs de parâmetros manualmente, como indicar a presença de períodos sazonais (semanal ou anual), feriados e changepoints.

No artigo em que os autores descrevem seus métodos (Taylor & Letham, 2017), observa-se uma abordagem prática para a previsão “em escala”, que combina modelos configuráveis com análises de desempenho analíticas em loop. É usado um modelo de regressão modular com parâmetros interpretáveis que podem ser intuitivamente ajustados por analistas com conhecimento de domínio sobre as séries temporais. Segundo a própria equipe de desenvolvimento, o Prophet funciona melhor com séries temporais de frequência diária, com pelo menos um ano de dado, sendo robusto a dados ausentes (NA), mudanças na tendência e outliers. Caso queira usar outra frequência, ela deve ser indicada dentro da função “Prophet()” como “freq=” – por exemplo, W para semanal.

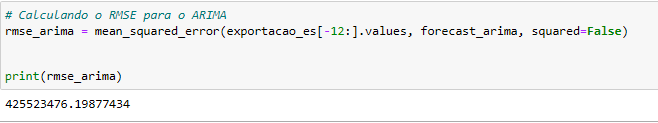
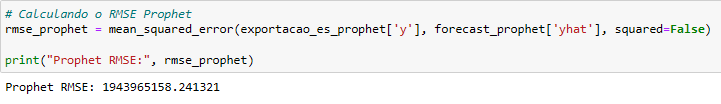
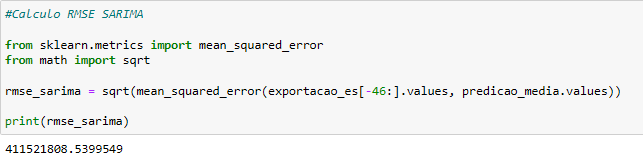


ARIMA



# 6. Interpretação dos Resultados

Calculo de RMSE:



Pelo resultado o modelo mais indicado seria o SARIMA, pois tem o menor RMSE.

# 7. Apresentação dos Resultados

Filtrar apenas os dados do estado do Espírito Santo, que utilizem via marítima e que cujo NCM seja Minério de ferro e seus concentrados

Comprara o RMSE (Root Mean Squared Error) - O RMSE é uma variante do MSE que é calculada como a raiz quadrada do MSE. Isso permite que a métrica seja expressa na mesma unidade que os dados originais, facilitando a interpretação, de cada modelo.

Utilizar modelos que tratem de Séries Temporais, tais como ARIMA, SARIMA e Prophet

Dados públicos de comércio exterior coletados do site [www.gov.br](http://www.gov.br).

Verificar se haverá aumento exponencial nas exportações de minério de ferro pelo estado do Espírito Santo que seja necessário maior investimento na logística

Analisar as exportações de minério de ferro pelo estado do Espírito Santo



# 8. Links

Link para o vídeo: <https://youtu.be/UQdcpgK3IZU>

Link para o repositório: [fkomati/PUC-MG: Arquivos da Pós (github.com)](https://github.com/fkomati/PUC-MG)

# REFERÊNCIAS

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2 ed. São Paulo: Edgard

Blücher, 2006.

TAYLOR, S. J.; LETHAM, Benjamin. **Forecasting at Scale.** Computer Science - The American Statistician, 2018

SOUZA, G. P.; SAMOHYL, R. W.; MIRANDA, R. G. **Métodos simplificados de previsão empresarial**. Rio de Janeiro: Ciência Moderna, 2008. 181 p.

SOUZA, R. C.; CAMARGO, M. E. **Análise e previsão de séries temporais: os modelos ARIMA**. 2. ed. Rio de Janeiro: Regional, 2004.

TUBINO, D. F. **Planejamento e controle da produção**. São Paulo: Atlas, 2007. 190 p.

# APÊNDICE

# Carregamento das bibliotecas para tratamento das informações

from datetime import datetime

from statsmodels.tsa.statespace.sarimax import SARIMAX

from statsmodels.tsa.seasonal import seasonal\_decompose

from statsmodels.tsa.stattools import adfuller

from prophet import Prophet

import statsmodels.api as sm

import pmdarima as pmd

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

import csv

import datetime

# Carregamento das bases de dados

exportacao22 = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\EXP\_2022.csv')

exportacao21 = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\EXP\_2021.csv', delimiter=';')

exportacao20 = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\EXP\_2020.csv', delimiter=';')

exportacao19 = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\EXP\_2019.csv', delimiter=';')

exportacao18 = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\EXP\_2018.csv', delimiter=';')

# Exibindo os cabeçalhos

print(exportacao22.dtypes)

exportacao22.describe()

exportacao18.head()

# Criando uma lista com as informações importadas

lista\_exportacao = [exportacao22, exportacao21, exportacao20, exportacao19, exportacao18]

# Concatenando as informações em um dataframe

exportacao = pd.concat(lista\_exportacao)

print(exportacao)

# Carregamento das tabelas auxiliares - Paises

paises = pd.read\_csv('D:\OneDrive\Desktop\Python\TCC\Paises.csv', sep=";", encoding = 'utf-8')

# Carregamento das tabelas auxiliares - NCM

NCM = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\NCM.csv', sep=";", encoding = 'utf-8')

# Carregamento das tabelas auxiliares - Via

via = pd.read\_csv(r'D:\OneDrive\Desktop\Python\TCC\via.csv', sep=";", encoding = 'utf-8')

print(paises)

print(NCM)

print(via)

# Criando um merge da tabela principal com a tabela auxiliar - NCM

exportacao = pd.merge(exportacao, NCM, how='left', on='CO\_NCM')

print(exportacao)

# Criando um merge da tabela principal com a tabela auxiliar - Pais

exportacao = pd.merge(exportacao, paises, how='left', on='CO\_PAIS')

print(exportacao)

# Criando um merge da tabela principal com a tabela auxiliar - Via

exportacao = pd.merge(exportacao, via, how='left', on='CO\_VIA')

print(exportacao)

#Principais estados exportadores

exportacao['date\_exportacao'] = pd.to\_datetime(exportacao['CO\_ANO'].astype(str) +

'-' + exportacao['CO\_MES'].astype(str).str.zfill(2) + '-01')

print(exportacao)

exportacao\_prod = exportacao[(exportacao["SG\_UF\_NCM"] == "ES") &

(exportacao["CO\_VIA"] == 1) &

(exportacao["KG\_LIQUIDO"] > 0)]

print(exportacao\_prod)

exportacao\_prod = exportacao\_prod.groupby(['CO\_ANO', 'NO\_PPE', 'SG\_UF\_NCM'])['VL\_FOB'].sum().sort\_values(ascending=False).to\_frame()

print(exportacao\_prod)

lista\_exportacao\_es = [exportacao\_prod.query('SG\_UF\_NCM == "ES" & CO\_ANO == 2018').head(5),

exportacao\_prod.query('SG\_UF\_NCM == "ES" & CO\_ANO == 2019').head(5),

exportacao\_prod.query('SG\_UF\_NCM == "ES" & CO\_ANO == 2020').head(5),

exportacao\_prod.query('SG\_UF\_NCM == "ES" & CO\_ANO == 2021').head(5),

exportacao\_prod.query('SG\_UF\_NCM == "ES" & CO\_ANO == 2022').head(5)]

print(lista\_exportacao\_es)

exportacao\_grafico = pd.concat(lista\_exportacao\_es)

print(exportacao\_grafico)

axis = exportacao\_grafico.plot.bar(y='VL\_FOB', rot=0, color=['r', 'b', 'g', 'y', 'm'])

print(axis)

plt.show()

# Filtrando dados - Selecionando apenas as informações de exportação de minério de ferro e seus

# concentrados pelos ES via Marítima

exportacao\_es\_t = exportacao[(exportacao["NO\_PPE"] == "MINERIOS DE FERRO E SEUS CONCENTRADOS") &

(exportacao["SG\_UF\_NCM"] == "ES") &

(exportacao["CO\_VIA"] == 1) &

(exportacao["KG\_LIQUIDO"] > 0)]

print(exportacao\_es\_t)

exportacao\_es\_dest = exportacao\_es\_t.groupby(['NO\_PAIS'])['KG\_LIQUIDO'].sum().sort\_values(ascending=False).to\_frame()

exportacao\_es\_dest.head(10)

df\_10 = exportacao\_es\_dest.reset\_index().head(10)

fig, ax = plt.subplots()

ax.bar(df\_10["NO\_PAIS"], df\_10["KG\_LIQUIDO"])

ax.set\_xlabel("País")

ax.set\_ylabel("Peso em milhões de Toneladas")

ax.set\_title("Princpais destinos das exportações de minério de ferro")

plt.show()

# Verificando os outliers

exportacao\_es\_t.boxplot('KG\_LIQUIDO')

print(exportacao\_es\_t)

# Removendo os outliers calculando o intervalo interquartil (IIQ) e os limites inferior e superior, utilizando uma constante

# que geralmente recebe o valor 1,5.

def drop\_outliers(df, field\_name):

distance = 1.5 \* (np.percentile(df[field\_name], 75) - np.percentile(df[field\_name], 25))

df.drop(df[df[field\_name] > distance + np.percentile(df[field\_name], 75)].index, inplace=True)

df.drop(df[df[field\_name] < np.percentile(df[field\_name], 25) - distance].index, inplace=True)

drop\_outliers(exportacao\_es\_t, "KG\_LIQUIDO")

exportacao\_es\_t.boxplot('KG\_LIQUIDO')

print(exportacao\_es\_t)

print(exportacao\_es\_t.dtypes)

#Criando um campo data que será o indíce a ser utilizado no dataframe

exportacao\_es\_t['date\_exportacao'] = pd.to\_datetime(exportacao\_es\_t['CO\_ANO'].astype(str) +

'-' + exportacao\_es\_t['CO\_MES'].astype(str).str.zfill(2) + '-01')

#Agrupando pela quantidade

exportacao\_es = exportacao\_es\_t.groupby(['date\_exportacao', 'CO\_ANO', 'CO\_MES', 'NO\_PPE',

'SG\_UF\_NCM', 'NO\_VIA'])['KG\_LIQUIDO'].sum().sort\_values(ascending=False).to\_frame()

exportacao\_es.describe()

print(exportacao\_es)

print(exportacao\_es.dtypes)

exportacao\_es = exportacao\_es.sort\_values(by="date\_exportacao")

exportacao\_es = exportacao\_es.groupby(['date\_exportacao']).KG\_LIQUIDO.sum()

exportacao\_es.describe()

exportacao\_es

exportacao\_es\_prophet = exportacao\_es.reset\_index(name='y')

exportacao\_es\_prophet.rename(columns={'date\_exportacao': 'ds'}, inplace=True)

print(exportacao\_es\_prophet)

exportacao\_es.plot(figsize=(10,6))

plt.xlabel('date\_export')

plt.ylabel('KG\_LIQUIDO')

from statsmodels.tsa.seasonal import seasonal\_decompose

resultado\_decom = seasonal\_decompose(exportacao\_es, period=12, extrapolate\_trend=5)

fig = plt.figure(figsize=(100,20))

fig = resultado\_decom.plot()

# Plotando gráfico: Total de recebimentos x Ano

ax = exportacao\_es.plot(kind='bar',color='lightblue',title='Total de Exportação x Ano')

ax.set\_ylabel('Total em Bilhões')

ax.set\_xlabel('Ano')

plt.show()

#

from pmdarima import auto\_arima

fit\_arima = auto\_arima(exportacao\_es, d=1, start\_p=1, start\_q=1, max\_p=3, max\_q=3,

seasonal=True,

m=6,

D=1,

start\_P=1,

start\_Q=1,

max\_P=2,

max\_Q=2,

information\_criterion='aic',

trace=True,

error\_action='ignore',

stewise=True)

modelo = SARIMAX(exportacao\_es,

order=(1,1,0),

enforce\_invertibility=False,

enforce\_stationarity=True,

seasonal\_order=(1,1,2,6))

resultado = modelo.fit()

resultado.summary()

resultado.plot\_diagnostics(figsize=(15, 12))

plt.show()

predicoes = resultado.get\_prediction(start=12)

predicao\_media = predicoes.predicted\_mean

print(predicao\_media)

datas\_previsao = np.asarray(predicao\_media.index)

datas = np.asarray(exportacao\_es.index)

plt.figure(figsize=(10,6))

plt.plot(datas\_previsao, predicao\_media.values, color='red', label='prediction')

plt.plot(datas, exportacao\_es.values, label='real')

plt.legend()

plt.show()

#Calculo RMSE SARIMA

from sklearn.metrics import mean\_squared\_error

from math import sqrt

rmse\_sarima = sqrt(mean\_squared\_error(exportacao\_es[-46:].values, predicao\_media.values))

print(rmse\_sarima)

forecast = resultado.get\_forecast(steps=12)

forecast\_medio = forecast.predicted\_mean

forecast\_medio.size

forecast\_medio[0]

datas\_forecast = np.asarray(forecast\_medio.index)

datas = np.asarray(exportacao\_es.index)

datas\_previsao = np.asarray(predicao\_media.index)

plt.figure(figsize=(10,6))

plt.plot(datas\_forecast, forecast\_medio.values, color='green', label='forecast')

plt.plot(datas\_previsao, predicao\_media.values, color='red', label='prediction')

plt.plot(datas, exportacao\_es.values, label='real')

plt.legend(loc='upper left')

plt.show()

# Utilizando o Prophet

from sklearn.model\_selection import train\_test\_split

exportacao\_es\_prophet = exportacao\_es.reset\_index(name='y')

exportacao\_es\_prophet.rename(columns={'date\_exportacao': 'ds'}, inplace=True)

train\_data\_prophet, test\_data\_prophet = train\_test\_split(exportacao\_es\_prophet, test\_size=0.2, shuffle=False)

model\_prophet = Prophet()

model\_prophet.fit(train\_data\_prophet)

future\_prophet = model\_prophet.make\_future\_dataframe(periods=len(test\_data\_prophet))

forecast\_prophet = model\_prophet.predict(future\_prophet)

fig = model\_prophet.plot(forecast\_prophet)

fig.show()

# Calculando o RMSE Prophet

rmse\_prophet = mean\_squared\_error(exportacao\_es\_prophet['y'], forecast\_prophet['yhat'], squared=False)

print("Prophet RMSE:", rmse\_prophet)

# Encontre os melhores parâmetros para o modelo ARIMA

model\_arima = auto\_arima(exportacao\_es, start\_p=1, start\_q=1, max\_p=3, max\_q=3, m=12, seasonal=True,

trace=True, error\_action='ignore', suppress\_warnings=True, stepwise=True)

# Imprima os parâmetros do modelo ARIMA encontrado

print(model\_arima.summary())

# Crie o modelo ARIMA com parâmetros p, d, e q

model\_arima = sm.tsa.ARIMA(exportacao\_es, order=(1, 1, 0))

# Ajuste o modelo aos dados

results\_arima = model\_arima.fit()

# Imprima um resumo dos resultados

print(results\_arima.summary())

residuals\_arima = results\_arima.resid[1:]

fig, ax = plt.subplots(1,2)

residuals\_arima.plot(title='Residuals', ax=ax[0])

residuals\_arima.plot(title='Density', kind='kde', ax=ax[1])

plt.show()

from statsmodels.graphics.tsaplots import plot\_acf, plot\_pacf

acf\_res = plot\_acf(residuals\_arima)

pacf\_res = plot\_pacf(residuals\_arima)

# Faça previsões para os próximos 12 períodos

forecast\_arima = results\_arima.predict(start=len(exportacao\_es), end=len(exportacao\_es)+11, dynamic=False)

# Visualize as previsões

plt.plot(exportacao\_es)

plt.plot(forecast)

plt.show()

# Calculando o RMSE para o ARIMA

rmse\_arima = mean\_squared\_error(exportacao\_es[-12:].values, forecast\_arima, squared=False)

print(rmse\_arima)