UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JOSÉ DANIEL DOS SANTOS

SÉRIES TEMPORAIS DE RECEITAS ORÇAMENTÁRIAS DA PREFEITURA DE CAMPINA GRANDE E PREDIÇÃO DE VALORES.

CAMPINA GRANDE
2021

JOSÉ DANIEL DOS SANTOS

SÉRIES TEMPORAIS DE RECEITAS ORÇAMENTÁRIAS DA PREFEITURA DE CAMPINA GRANDE E PREDIÇÃO DE VALORES.

Trabalho de Conclusão de Curso em Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.

Área de concentração: Ciência de dados.

Orientador: Prof. Dr. Eduardo Jorge Valadares Oliveira

CAMPINA GRANDE
2021

S237s Santos, José Daniel dos.
Séries temporais de receitas orçamentárias da Prefeitura de Campina Grande e predição de valores [manuscrito] / Jose Daniel dos Santos. - 2021.
29 p.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia , 2021.
"Orientação : Prof. Dr. Eduardo Jorge Valadares Oliveira , Coordenação do Curso de Computação - CCT."

1. Séries temporais. 2. Receitas públicas. 3. Modelo ARIMA. 4. Python. I. Título

21. ed. CDD 519.232

Elaborada por Lucas H. A. da Silva - CRB - 15/898 BC/UEPB

JOSÉ DANIEL DOS SANTOS

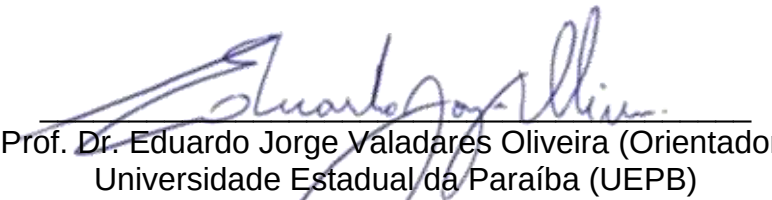
SÉRIES TEMPORAIS DE RECEITAS ORÇAMENTÁRIAS DA PREFEITURA DE CAMPINA GRANDE E PREDIÇÃO DE VALORES.

Trabalho de Conclusão de Curso em Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.

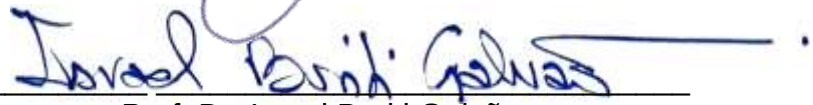
Área de concentração: Ciência de dados.

Aprovado em: 30 / 11 / 2021.

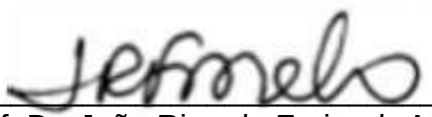
BANCA EXAMINADORA



Prof. Dr. Eduardo Jorge Valadares Oliveira (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Israel Buriti Galvão
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. João Ricardo Freire de Melo
Instituto Federal da Paraíba (IFPB)

AGRADECIMENTOS

Inicialmente, gostaria de agradecer a minha família e amigos. Especialmente, meus pais e minha irmã que sempre me apoiaram com tudo que eu precisava durante a minha vida. Também gostaria de agradecer a minha vó por todo cuidado e carinho.

A todos os amigos que direta ou indiretamente participaram da minha formação eu agradeço com um forte abraço.

Por fim, gostaria de expressar a minha gratidão a todos os meus professores que contribuíram com a minha formação acadêmica e profissional. Em especial ao Prof. Dr. Eduardo Jorge Valadares Oliveira pela oportunidade e apoio durante todo o processo de construção desse TCC.

RESUMO

O acompanhamento da evolução de receitas na esfera pública é de interesse de toda a população, visto que os recursos são de caráter público e, com isso, o acesso deve ser facilitado. Portanto, este trabalho tem como objetivo geral, fazer uma análise sobre os valores recebidos pela Prefeitura do Município de Campina Grande e a partir deles, fazer uma prospecção anual do quantitativo necessário para as próximas receitas. A análise será feita utilizando uma visão estatística a partir de séries temporais e posteriormente será apresentada uma proposta de modelo para predição de valores orçamentários. A análise irá expor como a receita da Prefeitura Municipal de Campina Grande é composta, além do montante recebido pelo referido órgão, despertando assim, o interesse particular pela gestão do orçamento público do município em questão. Será feita uma análise, utilizando uma série temporal das receitas orçamentárias em datas anteriores a este trabalho, baseadas nos dados obtidos pelo Tribunal de Contas do Estado da Paraíba, a fim de que a partir delas, seja realizada uma análise preditiva das receitas posteriores. A proposta é analisar os dados respondendo a algumas perguntas, como: Qual foi o ano em que o município recebeu mais receitas? Qual mês se destacou mais no recebimento das receitas? Houve meses em que o recebimento das receitas foi considerado anormal? Com as respostas a essas perguntas é possível se aprofundar no dataset e partir pra próxima etapa será feita uma análise da série temporal das receitas do orçamento da prefeitura e a predição utilizando as estratégias aplicando o modelo auto-regressivo integrado de médias móveis (Autoregressive Integrated Moving Average ou ARIMA). Com isso, será feita uma divisão de aproximadamente 80% dos valores, contidos no dataset, como base de treino e os demais 20% como base de testes, assim o gráfico com a predição terá os dados da base de teste para comparativo com os resultados obtidos.

Palavras-Chave: Receitas Públicas. Séries Temporais. ARIMA.

ABSTRACT

Monitoring the evolution of revenues in the public sphere is of interest to the entire population, since the resources are of a public nature and, therefore, access should be facilitated. Therefore, this work has the general objective of analyzing the amounts received by the Municipality of Campina Grande and, based on them, making an annual survey of the amount needed for the next revenue. An analysis will be performed using a statistical analysis from time series and later a proposal for a model for predicting budget values will be presented. The analysis will show how the revenue of the Municipality of Campina Grande is composed, in addition to the amount paid by the aforementioned agency, thus arousing the particular interest in managing the public budget of the municipality in question. An analysis will be carried out, using a time series of budget revenues in data prior to this work, based on data obtained by the Court of Auditors of the State of Paraíba, so that from them, a predictive analysis of subsequent revenues is carried out. The proposal is to analyze the data by answering some questions, such as: What was the year in which the municipality spent more revenue? Which month stood out the most in receiving revenue? Were there months when receipt of receipts was considered abnormal? With the answers to these questions, it is possible to go deeper into the dataset and from the next step onwards, an analysis of the time series of the city's budget revenue will be made and the prediction using the strategies applying the autoregressive integrated moving average model (Autoregressive Integrated Moving Average or ARIMA). With this, a division of approximately 80% of the values, contained in the dataset, will be made as a training base and the remaining 20% as a test base, so the graph with the prediction will have the test base data for comparison with the results Granted.

Keywords: Public Revenues. Time Series. ARIMA.

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

O acompanhamento da evolução de receitas na esfera pública é de interesse de toda a população, visto que os recursos são de caráter público e, com isso, o acesso deve ser facilitado Neste sentido, este trabalho tem como objetivo geral, fazer uma análise sobre os valores recebidos, entre 2003 e 2021, pela Prefeitura do Município de Campina Grande e a partir deles, fazer uma prospecção anual do quantitativo necessário para as próximas receitas. A análise será feita utilizando uma visão estatística a partir de séries temporais e posteriormente será apresentada uma proposta de modelo para predição de valores orçamentários.

1.2 O problema proposto

A motivação desta análise é expor como o orçamento de uma prefeitura municipal pode ser composta e quais os valores são recebidos pela entidade. Neste caso a entidade escolhida foi a Prefeitura Municipal de Campina Grande, cidade do Campus sede da Universidade Estadual da Paraíba onde foi idealizado esse trabalho, algo que traz bastante interesse é a gestão do orçamento público do município. Além do estudo dos dados, será feita uma análise preditiva utilizando uma série temporal das receitas orçamentárias baseadas nos dados obtidos pelo Tribunal de Contas do Estado da Paraíba.

A proposta é analisar os dados respondendo a algumas perguntas, tais como: Qual foi o ano em que o município recebeu mais receitas? Qual mês se destacou mais recebimento das receitas? Houve meses em que o recebimento das receitas foi considerado anormal? Quais seriam os prováveis motivos? Respondendo a perguntas assim, podemos nos aprofundar mais no *dataset* escolhido.

Na próxima etapa será feita uma análise da série temporal das receitas do orçamento da prefeitura e a predição utilizando estratégias estudadas ao longo do curso aplicando o modelo ARIMA (Autoregressive Integrated Moving Average).

Será utilizado o modelo ARIMA pois esses modelos geralmente são aplicados em casos em que o *dataset* mostra pontos de não estacionariedade, em que um passo inicial de diferenciação pode ser aplicado uma ou mais vezes para torna-lo

estacionário, é dito estacionário quando os dados possuem a propriedade de que a variância, a média e a autocorrelação não mudam no decorrer do tempo.

2 FUNDAMENTAÇÃO TEÓRICA

O foco do trabalho não é analisar a fundo a composição da receita pública municipal, porém é importante ter esses conceitos em mente para que a análise estatística fique mais clara. São conceitos que estarão dentro do trabalho e que precisam ser conceituados a fim de uma melhor compreensão teórica sobre o que será tratado ao decorrer dos pontos abordados dentro do trabalho acadêmico.

2.1 Orçamento Público

O Orçamento Público é, simplificadamente, um meio de transformar as arrecadações públicas em uma determinada estrutura de gastos, terminando por espelhar decisões políticas, pois estabelece ações prioritárias para atendimento das demandas sociais, em face da escassez de recursos (ALVES, 2013).

A importância do orçamento público pode ser avaliada por diversas dimensões: histórica, democrática, de gestão, entre outras. Porém, de um ponto de vista prático, o orçamento é importante porque, sem ele, a Administração Pública fica quase completamente impedida de agir. Num país com tantas regras e exceções, uma coisa é certa: não existe despesa orçamentária sem prévia autorização legislativa (exceto, aquelas realizadas por meio de créditos extraordinários, destinados a despesas urgentes e imprevistas, em caso de guerra, comoção intestina ou calamidade pública. Tais créditos são abertos por decreto do Poder Executivo, que deles dará imediato conhecimento ao Poder Legislativo.

2.1.1 Receita Pública Municipal

Segundo a Confederação Nacional de Municípios (2008) na Receita pública Municipal incluem-se recursos financeiros oriundos dos tributos municipais e preços pela utilização de bens ou serviços, e demais ingressos que o município recebe em caráter permanente, como a sua participação nas transferências constitucionais estaduais e federais (ICMS Art. 159, inc. I, "b" (BRASIL, 1988), FPM), ou eventuais, como os advindos de financiamentos, empréstimos, subvenções, auxílios e doações de outras entidades ou pessoas físicas.

2.2 Análise Estatística

Durante o trabalho será realizada uma análise estatística da receita municipal de Campina Grande, portanto um ponto muito importante para se ter em mente é a definição de análise estatística, que segundo a SAS (2021) é a ciência de coletar, explorar e apresentar grandes quantidades de dados utilizados para descobrir padrões e tendências. Esta por sua vez é aplicada todos os dias em pesquisas, indústrias e governos, tornando assim o processo de tomada de decisão um processo mais científico.

2.2.1 Computação Estatística

Os Métodos tradicionais de análise estatística que vão desde a definição do problema à análise e interpretação dos resultados, têm sido usados por cientistas há vários séculos, visto que a informação advinda dos dados tende a ser mais precisa, fazendo com que o volume de dados de hoje torne a estatística ainda mais valiosa e poderosa, prova disso é que algoritmos avançados, computadores poderosos e o armazenamento de baixo custo estão elevando o aumento do uso de estatística computacional. Sendo que para trabalhar com grandes volumes de dados ou executar permutações múltiplas em seus cálculos, a estatística computacional é essencial para os estatísticos.

Neste contexto, uma das linguagens com bastante destaque atualmente por fornecer muitas bibliotecas e ferramentas destinada a análise de dados é a linguagem python, que será utilizada durante nesta análise

2.2.2 Python

Python é conhecida por ser uma linguagem mais limpa do que outras e mais intuitiva, devido ao uso de palavras em inglês ao invés de pontuações frequentemente usadas em outras linguagens. A origem de seu nome surgiu como uma homenagem ao grupo de comédia britânico Monty Python, que inspirou os desenvolvedores a tornar a linguagem o mais divertida possível de se utilizar.

A linguagem Python é uma linguagem muito popular na comunidade científica. Python é uma linguagem de programação de alto nível, de script,

orientada a objetos e de tipagem dinâmica e forte. É uma linguagem *open source*, gratuita e tem uma ativa comunidade de programadores (MENEZES, 2014)

Python possui uma grande quantidade de bibliotecas focadas com foco em estatística e análise de dados, durante a análise serão utilizadas algumas que são elas:

NumPy: Operações com arrays n-dimensionais, suporta o processamento de grandes quantidades de dados fornecendo funções matemáticas de alto nível para operar esses dados;

Pandas: Armazenamento, operações e análises usando estruturas de dados como DataFrames e Series, oferece operações e estruturas para manipular séries temporais

Matplotlib: Utilizada para a Visualização de dados, fornecendo todos os gráficos que serão usados durante a análise.

3 COLETA DE DADOS

A fonte de dados utilizada foi adquirida através do site do Tribunal de Contas do Estado da Paraíba (https://sagresonline.tce.pb.gov.br/), gerado pelo sistema SAGRES, um sistema voltado a transparência dos gastos públicos onde pode ser encontrado todos os dados relacionados as contas públicas do estado e seus municípios.

O *dataset* de Receitas Orçamentárias da Esfera Municipal possui todos os recebimentos de entidades públicas do estado da Paraíba desde o mês de janeiro de 2003 até o mês atual, sendo atualizado mensalmente pelo Tribunal. Para este estudo, foi filtrado apenas os dados referentes a Prefeitura Municipal de Campina Grande, direcionando assim, o estudo para uma entidade. A descrição dos dados é disponibilizada no site do TCE-PB e pode ser conferida também abaixo.

Nome da coluna/campo	**Descrição**	**Tipo**
cd_ugestora	Identificador da unidade gestora.	Numérico inteiro
de_gestora	Nome da unidade gestora.	Texto
dt_ano	Ano.	Numérico inteiro
cd_receitaorcug	Código da unidade gestora.	Texto
de_receitaorcug	Descrição da receita.	Texto
tp_atualizacaoreceita	Código para a atualização/lançamento da receita.	Numérico inteiro
de_atualizacaoreceita	Descrição da atualização.	Texto
vl_lancamentoorc	Valor do lançamento, sendo o ponto(.) o separador decimal	Numérico monetário
dt_mesano	Mês e ano, na forma MMYYYY	Texto

Figura 1: Descrição das colunas presentes do dataset

4 PROCESSAMENTO/TRATAMENTO DOS DADOS

O *dataset* original, contendo todas as entidades públicas do estado da paraíba, possui 1.641.716 registros com dados de receitas recebidas pelas entidades do ano de 2003 até julho de 2021, mês do inicio do estudo deste trabalho. Alguns passos serão demonstrados neste documento, mas todo o código está disponibilizado em um Jupyter Notebook, onde toda a análise foi realizada. Para ter acesso ao Notebook, acesso o link disponibilizado neste documento.

```
# Definindo tipos das colunas antes de ler o dataset
columns_types = {'cd_ugestora': str, 'dt_mesano': str, 'dt_ano': int}
# Carrega os dados usando pandas
Receitas = pd.read_csv('datasets/TCE-PB-SAGRES-
Receita_Orcamentaria_Esfera_Municipal.
txt',sep='|',dtype=columns_types,encoding='utf-8')
```

Figura 2: Leitura do *dataset* utilizando Pandas

Como o foco principal deste trabalho é na Prefeitura Municipal de Campina Grande, é necessário aplicar um filtro informando o valor da unidade gestora, no caso sendo o valor da cd_unidade_gestora igual a 201050. Assim, nosso dataset fica com 10831 registros.

```
# Obtém todos os dados da Prefeitura Municipal de Campina Grande (cd_ugestora=201050)

pref_mun_CG = receitas[receitas['cd_ugestora'] == '201050']
```

Figura 3: Filtro no dataset obtendo apenas os dados da unidade gestora de interesse

Após uma análise exploratória no *dataset,* foi encontrado alguns registros que demonstraram conter dados nulos na coluna de_receitaorcug. Essa coluna é uma descrição do tipo de receita recebido pela entidade. Então foi necessário preencher esses dados com o texto ‘Sem descrição’, substituindo-os em todos os registros.

```
# Preenchendo os valores nulos

pref_mun_CG.de_receitaorcug.fillna('Sem descrição', inplace=True)
```

Figura 4: Substituição dos valores nulos

A outra verificação foi validar se as informações de mês, na coluna dt_mesano, estavam de acordo com o padrão, padrão este composto por um total de 12 meses. Sabendo disso, dentro da análise exploratória foi notado que algumas entidades possuíam, o campo dt_mesano valores como o mês 13 e, para sanar tal incoerência, foi feita uma execução do comando abaixo, a fim de obter a seguinte saída:

```
#Certificando se possuem meses incoerentes
pref_mun_CG.dt_mesano.str.slice(0,2).value_counts()
```

```
07 941
04 938
12 927
10 920
08 909
01 902
03 897
02 889
06 882
09 881
11 877
05 860
13 8
Name: dt_mesano, dtype: int64
```

Figura 5: Comando para verificar meses incoerentes e resultado da execução

Como foi identificado 8 entradas com mês 13, foi necessária uma análise exploratória do documento para tentar entender e foi constatado erro de digitação pois os meses que constavam com 13, estavam no meio de entradas do mês 12 ou entradas do mês de março, logo depois foi corrigido manualmente para que as informações possam ser utilizadas corretamente.

5 ANÁLISE E EXPLORAÇÃO DOS DADOS

Inicialmente para a exploração dos dados, é importante responder a perguntas provenientes da natureza do dataset. Quais os tipos de receitas que mais aparecem? Quais os anos que mais geraram receita? Quais foram os meses? Houve meses que se apresentam como outliers? Para isso, foi utilizado o Pandas, uma biblioteca utilizada para manipulação de dados com ferramentas de análise, e o *Matplotlib*, outra biblioteca para exibir gráficos com os dados manipulados utilizando o *Pandas*.

Iniciando pelos tipos de receitas que mais aparecem no dataset, utilizei um método personalizado que exibe um gráfico que utiliza o método value_counts do Pandas. Assim, abaixo estão os 6 tipos de receitas que mais se repetem no dataset:

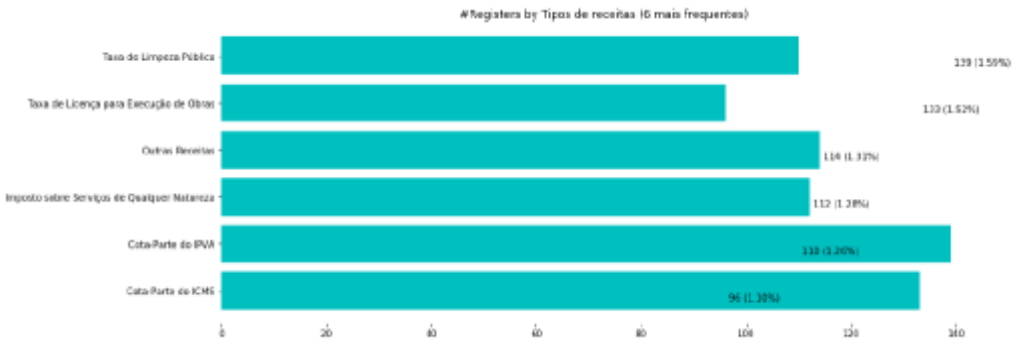


Figura 6: 6 tipos de receitas que mais aparecem no dataset

Como mostrado, acima estão alguns impostos bem conhecidos como Cota do IPVA (Imposto sobre a propriedade de Veículos Automotores), ICMS (Imposto sobre Circulação de Mercadorias e Serviços) e algumas taxas como a taxa de limpeza pública, onde os impostos citados estariam bem presentes, sendo os principais contribuidores das receitas dos municípios, assim como aportes da União.

Realizando agrupamento com a soma das receitas por ano, é obtido o gráfico abaixo com os totais anuais das receitas:

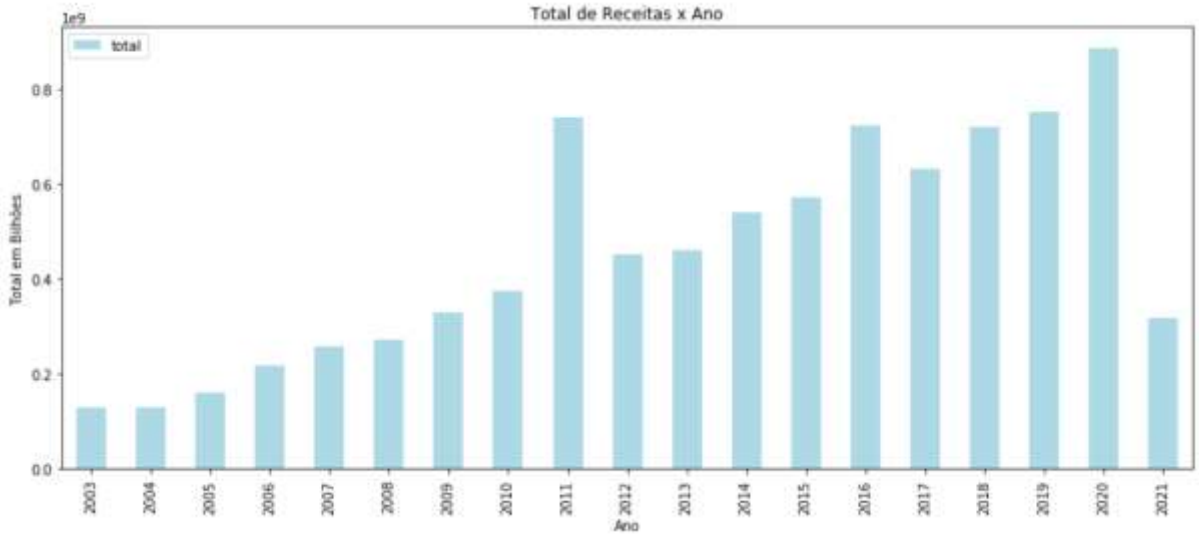


Figura 7: Totais das receitas por ano

O gráfico se comporta numa crescente, em que o ano de 2020 fica em destaque logo seguido do ano de 2019 e 2011. O ano de 2021 está bem abaixo pois esse estudo foi realizado em julho de 2021 e os dados para esse ano ainda serão gerados. O valor total recebido em 2020 foi **R$ 887.014.070,15**.

5.1 Outliers

Um dos pontos importantes é analisar o *dataset* com a finalidade de identificar possíveis valores muito distantes dos demais dados. Esses tipos de dados são considerados *outliers* e atrapalham a criação de modelos de predição de valores, pois geralmente eles não possuem um padrão, e no orçamento pode ser causado por exemplo por uma entrada ou uma saída que não é constante no orçamento.

Um gráfico bloxplot com esse agrupamento mostra que os dados se comportam dentro de um intervalo não considerado outlier.

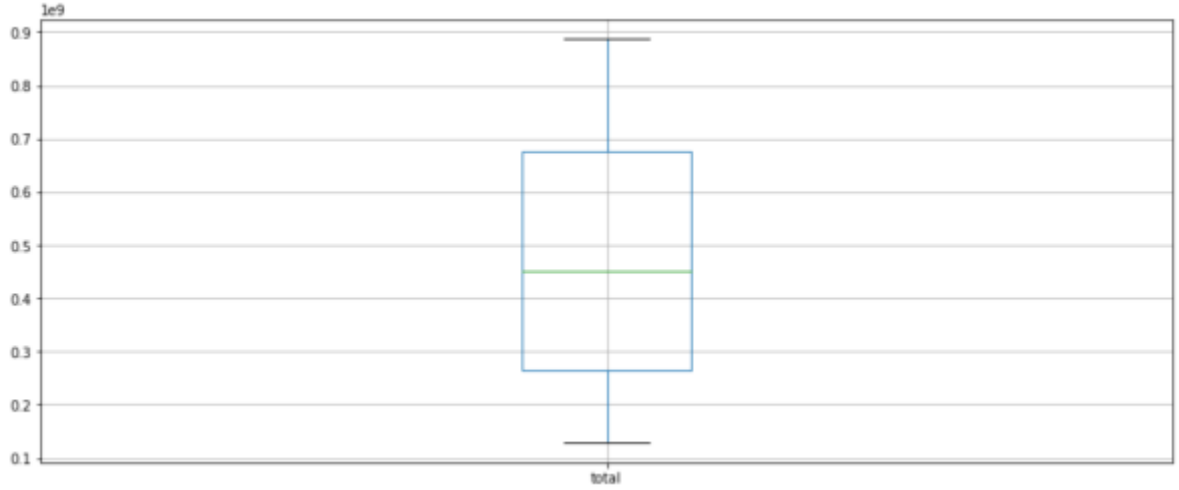


Figura 8: Boxplot dos totais recebidos por ano

Diante do exposto, poderia aparecer o seguinte questionamento: Mas como seria o comportamento dos dados se o agrupamento desses valores fosse por mês, em vez de anos? Mudando o agrupamento dessa forma teremos muito mais pontos a serem informados no gráfico em comparação ao agrupamento em anos. Assim, o próximo gráfico será exibido utilizando linhas em vez de barras.

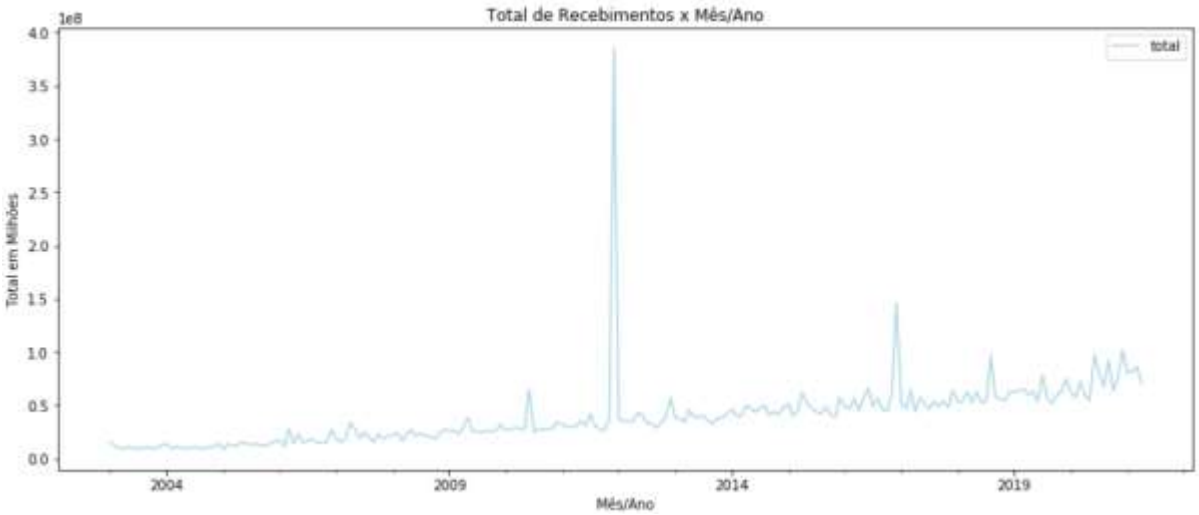


Figura 9: Totais das receitas por mês

Na figura acima fica mais nítido que ocorrem diferenças mais abruptas nas receitas recebidas por mês. Pode ser observado, por exemplo, que o final de 2011 teve um valor bem diferente dos demais, chegando a ser maior que o maior valor de 2020. É importante observar o gráfico *bloxplot* do *dataset* na visão das receitas por mês.

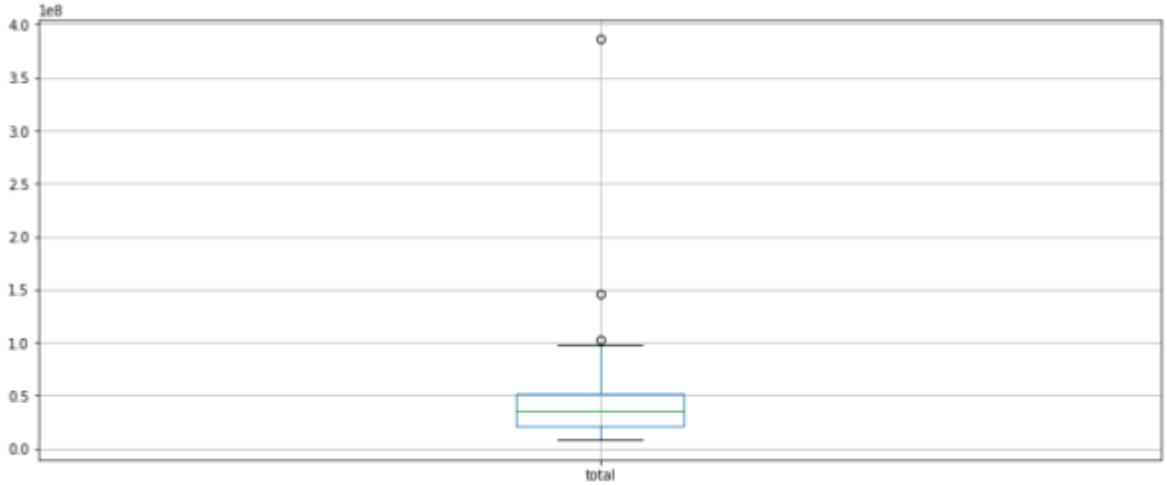


Figura 10: Aparecimento de valor considerado outlier no dataset das receitas por mês

Aqui o gráfico mostra que a média dos valores se encontra bem mais próxima do limite inferior do que o limite superior. Isso mostra que quando observa-se os dados dentro da perspectiva mensal, os dados podem ter um comportamento diferente que o anual, que demonstrou ser mais distribuída ao longo do tempo.

Pode ser notado que um valor se mostra fora dos limites, inicialmente é considerado um valor de *outlier* em relação aos demais do *dataset*, todos eles desde janeiro de 2003 até julho de 2021. O valor identificado se trata dos recebimentos de dezembro de 2011, um total no valor de **R$ 385.871.038,15**.

Diante desse dado tão valioso para o trabalho, surge as seguintes hipóteses: Mas será que este é o único valor que está fora dos limites? Como esse gráfico se comportaria se analisarmos os dados contidos em um único ano individualmente? Dentro de um mesmo ano não existem mais valores fora dos intervalos de máximo e mínimo? Abaixo este gráfico é mostrado.

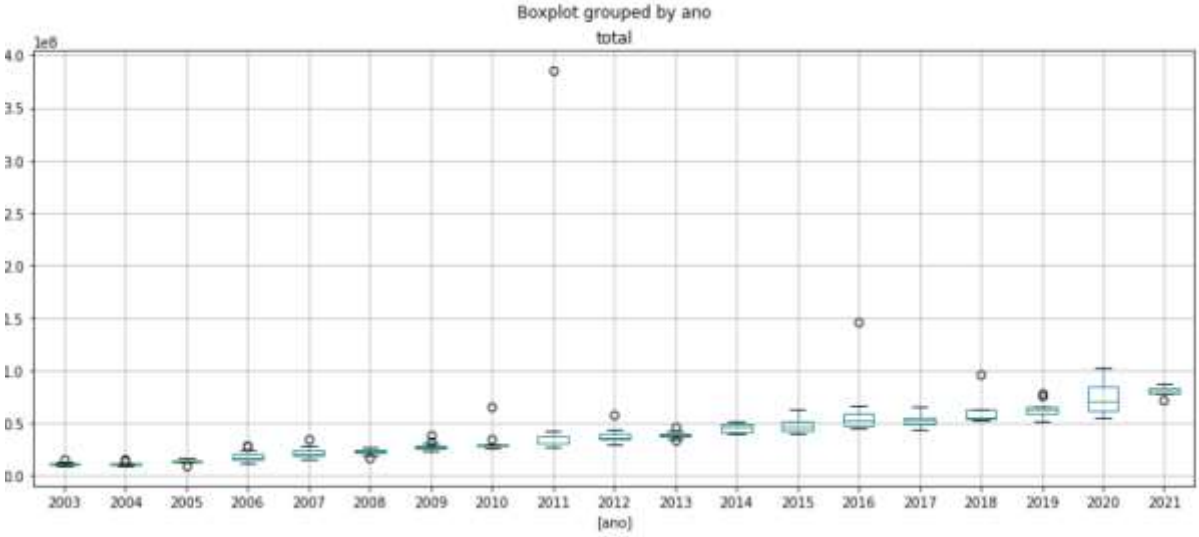


Figura 11: Boxplot dos dados contidos e agrupados por ano

De fato, vários outros pontos são identificados. Esse gráfico mostra que ao analisarmos os dados agrupados por ano, individualmente é possível identificar mais valores que estão fora do máximo e mínimo da distribuição autocontido no ano e também representa como esses valores oscilam vertiginosamente dentro dos anos.

Mas como será essa distribuição visualizando os meses dentro dos anos?

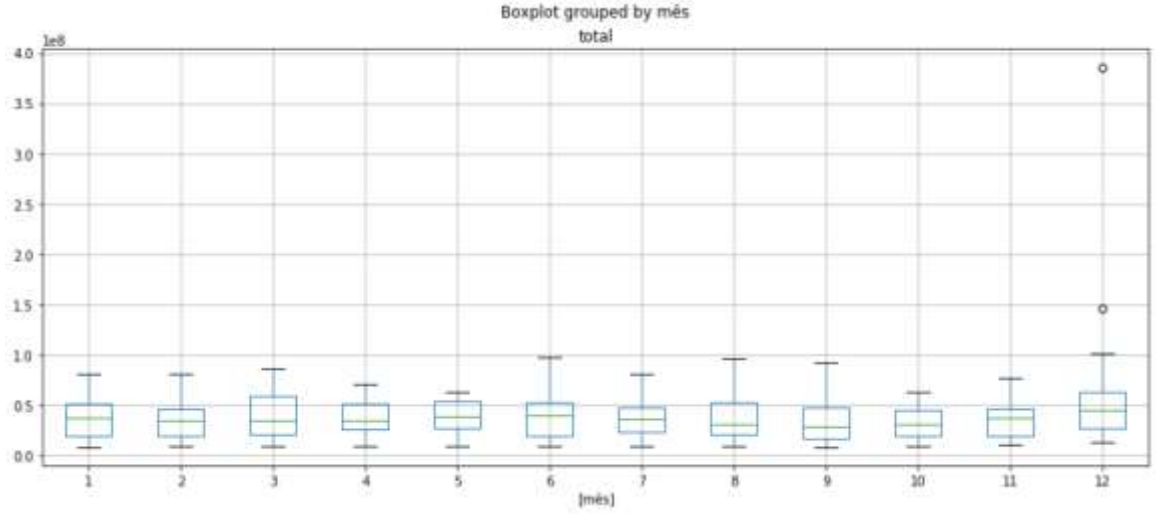


Figura 12: Visualizando a distribuição das receitas dentro de cada mês com passar dos anos

A leitura que pode ser feita do gráfico é que, dentro de cada mês, ao passar dos anos, os valores oscilam e possuem comportamento aceitável, tendo valores de quartis, de máximas e mínimas variadas. Isso pode ser explicado pela natureza do *dataset*. Alguns impostos que impactam nas receitas do município, como IPVA e IPTU, possuem calendários e formas de pagamentos flexíveis, o que pode acontecer

que em determinado ano a maioria das pessoas escolheram vencimentos ou formas de pagamentos que se acumularam em determinados meses que afetam o comportamento da série. Ou recebimentos de receitas com valores impactantes venham de forma assíncrona e arbitrária.

Como a análise da série é levando em consideração os dados mensais dos totais por mês, é importante considerar a figura 9, onde apenas um dos valores foi apontado como outlier relativo aos demais dados. Para o tratamento do dado e com o intuito de não impactar os valores da série, será utilizada a média do mês de dezembro do ano de 2010 (anterior) e do ano de 2012 (posterior). Assim, o valor de dezembro de 2011 passará de R$ **R$ 385.871.038,15** para **R$ 45.781.954,27**.

Assumindo esses valores como os aceitáveis para continuar a análise da série é possível cobrir as características que faltam para realizar a predição dos dados.

5.2 Tendência, Autocorrelação e Médias móveis

Para descobrir os valores a serem utilizados no modelo preditivo e validar a predição a ser realizada o *dataset* utilizado foi divido em 2 partes. Uma base de treino com 80% dos valores e uma base de teste com 20% restantes dos dados. Ao final, o gráfico com a predição terá os dados da base de teste para comparativo com os resultados obtidos. Para que seja possível criar modelos preditivos, é importante que garantir que a série seja considerada estacionária.

Para isso é necessário ter 3 elementos constantes, a média, a variância e a autocorrelação. A primeira verificação que deve ser realizada é checar se a série possui tendência.

5.2.1 Tendência

Para identificar se a série possui tendência, a biblioteca de modelos de estatísticas do Python possui um método para decomposição de séries temporais chamado seasonal_decompose().

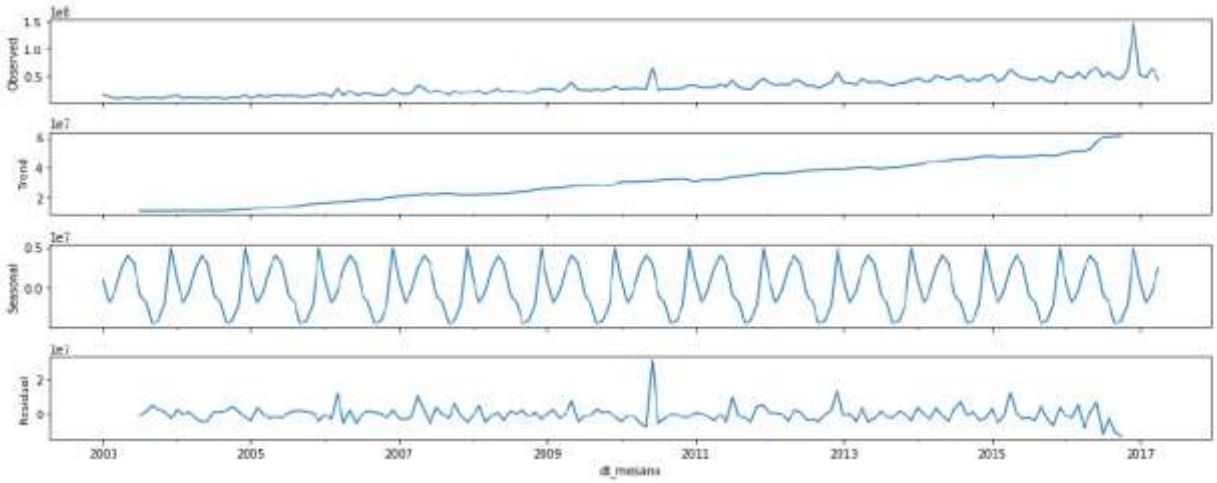


Figura 13: Confirmando a tendência da série

Como pode-se confirmar olhando o gráfico Trend, a série possui uma tendência crescente, assim torna-se necessário aplicar técnicas para tentar anular essa tendência. Existem testes que podem ser aplicados para comprovar que, estatisticamente, a série é ou não estacionária. Aqui foi usado o *Augmented Dickey-Fuller Test*, aqui caso o p-value calculado seja abaixo de 5%, pode se considerar que a série pode ser considerada estacionária.

Utilizando a série original de treinamento temos o resultado abaixo:

```
ADF Statistic: 1.696383
p-value: 0.998120
Critical Values:1%: -3.472
5%: -2.880
10%: -2.576
```

Figura 14: Teste ADF da série original

Como o valor de *p* foi de 99% está comprovado que a série é não estacionária. Com o intuito de corrigir esta tendência foi aplicado uma diferenciação à série. Criando novamente o gráfico, pode ser notado como fica a tendência da nova série.

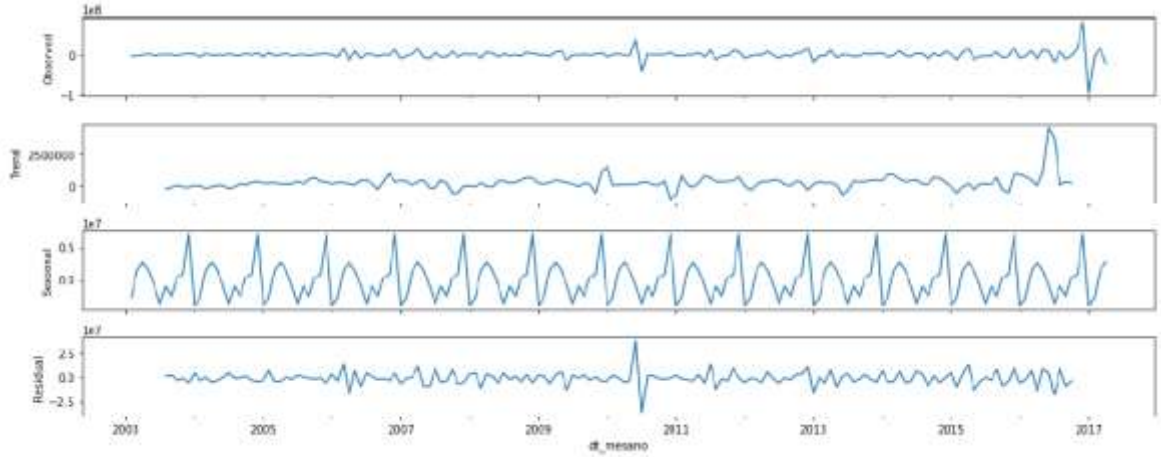


Figura 15: Tendência da série reduzida aplicando uma diferenciação

Agora é possível notar que a tendência (gráfico Trend) varia em torno de 0. Para validar isso, é necessário executar novamente o teste e validar o p-value. Reexecutando o teste de Dickey-Fuller na série com uma diferenciação temos o seguinte valor de p.

```
ADF Statistic: -6.704302
p-value: 0.000000
Critical Values:1%: -3.472
5%: -2.880
10%: -2.576
```

Figura 16: Teste ADF da série com 1 diferenciação

Com o valor de p sendo 0 posso considerar que uma diferenciação foi suficiente para estacionar a série.

5.2.2 Autocorrelação

A autocorrelação informará qual o grau de relação que cada dado tem com os demais dados passados da série. Isso ajudará a definir o valor p para ser utilizado no modelo preditivo.

Para isso, será utilizado o gráfico ACF (*Autocorrelation Factor*), que irá mostrar o número de *lags* que estão fora do intervalo de confiança. Esse valor será o nosso AR.

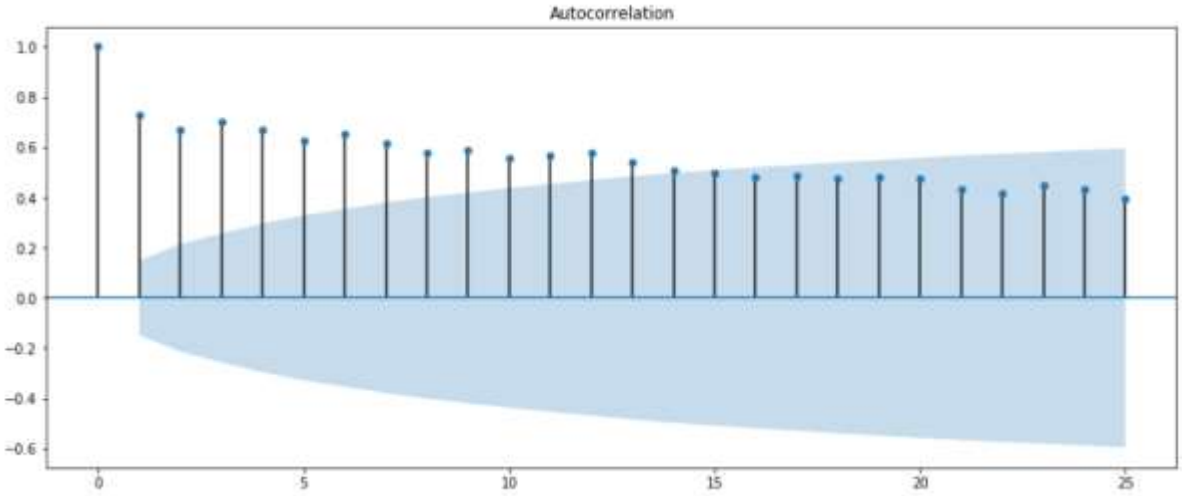


Figura 17: Gráfico ACF da série

Como o gráfico mostra, os valores ficam contidos no intervalo a partir do lag 14, então os valores a serem utilizados estará perto do valor 13. Após alguns testes e verificando o AIC (Critério de informação de Akaike) e BIC (Critério Bayesiano de Schwarz) do modelo, foi decidido usar o valor 10.

5.2.3 Médias móveis

Para obter o valor das médias móveis e termos uma pista sobre o valor *q*, utilizaremos o gráfico PACF (*Partial Autocorrelation Factor*).

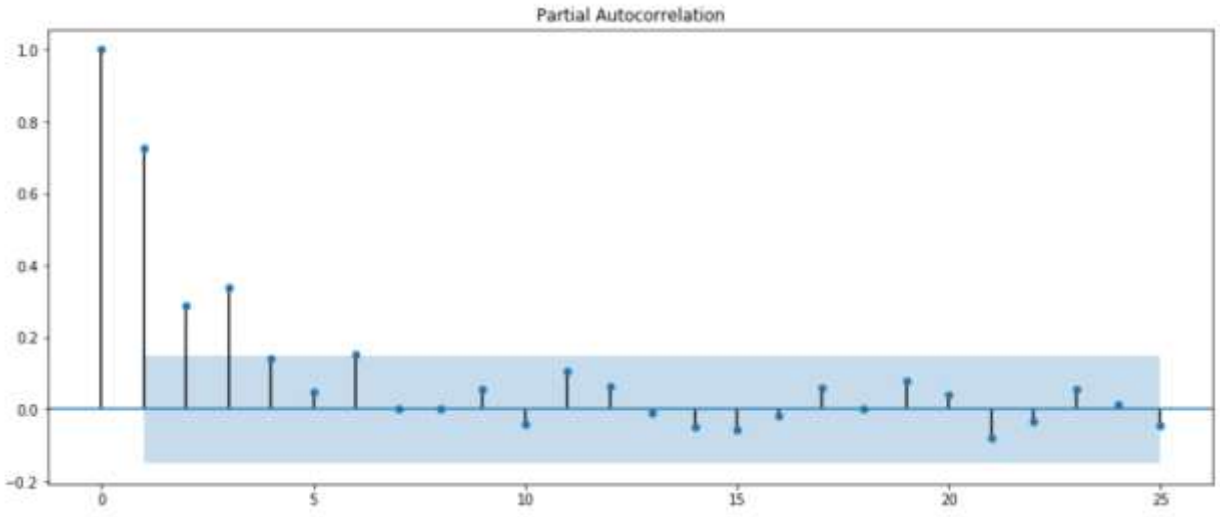


Figura 18: Gráfico PACF da série

Pode-se notar a partir desse gráfico que os três primeiros valores têm uma significância correlativa alta em relação aos demais. Então usarei o valor 3 inicialmente para definir o valor *q* do modelo preditivo.

6 MODELO DE PREDIÇÃO ORÇAMENTÁRIA

Após a construção do cenário, o modelo ficou com os seguintes parâmetros presentes na tabela:

```
# Criando o modelo

modelo = ARIMA(ts_treino, order=(10,1,3), freq=ts_treino.index.inferred_freq)
```

Figura 19: Parâmetros para o modelo ARIMA

Ao ajustar os parâmetros do modelo, é possível obter o seguinte resultado para os valores de AIC e BIC:

ARIMA Model Results

Dep. Variable:	D.total	**No. Observations:**	171
Model:	ARIMA(10, 1, 3)	**Log Likelihood**	-2971.705
Method:	css-mle	**S.D. of innovations**	8256716.868
Date:	Tue, 31 Aug 2021	**AIC**	5973.410
Time:	16:09:05	**BIC**	6020.535
Sample:	02-01-2003	**HQIC**	5992.532
	- 04-01-2017		

Figura 20: Resultados do modelo ARIMA(10,1,3) definido pelo estudo

Com os parâmetros do modelo ajustados, pode-se obter informações sobre os resíduos do modelo. Assim sendo possível averiguar o comportamento dos residuais e sua densidade.

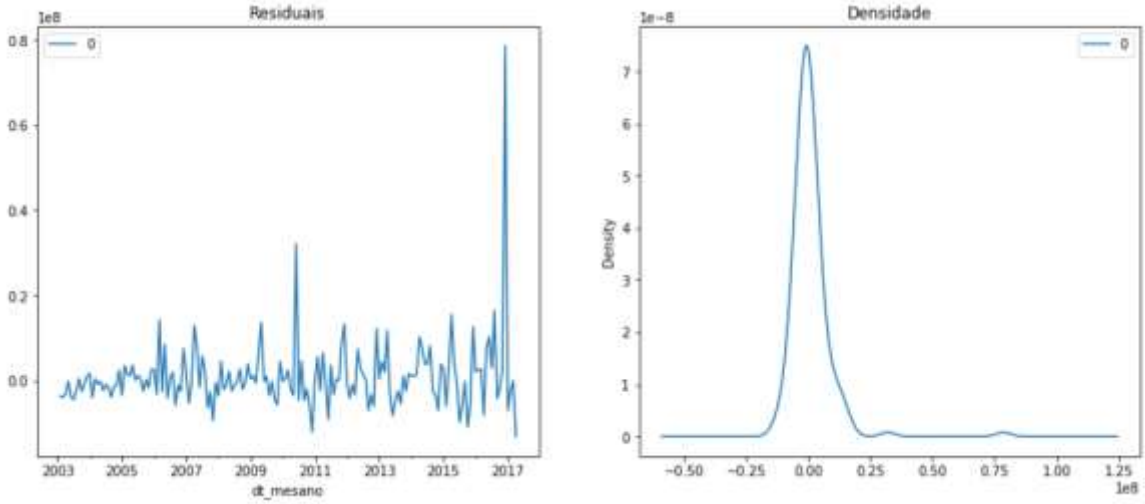


Figura 21: Dados residuais e sua densidade

Os resíduos ficam em torno de 0 com uma variação satisfatória, com os pontos tendendo a permanecer com média em torno de 0. E a densidade dos dados se comporta como uma distribuição normal

Agora é possível ver como os dados do modelo e os dados reais de treino exibidos em um mesmo gráfico nos dará uma ideia do comportamento do modelo.

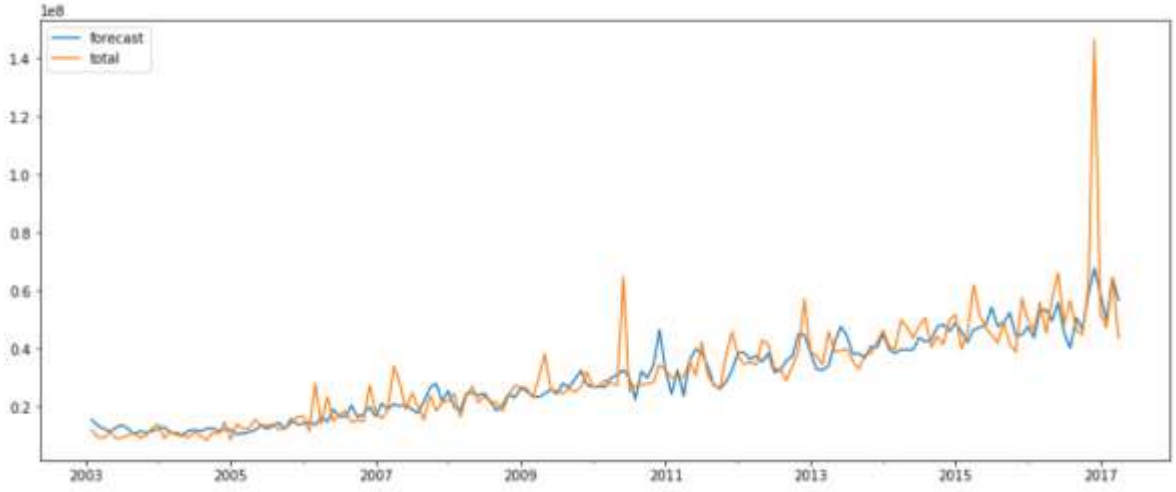


Figura 22: Dados reais x modelo ajustado

Com isso é notável aqui que os dados do modelo se comportam de maneira similar aos dados reais, todas as variações e de alta de baixa dos dados reais são acompanhadas pelo modelo, não assumindo exatamente os mesmos valores de pico e queda, mas sendo suficiente para ser utilizado como objeto deste estudo.

6.1 Predição

Após a realização de todo o estudo, é possível agora realizar a predição dos dados em comparação aos dados da base de teste validando assim o modelo proposto.

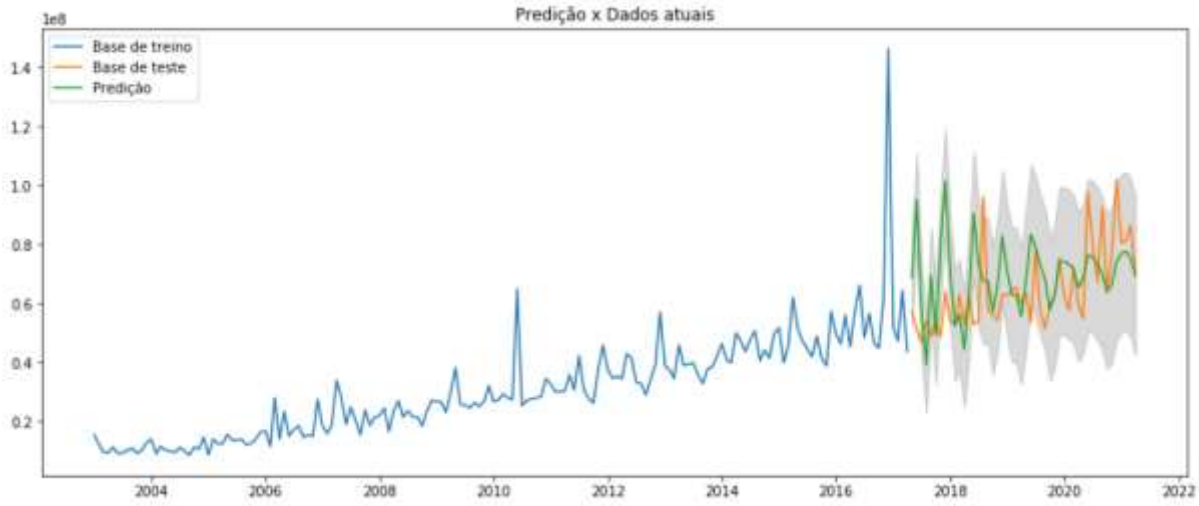


Figura 23: Dados reais x Predição

Os dados obtidos neste gráfico mostram que a predição utilizando o modelo ARIMA(10,1,3) exibe pontos bem próximos dos reais, com exceção dos picos que estão presentes fora do intervalo de confiança do modelo, o comportamento da curva acompanha os dados reais da base de teste.

7 CONCLUSÕES

O estudo de séries temporais é um dos assuntos mais interessantes na área de ciência de dados, e a medida em que se aprofunda no assunto mais se destaca a importância da matemática estatística nesse universo. Com isso é possível notar o quanto este mundo e o da informática devem e tendem a se fundir para atender às necessidades atuais sobre o estudo analítico dos dados gerados diariamente pelas pessoas e pelos sistemas.

Sobre este estudo, todas as etapas, desde a obtenção dos dados, o tratamento das informações e a análise do *dataset* para que seja possível gerar informação, tem uma profunda importância nos passos futuros. A fase mais fascinante foi a de análise da série e como obter os parâmetros para a predição de dados futuros de acordo com o modelo. Nesta etapa é onde se destacam os profissionais matemáticos que conseguem analisar os gráficos levando em consideração sua natureza e seu comportamento em uma linha do tempo. A análise não acontece friamente com base apenas em números, mas sim em o que esses números representam em um ambiente computacionalmente vivo e evolutivo.

Como pontos de melhoria no estudo seria interessante um tratamento mais minucioso nos *outliers* que pode melhorar o desempenho do modelo assim como experimentar outros tipos de modelo além do ARIMA. Uma dificuldade encontrada foi devido à natureza do dataset. Diferentemente de um dataset que poderia ter dados de temperaturas de um lugar onde a variância é conhecida e a identificação de outliers derivados de medições problemáticas ou defeitos de equipamentos, os dados de receitas orçamentárias de um município que são validadas por um órgão de competência como o Tribunal de Contas do Estado podem sofrer oscilações bem consideráveis, mas certamente justificáveis, como o recebimento de aportes da União para resolver questões momentâneas, e para isso cabe um estudo histórico mais aprofundado e detalhado para definir o que realmente é outlier e como tratá-los.

REFERÊNCIAS

ALVES, F. O. et. al. **O Plano Plurianual (PPA)**. Atividade acadêmica. Curso de Pós-graduação em Gestão e Contabilidade Pública. Universidade Estadual do Piauí, Parnaíba, PI, 2013.

ANÁLISE ESTATÍSTICA OLHE AO SEU REDOR. A ESTATÍSTICA ESTÁ EM TODOS OS LUGARES. SAS, 2021. Disponível em: <https://www.sas.com/pt_br/insights/analytics/analise-estatistica.html>. Acesso em: 25 de out. de 2021.

BRASIL. **Constituição** (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

BROCKWELL, Peter J; DAVIS, Richard A. **Introduction to Time Series and Forecasting**. 2. Ed. Nova York: Springer texts in statistics, 2002.

Confederação Nacional dos Municípios - CNM. **Finanças Públicas: Noções Básicas para os Municípios**. Brasília: CNM, 2008.

HYNDMAN, Rob J.; Athanasopoulos, George. **Forecasting: Principles and Practice.** 2. Ed. Melbourne, Austrália: Otexts, 2018.

MENEZES, N. N. C, **Introdução a programação com Python**. São Paulo: Novatec, 2014

PYTHON SOFTWARE FOUNDATION. **Python Language** Site: Documentation, 2021. Página de documentação. Disponível em: <https://www.python.org/doc/>. Acesso em: 06 de jul. de 2021.