

# ANÁLISE E PREVISÃO DA VARIAÇÃO DE EMPREGOS NO COMÉRCIO NO MUNICÍPIO DO RIO DE JANEIRO

**Bruno Felipe Penido Maciel Reis (UFF)**

bruno\_felipe@id.uff.br

**Eliane da Silva Christo (UFF)**

**Kelly Alonso Costa (UFF)**

**Hugo Pimentel Tavares (UFF)**

**Cecilia Toledo Hernandez (UFF)**



*O setor do comércio exerce um papel fundamental em qualquer município, sendo importante para o desenvolvimento econômico e a qualidade de vida da população local. Muitos negócios nesse setor contribuem significativamente para a geração de empregos na comunidade, por isso se faz importante entender essa dinâmica ao longo do tempo. Ter uma visão analítica dos dados à respeito da variação de empregos pode ajudar os governos a tomar decisões mais estratégicas e estimular de maneira correta o setor. A melhor maneira de realizar essa análise é através de séries temporais, que indicam a variação de empregos no comércio no decorrer do tempo. Portanto, a série temporal contendo o saldo de admissões e demissões recebeu um trato analítico com três modelos de previsão diferentes: Amortecimento Exponencial Holt-Winters, Modelo ARIMA e SARIMA. Através da análise dos erros de cada previsão foi possível encontrar aquele com melhor potencial para previsões.*

*Palavras-chave: ARIMA, SARIMA, HOLT-WINTERS, EMPREGOS, PREVISÃO.*

## **1. Introdução**

A geração de empregos é uma variável extremamente importante, principalmente pelo fato de ser utilizada como termômetro econômico e de desenvolvimento social. De acordo com Ministério do Trabalho e Emprego, o CAGED (Cadastro Geral de Empregados e Desempregados) é um registro permanente de admissões e demissões, tendo sua base de dados disponibilizada para consulta e pesquisas. Segundo o Ministério do Trabalho e Emprego (2015), esse Cadastro serve como um direcionador de tomadas de decisões para diversas ações públicas.

Portanto, foi com base nessa base de dados que o presente documento foi elaborado. Este artigo tem como objetivo aplicar três modelos de previsão para analisar a geração de empregos no município do Rio de Janeiro, partindo de janeiro de 2016 até dezembro de 2023. A estatística analisada parte do saldo entre admissões e demissões, ou seja, a subtração entre o total de admissões em um mês e o total de demissões desse mesmo mês. De acordo com o Ministério do Trabalho e Emprego (2016), esse saldo indica a variação absoluta  $A - D$ , que é muito importante para entender a dinâmica de geração de empregos.

## **2. Metodologia**

A pesquisa que foi realizada, é de natureza básica, uma vez que visa ampliar os estudos já existentes. Possui abordagem quantitativa, pois realizou-se a análise estatística para se chegar nos resultados. Possui um objetivo exploratório que permite familiarizar os envolvidos quanto ao tema principal, buscando ampliar os estudos existentes sobre determinadas ferramentas. Por último, quanto aos procedimentos, é uma pesquisa documental, analisando dados e realizando um tratamento analítico.

A coleta de informações e dados foi realizada integralmente através de sites e bancos de dados online. A base de dados, na qual foi objeto de estudo, foi coletada no site DATA.RIO, que disponibiliza diversos estudos, relatórios e bases de dados referentes a diferentes aspectos sociais, ambientais e demográficos.

Após a coleta dos dados, foi realizada sua análise através de dois softwares, sendo eles o Microsoft Excel e o Google Colab, onde é possível utilizar as bibliotecas da linguagem Python. Os resultados foram expostos e ponderados utilizando medidas estatísticas de erro.

## **3.Referencial Teórico**

### **3.1. Amortecimento Exponencial Holt-Winters**

O amortecimento exponencial triplo é utilizado para séries temporais que possuem tendência e sazonalidade. Segundo Yaffee e McGee (2000), o método Holt-Winters possui três constantes, sendo elas:  $\alpha$ (constante de ajustamento),  $\beta$ (constante de tendência) e  $\gamma$ (constante de sazonalidade).

A equação utilizada para este modelo sugere uma previsão baseada em três componentes, sendo eles N(componente nível), T(componente tendência) e I(componente de sazonalidade):

$$T_t = \beta * (N_t - N_{t-1}) + (1-\beta) * T_{t-1}$$

$$N_t = \alpha * (R_t / I_{t-1}) + (1-\alpha) * (N_{t-1} + T_{t-1})$$

$$I_t = \gamma * (R_t / N_t) + (1-\gamma) * I_{t-L}$$

Após obter os componentes, é realizado a previsão:

$$P_{t+1} = (N_t + T_t) * I_{t-L+1}$$

### 3.2 ARIMA(Autoregressive Integrated Moving Average)

Segundo Yaffee e McGee (2000), o modelo ARIMA possui três diferentes componentes: AR- Auto regressivo, I- Integrado e MA- Média Móvel. Sendo assim, esse método combina esses três componentes para realizar a análise.

AR( $p$ ) - O valor atual depende de valores passados.

I( $d$ ) - Quando a série necessita de diferenciação para se tornar estacionária.

MA( $q$ ) - A parte do modelo representa a média móvel.

Segundo Morettin e Toloi (1981), para usar o modelo ARIMA, a série deve ser estacionária, com isso podendo garantir uma certa estabilidade com o decorrer do tempo. Para torná-la estacionária é necessário realizar a diferenciação, representado pelo componente I( $d$ ). Geralmente se realiza uma diferenciação, raramente passando disso. Para definir os melhores componentes AR( $p$ ) e MA( $q$ ) é necessário examinar as funções de autocorrelação e autocorrelação parcial da série estudada.

### 3.3 SARIMA(Seasonal Autoregressive Integrated Moving Average)

O modelo SARIMA se caracteriza como o ARIMA sazonal, ou seja, possui o mesmo princípio porém com a parte sazonal. De acordo com Yaffee e McGee (2000), esse modelo é representado como ARIMA( $p,d,q$ )( $P,D,Q$ )<sub>s</sub>, sendo  $P$ ,  $D$  e  $Q$  maiúsculos os componentes sazonais. O componente  $s$  representa a ordem de sazonalidade.

## 4. Resultados

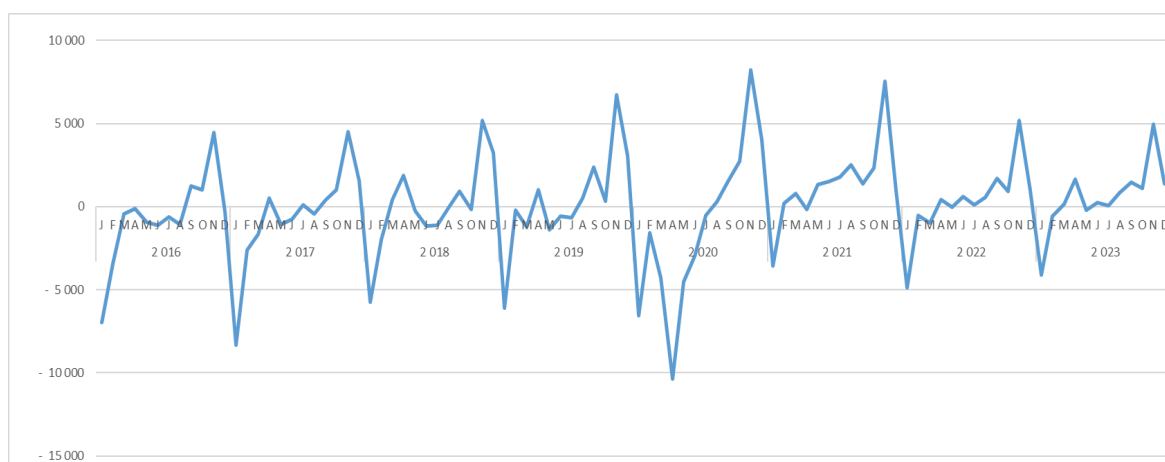
A série temporal utilizada como base de dados foi o saldo de admissões e desligamentos no município do Rio de Janeiro no setor do comércio, desde janeiro de 2016 a dezembro de 2023. Este saldo é calculado subtraindo o total de desligamentos do total de admissões, sendo esse saldo mensal a variação absoluta.

Os métodos de análise utilizados foram o amortecimento exponencial triplo, ou Holt-Winters, ARIMA e SARIMA. Com essa análise de dados, foi gerado o indicador de erro percentual médio absoluto e assim escolher o melhor modelo.

### 4.1. Base de Dados

Como foi dito anteriormente, a base de dados utilizada foi a série histórica com o saldo total de admissões e desligamentos no município do Rio de Janeiro, mais especificamente no setor do comércio, desde janeiro de 2016 a dezembro de 2023. A Figura 1 apresenta a série estudada:

Figura 1: Empregos no comércio.



Fonte: Autoria própria

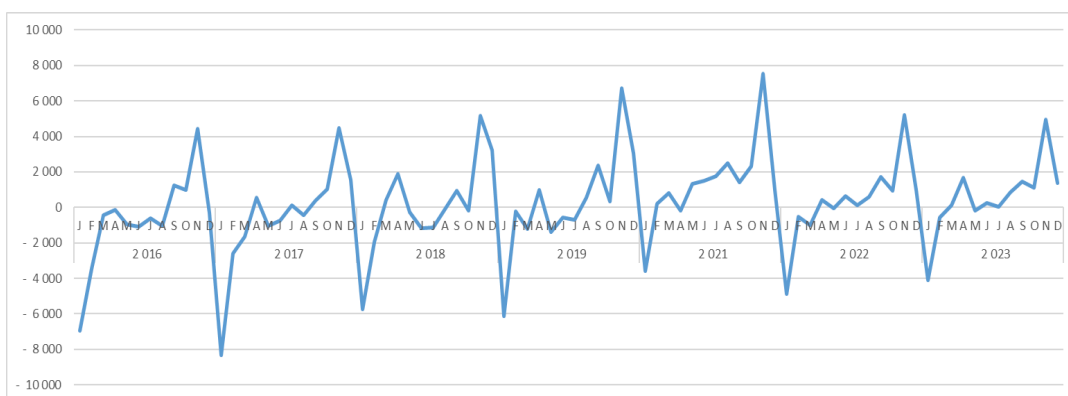
Visualmente, é possível identificar uma repetição de ciclos, ou seja, o fenômeno estudado possui natureza sazonal. Também se revelam diversos picos e vales acentuados na série, além de uma tendência crescente que é possível visualizar no software Microsoft Excel. Essas duas informações são de extrema importância pois permite logo no início selecionar aqueles modelos mais adequados.

É possível identificar uma mudança nos padrões no período de fevereiro até agosto de 2020. Este intervalo atípico pode ser explicado pelo período de pandemia de COVID-19, na

qual houve perdas humanas irreparáveis mas também muitas consequências socioeconômicas. Sendo assim, é compreensível a queda acentuada de geração de empregos nesse período.

Portanto, dado que este período demonstra um fenômeno totalmente fora dos padrões, foi optado pelo recorte do ano de 2020. Tal decisão, apesar de diminuir a amostra, torna a análise mais realista. Esse recorte foi crucial para a modelagem utilizando os modelos autoregressivos. A Figura 2 contém a série com o recorte:

Figura 2: Empregos no comércio com recorte.



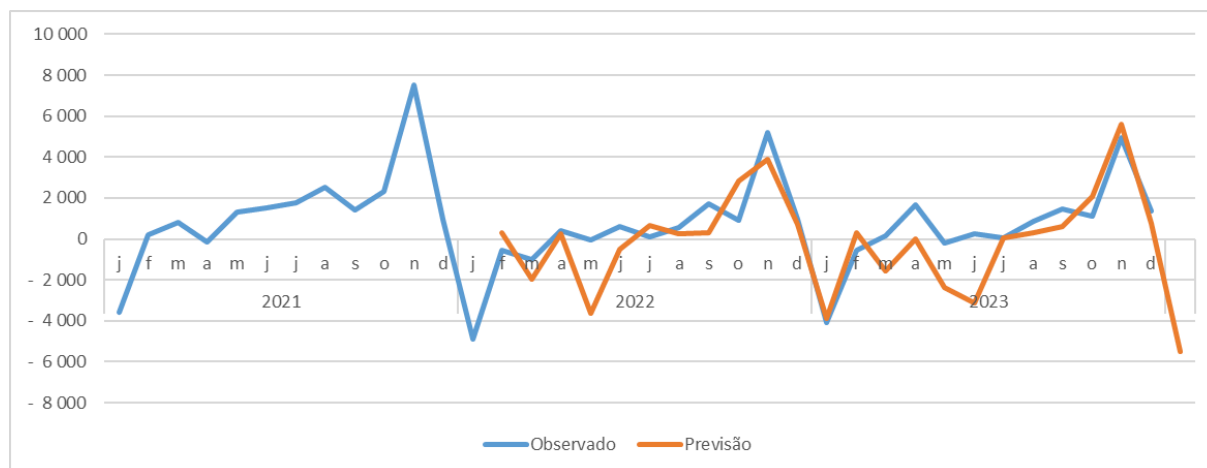
Fonte: Autoria própria

#### 4.2. Análise com o método Holt-Winters

A primeira análise foi feita direcionada para algum modelo que se adequasse a tendência e sazonalidade. Sendo assim, o primeiro teste ocorreu utilizando o método de amortecimento exponencial triplo ou Holt-Winters. A modelagem foi feita no Microsoft Excel, sendo utilizado para a modelagem os últimos 3 anos da série temporal, 2021, 2022 e 2023. Os coeficientes utilizados foram:  $\alpha = 0,9$ ,  $\beta = 0,1$  e  $\gamma = 0,9$  pois apresentaram os menores erros percentuais.

O processo de aplicação do modelo se deu, primeiramente, pela definição dos índices de sazonalidade iniciais com os dados observados do ano de 2021. Para esses índices iniciais utilizou-se: (dado registrado no mês / média dos dados do ano). Para 2022 e 2023 foi utilizado a equação  $I_t = \gamma \cdot (R_t / N_t) + (1-\gamma)I_{t-L}$ . Para o primeiro valor de  $N_t$  foi utilizado o valor real do período de dezembro de 2021 que é 981, daí em diante foi utilizado a equação  $N_t = \alpha(R_t / I_{t-L}) + (1-\alpha)(N_{t-L} + T_{t-L})$ . Para o primeiro valor de  $T_t$  foi utilizado 0, daí em diante foi utilizado a equação  $T_t = \beta \cdot (N_t - N_{t-L}) + (1-\beta)T_{t-L}$ . Foi feita uma previsão para o período de janeiro de 2024 tendo como resultado um saldo negativo de -5522. A Figura 3 contém a análise feita:

Figura 3: Previsão Holt-Winters



Fonte: Autoria própria

Por fim, foi realizado o registro e análise do erro percentual associado à previsão dos últimos 5 períodos já observados de 2023, tendo um erro percentual médio absoluto de 53%. Na Tabela 1 se encontra o resultado do MAPE elaborado no Excel:

Tabela 1: Erro Holt-Winter

Holt-Winters					
Mês(2023)	Observado	Previsão	Erro Absoluto	Erro percentual absoluto	MAPE
Ago	848	302,94	545,0569471	0,642755834	53%
Set	1461	599,14	861,8590913	0,589910398	
Out	1126	2082,78	956,7772798	0,849713392	
Nov	4962	5619,47	657,474427	0,1325019	
Dez	1385	752,60	632,3997308	0,456606304	

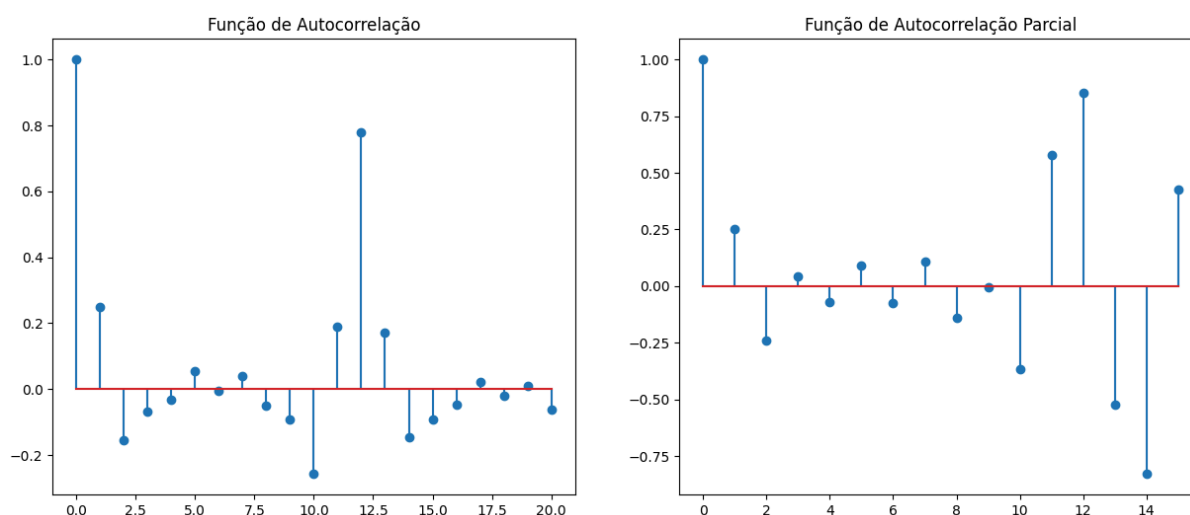
Fonte: Autoria própria

## 4.2. Análise com ARIMA

A segunda análise se deu em razão da busca por um modelo mais preciso, que gerasse resultados com um menor erro percentual. Na busca por um método de previsão mais robusto, optou-se por utilizar o método ARIMA, sendo adequado para séries que apresentam picos mais acentuados. A escolha também se deu pela análise da FAC e FACP, que será detalhada mais à frente.

Para o uso do modelo ARIMA foi utilizada a linguagem Python na plataforma Google Colab, que permite utilizar bibliotecas variadas do Python em um ambiente colaborativo. Primeiramente, foi realizada a análise dos *lags* da FAC (Função de Autocorrelação Parcial) e FACP (Função de Autocorrelação Parcial) para analisar o comportamento de dependência serial.

Figura 4: FAC e FACP



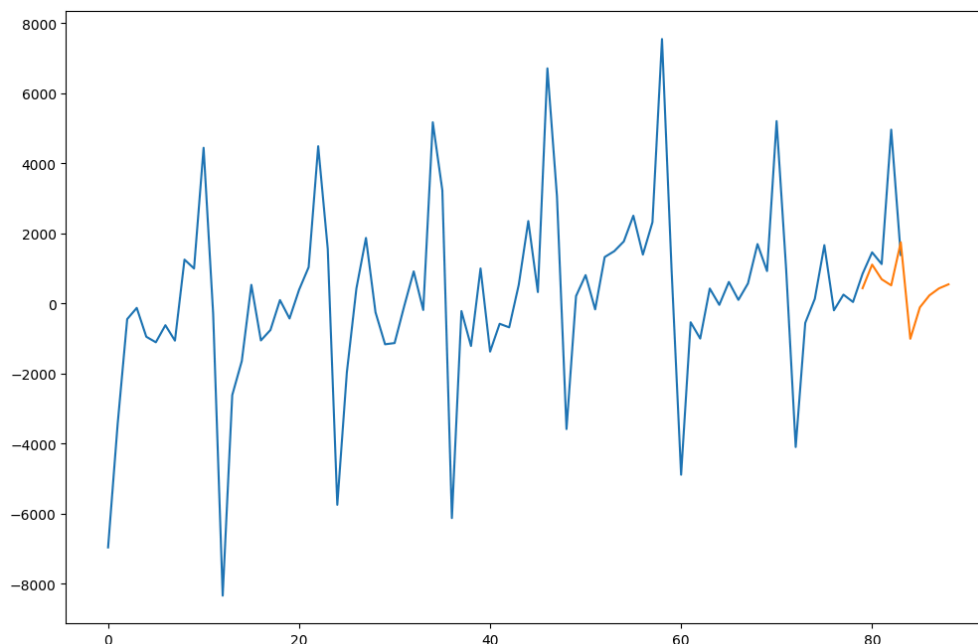
Fonte: Autoria própria

Logo após, foi feito o teste de Fuller para estacionariedade, sendo necessário saber se será utilizado a etapa de diferenciação. Com o teste, foi indicado a não estacionariedade, sendo necessário realizar a etapa de diferenciação.

Em sequência, foi utilizado o recurso auto arima, que consiste no programa gerar os resultados otimizados de  $p$ ,  $d$  e  $q$  para o modelo, sendo esses  $p = 1$ ,  $q = 1$  e  $d = 3$ . Portanto, foi utilizado o modelo ARIMA(1,1,3). Com base nisso, foi feita uma previsão para 5 períodos no futuro, o que equivale no momento de escrita deste artigo aos meses de janeiro, fevereiro, março, abril e maio do ano de 2024. Para medir o desempenho do modelo foi analisado o erro

percentual médio absoluto nos últimos 5 períodos de 2023. A Figura 5 contém a análise feita:

Figura 5: Previsão ARIMA(1,1,3)



Fonte: Autoria própria

A Tabela 2 apresenta os valores previsto no Python para os 5 períodos futuros para o ano de 2024:

Tabela 2: Previsão Arima

Previsão ARIMA		
Mês(2024)	Período	Previsão
Jan	84	-1005,96
Fev	85	-109,77
Mar	86	232,35
Abr	87	431,67
Mai	88	547,78

Fonte: Autoria própria

Por fim, foi realizado o registro e análise do erro percentual associado à previsão dos últimos 5 períodos já observados de 2023, tendo um erro percentual médio absoluto de 45%. Na Tabela 3 se encontra o resultado do MAPE elaborado no Excel:



Tabela 3: Erro Arima

Arima					
Mês(2023)	Observado	Previsão	Erro Absoluto	Erro percentual absoluto	MAPE
Ago	848	437,289531	410,710469	0,484328383	45%
Set	1461	1110,81603	350,18397	0,239687864	
Out	1126	692,508078	433,491922	0,384983945	
Nov	4962	518,965672	4443,034328	0,895411997	
Dez	1385	1746,837775	361,837775	0,261254711	

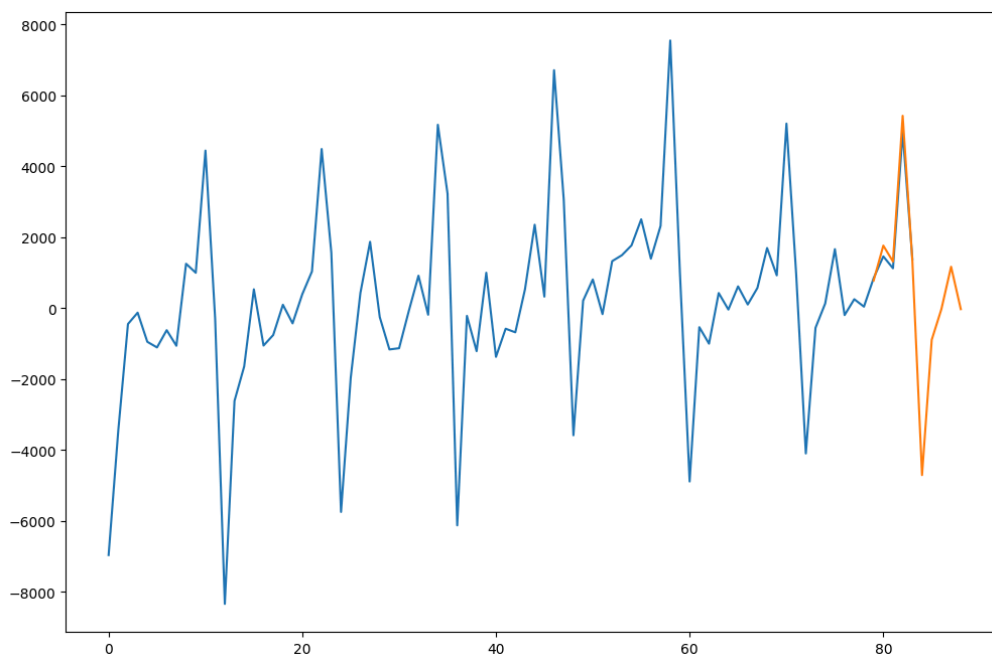
Fonte: Autoria própria

## 4.2. Análise com SARIMA

A terceira análise, assim como a segunda, teve como objetivo encontrar um modelo mais preciso que o anterior. Porém, considerando a sazonalidade da série, foi escolhido para teste o modelo SARIMA, que é apropriado pois a série temporal analisada possui uma repetição de padrões em ciclos.

Para o uso do modelo SARIMA, também foi utilizado a linguagem Python no ambiente do Google Colab. O processo de aplicação do modelo foi similar ao anterior, sendo a análise das FAC e FACP já realizadas na aplicação do ARIMA assim como o teste Fuller e a diferenciação. Os valores de  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$  e  $Q$  foram ajustados para o melhor resultado possível, sendo  $p = 1$ ,  $d = 1$ ,  $q = 2$ ,  $P = 1$ ,  $D = 1$  e  $Q = 1$ . Como a sazonalidade é de 12 períodos, temos o modelo SARIMA(1,1,2)(1,1,1)<sub>12</sub>. Com base nisso, foi feita uma previsão para 5 períodos no futuro, o que equivale no momento de escrita deste artigo aos meses de janeiro, fevereiro, março, abril e maio do ano de 2024. Para medir o desempenho do modelo foi analisado o erro percentual médio absoluto nos últimos 5 períodos de 2023. A Figura 6 contém a análise feita:

Figura 6: Previsão ARIMA(1,1,2)(1,1,1)<sub>12</sub>



Fonte: Autoria própria

A Tabela 4 apresenta os valores previsto no Python para os 5 períodos futuros para o ano de 2024:

Tabela 4: Previsão Arima

Previsão SARIMA		
Mês(2024)	Período	Previsão
Jan	84	-4705,86
Fev	85	-891,13
Mar	86	-24,66
Abr	87	-1169,75
Mai	88	-26,41

Fonte: Autoria própria

Por fim, foi realizado o registro e análise do erro percentual associado à previsão dos últimos 5 períodos já observados de 2023, tendo um erro percentual médio absoluto de 12%. Na Tabela 5 se encontra o resultado do MAPE elaborado no Excel:

Tabela 5: Erro Sarima

Sarima					
Mês(2023)	Observado	Previsão	Erro Absoluto	Erro percentual absoluto	MAPE
Ago	848	779,94	68,06	0,080259434	12%
Set	1461	1765,32	304,32	0,208295688	
Out	1126	1315,85	189,85	0,168605684	
Nov	4962	5426,2	464,2	0,093550988	
Dez	1385	1345,14	39,86	0,028779783	

Fonte: Autoria própria

## 5. Conclusão

A análise realizada demonstrou que o modelo SARIMA se mostra superior ao amortecimento exponencial Holt-Winters e ao ARIMA. Porém, tal resultado não torna os outros modelos inválidos pois são muito úteis em outras situações. O método Holt-Winters poderia ter um melhor desempenho caso a série tivesse picos menos acentuados. Já o modelo ARIMA teria um melhor desempenho caso o efeito da sazonalidade fosse menor.

Com o recorte realizado no período pandêmico, obteve-se uma análise livre de distorções não representativas, o que torna os resultados mais precisos, principalmente em modelos autorregressivos. Portanto, é possível concluir que no município do Rio de Janeiro, o setor do comércio tem um pico de contratações no período de setembro a dezembro, em razão das datas comemorativas de natal e ano novo, período em que as vendas no comércio são maiores e os empregos temporários tendem a crescer. Em contrapartida, no período de janeiro a fevereiro há uma queda acentuada, havendo mais desligamentos do que contratações.

## REFERÊNCIAS

MAYWORM, F. et al. ANÁLISE E PREVISÃO DA EMISSÃO DE GÁS CARBÔNICO NA CIDADE DE SÃO PAULO. XLIII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2023. Disponível em: <[https://www.abepro.org.br/biblioteca/TN\\_ST\\_401\\_1974\\_46408.pdf](https://www.abepro.org.br/biblioteca/TN_ST_401_1974_46408.pdf)>. Acesso em: 10 dezembro. 2023.

MINISTÉRIO DO TRABALHO E EMPREGO. O que é CAGED?. 2016. Disponível em: <<https://pdet.mte.gov.br/o-que-e-caged>>. Acesso em: 2 maio 2024.

MINISTÉRIO DO TRABALHO E EMPREGO. Cadastro Geral de Empregados e Desempregados (CAGED). 2015. Disponível em: <<https://www.gov.br/trabalho-e-emprego/pt-br/servicos/empregador/caged>>. Acesso em: 2 maio 2024.

MORETTIN, Pedro. Análise de Séries Temporais. 2. ed. São Paulo: Blucher, 2006.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. Total de admissões e desligamentos, por setores de atividade de trabalho, no Município do Rio de Janeiro entre janeiro/2000 e dezembro/2023. 2024. Disponível em: <<https://www.data.rio/documents/89ca68d48a844c4da943c9e34492111a/about>>. Acesso em: 29 abr. 2024.

SOUZA, A. et al. UTILIZAÇÃO DO MÉTODO BOX- JENKINS (ARIMA) NA PREVISÃO DE DEMANDAS DE UM PRODUTO DE UMA EMPRESA DE BENEFICIAMENTO DE AÇAÍ. XXXIII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2013. Disponível em: <[https://abepro.org.br/biblioteca/enegep2013\\_TN\\_STO\\_177\\_008\\_22716.pdf](https://abepro.org.br/biblioteca/enegep2013_TN_STO_177_008_22716.pdf)>. Acesso em: 24 abril. 2024.

YAFFEE, Robert. Introduction to Time Series Analysis and Forecasting. San Diego: Academic Press, 2000.