FUNDAÇÃO GETULIO VARGAS

MESTRADO PROFISSIONAL EM ECONOMIA EMPRESARIAL E FINANÇAS

EPGE ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS – FGV EPGE

JOÃO GABRIEL PEREZ TAVARES

Análise e aplicação de modelos preditivos no fluxo de turistas em território brasileiro frente a aspectos econômicos

RIO DE JANEIRO

2024

JOÃO GABRIEL PEREZ TAVARES

Análise e aplicação de modelos preditivos no fluxo de turistas em território brasileiro frente a aspectos econômicos

Dissertação para obtenção do grau de Mestre apresentada à Escola Brasileira de Economia e Finanças

Área de Concentração: Ciência de Dados

Orientador: Rafael Martins de Souza

**FICHA CATALOGRÁFICA**

**FOLHA DE APROVAÇÃO**

**AGRADECIMENTOS**

**RESUMO**

O turismo desempenha um papel fundamental na economia brasileira, impulsionando o crescimento e gerando empregos em diversas regiões do país. Compreender os fatores que influenciam a demanda turística é crucial para o planejamento estratégico do setor e para o desenvolvimento econômico sustentável.

O presente trabalho tem como objetivo analisar a evolução da demanda por turismo, a partir do número de passagens comercializadas no território brasileiro, com foco na sua correlação com seis variáveis econômicas relevantes: tarifa média praticada em passagens aéreas, taxa Selic, IPCA, cotação do dólar, renda média e desemprego. O estudo se baseia em dados coletados ao longo de um período significativo, abrangendo anos anteriores, e utiliza análises estatísticas e métodos econométricos para identificar padrões e relações entre as variáveis mencionadas.

Inicialmente, são apresentados dados históricos da demanda por turismo no Brasil, permitindo uma visão geral do comportamento do setor ao longo do tempo. Após isso, será realizado aplicação de métodos preditivos para buscar uma previsão do fluxo de passageiros com base nas variáveis selecionadas. Visando alcançar esse objetivo, foram utilizados o Modelo Autorregressivo e Integrado de Médias Móveis (ARIMA), o Modelo Autorregressivo e Integrado de Médias Móveis com Sazonalidade(SARIMA), Modelo Vetorial Autorregressivo(VAR) e o *Long Short Term Memory*(LSTM).

**Palavras Chaves:** Turismo, economia, fluxo de passageiros, variáveis macroeconômicas, análise estatística, ARIMA, SARIMA,VAR, LSTM.

**ABSTRACT**

Tourism performs a fundamental role in the Brazilian economy, driving growth and generating employment in various regions of the country. Understanding the factors that influence tourist demand is crucial for the strategic planning of the sector and sustainable economic development.

The present study aims to analyze the evolution of tourism demand based on the number of tickets sold within Brazilian territory, with a focus on its correlation with four relevant economic variables: average airfare, Selic rate, IPCA (Consumer Price Index), and the exchange rate of the dollar, average income, and unemployment. The study relies on data collected over a significant period, covering previous years, and employs statistical analyses and econometric methods to identify patterns and relationships among the mentioned variables.

Initially, historical data on tourism demand in Brazil is presented, providing an overview of the sector's behavior over time. After that, predictive methods will be applied to seek a forecast of passenger flow based on the selected variables. To achieve this goal, the Autoregressive Integrated Moving Average (ARIMA) model, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, the Vector Autoregressive (VAR) model, and the Long Short Term Memory (LSTM) model have been utilized.

**Keywords:** Tourism, economy, passenger flow, macroeconomic variables, statistical analysis, ARIMA, SARIMA, VAR, LSTM.

**LISTA DE TABELAS**

[Tabela 1: Amostra base consolidada dos dados levantados 18](#_Toc163479156)

[Tabela 2: Seleção do número de defasagens do modelo 24](#_Toc163479157)

[Tabela 3: Teste Drubi-Watson 25](#_Toc163479158)

[Tabela 4: Resultados VAR 40](#_Toc163479159)

[Tabela 5: Resultados LSTM 41](#_Toc163479160)

**LISTA DE GRÁFICOS**

[Gráfico 1 - ACF e PACF 22](#_Toc167034069)

[Gráfico 2 - Evolução Demanda Turística 31](#_Toc167034070)

[Gráfico 3 - Evolução com média móvel (12M) da demanda turística 31](#_Toc167034071)

[Gráfico 4 - Gráfico de Linhas da Diferencial - TotalPassageiros 32](#_Toc167034072)

[Gráfico 5 - Histograma da Diferencial - TotalPassageiros 32](#_Toc167034073)

[Gráfico 6 - Evolução Tarifa Média 33](#_Toc167034074)

[Gráfico 7 - Dispersão Total Passageiros x Tarifa Média 34](#_Toc167034075)

[Gráfico 8 - Gráfico de Linhas da Diferencial - Tarifa Média 34](#_Toc167034076)

[Gráfico 9 - Histograma da Diferencial – Tarifa Média 34](#_Toc167034077)

[Gráfico 10 – Evolução Dólar 35](#_Toc167034078)

[Gráfico 11- Dispersão Total Passageiros x Dólar 35](#_Toc167034079)

[Gráfico 12- Gráfico de Linhas da Diferencial - Dólar 36](#_Toc167034080)

[Gráfico 13- Histograma da Diferencial - Dólar 36](#_Toc167034081)

[Gráfico 14- Evolução IPCA 37](#_Toc167034082)

[Gráfico 15- Dispersão Total Passageiros x IPCA 37](#_Toc167034083)

[Gráfico 16- Gráfico de Linhas da Diferencial - IPCA 38](#_Toc167034084)

[Gráfico 17 Histograma da Diferencial - IPCA 38](#_Toc167034085)

[Gráfico 18 - Evolução Taxa de Juros - SELIC 39](#_Toc167034086)

[Gráfico 19- Dispersão Total Passageiros x SELIC 40](#_Toc167034087)

[Gráfico 20- Gráfico de Linhas da Diferencial - SELIC 40](#_Toc167034088)

[Gráfico 21- Histograma da Diferencial - SELIC 40](#_Toc167034089)

[Gráfico 22- Evolução Renda Média 41](#_Toc167034090)

[Gráfico 25 Histograma da Diferencial – Renda Média 42](#_Toc167034091)

[Gráfico 23- Dispersão Total Passageiros x Renda Média 42](#_Toc167034092)

[Gráfico 24- Gráfico de Linhas da Diferencial – Renda Média 42](#_Toc167034093)

[Gráfico 26- Evolução Taxa de Desocupação 43](#_Toc167034094)

[Gráfico 27- Dispersão Total Passageiros x Taxa de Desocupação 43](#_Toc167034095)

[Gráfico 28 Gráfico de Linhas da Diferencial – Taxa de Desocupação 44](#_Toc167034096)

[Gráfico 29 Histograma da Diferencial – Taxa de Desocupação 44](#_Toc167034097)

[Gráfico 30 Matriz de Correlação 45](#_Toc167034098)

[Gráfico 31- Matriz de Correlação - TotalPassageiros 45](#_Toc167034099)

[Gráfico 32-xxx 50](#_Toc167034100)

[Gráfico 33-xxx 51](#_Toc167034101)

[Gráfico 34-xxx 52](#_Toc167034102)

**LISTA DE IMAGENS**

[Figura 1: Ilustração Neurônio LSTM 26](#_Toc163547858)

# SUMÁRIO

[**SUMÁRIO 9**](#_heading=h.ltojpzzemfq2)

[**1. INTRODUÇÃO 10**](#_heading=h.gjdgxs)

[**2. REVISÃO DE LITERATURA 13**](#_heading=h.30j0zll)

[**2.1 Demanda Turística 13**](#_heading=h.1fob9te)

[**2.2 Levantamento de métodos quantitativos para a mensuração e previsão de fluxo turístico 14**](#_heading=h.s8buoqsz5dem)

[**3. METODOLOGIA 18**](#_heading=h.tyjcwt)

[**3.1 Variáveis Consideradas 18**](#_heading=h.by04jdpcpsbc)

[**3.2 Modelos Preditivos Estimados 20**](#_heading=h.cnqbbra1mkut)

[**3.2.1 ARIMA e SARIMA 21**](#_heading=h.zg83e9ey343)

[**3.2.2 VAR 24**](#_heading=h.f31y7gl0kp3p)

[**3.2.3 LTSM 29**](#_heading=h.1jdb1bjry7o9)

[**4. ANÁLISE DOS DADOS LEVANTADOS 31**](#_heading=h.2s8eyo1)

[**4.1 Evolução Demanda Turística 31**](#_heading=h.xymf14x3xhtv)

[**4.2 Tarifa Média 32**](#_heading=h.3rdcrjn)

[**4.3 Dólar 33**](#_heading=h.26in1rg)

[**4.4 IPCA 34**](#_heading=h.lnxbz9)

[**4.5 SELIC 35**](#_heading=h.35nkun2)

[**4.6 Renda média do trabalho 36**](#_heading=h.rid383ddmxlq)

[**4.7 Taxa de desocupação 36**](#_heading=h.n4jth69qn4p5)

[**4.8 Correlações 38**](#_heading=h.3yxgcmyc3k05)

[**4.9 Regressão 40**](#_heading=h.nscqm5r5dy93)

[**5. RESULTADOS 42**](#_heading=h.jccgeg5wze2w)

[**5.1 MODELO ARIMA e SARIMA 42**](#_heading=h.z337ya)

[**5.2 MODELO VAR 43**](#_heading=h.p5u4pr8enn6i)

[**5.3 MODELO LSTM 44**](#_heading=h.acw4yyb81kvs)

[**6. CONSIDERAÇÕES FINAIS 45**](#_heading=h.1y810tw)

[**7. REFERÊNCIAS BIBLIOGRÁFICAS: 47**](#_heading=h.4i7ojhp)

# INTRODUÇÃO

A Organização Mundial do Turismo (2023) define turismo como o fenômeno de aspecto social, cultural e econômico relacionado ao deslocamento de pessoas para lugares fora do seu ambiente pessoal. Com isso, o turismo apresenta uma enorme influência na geração de empregos e movimentação da economia em diversos setores de forma direta e indiretamente. Além disso, com o avanço tecnológico, surgimento de novos meios de comunicação e transporte, o turismo se apresenta como setor cada vez mais importante para a economia mundial e para a sociedade.

O setor de turismo possui um papel de grande importância na atual economia mundial e principalmente na brasileira, devido às suas paisagens naturais e diversidade cultural. No Brasil, o turismo é um importante agente na criação de empregos e forte dinamismo do setor terciário, uma vez que movimenta fortemente o setor de serviços. Evidência disso, segundo o Instituto Brasileiro de Turismo, o setor de turismo representa 3,6% do PIB do país e ainda segundo o Ministério do Turismo (2023), no mês de junho de 2023 a indústria do turismo foi responsável por 10% da geração de empregos com carteira assinada. Além disso, segundo estudo realizado pela Fundação Getúlio Vargas (FGV) aponta que, a cada 1 real investido na promoção do turismo, cerca de 20 reais são inseridos na economia do Brasil por meio do consumo dos visitantes estrangeiros, assim há um retorno de até 2.000% para a economia (Embratur, 2023).

Segundo a Organização Mundial do Turismo, o turismo pode ser visto como o conjunto de serviços que realizam o atendimento ao turista. No cenário brasileiro, de acordo com a Gerência de Informação e Inteligência de Dados da Embratur, o consumo médio de um turista no Brasil movimenta 571 atividades econômicas, sendo 21 serviços diretamente ligados ao turismo, como agências de turismo, hotéis e companhias aéreas. Além disso, há 191 atividades voltadas a moradores locais, como bares, restaurantes, táxis e serviços médicos com atendimentos a turistas. Ainda há 142 atividades atreladas aos setores de fornecimento de bens e serviços que atendem as empresas de turismo, como agricultura familiar, terceirização de serviços de limpeza e contabilidade. O restante não está ligado ao turismo, mas se aquecem quando há mais turistas no país, como mecânicas e construção civil (Embratur, 2023).

Além dos aspectos econômicos, a evolução do turismo, quando feito de forma bem planejada, pode beneficiar a preservação de espaços públicos em cidades, seja através de manutenção de praças e parques ou em manutenção de locais históricos (Barbosa, 2005). Já no contexto sociocultural, o turismo desempenha um papel central na satisfação das necessidades psicológicas e sociológicas dos viajantes, que são motivados por uma variedade de razões, como a busca por quebra de rotina, descanso, lazer, ou simplesmente o desejo de explorar novos lugares e conhecer novas pessoas (Dias e Cassar, 2005).

Entender a motivação e comportamento humano é de suma importância para entender a evolução do fluxo de turistas e aos estudos nessa área de conhecimento. Nesse sentido, fica claro que, em termos práticos, não há turismo sem a presença do turista. Esse, por sua vez, é movido por desejos e vontades em viajar, busca por conhecimento e desfrutar do lazer (Hirata, 2017).

Diante do grande potencial e impacto sociais, o presente estudo se propõe a desenvolver uma análise para entender a relação entre a demanda turística e índices econômicos e sociais. Além disso, a dissertação também visa aplicar modelos preditivos para prever a demanda turística no território brasileiro e auxiliar empresas do setor. A contribuição deste estudo reside em enriquecer os estudos sobre o fluxo de turismo em solo brasileiro, adicionando índices macroeconômicos ao tema e aos modelos preditivos. Destaca-se a novidade da correlação entre o fluxo de turismo com índices macroeconômicos e modelos preditivos univariados e multivariados em voos operados em solo brasileiro, tendo origem e/ou destino no Brasil. Desse modo, é possível identificar quais variáveis macroeconômicas mais afetam o fluxo de turismo.

Os estudos e previsões focados na demanda turística se mostram importantes visto o cenário atual de crescente popularização de viagens e grandes dificuldades enfrentadas por esse setor com a pandemia do Covid19. Nesse contexto, o presente estudo possui grande valor para o setor público e privado na ótica de tomada de decisão, alocação de recursos e planejamento adequado de hoteleiros, operadores de turismo e demais indústrias relacionadas (Chetty, 2011). Fica evidente, portanto, o impacto potencial que o estudo pode gerar na eficiência dos serviços e produtos relacionados ao turismo, aumentando o bem-estar do consumidor final.

# REVISÃO DE LITERATURA

## 2.1 Demanda Turística

As características e definições de demanda turística podem variar de acordo com o objetivo da análise e formação do pesquisador. Desse modo, estudiosos desse tema como formação em economia e sociologia podem oferecer visões e contribuições valiosas distintas para o entendimento da demanda turística (Hirata, 2017).

Diante disso, vale ressaltar a definição de demanda turística de Mathieson e Wall (1988) definida como “número de pessoas que viajam, ou desejam viajar, para utilizar facilidades e serviços turísticos em lugares distantes do seu local de trabalho e residência”. Nessa definição são utilizadas a demanda efetiva, ou seja, pessoas que praticam turismo em determinado período, e a demanda reprimida, aqueles que não viajaram, mas possuem interesse caso ocorra alguma mudança nas condições para realizar a ação (Cooper, 2007).

Já Beni (2007), de maneira mais aderente com o presente estudo, define demanda turística: “Sendo a viagem um movimento entre dois pontos (emissor e receptor), verifica se que a demanda turística é a demanda por esse movimento, e depende tanto das características relativas a esses dois pontos quanto do custo desse movimento”. Nesse cenário, a demanda se dá pelo fluxo de pessoas entre determinada origem e destino, ou seja, é considerada a demanda efetiva.

A definição de Beni (2007) é focada no fluxo de pessoas sem considerar a natureza do deslocamento. No entanto, a identificação do tipo de turismo adotada pelos autores e estudiosos ainda gera grandes debates no que diz respeito ao turismo de negócios. Muitos especialistas no tema não consideram as viagens a trabalho como atividade turística. A OMT, órgão oficial do setor, não considera esse tipo de deslocamento como turismo, pois desconsidera fluxos motivados por razões de cunho profissional (Barreto, 1995).

Nesse contexto, o estudo acerca do comportamento do viajante a fim de entender e prever a demanda por serviços e produtos voltados ao turismo é um campo vasto e complexo. São diversas variáveis como potencial para agregar o tema, desde o âmbito sociológico e psicologia humana até a fatores macroeconômicos pertencentes a países e cidades de origem e destino de turistas.

## 2.2 Levantamento de métodos quantitativos para a mensuração e previsão de fluxo turístico

Diversos trabalhos e artigos utilizam modelos variados para tentar prever preços, demandas, custos em muitos setores econômicos e áreas de pesquisa. No setor de turismo, essa realidade não é diferente, onde há inúmeros estudos e abordagens sobre o tema. Além disso, pode se destacar as aos diversos modelos e abordagens devido a diferentes naturezas, objetivos, base de dados e complexidades, resultando em resultados diversos, mas todos com sua parcela de contribuição.

Segundo Moro e Rita (2022), os artigos relacionados à “previsão de turismo” aumentaram drasticamente entre os anos de 1980 e 2010. Além disso, Song e Li (2008) revisaram 121 estudos publicados de 2000 até 2007 (acredita-se que esse número seja ainda maior). Diante desses levantamentos, ficou evidente que a chegada de turistas foi a medida mais popular para representar a demanda e fluxo do turismo. Sendo essa variável medida pelo total de chegadas de turistas de uma origem para um destino e alguns casos podendo ser classificado de acordo com o objetivo da viagem, como lazer, visita a familiares e/ou amigos ou negócios. Outros estudos utilizaram os gastos dos turistas no destino como variável de estudo (tais como Li et al., 2004, 2006a e 2006b) e outros utilizaram as despesas turísticas em categorias específicas de produtos turísticos, tais como despesas com refeições (Au e Law, 2002), despesas em pontos turísticos (Au e Law, 2000) e compras (Law e Au, 2000). Outras variáveis relacionadas à procura turística utilizadas na literatura incluem as receitas do turismo (Akal, 2004), o emprego no turismo (Witt et al., 2004) e importação e exportação de turismo (Smeral, 2004).

Vale ainda destacar que, segundo Khaidi et al. (2019), existem 14 estudos sobre previsão de demanda turística publicados entre 2010 e 2018 no banco de dados do Google Scholar. O foco dos artigos abrange diversas partes do mundo, como países asiáticos, países europeus e países de regiões africanas. O Reino Unido e os Estados Unidos da América são os países que mais são alvos de estudos desse tema, sendo seis estudos focados nos EUA e quatro no Reino Unido. O artigo ainda apontou que o modelo de série temporal se destaca por sua capacidade de proporcionar maior precisão, mas requer um amplo conjunto de dados históricos e se mostra especialmente eficaz na identificação de sazonalidade nos dados. Enquanto isso, a inteligência artificial demonstra sua utilidade mesmo quando há uma quantidade limitada de dados disponíveis. No entanto, a comparação de desempenho entre os diferentes métodos é desafiadora, uma vez que diferentes modelos são aplicados para analisar variáveis explicativas diversas em estudos distintos. Portanto, a sugestão é considerar a combinação de dois ou até mesmo dos três métodos, quando viável, a fim de otimizar o potencial de cada modelo de previsão.

Khaidi et al. (2019) promovem ainda um rico levantamento das principais variáveis explicativas utilizadas nos estudos. Os principais fatores que afetam a procura por turismo são o PIB, taxa de câmbio, índice de preço ao consumidor, custo de vida, preço médio de produtos e serviço nos locais de origem e destino, além de variáveis climáticas em geral.

Em relação à metodologia, na previsão da procura turística, são normalmente praticados métodos quantitativos, qualitativos ou a combinação de ambos. Os modelos de séries temporais têm sido amplamente utilizados para a previsão da procura turística nas últimas quatro décadas. Esse domínio é evidenciado pelos modelos de médias móveis autoregressivas integradas (ARIMAs) desenvolvidos por Box e Jenkins em 1970. Mais de dois terços dos estudos realizados após o ano 2000 que empregaram técnicas de previsão de séries temporais optaram por utilizar diferentes versões desses modelos ARIMA. A escolha entre modelos ARIMA simples e modelos ARIMA sazonais (SARIMA) depende da frequência das séries temporais. Recentemente, os modelos SARIMA ganharam crescente popularidade, especialmente devido à marcante sazonalidade na indústria do turismo, que atrai o interesse significativo dos tomadores de decisão em relação às variações sazonais na demanda turística (SONG and LI, 2008).

Yang et al. (2015) aplicaram a técnica de média móvel auto-regressiva (ARMA) em uma tentativa de analisar a relação entre os dados de pesquisa no mecanismo de busca e o número de visitantes na província de Hainan, localizada na China. Liang (2014) inovou ao sugerir a fusão entre o modelo sazonal ARIMA (SARIMA) e a técnica de heterocedasticidade condicional autoregressiva generalizada (GARCH) (SARIMA-GARCH) com o objetivo de antecipar o fluxo turístico, com base nas concessões para turistas da China Continental que viajam para Taiwan. A capacidade preditiva dessa combinação, que une um modelo de séries temporais e técnicas de Inteligência Artificial (IA), é comparada a outros modelos, como regressão, suavização exponencial, suavização exponencial Holt-Winter e redes neurais de retropropagação com uma abordagem de algoritmo geral, e se destaca como a mais eficaz. Ao analisar os resultados, é evidenciado que a demanda turística de Taiwan exibe uma sazonalidade em um ciclo e variações ao longo do tempo

No Brasil, destaca-se o trabalho realizado por Casanova (2011), que promoveu um estudo de grande valor ao realizar uma aplicação do modelo preditivo ARIMA com intervenção e ARIMA X-12 para realizar uma previsão da demanda turística para a cidade de Foz do Iguaçu. O artigo utiliza dados coletados pela Secretaria de Estado do Turismo do Paraná referente a chegada de passageiros no aeroporto e na rodoviária. A aplicação demonstrou ser eficaz na projeção da demanda, fornecendo informações para orientar o planejamento de recursos no setor de turismo na região examinada.

Divino e McAleer (2009) conduziram uma análise da demanda de turismo sustentável nos estados do Amazonas e do Pará, empregando a modelagem de Box-Jenkins. Eles examinaram o número de chegadas de turistas, tanto domésticos quanto internacionais, com base em dados mensais e anuais. Suas previsões para os anos de 2006 e 2007 não apresentaram diferenças significantes, indicando que um aumento notável na chegada de turistas internacionais na região amazônica era improvável. Por outro lado, Martins, Werner e Belleza (2013) adotaram uma abordagem que envolveu a combinação de previsões utilizando a variância mínima e a aplicação de ajustes matemáticos para estimar o número de turistas esperados durante a Copa do Mundo (FIFA) em 2014.

Gouveia et al. (2022) também promovem um estudo focado no Brasil ao estudar medir a dependência entre a demanda turística internacional e a taxa de câmbio no Brasil. Desse modo, para alcançar esse objetivo foram utilizados dados mensais referentes à taxa de câmbio e número de chegadas da Argentina, Estados Unidos e Alemanha entre 1999 e 2018. Além disso, foram utilizados modelos ARMA-GARCH para tratar a dependência temporal e resíduos.

# METODOLOGIA

## 3.1 Variáveis Consideradas

Diante da definição e objetivos levantados, faz se necessário a coleta e tratamento de alguns dados econômicos e sociais que servirão como base para o estudo e atingimento das finalidades propostas. Vale ressaltar, que para o presente estudo será considerado a definição de Beni (2007) demanda turística, ou seja, será o fluxo total de passageiros entre uma origem e destino, sem descriminação sobre a natureza do deslocamento.

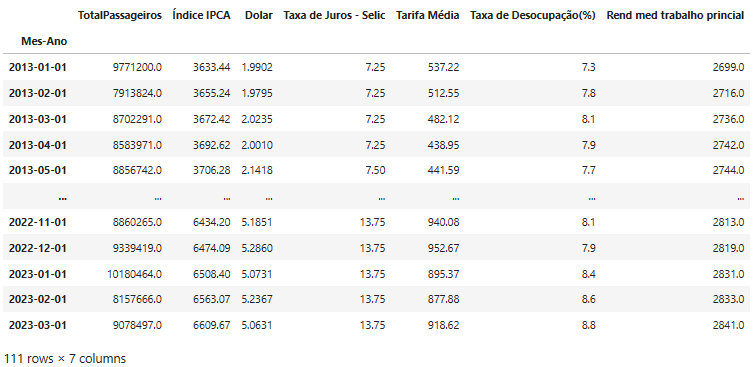
Dados para o estudo:

Variável explicada:

* Assentos Vendidos: Número total de assentos comercializados por mês para voos domésticos no Brasil fornecidos pela Agência Nacional de Aviação Civil (ANAC).

Variáveis explicativas:

* IPCA (Índice Nacional de Preços ao Consumidor Amplo): Foi considerado o índice do IPCA mensal. O IPCA é o índice utilizado para medir a inflação de um grupo de produtos e serviços comercializados no varejo, referentes ao consumo das famílias com rendimentos entre 1 a 40 salários mínimos. Para o presente estudo foram utilizados dados fornecidos pelo IBGE.
* Cotação Dólar em relação ao Real: Foi considerada a cotação do dólar no dia de fechamento de cada mês. Dados retirados do site Investing.com.
* SELIC: Taxa de juros - Selic - fixada pelo Comitê de Política Monetária (COPOM). É a taxa básica de juros da economia, influenciando todas as taxas de juros do país, como as taxas de juros de empréstimos, dos financiamentos e aplicações financeiras. Foram utilizados dados divulgados pelo Banco Central do Brasil, Boletim, Seção mercado financeiro e de capitais (BACEN/Boletim/M.Finan.).
* Tarifa Média: Valor retirado dos dados disponibilizados pela ANAC. As informações são organizadas por data de comercialização e tarifa praticada no momento da compra e número de assentos vendido por tarifa. O estudo em questão realizou um cálculo de média simples para considerar a tarifa média praticada em cada mês e ano.
* Renda média: Foi considerado o rendimento bruto nominal médio efetivamente recebido no mês de referência em todos os trabalhos, efetivamente recebido por mês pelas pessoas de 14 anos ou mais de idade, ocupadas na semana de referência. Foram utilizados dados divulgados pelo Instituto Brasileiro de Geografia e Estatística, Pesquisa Nacional por Amostra de Domicílios Contínua (IBGE/PNAD Contínua)
* Taxa de desocupação: Foi considerado o percentual de pessoas desocupadas com mais de 14 anos em relação às pessoas na força de trabalho na semana de referência. Percentual das pessoas economicamente ativas que estavam procurando trabalho. O presente estudo utilizou os dados fornecidos pelo Instituto Brasileiro de Geografia e Estatística, Pesquisa Nacional por Amostra de Domicílios Contínua (IBGE/PNAD Contínua)

Tabela 1: Amostra base consolidada dos dados levantados

*Fonte: Elaboração própria*

## 3.2 Modelos Preditivos Estimados

Com os dados selecionados formou-se uma tabela com as informações agrupadas com periodicidade mensal, de 2002 até 2021. Desse modo, foi realizado uma análise descritiva de dados através de gráficos, tabelas e histogramas a fim de possuir um maior conhecimento dos elementos. É possível observar os resultados encontrados dessa parte no capítulo 4, onde é realizada uma análise dos dados encontrados. Após isso, foram realizados estudos de correlação e regressões entre a variável explicadas e variáveis explicativas. Depois de realizar uma análise aprofundada dos dados, chegou-se à última etapa do estudo referente à estimação da demanda turística a partir da aplicação de modelos preditivos. Desse modo, foram utilizados o Modelo Autorregressivo Integrado de Médias Móveis (ARIMA) e o Modelo Autorregressivo e Integrado de Médias Móveis com Sazonalidade (SARIMA) para realizar a predição univariada da variável de Assentos Vendidos. Além dos modelos univariados, foram utilizados modelos multivariados como o Modelo Vetorial Autorregressivo (VAR) e o *Long Short Term Memory (LSTM)* utilizando as demais variáveis.

### 3.2.1 ARIMA e SARIMA

O modelo de Média Móvel Integrada Auto-Regressiva (ARIMA) analisa uma série temporal com base em seus valores anteriores, considerando defasagens e erros de previsão passados. É especialmente útil para prever o futuro de séries temporais não estacionárias que apresentam padrões discerníveis, em oposição a serem simplesmente ruídos branco irregulares. O modelo ARIMA é caracterizado por três termos principais, representados pelos parâmetros (*p*, *d*, *q*). Cada um desses termos corresponde às ordens dos termos AR, à diferenciação necessária para tornar a série temporal estacionária e ao termo MA, respectivamente, Takahashi et al.,2000).

O termo "AR" em ARIMA indica que é um modelo de regressão linear que utiliza defasagens para realizar previsões. Modelos de regressão linear proporcionam resultados mais precisos quando não há correlação entre os preditores e sua independência entre si. Uma série temporal é considerada estacionária quando suas propriedades não se alteram ao longo do tempo. Por exemplo, as temperaturas de um mês específico ao longo dos anos formam uma série temporal estacionária, ao passo que as temperaturas de todos os meses ao longo de um ano não são estacionárias, devido à variação sazonal. Para construir um modelo de previsão é necessário que as séries temporais sejam estacionárias. Visando eliminar a não estacionariedade de uma série, normalmente é realizada a diferenciação da série (Sirisha et al, 2020).

Caso uma série temporal seja univariada e apresente elementos de tendência e/ou sazonalidade, a escolha adequada seria o emprego do Modelo ARIMA Sazonal (SARIMA). Se houver a inclusão de um preditor externo, denominado 'variável exógena', no modelo SARIMA, este passa a ser identificado como modelo SARIMAX O uso de uma variável exógena demanda o conhecimento prévio do valor desta variável durante o período de previsão, Deretić et al, 2022).

O SARIMA possui dois tipos de parâmetros, nomeadamente ordem e ordem sazonal (p,d,q) e (P,D,Q,s). Semelhante ao ARIMA, a ordem deste modelo consiste no número de parâmetros de AR, ordem de diferenciação e parâmetros de MA como termos p, d, q. A ordem sazonal consiste no elemento sazonal deste modelo para as unidades AR, diferenças, unidades MA e periodicidade como termos P, D, Q, s. D aqui deve ser o número inteiro que informa sobre a ordem de integração do processo que está sendo executado. P e Q devem ser valores integrais que indicam as ordens das unidades AR, MA que ajudam a incluir todas as defasagens até aquele ponto ou podem ser valores interativos que fornecem defasagens AR/MA específicas que precisam ser incluídas, (Sirisha et al, 2020).

A série temporal utilizada foi a referente ao número total de assentos comercializados por mês para voos domésticos no Brasil fornecidos pela ANAC que está organizada por mês e ano. Os dados para essa série estão disponíveis de janeiro de 2000 até março de 2023.

Visando amenizar os efeitos de outliers, em especial dados impactados pela pandemia do COVID-19, foram testadas duas alternativas. A primeira foi simplesmente utilizar a série temporal até o ano de 2019, desconsiderando dados do ano de 2020 em diante. A segunda opção foi a utilização da técnica *winsorize*. A função *winsorize* é uma técnica estatística utilizada para mitigar o impacto de valores extremos em uma distribuição de dados. Essa abordagem é aplicada para limitar a influência de observações muito altas ou muito baixas, substituindo-as por valores mais próximos. A operação é realizada definindo limites superior e inferior, e os valores que excedem esses limites são substituídos pelos limites em si. Isso ajuda a suavizar a influência de observações extremas sem removê-las completamente, preservando a integridade do conjunto de dados.

Após isso, um teste Augmented Dickey-Fuller (ADF) foi conduzido na série temporal original de assentos comercializados. Os resultados indicaram que a série original é não estacionária. A série temporal foi diferenciada de primeira ordem, buscando conferir estacionariedade à mesma. Um novo teste ADF foi realizado na série temporal diferenciada, revelando estacionariedade.

Funções de autocorrelação (ACF) e autocorrelação parcial (PACF) foram calculadas para a série temporal diferenciada.

As duas linhas pontilhadas nos gráficos de ACF e PACF, gráfico 1, são os intervalos de confiança. Usamos essas linhas para determinar os valores de 'p' e 'q'. Com isso, p é definido como o valor do atraso no qual o gráfico PACF cruza o intervalo de confiança superior pela primeira vez. Assim, p = 1. Utilizando uma lógica parecida, q é definido como o valor do atraso no qual o gráfico ACF cruza o intervalo de confiança superior pela primeira vez, portanto, q = 1.

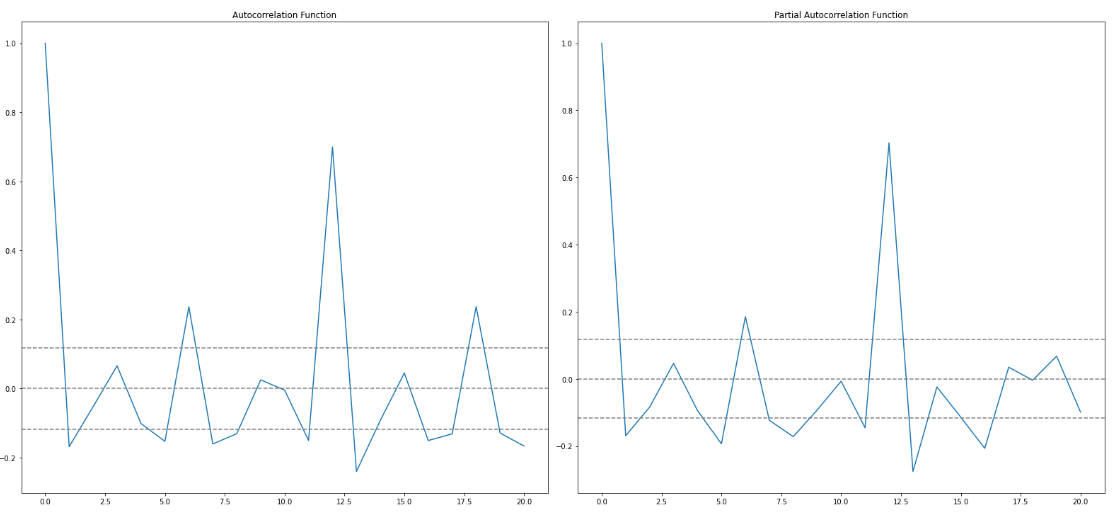


Gráfico 1 - ACF e PACF

### 3.2.2 VAR

O uso dos Modelos Vetoriais Auto Regressivos (VAR) começou a ter um uso mais frequente quando Sims (1980) defendeu o seu uso para análises macroeconômicas como alternativa para os modelos tradicionais da época (Lutkpohl, 2013).

O modelo VAR tem como objetivo estimar e realizar previsões usando séries temporais multivariadas. Os sistemas econômicos em sua grande maioria são expressos através de diversas variáveis. Nesse contexto, os modelos univariados são limitados para alcançar os objetivos do presente estudo.

Sendo uma série temporal multivariada composta por *k* elementos no tempo *t*, representada por , considera se um processo VAR de ordem *p*, VAR(p) o seguinte modelo:

onde ϕ é um vetor de dimensão k,  são matrizes k x k para j = 1, 2, ..., p e é considerado um ruído branco composto por uma sequência de vetores aleatórios independentes e identicamente distribuídos com média 0 e matriz de covariância ∑. Já representa o número de defasagens auto regressivas no sistema.

Um exemplo de VAR(1) com 2 variáveis(n=2):

Onde:

​ e são as duas variáveis no tempo t.

, , ​ e ​ são os coeficientes dos termos defasados.

​ e ​ são os termos de erro.

Algumas condições devem ser respeitadas para realização a estimação dos parâmetros:

* O termo de erro tem uma média condicional zero;
* As variáveis do modelo devem ser estacionárias;
* Grandes valores discrepantes são improváveis;
* Não há multicolinearidade perfeita.

Inicialmente, realizou-se a verificação da estacionaridade de cada série temporal representada pela variável explicada e as variáveis explicativas. Uma série estacionária, também denominada processo estacionário, constitui um conceito fundamental em estatística e análise de séries temporais. Uma série temporal é considerada estacionária quando os dados aleatórios oscilam em torno de um valor fixo. Nesse contexto, uma série estacionária é caracterizada pela constância ao longo do tempo das propriedades estatísticas, como média, variância e autocorrelação. Em outras palavras, a série não apresenta tendência ou sazonalidade e exibe comportamento previsível (Box, Jenkins, & Reinsel, 2016). Para realizar a verificação de estacionariedade foi utilizado o teste Augmented Dickey-Fuller (ADF). O referido teste é de natureza estatística, o que implica que os resultados são expressos em termos de testes de hipóteses, com hipóteses nulas e alternativas. Consequentemente, obtém-se um valor p, a partir do qual é possível realizar inferências sobre a série temporal, seja ela estacionária ou não (Verma, 2021). Em conformidade com o resultado do teste ADF, foi necessário proceder a duas diferenciações nas séries temporais, com o intuito de transformá-las todas em estacionário.

A determinação do comprimento de defasagem em modelos VAR é um aspecto crítico. Sua importância é destacada por Braun e Mittnik (1993), que evidenciam que estimativas de um VAR com um comprimento de defasagem diferente do real resultam em estimativas inconsistentes, afetando também as funções de resposta a impulsos e as decomposições de variância derivadas do modelo estimado. Lütkepohl (1993) aponta que o *overfitting* (selecionar um comprimento de defasagem maior do que o real) aumenta os erros médios quadrados das previsões do VAR, enquanto o *underfitting* do comprimento de defasagem geralmente leva a erros autocorrelacionados. Além disso, Hafer e Sheehan (1989) constatam que a precisão das previsões em modelos VAR varia significativamente dependendo do comprimento de defasagem escolhido.

A maioria dos modelos VAR são estimados com defasagens simétricas, ou seja, o mesmo número de *lags* é aplicado a todas as variáveis em todas as equações do modelo. A escolha desse número de *lags* é geralmente baseada em critérios estatísticos bem definidos, como o Critério de Informação de Akaike (AIC) ou o Critério de Informação Bayesiano Schwartz (BSIC).

*Fonte: Elaboração própria*

Tabela 2: Seleção do número de defasagens do modelo

Texto, Tabela

Descrição gerada automaticamente com confiança médiaDiante desse cenário, foi considerado o *lag* que apresentou o menor AIC, como pode ser verificado na Tabela 2:Seleção do número de defasagens do modelo.

|  |  |
| --- | --- |
| **Variáveis** | **Teste Durbin-Watson** |
| Total Passageiros | 1,76 |
| Índice IPCA | 2,1 |
| Dólar | 1,86 |
| Taxa de Juros (SELIC) | 2,13 |
| Tarifa Média | 1,81 |
| Taxa de Desocupação (%) | 2,04 |
| Renda Média Trabalho | 2,38 |

Tabela 3: Teste Drubi-Watson

*Fonte: Elaboração própria*

Outra verificação realizada foi o Teste de Durbin Watson. O teste de Durbin-Watson possibilita a análise da presença de correlação nos resíduos. A existência de correlação residual indica que há padrões nas séries temporais que ainda não foram captados pelo modelo. Os valores desta estatística estão compreendidos entre 0 e 4. Quando se aproxima de 2, indica a ausência de correlação entre os erros da série. Valores mais próximos de 0 sugerem correlação positiva, enquanto valores mais próximos de 4 indicam correlação negativa. Como pode ser observado na Figura 1, as estatísticas foram próximas a 2, assim é possível considerar que não há correlação entre os erros da série.

Foi realizado o teste causalidade, utilizando o teste Granger. Granger (1969) desenvolveu uma metodologia para analisar as relações causais entre séries temporais. Segundo Granger(1969), dado duas séries estacionárias e , podemos utilizar o modelo para testa se X causa influência em Y. Se os valores passados de x forem preditores significativos do valor atual de y, mesmo quando os valores passados de y foram incluídos no modelo, então x exerce uma influência causal sobre y. Utilizando o modelo é possível investigar a causalidade com base em um Teste F, com a hipótese nula: . Se for rejeitada, pode se concluir que a causalidade de X para Y existe (Lopez and Weber, 2017). Desse modo, aplicando o Teste de Granger nas séries levantadas para o presente estudo, foram encontrados os p-valores apresentados na Tabela 4. Utilizando um nível de significância de 0,05 e observando os p-valores atrelados a variáveis interesses, não é possível rejeitar a hipótese nula de nenhum par de variáveis. O ideal seria que todas os p-valores fossem menores que o nível de significância, pois as series seriam totalmente correlacionadas.

*Fonte: Elaboração própria*



Tabela 4 Resultado teste de Granger

Além disso, o ano de 2020 por se tratar do período mais afetado pela pandemia do Covid 19, sendo assim considerado um *outlier*. Com isso, o modelo foi treinado com base sendo diferenciada duas vezes, com a ordem 12 selecionada e desconsiderando o ano de 2020.

### 3.2.3 LTSM

Dado que as séries temporais exibem uma dependência sequencial entre suas variáveis de entrada, uma abordagem eficaz para análise seria a utilização de redes neurais (NN) capazes de lidar com tais propriedades interdependentes. Nesse contexto, as redes neurais recorrentes (RNN) se destacam como uma escolha ideal. Em particular, a rede de *Long Short Term Memory* (LSTM) surge como uma variante de RNN aplicada em *Deep Learning*.

Sua eficácia reside na capacidade de treinar extensos conjuntos de dados, resultando em altas precisões. O modelo LSTM incorpora um mecanismo de aprendizagem que memoriza e compreende a relação entre variáveis de entrada e saída, identificando dinamicamente o contexto relevante derivado dos dados de entrada. Esse contexto pode ser adaptado de maneira dinâmica de acordo com as necessidades do problema em questão. (Sirisha et al, 2020)

Os modelos *LSTMs* são arquiteturas que compõem o grupo de modelos de RNN. As RNNs são modelos de programação que imitam as atividades de redes neurais do cérebro humano. Desse modo, sua estrutura é composta por unidades de processamento simples que se relacionam através de conexões com e pesos e processam os dados por meio de estímulos externos. (Haykin, 2001)

As redes neurais recorrentes LSTM passaram a ganhar notoriedade quando ficou visível a sua capacidade de aprendizado de longo prazo. O neurônio de uma LSTM possui portas com habilidade de adicionar, descartar ou modificar as informações no tempo, sendo assim capaz de prever melhor os cenários futuros (Nelson et al., 2018). Cada célula de memória é composta por três portas: uma porta de entrada, uma porta de esquecimento e uma porta de saída. Essas portas servem para regular as interações entre a célula de memória e seu ambiente. A porta de entrada controla se o sinal recebido altera o estado da célula de memória ou é bloqueado. A porta de saída permite que o estado da célula de memória influencie outros neurônios ou os evite. Por fim, a porta de esquecimento modula a conexão autorreferente da célula de memória, permitindo que ela lembre ou esqueça seu estado anterior conforme necessário (CARRIER, 2020).

Em síntese, as funções das portas são:

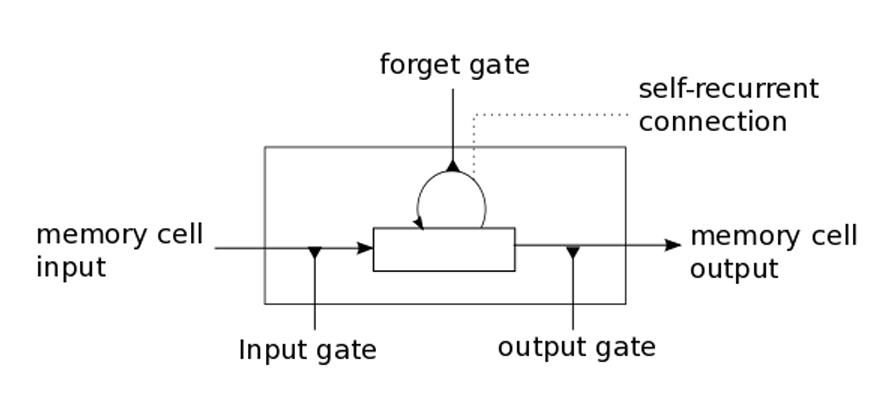
* Porta de entrada: regula a quantidade pela qual uma nova entrada afeta o valor da célula;
* Porta de esquecimento: regula a quantidade pela qual um valor permanece no valor da célula;
* Porta de saída: regula a quantidade pela qual o valor da célula é utilizado na saída.

Figura 1: Ilustração Neurônio LSTM

*Fonte: CARRIER, 2020*

Para a aplicação proposta do presente artigo, primeiramente os dados da base foram normalizados. A normalização dos dados foi realizada através da função MinMaxScaler do pacote sklearn e selecionando a faixa entre 0 e 1.

Após isso, a base de dados foi dividida em conjuntos de treinamento e teste e, em seguida, há a separação dos conjuntos de treinamento e teste em variáveis de entrada e saída.

O modelo LSTM foi definido com 50 neurônios na primeira camada oculta e 1 neurônio na camada de saída para previsão total de passageiros. Foi utilizado a função de perda do erro médio absoluto (MAE) e a versão eficiente de Adam da descida gradiente estocástica. O modelo será adequado para 50 épocas de treinamento com tamanho de lote de 72.

# ANÁLISE DOS DADOS LEVANTADOS

## 4.1 Resumo dos Dados Originais e Diferenciados

Tabela

Descrição gerada automaticamente

Tabela 5 Resumo Dados Originais

*Fonte: Elaboração própria*

Tabela

Descrição gerada automaticamente

Tabela 6 Resumo Dados Diferenciados

*Fonte: Elaboração própria*

## 4.2 Evolução Demanda Turística

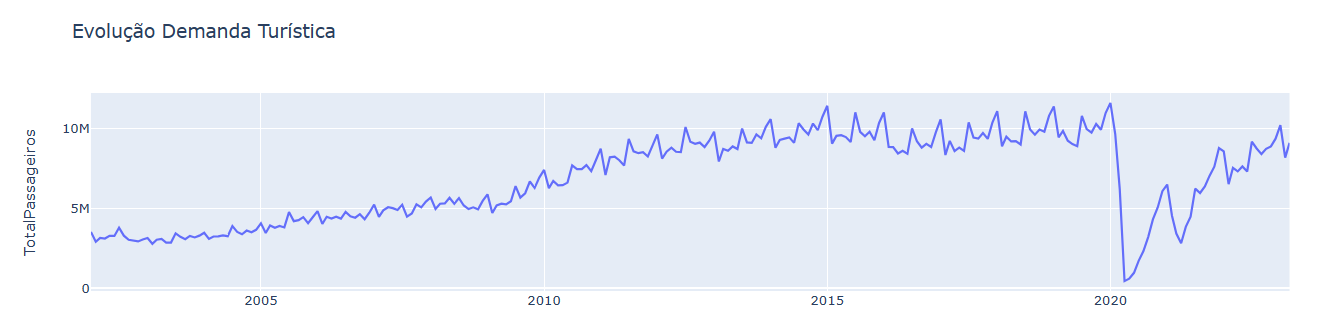
Base retirada do site da ANAC contendo o total de passageiros que ocuparam assentos comercializados ao público e que geraram receita para a empresa de transporte aéreo.

Para o estudo proposto foi considerado esse dado como indicador para representa a demanda turística.

A partir dos gráficos 3 e 4 podemos observar bem definida a sazonalidade esperada da demanda por turismo. Os picos são referentes aos meses de dezembro, janeiro e julho, nos quais há uma demanda maior por passagem aéreas devido a estação do ano e férias escolares. Vale ressaltar, também, a mínima história durante a pandemia do Covid-19.

Conforme indicado no Capítulo 3, referente a metodologia dos modelos preditivos, foi necessário aplicar duas diferenciais a series de dados para torná-las estacionarias. Desse modo, é possível observar a o comportamento da série duplamente diferenciada nos gráficos 5 e 6.

O gráfico de linhas da série temporal diferenciada duas vezes exibe as mudanças na taxa de variação da série original ao longo do tempo. Diferenciar a série duas vezes geralmente ajuda a tornar a série mais estacionária, removendo tendências e volatilidade de longo prazo. A linha traçada reflete como a segunda diferença varia ao longo do tempo, revelando padrões de curto prazo e possíveis sazonalidades mais claramente. No gráfico é possível ver as sazonalidades referentes aos meses de dezembro, janeiro e julho com picos mais acentuados. O histograma da série temporal diferenciada duas vezes ilustra a distribuição dos valores diferenciados. No histograma, o eixo x representa os intervalos de valores diferenciados, enquanto o eixo y indica a frequência ou a contagem de ocorrências desses valores dentro de cada intervalo. Diferenciar a série duas vezes geralmente resulta em uma distribuição mais centrada em torno de zero, com a maioria dos valores concentrados próximos ao centro e uma menor dispersão, indicando que a série está mais estabilizada em termos de variabilidade.



*Fonte: Elaboração própria*

Gráfico 2 - Evolução Demanda Turística

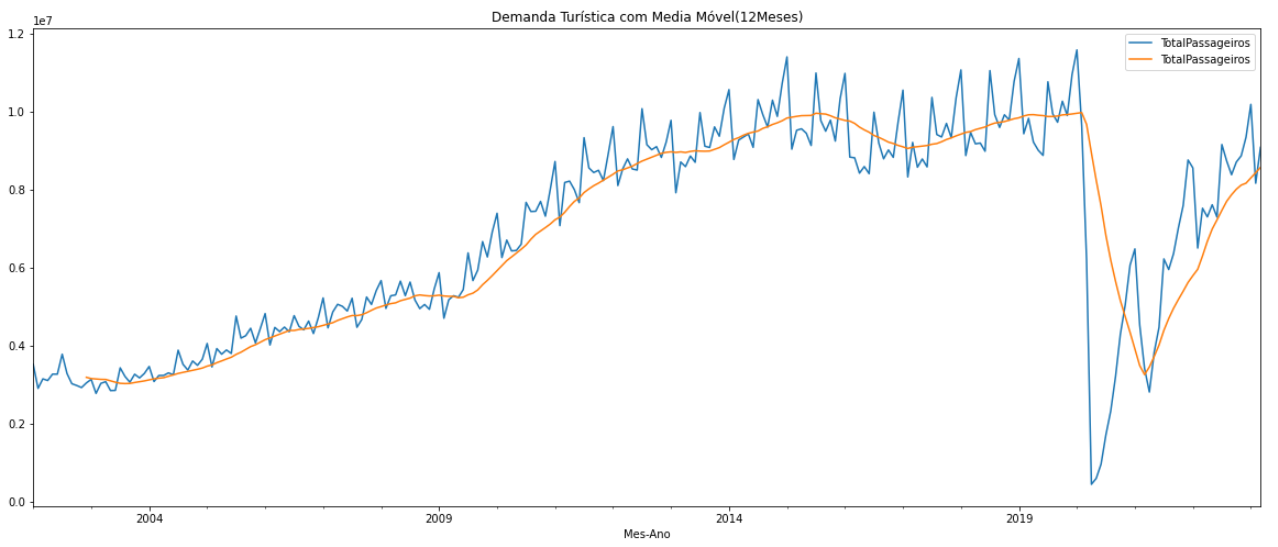
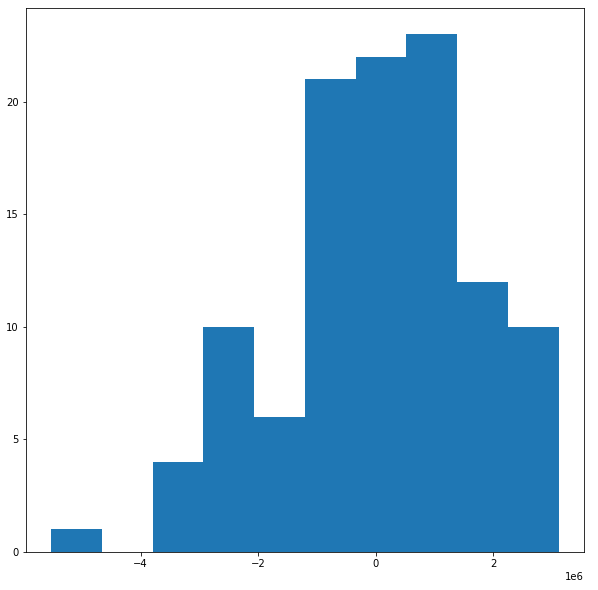


Gráfico 3 - Evolução com média móvel (12M) da demanda turística

*Fonte: Elaboração própria*

**Uma imagem contendo Gráfico de linhas

Descrição gerada automaticamente**

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

Gráfico 4 - Gráfico de Linhas da Diferencial - TotalPassageiros

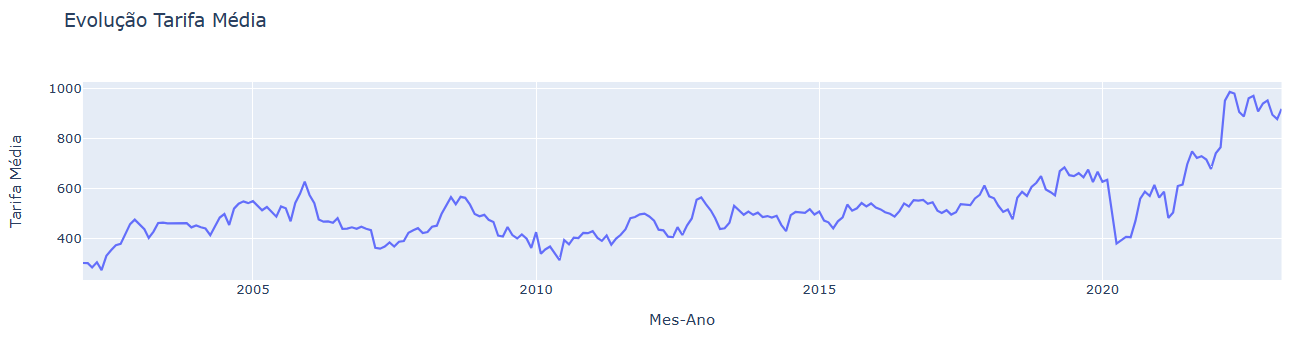
Gráfico 5 - Histograma da Diferencial - TotalPassageiros

## 4.3 Tarifa Média

Informações de tarifas aéreas domésticas fornecidas pela ANAC, foram considerados apenas voos com origem e destino em território brasileiro. São consideradas as tarifas efetivamente vendidas ao público, sendo consideradas com as datas em que foram comercializadas.

Para o estudo os dados foram agrupados por mês, considerando a tarifa média nominal praticada em todos os assentos comercializados por período.

A partir do gráfico de dispersão é possível observar uma grande concentração do preço médio na faixa de 400 a 600 reais e com uma variação mais acentuada em relação à demanda turística. Vale ressaltar, também, as amostras mais recentes mostrando dados de tarifa média e demanda em níveis elevados.

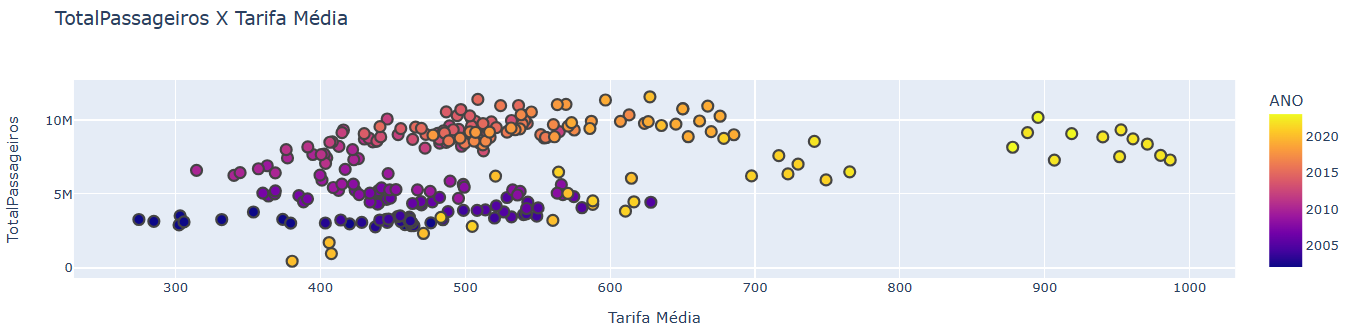
O gráfico de linhas da segunda diferença das tarifas aéreas exibe as mudanças na taxa de variação diária dos preços médios das passagens. Este gráfico é mais volátil, com flutuações mais acentuadas em torno de zero. As oscilações frequentes indicam variações rápidas e intensas nos preços, refletindo uma maior sensibilidade do mercado a alterações de curto prazo. O histograma da segunda diferença das tarifas aéreas apresenta uma distribuição mais ampla e volátil, centrada em torno de zero. A maioria dos valores está próxima de zero, mas com flutuações maiores, indicando que a série temporal apresenta uma maior volatilidade residual após a segunda diferenciação. Esta maior volatilidade pode ser atribuída à sensibilidade do mercado a mudanças rápidas na demanda, promoções e outros fatores econômicos.

*Fonte: Elaboração própria*

Gráfico 6 - Evolução Tarifa Média

Gráfico, Histograma

Descrição gerada automaticamenteGráfico, Gráfico de linhas

Descrição gerada automaticamente 

*Fonte: Elaboração própria*

Gráfico 9 - Histograma da Diferencial – Tarifa Média

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

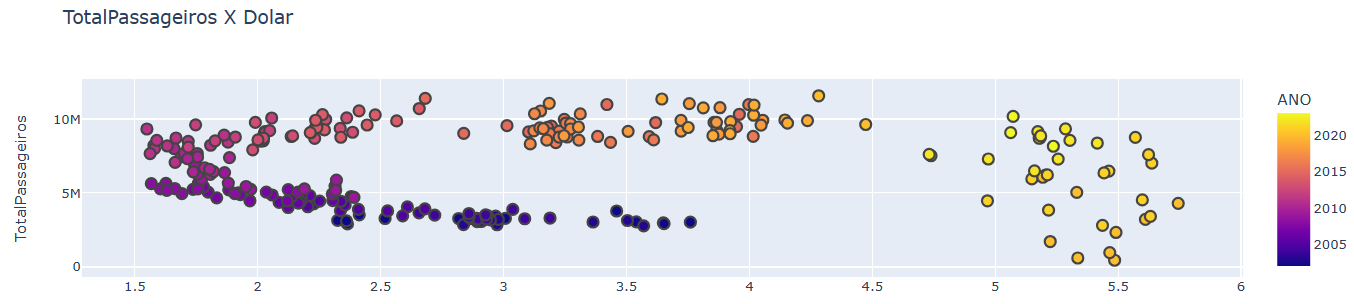
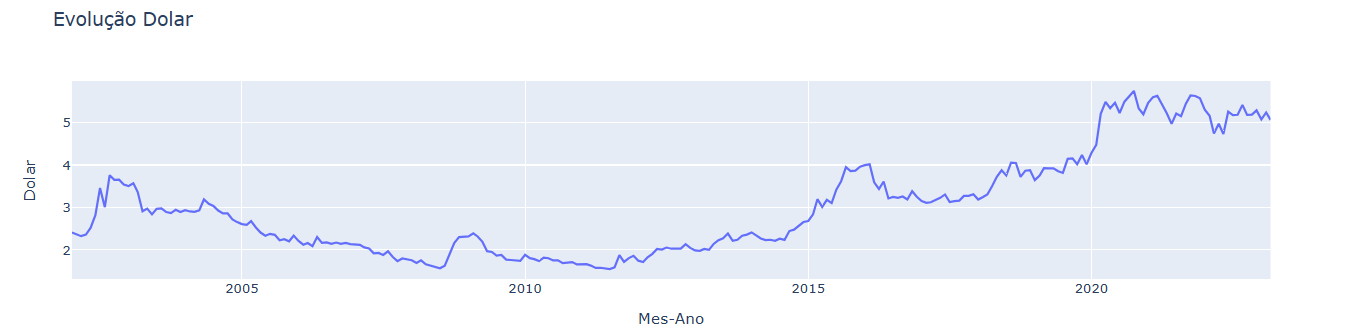
Gráfico 7 - Dispersão Total Passageiros x Tarifa Média

Gráfico 8 - Gráfico de Linhas da Diferencial - Tarifa Média

## 4.4 Dólar

Em relação aos dados encontrados para a cotação do dólar, o gráfico de dispersão mostrou 4 concentrações distintas de casos. Além disso, é interessante notar que os dados mais recentes apresentam o nível de cotação do dólar mais elevado dentro do período analisado. O gráfico de dispersão apresentou dados mais dispersos de total de passageiros em um patamar de dólar mais elevado.

O gráfico de linhas e histograma da diferença da cotação do dólar mostram a distribuição das variações diárias da taxa de câmbio. A maioria dos valores se concentra em torno de zero, refletindo a volatilidade diária da cotação, apesar de possuir pico superiores e inferiores. Este histograma ajuda a visualizar a frequência e magnitude das variações diárias da cotação do dólar.



*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

Gráfico 10 – Evolução Dólar

Gráfico 11- Dispersão Total Passageiros x Dólar

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

*Fonte: Elaboração própria*

Gráfico 12- Gráfico de Linhas da Diferencial - Dólar

Gráfico, Histograma

Descrição gerada automaticamente

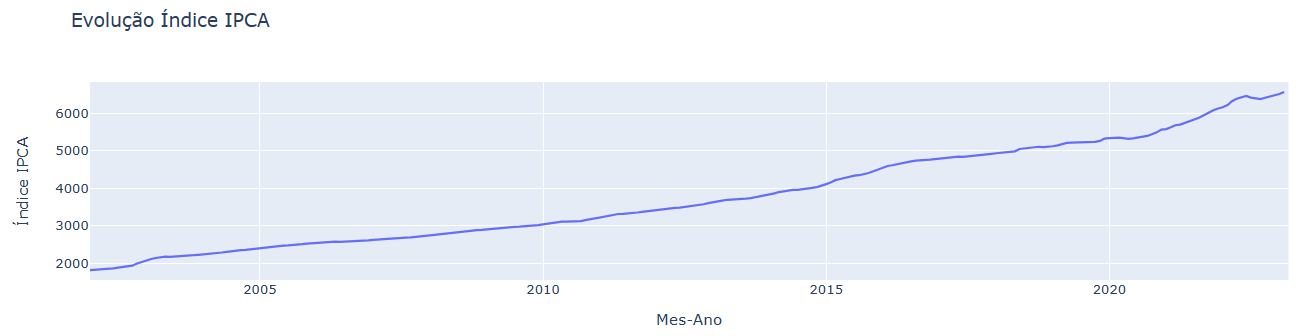
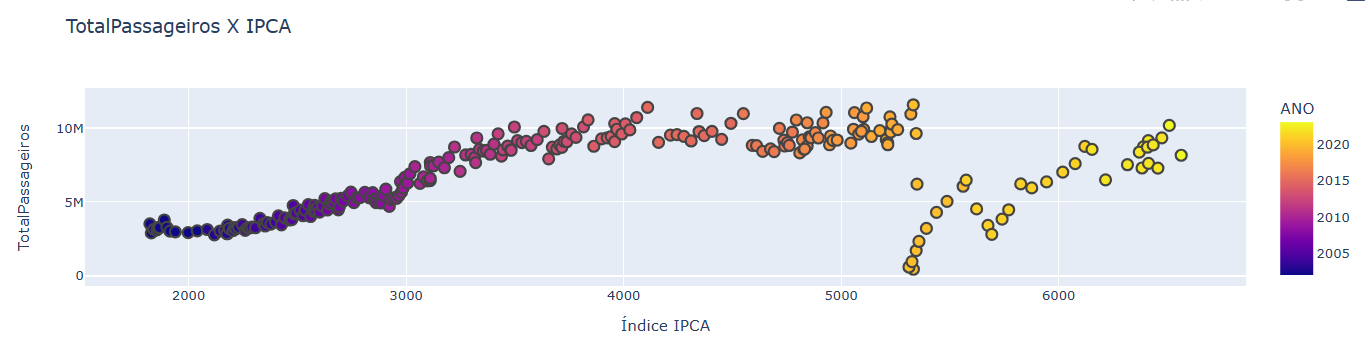
*Fonte: Elaboração própria*

Gráfico 13- Histograma da Diferencial - Dólar

## 4.5 IPCA

Os gráficos 8 e 9 apresentaram os resultados mais uniformes em relação às demais variáveis. Desse modo, observamos uma correlação clara entre o total de passageiros e o índice IPCA.

Em relação ao gráfico e histograma da diferencial, é possível observar dados mais bem comportados, com dados próximos ao nível 0. Isso indica que a série está mais estabilizada em termos de variabilidade após a segunda diferenciação. Tanto o gráfico de linhas quanto o histograma da segunda diferença destacam a volatilidade de curto prazo na inflação, facilitando a análise das mudanças rápidas na taxa de variação.



*Fonte: Elaboração própria*

Gráfico 14- Evolução IPCA

*Fonte: Elaboração própria*

Gráfico 15- Dispersão Total Passageiros x IPCA

**Gráfico, Gráfico de linhas

Descrição gerada automaticamente**

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

Gráfico 16- Gráfico de Linhas da Diferencial - IPCA

**Gráfico, Histograma

Descrição gerada automaticamente**

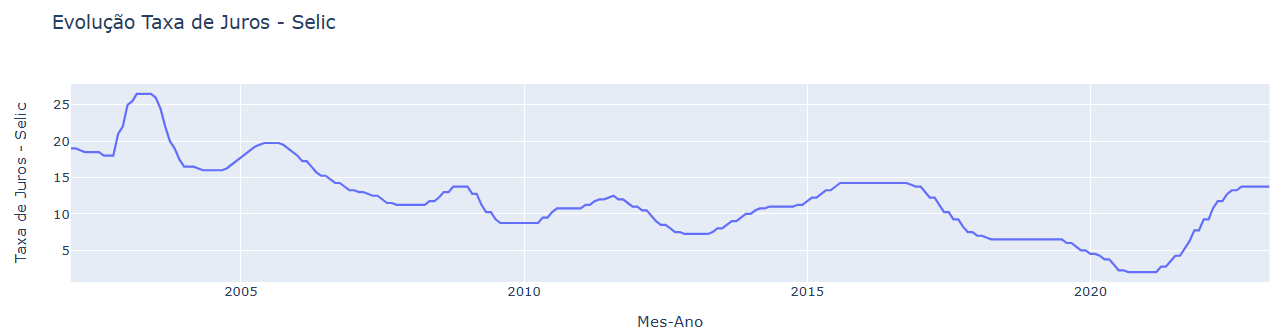
Gráfico 17 Histograma da Diferencial - IPCA

## 4.6 SELIC

Como a taxa SELIC é determinada pelas reuniões do COPOM realizadas periodicamente, observamos taxas constantes em determinado período do gráfico de Evolução e de dispersão. No gráfico de dispersão é possível observar que a uma taxa SELIC menor (entre 5% e 14%) obtivemos número de passageiros maiores, na faixa de 10 milhões. Já quando observamos casos em que a taxa SELIC estava mais elevada (entre 15% e 25%), percebemos que o total de passageiros permaneceu em níveis mais baixos, próximo aos 5 milhões de passageiros.

O gráfico de linhas da série temporal da taxa de juros SELIC, diferenciada duas vezes, exibe as mudanças na taxa de variação da SELIC ao longo do tempo. O gráfico da segunda diferença é mais centrado em torno de zero, destacando as acelerações e desacelerações na mudança da taxa SELIC. A maioria dos valores estará mais próxima de zero, com menores flutuações, mostrando a estabilidade relativa após a segunda diferenciação. Observando os gráficos diferenciados, economistas e analistas podem identificar períodos de políticas monetárias mais agressivas ou conservadoras e entender melhor as respostas do mercado às mudanças na taxa SELIC.

Estas visualizações e suas descrições ajudam a capturar a essência da dinâmica da taxa SELIC e sua transformação ao longo do tempo, fornecendo uma base sólida para análise econômica e decisões políticas.

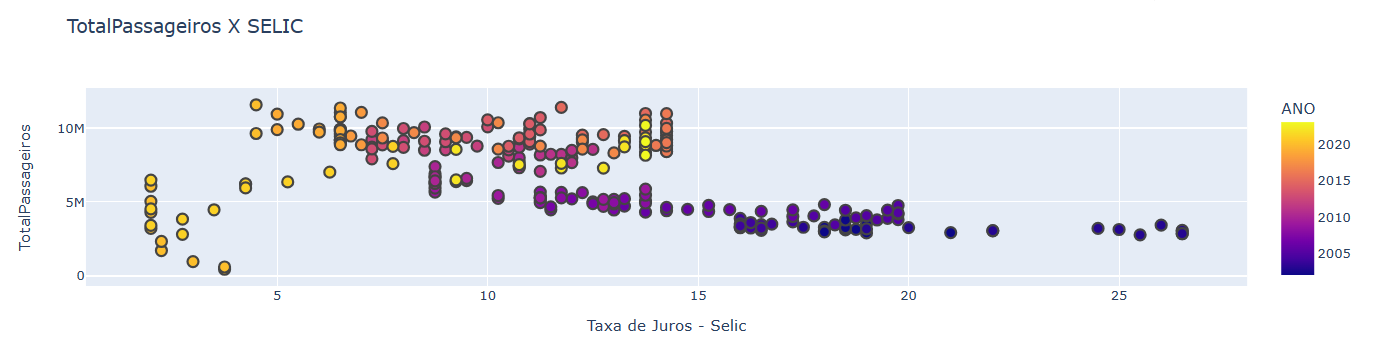


*Fonte: Elaboração própria*

Gráfico 18 - Evolução Taxa de Juros - SELIC

**Gráfico, Histograma

Descrição gerada automaticamenteDiagrama

Descrição gerada automaticamente**

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

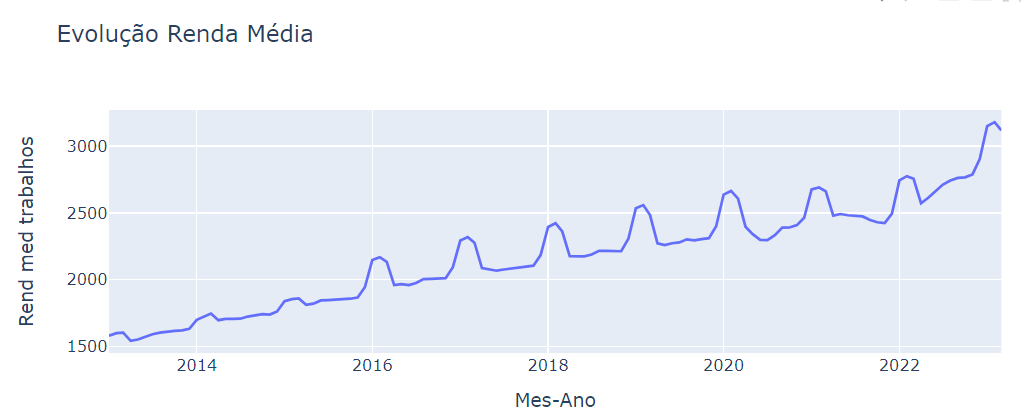
Gráfico 19- Dispersão Total Passageiros x SELIC

Gráfico 20- Gráfico de Linhas da Diferencial - SELIC

Gráfico 21- Histograma da Diferencial - SELIC

## 4.7 Renda média do trabalho

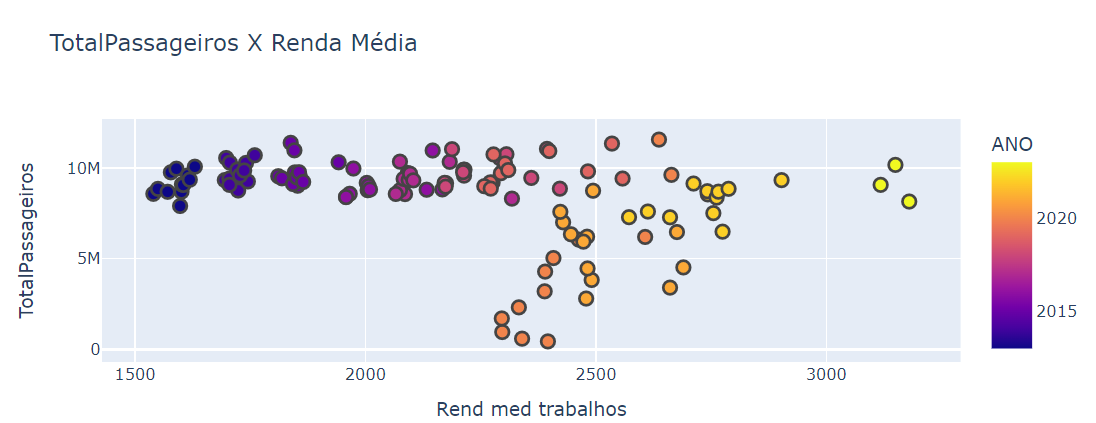
No Gráfico 12, referente à evolução da renda média, podemos observar que a renda apresentou um leve aumento ao longo do tempo, com picos nos meses de janeiro. Durante a pandemia do Covid-19 foi possível observar uma leve queda na renda, com uma retomada acentuada a partir do ano de 2022. Em relação ao Gráfico 13, observamos uma concentração de casos na faixa de 10 milhões no eixo X, sem concentração bem definida no eixo Y. Os casos fora da faixa próxima à linha de 10 milhões no eixo X são referentes ao período da pandemia do Covid19.

O gráfico de linhas da segunda diferença da renda média da população exibe as mudanças na taxa de variação periódica da renda média. Este gráfico é mais volátil, com flutuações acentuadas em torno de zero. As oscilações frequentes indicam variações rápidas e intensas na renda média, refletindo uma maior sensibilidade a mudanças de curto prazo. O histograma da segunda diferença da renda média apresenta uma distribuição mais ampla e volátil, centrada em torno de zero. A maioria dos valores está próxima de zero, mas com flutuações maiores, indicando que a série temporal apresenta uma maior volatilidade residual após a segunda diferenciação. Esta maior volatilidade pode ser atribuída à sensibilidade da renda média a mudanças rápidas em fatores econômicos e sociais.

*Fonte: Elaboração própria*

Gráfico 22- Evolução Renda Média

Diagrama

Descrição gerada automaticamente com confiança médiaGráfico, Histograma

Descrição gerada automaticamente

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

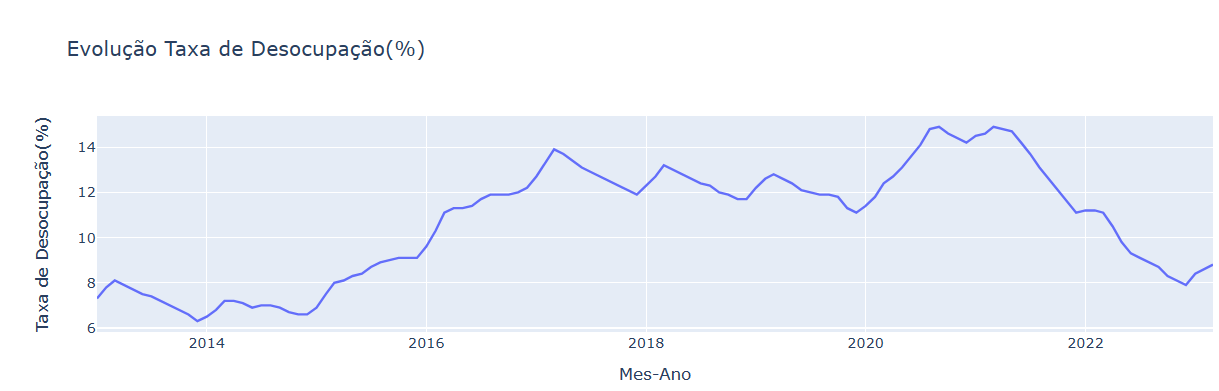
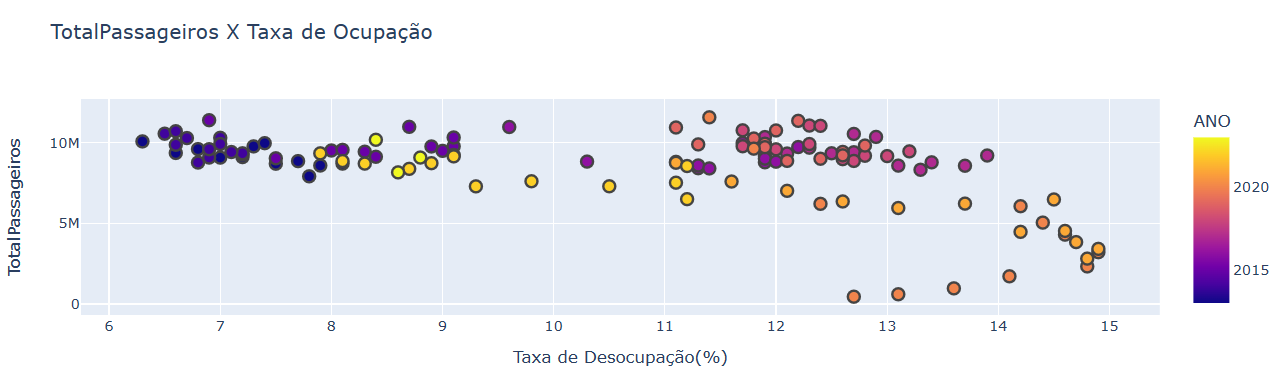
Gráfico 25 Histograma da Diferencial – Renda Média

Gráfico 23- Dispersão Total Passageiros x Renda Média

Gráfico 24- Gráfico de Linhas da Diferencial – Renda Média

## 4.8 Taxa de desocupação

Em relação à evolução da taxa de desemprego, podemos destacar o grande crescimento da taxa entre os anos de 2015 até 2019 e o pico da taxa por volta do ano de 2021. Enquanto isso no gráfico de dispersão não é possível observar uma correlação clara entre os dados de desocupação e número de passageiros.

O gráfico de linhas da segunda diferença da taxa de desemprego exibe as mudanças na taxa de variação periódica da taxa de desemprego. Este gráfico é mais bem comportado, com flutuações suaves e centradas em torno de zero. A série temporal apresenta pouca variação, indicando estabilidade após a segunda diferenciação. Tanto o gráfico de linhas quanto o histograma da segunda diferença destacam uma série temporal mais controlada e previsível. Isso facilita a análise das tendências de curto prazo no mercado de trabalho e na taxa de desemprego.

*Fonte: Elaboração própria*

*Fonte: Elaboração própria*

Gráfico 26- Evolução Taxa de Desocupação

Gráfico 27- Dispersão Total Passageiros x Taxa de Desocupação

**Gráfico, Gráfico de linhas

Descrição gerada automaticamente**

*Fonte: Elaboração própria*

Gráfico 28 Gráfico de Linhas da Diferencial – Taxa de Desocupação

*Fonte: Elaboração própria*

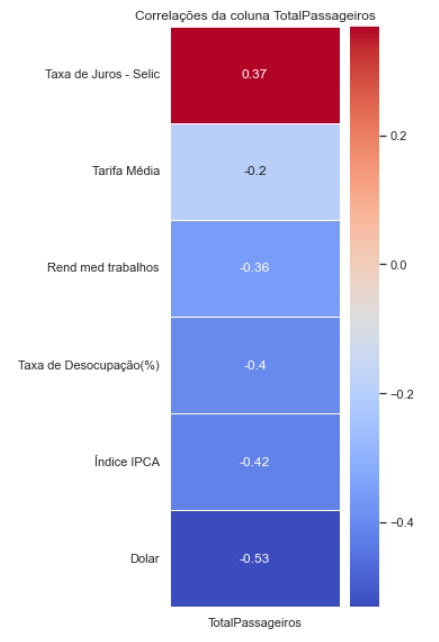
**Gráfico, Histograma

Descrição gerada automaticamente**

Gráfico 29 Histograma da Diferencial – Taxa de Desocupação

## 4.9 Correlações

Gráfico 30 Matriz de Correlação



*Fonte: Elaboração própria*

Gráfico 31- Matriz de Correlação - TotalPassageiros

*Fonte: Elaboração própria*

**Taxa de Juros - Selic (0,37):** A correlação de 0,37 com a Taxa de Juros - Selic indica uma correlação positiva moderada. Isso sugere que um aumento na Taxa de Juros - Selic está associado a um aumento moderado no número de passageiros pagos. Portanto, o aumento nas taxas de juros está positivamente relacionado ao aumento da demanda turística.

**Renda Média do Trabalho (-0,36):** A correlação negativa de -0,36 com a Renda Média sugere que um aumento na renda média está associado a uma redução moderada no número de passageiros pagos. A influência da renda no turismo é negativa, mas a correlação é relativamente moderada, indicando que a renda tem um impacto significativo na demanda turística.

**Tarifa Média (-0,2):** A correlação negativa de -0,2 com a Tarifa Média a ser negativa, indicando que um aumento na tarifa média está associado a uma redução leve no número de passageiros pagos. A influência das tarifas médias no fluxo turístico é negativa.

**Taxa de Desocupação (%) (-0,4):** A correlação de -0,4 com a Taxa de Desocupação indica uma correlação negativa moderada. Isso sugere que um aumento na taxa de desocupação está associado a uma redução moderada no número de passageiros pagos. Portanto, um mercado de trabalho mais fraco (representado por uma taxa de desocupação mais alta) pode levar a uma redução moderada na demanda turística.

**Índice IPCA (-0,42):** A correlação negativa de -0,42 com o Índice IPCA indica uma correlação negativa moderada. Isso sugere que um aumento no Índice IPCA (representando uma maior inflação) está associado a uma redução moderada no número de passageiros pagos. Portanto, um aumento na inflação pode levar a uma redução na demanda turística.

**Dólar (-0,53):** A correlação de -0,53 com o dólar indica uma correlação negativa moderada mais forte. Isso sugere que um aumento no valor do dólar está associado a uma redução moderada no número de passageiros pagos. O impacto do câmbio (representado pelo dólar) no fluxo de passageiros é mais acentuado, com uma correlação negativa mais forte.

Com base nos novos dados, a análise revela que a Taxa de Juros - Selic, o Índice IPCA, o Dólar, Taxa de Desocupação e Renda Média têm correlações mais fortes com a demanda turística, enquanto a Tarifa Média apresentou correlação mais fraca. A relação entre as variáveis e a demanda turística pode ser usada para tomar decisões e prever o comportamento do setor de turismo com mais precisão.

## 4.10 Teste ADF

Conforme mencionado no capítulo 3, para verificar a estacionariedade das séries temporais foi utilizado o teste ADF. Desse modo, foi utilizado o nível de significância de 0,05 para rejeita a hipótese nula de que as séries são não estacionarias, ou seja, se o p-valor encontrado for menor do que 0,05 podemos rejeitar a hipótese nula e dizer a que a série é estacionaria.

Analisando a tabela 5, verificamos que todas as series demostraram comportamento não estacionário em seu estado original. Após a 1ª diferenciação, dois p-valores encontrados ainda permaneceram maior que o nível de significância. Após a 2 diferenciação, todos os p-valores encontrados foram abaixo do nível de significância, demonstrando assim que todas as séries agora se tornaram estacionários.



Tabela 7 Teste ADF - P-valores

*Fonte: Elaboração própria*

# RESULTADOS

## 5.1 Regressão

Visando complementar a análise foi aplicado uma regressão linear simples a fim de representar os coeficientes da regressão para cada uma das variáveis independentes (preditoras) que foram incluídas no modelo, incluindo o termo de intercepto.

A regressão linear é uma técnica estatística utilizada para modelar a relação entre uma variável dependente (ou variável de resposta) e uma ou mais variáveis independentes (ou variáveis preditoras). O objetivo principal da regressão linear é estabelecer uma equação linear que melhor descreva a relação entre essas variáveis, permitindo realizar previsões ou inferências sobre a variável dependente com base nas variáveis independentes.

A equação da regressão linear para o presente estudo é dada por:

onde:

​ é o coeficiente do intercepto (ou constante).

, ​, , ​, , ​ são os coeficientes das variáveis independentes (Índice IPCA, Dólar, Taxa de Juros - Selic e Tarifa Média, Renda Média do Trab. Principal e Taxa de Desocupação).

Os valores encontrados para os coeficientes na regressão são:

​ (Intercepto): -4.958118e+06

(Índice IPCA): -3.301504e+02

​(Dólar): -1.035609e+06

(Taxa de Juros - Selic): 1.273191e+05

(Tarifa Média): 5.602617e+03

(Renda Média do Trab. Principal): 4.997586e+03

(Taxa de Desocupação): 8.430587e+04

Esses coeficientes representam o efeito das variáveis independentes no valor previsto pela regressão. Por exemplo, para cada unidade de aumento no Índice IPCA, o valor previsto (número total de assentos vendidos por companhia aérea) diminui em 330.1504 unidades, mantendo todas as outras variáveis constantes. As interpretações dos coeficientes fornecem informações valiosas sobre como as variáveis independentes afetam a demanda no setor de turismo com base no modelo de regressão.

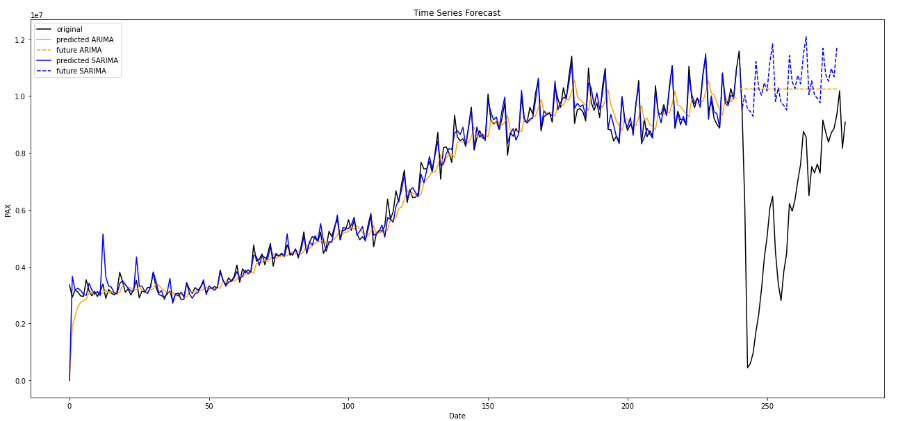
## 5.2 MODELO ARIMA e SARIMA

Com os parâmetros definidos no capítulo de metodologia, o Erro Médio Quadrático (RMSE) encontrados para a base original (período de janeiro de 2000 até Março de 2023), base tratada com Winsorize e base considerando o período até o ano 2000, podem ser observados na Tabela 4. Desse modo, é possível observar que ao desconsiderarmos o período mais afetado pela pandemia do COVID-19 foi possível obter melhores resultados, ou seja, menor diferença entre o valor real e valor previsto pelo modelo. Texto

Descrição gerada automaticamente com confiança média

*Fonte: Elaboração própria*

Tabela 8 Resultados ARIMA/SARIMA

No Gráfico 17 foi utilizado a base com dados até 2020, sendo possível observar a evolução real (linha preta) e a evolução prevista pelos modelos ARIMA (linha azul) e SARIMA (linha laranja), bem como os dados posteriores a 2020 para ambos os modelos. Foi utilizado a base que apresentou o melhor nível de RMSE.

*Fonte: Elaboração própria*

Gráfico 32- Real vs Previsto ARIMA/SARIMA

## 5.3 MODELO VAR

Definidos os parâmetros para o modelo VAR, foi possível realizar a previsão do fluxo de passageiros e a comparação com os dados originais, como pode ser visto no Gráfico 18. Diante desse cenário, foi possível observar que a inserção de novas variáveis ao problema não trouxe melhores resultados a modelagem proposta, como é possível perceber de forma visual no Gráfico 18 e no aumento do RMSE na Tabela 5.

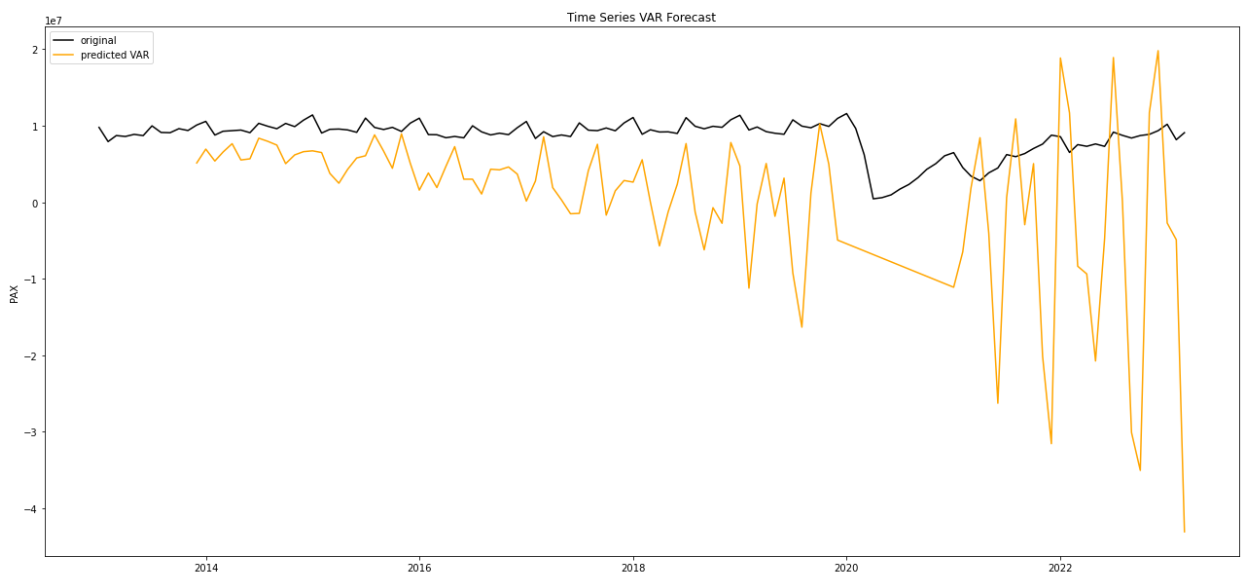
|  |  |  |
| --- | --- | --- |
| **Método** |  | **Resultados** |
| MAPE |  | 0,4249 |
| MAE |  | 2.988.625,20 |
| RMSE |  | 3.613.002,79 |

Tabela 9: Resultados VAR

*Fonte: Elaboração própria*

Gráfico 33- Real vs Previsto VAR

*Fonte: Elaboração própria*



## 5.3 MODELO LSTM

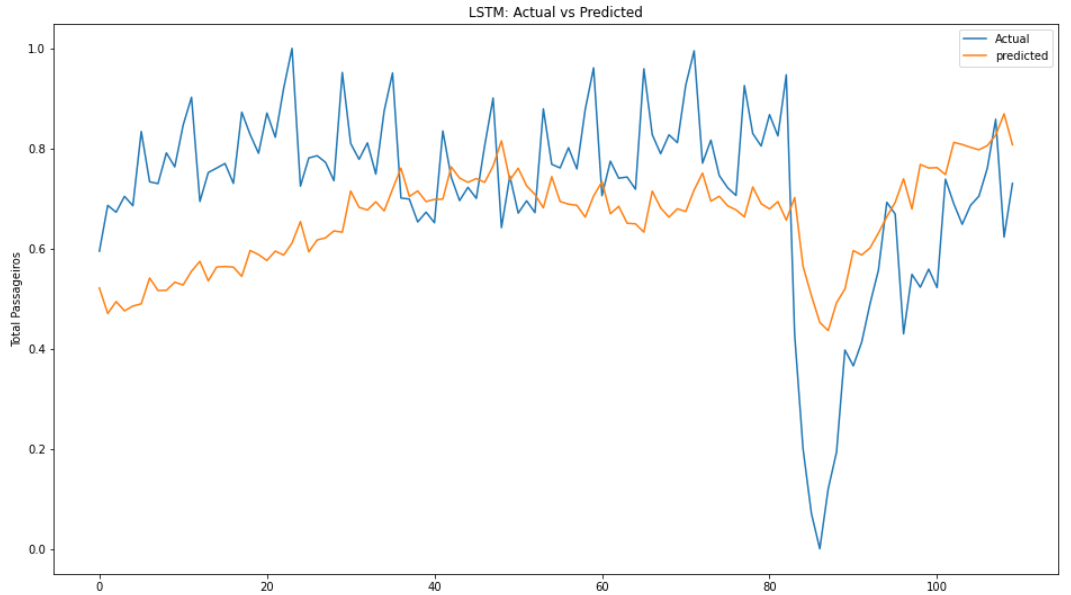
Diante dos parâmetros definidos, o modelo LSTM foi aplicado. No Gráfico 19 é possível observar o comparativo entre a evolução real e a prevista, ambos estão apresentados na sua forma normalizada devido aos requisitos do modelo. Nesse contexto, é possível destacar que o modelo de redes neurais apresentou resultados mais harmônicos do que o modelo VAR.

Gráfico 34- Real vs Previsto LSTM

|  |  |
| --- | --- |
| **Método** | **Resultados** |
| MAPE | 0,03 |
| MAE | 0,15 |
| RMSE | 0,18 |

Tabela 10: Resultados LSTM

*Fonte: Elaboração própria*

# CONSIDERAÇÕES FINAIS

O contexto atual do turismo e diversos setores, repleto de dados e capacidade de processamento, é um cenário propício para análises explorativas de dados e aplicação de modelos preditivos. No presente estudo foi utilizado como base principal de movimentação aeroportuária fornecida pela ANAC, onde foi possível observar o fluxo de passageiros em voos operados em aeroportos em território brasileiro. Além disso, outras variáveis foram consultadas com o objetivo de dar maior robustez ao estudo e obter melhores resultados. Desse modo, foi possível observar que no período observado o fluxo de passageiros foi mais influenciado e correlacionado com a cotação do dólar e o índice de preço IPCA. Esses dois fatores estão diretamente relacionados aos preços dos serviços atrelados ao turismo e à renda da população, nas quais segundo Rebahy(2019) possuem uma relação elástica em relação ao fluxo turístico. Além disso, podemos destacar os efeitos da pandemia da covid 19 que apresentou uma queda muito acentuada nos fluxo de passageiros no ano 2020. Esse período foi tratado como *outlier* e quando foi retirado das bases apresentou melhorias dos resultados dos modelos.

Após uma profunda análise exploratória dos dados coletados, foram aplicados alguns modelos preditivos, sendo eles univariados e multivariados. Os resultados obtidos mostram o modelo SARIMA com os melhores resultados entre a classe dos univariados. Já na classe dos modelos multivariados o modelo LSTM, que utiliza uma arquitetura de redes neurais, apresentou uma melhor performance, no entanto o resultado ainda foi inferior ao modelo SARIMA. Adicionalmente, de acordo com o estudo de Cerqueira, Torgo e Soares (2019), observa-se uma melhoria no desempenho preditivo dos modelos de aprendizado de máquina quando a amostra é significativamente extensa. No entanto, esse não foi o cenário abordado neste trabalho, uma vez que ele se baseou em apenas 111 amostras, representadas pelos meses de cada ano.

Vale ressaltar que a inclusão de novas variáveis ao problema não resultou em uma melhora de resultado. A introdução de novas métricas e índices, no entanto, trouxe aspectos interessantes à análise exploratória e pode ser de grande valor a futuros estudos e aplicações de modelos preditivos multivariados no ramo do turismo.

A realização de uma investigação mais aprofundada e elaborada, como a elaborada na presente dissertação, depara-se com o desafio da obtenção dos dados necessários para a análise. Isso ocorre devido à dificuldade em encontrar informações sistematizadas, organizadas e em períodos longos, as quais nem sempre estão prontamente disponíveis para os pesquisadores da área. Além disso, como o presente estudo utilizou dados referente a todo o fluxo de passageiros em aeroportos brasileiros, o grau de complexidade aumentou e assim diminuiu a assertividade da previsão. Desse modo, para futuros estudos seria de grande valor selecionar algumas regiões ou estados brasileiros para reproduzir uma análise mais focada. Assim, é esperado que este trabalho sirva como um estímulo para a realização de estudos adicionais sobre o tema, principalmente voltados ao turismo brasileiro.

# REFERÊNCIAS BIBLIOGRÁFICAS:

Casanova, S., Moreira Guedes de Araujo, V., Vieira da Silva, W., & Torres da Rocha, D. (2012). Previsão da Demanda Turística da Cidade de Foz do Iguaçu: Uma Aplicação com Os Modelos Arima. Turismo - Visão e Ação, 14(3), 366-385.

CHETTY, Pryia; Advantages of Forecasting the Demand for Tourism. Disponível em: http://www.projectguru.in/publications/advantages-of-forecasting-the-demand-for-tourism/. Acesso em:01/02/2024.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2016). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

ANAC. Plano de Dados Abertos da ANAC — Agência Nacional de Aviação Civil ANAC. Disponível em: https://www.anac.gov.br/acesso-a-informacao/dados-abertos/plano-de-dados-abertos-vigentes. Acesso em:01/02/2024

GOUVEIA, B. V. L. et al.. Demanda turística internacional e taxa de câmbio: modelagem de dependência baseada no modelo copula-GARCH.. Revista Brasileira de Pesquisa em Turismo, v. 16, p. e–2263, 2022.

FOHB. Estudos & Pesquisas. Disponível em: https://fohb.com.br/estudos\_e\_pesquisas/. Acesso em: 01/12/2023.

Ministério do Turismo (Brazil). Number of international tourist arrivals in Brazil from 2010 to 2021 (in 1,000s). Statista. Disponível em: Inc.https://www-statista-com.sbproxy.fgv.br/statistics/805683/number-international-tourist-arrivals-brazil/?locale=en. Acesso em: 10/06/2023.

Bastian Herre, Veronika Samborska and Max Roser (2023) - "Tourism". Published online at OurWorldInData.org. Disponível em: 'https://ourworldindata.org/tourism'

Bora Investir B3. Dólar: entenda por que ele sobe e desce e como afeta os brasileiros. Disponível em: https://borainvestir.b3.com.br/objetivos-financeiros/dolar-entenda-como-a-moeda-como-afeta-os-brasileiros/. Acesso em: 05/10/2023.

HIRATA, Fernanda Akemi; BRAGA, Debora Cordeiro. Demanda turística e o estudo sobre motivação. Boa Vista: UFRR, 2017.

BENI, M C. Análise estrutural do turismo. São Paulo. Editora SENAC, 2007.

MATHIESON, Alister, WALL, Geoffrey. Turism: economic, physical and social impacts. London: Longman, 1998

BARRETO, Margarita. Manual de Iniciação ao estudo do turismo. Campinas, SP. Papirus, 1995.

LÜTKEPOHL, Helmut. Vector autoregressive models. Handbook of research methods and applications in empirical macroeconomics, v. 30, 2013.

HAYKIN, S. Neural networks: a comprehensive foundation. New Delhi: Pearson Prentice Hall, 2001

GRAVES A.Towards end-to-end speech recognition with recurrent neural networks.In: 31st International Conference on Machine Learning, Proceedings [...]. Beijing: ICML-14, 2014.

NELSON, M. Q.; PEREIRA, A.C. M.; OLIVEIRA R. A. Stock market’s price prediction with LSTM neural networks.In: International Joint Conference of Neural Networks. Proceedings [...]. Anchorage: IJCNN , Alaska, 2017.

RITA, Paulo; MORO, Sérgio; CAVALCANTI, Gabriel. The impact of COVID-19 on tourism: Analysis of online reviews in the airlines sector. Journal of Air Transport Management, v. 104, p. 102277, 2022.

SONG, Haiyan; LI, Gang. Tourism demand modelling and forecasting—A review of recent research. Tourism management, v. 29, n. 2, p. 203-220, 2008.

KHAIDI, S.M.; Abu, N.; Muhammad, N. Tourism demand forecasting—A review on the variables and models. J. Phys. Conf. Ser. 2019.

GOUVEIA, B. V. L.; COELHO, M. de F.; SILVA JÚNIOR, J. C. A. da; LACERDA, M. S. Demanda turística internacional e taxa de câmbio: modelagem de dependência baseada no modelo copula-GARCH. Revista Brasileira de Pesquisa em Turismo, [S. l.], v. 16, p. 2263, 2022. DOI: 10.7784/rbtur.v16.2263. Disponível em: https://rbtur.org.br/rbtur/article/view/2263. Acesso em: 17 out. 2023.

DIVINO, J.A; MCALEER, M. Modelling sustainable international tourism demand to the Brazilian Amazon. Environmental Modelling & Software, v.24, n.12, p.1411-1419, 2009

MARTINS, V.L.M; WERNER, L; BELLEZA, M.R. Tourism Demand In The World Cup 2014 In Brazil: An estimate based on the combination of forecasts and mathematical adjustment. Espacios. v.34, n.8, p.5, 2013.

LIANG Y H 2014 Comput. Ind. Eng. 74 111–9

YANG X, PAN B, EVANS JA and Lv B 2015 Tour. Manag. 46 386–97

VERMA, Yugesh. Augmented Dickey-Fuller (ADF) Test In Time-Series Analysis. In: Analytics India Magazine, [s.l.], Aug. 2021. Disponível em: https://analyticsindiamag.com/complete-guide-to-dickey-fuller-test-in-time-series-analysis/. Acesso em: 17 dez. 2023.

Y. Takahashi, H. Aida, and T. Saito, ‘‘ARIMA model’s superiority over f-ARIMA model,’’ in Proc. Int. Conf. Commun. Technol. (WCC-ICCT), vol. 1, 2000, pp. 66–69

U. M. Sirisha, M. C. Belavagi and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," in IEEE Access, vol. 10, pp. 124715-124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

N. Deretić, D. Stanimirović, M. A. Awadh, N. Vujanović, and A. Djukić, ‘‘SARIMA modelling approach for forecasting of traffic accidents,’’ Sustainability, vol. 14, no. 8, p. 4403, Apr. 2022.

RABAHY, W. A (2019). Análise e perspectivas do turismo no Brasil. Revista Brasileira de Pesquisa em Turismo, São Paulo, 14 (1), p. 1 - 13, jan./abr. http://dx.doi.org/10.7784/rbtur.v14i1.1903

EMBRATUR. Fortalecer o turismo é compromisso com o Brasil. Disponível em:https://embratur.com.br/2023/04/27/fortalecer-o-turismo-e-compromisso-com-o-brasil/. Acesso em: 21/02/2023.

CARRIER, P. C. LSTM network for sentiment analysis.Disponível em: < http://deeplearning.net/tutorial/lstm.html>. Acesso em: 03 abr. 2024

LOPEZ, L., & WEBER, S. (2017). Testing for Granger Causality in Panel Data. The Stata Journal, 17(4), 972-9 84. Disponível em: https://doi.org/10.1177/1536867X1801700412. Acesso em: 13/05/2024