



Observatório
Portuário



2023

RELATÓRIO METODOLÓGICO

V. 1, N. 3



UNIVERSIDADE FEDERAL DO MARANHÃO
Reitor: Prof. Dr. Natalino Salgado
Vice-Reitor: Prof. Dr. Marcos Fábio Belo Matos

Expediente

Observatório Portuário

Coordenação

Dr. Sérgio Sampaio Cutrim

Vice-coordenação

Dr. Tadeu Gomes Teixeira

Pesquisadores

Dra. Darliane Ribeiro Cunha

Dr. Leo Tadeu Robles

Cientista de dados

Francisco Jadson Brito de Oliveira

Assistente de pesquisa

Karoline Oliveira Soares

Luciane Maramaldo Oliveira

Mariana Esthefane Mendes Ribeiro

Pedro Vinícius Bernard

Marketing e Redes Sociais

Letícia Oliveira Conceição

Endereço

Av. dos Portugueses, 1996.

Vila Bacanga, São Luís - MA

CEP: 65080-805

Centro de Ciências Sociais

Departamento de Ciências Contábeis e Administração

Observatório Portuário

www.observatorioportuario.ufma.br

observatorioportuario@ufma.br

**Observatório
Portuário**

Copyright © 2023 by UFMA



Baum Soluções Informacionais
Projeto gráfico

SUMÁRIO

02

APRESENTAÇÃO

03

QUEM SOMOS

04

1 IMPORTÂNCIA DA METODOLOGIA NA ELABORAÇÃO DE RELATÓRIOS TÉCNICOS

07

2 ANÁLISE QUANTITATIVA: CARACTERÍSTICAS E FERRAMENTAS

09

3 ANÁLISE DESCRITIVA DOS DADOS: TÉCNICAS USADAS NOS RELATÓRIOS

12

4 LINGUAGENS E SOFTWARES USADOS NO OBSERVATÓRIO PORTUÁRIO

13

4.1 Python para análise de dados

14

4.2 R para análise de dados

15

4.3 Visualização de dados com Power BI

17

5 CONSIDERAÇÕES FINAIS

20

6 CONVÊNIO

APRESENTAÇÃO

O Observatório Portuário é um *think thank* originado do Grupo de Pesquisa Labportos da Universidade Federal do Maranhão (UFMA) e financiado pela Empresa Maranhense de Administração Portuária (EMAP).

Nosso propósito é subsidiar os atores do setor portuário (gestores públicos, empresários, trabalhadores e sociedade) com informações analisadas a partir do contexto regional maranhense, sobretudo em aspectos relacionados a relação portocidade, impactos econômicos da atividade portuária, indicadores da atividade, dentre outros cenários e perspectivas que influenciam as estratégias para o setor.

Esperamos que nossos produtos, disponibilizados gratuitamente, fomentem discussões e estratégias para o setor.

Boa Leitura!

Prof. Dr. Sergio Sampaio Cutrim





1 Prof. Dr. Sérgio Cutrim
Coordenador

2 Prof. Dr. Tadeu Gomes Teixeira
Vice - coordenador

3 Prof. Dr. Leo Tadeu Robles
Pesquisador

4 Profa. Dra. Darliane Ribeiro Cunha
Pesquisadora

5 Francisco Jadson de Oliveira
Cientista de Dados

6 Leticia Oliveira
Social Media

7 Mariana Esthefane Ribeiro
Assistente de Pesquisa

8 Luciane Maramaldo
Assistente de Pesquisa

9 Karoline Soares
Assistente de Pesquisa



1 IMPORTÂNCIA DA METODOLOGIA NA ELABORAÇÃO DE RELATÓRIOS TÉCNICOS

A análise de dados tem se tornado cada vez mais relevante e imprescindível em diversas áreas, desde a ciência e a tecnologia até os negócios e as políticas públicas. Porém, a obtenção de insights significativos e a tomada de decisões confiáveis baseadas em dados requerem mais do que apenas ferramentas e técnicas estatísticas avançadas. A metodologia empregada nesse processo desempenha um papel crucial para garantir a qualidade, a precisão e a interpretação correta dos resultados obtidos, preocupação que orientou o processo de elaboração dos relatórios do Observatório Portuário.

A metodologia, nesse contexto, refere-se ao conjunto de procedimentos, técnicas e etapas adotadas para conduzir a análise de dados de forma sistemática e rigorosa. Ela estabelece as diretrizes que norteiam a coleta, o processamento, a organização e a interpretação dos dados, permitindo que os resultados sejam reproduzíveis e generalizáveis. Além disso, a metodologia adequada assegura a confiabilidade dos resultados e a minimização de vieses que poderiam afetar a validade das conclusões.



Um dos aspectos cruciais da metodologia na análise de dados é a definição clara e precisa das perguntas de pesquisa ou dos objetivos a serem alcançados. Uma formulação adequada desses questionamentos orienta todo o processo de análise, evitando desvios ou explorações desnecessárias. Isso também permite a escolha adequada das variáveis relevantes, que devem ser coletadas e consideradas durante o processo.

A seleção adequada das fontes de dados também é uma etapa fundamental da metodologia. É importante garantir que as informações obtidas sejam representativas e confiáveis, evitando vieses amostrais que possam comprometer os resultados. Além disso, a coleta de dados deve ser feita de forma ética e respeitando as normas de privacidade, protegendo a identidade dos indivíduos envolvidos.

Após a coleta de dados, a metodologia determina as técnicas de análise a serem aplicadas. Isso pode envolver o uso de métodos estatísticos, técnicas de aprendizado de máquina, análise de séries temporais ou outras abordagens específicas, dependendo do tipo de dados e das perguntas de pesquisa. A escolha das técnicas adequadas é essencial para obter resultados válidos e relevantes.



Outro aspecto importante da metodologia é o tratamento dos dados ausentes ou inconsistentes. Erros de coleta, valores faltantes e outliers podem ocorrer em qualquer conjunto de dados. A metodologia deve propor estratégias para lidar com essas questões de maneira adequada, seja por meio de imputação de dados, técnicas de limpeza ou identificação e correção de erros.

A interpretação correta dos resultados é outro ponto-chave proporcionado pela metodologia sólida. A análise de dados pode gerar insights valiosos, mas sem uma interpretação cuidadosa, esses resultados podem ser mal interpretados ou usados de forma inadequada. A metodologia auxilia na contextualização dos resultados, compreendendo suas limitações e incertezas, e fornecendo conclusões que possam ser devidamente comunicadas a outros profissionais, gestores, tomadores de decisão ou ao público em geral.

Além disso, a metodologia também desempenha um papel importante na replicabilidade dos estudos. A capacidade de outros pesquisadores ou profissionais em reproduzir os mesmos resultados é fundamental para validar as conclusões e fortalecer a confiança nas descobertas das análises.



2 ANÁLISE QUANTITATIVA: CARACTERÍSTICAS E FERRAMENTAS

A análise quantitativa é uma abordagem essencial na interpretação de dados em diversas áreas do conhecimento, desde a ciência até os negócios e as políticas públicas. Ela se concentra na utilização de métodos e técnicas estatísticas para descrever, analisar e interpretar dados numéricos, com o objetivo de extrair insights, identificar padrões e realizar previsões com base em evidências concretas.

A principal característica da análise quantitativa é o foco nos números e nos dados mensuráveis. Ela se baseia em informações quantificáveis, como valores numéricos, medidas, índices e proporções, o que a torna especialmente adequada para tratar de questões objetivas e mensuráveis. Essa abordagem permite uma visão mais precisa e rigorosa dos fenômenos estudados, contribuindo para a obtenção de resultados confiáveis.



Uma das principais ferramentas da análise quantitativa é a estatística descritiva, que busca resumir e apresentar os dados de forma concisa e significativa. Nessa etapa, medidas como média, mediana, moda, desvio padrão e percentis são utilizadas para fornecer uma visão geral do conjunto de dados, ajudando a identificar tendências, variações e características relevantes.

Além da estatística descritiva, a análise quantitativa emprega técnicas de inferência estatística, que permitem fazer generalizações sobre uma população maior a partir de uma amostra menor de dados. Isso é especialmente útil quando se deseja tirar conclusões sobre um grupo maior de indivíduos ou observações sem ter que examiná-los individualmente. Através da inferência estatística, é possível estabelecer a probabilidade de certos eventos ocorrerem e medir a confiança nos resultados obtidos.

A análise quantitativa é amplamente utilizada em pesquisas científicas para testar hipóteses, validar teorias e obter evidências empíricas para suportar ou refutar argumentos. Na área de negócios, ela desempenha um papel crucial na tomada de decisões estratégicas, como na análise de mercado, na previsão de demanda e na avaliação de desempenho financeiro. Também é amplamente aplicada em pesquisas de opinião, levantamentos populacionais e análises de políticas públicas.

Contudo, é importante reconhecer que a análise quantitativa tem suas limitações. Ela pode fornecer respostas precisas para questões específicas, mas nem sempre é capaz de capturar a complexidade total dos fenômenos estudados. Além disso, é fundamental que os dados utilizados sejam representativos e coletados de forma adequada, a fim de evitar distorções nos resultados.

Outro aspecto relevante é que a análise quantitativa, por si só, não fornece explicações causais. Ela pode mostrar correlações entre variáveis, mas não necessariamente revela as razões subjacentes a essas relações. Para isso, é frequentemente necessário combinar a análise quantitativa com abordagens qualitativas, que buscam compreender os processos e os contextos sociais mais amplos que influenciam os resultados observados.

3 ANÁLISE DESCRITIVA DOS DADOS: TÉCNICAS USADAS NOS RELATÓRIOS

A análise descritiva, também conhecida como estatística descritiva, é uma das etapas fundamentais da análise de dados. Essa abordagem estatística tem como objetivo descrever e resumir as características principais de um conjunto de dados, fornecendo uma visão geral do comportamento das variáveis estudadas. É uma das técnicas mais básicas e utilizadas na análise quantitativa.

A análise descritiva utiliza uma série de medidas e gráficos para resumir e visualizar os dados de maneira concisa e informativa. Entre as principais técnicas utilizadas estão:

1. Medidas de tendência central:

As medidas de tendência central são utilizadas para encontrar um valor central que represente o conjunto de dados. São três as principais medidas:

- Média aritmética: É obtida somando todos os valores do conjunto de dados e dividindo pelo número total de observações. A média é sensível a valores extremos, o que significa que um único valor muito alto ou muito baixo pode distorcer o resultado. É amplamente utilizada quando se tem uma distribuição aproximadamente simétrica dos dados.
- Mediana: É o valor que separa o conjunto de dados em duas partes iguais. Para calcular a mediana, os dados devem ser organizados em ordem crescente ou decrescente, e a mediana será o valor do meio. Se houver um número par de observações, a mediana é a média dos dois valores centrais. A mediana é menos sensível a valores extremos, sendo uma medida mais robusta em relação à média.
- Moda: Representa o valor que ocorre com maior frequência no conjunto de dados. Pode haver mais de uma moda (moda múltipla) ou nenhum valor repetido (sem moda). A moda é útil para identificar o valor mais frequente em uma distribuição, mas nem sempre é uma medida representativa do conjunto de dados, especialmente quando a distribuição é ampla e sem uma clara tendência central.

2. Medidas de dispersão:

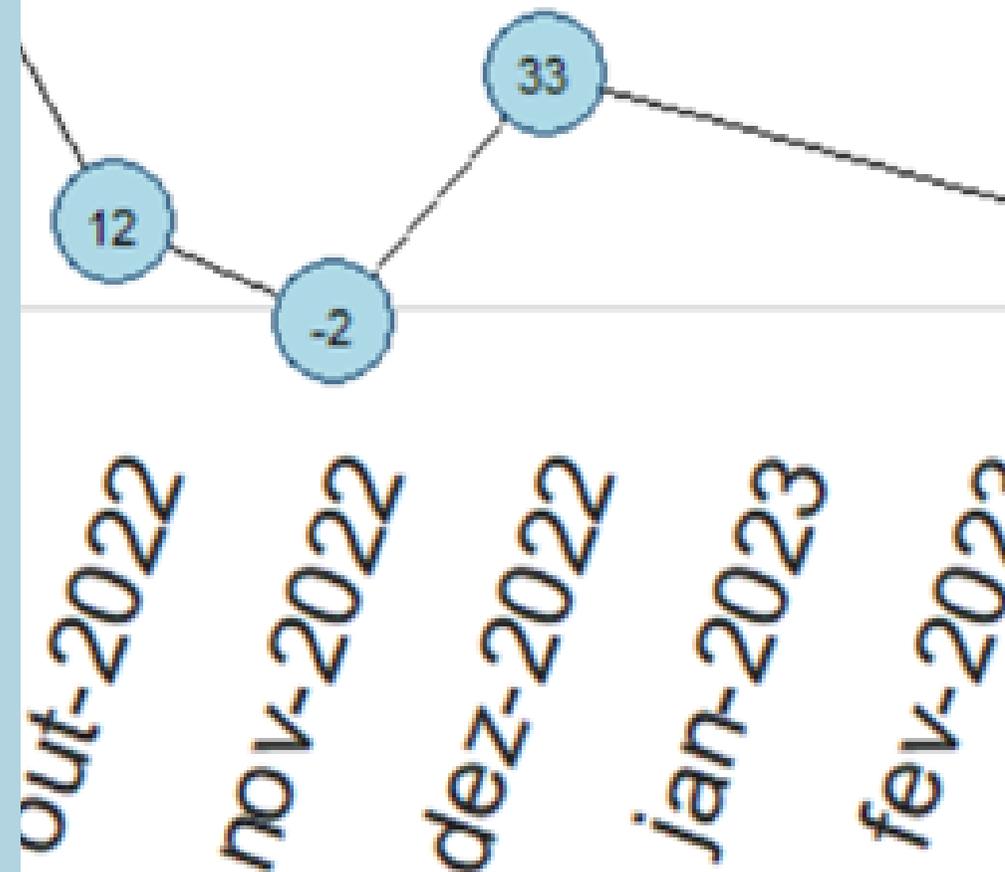
As medidas de dispersão indicam o quão espalhados os dados estão em relação à medida de tendência central. Elas fornecem informações sobre a variabilidade dos dados. Algumas das principais medidas de dispersão são:

- Desvio padrão: É a medida mais comum de dispersão e representa a raiz quadrada da variância. O desvio padrão mostra o quanto os valores do conjunto de dados tendem a se afastar da média. Um desvio padrão maior indica maior dispersão dos dados, enquanto um desvio padrão menor indica maior proximidade dos dados à média.

- Amplitude: É a diferença entre o maior e o menor valor do conjunto de dados. A amplitude indica o intervalo total dos dados, mas pode ser influenciada por valores extremos e não reflete a variabilidade interna dos dados.

- Variância: É uma medida que quantifica a dispersão dos dados em relação à média. A variância é obtida calculando a média dos quadrados das diferenças entre cada valor e a média. Uma variância maior indica maior dispersão dos dados.

- Coeficiente de variação: É uma medida relativa de dispersão, expressa como uma porcentagem do desvio padrão em relação à média. É útil para comparar a variabilidade de diferentes conjuntos de dados, especialmente quando as médias são diferentes.



3. Gráficos descritivos:

Os gráficos são uma maneira poderosa de visualizar os dados e identificar padrões visuais rapidamente. Diferentes tipos de gráficos são usados dependendo do tipo de dado e do objetivo da análise:

- Histograma: É um gráfico de barras utilizado para representar a distribuição dos dados numéricos em intervalos. Cada barra representa um intervalo e a altura da barra mostra a frequência dos valores dentro daquele intervalo.
- Gráfico de barras: Utilizado para dados categóricos, mostra a frequência de cada categoria em forma de barras. É especialmente útil para comparar a frequência de diferentes categorias.
- Gráfico de dispersão: Mostra a relação entre duas variáveis numéricas. Cada ponto no gráfico representa uma observação e sua posição no eixo x e y representa os valores das duas variáveis. O gráfico de dispersão é útil para identificar padrões de correlação entre as variáveis.
- Box plot: Também conhecido como diagrama de caixa, é um gráfico que apresenta a distribuição dos dados de forma resumida. O box plot mostra a mediana, os quartis e os valores extremos dos dados, permitindo identificar a dispersão e a presença de outliers.
- Gráfico de linha: É usado para representar a variação de uma variável ao longo do tempo ou de um outro parâmetro contínuo. Cada ponto é conectado por uma linha, facilitando a identificação de tendências ou padrões ao longo do tempo.

A análise descritiva é uma etapa crucial na exploração inicial dos dados. Ela oferece uma visão geral dos principais aspectos do conjunto de dados, fornecendo informações que podem guiar análises mais aprofundadas e ajudar a formular hipóteses para investigações posteriores. A partir da análise descritiva identificam-se padrões, tendências, discrepâncias e outliers, o que pode levar a insights importantes e suportar a tomada de decisões informadas em diferentes áreas de estudo.

4 LINGUAGENS E SOFTWARES USADOS NO OBSERVATÓRIO PORTUÁRIO

Python e R são duas das linguagens de programação mais populares e poderosas utilizadas para análise de dados. Ambas possuem bibliotecas e pacotes extensos que oferecem uma ampla gama de funcionalidades estatísticas, gráficas e de manipulação de dados, tornando-as escolhas populares entre cientistas de dados, analistas e pesquisadores.



4.1 Python para análise de dados

Python é uma linguagem de programação versátil e fácil de aprender, o que a torna uma escolha popular para iniciantes e profissionais experientes. Algumas das principais bibliotecas Python para análise de dados são:

- Pandas: É uma biblioteca poderosa para manipulação e análise de dados em formato tabular. O Pandas permite ler, limpar e filtrar dados, realizar agregações, transformações e combinações de datasets, além de lidar com dados ausentes de forma eficiente.
- NumPy: É uma biblioteca fundamental para computação numérica em Python. Ela oferece suporte para arrays multidimensionais e operações matemáticas eficientes, permitindo cálculos rápidos em grandes conjuntos de dados.
- Matplotlib e Seaborn: São bibliotecas para criação de gráficos e visualizações. O Matplotlib oferece uma ampla variedade de gráficos, enquanto o Seaborn é especialmente projetado para criar gráficos estatísticos atraentes e informativos.
- SciPy: É uma biblioteca que estende as funcionalidades do NumPy, oferecendo ferramentas adicionais para otimização, estatísticas, interpolação e outras operações científicas.
- Scikit-learn: É uma biblioteca de aprendizado de máquina (machine learning) muito popular em Python. Ela oferece uma ampla gama de algoritmos de classificação, regressão, clustering, entre outros, além de ferramentas para pré-processamento e avaliação de modelos de aprendizado de máquina.



4.2 R para análise de dados

R é uma linguagem de programação estatística amplamente utilizada em ambientes acadêmicos e pesquisas, especialmente em ciências sociais e bioinformática. Algumas das principais bibliotecas R para análise de dados são:

- dplyr: É uma biblioteca que fornece funções eficientes para manipulação e transformação de dados. Com o dplyr, é possível realizar operações como filtragem, seleção de colunas, agregações e junções de forma simples e concisa.

- ggplot2: É uma biblioteca para criação de gráficos e visualizações. O ggplot2 utiliza a gramática dos gráficos, o que permite criar visualizações complexas com poucas linhas de código. É especialmente reconhecido por criar gráficos estatísticos de alta qualidade.

- tidyr: É uma biblioteca que facilita a limpeza e organização dos dados em um formato mais "arrumado". Ela permite transformar os dados para o formato "tidy data", que é uma forma estruturada e consistente de organizar dados tabulares.

- base: É a biblioteca base do R e oferece funções essenciais para manipulação de vetores, matrizes e listas. Embora muitas funcionalidades já estejam disponíveis no R, outras bibliotecas como dplyr e tidyr tornaram-se populares por oferecer uma sintaxe mais intuitiva e eficiente.

- caret: É uma biblioteca para aprendizado de máquina em R. Assim como o Scikit-learn em Python, o caret oferece uma variedade de algoritmos de aprendizado de máquina e ferramentas para avaliação e ajuste de modelos.

Foram usadas as duas linguagens de programação nas atividades do Observatório, complementando-se.

4.3 Visualização de dados com Power BI

O Power BI é uma ferramenta de visualização de dados desenvolvida pela Microsoft que tem se tornado cada vez mais popular em empresas e organizações de diferentes setores. Ele oferece um conjunto abrangente de recursos para criar, compartilhar e interagir com relatórios e painéis interativos, permitindo que os usuários explorem e entendam melhor seus dados de forma rápida e intuitiva.

Características do Power BI:

1. Conexão com diversas fontes de dados: O Power BI permite conectar-se a uma ampla variedade de fontes de dados, como bancos de dados SQL, Excel, SharePoint, serviços em nuvem (Azure, Google Analytics, Salesforce) e muitos outros. Isso torna a integração de dados de diferentes origens mais fácil e eficiente.
2. Criação de relatórios interativos: Com o Power BI, é possível criar relatórios interativos e visualmente atraentes usando uma variedade de gráficos, tabelas, mapas, medidores e outros elementos visuais. A interface amigável e a capacidade de arrastar e soltar facilitam a criação e personalização dos relatórios de acordo com as necessidades do usuário.
3. Painéis de controle dinâmicos: Os painéis do Power BI permitem agrupar relatórios e gráficos relacionados em uma única página, proporcionando uma visão geral dos principais indicadores de desempenho. Os painéis são dinâmicos e atualizam-se automaticamente à medida que os dados mudam, permitindo um acompanhamento em tempo real de métricas importantes.
4. Capacidades de análise avançada: O Power BI oferece recursos de análise avançada, como a criação de medidas personalizadas usando a linguagem DAX (Data Analysis Expressions). Isso permite que os usuários realizem cálculos complexos e desenvolvam métricas personalizadas para atender a requisitos específicos de negócios.



O Power BI tem um enorme potencial de análise, pois permite reunir e consolidar dados de diferentes fontes em um único lugar, transformá-los em informações visuais e interativas e compartilhar essas análises com colegas e stakeholders. Além disso, suas capacidades de análise avançada permitem a criação de modelos e cálculos personalizados, fornecendo insights mais profundos e detalhados.

Com a possibilidade de criar painéis e relatórios atualizados em tempo real, o Power BI facilita a tomada de decisões com base em dados atualizados e precisos. Ele também permite a exploração interativa dos dados, permitindo que os usuários explorem diferentes cenários e visualizem dados de diferentes perspectivas.

O potencial de análise do Power BI é ainda mais aprimorado quando combinado com outras ferramentas e linguagens de programação, como Python e R. Essa integração permite a criação de análises mais avançadas, modelos de machine learning e outras técnicas analíticas complexas, ampliando ainda mais o poder da ferramenta na obtenção de insights valiosos e na tomada de decisões fundamentadas.



5 CONSIDERAÇÕES FINAIS

Neste relatório, exploramos a importância da metodologia na análise de dados, destacando como uma abordagem sistemática e rigorosa é essencial para garantir a qualidade e a validade dos resultados obtidos. Além disso, enfatizamos que a definição clara das perguntas de pesquisa, a seleção adequada das fontes de dados, o tratamento de dados ausentes e a interpretação correta dos resultados são aspectos cruciais da metodologia.

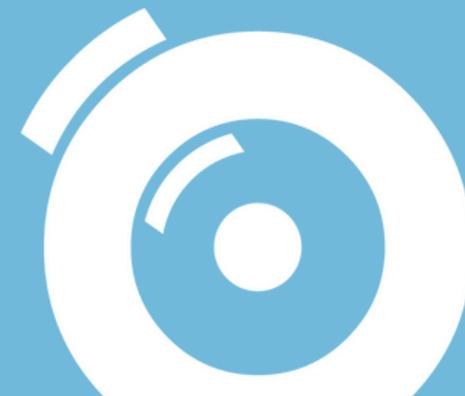
Discorremos também sobre a análise quantitativa, ressaltando como essa abordagem estatística é fundamental para a compreensão objetiva e precisa de diferentes fenômenos. Ao utilizar técnicas como medidas de tendência central, medidas de dispersão e gráficos descritivos, a análise quantitativa permite extrair insights significativos e identificar padrões nos dados.

Além disso, destacamos o papel das ferramentas Python e R na análise de dados, enfatizando como essas linguagens de programação têm se mostrado poderosas e versáteis para lidar com diversas tarefas analíticas. Suas bibliotecas e pacotes fornecem uma ampla gama de recursos estatísticos, gráficos e de manipulação de dados, permitindo que os usuários explorem, visualizem e entendam melhor suas informações.

Finalmente, abordamos o Power BI, uma ferramenta de visualização de dados que se destaca pela facilidade de uso e recursos interativos. Suas características de criação de relatórios e painéis interativos, combinadas com a capacidade de análise avançada, tornam o Power BI uma escolha popular para análise de dados em diferentes cenários, desde análises de desempenho de negócios até dados de marketing e recursos humanos.



A análise de dados é uma área fundamental para a tomada de decisões informadas e a obtenção de insights valiosos na área portuária. A metodologia adequada, a abordagem quantitativa e o uso de ferramentas poderosas, como Python, R e Power BI, desempenham um papel crucial no processo de análise de dados, permitindo que profissionais e pesquisadores enfrentem desafios complexos e transformem dados em conhecimento significativo. Com uma abordagem cuidadosa e o uso apropriado das ferramentas disponíveis, é possível explorar o potencial dos dados portuários e subsidiar o processo de tomada de decisão.



6 CONVÊNIO



O presente relatório do Observatório Portuário é fruto de um Convênio de Educação, Ciência, Tecnologia e Inovação (ECTI), celebrado com a Empresa Maranhense de Administração Portuária, administradora do Porto do Itaqui. Cabe registrar informações institucionais sobre o Porto do Itaqui.

Estrategicamente localizado próximo aos principais mercados mundiais, o porto público do Maranhão possui infraestrutura para movimentar granéis sólidos vegetais e minerais, líquidos, cargas gerais e contêineres. Em 2021 foram movimentadas mais de 31 milhões de toneladas de cargas, ano em que 24 marcas históricas foram superadas, com destaque para soja, combustíveis e fertilizantes.

Principal exportador de grãos na região Centro-Norte, o Itaqui é o único porto público do Brasil com quatro certificações ISO: 9001 (Qualidade), 14001 (Meio Ambiente), 27001 (Segurança da Informação) e 45001 (Segurança e Saúde Ocupacional).

Conectado ao restante do país por modernas ferrovias e rodovias, o Itaqui destaca-se como o principal porto do Corredor Centro-Norte do país, um trabalho reconhecido com nota máxima no IGAP - Índice da Gestão das Autoridades Portuárias, categoria do Prêmio Portos + Brasil 2022, do Ministério de Infraestrutura.

Porto do Itaqui, patrimônio do Maranhão e do Brasil



SIGA-NOS NAS REDES SOCIAIS



TWITTER

@obs_portuario



INSTAGRAM

@observatorioportuario