

VARIAÇÕES DE SENSAÇÃO TÉRMICA EM SANTOS: Aplicando Regressão Linear Múltipla e Modelagem Estatística a Fatores climáticos

THERMAL SENSATION VARIATIONS IN SANTOS: Applying Multiple Linear Regression and statistical modeling to climatic factors

SERGIO LOPES DA SILVA JUNIOR (FATEC RL)
sergio.silva116@fatec.sp.gov.br

RESUMO

O estudo tem como objetivo desenvolver um modelo empírico de estimativa da sensação térmica (índice de calor) com base em variáveis meteorológicas obtidas em uma estação automática. Para isso, aplicou-se regressão linear múltipla utilizando a biblioteca scikit-learn, considerando as variáveis temperatura do ar, umidade relativa, velocidade do vento, pressão atmosférica e radiação solar como preditores. A modelagem foi conduzida com enfoque exploratório e segmentação sazonal, a fim de verificar a sensibilidade dos coeficientes e métricas de desempenho (R^2 e RMSE) ao longo dos trimestres. Foram utilizados gráficos gerados com matplotlib e análises de correlação com seaborn para melhor compreensão visual dos dados. Os resultados apontaram alta acurácia do modelo nos trimestres de menor variabilidade térmica (Q2 e Q3), com coeficientes ajustados em conformidade com a literatura sobre conforto térmico humano. A matriz de correlação evidenciou forte dependência entre temperatura e as demais variáveis, indicando que o calor sentido pelo corpo humano pode ser significativamente modulado por fatores ambientais além da temperatura isolada. No verão (Q1), observou-se aumento acentuado no índice de calor estimado, reforçando a hipótese de que a combinação de alta umidade e radiação amplifica a percepção térmica. As análises indicam que o modelo é útil como ferramenta de apoio à caracterização do microclima urbano e avaliação do desconforto térmico. Conclui-se que a inclusão de múltiplas variáveis meteorológicas melhora significativamente a estimativa do índice de calor em relação ao uso apenas da temperatura, permitindo melhor suporte à tomada de decisão em saúde pública e planejamento ambiental.

Palavras-chave: Análise de dados; Ciências de dados; Mudanças climáticas; Modelagem estatística; Sensação térmica.

ABSTRACT

This study aims to develop an empirical model to estimate thermal sensation (heat index) based on meteorological variables collected from an automatic weather station. Multiple linear regression was applied using the scikit-learn library, with air temperature, relative humidity, wind speed, atmospheric pressure, and solar radiation as independent variables. The modeling followed an exploratory approach, segmented by season, to assess the sensitivity of regression coefficients and performance metrics (R^2 and RMSE) throughout the quarters. Visual analyses were performed using matplotlib plots and correlation matrices generated with seaborn,

enhancing data interpretation. The results showed high model accuracy during periods of lower thermal variability (Q2 and Q3), with coefficients aligned with thermal comfort literature. The correlation matrix revealed a strong relationship between temperature and other atmospheric factors, suggesting that the human perception of heat is significantly influenced by combined environmental conditions. In the summer quarter (Q1), a marked increase in the estimated heat index was observed, supporting the hypothesis that high humidity and solar radiation amplify thermal discomfort. These findings indicate the model's potential as a tool for urban microclimate characterization and thermal stress assessment. It is concluded that incorporating multiple meteorological variables significantly improves the accuracy of heat index estimation compared to using temperature alone, thus offering valuable insights for public health strategies and urban planning.

KEYWORDS: *Data analysis; Data sciences; Climate change; Statistical modeling; Thermal sensation.*

1 INTRODUÇÃO

A sensação térmica representa a temperatura efetivamente percebida pelo corpo humano, sendo determinada não apenas pela temperatura do ar, mas também por fatores como umidade relativa. Segundo Steadman (1979), essa percepção não é um cálculo exato, já que diferentes corpos de massas podem sentir temperaturas distintas, onde a combinação desses elementos pode gerar desconforto térmico significativo e impactar diretamente a saúde pública e a qualidade de vida.

Segundo o sistema de classificação climática global de Köppen-Geiger (Peel, 2007), Santos é identificada como região 'Af', correspondente ao clima equatorial úmido, caracterizado por elevada pluviosidade, temperaturas elevadas e ausência de estação seca. Essa condição climática, embora comum em regiões tropicais, pode intensificar a sensação de abafamento, especialmente em áreas urbanizadas com menor circulação de ar. Como destaca Parsons (2014), ambientes quentes e úmidos favorecem o acúmulo de calor corporal, o que pode levar a estresse térmico, fadiga, insolação e, em casos extremos, riscos à saúde cardiovascular.

Com o avanço de modelos estatísticos, tornou-se viável modelar essas interações climáticas utilizando ferramentas estatísticas e computacionais. A regressão linear múltipla, por exemplo, permite estimar o efeito simultâneo de diversas variáveis independentes sobre uma variável dependente, como a sensação térmica. Esse método é amplamente aplicado em estudos ambientais por sua capacidade de gerar modelos explicativos robustos e previsões confiáveis (Montgomery & Runger, 2018; Wilks, 2019). Além disso, o uso de algoritmos como o Heat Index, proposto por Steadman (1979), permite refinar a percepção de calor ao incorporar os efeitos combinados da temperatura e da umidade.

Neste estudo, propõe-se a construção de um modelo estatístico baseado em regressão linear múltipla, com o objetivo de prever a sensação térmica na cidade de Santos ao longo de um ano. Os dados utilizados foram obtidos da Companhia Ambiental do Estado de São Paulo (CETESB), por meio da plataforma QualAr, e incluem as variáveis: temperatura média, umidade relativa do ar, pressão atmosférica, velocidade do vento e radiação solar. Para análise, serão empregadas linguagens e ferramentas como Python, utilizando das seguintes bibliotecas: pandas, numpy, scikit-learn, matplotlib e seaborn, que possibilita visualização clara dos

resultados estatísticos, assim como a previsão e regressão dos dados adquiridos, em um escopo de 1 ano, de abril de 2024 a abril de 2025.

Este trabalho tem como objetivo desenvolver um modelo estatístico baseado em regressão linear múltipla para prever a sensação térmica em Santos-SP, utilizando dados meteorológicos da CETESB (QualAr). A modelagem estatística permitirá identificar quais variáveis apresentam maior peso na construção da temperatura percebida, bem como verificar padrões sazonais e flutuações temporais ao longo do período analisado. A aplicação deste estudo se estende a áreas como urbanismo, saúde pública e planejamento climático. Esta abordagem demonstra como modelos estatísticos aplicados podem transformar dados meteorológicos em políticas públicas eficazes, particularmente em cidades costeiras com clima tropical úmido como Santos.

2 FUNDAMENTAÇÃO TEÓRICA

A sensação térmica, ou temperatura aparente, é o resultado da interação entre variáveis ambientais que afetam a percepção subjetiva do calor pelo corpo humano. Entre os fatores mais relevantes destacam-se a temperatura do ar, a umidade relativa, a velocidade do vento, a pressão atmosférica e, em alguns contextos, a radiação solar. Segundo Steadman (1979), o *Heat Index* (HI) índice amplamente utilizado para estimar a sensação térmica, é calculado a partir de uma fórmula empírica que incorpora a temperatura e a umidade relativa para refletir o desconforto térmico percebido. Embora seja um modelo padronizado, sua aplicabilidade pode ser estendida por meio de técnicas estatísticas que considerem outras variáveis ambientais relevantes ao contexto regional.

Estudos como os de Parsons (2014) apontam que ambientes urbanos com clima quente e úmido favorecem o acúmulo de calor corporal, especialmente quando há baixa circulação de ar e alta radiação solar. Esse fenômeno pode intensificar o estresse térmico, acarretando sintomas como fadiga, tontura, desidratação e, em casos extremos, agravamento de doenças cardiovasculares. Cidades costeiras como Santos, no litoral paulista, caracterizam-se por um clima tropical úmido (Af na classificação de Köppen-Geiger), com temperaturas elevadas e alta umidade durante praticamente todo o ano. Tais características justificam a escolha de Santos como objeto de estudo para modelagem da sensação térmica.

Do ponto de vista estatístico, a regressão linear múltipla configura-se como uma ferramenta robusta e amplamente adotada na modelagem de fenômenos ambientais complexos. Conforme discutido por Montgomery e Runger (2018), essa técnica permite avaliar o efeito simultâneo de diversas variáveis independentes sobre uma variável resposta, tornando-se ideal para estimar a sensação térmica a partir de múltiplos fatores meteorológicos. Além disso, a regressão linear permite interpretar a influência de cada variável por meio de seus coeficientes, facilitando a identificação de quais fatores mais impactam a temperatura percebida.

A abordagem adotada neste estudo visa utilizar a regressão linear múltipla não apenas como ferramenta preditiva, mas também explicativa, permitindo análises como: a avaliação da importância relativa das variáveis (por meio dos coeficientes padronizados), a análise de resíduos (para verificar a adequação do modelo) e a avaliação do desempenho por trimestre, visando compreender possíveis padrões sazonais. A escolha por Python e bibliotecas como pandas, scikit-learn e seaborn deve-se à flexibilidade que essas ferramentas oferecem para manipulação, visualização e modelagem de dados, possibilitando uma análise clara e reprodutível.

2.1 Ferramentas da ciência de dados

Para esta pesquisa, as principais ferramentas usadas foram: Google Colab, Python junto a bibliotecas de códigos como *Pandas*, *Numpy*, *Scikit-learn*, *Matplotlib* e *Seaborn*.

O Google Colab é uma plataforma online gratuita que permite a execução de notebooks Jupyter na nuvem, com suporte nativo para Python e bibliotecas como *pandas*, *sklearn* e *matplotlib*. Ele oferece um ambiente interativo com recursos computacionais escaláveis e integração com o Google Drive, o que facilita o carregamento, armazenamento e processamento de grandes conjuntos de dados. Neste projeto, o Google Colab foi utilizado como ambiente principal para leitura dos dados meteorológicos da CETESB, desenvolvimento do modelo de regressão múltipla e visualização dos resultados ao longo do período analisado.

Python é uma linguagem de programação muito popular em aplicações web, desenvolvimento de softwares, ciência de dados e aprendizado de máquina, permitindo o usuário, entre outras funcionalidades, ênfase em extrair, transformar e carregar dados, diminuindo o tempo e aperfeiçoando uma determinada função. Utilizando de base para o uso das seguintes bibliotecas:

A biblioteca *Pandas* é usada para manipulação e análise de dados em Python. É amplamente utilizada para tarefas como limpeza, transformação e organização de dados tabulares em estruturas chamadas *DataFrames*. Com o *Pandas*, os usuários podem carregar conjuntos de dados de diversos formatos, como CSV, Excel e SQL, para explorar e manipular os dados de maneira eficiente. Ele oferece métodos para filtragem, agrupamento, preenchimento de valores ausentes e cálculos estatísticos.

A biblioteca *NumPy* é fundamental para a computação numérica em Python. Ela oferece suporte a *arrays* multidimensionais, que são estruturas altamente eficientes para armazenar e manipular grandes volumes de dados numéricos. *NumPy* também fornece uma variedade de funções matemáticas para operações lineares, e transformações estatísticas.

O *Scikit-learn*, ou *sklearn*, é uma biblioteca voltada para aprendizado de máquina e análise de dados. Ela oferece ferramentas para tarefas como classificação, regressão, agrupamento, redução de dimensionalidade e avaliação de modelos. Uma funcionalidade central do *sklearn* é a implementação de algoritmos como regressão linear múltipla, que permite modelar a relação entre uma variável dependente e várias variáveis independentes simultaneamente, sendo útil para prever fenômenos complexos como a sensação térmica a partir de dados meteorológicos.

A biblioteca *Matplotlib* utiliza suas ferramentas em linguagem Python para a criação de visualizações gráficas e análise de dados. Ela permite a geração de gráficos 2D, como linhas, barras, dispersão, histogramas, e gráficos de setores, além de possibilitar personalizações detalhadas em cada aspecto das representações visuais.

O *Seaborn* é uma biblioteca de visualização de dados baseada no *matplotlib*, projetada para tornar os gráficos estatísticos mais acessíveis e informativos. Ela fornece interfaces de alto nível para a construção de gráficos como distribuições, correlações, regressões lineares e mapas de calor. Neste estudo, o *Seaborn* foi utilizado para construir visualizações como a matriz de correlação entre variáveis meteorológicas, facilitando a identificação de relações estatísticas entre os fatores que influenciam a sensação térmica.

2.2 Fontes de dados utilizados no estudo

Os dados utilizados neste estudo foram obtidos a partir da Companhia Ambiental do Estado de São Paulo (CETESB), por meio da plataforma QualAr, que disponibiliza registros históricos de variáveis meteorológicas coletadas por estações automáticas de monitoramento ambiental. A estação selecionada para esta pesquisa está localizada no município de Santos-SP, cidade litorânea que apresenta condições climáticas específicas de um clima equatorial úmido (classificação "Af" segundo Köppen-Geiger).

O conjunto de dados cobre o período de abril de 2024 a abril de 2025, com registros em frequência horária. As variáveis meteorológicas incluídas no banco de dados foram:

- Data (dia/mês/ano)
- Hora (registro em hora cheia)
- Temperatura (°C)
- Umidade relativa do ar (%)
- Velocidade do vento (m/s)
- Pressão atmosférica (hPa)
- Radiação solar (W/m²)

Essas variáveis foram tratadas e integradas a um dataframe em Python utilizando bibliotecas como pandas e numpy, possibilitando a manipulação temporal, identificação de valores ausentes e preparação dos dados para modelagem estatística. Com base nos valores de temperatura e umidade, foi calculado o índice de calor (Heat Index) em graus Celsius, utilizando a fórmula proposta por Steadman (1979), conforme abaixo:

$$\text{HIF} = -42,379 + 2,04901523 \cdot T + 10,14333127 \cdot \text{RH} - 0,22475541 \cdot \text{TR} \cdot \text{H} - 6,83783 \times 10^{-3} \cdot T^2 - 5,481717 \times 10^{-2} \cdot \text{RH}^2 + 1,22874 \times 10^{-3} \cdot T^2 \cdot \text{RH} + 8,5282 \times 10^{-4} \cdot T \cdot \text{RH}^2 - 1,99 \times 10^{-6} \cdot T^2 \cdot \text{RH}^2$$

Onde:

HIF = Heat Index

T = temperatura do ar em Fahrenheit

RH = umidade relativa do ar (%)

O resultado foi posteriormente convertido para graus Celsius, viabilizando a construção da variável dependente "Heat_Index", que representa a sensação térmica percebida. Essa variável foi utilizada como alvo (variável dependente) em um modelo de regressão linear múltipla, cuja modelagem considera como preditores as demais variáveis meteorológicas.

3 DESENVOLVIMENTO METODOLÓGICO

O desenvolvimento metodológico deste estudo baseia-se na construção de um modelo estatístico de regressão linear múltipla, com o objetivo de estimar a sensação térmica na cidade de Santos-SP a partir de variáveis meteorológicas observadas ao longo de um ano. A regressão linear múltipla foi escolhida por sua capacidade de modelar a relação entre uma variável dependente contínua (neste

caso, o Heat Index já convertido para celsius) e múltiplas variáveis independentes (as variáveis meteorológicas).

3.1 Tratamento e Preparação dos Dados

Os dados foram importados em ambiente Google Colab, utilizando a linguagem de programação Python. A manipulação inicial foi realizada com a biblioteca pandas, que permitiu o tratamento de valores ausentes, transformação de colunas de data e hora em índices temporais e filtragem de registros com valores inválidos ou extremos.

As colunas do DataFrame final utilizado na regressão incluíram:

data: data do registro

hora: hora do registro

temperatura: temperatura média (°C)

umidade: umidade relativa do ar (%)

velocidade_vento: velocidade do vento (m/s)

pressao: pressão atmosférica (hPa)

radiacao_solar: radiação solar (W/m²)

heat_index: sensação térmica (°C), calculada com base na fórmula de Steadman (1979)

Para valores de velocidade do vento e radiação solar iguais a zero, como períodos noturnos, adotou-se um valor mínimo artificialmente positivo, visando evitar distorções nos coeficientes da regressão e preservar a significância estatística das variáveis.

A modelagem preditiva da sensação térmica foi conduzida por meio da biblioteca scikit-learn, que fornece ferramentas eficientes para regressão linear múltipla. Esse tipo de regressão permite avaliar o impacto simultâneo de diversas variáveis independentes sobre uma variável dependente — neste caso, o Heat Index, que representa a temperatura percebida pelo corpo humano.

As variáveis independentes selecionadas foram:

Temperatura

Umidade

Velocidade vento

Pressão Atmosférica

Radiação Solar

Essas colunas compõem o conjunto de entrada X, enquanto a variável resposta y corresponde ao Heat Index calculado com base na fórmula de Steadman (1979), já incorporada previamente ao DataFrame. modelo foi ajustado utilizando o objeto LinearRegression() do scikit-learn

Figura 1 – Modelo ajustado utilizando o objeto 'LinearRegression()' e a definição dos conjuntos de entrada X e resposta y.

```
df.dropna(subset=['Temperatura', 'Umidade', 'Velocidade_Vento', 'Pressao', 'Radiacao_Solar'], inplace=True)
X = df[['Temperatura', 'Umidade', 'Velocidade_Vento', 'Pressao', 'Radiacao_Solar']]
y = df['Heat_Index']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression().fit(X_train, y_train)
```

Fonte: elaborado pelo autor via Google Colab

Complementando a modelagem, foi realizada uma análise estatística descritiva tanto do Heat Index calculado quanto do estimado. Os seguintes parâmetros foram avaliados:

Média
Desvio padrão
Valor mínimo
Valor máximo

Essas estatísticas serviram como base para validar a consistência do modelo e a capacidade preditiva frente às variações térmicas observadas ao longo de um ano de dados meteorológicos, cujo foram separados por quartis (trimestrais).

Figura 2 – Resumo Estatístico Trimestral, Heat Index (Fórmula Steadman).

Resumo Estatístico Trimestral:				
Trimestre	Heat_Index			
	mean	std	min	max
2024Q1	24.325884	2.841706	20.5	31.486019
2024Q2	23.662492	4.025677	15.5	37.712252
2024Q3	19.320702	3.228230	9.3	36.073790
2024Q4	23.994500	3.563241	16.8	38.712328
2025Q1	28.430478	5.048004	19.6	43.535789

Fonte: elaborado pelo autor via Google Colab

Figura 3 – Resumo Estatístico Trimestral, Heat Index (Estimado).

Resumo Estatístico Trimestral:				
Trimestre	HI_Estimado			
	mean	std	min	max
2024Q1	24.816434	2.696558	20.519759	31.439242
2024Q2	23.979991	4.261966	14.219452	39.929543
2024Q3	19.141950	3.848150	6.810125	40.804050
2024Q4	24.236952	3.518082	15.870116	38.466994
2025Q1	27.998294	4.006710	19.096772	46.354220

Fonte: elaborado pelo autor via Google Colab

O modelo de regressão múltipla apresentou estimativas próximas dos valores reais em todos os trimestres, com diferenças médias menores que 0.5 °C — o que é um bom indicativo de precisão. A maior diferença ocorre em 2025Q1, quando o modelo subestimou o HI em média 0.44 °C.

A maior variabilidade para ambos os índices ocorre em 2025Q1, indicando um trimestre mais instável em termos de sensação térmica. O modelo foi capaz de reproduzir bem os limites máximos e mínimos de cada trimestre.

Essas tabelas indicam que o modelo de regressão linear múltipla obteve um bom desempenho preditivo:

- Médias trimestrais próximas entre os valores reais e estimados.
- Desvios padrão similares, refletindo variações sazonais coerentes.

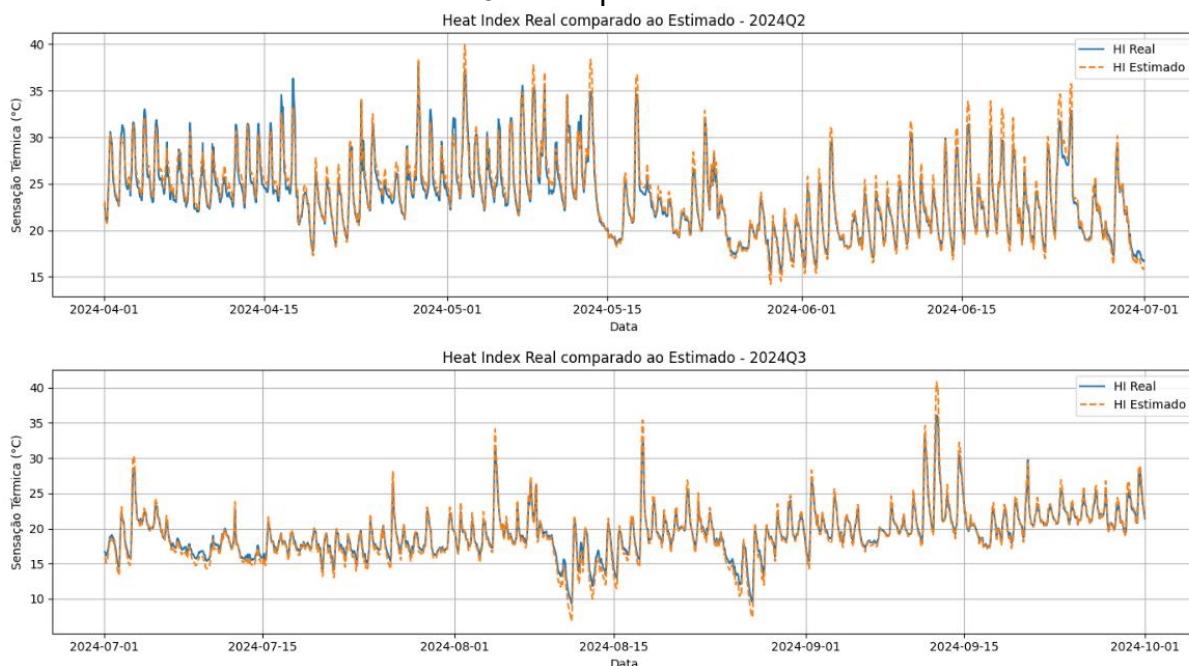
Pequenas diferenças nos extremos que podem ser associadas ao aumento da sensação de calor por variações de estações.

O bom alinhamento reforça que os preditores escolhidos (temperatura, umidade, vento, radiação, pressão) são relevantes para estimar a sensação térmica em Santos.

4 ANÁLISE TEMPORAL HEAT INDEX (STEADMAN) COMPARADO AO ESTIMADO

Para a visualização dos dados e validação gráfica da modelagem estatística, foi utilizada a biblioteca Matplotlib, amplamente empregada em ambientes científicos para criação de gráficos estáticos e comparações temporais. Com essa ferramenta, foram gerados gráficos de linhas que comparam, ao longo dos diferentes trimestres, os valores de Steadman da sensação térmica (Heat Index) com os valores estimados por meio da regressão linear múltipla. A visualização permitiu identificar o comportamento sazonal do índice, bem como avaliar visualmente a precisão da estimativa estatística aplicada.

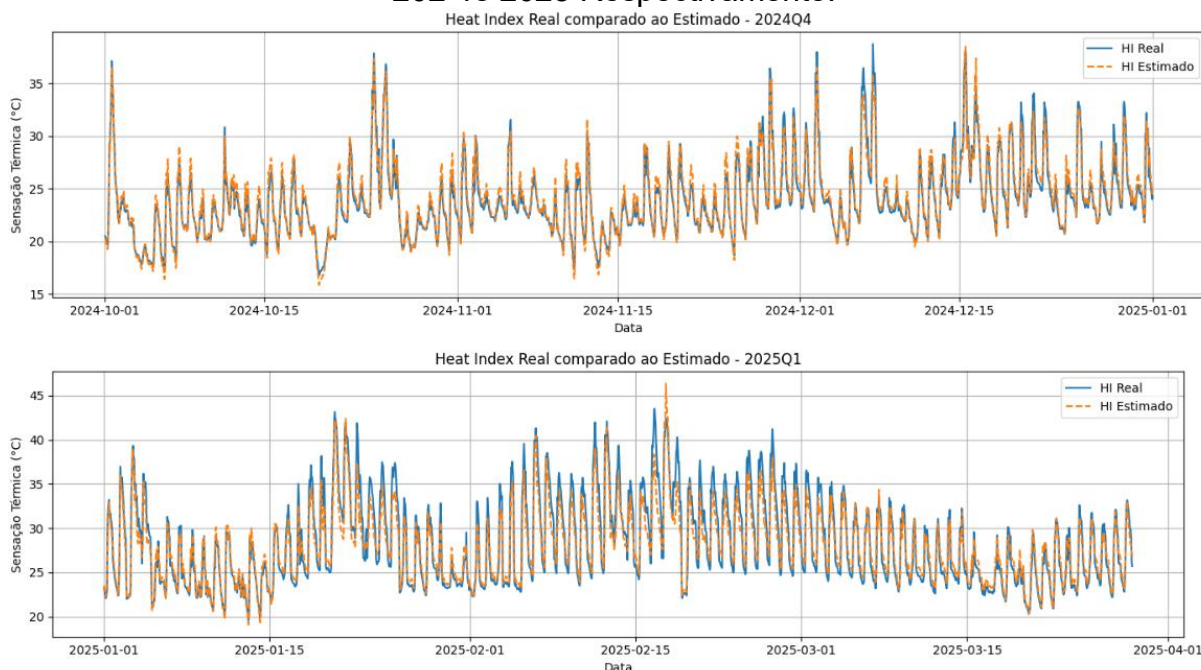
Figura 4 - Gráfico de Linha Temporal do Heat Index, do Segundo e Terceiro Quartil de 2024 Respectivamente.



Fonte: elaborado pelo autor via Google Colab

Observa-se que nos trimestres 2024Q2 e 2024Q4 (Figura 4), a estimativa do modelo acompanha de forma bastante próxima os picos e vales do índice de Steadman, demonstrando boa aderência do modelo, mesmo durante períodos de variações abruptas de temperatura.

Figura 5 - Gráfico de Linha Temporal do Heat Index, do Quarto e Primeiro Quartil de 2024 e 2025 Respectivamente.



Fonte: elaborado pelo autor via Google Colab

Na Figura 5, referente aos trimestres 2024Q3 (inverno) e 2025Q1 (verão), é possível identificar um desempenho ainda mais eficiente no verão de 2025, com as curvas quase sobrepostas na maioria dos dias. Já no trimestre 2024Q3, embora o modelo ainda siga as tendências gerais da curva real, ocorrem maiores discrepâncias em momentos de queda acentuada do Heat Index, evidenciando a sensibilidade do modelo a variações mais bruscas em condições menos extremas de calor.

As Figuras 4 e 5 reforçam a validade do modelo de regressão linear múltipla para estimar a sensação térmica em Santos, com alta correlação visual entre os valores reais e os previstos. O modelo mostra maior acurácia durante períodos de maior calor (verão), o que é consistente com a lógica do Heat Index de Steadman, cujo é mais sensível a altas temperaturas e alta umidade.

As discrepâncias mais acentuadas observadas no inverno (2024Q3) podem ser atribuídas a fatores como menor impacto da umidade e da radiação solar no cálculo do Heat Index, bem como à maior variabilidade de outros fatores meteorológicos.

Tais evidências sugerem que o modelo se comporta de maneira robusta na maioria das estações, e pode ser uma ferramenta confiável para estudos de conforto térmico e planejamento urbano em contextos tropicais úmidos.

4.1 Análise dos coeficientes trimestrais

Para a modelagem preditiva da sensação térmica, foi utilizada a biblioteca Scikit-learn, que disponibiliza ferramentas eficientes de regressão linear múltipla, permitindo calcular os coeficientes estimados para cada variável independente. A modelagem visou avaliar a influência conjunta de fatores meteorológicos por cada quartil (trimestre). A análise retornou duas métricas fundamentais: o R^2 (coeficiente de determinação) e o *Root Mean Squared Error* (RMSE).

A função de R^2 indica a proporção da variabilidade da sensação térmica que é explicada pelas variáveis preditoras; quanto mais próximo de 1, maior o poder explicativo do modelo.

Enquanto a função RMSE traduzindo para Raiz do Erro Quadrático Médio, medido em graus Celsius. Cujo representa a média dos erros entre os valores reais e estimados. No contexto da análise, é quando a sensação térmica estimada se distancia da real. Essas métricas são essenciais para captar como o corpo humano percebe o calor quando múltiplos fatores além da temperatura são considerados.

Figura 6 - Saída dos Coeficientes Trimestrais do Segundo e Terceiro quartil de 2024 respectivamente.

Trimestre: 2024Q2		Trimestre: 2024Q3	
R²: 0.9725 RMSE: 0.6175 °C		R²: 0.9976 RMSE: 0.1590 °C	
Coeficientes:		Coeficientes:	
Temperatura	= 1.0969	Temperatura	= 0.9984
Umidade	= 0.0165	Umidade	= 0.0019
Velocidade_Vento	= 0.1448	Velocidade_Vento	= 0.0059
Pressao	= -0.0227	Pressao	= -0.0018
Radiacao_Solar	= 0.0016	Radiacao_Solar	= 0.0002

Fonte: elaborado pelo autor via Google Colab

No trimestre 2024Q2, observou-se um excelente ajuste do modelo, com $R^2 = 0,9725$ e um RMSE de apenas $0,6175^{\circ}\text{C}$, indicando que o modelo conseguiu estimar com precisão a sensação térmica com base nas variáveis envolvidas. O coeficiente da temperatura foi o mais expressivo (1,0969), seguido pela leve influência da umidade (0,0165) e uma pequena contribuição da velocidade do vento e pressão atmosférica.

Já no trimestre 2024Q3, o desempenho foi ainda superior, com $R^2 = 0,9976$ e um RMSE extremamente baixo de $0,1590^{\circ}\text{C}$, o que evidencia uma quase perfeita correspondência entre os valores estimados e reais. O coeficiente da temperatura ficou próximo de 1 (0,9984), sugerindo dominância dessa variável, enquanto os demais fatores mantiveram baixa influência.

Figura 7 - Saída dos Coeficientes Trimestrais do Quarto e Primeiro quartil de 2024 e 2025 respectivamente.

Trimestre: 2024Q4		Trimestre: 2025Q1	
R²: 0.9673 RMSE: 0.6555 °C		R²: 0.9455 RMSE: 1.1198 °C	
Coeficientes:		Coeficientes:	
Temperatura	= 1.1831	Temperatura	= 1.6912
Umidade	= 0.0030	Umidade	= 0.0709
Velocidade_Vento	= -0.1932	Velocidade_Vento	= 0.3309
Pressao	= -0.0099	Pressao	= 0.0386
Radiacao_Solar	= 0.0010	Radiacao_Solar	= 0.0007

Fonte: elaborado pelo autor via Google Colab

Para o trimestre 2024Q4, o modelo continuou apresentando bons resultados com $R^2 = 0,9673$ e $RMSE = 0,6555^{\circ}C$. A temperatura manteve papel predominante (coef. = 1,1831), mas as demais variáveis também mostraram influência significativa, sobretudo a pressão atmosférica e a radiação solar com coeficientes negativos, o que indica possível efeito de atenuação do Índice de calor.

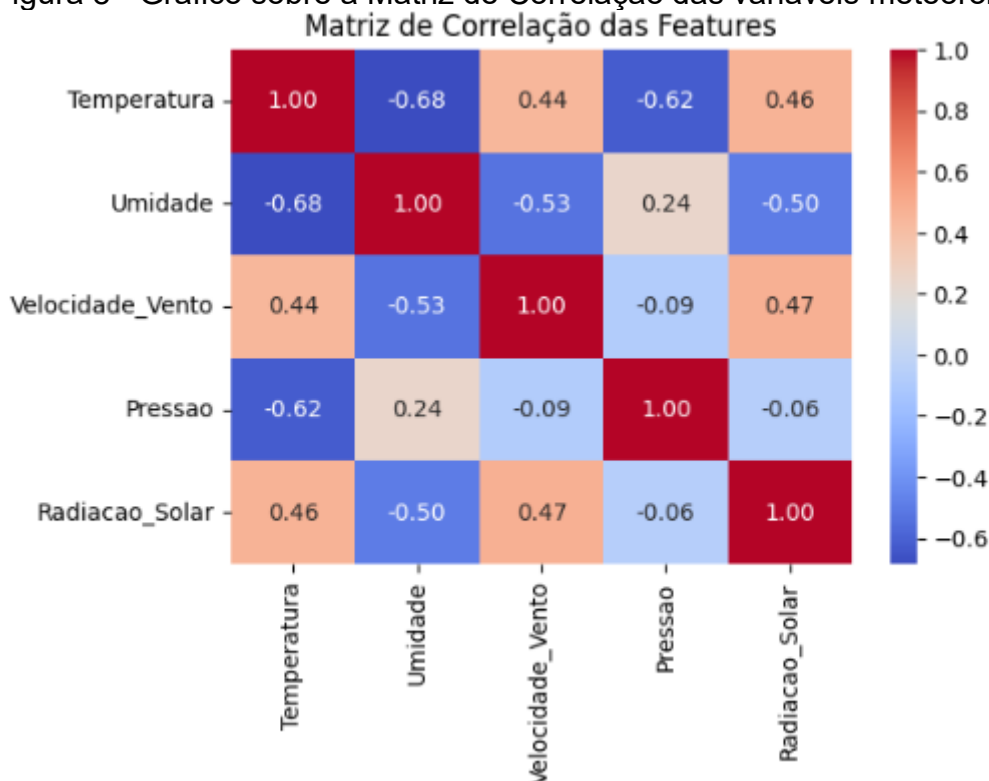
No trimestre 2025Q1, embora o modelo continue com desempenho elevado ($R^2 = 0,9455$), o RMSE subiu para 1,1198°C, sugerindo maior quantidade de variações extremas desse período. Isso pode estar relacionado a maior amplitude térmica no trimestre. O coeficiente de temperatura aumentou para 1,6912, refletindo a maior dominância desse fator específico nesse período. A umidade também apresentou maior participação (0,0709), sugerindo que o desconforto térmico pode estar mais relacionado à combinação de calor e umidade nos meses iniciais do ano, durante o verão.

4.2 Análise da Matriz de Correlação

Para examinar a relação entre as variáveis meteorológicas envolvidas na estimativa da sensação térmica, foi utilizada a biblioteca Seaborn, ferramenta baseada no Matplotlib que facilita a criação de visualizações estatísticas em Python. Especificamente, foi empregada uma matriz de correlação (heatmap), representando visualmente os coeficientes de correlação de Pearson (2014), entre as variáveis preditoras: temperatura, umidade, velocidade do vento, pressão atmosférica e radiação solar.

O coeficiente de correlação de Pearson varia de -1 a +1: valores próximos de +1 indicam forte correlação positiva (ambas variáveis aumentam juntas), enquanto valores próximos de -1 indicam forte correlação negativa (uma variável aumenta enquanto a outra diminui). Esse tipo de análise pode indicar possíveis multicolinearidades no modelo de regressão linear, além de fornecer insights sobre os fenômenos ambientais.

Figura 8 - Gráfico sobre a Matriz de Correlação das variáveis meteorológicas.



Fonte: elaborado pelo autor via Google Colab

A Figura 8 mostra a matriz de correlação entre as cinco variáveis climáticas utilizadas como preditoras do Heat Index. A temperatura apresenta correlação negativa moderada com a umidade (-0,68), indicando que, em geral, períodos mais quentes tendem a ser menos úmidos.

A temperatura também apresenta correlação moderada e positiva com a velocidade do vento (+0,44) e com a radiação solar (+0,46), sugerindo que dias mais quentes tendem a coincidir com maior incidência solar e movimento de ar, o que é coerente com padrões diurnos e possíveis condições de céu limpo. Há ainda correlação negativa com a pressão atmosférica (-0,62), o que indica que temperaturas elevadas ocorrem com pressões relativamente mais baixas.

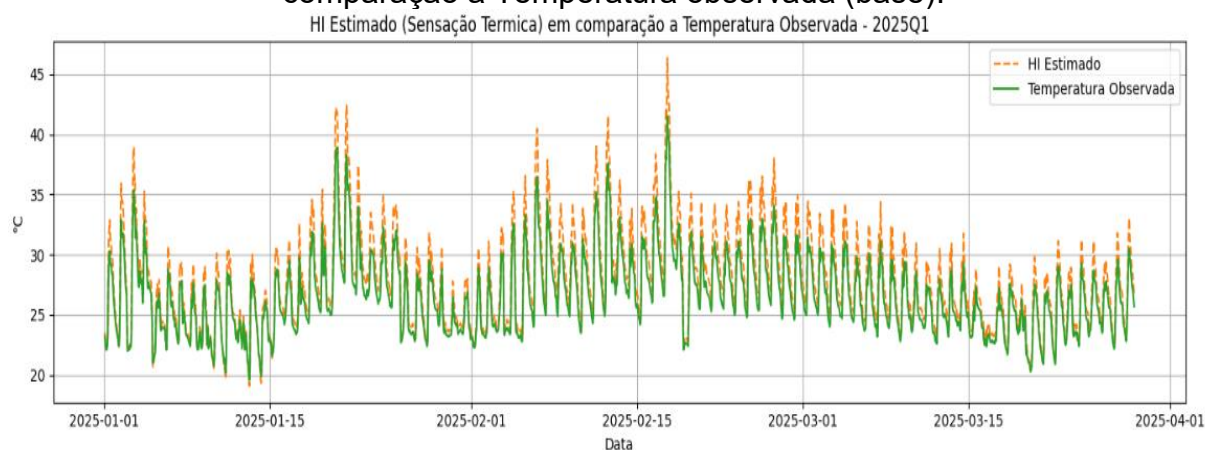
A umidade exibe correlações negativas com quase todas as variáveis: velocidade do vento (-0,53), radiação solar (-0,50) e temperatura (-0,68), sendo coerente com a ideia de que massas de ar seco tendem a dominar em condições ensolaradas e ventiladas. A única correlação positiva leve da umidade é com a pressão (+0,24), reforçando o papel de massas de ar úmido associadas a sistemas de alta pressão.

4.3 Análise do Índice de Calor estimado (Sensação Térmica) comparado a temperatura observada

Para avaliar a eficácia do modelo de regressão linear na estimativa da sensação térmica (Heat Index - HI), foi utilizada a biblioteca Matplotlib, por meio do módulo pyplot (acessado via plt), para plotar comparativamente os valores observados de temperatura e os valores estimados de HI ao longo do tempo.

Neste contexto, a visualização permite verificar o comportamento dinâmico da variável alvo (HI estimado) frente à variável base (temperatura observada), evidenciando as flutuações térmicas que não são perceptíveis apenas com base na temperatura do ar, ao considerar os efeitos combinados de umidade, radiação solar, vento e pressão atmosférica.

Figura 9 - Gráfico de Linha Temporal do da Sensação Térmica estimada em comparação a Temperatura observada (base).



Fonte: elaborado pelo autor via Google Colab

A Figura 9 apresenta a comparação entre a temperatura observada (linha verde contínua) e o Heat Index estimado (linha laranja tracejada) ao longo do primeiro trimestre de 2025 (2025Q1), que corresponde à estação do verão no hemisfério sul. Observa-se que, consistentemente, o HI estimado é superior à temperatura do ar em praticamente todos os dias do período analisado.

Esse comportamento era hipoteticamente esperado, já que o índice de calor é justamente uma medida que leva em consideração não apenas a temperatura, mas também o impacto da umidade relativa do ar, da radiação solar, da velocidade do vento e da pressão atmosférica sobre o conforto térmico humano. A influência da umidade é especialmente importante no verão, pois, temperaturas mais altas diminui a umidade do ar respectivamente, acentuando a sensação de calor percebida.

A curva de HI apresenta picos mais pronunciados durante os períodos de maior incidência solar e estabilidade atmosférica, sugerindo que a combinação dos fatores meteorológicos atua sinergicamente para intensificar o desconforto térmico. Isso justifica os maiores valores de RMSE registrados nesse trimestre, como observado nos resultados do modelo (Figura 7), e comprova que a sensação térmica não pode ser explicada apenas pela temperatura.

A comparação entre a temperatura observada e o índice de calor estimado evidencia que o verão apresenta as condições climáticas mais críticas do ponto de vista do conforto térmico. O modelo captura essa tendência ao incorporar variáveis adicionais além da temperatura, o que demonstra a validade da abordagem

multivariada para avaliar com mais precisão os impactos do calor sobre o corpo humano.

5 CONSIDERAÇÕES FINAIS

A comparação entre a temperatura observada e o índice de calor estimado evidencia que o verão apresenta as condições climáticas mais críticas do ponto de vista do conforto térmico. O modelo captura essa tendência ao incorporar variáveis adicionais além da temperatura, o que demonstra a validade da abordagem multivariada para avaliar com mais precisão os impactos do calor sobre o corpo humano.

Os coeficientes obtidos a cada trimestre permitiram observar a variação da influência das variáveis ao longo do tempo. No segundo (Q2) e terceiro trimestres (Q3), observou-se maior precisão no modelo ($R^2 > 0,97$), com baixa margem de erro ($RMSE < 0,2 \text{ }^\circ\text{C}$), o que evidencia maior regularidade e correlação entre os fatores climáticos nesse período. Já no verão (Q1), a elevação do RMSE ($\approx 1,12 \text{ }^\circ\text{C}$) e a intensificação dos coeficientes de temperatura e umidade indicam que esse trimestre apresenta maior variabilidade e impacto térmico, destacando-se como o mais crítico em termos de desconforto térmico.

A matriz de correlação entre variáveis demonstrou que temperatura e umidade são as principais influências no modelo, seguidas por velocidade do vento e radiação solar, que também apresentaram correlações moderadas entre si. Evidenciando que, em ambientes mais quentes, a umidade relativa tende a ser menor, impactando diretamente na sensação térmica percebida. A temperatura apresentou correlação positiva com velocidade do vento ($r = 0,44$) e radiação solar ($r = 0,46$), enquanto a pressão atmosférica teve correlação negativa com praticamente todas as variáveis, especialmente com a temperatura ($r = -0,62$). Isso indica que dias mais quentes, com maior radiação e menor pressão, tendem a coincidir com condições de maior desconforto térmico.

A comparação gráfica entre o HI estimado e a temperatura observada reforçou a hipótese de que a sensação térmica é sistematicamente superior à temperatura real durante o verão. Isso demonstra a limitação de análises baseadas exclusivamente na temperatura do ar, sobretudo em contextos de planejamento urbano, saúde pública e prevenção de estresse térmico em populações vulneráveis, principalmente em regiões de clima equatorial como Santos.

REFERÊNCIAS

- GOOGLE. **Google Colaboratory**. [software]. Disponível em: <https://colab.research.google.com/>. Acesso em: 28 mai. 2025.
- MATPLOTLIB DEVELOPMENT TEAM. **Matplotlib: comprehensive library for creating static, animated, and interactive visualizations in Python**. [software]. Disponível em: <https://matplotlib.org/>. Acesso em: 22 nov. 2024.
- MONTGOMERY, Douglas C.; RUNGER, George C. **Applied statistics and probability for engineers**. 7. ed. Hoboken: Wiley, 2018.
- NUMPY DEVELOPERS. **NumPy: Fundamental package for numerical computations in Python**. [software]. Disponível em: <https://numpy.org/>. Acesso em: 28 mai. 2025.
- PARSONS, Ken. **Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort and performance**. 3. ed. Boca Raton: CRC Press, 2014.
- PEEL, M. C.; FINLAYSON, B. L.; MCMAHON, T. A. **Updated world map of the Köppen-Geiger climate classification**. *Hydrology and Earth System Sciences*, [S. l.], v. 11, n. 5, p. 1633–1644, 2007. DOI: <https://doi.org/10.5194/hess-11-1633-2007>. Acesso em: 5 jun. 2025.
- PYTHON SOFTWARE FOUNDATION. **Python**: versão 3.12. [software]. Disponível em: <https://www.python.org/>. Acesso em: 28 mai. 2025.
- WASKOM, M. L. **Seaborn: statistical data visualization**. *Journal of Open Source Software*, v. 6, n. 60, p. 3021, 2021. [software]. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 28 mai. 2025.
- SCIKIT-LEARN DEVELOPERS. **Scikit-learn: Machine Learning in Python**. [software]. Disponível em: <https://scikit-learn.org/>. Acesso em: 28 mai. 2025.
- STEADMAN, Robert G. **The assessment of sultriness: Part I. A temperature-humidity index based on human physiology and clothing science**. *Journal of Applied Meteorology*, Boston, v. 18, n. 7, p. 861–873, 1979.
- THE PANDAS DEVELOPMENT TEAM. **Pandas: Python Data Analysis Library**. [software]. Disponível em: <https://pandas.pydata.org/>. Acesso em: 28 mai. 2025.
- WILKS, D. S. **Statistical methods in the atmospheric sciences**. 4. ed. Cambridge: Academic Press, 2019. (International Geophysics Series, v. 100).