

# **Relatório Técnico - Equipe Salvador**

**Autor(es):** João Vítor Da Conceição de Almeida, André Almeida Gomes Neto, Deivid Souza Dos Santos Oliveira.

Instituição: Centro Universitário de Excelência / Sistemas de Informação

23/08/2025

## **Abstract**

This study developed a statistical metrics application using the Python programming language. The goal was to implement a recommendation model for food delivery app users, using statistics as a basis. Statistics is the branch of mathematics that aims to gather data on a given phenomenon or population, thus relating facts and figures. Based on these factors, the code was implemented, performing validations and testing to ensure the algorithm's proper functioning.

**Keywords:** Statistics, Python, Data Science, Recommendation, FoodDelivery

## **Resumo**

Neste estudo foi desenvolvido aplicação de métricas estatísticas utilizando a linguagem de programação Python. Com o objetivo de implementar um modelo de recomendações para os usuários do app food Delivery, utilizando-se como base a Estatística que é a área da matemática que tem por objetivo levantar dados de determinado fenômeno ou população, assim, relacionando fatos e números. A partir destes fatores foi realizado a implementação do código, realizando validações e testes para o devido funcionamento do algoritmo.

**Palavras-chave:** Estatística, Python, Ciência de Dados, Recomendação, FoodDelivery

## **1. Introdução**

O avanço da ciência de dados impulsionou e possibilitou transformar a maneira como as empresas ofertam serviços ou produtos personalizados aos seus respectivos clientes. Para que seja possível ter uma maior assertividade em atrair um usuário à finalizar um pedido, traçar padrões para identificar o seu perfil é de extrema importância. A FoodDelivery como uma plataforma, utiliza de funções estatísticas e computacionais para obter métricas e a partir delas gerar recomendações de acordo ao perfil de cada usuário.

## 2. Referencial Teórico

A estatística fornece os instrumentos matemáticos necessários para a análise, interpretação e síntese de dados, sendo fundamental para a ciência de dados e o desenvolvimento de sistemas de recomendação. A seguir, são descritas as principais métricas implementadas no presente trabalho, com suas definições e aplicações.

### 2.1 Média Aritmética

Fórmula: 
$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

A média aritmética representa a soma de todos os valores dividida pela quantidade de elementos. É usada como uma medida central dos dados.

### 2.2 Mediana

A mediana corresponde ao valor central de um conjunto de dados ordenado. É uma medida de tendência central menos sensível a valores extremos.

### 2.3 Moda

A moda é o valor que ocorre com maior frequência em um conjunto de dados. Pode haver uma ou mais modas em uma distribuição.

### 2.4 Variância

Fórmula: 
$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

A variância mede o quanto os dados se afastam da média, sendo uma medida de dispersão. Quanto maior a variância, mais espalhados estão os valores.

### 2.5 Desvio Padrão

Fórmula: 
$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$$

O desvio padrão é a medida de dispersão que indica o quanto, em média, os valores se afastam da média. É amplamente utilizado por estar na mesma unidade dos dados originais.

### 2.6 Covariância

Fórmula: 
$$\text{cov}(X, Y) = \frac{\sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)}{N}$$

A covariância avalia o grau de associação entre duas variáveis. Quando positiva, indica que as variáveis tendem a crescer juntas; quando negativa, indica que uma tende a crescer enquanto a outra diminui.

### 2.7 Conjunto de Itens Únicos (Itemset)

O itemset corresponde ao conjunto de valores distintos presentes em uma variável. É útil para identificar categorias e preparar dados para algoritmos de mineração.

## **2.8 Frequência Absoluta**

A frequência absoluta indica quantas vezes cada valor aparece em um conjunto de dados. É uma das formas mais simples de análise descritiva.

## **2.9 Frequência Relativa**

A frequência relativa mostra a proporção de ocorrência de cada valor em relação ao total, permitindo comparações mais justas entre diferentes conjuntos de dados.

## **2.10 Frequência Acumulada**

A frequência acumulada corresponde à soma progressiva das frequências, seja absoluta ou relativa. É utilizada para identificar percentis e entender a distribuição dos dados.

## **2.11 Probabilidade Condicional**

A probabilidade condicional mede a chance de um evento ocorrer dado que outro já ocorreu. Essa métrica é essencial em modelos probabilísticos e em regras de associação.

# **3. Metodologia**

A metodologia adotada baseou-se nos seguintes passos:

1. **Definição da estrutura de dados:** o conjunto de dados foi representado como um dicionário em Python, no qual cada chave corresponde a uma coluna e cada lista, aos valores dessa coluna.
2. **Validação da entrada:** garantiu-se que todas as chaves possuíam listas do mesmo tamanho e que cada coluna apresentava dados homogêneos em termos de tipo.
3. **Implementação das funções estatísticas:** foram implementadas funções para média, mediana, moda, variância, desvio padrão e média ponderada.
4. **Testes de consistência:** os resultados das funções implementadas foram
5. Validados através de testes unitários para garantir as métricas estabelecidas atendam as funções estabelecidas
6. **Versionamento e colaboração:** todo o código foi desenvolvido e versionado em repositório Git, conforme o modelo de workflow centralizado proposto.

# **4. Resultados e Discussão**

A implementação das métricas demonstrou resultados consistentes, um dataset de teste contendo valores numéricos simples. Além disso, o processo de implementação permite compreender a lógica por trás de cálculos.

## 5. Conclusão

Este relatório apresentou a primeira etapa do programa de trainee da FoodDelivery, centrada na implementação de métricas estatísticas em Python. A atividade reforçou conceitos fundamentais de estatística e programação, além de estimular práticas de versionamento e colaboração em repositórios compartilhados.

Como próximos passos, pretende-se aplicar esses conceitos no desenvolvimento de modelos de recomendação baseados em regras de associação e, posteriormente, na segmentação de clientes. Entre os pontos mais desafiadores, destacaram-se a implementação utilizando apenas os recursos nativos da linguagem e a validação dos tipos de dados.

A experiência evidenciou a importância de compreender o funcionamento interno dos algoritmos antes de utilizar bibliotecas prontas, o que contribui para a formação de profissionais mais críticos e capazes de resolver problemas complexos.

## Referências

[1] OLIVEIRA, Francisco Estevam Martins de. **Estatística e Probabilidade - Exercícios Resolvidos e Propostos, 3ª edição**. Rio de Janeiro: LTC, 2017. *E-book*. pág.108. ISBN 9788521633846

[2] **MENEZES, Nilo Ney Coutinho**. *Introdução à programação com Python*. 4. ed. São Paulo: Novatec, 2024. ISBN 978-85-7522-886-9.