

Mineração de Dados – Estatística Básica aplicada à FoodDelivery

Autores: Anna Beatriz Lima
Ian Carneiro
Valéria Santos

Agenda

Apresentar a implementação de operações estatísticas básicas em Python, desenvolvidas sem bibliotecas externas, e demonstrar sua aplicação prática no projeto FoodDelivery .

1. Introdução e Objetivos:

Importância da estatística na mineração de dados e o propósito do projeto.

2. Fundamentação Teórica:

Breve explicação das principais operações estatísticas.

3. Implementação da

Classe *Statistics*: Detalhes da classe e demonstração das principais funções.

4. Metodologia: Estratégias e ferramentas usadas na implementação.

5. Resultados e Desafios:


Avaliação da implementação e desafios enfrentados.

6. Conclusão:

Aprendizados obtidos e os próximos passos na disciplina de Mineração de Dados.

Introdução e Objetivos



- **Dados**  recurso estratégico (“petróleo do século XXI”).
- A **Mineração de dados** transforma dados brutos em informação útil.
- **Estatística** é a base para organizar, interpretar e gerar conhecimento.
- **Nosso objetivo:**
 - Implementar funções estatísticas básicas em Python sem bibliotecas externas.
 - Aplicar no cenário simulado de Algoritmo de Recomendação da FoodDelivery.

Mineração de Dados – Estatística
Básica aplicada à FoodDelivery

Fundamentação Teórica

Operações estatísticas:

- **Média:** valor central (ex: gasto médio do cliente).
- **Moda:** valor mais frequente (ex: prato mais pedido).
- **Mediana:** ponto central da distribuição.
- **Variância e Desvio Padrão:** variação de dados (ex: diferença entre gastos de clientes)
- **Covariância:** relação entre duas variáveis (ex: idade \times gasto).
- **Frequências:**
 - **Absoluta** \rightarrow contagem direta
 - **Relativa** \rightarrow proporção em relação ao total
 - **Acumulada** \rightarrow Soma progressiva das ocorrências
- **Itens Únicos:** lista de valores distintos (ex: quantos pratos diferentes foram pedidos)
- **Probabilidade Condicional:** chance de um evento acontecer dado que outro já ocorreu (ex: chance de pedir sobremesa depois de pedir um prato principal)

agui
talvez
tivesse
tido
logo
mostrar
sobre
métrica
de
dispersão,
análise
descritiva,
etc

Implementação da Classe Statistics



Estrutura geral:

- Classe Statistics recebe dataset em forma de dicionário.
- Cada operação é implementada como método independente.
- Validações: colunas com mesmo tamanho, listas vazias, múltiplas modas.

Demonstração detalhada das principais funções:

- **Média:** soma dos valores \div quantidade (ex: gasto médio).
- **Covariância:** relação entre duas variáveis (ex: idade \times gasto médio).
- **Probabilidade Condicional:** chance de um valor ocorrer dado que outro ocorreu antes (ex: probabilidade de um prato ser pedido após outro prato).

Exemplo de implementação

unex

```
1 def absolute_frequency(self, column):
2     self._validate_column(column)
3     data = self.dataset[column]
4
5     if data == []:
6         return {}
7
8     absolute_frequency = {}
9
10    for value in data:
11        if value not in absolute_frequency:
12            absolute_frequency[value] = 0
13
14            absolute_frequency[value] += 1
15
16    return absolute_frequency
17
```

era bom
colocar
a eq.
do
lado
de imple-
mentação

Metodologia



- **Trabalho colaborativo** → cada integrante responsável por parte das funções.
- **Revisão cruzada** entre colegas.
- **Ferramentas:**
 - Python puro (sem numpy/pandas).
 - GitHub para versionamento.
 - Arquivo tests.py fornecido pelo professor para validação.

Resultados e Desafios



Resultados:

- Todas as operações passaram nos testes.
- Resultados condizentes com a teoria.

Desafios:

- Implementar sem bibliotecas externas.
- Tratar listas vazias e múltiplas modas.
- Garantir clareza do código para todos os integrantes.

Conclusão



- **Estatística básica** é pilar da mineração de dados.
- O projeto reforçou lógica, manipulação de dados e aplicação prática da teoria.
- A FoodDelivery mostrou como métricas simples já dão insights úteis (ex: recomendar pratos).

Referências



- **SAS INSTITUTE INC.** O que é mineração de dados? SAS, 2025. Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html. Acesso em: 21 ago. 2025.
- **AMAZON WEB SERVICES, INC.** O que é mineração de dados? AWS, 2025. Disponível em: <https://aws.amazon.com/pt/what-is/data-mining/>. Acesso em: 21 ago. 2025.
- **DATA CAMP.** Média aritmética. DataCamp, 2025. Disponível em: <https://www.datacamp.com/pt/tutorial/arithmetic-mean>. Acesso em: 21 ago. 2025.
- **TEACHY.** Moda – Medida de tendência central. Teachy, 2025. Disponível em: <https://www.teachy.com.br/resumos/ensino-fundamental/8ano/matematica/estatistica-moda-compreendendo-e-aplicando-a-medida-de-tendencia-central-Expositiva>. Acesso em: 21 ago. 2025.
- **KHAN ACADEMY.** Média, mediana e moda. Khan Academy, 2025. Disponível em: <https://pt-pt.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/mean-median-basics/a/mean-median-and-mode-review>. Acesso em: 21 ago. 2025.
- **SUNO.** O que é covariância? Suno, 2025. Disponível em: <https://www.suno.com.br/artigos/covariancia/>. Acesso em: 21 ago. 2025.
- **DATA CAMP.** Covariância. DataCamp, 2025. Disponível em: <https://www.datacamp.com/pt/tutorial/covariance>. Acesso em: 21 ago. 2025.
- **MATEMÁTICA HOJE.** Frequência absoluta, relativa e acumulada. Matemática Hoje, 2025. Disponível em: <https://matematicahoje.blog/frequencia-absoluta-relativa-acumulada-e-relativa-acumulada/>. Acesso em: 21 ago. 2025.
- **ESTATÍSTICA FÁCIL.** Frequência acumulada. Estatística Fácil, 2025. Disponível em: <https://estatisticafacil.org/glossario/o-que-e-frequencia-acumulada/>. Acesso em: 21 ago. 2025.

A apresentação foi feita dentro do tempo estabelecido.

Fui gostei bastante do encaideamento das ideias, mas acredito que alguns pontos precisam ser considerados:

- Senti falta de um aprofundamento sobre as métricas de estatística e suas aplicações. Definir quais eram métricas de tendência central, dispersão, dentre outras.
- Seria interessante correlacionar o código desenvolvido com a equação matemática.
- acho que nos resultados poderia ser apresentado os testes passando.

Sobre as perguntas feitas:

- Senti a apresentadora um pouco insegura quanto as métricas
- A equipe conseguiu elaborar, mas talvez a divisão do trabalho com uma revisão detalhada pode nos trazer uma base clara.

Sobre a apresentação geral:

- Precisa de mais imagens
- exemplificações!!!

O peso da apresentação é:

2,3 pontos