

Dendê Statistics

Autor: Bruno Sampaio Silva
Levi Falcão de Queiroz
Moisés de Souza Oliveira
Raphael dos Santos Moreira

Agenda

O objetivo desta apresentação é demonstrar as métricas de estatísticas fundamentais para extrair dados relevantes para uma empresa ou segmento e como elas foram implementadas em python

1. Introdução:

Importância da análise de dados e apresentação da biblioteca

2. Média Aritmética:

Demonstração da métrica de tendência central

3. Frequência:

Explicação das medidas de frequência absoluta e relativa

4. Moda:

Explicação de moda e como foi implementado

5. Spotify (AED):

Análise exploratório do dataset, utilizando a biblioteca criada

6. Considerações Finais:

Agradecimentos e o que poderia ter sido feito, e problemas

Introdução



- Volumes massivos de informação gerados globalmente
- Objetivo da AED
- Biblioteca Dendê Statistics

- Média Aritmética:
 - Medida de Tendência Central
 - Fórmula: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
 - Tratar entrada de dados
 - Apenas Números

```
Title
1  def mean(self, column):
2      if column not in self.dataset:
3          return None
4
5      dados = self.dataset[column]
6
7      quantidade = len(dados)
8      if quantidade == 0:
9          return 0
10
11     if not isinstance(dados[0], (int, float)):
12         return None
```

Média

```
Title

1     soma_total = 0
2     for valor in dados:
3         soma_total += valor
4
5     if quantidade == 0:
6         return 0
7
8     mean = soma_total / quantidade
9     return mean
```

Frequência

- Frequência Absoluta:
 - Medida de frequência
 - Não tem uma fórmula
 - Números e Categorias

```
Title
1  def absolute_frequency(self, column):
2      if column not in self.dataset:
3          return {}
4
5      data_column = self.dataset[column]
6      frequency = {}
7
8      for value in data_column:
9          if value in frequency:
10             frequency[value] += 1
11         else:
12             frequency[value] = 1
13
14     return frequency
```

Frequência

- Frequência Relativa:
 - Medida de frequência
 - Fórmula: $f_r = \frac{f_i}{n}$
 - Números e Categorias

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

Moda

- Moda:
 - Medida de Tendência Central
 - Elemento que mais se repete
 - Retorna uma lista caso empate

```
def mode(self, column):  
    frequencias = self.absolute_frequency(column)  
  
    if not frequencias:  
        return []  
  
    maior_frequencia = max(frequencias.values())  
  
    modas = []  
    for valor, qtd in frequencias.items():  
        if qtd == maior_frequencia:  
            modas.append(valor)  
  
    return modas
```


Spotify (AED)

- Usuário com atenção curta
- Artista mais ouvida: Taylor Swift
- Catálogo Family Friend
- Recomendação acima de 71 de nota
- Albuns completos Predominam

[Duração da Faixa (Minutos)]

Média: 3.49

Mediana: 3.45

[Popularidade da Faixa (0 a 100)]

Q1: 39.0

Q2: 58.0 (Mediana)

Q3: 71.0

[Seguidores do Artista]

Média: 24034719.33

Desvio Padrão: 38029589.10

[Covariância]

Fama do Artista x Sucesso da Música: 218.43

[Nome do Artista]

Moda: Taylor Swift

Frequência Absoluta (da Moda): 324

[Tipo de Álbum]

Moda: album

Freq. Relativa ('album'): 0.6824 (68.24%)

[Conteúdo Explícito (True/False)]

Freq. Relativa ('TRUE'): 0.2503 (25.03%)

Freq. Relativa ('FALSE'): 0.7497 (74.97%)

Considerações Finais



- Tratamento de dados nos métodos
- Lidar com tipos mistos na coluna do spotify
- Teste de código, principal na condicional e mediana

Referências

- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Montgomery, D. C.; Runger, G. C. (2014). *Applied Statistics and Probability for Engineers*. 6th ed. Wiley.
- Ross, S. M. (2014). *A First Course in Probability*. 9th ed. Pearson.