

Programação em Python para Data Science

Mariana P Silva

full backend developer



Formada em Análise e Desenvolvimento de Sistemas na FATEC de Cruzeiro-SP.

Cursando Pós em Software Architecting na FIAP. Completando 5 anos de experiência na área em 2025.

Instrutor de Formação Profissional III no SENAI (desde 2022).

Backend Developer na Venturus (desde 2022).

A close-up photograph of a person's hand holding a purple marker, drawing on a whiteboard. The background is blurred, showing some bokeh lights. The text 'Chamada Interativa' is overlaid on the left side of the image.

Chamada Interativa

1. Qual seu maior defeito;
2. Qual sua maior qualidade?
3. Qual o maior desafio profissional que marcou sua carreira?
4. Descreva sua personalidade.

Regras

[link](#)



.....

2569 5413

menti.com

Sobre o Curso

Carga Horária: 60h

Objetivo: O desenvolvimento de competências relativas programação utilizando codificações na linguagem Python, com vistas à coleta, limpeza, organização e transformação de dados e seguindo boas práticas, procedimentos e normas.

Calendário

AUGUST							SEPTEMBER						
S	M	T	W	T	F	S	S	M	T	W	T	F	S
					1	2		1	2	3	4	5	6
3	4	5	6	7	8	9	7	8	9	10	11	12	13
10	11	12	13	14	15	16	14	15	16	17	18	19	20
17	18	19	20	21	22	23	21	22	23	24	25	26	27
24	25	26	27	28	29	30	28	29	30				
31													

.....

Calendário

AUGUST

S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

SEPTEMBER

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

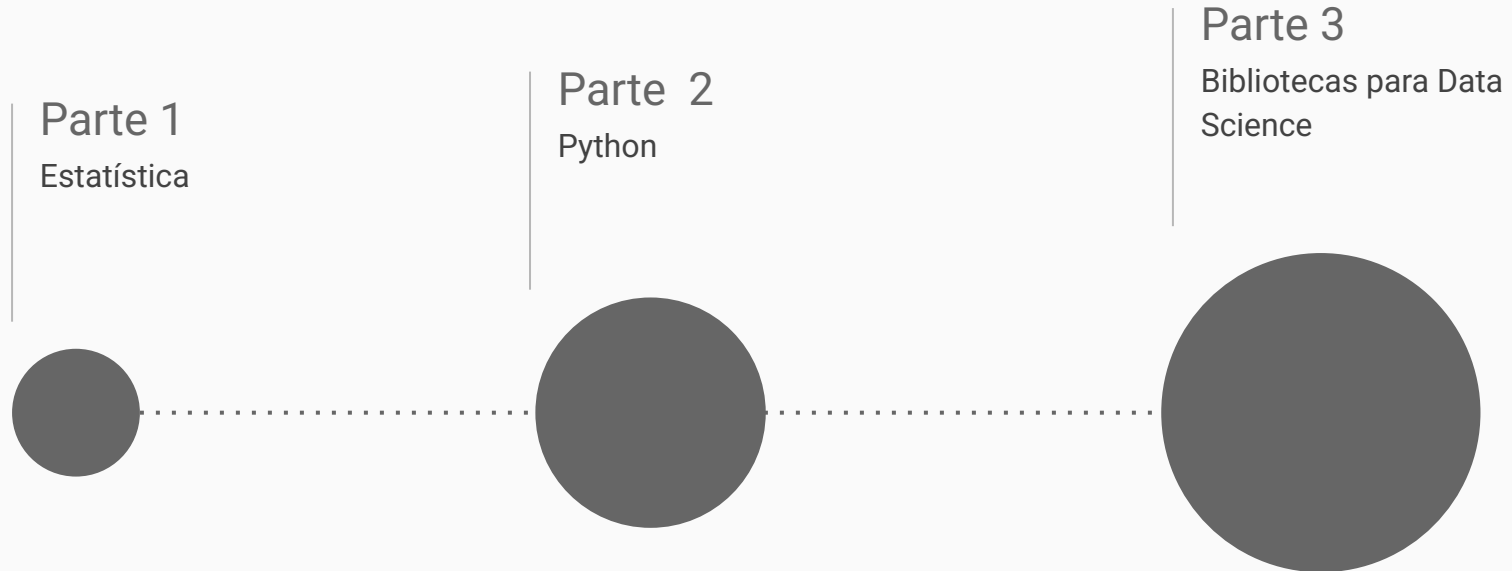
O que preciso para ter meu certificado?

Nota maior ou igual a 50 - Participação nas aulas e projeto avaliativo.

Frequência maior ou igual a 75% - Faltar em último caso.

Deem o melhor de vocês sempre!

Como será dividido?



O que é Data Science?

Data Science, ou Ciência de Dados, é uma área multidisciplinar que combina estatística, programação, conhecimento do domínio e técnicas de visualização para extrair conhecimento e insights de dados brutos.

Em outras palavras:

É o processo de transformar dados em informações úteis para a tomada de decisão.

Componentes principais:

Estatística: Ajuda a entender padrões e testar hipóteses.

Programação: Permite manipular e processar grandes volumes de dados (ex: Python, R).

Banco de Dados: Coleta e armazena dados estruturados e não estruturados.

Visualização Comunica os resultados de forma clara e acessível (ex: gráficos, dashboards).

Componentes principais:

Machine Learning: Usa algoritmos para prever ou classificar dados com base em padrões.

Conhecimento de Domínio Ajuda a interpretar os dados de forma contextualizada.

.....

Etapas comuns de um projeto de Data Science:

Coleta de dados (ex: APIs, bancos de dados, web scraping)

Limpeza e tratamento dos dados (ex: remoção de nulos, normalização)

Análise exploratória (ex: médias, dispersão, correlações)

Modelagem e predição (usando estatística ou machine learning)

Interpretação e visualização dos resultados

Tomada de decisão baseada em dados

Qual carreira eu posso seguir estudando esses conceitos?

1. Cientista de Dados (Data Scientist)

O que faz: Coleta, limpa, analisa e modela dados para gerar insights estratégicos.

Habilidades: Estatística, Python/R, machine learning, SQL, visualização de dados.

Diferencial: Capacidade de comunicar achados complexos para públicos não técnicos.

2. Analista de Dados (Data Analyst)

O que faz: Cria relatórios, dashboards e análises descritivas para orientar decisões.

Habilidades: Excel, SQL, Power BI/Tableau, estatística básica.

Perfil: Ótima escolha para começar na área.

Analista X Cientista de Dados





Estatística

É um ramo da matemática
aplicada que se dedica a coletar,
analisar e interpretar dados

O que é?

Conceitos de Estatística

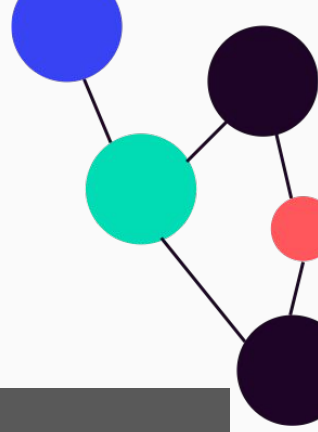
Dado

Informação

População
X
Amostra

Variáveis
Qualitativa
Quantitativa

Representação de
Dados
(Tipos de Gráfico)



Dado

São registros do que acontece no nosso dia-a-dia. Eles são a forma que encontramos de capturar, guardar e comunicar os fatos do nosso entorno.

Informação

Ordenação e organização dos dados de forma a transmitir significado e compreensão dentro de determinado contexto. Conjunto ou consolidação dos dados.

Conhecimento

Conjunto de informações e princípios aprendidos. É quando você usa a informação para refletir, aprender ou tomar decisões.



Exemplo:

Situação completa:

Ana tirou nota 4 na prova de matemática, que tinha peso 2. A média necessária para passar é 6.

Dados:

Nota: 4

Matéria: Matemática

Peso: 2

Média para passar: 6



Exemplo:

Situação completa:

Ana tirou nota 4 na prova de matemática, que tinha peso 2. A média necessária para passar é 6.

Informação:

É quando os dados são combinados com contexto e interpretados:

"Ana foi mal na prova de matemática e não atingiu a média."

Chave da informação: já responde "o que significa?"



Exemplo:

Situação completa:

Ana tirou nota 4 na prova de matemática, que tinha peso 2. A média necessária para passar é 6.

Conhecimento:

É quando você usa a informação para refletir, aprender ou tomar decisões:

"Como Ana tirou 4 na prova com peso 2, ela precisará de uma nota maior nas próximas avaliações para conseguir passar. Talvez precise de reforço em matemática."

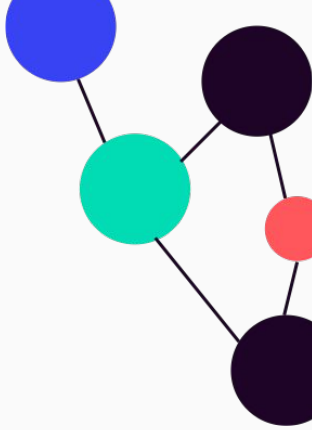
Chave do conhecimento: responde "o que fazer com isso?" ou "o que posso concluir?"



Quizzinho

"38°C"

- a) Dado
- b) Informação
- c) Conhecimento



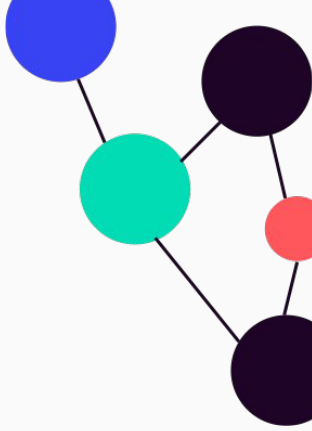
Quizzinho

"38°C"

a) Dado

b) Informação

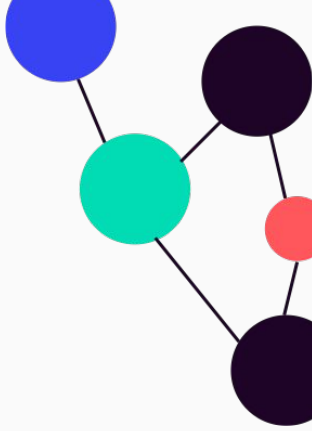
c) Conhecimento



Quizzinho

"João - 8h45 - Entrada"

- a) Dado
- b) Informação
- c) Conhecimento



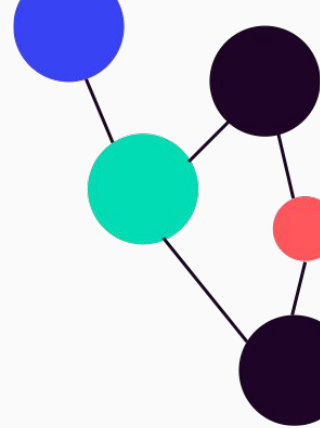
Quizzinho

"João - 8h45 - Entrada"

a) Dado

b) Informação

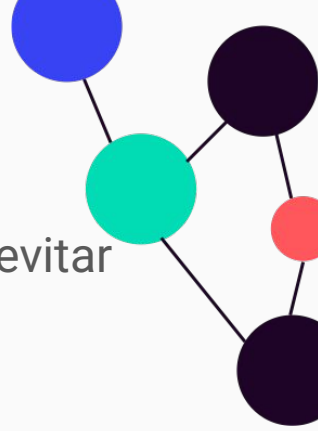
c) Conhecimento



Quizzinho

"Quando a temperatura ultrapassa os 37°C, é importante se hidratar e evitar exposição ao sol entre 10h e 16h."

- a) Dado
- b) Informação
- c) Conhecimento



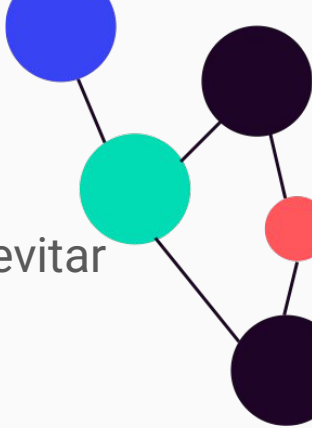
Quizzinho

"Quando a temperatura ultrapassa os 37°C, é importante se hidratar e evitar exposição ao sol entre 10h e 16h."

a) Dado

b) Informação

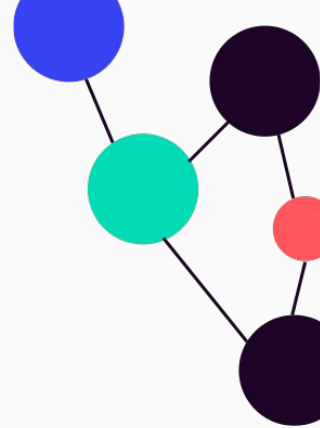
c) Conhecimento



Quizzinho

"João chegou atrasado ao trabalho, pois o horário de entrada é 8h."

- a) Dado
- b) Informação
- c) Conhecimento



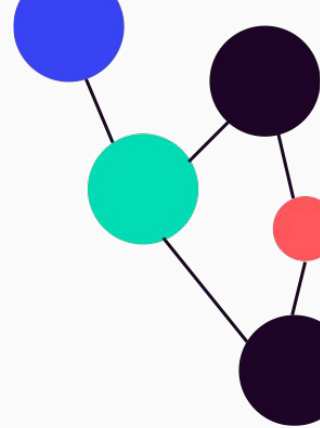
Quizzinho

"João chegou atrasado ao trabalho, pois o horário de entrada é 8h."

a) Dado

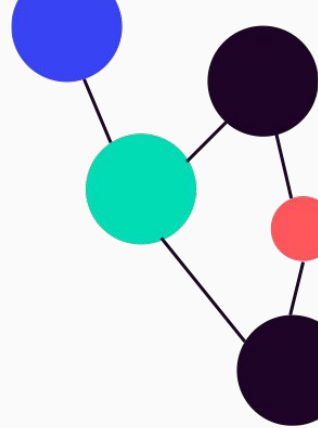
b) Informação

c) Conhecimento



População X Amostra

A população é o conjunto completo de todos os elementos que estão sendo estudados, enquanto a amostra é uma porção selecionada dessa população



Variáveis

1. Variáveis Qualitativas (Categóricas):

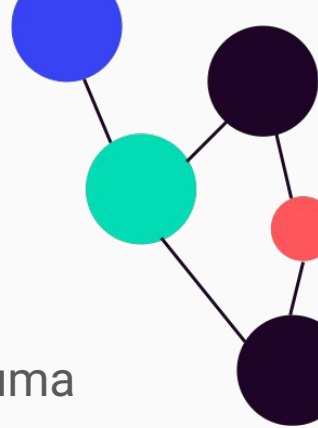
Essas variáveis descrevem qualidades ou categorias e não possuem uma ordem numérica inerente.

- Nominais: Não têm uma ordem específica.

Exemplo: Cores favoritas (vermelho, azul, verde)

- Ordinais: Têm uma ordem específica, mas as diferenças entre os valores não são mensuráveis.

Exemplo: Nível de satisfação (muito insatisfeito, insatisfeito, neutro, satisfeito, muito satisfeito)



Variáveis

2. Variáveis Quantitativas (Numéricas):

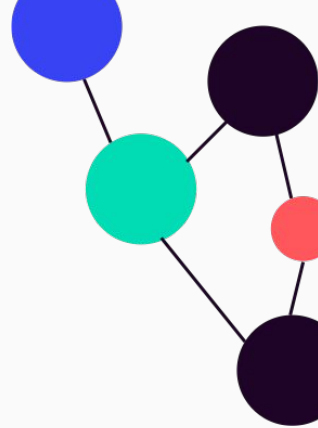
Essas variáveis representam quantidades e podem ser medidas numericamente.

- Discretas: Assumem valores inteiros e contáveis.

Exemplo: Número de filhos em uma família (0, 1, 2, 3)

- Contínuas: Podem assumir qualquer valor dentro de um intervalo contínuo e são medidas com precisão.

Exemplo: Tempo de espera (em minutos) para atendimento (2,5 minutos, 7,8 minutos)

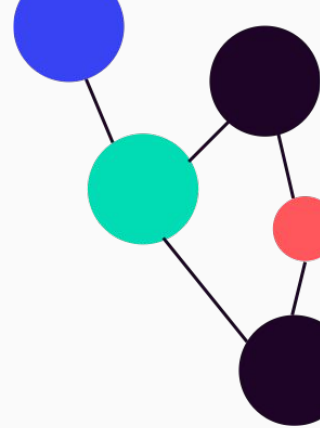


Quizzinho

Cor dos olhos de uma pessoa

a) Quantitativa

b) Qualitativa

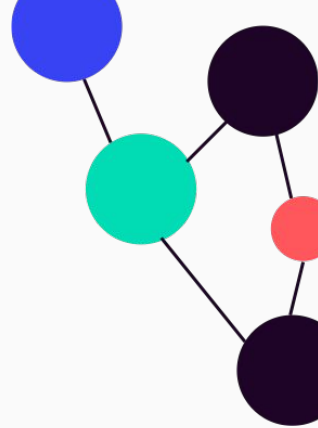


Quizzinho

Cor dos olhos de uma pessoa

a) Quantitativa

b) Qualitativa

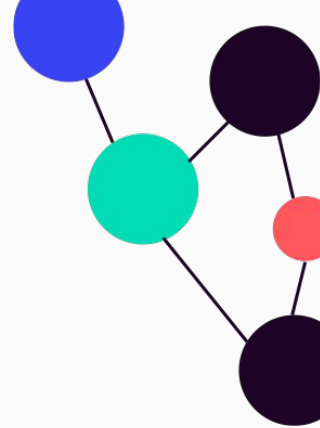


Quizzinho

Idade de um aluno

a) Quantitativa

b) Qualitativa

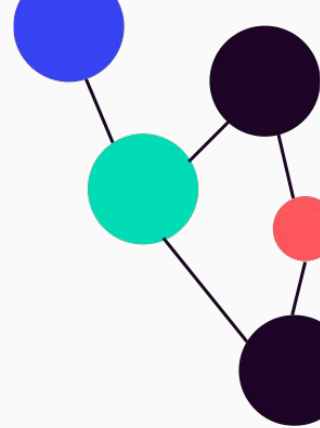


Quizzinho

Idade de um aluno

a) Quantitativa

b) Qualitativa

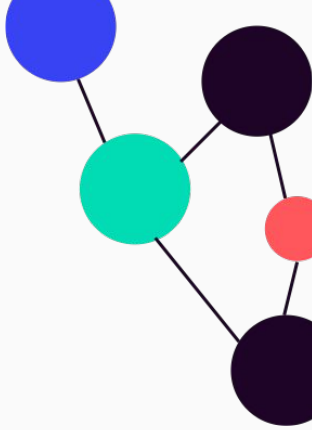


Quizzinho

Tipo sanguíneo (A, B, AB, O)

a) Quantitativa

b) Qualitativa

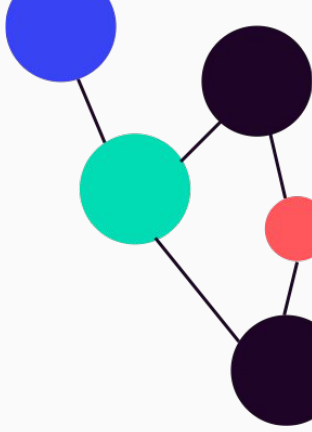


Quizzinho

Tipo sanguíneo (A, B, AB, O)

a) Quantitativa

b) Qualitativa



Tipos de Gráficos

COMPARAÇÃO DE GRUPOS

Gráfico de barra

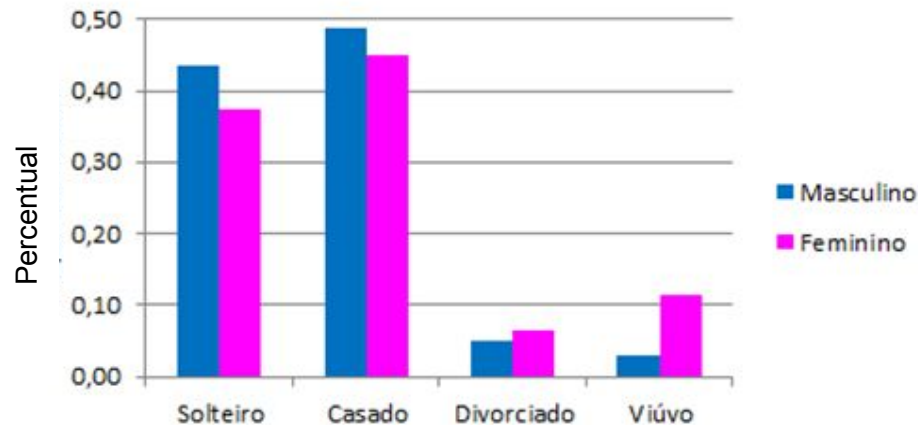
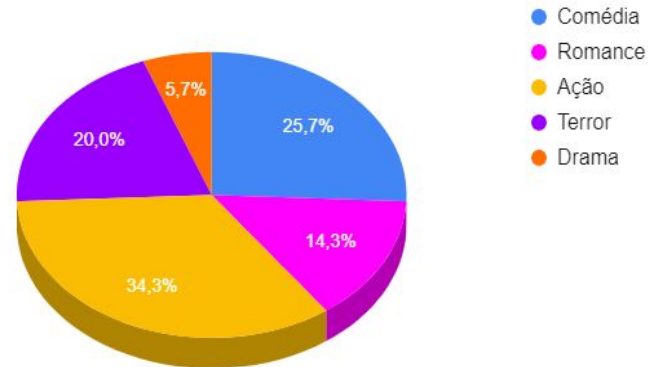


Gráfico de pizza

Preferência por filmes - alunos do 3º ano



Tipos de Gráficos

COMPARAÇÃO DE GRUPOS

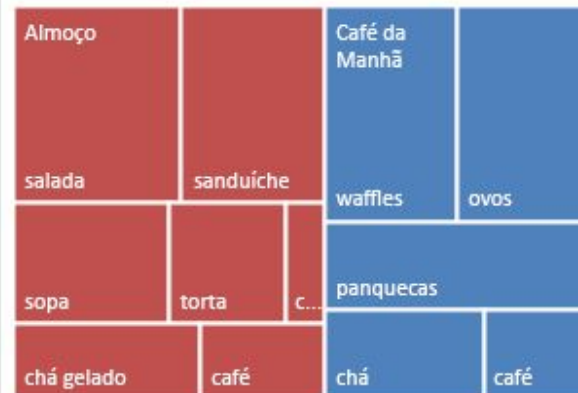
Nuvem de Palavras



Diagrama de Árvore

Vendas diárias de alimentos

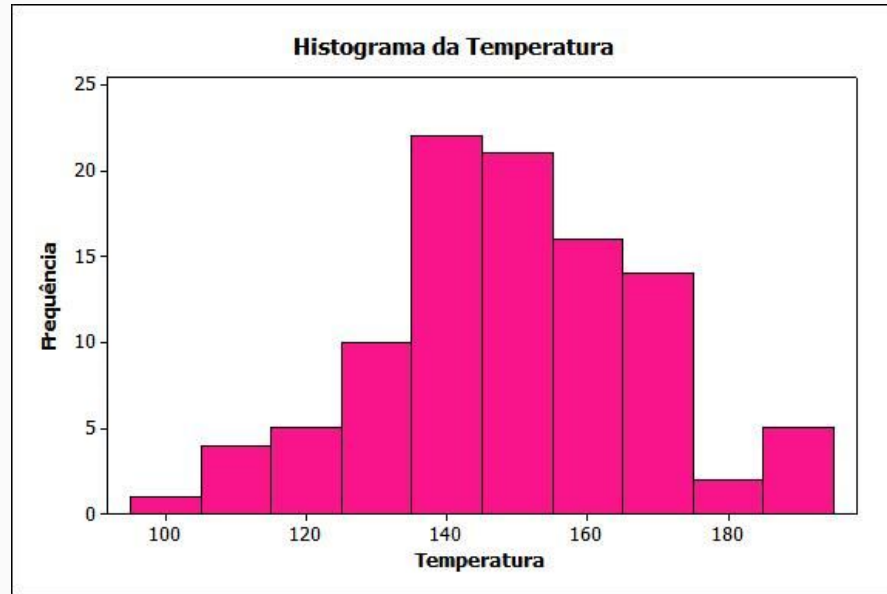
■ Café da Manhã ■ Almoço



Tipos de Gráficos

DISTRIBUIÇÃO DOS DADOS

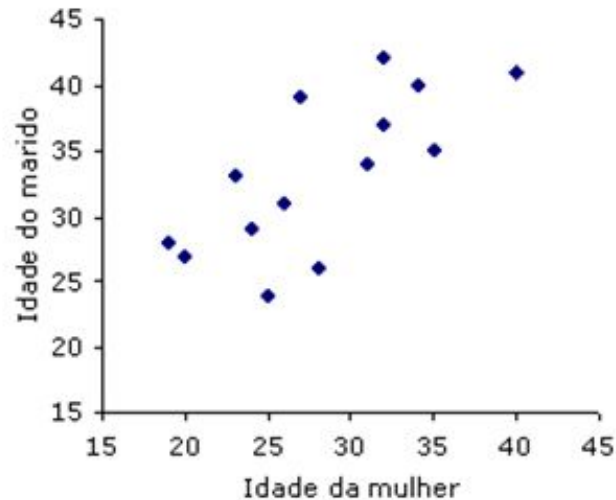
Histograma



Tipos de Gráficos

RELAÇÃO ENTRE VARIÁVEIS

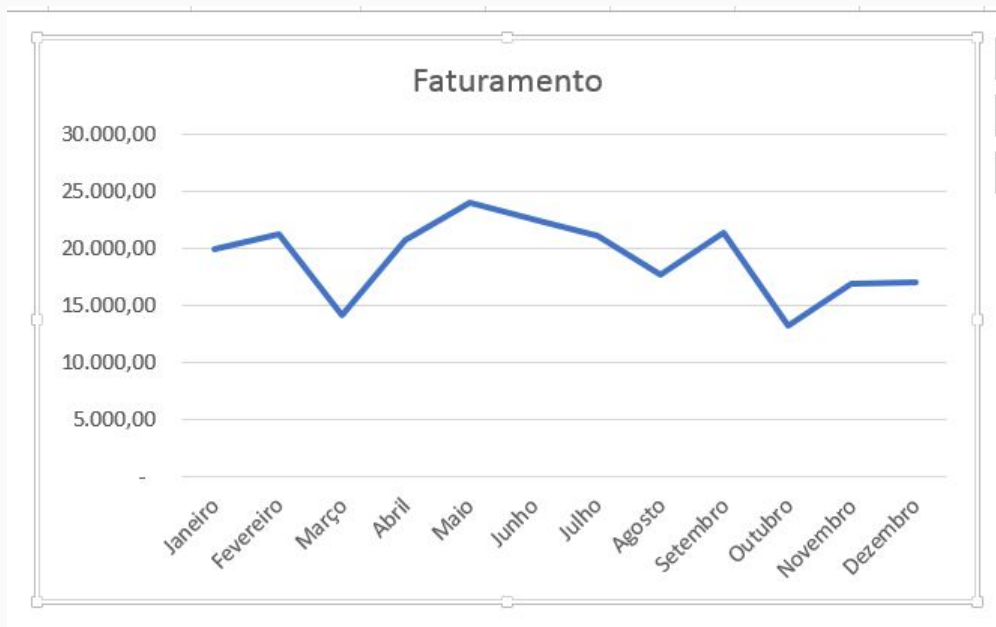
Gráfico de dispersão



Tipos de Gráficos

EVOLUÇÃO

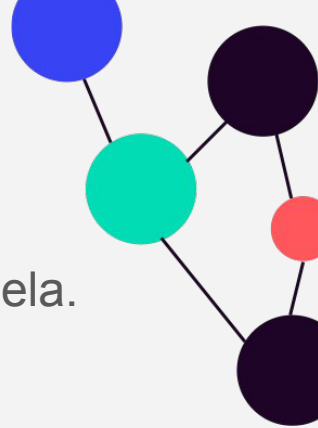
Gráfico de linha



Distribuição de Frequências

É uma forma de organizar e resumir um conjunto de dados em uma tabela.

1. Frequência absoluta
2. Frequência acumulada
3. Frequência relativa
4. Frequência relativa acumulada



Regras para calcular o número de classes

A regra de Sturges

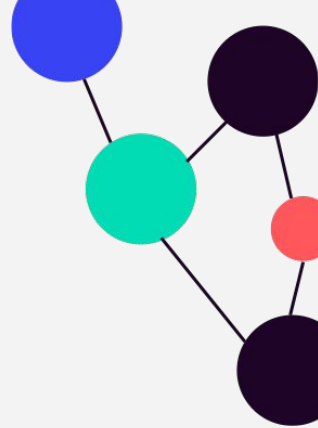
$$k = 1 + 3,3 \log (n)$$

A regra da Raíz

$$k = \sqrt{n}$$

Intervalo de Classes

$$h = AMP / k$$

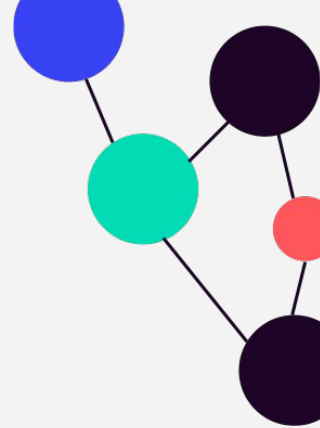


Medidas de Tendência Central

Média - Média Ponderada

Mediana

Moda



Medidas de Dispersão

Variância σ^2

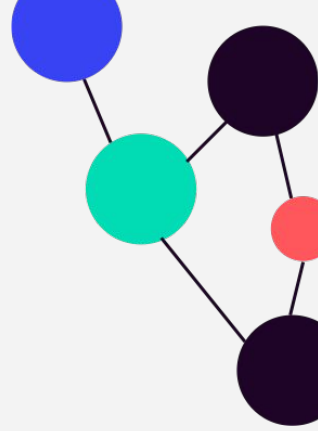
$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Desvio Padrão σ

$$\sigma = \sqrt{\sigma^2}$$

Coeficiente de Variação cv

$$cv = \sigma / \bar{x}$$

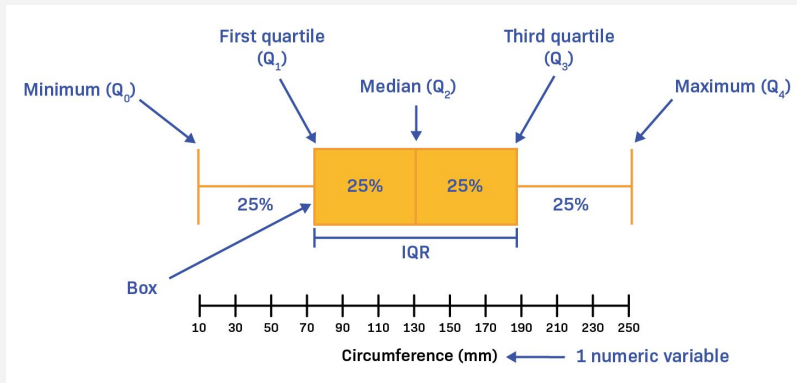
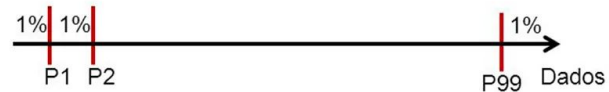


Quartis e Percentis

Quartil



Percentil
Ou Centil



Atividade em Grupo: Aula invertida

Em grupos de 3 a 5 pessoas, vocês deverão criar uma apresentação clara e didática que aborde os tópicos de Estatística Descritiva.

Todos devem anexar o slide no Classroom.

Não tem máximo nem mínimo de slides.

Irei sortear dois grupos para apresentarem. Se preparem!

Atividade em Grupo: Aula invertida

Em grupos de 3 a 5 pessoas, vocês deverão criar uma apresentação clara e didática que aborde os tópicos de Estatística Descritiva.

Todos devem anexar o slide no Classroom.

Não tem máximo nem mínimo de slides.

Irei sortear dois grupos para apresentarem. Se preparem!

Principais ferramentas para Data Science

1. Linguagens de programação

Python: A mais popular, versátil e com uma enorme biblioteca para análise, manipulação, visualização de dados e machine learning (ex: pandas, NumPy, scikit-learn, TensorFlow, PyTorch).

R: Muito usada para estatística, análises exploratórias e visualizações sofisticadas (ex: ggplot2, dplyr).

SQL: Linguagem padrão para consulta em bancos de dados relacionais.

Principais ferramentas para Data Science

2. Ambientes e Notebooks

Jupyter Notebook / JupyterLab: Ambiente interativo para desenvolvimento com células de código, texto e gráficos.

Google Colab: Versão cloud do Jupyter, com GPUs grátis e integração com Google Drive.

RStudio: IDE para linguagem R, com suporte para scripts, notebooks e visualizações.
