ACISSP
Associação Comercial, Industrial, Agropecuária e de Serviços de São Sebastião do Paraíso-MG

UFLA
UNIVERSIDADE FEDERAL DE LAVRAS

Libertas
FACULDADES INTEGRADAS

UAL Tech
Inteligência Artificial e Ciência de Dados (3horas)

Workflow do Cientista de Dados:
Uma abordagem prática

Prof. Dr. Leonilson Kiyoshi Sato de Herval

Sobre Mim

Pai do Isaac, Marido da Ana e Professor.
Apaixonado por ensinar, aprender, programar, pelo mercado financeiro, pela ciência de dados, empreendedorismo e inovação.

Pesquisas Científicas

Física Aplicada ao Mercado Financeiro
Ciência de Dados Aplicada (Machine Learning)
Parceria com UpVendas – Análise de Risco

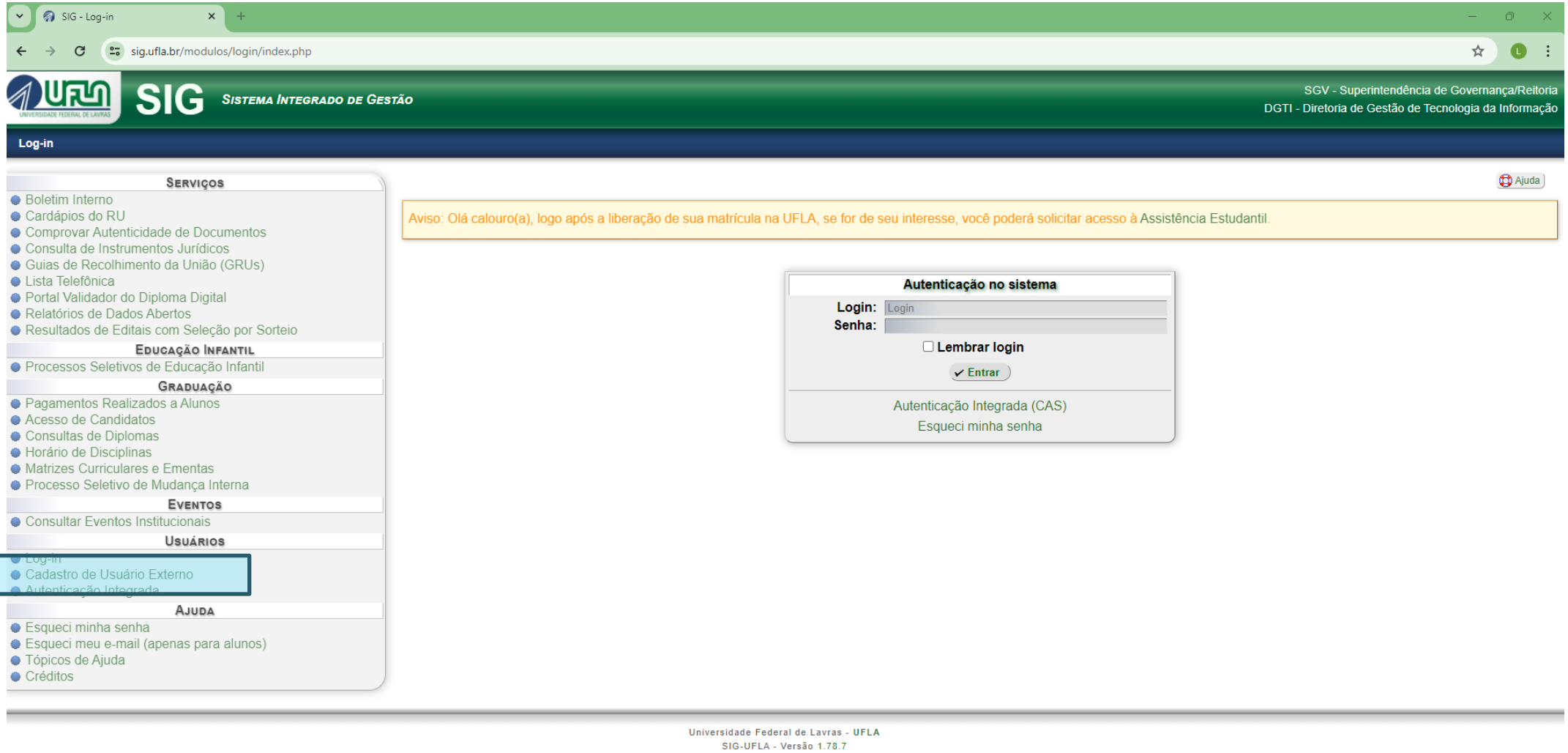
Experiência Profissional

Bacharel em Física Biológica
Mestre em Ciências (Física)
Doutor em Ciências (Física)
Colaboração Esporádica com Research focada no mercado de opções

Orientações

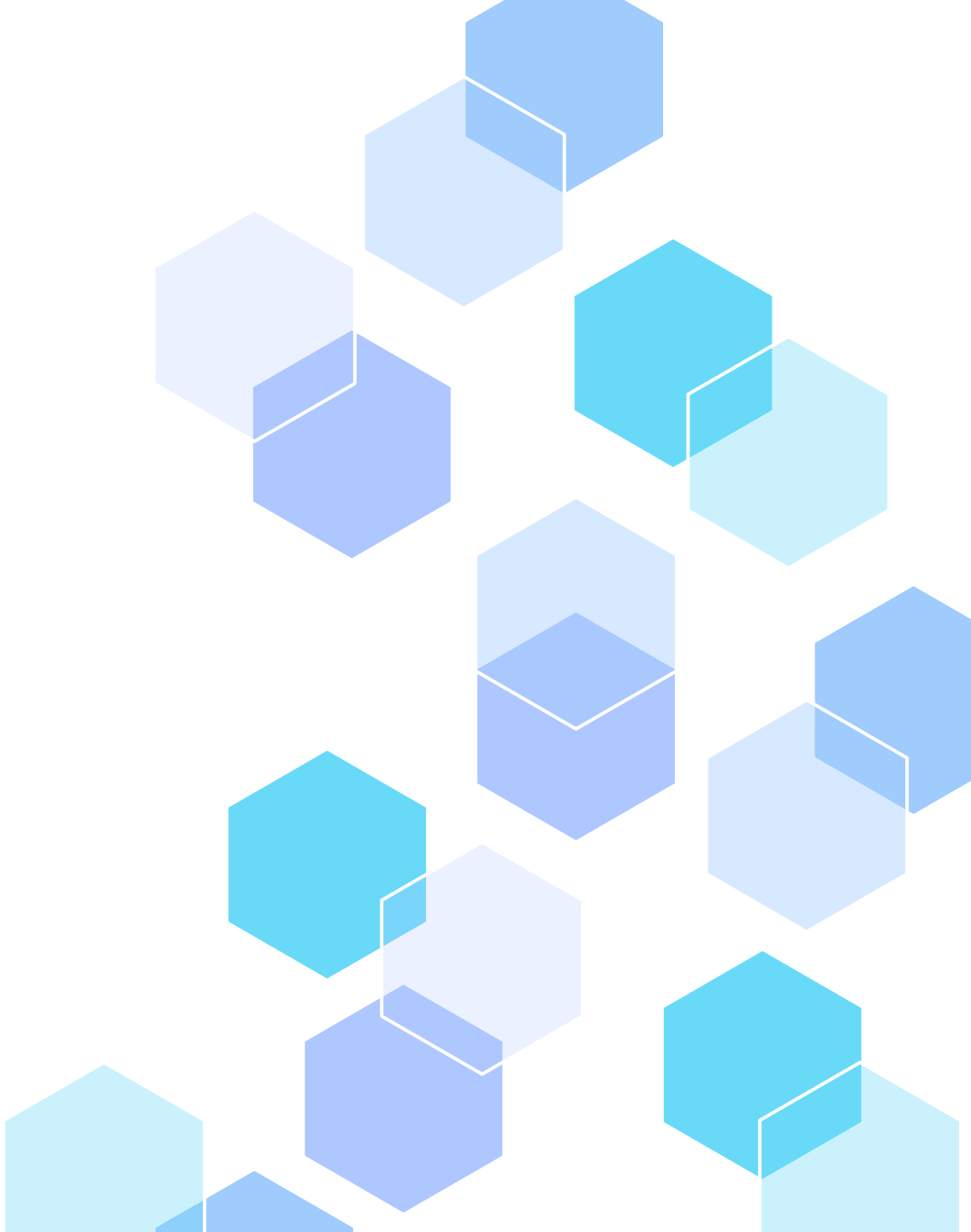
Cadastro como usuário externo no SIG;
Assinar Lista de Presença nos dois dias;

sig.ufla.br


SIG - Log-in
sig.ufla.br/modulos/login/index.php
SIG
Sistema Integrado de Gestão
SGV - Superintendência de Governança/Reitoria
DGTI - Diretoria de Gestão de Tecnologia da Informação
Log-in
Serviços
Boletim Interno
Cardápios do RU
Comprovar Autenticidade de Documentos
Consulta de Instrumentos Jurídicos
Guias de Recolhimento da União (GRUs)
Lista Telefônica
Portal Validador do Diploma Digital
Relatórios de Dados Abertos
Resultados de Editais com Seleção por Sorteio
Educação Infantil
Processos Seletivos de Educação Infantil
Graduação
Pagamentos Realizados a Alunos
Acesso de Candidatos
Consultas de Diplomas
Horário de Disciplinas
Matrizes Curriculares e Ementas
Processo Seletivo de Mudança Interna
Eventos
Consultar Eventos Institucionais
Usuários
Log-in
Cadastro de Usuário Externo
Autenticação Integrada
Ajuda
Esqueci minha senha
Esqueci meu e-mail (apenas para alunos)
Tópicos de Ajuda
Créditos
Ajuda
Aviso: Olá calouro(a), logo após a liberação de sua matrícula na UFLA, se for de seu interesse, você poderá solicitar acesso à Assistência Estudantil.
Autenticação no sistema
Login:
Login
Senha:
Lembrar login
Entrar
Autenticação Integrada (CAS)
Esqueci minha senha
Universidade Federal de Lavras - UFLA
SIG-UFLA - Versão 1.78.7

Alinhar Expectativas

De qual área vocês são?
Qual a profissão?
O que vocês esperam deste Mini-Curso?



Guia do Mini-Curso

01
Introdução

O que é Ciência de Dados

02
Diferenças

Analista x Cientista x BI

03
Machine Learning

Conceitos Gerais

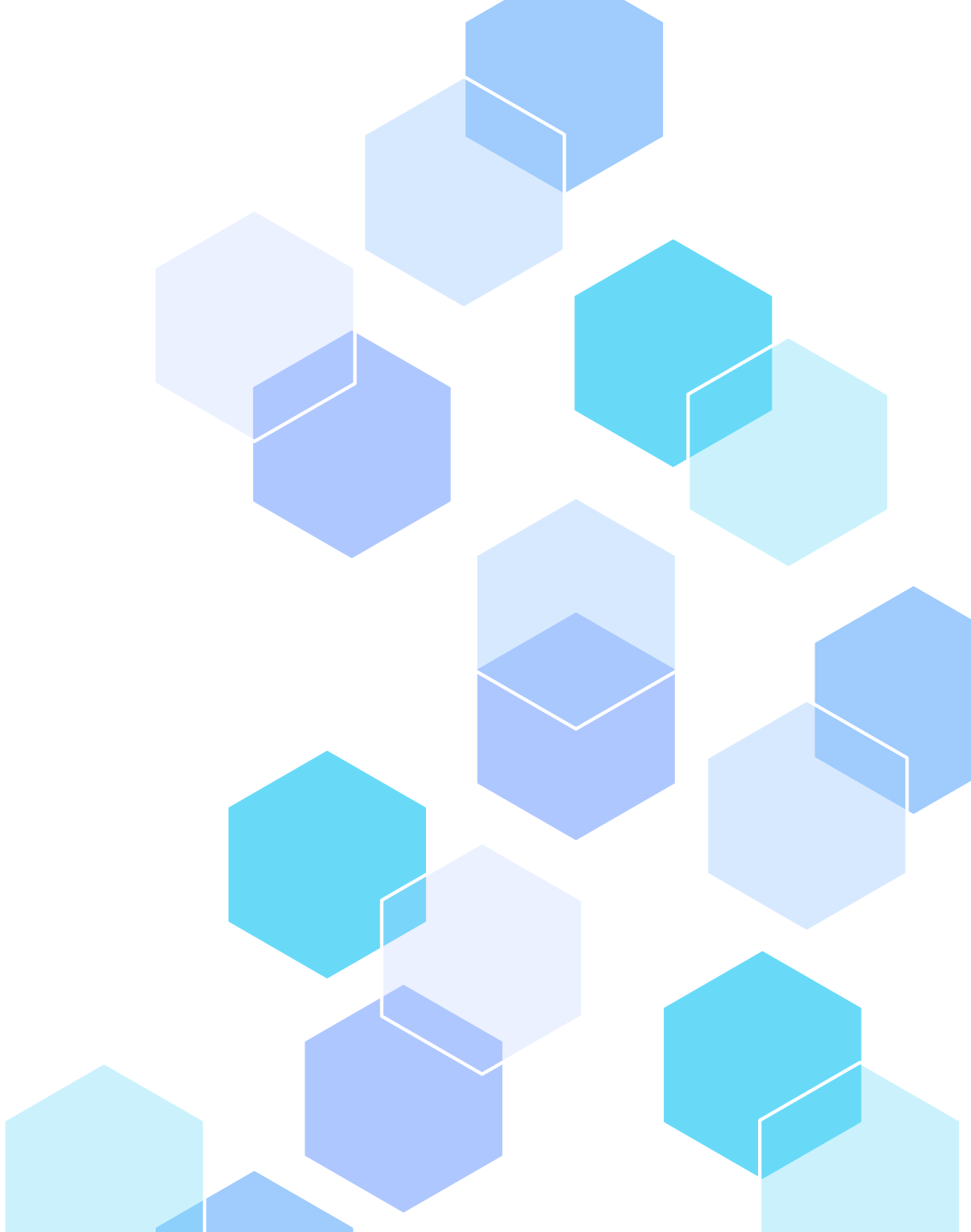
04
Prática

WorkFlow em uma base de dados de exemplo

01

Introdução

O que é Ciência de Dados?



O que é Ciência de Dados?

Ciência

1. Conhecimento atento e aprofundado de algo.
2. Corpo de conhecimentos sistematizados adquiridos via observação, identificação, pesquisa e explicação de determinadas categorias de fenômenos e fatos, e formulado metódica e racionalmente.

Definição de Oxford Languages

O que é Ciência de Dados?

Ciência

1. Conhecimento atento e aprofundado de algo.
2. Corpo de conhecimentos sistematizados adquiridos via observação, identificação, pesquisa e explicação de determinadas categorias de fenômenos e fatos, e formulado metódica e racionalmente.

Definição de Oxford Languages

Dados

- São observações documentadas, ou resultado da medição que gera informações que identifica algo, alguém ou sobre si mesmo.
- O que se consegue processar e decodificar a partir de um computador.

IME - UNICAMP

O que é Ciência de Dados?

É o processo de **exploração, manipulação e análise** dos dados para descoberta e previsão através da **criação de hipóteses, testes e validações** com objetivo de responder perguntas e/ou fazer recomendações.

O que é Ciência de Dados?

É o processo de **exploração, manipulação e análise** dos dados para descoberta e previsão através da **criação de hipóteses, testes e validações** com objetivo de responder perguntas e/ou fazer recomendações.

Requisitos do processo:

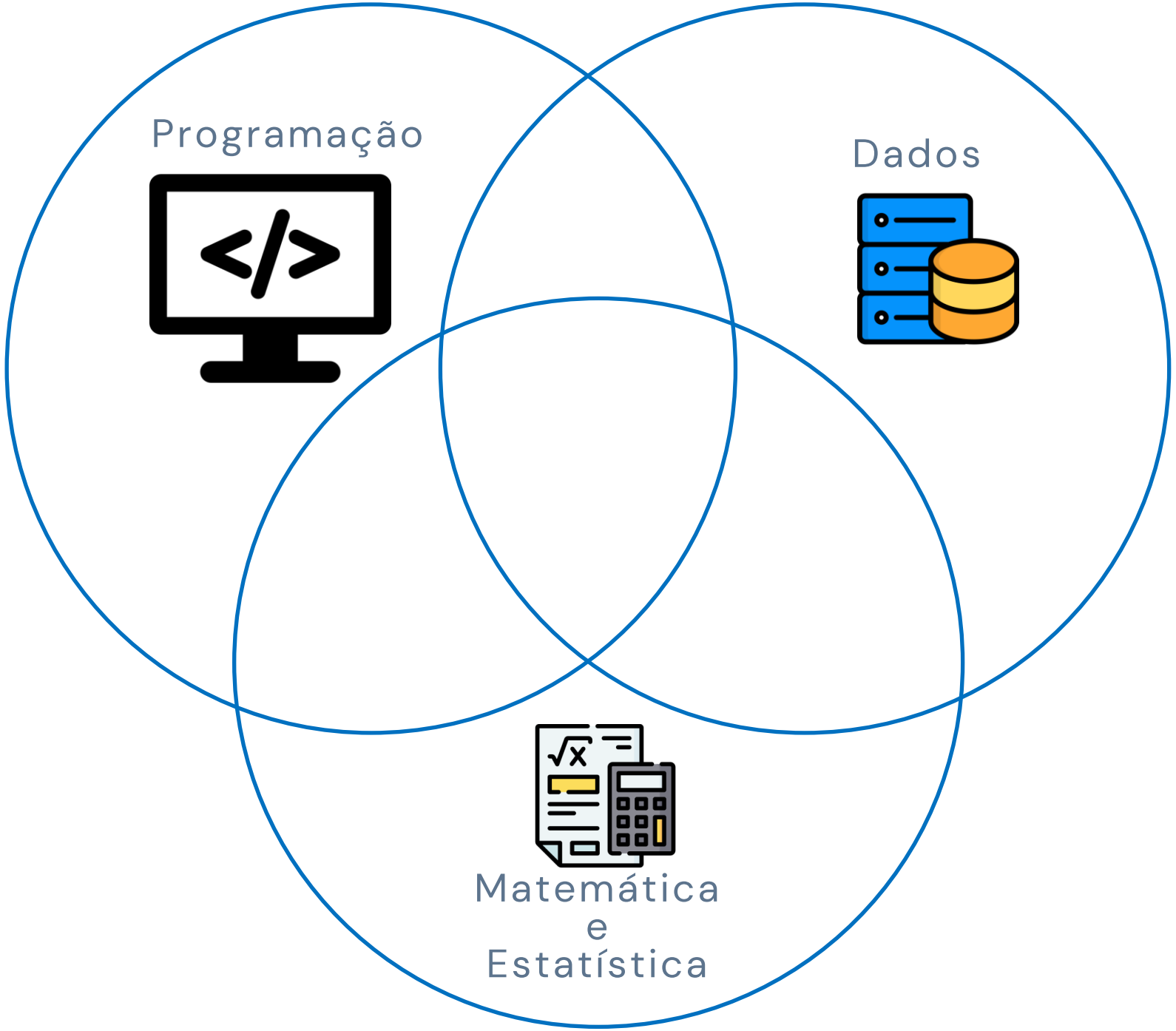
Embasamento estatístico e matemático

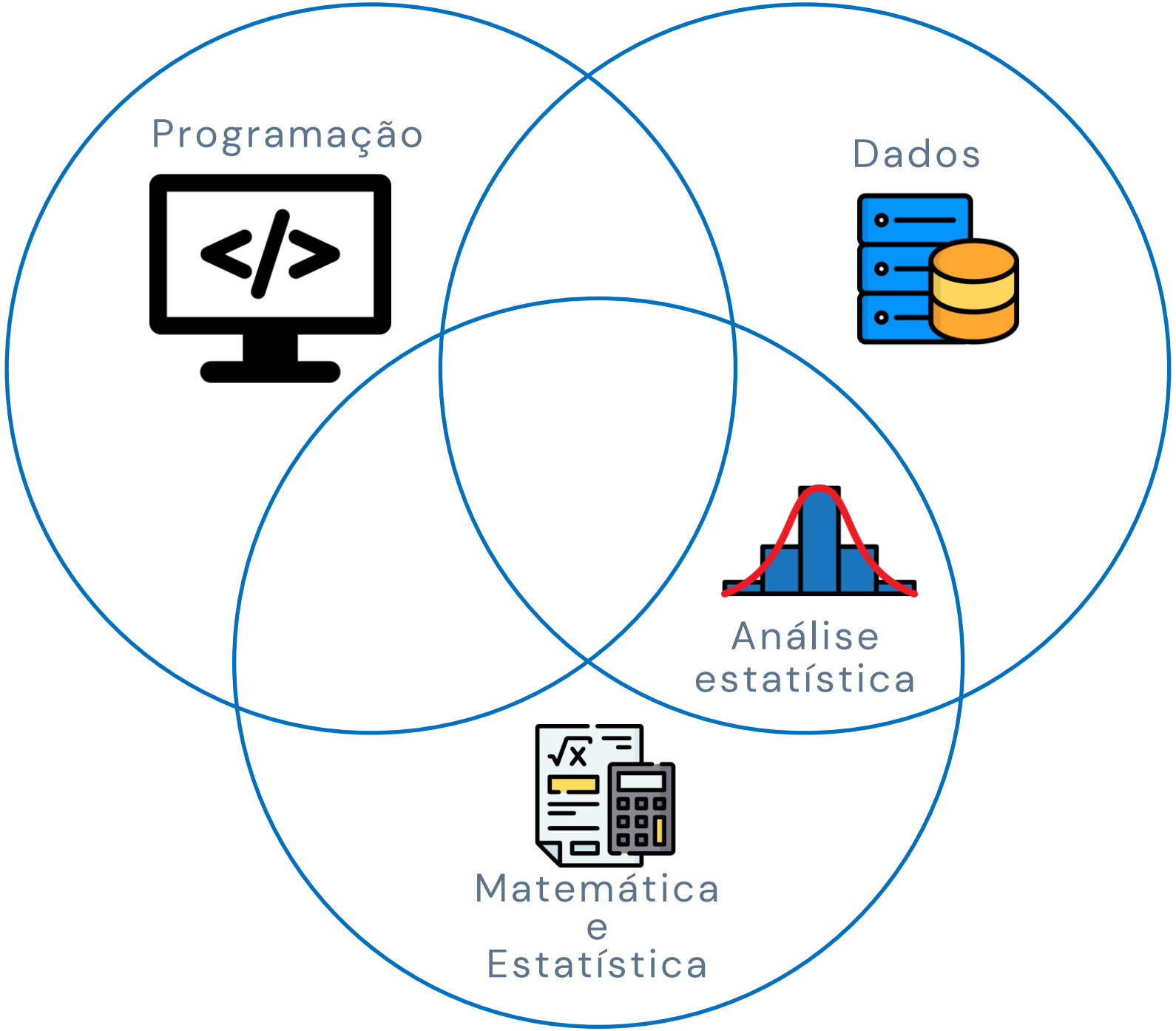
Ser escalável, replicável e viável

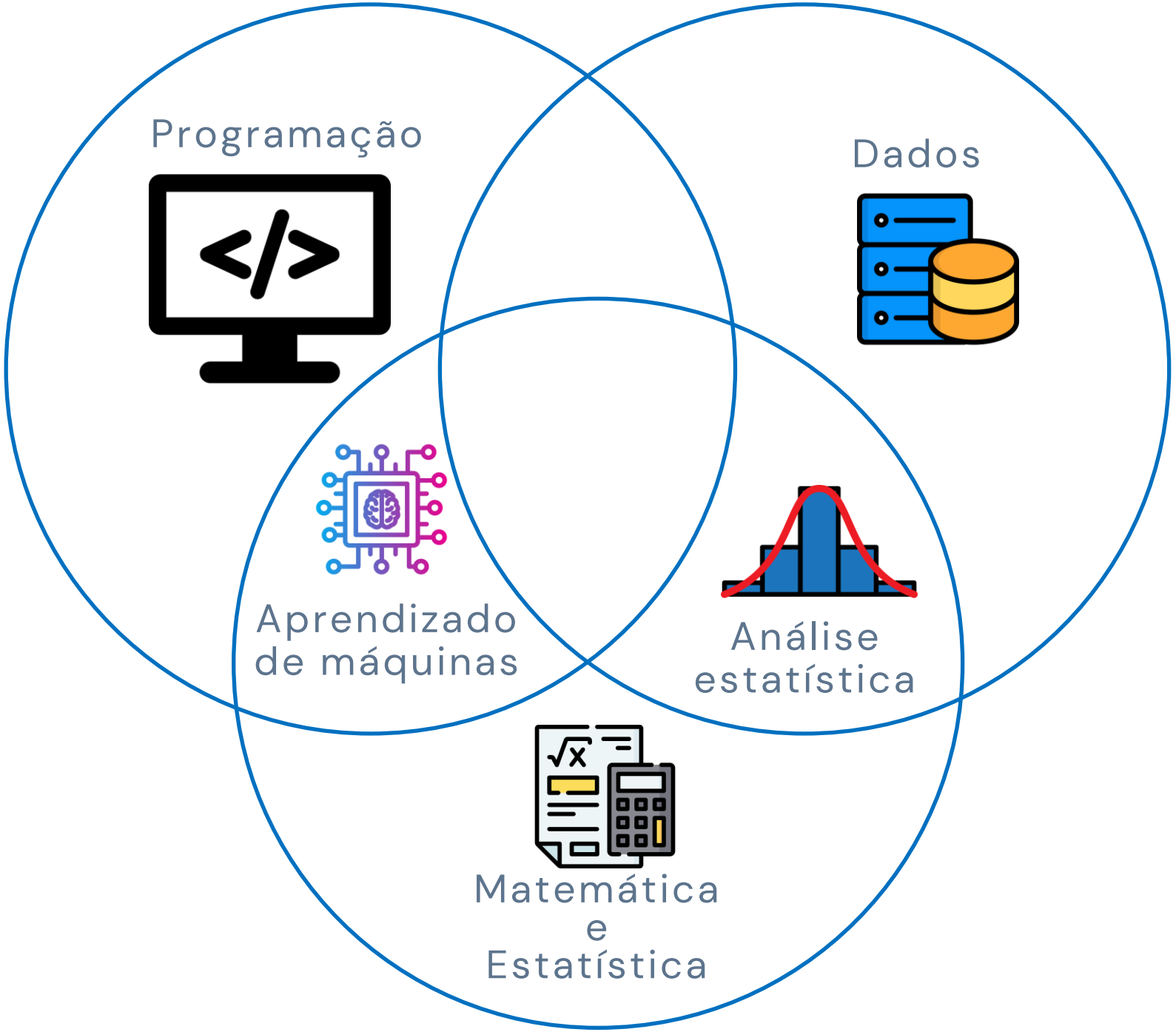
Observação, hipóteses, testes, validação, análise e monitoramento

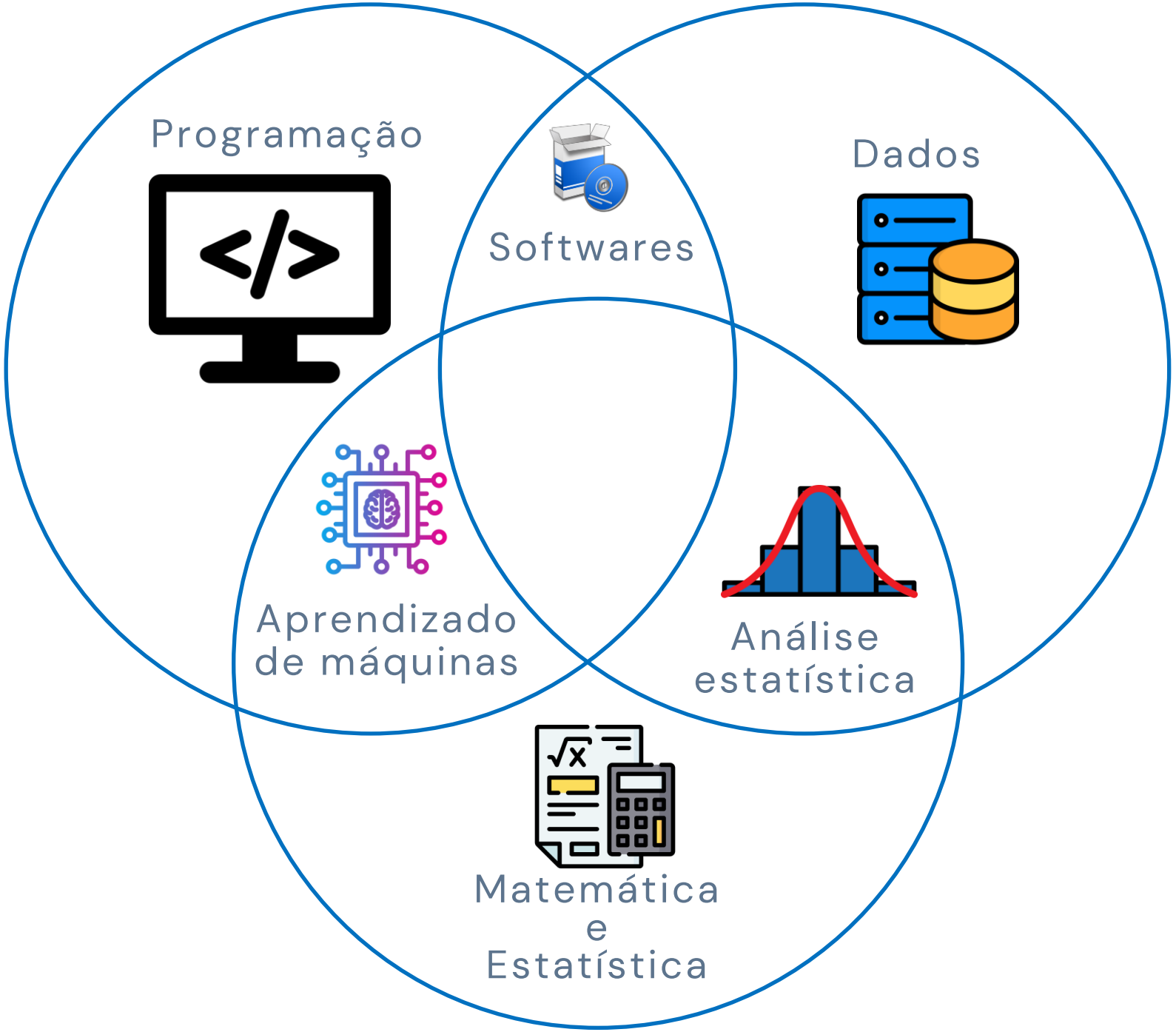
O que é Ciência dos Dados?

Dados armazenados, processamento e visualização

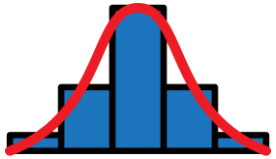

Programação
Dados
Matemática
e
Estatística


Programação
Dados
Análise
estatística
Matemática
e
Estatística


Programação
Dados
Aprendizado de máquinas
Análise estatística
Matemática e Estatística


Programação
Softwares
Dados
Aprendizado
de máquinas
Análise
estatística
Matemática
e
Estatística

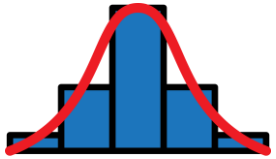
De maneira isolada



Análise estatística

Estatística Descritiva		
Parâmetro	**Turma (T1)**	**Turma (T2)**
Média amostral	4,6	4,5
Coeficiente de Variação	0,5348	0,5044
Mediana	4,5	4,6
Desvio padrão amostral	2,46	2,27
Variância Amostral	6,0516	5,1529
Número total de alunos	68	48

De maneira isolada



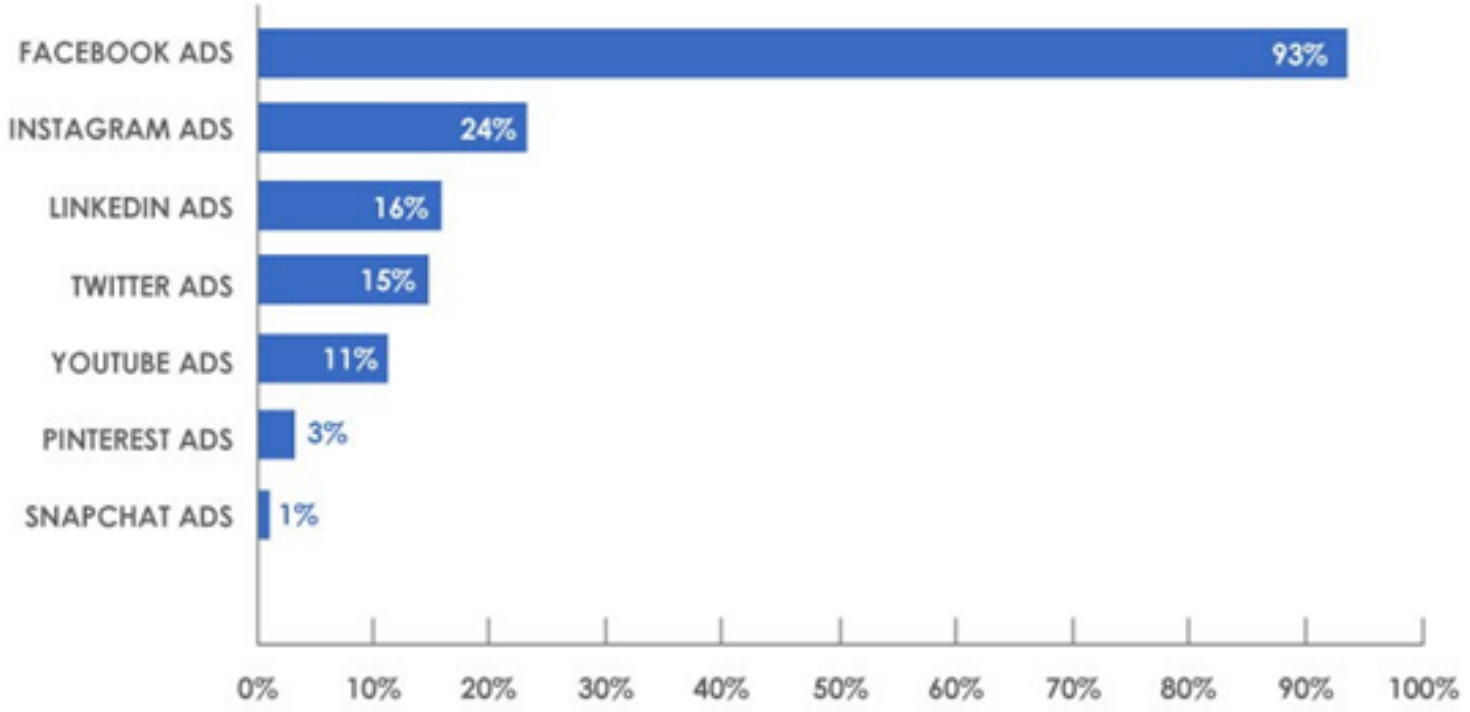
Análise estatística

Estatística Descritiva		
Parâmetro	**Turma (T1)**	**Turma (T2)**
Média amostral	4,6	4,5
Coeficiente de Variação	0,5348	0,5044
Mediana	4,5	4,6
Desvio padrão amostral	2,46	2,27
Variância Amostral	6,0516	5,1529
Número total de alunos	68	48

Por que eu preciso dessas informações?

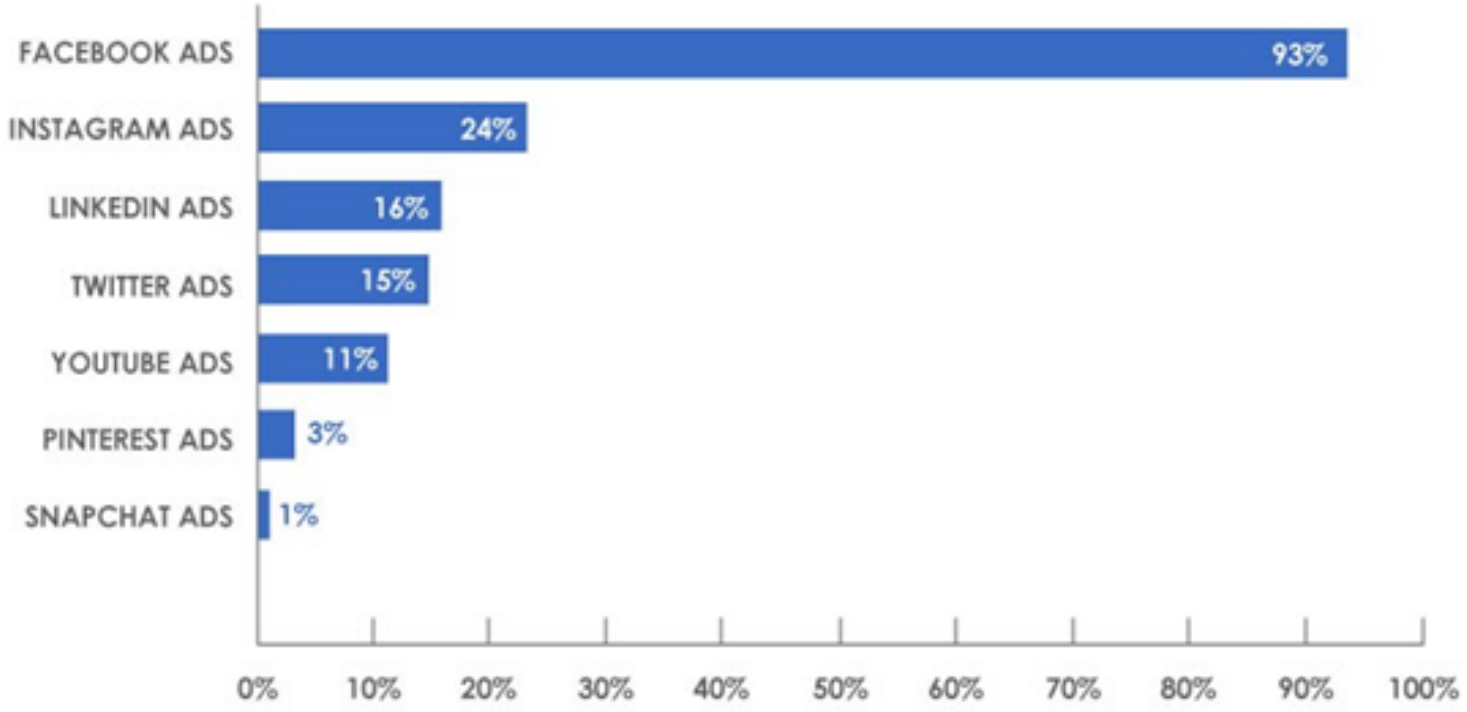
De maneira isolada

Aprendizado de máquinas



Aprendizado de máquinas

De maneira isolada



Qual o Período destes Dados?

Softwares

De maneira isolada

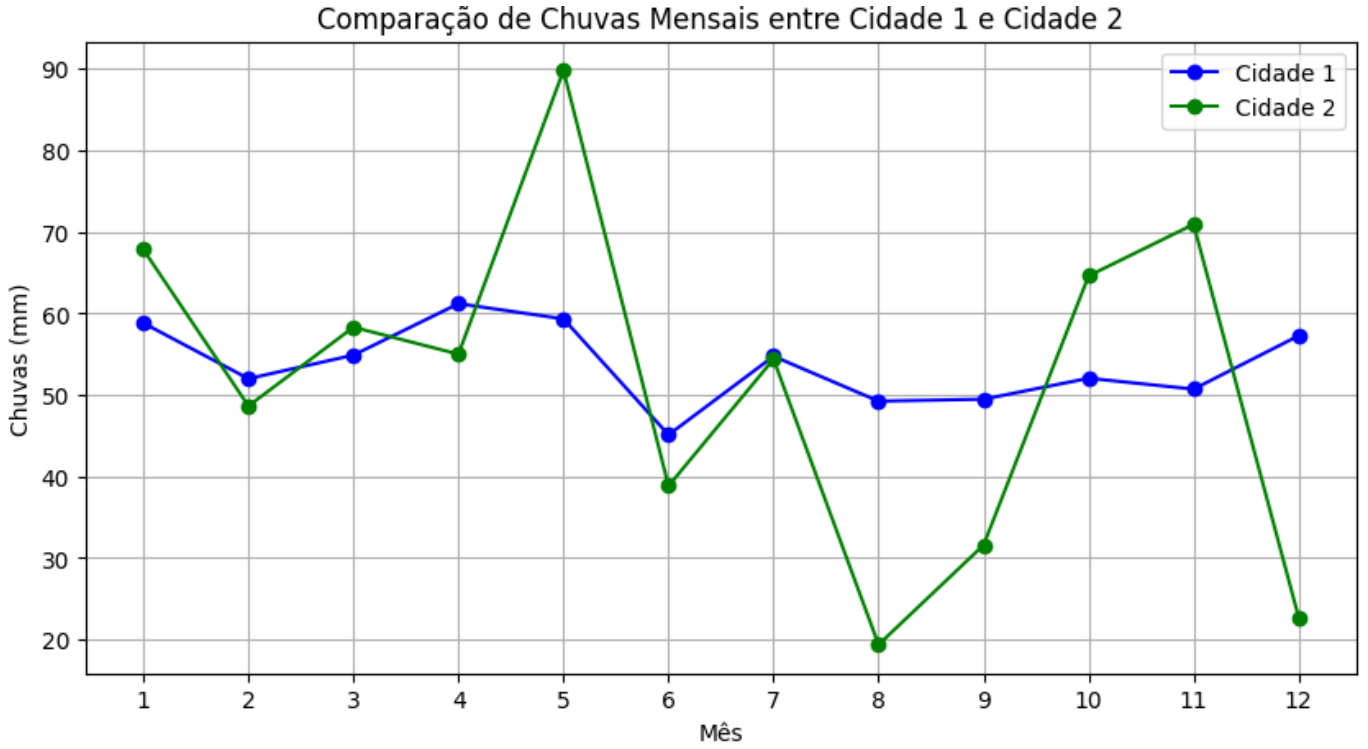
Cidade	Média de Chuvas (mm)
Cidade 1	55.741062
Cidade 2	51.842557

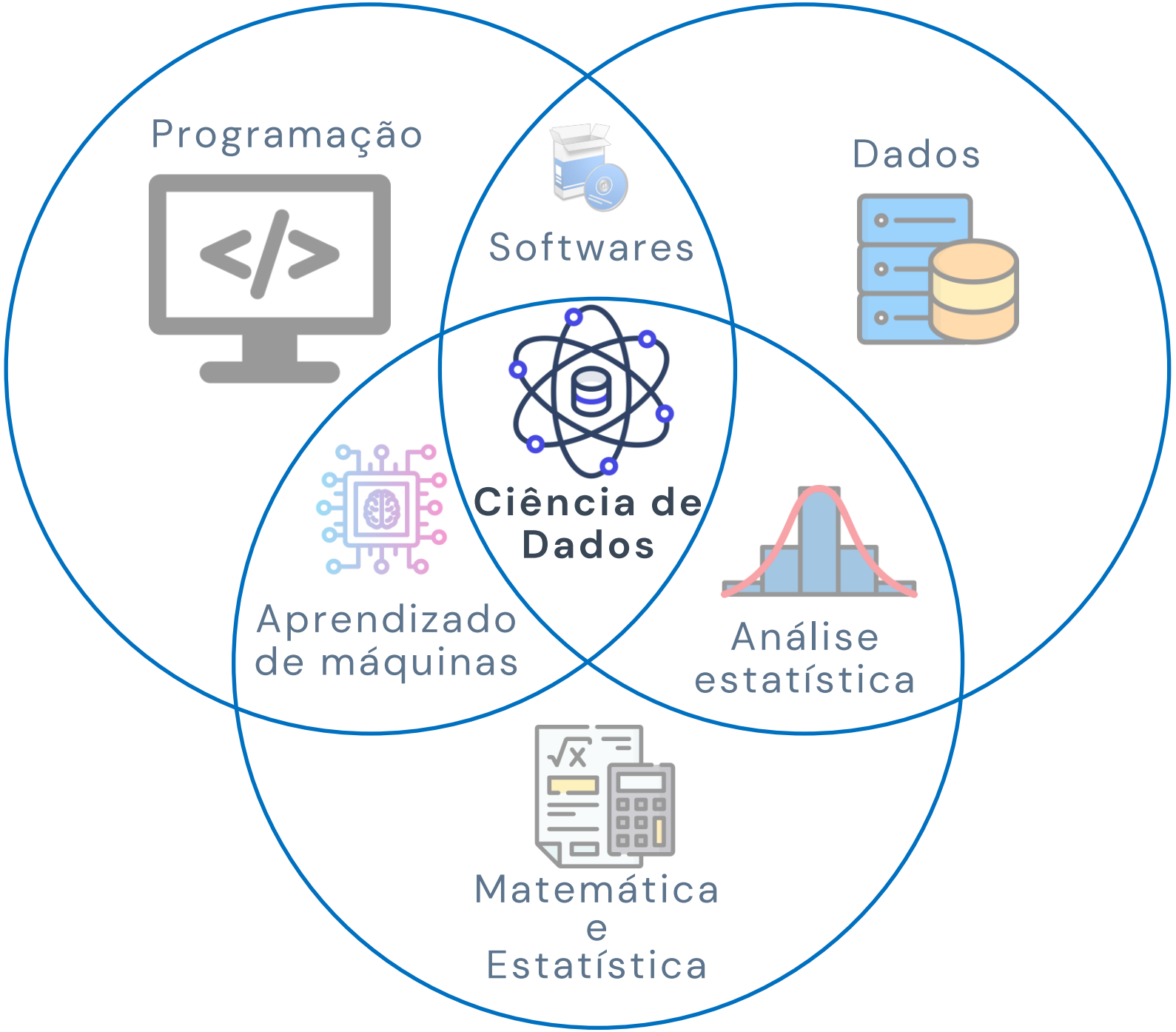
Qual a melhor cidade visitar para diminuir o risco de chuvas?

De maneira isolada

Softwares

Cidade	Média de Chuvas (mm)
Cidade 1	55.741062
Cidade 2	51.842557

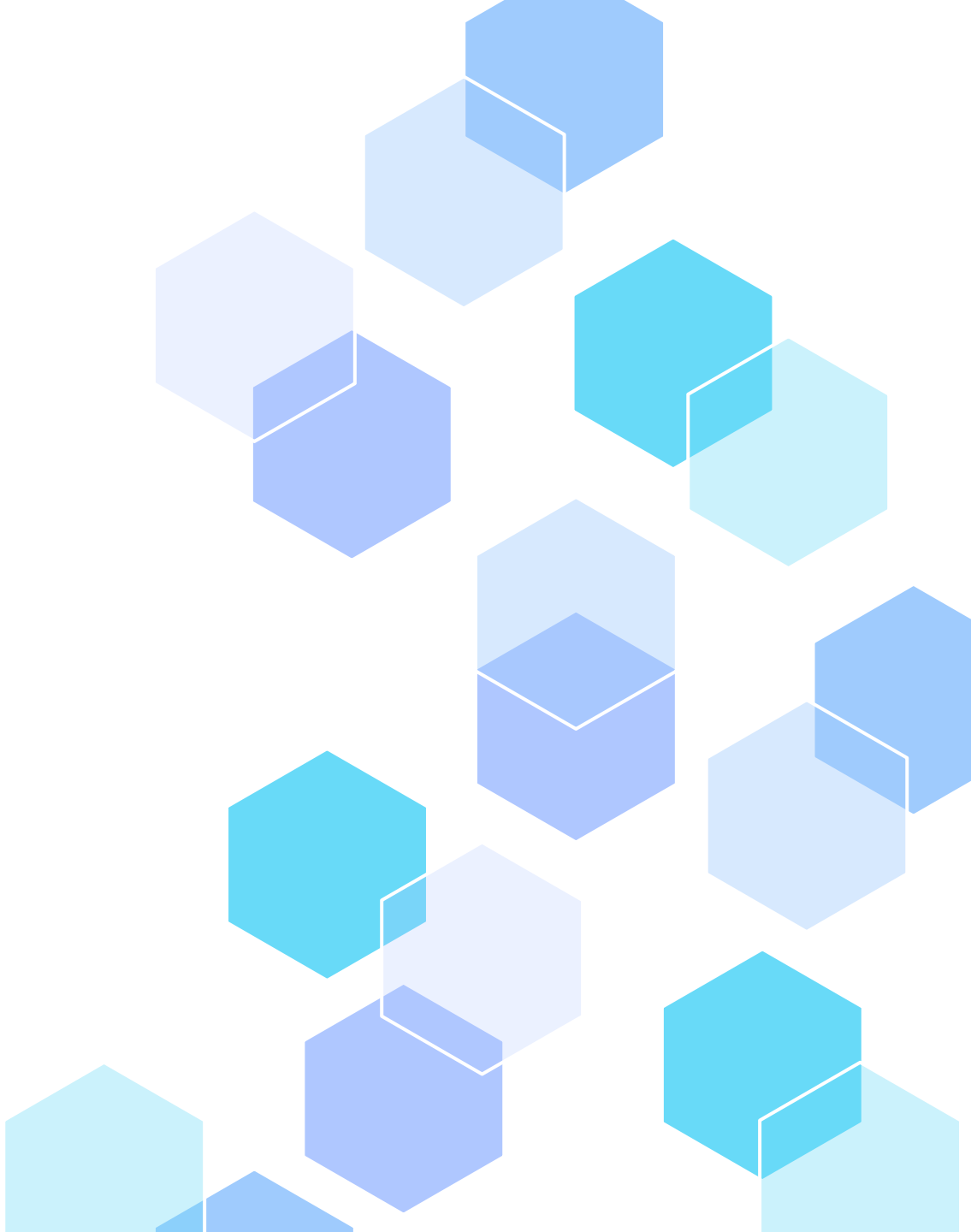



Programação
Softwares
Dados
Ciência de Dados
Aprendizado de máquinas
Análise estatística
Matemática e Estatística

02

Diferenças

BI x Analista de Dados x Cientista de Dados



BI

- Auxilia o time responsável por tomadas de decisões, gerencia e acompanha as melhorias.
- Recebe os dados já tratados (Engenheiros de Dados e DBA's).
- Gera visualizações e relatórios que contam uma história

Analista

- Extrai infomações que podem ser usada diretamente para resolver problemas;
- Pega os dados de forma bruta;
- Faz inferências baseada em processos estatísiticos para validar hipóteses.

Cientista

- Identifica, analisa e propõe soluções para problemas baseado em estatística;
- Possui habilidades matemáticas, pensamento lógico e comunicação eficiente;
- Busca entender o futuro usando técnicas de Machine Learning.

Exemplo: Vendas do Produto e Payback

BI

- Monitoramento de vendas;
- Exibir informações de total de vendas, número de compras, número de clientes, margem de vendas, etc.
- Geralmente transforma os dados em Dashboards.

Analista

- Usa programação para aplicar técnicas estatísticas;
- Realiza a limpeza e agrupamento dos dados para verificar possíveis problemas ou ganhos;
- Transmite a informação para equipe de gestores;
- Geralmente usa relatórios e Dashboards.

Cientista

- Aplica técnicas de estatística, machine learning, inteligência artificial ou correlatas;
- Busca detectar e prever o ponto mais eficaz para efetuar a operação, de modo à retornar mais investimento para a empresa;
- Faz previsões de venda na loja que ele considere ser a melhor para esse determinado produto.

03

Machine Learning

Conceitos Gerais

Gato

Gato

Gato

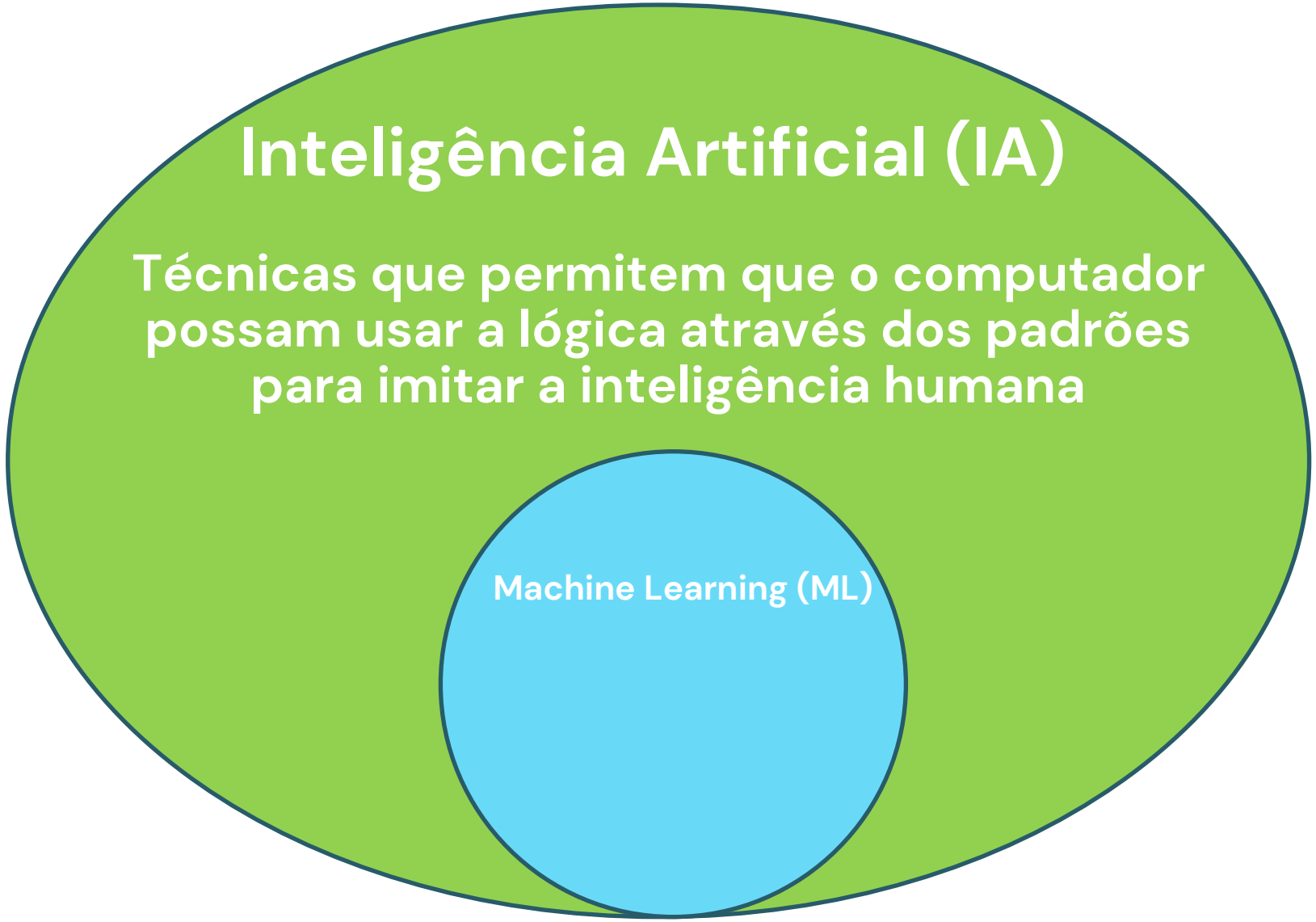
Tigre

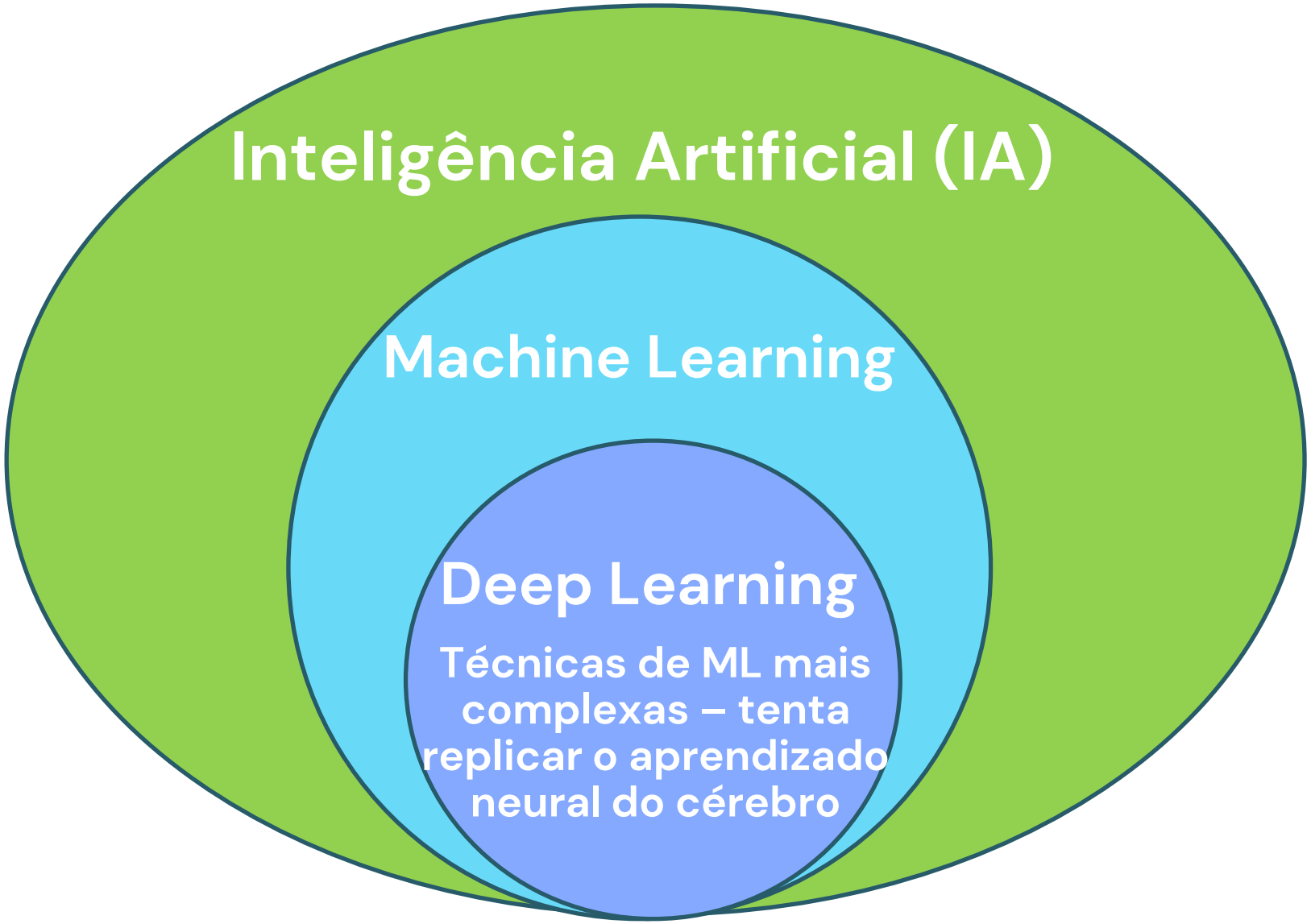
Qual a diferença?

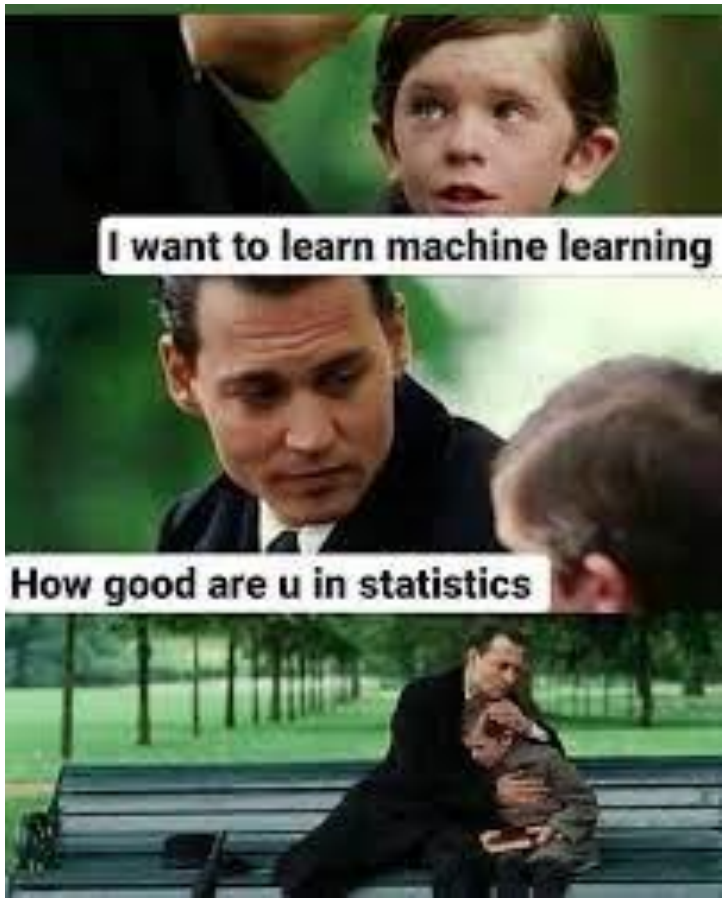
Aprendizado dos Padrões

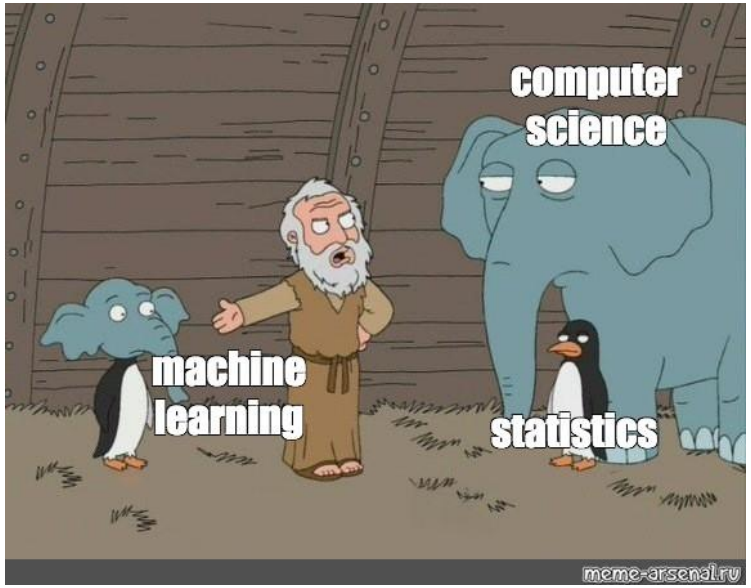
Machine Learning (ML)

Conjunto de técnicas estatísticas que permitem a máquina identificar padrões, melhorar tarefas por meio da experiência


Inteligência Artificial (IA)
Técnicas que permitem que o computador possam usar a lógica através dos padrões para imitar a inteligência humana
Machine Learning (ML)


Inteligência Artificial (IA)
Machine Learning
Deep Learning
Técnicas de ML mais complexas – tenta replicar o aprendizado neural do cérebro


I want to learn machine learning
How good are u in statistics


computer science
machine learning
statistics
meme-arsenal.ru


MACHINE LEARNING
Let's see who you really are!
STATISTICS

Estatística

Utiliza dados para identificar padrões e/ou correlações.

Machine Learning

Utiliza dados para aprender e fazer predições, através de dados para identificar padrões e/ou correlações.

Estatístico → Procura entender as relações das variáveis ao fazer previsões;

> O modelo prediz o resultado de y com 95% de precisão com base em nossos dados que possui nível de confiança de 98%

Machine Learner → Preocupa-se no resultado da previsão;

> O modelo prediz o resultado de y com 95% de precisão com base em nossos dados

Teoria

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2$$

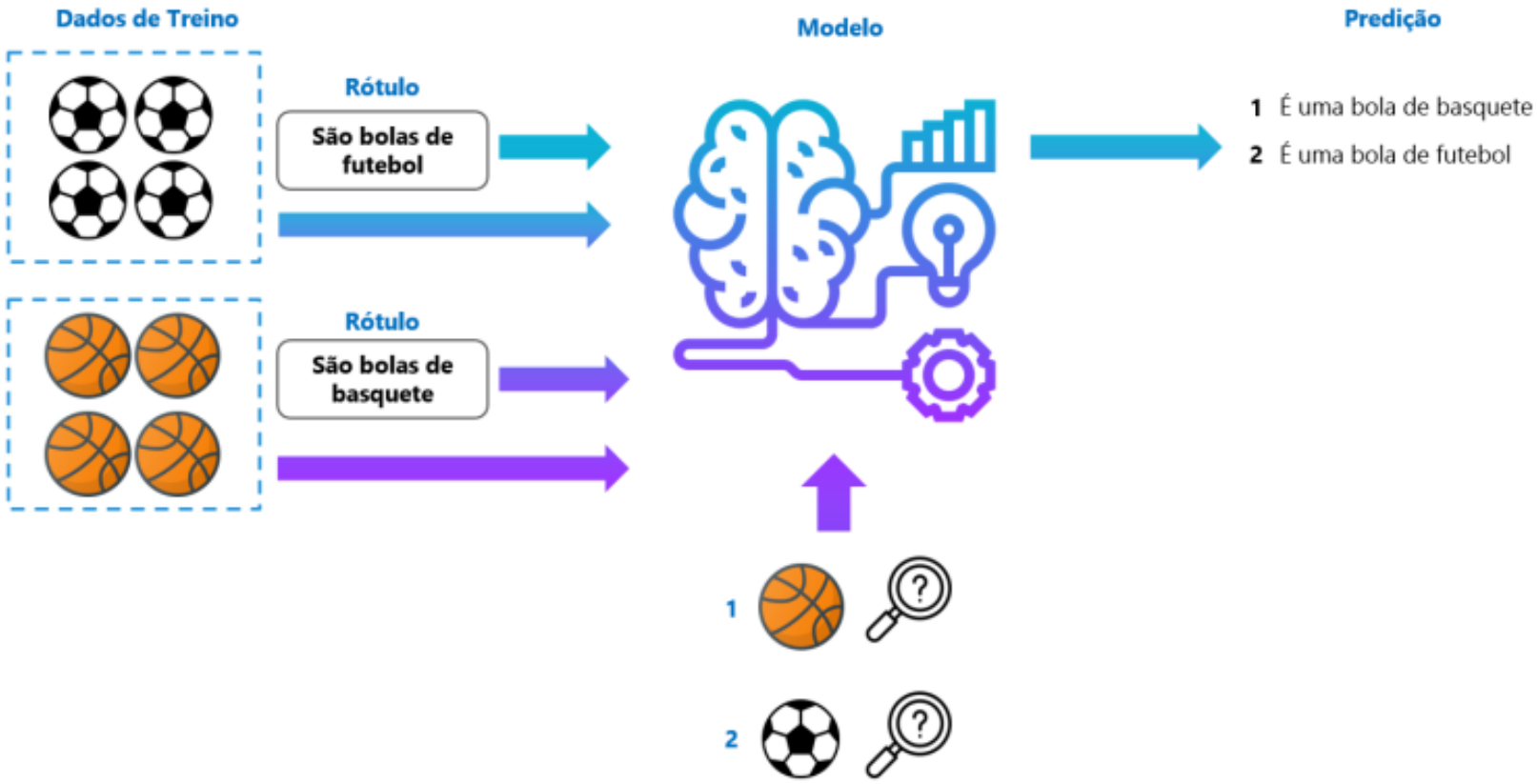
Prática

```
# Criando o classificador de árvore de decisão
clf = tree.DecisionTreeClassifier()

# Treinando o classificador com os dados
clf = clf.fit(X, Y)
```

Tipos de Aprendizado

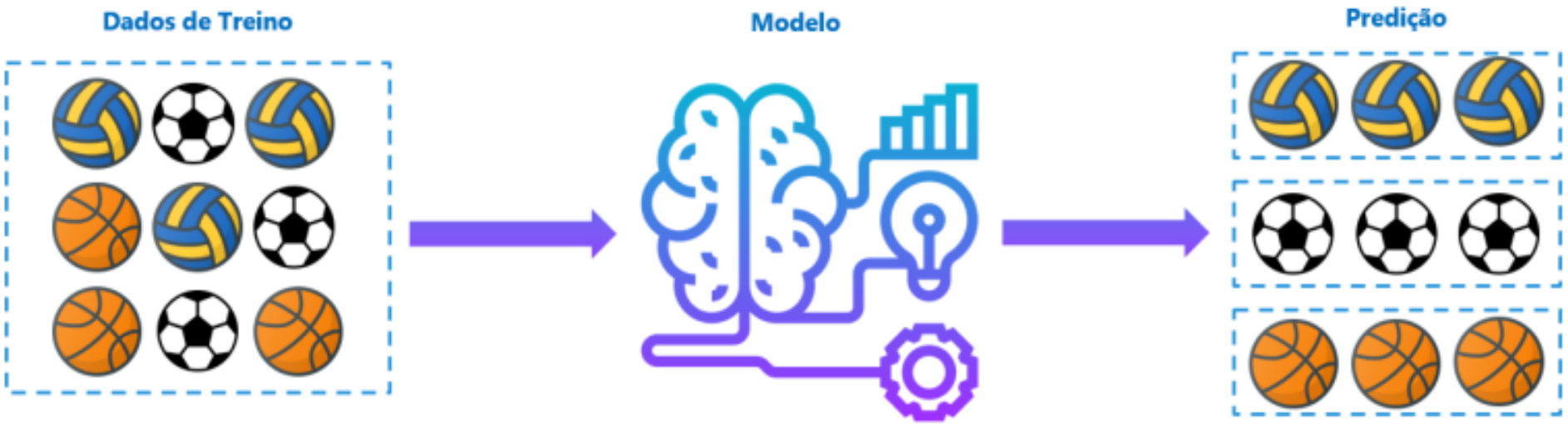
Supervisionado



Fonte: Machine Learning: Conceitos, capacidades e oportunidades 1ª edição:

Tipos de Aprendizado

Não Supervisionado



Fonte: Machine Learning: Conceitos, capacidades e oportunidades 1ª edição:

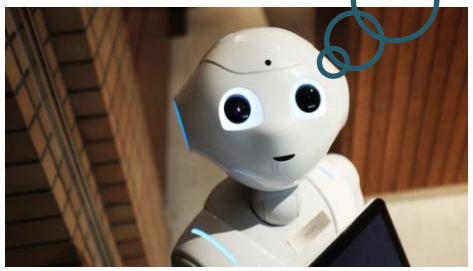
Tipos de Aprendizado

Por reforço

Tipos de Aprendizado

Por reforço

$$\sum_{t=0}^{t=\infty} \gamma^t r(x(t), a(t))$$

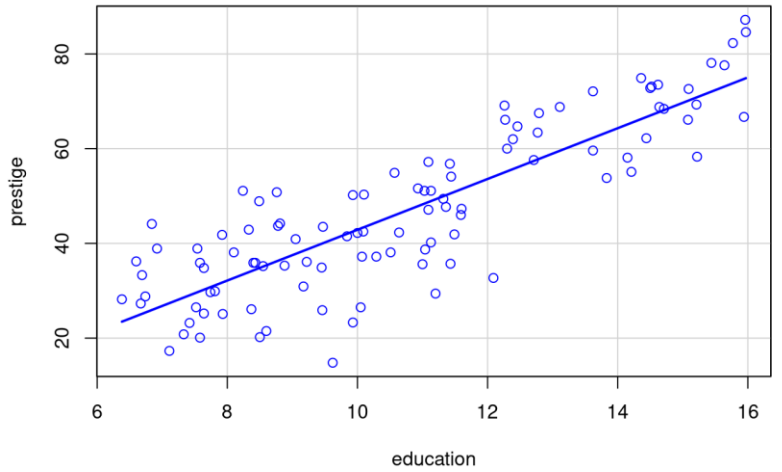


Programamos o algoritmo para maximizar uma função

Modelos de Aprendizado

Regressão

- Usada para prever um valor contínuo com base em variáveis independentes;
- Útil para entender relações entre variáveis;
- Prevê resultados numéricos.



Aplicação

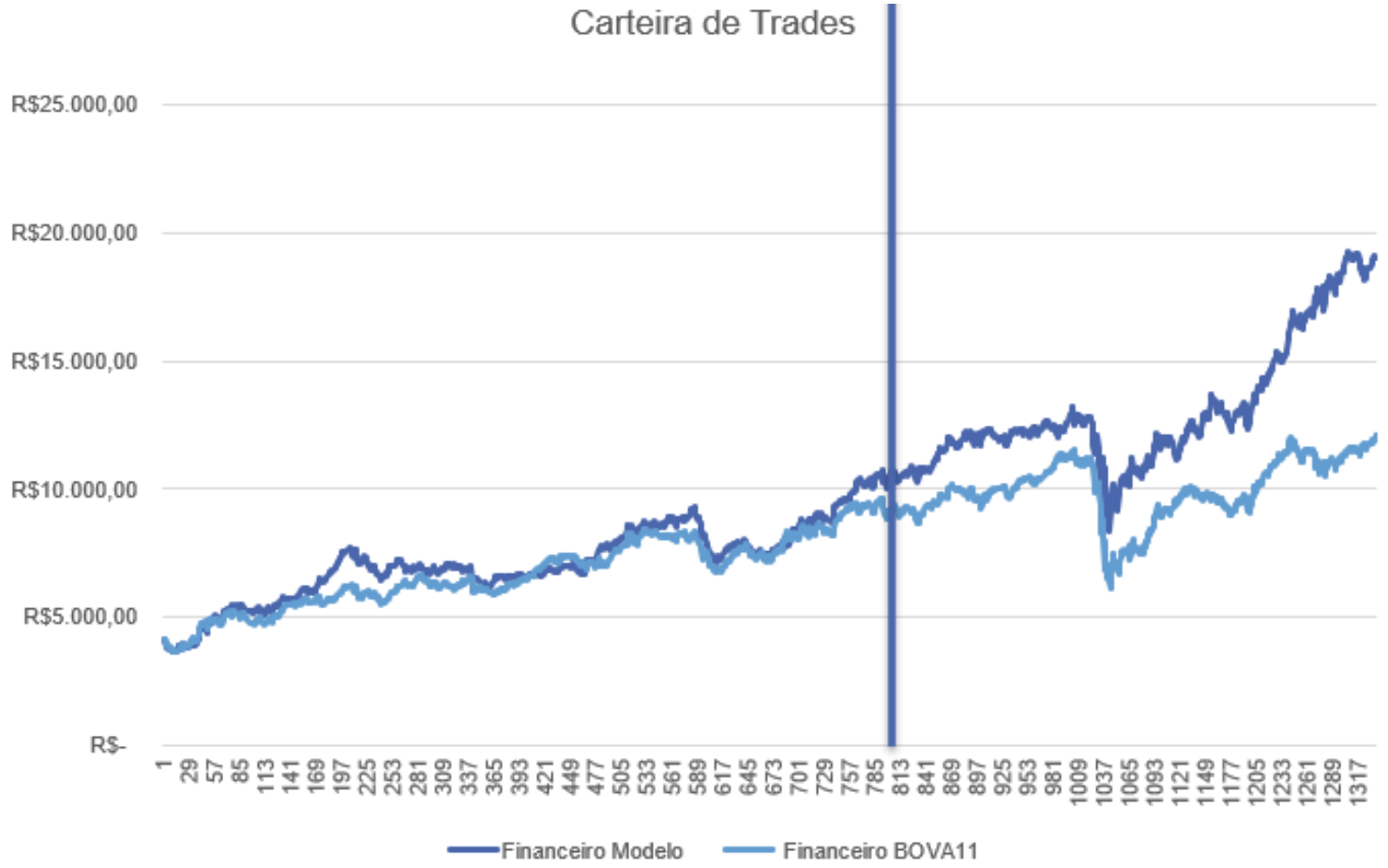
- Previsão de preços de imóveis;

Tamanho – localização – número de quartos

- Modelos de Investimento;

Modelos de Aprendizado

Regressão



Modelos de Aprendizado

Classificação

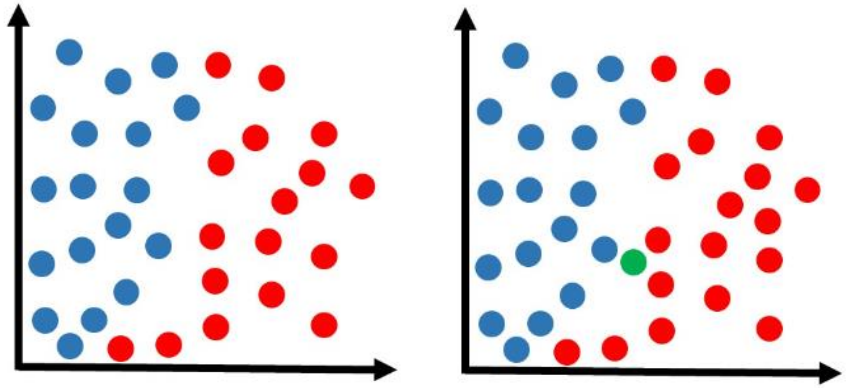
- Usada para categorizar dados em classes definidas;
- Identifica a classe à qual uma nova observação pertence.

Aplicação

- Detecção de Spam;
- Diagnóstico Médico;
- Carros autônomos.

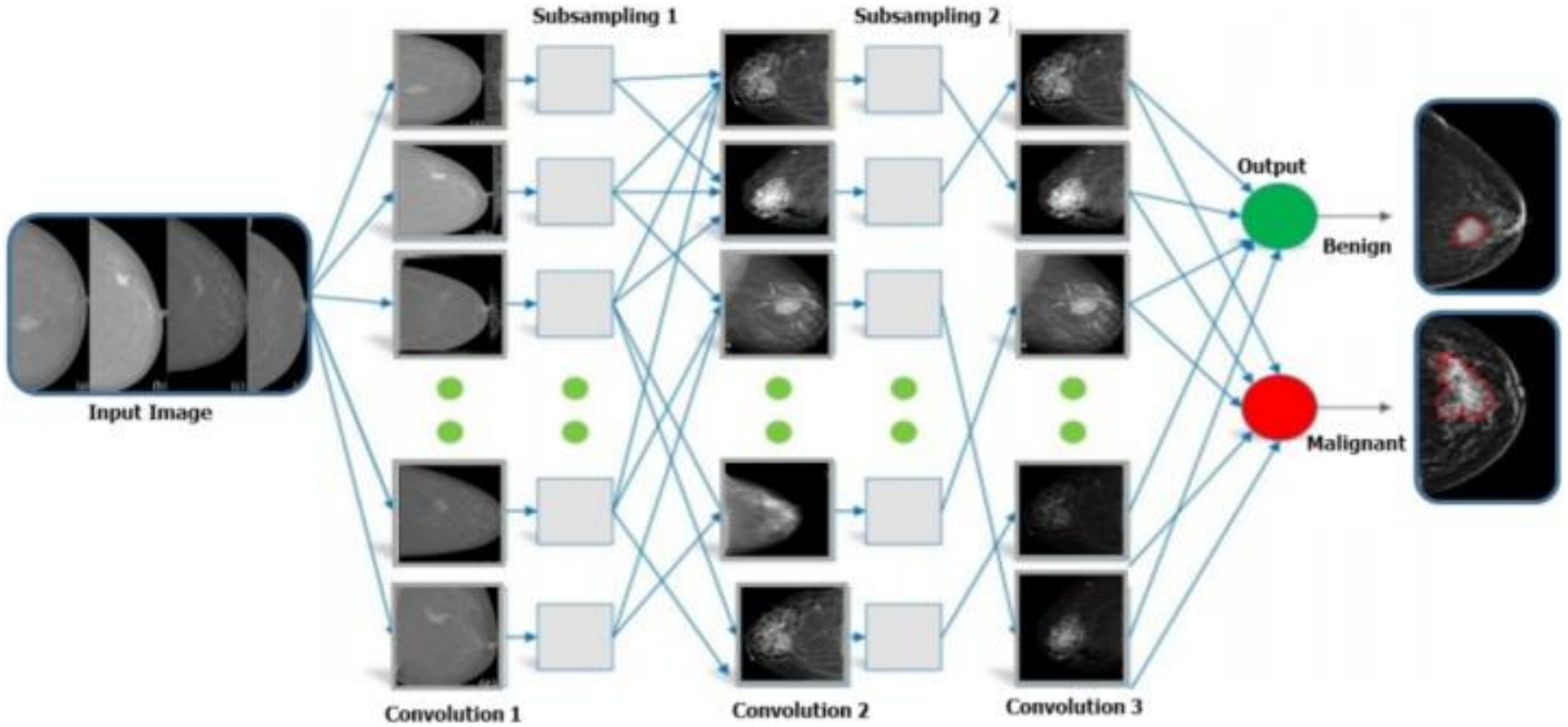
Modelos de Aprendizado

Classificação



Modelos de Aprendizado

Classificação



Modelos de Aprendizado

Clusterização

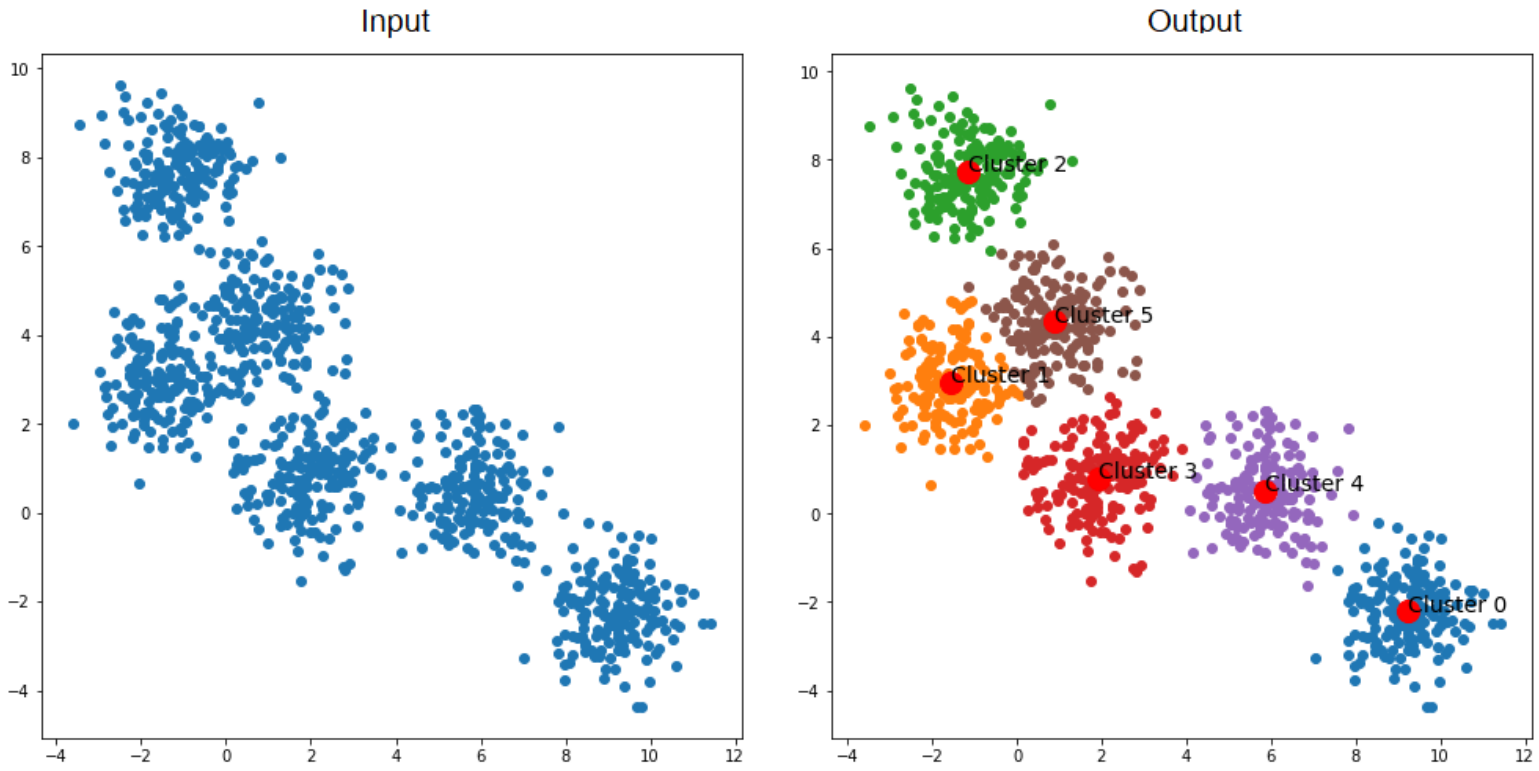
- Usada para agrupar dados similares em clusters;
- Não precisa de rótulos definidos (pode ser não supervisionado)

Aplicação

- Segmentação de Clientes (comportamento de compra, preferência de produtos, frequência de compra);
- Análise de Redes Sociais;

Modelos de Aprendizado

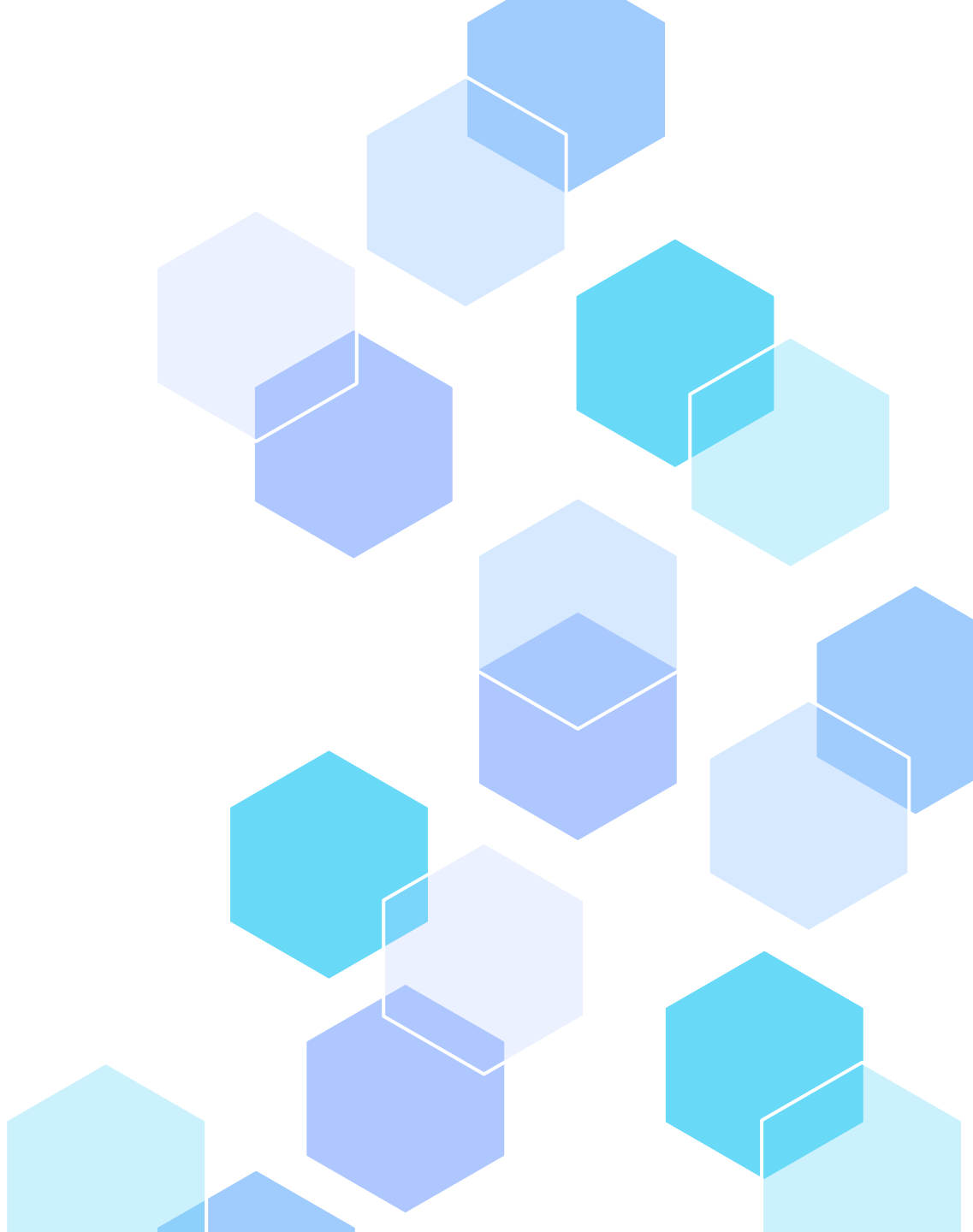
Clusterização



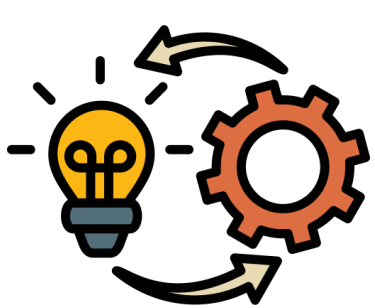
04

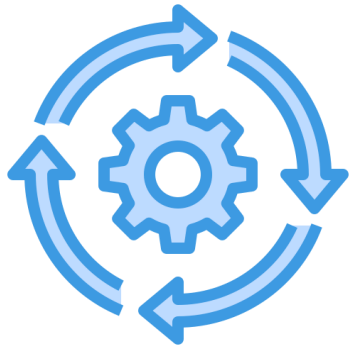
Prática

WorkFlow em uma base de dados de exemplo



Workflow





Entendimento do Problema

Coleta dos Dados

Entendimento dos dados

Organização e Preparação dos dados

Análise Exploratória dos Dados (EDA)

Feature Engineering

Formulação de Hipóteses

Escolha do Modelo

Validação do Modelo

Interpretação dos Resultados

Comunicação dos Resultados

Tomada de Decisão

Implementação

Workflow

Entendimento do Problema

Coleta dos Dados

Entendimento dos dados

Organização e Preparação dos dados

Análise Exploratória dos Dados (EDA)

Feature Engineering

Formulação de Hipóteses

Escolha do Modelo

Validação do Modelo

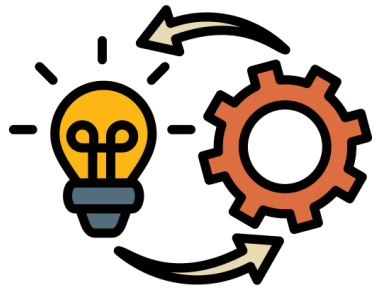
Interpretação dos Resultados

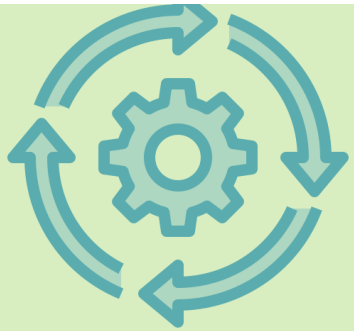
Comunicação dos Resultados

Tomada de Decisão

Implementação

70% a 80% do tempo





Caso Walmart

Caso Walmart

Caso Walmart

Hands On

O que é Ciência de Dados

“This is a quote, words full of wisdom that someone important said and can make the reader get inspired”

Whoa!

This can be the part of the presentation where you introduce yourself, write your email...

Sobre Mim

You can delete this slide when you're done editing the presentation

Fonts	To view this template correctly in PowerPoint, download and install the fonts we used
Used and alternative resources	An assortment of graphic resources that are suitable for use in this presentation
Thanks slide	You must keep it so that proper credits for our design are given
Colors	All the colors used in this presentation
Icons and infographic resources	These can be used in the template, and their size and color can be edited
Editable presentation theme	You can edit the master slides easily. For more info, click **here**

For more info:
SLIDESGO | BLOG | FAQs

You can visit our sister projects:
FREEPIK | FLATICON | STORYSET | WEPIK | VIDEVO

Introduction

Mercury is the closest planet to the Sun and the smallest one in the entire Solar System. Contrary to popular belief, this planet's name has nothing to do with the liquid metal. Mercury was, instead, named after the famous Roman messenger god Mercurius

Mercury takes a little more than 58 days to complete its rotation, so try to imagine how long days must be there! Since the temperatures are so extreme, albeit not as extreme as on Venus, Mercury has been deemed to be non-habitable for humans

Data collector & analysis

Do you know what helps you make your point crystal clear? Lists like this one:

- They’re simple
- You can organize your ideas clearly
- You’ll never forget to buy milk!

And the most important thing: the audience won’t miss the point of your presentation

Data collection techniques

Surveying

Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than the Moon

Observation

Venus has a beautiful name and is the second planet from the Sun. It's hot and has a very poisonous atmosphere

Data analysis demystified

Descriptive

Mercury is the closest planet to the Sun and the smallest of them all

Inferential

Venus has a beautiful name and is the second planet from the Sun

Predictive

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

Unlocking the power

Data insights

Mercury is the closest planet to the Sun and the smallest of them all

Predictive modeling

Venus has a beautiful name, but also very high temperatures

Optimization

Saturn is the second-largest planet in the Solar System

Data visualization

Jupiter is a gas giant and the biggest planet in the Solar System

Data analysis for beginners

Collection

Despite being red, Mars is a cold place

Cleaning

Venus has extremely high temperatures

Exploratory

Neptune is the farthest planet from the Sun

Statistical

Mercury is the closest planet to the Sun

Visualization

Saturn is a gas giant and has several rings

Hypothesis

Jupiter is the biggest planet of them all

Awesome words

“This is a quote, words full of wisdom that someone important said and can make the reader get inspired”

—Someone Famous

A picture is worth a thousand words

A picture always reinforces the concept

Images reveal large amounts of data, so remember: use an image instead of a long text. Your audience will appreciate it

98,300,000

Big numbers catch your audience’s attention

9h 55m 23s

Jupiter’s rotation period

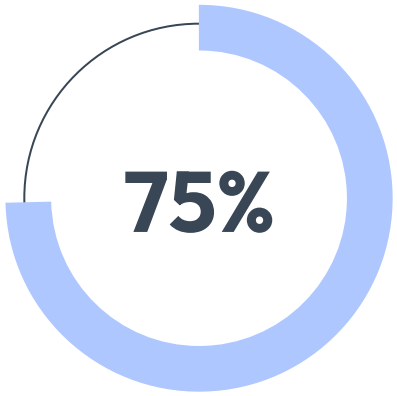
333,000

The Sun’s mass compared to Earth’s

386,000 km

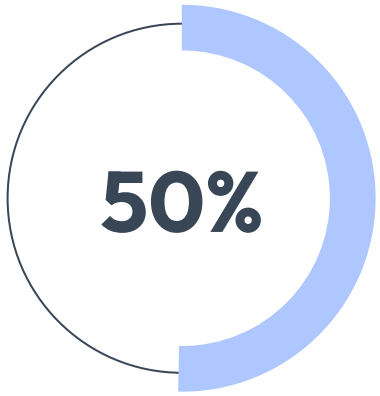
Distance between Earth and the Moon

Three percentages related



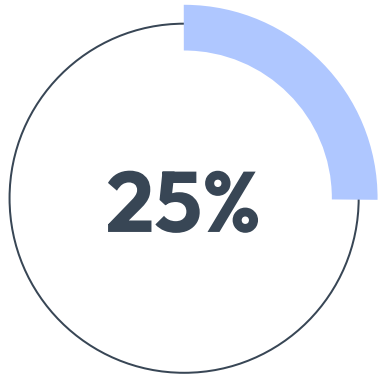
Collection

Mercury is the closest planet to the Sun and the smallest of them all



Quality

Venus has a beautiful name and is the second planet from the Sun

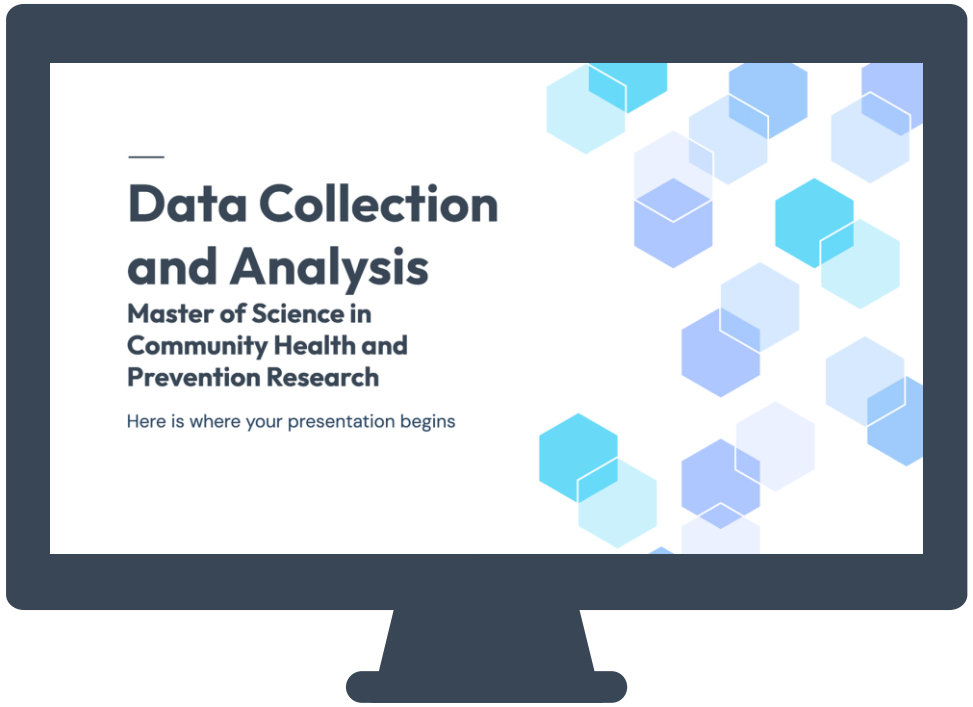


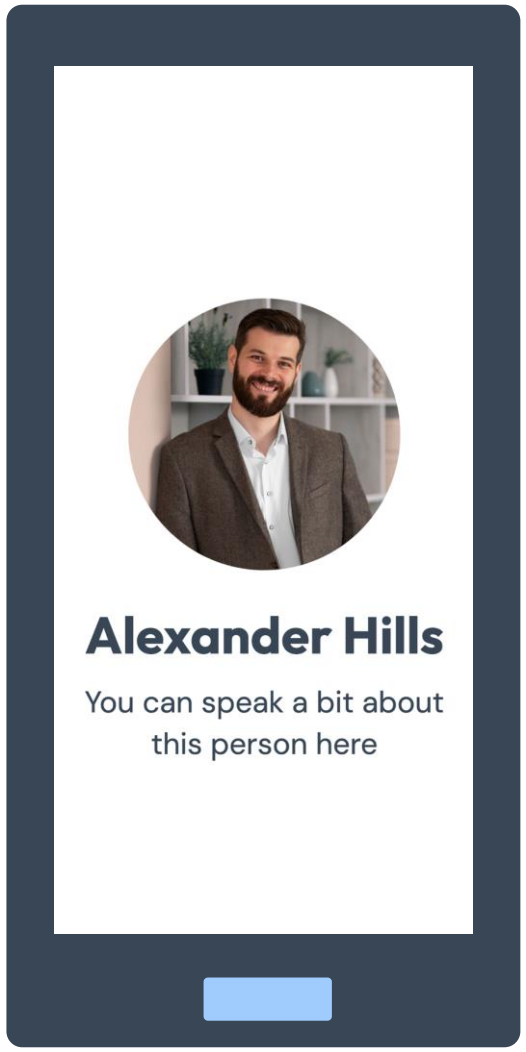
Insight

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

Computer mockup

You can replace the image on the screen with your own work. Just right-click on it and select “Replace image”

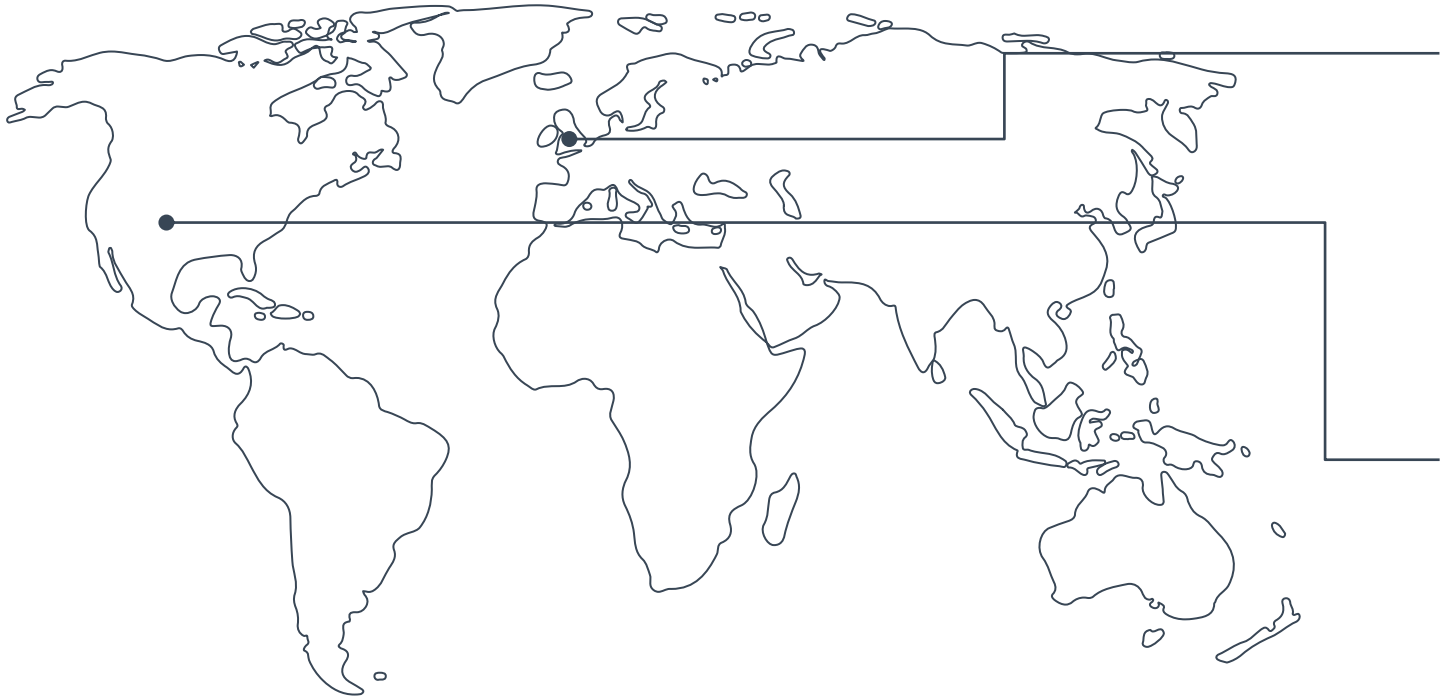




Phone mockup

You can replace the image on the screen with your own work. Just right-click on it and select “Replace image”

Where did we take the sample from?



United Kingdom

Mercury is the closest planet to the Sun and the smallest of them all

United States

Venus has a beautiful name and is the second planet from the Sun

Historical

1XXX

Venus has a beautiful name

1XXX

Mercury is very small planet

1XXX

Mars is full of iron oxide dust

1XXX

Jupiter is an enormous planet

2XXX

Saturn is a gas giant and has rings

2XXX

Uranus is one of the ice giants

2XXX

Earth is the planet that harbors life

Now

Neptune is very far from the Sun

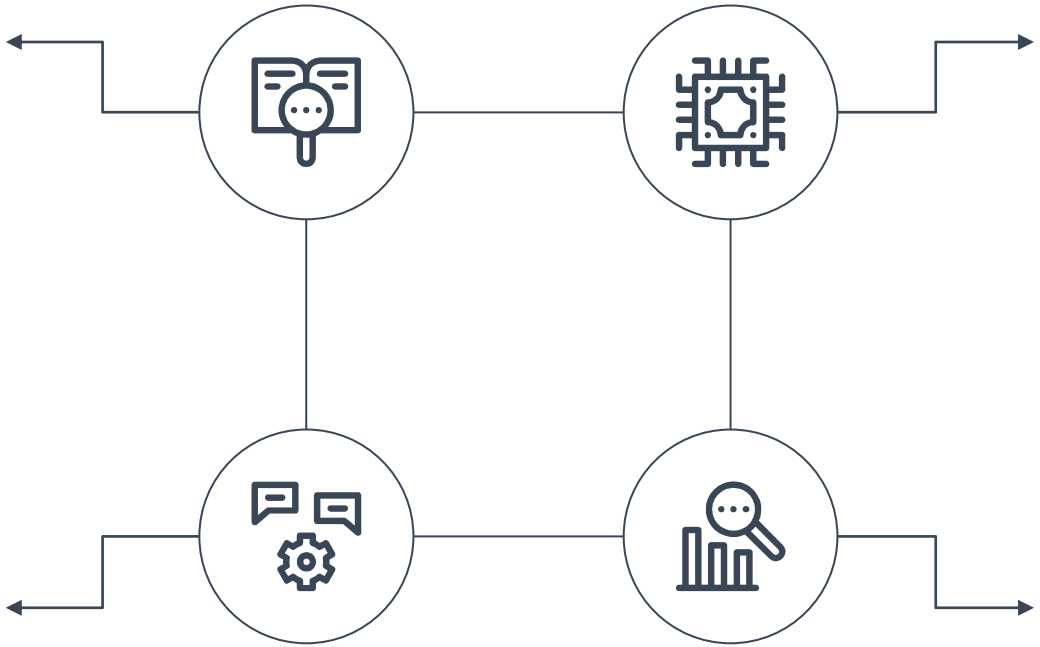
Data collection: best practices

Clarity

Mars is actually a very cold place

Precision

Saturn is a gas giant and has several rings



Privacy

Mercury is the closest planet to the Sun

Variety

Venus has extremely high temperatures

Example of data collection tasks

Tasks	Descriptions	Timeline
Task A	You can add a description here	1 week
Task B	You can add a description here	2 weeks
Task C	You can add a description here	3 weeks
Task D	You can add a description here	4 weeks

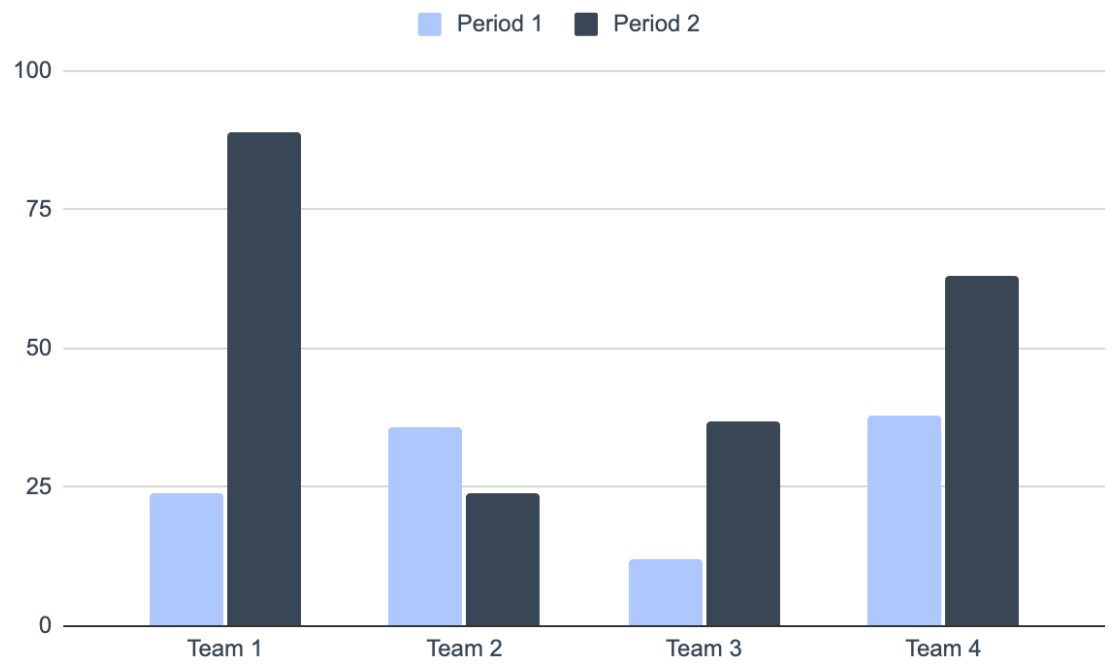
You can use this graph

- **Mercury**

  Mercury is the closest planet to the Sun and the smallest of them all

- **Venus**

  Venus has a beautiful name and is the second planet from the Sun



Follow the link in the graph to modify its data and then paste the new one here. **For more info, click here**

Our team

Alexander Hill

You can speak a bit about this person here

Maria Karla

You can speak a bit about this person here

Data collection vs data analysis

Data collection

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

Venus has a beautiful name and is the second planet from the Sun

Neptune is the farthest planet from the Sun and the fourth-largest

Data analysis

Mercury is the closest planet to the Sun and the smallest of them all

Saturn has several rings. It’s composed mostly of hydrogen and helium

Jupiter is a gas giant and the biggest planet in the Solar System

Project planner

November 2XXX							
Project	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Project 1							
Project 2							
Project 3							
Project 4							
Project 5							

Assignment brief

Qualification	Mercury is a small planet
Unit number & tittle	Venus is a hot planet
Learning aim(s)	We all live on Earth
Assignment title	Saturn is a gas giant and has rings. It's composed mostly of hydrogen and helium
Assessor	Jupiter is a huge gas giant
Issue date	12/10/2XXX
Hand in deadline	12/12/2XXX

Scenario	Community health and prevention research relies heavily on the collection and analysis of data to inform effective interventions. In this essay, critically examine the impact of different data collection methods on the success of community health initiatives

Task 1	Discuss the advantages and limitations of various data collection techniques, such as surveys, interviews, observational studies, and secondary data analysis, and how they influence the development and implementation of prevention strategies

Rubric

	Level 1	Level 2	Level 3	Level 4
Performance 1	You can add a description here	You can add a description here	You can add a description here	You can add a description here
Performance 2	You can add a description here	You can add a description here	You can add a description here	You can add a description here
Performanc	You can add a	You can add a	You can add a	You can add a

Exercise 1

What is the term for the process of ensuring that data collected is accurate and reliable?

A	Data analysis
B	Data interpretation
C	Data validation
D	Data visualization

In data collection, what is the term for a representative subset of a population?

A	Sample
B	Survey
C	Experiment
D	Observation

Exercise 2

Explain what is the difference between data collection and data analysis?	How to unlock the power of data collection and analysis?
You can answer here	You can answer here

What’s the difference?

Data collection

Survey

Mars is actually a very cold place

Interview

Venus has high temperatures

Observation

Earth is the planet with life

Data analysis

Descriptive

Jupiter is a huge gas giant

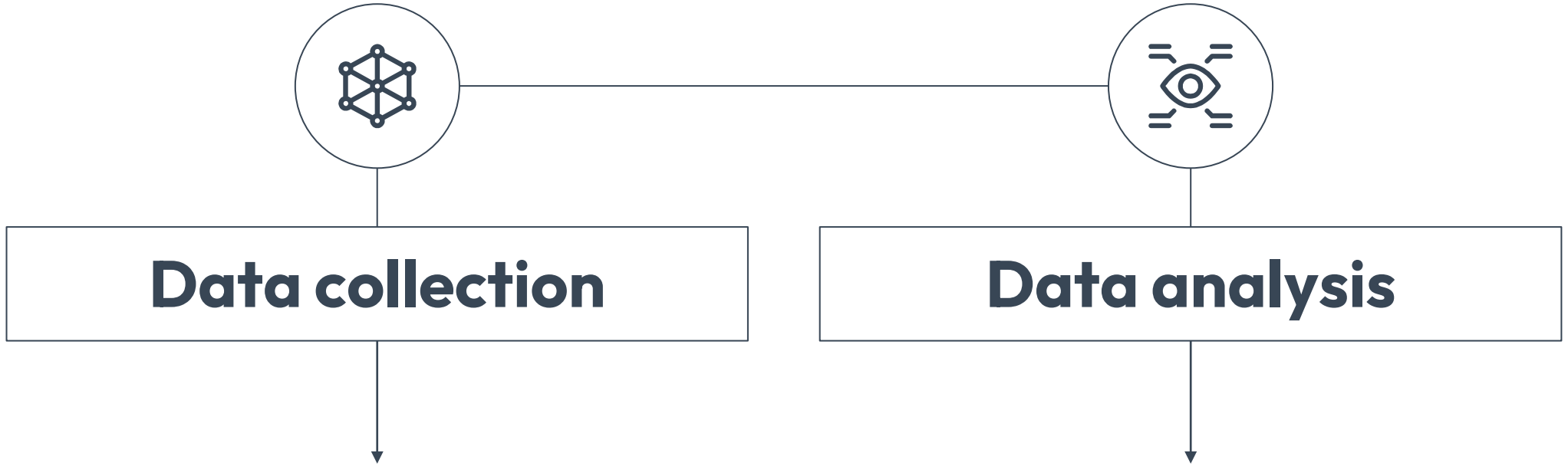
Inferential

Mercury is a small planet

Hypothesis

Saturn has several rings

Let’s review before wrapping up



Information-gathering

Venus has a beautiful name and is the second planet from the Sun

Insights-generation

Despite being red, Mars is actually a cold place. It’s full of iron oxide dust

Thanks!

Do you have any questions?

youremail@freepik.com
+34 654 321 432
yourwebsite.com

Icon pack

Alternative resources

Here’s an assortment of alternative resources whose style fits that of this template:

Vectors

- Geometric models background in flat design

Resources

Did you like the resources used in this template? Get them on these websites:

Vectors

- International nurses day template

Icons

- Icon Pack: Research and Development | Lineal

Photos

- Student posing during a group study session with colleagues
- Friends learning in a study group
- Front view smiley man holding laptop
- Portrait of senior woman in professional blazer outdoors

Instructions for use

If you have a free account, in order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide. Please refer to the next slide to read the instructions for premium users.

As a Free user, you are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit our blog:
https://slidesgo.com/faqs and **https://slidesgo.com/slidesgo-school**

Instructions for use (premium users)

As a Premium user, you can use this template without attributing **Slidesgo** or keeping the **Thanks** slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the “Thanks” slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit our blog:
https://slidesgo.com/faqs and **https://slidesgo.com/slidesgo-school**

Fonts & colors used

This presentation has been made using the following fonts:

Outfit
(https://fonts.google.com/specimen/Outfit)

DM Sans
(https://fonts.google.com/specimen/DM+Sans)

#384655 #ffffff #afc7ff #9fcbfd #68daf8

Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out **how it works**.

Pana

Amico

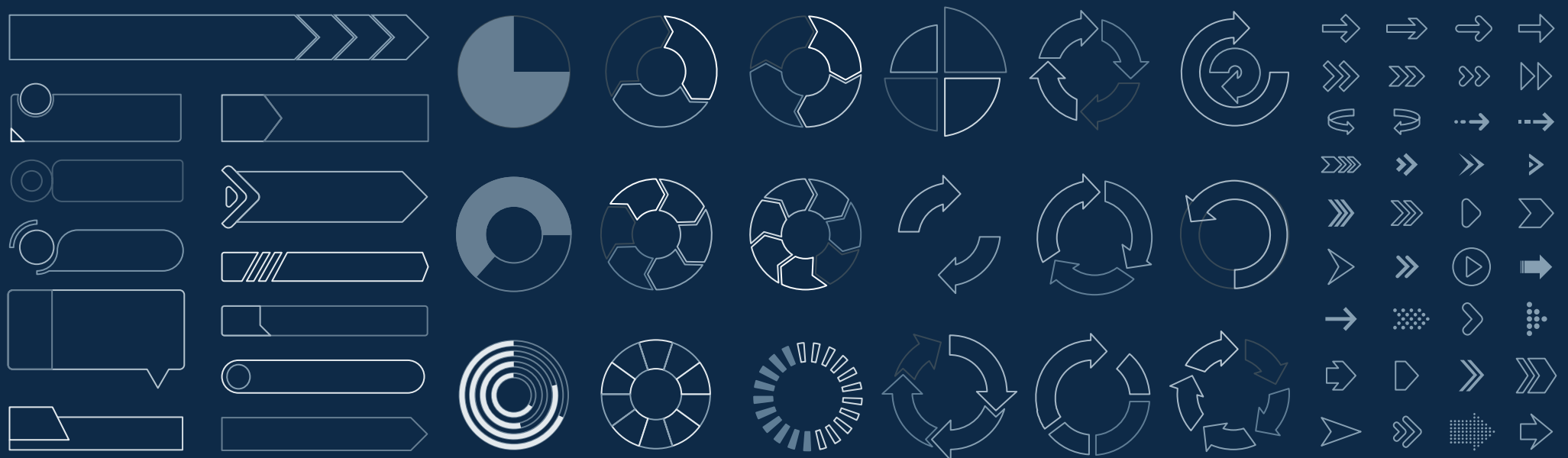
Bro

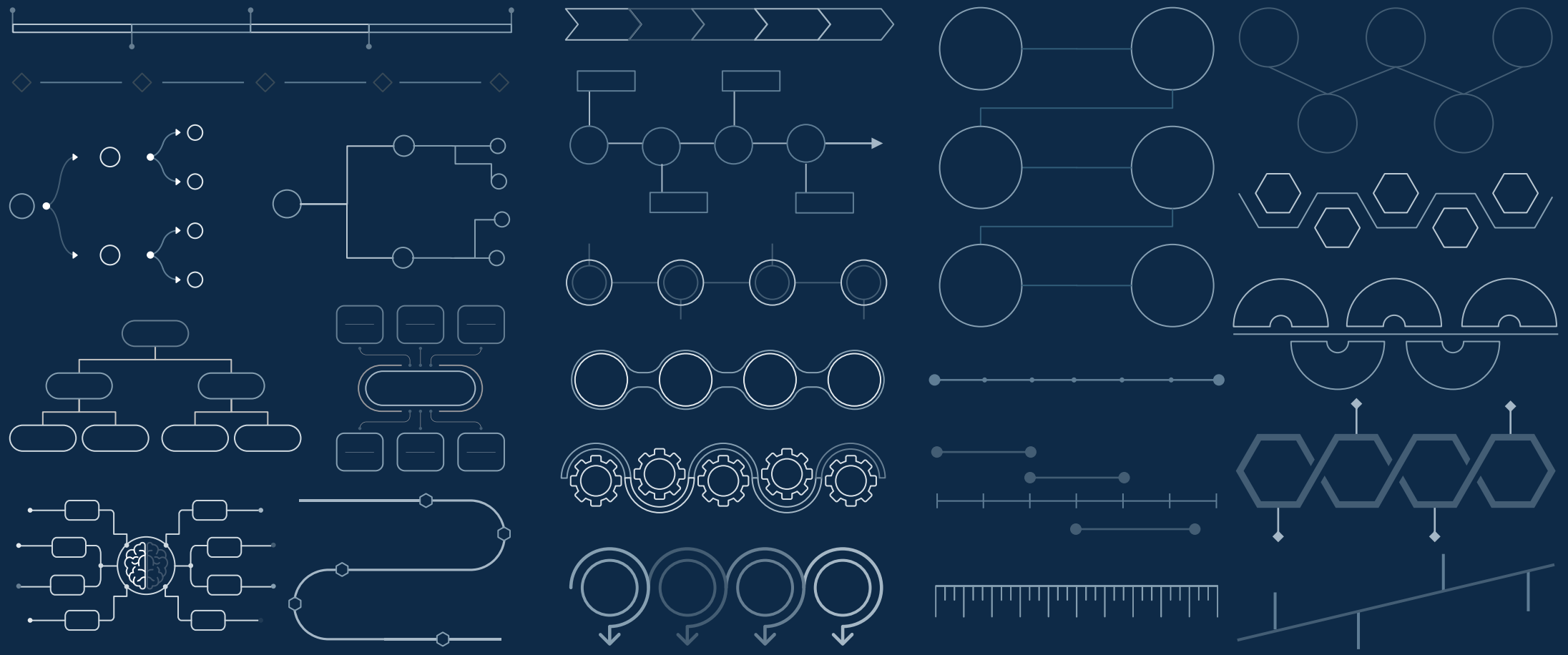
Rafiki

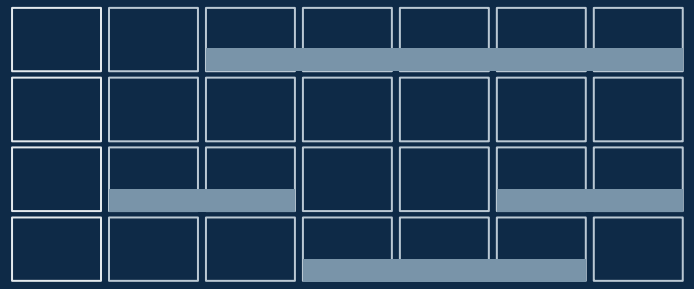
Cuate

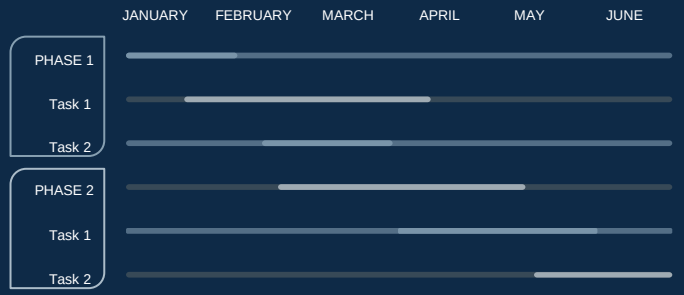
Use our editable graphic resources...

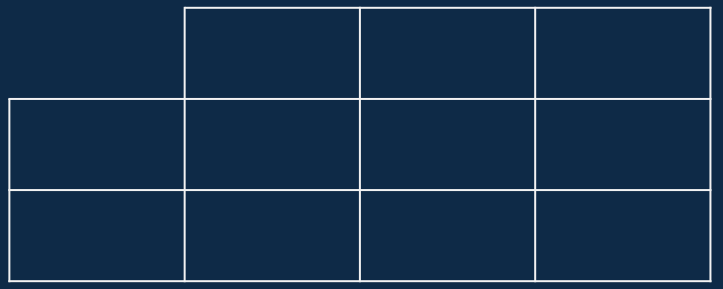
You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.

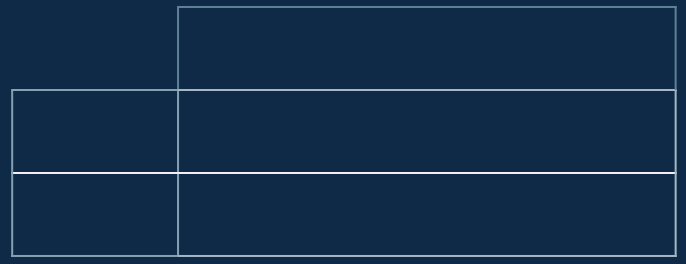


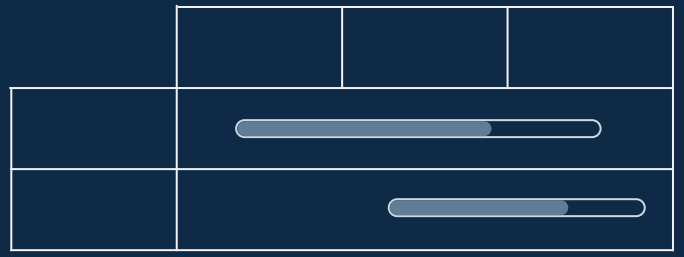


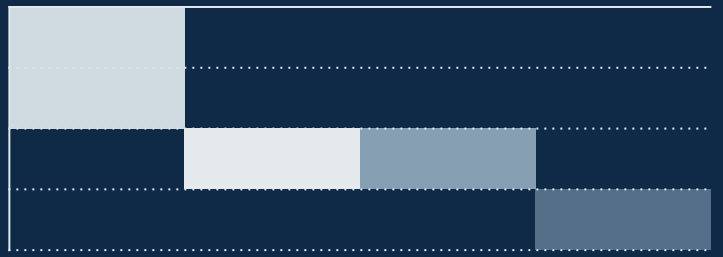


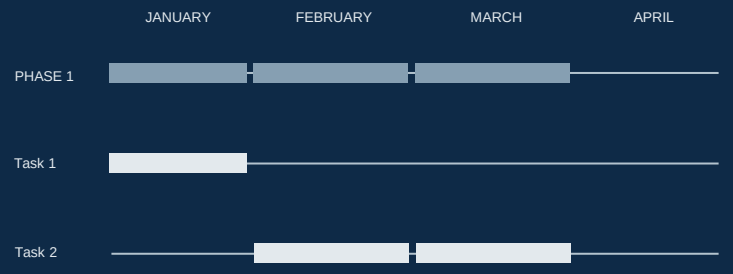

JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2

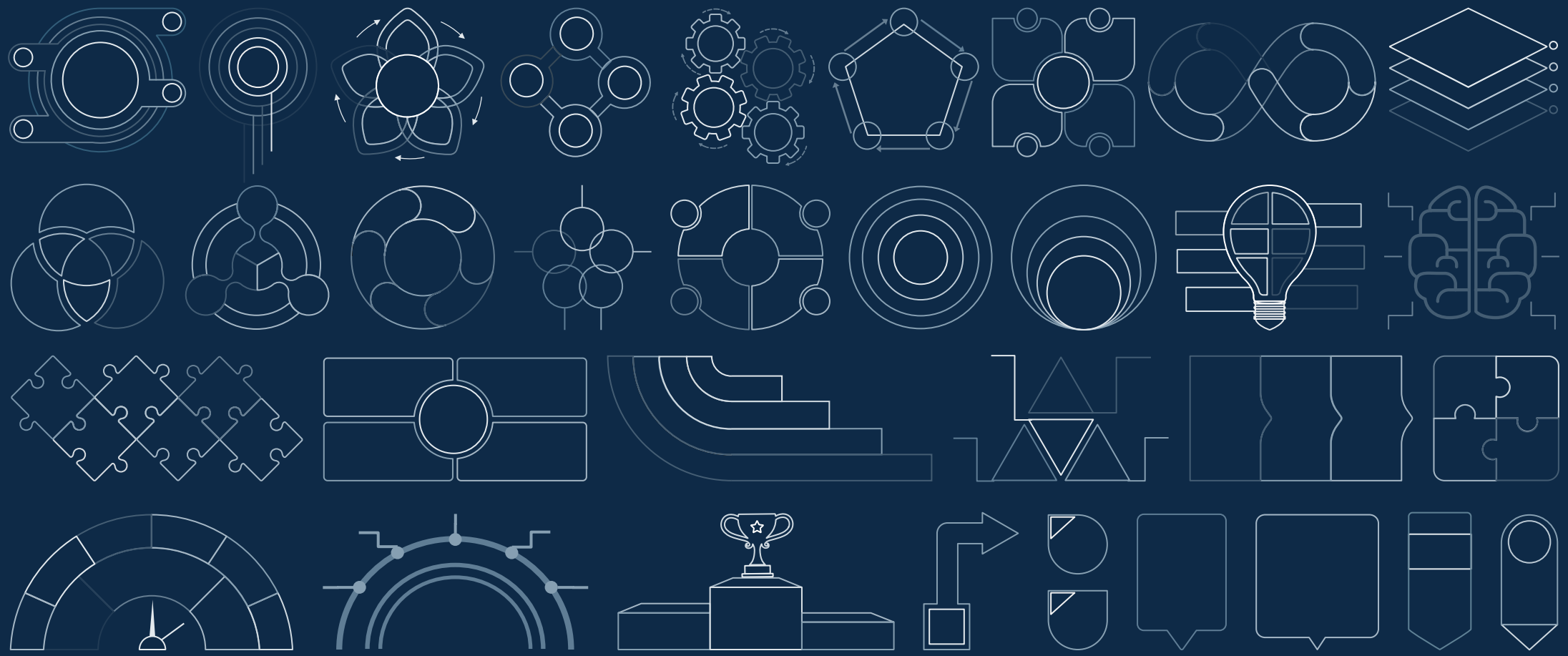


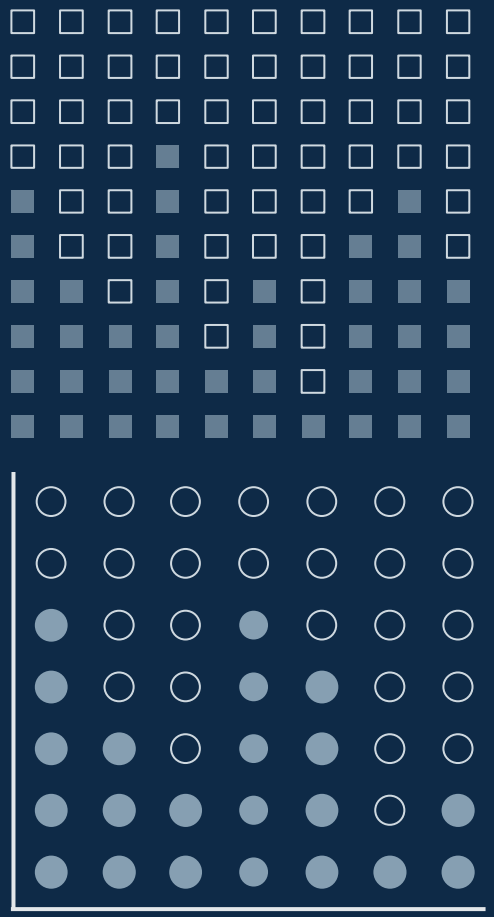
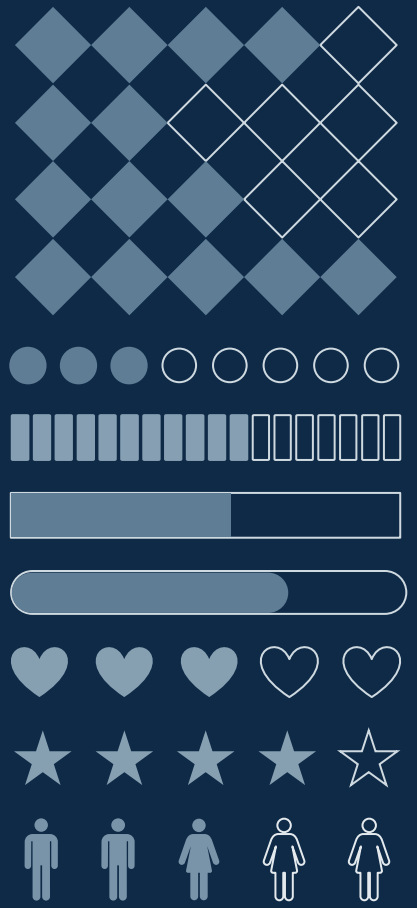
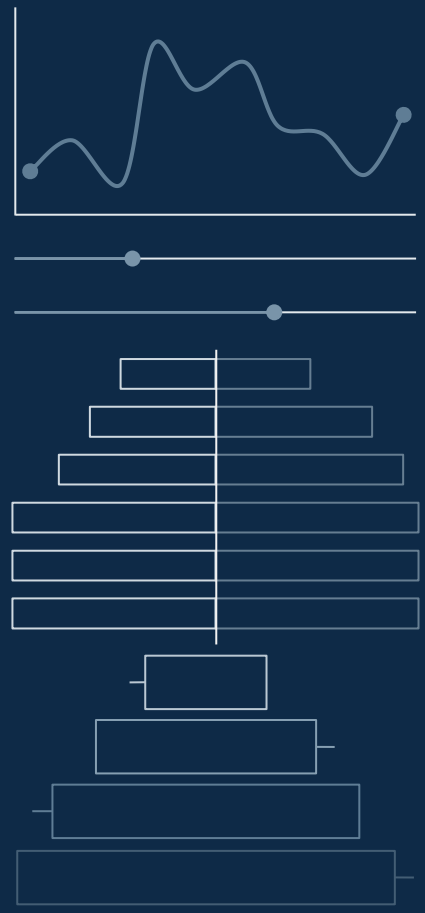
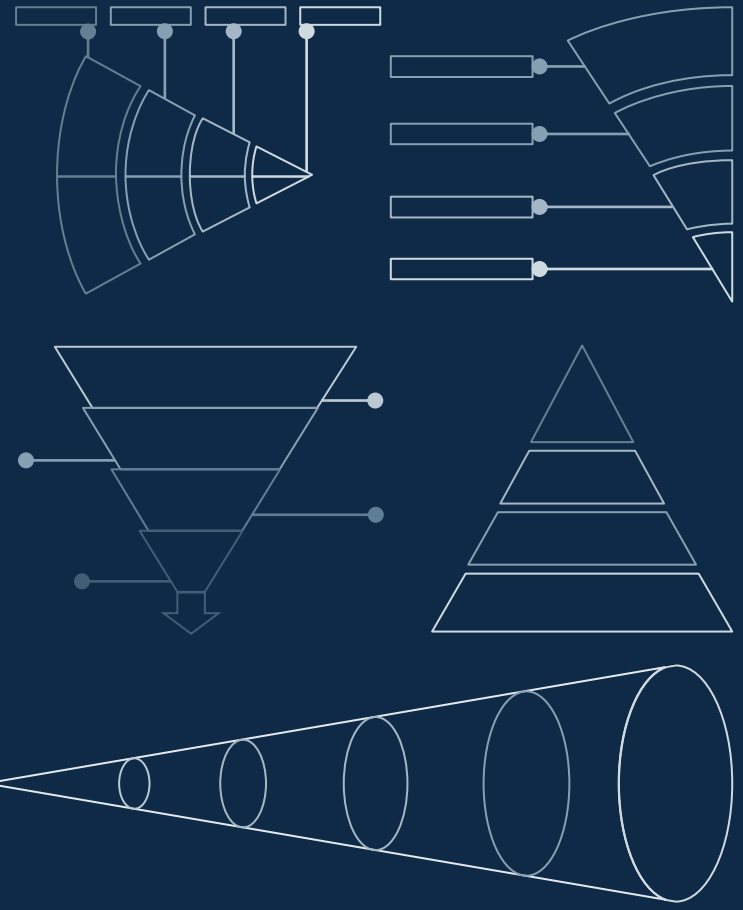







JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2



...and our sets of editable icons

You can **resize** these icons without losing quality.
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.
In Google Slides, you can also use **Flaticon's extension**, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons


slidesgo