



# Apresentação do Projeto Grupal

Abordagem para Identificação dos Fatores  
Determinantes no Abandono Acadêmico no Ensino  
Superior Utilizando Algoritmos de IA



# O QUE SERÁ ABORDADO?

01

## Conjunto de Dados

Qual o conjunto escolhido e porquê

02

## Abordagem do Problema

Nosso processo de pesquisa e análise de dados

03

## Treinamento dos Modelos

Treinamento, validação e escolha de modelos

04

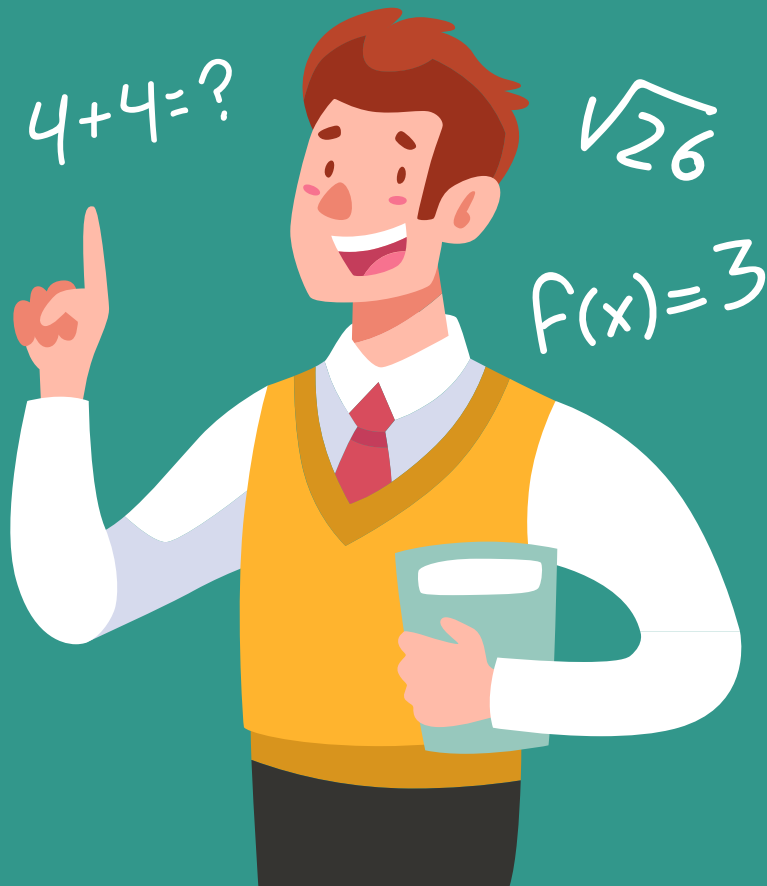
## Interpretabilidade

Quais foram as características mais impactantes

# Problema da Evasão

Considera-se o fenómeno uma das principais preocupações do Ministério da Educação, a evasão realizada antes da formação completa do curso e é visto como um alvo a ser combatido ou um índice a ser reduzido

(Lima Coimbra and Costa 2021)




# INTRODUÇÃO

## Por quê desse tema?

- 4º Item dos Objetivos Sustentáveis
- Tema relevante em todo o mundo
  - Mesmo que os motivos sejam geralmente diferentes
- Taxa de evasão alta no Brasil
- Ministério da Educação:
  - (a) abandono (deixar de se matricular);
  - (b) desistência (formalização junto da universidade)
  - (c) exclusão por norma institucional
  - (d) transferência (mudança de curso)



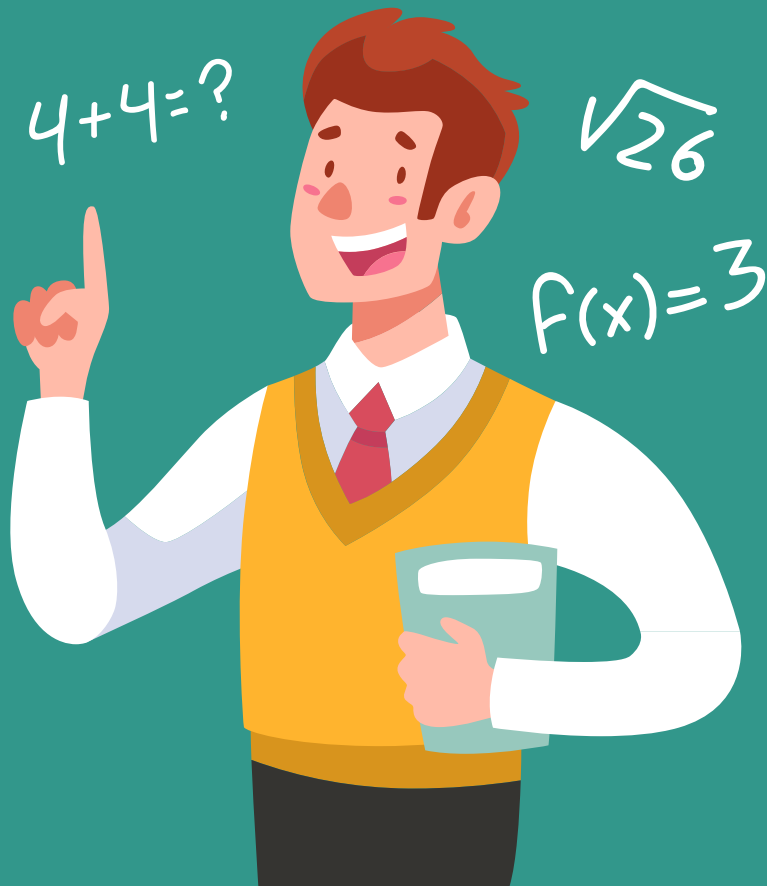


Alguns dos motivos que podem explicar este fenômeno são a dificuldade de conciliar os estudos com o seu emprego, perda de interesse no curso, uma infraestrutura precarizada por parte da instituição de ensino que não atende aos interesses e expectativas do aluno, entre outros.

(Cunha and Morosini 2014)

# O QUE FIZEMOS?

Uma análise exploratória e o desenvolvimento de um modelo de aprendizado de máquina preditiva, a fim de compreender o problema do êxodo de estudantes no ensino superior. Assim sendo, é possível relacionar os motivos mais prováveis de tal fenômeno para contribuir à discussão sobre problema.





01

# CONJUNTO DE DADOS

Qual o conjunto escolhido



# 4.424

Instâncias contidas no conjunto de dados original, cada registro contém informações de alunos de graduação, tais como fatores internos da universidade, rendimento acadêmico e de fatores socioeconômicos que os estudantes enfrentam durante a graduação





# LOCAL DOS DADOS



Dados obtidos de um artigo intitulado “**Early prediction of student’s performance in higher education: a case study**” realizado por alunos do **Instituto Politécnico de Portalegre (Portugal)**. Consultar referências ao final da apresentação

# CARACTERÍSTICAS DO CONJUNTO DE DADOS

4.424

Alunos

36

Colunas contendo fatores  
internos da universidade e  
fatores socioeconômicos

0

Valores nulos



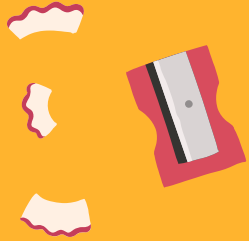
02

## ABORDAGEM DO PROBLEMA

Nosso processo de pesquisa e  
análise de dados



# PROCESSO DE PESQUISA



## **Análise Exploratória**

Compreensão dos dados e  
remoção de colunas

## **Pré-Processamento**

Adequação dos dados para  
treinamento dos modelos

## **Modelagem e Armazenamento**

Treinamento, avaliação,  
escolha e armazenamento



# SEPARAÇÃO E ANÁLISE DOS DADOS

## NUMÉRICAS

16 colunas com  
valores numéricos  
discretos e contínuos

## CATEGÓRICAS

20 colunas com  
valores categóricos  
pré codificados

# ANÁLISE DAS VARIÁVEIS NUMÉRICAS



## CORRELAÇÃO DE PEARSON

Usamos para remover colunas que possuíam pouca correlação com a variável alvo



## Fator de Inflação da Variância - VIF

Utilizamos o Fator de Inflação da Variância (VIF) para eliminar variáveis com  $VIF > 5$  e alta colinearidade

# VARIÁVEIS NUMÉRICAS REMOVIDAS

1

Curricular units 1st sem (credited)  
Curricular units 1st sem (enrolled)  
Curricular units 1st sem (grade)  
Curricular units 1st sem (evaluations)  
Curricular units 1st sem (without evaluations)

2

Curricular units 2nd sem (credited)  
Curricular units 2nd sem (enrolled)  
Curricular units 2nd sem (grade)  
Curricular units 2nd sem (evaluations)  
Curricular units 2nd sem (without evaluations)

# ANÁLISE DAS VARIÁVEIS CATEGÓRICAS



## CHI2 EM TABELA DE CONTINGÊNCIA

Avaliamos a associação entre variáveis categóricas e a variável alvo



## VALOR V DE CRAMER

Foi usado para medir a força da associação das variáveis



# VARIÁVEIS CATEGÓRICAS REMOVIDAS

- 1 Nacionalidade
- 2 Educational special needs
- 3 International



**03**

# **TREINAMENTO DOS MODELOS**

Treinamento, validação e  
escolha de modelos



# Pré-Processamento



## One Hot Encoding

Aplicado nas  
variáveis  
categóricas



## Normalização

Aplicado nos variáveis  
numéricas  
(usamos min-max)



## Rebalanceamento

Ligações de TOMEK  
(classe majoritária).

SMOTE (não Majoritária).



## Agrupamento

Optamos por juntar as  
classes “Graduate” e  
“Enrolled”, para converter  
o problema em um caso  
binário.

Observação: Foram testadas diferentes combinações dessas técnicas de  
pré-processamento.

# PIPELINE DE TREINAMENTO DOS MODELOS

## Preparação do treino e teste

Separação e seleção de  
features

01



02



03



04



## Treino e seleção de melhor modelo

Grid Search de  
Hiperparâmetros

## Avaliação do melhor modelo

Medidas de  
desempenho no teste

## Armazenamento

Armazenamento dos  
melhores modelos

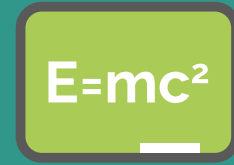
# ALGORITMOS USADOS



Naive Bayes



Árvore de Decisão



Regressão Logística

Observação: Todo o pipeline foi montado para ser executado em cada um desse algoritmos

# PREPARAR TREINO E TESTE



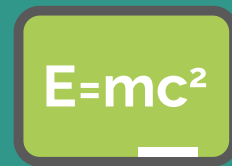
## Treino/Validação e Teste

Separação de 80/20 para treino e teste.  
[Validação foi feita dentro do conjunto de treino]



## Seleção de Features

Dentro do conjunto de treino temos o uso da função Select K-Best.  
[K final = 60]



## Refletindo Mudanças

Refletimos as mudanças no conjunto de teste

# AVALIAÇÃO E SELEÇÃO DOS MODELOS

Algoritmo	Conjunto de Dados	Hiperparâmetros	Acurácia* (%)	F1* (%)
Naive Bayes	Balanceamento original	'alpha': 0.1	74,37	57,74
	Rebalanceado	'alpha': 0.1	76,26	76,02
Árvore de Decisão	Balanceamento original	'max_depth': 5, 'min_samples_leaf': 4, 'min_samples_split': 2	85,50	75,54
	Rebalanceado	'max_depth': 3, 'min_samples_leaf': 1, 'min_samples_split': 2	87,46	87,78
Regressão Logística	Balanceamento original	'C': 10, 'class_weight': None, 'penalty': 'l1', 'solver': 'liblinear'	86,12	76,74
	Rebalanceado	'C': 10, 'class_weight': None, 'penalty': 'l1', 'solver': 'liblinear'	86,54	86,35

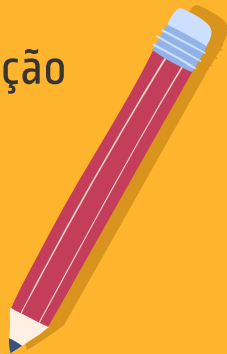
**Acurácia\* e F1\*:** Média das medidas de cada fold de uma validação cruzada de 5 folds dentro do conjunto de treinamento.

# ARMAZENAMENTO DOS MODELOS



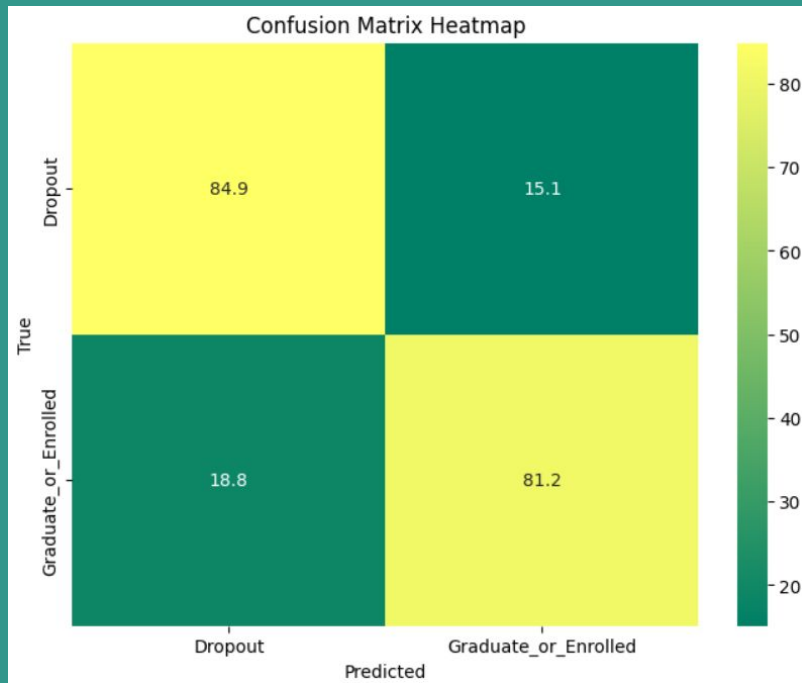
Optamos por armazenar o melhor modelo em arquivos .pkl

Com seu nome obedecendo a seguinte codificação:  
a{acurácia}\_f{medida f1}.pkl, para facilitar a identificação





# MODELO ESCOLHIDO



Matriz de confusão

## Regressão Logística

- **Conjunto de dados:** Rebalanceado
- **Hiperparâmetros:** 'C': 10, 'class\_weight': 'balanced', 'penalty': 'l1', 'solver': 'liblinear'
- **Justificativa:** Melhor desempenho nas medidas acurácia e F1, além de ter melhor desempenho em ambas as classes



04

## INTERPRETABILIDADE

Quais foram as características mais impactantes

# Variáveis mais importantes usando SHAP values



Curricular units 2nd  
sem (approved)



Tuition fees up to  
date



Scholarship holder



Previous qualification



Mother's qualification  
and occupation



Father's qualification  
and occupation

# Obrigado!

## Integrantes:

Daniel Yuji Yamada

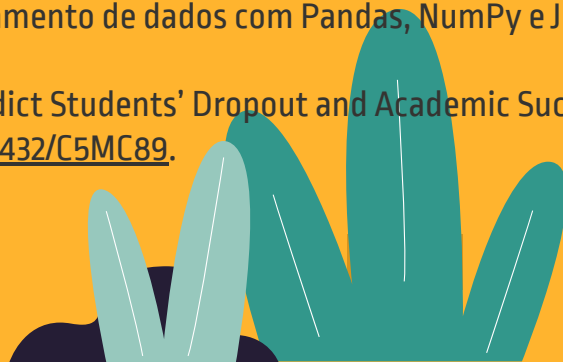
Felipe Mateos Castro de Souza



**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

# REFERÊNCIAS

- Bowyer, K. W.; Chawla, N. V.; Hall, L. O.; and Kegelmeyer, W. P. 2011. SMOTE: synthetic minority over-sampling technique. CoRR abs/1106.1813
- Cunha, E. R., and Morosini, M. C. 2014. Evasão na educação superior: uma temática em discussão. Revista Cocar 7[14]:82–89.
- Jeatrakul, P.; Wong, K. W.; and Fung, C. C. 2010. Classification of imbalanced data by combining the complementary neural network and smote algorithm. In Wong, K. W.; Mendis, B. S. U.; and Bouzerdoun, A., eds., Neural Information Processing. Models and Applications, 152–159. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lima Coimbra, Camila, V. M. M. B. e. S. L., and Costa, N. C. D. 2021. A evasão na educação superior: definições e trajetórias. SciELO Brazil. DOI: <https://doi.org/10.1590/S1678-4634202147228764>.
- McKinney, W., ed. 2023. Python para análise de dados: Tratamento de dados com Pandas, NumPy e Jupyter. O'Reilly.
- Realinho, Valentim, V. M. M. M. J., and Baptista, L. 2021. Predict Students' Dropout and Academic Success. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5MC89>.





# YOU CAN USE AN INFOGRAPHIC

## Reliability

Venus has a beautiful name, but it's hot

Despite being red, Mars is actually a cold place



Mercury is the closest planet to the Sun

Neptune is the farthest planet from the Sun

# 100,000

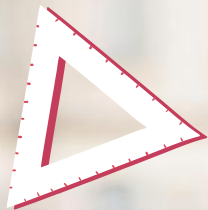
Big numbers catch your audience's attention





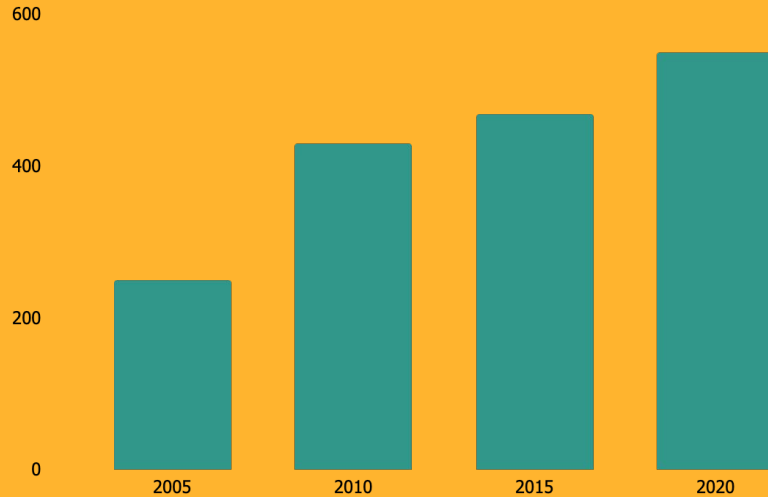


**A picture is worth  
a thousand words**



# STATISTICS

## Student Growth



To modify this graph, click on it, follow the link,  
change the data and paste the new graph here

## Student Hobbies



# OUR TEAM

**John James**

You can replace the  
image on the screen  
with your own



**Jenna Doe**

You can replace the  
image on the screen  
with your own



**Jane Patterson**

You can replace the  
image on the screen  
with your own



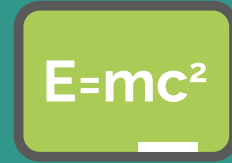
# OUR GOALS



Venus has a beautiful name, but it's really hot



Despite being red, Mars is actually a very cold place



Neptune is the farthest planet from the Sun

# PIPELINE

## Preparar treino e teste

Mercury is the closest planet to the Sun

01



02



## Treino e seleção de melhor modelo

Despite being red, Mars is a cold place

03



## Avaliação do melhor modelo

Venus is the second planet from the Sun

04



## Armazenamento

Saturn is composed of hydrogen and helium

# FINAL WORDS

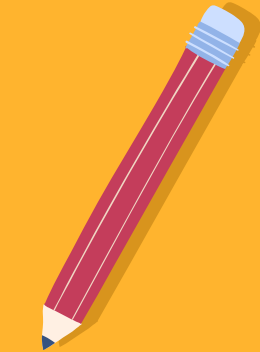


- 1 Venus has a beautiful name, but it's hot
- 2 Mercury is the closest planet to the Sun
- 3 Neptune is the farthest planet from the Sun

# REMINDERS



- Despite being red, Mars is actually a cold place. It's full of iron oxide, which gives the planet its reddish cast
- Jupiter is a gas giant and the biggest planet in our Solar System. It's also the fourth-brightest object in the sky
- Neptune is the farthest planet from the Sun and the fourth-largest by diameter in the Solar System





# Thanks!



Does anyone have any questions?



your-email@freepik.com

+91 620 421 838

yourcompany.com

Please keep this slide for attribution.



**CREDITS:** This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.



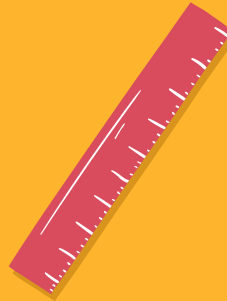
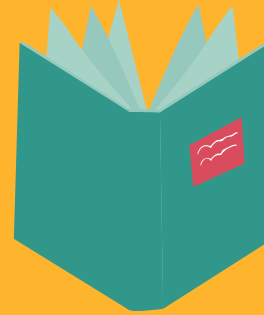
# ALTERNATIVE RESOURCES

## PHOTOS

- Close-up male student checking mobile phone

## VECTORS

- Hand drawn back to school squared sales banner



# RESOURCES

## PHOTOS

- Group of students happy to be back at university
- Stylish young student reading book
- Stylish student at campus looking away
- Portrait of handsome student at campus
- Back to school concept with woman studying in library
- Handsome male with marker and looking at camera
- Student with digital tablet at a library



## VECTORS

- Hand drawn back to school timetable
- School elements Icon Pack
- Happy teacher holding a book in front of chalkboard
- Hand-drawn creativity infographics template
- Flat design back to school timetable
- Appointment booking with calendar and man



# Instructions for use (free users)

In order to use this template, you must credit [Slidesgo](#) by keeping the Thanks slide.

## You are allowed to:

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

## You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- 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.
- Delete the “Thanks” or “Credits” slide.
- 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 Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Instructions for use (premium users)

In order to use this template, you must be a Premium user on [Slidesgo](#).

## 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 Slidesgo School:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

# Fonts & colors used

This presentation has been made using the following fonts:

## **Chau Philomene One**

(<https://fonts.google.com/specimen/Chau+Philomene+One>)

## **Ropa Sans**

(<https://fonts.google.com/specimen/Ropa+Sans>)

#353431

#ffffff

#97c8bd

#a7d3c7

#32978b

#ffb42e

#d8941d

#d84c5d

#d84c5d

# Stories by Freepik

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

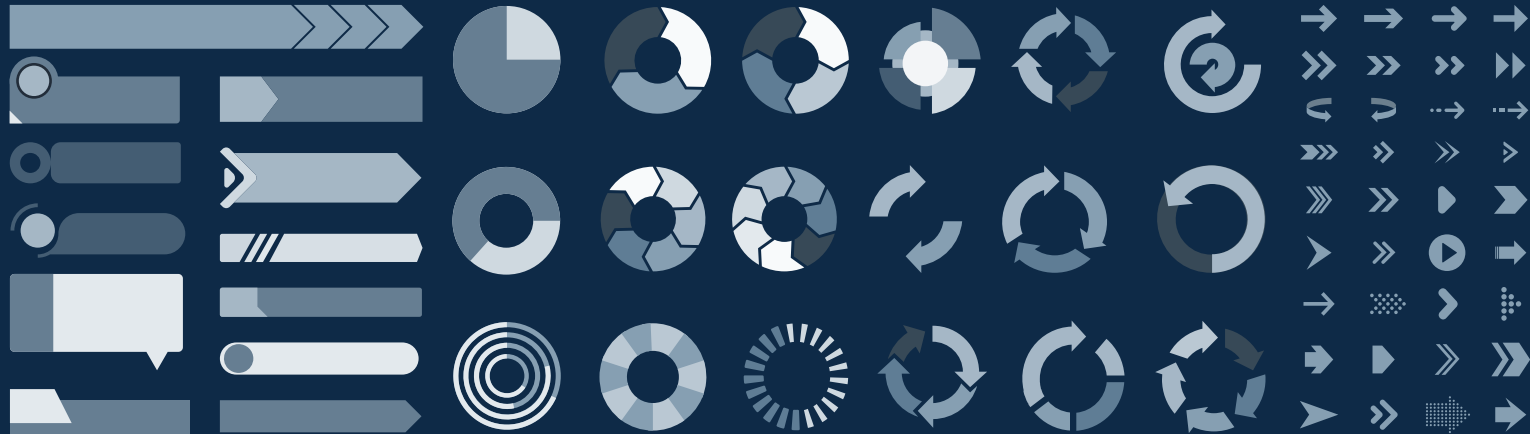


Cuate

# Use our editable graphic resources...

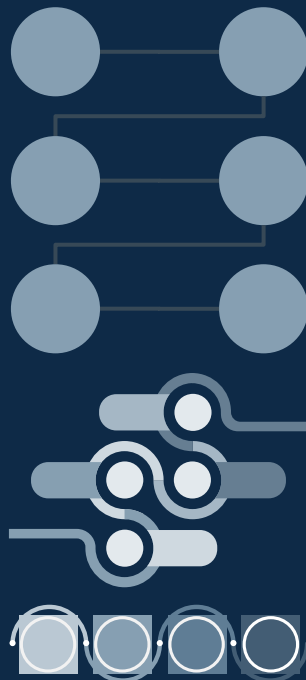
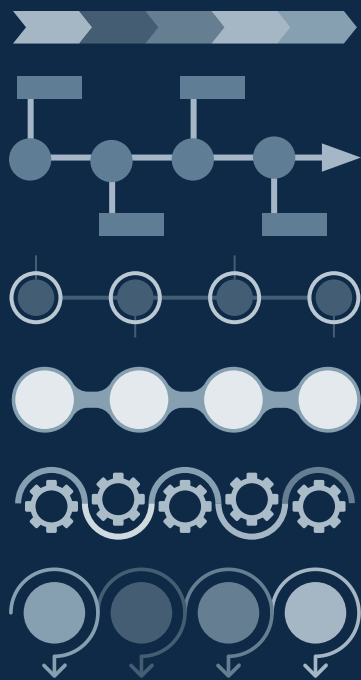
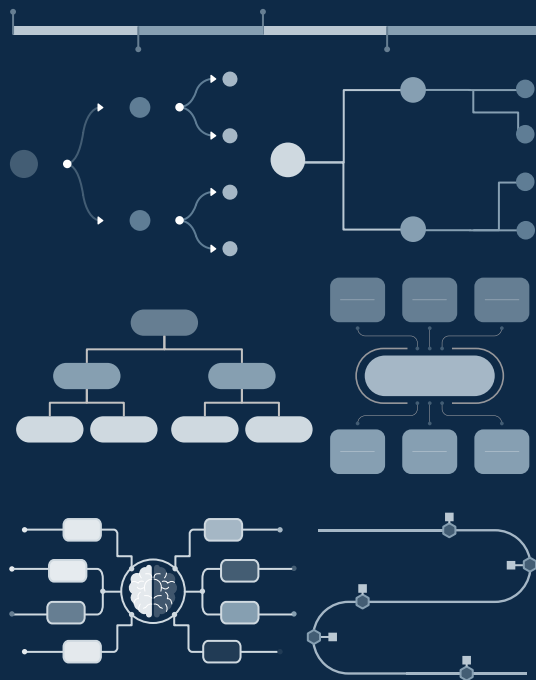
You can easily [resize](#) these resources without losing quality. To [change the color](#), just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

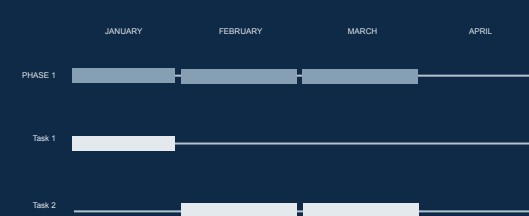
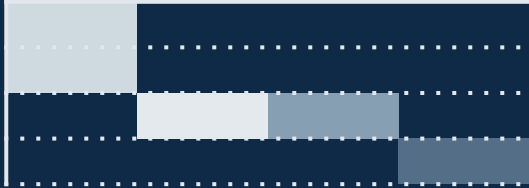
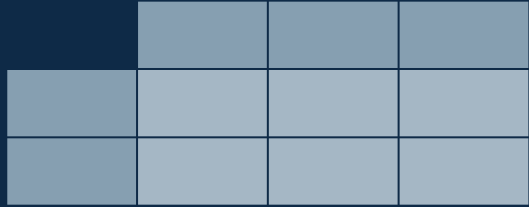
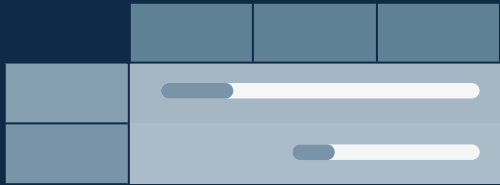
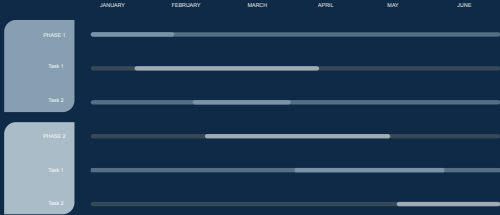
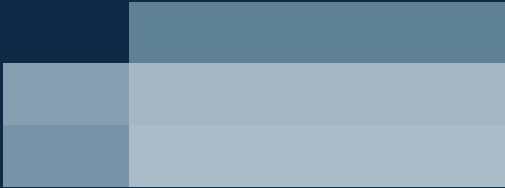
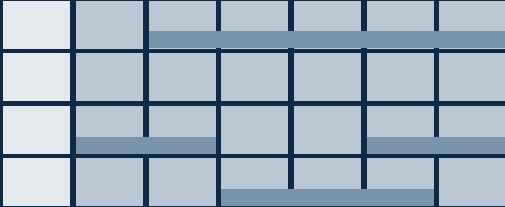
Group the resource again when you're done. You can also look for more [infographics](#) on [Slidesgo](#).

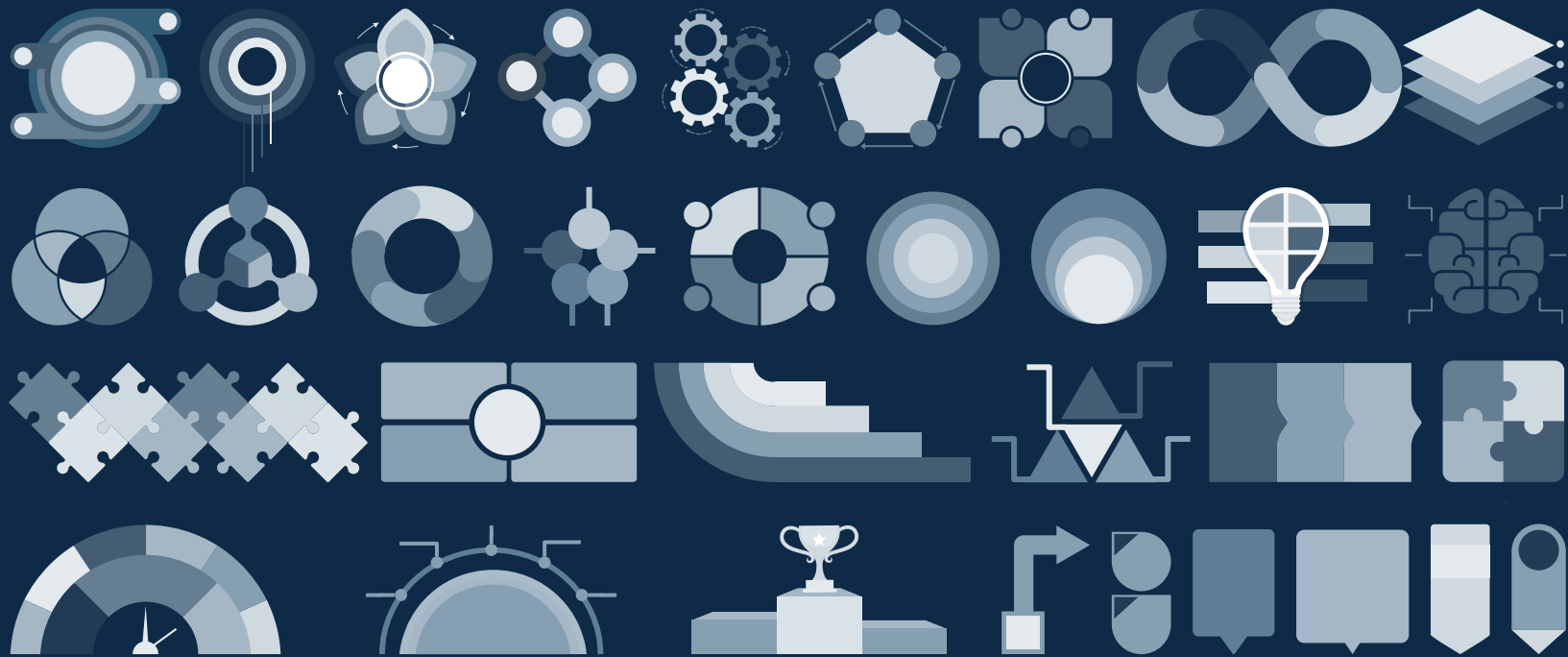


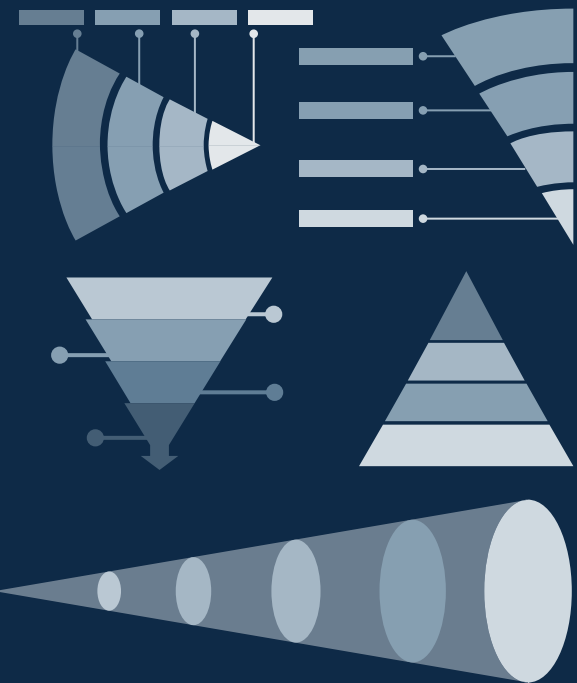
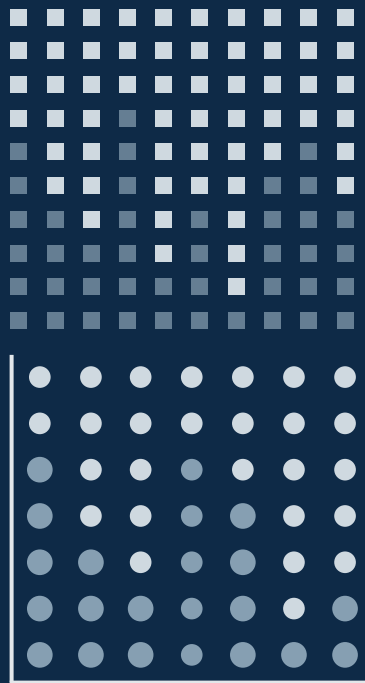












## ...and our sets of editable icons

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



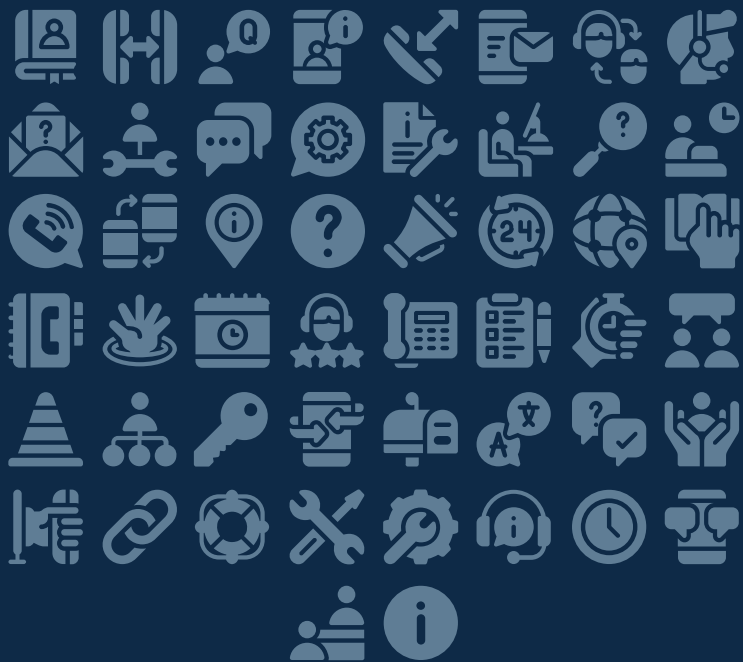
## Business Icons



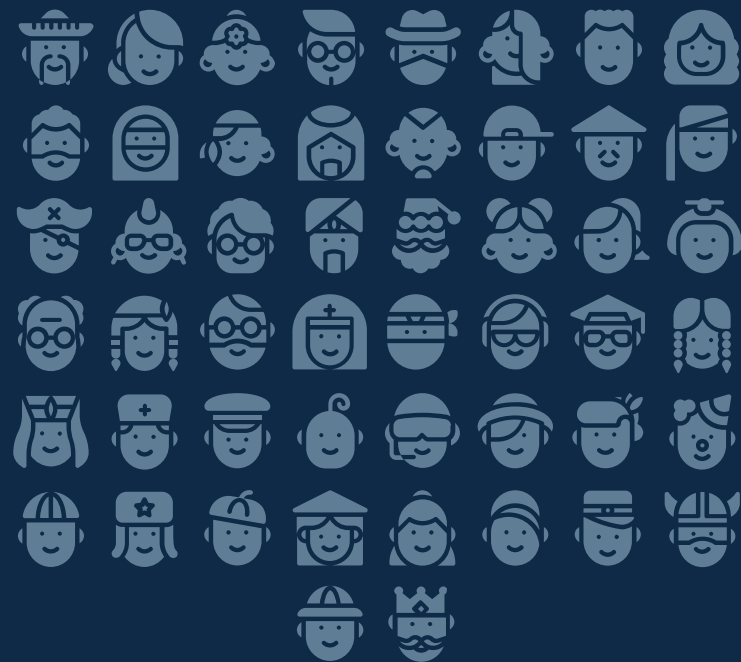
## Teamwork Icons



# Help & Support Icons



# Avatar Icons





## A large grid of 60 white icons representing various digital marketing and technology concepts, arranged in 6 rows and 10 columns. The icons include symbols for analytics, design, communication, social media, and general business operations.

[illegible]

# Nature Icons



# SEO & Marketing Icons



