THE ULTIMATE MARKET PREDICTOR

FEL1-Team 9: Audric Yap, Joshua Chin, Gabriel Lim

TABLE OF CONTENTS

I

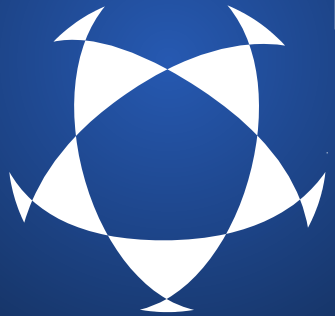
EDA

Exploration of raw data

II

DATA PROCESSING

Manipulation and cleaning of data



III

MACHINE LEARNING

ML algorithms used and analysing their results

IV

INSIGHTS

What we gathered at the end of the project

PROBLEM STATEMENT

Predict the current market value of football players to better understand what drives the value of players, using available personal and game statistics

I

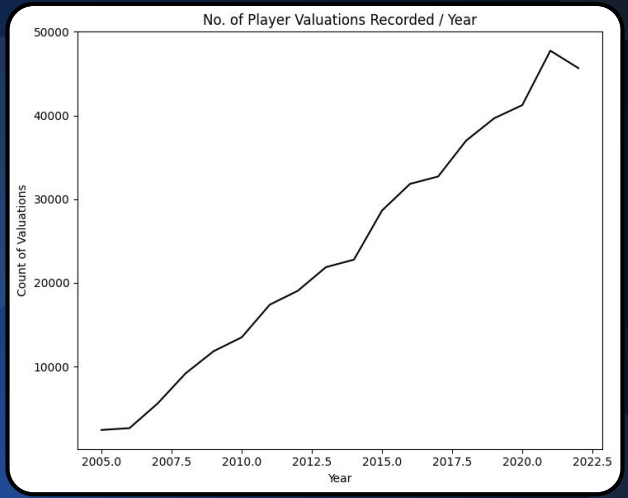
EDA

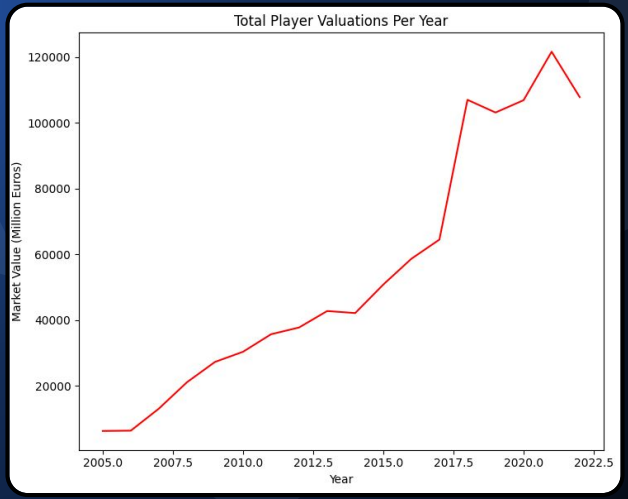
THE DATASET

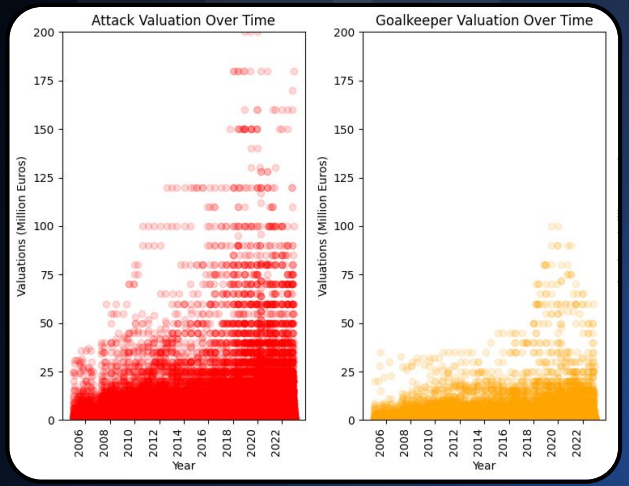
- Dataset obtained from **Kaggle**
- **Scraped from the TransferMarkt** website for its reliability and consistency
- Contains **detailed information** on player and game statistics, valuations and more

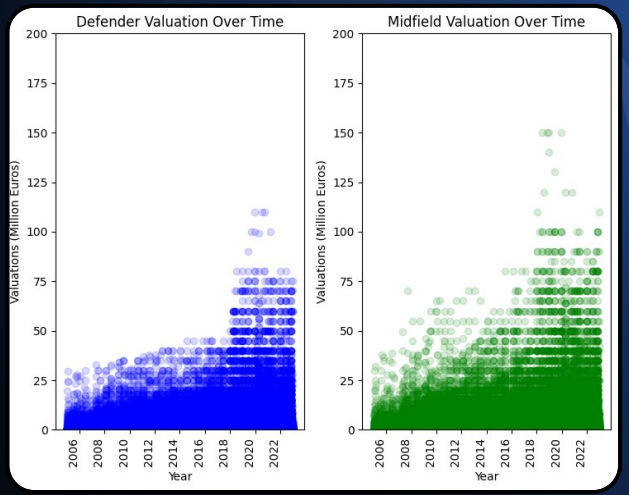
NO. OF MARKET VALUATIONS

- **Number of player valuations increases** consistently over time
- **Big spike** in the sum of player valuations past 2017, followed by an **inconsistent rise** till current day
- This could be attributed to a multitude of factors such as **sudden rising stars and the COVID-19 pandemic,** as well as **inflation**





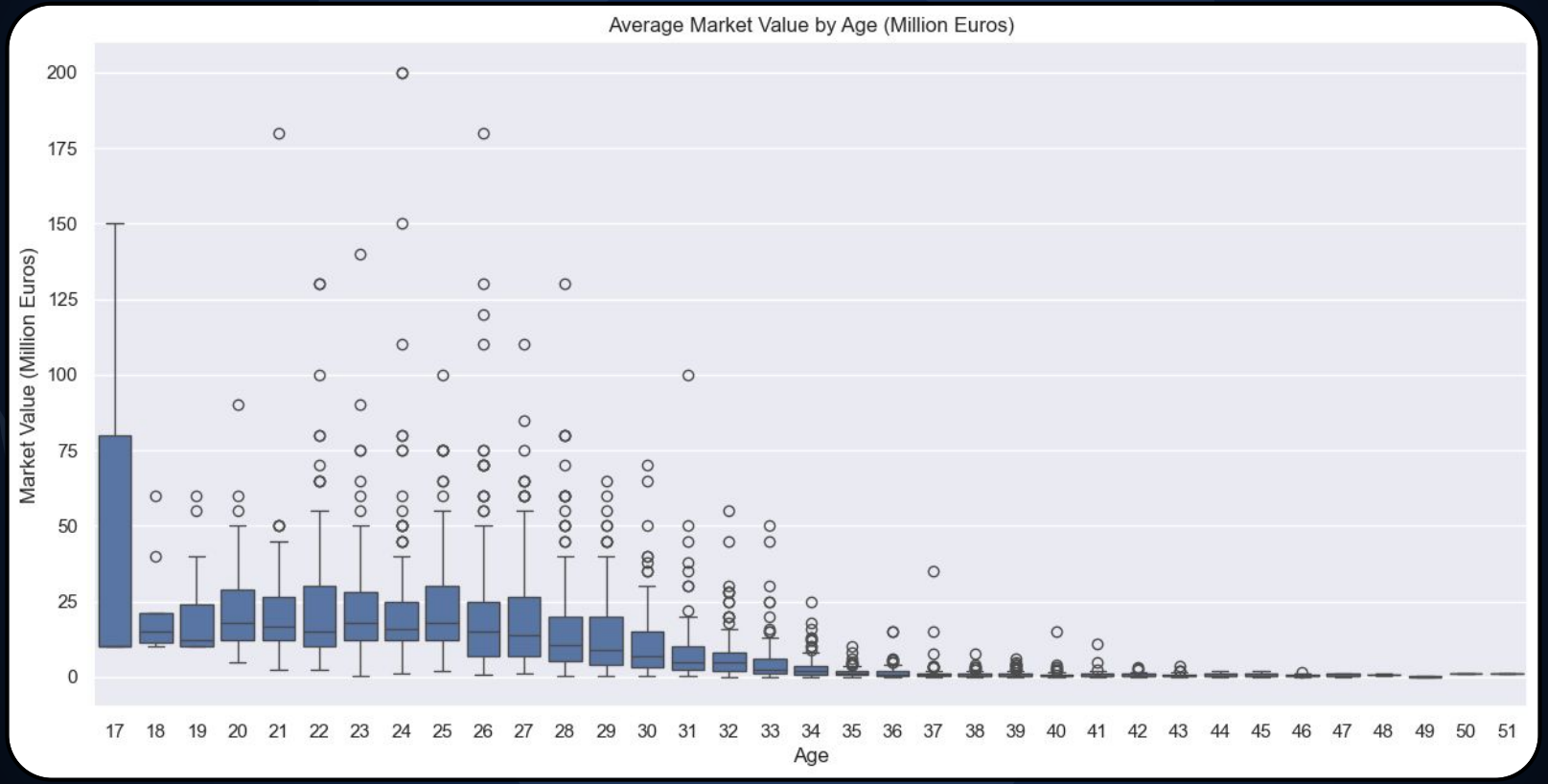




MARKET VALUATIONS BASED ON POSITION

- Generally as time progresses, the players' valuations in **all positions increases**, particularly during **2018 onwards**
- **Attackers seem to be valued more**, followed by Midfielders, Defenders and lastly Goalkeepers
- This is **reflective of real-world scenarios:**
  - Vinicius Jr., a world-class Attacker, is worth **€200 Million**
  - William Saliba, a world-class Defender, is contrastingly worth only **€80 Million**

MARKET VALUATIONS BASED ON AGE



II

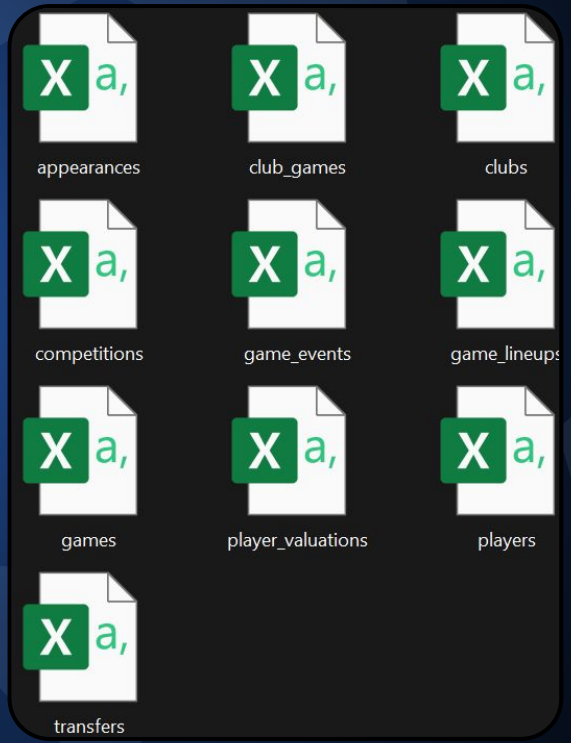
DATA PROCESSING

MERGING THE DATA

Data is split between multiple .csv files. We would need to **merge them together** to one data frame for **easier training**

After **cutting out unimportant data**, we decided to merge the following .csv files:

- players.csv
- appearances.csv
- games.csv
- competitions.csv



FEATURE ENGINEERING

- Mapped each player's league competition to a **ranking based on UEFA coefficients**
- **Compiled game statistics** for all players from 2020 to 2023:
  - Games Played
  - Minutes Played
  - Goals and Assists (Individual and Team)
  - Yellow and Red Cards
- Obtained the **current age** of players from based on current day
- **OneHotEncoded player positions** for more meaningful analysis

FINAL DATA FOR TRAINING

PERSONAL STATS

The current age and height of players

POSITION

The players' preferred playing positions

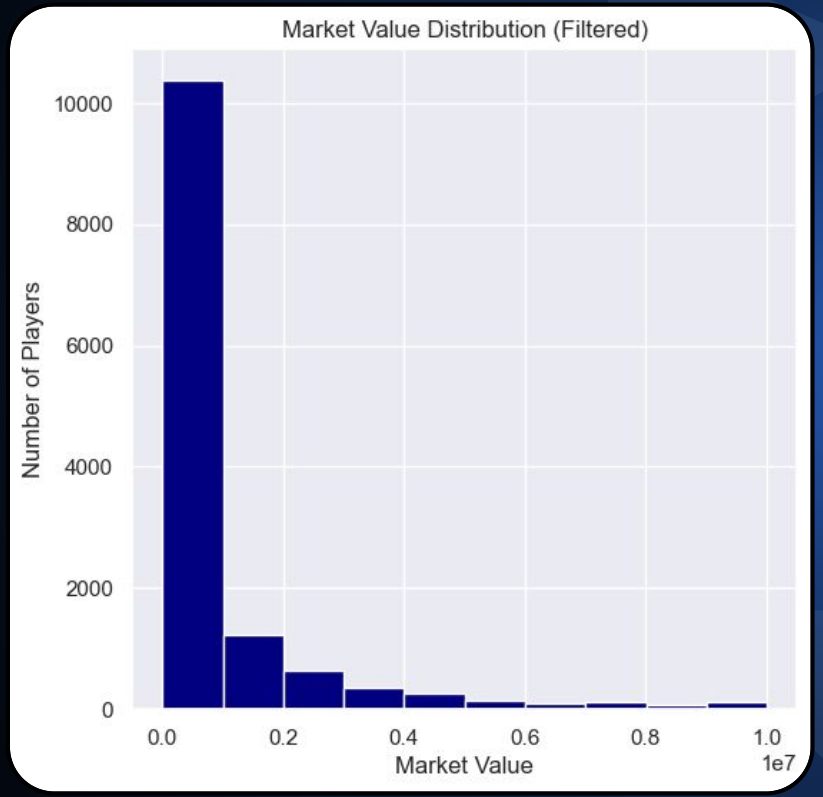
GAME STATS

Game statistics from 2020 season till 2023 season, as well as competition ranks

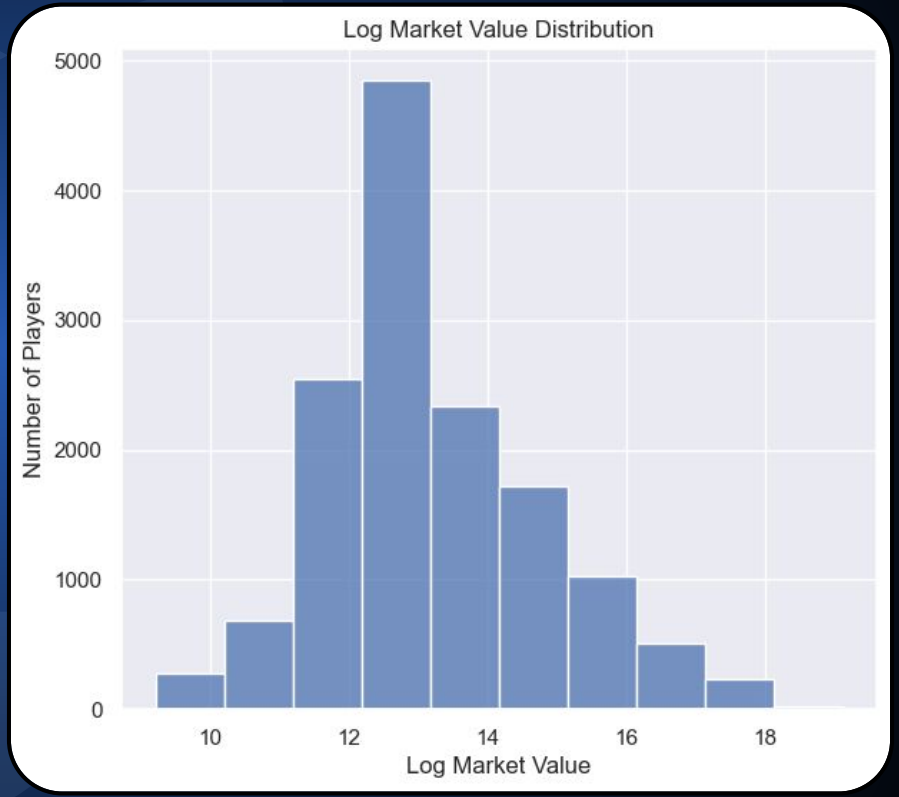
REGION

The regions in which players are born in

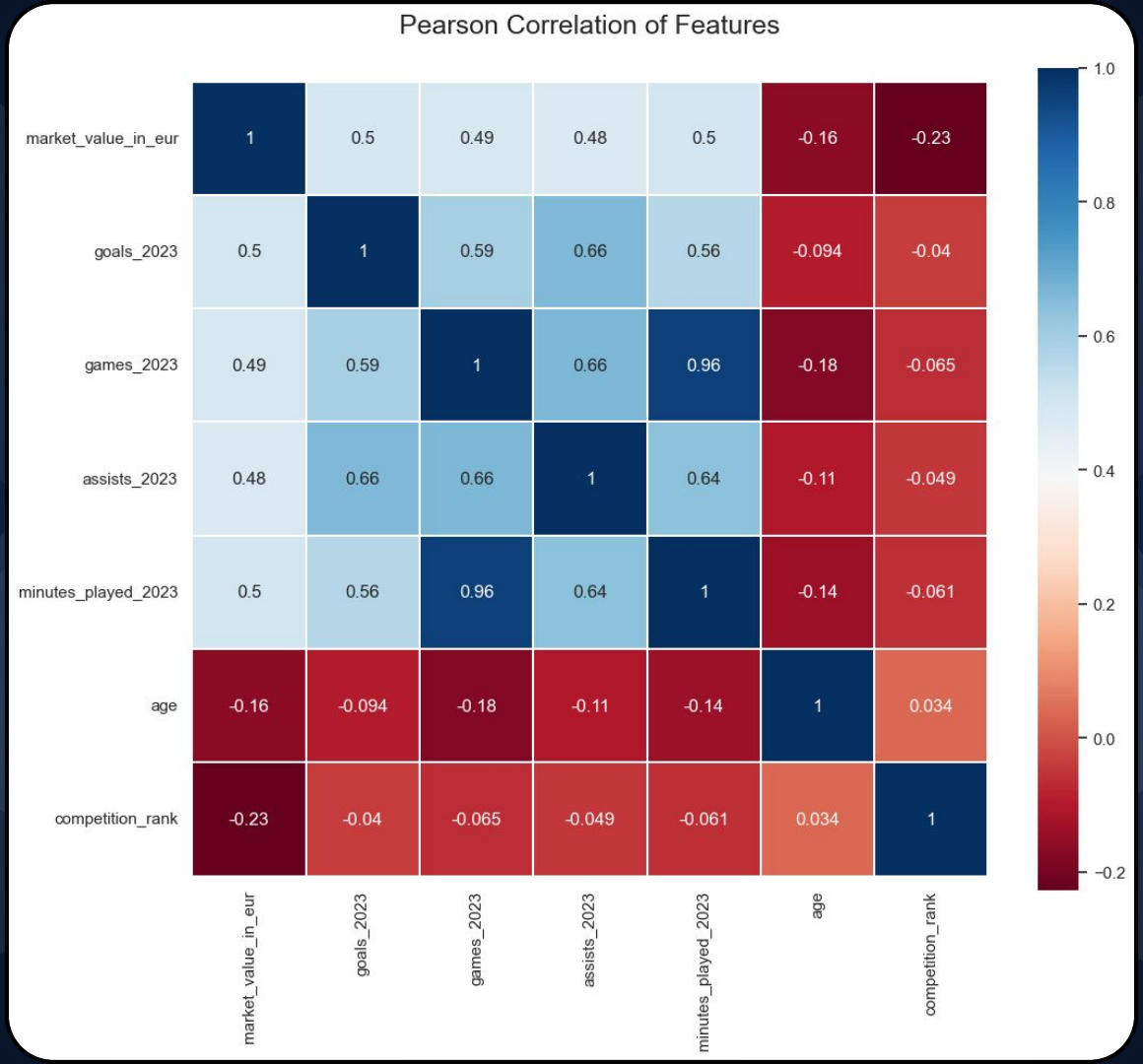
TRANSFORMING MARKET VALUE “y”



Before Log



After Log


Pearson Correlation of Features
market_value_in_eur
goals_2023
games_2023
assists_2023
minutes_played_2023
age
competition_rank
1 0.5 0.49 0.48 0.5 -0.16 -0.23
0.5 1 0.59 0.66 0.56 -0.094 -0.04
0.49 0.59 1 0.66 0.96 -0.18 -0.065
0.48 0.66 0.66 1 0.64 -0.11 -0.049
0.5 0.56 0.96 0.64 1 -0.14 -0.061
-0.16 -0.094 -0.18 -0.11 -0.14 1 0.034
-0.23 -0.04 -0.065 -0.049 -0.061 0.034 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

III

MACHINE LEARNING

METRIC IMPORTANCE

In the football transfer market, the focus is primarily on **interpretability as well as reliability**. As such, our metric focuses will be as such:

- **Primary Metric: Mean Absolute Error (MAE)**
  - Chosen as football market values are expressed in **real-world currency**, in this case Euros, so decision makers like club analysts or agents are **easily able to understand** the average deviation between predicted and actual market values

- **Secondary Metric: R^2**
  - Chosen so that the model is able to **explain variability in market values**, thus increasing its **reliability** in capturing market trends

TESTED MODELS

LINEAR REGRESSION

Fits a **linear model** with coefficients to **minimize the residual sum of squares** between the observed targets in the dataset

RANDOM FOREST REGRESSOR

Fits decision tree regressors on various sub-samples and uses averaging to **improve the predictive accuracy** and **control over-fitting**

ELASTIC NET

Uses the **penalties** from both the **lasso and ridge** techniques to **regularize** regression models.

XGBOOST REGRESSOR

Builds an **ensemble of decision trees**, where **each tree is trained** to make predictions based on a subset of the available data

BASELINE MODEL RESULTS

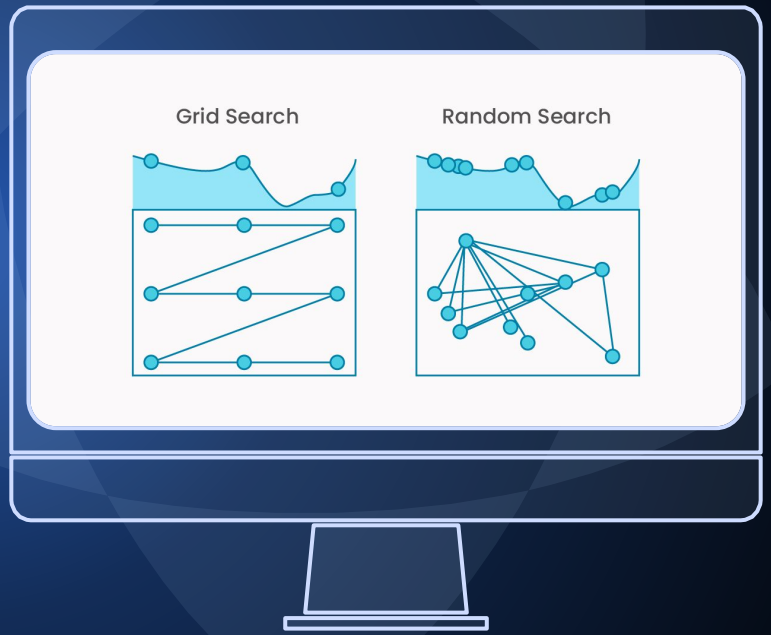
Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	R^2 Score
Linear Regression	0.7873	1.0353	0.5948
Elastic Net	0.8272	1.1457	0.5516
Random Forest	0.6981	0.8513	0.6668
XGBoost	0.6548	0.7546	0.7047

HYPER-PARAMETER TUNING

Given that **XGBoost** is currently the **best model**, having MAE and MSE closest to 0 and R^2 closest to 1, we want to tune it using **GridSearchCV**:

Here are our chosen **optimal hyper-parameters**:

- colsample_bytree: 0.6
- learning_rate: 0.03
- max_depth: 6,
- n_estimators: 500
- subsample: 0.9



XGBOOST POST-TUNING

MAE

-0.013

MSE

-0.024

R^2

+0.010

IV

INSIGHTS

INTERPRETING THE RESULTS

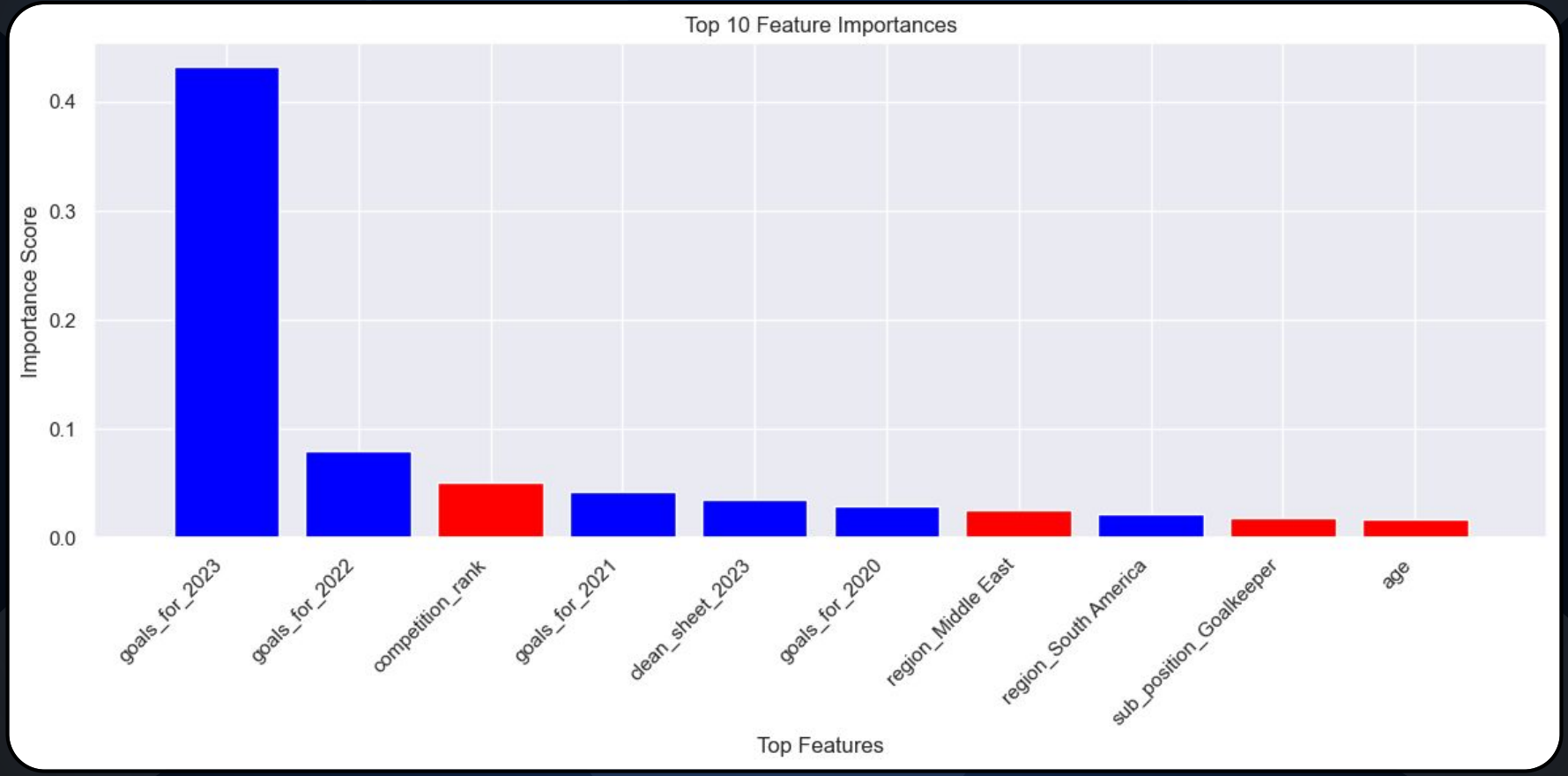
From our XGBoost Model, we obtained a **MAE of 0.643**

- Recomputing MAE using the original scale, we found that the MAE is around **€1.14 Million**
- On average, the model predicts market values with an **absolute error of €1.14 Million**, so it is **relatively accurate** for predicting market values in a domain where values can range widely up to tens of millions

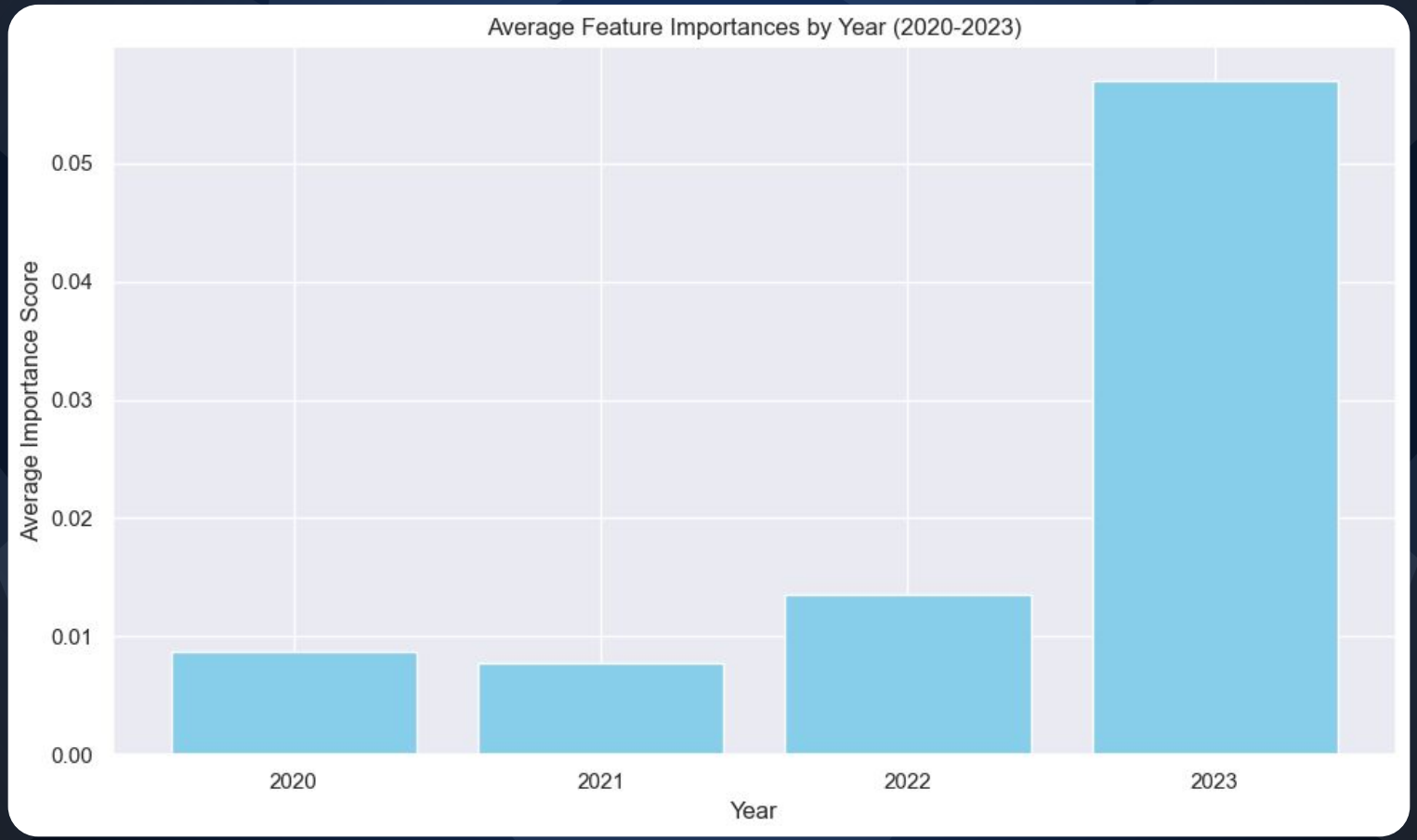
Our model also obtained an **R^2 score of 0.714**

- Explains **71.4% of variability** in football players' market values
- This is a **good result** as other **external factors like club/player sentiments that is not captured** in the model can be attributed to this result

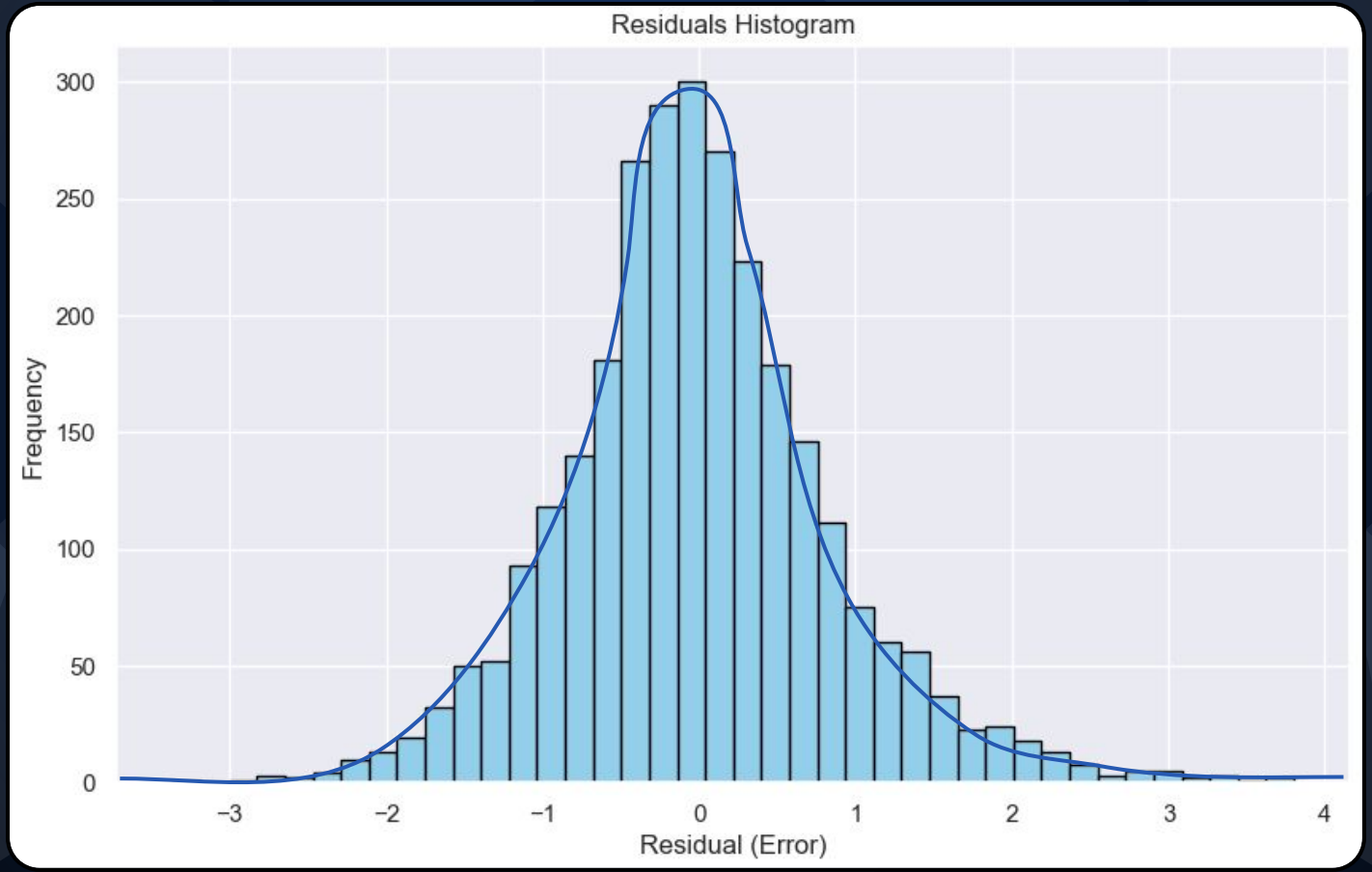
XGBOOST TOP FEATURE IMPORTANCE



XGBOOST FEATURE IMPORTANCE BY YEAR



XGBOOST ERROR RESIDUALS



CONCLUSION

- Our model is able to predict, with a **low error margin**, the current market prices of football players based on their **past and current game statistics and personal traits**

- The model also has a **variability of 71.4%** of the market captured, allowing it to **pick up on market trends** reliably and easily

- This displays the **robustness** of our model in predicting football market values, which is a **useful tool for any football club** looking to make a player investment

THANK YOU

SLIDES AFTER THIS ARE TEMPLATES

MAYBE YOU NEED TO DIVIDE THE CONTENT

MERCURY

Mercury is the closest planet to the Sun and the smallest one

VENUS

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

…OR MAYBE YOU COULD USE THREE COLUMNS

MARS

Despite being red, Mars is a very cold place full of iron oxide dust

VENUS

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

MERCURY

Mercury is the closest planet to the Sun and the smallest one

YOU COULD USE FOUR COLUMNS, WHY NOT?

SATURN

Saturn is composed mostly of hydrogen and also of helium

VENUS

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

MERCURY

Mercury is the closest planet to the Sun and the smallest one

MARS

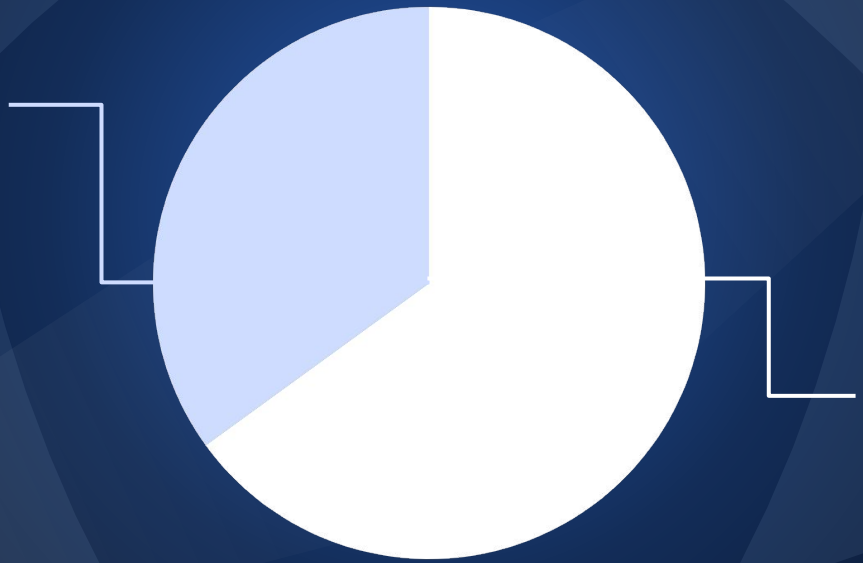
Despite being red, Mars is a very cold place full of iron oxide dust

AWESOME WORDS

THIS IS A GRAPH

MARS

Despite being red, Mars is a very cold place, not hot



VENUS

Venus has a very beautiful name, but it's terribly hot

To modify this graph, follow the link, change the data and replace it

6,777,000

Big numbers catch your audience's attention

SOMETIMES, REVIEWING CONCEPTS IS A GOOD IDEA

EARTH

Earth is the third planet from the Sun

VENUS

Venus is the second planet from the Sun

JUPITER

Jupiter is the fifth planet from the Sun

PLUTO

Pluto is now considered a dwarf planet

MARS

Mars is the fourth planet from the Sun

URANUS

Uranus is the seventh planet from the sun

2

PRESENTATION

You can enter a subtitle here
if you need it

INFOGRAPHICS MAKE YOUR IDEA UNDERSTANDABLE...

EARTH

Earth is the third planet from the Sun

MARS

Despite being red, Mars is a very cold place

SATURN

Saturn is composed mostly of hydrogen

MERCURY

Mercury is the closest planet to the Sun

HOW ABOUT PERCENTAGES?

VENUS

Venus has a beautiful name, but it's hot

MARS

Despite being red, Mars is a very cold place

MERCURY

Mercury is the closest planet to the Sun

THIS IS A TABLE

	TEAM 1	TEAM 2	TEAM 3	TEAM 4
GROUPS	20 times	45 times	33 times	12 times
QUARTERFINAL	8 times	20 times	4 times	4 times
SEMIFINAL	2 times	20 times	20 times	5 times
FINAL	5 times	2 times	7 times	3 times

FOUR COLUMNS

SATURN

Saturn is composed mostly of hydrogen

MERCURY

Mercury is the closest planet to the Sun

VENUS

Venus has a beautiful name, but it's hot

EARTH

Earth is the third planet from the Sun

MAYBE YOU COULD USE THESE FIVE COLUMNS

SATURN

Saturn is composed mostly of hydrogen and helium

VENUS

Venus has a beautiful name, but it's hot

MERCURY

Mercury is the closest planet to the Sun

MARS

Despite being red, Mars is a very cold place, not hot

EARTH

Earth is the third planet from the Sun

THIS IS A TIMELINE

GROUPS

Despite being red, Mars is a very cold place

QUARTERFINAL

Saturn is composed mostly of hydrogen

SEMIFINAL

Earth is the third planet from the Sun

FINAL

Mercury is the closest planet to the Sun

THIS IS A MAP

VENUS

Venus has a beautiful name

MARS

Despite being red, Mars is a cold place

SATURN

It is composed mostly of hydrogen

MERCURY

Mercury is the smallest planet

THIS TEXT SEEMS TO BE IMPORTANT

Venus has a beautiful name and is the second planet from the Sun. It's terribly hot—even hotter than Mercury

3,957,493

Earths is the Sun's mass

5,666,046

It's Jupiter's rotation period

4,555,375

It's the distance between planets

OUR TEAM

HELEN MARK

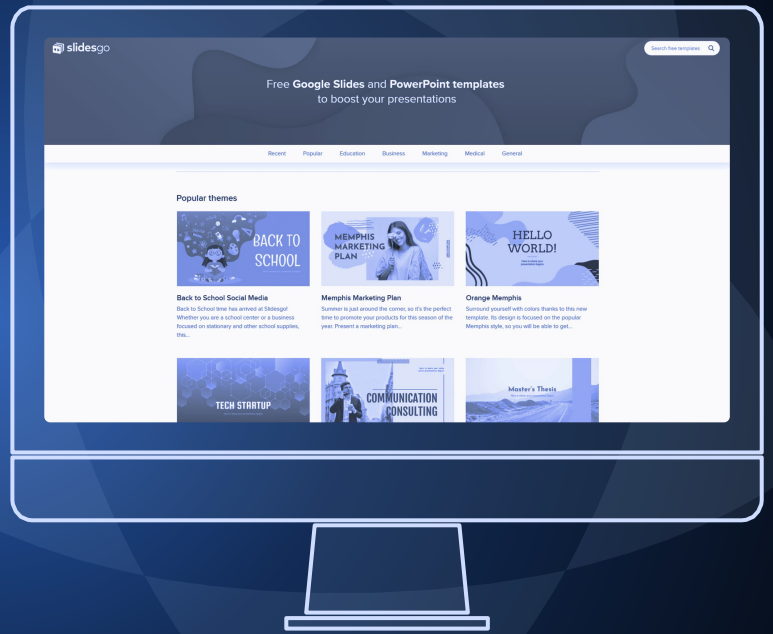
Here you could talk a bit about this person

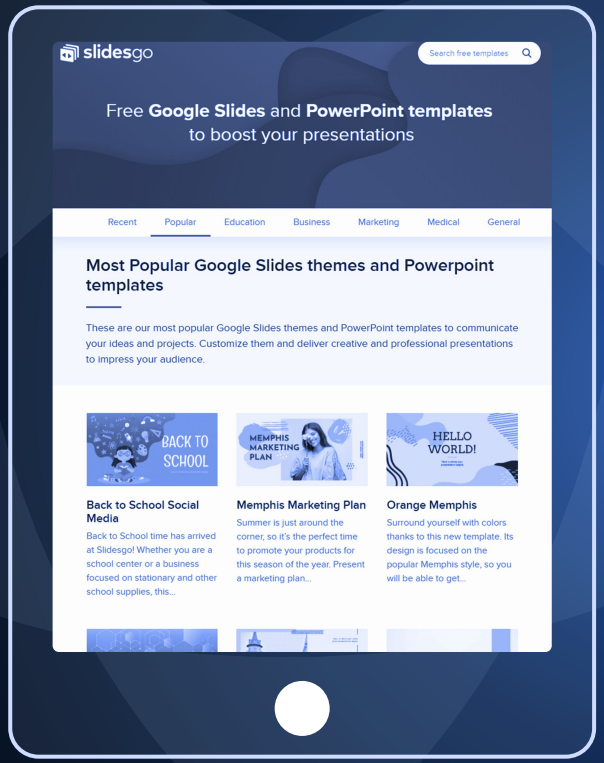
JOHN DOE

Here you could talk a bit about this person

DESKTOP SOFTWARE

You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly





TABLET APP

You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly

MOBILE WEB

You can replace the image on the screen with your own work. Just delete this one, add yours and center it properly

THANKS!

Do you have any questions?
addyouremail@freepik.com
+91 620 421 838
yourcompany.com

ALTERNATIVE RESOURCES

VECTOR

- European soccer championship background with waves
- Football poster template or flyer design
- Flat sport landing page template
- Soccer player background in abstract style
- Soccer balls with green brush strokes
- Bright european soccer championship with a ball and winners

PHOTO

- Young woman in sportswear playing soccer
- Soccer player celebrating victory
- Sportsman celebrating victory and screaming
- Close-up leg kicking ball
- Crop hand with soccer ball

RESOURCES

VECTOR

- Open air concert illustration
- Set of soccer balls in flat style
- Bright sport silhouettes background
- Soccer league flyer with player silhouette
- Football background with silhouette
- Soccer league flyer with player silhouette
- Runner infographic
- Sports runner background
- Flat european map with blue background
- Technology elements background in flat style

PHOTO

- Crop legs shooting ball
- Fit woman playing with soccer ball
- Front view woman holding soccer ball
- Football player with ball under arm near wall

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:

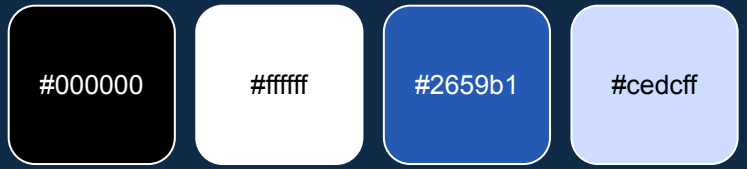
https://slidesgo.com/faqs and https://slidesgo.com/slidesgo-school

Fonts & colors used

This presentation has been made using the following fonts:

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

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



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

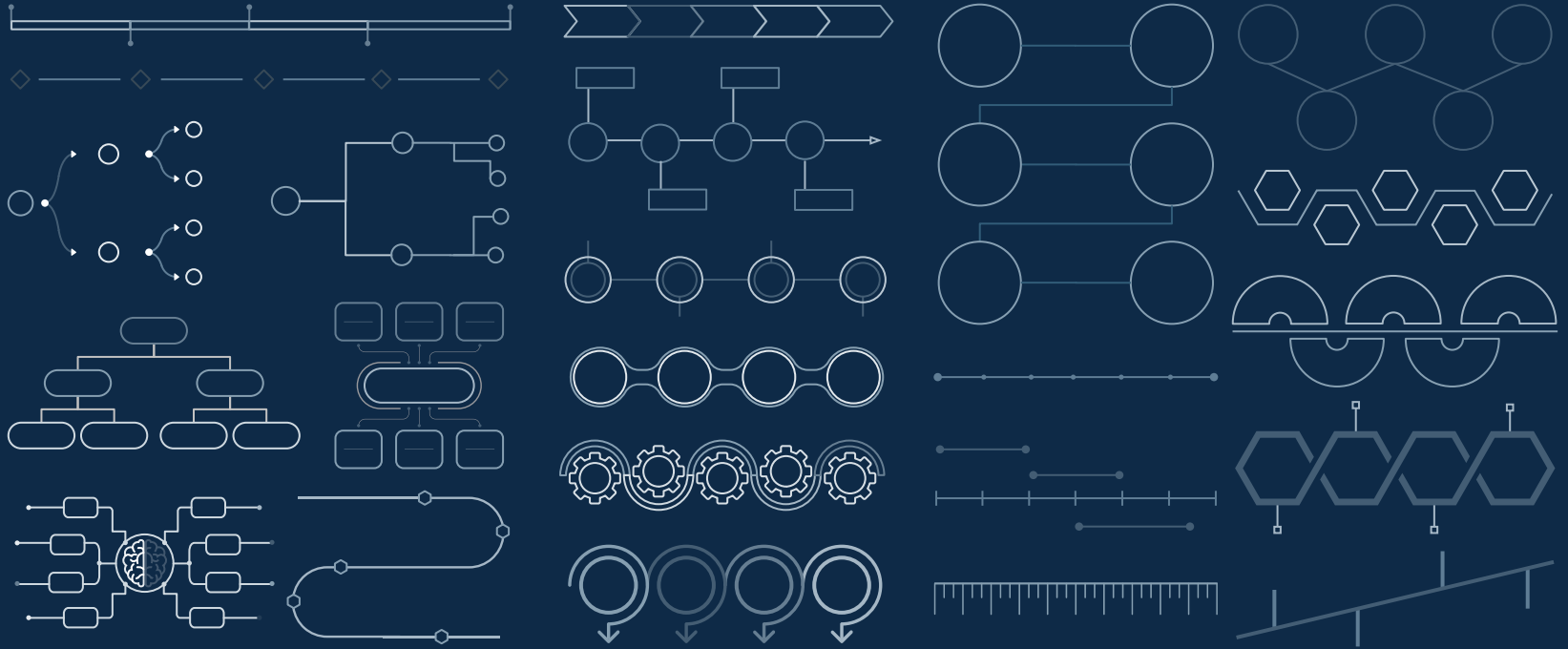
Bro

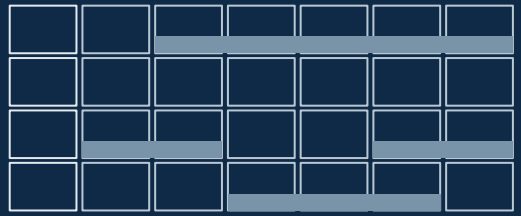
Rafiki

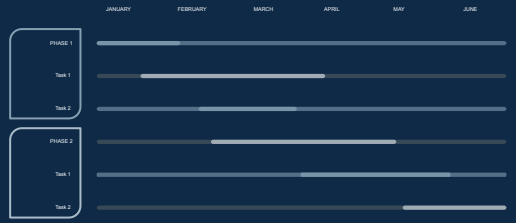
Cuate

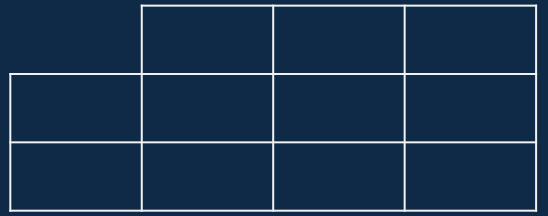
Use our editable graphic resources...

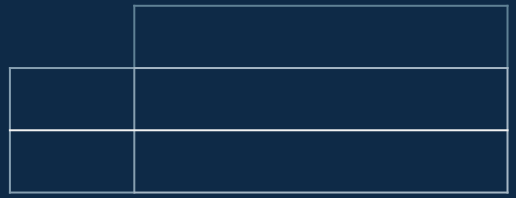
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more infographics on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

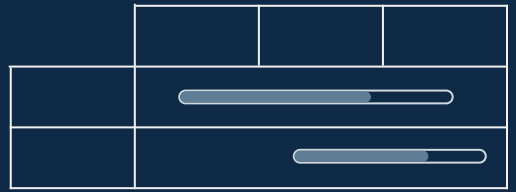


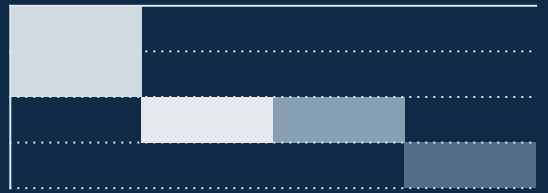


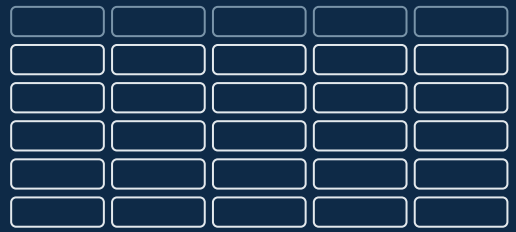

JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2

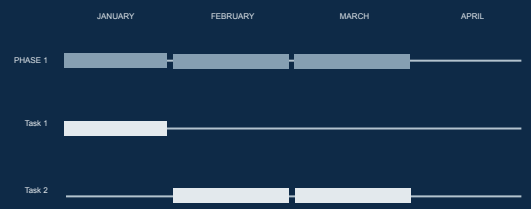


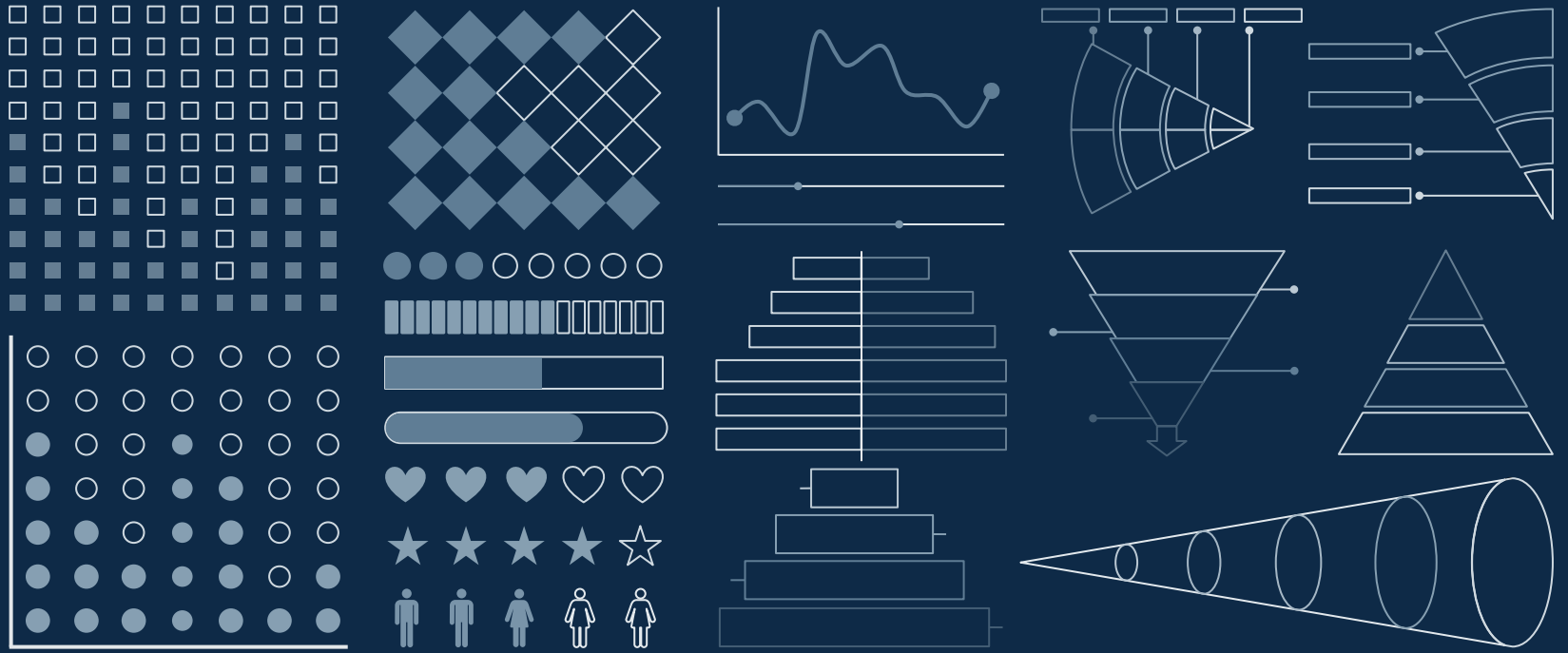









JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2



...and our sets of editable icons

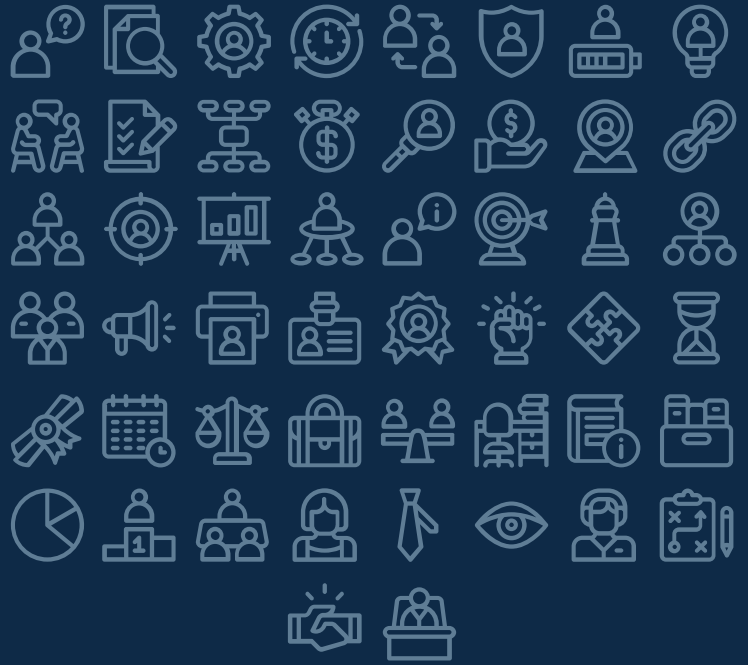
You can resize these icons without losing quality.
You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.
In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

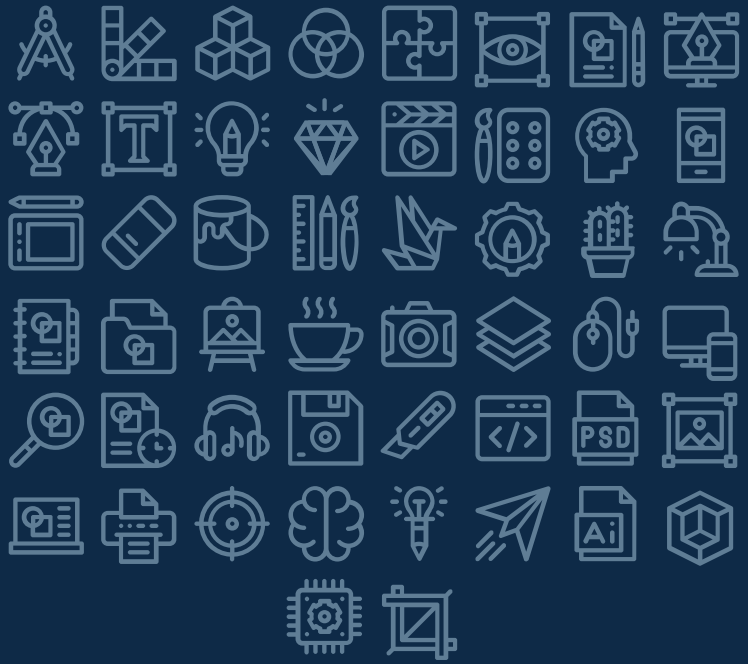
Teamwork Icons



Help & Support Icons

Avatar Icons

Creative Process Icons



Performing Arts Icons

Nature Icons

SEO & Marketing Icons


slidesgo