FIFA World cup Data Analysis and Visualization

ECE 143 Group 1 Project
Winter 2023
Group members:
    Yiting Shi, Jean Gorby Calicdan,
    Zhongkang Fang, Zhen Wang, Puhung Li

Motivation

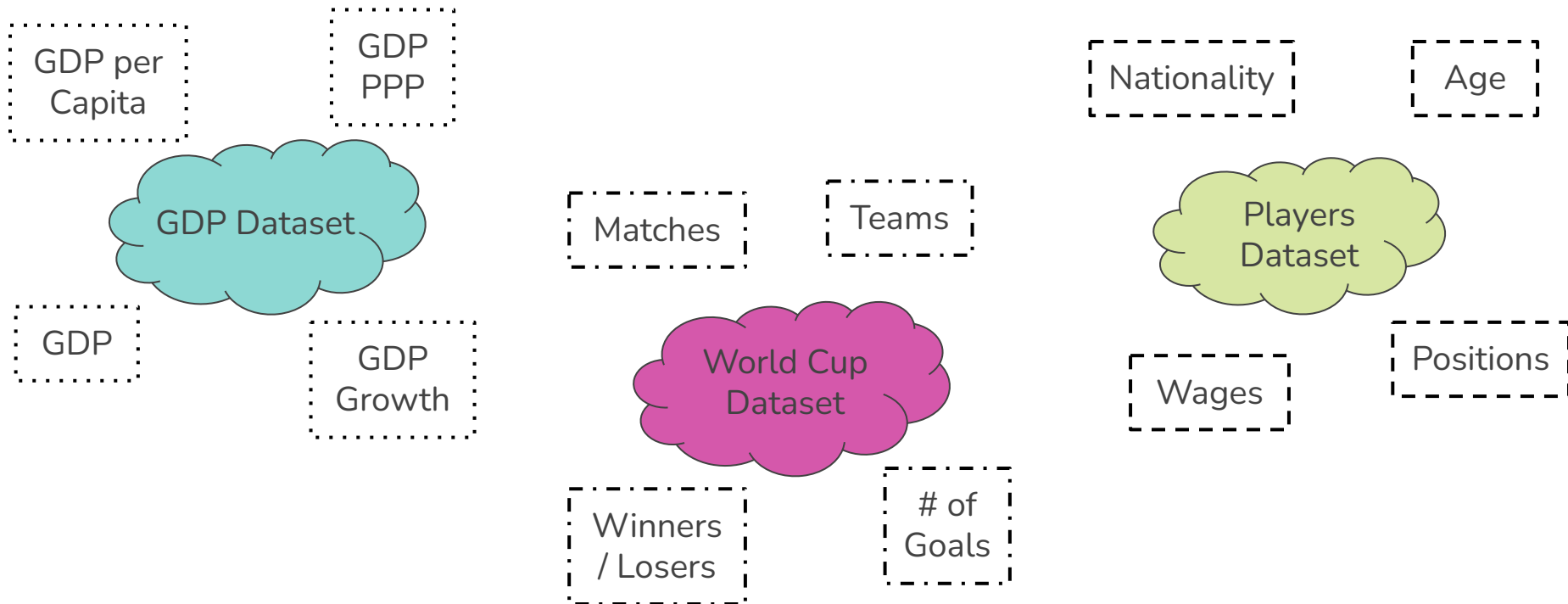Offer a different perspective within a data analysis problem that has been done over and over.

Understand how external factors could influence internal numbers and statistics

Examine keys that show strong offensive ability

Data Overview

GDP per Capita

GDP PPP

GDP Dataset

GDP

GDP Growth

Matches

Teams

World Cup Dataset

Winners / Losers

# of Goals

Nationality

Age

Players Dataset

Wages

Positions

Data Cleaning Principles

- ❏ Ensure country names are the same throughout datasets
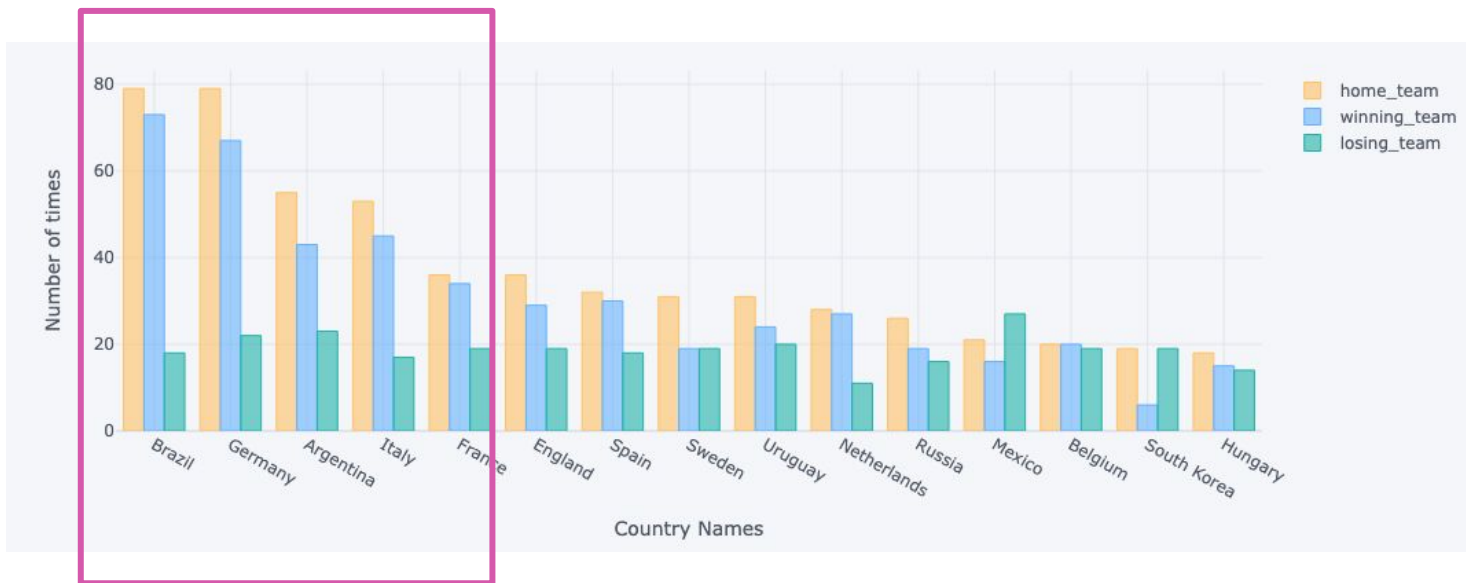- ❏ Reformat attribute names

Efficient Data
- ❏ Drop columns that are not relevant
- ❏ Filter countries and players that don't appear in the world cup
- ❏ Match some data points to other relevant values in other datasets

History Analysis

❏ Top 15 countries home-team*, winning, and losing stats from 1930 - 2018

* note: home team meaning that the team is playing in the usual area that they play in, as opposed to the visitor team
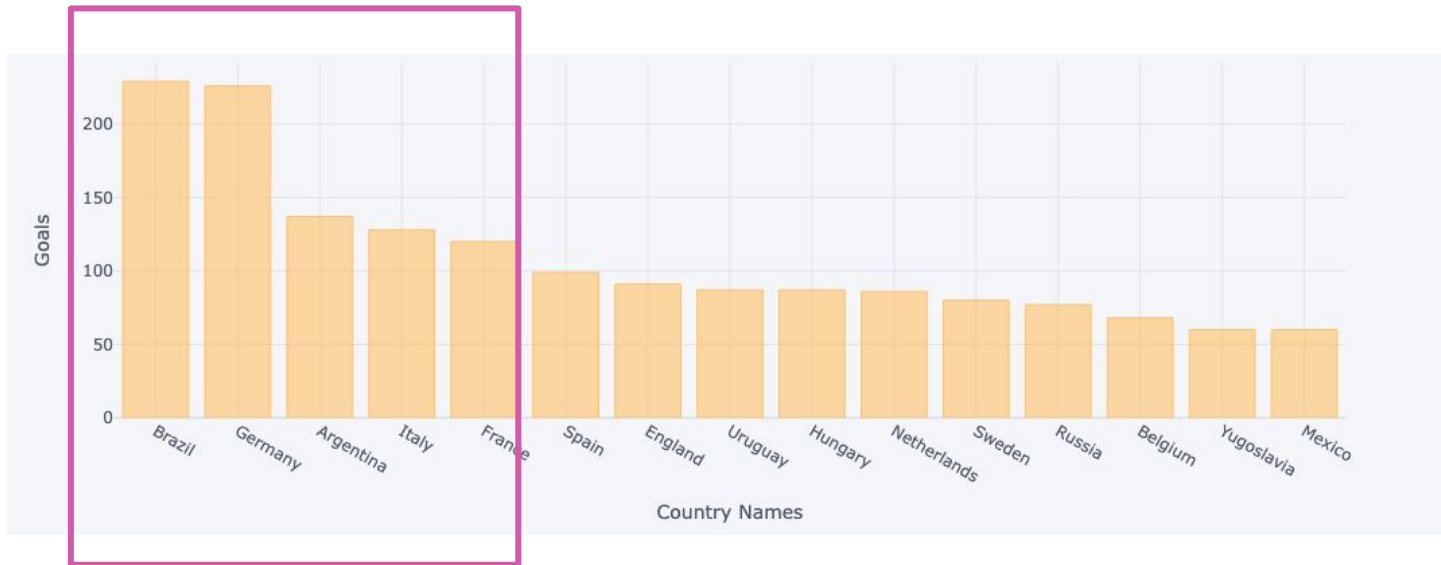
Winning Analysis

❏ Top 15 countries winning count from 1930 - 2018
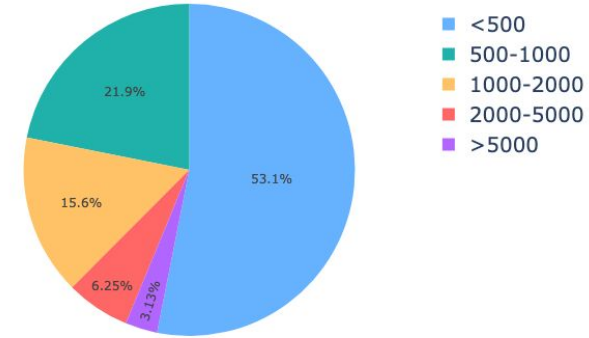
Goal-Scoring Analysis

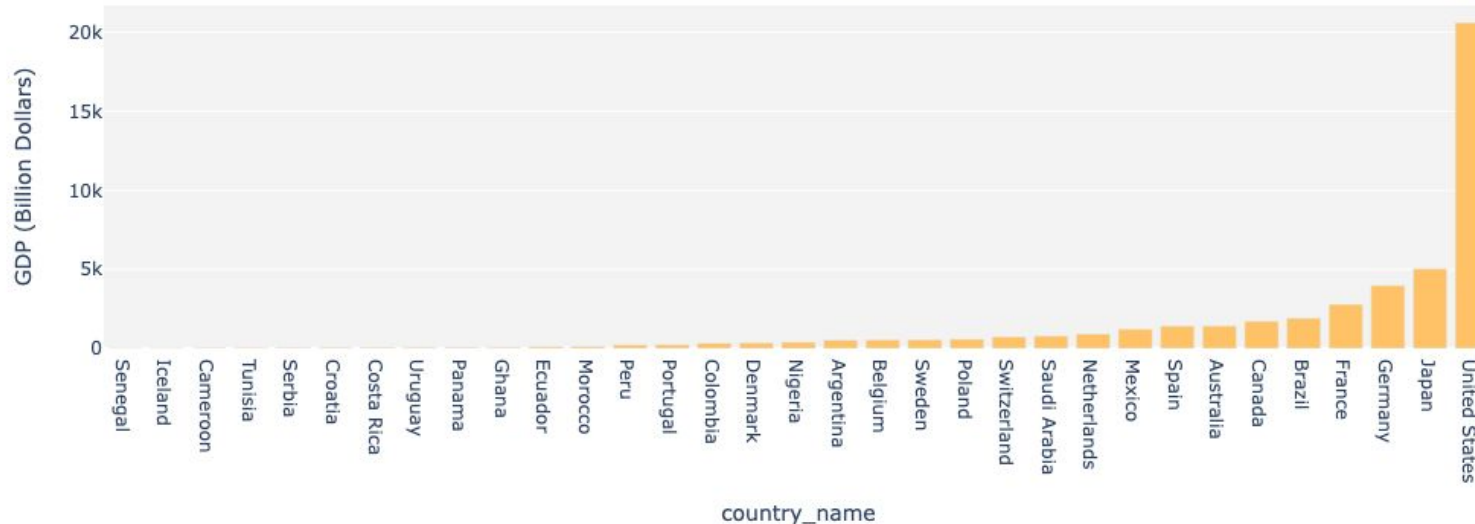❏ Top 15 Countries that hit number of goals 1930 - 2018

GDP Influences #Goals?

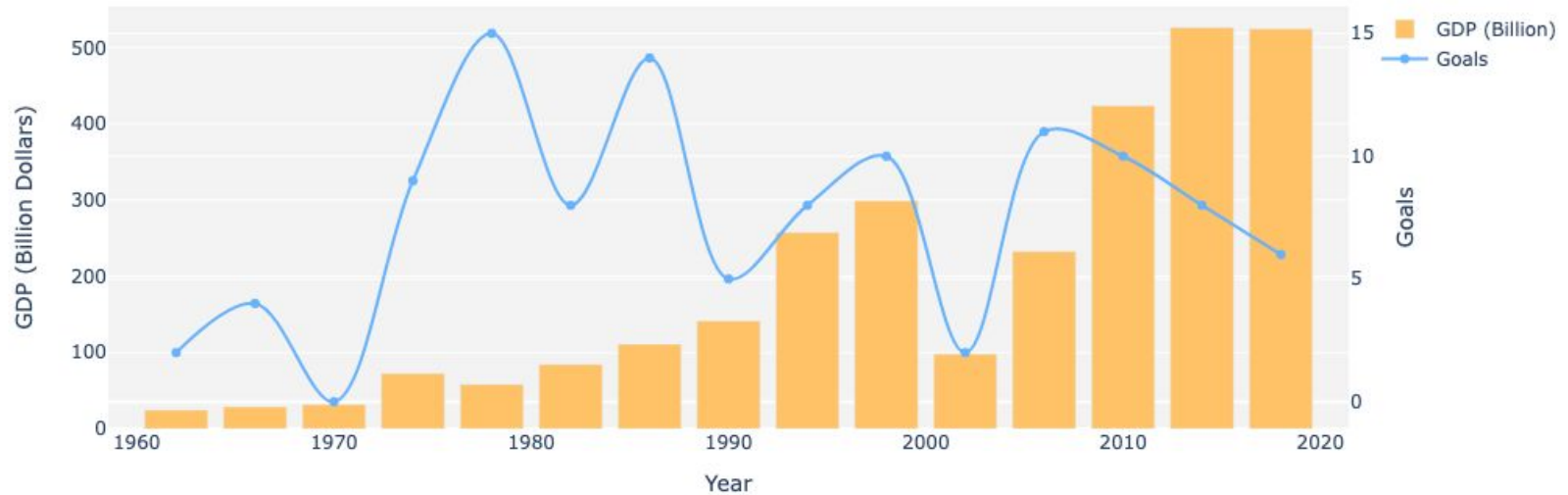❏ GDP overview of the countries participating in the 2018 World Cup
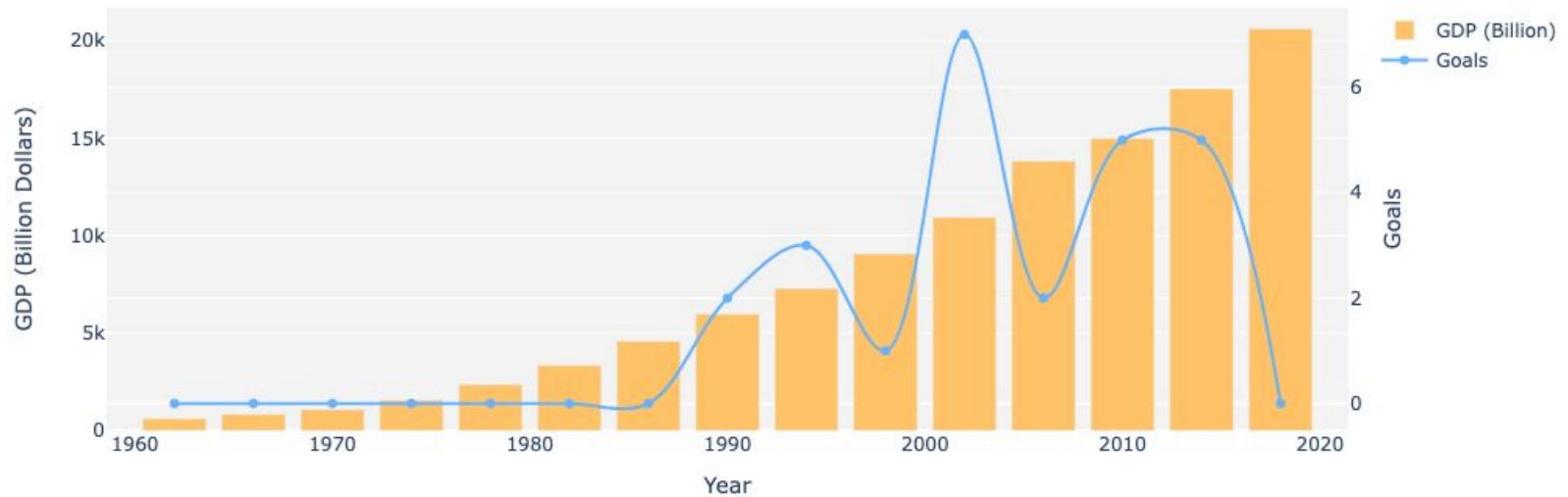

GDP Ranges by Country

GDP by Country

Goal-Scoring and GDP

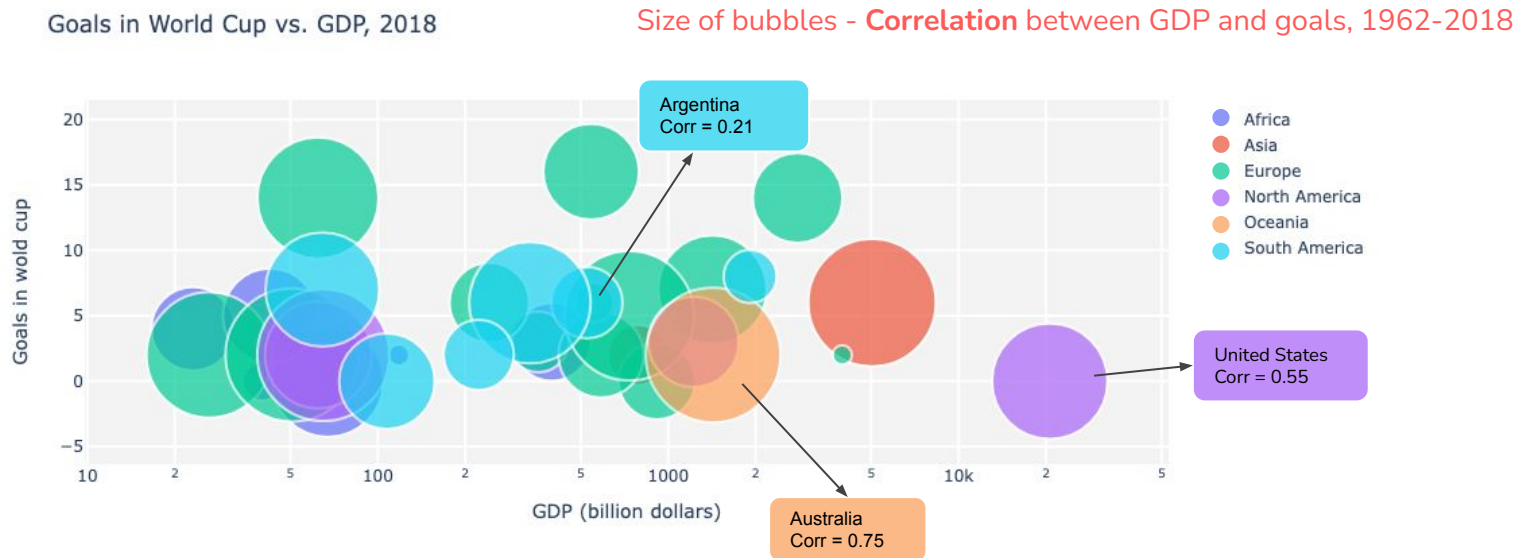❏ Number of goals and GDP from 1962 to 2018 - Argentina

Goal-Scoring and GDP

❏ Number of goals and GDP from 1962 to 2018 - United States

Correlation between #Goals and GDP

- ❏ Correlation between #goals and GDP varies over countries.
- ❏ No strong relationship between #goals and GDP

Goals in World Cup vs. GDP, 2018

Size of bubbles - **Correlation** between GDP and goals, 1962-2018

Argentina
Corr = 0.21

United States
Corr = 0.55

Australia
Corr = 0.75

Africa
Asia
Europe
North America
Oceania
South America

Keys to Show Strong Offensive Ability

Take 2022 Argentina World Cup team as an example
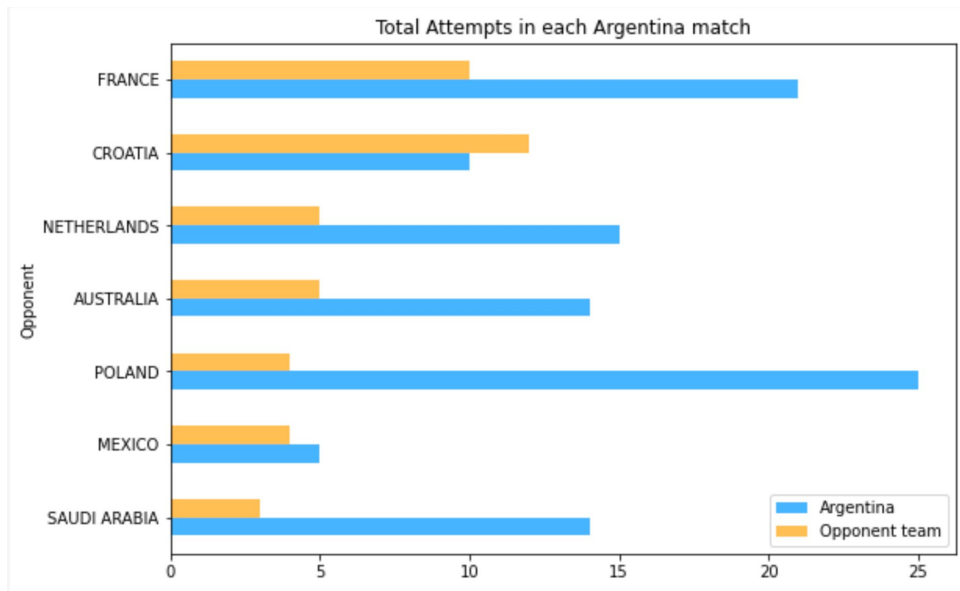
❏   Total Goal Attempts
❏   Goal Attempt Accuracy and Efficiency
❏   Total Offers to Receive

Total Goal Attempts

❏ Total Goal Attempts: A team's attacking ability to break through the opponent defense line
❏ More Total Goal Attempts -> Better attacking ability

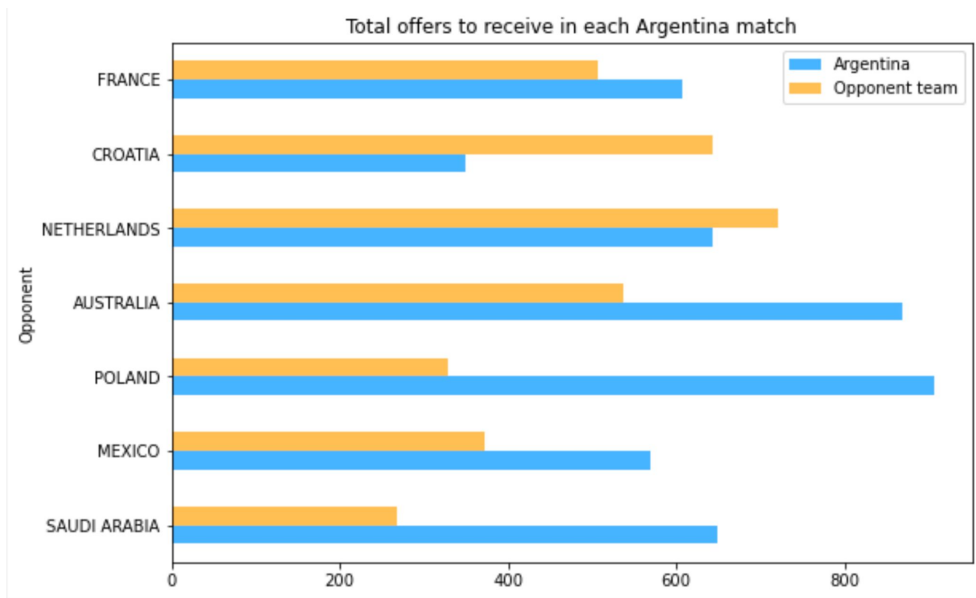Total Attempts in each Argentina match

Goal Attempt Accuracy

❏ Total attempts: The average goal attempts of each country
❏ Accuracy = (On target attempts - off target attempts) / total attempts
❏ Higher goal accuracy -> Greater threat to the opponent

	total attempts	Accuracy
GERMANY	22.333333	-0.134328
FRANCE	14.428571	-0.069307
MEXICO	14.000000	0.000000
ARGENTINA	14.857143	0.057692
BRAZIL	18.000000	0.077778

Total Offers to Receive

❏ Offering to receive: The action performed in an attempt to receive the ball
❏ More offers to receive -> Longer ball possession periods and more scoring opportunities

Total offers to receive in each Argentina match

Conclusion

- ❏ Goal-scoring ability is strongly correlated with winning and **hosting** the game

- ❏ Goal-scoring ability is **not** strongly correlated with GDP of the countries

- ❏ Strong offensive abilities are shown by the following keys: **Total Goal Attempts**, **Goal Attempt Accuracy**, and **Total Offers to Receive**

Thank you!