Project Report

FIFA WORLD CUP

Carlos Soda
Bharat Guturi
Balvinder Rajbans

Sources of Data

The data is sourced from ("Joshua C. Fjelstul, Ph.D."), a notice that the database is copyrighted ("© 2022 Joshua C. Fjelstul, Ph.D."), a link to the CC-BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/legalcode), and a link to this repository (https://www.github.com/jfjelstul/worldcup)

No	Source	Link	File
1	Git Hub	https://github.com/jfjelstul/worldcup/tree/master/data-csv	goals.csv
2	Git Hub	https://github.com/jfjelstul/worldcup/tree/master/data-csv	tournament_standings.csv
3	The Sporting News	https://www.sportingnews.com/uk/soccer/news/fifa-world-cup-which-teams-have-qualified/86nbyru9dkh41ii7s800gjwav	

Modifications

The csv file goals.csv was modified to the requirements of the project and the following columns were dropped: 'match_id','stage_name','group_name','shirt_number','player_team_id','player_team_code','minute_label','minute_stoppage','own_goal','penalty'

The csv file tournament_standings.csv was NOT modified as it met the requirements of the project.

1.1 Database - Introduction

A structured database has been selected, postgresSQL, and pgAdmin 4 as it is a structured database providing the necessary functionality required to create and analyse the data.

1.2 Database Identification

Database Name: project_3

Table Name: goals
Primary Key: key_id

Table Name: tournament_standings
Primary Key: key_id

1.3 PostgresSQL Schema Information

```
CREATE TABLE goals(
  key_id INTEGER,
  goal_id TEXT NOT NULL,
  tournament_id TEXT NOT NULL,
  tournament_name TEXT NOT NULL,
  match_name TEXT NOT NULL,
  match_date TEXT NOT NULL,
  team_id TEXT NOT NULL,
  team_name TEXT NOT NULL,
  team_code TEXT NOT NULL,
  home_team INTEGER,
  away_team INTEGER,
  player_id TEXT NOT NULL,
  family_name TEXT NOT NULL,
  given_name TEXT NOT NULL,
  player_team_name TEXT NOT NULL,
  minute_regulation INTEGER,
  match_period TEXT,
  match_id VARCHAR,
  PRIMARY KEY (key_id)
);
CREATE TABLE tournament_standings(
  key_id INTEGER,
  tournament_id TEXT NOT NULL,
  tournament_name TEXT NOT NULL,
  position INTEGER,
  team_id TEXT NOT NULL,
  team_name TEXT NOT NULL,
  team_code TEXT NOT NULL,
  PRIMARY KEY (key_id) );
```

2.0 Data Extraction, Transformation and Loading

2.1 Database - Creation

Initially start by creating the database inside of pgAdmin.

The name chosen for the database is ‘project_3’.

NOTE: It is vital to take note of the access details such as username and password used to access pgAdmin - this is needed to update the .env file

2.2 Database - Table Creation

Run the commands located inside ‘postgresSQL_project_3_schema.sql’ to create the tables needed. These table creation commands are also located in section *1.3 PostgresSQL Schema Information.*

2.3 Data Extraction

The data used is in csv format.

In order to extract the data from the csv files, Jupyter Notebook is used and the library is Pandas.

Starting off by using the ‘read_csv’ function inside Pandas to read in the csv’s as Dataframes, the relevant data can be extracted specific to our needs. The files goals.csv and tournament_standings.csv have been used.

2.4 Data Transformation

2.4.1 Goals Data

- Narrow Down Dataset

The goals dataframe is created by reading into the goals csv file and the extraction of data is processed and additional columns not required are dropped ('match_id', 'stage_name', 'group_name', 'shirt_number', 'player_team_id', 'player_team_code', 'minute_label', 'minute_stoppage' ,'own_goal', 'penalty').
The extracted data is stored in a new Dataframe and will be used to load into the database.

2.4.2 Tournament Standings Data

This dataset does not require any transformation as it meets the structure and requirements of the project.

2.5 Data Loading

As the Dataframes are now cleaned and transformed ready to be loaded into the postgreSQL database, the library SQLAlchemy is utilised.

- PostgreSQL Database connection via SQLAlchemy

Initially connecting to our postgres database using a connection string composed of our server's Host Name (host), Username (username), Password (password), Port Number (port) and Database name (database_name).

Once a connection to the postgreSQL server is made, inspect the Database and inspect the tables. This step verifies that the tables are visible and that our connection to the database is true.

- Load Dataframes

Finally, load the Dataframes created into their respective tables inside of our postgreSQL database.

- Verifying Tables

Open pgAdmin and run the following commands to verify the data exists in the tables created.

SELECT * FROM goals

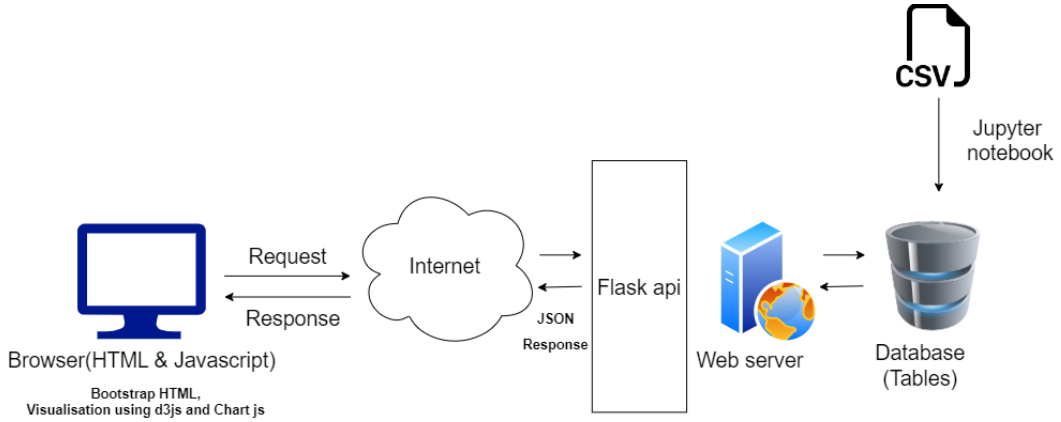
key_id [PK] integer	goal_id text	tournament_id text	tournament_name text	match_name text	match_date text	team_id text	team_name text	team_code text	home_team integer
1	G-0001	WC-1930	1930 FIFA World C...	France v Mex...	1930-07-13	T-28	France	FRA	1
2	G-0002	WC-1930	1930 FIFA World C...	France v Mex...	1930-07-13	T-28	France	FRA	1
3	G-0003	WC-1930	1930 FIFA World C...	France v Mex...	1930-07-13	T-28	France	FRA	1
4	G-0004	WC-1930	1930 FIFA World C...	France v Mex...	1930-07-13	T-28	France	FRA	1
5	G-0005	WC-1930	1930 FIFA World C...	France v Mex...	1930-07-13	T-44	Mexico	MEX	0
6	G-0006	WC-1930	1930 FIFA World C...	United State...	1930-07-13	T-79	United States	USA	1
7	G-0007	WC-1930	1930 FIFA World C...	United State...	1930-07-13	T-79	United States	USA	1
8	G-0008	WC-1930	1930 FIFA World C...	United State...	1930-07-13	T-79	United States	USA	1
9	G-0009	WC-1930	1930 FIFA World C...	Yugoslavia v ...	1930-07-14	T-83	Yugoslavia	YUG	1
10	G-0010	WC-1930	1930 FIFA World C...	Yugoslavia v ...	1930-07-14	T-83	Yugoslavia	YUG	1
11	G-0011	WC-1930	1930 FIFA World C...	Yugoslavia v ...	1930-07-14	T-09	Brazil	BRA	0
12	G-0012	WC-1930	1930 FIFA World C...	Romania v P...	1930-07-14	T-58	Romania	ROU	1

SELECT * FROM tournament_standings

key_id [PK] integer	tournament_id text	tournament_name text	position integer	team_id text	team_name text	team_code text
13	WC-1950	1950 FIFA World C...	1	T-80	Uruguay	URY
14	WC-1950	1950 FIFA World C...	2	T-09	Brazil	BRA
15	WC-1950	1950 FIFA World C...	3	T-71	Sweden	SWE
16	WC-1950	1950 FIFA World C...	4	T-70	Spain	ESP
17	WC-1954	1954 FIFA World C...	1	T-82	West Germany	DEU
18	WC-1954	1954 FIFA World C...	2	T-34	Hungary	HUN
19	WC-1954	1954 FIFA World C...	3	T-05	Austria	AUT
20	WC-1954	1954 FIFA World C...	4	T-80	Uruguay	URY
21	WC-1958	1958 FIFA World C...	1	T-09	Brazil	BRA
22	WC-1958	1958 FIFA World C...	2	T-71	Sweden	SWE
23	WC-1958	1958 FIFA World C...	3	T-28	France	FRA
24	WC-1958	1958 FIFA World C...	4	T-82	West Germany	DEU
25	WC-1962	1962 FIFA World C...	1	T-09	Brazil	BRA

3.0 Visualisations

Architecture Diagram



Run the Flask API

Flask api


```
@app.route("/api/v1.0/goals")
def goals():
    session = Session(engine)
    results = session.query(Goals).all()
    session.close()
    goals = []
    for x in results:
        goals_dict = {}
        goals_dict["key_id"] = x.key_id
        goals_dict["tournament_id"] = x.tournament_id
        goals_dict["match_id"] = x.match_id
        goals_dict["tournament_name"] = x.tournament_name
        goals_dict["match_name"] = x.match_name
        goals_dict["match_date"] = x.match_date
        goals_dict["team_id"] = x.team_id
        goals_dict["team_name"] = x.team_name
        goals_dict["team_code "] = x.team_code
        goals_dict["home_team"] = x.home_team
        goals_dict["away_team"] = x.away_team
        goals_dict["player_id"] = x.player_id
```


```
[
  {
    "away_team": 0,
    "family_name": "Laurent",
    "given_name": "Lucien",
    "home_team": 1,
    "key_id": 1,
    "match_date": "1930-07-13",
    "match_id": "M-1930-01",
    "match_name": "France v Mexico",
    "match_period": "first half",
    "minute_regulation": 19,
    "player_id": "P-09831",
    "player_team_name": "France",
    "team_code ": "FRA",
    "team_id": "T-28",
    "team_name": "France",
    "tournament_id": "WC-1930",
    "tournament_name": "1930 FIFA World Cup"
  },
  {
```


Request the data for goals and tournament_standings

```
@app.route("/api/v1.0/goals")
```

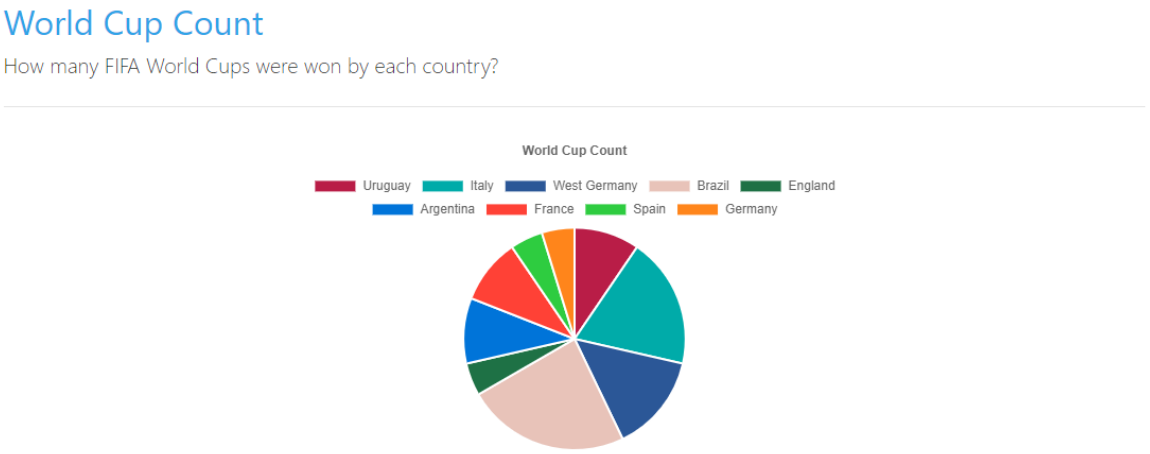
127.0.0.1:5000/api/v1.0/goals


```
{
    "away_team": 0,
    "family_name": "Laurent",
    "given_name": "Lucien",
    "home_team": 1,
    "key_id": 1,
    "match_date": "1930-07-13",
    "match_name": "France v Mexico",
    "match_period": "first half",
    "minute_regulation": 19,
    "player_id": "P-09831",
    "player_team_name": "France",
    "team_code ": "FRA",
    "team_id": "T-28",
    "team_name": "France",
    "tournament_id": "WC-1930",
    "tournament_name": "1930 FIFA World Cup"
  },
  {
    "away_team": 0,
    "family_name": "Langiller",
    "given_name": "Marcel",
    "home_team": 1,
    "key_id": 2,
    "match_date": "1930-07-13",
    "match_name": "France v Mexico",
```

```
@app.route("/api/v1.0/standings")
```

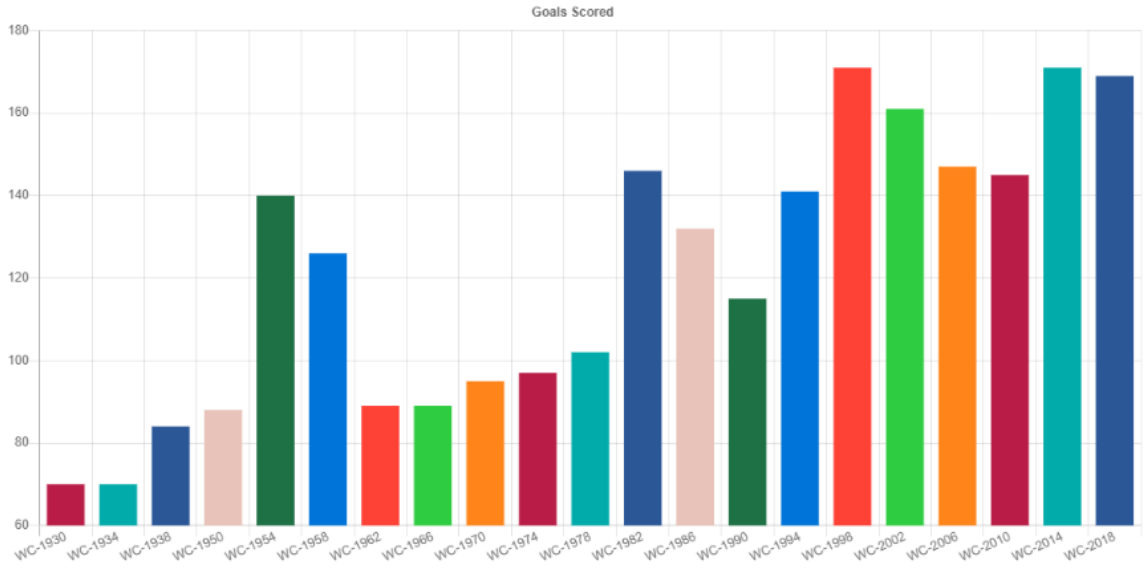

```
[
  {
    "key_id": 1,
    "position": 1,
    "team_code": "URY",
    "team_id": "T-80",
    "team_name": "Uruguay",
    "tournament_id": "WC-1930",
    "tournament_name": "1930 FIFA World Cup"
  },
  {
    "key_id": 2,
    "position": 2,
    "team_code": "ARG",
    "team_id": "T-03",
    "team_name": "Argentina",
    "tournament_id": "WC-1930",
    "tournament_name": "1930 FIFA World Cup"
  }
```


From the templates folder open the index html



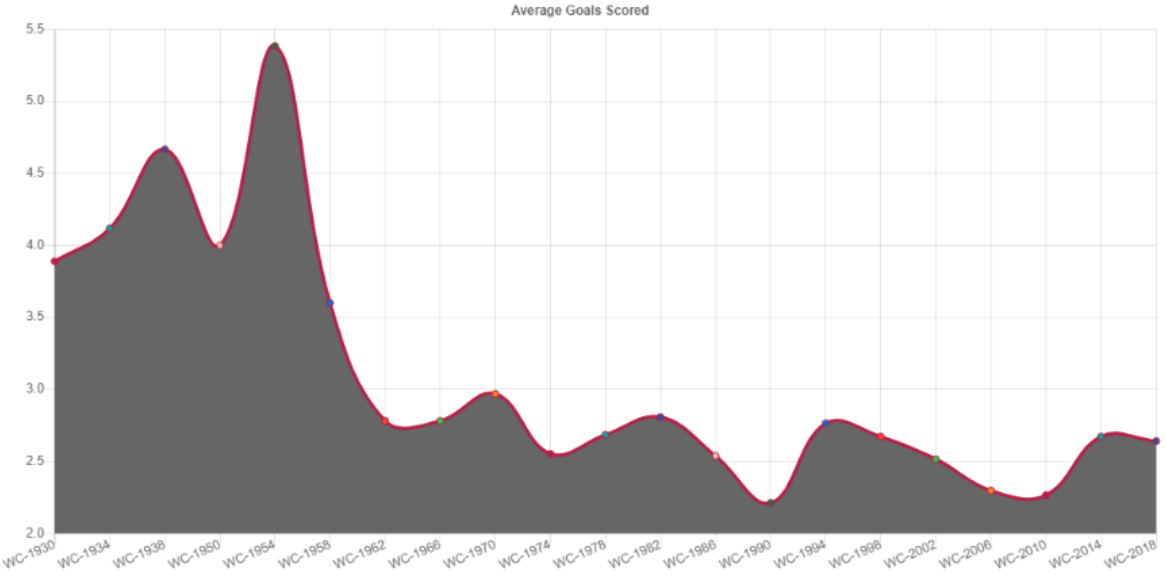
Goals Scored

How many goals are scored in each world cup?

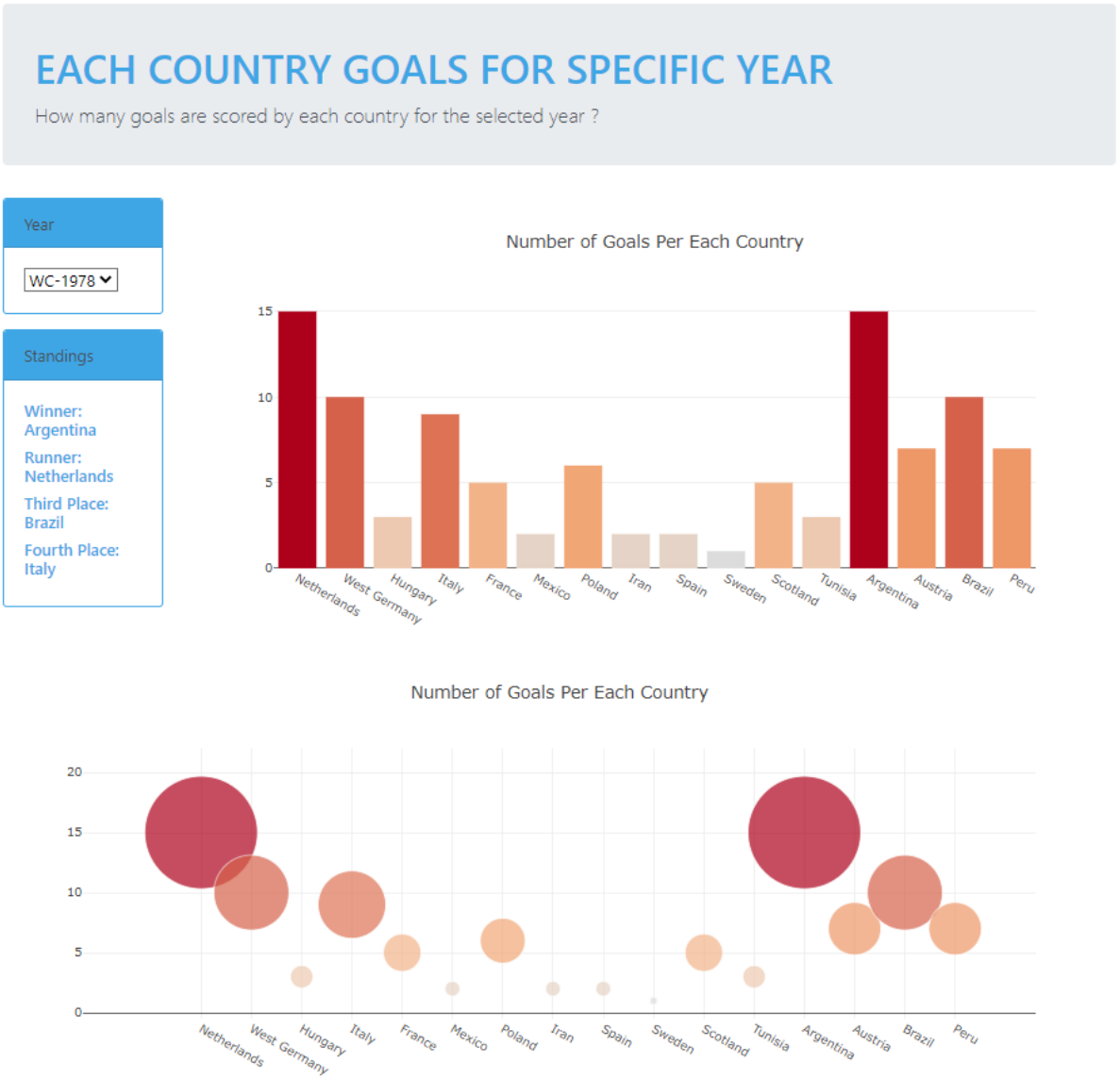


Average Goals Per Match

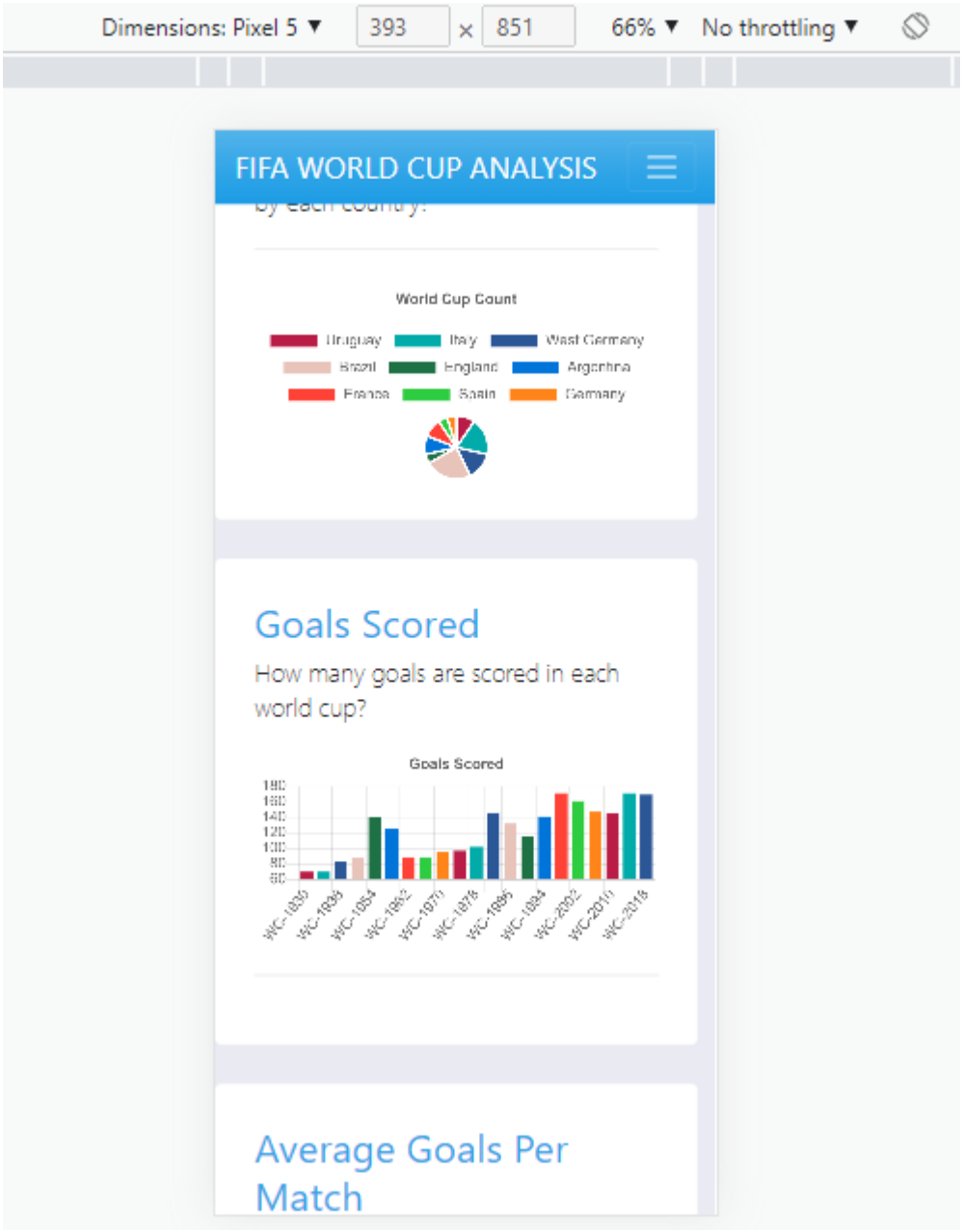
What is the average number of goals scored per match in each World Cup?

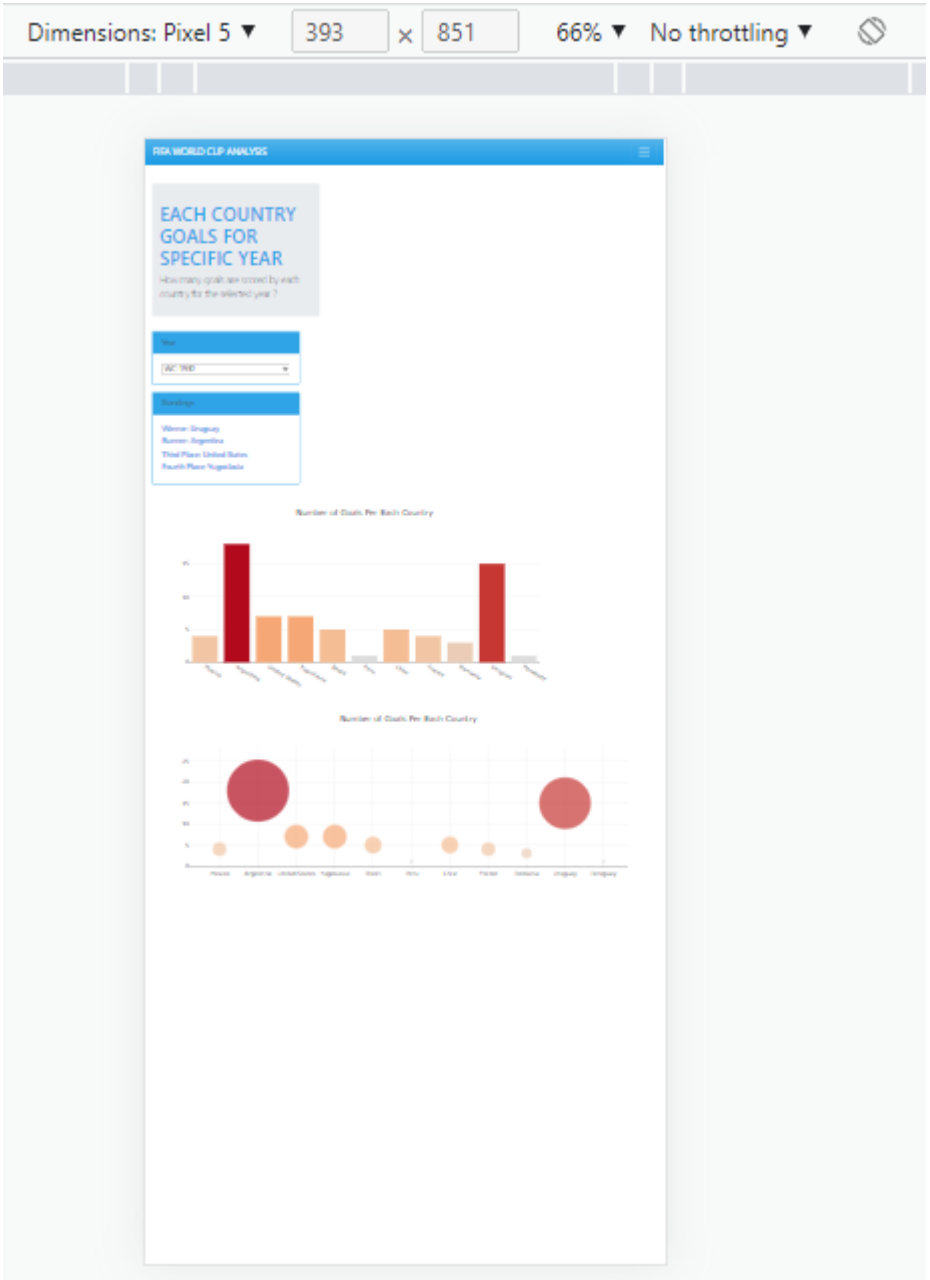


Click on Country Goals to get below:

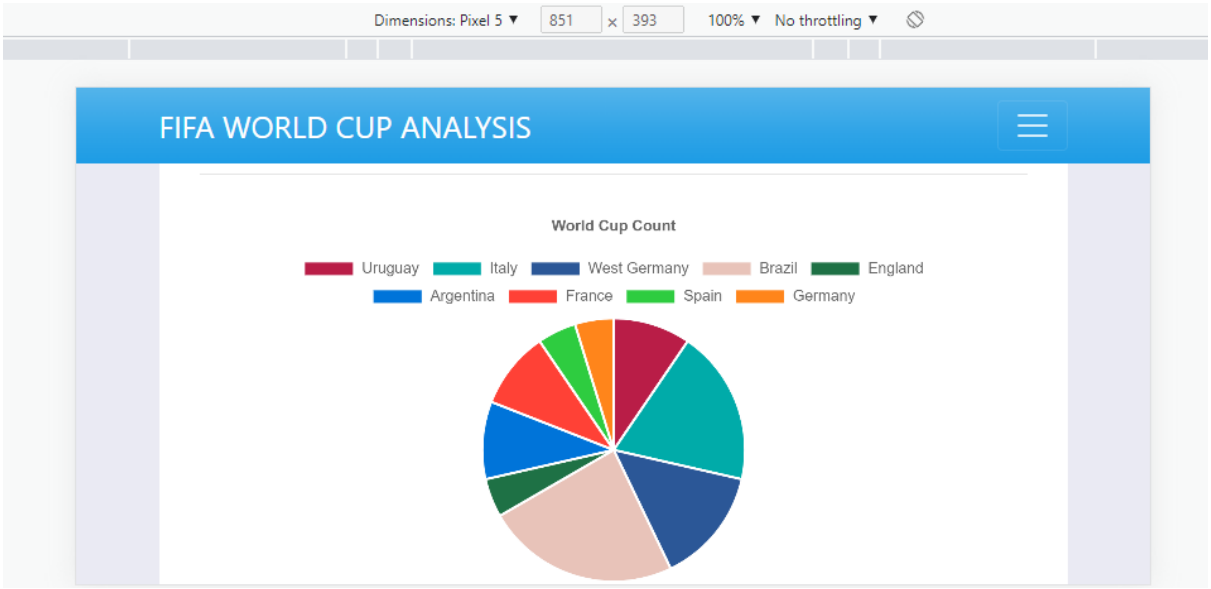


Responsive View:
Portrait View:


Dimensions: Pixel 5
393
851
66%
No throttling
FIFA WORLD CUP ANALYSIS
World Cup Count
Uruguay
Italy
West Germany
Brazil
England
Argentina
France
Spain
Germany
Goals Scored
How many goals are scored in each world cup?
Goals Scored
Average Goals Per Match



Landscape View:



Click on FIFA World Cup 2022 Teams to see below: