

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ  
ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΩΝ  
ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

---

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ  
ΑΚΑΔΗΜΑΪΚΟ ΈΤΟΣ 2024-2025

---

---

ΟΜΑΔΑ 4669-2962

ΕΜΜΑΝΟΥΗΛΙΔΗΣ ΕΜΜΑΝΟΥΗΛ, 4669

ΚΩΝΣΤΑΝΤΙΝΟΣ ΔΕΔΙΚΟΥΣΗΣ, 2962

---

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ 2025

## ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
16/03/2025	1.0	Σχεδιασμός της βάσης και δημιουργία data frame για την διαχείριση των δεδομένων	Κωνσταντίνος Δεδικούσης. Εμμανουηλίδης Εμμανουήλ
11/04/2025	2.0	Ολοκλήρωση της 2ης Φάσης	Κωνσταντίνος Δεδικούσης. Εμμανουηλίδης Εμμανουήλ
25/05/2025	3.0	Ολοκλήρωση της 3ης Φάσης	Κωνσταντίνος Δεδικούσης. Εμμανουηλίδης Εμμανουήλ

# 1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Στην παρούσα ενότητα περιγράφονται τα σχήματα της βάσης (ή βάσεων, αν είναι παραπάνω από μία) δεδομένων που χρησιμοποιούνται στο project.

## 1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ



Σχήμα 1.1 Σχεσιακό σχήμα της βάσης δεδομένων του συστήματος

```
CREATE TABLE IF NOT EXISTS countries (
    country_id INT AUTO_INCREMENT PRIMARY KEY,
    iso_code VARCHAR(10) NOT NULL,
    iso3_code VARCHAR(10) NOT NULL,
    display_name VARCHAR(100) NOT NULL,
    region_name VARCHAR(50),
    status VARCHAR(200),
    developed_status VARCHAR(20),
    population BIGINT
)
```

```
CREATE TABLE IF NOT EXISTS country_stats (  
    country_stat_id INT AUTO_INCREMENT PRIMARY KEY,  
    country_id INT,  
    country_name VARCHAR(100),  
    match_year INT,  
    wins INT,  
    losses INT,  
    draws INT,  
    home_wins INT,  
    home_losses INT,  
    home_draws INT,  
    away_wins INT,  
    away_losses INT,  
    away_draws INT,  
    shootout_participations INT DEFAULT 0,  
    FOREIGN KEY (country_id) REFERENCES countries (country_id)  
)  
  
CREATE TABLE IF NOT EXISTS matches (  
    match_id INT AUTO_INCREMENT PRIMARY KEY,  
    home_team_country_id INT,  
    away_team_country_id INT,  
    match_date DATE,  
    year INT,  
    home_score INT,  
    away_score INT,  
    tournament VARCHAR(100),  
    city VARCHAR(100),  
    country VARCHAR(100),  
    neutral BOOLEAN,  
    draw BOOLEAN,  
    had_shootout BOOLEAN DEFAULT FALSE,  
    FOREIGN KEY (home_team_country_id) REFERENCES countries (country_id),  
    FOREIGN KEY (away_team_country_id) REFERENCES countries (country_id)  
)  
  
CREATE TABLE IF NOT EXISTS player_goals (  
    player_id INT AUTO_INCREMENT PRIMARY KEY,  
    player_name VARCHAR(100),  
    match_date DATE,  
    year INT,  
    goals INT,  
    most_goals_scored INT,  
    team_id INT,  
    FOREIGN KEY (team_id) REFERENCES countries (country_id)  
)
```

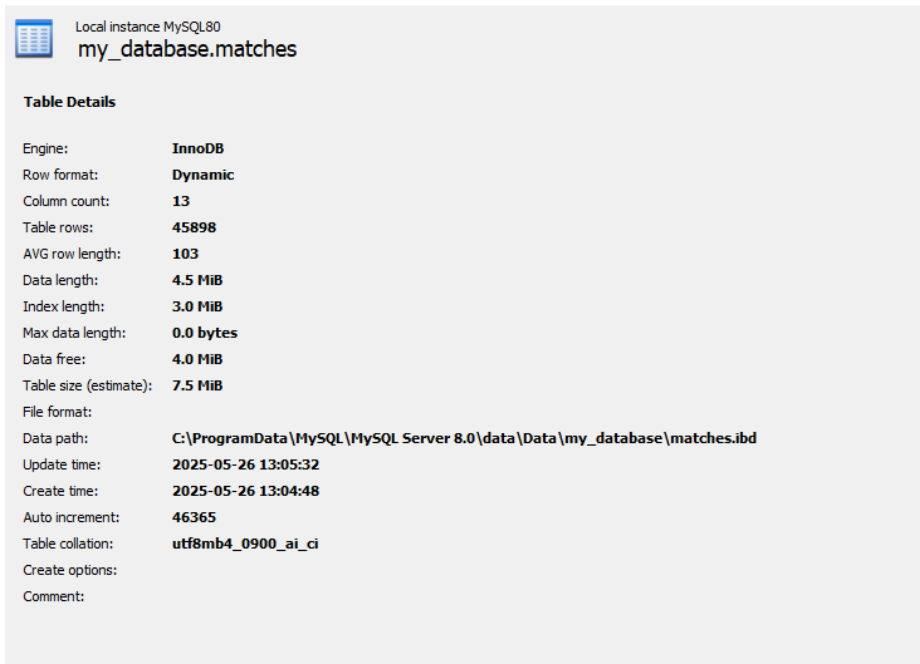
---

## 1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

---

Local instance MySQL80	
my_database.countries	
<b>Table Details</b>	
Engine:	InnoDB
Row format:	Dynamic
Column count:	8
Table rows:	252
AVG row length:	195
Data length:	48.0 KiB
Index length:	0.0 bytes
Max data length:	0.0 bytes
Data free:	0.0 bytes
Table size (estimate):	48.0 KiB
File format:	
Data path:	C:\ProgramData\MySQL\MySQL Server 8.0\data\Data\my_database\countries.ibd
Update time:	2025-05-26 13:04:48
Create time:	2025-05-26 13:04:48
Auto increment:	253
Table collation:	utf8mb4_0900_ai_ci
Create options:	
Comment:	

Local instance MySQL80	
my_database.country_stats	
<b>Table Details</b>	
Engine:	InnoDB
Row format:	Dynamic
Column count:	14
Table rows:	12978
AVG row length:	122
Data length:	1.5 MiB
Index length:	400.0 KiB
Max data length:	0.0 bytes
Data free:	4.0 MiB
Table size (estimate):	1.9 MiB
File format:	
Data path:	C:\ProgramData\MySQL\MySQL Server 8.0\data\Data\my_database\country_stats.ibd
Update time:	2025-05-26 13:05:32
Create time:	2025-05-26 13:04:48
Auto increment:	12797
Table collation:	utf8mb4_0900_ai_ci
Create options:	
Comment:	



Local instance MySQL80  
**my\_database.matches**

**Table Details**

Engine:	InnoDB
Row format:	Dynamic
Column count:	13
Table rows:	45898
AVG row length:	103
Data length:	4.5 MiB
Index length:	3.0 MiB
Max data length:	0.0 bytes
Data free:	4.0 MiB
Table size (estimate):	7.5 MiB
File format:	
Data path:	C:\ProgramData\MySQL\MySQL Server 8.0\data\Data\my_database\matches.ibd
Update time:	2025-05-26 13:05:32
Create time:	2025-05-26 13:04:48
Auto increment:	46365
Table collation:	utf8mb4_0900_ai_ci
Create options:	
Comment:	



Local instance MySQL80  
**my\_database.player\_goals**

**Table Details**

Engine:	InnoDB
Row format:	Dynamic
Column count:	7
Table rows:	36638
AVG row length:	71
Data length:	2.5 MiB
Index length:	1.5 MiB
Max data length:	0.0 bytes
Data free:	4.0 MiB
Table size (estimate):	4.0 MiB
File format:	
Data path:	C:\ProgramData\MySQL\MySQL Server 8.0\data\Data\my_database\player_goals.ibd
Update time:	2025-05-26 13:05:52
Create time:	2025-05-26 13:05:33
Auto increment:	36515
Table collation:	utf8mb4_0900_ai_ci
Create options:	
Comment:	

---

### 1.2.1 ΠΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

---

Για την υλοποίηση της βάσης δεδομένων επιλέχθηκε το DBMS MySQL 8.0, με InnoDB ως τον κύριο storage engine

Η διαχείριση μνήμης γίνεται μέσω των προκαθορισμένων buffer pools του InnoDB.

## 2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ

---

## APPLICATION TECH STACK

### Apollo Client

Ο **Apollo Client** είναι μια βιβλιοθήκη που διευκολύνει την επικοινωνία ενός React frontend με ένα GraphQL backend. Μέσω των custom hooks που δημιουργεί (useQuery, useMutation), επιτρέπει την εύκολη υποβολή ερωτημάτων και την παρακολούθηση της κατάστασης (loading, error, data). Συνδυάζεται με caching, error handling, polling, και επιπλέον παρέχει ενιαίο context διαχείρισης δεδομένων. Στο παρόν έργο, ο Apollo Client ενσωματώνεται με τον GraphQL Code Generator, ώστε τα queries να χρησιμοποιούνται με πλήρη type safety.

---

### Apollo Server

Ο **Apollo Server** είναι ένας έτοιμος προς χρήση HTTP server για GraphQL APIs. Υλοποιεί τη λειτουργία του backend, επεξεργαζόμενος τα εισερχόμενα queries από τον client, και επιστρέφει δομημένες απαντήσεις βάσει του σχήματος (schema.graphql) και των resolver συναρτήσεων (resolver.js). Παρέχει debugging, performance tracing, και είναι πλήρως επεκτάσιμο. Στην παρούσα εργασία, ο Apollo Server χρησιμοποιείται σε Node.js περιβάλλον με ES modules και ενοποιείται με τη βάση δεδομένων μέσω Sequelize.

---

### GraphQL

Η **GraphQL** είναι γλώσσα ερωτήσεων και runtime για APIs, σχεδιασμένη από το Facebook. Αντί για πολλά endpoints όπως στο REST, όλα τα δεδομένα ζητούνται μέσω ενός μόνο endpoint, με queries που καθορίζουν ακριβώς τι ζητά ο client. Προσφέρει αποδοτικότητα, προβλεψιμότητα και ευελιξία. Στην παρούσα εργασία, το schema ορίζεται με SDL (schema.graphql) και περιγράφει όλους τους τύπους δεδομένων και τα queries (π.χ. getAllMatchesByCountryId). Το backend υλοποιεί τα resolvers και το frontend τα καταναλώνει μέσω του Apollo Client.

---

### Sequelize

Το **Sequelize** είναι ORM (Object-Relational Mapping) για Node.js, που επιτρέπει την αλληλεπίδραση με σχεσιακές βάσεις δεδομένων όπως η MySQL, μέσω αντικειμενοστραφούς σύνταξης. Υποστηρίζει μοντέλα, σχέσεις, migrations και validation. Στην παρούσα εργασία χρησιμοποιήθηκε για την ορισμό των πινάκων countries, matches, country\_stats, player\_goals, και την εκτέλεση queries από τους resolvers του GraphQL API.

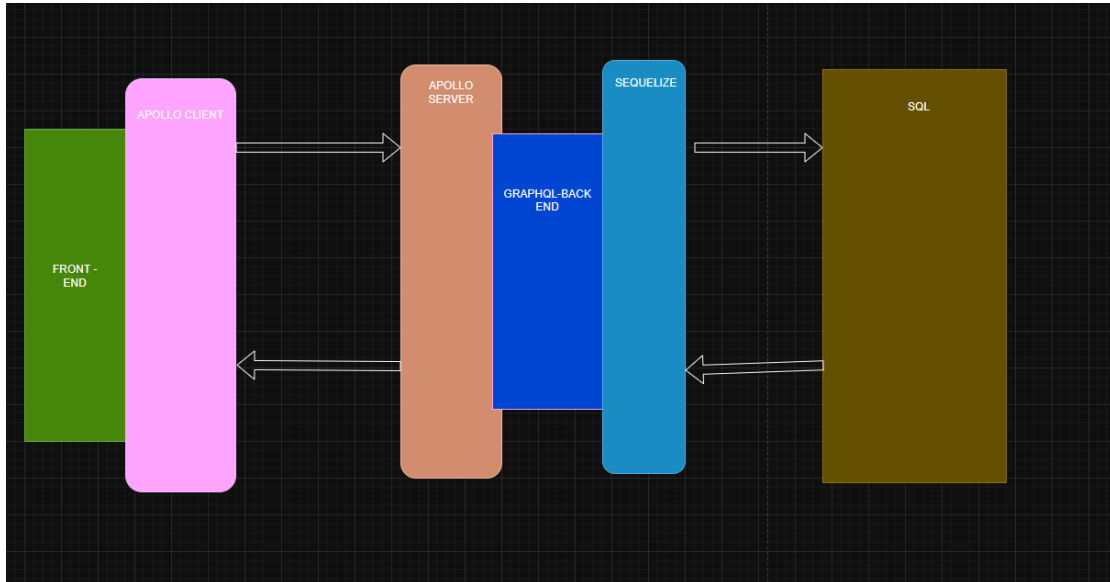
---

### React

Η **React** είναι βιβλιοθήκη της JavaScript για την κατασκευή user interfaces. Λειτουργεί με συνθετικά components και virtual DOM, προσφέροντας υψηλή απόδοση και modular

ανάπτυξη. Στην παρούσα εργασία χρησιμοποιήθηκε για την υλοποίηση όλων των διαδραστικών στοιχείων του frontend (πίνακες, γραφήματα, modal, φίλτρα), με Ant Design ως UI framework. Η εφαρμογή υλοποιείται με μοντέρνα hooks (useState, useEffect) και ενσωματώνει αποδοτικά τα δεδομένα από το backend μέσω Apollo hooks.

### TECH STACK OVERVIEW



### Χρήση GraphQL Code Generator

Για τη διασύνδεση μεταξύ του frontend (React + Apollo Client) και του GraphQL backend, χρησιμοποιήθηκε το εργαλείο **GraphQL Code Generator** (@graphql-codegen/cli). Το εργαλείο αυτό επιτρέπει την αυτόματη δημιουργία **τύπων TypeScript** και **custom React hooks** (όπως useGetAllCountriesQuery) βασισμένων στα .graphql αρχεία των queries και στο schema του backend.

Η λειτουργία του βασίζεται στην παρακάτω ροή:

1. **Καθορίζεται το schema** του GraphQL API (π.χ. μέσω του URL `http://localhost:4000` ή ενός τοπικού αρχείου .graphql).
2. **Συλλέγονται όλα τα .graphql αρχεία** με τα queries και τα fragments που χρησιμοποιεί ο client.
3. Το εργαλείο παράγει **τύπους TypeScript για όλα τα queries και τα αποτελέσματά τους**, καθώς και **έτοιμα hooks** για χρήση στον React κώδικα.
4. Ο developer μπορεί να χρησιμοποιεί αυτά τα hooks (π.χ. useGetAllMatchesByCountryIdQuery) με πλήρη **type-safety**, χωρίς να γράφει χειροκίνητα το query ή να ανησυχεί για το shape της απάντησης.



Το αρχείο ρυθμίσεων `codegen.ts` ορίζει το `schema`, τα `documents`, και τα `plugins` (`typescript`, `typescript-operations`, `typescript-react-apollo`). Έτσι διασφαλίζεται:

- Ασφαλής και τυποποιημένη πρόσβαση στα δεδομένα του `backend`,
- Πλήρης συμβατότητα με το `schema`, ακόμα και αν αυτό αλλάξει,
- Μείωση των λαθών κατά την ανάπτυξη του `frontend`.

---

## 2.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΔΟΜΗ ETL

---

Για την ETL διαδικασία χρησιμοποιήσαμε ένα `python` script (`testscript.py`) το οποίο υλοποιεί μια ολοκληρωμένη διαδικασία εισαγωγής, καθαρισμού και επεξεργασίας δεδομένων από αρχεία CSV που αφορούν διεθνείς ποδοσφαιρικούς αγώνες, σκόρερς και πληροφορίες χωρών. Στόχος είναι η δημιουργία και ο πληθυσμός σχεσιακών πινάκων σε βάση δεδομένων MySQL, καθώς και ο υπολογισμός στατιστικών στοιχείων ανά χώρα και παίκτη.

Παρακάτω παρατίθεται η αναλυτική λειτουργία και λογική:

### 1. Είσοδοι – Αρχεία

Το script διαβάζει τα εξής αρχεία CSV:

- `countries.csv`: Πληροφορίες χωρών (κωδικοί ISO, πληθυσμός, όνομα κ.ά.).
- `results.csv`: Αποτελέσματα διεθνών ποδοσφαιρικών αγώνων.
- `shootouts.csv`: Αγώνες που κρίθηκαν με πέναλτι.
- `goalscorers.csv`: Πληροφορίες για παίκτες που σκόραραν.

Τα αρχεία καθαρίζονται από `null bytes` και διαβάζονται με έλεγχο πολλών `encodings`

### 2. Σύνδεση και Διαχείριση MySQL

Το script συνδέεται στη βάση `my_database` και αρχικά διαγράφει (αν υπάρχουν) τους πίνακες:

- `countries`
- `country_stats`
- `matches`
- `player_goals`

Στη συνέχεια, δημιουργούνται οι πίνακες με τη χρήση SQL εντολών, με κατάλληλους τύπους δεδομένων και ξένους κλειδιά (FOREIGN KEY constraints).

### 3. Αντιστοίχιση και Καταγραφή Δεδομένων

#### 3.1 Χώρες (countries)

- Κάθε χώρα από το countries.csv εισάγεται στον αντίστοιχο πίνακα.
- Δημιουργείται map display\_name → country\_id για χρήση σε επόμενα στάδια.

#### 3.2 Αποτελέσματα Αγώνων & Πέναλτι (matches, shootouts)

- Γίνεται fuzzy αντιστοίχιση των ομάδων με ονόματα χωρών (χρήση βιβλιοθήκης RapidFuzz).
- Οι αγώνες καταγράφονται στον πίνακα matches με πληροφορίες όπως σκορ, ημερομηνία, διοργάνωση, ουδέτερη έδρα, κ.λπ.
- Εντοπίζονται αγώνες που κρίθηκαν στα πέναλτι με βάση το shootouts.csv.

#### 3.3 Στατιστικά Χωρών (country\_stats)

- Για κάθε αγώνα υπολογίζονται:
  - Νίκες/Ηττες/Ισοπαλίες
  - Νίκες/Ηττες/Ισοπαλίες ως γηπεδούχος ή φιλοξενούμενη
  - Συμμετοχές σε πέναλτι
- Τα στατιστικά αποθηκεύονται στον πίνακα country\_stats ανά χώρα και έτος.

### 4. Καταγραφή Γκολ Παικτών (player\_goals)


- Τα own goals αγνοούνται.
- Για κάθε παίκτη και ημερομηνία, υπολογίζεται ο αριθμός γκολ που σημείωσε.
- Καταγράφεται και το μέγιστο πλήθος γκολ που πέτυχε ένας παίκτης σε έναν αγώνα (most\_goals\_scored).
- Τα δεδομένα εισάγονται στον πίνακα player\_goals.


## 2.2 APPLICATION COMPONENT DIAGRAM




### 3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

Η αρχική σελίδα της εφαρμογής είναι ως εξής:

 Stats Per Country/Year

 Countries Leaderboard

 Scorers Leaderboard

---

#### Stats Per Country

Select a Country ▼

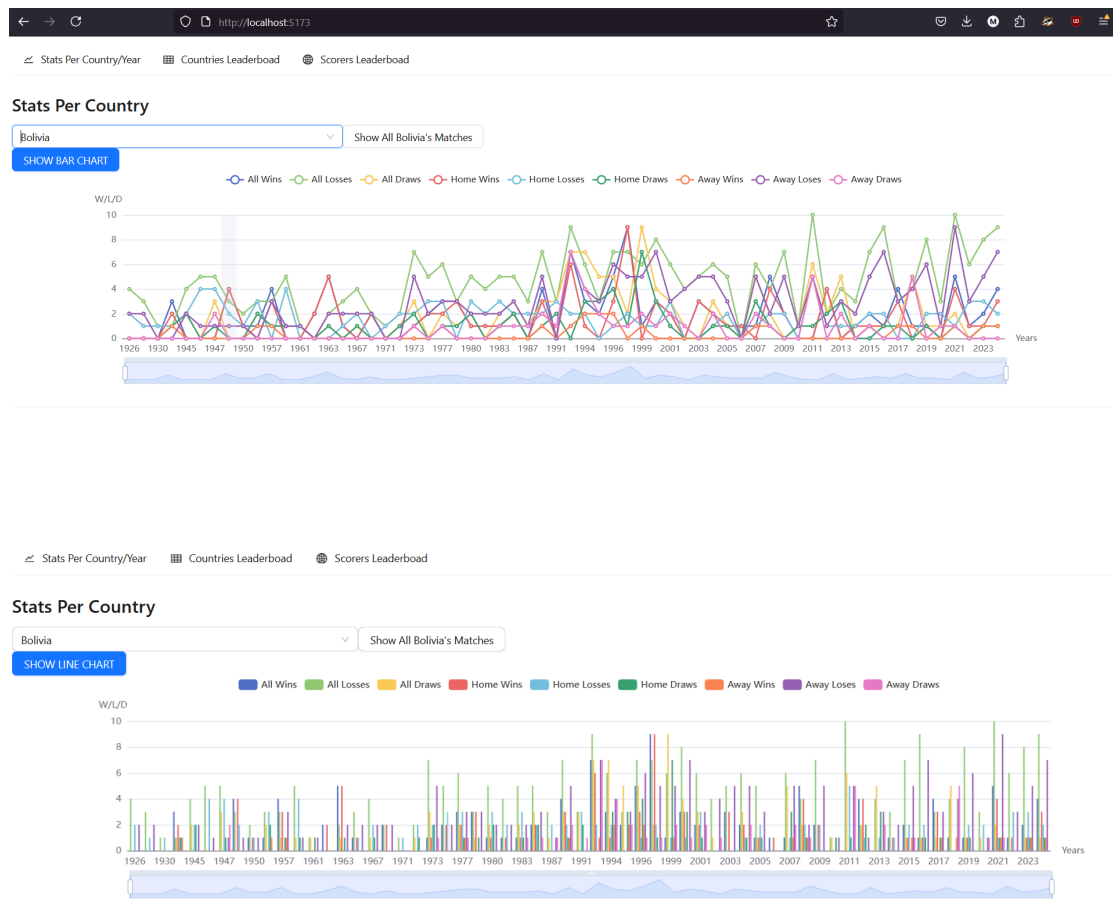
---

#### Stats Per Year

Select a Year ▼

Στην αρχική σελίδα ο χρήστης βρίσκει δυο selectors. Εναν για τις χώρες και έναν για τις χρονιές έτσι ώστε να μπορεί να επιλέξει την χώρα όπως και την χρονιά της οποίας θέλει να δει τα στατιστικά. Οι παραπάνω selectors είναι φτιαγμένοι ώστε να παρέχουν ταυτόχρονα input το οποίο φιλτράρει real-time τις παρεχόμενες επιλογές.

Όταν γίνει μια επιλογή χώρας, εμφανίζεται το linechart με επιλογή αλλαγής σε bar chart. Το chart δείχνει διαφορά στατιστικά (W/L/D) ανά χρονιά για μια συγκεκριμένη ομάδα.



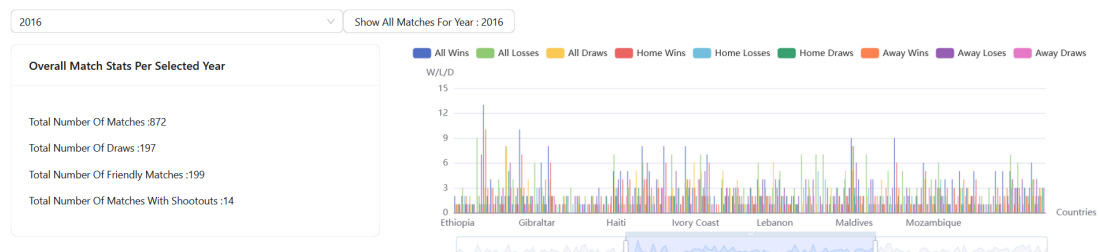
Να σημειωθεί ότι σε κάθε γράφημα υποστηρίζεται zoom in/out και με το scroll αλλά και με βοηθητικούς sliders στο κάτω μέρος του ώστε να μπορεί ο χρήστης να δει πιο εύκολα στατιστικά που τον ενδιαφέρουν. Επίσης τα legends στο πάνω μέρος (All Wins etc.) είναι togglable παρέχοντας την δυνατότητα στον χρήστη απομόνωσης ορισμένων στατιστικών.

Επίσης, πατώντας το κουμπί Show All <Country Name> 's Matches εμφανίζεται ένας πίνακας με όλα τα καταγεγραμμένα παιχνίδια της ομάδας αυτής ως Home Team, με επιλογή φιλτραρίσματος του αντιπάλου ως προς κάποια στοιχεία (Developed/Developing, Status, Region).

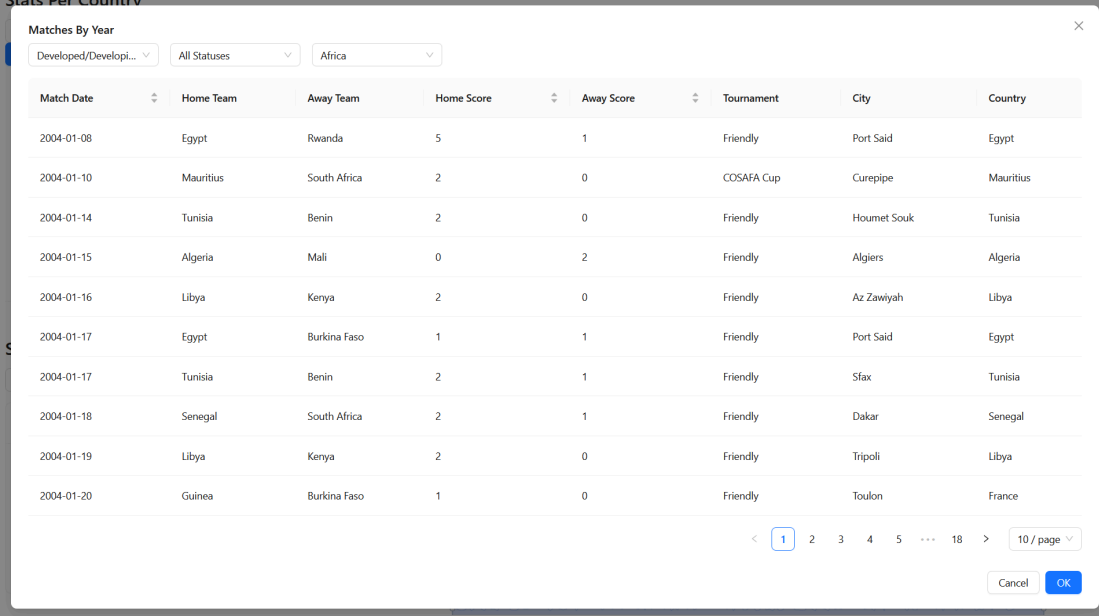
Matches By Year								
Developed/Developi...		All Statuses		All Regions				
Match Date	Home Team	Away Team	Home Score	Away Score	Tournament	City	Country	
1926-10-23	Bolivia	Paraguay	1	6	Copa América	Santiago	Chile	
1926-10-28	Bolivia	Uruguay	0	6	Copa América	Santiago	Chile	
1927-11-06	Bolivia	Uruguay	0	9	Copa América	Lima	Peru	
1930-07-20	Bolivia	Brazil	0	4	FIFA World Cup	Montevideo	Uruguay	
1938-08-08	Bolivia	Ecuador	1	1	Bolivarian Games	Bogotá	Colombia	
1938-08-11	Bolivia	Venezuela	3	1	Bolivarian Games	Bogotá	Colombia	
1938-08-14	Bolivia	Peru	0	3	Bolivarian Games	Bogota	Colombia	
1938-08-22	Bolivia	Ecuador	2	1	Bolivarian Games	Bogotá	Colombia	
1945-01-28	Bolivia	Brazil	0	2	Copa América	Santiago	Chile	
1945-02-11	Bolivia	Ecuador	0	0	Copa América	Santiago	Chile	

Όταν γίνει μια επιλογή χρονιάς, εμφανίζονται τα καθολικά στατιστικά της και το bar chart με όλες τις ομάδες που έχουν παίξει ματς την χρονιά αυτή.

#### Stats Per Year



Πατώντας το κουμπί Show All Matches For <Year> εμφανίζεται ένας πίνακας με όλα τα καταγεγραμμένα παιχνίδια της χρονιάς αυτής, με επιλογή φιλτραρίσματος των χωρών που εμφανίζονται ως προς κάποια στοιχεία (Developed/Developing, Status, Region). Το φιλτράρισμα γίνεται ΚΑΙ στις δύο χώρες.



Match Date	Home Team	Away Team	Home Score	Away Score	Tournament	City	Country
2004-01-08	Egypt	Rwanda	5	1	Friendly	Port Said	Egypt
2004-01-10	Mauritius	South Africa	2	0	COSAFA Cup	Curepipe	Mauritius
2004-01-14	Tunisia	Benin	2	0	Friendly	Houmet Souk	Tunisia
2004-01-15	Algeria	Mali	0	2	Friendly	Algiers	Algeria
2004-01-16	Libya	Kenya	2	0	Friendly	Az Zawiyah	Libya
2004-01-17	Egypt	Burkina Faso	1	1	Friendly	Port Said	Egypt
2004-01-17	Tunisia	Benin	2	1	Friendly	Sfax	Tunisia
2004-01-18	Senegal	South Africa	2	1	Friendly	Dakar	Senegal
2004-01-19	Libya	Kenya	2	0	Friendly	Tripoli	Libya
2004-01-20	Guinea	Burkina Faso	1	0	Friendly	Toulon	France

Στην κορυφή του παραθύρου, μέσω του navigation bar, ο χρήστης μπορεί να πατήσει το κουμπί Countries Leaderboard για να δει ταξινομημένες όλες τις χώρες, με βάση το στατιστικό που θέλει.

Αρχικά έχει οριστεί ως default φίλτρο το Score. Ο χρήστης μπορεί να επιλέξει διαφορετικό φίλτρο μέσω των παρεχόμενων φίλτρων (switches) αλλά μόνο ένα κάθε φορά. Αν ο χρήστης δεν επιλέξει κάποιο φίλτρο δεν γίνεται καμία ταξινόμηση και απλά γίνεται ο πίνακας displayed.

Επίσης μπορεί να φιλτράρει τις χώρες με βάση κάποια στοιχεία τους όπως Developed, Status, Region. Να σημειωθεί εδώ ότι οι επιλογές των φίλτρων αυτών παρέχονται από τα ίδια τα δεδομένα δυναμικά.

Παρακάτω φαίνεται η ταξινομημένη λίστα, καθώς και τα σχετικά barcharts και scatterplots

[Stats Per Country/Year](#) [Countries Leaderboard](#) [Scorers Leaderboard](#)

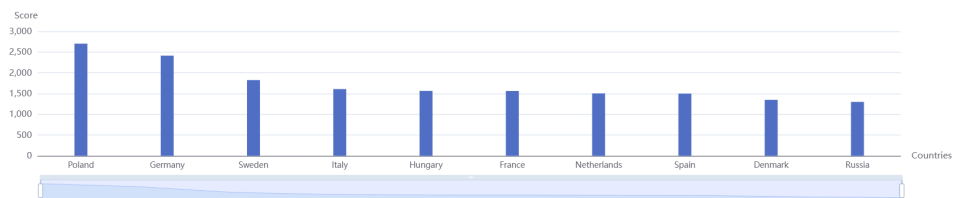
## Global Scoring Stats

☐ Total Wins ☒ Score ☐ Avg Wins/Year ☐ Avg Score/Year  

Country Name	Region	Status	Developed	Total Wins	Score	Avg Wins/Year	Avg Score/Year
Poland	Europe	Independent	Developed	768	2701	5.408	19.021
Germany	Europe	Independent	Developed	713	2412	6.790	22.971
Sweden	Europe	Independent	Developed	532	1824	4.626	15.861
Italy	Europe	Independent	Developed	457	1607	4.193	14.743
Hungary	Europe	Independent	Developed	451	1563	3.697	12.811
France	Europe	Independent	Developed	458	1561	3.983	13.574
Netherlands	Europe	Independent	Developed	437	1503	3.937	13.541
Spain	Europe	Independent	Developed	442	1498	4.465	15.131
Denmark	Europe	Independent	Developed	389	1348	3.473	12.036
Russia	Europe	Independent	Developed	369	1298	4.500	15.829

&lt; 1 2 3 4 5 &gt;

## Stats Per Country



## Score X Population



Τέλος, πατώντας το 3ο κουμπί στην πλοήγηση, ο χρήστης βλέπει έναν Selector με όλους τους παίκτες. Στην συνέχεια μπορεί να επιλέξει τον παίκτη που επιθυμεί και έτσι να εμφανίσει τα επιθυμητά στατιστικά και το linechart για αυτόν τον παίκτη.

Τα στατιστικά του παίκτη γίνονται display σε ένα Card Component το οποίο δείχνει και τις χρονιές του παίκτη με την μορφή slider.

Αν ο χρήστης μετακινήσει τον slider το γραφικό γίνεται rerendered για το καινούριο εύρος χρονιών που διάλεξε ο χρήστης. Ακολούθως επανυπολογίζονται τα totalNumberOfGoals και τα averageGoalsPerMatch

Τέλος το γράφημα δείχνει Goals ανά χρονιά του παίκτη.

