## **Introduction**

We propose a three-phase development approach for "Awareia," an AI-powered youth soccer scouting and analytics platform. The project involves:

1. **Data Extraction & Scraping** – Extracting match data from external sources into a structured database.
2. **Web Application Development** – Building a web-based platform for coaches to access insights, submit reports, and interact with AI-generated data.
3. **AI Integration** – Implementing AI-powered analysis for scouting reports, SWOT assessments, and strategic recommendations.

#### **Phase 1: Web Scraping & Data Extraction**

**Objective:** Extract match data from league sources (EDP & WYSL) and store it in a structured database for further processing.

**Scope:**

* Develop a **web scraper** to extract match history, team rankings, goal differentials, and other relevant statistics from:
  1. **EDP**: <https://newyorkclubsoccer.com/nycsl/> [Available history all the way up to Spring 2025 alignments]
  2. **WYSL**: <https://system.gotsport.com/org_event/events/34449/clubs> [Available history all the way up to Spring 2025 alignments]
* Store data in a **PostgreSQL/MySQL database**, structured into at least four tables:
  1. **Clubs**– Club Name, Number of teams, contactInfo etc.
  2. Game # — Very important (That will come in match details)
  3. **Teams**– TeamName, Gender etc.
  4. **StandingTable** – Team, GA, GD etc.
  5. **Matches** – complete match details.
  6. You should include other tables as well so data can be filtered as much as possible like: Gender, Location, State, League, etc. (Okay I will try to add as much information as its available)
  7. Review what these guys are doing on their site because they scrape from the other sites as wel to build theirsl: <https://youthsoccergames.com/#gsc.tab=0> (Sure I will inspect that in detail)
* Set up an **automated data refresh** process to keep information up to date.
* Ensure data accuracy and format consistency for AI processing.

**Timeline:** 3 to 4 days  
**Estimated Cost:** **XXXX [Started 2.12.2025]**

#### **Phase 2: Web Development & Platform Build**

**Objective:** Build the **Awareia web platform** for coaches to access, submit, and interact with match data and scouting insights.

**Scope:**

* **User Registration & Team Association**
  + Coaches sign up and associate with their team.
  + Access past match data and scouting reports.
* **Dashboard & Data Display**
  + Interactive league tables, match histories, and team analytics
  + Searchable database of opponent scouting reports.
  + Should also give coaches the ability to easily data enter their roster of players by name 1st preferred position and 2nd preferred position and allow them to rate easily rate the the player on 1) Technical skills, 2) Mental toughness, 3) Game IQ 4) Athleticism — This is a structured data where the AI can suggest players based on this input (This will also given by the coaches as input?)
* **Scouting Report Submission**
  + Text-based report submission.
  + Voice-to-text integration for easy input.
* **SWOT Analysis & Performance Tracking**
  + Team and player performance summaries based on data trends.
  + Historical performance comparison.
* **Subscription & Monetization**
  + Free tier with basic access.
  + Paid tiers with in-depth AI-driven insights and opponent analysis.

**Tech Stack:**

* **Frontend:** React.js
* **Backend:** Node.js
* **Database:** PostgreSQL / MongoDB

**Timeline:** 4-5 weeks  
**Estimated Cost:** **XXXX$**

#### **Phase 3: AI Integration & Advanced Features**

**Objective:** Implement AI-powered insights, opponent scouting reports, and voice-based input processing.

**Scope:**

* **AI-Generated Opponent Insights**
  + Aggregation of scouting reports into structured summaries.
  + AI analysis of opponent strengths, weaknesses, and tactics.
* **Voice-to-Text AI for Coach Input**
  + Whisper AI integration for transcribing coach assessments.
  + AI-assisted structuring of reports.
* **SWOT Analysis & Automated Match Insights**
  + AI compiles team strengths and weaknesses based on match data.
  + Personalized reports before each match.
* **Push Notifications & Match Alerts**
  + AI-generated scouting reports delivered before match days.
  + Performance trend alerts for teams and players.

**Tech Stack:**

* **AI Model:** OpenAI GPT-4 for text summarization & match insights.
* **Voice Processing:** Whisper AI for voice input transcription.

**Timeline:** 3-4 weeks  
**Estimated Cost:** **XXXX**

### **Total Project Timeline & Cost**

* **Phase 1: Web Scraping & Data Extraction** – 3**-4 days**
* **Phase 2: Web Development** – 4**-5 weeks**
* **Phase 3: AI Integration** – **3-4 weeks**
* **Total Estimated Duration:** **-7-8 weeks**
* **Total Estimated Cost:** **XXXX$**

### **Next Steps**

1. Review and finalize project scope and cost breakdown.
2. Begin development of Phase 1 (scraping).
3. Set up project milestones and deliverables for each phase.

Let me know if you have any questions or would like adjustments to the approach. Looking forward to working together on this!