Lafayette College Senior Project 2020

Cameron Zurmuhl, Joseph Teddick, Matt Gerber, and Xingwen Wei

Soccer Hotspots

The Project

- Develop a web application that will display high schools of players from 34 institutions across 11 conferences with Division-One soccer programs on a Google map.
- Schools of interests were those of high academic caliber. Data collection starts 2006.
- The user can filter the data to observe those high schools and players that have distinguished accolades
    - Number of years the player obtained an all-conference award
    - Number of years starter

Examples of Data

- School Roster Sites
- Starter Data
- All-Conference Data
- Acquired high school geography information from College Board and the National Center for Education Statistics

Soccer Hotspots

Project History

- Coach Dennis Bohn of Lafayette College's Men's Soccer Team had the idea in Fall of 2018, inspired by literature that researches how talent appears in "hotspots". For example, Westlake High School in Austin, TX.
- Cameron Zurmuhl started the project as an EXCEL Scholar under Professor Trent Gaugler from Fall 2018-Spring 2019.
- Due to its scale, the project was transferred to the CS department as a class project. Last semester, Professor Ordille's databases class worked on a prototype, which was enhanced this term in the senior project class.

Soccer Hotspots

Problems we Inherited

- Database did not account for international students
- High School matching scheme across tables was not accurate nor robust
  - A player from California was displayed as being in Jamaica
- Filters were semantically confusing
- UI had individual markers per player, but the markers did not group

Soccer Hotspots

Our Mission

- Create a new, inclusive database
  - Correct matching and grouping
  - Contains all information Dennis Bohn wants for filters
  - Get international students in the database
- Improve the front-end and make it more user-friendly
  - Mark the high schools on map with the ability to see more info about players
- Consider how future data integration will be accomplished

Soccer Hotspots

Tools and Platforms We Used

- Python - Used for scraping the data and our website
- PostgreSQL - Database management system
- Django - Frontend framework
- Javascript - Google Maps and front end scripting
- HTML & CSS - Frontend UI and design

Agile Development - Story Backlog

Name (Green=completed, yellow=in progress)	Description	Point Value	Status
Production Launch	Launch to server	8	Completed
Manual Data Upload	Add the ability, that Dennis requested, to manually enter in new matched data (that you could confirm through your own research).	8	Completed
Waiting and Loading Functionality	Add waiting and loading functionality to the data upload and dump restore screens, so that the user cannot interact with the site until it is complete. This will be indicated by a pop up message.	5	Completed
Documentation and Comments	Make documentation and comment our code more	5	Completed
More robust data matching process	Come up with a way to get more matches uses methods such as levenshtein distance and looking at boarding schools	5	Completed
Database Version Control	Implement a version control system for our back-end database	8	Completed
CSS Refinements	make the site beautiful	3	Completed
UI Enhancements	Display the data in a better way with the table. When a user zooms in on the map of an area, the tables should update accordingly. Add in appropriate filters that the customer wants. Clean up the drop pins so there is only one drop pin per high school, and display the player information more coherently and in an organized way. Display a clickable bio links of players. **UI necessities: Figure out how to post data with ajax so page doesn't refresh. If a league is selected, all colleges in that league should be preselected on the UI. When an ajax call**	5	Completed
Django & Database Integration	Review Django framework that the team last semester used. Make the necessary modifications to fit with out new database.	2	Completed
Matching High School Names across tables	Figure out a robust, accurate, and persistent way to match the high school names in the public database to our tables so we know locations of the schools.	8	Completed
Building a Database	We need to rebuild the current database and make schema to incorporate cleaner tables for our project. We also need to find a way to consolidate the players in our roster table so there are no duplicate players across roster years. At the same time, we need to makes sure all data persists after grouping.	5	Completed
GPS Coordinates for Foreign High Schools	Geocode the foreign high schools csv data	3	Completed

Soccer Hotspots

Implemented Features!

- Gathered International High School Data and geocoded it
- Created a master High School information file
- Rebuilt database to include all information Coach Dennis wanted
- Completed 100% accurate Player and High School matching strategy applying special techniques for Public, Private, International, and Boarding schools.
- Grouped up player data to combine into one entry for display
- Built an intuitive and attractive front end to integrate with the data
- Included a Google Map to display a heatmap, hotspots, and high schools with player counts
- Ability to export data to a .csv file
- Ability to upload scraped data in future years
- Ability to upload manual data for personal investigation into matching players
- Version control to ensure data is backed up and can be restored (with “loading screen” functionality)
- Deploying to server

Player and High School Matching

- Matching Logic Flow
- Originally had 14.8% increased it to 15.8% by adding boarding and alternative schools to the matching strategy (592 unique players).
- Designed around the ability to upload more data to make application more intelligent

Uploading New Data

- Scraped Data
  - Upload 3 files: roster, starter, and accolade data
  - Trigger in Database will match (according to our logic), group up, and union all the data automatically
- Manual Entry
  - Bypass the matching process (user guarantees match)
  - Intended for personal investigation
  - Trigger will handle grouping and unioning the data automatically

Soccer Hotspots

Filtering

- Selecting one college league will select all the colleges in that league
  - User can then deselect/select more colleges in the college dropdown
- Selecting nothing is equivalent to selecting everything in the dropdown. It is as if the user does not care to filter that dropdown.

Soccer Hotspots

Google Map - Pricing

- We use the Google Map JavaScript API for displaying the icons
- Pricing
  - At most $14 per 1,000 page loads
- $300 credit for first year of being a new Google Cloud user
- $200 recurring monthly credit for Maps API
- So to have a charge, there would need to be more than 14,285 (conservative estimate) page loads per month, not taking the $300 into account.
  - When developing the site, we used 4,000 API requests

Google Map - Features

- Marker Clustering / Infowindows
- Heat Map
- Map / satellite view
- Reset table

Future Enhancements

- Adding in scraped data for more years
- Enhance matching algorithm
- Extend the backup/save feature for manual uploads

Demo!

Thank You!

- Coach Bohn
- Dr. Joann Ordille