Project: Recruit

Comprehensive Data-Driven Analysis of High-School Football Talent Pipelines (2015–2020)

Executive Summary

This project integrates over **700 athlete records** spanning the 2015–2020 recruiting classes to identify the **most reliable high-school talent pipelines in the United States**. Using JavaScript-based data extraction and multi-table SQL analysis, the study evaluates three core dimensions of talent production:

1. **Top-150 Recruit Output**
2. **Position-Specific Pipeline Strength**
3. **NFL Draft Conversion Rates**

The objective is to provide **actionable recruiting intelligence** by identifying high schools and regions that consistently produce high-impact prospects. The analysis confirms that certain programs—most notably **IMG Academy, St. Thomas Aquinas, Bishop Gorman, Allen (TX), and Trinity Christian Academy**—function as statistically validated recruiting hubs with outsized influence on college and NFL talent pipelines.

1. Methodology Overview

Data Acquisition and Processing

- Extracted approximately **700 rows** of structured athlete data using custom JavaScript scraping methods.
- Imported datasets into a **multi-table SQL relational schema**, separated by recruiting class year.
- Standardized all attributes (player name, school, position, hometown, ranking) and validated integrity through cross-year normalization.

Analytical Framework

The project applied SQL-based aggregation to measure:

- **Absolute talent output** (top-150 and top-32 counts)
- **Positional concentration** (e.g., CB, WR, EDGE pipelines)
- **NFL draft success metrics** (flagged via cross-table tagging)

This enables **horizontal comparisons across years** and **vertical comparisons within schools**, producing a multidimensional view of program performance.

2. Top 100–150 Talent Production (Macro Pipeline Strength)

Across all classes (2015–2020), the analysis shows that a small number of schools consistently produce elite talent. These programs represent the highest-yield recruiting targets nationally.

Key Findings

- **IMG Academy** produced **19 top-150 recruits**, the highest total in the dataset.
- **St. Thomas Aquinas** and **Bishop Gorman** follow as the strongest *native* pipelines (programs that primarily draw local student-athletes).
- The repeated appearance of these schools across six independent datasets indicates **pipeline reliability**, not random fluctuation.

Interpretation for Recruiters

- Schools with multi-year consistency provide **low-variance recruiting outcomes**, reducing the risk associated with prospect evaluation.
- These programs demonstrate stable coaching infrastructures and advanced development environments, increasing the likelihood that athletes arriving at the collegiate level are physically, mentally, and technically prepared.

3. Positional Pipeline Analysis (Micro-Level Talent Specialization)

The SQL grouping of all tables by position reveals which high schools specialize in producing certain types of athletes.

Example: Cornerback (CB) Production

- **IMG Academy** and **Trinity Christian Academy** each produced **three top-150 CBs** during the analysis window.
- Cornerback is a position with high developmental variance, making consistent production extremely valuable.

Analytical Implications

This pattern signals:

- Specialized coaching for defensive back technique
- Systematic exposure to elite competition
- Repeatable development traits (e.g., frame, speed, change-of-direction capabilities)

Strategic Use

When a recruiting class requires targeted reinforcements (e.g., defensive backs), these programs should be considered **priority evaluation sites** due to positional reliability.

4. Five-Star Athlete Production (Top-32 Output)

This segment of the analysis focuses on the rarest subset of prospects: top-32 athletes, commonly classified as **five-star recruits**.

Key Findings

- **IMG Academy: 9 five-star recruits** between 2015–2020

- **Allen High School (TX): 2 five-star recruits**, placing it among the top non-IMG programs in the country

Interpretation

Producing a five-star athlete is an extremely rare event—most high schools produce **zero** across multiple decades. IMG's nine five-stars in six years represent an **institutional anomaly** unmatched by traditional programs.

Strategic Implications

- IMG's disproportionate concentration of top-tier talent positions it as the **single most valuable recruiting hub** in the United States.
- Allen High School's output, though smaller, reflects **organic regional dominance**, especially at quarterback—historically one of the most predictive positions of collegiate success.

5. NFL Draft Conversion Metrics (Ultimate Development Indicator)

Evaluating which high schools produce **NFL draftees** provides the most meaningful long-term validation of a program's developmental quality.

Key Findings

- IMG Academy produced **seven (7) NFL draftees** in the observed five-year window.
- Given that **<1% of all high school athletes reach the NFL**, this output places IMG at the **extreme upper boundary** of development efficiency.

Why This Matters

NFL draft conversion reflects:

- High-level strength and conditioning foundations

- Robust coaching and technical development
- Strong competitive environments
- Player readiness for college systems (early contributors, low-bust risk)

Recruiter Interpretation

IMG's conversion rate confirms it as a **low-risk, high-certainty recruiting environment**. Even average IMG recruits tend to outperform their non-IMG peers in college readiness and adaptability.

6. Advanced Enhancement Opportunities

Your dataset supports several analytical expansions that would elevate this project into a professional-grade recruiting intelligence tool.

A. High School Efficiency Index (HEI)

Create an index combining:

- Top-150 production
- Five-star production
- NFL draft success
- Positional specialization weighting

This normalizes schools on a 0–1 scale, enabling **objective comparisons** across programs.

B. Regional and Geographic Heat Mapping

Using hometown fields, produce maps showing:

- Talent density clusters
- State and metro-level production trends
- Regional volatility or consistency

This becomes highly valuable for travel planning and staff resource allocation.

C. Temporal Trend Analysis

Measuring early vs. late-year changes can highlight:

- Emerging programs
- Declining pipelines
- Shifts in development structures

D. Position Scarcity Index

Assign higher weights to positions that are harder to evaluate or typically lower in supply (e.g., QB, EDGE, OT).

7. Final Conclusion

Across all metrics—top-150 output, positional concentration, five-star production, and NFL draft conversion—the data demonstrates that:

- **IMG Academy** is the most dominant talent pipeline in the nation, both in quantity and quality.
- **St. Thomas Aquinas**, **Bishop Gorman**, **Allen**, and **Trinity Christian Academy** form the next tier of highly reliable recruiting environments.
- Certain programs offer **position-specific advantages** that can inform targeted recruiting strategies.
- NFL conversion analysis further validates which schools produce not merely elite high-school athletes, but **long-term high-performance outcomes**.

Your SQL-driven methodology presents a replicable model for evaluating recruiting efficiency and can serve as a foundation for a full-scale recruiting analytics platform.