NCAA Football Analysis Case Study (2013–2023)

Executive Summary

This project analyzes NCAA football team statistics from 2013 to 2023 to identify which performance metrics most strongly predict team success. Using data from the NCAA via Kaggle, the analysis explores offensive and defensive performance, third-down and red-zone efficiency, turnover margins, and special teams metrics.

- Efficiency metrics (e.g., third-down conversion) explain 72% of win percentage variation.
- Balanced teams (strong offense and defense) have an 84.7% win rate.
- Special teams have minimal predictive power (only 2%).

Recommendations focus on training for efficiency, balanced development, and conference-specific strategies. This analysis supports data-informed decision-making for recruiting, practice planning, and game strategy.

Table of Contents

Key Metrics Snapshot

- **Seasons Analyzed:** 2013–2023
- **Total Teams:** ~130 FBS programs
- **Total Metrics:** 145+ team stats per season
- **Models Built:** Full, Offense-only, Defense-only, Special Teams, Efficiency
- **Best R^2:** 0.74 (Full Model), 0.72 (Efficiency Model)
- **Most Predictive Metrics:** Third Down Conversion, Turnover Margin

Phase 1: Ask

- **Client:** College sports analytics firm working with NCAA programs.
- **Business Task:** Identify performance metrics that predict team success.
- **Stakeholders:** Coaches, athletic directors, recruitment analysts.
- **Guiding Questions:**
    - Which metrics most predict wins?
    - Is defense more important than offense?
    - Which conferences dominate?
    - How important are special teams?
    - What role does third-down efficiency play?

Phase 2: Prepare

- **Source:** Kaggle Dataset (College Football Team Stats)
- **Coverage:** 2013–2023, all FBS teams
- **Metrics:** Offense, Defense, ST, Turnovers, Red Zone, Efficiency
- **Credibility:** Scraped from NCAA.com, reviewed annually
- **License:** Open Database Commons

Phase 3: Process

- **Tools:** Google Sheets, Excel
- **Cleaning Tasks:**
    - Imputed missing values using season medians
    - Removed 8 duplicate entries
    - Standardized column names
    - Created derived metrics (e.g., efficiency ratios)

Step	Tool	Impact	Details
Missing Values	Google Sheets	~3%	Median imputation
Duplicate Removal	Google Sheets	8 records	Removed by unique team-year
Column Renaming	Google Sheets	All	Acronyms clarified
Metric Creation	Google Sheets	10+ fields	Efficiency & performance rates

Phase 4: Analyze

- **Methods:** EDA, correlation analysis, regression, ANOVA
- **Key Insights:**
    - Efficiency metrics = strongest predictors (72% variance explained)
    - High offense + high defense = 84.7% win rate
    - Special teams = minor impact
- **Surprises:** Balance outperforms specialization; third-down performance highly predictive.

Phase 5: Share

- **Format:** HTML dashboard-style report with visual insights

- **Tools:** Google Sheets, Canva
- **Visuals:** Heatmaps, model charts, win rate boxes
- **Design Choices:** Clear labels, team-focused insights, accessible formatting

Phase 6: Act

Recommendations

- Prioritize third-down/red-zone efficiency training
- Balance development between offense and defense
- Tailor strategies based on conference competition
- Deploy special teams tactically—not strategically

Next Steps

- Integrate player-level data (QB performance, etc.)
- Analyze recruiting class outcomes
- Use game-level and play-by-play data in future models

Tools & Technologies Used

- Google Sheets – cleaning, transformation, modeling
- Excel – conversions and formatting
- Canva – visual design and annotations
- HTML/CSS – dashboard presentation

Reflection

This project strengthened my ability to work with large datasets, translate statistics into strategy, and communicate findings effectively. If I were to extend it, I would explore advanced modeling

(e.g., logistic regression), bring in player-level data, and build interactive dashboards.