NBA Fantasy Basketball Predictor

A Data Management System for Sports Analytics

Harsh and Aryan

The Problem: Data Fragmentation

CURRENT STATE:

Managers must manually toggle between 5+ tabs (Box Scores, Injury Reports, Betting Odds) to make decisions.

SOLUTION:

A Unified Data Management System that centralizes ingestion, storage, and analysis into a single “Source of Truth”.

System Architecture: The ETL Pipeline

Extraction

nbastats.csv

- Ingesting raw nbastats.csv with 500+ player records

Transformation

- Cleaning null values
- handling UTF-8 encoding (e.g., 'Dončić')
- applying ESPN scoring weights

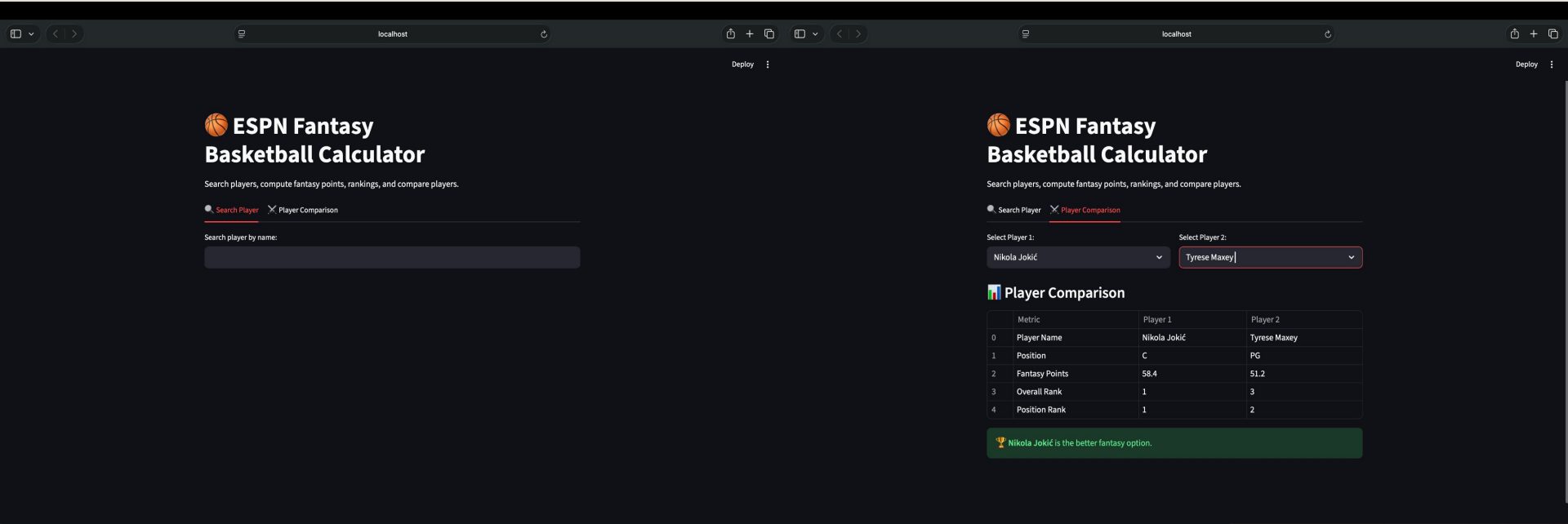
Loading

fantasy.db

- Storing structured data into persistent SQLite database named fantasy.db

How does it work?

The following screenshot shows how our project has two interfaces one where you can search any nba player and get their Overall rank , position rank and fantasy points average , the other tab helps in comparing players and tells you which player is better



Database Design: SQL Implementation

The system transitions from flat files (like CSVs) to a robust Relational Database (2NF), improving data structure and manageability.

Players Table (Dimension)

Player_ID (PK)	Name	Team	Position
1	LeBron James	LAL	SF
2	Stephen Curry	GSW	PG

Player_Stats Table (Fact)

Stat_ID (PK)	Player_ID (FK)	Points	Rebounds	Assists
101	1	25.4	7.8	10.2
102	2	29.1	6.1	6.3

- Normalization ensures data integrity and reduces redundancy.

Methodology: SQL-Based Analysis

1. True Value Calculator

Ranking players by actual production, not reputation.

2. Efficiency Hunter

Calculating Fantasy Points Per Minute to find “Sleepers”.

3. Positional Scarcity

Using GROUP BY to find which positions are hardest to fill.

Analytical Results

1. Undervalued Asset

Defensive stats (Stocks) are the most undervalued asset.

2. Top Waiver-Wire Targets

The ‘Efficiency Per Minute’ query identified backup Centers as top waiver-wire targets.

3. ‘Big Man First’ Strategy

Positional Analysis reveals a surplus of Guards but a scarcity of elite Centers, dictating a ‘Big Man First’ draft strategy.

Visualization: Streamlit Dashboard

The user-facing web application provides an interactive interface to the SQL-based analysis.

1. Dynamic Search

Real-time SQL querying of player ranks.

2. Comparator Engine

A Head-to-Head tool compares two players and algorithmically declares a ‘Winner’.

This democratizes access to the data without requiring SQL knowledge.

Future Roadmap

Phase 1: Real-Time API

Replacing CSVs with live nba_api data.

Phase 2: Predictive ML

Training a Random Forest model to forecast future games.

Phase 3: Cloud Deployment

Hosting the database on AWS.

This project successfully transformed raw, fragmented data into a competitive strategic advantage.

Thank you