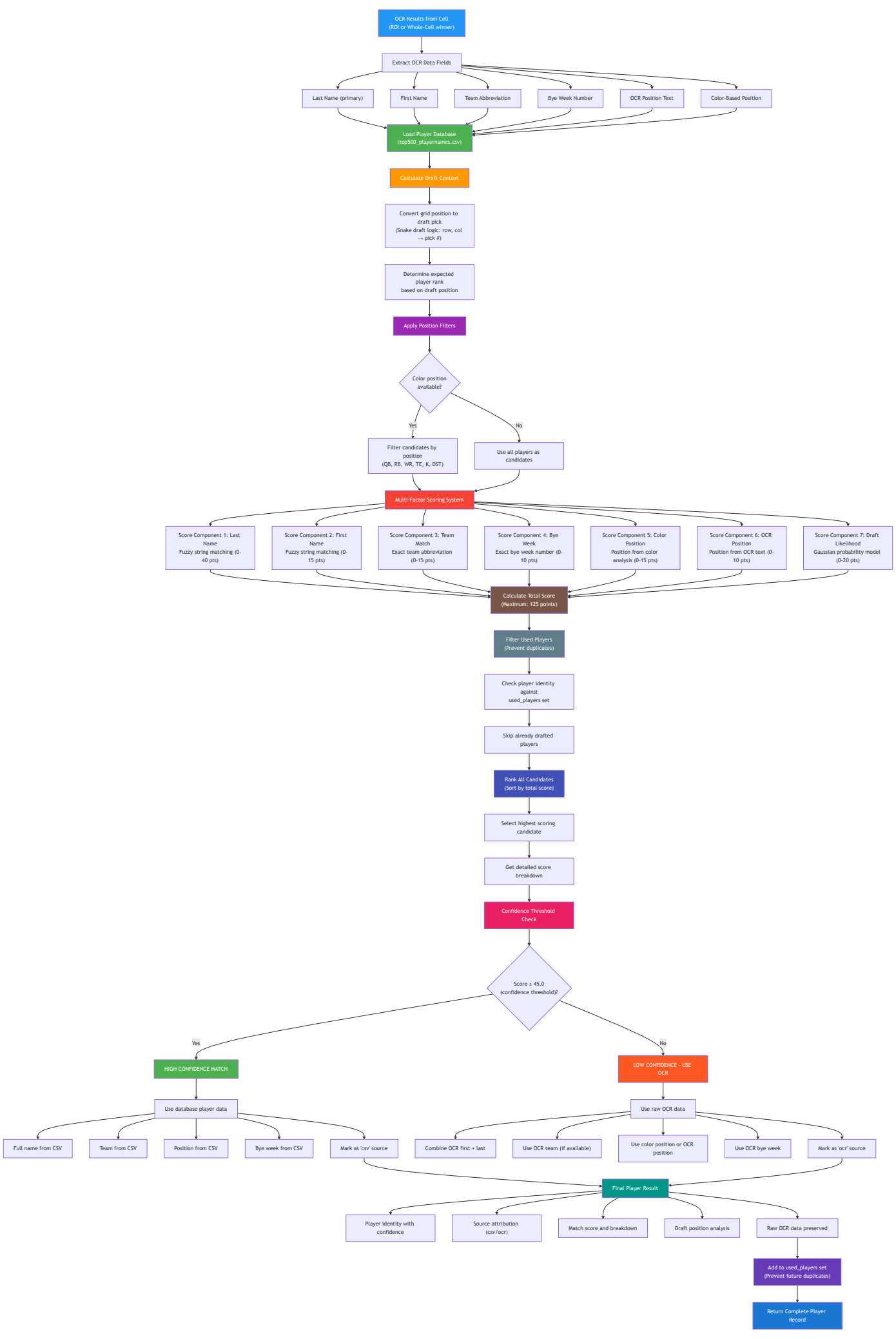
Player Name Prediction and Reconciliation Process



Process Stages Legend

- **OCR Input:** Winning OCR results from dual-strategy pipeline
- **Database Integration:** Load and process player database with rankings
- **Draft Context:** Calculate draft position and expected player rank
- **Position Filtering:** Use color-based position to narrow candidates
- **Multi-Factor Scoring:** 7-component scoring system (125 max points)
- **Score Calculation:** Combine all scoring components
- **Duplicate Prevention:** Filter out already-used players
- **Candidate Ranking:** Sort and select best match
- **Confidence Decision:** Choose database vs OCR data based on score
- **High Confidence:** Use verified database player data (≥45 points)
- **Low Confidence:** Fall back to raw OCR data (<45 points)
- **Final Result:** Complete player record with metadata
- **Usage Tracking:** Add to used players set
- **Complete Record:** Return final player data

Multi-Factor Scoring System (125 Points Maximum)

Last Name (40 pts)
Fuzzy string matching using token_set_ratio
Handles OCR errors and name variations

First Name (15 pts)
Additional fuzzy matching when available
Provides extra validation

Team Match (15 pts)
Exact team abbreviation matching
High confidence when teams align

Bye Week (10 pts)
Exact bye week number matching
Additional validation factor

Color Position (15 pts)
Position from color analysis
Strong positional validation

OCR Position (10 pts)
Position from OCR text recognition
Secondary positional validation

Draft Likelihood (20 pts)
Gaussian probability model based on Average Draft Position (ADP)
Considers draft position context and player rankings
Sigma grows with rank to reflect real draft variance

Key Intelligence Features:

- **Fuzzy Matching:** Handles OCR errors using Levenshtein distance algorithms
- **Draft Context:** Uses snake draft logic and ADP rankings for realistic predictions
- **Multi-Modal Validation:** Combines text, color, positional, and statistical data
- **Confidence Thresholding:** 45-point threshold ensures data quality
- **Graceful Degradation:** Falls back to OCR when database matching fails
- **Duplicate Prevention:** Sophisticated identity tracking prevents double-drafting
- **Transparency:** Detailed score breakdown enables debugging and validation
- **Source Attribution:** Tracks whether data came from database or OCR

Draft Likelihood Model:

The system uses a sophisticated Gaussian probability model:

- **Variable Sigma:** $\sigma = 2.0 + 0.1 \times \text{player\_rank}$ (uncertainty grows with rank)
- **Z-Score Calculation:** $z = (\text{draft\_pick} - \text{player\_rank}) / \sigma$
- **Probability Score:** $100 \times e^{(-0.5 \times z^2)}$
- **Real Draft Variance:** Early picks are more predictable than late picks
- **Context Awareness:** Considers both player ADP and actual draft position