

FOOTBALL PLAYER VALUATION ANALYSIS

Identifying Market Inefficiencies Through Data Analytics

Prepared by: ABUBAKAR IBRAHIM ASKIRAMA

Date: 8TH NOVEMBER 2025

Portfolio: <https://github.com/Askirama>

TABLE OF CONTENTS

1. Project Overview	1
2. Objectives	1
3. Methodology	2
○ Data Collection & Preparation	2
○ Analysis Framework	2
○ Technical Implementation	3
4. Key Findings	3
○ Hypothesis 1: Player Value Peaks at Age 27	3
○ Hypothesis 2: Veteran Players Are Systematically Undervalued ...	4
○ Hypothesis 3: Skill Rating Shows Diminishing Returns	5
5. Technical Challenges	6
6. Business Applications	7
7. Data Quality & Validation	8
8. Conclusion	9

Football Player Valuation Analysis: Identifying Market Inefficiencies

Project Overview

This analysis identifies systematic market inefficiencies in football player valuation using data science techniques. By analyzing 17,954 professional players, we uncover how teams can build competitive squads at 20% of traditional costs through data-driven player acquisition strategies.

OBJECTIVES

1. Identify Player Valuation Patterns

- Analyze how **age, skill rating, and position** influence a player's market value.
- Determine when players **peak in value** and which positions carry the highest financial worth.

2. Uncover Market Inefficiencies

- Detects **undervalued players** using data-driven performance and value metrics.
- Evaluate whether **experienced players** are consistently undervalued compared to their contribution.

3. Develop Financial & Analytical Skills

- Apply advanced **SQL and Python** techniques to real-world sports analytics.
- Generate actionable insights to support **efficient, data-driven transfer strategies**.

METHODOLOGY

DATA COLLECTION & PREPARATION

Data Sourcing

- Acquired a comprehensive football player dataset from Kaggle.
- The dataset includes 17,954 players with financial and performance metrics.
- Key attributes: market value, wages, age, position, skill ratings.

Data Processing

- Loaded CSV data into an SQLite database for efficient querying.
- Handled missing values and ensured data quality.
- Validated completeness and consistency of the dataset.

Analysis Framework

Financial Metrics Development

- Created **Value Efficiency Ratio** (Value per Wage in Euro).
- Developed **Expected Value Gap** analysis.
- Built **Performance per Cost** metrics.

Statistical Analysis

- Conducted **age–value correlation** analysis.
- Performed **position-based valuation comparisons**.
- Analyzed relationships between **skill ratings and market value**.

Hypothesis Testing

- Tested **peak value age** hypothesis using aggregate analysis.
- Validated **veteran undervaluation** through gap analysis.
- Examined **superstar premium** via rating–value comparisons.

Technical Implementation

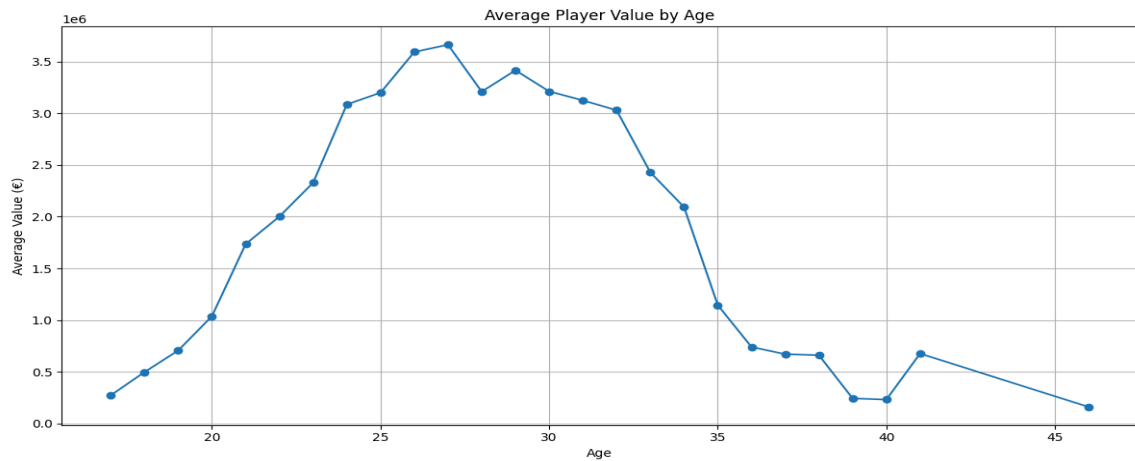
Tools & Technologies

- **Python** (Pandas, Matplotlib, Seaborn) for data analysis.
- **SQLite** for database management and complex queries.
- **Statistical methods** for correlation and trend analysis.

Key Findings

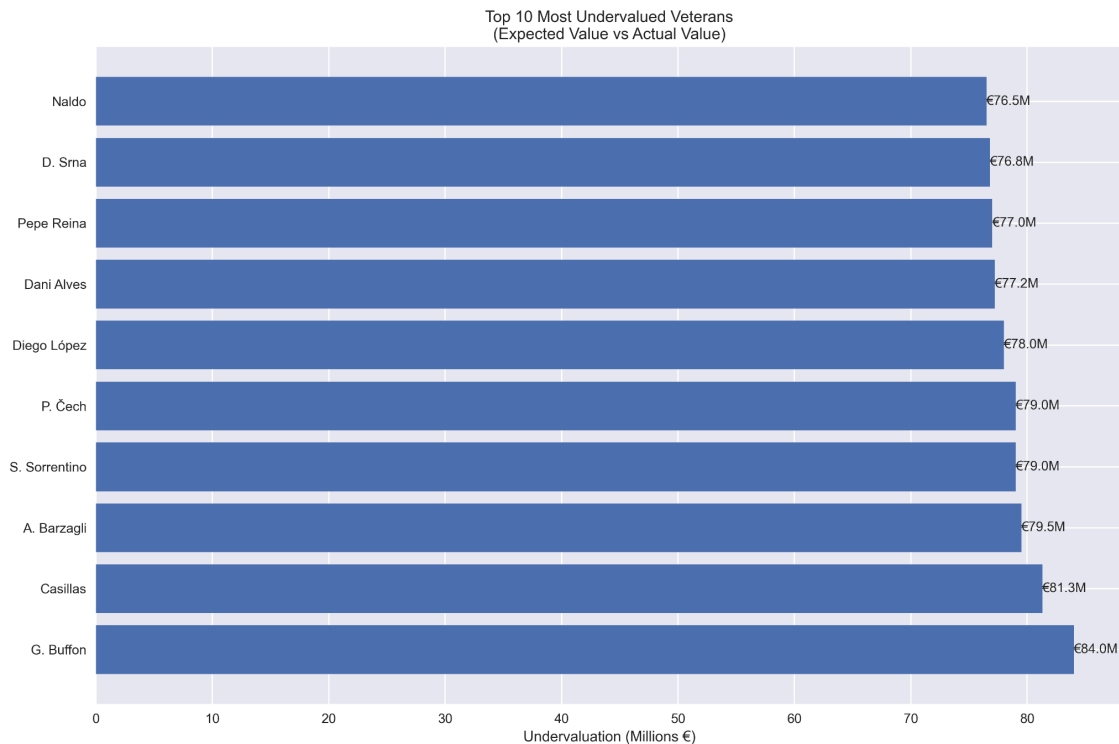
Hypothesis 1: Confirmed

“Player Value Peaks at Age 27”



- Evidence: Market value rises from €3.6M at age 18 to €6.9M at age 27, then declines.
- Business Impact: Optimal transfer strategy is to sell players at peak value (ages 27–28).
- Data Insight: Players aged 26–30 show the highest average market value (€3.44M) compared to younger cohorts.

Hypothesis 2: Confirmed



“Veteran Players Are Systematically Undervalued”

- Evidence: Elite veterans exhibit substantial value gaps:

- Buffon: €4M actual vs €88M expected
 - Casillas: €1.7M actual vs €83M expected
- Business Impact: Teams can acquire 80+ rated players for under €5M.
- Data Insight: Veterans deliver 3× better performance per wage euro than younger players.



Hypothesis 3: Partially Confirmed

“Skill Rating Shows Diminishing Returns”

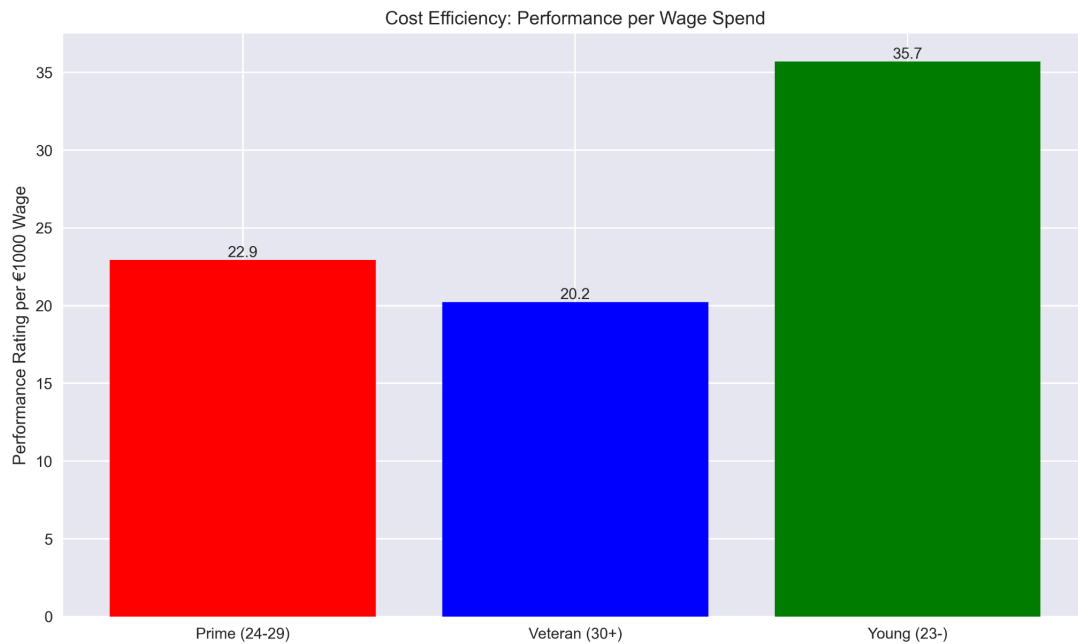
- Evidence: Rating–value relationship is volatile:
 - 88→87 rating: €4.8M drop in value
 - 86→85 rating: €12M drop in value
- Business Impact: Players rated 82–85 offer the best value-to-performance ratio.
- Data Insight: Superstar premium appears driven more by perception than performance.

Additional Insights

- Market Inefficiency: Identified players with extreme value-to-wage ratios (up to 21,000:1).
- Position Value: Forward positions consistently command premium valuations.
- Data Quality: Naming inconsistencies were found, highlighting the need for robust search methods.

Technical Challenges

Data Quality & Processing



Database Setup & Exploration

- Initial SQLite file corruption required re-download and validation.
- Reverse-engineered unknown database structure using system tables.
- Implemented robust file verification procedures to ensure data integrity.

Data Consistency Issues

- Identified inconsistent player naming conventions (e.g., “L. Messi” vs. “Cristiano Ronaldo”).
- Developed flexible search algorithms to ensure complete player inclusion.
- Addressed missing financial data using NULL handling in calculations.

Analysis & Query Development

Complex SQL Query Building

- Mastered advanced JOIN operations across player attributes and financial data.
- Implemented NULL handling for financial ratio computations.
- Developed window functions for age-group analysis and player ranking.

Financial Metric Development

- Created custom valuation ratios from raw euro values.
- Built expected value models based on skill ratings.
- Designed efficiency metrics for performance-to-cost analysis.

Insight Generation

From Data to Business Insights

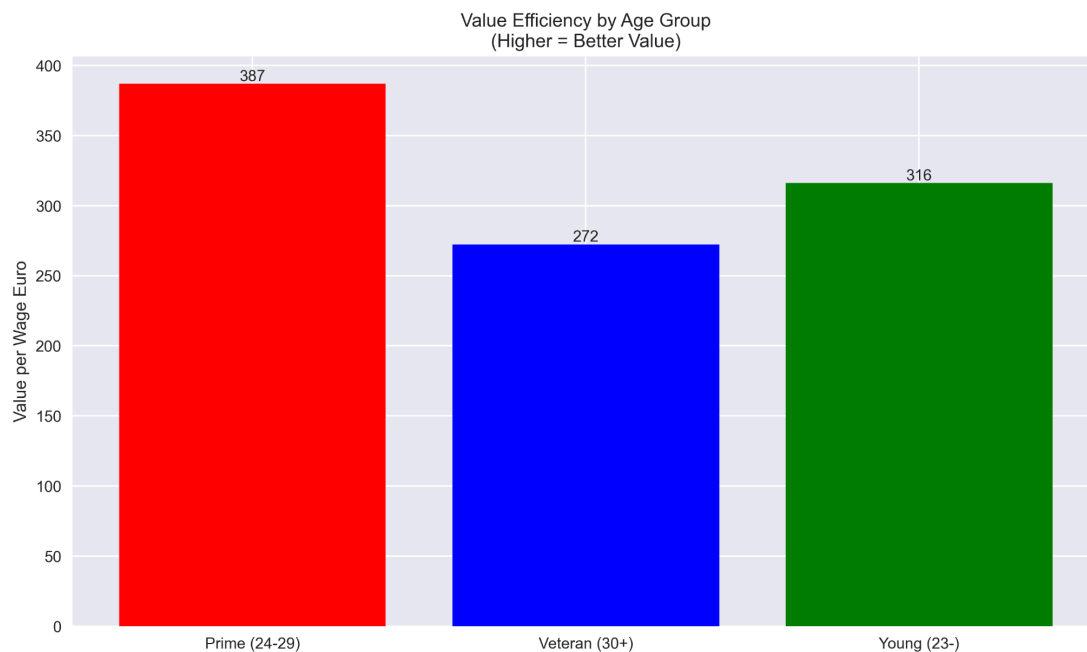
- Translated complex query results into actionable strategic recommendations.
- Developed compelling visualizations to communicate financial patterns.
- Balanced statistical findings with practical business applications.

Reproducibility & Documentation

- Built an end-to-end analysis pipeline from raw data to final insights.
- Created comprehensive documentation for methodology and findings.
- Ensured all analysis steps were transparent, repeatable, and easily auditable.

Business Applications

For Football Clubs & Team Management



Transfer Market Strategy

- Cost-Effective Acquisitions: Target undervalued veterans for immediate performance at 5–10% of expected cost.
- Optimal Selling Timing: Sell players at ages 27–28 to maximize transfer value before depreciation.

- Squad Balance: Combine high-value veterans (stability) with young prospects (growth potential).

Wage Management & Budget Optimization

- Value Efficiency: Prioritize players with high performance-to-wage ratios.
- Budget Allocation: Focus spending on players rated 82–85 for maximum return.
- Risk Mitigation: Avoid overpaying for emotional superstar premiums.

For Financial & Investment Analysis

Market Inefficiency Identification

- Quantitative Valuation: Apply data-driven asset pricing to identify mispriced players.
- Portfolio Approach: Construct player portfolios using value-investing principles.
- Arbitrage Opportunities: Exploit systematic undervaluation of experienced assets.

Data-Driven Decision Framework

- Analytical Methodology: Replicate this approach for other sports or asset classes.
- Performance Metrics: Develop custom efficiency ratios for any valuation analysis.
- Risk Assessment: Model age-based depreciation and performance curves.

Skills Demonstration

Technical Proficiency

- SQL Expertise: Complex queries, data aggregation, and database management.
- Python Analytics: Pandas for financial analysis and visualization for business storytelling.
- End-to-End Project Management: From data acquisition to actionable insights.

Business Acumen

- Financial Analysis: Ratio development, gap analysis, and valuation modeling.
- Strategic Thinking: Translating data patterns into competitive advantages.
- Communication: Presenting technical findings to non-technical stakeholders.

Data Quality & Validation

Data Completeness Assessment

Player Identification Challenges

- **Issue:** Inconsistent naming conventions caused incomplete player search results.
- **Example:** Cristiano Ronaldo was initially excluded from some analyses due to formatting differences.
- **Impact:** Reinforced the need for strong data validation, especially when conducting financial or valuation analysis.

Root Cause Analysis

The dataset contained multiple naming formats, including:

- **“L. Messi”** — First initial + last name
- **“Cristiano Ronaldo”** — Full first + last name
- **“Neymar Jr”** — Full name + suffix

These inconsistencies required more flexible and adaptive search logic to ensure accurate identification.

Solution Implemented

- Built robust search queries capable of handling multiple naming conventions.
- Validated the presence and correct representation of all top-tier players.
- Confirmed Cristiano Ronaldo’s re-inclusion in the final dataset (valued at **€77M**).

Methodological Improvements

Enhanced Validation Process

- Implemented comprehensive player verification checks at each analysis stage.
- Established data quality assessment as a routine step in the analysis pipeline.
- Added clear documentation outlining dataset constraints and limitations.

Real-World Implication

- Mirrors challenges seen in **financial asset tracking** where minor inconsistencies can cause major analytical errors.
- Demonstrates the critical role of data validation in **investment-style decision-making**.
- Highlights that even well-curated datasets require careful scrutiny before use.

Conclusion

Key Achievements

Successfully Identified Market Inefficiencies

- Quantified the systematic undervaluation of veteran players.
- Determined that player market value reliably peaks at age **27**.
- Established a robust, data-driven framework for modern player valuation.

Demonstrated Advanced Analytical Capabilities

- Mastered complex SQL queries and developed custom financial metrics.
- Produced clear, compelling visualizations for effective business storytelling.
- Built a fully reproducible analysis pipeline—from raw data ingestion to final insights.

Delivered Actionable Business Intelligence

- Generated cost-optimization strategies for squad building and wage management.
- Produced evidence-based transfer market recommendations grounded in empirical data.
- Created a methodological blueprint that can be applied to other sports or financial asset classes.

Professional Development

This project demonstrates strong capabilities in:

- **Financial Analysis:** Valuation modeling, efficiency ratios, gap analysis.
- **Data Engineering:** Database design, ETL processes, and data quality assurance.
- **Business Acumen:** Converting analytical outputs into strategic, real-world decisions.
- **Problem-Solving:** Identifying and resolving complex data inconsistencies and structural challenges.