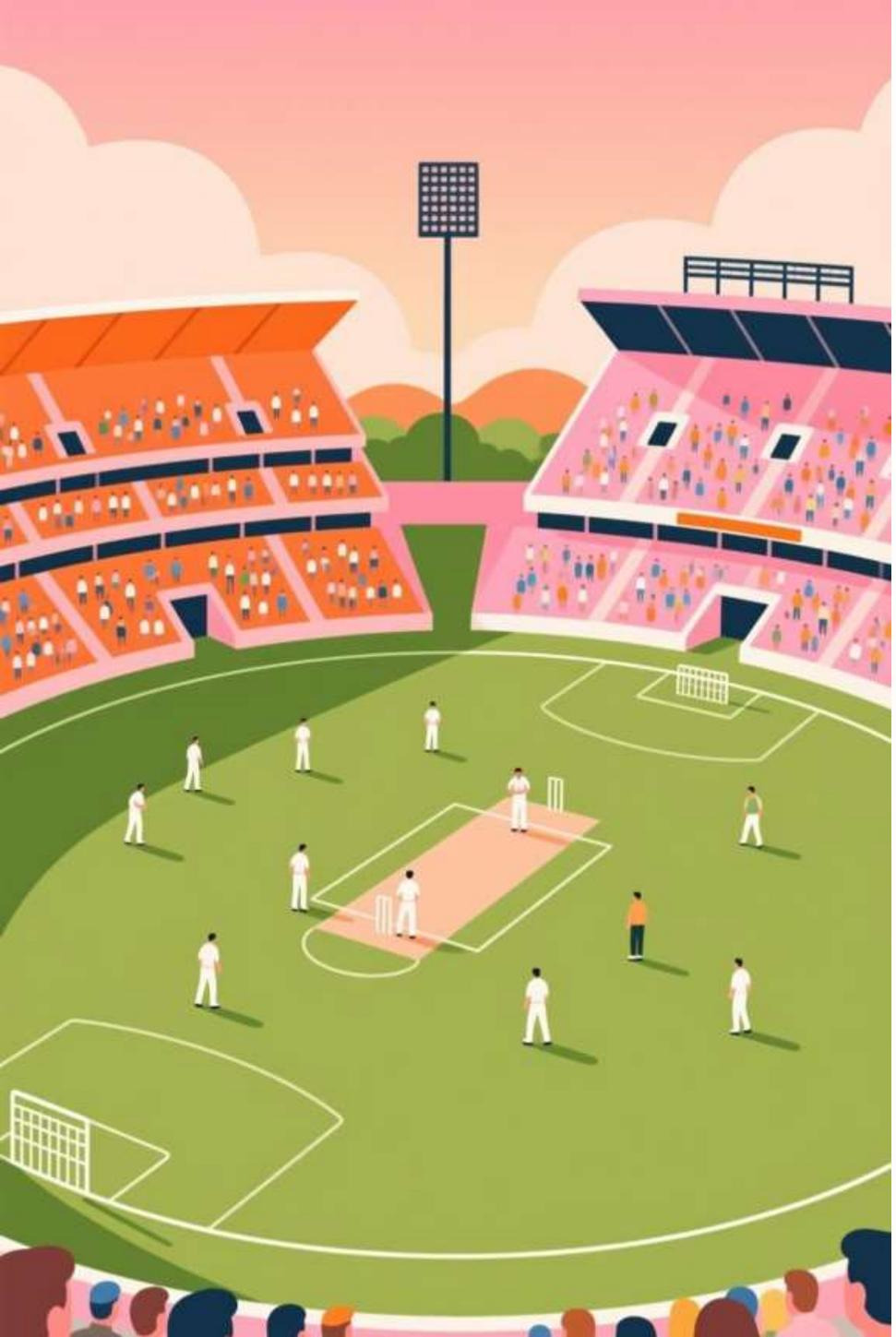


Cricket Best 11: Data-Driven Team Selection

Leveraging Python analytics and Power BI visualization to identify the optimal T20 cricket team based on comprehensive performance metrics. This Real world project demonstrates an end-to-end data pipeline that transforms raw match statistics into actionable insights for evidence-based player selection.

Technology Stack & Implementation

Python & Pandas

Core data processing engine for cleaning, transformation, and feature engineering of player statistics

JSON Library

Parsing and structuring nested match data from complex tournament hierarchies

Power BI

Interactive visualization platform enabling dynamic player comparison and team selection

Jupyter Notebook

Exploratory analysis environment for iterative hypothesis testing and metric validation

This technology combination enables seamless data flow from raw sources through analytical processing to executive-ready dashboards. The Python-to-Power BI pipeline ensures reproducibility and scalability for future tournament analyses.

Dataset: International T20 Match Statistics

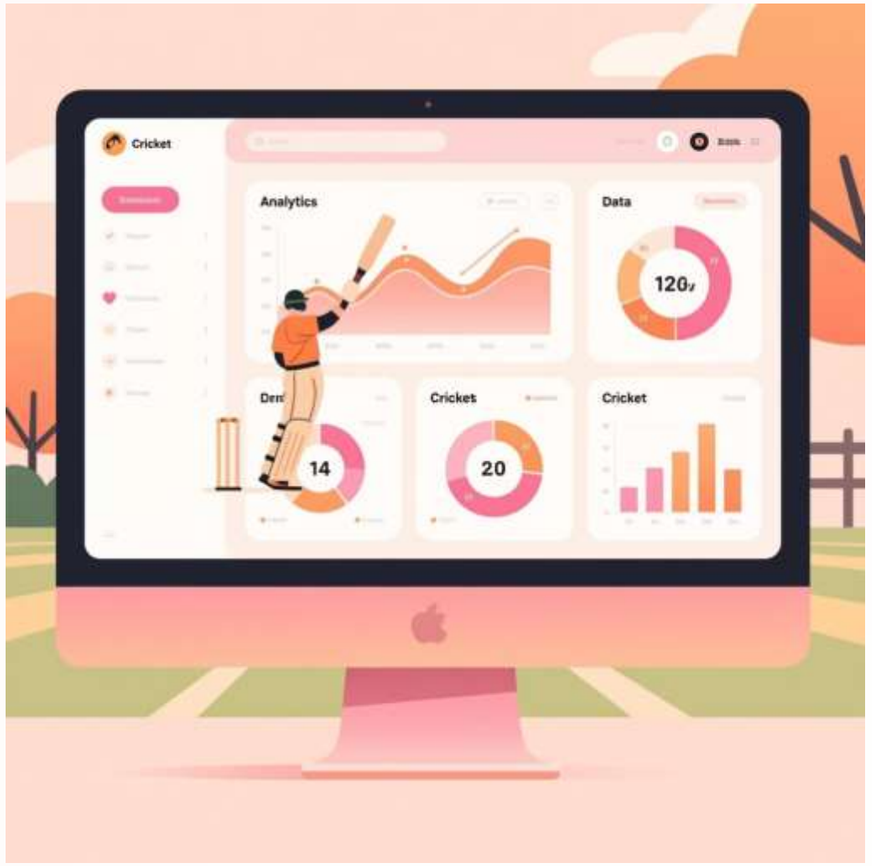
Data Source & Scope

*Player-level performance data was sourced from **Bright Data**, encompassing comprehensive statistics from international T20 cricket matches. The dataset captures granular details across batting, bowling, and all-round performance dimensions.*

This rich dataset includes match-by-match player contributions, enabling calculation of advanced metrics beyond simple averages. The data structure supports multi-dimensional analysis across player roles, match contexts, and tournament phases.

Key data categories include:

- *Batting statistics: runs, balls faced, boundaries, strike rates*
- *Bowling figures: wickets, economy rates, maiden overs*
- *Fielding contributions: catches, run-outs, dismissals*
- *Contextual metadata: match venues, opposition strength, match outcomes*



Data Preprocessing Pipeline

*All preprocessing operations were executed in **t20_data_preprocessing.ipynb**, establishing a reproducible analytical foundation. The pipeline transforms nested JSON structures into analysis-ready tabular formats optimized for Power BI consumption.*

Data Ingestion

Loaded JSON files and flattened nested structures using Pandas

Data Cleaning

Standardized formats, removed duplicates, handled missing values

Feature Engineering

Calculated advanced performance metrics and composite indices

Export for BI

Structured final dataset for Power BI integration

Calculated Performance Metrics

Batting Indicators

- ***Average:*** *Runs per dismissal*
- ***Strike Rate:*** *Run-scoring velocity*
- ***Boundary %:*** *Percentage of runs from 4s and 6s*
- ***Consistency Index:*** *Standard deviation of match scores*

Bowling Indicators

- ***Economy Rate:*** *Runs conceded per over*
- ***Bowling Strike Rate:*** *Balls per wicket*
- ***Wicket Frequency:*** *Probability of dismissal per delivery*

Power BI Dashboard: Interactive Team Selection

*The **Cricket-Best-11.pbix** dashboard transforms processed data into an interactive selection tool that enables dynamic comparison and evidence-based decision-making for team composition.*

1. ## Player Performance Comparison

   Side-by-side visual analysis across multiple dimensions with drill-down capabilities

2. ## Role-Based KPIs

   Customized performance indicators for batsmen, bowlers, all-rounders, and wicket keepers

3. ## Dynamic Filtering System

   Interactive slicers enabling analysis by player role, match type, and time period

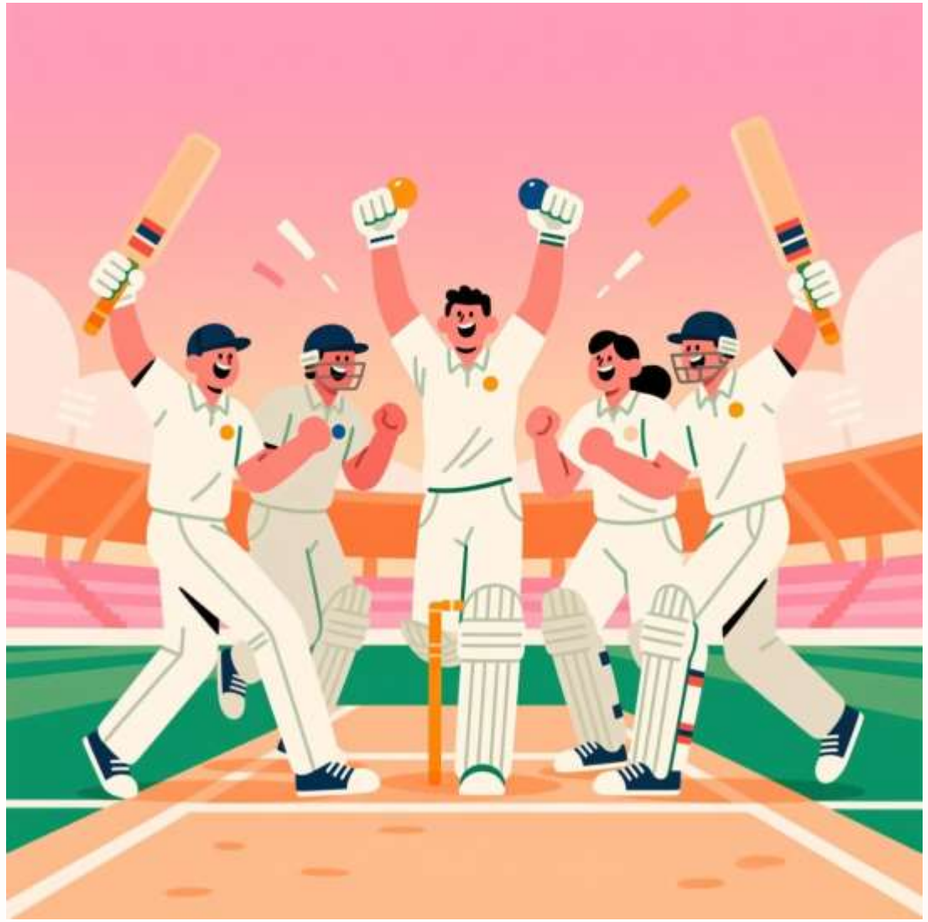
4. ## Automated Ranking Logic

   Algorithm-driven selection of top 11 players based on weighted performance criteria

5. ## Rich Visual Suite

   Bar charts, KPI cards, scatter plots, and custom visuals for comprehensive insights

Key Insights & Findings



Data-Driven Player Selection

The analytical framework identified high-impact performers who consistently deliver across multiple performance dimensions. By moving beyond simple averages, the model captured player value in high-pressure situations and against quality opposition.

Consistency Matters

Identified players with low performance variance—reliable contributors who minimize selection risk across varying match conditions

Balanced Best 11

Selected optimal team composition balancing batting depth, bowling variety, and fielding capability using multi-criteria optimization

Impact Players

Evidence-Based Selection

Project Outcome & Impact

End-to-End Analytical Workflow

This project successfully demonstrates a complete data analytics pipeline—from raw JSON ingestion through Python-based preprocessing to interactive Power BI visualization. The workflow showcases the synergy between Python's data manipulation capabilities and Power BI's enterprise visualization power.

Raw Data Acquisition

Structured tournament data from Bright Data

Python Processing

Cleaned, transformed, and engineered features

Power BI Dashboard

Interactive visualization for team selection

Optimal Best 11

Data-driven team composition ready for deployment

The deliverable provides a reusable framework applicable to future tournaments, different cricket formats, or alternative sports analytics scenarios. This methodology bridges the gap between raw sports data and actionable coaching insights.