# Sports Data Warehouse & Predictive Modeling Roadmap

**Target sports: Football (Soccer) and Cricket. Then extend to other sport later.**

Purpose: Create a scalable data platform and ML lifecycle to ingest historical + real-time data, store odds and statistical information, and build models to predict odds, match outcomes, and player/team performance.

## Executive Summary

This document lays out a phased roadmap, detailed architecture design, data model, ETL approach, ML lifecycle, deployment strategy, and governance necessary to build a sports data warehouse and predictive modeling system intended for commercial use (e.g., for a buyer like ESPN). It focuses initially on Football and Cricket with markets including match winner, point spreads, and over/under.

## Objectives & Key Results (OKRs)

* Ingest and normalize historical and real-time match & odds data for Football & Cricket.
* Provide clean, timestamped odds history across bookmakers for value-detection and odds forecasting.
* Deliver predictive models that estimate match outcomes, fair odds, and player/team performance metrics.
* Design production-ready APIs and dashboards for analytics and external customers.
* Achieve data quality SLAs: completeness > 99%, latency for live odds < 5s (target), documentation coverage 100%.

## Scope & Deliverables

* Canonical data model (Sport-agnostic, extensible).
* ETL pipelines for historical backfill and live ingestion.
* Time-series storage for odds and event-streams.
* Data transformations and analytical views (dbt-like).
* Set of baseline ML models (classification and time-series forecasting).
* Backtesting and model evaluation framework.
* Deployment pipelines, monitoring, and alerts.
* Technical documentation and a concise executive summary for sales.

## High-Level Architecture

Core components and responsibilities:

* Data Sources: public & free APIs, scraped sources where legal, historical datasets (Kaggle, open repos).
* Ingestion Layer: API connectors, web scrapers (where permitted), streaming adapters for live odds.
* Raw Data Lake: S3 (or equivalent) for raw JSON/CSV + immutable storage.
* Processing & ETL: orchestrated jobs (Airflow/Prefect) + transformation engine (dbt or SQL).
* Primary Datastore: PostgreSQL with TimescaleDB (or dedicated Time-Series DB) for odds & event time-series.
* Analytics Warehouse: star-schema views for fast BI queries (can be same Postgres or Redshift/Snowflake later).
* Feature Store: online & offline features for model training and low-latency serving.
* Model Training & Experimentation: notebooks → reproducible pipelines (MLflow, DAGs).
* Model Serving: REST/gRPC endpoints + batch inference pipelines.
* Monitoring & Observability: Prometheus/Grafana + data quality checks, model drift detectors.
* Access Layer: internal API (FastAPI), BI dashboards (Metabase/Looker) and export endpoints for customers.

## Data Model (Canonical Entities)

Key tables and time-series collections:

* sports(sport\_id, name)
* leagues(league\_id, sport\_id, name, country)
* teams(team\_id, league\_id, name, canonical\_external\_ids)
* players(player\_id, team\_id, name, role, canonical\_external\_ids)
* matches(match\_id, league\_id, home\_team\_id, away\_team\_id, start\_time, venue, status)
* match\_events(event\_id, match\_id, timestamp, event\_type, player\_id, metadata) -- minute/ball-by-ball
* odds(odds\_id, match\_id, bookmaker, market\_type, selection, odds\_value, timestamp, source\_id)
* statistics(stat\_id, match\_id, team\_id, player\_id, stat\_type, value)
* results(result\_id, match\_id, home\_score, away\_score, winner)

## Ingestion & ETL Design

Principles: idempotency, immutability for raw ingestion, schema evolution support, observability, and lineage.

* Historical Backfill: batch jobs that pull bulk files/APIs and write to raw lake (S3).
* Real-time Ingestion: lightweight connectors that poll APIs or subscribe to streaming feeds, publish to Kafka or Kinesis.
* Transformation: incremental transforms (dbt or SQL) that materialize cleaned tables and analytics views.
* Timestamps: every record stores source\_timestamp and ingestion\_timestamp.
* Change Data Capture: for external relational sources, adopt CDC connectors to capture updates.

## Time-series & Odds Handling

Odds require special treatment because they change rapidly and historical tracking is essential.

* Use TimescaleDB extension or a dedicated TSDB to store odds with efficient retention & compression.
* Partition by sport/league/date to optimize queries.
* Normalize odds into a canonical decimal format and store implied probability.
* Compute bookmaker margin (overround) per snapshot and store it.
* Keep full snapshot history; do not overwrite raw odds.

## Feature Engineering & Feature Store

Features should be produced offline for training and online for serving:

* Aggregate temporal features: rolling means, form metrics (last N matches), head-to-head aggregates.
* Market features: current odds, odds movement delta over specific windows, average bookmaker probability, liquidity proxies.
* Contextual features: venue, weather (optional), rest days, travel distance, injuries.
* Player-level features for cricket: recent strike rate, average, wicket rate. For football: xG, goals/90, assists/90.
* Use Feast or a homegrown store to keep online feature retrieval latencies low (<50ms target).

## Modeling Strategy

Models to build (baseline → advanced):

* Baseline Outcome Model: Logistic Regression / XGBoost for match winner / draw classification.
* Odds Calibration Model: regression to predict 'fair' odds from features (detect mispriced markets).
* Odds Movement Forecast: time-series models (Prophet, LSTM, or TFT) to predict odds trajectory pre-match.
* Player Performance Models: regression/classification for runs/goals, wickets, assists.
* Ensemble & Stacking: combine market-implied probabilities with model outputs for final forecasts.

## Training, Validation & Backtesting

Robust evaluation is crucial for credibility when selling to a partner like ESPN:

* Train/Validation/Test split using time-aware splits (no leakage): rolling window CV.
* Backtesting engine to simulate trading/betting strategies: slippage, stake sizing, bookmaker limits.
* Metrics: AUC/Accuracy for classification, Brier score for probabilistic forecasts, RMSE for regression, ROI for trading sims.
* Calibration checks: reliability diagrams, isotonic calibration when needed.

## MLOps & Deployment

Productionization checklist: reproducible training, automated CI/CD, model governance.

* Experiment Tracking: MLflow or Weights & Biases.
* CI/CD: containerized model pipelines, automated tests, canary rollout for models.
* Model Serving: FastAPI endpoints or KFServing for scalable inference.
* Batch Scoring: nightly re-score for upcoming matches and refresh feature snapshots.
* Monitoring: data quality checks, model performance (drift), latency/throughput monitoring.

## APIs, Dashboards & Exports

* Internal APIs: match/odds/stats endpoints for downstream consumers.
* Customer APIs: authenticated endpoints for external buyers (rate-limited, SLA-backed).
* Dashboard: interactive dashboards for exploration (Metabase / Superset).
* Export formats: JSON, CSV, Parquet; scheduled data dumps for partners.
* Documentation: OpenAPI spec, data dictionary, SLAs and terms of use.

## Security, Compliance & Legal

* Ensure compliance with Terms of Service for scraped sources; prefer licensed feeds for commercial use.
* Data privacy: limit PII, encrypt data at rest and in transit.
* Access controls: RBAC, audit logs, API keys for external access.
* IP & Licensing: retain records of source licenses and attributions; consider legal counsel before monetization.

## Scalability & Cost Considerations

Initial (MVP) vs. Enterprise considerations:

* Start with single-region cloud infra, use managed DB (RDS/Postgres) + S3.
* Scale to multi-region and add read-replicas and query warehouses (Redshift/Snowflake) as needed.
* Consider caching for frequently requested endpoints (Redis).
* Cost drivers: API usage, storage for full odds history, compute for model training. Budget planning recommended.

## Phased Roadmap & Timeline (Estimate)

Assumes a small cross-functional team (2 data engineers, 1 backend, 1 ML engineer, 1 product/PM).

* Phase 0 — Planning: finalize scope, select initial APIs, legal check, infra choices.
* Phase 1 — MVP Data Platform raw ingestion, schema, historical backfill for Football & Cricket, basic ETL.
* Phase 2 — Analytics & Basic Models analytics views, simple baseline models, dashboards.
* Phase 3 — Real-time & Feature Store live odds ingestion, feature store, online serving.
* Phase 4 — Advanced Models & Productionization advanced models, backtesting, CI/CD, monitoring.
* Phase 5 —API SLA setup, documentation, legal sign-off, pilot with customer.

## Risks & Mitigations

* Data Licensing Risk: Mitigation — prioritize free/open sources for development, obtain licensed feeds for production.
* Data Quality & Completeness: Mitigation — implement validation rules and alerts, fallback sources.
* Model Overfitting / Non-stationarity: Mitigation — frequent re-training, drift detection, conservative backtesting.
* Legal/Regulatory: Mitigation — legal review, limit distribution in regulated jurisdictions.

## Next Steps & Immediate Recommendations

* Run a 2-week spike to ingest a 2-year historical dataset for Football and Cricket and validate schema.
* Create canonical ID mapping for teams & players (avoid duplicates across sources).
* Build baseline models and backtesting framework in parallel with the MVP ETL.
* Prepare an executive one-pager and tech deck for potential buyers (ESPN).

## One-Page Summary (Executive)

Summary: Build a reliable, auditable sports data warehouse that captures full odds history and rich statistical features for Football and Cricket. The product targets both analytics and predictive modeling to forecast outcomes and fair odds. Start with a focused MVP that proves data quality and model signal, then expand into real-time capabilities and commercial APIs for customers like ESPN.

Here are several good downloadable datasets for **football** + **cricket** historical data — with direct links.

## 🏏 Cricket Datasets

| **Name** | **Description** | **What data / fields included** | **Link & Notes** |
| --- | --- | --- | --- |
| **Cricsheet** | Full ball-by-ball data for many major cricket competitions (internationals + domestic T20, ODI, Test). ([Cricsheet](https://cricsheet.org/?utm_source=chatgpt.com" \o "Cricsheet)) | JSON/YAML (and experimental CSV/XML). Match metadata, players, deliveries, bowler/pitcher, runs, wickets, etc. All matches broken down by types (Test, ODI, T20) and by competition. ([Cricsheet](https://cricsheet.org/matches/?utm_source=chatgpt.com" \o "Match data)) | [Download page for Cricsheet match data](https://cricsheet.org/downloads/) ([Cricsheet](https://cricsheet.org/downloads/?utm_source=chatgpt.com" \o "Available match data downloads)) |
| **cricketdata R package** | Tools to fetch & load cricket data from Cricsheet + ESPN-Cricinfo in R. Very useful for quick prototyping. ([pkg.robjhyndman.com](https://pkg.robjhyndman.com/cricketdata/?utm_source=chatgpt.com)) | Match data, player innings, player metadata, etc., in tidy “tibbles” (R data frames) ready for analysis. ([pkg.robjhyndman.com](https://pkg.robjhyndman.com/cricketdata/?utm_source=chatgpt.com)) | [cricketdata on CRAN](https://pkg.robjhyndman.com/cricketdata/) ([pkg.robjhyndman.com](https://pkg.robjhyndman.com/cricketdata/?utm_source=chatgpt.com)) |

## ⚽ Football (Soccer) Datasets

| **Name** | **Description** | **What data / fields included** | **Link & Notes** |
| --- | --- | --- | --- |
| **International Football Results (1872-to-present)** | Very long history of international matches from many countries. ([Kaggle](https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017?utm_source=chatgpt.com)) | Match date, teams, result (win/draw/loss), possibly score; useful for historical modeling. No odds included. ([Kaggle](https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017?utm_source=chatgpt.com)) | [Kaggle dataset: International Football Results](https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017) ([Kaggle](https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017?utm_source=chatgpt.com)) |
| **European Soccer Database** | Matches from many European leagues + player info etc. ([Kaggle](https://www.kaggle.com/datasets/hugomathien/soccer?utm_source=chatgpt.com)) | Match data (date, home/away, goals), player rosters, player info. Useful for team/player stats. May need cleaning. ([Kaggle](https://www.kaggle.com/datasets/hugomathien/soccer?utm_source=chatgpt.com)) | [Kaggle: European Soccer Database](https://www.kaggle.com/datasets/hugomathien/soccer) ([Kaggle](https://www.kaggle.com/datasets/hugomathien/soccer?utm_source=chatgpt.com)) |
| **Club Football Matches 2000-2025 (42 leagues, 27 countries)** | More recent data for club matches across many countries. ([Kaggle](https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025?utm_source=chatgpt.com)) | Match date, home/away teams, score, league, country. Probably fewer advanced stats, no odds by default. ([Kaggle](https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025?utm_source=chatgpt.com)) | [Kaggle: Club Football Match Data 2000-2025](https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025) ([Kaggle](https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025?utm_source=chatgpt.com)) |
| **Worldwide Football Results (1872-2022)** | Very large dataset for international matches (FIFA + others). ([Kaggle](https://www.kaggle.com/datasets/zeesolver/fifa-results?utm_source=chatgpt.com)) | Matches, date, home & away, results. Similar to “International Football Results” above. Useful for long-term trends. ([Kaggle](https://www.kaggle.com/datasets/zeesolver/fifa-results?utm_source=chatgpt.com)) | [Kaggle: Worldwide Football Results](https://www.kaggle.com/datasets/zeesolver/fifa-results) ([Kaggle](https://www.kaggle.com/datasets/zeesolver/fifa-results?utm_source=chatgpt.com)) |

## ⚠️ Considerations & Gaps

* **Odds data is usually NOT included** in these historical datasets. If we need historical odds, likely need sportsbook API archives or specialized paid datasets.
* **Advanced stats** like expected goals (xG), possession, shot location, etc., are less commonly included in general datasets and may require either scraping specialized websites or subscribing to paid services.
* **Data cleaning will be needed**: team / player names may differ across sources; missing values; different date formats.
* **Licensing / usage terms**: especially for soccer data from Kaggle or other sources, check whether commercial use is permitted (if we want to sell/sub-license). Same for cricket data.

# 📜 ****Odds-Based Dataset & API List (Download/Access Checklist)****

# 📊 Comprehensive Sports Betting Odds Datasets & Sources

## 🟢 ****FREE DATASETS - Direct Download****

### ****⚽ Football/Soccer****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **Beat-The-Bookie (Kaggle)** | CSV | 500k+ matches, 11 years, 1,005 leagues, 32 bookmakers | Opening, Closing, Average | Match results, team stats | [Kaggle Link](https://www.kaggle.com/datasets/austro/beat-the-bookie-worldwide-football-dataset) | ✅ Active |
| **Club Football Match Data (2000-2025)** | CSV | 27 leagues worldwide, 25 years | Opening, Peak, Closing from multiple bookies | Goals, shots, possession, cards, corners | [Kaggle Link](https://www.kaggle.com/datasets/adamgbor/club-football-match-data-2000-2025) | ✅ Active (June 2025) |
| **Historical Football Results/Betting Odds** | CSV | 31 seasons results, 24 seasons betting data | Match odds, Over/Under, Asian Handicap | Match stats, league tables | [Kaggle Link](https://www.kaggle.com/datasets/mexwell/historical-football-resultsbetting-odds-data) | ✅ Active |
| **Soccer Bookmaker Odds (GitHub)** | JSON/CSV | European leagues 2005-2019 | Average bookmaker odds | Team scores, match results | [GitHub Link](https://github.com/iredchuk/soccer-bookmaker-odds) | ✅ Active |
| **FootballData Collection (GitHub)** | CSV/JSON | Multiple leagues, historical | Various bookmaker odds | Results, statistics | [GitHub Link](https://github.com/jokecamp/FootballData) | ✅ Active |
| **Football-Data.co.uk** | CSV | 16+ worldwide leagues, 20+ years | Home/Draw/Away, O/U, Asian Handicap | Half-time results, shots, corners, cards | [Direct Link](https://www.football-data.co.uk/data.php) | ✅ Active |
| **Football-Bet-Data.com** | CSV/API | 65+ leagues including Premier League | Multiple bookmaker odds | Advanced statistics, xG | [Website Link](https://www.football-bet-data.com/) | ✅ Free tier |

### ****🏀 Basketball (NBA)****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **NBA Historical Stats and Betting Data** | CSV | Historical NBA games | Match odds, totals, spreads | Box scores, player stats, advanced metrics | [Kaggle Link](https://www.kaggle.com/datasets/ehallmar/nba-historical-stats-and-betting-data) | ✅ Active |
| **BigDataBall NBA Dataset** | Excel/CSV | Current & historical seasons | Moneyline, spreads, totals | Box scores, player stats, DFS data | [BigDataBall Link](https://www.bigdataball.com/datasets/) | ✅ Paid/Free tier |

### ****🏈 American Football (NFL)****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **NFL Scores and Betting Data** | CSV | Historical NFL games | Point spreads, totals, moneylines | Game scores, weather, attendance | [Kaggle Link](https://www.kaggle.com/datasets/tobycrabtree/nfl-scores-and-betting-data) | ✅ Active (Updated 4 days ago) |
| **BigDataBall NFL Dataset** | Excel/CSV | Current & historical seasons | All major bet types | Advanced stats, play-by-play | [BigDataBall Link](https://www.bigdataball.com/datasets/) | ✅ Paid/Free tier |

### ****🎾 Tennis****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **BigDataBall Tennis Dataset** | Excel/CSV | ATP & WTA tournaments | Match winner odds | Match stats, point-by-point data | [Tennis Data Link](https://www.bigdataball.com/datasets/tennis-data/) | ✅ Active |
| **Tennis Betting Odds (OSU)** | CSV | Men's & women's tennis | Pre-match betting odds | Match results | [OSU Sports Data](https://sportsandsociety.osu.edu/sports-data-sets) | ✅ Active |

### ****🥊 Mixed Martial Arts (UFC/MMA)****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **UFC-Fight Historical Data (1993-2021)** | CSV | Comprehensive UFC fights | Basic betting odds | Fighter stats, fight details | [Kaggle Link](https://www.kaggle.com/datasets/rajeevw/ufcdata) | ✅ Active |
| **Ultimate UFC Dataset** | CSV | Merged UFC datasets | Various odds | Fighter details, fight outcomes | [Kaggle Link](https://www.kaggle.com/datasets/mdabbert/ultimate-ufc-dataset) | ✅ Active (Dec 2024) |
| **UFC Complete Dataset (1996-2024)** | CSV | All UFC events | Betting odds included | Complete fight stats | [Kaggle Link](https://www.kaggle.com/datasets/maksbasher/ufc-complete-dataset-all-events-1996-2024) | ✅ Active |
| **UFC Data with Betting Odds (GitHub)** | CSV | 30 years match history + 9 years odds | Historical betting odds | Fighter statistics | [GitHub Link](https://github.com/jansen88/ufc-match-predictor) | ✅ Active |
| **UFC Stats Scraper (GitHub)** | CSV | Complete UFC database | Not included | Fight details, fighter stats | [GitHub Link](https://github.com/Greco1899/scrape_ufc_stats) | ✅ Active |

### ****🏒 Hockey (NHL)****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **BigDataBall NHL Dataset** | Excel/CSV | Current & historical seasons | Moneyline, puck line, totals | Game stats, player performance | [BigDataBall Link](https://www.bigdataball.com/datasets/) | ✅ Paid/Free tier |

### ****⚾ Baseball (MLB)****

| **Dataset Name** | **Format** | **Coverage** | **Odds Types** | **Other Statistics** | **Download Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **BigDataBall MLB Dataset** | Excel/CSV | Current & historical seasons | Moneyline, run line, totals | Box scores, advanced sabermetrics | [BigDataBall Link](https://www.bigdataball.com/datasets/) | ✅ Paid/Free tier |

## 🟡 ****FREEMIUM/API SERVICES**** (Free Tier Available)

| **Service Name** | **Format** | **Coverage** | **Odds Types** | **Free Tier Limits** | **Access Link** | **Status** |
| --- | --- | --- | --- | --- | --- | --- |
| **The Odds API** | JSON | 40+ sports, multiple bookmakers | All major markets | 500 requests/month | [API Link](https://the-odds-api.com/) | ✅ Active |
| **OddAlerts Downloads** | CSV | Football (6 months history) | Opening, Closing, Peak odds | Free registration required | [OddAlerts Link](https://oddalerts.com/downloads) | ✅ Active |
| **API-Football/API-Sports** | JSON | Football + other sports | Pre-match odds from multiple bookies | 100 requests/day | [API-Sports Link](https://www.api-sports.io/) | ✅ Active |
| **Sports Betting Predictive Analysis Dataset** | CSV | Multi-sport predictions | ML-ready odds data | Kaggle free tier | [Kaggle Link](https://www.kaggle.com/datasets/pratyushpuri/sports-betting-predictive-analysis-dataset) | ✅ Active (July 2025) |

## 🔴 ****PAID/COMMERCIAL DATASETS****

| **Service Name** | **Format** | **Coverage** | **Odds Types** | **Pricing Model** | **Access Link** | **Notes** |
| --- | --- | --- | --- | --- | --- | --- |
| **Betfair Historical Data** | CSV/TAR | Multi-sport exchange data | Tick-by-tick price & volume | Per market/competition | [Betfair Data](https://historicdata.betfair.com/) | Official exchange data |
| **OddsWarehouse** | CSV/DB | Multi-sport comprehensive | All major bookmaker odds | Seasonal/annual license | Contact for pricing | Bulk historical data |
| **SportsDataIO** | API/Export | Multi-sport | Pre-match & live odds | Monthly/annual plans | [SportsDataIO](https://sportsdata.io/) | Enterprise-grade |
| **Sports Insights Historical** | Various | Multi-sport | Line movements, steam moves | Subscription-based | [Sports Insights](https://www.sportsinsights.com/) | Professional betting data |
| **OddsMatrix** | API | Multi-sport including UFC/MMA | Real-time & historical | Enterprise pricing | [OddsMatrix](https://oddsmatrix.com/) | Developer-focused API |

## 🔍 ****BROWSER-ACCESSIBLE DATABASES**** (No Download)

| **Source Name** | **Coverage** | **Odds Types** | **Access Method** | **Link** | **Status** |
| --- | --- | --- | --- | --- | --- |
| **OddsPortal Results** | Multi-sport historical | Opening vs Closing odds | Web browsing/scraping | [OddsPortal](https://www.oddsportal.com/results/) | ✅ Active |
| **Odds Shark Database** | NBA, NFL, College sports | Various betting lines | Web interface | [Odds Shark](https://www.oddsshark.com/) | ✅ Active |
| **OddsBase.net** | 20+ years historical | Multiple bookmakers | Web access | [OddsBase](https://oddsbase.net/) | ✅ Active |

## 📝 ****DATASET FEATURE COMPARISON****

### ****Most Comprehensive Free Datasets:****

1. **Beat-The-Bookie (Kaggle)** - 500k+ matches, 11 years, 32 bookmakers
2. **Club Football Data 2000-2025** - 25 years, advanced statistics
3. **Football-Data.co.uk** - 20+ years, multiple leagues, reliable source

### ****Best for Machine Learning Research:****

1. **Sports Betting Predictive Analysis Dataset** - ML-ready format
2. **Beat-The-Bookie** - Large sample size, multiple features
3. **BigDataBall datasets** - Clean, structured data

### ****Best Multi-Sport Coverage:****

1. **The Odds API** - 40+ sports
2. **BigDataBall** - NBA, NFL, MLB, NHL, Tennis
3. **OddsPortal** - Comprehensive but no direct download

### ****Most Recent Data:****

1. **Club Football Match Data** - Updated June 2025
2. **NFL Scores and Betting Data** - Updated weekly
3. **The Odds API** - Real-time updates

## 🎯 ****Recommended Starting Points for Research:****

**For Football/Soccer Research:**

* Start with Beat-The-Bookie for historical analysis
* Use Club Football Data 2000-2025 for recent comprehensive data
* Supplement with Football-Data.co.uk for specific leagues

**For Multi-Sport Analysis:**

* BigDataBall datasets for clean, structured data
* The Odds API for current data collection
* Kaggle datasets for historical analysis

**For MMA/UFC Research:**

* Ultimate UFC Dataset (most recent compilation)
* UFC Complete Dataset 1996-2024
* GitHub UFC repositories for additional features