**PROJECT CHARTER: PREDICTING THE IDEAL STRIKER FOR ARSENAL**

**1. Problem Statement**

Arsenal is in the market for a new striker who not only delivers consistent goal-scoring performances but also fits within the club’s overall playing philosophy and team structure. The challenge is to identify a striker who:

* Exhibits a playing style that complements Arsenal’s tactical setup.
* Demonstrates consistent performance over the last five years (goals, assists, hold-up play, link-up ability, and decisive contributions in key moments).
* Maintains high availability (minimal injuries or suspensions) and shows strong conversion rates.
* Is realistically attainable during the summer 2025 transfer window given market availability and cost constraints.

**2. Scope**

**In-Scope**

* **Player Performance Metrics:** Historical data on goals, assists, hold-up/link-up play metrics, conversion rates, and clutch performances.
* **Player Attributes:** Age, behavior/character assessments, injury records, availability percentage per season.
* **Market Analysis:** Transfer availability, cost analysis (market value, transfer fees, wage demands), and league difficulty ratings.
* **Data Integration:** Aggregating data from multiple sources to ensure a holistic view of each candidate’s performance and market conditions.
* **Predictive Analytics:** Statistical and machine learning models to forecast performance, consistency, and fit with Arsenal’s tactical style.
* **Visualization:** Dashboard development in PowerBI to visualize comparative metrics, player performance trends, and cost/benefit analyses.

**Out-of-Scope**

* Detailed tactical analysis of match-day strategies (beyond linking striker performance to existing team structure).
* Real-time transfer negotiations or proprietary market intelligence that isn’t publicly available.

**3. Data Collection Strategy**

**3.1 Types of Data Needed**

* **Performance Data:**
  + Goals, assists, shots on target, conversion rates.
  + Advanced metrics: expected goals (xG), expected assists (xA), and pressing statistics.
* **Player Attributes:**
  + Age, injury history, minutes played, suspension records.
  + Behavioral/character ratings (possibly from scouting reports or expert evaluations).
* **Market Data:**
  + Transfer market values, wage details, historical transfer fees.
  + League difficulty metrics, potentially quantified through average team performance metrics.
* **Availability Data:**
  + Percentage of games played (indicating health and reliability).
  + Injury and suspension frequency.

**3.2 Data Sources & Collection Methods**

* **Web Scraping & APIs:**
  + **Sports Data Providers:** Utilize APIs from sources such as Opta, StatsBomb, or WyScout for in-depth match and player statistics.
  + **Transfer Data:** Scrape data from reputable websites like Transfermarkt, Soccerway, or club official sites for market values and availability.
  + **Injury & Availability:** Leverage platforms like InStat or publicly available datasets from sports analytics communities.
* **Third-Party Datasets:**
  + Purchase or license historical performance datasets from sports analytics firms.
* **Manual Data Collection:**
  + Supplement gaps using expert scouting reports, interviews, or published reports on player behavior/character.

**3.3 Data Aggregation and Integration**

* **Cleaning & Normalization:**
  + Standardize metric units (e.g., per 90 minutes or per season averages).
  + Normalize data across different leagues by applying league difficulty adjustments or weighting factors.
* **Merging Data:**
  + Create a robust pipeline to merge datasets on a unique player identifier.
  + Alternatively, keep datasets separate but build relational links (e.g., SQL joins, ETL pipelines) to handle cross-references.

**4. Analysis Flow**

**4.1 Data Processing Pipeline**

* **Step 1:** Data Ingestion
  + Import datasets from multiple sources using APIs, web scraping scripts, or direct file uploads.
* **Step 2:** Data Cleaning & Transformation
  + Remove duplicates, handle missing values, and convert formats as needed.
  + Normalize statistics to a common scale (e.g., per season or per 90 minutes).
* **Step 3:** Data Integration
  + Merge multiple datasets into a unified data warehouse or use relational database management to maintain dataset integrity.
* **Step 4:** Feature Engineering
  + Create derived metrics such as availability percentage, performance consistency scores, and cost-effectiveness indices.
  + Engineer a “fit score” that combines performance, availability, and market affordability.

**4.2 Mode of Analysis**

* **Descriptive Analysis:**
  + Use statistical summaries and visualizations to understand the distribution of key performance metrics.
* **Predictive Analysis:**
  + Apply regression models and machine learning techniques to predict future performance based on historical trends.
  + Utilize classification models to determine whether a player’s profile matches Arsenal’s style.
* **Comparative Analysis:**
  + Benchmark players against each other using key performance indicators (KPIs) and league-adjusted statistics.
* **Cost-Benefit Analysis:**
  + Model the potential return on investment (ROI) based on performance metrics versus market cost.
* **Sensitivity Analysis:**
  + Test how changes in market availability, injury risk, or performance consistency affect the “ideal striker” predictions.

**5. Insights Extraction**

**Key Insights to Extract**

* **Player Performance Trends:**
  + Identify players with consistent high performance over five seasons.
* **Availability & Reliability:**
  + Evaluate players based on the percentage of minutes played and their injury/suspension records.
* **Cost Efficiency:**
  + Determine the most cost-effective player by balancing transfer fees, wages, and on-field contributions.
* **Style Fit Analysis:**
  + Quantify how closely a striker’s style of play (e.g., movement, link-up play) aligns with Arsenal’s tactical blueprint.
* **Market Feasibility:**
  + Assess which players are realistically obtainable within the transfer window constraints.

**Decision Criteria**

* **Performance Metrics Threshold:**
  + Strikers scoring 20–25 goals per season with supporting assist data.
* **Reliability Benchmark:**
  + Average availability > 90% per season.
* **Cost Analysis:**
  + Evaluation of market value against predicted contributions.
* **Behavior/Character Assessment:**
  + Qualitative insights combined with quantitative performance to ensure the right cultural and tactical fit.

**6. Data Visualization with PowerBI**

**Dashboard Components**

* **Overview Dashboard:**
  + Summary of key metrics for all shortlisted strikers.
  + Interactive filters for league, age, and performance indicators.
* **Performance Trend Analysis:**
  + Time series visualizations showing goals, assists, and conversion rates over the last 5 years.
* **Reliability & Availability Metrics:**
  + Visualizations depicting minutes played, injury records, and overall availability percentages.
* **Cost vs. Performance Analysis:**
  + Scatter plots and bubble charts to visualize ROI, balancing market cost against performance and reliability.
* **Style & Fit Analysis:**
  + Custom visuals (e.g., radar charts) comparing tactical attributes against Arsenal’s style benchmarks.

**Interactivity & Drill-Down Features**

* **Dynamic Filtering:**
  + Allow users to filter by specific criteria (e.g., league, age range, performance metrics).
* **Drill-Down Analysis:**
  + Enable detailed views on individual player profiles and historical performance trends.
* **Scenario Modeling:**
  + Use what-if parameters to simulate changes in market availability or performance to aid transfer decision-making.

**7. Project Timeline and Milestones**

**Phased Approach**

* **Phase 1: Project Initiation & Planning (1 month)**
  + Finalize project scope, define data sources, and set up infrastructure.
* **Phase 2: Data Collection & Integration (2 months)**
  + Develop data ingestion pipelines, perform initial cleaning, and integrate multiple datasets.
* **Phase 3: Analysis & Model Building (3 months)**
  + Conduct exploratory data analysis, feature engineering, and build predictive models.
* **Phase 4: Insights Extraction & Validation (1 month)**
  + Validate insights with domain experts and refine models.
* **Phase 5: Dashboard Development & Deployment (2 months)**
  + Develop PowerBI dashboards and integrate interactive elements.
* **Phase 6: Review & Final Reporting (1 month)**
  + Compile findings, prepare final reports, and present recommendations.

**8. Risk Management & Mitigation**

* **Data Quality Issues:**
  + Mitigation: Implement robust cleaning protocols and use multiple sources for cross-validation.
* **Model Uncertainty:**
  + Mitigation: Use ensemble methods and sensitivity analyses to gauge robustness.
* **Integration Challenges:**
  + Mitigation: Design modular pipelines that allow for incremental integration and testing.
* **Market Dynamics:**
  + Mitigation: Regularly update market data and include contingency plans for rapid market changes.

**9. Conclusion**

This project aims to deliver a data-driven recommendation on the most suitable striker for Arsenal ahead of the summer 2025 transfer window. By leveraging multiple data sources, rigorous analytical methods, and interactive PowerBI visualizations, the project will not only identify top candidates but also provide actionable insights regarding performance, cost, and tactical fit. This comprehensive approach ensures that all critical dimensions—from statistical performance and reliability to market feasibility—are fully addressed, paving the way for an informed transfer strategy.