**EXPERIMENT – 9**

**T20 Cricket Data Analytics Project**

**About The Project**

The primary objective of this project is to conduct a comprehensive analysis of T20 World Cup cricket data to identify and select the best 11 players. This initiative aims to demonstrate the power of data analytics in sports, particularly in making informed decisions regarding player selection based on a variety of performance metrics.

**Key Objectives:**

1. **Data Acquisition:** Efficiently scrape quality match, player, and scorecard data from ESPNcricinfo.
2. **Data Processing:** Clean, standardize, and transform data using Python/Pandas, creating key performance metrics.
3. **Performance Analysis:** Define and analyze KPIs (batting average, strike rate, economy) to evaluate player performance.
4. **Player Selection:** Develop an algorithm to rank and select the optimal 11, potentially simulating match scenarios.
5. **Data Visualization:** Create interactive Power BI dashboards to present player performance and match statistics.
6. **Engagement:** Involve the audience through challenges and gather feedback for model improvement.

**Expected Outcomes**

* **Informed Player Selection**: Achieve a well-reasoned selection of the best 11 players based on data-driven insights.
* **Enhanced Understanding of Data Analytics in Sports**: Showcase how data analytics can influence sports strategies and decisions.
* **Documentation and Knowledge Sharing**: Produce comprehensive documentation of methodologies and findings for future reference in sports analytics.

**Workflow for T20 Cricket Data Analytics Project**

 **Initiation:**

* Define goal: Select best 11 players.
* Identify stakeholders

 **Data Collection:**

* **Setup:** Create Bright Data collectors for match results, scorecards, and player info on ESPNcricinfo.
* **Extraction:** Run collectors and save data as JSON.

 **Data Processing:**

* **Load:** Use Python json to load JSON.
* **Transform:** Convert to Pandas DataFrames.
* **Clean:** Rename columns, create player status, remove special characters.
* **Link:** Create a dictionary linking match IDs and player data.
* **Export:** Save cleaned DataFrames as CSV for Power BI.

 **Power BI Analysis:**

* **Import:** Load CSV files.
* **Model:** Establish relationships between fact (matches, batting) and dimension (players) tables.
* **DAX Measures:** Develop measures for total runs, innings batted/dismissed.
* **Calculated Columns:** Calculate boundary runs.

 **Visualization:**

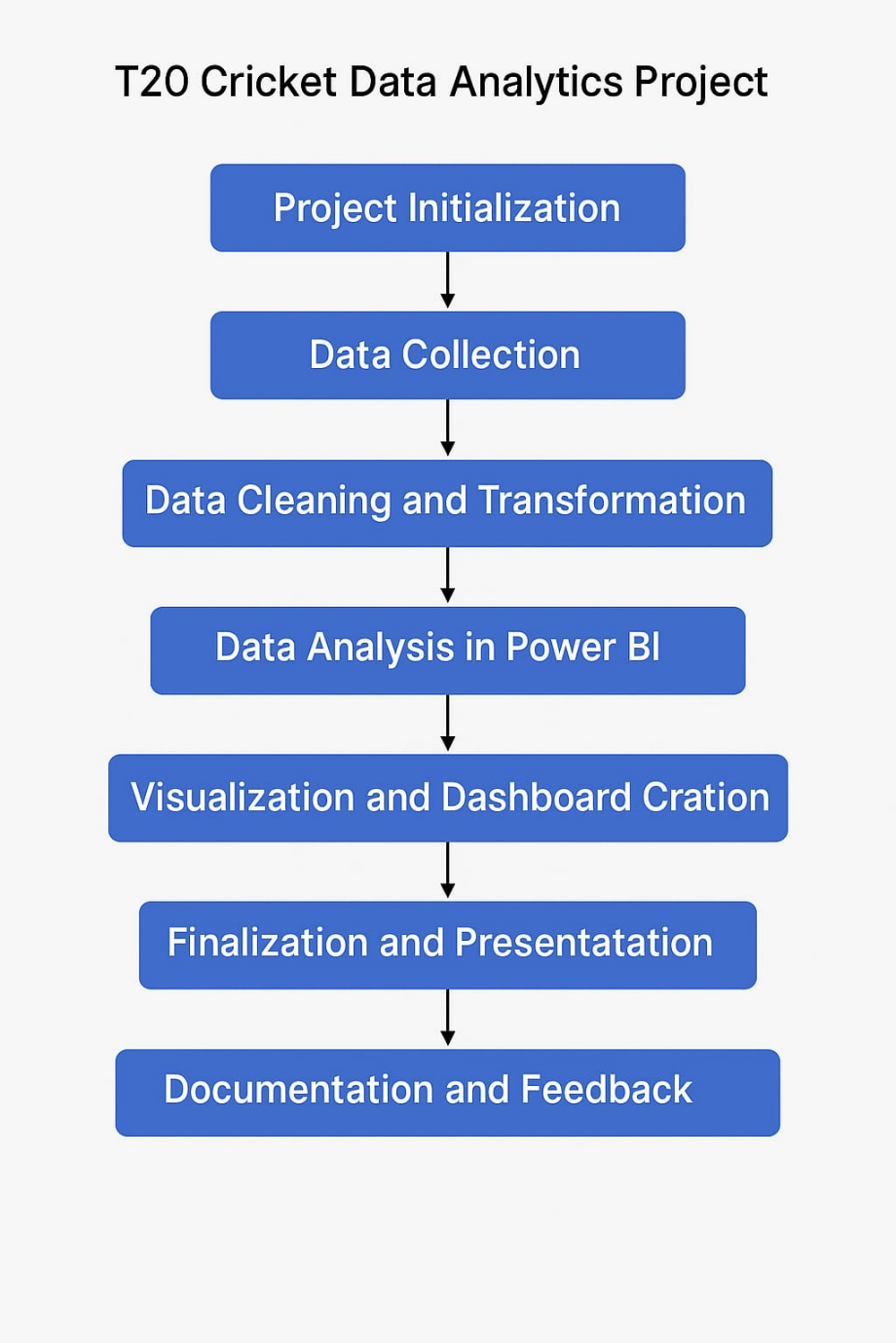
* Build charts/graphs for player performance and match stats.
* Review dashboard for insights and adjustments.

 **Finalization & Presentation:**

* Finalize DAX measures and calculated columns.
* Present findings and introduce an engaging challenge.

 **Documentation & Feedback:**

* Document workflow, methodologies, and findings.
* Collect feedback for future improvements.



**Work Flow Diagram**

**ABOUT THE DATA**

The dataset for this T20 World Cup Cricket Data Analytics project was meticulously gathered from **ESPNcricinfo**, a premier source for cricket statistics. The data includes rich and granular information across multiple dimensions of the game, enabling a robust performance evaluation and player selection process.

**Data Sources:**

* **Match Details**: Match IDs, dates, venues, teams, toss results, and outcomes.
* **Scorecards**: Innings-wise batting and bowling statistics, including runs, balls faced, boundaries, wickets, and economy rates.
* **Player Profiles**: Career statistics, player roles (batsman, bowler, all-rounder), batting and bowling styles, and team affiliations.

**Data Characteristics:**

* Format: Extracted as **JSON** using Bright Data collectors.
* Volume: Covers all matches from recent T20 World Cups.
* Structure: Segregated into distinct datasets such as **Match Info**, **Batting Stats**, **Bowling Stats**, and **Player Details**.

**Preprocessing Highlights:**

* Data converted into **Pandas Data Frames** for manipulation.
* Columns renamed and normalized for consistency.
* Special characters removed and missing values handled.
* Custom metrics such as **batting impact score**, **bowling efficiency**, and **fielding contribution index** were derived for comprehensive evaluation.

**Usage in Power BI:**

* Cleaned CSV files were imported to Power BI.
* Relationships were established to facilitate dynamic and interactive reporting.
* DAX was used to generate calculated insights such as:
  + Strike Rate = (Total Runs / Balls Faced) \* 100
  + Batting Average = Total Runs / Dismissals
  + Economy Rate = Runs Conceded / Overs Bowled

This structured, multidimensional data foundation enabled a data-driven and transparent approach to forming the best 11 T20 players, maximizing both statistical rigor and interpretability.