

# ***Tool Documentation***

## **Python**

**Purpose:** Programming language used for synthetic data generation.

**Justification:**

- Automates the creation of realistic fake data using libraries like Faker, random, and numpy.
- Handles data processing and organization through pandas.
- Enables export of clean datasets in .csv format for further analysis.

**Integration:**

- Generates simulated datasets (e.g., players, matches, weather conditions).
- Ensures logical relationships between entities.
- Exports structured data to CSV files for visualization in Power BI.

## **Python Libraries Used**

- **Faker:** Generates realistic fake data such as names, nationalities, and dates.
- **pandas:** Structures and manages data in table format, ready for CSV export.
- **numpy:** Creates statistical distributions and complex random values.
- **random:** Adds variability through basic random selections.
- **datetime:** Generates and handles time-based data like birthdates and match dates.

## **SQL**

**Purpose:** Language used to define, structure, and relate the tables in a relational database.

**Justification:** SQL is the standard for managing structured data. It is essential for designing databases, executing complex queries, ensuring referential integrity, and establishing relationships between domain entities.

**Integration:**

- Defines the database structure using CREATE TABLE statements.
- Establishes primary and foreign keys to ensure referential integrity.
- Models relationships between entities like players, coaches, teams, and matches.
- Lays the foundation for later analysis using SQL queries (SELECT, JOIN, etc.).

**Power BI**

**Purpose:** Data visualization and business intelligence tool.

**Justification:**

- Transforms raw CSV files into interactive dashboards and visual reports.
- Enables exploration of patterns, trends, and key performance metrics.

**Integration:**

- Imports CSV files generated by Python.
- Builds relationships between tables, replicating the relational structure modeled in SQL.
- Visualizes key metrics such as goals per player, team performance, stadium attendance, and weather impact on matches.
- Uses filters, slicers, and visual elements to enable deep and intuitive data analysis.