

FIFA World Cup Data Analysis

**High Level Design (HLD)
Syeda Viqar Sultana**

**Revision Number: 1.0
Last date of revision: 15/06/2024**

Document Version Control

| Date Issued | Version | Description | Author |
|-------------|---------|---|---------------------|
| 15/06/2024 | 1.0 | First Version of Complete High-Level Design | Syeda Viqar Sultana |

Contents

| | |
|---|----|
| Document Version | 2 |
| Abstract..... | 4 |
| Introduction | 5 |
| • Why this High-Level Design Document?..... | 5 |
| • 1.2 Scope | 6 |
| General Description | 6 |
| • 2.1 Product Perspective & Problem Statement | 6 |
| • 2.2 Tools used..... | 7 |
| Design Details..... | 8 |
| • 3.1 Functional Architecture | 8 |
| • 3.2 Optimization | 9 |
| KPIs..... | 10 |
| Deployment..... | 11 |

Abstract

This comprehensive analysis explores the rich history of the FIFA World Cup, the world's most prestigious football tournament, through the lens of Exploratory Data Analysis (EDA). The dataset spans from the inaugural tournament in 1930 to the most recent events, encompassing match results, team performances, player statistics, and tournament outcomes. The primary objectives of this analysis are to identify patterns and trends that have emerged over the decades, understand the factors contributing to the success of dominant teams, and analyse goal-scoring behaviors and defensive strategies. By employing a variety of EDA techniques, including data visualization, statistical summaries, and correlation analysis, we aim to provide a nuanced understanding of the evolution of the World Cup. Key findings from this analysis include the identification of historically dominant teams, the impact of geographical and climatic factors on match outcomes, and the evolution of playing styles and strategies. Additionally, the analysis highlights significant upsets and underdog victories, providing insights into the unpredictability and excitement that characterize the World Cup. Visualizations such as time series plots, heatmaps, and bar charts are utilized to illustrate trends and patterns, while statistical tests are employed to validate hypotheses and uncover underlying relationships within the data. This analysis not only celebrates the rich history of the FIFA World Cup but also offers valuable insights for teams, coaches, and analysts looking to understand the factors that contribute to success on the world stage. In conclusion, this EDA-driven exploration of FIFA World Cup data provides a detailed and insightful look into the tournament's history, offering a deeper appreciation of the beautiful game's global impact and the myriad factors that influence its outcomes.

1. Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like:
 - Security
 - Reliability
 - Maintainability
 - Portability
 - Reusability
 - Application compatibility
 - Resource utilization
 - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

1.3 Definitions

- ETL – Extract, Transform, Load.
- EDA – Exploratory Data Analysis.
- CSV file – Comma-separated values file, opened in MS Excel.
- Power Query – With Power Query, you can connect to many different data sources and transform the data into the shape you want.
- Power BI - An interactive data visualization software company focused on business intelligence.

2. General Description

2.1 Product Perspective & Problem Statement

With FIFA is in the blood of many people of the world. You are tasked to tell the story of unsung analysts who put great efforts to provide accurate data to answer every question of fans. The FIFA World Cup is a global football competition contested by the various football-playing nations of the world. It is contested every four years and is the most prestigious and important trophy in the sport of football. The main objective of the project is to analyse the various aspects with different use cases which covers many aspects of World Cup stats. It helps in not only understanding the meaningful relationships between attributes but also allows us to do our own research and come up with our findings. This project aims to provide visual understanding of the data using Microsoft Power Bi.

2.2 Tools used

Microsoft Excel

Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications. Excel forms part of the Microsoft Office suite of software. Microsoft Excel is used for loading the data in CSV format, basic data cleaning and filter operations to execute the program. MS Excel file was loaded into Power BI software.

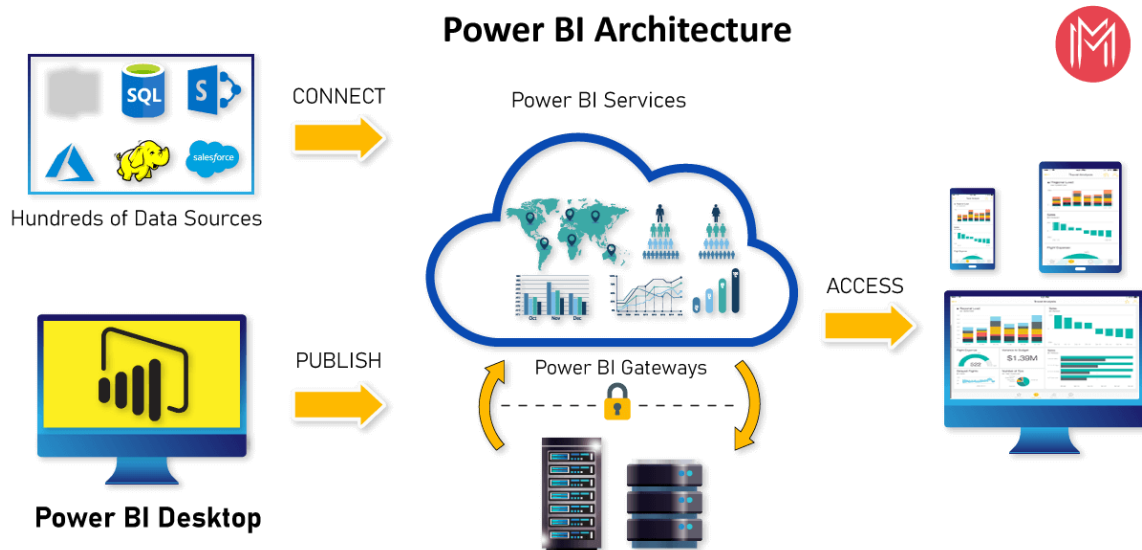
Microsoft Power BI

Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence. It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Data may be input by reading directly from a database, webpage, or structured files such as spreadsheets, CSV, XML, and JSON. With the help of Power BI, we have done the complete analysis and visualization.



3. Design Details

3.1 Functional Architecture



Power BI

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports. It provides interactive visualizations with self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on any information technology staff or database administrator.

Benefit of Business Intelligence

The benefits of Business intelligence are as follows:

- Business intelligence is faster more accurate process of reporting critical information.
- Business intelligence facilitates better and efficient decision-making process.
- Business intelligence provides timely information for better customer relationship management.
- Business intelligence improves profitability of the company.
- Business intelligence provides a facility of assessing organization's readiness in meeting new business challenges.
- Business intelligence supports usage of best practices and identifies every hidden cost

3.2 Optimization

Your data strategy drives performance

- Minimize the number of fields.
- Minimize the number of records.
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views.

Reduce the marks (data points) in our view.

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views. Limit your filters by number and type

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
- Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much

faster than an exclude filter, especially for dimensions with many members.

- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.
- Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.
- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database.
- Reduce the number of nested calculations.
- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
 - LODs - Look at the number of unique dimension members in the calculation.
- Table Calculations - the more marks in the view, the longer it will take to calculate. Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.

4.KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators about the stats of the tournament in various years.

4.1 KPIs (Key Performance Indicator)

Key findings from this analysis include the identification of historically dominant teams, the impact of geographical and climatic factors on match outcomes, and the evolution of playing styles and strategies. Additionally, the analysis highlights significant upsets and underdog victories, providing insights into the unpredictability and excitement that characterize the World Cup.

Visualizations such as time series plots, heatmaps, and bar charts are utilized to illustrate trends and patterns, while statistical tests are

employed to validate hypotheses and uncover underlying relationships within the data. This analysis not only celebrates the rich history of the FIFA World Cup but also offers valuable insights for teams, coaches, and analysts looking to understand the factors that contribute to success on the world stage.

5. Deployment

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analysing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

