

FIFA WorldCup **Data Analysis**

LOW LEVEL DESIGN(LLD)

Syeda Viqar Sultana

VERSION: 1.0

DATE: 15/06/2024

Document Version Control

Date Issued	Version	Description	Author
15-06-2024	1.0	First Version on Complete Low-Level Design	Syeda Viquar Sultana

Project Introduction

This project offers a comprehensive analysis of the FIFA World Cup spanning from 1930 to 2014. Leveraging data sourced from Kaggle, it underwent ETL processes for accuracy and relevance. The primary aim was to provide a visually compelling and informative Power BI dashboard that encapsulates the essence of the tournament's history. The Dataset determines the countries who hosted the tournament, winners, qualifiers, and the attendance at the stadium. The project helps determine key metrics and factors that influence the World Cup win.

Table of Contents

Document Version Control.....	1
Project Introduction.....	2
Abstract	4
1.Introduction	5
1.1 What is Low-Level Design Document	5
1.2 Scope	5
2. Architecture.....	6
3. Architecture Description	7
3.1 Data Sourcing	7
3.2 Data Overview	7
3.3 Data Description	7
3.4 Data Loading in Power BI Query Editor	11
3.5 Data to Insights through Analysis and Visualizations	12

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 Low-Level Design Document?

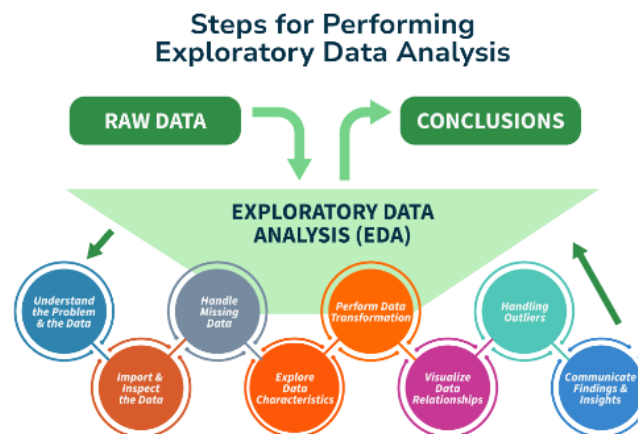
The purpose of this LLD or a Low-Level Design (LLD) document is to give the internal logical design of the actual program code for FIFA World Cup Data Analysis project. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document. This document is intended for both the stakeholders and the developers of this project and will be proposed to the higher management for its approval. 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.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

This study demonstrates how different analysis helps to make determine the different scenarios and probabilities of different teams qualifying and winning the tournament.

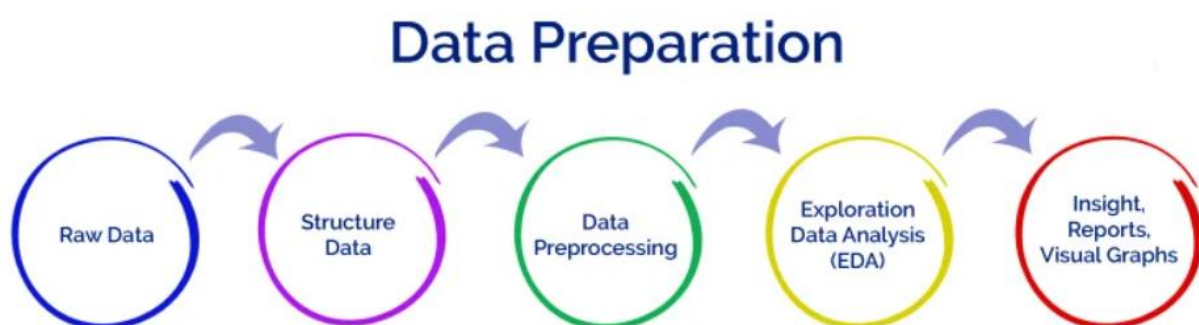
2.Architecture



Exploratory Data Analysis (EDA) is a technique used to analyze and understand data by summarizing its main characteristics. This technique involves the use of statistical and visualization tools to gain insights into the data. EDA is a powerful method for identifying patterns and trends, detecting outliers, and understanding the distribution of data. Power BI can be used for exploratory data analysis in a variety of ways, including data modelling, data visualization, and data analysis. These features allow businesses to gain insights into their data, identify patterns and trends, detect outliers, and make data-driven decisions.

Below are the following steps to follow for ETL:

1. Data Sourcing
2. Data Cleaning
3. Data Preprocessing
4. Exploratory Data Analysis (EDA)
4. Data Visualization



3. Architecture Description

3.1 Data Sourcing

The 3 datasets are in CSV (Comma Separated Values) format. MS Excel is to load the data.

These datasets are publicly available for research purposes.

Title: WorldCups.csv

WorldCupPlayers.csv

WorldCupMatches.csv

Source: [GitHub](#)

3.2 Data Overview

- The datasets are of size 1.38 KB, 2.05 MB, 233 KB respectively
- It includes three files in “.csv” format.
- Number of rows/records: 20, 37784, 854.
- Number of attributes: 10, 9, 20.

3.3 Data Description

The following attributes describes the dataset.

WorldCup.csv

- Year
 - Description: Year
 - Datatype: Int
- Country
 - Description: Name of different Countries
 - Datatype: string
- Winner
 - Description: Name of winning Countries
 - Datatype: string
- Runners-Up
 - Description: Name of different Countries
 - Datatype: string

- Third
 - Description: Name of different Countries
 - Datatype: string
- Fourth
 - Description: Name of different Countries
 - Datatype: string
- GoalsScored
 - Description: Number of Goals Scored
 - Datatype: Int
- QualifiedTeams
 - Description: Number of Teams qualified for the tournament
 - Datatype: Int
- MatchesPlayed
 - Description: Number of matches played in the tournament
 - Datatype: Int
- Attendance
 - Description: Number of spectators
 - Datatype: Decimal

WorldCupPlayers.csv

- RoundID
 - Description: ID number of the round
 - Datatype: Int
- MatchID
 - Description: ID number of the round
 - Datatype: Int
- Team Initials
 - Description: Initial letters of the team
 - Datatype: string
- Line-up
 - Description: Line-up
 - Datatype: char

- Shirt Number
 - Description: Shirt number of the player
 - Datatype: Int
- Player Name
 - Description: Name of the player
 - Datatype: string
- Position
 - Description: Position of the Player
 - Datatype: string
- Event
 - Description: Name of the event
 - Datatype: string

WorldCupMatches

- Year
 - Description: Year
 - Datatype: Int
- Datetime
 - Description: Date and time
 - Datatype: date
- Stage
 - Description: Name of stage
 - Datatype: string
- City
 - Description: Name of different Cities
 - Datatype: string
- Home team Name
 - Description: Name of different home teams
 - Datatype: string
- Home Team Goals
 - Description: Number of goals by home teams
 - Datatype: Int

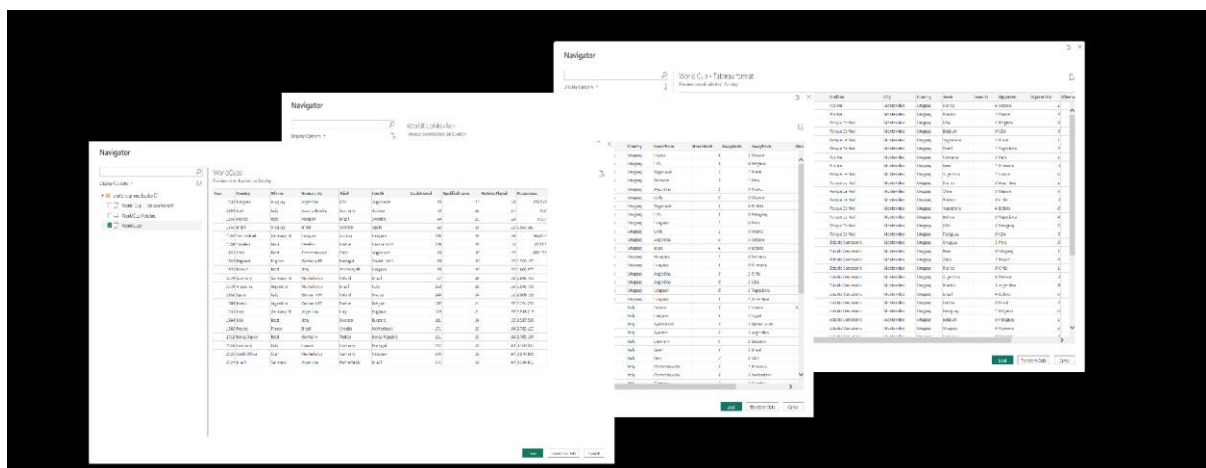
- Away Team Goals
 - Description: Number of goals by away teams
 - Datatype: Int
- Away team Name
 - Description: Name of different away teams
 - Datatype: string
- Win conditions
 - Description: Win conditions of the teams
 - Datatype: string
- Attendance
 - Description: Number of spectators
 - Datatype: Decimal
- Half-time Home Goals
 - Description: Goals scored by home teams in half-time
 - Datatype: Int
- Half-time Away Goals
 - Description: Goals scored by away teams in half-time
 - Datatype: Int
- Referee
 - Description: Name of the referee
 - Datatype: string
- Assistant 1
 - Description: Name of assistant 1
 - Datatype: string
- Assistant 2
 - Description: Name of assistant 2
 - Datatype: string
- RoundID
 - Description: ID number of the round
 - Datatype: Int
- MatchID
 - Description: ID number of the round
 - Datatype: Int

- Home team Initials
 - Description: Initials of different home teams
 - Datatype: string
- Away team Initials
 - Description: Initials of different away teams
 - Datatype: string

3.4 Data Loading in Power BI Query Editor

Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data warehouse, and more with the following characteristics.

- There can be multiple rows and columns in the data.
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data like numbers, strings, dates, Boolean etc.



3.5 Data to Insights through Visualizations and Data Analysis

