World Cup Soccer Data Games from 1930-2014

Authors : Ricardo Jose Molina Gonzalez and Enghel Lopez

Course : CAP1788 Intro to Data Analytics

Professor: Ralph De Arazoza

Data Set Overview

Complete Game Records

Details of matches from group stage to finals, spanning 1930 to 2014.

Tableau Public Source

The original data set as shared on Tableau Public platform.

Key Data Features

Includes date, host countries, cities, stadiums, scores, teams, and observations.

Data Cleaning Process

Deleting Unnecessary Columns

Removed year columns in Excel as date column was sufficient and less redundant.

Creating Meaningful Columns

Used IF conditions to extract winning countries and penalty results from observations.

Data Standardization

Standardized stadium names by filtering special characters for readability.

Removing Duplicates and Handling Missing Data

Deleted duplicated rows for Brazil 2014; identified missing eliminatory data in 1950 World Cup.

Made with GAMMA

Objective of the Analysis

- Teams that have won the World Cup while hosting the tournament.

- Examine the trend of average goals scored annually

- Analyze average goals scored during the Eliminatory rounds

- Investigate the relationship between Away and Home goals in the Group Stage

Made with GAMMA

Conclusion

Future Directions

- Conduct regional analysis to compare World Cup performance by continent and country

- Calculate and analyze the percentage of games decided by penalties across all tournaments

- Identify and rank the top 10 most offensive teams based on average goals scored per game

Sources

- The original data set was obtained from the Tableau Public platform, providing comprehensive records of all matches from 1930 to 2014.

- https://public.tableau.com/app/profile/robert.mundigl/viz/fifa_world_cup_statistics/WorldCupStatsperTeam

- Fifa.com to verify the source

Made with GAMMA