

A dark blue vertical bar is positioned on the left side of the page. A blue arrow-shaped banner points to the right from this bar, containing the date. In the bottom-left corner, there are several thin, curved lines in shades of blue and grey, resembling stylized grass or abstract brushstrokes.

4/11/2024

European Soccer Database Analysis

Ahmed Mohamed Fawzy Abdelfattah

Ahmed Mohamed Fawzy Abdelfattah

Supervised by:

Eng. Ahmed Mostafa

11/4/2024

Table of Contents

- Project Summary
- Tools and Technologies
- Project Steps
 - Data Acquisition and Exploration
 - Data Preprocessing and Feature Engineering
 - Data Cleaning
 - Exploratory Data Analysis (EDA)
 - Web Application Development
- Key Deliverables
- Project Impact

Project Summary

This project explored and analyzed data from the European Soccer Database available on Kaggle. The database covers professional soccer matches, players, and teams from various European leagues between 2008 and 2016. The goal was to create a user-friendly web application for visualizing and analyzing team performance, formations, and overall trends.

Tools and Technologies

- **Data Exploration and Cleaning:**
 - DB Browser for SQLite access
 - Python for data manipulation
 - SQLite3 for database handling
 - NumPy and Pandas for data analysis
- **Data Visualization:**
 - Plotly for creating interactive charts
- **Web Application Development:**
 - Streamlit for building the web app

Project Steps

1. **Data Acquisition and Exploration**
 - Downloaded the European Soccer Database from Kaggle.
 - Connected to the database using SQLite and loaded data tables into Pandas dataframes.
 - Analyzed each table individually to understand its structure and content.
 - Performed initial data exploration to get an overview of the data.

2. Data Preprocessing and Feature Engineering

- Extracted new features from the match data, including formations, results, and points earned.
- Reshaped the "Match" table using melt function for easier merging with other tables.
- Merged the "Match" table with other relevant tables in the database.

3. Data Cleaning

- Cleaned the data by handling missing values and inconsistencies.
- Removed unnecessary columns to improve data efficiency.

4. Exploratory Data Analysis (EDA):

- Performed univariate and bivariate analysis on teams, formations, and overall season performance.
- Visualized key insights using Plotly to create interactive charts.

5. Web Application Development:

- Built interactive dashboards using Streamlit to showcase the data analysis.
- Developed a user-friendly web application to allow anyone to explore the data and analyze their favorite teams.

Key Deliverables

- A Streamlit web application for interactive exploration of the European Soccer Database (<https://lnkd.in/d3UuR3it>).
- A GitHub repository containing the project code (<https://lnkd.in/duiVksk>).

Project Impact

This project provides a valuable tool for soccer fans and analysts to explore historical data, analyze trends in team performance, and gain insights into formations and overall league dynamics. The user-friendly web application interface allows anyone to easily access and interact with the data, regardless of their technical expertise.