Capstone **Project**

Utiva
...the future of learning

Case Scenario

The European Soccer data was gotten from Kaggle. As part of an open source data crawling project, this real life data is available for reuse by researchers and applications for analysis and storytelling.

You have been employed by the Fédération Internationale de Football Association (FIFA) as a data analyst to review European Football between July 2008 and May 2016. You're required to provide answers to certain football statistical questions and to come up with insights concerning European Football and her players.

Find more information on the match tables HERE.

Case Questions

1. Which two teams have the highest total sum of number of goals scored in all their face-offs with each other (regardless of who is at home or away)?

2. Which young player (under or equal 21 years old as of start of 2011/2012 season) has the highest average overall rating in the 2011/2012 season? Assuming 2011/2012 season started on August 13th, 2011 and ended on May 13th, 2012.

3. Write a query to return the final English Premier League table for 2010/2011 season. Return the teams names, total points (win=3 points, draw=1 point, lose=0 point), total goal scored, total goal against, and total goal difference. Sort the resulting table by total points and then by goal difference (both in descending order).

Case Questions

4. Role 6 and 4 are considered as center back and central midfielder respectively by some people. Conversely, others consider role 6 and 4 as central midfielder and center back respectively. Using only the home players, compare the y coordinate with the highest number of matches for these two roles and tell us which of these two roles is considered as central back in the dataset.

5. Which team(s) has the maximum chance creation shooting in the database?

6. Who are the top 10 players with the highest average sprint_speed in the database?

7. Which goalkeeper (role 1) has the most number of appearance in the database?

8. Which league has the highest goals per match in the database?

9. Which team has the lowest Bet365 home win odds in the database?

Case Questions

10.   There is a constant debate of who the greatest of all time (GOAT) is. Compare Lionel Messi and Cristiano Ronaldo's average, median, maximum, and minimum overall rating.

11.   Find the correlation coefficient and slope of a linear regression line fitted on finishing and overall_rating (using finishing as the independent variable and overall_rating as the dependent variable) for all players that had one time played role number 9.
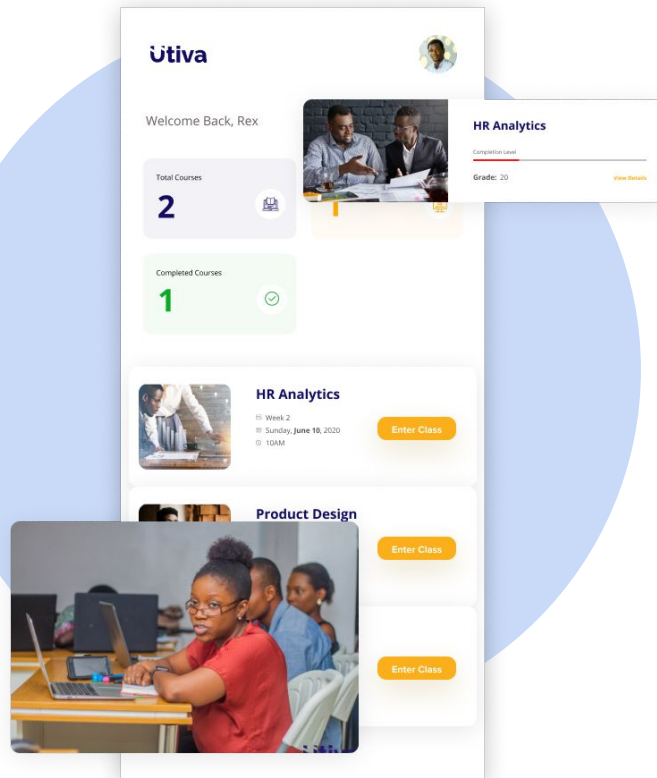
12.   Report any other insights you obtained from the data (exclusive of what could be gotten from any of the tasks above).

Dataset

Access the dataset [HERE](#)

academy@utiva.io
767 Somide Odujirin Ave, Omole Phase 1, Lagos
www.utiva.io