

Diogo Almeida
Madavane Anbane
Aroune Krishnaraj

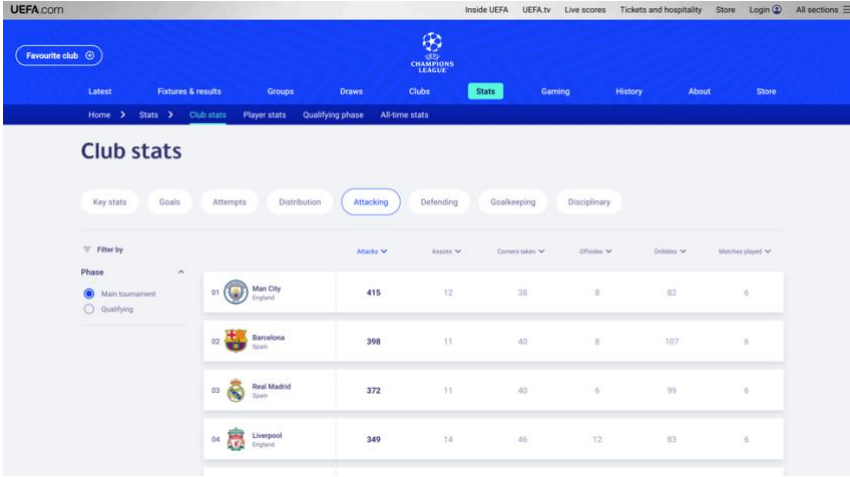
Sujet :

Notre projet vise à explorer l'influence des statistiques des clubs et des joueurs de football sur les interactions sur les réseaux sociaux, plus précisément sur le réseau social Twitter, comme le nombre de tweets. Par exemple, nous souhaitons déterminer combien de tweets sont générés lorsqu'un joueur marque un but ou un club gagne. Ce thème a été choisi car il s'agit de notre point commun

Les données :

Notre projet utilise deux sources de données : une méthode de Scrapping et une API. En utilisant la technique de Scrapping, nous accédons au site de l'UEFA pour récupérer les statistiques des joueurs qui attaquent (buts, assists, etc.) et les clubs les plus performants dans les compétitions de la Ligue des champions et de l'Europa League.

Voici à quoi ressemblent les sites scrappés :



Club	Goals	Assists	Goals per game	Assists per game	Goals per assist	Goals per game
01 Man City (England)	415	12	38	8	82	6
02 Barcelona (Spain)	398	11	40	8	107	6
03 Real Madrid (Spain)	372	11	40	6	99	6
04 Liverpool (England)	349	14	46	12	83	6

Player	Assists	Goals taken	Offsides	Matches played
01 Evander Middlesbrough - Middlesbrough	5	12	1	6
02 Bakasetas Preston - Middlesbrough	4	17	0	6
03 Pelagiri Preston - Middlesbrough	3	29	1	5
03 Gokpo	3	10	5	5

Grâce au scrapping, les datasets sont créés sous forme de 4 fichiers CSV. 2 LDC et 2 EL. Ces 4 fichiers vont être transférés sur Mongo DB Atlas.

Nous utilisons l'API de Twitter pour récupérer les tweets et les nombres de likes. Pour alimenter la BDD avec l'API de Twitter, on a utilisé une librairie qui se nomme Tweepy qui permet intégrer des fonctionnalités liées à Twitter telles que la récupération de tweets, la publication de tweets, la récupération des informations sur les utilisateurs. Le but est de mettre en corrélation avec les statistiques des joueurs ou des clubs.

À l'aide de clés d'accès et autorisations, l'API twitter pourra se connecter à la base de données MongoDB.

Le fichier Requierement.txt est l'endroit où nos versions des différentes librairies sont affichées.

8 collections sont obtenues à la fin, 4 liées à l'api twitter et 4 dernières collections pour les scrappings.

The screenshot displays the Atlas Cluster0 dashboard. The top navigation bar includes the Atlas logo, user information (Aroune's Org), and links to Access Manager and Billing. The main header shows 'Project 0' and 'Data Services'. The left sidebar contains sections for DEPLOYMENT, SERVICES, and SECURITY. The main content area is titled 'Cluster0' and shows 'DATABASES: 1' and 'COLLECTIONS: 8'. A search bar is present. The 'Collections' tab is active, showing a list of collections under 'Cluster0', including 'EL-Clubs'. The 'EL-Clubs' collection is selected, displaying its details: STORAGE SIZE: 36KB, LOGICAL DATA SIZE: 4.79KB, TOTAL DOCUMENTS: 30, INDEXES TOTAL SIZE: 36KB. Below this, there are tabs for Find, Indexes, Schema Anti-Patterns, Aggregation, and Search Indexes. A filter bar shows '{ field: "value" }'. The 'QUERY RESULTS: 1-20 OF MANY' section displays two JSON documents. The first document is for a player from Real Sociedad, and the second is for a player from Roma. The URL at the bottom is a long, complex string.

Grâce à nos graphiques sur notre Dashboard que nous avons réalisé avec la librairie Streamlit. Nous pouvons conclure que en effet les performances des joueurs ont une petite influence sur l'interaction avec les réseaux sociaux (twitter) mais il ne faut pas oublier la notoriété du joueur, du club et aussi des compétitions dans lesquelles ils jouent.