Data science assignment on seo

I will make a video for explanation on what i did how i did my thought process
Link-https://www.tella.tv/video/seo-data-insights-recommendations-1p0n
Ppt link-https://gamma.app/docs/SEO-Data-Science-Assignment-lzjv13h8k35vso5

Problem Statement
Your mission is to analyze SEO data and translate the insights into business recommendations that drive measurable results.
Using the provided dataset, you need to accomplish the following:
1. Analyze Current SEO Performance

- Identify which keywords perform the best.
- Determine which markets (FR, EN, IT, BR) show the strongest potential.

2. Identify SEO Growth Opportunities

- Find keywords with high search volume but an average position greater than 5
- Highlight cases where competitors are ranking higher than your site.

3. Provide Actionable Business Recommendations

- Prioritize improvements (e.g., content, technical fixes, backlinks).
- Suggest new keywords or clusters to target for growth.

Expected Deliverable

- A short report (2–3 pages or slides) that includes:
- Methodology
- Key visualizations (charts, tables, etc.)
- Key insights
- Clear, prioritized business recommendations

Optional
You may propose a simple estimation or model showing the impact of ranking improvements (e.g., moving from position 8 to 3) on organic traffic.
The focus should remain on reasoning and business understanding

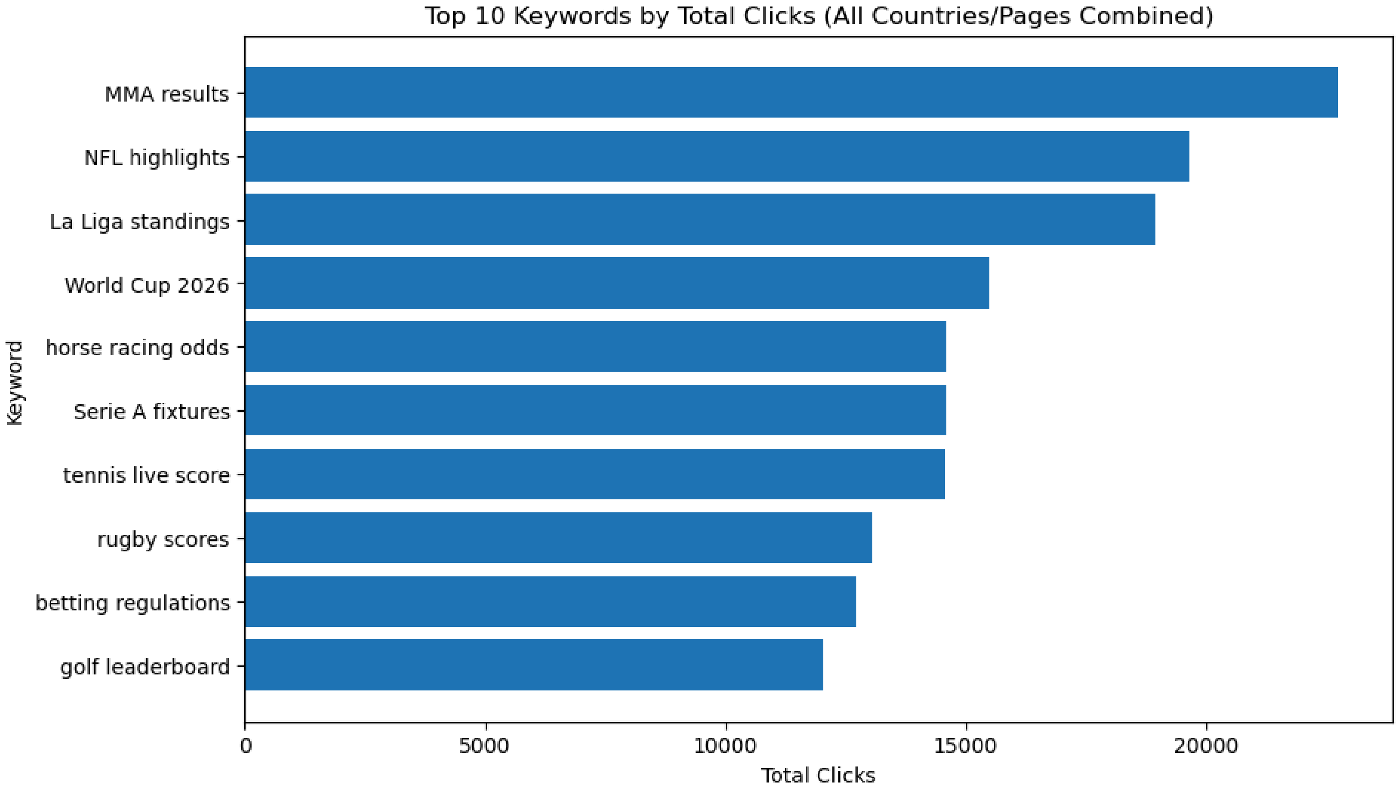
1.Methodology

1.first i read the dataset using pd.read_csv imported the libraries too.
2.understood the data first i went to google understood the metrics of seo and the columns present and how it impacts and then i went to solve the basic eda and all.
3.i did check for any columns with spaces or something and then i checked the data summary using the pd.info and pd.describe and then the shape of data and then the unique value present in each columns.
4.converting to lower and checking if all teh numeric columns are numeric by looping and converting them
5.no duplicate elements were found but there were repeated elements in keywords columns but for different country so they are unique
6.no nan values were found .i here by confirm all the data was right maybe a little outlier in ctr but all good.
7.and then i went to the data and collect insights from it .

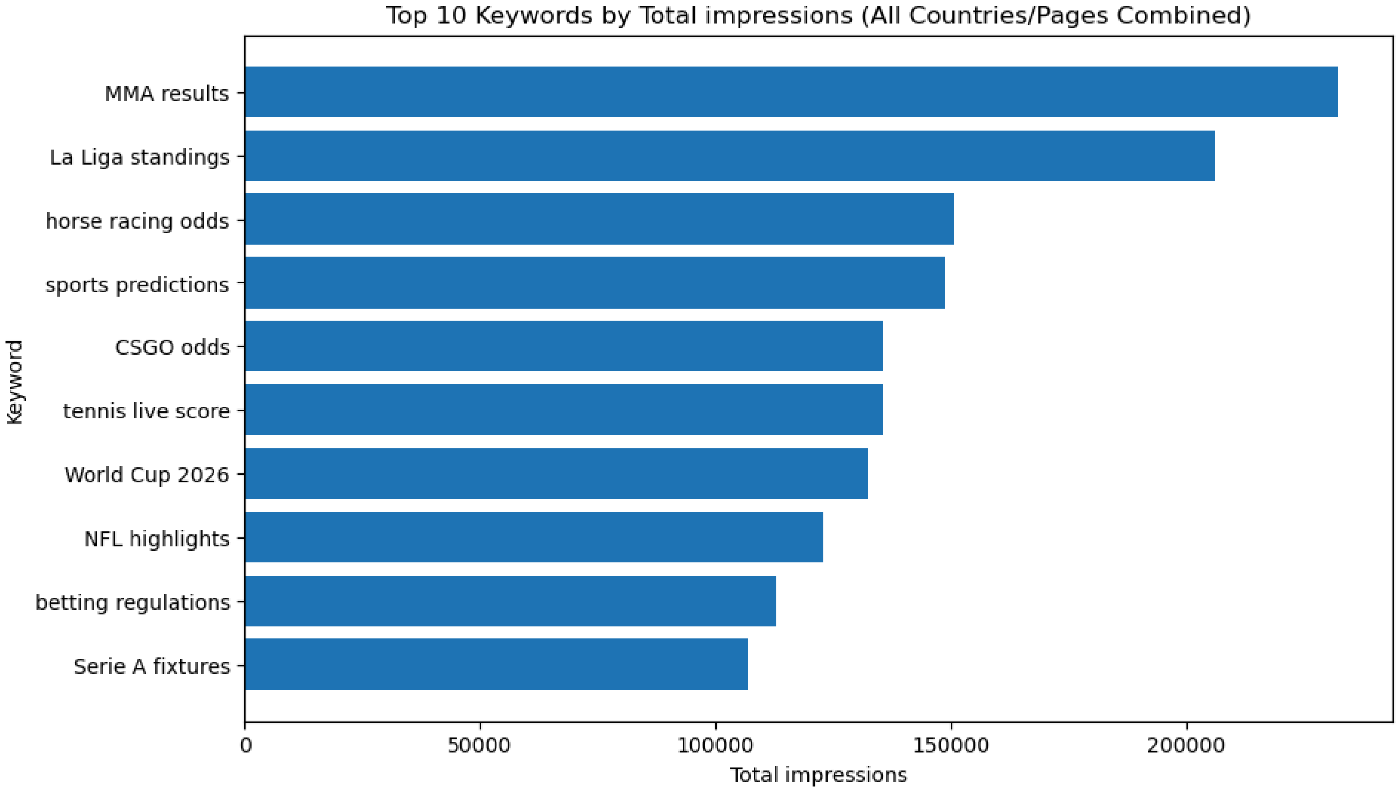
2.Insights

1.i found out that best performing keywords i grouped them as i said there were repeated and for each country so i grouped the keywords and used mean ,sum on clicks and impressions and ctr . so based on these 3 we have data

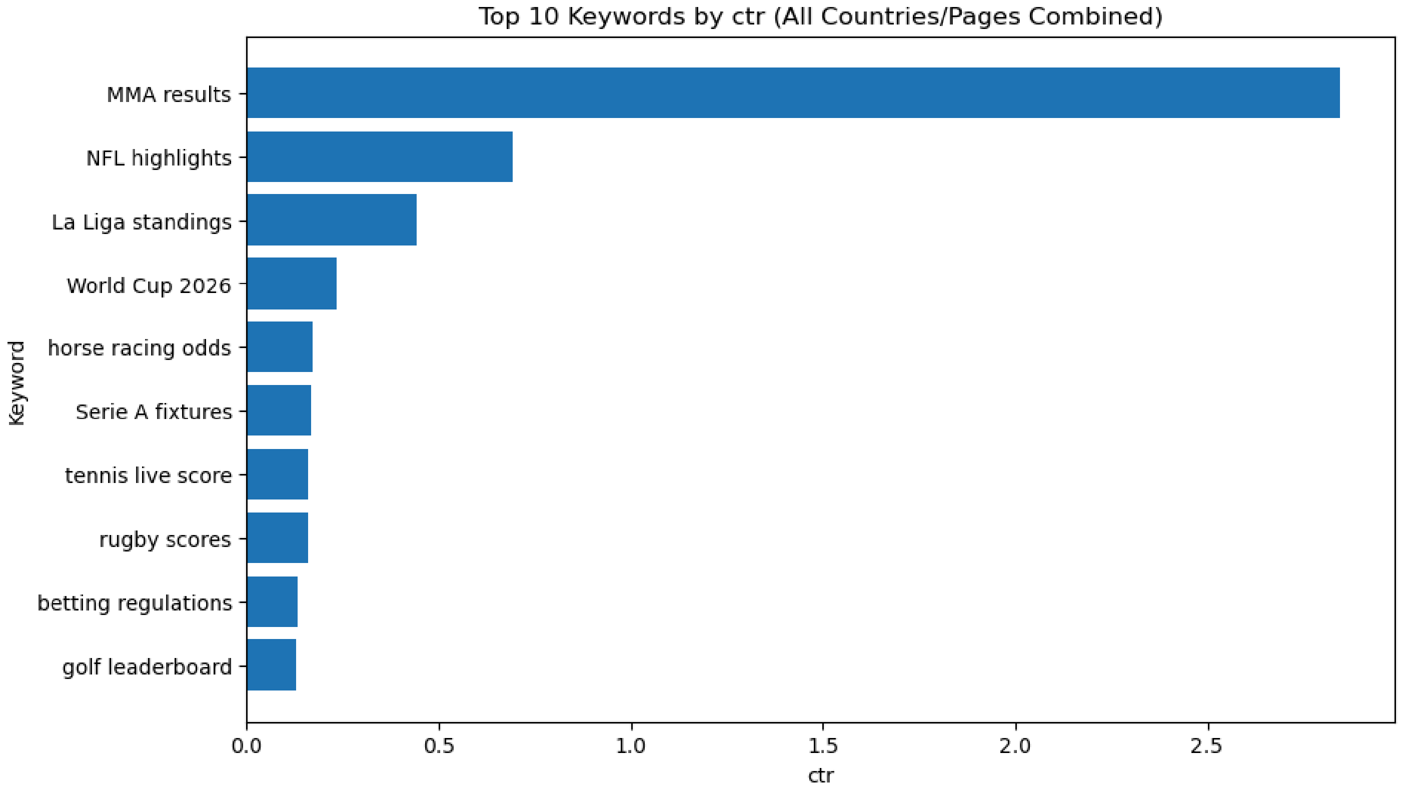
on clicks



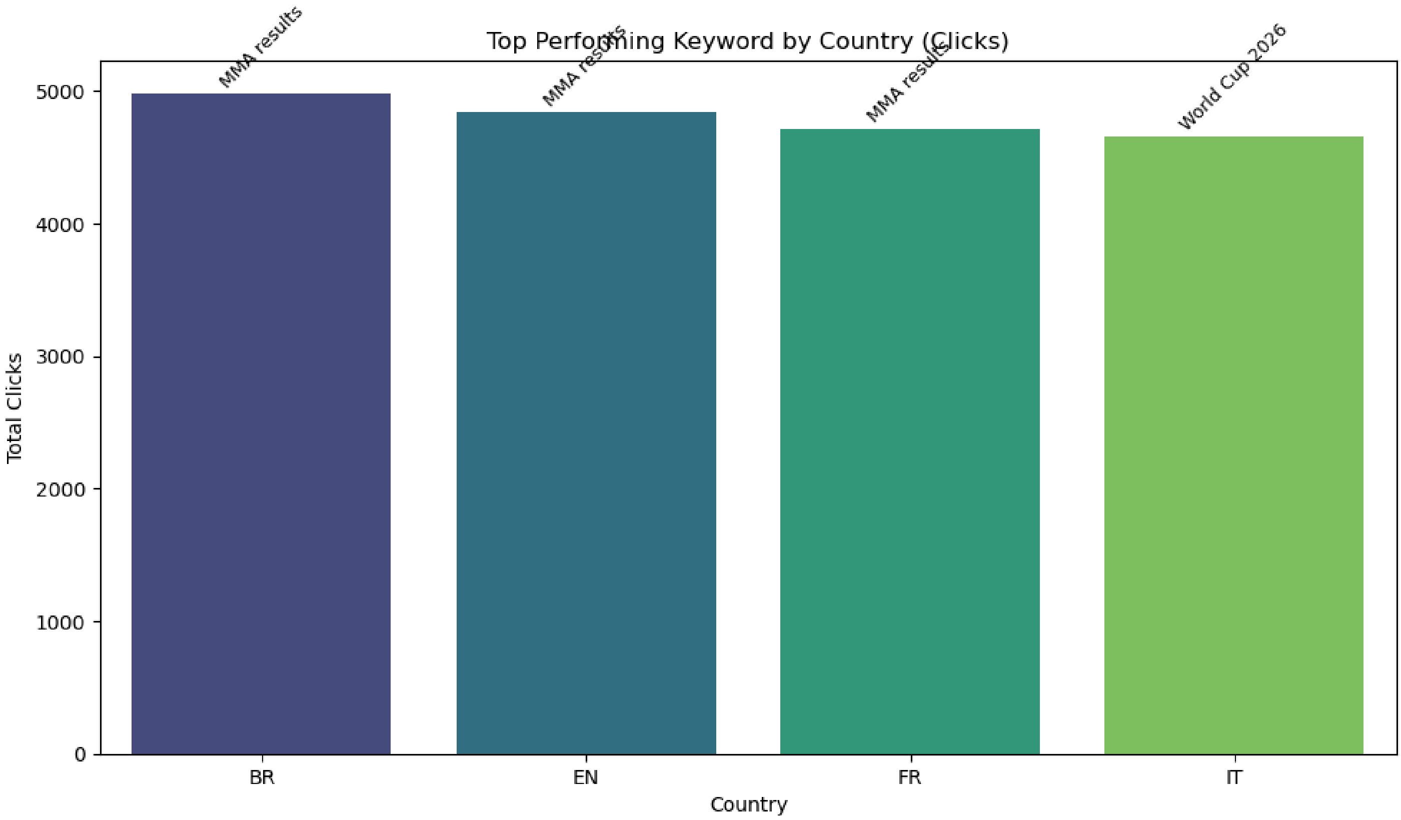
On impressions



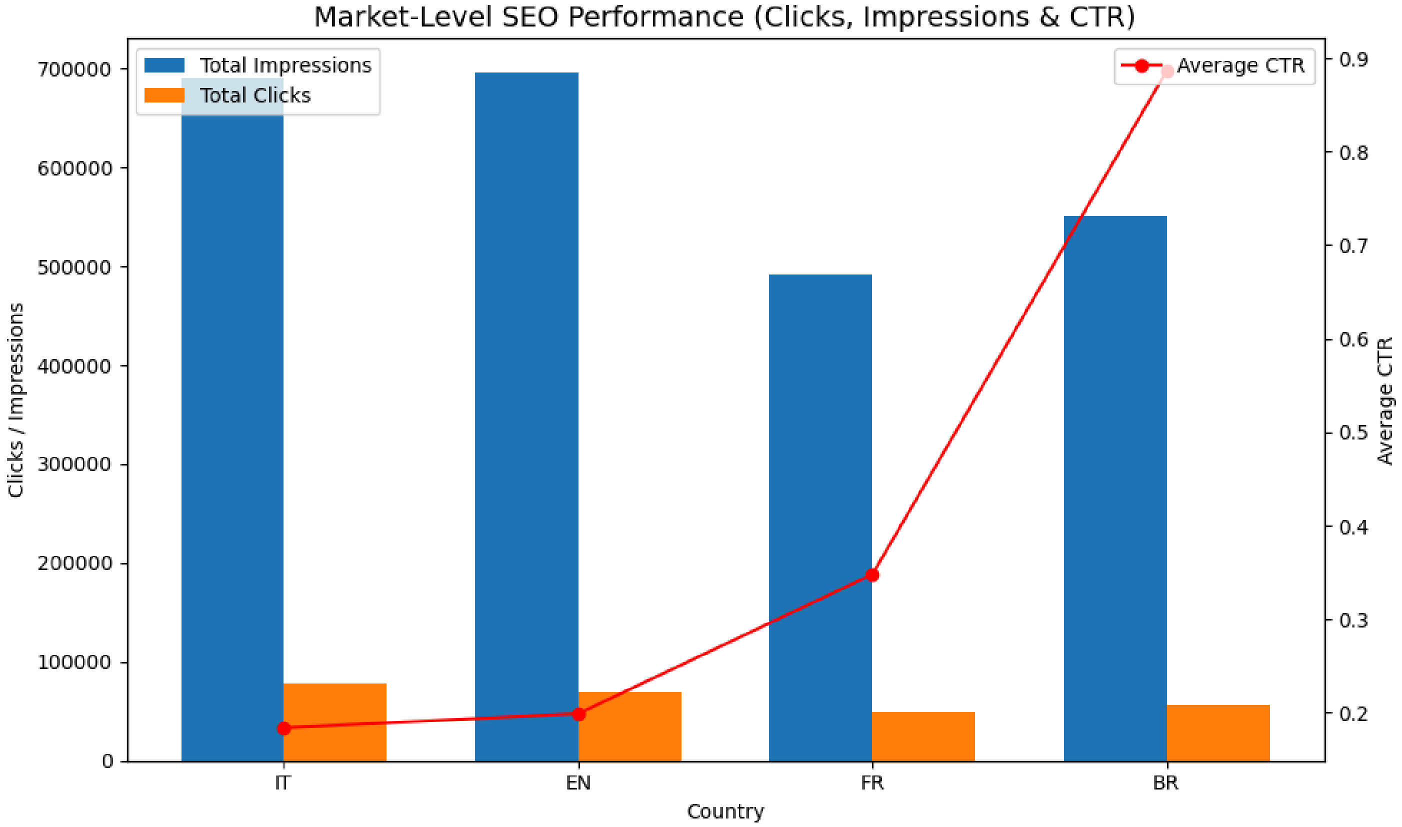
On ctr



2.top keywords in each country

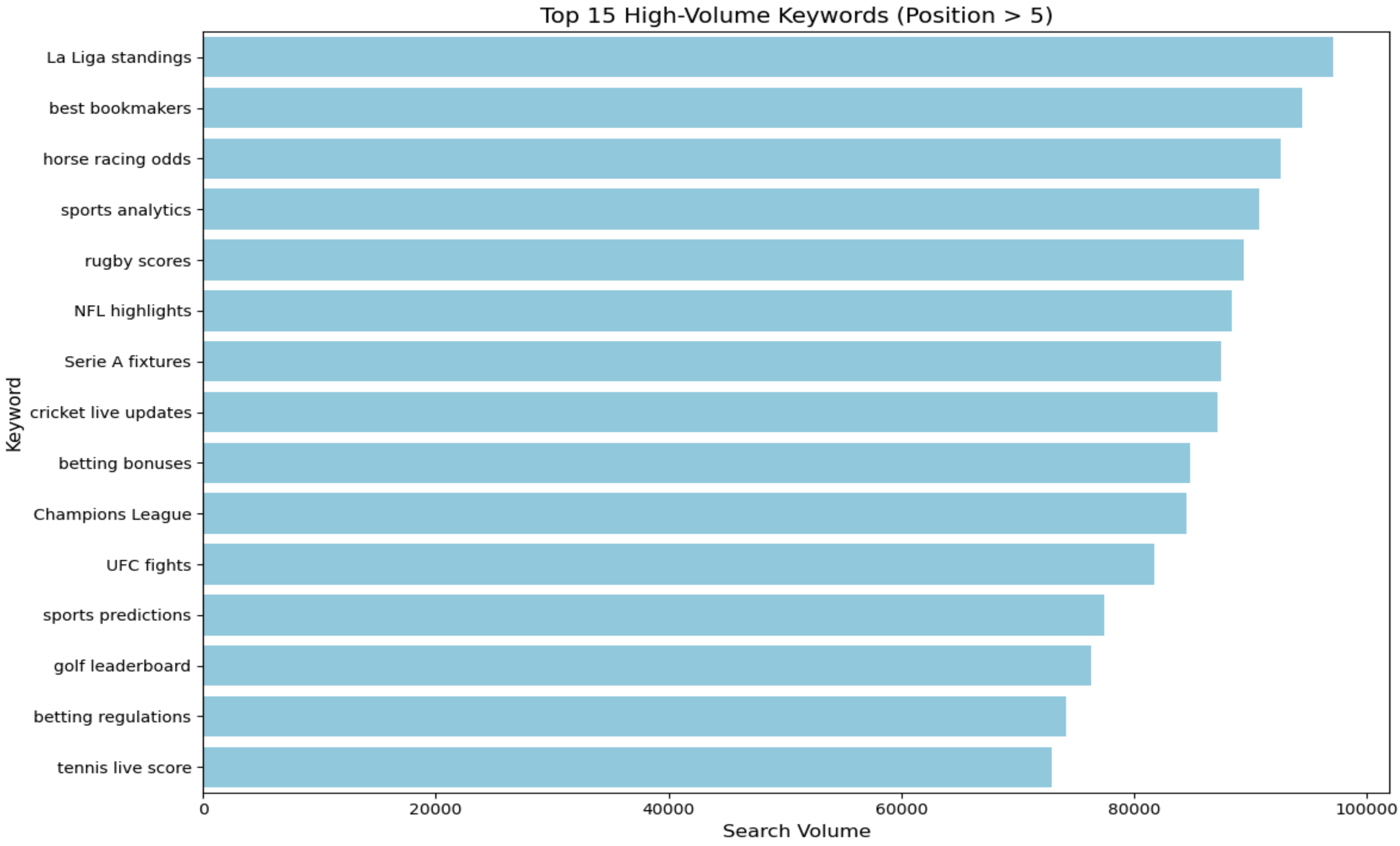


2.Determine which markets (FR, EN, IT, BR) show the strongest potential.

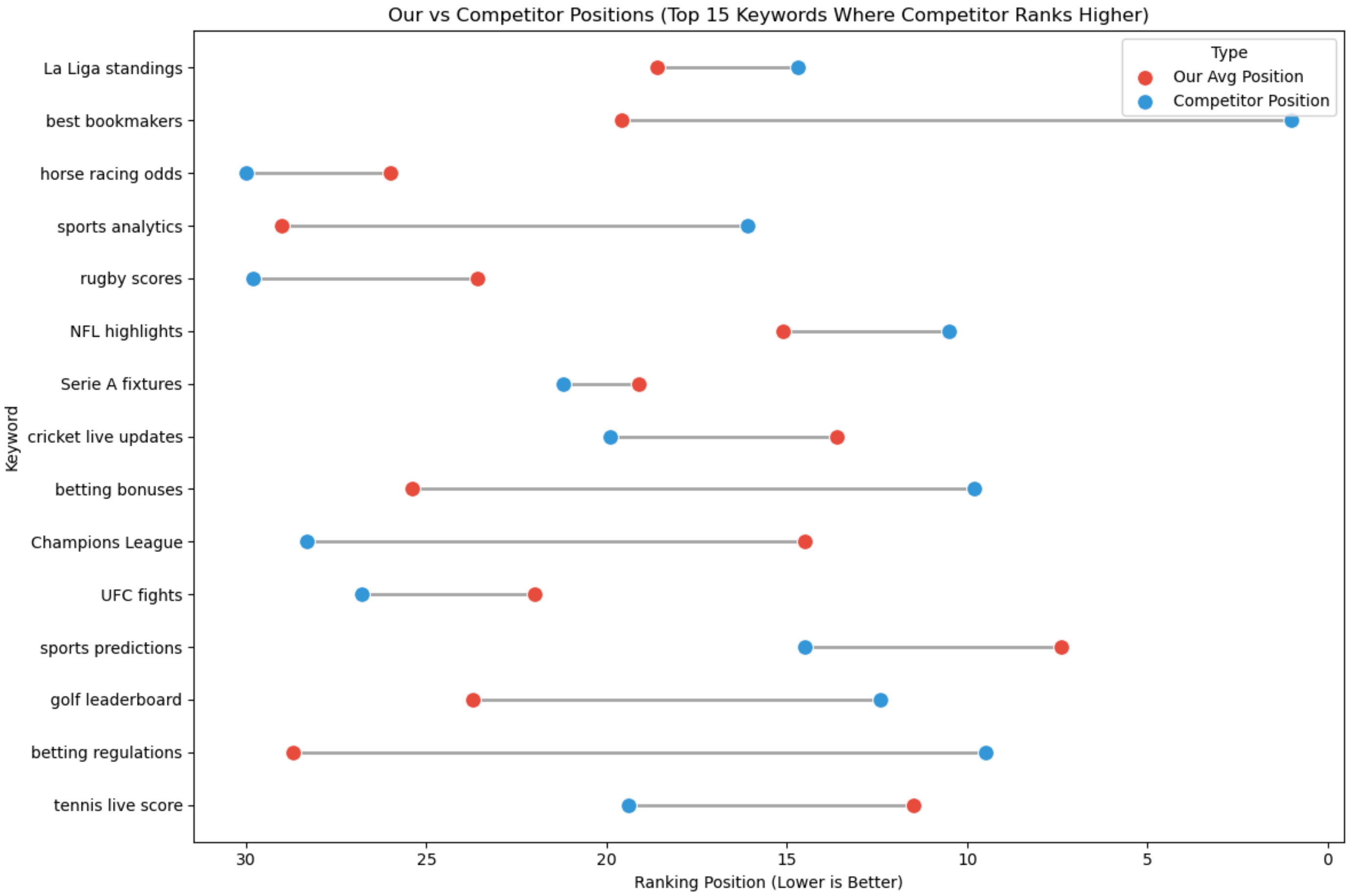


so here italy has average of .18 ctr which means that only 18 percent people click the link but many people see our links that is 689835 people so we have an high potential opportunity here we can write some good lines over here to increse that ctr.
and same goes for england

2.Find keywords with high search volume but and average position greater than 5 threshold is 0.75



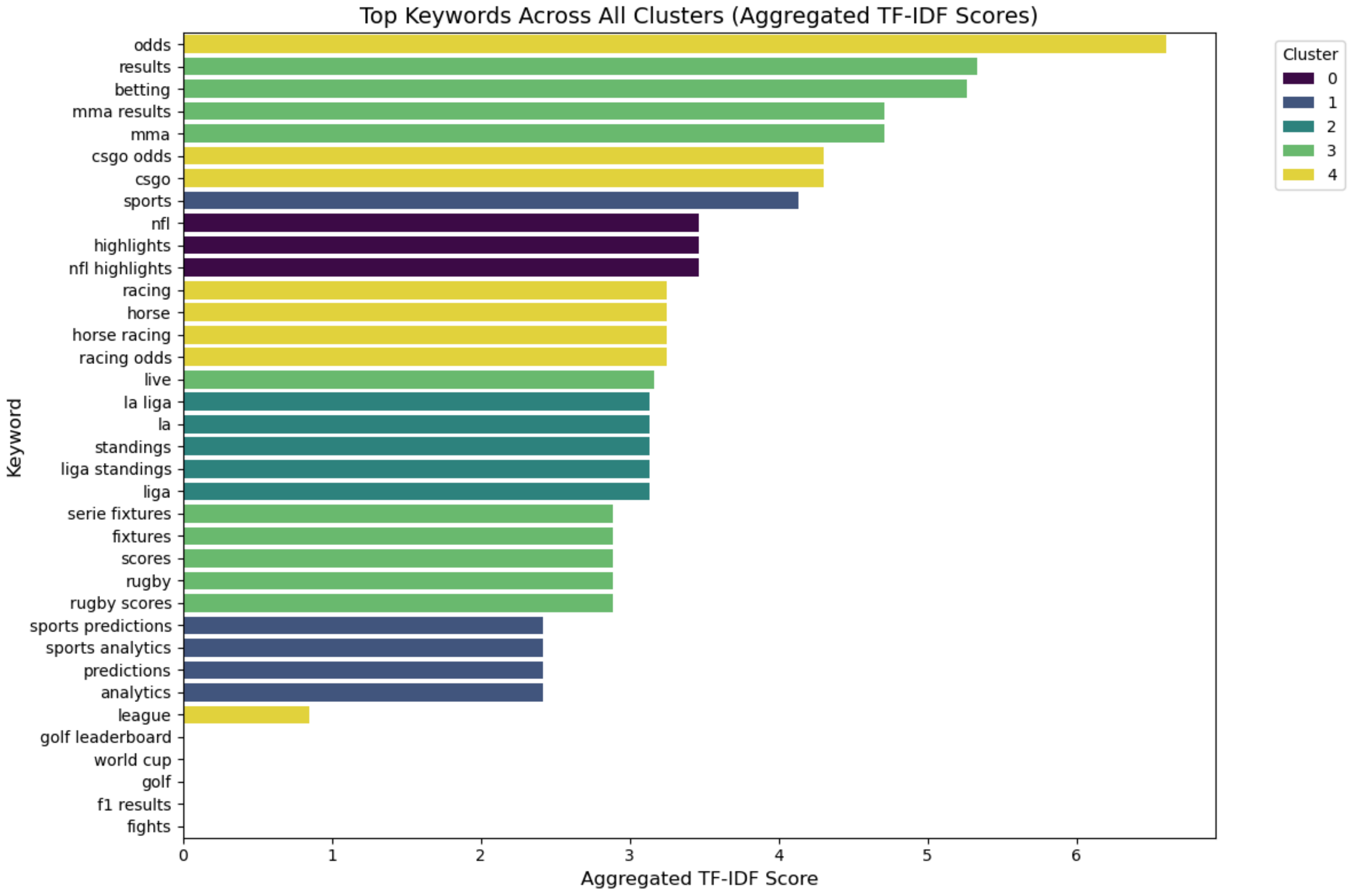
b.- Highlight cases where competitors are ranking higher than your site.



3. Provide Actionable Business Recommendations

Priority	Focus Area	Impact	Effort	Why
1	Optimize EN & IT content (CTR + title/meta)	High	Medium	Biggest markets, low CTR
2	Update mid-ranking high-volume keywords	High	Low	Quick wins (position > 5)
3	Fix technical issues (speed, schema)	Medium	Medium	Improves site quality
4	Backlink & authority building	High	High	Long-term impact
5	Localization & translations	Medium	Medium	Better local targeting

b. suggesting new keywords which shows growth



	keyword	cluster
0	F1 results	3
1	UFC fights	3
2	NBA playoffs	3
3	tennis live score	3
4	La Liga standings	2
5	UFC fights	3
6	Serie A fixtures	3
7	tennis live score	3
8	MMA results	3
9	La Liga standings	2
10	sports predictions	1
11	MMA results	3

Also made a model to .first i went to use a power law one a d then increased complexity and using multiple linear regression and now the r^2 is 0.667 not that good but if we have more data and shift to a nicer model we can get better results .