1. **Bar chart displaying the average mentions of a specific word in a journal per article**

SELECT

JournalTitle,

ROUND( SUM ((( length(text) - length(replace(lower(text), {{ Subject }}, '')) ) / length({{ Subject }}))) / count(\*), 2) as AvgMentionPerArticle,

count(\*) as NumberOfArticles

FROM century1\_avro

GROUP BY JournalTitle

ORDER BY AvgMentionPerArticle DESC

LIMIT 10



**Variations**

1. Normalize for article length
2. Include a time-frame

SELECT

JournalTitle,

100 \* SUM (((length(text) - length(replace(lower(text), {{ Subject }}, ''))) / length({{ Subject }}))) / count(\*) / SUM (wordCount) as PercentOfWords,

count(\*) as NumberOfArticles

FROM century1\_avro

GROUP BY JournalTitle

ORDER BY PercentOfWords DESC

LIMIT 10

I think if I am going to use the SIZE, SPLIT, you are going to have to subtract by one

* It will always split up the string into 1 more section than the number of word