

Exploratory Data Analysis Report

Dataset

I used this [dataset](#) to perform Data analysis on. This data consists of 11 .csv files containing stories from Hespress with details such as the author, the publishing date, the topic. In order to achieve this analysis, I used Microsoft PowerBI to remove any duplicates, clean any unnecessary values, merge the csv files, and visualize the data to gain insights.

Insights

- **Number of authors in each topic**

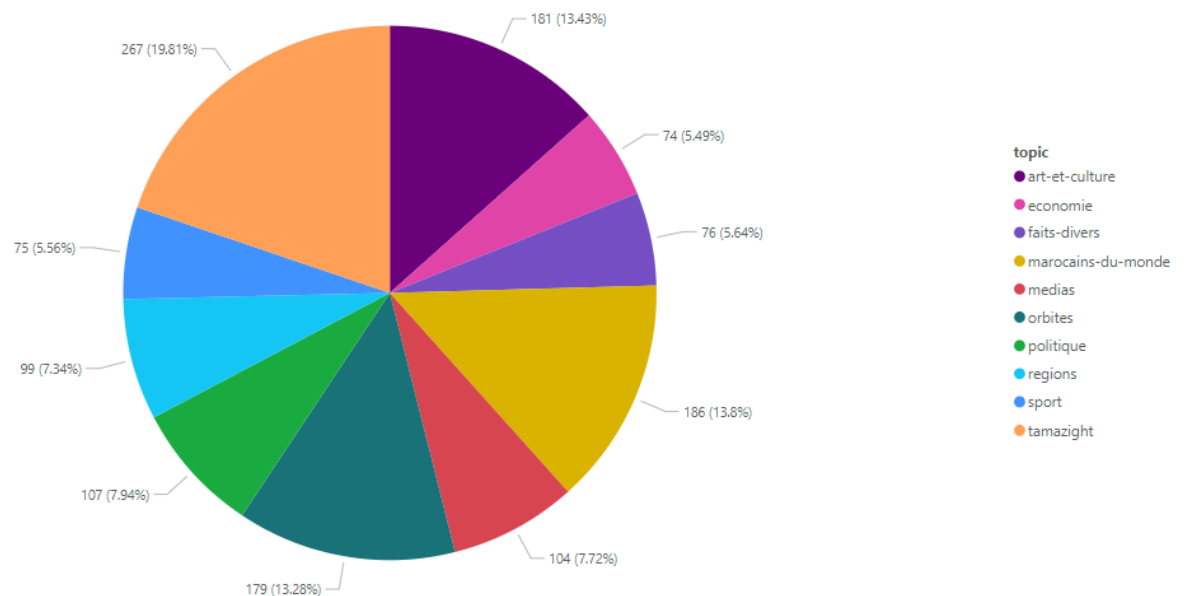


Figure 1

As shown in the above pie chart, we can see that most authors are interested in writing about marocains du monde topic.

- **Number of examples per topic**

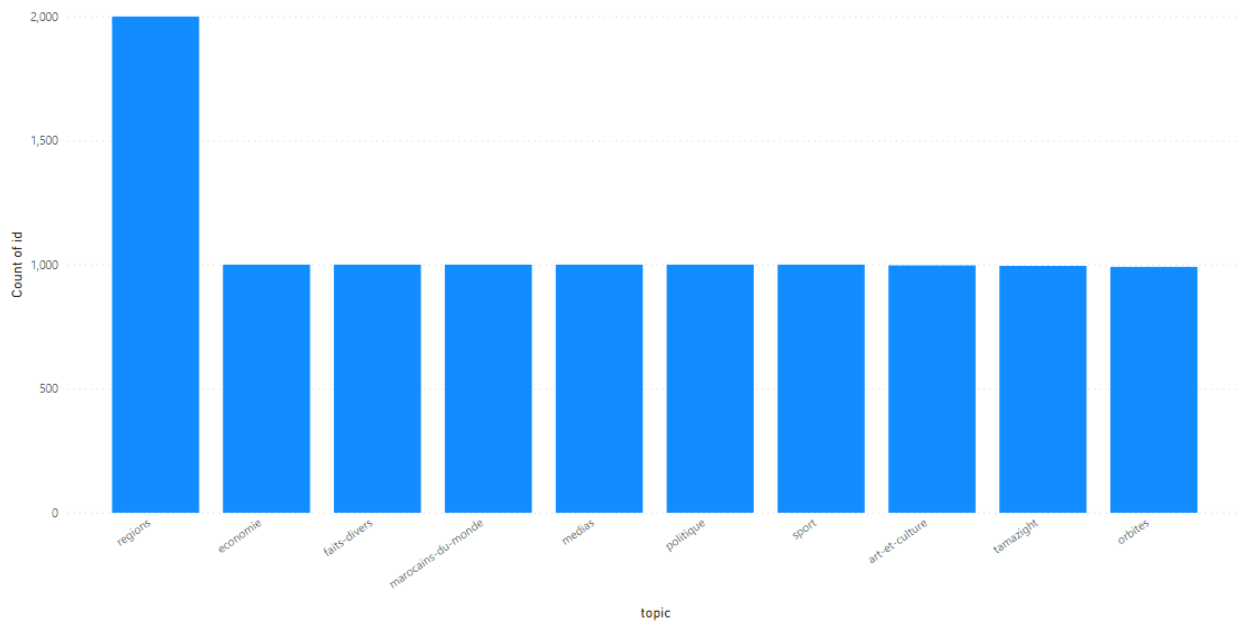


Figure 2

As shown in the above column chart, we can see that our data mostly consists of examples in regions topic and all the other topics have equal amount of data. Even though regions topic has the largest number of examples in our dataset it doesn't attract many writers as shown in figure1.

- Number of words in an example.

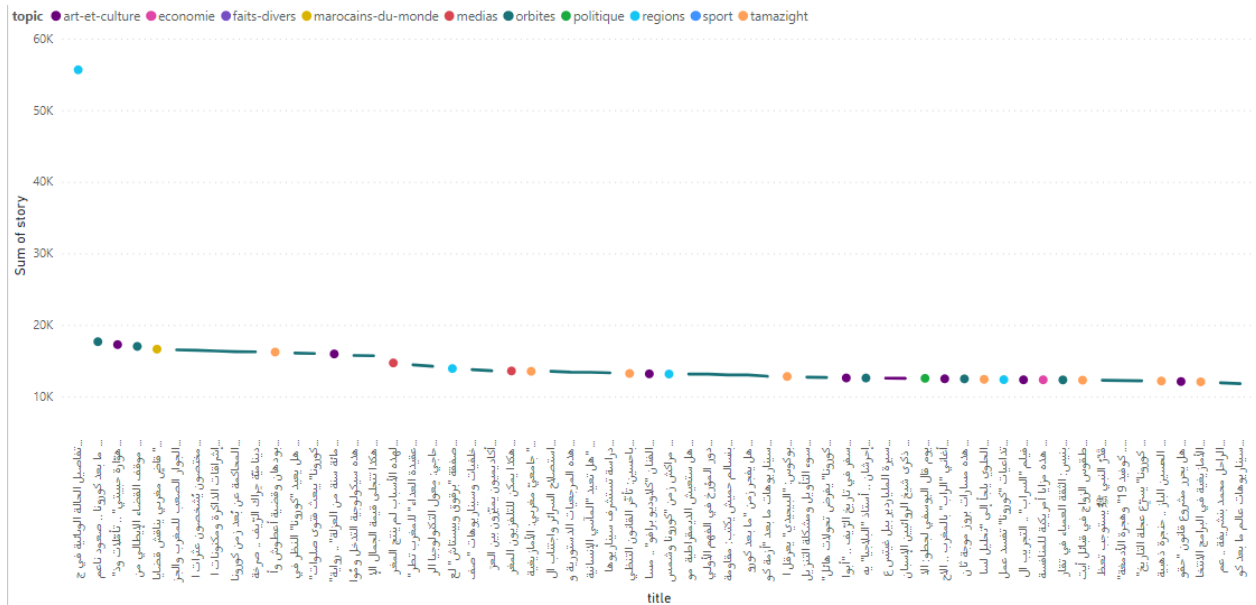


Figure 3

As shown in the above line chart, we can see that the example that has most text in it is from the regions topic and most of the writings are between 10,000 and 20,000 words.