

Identifying content topics of opportunity for **PlateaPR**



presented by **Tania Hernandez Colon**



Goal

- Better understand our audience.
- Identify topics that resonates the most with the audience.

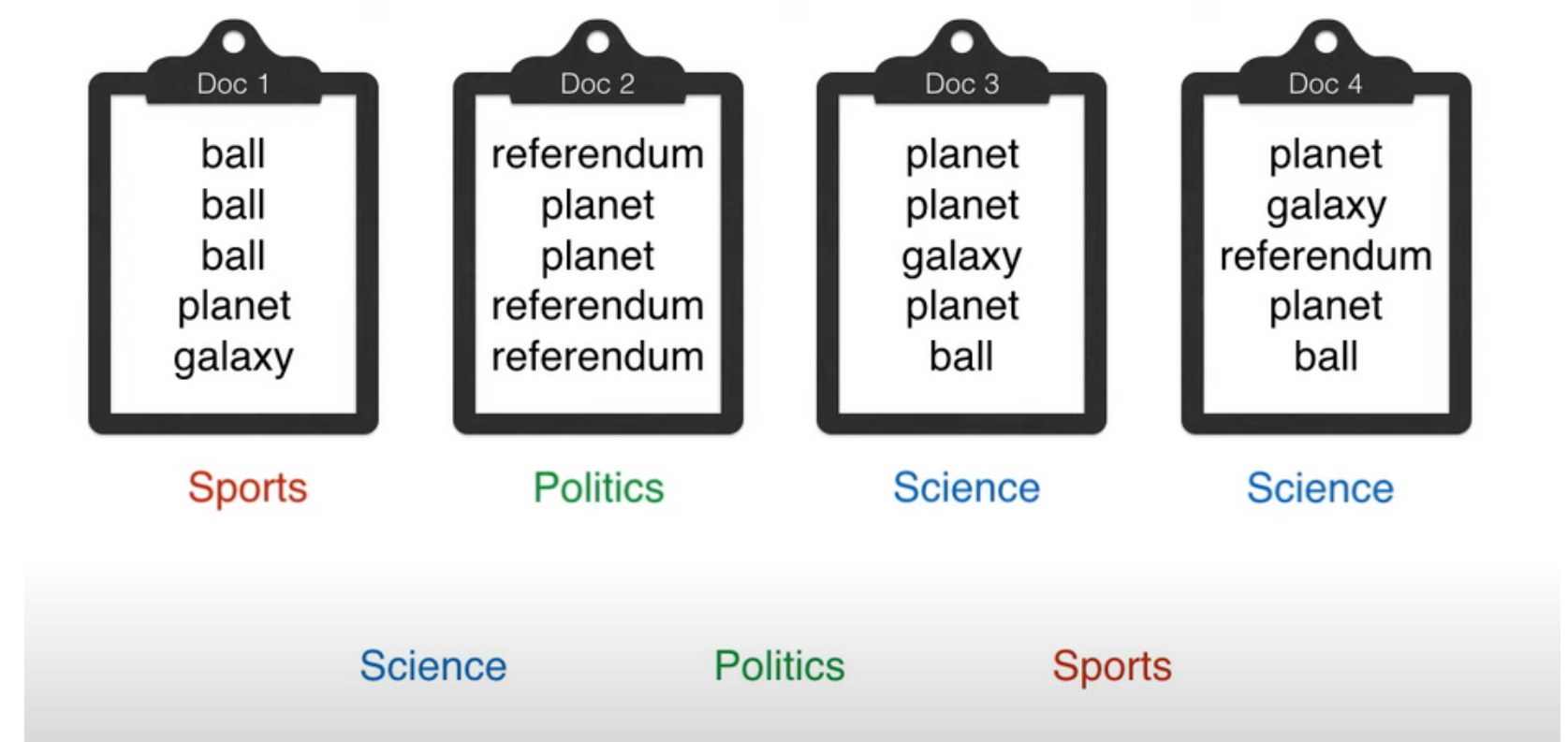
Approach

- Obtain the blog articles and sessions associated with those
- Classify articles with topics using Latent Dirichlet Allocation (LDA)
- Identify which topics bring more sessions

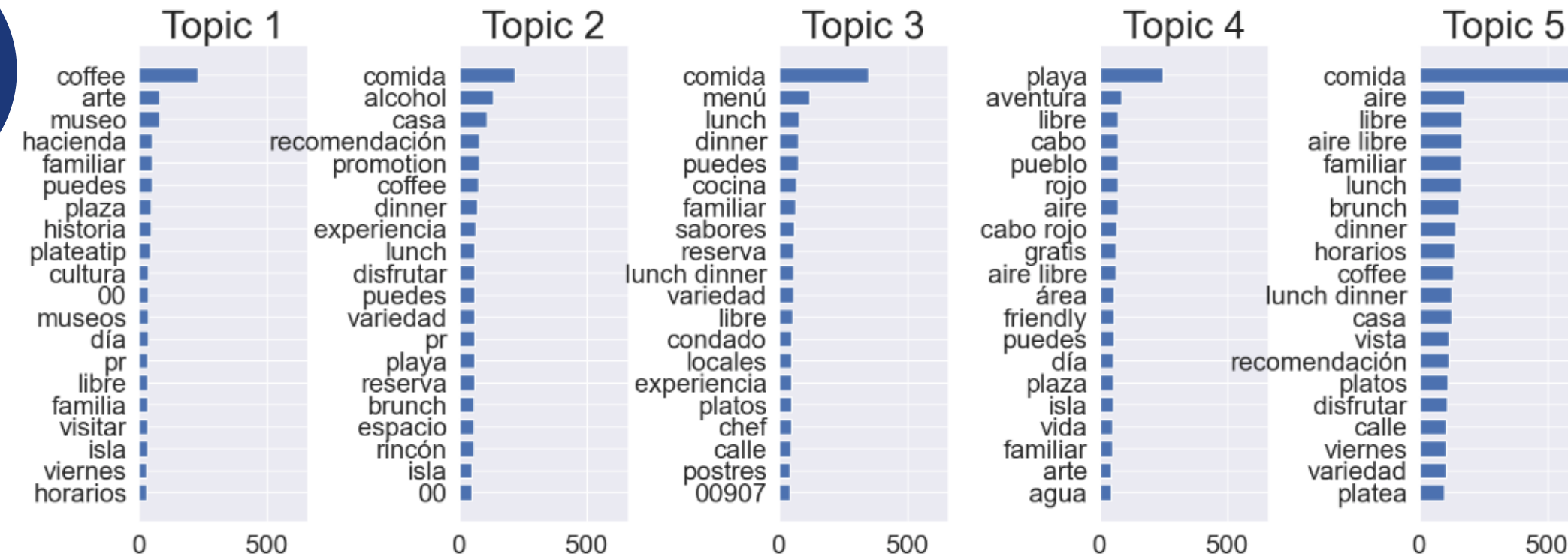
LDA

At a high level the LDA uses two probability distributions given data that is "words present in documents".

1. The probability of a topic over words
2. The probability of a document over topics

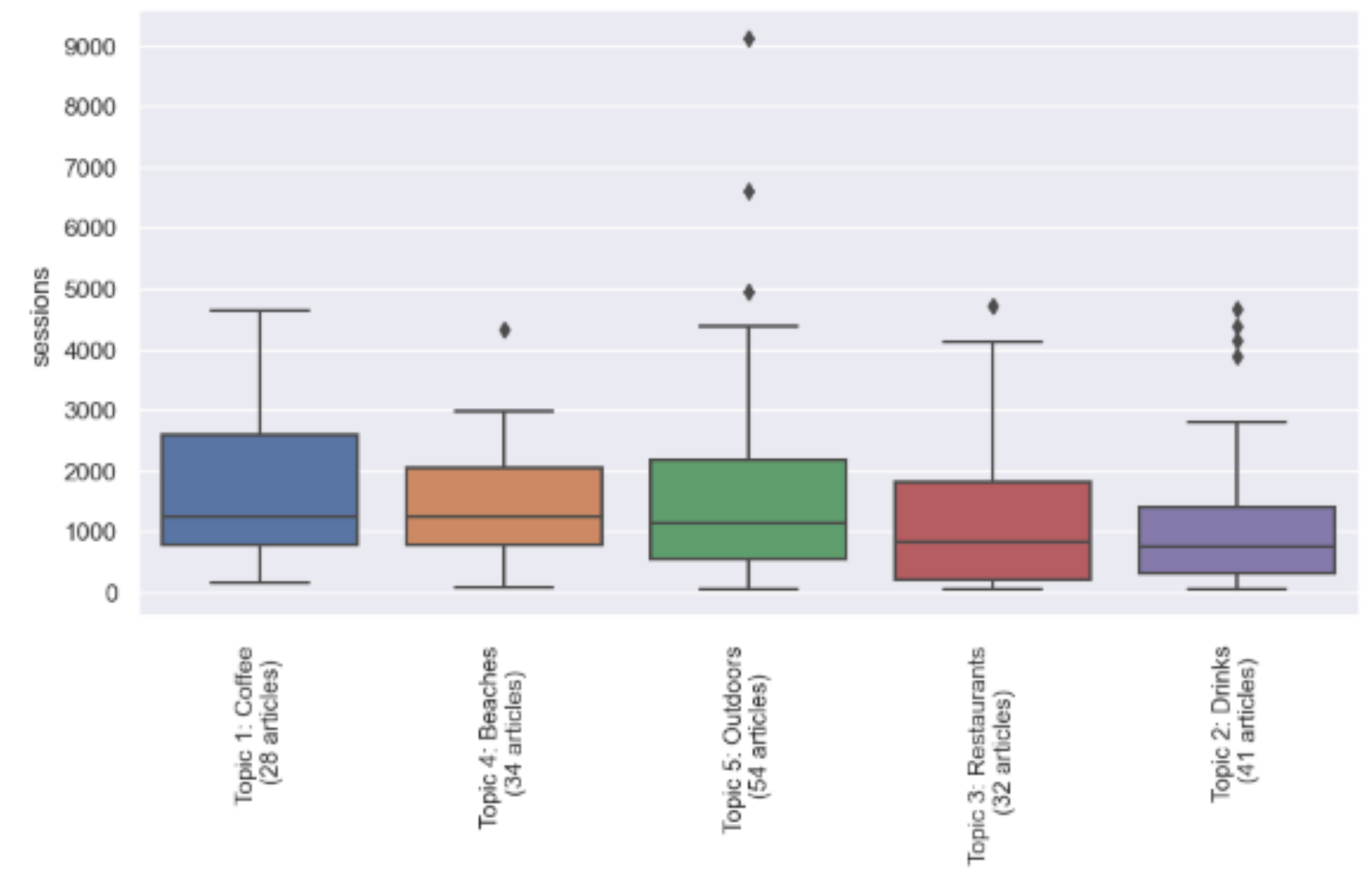
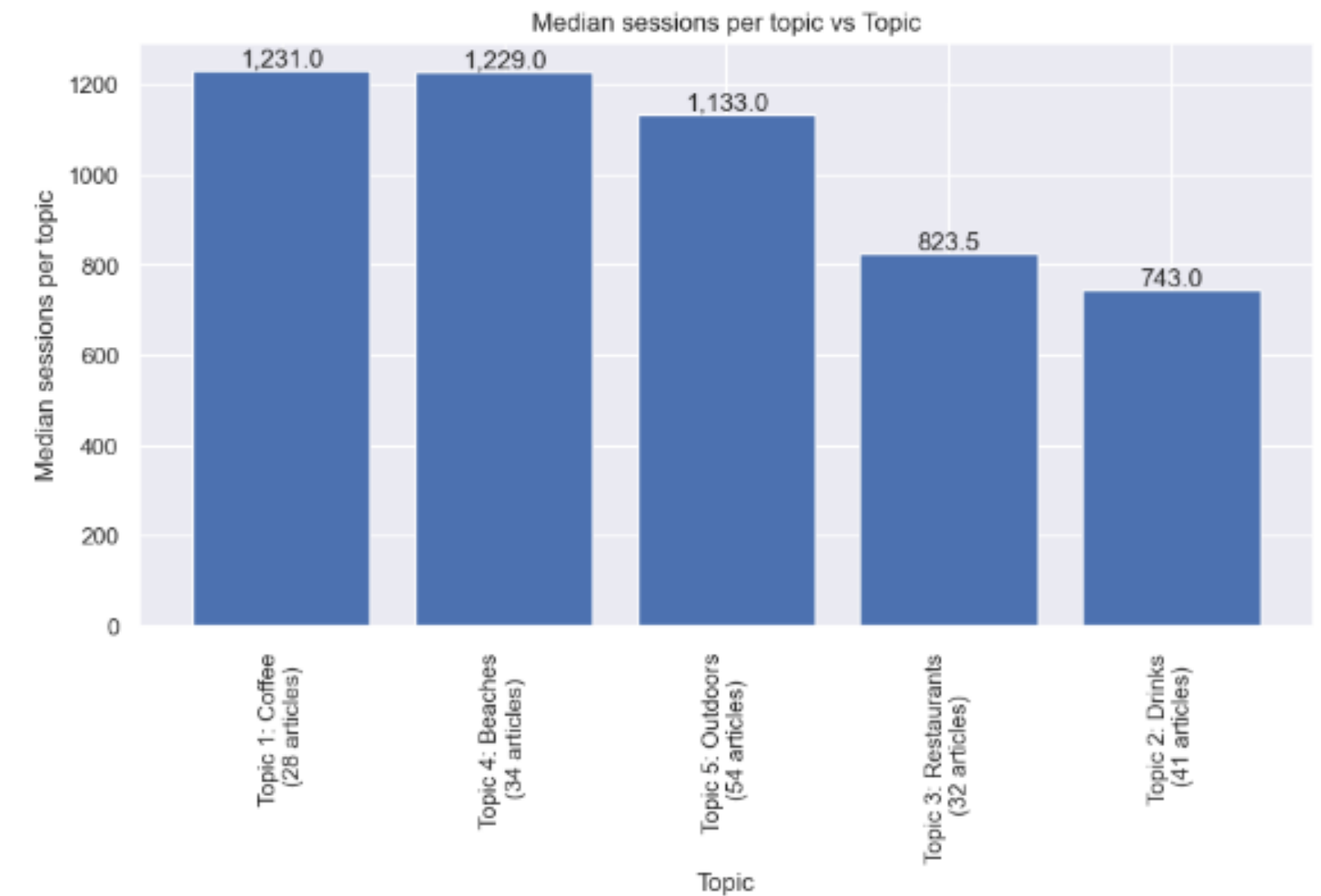
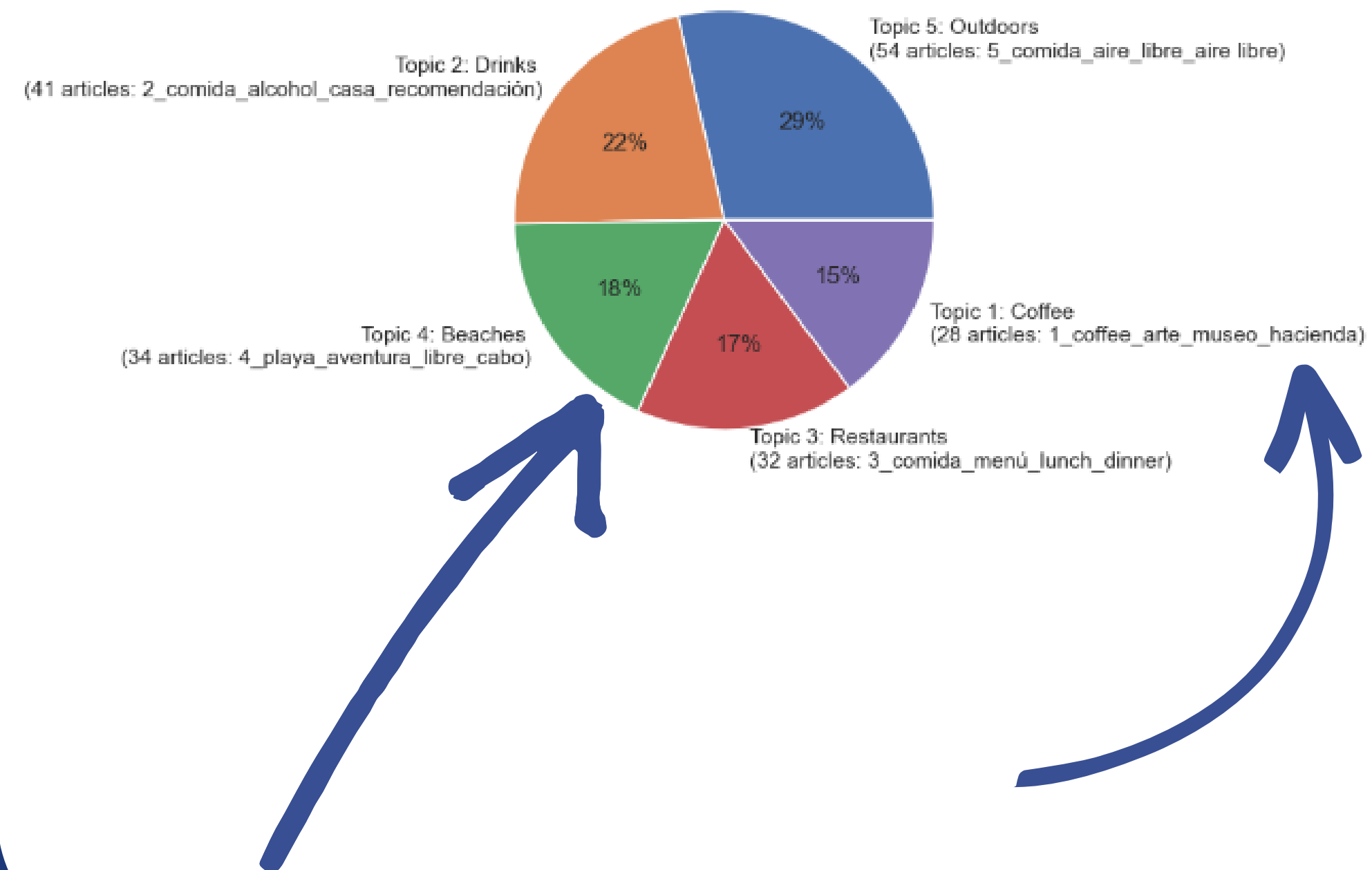


Results



Topic 1	Featuring coffee, art, museum. A broader topic could be “Coffee”.
Topic 2	Featuring the food, alcohol, house. A broader topic could be “Drinks”
Topic 3	Featuring food menu lunch, dinner “Restaurants”
Topic 4	Featuring the words beach, adventures, outdoors, “Beaches”
Topic 5	Featuring outdoors 3 times, food, familiar. A broader topic could be “Outdoors”

Results





Questions?