Capstone Final Paper:

Identifying content topics of opportunity for PlateaPR

Data Science Accelerator

By: Tania Hernandez Colon

Goals

The goal for this capstone project is to identify what are the topics that have more impact on the PlateaPR blog. Identifying topics that are more impactful and have less articles highlighting topics of opportunities.

Procedure

The procedure was to get the articles in the PlateaPR blog and use Latent Dirichlet Allocation (LDA) topic model to cluster those in 5 broader topics. Then cross reference those topics with the sessions they would generate to identify clusters that bring more sessions. In the data cleaning process, some articles/pages were removed, for example: Terms and conditions, What is Platea?, Fiona Hurricane associated articles since it does not represent typical Platea'sPR content, as well as, an outlier article that is updated on a weekly basis.
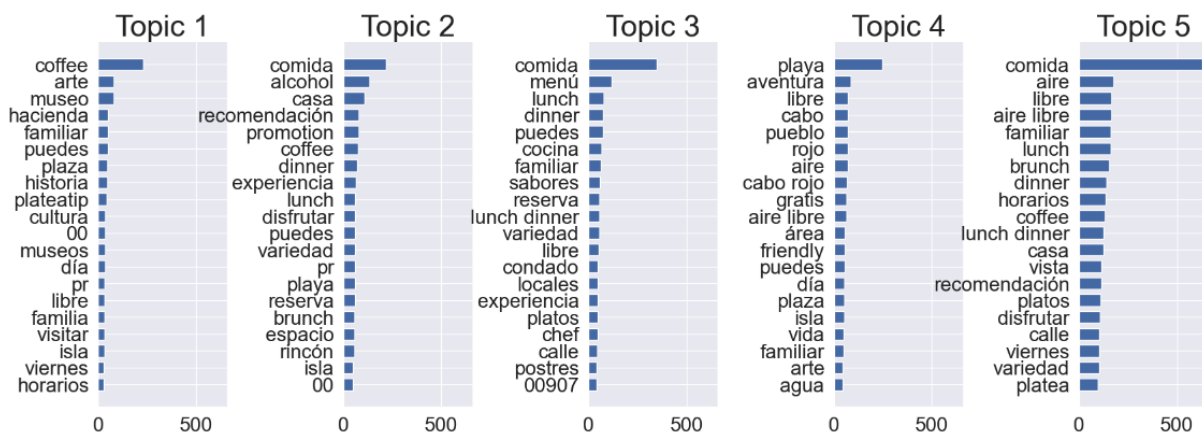
Results

Figure 1. Top words in topics

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”**

Disclaimer: These broader topics are made based on the top in words topics and a view of the titles on the cluster . There may be articles that are not coffee related under the coffee topic and so on.
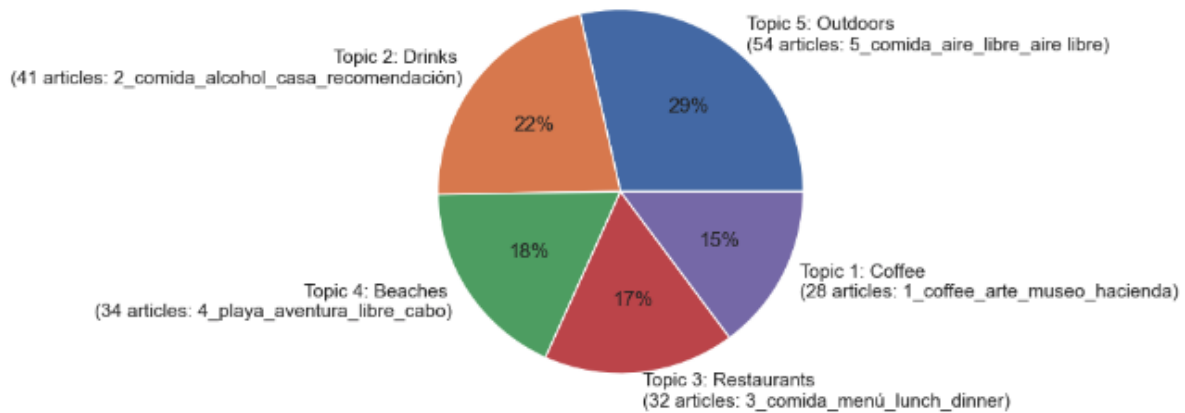
Figure 2. Articles by Topics

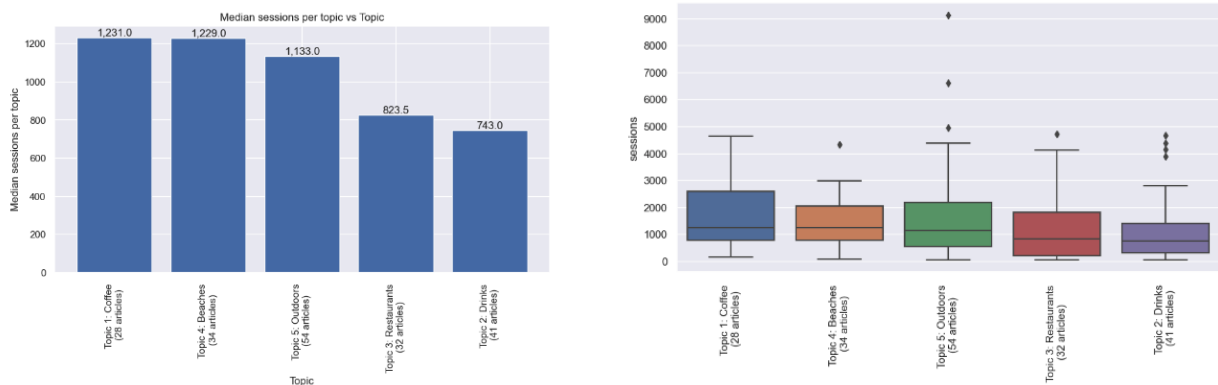The group with the majority of articles is the Outdoors topic.

Figure 3. Median sessions per topic

To get a measure of impact we are using the median number of sessions per article on that topic. The topic **Coffee** followed by **Beaches** are the two topics that have the highest median sessions making more impact..

Recommendations

The recommendation would be to increase focus or importance on content related to **coffee or beach** topics, as they bring more sessions per article. Therefore the opportunity would be to focus on the generation of articles similar to the ones under the coffee/beach clusters. More details on what are those on the jupyter notebook for this project.