

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

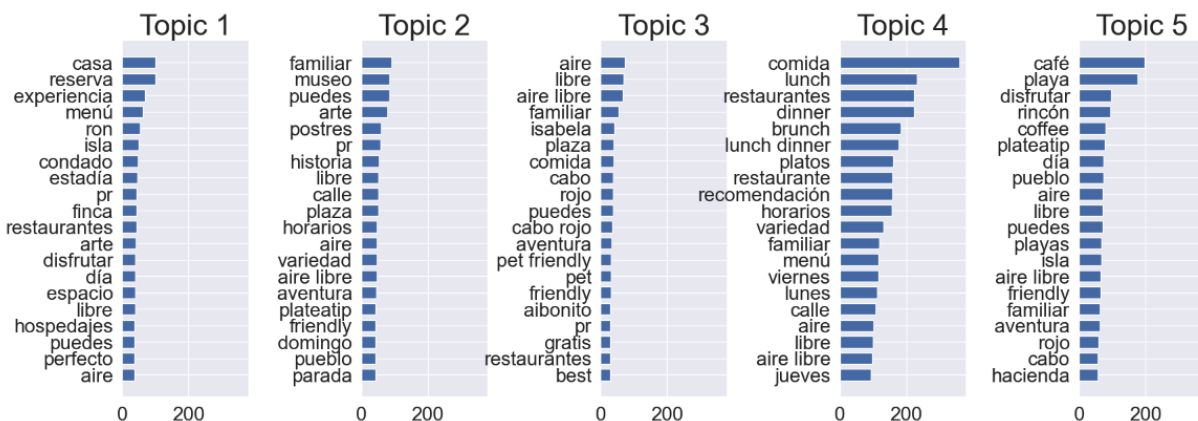


Figure 1. Top words in topics

Topic 1	Featuring house, reserve, experience, stay. A broader topic could be <b>“Trips”</b> .
Topic 2	Featuring the words familiar, museum, art. Since this topic had words that did not make a lot of sense together, we proceeded to investigate titles of the articles on this cluster. Based on that we concluded a good topic could be <b>“Timely content”</b>
Topic 3	Featuring outdoors 3 times, family A broader topic could be <b>“Outdoors”</b>
Topic 4	Featuring the words food, lunch, restaurants, dinner. A broader topic would be <b>“Restaurants”</b>
Topic 5	Featuring coffee (spanish and english), beach. A broader topic could be <b>“Coffee”</b>

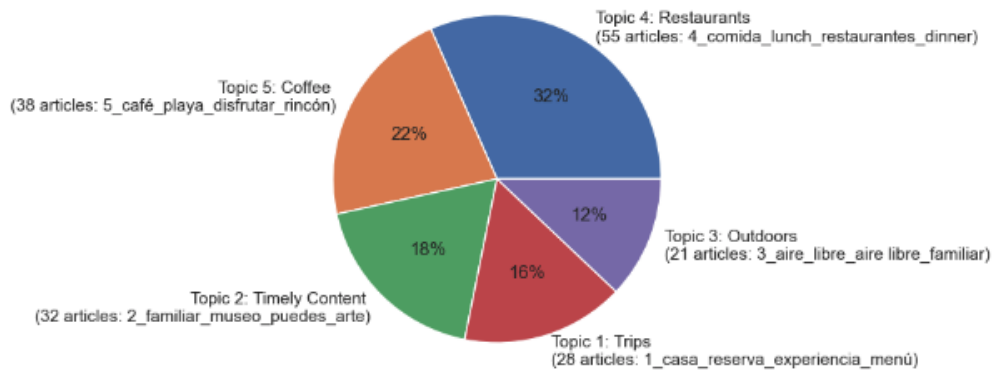


Figure 2. Articles by Topics

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

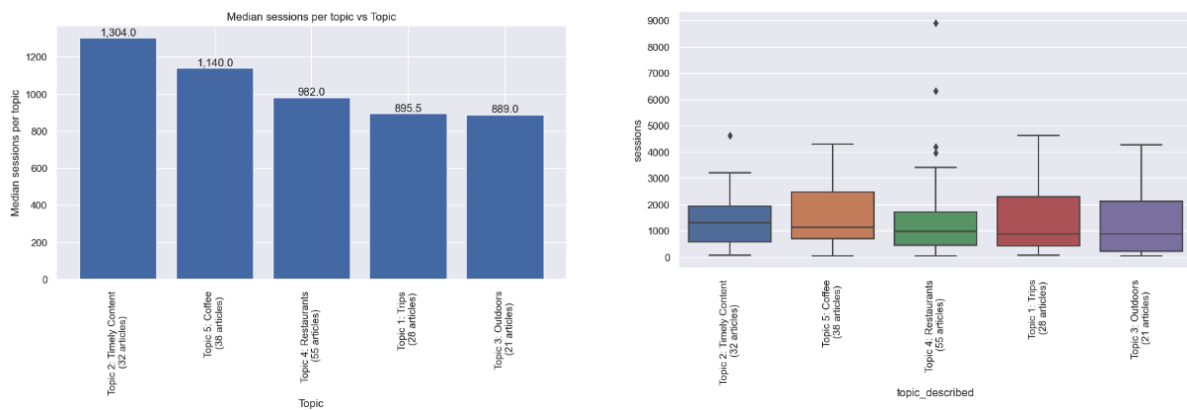


Figure 3. Median sessions per topic

To get a measure of impact we are using the median number of sessions per article on that topic. The topic **Timely content** followed by **Coffee** 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 that may be **timely content** or related to **coffee** topics, as they bring more sessions per article. Note that in the cleaning process the series of “What to do next weekend” was removed, as the article is updated weekly that made it an outlier. This implies there is effort already in place to generate timely content. Therefore the opportunity would be to focus on the generation of articles similar to the ones under the coffee cluster. More details on what are those on the jupyter notebook for this project.