

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 brand blog. Possibly identifying topics that made more impact have less articles highlighting topics of opportunities.

Procedure

The procedure for this is to get all the articles in PlateaPR blog using Latent Dirichlet Allocation (LDA) topic model to cluster the topics associated with the articles. In order to implement this some articles/pages were removed, ie, Terms and conditions, What is Platea?, and Fiona Hurricane associated articles because they do not represent the content Platea'sPR typically covers. The resulting articles' content were used to train the LDA model with the goal of getting 5 general topics of the data.

Results

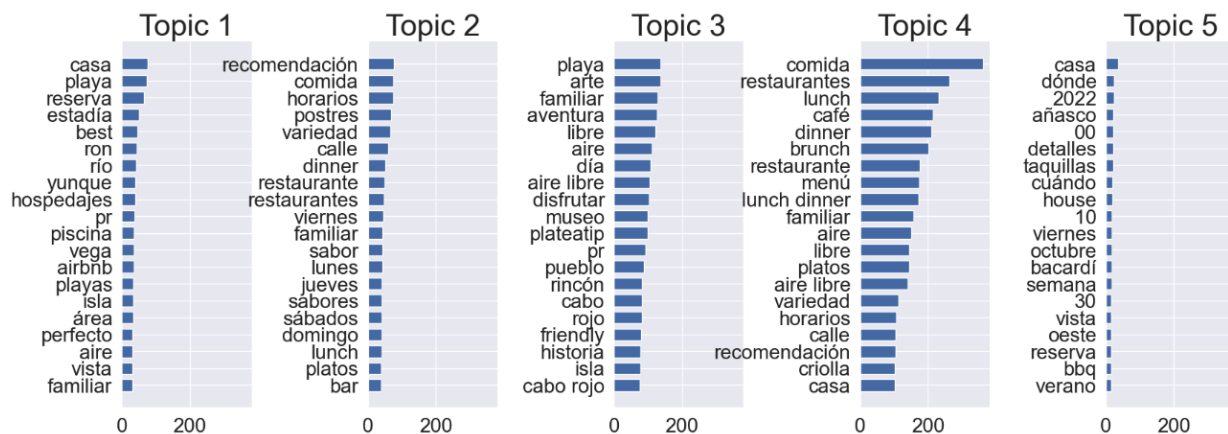


Figure 1. Top words in topics

Topic 1	Featuring house, beach, reserve, stay. A broader topic could be "Trips".
Topic 2	Featuring recommendation, food, hours, desserts. A broader topic could be "Food/Desserts"
Topic 3	Featuring beach, art, family, adventures, outdoors. A broader topic could be "Family/Adventures"
Topic 4	Featuring food, restaurants, lunch, coffee dinner "Food/Coffee"
Topic 5	Featuring the words menu, rum, house, food, bacardi (a rum brand). Since this one had words that did not make a lot of sense together, we proceeded to investigate the topic to see a trend. Based on that we concluded a good topic could be "Timely content"

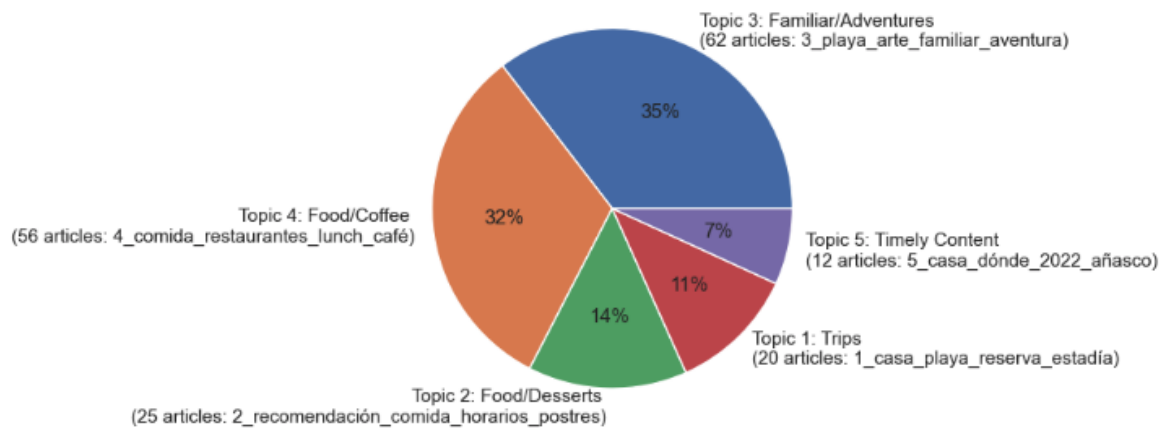


Figure 2. Articles by Topics

The group with the majority of articles is the group that we Familiar Adventures which can be represent things to do as a family in Puerto Rico

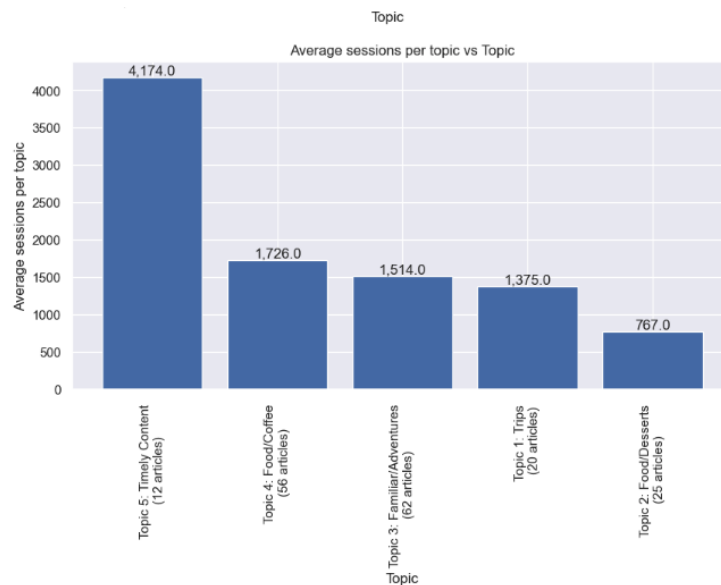


Figure 3. Average sessions per topic

To get a fair measure of impact we are using the average number of sessions per article on that topic. This value is calculated by adding up all the sessions of the articles with a given topic and then dividing that up by the number of articles with the topic. We can see that the topic Timely content seems to have the most impact by article.

Recommendations

Based on results we can recommend to increase focus or importance on content that may be timely. Since this content is bringing more sessions per article. For example, what to do on the weekend (which we already do), summer, thanksgiving, and any other holidays.