

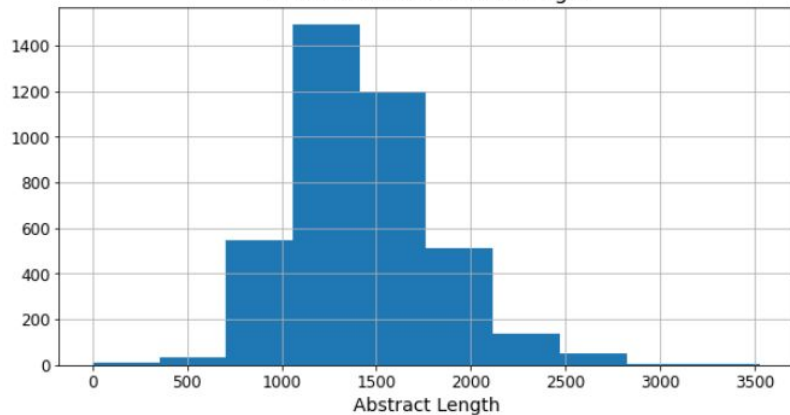


Information Extraction from Food Journals

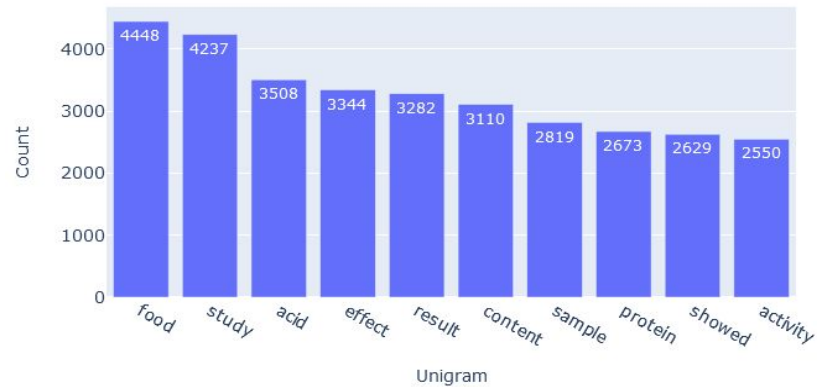
Deepak Tripathi

Abstract Stats

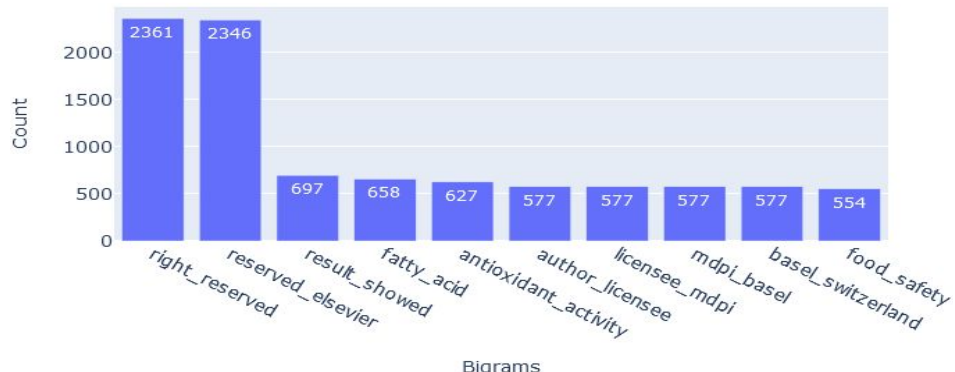
Distribution of Abstract Length



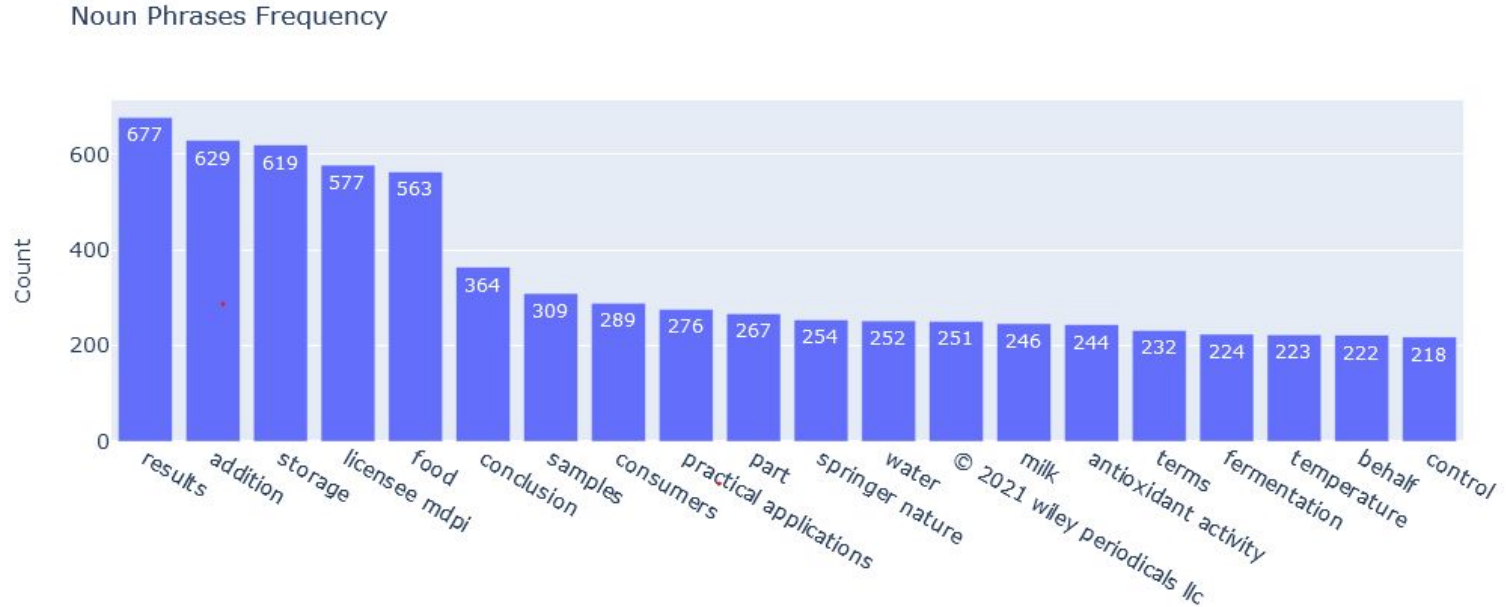
Unigram Frequency



Bigrams Frequency



Noun Phrases in Abstract



Topic Modelling Using LDA

Topic Count

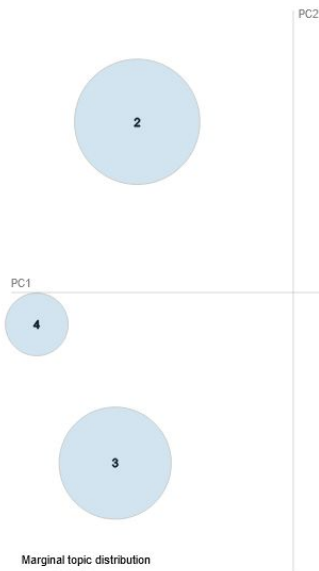
Selected Topic: 0

Slide to adjust relevance metric:⁽²⁾

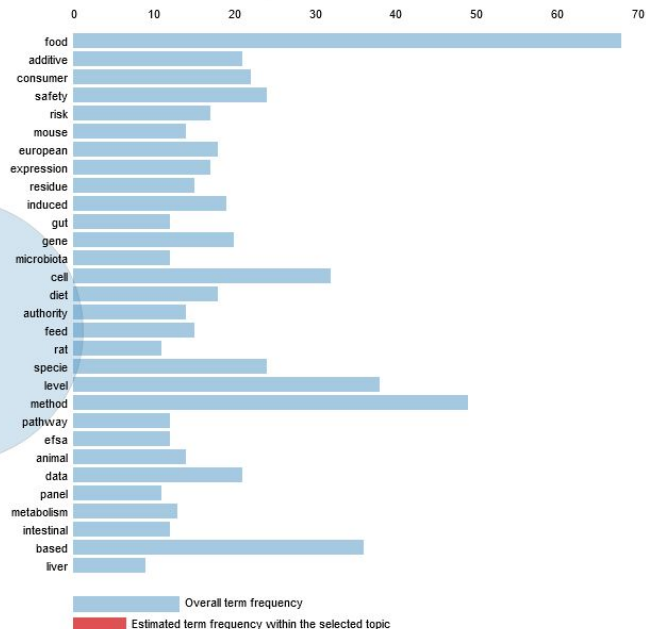
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



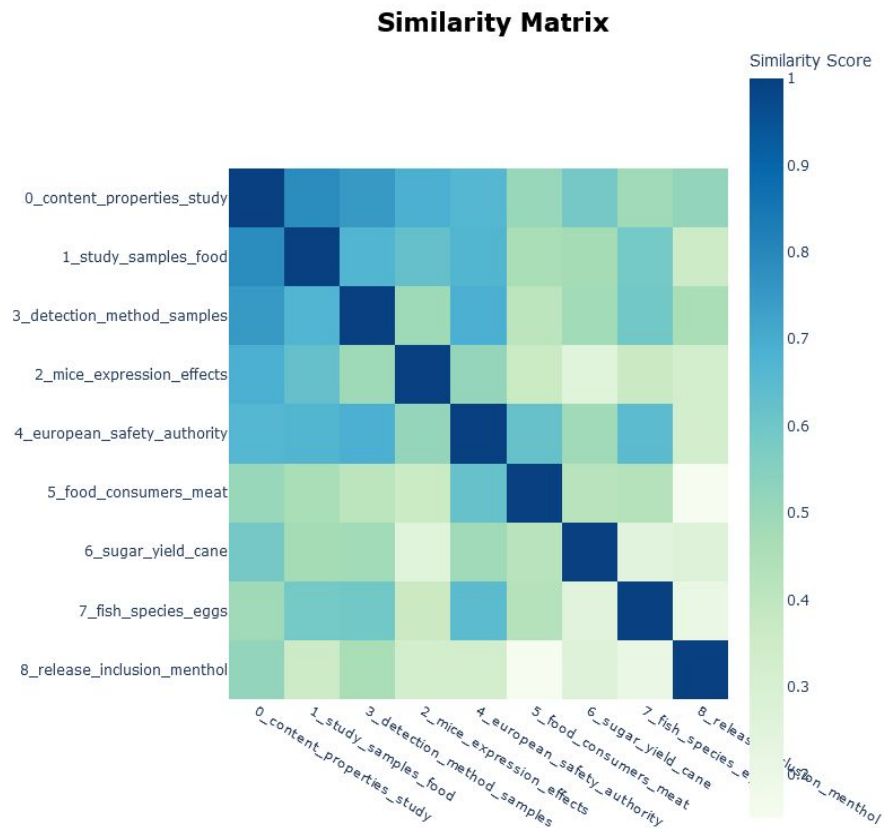
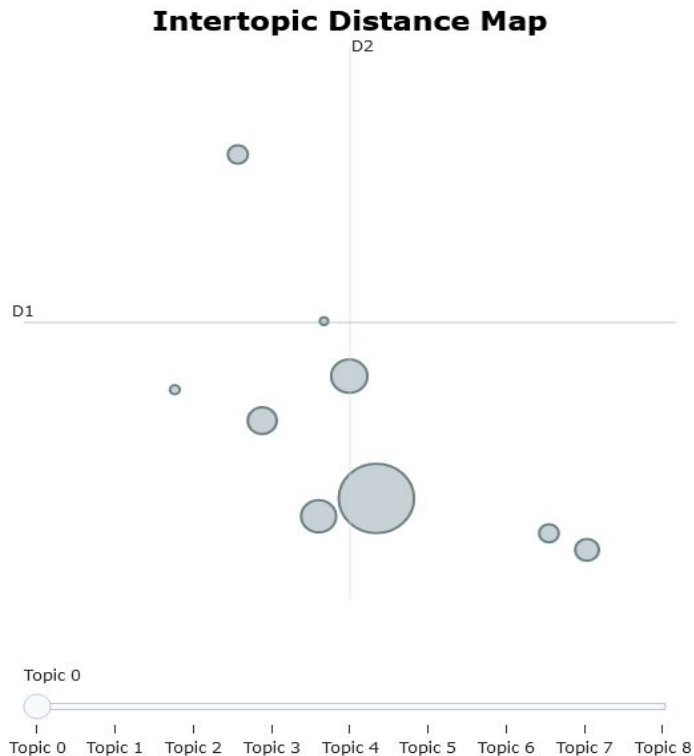
Top-30 Most Salient Terms¹



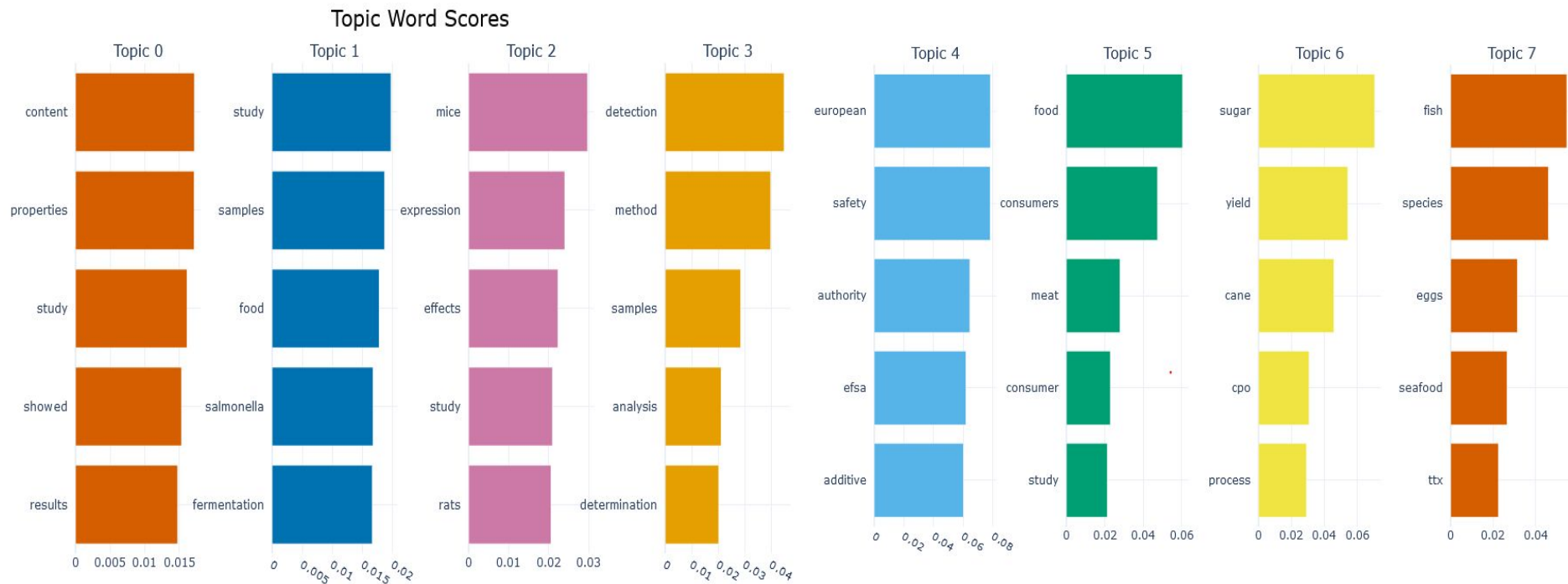
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t ; see Chuang et. al (2012)

2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Examples

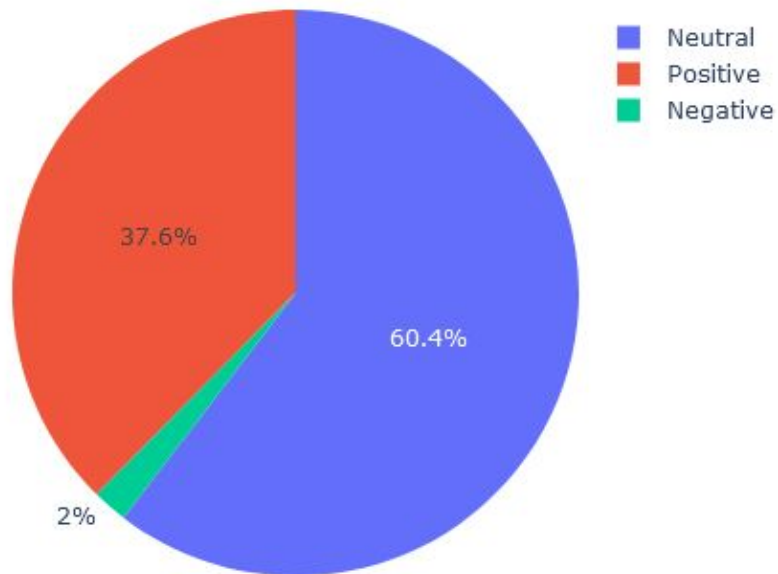


Topic Keywords



Sentiment Analysis

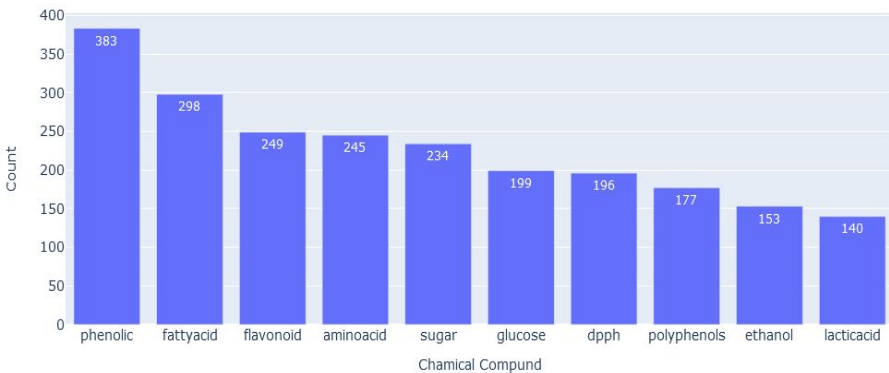
Sentiments Frequency



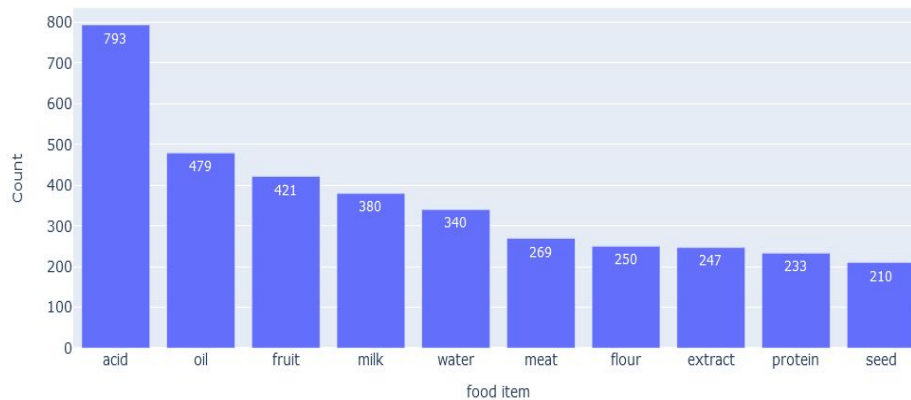
Examples

Named Entity Recognition

Chemical Compound Frequency

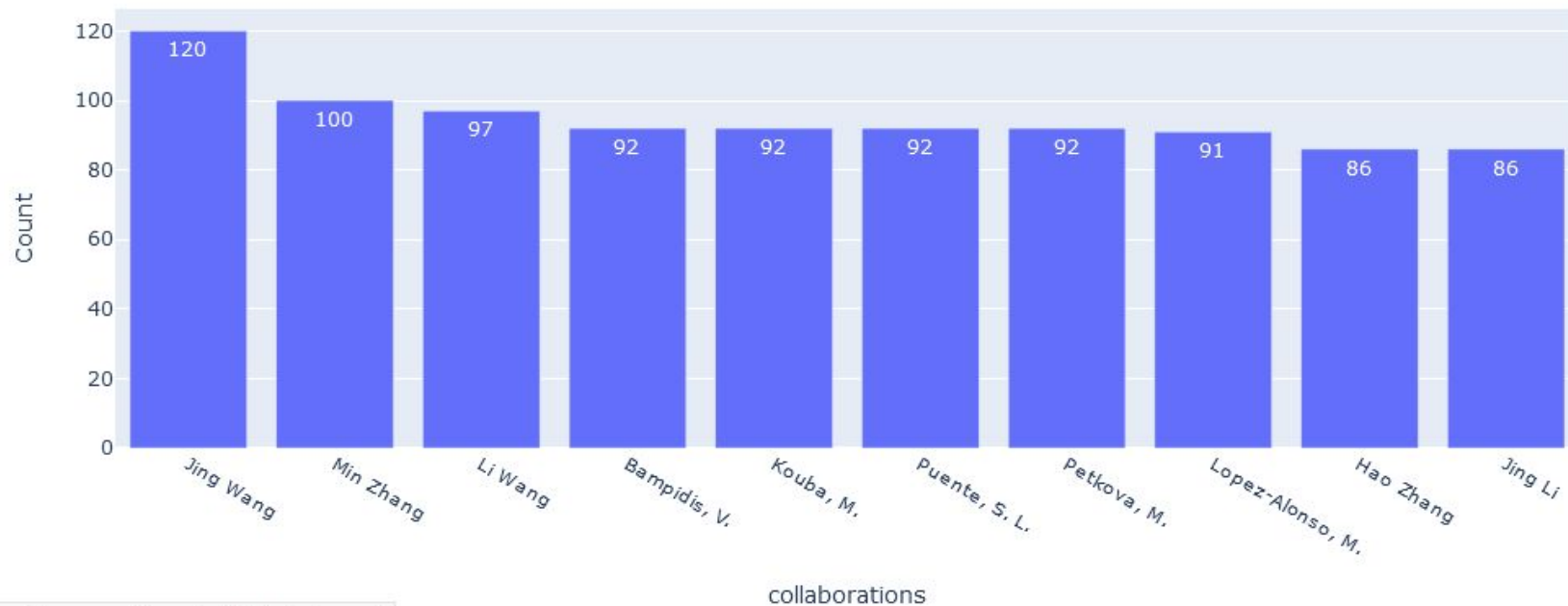


food item Frequency



Collaborations

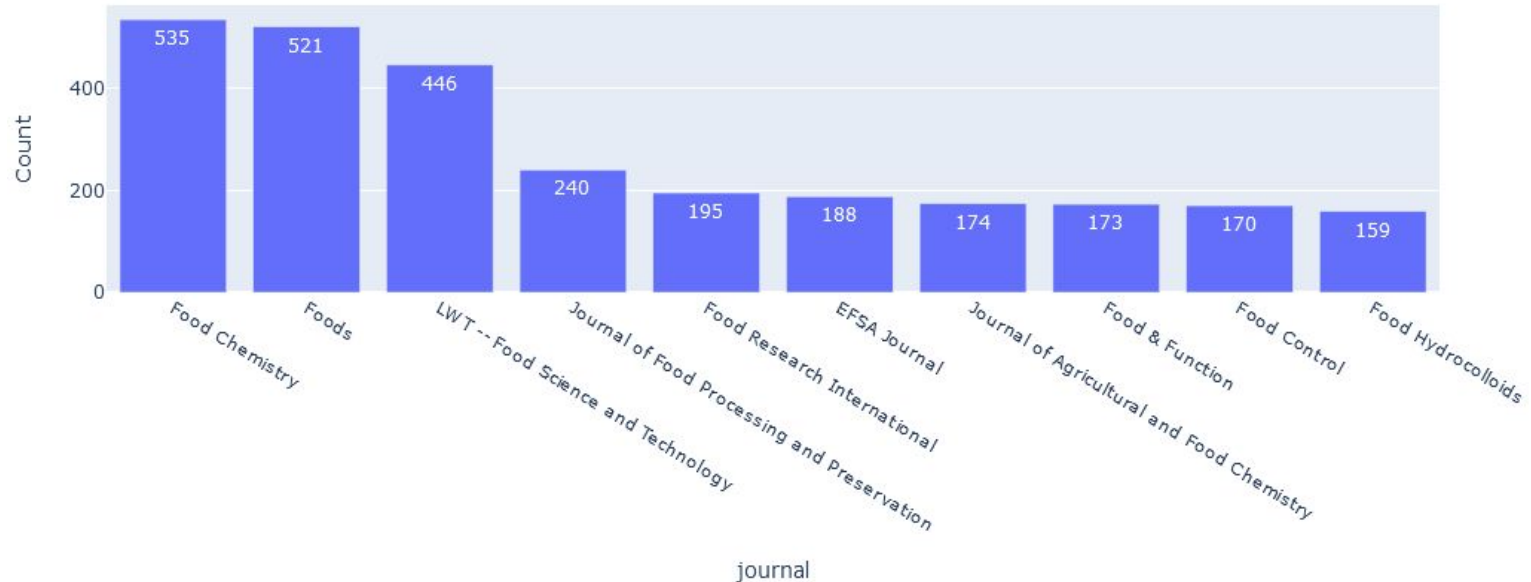
collaborations Frequency



Articles by Journals



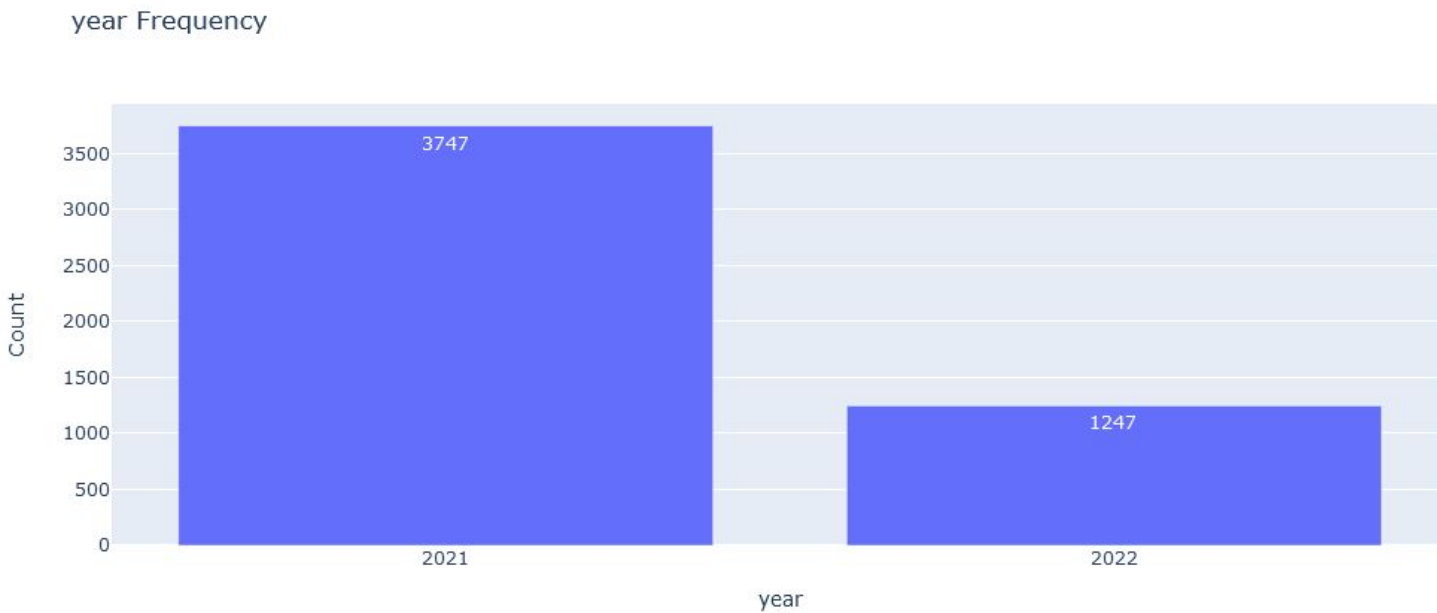
journal Frequency



Top Authors



Articles by year





Thank You