

Data Science Foundations

Master in Big Data Solutions 2017-2018



Liana Napalkova
liana.napalkova@bts.tech

Ludovico Boratto
ludovico.boratto@bts.tech

Francisco Gutierrez
francisco.gutierrez@bts.tech

Session 7

Assignments

Assignments

1. Given the lyrics you selected in your previous homework, compute their similarity.
2. Improve the exercise by first removing the stop words and by stemming the lyrics. Recompute the similarity. Did anything change, apart from the values (e.g., the two most similar songs changed)? Write a markdown with some comments.
3. Run the “Word association” task on the `treebank` corpus. Find the most co-occurrent word of the following:
 - chairman
 - profit
 - president
 - investors
 - city
 - chiefWhich considerations can you make about “chief”?