LDA Topic Modelling

<u>My Output:</u>

[(0, '0.344*"pike" + 0.209*"bass" + 0.140*"horn" + 0.135*"deep" + 0.121*"catapult" + 0.052*"tuba"') – Corresponds to **Topic 2** of Beta Vector because probabilities for words deep, tuba and bass match closely with beta vectors of these words for topic 2.

(1, '0.304*"deep" + 0.280*"pike" + 0.249*"bass" + 0.073*"catapult" + 0.063*"horn" + 0.031*"tuba"') – Corresponds to **Topic 1** of Beta Vector because tube, horn and catapult have probabilities close to 0 matching topic 1 closely.

(2, '0.230*"bass" + 0.202*"horn" + 0.171*"deep" + 0.163*"pike" + 0.129*"tuba" + 0.106*"catapult"')] – Corresponds to **Topic 3** of Beta Vector because bass and horn are both present with high probabilities matching topic 3 beta vectors for these words.

There is some noise but that is expected. This is the closest the above output matches to the beta vectors below.

<u>Beta Vectors:</u>

Word Order: "bass", "pike", "deep", "tuba", "horn", "catapult"

[0.4, 0.4, 0.2, 0.0, 0.0, 0.0] – Topic 1
[0.0, 0.3, 0.1, 0.0, 0.3, 0.3] – Topic 2
[0.3, 0.0, 0.2, 0.3, 0.2, 0.0] – Topic 3