TABLE SHOWING HYPERPARAMETER VALUES, COHERENCE SCORE AND COMMENTS ON LDA VISUALIZATION/OBSERVATION

	Num_ of _Topics	Alpha	Eta	Coherence score	LDA visualization observation/reason why I chose hyperparameter values
Baseline parameters	7	auto	0.61	0.3993986621605316	We notice an overlap in topic 7 and 8 so we will try changing values for each of our parameters to see if we can obtain a cleaner visualization. We wouldn't want overlaps but rather have our topics spaced out
Case 1	7	symmetric	0.61	0.4040016372690112	For the first case I decided to maintain the eta and num_of_topics parameters but change the alpha parameter to see if alpha will have a big effect on our results. We notice an overlap with topics 4 and 8 as well as topics 1 and 3. Also the coherence score wasn't much affected so lets reduce the number of topics and see if we solve the overlap issue or get better coherence score
Case 2	5	auto	0.61	0.4137225588970958	Reducing the number of topics improves on the visualization. We do not have any overlap and the coherence score changes only by a %. Let's maintain these values but for the eta value. I will reduce it drastically and see how it affects our model
Case 3	5	auto	0.01	0.42416637391739564	Reducing the eta value from 0.61 to 0.01 might not be a good idea as we can notice a slight overlap between topics 4 and 5. So lets increase the eta value and see what happens
Case 4	6	auto	0.91	0.4595378315330532	We can observe an increase in the coherence score but Increasing the eta value causes overlap in topic 1 and 3 as well as topic 5 and 6.
Case 5	6	asymmetric	0.61	0.4334035307528133	Using alpha as asymmetric, we notice a better spacing in the LDA visualization and no overlap. Topic 1 and 2 as the largest topics with highest percentage of tokens. Let's maintain alpha as asymmetric and number of topics as 6

					and rather reduce eta to see if we get better results for the next case
Case 6	7	asymmetric	0.91	4845389360239792	Highest coherence score I have seen so far which was as a result of increasing the eta value and leaving alpha as asymmetric. Topic 1 carries 54% of the tokens and we notice an overlap with three topics 5,6,7. From here I can now see how asymmetric value for alpha and high values of eta are affecting our model
Case 7	10	asymmetric	0.91	0.41328583095727217	Topic 1 covers most of the tokens, topic 5,6,7,8,9,10 overlaps. Let's try these same parameters but this time put alpha as auto to see if we get some improvement
Case 8	10	auto	0.91	0.46061487147544977	Topic 1 covering over 60% of the tokens and a lot of overlap with topics 5,6,7,8,9,10. However the coherence score was better. Let's reduce the number of topics and play around with the eta values while maintaining alpha as auto
Case 9	8	auto	0.71	0.41796292329158197	This is not a good one, there is a lot of overlap in almost all the topics except topic 1 and 4.
Case 10	6	asymmetric	1.0	0.5000019887084983	I decided to bump the eta value to a 1.0 and we notice a great increase in the coherence score but for the lda visualization, we notice there is an overlap with topic 5 and 6. So let's play around with the alpha value while maintaining the eta value at 1.0 to see if we get any improvement
Case 11	6	auto	1.0	0.42359885655690394	Changing alpha to auto removes the overlap in topic 5 and 6 but overlaps topic 4 and 3. Let's change alpha again while maintaining the same values for the other parameters and see if it makes any difference
Case 12	6	symmetric	1.0	0.4107871606381357	Changing the value of alpha to symmetric kind of makes it worse as we seeing the topics closer to each other and an overlap with topics 4,5 and 6. The coherence score is not that different from when we had alpha as symmetric so I can conclude that with this dataset, alpha as

					asymmetric produces a better visualization
Case 13	7	asymmetric	1.2	0.40146815011602827	Increasing the value of eta makes it worse as we notice topic 1 covering almost all the tokens and 4topics overlapping with very small percentages. Since I got a better coherent score when eta was at 1.0, I will change the value back to 1.0 and rather change the number of topics parameter
Case 14	5	asymmetric	1.0	0.3635594493196759	Since I noticed from my previous parameter alterations that when I set eta to a higher value but not too high precisely 1.0 and alpha as asymmetric, I see some good results I decided to reduce the number of topics to see if I get improvements on my visualization and we can see here that , changing the number of topics to 5 and using the values stated above, we notice we are getting the lowest coherent score from all my runs but the lda visualization looks good, no overlaps and with topic 1 covering 62% of the tokens