

ASSIGNMENT 2

1. In this assignment, I have narrowed down the insights and findings for a particular stock. Therefore, instead of utilizing the entire dataset, I've taken a subset for the stock of "ABX" and constructed topic modeling around that.
2. The number of topics was modified to 10 and the number of words was changed to 6 for the first iteration of the LDA model.
3. The final iteration of topic modeling has been set to a total of 20 words, i.e we can see the weight of 20 topics that dominate a document.
4. **Finding 1:** There were a couple of keywords that dominated each document such as: "gold", "mining", and "Barrick".
5. **Insight 1:** The first insight derived on the basis of finding 1 was that the stock ABX was related to either the price of gold, or mining, or the word "Barrick".
6. **Finding 2:** Based on the output of the weight of each topic in the document we can again notice a pattern of these three keywords weighing the highest in most of the documents. (gold- approx 70%, barrick approx 30% and mine/mining approx 20% in most of the documents).
7. **Insight 2:** Based on research we find, that the stock "ABX" belongs to a gold mining company called "Barrick Gold Corp" in Canada, which ties back to the three topics that had the highest weightage in each of our documents. Link: <https://www.barrick.com/English/home/default.aspx>
8. **Finding 3:** The word "newmont" and "merger" stand out as well, with a low weightage of 0.011 and 0.007 respectively, but a significant frequency in some documents to draw curiosity.
9. **Insight 3:** Based on research between the words "Barrick", "Newmont" and "Merger", I found a research article that stated that Barrick and Newmont were mining companies that did a joint venture together. Link: <https://www.newmont.com/investors/news-release/news-details/2019/Barrick-and-Newmont-Forge-Nevada-Joint-Venture-Agreement/default.aspx>. This shows how topic modelling can help us understand real-world aspects and help draw current trends that are current in an industry.
10. **Finding 4:** We can notice also the word "Tanzania" having a weight of 0.005 which can help us further understand whether there are any gold mining opportunities in Tanzania or if many people from Tanzania have purchased this stock.
11. **Insight 4:** Based on further research, I realized that the company Barrick Gold had plans of expanding mining in Tanzania. This again shows a strong correlation between the real world and our topics. Link: <https://www.reuters.com/article/us-barrick-gold-tanzania-idUSKBN2781YJ>.

Hence, to sum up, the topic modeling was channelized towards the stock "ABX" which on the basis of our results and research was found to belong to the Barrick Gold company in Canada, which we can observe is one of our prominent topics with high weightage in our outputs. In our findings, we mainly observe the occurrences of prominent keywords such as "gold", "Barrick", and "mining", etc which we tie back to real-world insights that state that Barrick is a gold mining company. This shows us how topic modelling can be utilized to not only understand how topics dominate in a corpus but also can be utilized to derive trends, insights and further understand future opportunities in an industry.