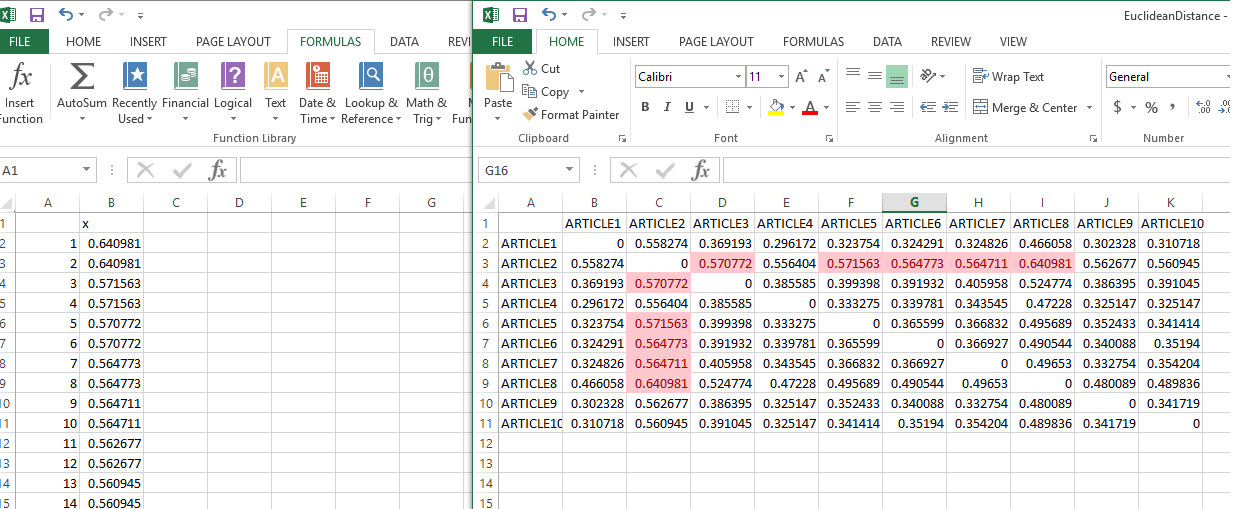
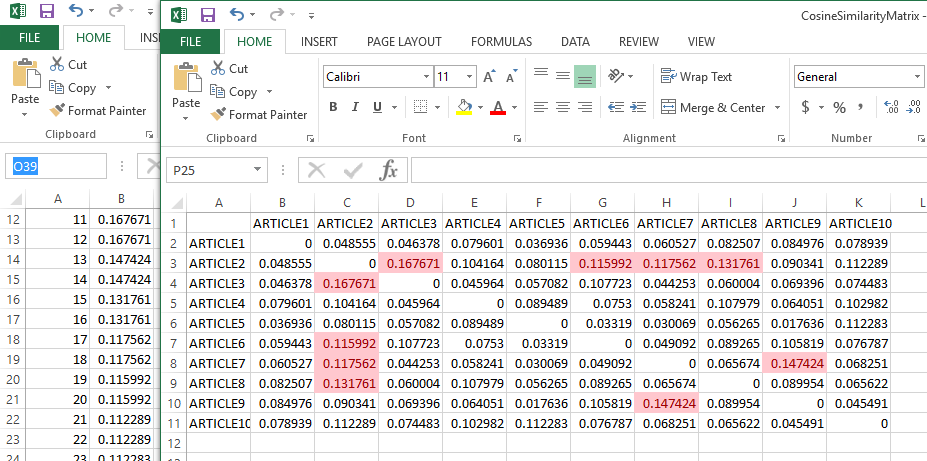
Assignment 1-

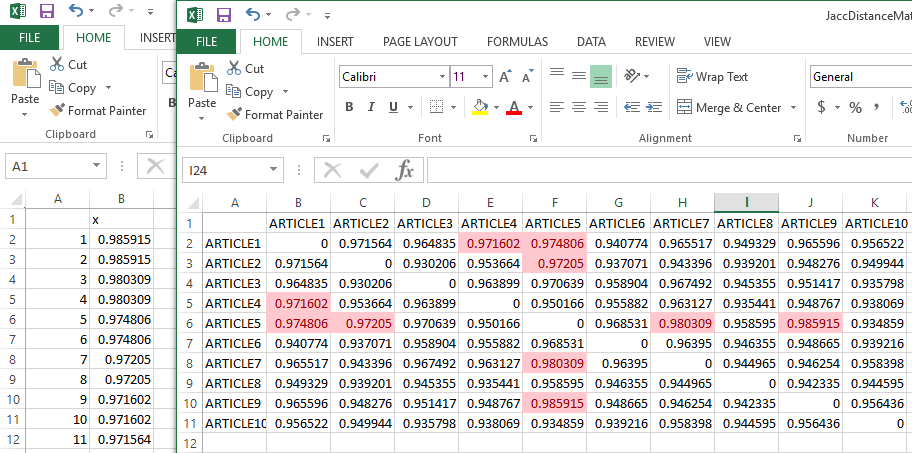
Euclidean Distance: Looking at the Euclidean Distance data matrix in Excel the top 10% can be highlighted. From the sorted Euclidean Distance matrix the highest value is clearly shown to be the first value, .640981, between Article 8 and Article 2. The remaining highest percentage relationships are among these, (Article 5, Article 2), (Article 2, Article 3), (Article 6, Article 2), (Article 7, Article 2). From this it can be shown that Article 2 is found in each of the top 10% relationships.



Cosine Similarity: When looking at the cosine similarity matrix, (Article2, Article3), had the highest cosine distance value. The remaining articles in the top 10% were, (Article9, Article7), (Article8, Article2), (Article7, Article2), (Article6, Article2).



**Jaccard Distance: Using the Jaccard Distance, (Article9, Article5), had the highest value. The remaining pairs in the top 10% were, (Article5, Article7), (Article1, Article5), (Article2, Article7), (Article4, Article1).**



**In summary, the cosine similarity looks to be the most accurate. The Euclidean Distance looks as though it may have been effected by outliers, being that Article 2 was present in all the highest relationships. Then the top Jaccard pairs contained very different articles than the other two distances.**