DAM Assignment 3– Analysis of Unstructured Data

Part A: The Anatomy of an Unknown Corpus (Weightage; 25%)

Task Description

Buried deep in your manager's computer is a directory helpfully named "**docs**." He has no recollection of what they're about and is disinclined to wade through them himself. Knowing of your interest in data analysis, he sends you the folder with the challenge to "provide insights into the contents and themes of the documents in the directory" (words excerpted verbatim from his email). As it happens, last weekend you learnt a bunch of techniques that could help you do just that. So you accept the challenge, pretty confident that you can deliver.

As you are about to start your analysis, you realise that there could be more at stake here than just a one-off data challenge. If you do a good job, your manager may even keep you in mind when he discusses that new data science role with the CIO. The task thus presents an opportunity to highlight not just your analytical skills but also your ability to weave a narrative around your findings, supported by appropriate visualisations.

Deliverables

- A written report including visualisations and highlights. **Length**: 1000-1500 words.
- Working code in separate file.

Due Date

11:59 pm on Sunday, Jun 7th.

Part B: Blog (Weightage; 5%)

Task Description

Write a **reflective** blog post on CIC Around on the use of the techniques of text analysis in the workplace. For example: what can it be used for and what are its implications? Tag your post with "**DAM-AUT-2020**"

Deliverables

Blog on CIC Around. **Length**: 750-1000 words.

Due Date

11:59 pm on Sunday Jun 7th.