Report Template – DCPP Group Assignment :

* **Executive Summary:** Problem Statement, Proposed Solution, Brief understanding of the challenges (1 page)
* **The chosen domain and Seed sources:** What analysis was conducted in this part? Why did you make those choices?
* **The structured and unstructured sources from open domain/ internal sources:** What are the sources chosen and explain the reasoning behind the choices.
* **Download/ crawl/ collect data from all the sources:** What are the methods used for this process? What are the challenges and how did you overcome them?
* **Convert data from original sources (Webpages, pdf files, CSV files, …) to structured data fields:** Explain the process chosen and what are all the challenges taken?
* **Data cleaning/pre-processing as needed**: What data cleaning/ pre-processing techniques were taken up for this stage and why did you feel the requirement for it?
* **Observations/ Insights and Analysis on the data collected**
* **Strategy to enhance the data with crowd sourcing methods:** What strategy did you implement to improve on the crowd sourcing methods and why?
* **References and Sources used for this Assignment**