**Project Milestone 4 Individual Report:**

As a team, we strategically divided tasks to streamline our efforts. Recognizing the complexity of configuring software for data ingestion and cleaning, my focus, alongside another team member, involved finalizing columns for algorithm development and integrating Hive with Pyspark. Specifically, I concentrated on formulating k-means algorithms and employing the elbow method to determine optimal k-values using sample data and specified columns. Subsequently, my code seamlessly transitioned to Gaofeng, who incorporated it for page rank. The integration of the first pipeline encompassed algorithmic refinement and visualization, culminating in a unified body of work. For our project report, we collaboratively documented challenges faced, with me taking the lead in formatting and structuring the content within a Google document. The process ensured equitable participation and a coherent presentation of our collective efforts.