**DATASET CREATION -**

* First, I created the dataset manually using the tweets from twitter dataset comprising of three categories i.e., TheSocialDilemma, FIFA World Cup and Religion.
* It consists of various columns like texts which are actually tweets, hashtags, user name, user location, etc.
* Since, Unsupervised Learning is asked in the task, I did not provide a label column.

**CODE –**

* I used only text and hashtag column to train the data, since other columns are not so relevant.
* Since the text and hashtag column contains text data, so NLP is used to clean the text and then create Bag of Words Model. Then both the dataframes are combined.
* Since there are no labels in the dataset, so Unsupervised Learning Algorithm or KMeans Clustering Algorithm is used.
* The number of clusters to be used is taken care of by Elbow Method.
* This is then used to train the data and form the clusters.
* The Cluster Number is then appended to the dataframe.
* The results are somewhat relatable.
* It can be further improved by using Word Embedding Layer.
* The count of cluster number will help in the finding the percentage of a given label from the whole dataset and thus we will be able to provide ranking to it based on the percentage.