

**Final Dataset:**

Final dataset contains a scaled-up dataset of up to 100 videos with the same set of variables which we maintained for the initial dataset. This is the dataset which we used for both Unsupervised and Semi supervised topic modelling whereas Unsupervised modelling takes into consideration the variable “topics” whereas the semi supervised modelling does not take this into consideration, Bag of words is the input for this modelling.