# Report on Twitter Question 1

This report analyzes twitter dataset question 1 where we need to find the hour of the day PrezOno tweets the most on average. We use Hadoop framework to analyze the trends of twitter like finding statistics of the hour of the day most tweets are tweeted by a particular user.

The MapReduce framework is used to code this. I used the key-value approach to find the summation of all tweets tweeted for that hour. The data is fetched through hdfs and its patterns were carefully observed. It was analyzed that the total number of tweets by PrezOno was 341 for the given year.

The first part of this implementation was analyzing the collection of tweets over the time. The second part was to obtain the key value pairs and map them so that the keys are distributed in the reducer. The mapper assigns input key value, then its run for each key value, then shuffling pattern approach is used and finally the reducer is run for each key value. Finally the code was run on cluster by Hadoop streaming.

The output file obtained helped me to plot the graph of the expected number of tweets for each hour of the day, for those PrezOno tweeted.

.     Maximum number of tweets: **26** for the **17th** hour