Fonseca, Menor, Paul
CS 129.1 B

Final Project Report

A. Statement of the Problem

Twitter is one of the internet's most popular social media platforms, and is a convenient repository of data on news and trending topics. This research seeks to explore the effect of time on tweet performance. Focusing on X number of twitter users, with a combined total of Y tweets, this research aims to answer the question: Given a certain number of Twitter users, how can we find out what time of the day is most successful for them to tweet?

Tweet performance is measured through engagement - that is, favorites and retweets from other users. The study measures time in one hour intervals, thus, we will be measuring tweet performance the between 00:00h and 23:59h. Retweets are not considered in this study as they originate from a third party user.

Finding the optimal time frame for tweet engagement can have positive implications in the fields of online digital marketing. Posting at the optimal time essentially broadens one's post reach, leading to a higher probability for responses from a firm's target market. It is important to note however, that this study is not an exhaustive analysis of tweet performance, but rather a demonstration of the applications of big data in the real world.

B. Description of Source Dataset

We have chosen to use five public twitter accounts affiliated with Ateneo de Manila University, namely - @ateneodemanilau, @ADMUSanggu, @GetBlued, @TheGUIDON, and @TheGUIDONSports. These accounts were chosen because of their follower count, number of tweets, but also because they are the twitter platforms that which the Loyola Schools community engages with. We believe that these are fair approximations of the school community, with the administration, student government, fan pages, university publication, and the sports community represented.The data was gathered using a Python script with the Tweepy library for accessing the Twitter API.

C. Relation of the output to the Big Data problem

Our output gives data regarding the average number of favorites of a user's tweets depending on the time of the day. Then, this was also put through another MapReduce to give an overall assessment of which hour yields the most number of favorites among all the 5 users. This is relevant because the data tells the user which hour of the day they should post so that they can yield get the maximum amount of favorites. This may also mean that more people access the

said site on that specific hour interval and notifications or public service announcements of the school should be tweeted at this or these hour intervals.

D. Visualization of the output