**Twitter Analysis**

*Points: 100*

**Description:** For this project, you will be expected to capture a few thousand tweets on a topic of your choice, store these tweets in a SQL Server database, process them through Microsoft’s Cognitive Services Text Analytics API, and visualize your findings. Specifically, you need to accomplish the following:

* Design a database to store key information about your tweets, including the ID, the text of the tweet, the user who sent the tweet, when it was sent, and the location that was tracked (if any).
* The database also needs to store the sentiment score and key phrases that are returned from the Microsoft API’s
* Using Python and either the Tweepy library or the actual Twitter API, collect and store 2000 to 3000 tweets. You do not want duplicate tweets or retweets.
* Using Python and pyodbc, store your Tweets in your SQL Server database. You can either build stored procedures (preferred) or dynamic SQL (make sure no SQL Injection is possible).
* Format your Twitter data to be passed to both the Sentiment API and the Key Phrases API’s provided by Microsoft.
* Store the data returned from the Microsoft API’s in your database. Again, you can use stored procedures or dynamic SQL
* Visualize the data you’ve collected, using either Excel, Power BI, or Python libraries. Specifically, the following need to be accomplished:
  + Tag/Word Cloud of your key phrases
  + Some visualization that shows a breakdown of your sentiment scores – a pie chart, bar chart, etc.
  + Top five most negative locations, top five most positive locations
  + Top five users and the count of how many tweets you collected on each
  + The most positive tweet
  + The most negative tweet
  + Volume of tweets by day

Turn in a zipped up copy of your Python project, the script to create your database, the file that contains your completed visualizations (screenshots, Excel sheet, Power BI document, etc.).

**What you’ll need to complete this project**

* A Twitter account with an email address attached to it. This will need to be sent to me to be added to the development account.
* A Microsoft account with Azure credits. If you follow the directions on this link, you’ll be able to sign up without a credit card: <https://azure.microsoft.com/en-us/free/students/>
* A Python IDE of your choice. I will use PyCharm or Visual Studios in class
* If you are on a Mac and are not running a Windows Virtual Machine, you’ll need to install ODBC drivers. The instructions for this can be found at: <https://docs.microsoft.com/en-us/sql/connect/odbc/linux-mac/installing-the-microsoft-odbc-driver-for-sql-server?view=sql-server-2017>
* Your SQL Server username and password

Useful links

* Tag Cloud using Python: <https://www.datacamp.com/community/tutorials/wordcloud-python>
* MathPlotLib Tutorial for Python: <https://matplotlib.org/users/pyplot_tutorial.html>
* Excel Chart Tutorials: <https://www.microsoft.com/en-us/microsoft-365/blog/2012/05/30/our-eight-best-tutorials-on-excel-charts/>
* Power BI Basics: <https://www.youtube.com/watch?v=h6AIAxMEDiw>
* QuickStart for Cognitive Services API & Python: <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/quickstarts/python>

I’ll also provide a number of useful code snippets to help you through this.