Armand Post

12/05/16

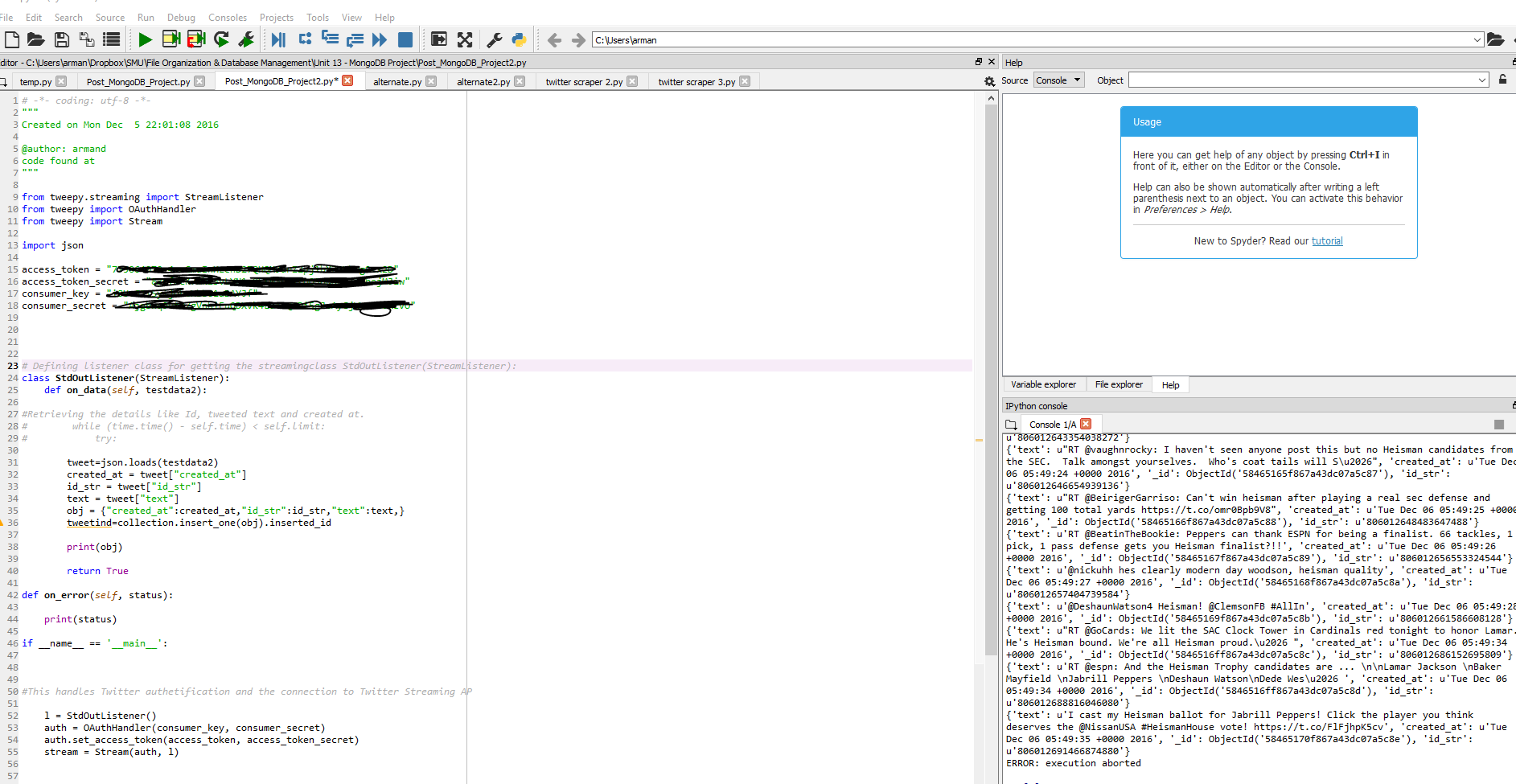
MongoDB Project

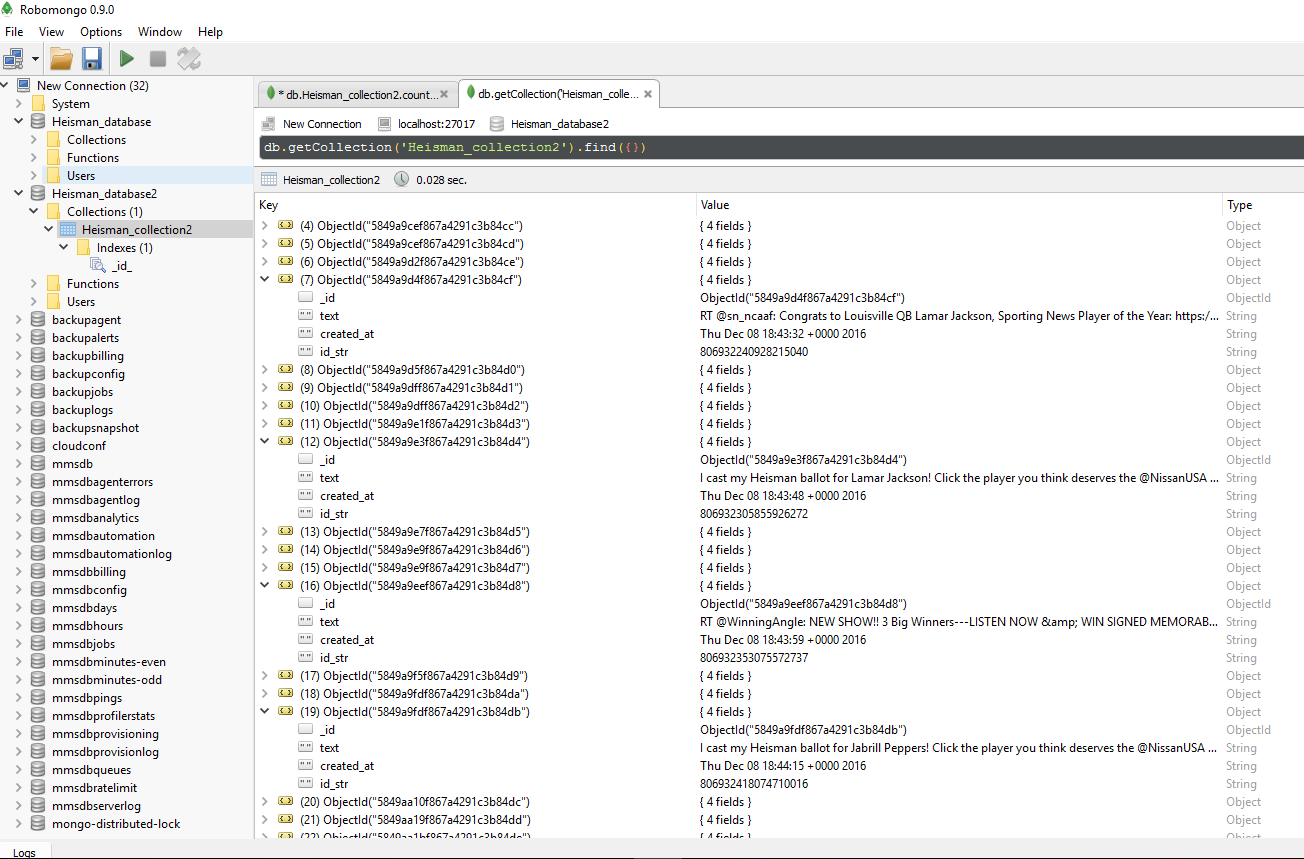
The purpose of this project was to collect tweets with the word “Heisman” and count the number of times the five finalists’ names showed up in the tweets to see if it could predict who wins the award on December 12th. The tweets were gather using the Tweepy package for Python and stored in a MongoDB database where they were analyzed. Tweets were collected on three different dates and compared to betting odds in Las Vegas. Summary statistics are below.



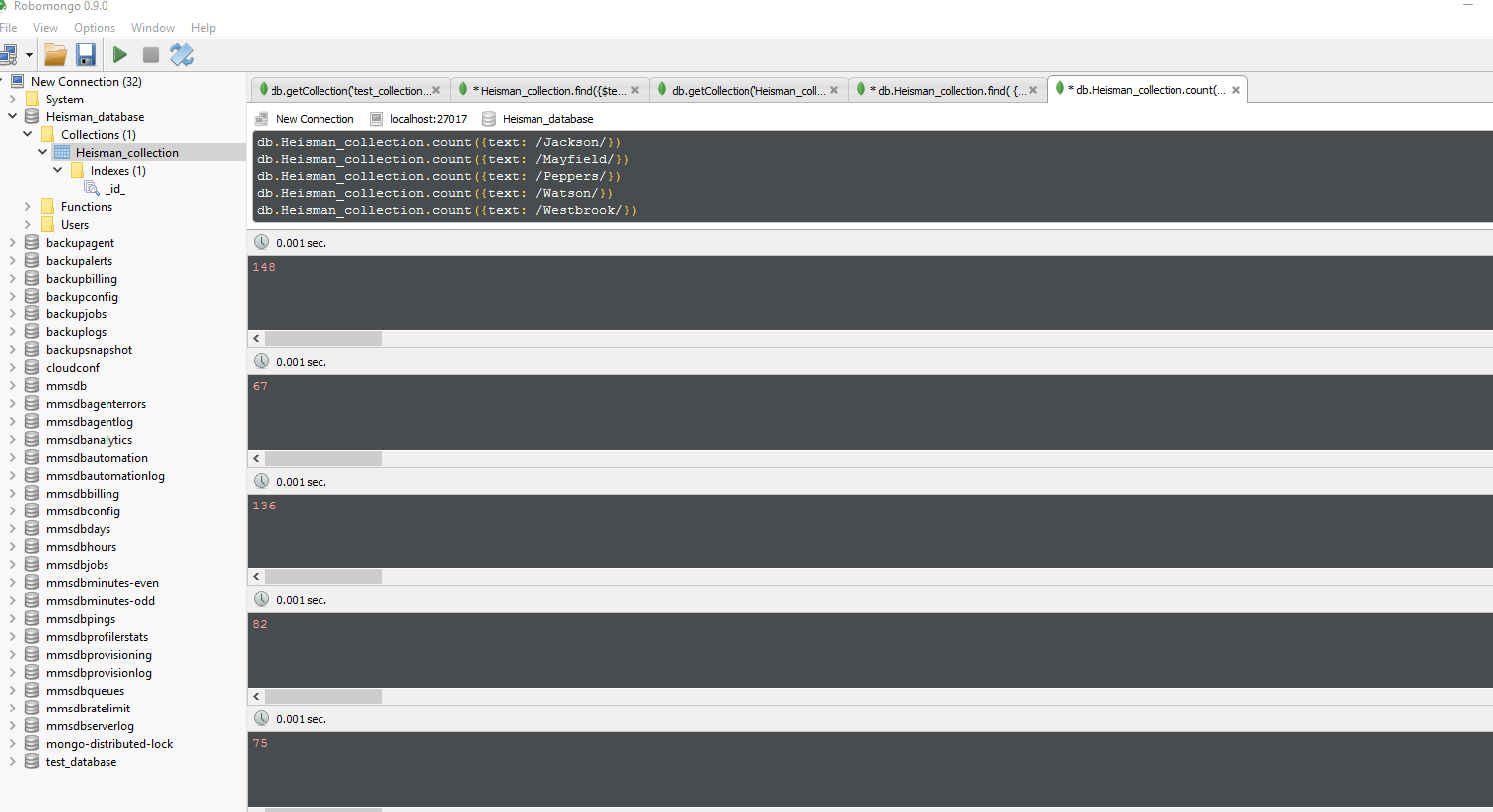
Analysis of the tweets yielded some interesting findings. On December 6th a former winner, Salaam, committed suicide and showed up in nearly half of the tweets. By the next day, Salaam showed up in a very small portion. The news of Salaam’s death didn’t skew the rank order of tweet volume with it remaining the same as the day before with number too narrowing the gap with number one. Surprisingly, by day three Peppers overtook Jackson for tweet volume. Jackson is a heavy favorite according to Vegas odds, with a $1 bet paying $0.40. Peppers on the other hand is a longshot with a $1 bet paying $50. At the time of this writing, the awards ceremony is still four days away, and despite Jackson being a heavy favorite, I think Peppers offers some enticing odds. However, the content of the tweets weren’t assessed past the occurrence of finalists’ names so the counts may not be a great indicator of the winner.

Screenshot 1 of console collecting tweets



Screenshot 2 of tweets stored in MongoDB

Screenshot of Analysis



# -\*- coding: utf-8 -\*-

"""

Created on Mon Dec 5 22:01:08 2016

@author: armand

code found at https://www.linkedin.com/pulse/collecting-twitter-stream-using-python-mongodb-shailendra

"""

from tweepy.streaming import StreamListener

from tweepy import OAuthHandler

from tweepy import Stream

import json

access\_token = ""

access\_token\_secret = ""

consumer\_key = ""

consumer\_secret = ""

# Defining listener class for getting the streamingclass StdOutListener(StreamListener):

class StdOutListener(StreamListener):

def on\_data(self, testdata2):

#Retrieving the details like Id, tweeted text and created at.

# while (time.time() - self.time) < self.limit:

# try:

tweet=json.loads(testdata2)

created\_at = tweet["created\_at"]

id\_str = tweet["id\_str"]

text = tweet["text"]

obj = {"created\_at":created\_at,"id\_str":id\_str,"text":text,}

tweetind=collection.insert\_one(obj).inserted\_id

print(obj)

return True

def on\_error(self, status):

print(status)

if \_\_name\_\_ == '\_\_main\_\_':

#This handles Twitter authetification and the connection to Twitter Streaming AP

l = StdOutListener()

auth = OAuthHandler(consumer\_key, consumer\_secret)

auth.set\_access\_token(access\_token, access\_token\_secret)

stream = Stream(auth, l)

# Below code is for making connection with mongoDB

from pymongo import MongoClient

client = MongoClient()

client = MongoClient('localhost', 27017)

db = client.Heisman\_database

collection = db.Heisman\_collection

#This line filter Twitter Streams to capture data by the keywords: ‘Heisman’

stream.filter(track=['Heisman'])