Section 2 – Data Processing

We've now added a loop to our code - everything tabbed in (lines 11-15, with white space at the start) is repeated with i taking all the values between 0 and 9, one after another.

Run the script, and make sure you understand how the code leads to the output you see in the console. Pay attention to everywhere i is used in the loop compared to what you see in the console.

Edit the code to display the likes for the 20 most recent tweets rather than 10.

For each tweet, display the number of retweets as well as the text and likes.

What does line 15 do? Try deleting it to check if you're not sure. Why is this useful?

Coding Challenge: Which account receives the highest mean number of likes and retweets?

If you know how to approach this problem, go ahead. If not:

To calculate the mean number of likes, we need two things –

- the number of tweets we're processing - this is set on line 6 of the example code
- the total number of likes those tweets received – this we have to calculate` and will store in the variable total_likes

Python gives us an instruction to add number b onto an existing variable a:

```
a += b
```

You should be able to use this inside your loop to total up likes. What are a and b in this case? Remember that each time the loop runs i increases by 1, starting from 0.

Don't forget that you need to add a tab at the start of the line to make sure your += statement is inside the loop (it should be level with the existing print statements).

Finally, print total_likes, tweets_to_process and the mean number of likes to the console. You will have to calculate the mean number of likes – the divide operator in python is /

Questions to Investigate:

- Who gets the most engagement calculated via mean likes?
- How many tweets should you be calculating your mean from to get an accurate number?
- You should be able to make a very small edit to this script to calculate the mean retweets rather than likes. Does the same person get the most retweets and the most likes?
- Is the proportional change the same (e.g. if one person gets twice as many mean likes, do they get twice as many mean retweets?)
- What factors might affect the levels of engagement? You can find more information about these two accounts on their twitter pages:
    - https://twitter.com/BarackObama
    - https://twitter.com/taylorswift13
- What does all of this tell us about people that follow these two people on twitter?

Extension Challenge 1: Rewrite your script so it always processes all of the tweets for an account without you hard coding this number in the script. One way to do this is with the len() function.

Extension Challenge 2: Write a script that loads both accounts, compares their levels of engagement and outputs to the console who gets the most engagement (along with the associated numbers)

Extension Challenge: Calculate the standard deviation instead of the mean (NOTE: replit gives you access to the statistics module - you don't have to write your own standard deviation function, although don't let me stop you). What more does this statistic tell us?