CS381 Data Analytic Quiz 6

Instruction: For multiple choice questions, clearly circle one of the choice; for all other questions, write your answer right below the questions. All questions carry the same weights.

Name:

Question 1: Fill in the blank with either "higher" or "lower" or "bigger" or "smaller" in the following statements

```
 1. The IDF of words in emails sent by the presidents of US should be
__higher_____  than the average IDF of words in emails sent by all US citizens.

 2. Speech from President Trump's should have a __lower___ IDF than that from
President Obama.

 3. When considering only all historical speeches from all the past Presidents of
United States, the avergae IDF from President Trump should be __lower____ than the
average when stops words are not excluded.

 4. When considering only all historical speeches from all the past Presidents of
United States, the set of stops word from President Trump should be _ bigger___
than that of the average from all the other presidents
```

Question 2: Write down the results from the statements.

```
  1. Results from running tozenizer of the sentence:  Good morning Americans, have
a great day!
```

Answer:

```
  2. Results from running stemming from the words:   jumping, jumped, jump
```

Answer: jump, jump, jump

```
  3. Results after removing stopwords of the sentence:  Cristiano Ronaldo was born
on February 5, 1985.
```

Answer: Cristiano Ronaldo born February 5, 1985.