

**START OF QUIZ**

**Student ID:**

**67446104,Nguyen,Minh**

## Question 1

Topic: Lecture 1

Source: Lecture 1

In class, we talked about how `.isdigit()` is insufficient for determining whether we can convert a string to a float. Write a short function `"isfloat"` that determines whether a provided string is a valid floating point number. (2)

## Question 2

Topic: Lecture 1

Source: Lecture 1

Vowels are often used as a proxy for syllables in words (it's not a perfect correspondence, but it's not bad). Write a function that counts the vowels in a word, without using a loop, using only the tools we went over in Lecture 1 (list comprehension counts as a loop). (2)

### Question 3

Topic: Lecture 4

Source: Lecture 4

Why does type-to-token ratio decrease as the size of the corpus increases? What does this suggest about long documents? (1)

## Question 4

Topic: Lecture 4

Source: Lecture 4

How would we sort a dictionary alphabetically by the reverse of its keys (assuming the keys are strings)? Write a short piece of code, and briefly explain your logic. (1)

## Question 5

Topic: Lecture 3

Source: Lecture 3

What is the Big O time complexity of finding the elements in a set that intersect with an iterable (ie, string, list, etc)? Briefly explain. (1)

## Question 6

Topic: Lecture 2

Source: Lecture 2

Do you think it's possible for a language not to follow a Zipfian curve? What consequences might that have on communication (if, let's say, if the curve were linear)? (2)

## Question 7

Topic: Lecture 3

Source: Lecture 3

Describe the concept of the "Minimum viable product", and how it relates to using lexicons.

(1)



## Question 8

Topic: Lecture 2

Source: Lecture 2

What are 2 benefits of providing a corpus reader with your corpus? (1)

## Question 9

Topic: Coding

Source: Lecture 2

Imagine we have a large corpus in an unknown language. We don't have any ML tools to analyze the data. How might we determine the stopwords in our corpus? How might we test our theory of stopwords? (I'll make it easy on you - the tokens are space separated, and we have some way of separating sentences.) (3)

**END OF QUIZ**