Using Spark-Scala complete following tasks

Tasks

1. Create RDD from file `1918NewYearHonours.txt`
   use `SparkContext.textFile()` method
2. Create RDD from file `ListOfAustralianTreaties.txt`
   use `SparkContext.textFile()` method
3. Tokenize (split string into words)
   String "1842 – Treaty 5 March 1856)[5]" should consists of following words: 1842, Treaty, 5, March, 1856, 5
4. Count words in RDD
   Given RDD[String]. You need tokenize it using method words() and count words
5. How many words are in ListOfAustralianTreaties.txt?
   Hint: use countWords() to count amount of words
6. How many words are in both .txt files?
   Hint: use countWords() to count amount of words
7. Transform RDD so that it should contain numbers only
   i.e. string "1842 – Treaty 5 March 1856)[5]" should consists of following numbers: 1842, 5, 1856, 5
8. How many unique numbers are in *ListOfAustralianTreaties.txt?*
9. Calculate average of all numbers in *ListOfAustralianTreaties.txt?*
   i.e. string "1842 – Treaty 5 March 1856)[5]» has average 927
10. Get word occurrences
    count how often each word repeats
11. What are 10 most frequent symbols in *ListOfAustralianTreaties.txt?*
12. Split word into 5 groups such as:
    Group 0: where D > 0 or A > 0 and B+C >0, name it “thrash”
    Group 1: where A > 0, name it “numbers”
    Group 2: where B == C, name it “balanced_words”
    Group 3: where B > C, name it “singing_words”
    Group 4: others, name it “grunting_words”
    Where
    A = Number of digits
    B = Number of vowels
    C = Number of consonants
    D = Number of other symbols
13. How many elements there are in each group in *ListOfAustralianTreaties.txt*
    a)
14. Print samples of each group with A, B, C, D values from *ListOfAustralianTreaties.txt*?