Bigdata Assignement 4.8

Create a list of tuples, where the 1st element of the tuple is an int and the second
element is a string.
Example - ((1, ‘alpha’), (2, ‘beta’), (3, ‘gamma’), (4, ‘zeta’), (5, ‘omega’))
- for the above list, print the numbers where the corresponding string length is 4
- find the average of all numbers, where the corresponding string contains alphabet ‘m’
or alphabet ‘z’

Solution -

1. Entered into spark shell by using command – **spark-shell**

2. Creatd the list of tuples

var tuple = List((1,”alpha”),(2,”beta”),(3,”gamma”),(4,”zeta”),(5,”omega”))

```
scala> var tuple = List((1,"alpha"),(2,"beta"),(3,"gamma"),(4,"zeta"),(5,"omega"))
tuple: List[(Int, String)] = List((1,alpha), (2,beta), (3,gamma), (4,zeta), (5,omega))

scala> println(tuple)
List((1,alpha), (2,beta), (3,gamma), (4,zeta), (5,omega))
```

Problem 1

Filtered the list where string length is equal to 4 and print its corresponding number

Ans – **tuple.filter(x=>x._2.length == 4).map(_._1).foreach(println)**

o/p :

2
4

```
scala> tuple.filter(x=>x._2.length == 4).map(_._1).foreach(println
2
4
```

Problem 2

Filtered the list if a string contains m or z and then calculated its sum and divided by its size to get the average.

Ans

var tup2 = tuple.filter(x=>x.-2.contains(“m”) || x._2.contains(“z”))
tup2.map(_._1).sum/tup2.size

Output:

4

```
scala> var tup2 = tuple.filter(x=>x._2.contains("m") || x._2.contains("z"))
tup2: List[(Int, String)] = List((3,gamma), (4,zeta), (5,omega))

scala> tup2.map(_._1).sum/tup2.size
res2: Int = 4
```