1. Write a **Spark program using Python**, to analyse the given **Insurance Data** and generate a statistics report with the construction building name and the count of building/ county name and its frequency.

**Code:**  insurance.py

import sys

from pyspark import SparkContext

if(len(sys.argv)!=4):

    print("Provide Input File and Output Directory")

    sys.exit(0)

sc =SparkContext()

f = sc.textFile(sys.argv[1])

# Construction building or Count of building

temp=f.map(lambda x: (x.split(',')[16],1))

data=temp.countByKey()

dd=sc.parallelize(data.items())

dd.saveAsTextFile(sys.argv[2])

# County name and its frequency

temp=f.map(lambda x: (x.split(',')[2],1))

data=temp.countByKey()

dd=sc.parallelize(data.items())

dd.saveAsTextFile(sys.argv[3])

**Execution:**

spark-submit insurance.py input-insurance.csv construction county

**Output:**

$ cat construction/\*

$ cat county/\*

