TECNOLÓGICO NACIONAL DE MEXICO

INSTITUTO TECNOLOGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

SEMESTRE FEBRERO-JUNIO 2022

MATERIA:

Datos masivos.

UNIDAD 2

Practica 1

DOCENTE:

JOSE CHRISTIAN ROMERO HERNANDEZ

INTEGRANTES:

López Higuera Saúl Alfredo #18210493

Ramos Rivera Manuel Isaí #17212931

Tijuana BC 05 de mayo del 2022

```
package org.apache.spark.examples.mllib


import org.apache.spark.{SparkConf, SparkContext}

// $example on$

import org.apache.spark.mllib.linalg._

import org.apache.spark.mllib.stat.Statistics

import org.apache.spark.rdd.RDD

object CorrelationsExample {


def main(){


   val conf = new SparkConf().setAppName("CorrelationsExample")

   val sc = new SparkContext(conf)


   // $example on$

   val seriesX: RDD[Double] = sc.parallelize(Array(1, 2, 3, 3, 5))  // a series

   // must have the same number of partitions and cardinality as seriesX

   val seriesY: RDD[Double] = sc.parallelize(Array(11, 22, 33, 33, 555))


   // compute the correlation using Pearson's method. Enter "spearman" for Spearman's method.
If a

   // method is not specified, Pearson's method will be used by default.

   val correlation: Double = Statistics.corr(seriesX, seriesY, "pearson")

   println(s"Correlation is: $correlation")
```

```
    val data: RDD[Vector] = sc.parallelize(

    Seq(

      Vectors.dense(1.0, 10.0, 100.0),

      Vectors.dense(2.0, 20.0, 200.0),

      Vectors.dense(5.0, 33.0, 366.0))

    )  // note that each Vector is a row and not a column


    // calculate the correlation matrix using Pearson's method. Use "spearman" for Spearman's
method

    // If a method is not specified, Pearson's method will be used by default.

    val correlMatrix: Matrix = Statistics.corr(data, "pearson")

    println(correlMatrix.toString)

    // $example off$


    sc.stop()

    }

}
```