/\*

\* Application to perform analysis on state development

\*/

package statedevelopment

import org.apache.spark.SparkContext

import org.apache.spark.SparkConf

import org.apache.spark.sql.SQLContext

import org.apache.spark.sql.SQLImplicits

import org.apache.spark.sql.types.\_

import org.apache.spark.sql.types.StringType

import org.apache.spark.sql.\_

object AnalysisStateDevelopment {

def main(args: Array[String]): Unit = {

//specify the configuration for the spark application using instance of SparkConf

val config = new SparkConf().setAppName("Assignment 20.2").setMaster("local")

//setting the configuration and creating an instance of SparkContext

val sc = new SparkContext(config)

//Entry point of our sqlContext

val sqlContext = new org.apache.spark.sql.SQLContext(sc)

//to use toDF method

import sqlContext.implicits.\_

/\*

\* create the rdd from textfile and get the columns and convert it to Dataframe using toDF method and register the same to temp table census

\*/

val census\_data = sc.textFile("/home/acadgild/Spark/assignment/census.csv").map(x => x.split(",")).map(x => (x(0), x(2), x(3), x(4), x(5), x(6), x(7), x(8), x(9), x(10), x(11), x(12), x(13), x(14), x(15), x(16), x(17), x(18), x(19), x(20), x(21), x(22))).toDF("State", "Persons", "Males", "Females", "Growth\_1991\_2001", "Rural", "Urban", "Scheduled\_Caste\_population", "Percentage\_SC\_to\_total", "Number\_of\_households", "Household\_size\_per\_household", "Sex\_ratio\_females\_per\_1000\_males ", "Sex\_ratio\_0\_6\_years", "Scheduled\_Tribe\_population", "Percentage\_to\_total\_population\_ST", "Persons\_literate", "Males\_Literate", "Females\_Literate", "Persons\_literacy\_rate", "Males\_Literatacy\_Rate", "Females\_Literacy\_Rate", "Total\_Educated").registerTempTable("census")

//1. Find out the state wise population and order by state

/\*

\* get the sum of persons and group by state to get the total population by state

\*/

val population = sqlContext.sql("select state,sum(persons) as total\_population from census group by state order by total\_population desc").show

//2. Find out the Growth Rate of Each State Between 1991-2001

/\*

\* get the growth rate by using aggregate function avg on Growth\_1991\_2001 column grouping by state

\*/

val growth\_rate = sqlContext.sql("select state,avg(Growth\_1991\_2001) as total\_growth from census group by state").show

//3. Find the literacy rate of each state

/\*

\* get the literacy rate of each state by using aggregate function avg on Persons\_literacy\_rate grouping by state

\*/

val literacy = sqlContext.sql("select state,avg(Persons\_literacy\_rate) from census group by state").show

//4. Find out the States with More Female Population

/\*

\* get the States with More Female Population by using aggregate function sum and grouping by state

\*/

val female\_pop = sqlContext.sql("select state, sum(Males)-sum(Females) from census group by state").show

//5. Find out the Percentage of Population in Every State

/\*

\* get the Percentage of Population in Every State by using aggregate sunction sum and window function sum() over

\*/

val percenet\_pop = sqlContext.sql("select state, (sum(persons) \* 100.0) / SUM(sum(persons)) over() as percent\_pop\_by\_state from census group by state").show

}

}









