**Name: Manish Patil**

**For this problem, you are to take the flight dataset for year 2001 and write a spark program which returns the top 10 flights with the highest departure delay in year 2001**

scala> val flightrdd=sc.textFile("Documents/2001.csv")  
flightrdd: org.apache.spark.rdd.RDD[String] = Documents/2001.csv MapPartitionsRDD[1] at textFile at <console>:24  
  
scala> flightrdd.take(2)  
res0: Array[String] = Array(Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay, 2001,1,17,3,1806,1810,1931,1934,US,375,N700��,85,84,60,-3,-4,BWI,CLT,361,5,20,0,NA,0,NA,NA,NA,NA,NA)  
  
scala> val headerrdd=flightrdd.first  
headerrdd: String = Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,LateAircraftDelay  
  
scala> val noheaderrdd=flightrdd.filter(line=>line!=headerrdd)  
noheaderrdd: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[4] at filter at <console>:28  
  
scala> noheaderrdd.take(2).foreach(println)  
2001,1,17,3,1806,1810,1931,1934,US,375,N700��,85,84,60,-3,-4,BWI,CLT,361,5,20,0,NA,0,NA,NA,NA,NA,NA  
2001,1,18,4,1805,1810,1938,1934,US,375,N713��,93,84,64,4,-5,BWI,CLT,361,9,20,0,NA,0,NA,NA,NA,NA,NA  
  
scala> val splitrdd=noheaderrdd.map(line=>line.split(","))  
splitrdd: org.apache.spark.rdd.RDD[Array[String]] = MapPartitionsRDD[5] at map at <console>:30  
  
scala> val removeNA="NA"  
removeNA: String = NA  
  
scala> val filteredOutput = splitrdd.map(values=>(values(8),values(16),values(17),values(0)+"/"+values(1)+"/"+values(2),values(15).replaceAll(removeNA, "0")))  
filteredOutput: org.apache.spark.rdd.RDD[(String, String, String, String, String)] = MapPartitionsRDD[15] at map at <console>:34  
  
scala> filteredOutput.takeOrdered(10)(Ordering[Int].reverse.on(x=>(x.\_5.toInt))).foreach(println)  
(NW,HNL,MSP,2001/1/19,1692)                                                       
(NW,SNA,DTW,2001/10/25,1592)  
(NW,MCO,MSP,2001/1/6,1561)  
(NW,HNL,MSP,2001/1/8,1499)  
(NW,ORD,MEM,2001/9/19,1477)  
(NW,HNL,MSP,2001/4/11,1466)  
(NW,PHX,DTW,2001/2/9,1435)  
(NW,PHX,DTW,2001/2/2,1435)  
(NW,PHX,DTW,2001/2/14,1435)  
(NW,PHX,DTW,2001/2/4,1435)