from pyspark.sql import SparkSession  
from pyspark.sql.functions import col, avg,sum,count  
  
spark=SparkSession=SparkSession.builder.master("local[1]").appName("bootcamp.com").getOrCreate()  
  
df = spark.read.csv("file:////home/takeo/citytemp", header=True, inferSchema=True)  
#df.show()  
  
#to display avg, total and number of measurement by using aggregation  
  
df\_aggregated = df.groupBy("city").agg(  
 avg("temperature").alias("avg\_temperature"),  
 sum("temperature").alias("total\_temperature"),  
 count("temperature").alias("num\_measurements")  
)  
#df\_aggregated.show()  
  
total\_tem\_less\_then\_30=df\_aggregated.filter(df\_aggregated.total\_temperature>30)  
#total\_tem\_less\_then\_30.show()  
  
sort\_by\_city\_asce=total\_tem\_less\_then\_30.orderBy("city")  
sort\_by\_city\_asce.show()  
  
sort\_by\_city\_desc=total\_tem\_less\_then\_30.orderBy("city",ascending=False)  
sort\_by\_city\_desc.show()  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 ()