a) Load the Weather Dataset and create DataFrame

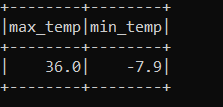
df = spark.read.options(delimiter=' ',header=False,inferSchema=True).csv("/home/miles/futurense\_hadoop-pyspark/labs/dataset/weather/weather\_fine.txt")

df = df.select('\_c1','\_c5','\_c6')

df = df.selectExpr("\_c1 as lst\_date","\_c5 as t\_daily\_max","\_c6 as t\_daily\_min")

b) Show Min and Max Temperature

df.select(max('t\_daily\_max').alias('max\_temp'),min('t\_daily\_min').alias('min\_temp')).show()



c) Show month wise Min and Max Temperature

df.groupBy(month('lst\_date').alias('months')).agg(max('t\_daily\_max').alias('max\_temp'),min('t\_daily\_min').alias('min\_temp')).orderBy(month('lst\_date')).show()

