

BDA Lab1 - SparkSQL

Student: Xuan Wang (xuawa284)

Student: Lepeng Zhang (lepzh903)

Question1:

year, station with the max, maxValue ORDER BY maxValue DESC

year, station with the min, minValue ORDER BY minValue DESC

1) Code: lab2-1.py

2) Partial results:

Max:

```
Row(year=1975, station='86200', MaxValue=36.1)
Row(year=1992, station='63600', MaxValue=35.4)
Row(year=1994, station='117160', MaxValue=34.7)
Row(year=2014, station='96560', MaxValue=34.4)
Row(year=2010, station='75250', MaxValue=34.4)
Row(year=1989, station='63050', MaxValue=33.9)
Row(year=1982, station='94050', MaxValue=33.8)
Row(year=1968, station='137100', MaxValue=33.7)
Row(year=1966, station='151640', MaxValue=33.5)
Row(year=1983, station='98210', MaxValue=33.3)
```

Min:

```
Row(year=1990, station='147270', minValue=-35.0)
Row(year=1990, station='166870', minValue=-35.0)
Row(year=1952, station='192830', minValue=-35.5)
Row(year=1974, station='179950', minValue=-35.6)
Row(year=1974, station='166870', minValue=-35.6)
Row(year=1954, station='113410', minValue=-36.0)
Row(year=1992, station='179960', minValue=-36.1)
Row(year=1975, station='157860', minValue=-37.0)
Row(year=1972, station='167860', minValue=-37.5)
Row(year=1995, station='182910', minValue=-37.6)
```

Question2:

year, month, value ORDER BY value DESC

year, month, value ORDER BY value DESC

1) Code: lab2-2.py

2) Partial results:

```
[x_xuawa@sigma total_count]$ head -10 part-00000
Row(year=2014, month=7, count=147681)
Row(year=2011, month=7, count=146656)
Row(year=2010, month=7, count=143419)
Row(year=2012, month=7, count=137477)
```

```

Row(year=2013, month=7, count=133657)
Row(year=2009, month=7, count=133008)
Row(year=2011, month=8, count=132734)
Row(year=2009, month=8, count=128349)
Row(year=2013, month=8, count=128235)
Row(year=2003, month=7, count=128133)

```

```
[x_xuawa@sigma distinct_count]$ head -10 part-00000
```

```

Row(year=1972, month=10, count=378)
Row(year=1973, month=6, count=377)
Row(year=1973, month=5, count=377)
Row(year=1972, month=8, count=376)
Row(year=1973, month=9, count=376)
Row(year=1972, month=5, count=375)
Row(year=1972, month=9, count=375)
Row(year=1972, month=6, count=375)
Row(year=1971, month=8, count=375)
Row(year=1971, month=6, count=374)

```

Question3:

year, month, station, avgMonthlyTemperature ORDER BY avgMonthlyTemperature DESC

1) Code: lab2-3.py

2) Partial results:

```
[x_xuawa@sigma lab2-3]$ head -10 part-00000
```

```

Row(year=2014, month=7, station='96000', avgMonthlyTemperature=26.3)
Row(year=1994, month=7, station='96550',
avgMonthlyTemperature=23.071052631578947)
Row(year=1983, month=8, station='54550', avgMonthlyTemperature=23.0)
Row(year=1994, month=7, station='78140',
avgMonthlyTemperature=22.970967741935485)
Row(year=1994, month=7, station='85280',
avgMonthlyTemperature=22.87258064516129)
Row(year=1994, month=7, station='75120',
avgMonthlyTemperature=22.85806451612903)
Row(year=1994, month=7, station='65450',
avgMonthlyTemperature=22.85645161290323)
Row(year=1994, month=7, station='96000',
avgMonthlyTemperature=22.80806451612903)
Row(year=1994, month=7, station='95160',
avgMonthlyTemperature=22.764516129032256)
Row(year=1994, month=7, station='86200',
avgMonthlyTemperature=22.711290322580645)

```

Question4:

station, maxTemp, maxDailyPrecipitation ORDER BY station DESC

1) Code: lab2-4.py

2) Partial results:

Empty, no stations fulfill such requirements.

Question5:

year, month, avgMonthlyPrecipitation ORDER BY year DESC, month DESC

1) Code: lab2-5.py

2) Partial results:

```
[x_xuawa@sigma lab2-5]$ head -10 part-00000
```

```
Row(year=2016, month=7, avgMonthlyPrecipitation=0.0)
```

```
Row(year=2016, month=6, avgMonthlyPrecipitation=47.6625)
```

```
Row(year=2016, month=5, avgMonthlyPrecipitation=29.250000000000007)
```

```
Row(year=2016, month=4, avgMonthlyPrecipitation=26.900000000000001)
```

```
Row(year=2016, month=3, avgMonthlyPrecipitation=19.962500000000002)
```

```
Row(year=2016, month=2, avgMonthlyPrecipitation=21.562500000000004)
```

```
Row(year=2016, month=1, avgMonthlyPrecipitation=22.325)
```

```
Row(year=2015, month=12, avgMonthlyPrecipitation=28.925)
```

```
Row(year=2015, month=11, avgMonthlyPrecipitation=63.887500000000001)
```

```
Row(year=2015, month=10, avgMonthlyPrecipitation=2.2625)
```