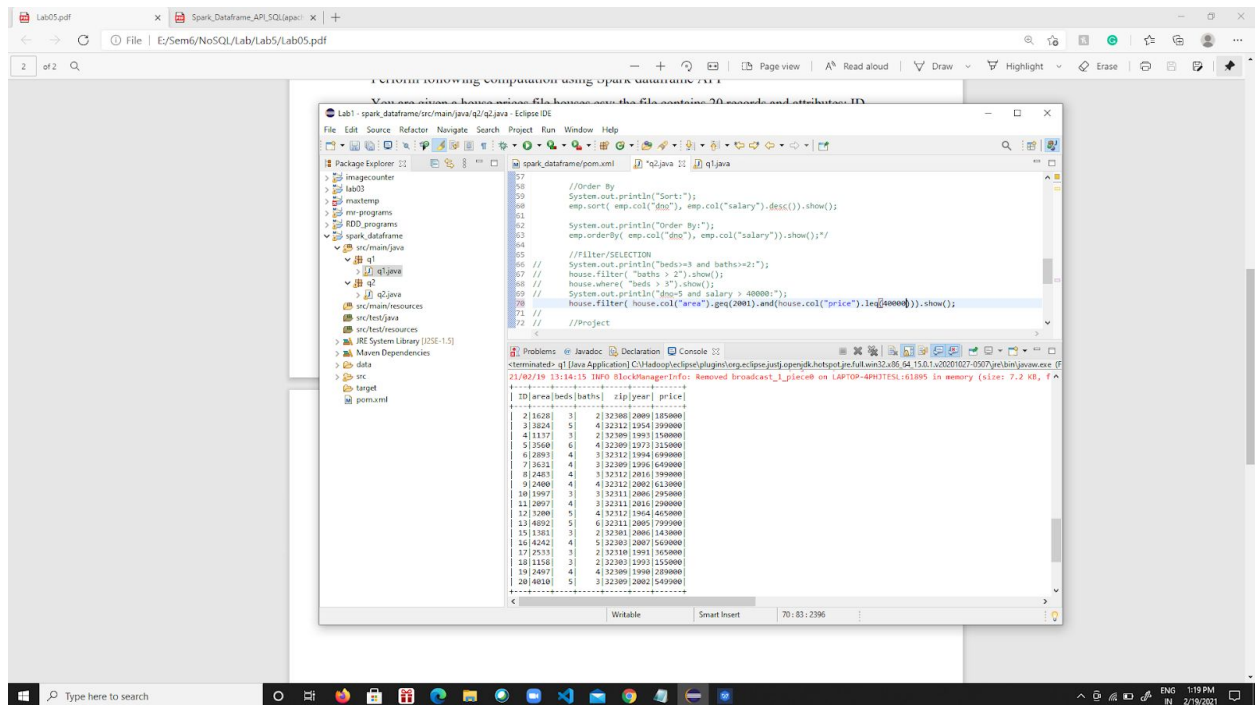
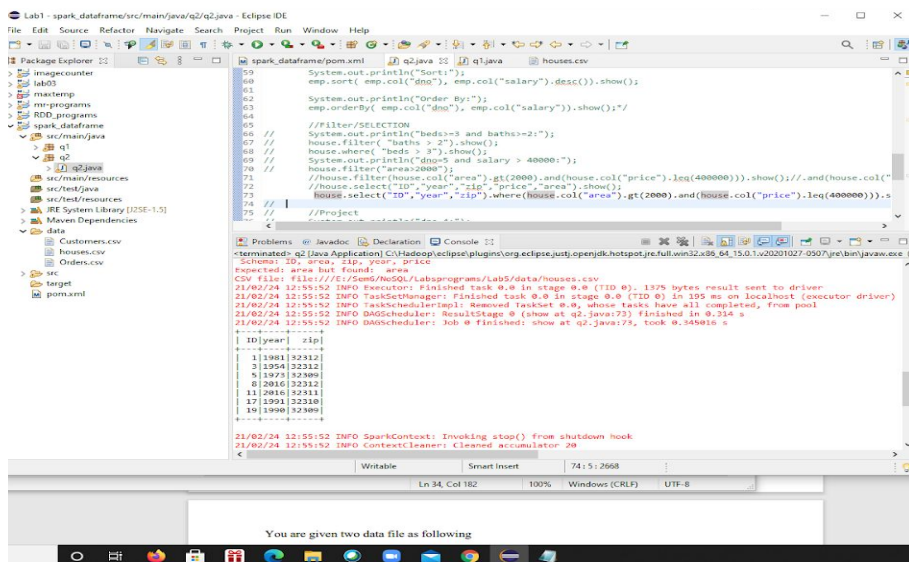


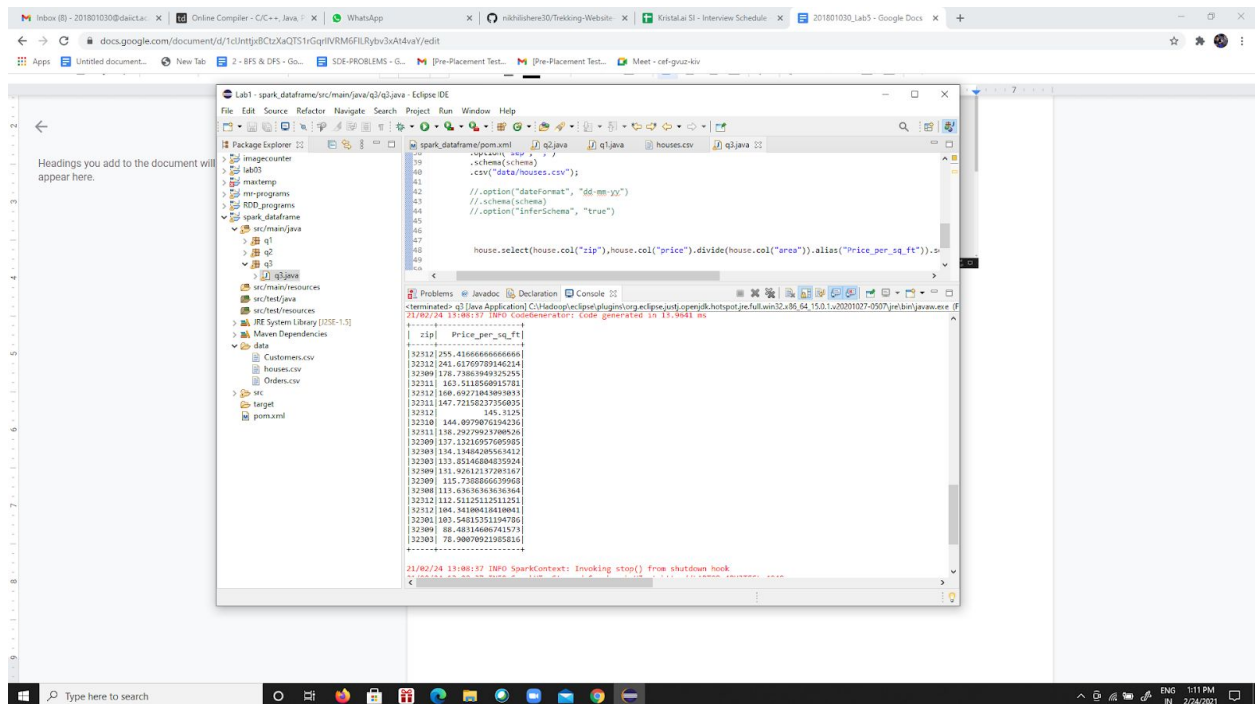
Q1 without sql



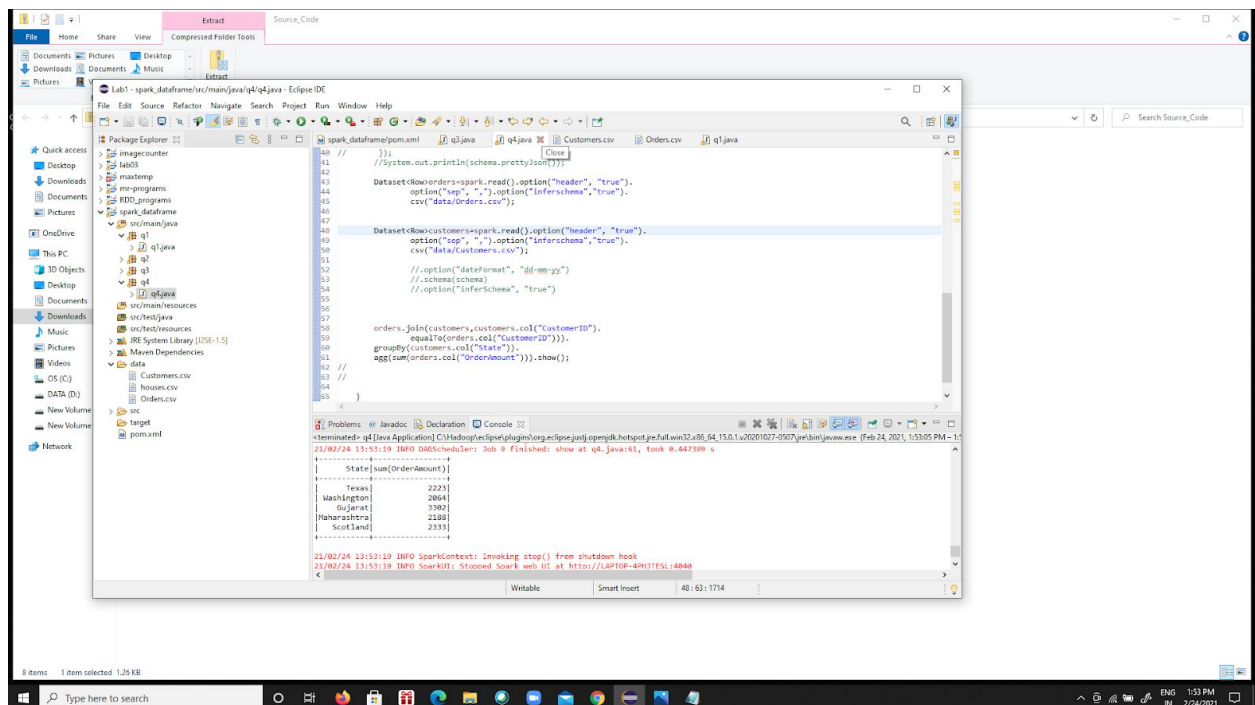
Q2 without sql



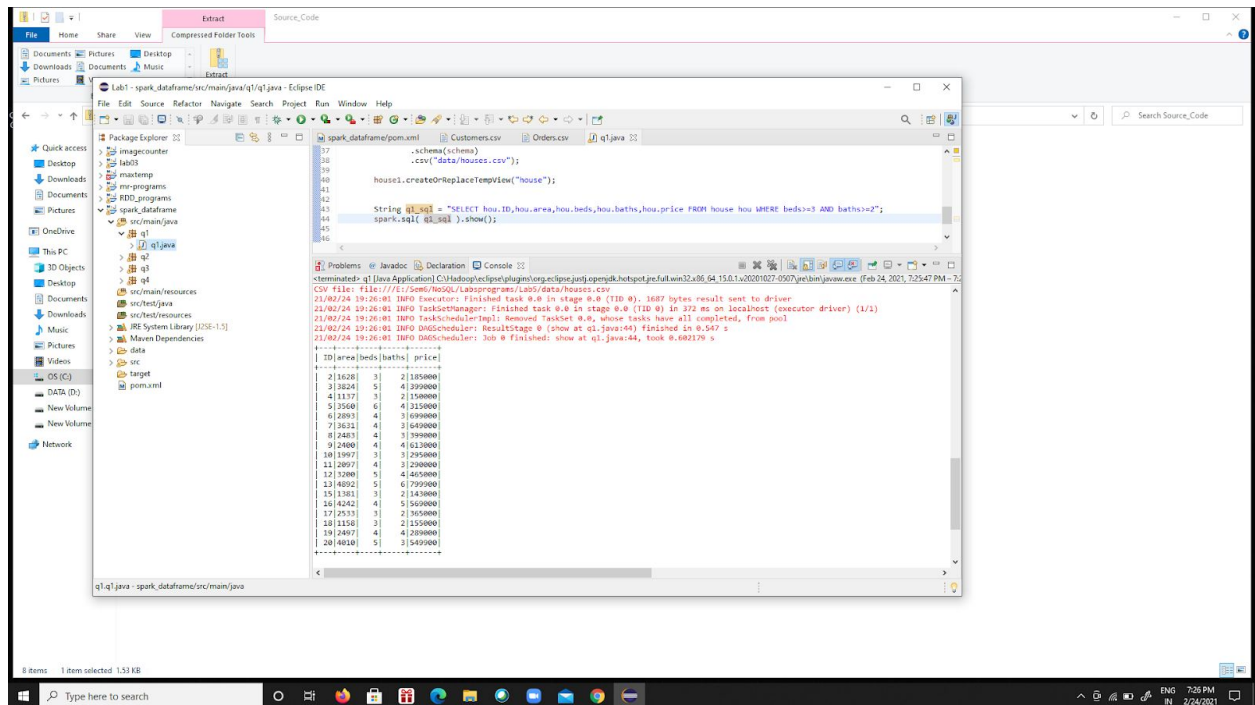
Q3 without sql



Q4 without sql



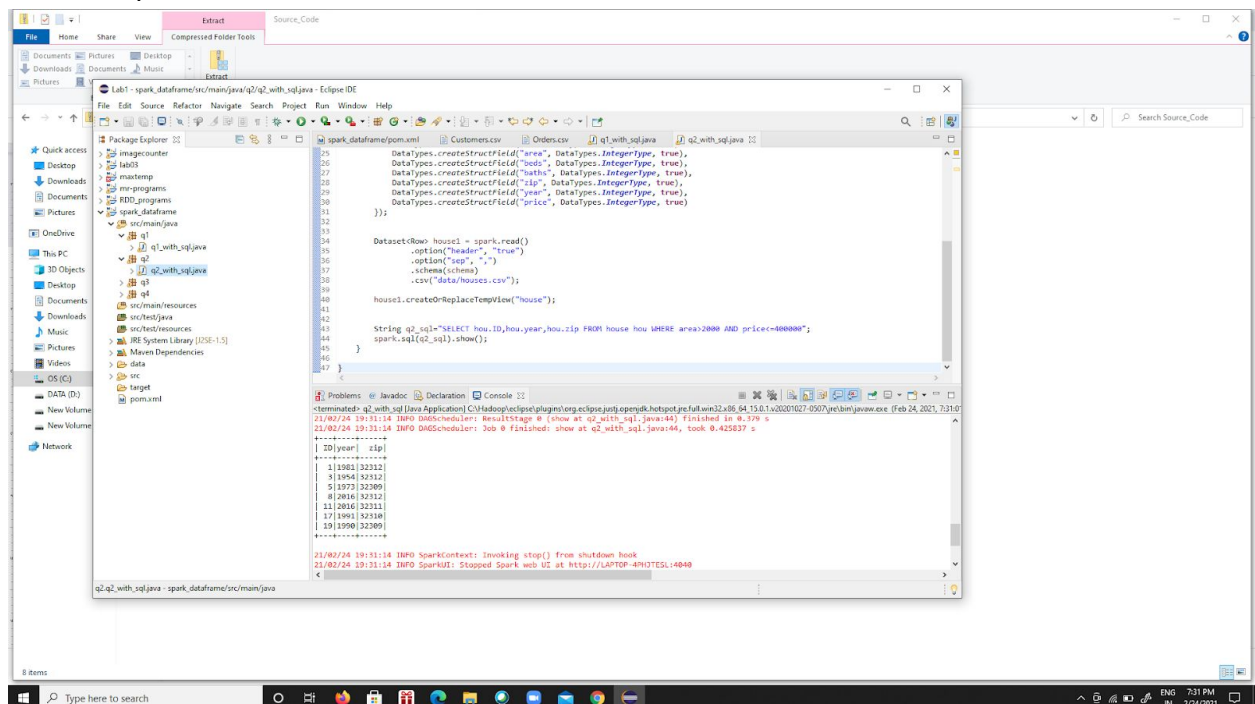
Q1 with sql



```
17 .schema(schema)
18 .csv("data/houses.csv");
19
20 house1.createOrReplaceTempView("house");
21
22 String sql = "SELECT hou.ID,hou.area,hou.beds,hou.baths,hou.price FROM house hou WHERE beds>3 AND baths>2";
23 spark.sql(sql).show();
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

```
21/02/24 19:26:01 INFO Executor: Finished task 0.0 in stage 0.0 (TID 0). 1687 bytes result sent to driver
21/02/24 19:26:01 INFO TaskScheduler: Finished task 0.0 in stage 0.0 (TID 0) in 372 ms on localhost (executor driver) (1/1)
21/02/24 19:26:01 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
21/02/24 19:26:01 INFO DAGScheduler: ResultStage 0 (show at q1.java:44) finished in 0.547 s
21/02/24 19:26:01 INFO DAGScheduler: Job 0 finished: show at q1.java:44, took 0.682179 s
-----
+-----+
| ID|area|beds|baths| price|
+-----+
| 1|21628| 3| 2|105000|
| 3|3824| 5| 4|399000|
| 4|1257| 5| 2|150000|
| 5|2560| 6| 4|315000|
| 6|2893| 4| 3|690000|
| 7|3631| 4| 3|640000|
| 8|2403| 4| 3|399000|
| 9|2400| 4| 4|615000|
|10|1997| 3| 3|295000|
|11|2097| 4| 2|290000|
|12|1200| 5| 4|460000|
|13|4892| 5| 6|799000|
|15|1361| 3| 2|145000|
|16|4242| 4| 5|560000|
|17|2533| 3| 2|365000|
|18|1158| 3| 2|155000|
|19|2497| 4| 4|289000|
|20|4010| 5| 3|549000|
+-----+
```

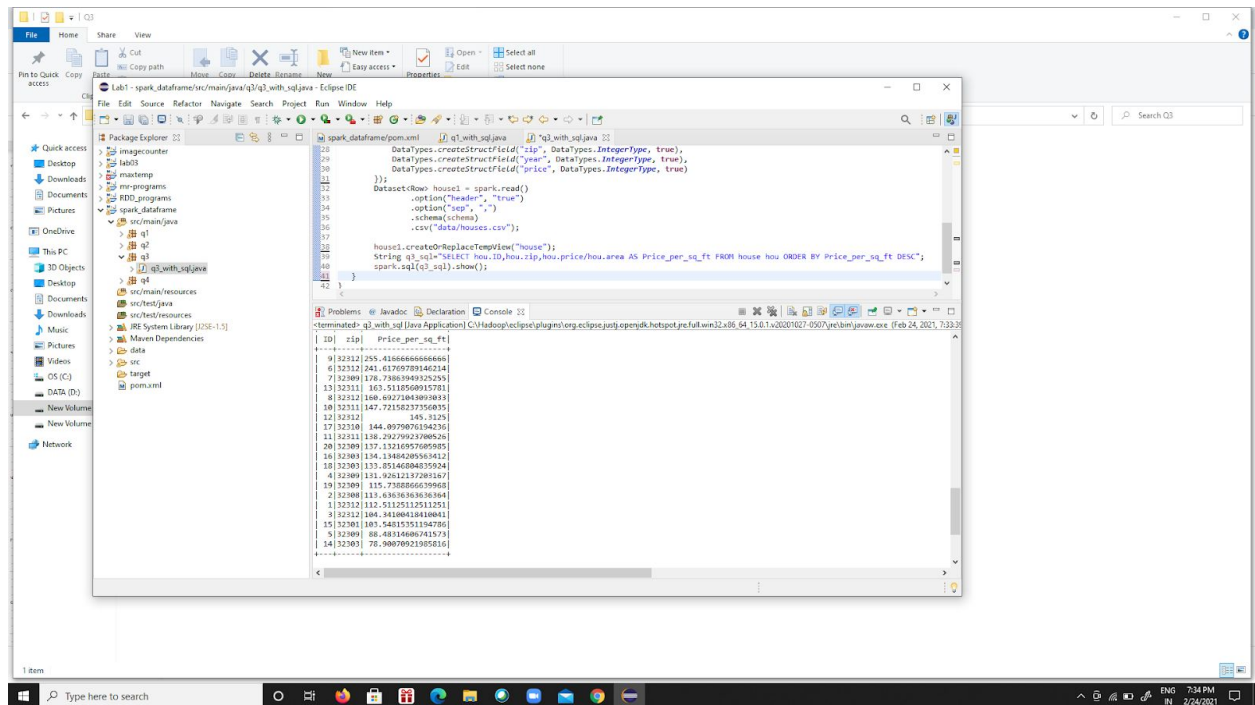
Q2 with sql



```
25 DataTypes.createStructField("area", DataTypes.IntegerType, true),
26 DataTypes.createStructField("beds", DataTypes.IntegerType, true),
27 DataTypes.createStructField("baths", DataTypes.IntegerType, true),
28 DataTypes.createStructField("zip", DataTypes.IntegerType, true),
29 DataTypes.createStructField("year", DataTypes.IntegerType, true),
30 DataTypes.createStructField("price", DataTypes.IntegerType, true)
31 ));
32
33
34 Dataset<Row> house1 = spark.read()
35 .option("header", "true")
36 .option("sep", ",")
37 .schema(schema)
38 .csv("data/houses.csv");
39
40 house1.createOrReplaceTempView("house");
41
42 String sql = "SELECT hou.ID,hou.year,hou.zip FROM house hou WHERE area>2000 AND price<800000";
43 spark.sql(sql).show();
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

```
21/02/24 19:31:14 INFO DAGScheduler: ResultStage 0 (show at q2_with_sql.java:44) finished in 0.179 s
21/02/24 19:31:14 INFO DAGScheduler: Job 0 finished: show at q2_with_sql.java:44, took 0.425537 s
-----
+-----+
| ID|year| zip|
+-----+
| 1|1981|32312|
| 3|1954|32312|
| 5|1973|32389|
| 8|2016|32312|
|11|2016|32313|
|17|1993|32310|
|19|1990|32389|
+-----+
```

Q3 with sql



Q4 with sql

