

- a. 583985
- b. 26216
- c. 238
- d. 125861
- e. LGA, DCA, LAX, MSP, ATL
- f. ATL

```
[>>> spark.sql("select count(*) from january_data").show()
+-----+
|count(1)|
+-----+
|  583985|
+-----+

[>>> spark.sql("select count(*) from january_data where ORIGIN = 'ORD' and dest = 'JFK']").show()
File "<stdin>", line 1
    spark.sql("select count(*) from january_data where ORIGIN = 'ORD' and dest = 'JFK']").show()
                        ^
SyntaxError: invalid character in identifier

[>>> spark.sql("select count(*) from january_data where ORIGIN = 'ORD' and dest = 'JFK']").show()
+-----+
|count(1)|
+-----+
|    238|
+-----+

[>>> spark.sql("select count(*) from january_data where ORIGIN = 'ORD']").show()
+-----+
|count(1)|
+-----+
|   26216|
+-----+

[>>> spark.sql("select count(*) from january_data where DEP_DEL15 = '1']").show()
+-----+
|count(1)|
+-----+
|   98927|
+-----+

[>>> spark.sql("select count(*) from january_data where ARR_DEL15 = '1']").show()
+-----+
|count(1)|
+-----+
|  105222|
+-----+

[>>> spark.sql("select count(*) from january_data where ARR_DEL15 = '1' or DEP_DEL15 = '1']").show()
+-----+
|count(1)|
+-----+
|  125861|
+-----+
```

```
[>>> spark.sql("select origin, dest, count(*) as NumFlights from january_data jd where origin = 'ORD' GROUP BY ORIGIN, DEST ORDER BY NumFlights desc limit 10").show()
```

```
+-----+-----+-----+
|origin|dest|NumFlights|
+-----+-----+-----+
|ORD|LGA|1230|
|ORD|DCA|657|
|ORD|LAX|648|
|ORD|MSP|623|
|ORD|ATL|621|
|ORD|DFW|567|
|ORD|DTW|525|
|ORD|BOS|496|
|ORD|PHX|491|
|ORD|DEN|479|
+-----+-----+-----+
```

```
[>>> spark.sql("select origin, count(cancelled) as NumCancels from january_data jd GROUP BY ORIGIN ORDER BY NumCancels desc limit 5").show()
```

```
+-----+-----+
|origin|NumCancels|
+-----+-----+
|ATL|31155|
|ORD|26216|
|DFW|23063|
|CLT|19100|
|DEN|18507|
+-----+-----+
```