

 default

FINISHED

Took 3 sec. Last updated by anonymous at July 07 2022, 10:01:46 PM.

SPARK JOB FINISHED

Took 17 sec. Last updated by anonymous at July 07 2022, 10:39:52 PM.

SPARK JOB FINISHED

Took 1 sec. Last updated by anonymous at July 07 2022, 7:30:31 PM.

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 8:17:10 PM.

SPARK JOB FINISHED

✕

Took 1 sec. Last updated by anonymous at July 07 2022, 8:16:24 PM. (outdated)

SPARK JOB FINISHED

✕

Took 13 sec. Last updated by anonymous at July 07 2022, 10:39:53 PM.

SPARK JOB FINISHED

Took 1 sec. Last updated by anonymous at July 07 2022, 7:45:20 PM

FINISHED

```
root
|-- datasetId: string (nullable = true)
|-- name: string (nullable = true)
|-- address1: string (nullable = true)
|-- address2: string (nullable = true)
|-- city: string (nullable = true)
|-- state: string (nullable = true)
|-- zip: string (nullable = true)
|-- tel: string (nullable = true)
|-- dataStart: string (nullable = true)
```

```
-- name: string (nullable = true)
-- dateEnd: string (nullable = true)
-- type: string (nullable = true)
-- risk: long (nullable = true)
-- rpt_area_desc: string (nullable = true)
-- seats: long (nullable = true)
-- sewage: string (nullable = true)
-- smoking_allowed: string (nullable = true)
-- status: string (nullable = true)
-- geolocation: array (nullable = true)
-- element: double (containsNull = true)
```

Took 0 sec. Last updated by anonymous at July 07 2022, 7:50:56 PM.

```
val dftrab = dfnew
    .withColumn("dateStart", to_date(col("dateStart"), "yyyy-MM-dd"))
    .withColumn("dateEnd", to_date(col("dateEnd"), "dd-MM-yyyy"))

dftrab: org.apache.spark.sql.DataFrame = [datasetId: string, name: string ... 16 more fields]
```

Took 2 sec. Last updated by anonymous at July 07 2022, 10:39:54 PM.

```
dftrab.printSchema
-- address1: string (nullable = true)
-- address2: string (nullable = true)
-- city: string (nullable = true)
-- state: string (nullable = true)
-- zip: string (nullable = true)
-- tel: string (nullable = true)
-- dateStart: date (nullable = true)
-- dateEnd: date (nullable = true)
-- type: string (nullable = true)
-- risk: long (nullable = true)
-- rpt_area_desc: string (nullable = true)
-- seats: long (nullable = true)
-- sewage: string (nullable = true)
-- smoking_allowed: string (nullable = true)
-- status: string (nullable = true)
-- geolocation: array (nullable = true)
| |-- element: double (containsNull = true)
```

Took 0 sec. Last updated by anonymous at July 07 2022, 10:39:58 PM.

z.show(dftrab)

SPARK JOB FINISHED

datasetId	name	address1	address2	city	state	zip	tel	dateStart	dateEnd	type
	BEGINNINGS									
59120	BRIDGES AT SOUTHPOINT APTS	7304 CALIBRE PARK DR	null	DURHAM	NC	27707	null	1996-04-02	null	53 - Year-Round Swimming Pool
59124	SHEARTON INN UNIVERSITY	2800 CAMPUS WALK AVE	null	DURHAM	NC	27705	null	1996-04-25	null	53 - Year-Round Swimming Pool
59263	SPA HEALTH CLUB	3419 HILLSBOROUGH RD	null	DURHAM	NC	27705	null	2000-05-30	null	55 - Year-Round Swimming Pool
58349	KROGER R 381 MEAT MARKET	3825 S ROXBORO ST	SUITE 101	DURHAM	NC	27707	(919) 361-0470	2003-03-03	null	30 - Meat Market
59249	LEONE	240 CAMFRET AVE	SUITE 100	DURHAM	NC	27704	(919) 288-2888	2004-08-17	null	20 - Meat Market

Output is truncated to 102400 bytes. Learn more about ZEPPELIN_INTERPRETER_OUTPUT_LIMIT

Took 0 sec. Last updated by anonymous at July 07 2022, 10:10:12 PM. (outdated)

¿Cuál es el restaurante más grande según el número de sillas (seats)?

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:45:49 PM.

```
dftrab.select("name", "seats").where(col("type") === "1 - Restaurant").orderBy(col("seats").desc).show(1, false)
```

SPARK JOB FINISHED

```
-----+-----+
|name      |seats|
-----+-----+
|JUMBO CHINA|2040 |
-----+-----+
only showing top 1 row
```

Took 0 sec. Last updated by anonymous at July 07 2022, 8:30:25 PM.

Muestre (en tabla) las ciudades y la cantidad de restaurantes que tiene (city).

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:48:12 PM.

```
dftrab.select("city").where(col("type") === "1 - Restaurant").groupBy(col("city")).count().show(false)
```

SPARK JOB FINISHED

```
-----+-----+
|city      |count|
-----+-----+
|MORRISVILLE|5 |
|RTP        |7 |
|RESEARCH TRIANGLE PARK|1 |
|CHAPEL HILL |13 |
|DURHAM      |660 |
|ROUGEMONT   |1 |
-----+-----+
```

Took 2 sec. Last updated by anonymous at July 07 2022, 8:35:34 PM.

Cuántos los restaurantes que permiten y no permiten fumadores (smoking_allowed)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:51:38 PM.

```
dftrab.select("smoking_allowed").where(col("type") === "1 - Restaurant").groupBy("smoking_allowed").count().show()
```

SPARK JOB FINISHED

```
-----+-----+
|smoking_allowed|count|
-----+-----+
|N/A            |396 |
|NO             |291 |
-----+-----+
```

Took 0 sec. Last updated by anonymous at July 07 2022, 8:45:27 PM.

¿Cuál es el restaurante más antiguo (año de apertura)? (dateStart)

FINISHED

¿Cuál es el restaurante más antiguo (dateStart)? (dateStart)

Took 0 sec. Last updated by anonymous at July 07 2022, 8:53:31 PM.

```
| dftrab.select($"name", col("dateStart")).sort(col("dateStart")).show(1)
+-----+
|          name| dateStart|
+-----+
|PIZZA HUT DELIVERY|1990-07-01|
+-----+
only showing top 1 row
```

SPARK JOB FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 10:45:06 PM.

¿Cuál fue el año en donde se abrieron más restaurantes? (dateStart)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:52:19 PM. (outdated)

```
| dftrab.select(year(col("dateStart")).as("año")).groupBy("año").count().sort(col("count").desc).show(1)
+-----+
| año|count|
+-----+
|2016| 253|
+-----+
only showing top 1 row
```

SPARK JOB FINISHED

Took 1 sec. Last updated by anonymous at July 07 2022, 10:45:20 PM.

¿Cuál fue el año en donde se abrieron menos restaurantes? (dateStart)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:54:16 PM.

```
| dftrab.select(year(col("dateStart")).as("año")).groupBy("año").count().sort(col("count")).show(1)
+-----+
| año|count|
+-----+
|1992| 12|
+-----+
only showing top 1 row
```

SPARK JOB FINISHED

Took 1 sec. Last updated by anonymous at July 07 2022, 10:45:29 PM.

¿Cuáles son los tipos de restaurantes que existen? (type)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:54:42 PM.

```
| z.show(dftrab.select("type").distinct().orderBy("type"))
```

SPARK JOB FINISHED



type
52 - Seasonal Spa
53 - Year-Round Swimming Pool
54 - Year-Round Wading Pool
55 - Year-Round Spa
6 - Edu. Food Service
61 - Tattoo Artists
72 - Summer Feeding Program
9 - Elderly Nutrition
4

Took 0 sec. Last updated by anonymous at July 07 2022, 8:03:28 PM.

¿Cuál es el sistema más común para procesar las aguas negras? (sewage)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:55:18 PM.

```
| dftrab.select("sewage").groupBy("sewage").count().orderBy(desc("count"))show(1, false)
+-----+
|sewage|count|
+-----+
|3 - Municipal/Community|2437 |
+-----+
only showing top 1 row
```

SPARK JOB FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 8:21:25 PM.

¿Cuántos restaurante cerraron en el año 2010? (dateEnd)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:55:38 PM.

```
| z.show(dftrab.select(year(col("dateEnd"))) ).where(year(col("dateEnd")) === "2010").count())
```

SPARK JOB FINISHED

ZeppelinContext doesn't support to show type: java.lang.Long
1

Took 0 sec. Last updated by anonymous at July 07 2022, 10:47:01 PM.

¿Cuántos restaurantes abrieron en el mes de diciembre? (dateStart)

FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 7:56:01 PM. (outdated)

```
| dftrab.select(month(col("dateStart")).as("month")).where(col("month") === "12").groupBy("month").count().show()
+-----+
|month|count|
+-----+
| 12| 123|
+-----+
```

SPARK JOB FINISHED

Took 0 sec. Last updated by anonymous at July 07 2022, 10:47:36 PM.

