Mod13_Sesion3_parte2_vista_de_SQL_dentro_de_pyspark

September 23, 2023

[1]:
```
df = spark.read.load("/tmp/dcd/pyspark/covid/parquet")

df.printSchema()
```

```
root
 |-- ORIGEN: integer (nullable = true)
 |-- SECTOR: integer (nullable = true)
 |-- ENTIDAD_UM: string (nullable = true)
 |-- SEXO: integer (nullable = true)
 |-- EDAD: integer (nullable = true)
 |-- EPOC: integer (nullable = true)
```

[2]:
```
df.show()

df.count()
```

```
+------+------+----------+----+----+----+
|ORIGEN|SECTOR|ENTIDAD_UM|SEXO|EDAD|EPOC|
+------+------+----------+----+----+----+
|     2|    12|        09|   2|  41|   2|
|     1|    12|        31|   1|  32|   2|
|     2|    12|        09|   2|  34|   2|
|     2|     4|        14|   2|  27|   2|
|     1|     4|        09|   2|  66|   1|
|     1|     4|        21|   2|  52|   2|
|     2|     4|        15|   1|  41|   2|
|     2|     4|        14|   2|  39|   2|
|     2|     4|        15|   1|  34|   2|
|     1|     4|        30|   1|  35|   2|
|     2|     4|        14|   1|  40|   2|
|     2|     4|        21|   2|  59|   2|
|     1|     4|        18|   2|  52|   2|
|     1|     4|        22|   1|  42|   2|
|     1|     4|        13|   2|  50|   2|
|     2|     4|        02|   1|  67|   2|
```

```
|     2|     4|        30|   2|  51|   2|
|     1|    12|        31|   2|  28|   2|
|     2|     4|        14|   2|  46|   2|
|     2|     4|        27|   2|  44|   2|
+------+------+----------+----+----+----+
only showing top 20 rows
```

[2]:
```
16787
```

[3]:
```
spark.catalog.listDatabases()
```

```
ivysettings.xml file not found in HIVE_HOME or
HIVE_CONF_DIR,/etc/hive/conf.dist/ivysettings.xml will be used
```

[3]:
```
[Database(name='default', description='Default Hive database',
locationUri='hdfs://cluster-sir-dcd05-m/user/hive/warehouse')]
```

[4]:
```
spark.catalog.listTables()
```

[4]:
```
[Table(name='covid_avro', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_csv', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_job', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_mes_part', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_parquet', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_sql', database='default', description=None,
tableType='MANAGED', isTemporary=False),
 Table(name='sirilo', database='default', description=None,
tableType='EXTERNAL', isTemporary=False)]
```

[5]:
```
df.createOrReplaceTempView("covid_df")

spark.catalog.listTables()
```

[5]:
```
[Table(name='covid_avro', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_csv', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_job', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_mes_part', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_parquet', database='default', description=None,
```

```
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_sql', database='default', description=None,
tableType='MANAGED', isTemporary=False),
 Table(name='sirilo', database='default', description=None,
tableType='EXTERNAL', isTemporary=False),
 Table(name='covid_df', database=None, description=None, tableType='TEMPORARY',
isTemporary=True)]
```

[6]:
```
#listar las tablas declaradas utilizando SQL

spark.sql("show tables").show()
```

```
+---------+-------------+-----------+
|namespace|     tableName|isTemporary|
+---------+-------------+-----------+
|  default|    covid_avro|      false|
|  default|     covid_csv|      false|
|  default|     covid_job|      false|
|  default|covid_mes_part|      false|
|  default| covid_parquet|      false|
|  default|     covid_sql|      false|
|  default|        sirilo|      false|
|         |      covid_df|       true|
+---------+-------------+-----------+
```

[7]:
```
#Ver el esquema de una tabla
spark.sql('describe covid_df').show()
```

```
+----------+---------+-------+
|  col_name|data_type|comment|
+----------+---------+-------+
|    ORIGEN|      int|   null|
|    SECTOR|      int|   null|
|ENTIDAD_UM|   string|   null|
|      SEXO|      int|   null|
|      EDAD|      int|   null|
|      EPOC|      int|   null|
+----------+---------+-------+
```

[8]:
```
#Sentecia SQL
spark.sql("SELECT SEXO, EDAD, EDAD * 12 AS MESES FROM covid_df").show()
```

```
+----+----+-----+
|SEXO|EDAD|MESES|
+----+----+-----+
|   2|  41|  492|
|   1|  32|  384|
```

```
|   2|  34|  408|
|   2|  27|  324|
|   2|  66|  792|
|   2|  52|  624|
|   1|  41|  492|
|   2|  39|  468|
|   1|  34|  408|
|   1|  35|  420|
|   1|  40|  480|
|   2|  59|  708|
|   2|  52|  624|
|   1|  42|  504|
|   2|  50|  600|
|   1|  67|  804|
|   2|  51|  612|
|   2|  28|  336|
|   2|  46|  552|
|   2|  44|  528|
+----+----+-----+
only showing top 20 rows
```

[9]:
```
sql = """
select sexo, edad, origen, count(*) as nreg from(
select sexo, edad, 'parquet' origen from covid_parquet
UNION ALL
select sexo, edad, 'avro' origen from covid_avro
UNION ALL
select sexo, edad, 'csv' origen from covid_csv
UNION ALL
select sexo, edad, 'csv' origen from covid_df) a
group by sexo, edad, origen
order by sexo, edad, origen;
"""
spark.sql(sql).show()
```

```
23/09/23 17:30:36 WARN SessionState: METASTORE_FILTER_HOOK will be ignored,
since hive.security.authorization.manager is set to instance of
HiveAuthorizerFactory.
+----+----+-------+----+
|sexo|edad| origen|nreg|
+----+----+-------+----+
|   1|   0|   avro|  28|
|   1|   0|    csv|  84|
|   1|   0|parquet|  28|
|   1|   1|   avro|  20|
|   1|   1|    csv|  60|
|   1|   1|parquet|  20|
```

```
|   1|   2|   avro|   8|
|   1|   2|    csv|  24|
|   1|   2|parquet|   8|
|   1|   3|   avro|  16|
|   1|   3|    csv|  48|
|   1|   3|parquet|  16|
|   1|   4|   avro|  12|
|   1|   4|    csv|  36|
|   1|   4|parquet|  12|
|   1|   5|   avro|  17|
|   1|   5|    csv|  51|
|   1|   5|parquet|  17|
|   1|   6|   avro|  16|
|   1|   6|    csv|  48|
+----+----+-------+----+
only showing top 20 rows
```

[12]:
```
#Leer catálogos CSV from URL COVID usando pandas
import pandas as pd

dfc = pd.read_csv("https://raw.githubusercontent.com/omarmendoza564/datos/main/
 ↪datos/201128CatalogosEntidades.csv" \
                  ,header = 0,dtype = {'CLAVE_ENTIDAD':␣
 ↪str,'ENTIDAD_FEDERATIVA': str, 'ABREVIATURA': str } \
                  ,keep_default_na=False)


dfc
```

[12]:

	CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA	ABREVIATURA
0	01	AGUASCALIENTES	AS
1	02	BAJA CALIFORNIA	BC
2	03	BAJA CALIFORNIA SUR	BS
3	04	CAMPECHE	CC
4	05	COAHUILA DE ZARAGOZA	CL
5	06	COLIMA	CM
6	07	CHIAPAS	CS
7	08	CHIHUAHUA	CH
8	09	CIUDAD DE MÉXICO	DF
9	10	DURANGO	DG
10	11	GUANAJUATO	GT
11	12	GUERRERO	GR
12	13	HIDALGO	HG
13	14	JALISCO	JC
14	15	MÉXICO	MC
15	16	MICHOACÁN DE OCAMPO	MN

16	17	MORELOS	MS
17	18	NAYARIT	NT
18	19	NUEVO LEÓN	NL
19	20	OAXACA	OC
20	21	PUEBLA	PL
21	22	QUERÉTARO	QT
22	23	QUINTANA ROO	QR
23	24	SAN LUIS POTOSÍ	SP
24	25	SINALOA	SL
25	26	SONORA	SR
26	27	TABASCO	TC
27	28	TAMAULIPAS	TS
28	29	TLAXCALA	TL
29	30	VERACRUZ DE IGNACIO DE LA LLAVE	VZ
30	31	YUCATÁN	YN
31	32	ZACATECAS	ZS
32	36	ESTADOS UNIDOS MEXICANOS	EUM
33	97	NO APLICA	NA
34	98	SE IGNORA	SI
35	99	NO ESPECIFICADO	NE

[13]:
```
dfcat = spark.createDataFrame(dfc)

dfcat.show()
type(dfcat)
```

```
+-------------+--------------------+-----------+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|ABREVIATURA|
+-------------+--------------------+-----------+
|           01|      AGUASCALIENTES|         AS|
|           02|     BAJA CALIFORNIA|         BC|
|           03| BAJA CALIFORNIA SUR|         BS|
|           04|            CAMPECHE|         CC|
|           05|COAHUILA DE ZARAGOZA|         CL|
|           06|              COLIMA|         CM|
|           07|             CHIAPAS|         CS|
|           08|           CHIHUAHUA|         CH|
|           09|    CIUDAD DE MÉXICO|         DF|
|           10|             DURANGO|         DG|
|           11|          GUANAJUATO|         GT|
|           12|            GUERRERO|         GR|
|           13|             HIDALGO|         HG|
|           14|             JALISCO|         JC|
|           15|              MÉXICO|         MC|
|           16| MICHOACÁN DE OCAMPO|         MN|
|           17|             MORELOS|         MS|
|           18|             NAYARIT|         NT|
|           19|          NUEVO LEÓN|         NL|
```

```
|          20|               OAXACA|         OC|
+------------+--------------------+-----------+
only showing top 20 rows
```

[13]:
```
pyspark.sql.dataframe.DataFrame
```

[14]:
```
#Registrar la vista temporal Estados

dfcat.createOrReplaceTempView("estados")

spark.sql("show tables").show()
```

```
+---------+-------------+-----------+
|namespace|    tableName|isTemporary|
+---------+-------------+-----------+
|  default|   covid_avro|      false|
|  default|    covid_csv|      false|
|  default|    covid_job|      false|
|  default|covid_mes_part|      false|
|  default| covid_parquet|      false|
|  default|    covid_sql|      false|
|  default|       sirilo|      false|
|         |     covid_df|       true|
|         |      estados|       true|
+---------+-------------+-----------+
```

[16]:
```
spark.sql("SELECT * FROM covid_df WHERE ENTIDAD_UM in('36', '97', '98')")
```

[16]:
```
DataFrame[ORIGEN: int, SECTOR: int, ENTIDAD_UM: string, SEXO: int, EDAD: int,
EPOC: int]
```

[15]:
```
sql = """
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA
ORDER BY 3 DESC
"""
spark.sql(sql).show(35)
```

```
[Stage 23:============================>                            (1 + 1) / 2]
+------------+-------------------+----+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|nreg|
+------------+-------------------+----+
|          09|    CIUDAD DE MÉXICO|5608|
|          19|          NUEVO LEÓN|1050|
|          08|           CHIHUAHUA| 957|
```

```
|           11|          GUANAJUATO| 807|
|           15|              MÉXICO| 623|
|           05|COAHUILA DE ZARAGOZA| 556|
|           14|             JALISCO| 526|
|           24|     SAN LUIS POTOSÍ| 521|
|           27|             TABASCO| 494|
|           21|              PUEBLA| 466|
|           10|             DURANGO| 447|
|           22|           QUERÉTARO| 394|
|           28|          TAMAULIPAS| 356|
|           16| MICHOACÁN DE OCAMPO| 335|
|           03| BAJA CALIFORNIA SUR| 321|
|           02|     BAJA CALIFORNIA| 315|
|           32|           ZACATECAS| 293|
|           31|             YUCATÁN| 288|
|           30|VERACRUZ DE IGNAC…| 277|
|           26|              SONORA| 275|
|           25|             SINALOA| 265|
|           01|      AGUASCALIENTES| 224|
|           12|            GUERRERO| 207|
|           13|             HIDALGO| 201|
|           20|              OAXACA| 190|
|           04|            CAMPECHE| 189|
|           23|        QUINTANA ROO| 125|
|           06|              COLIMA| 115|
|           17|             MORELOS| 110|
|           29|            TLAXCALA| 103|
|           07|             CHIAPAS|  93|
|           18|             NAYARIT|  56|
+-------------+--------------------+----+
```

[17]:
```
nr = 600
orden = "DESC"
sql = f"""
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA
HAVING nreg  >= {nr}
ORDER BY 3 {orden}
"""
spark.sql(sql).show(35)
```

```
[Stage 27:============================>                            (1 + 1) / 2]

+-------------+------------------+----+
|CLAVE_ENTIDAD|ENTIDAD_FEDERATIVA|nreg|
```

```
+-------------+------------------+----+
|           09|  CIUDAD DE MÉXICO|5608|
|           19|        NUEVO LEÓN|1050|
|           08|         CHIHUAHUA| 957|
|           11|        GUANAJUATO| 807|
|           15|            MÉXICO| 623|
+-------------+------------------+----+
```

[21]:
```
nr = 100
sexo = 1
orden = "DESC"
# OJO con la f antes del valor d ela cadena, le dice que va a recibir parámetros
sql = f"""
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
WHERE sexo= '{sexo}'
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA
HAVING nreg  >= {nr}
ORDER BY 3 {orden}
"""
spark.sql(sql).show(35)
```

```
+-------------+--------------------+----+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|nreg|
+-------------+--------------------+----+
|           09|    CIUDAD DE MÉXICO|2899|
|           19|          NUEVO LEÓN| 539|
|           08|           CHIHUAHUA| 479|
|           11|          GUANAJUATO| 422|
|           05|COAHUILA DE ZARAGOZA| 318|
|           15|              MÉXICO| 312|
|           24|     SAN LUIS POTOSÍ| 293|
|           14|             JALISCO| 280|
|           27|             TABASCO| 261|
|           10|             DURANGO| 250|
|           22|           QUERÉTARO| 222|
|           21|              PUEBLA| 221|
|           16| MICHOACÁN DE OCAMPO| 209|
|           28|          TAMAULIPAS| 200|
|           03| BAJA CALIFORNIA SUR| 168|
|           02|     BAJA CALIFORNIA| 165|
|           32|           ZACATECAS| 161|
|           30|VERACRUZ DE IGNAC…| 156|
|           25|             SINALOA| 154|
|           26|              SONORA| 153|
|           31|             YUCATÁN| 139|
```

```
|           12|            GUERRERO| 130|
|           01|      AGUASCALIENTES| 119|
|           13|             HIDALGO| 107|
|           20|              OAXACA| 104|
+-------------+--------------------+----+
```

[22]:
```
nr = 200
sexo = 2
orden = "DESC"
# OJO con la f antes del valor d ela cadena, le dice que va a recibir parámetros
sql = f"""
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
WHERE sexo= '{sexo}'
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA
HAVING nreg  >= {nr}
ORDER BY 3 {orden}
"""
spark.sql(sql).show(35)
```

```
+-------------+--------------------+----+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|nreg|
+-------------+--------------------+----+
|           09|    CIUDAD DE MÉXICO|2709|
|           19|          NUEVO LEÓN| 511|
|           08|           CHIHUAHUA| 478|
|           11|          GUANAJUATO| 385|
|           15|              MÉXICO| 311|
|           14|             JALISCO| 246|
|           21|              PUEBLA| 245|
|           05|COAHUILA DE ZARAGOZA| 238|
|           27|             TABASCO| 233|
|           24|     SAN LUIS POTOSÍ| 228|
+-------------+--------------------+----+
```

[ ]:

[19]:
```
sql = """
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, SEXO, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, SEXO
ORDER BY 4 DESC
"""
spark.sql(sql).show(35)
```

```
[Stage 31:============================>                            (1 + 1) / 2]
```

```
+-------------+--------------------+----+----+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|SEXO|nreg|
+-------------+--------------------+----+----+
|           09|    CIUDAD DE MÉXICO|   1|2899|
|           09|    CIUDAD DE MÉXICO|   2|2709|
|           19|          NUEVO LEÓN|   1| 539|
|           19|          NUEVO LEÓN|   2| 511|
|           08|           CHIHUAHUA|   1| 479|
|           08|           CHIHUAHUA|   2| 478|
|           11|          GUANAJUATO|   1| 422|
|           11|          GUANAJUATO|   2| 385|
|           05|COAHUILA DE ZARAGOZA|   1| 318|
|           15|              MÉXICO|   1| 312|
|           15|              MÉXICO|   2| 311|
|           24|     SAN LUIS POTOSÍ|   1| 293|
|           14|             JALISCO|   1| 280|
|           27|             TABASCO|   1| 261|
|           10|             DURANGO|   1| 250|
|           14|             JALISCO|   2| 246|
|           21|              PUEBLA|   2| 245|
|           05|COAHUILA DE ZARAGOZA|   2| 238|
|           27|             TABASCO|   2| 233|
|           24|     SAN LUIS POTOSÍ|   2| 228|
|           22|           QUERÉTARO|   1| 222|
|           21|              PUEBLA|   1| 221|
|           16| MICHOACÁN DE OCAMPO|   1| 209|
|           28|          TAMAULIPAS|   1| 200|
|           10|             DURANGO|   2| 197|
|           22|           QUERÉTARO|   2| 172|
|           03| BAJA CALIFORNIA SUR|   1| 168|
|           02|     BAJA CALIFORNIA|   1| 165|
|           32|           ZACATECAS|   1| 161|
|           30|VERACRUZ DE IGNAC…|   1| 156|
|           28|          TAMAULIPAS|   2| 156|
|           25|             SINALOA|   1| 154|
|           03| BAJA CALIFORNIA SUR|   2| 153|
|           26|              SONORA|   1| 153|
|           02|     BAJA CALIFORNIA|   2| 150|
+-------------+--------------------+----+----+
only showing top 35 rows
```

[20]:
```
sql = """
SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA,
SUM(IF(SEXO = 1, nreg, 0)) MASCULINO,
SUM(IF(SEXO = 2, nreg, 0)) FEMENINO
```

```
FROM
(SELECT CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, SEXO, COUNT(*) as nreg
FROM covid_df c JOIN estados e ON(c.ENTIDAD_UM = e.CLAVE_ENTIDAD)
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA, SEXO)
GROUP BY CLAVE_ENTIDAD, ENTIDAD_FEDERATIVA
ORDER BY 1
"""
spark.sql(sql).show(70)
```

```
+-------------+--------------------+---------+--------+
|CLAVE_ENTIDAD|  ENTIDAD_FEDERATIVA|MASCULINO|FEMENINO|
+-------------+--------------------+---------+--------+
|           01|      AGUASCALIENTES|      119|     105|
|           02|     BAJA CALIFORNIA|      165|     150|
|           03| BAJA CALIFORNIA SUR|      168|     153|
|           04|            CAMPECHE|       95|      94|
|           05|COAHUILA DE ZARAGOZA|      318|     238|
|           06|              COLIMA|       67|      48|
|           07|             CHIAPAS|       62|      31|
|           08|           CHIHUAHUA|      479|     478|
|           09|    CIUDAD DE MÉXICO|     2899|    2709|
|           10|             DURANGO|      250|     197|
|           11|          GUANAJUATO|      422|     385|
|           12|            GUERRERO|      130|      77|
|           13|             HIDALGO|      107|      94|
|           14|             JALISCO|      280|     246|
|           15|              MÉXICO|      312|     311|
|           16| MICHOACÁN DE OCAMPO|      209|     126|
|           17|             MORELOS|       73|      37|
|           18|             NAYARIT|       30|      26|
|           19|          NUEVO LEÓN|      539|     511|
|           20|              OAXACA|      104|      86|
|           21|              PUEBLA|      221|     245|
|           22|           QUERÉTARO|      222|     172|
|           23|        QUINTANA ROO|       69|      56|
|           24|     SAN LUIS POTOSÍ|      293|     228|
|           25|             SINALOA|      154|     111|
|           26|              SONORA|      153|     122|
|           27|             TABASCO|      261|     233|
|           28|          TAMAULIPAS|      200|     156|
|           29|            TLAXCALA|       51|      52|
|           30|VERACRUZ DE IGNAC…|      156|     121|
|           31|             YUCATÁN|      139|     149|
|           32|           ZACATECAS|      161|     132|
+-------------+--------------------+---------+--------+
```

[ ]:

[23]:
```
sql = """
SELECT ENTIDAD_UM, SEXO, SUM(EDAD) EDAD_SUMA, AVG(EDAD) EDAD_PROMEDIO,␣
 ↪MAX(EDAD) EDAD_MAXIMA, MIN(EDAD) EDAD_MINIMA
FROM covid_df
GROUP BY ENTIDAD_UM, SEXO
HAVING SUM(EDAD) > 0
ORDER BY 4 DESC
"""
spark.sql(sql).show()
```

```
+----------+----+---------+------------------+-----------+-----------+
|ENTIDAD_UM|SEXO|EDAD_SUMA|     EDAD_PROMEDIO|EDAD_MAXIMA|EDAD_MINIMA|
+----------+----+---------+------------------+-----------+-----------+
|        18|   1|     1426| 47.53333333333333|         83|         15|
|        18|   2|     1220| 46.92307692307692|         84|          8|
|        20|   2|     3790| 44.06976744186046|         82|         13|
|        14|   2|    10829|44.020325203252035|         97|          1|
|        07|   2|     1359| 43.83870967741935|         87|          0|
|        20|   1|     4511|            43.375|         94|          0|
|        25|   2|     4778| 43.04504504504504|         89|          0|
|        26|   2|     5235| 42.90983606557377|         90|          9|
|        30|   2|     5192| 42.90909090909091|         98|          0|
|        15|   1|    13333| 42.73397435897436|         89|          0|
|        13|   2|     3990|  42.4468085106383|         83|          7|
|        28|   2|     6613| 42.39102564102564|         98|          0|
|        06|   2|     2020|42.083333333333336|         78|         10|
|        17|   2|     1555|42.027027027027025|         84|          0|
|        30|   1|     6534| 41.88461538461539|         97|          5|
|        28|   1|     8333|            41.665|         85|          9|
|        08|   2|    19910| 41.65271966527197|         92|          0|
|        09|   2|   112557| 41.54928017718716|         97|          0|
|        13|   1|     4441|41.504672897196265|         79|         10|
|        29|   2|     2157| 41.48076923076923|         92|         17|
+----------+----+---------+------------------+-----------+-----------+
only showing top 20 rows
```

[ ]:

[ ]:

[ ]: