```
 1 /Users/gorgeousdan/Documents/projects/intellijidea/libs/spark_practice/bin/python /Users/gorgeousdan/Documents/projects/intellijidea/
   spark_practice/old_ETL_job.py
 2 Setting default log level to "WARN".
 3 To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
 4 24/12/10 12:25:08 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where
   applicable
 5 24/12/10 12:25:09 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
 6 Restaurant data schema:
 7 root
 8  |-- id: long (nullable = true)
 9  |-- franchise_id: integer (nullable = true)
10  |-- franchise_name: string (nullable = true)
11  |-- restaurant_franchise_id: integer (nullable = true)
12  |-- country: string (nullable = true)
13  |-- city: string (nullable = true)
14  |-- lat: double (nullable = true)
15  |-- lng: double (nullable = true)
16
17 Weather data schema:
18 root
19  |-- lng: double (nullable = true)
20  |-- lat: double (nullable = true)
21  |-- avg_tmpr_f: double (nullable = true)
22  |-- avg_tmpr_c: double (nullable = true)
23  |-- wthr_date: string (nullable = true)
24
25 Sample restaurant data:
26 +------------+------------+---------------+-----------------------+-------+----------+------+-------+
27 |          id|franchise_id| franchise_name|restaurant_franchise_id|country|      city|   lat|    lng|
28 +------------+------------+---------------+-----------------------+-------+----------+------+-------+
29 |197568495625|          10|The Golden Spoon|                 24784|     US|   Decatur|34.578|-87.021|
30 | 17179869242|          59|    Azalea Cafe|                  10902|     FR|     Paris|48.861|  2.368|
31 |214748364826|          27| The Corner Cafe|                 92040|     US|Rapid City| 44.08|-103.25|
32 |154618822706|          51|    The Pizzeria|                 41484|     AT|    Vienna|48.213| 16.413|
33 |163208757312|          65|   Chef's Corner|                 96638|     GB|    London|51.495| -0.191|
34 +------------+------------+---------------+-----------------------+-------+----------+------+-------+
35 only showing top 5 rows
36
37 Sample weather data:
38 +--------+-------+----------+----------+----------+
39 |     lng|    lat|avg_tmpr_f|avg_tmpr_c| wthr_date|
40 +--------+-------+----------+----------+----------+
41 | -111.09|18.6251|      80.7|      27.1|2017-08-29|
```

```
42 |-111.042|18.6305|        80.7|        27.1|2017-08-29|
43 |-110.995|18.6358|        80.7|        27.1|2017-08-29|
44 |-110.947|18.6412|        80.9|        27.2|2017-08-29|
45 |  -110.9|18.6465|        80.9|        27.2|2017-08-29|
46 +--------+-------+----------+----------+----------+
47 only showing top 5 rows
48
49 Rows with invalid coordinates:
50 +----------+------------+-------------+---------------------+-------+------+----+----+
51 |        id|franchise_id|franchise_name|restaurant_franchise_id|country|  city| lat| lng|
52 +----------+------------+-------------+---------------------+-------+------+----+----+
53 |85899345920|          1|      Savoria|                18952|     US|Dillon|NULL|NULL|
54 +----------+------------+-------------+---------------------+-------+------+----+----+
55
56 Checking coordinates for city: Dillon, US...
57 API response: lat=34.4014089, lng=-79.3864339
58 Restaurant data with geohashes:
59 +------------+------------+------------------+---------------------+-------+--------------+------+-------+-------+
60 |          id|franchise_id|      franchise_name|restaurant_franchise_id|country|          city|   lat|    lng|geohash|
61 +------------+------------+------------------+---------------------+-------+--------------+------+-------+-------+
62 |197568495625|          10|   The Golden Spoon|                24784|     US|       Decatur|34.578|-87.021|   dn4h|
63 | 17179869242|          59|       Azalea Cafe|                10902|     FR|         Paris|48.861|  2.368|   u09t|
64 |214748364826|          27|   The Corner Cafe|                92040|     US|    Rapid City| 44.08|-103.25|   9xyd|
65 |154618822706|          51|      The Pizzeria|                41484|     AT|        Vienna|48.213| 16.413|   u2ed|
66 |163208757312|          65|     Chef's Corner|                96638|     GB|        London|51.495| -0.191|   gcpu|
67 | 68719476763|          28|   The Spicy Pickle|                77517|     US|      Grayling|44.657|-84.744|   dpgw|
68 |223338299419|          28|   The Spicy Pickle|                36937|     US|        Oswego|43.452|-76.532|   dr9x|
69 |240518168650|          75|    Greenhouse Cafe|                93164|     NL|     Amsterdam| 52.37|  4.897|   u173|
70 |128849018936|          57|The Yellow Submarine|                 5679|     FR|         Paris|48.872|  2.335|   u09w|
71 |197568495635|          20|      The Brasserie|                24784|     US|Jeffersonville|39.616|-83.612|   dph9|
72 | 68719476768|          33|  The Blue Elephant|                77517|     IT|         Milan|45.479|  9.146|   u0nd|
73 | 51539607582|          31|         Bistro 42|                 6934|     IT|         Milan|45.444|  9.153|   u0nd|
74 | 94489280554|          43|    The Food House|                95399|     FR|         Paris|48.867|  2.329|   u09t|
75 |206158430215|           8|    The Green Olive|                53370|     US|   Haltom City|32.789| -97.28|   9vff|
76 |154618822657|           2|      Bella Cucina|                41484|     US|   Fort Pierce|27.412|-80.391|   dhyg|
77 | 17179869217|          34|    The Tasty Treat|                10902|     US|     Green Bay|44.476|-88.077|   dpcm|
78 |  8589934633|          42|    The Daily Scoop|                12630|     FR|         Paris|48.854|  2.343|   u09t|
79 |240518168596|          21|     The Lazy Daisy|                93164|     US|    Mendenhall| 39.86|-75.646|   dr44|
80 |171798691906|          67| Crimson and Clover|                65939|     NL|     Amsterdam|52.361|  4.894|   u173|
81 | 42949673022|          63|        Cafe Paris|                89646|     GB|        London|51.508| -0.107|   gcpv|
82 +------------+------------+------------------+---------------------+-------+--------------+------+-------+-------+
83 only showing top 20 rows
84
```

```
 85 Weather data with geohashes:
 86 +--------+-------+----------+----------+----------+-------+
 87 |     lng|    lat|avg_tmpr_f|avg_tmpr_c| wthr_date|geohash|
 88 +--------+-------+----------+----------+----------+-------+
 89 | -111.09|18.6251|      80.7|      27.1|2017-08-29|   9e31|
 90 |-111.042|18.6305|      80.7|      27.1|2017-08-29|   9e31|
 91 |-110.995|18.6358|      80.7|      27.1|2017-08-29|   9e34|
 92 |-110.947|18.6412|      80.9|      27.2|2017-08-29|   9e34|
 93 |  -110.9|18.6465|      80.9|      27.2|2017-08-29|   9e34|
 94 |-110.852|18.6518|      80.9|      27.2|2017-08-29|   9e34|
 95 |-110.804|18.6571|      80.9|      27.2|2017-08-29|   9e34|
 96 |-105.068|19.1765|      82.4|      28.0|2017-08-29|   9emm|
 97 | -105.02|19.1799|      82.0|      27.8|2017-08-29|   9emm|
 98 |-104.972|19.1832|      82.0|      27.8|2017-08-29|   9emm|
 99 |-104.924|19.1866|      82.0|      27.8|2017-08-29|   9emm|
100 |-104.876|19.1899|      82.0|      27.8|2017-08-29|   9emm|
101 |-104.828|19.1932|      81.6|      27.6|2017-08-29|   9emm|
102 | -104.78|19.1964|      81.6|      27.6|2017-08-29|   9emm|
103 |-104.732|19.1997|      81.6|      27.6|2017-08-29|   9emt|
104 |-104.684| 19.203|      77.8|      25.4|2017-08-29|   9emt|
105 |-104.636|19.2062|      77.8|      25.4|2017-08-29|   9emt|
106 |-104.588|19.2094|      75.2|      24.0|2017-08-29|   9emt|
107 | -104.54|19.2126|      75.2|      24.0|2017-08-29|   9emt|
108 |-104.492|19.2158|      75.2|      24.0|2017-08-29|   9emt|
109 +--------+-------+----------+----------+----------+-------+
110 only showing top 20 rows
111
112 Sample of joined data:
113 Checking coordinates for city: Dillon, US...
114 API response: lat=34.4014089, lng=-79.3864339
115 24/12/10 12:25:19 WARN GarbageCollectionMetrics: To enable non-built-in garbage collector(s) List(G1 Concurrent GC), users should
    configure it(them) to spark.eventLog.gcMetrics.youngGenerationGarbageCollectors or spark.eventLog.gcMetrics.
    oldGenerationGarbageCollectors
116 +-------+------------+------------+----------------+---------------------+-------+---------+------+--------+----------+----------+
    ----------+
117 |geohash|id          |franchise_id|franchise_name  |restaurant_franchise_id|country|city     |lat   |lng     |avg_tmpr_f|avg_tmpr_c|
    wthr_date |
118 +-------+------------+------------+----------------+---------------------+-------+---------+------+--------+----------+----------+
    ----------+
119 |dnqx   |103079215121|18          |The Harvest Room |4340                 |US     |Mount Airy|36.481|-80.602 |66.4      |19.1      |
    2017-08-29|
120 |9r13   |214748364814|15          |The Pearl Kitchen|92040                |US     |Chico    |39.726|-121.837|79.6      |26.4      |
    2017-09-10|
```

```
121 |9y7v    |103079215120|17              |The Rustic Table |4340                  |US     |Tulsa    |36.168|-95.851 |74.6    |23.7     |
    2017-08-29|
122 |dn4h    |197568495625|10              |The Golden Spoon |24784                 |US     |Decatur  |34.578|-87.021 |66.5    |19.2     |
    2017-09-10|
123 |gcpv    |171798691900|61              |Blue Water Bistro|65939                 |GB     |London   |51.519|-0.087  |59.8    |15.4     |
    2017-08-13|
124 +-------+------------+------------+----------------+----------------------+-------+---------+------+--------+---------+----------+
    ----------+
125 only showing top 5 rows
126
127 Checking coordinates for city: Dillon, US...
128 API response: lat=34.4014089, lng=-79.3864339
129 Joined data saved at /Users/gorgeousdan/Documents/projects/intellijidea/spark_practice/enriched_data.parquet
130 Running validation tests...
131 Checking coordinates for city: Dillon, US...
132 API response: lat=34.4014089, lng=-79.3864339
133 Checking coordinates for city: Dillon, US...
134 API response: lat=34.4014089, lng=-79.3864339
135 Checking coordinates for city: Dillon, US...
136 API response: lat=34.4014089, lng=-79.3864339
137 Checking coordinates for city: Dillon, US...
138 API response: lat=34.4014089, lng=-79.3864339
139 All tests passed!
140
141 Process finished with exit code 0
142
```