

Analysis of data about weather stations in Cordova

José Luis Puente Bodoque

June 21, 2023

Abstract

In this report I am going to show my configuration in Amazon Web Services to carry out the analysis of this assignment.

Contents

1	Amazon S3: directory hierarchy of my bucket	3
2	AWS Glue: Data Catalog	3
3	Analysis of data: queries	5
3.1	Query 1	5
3.2	Query 2	6
3.3	Query 3	6
3.4	Query 4	7
3.5	Query 5	8
3.6	Query 6	8
3.7	Query 7	9
3.8	Query 8	10
3.9	Query 9	12
3.10	Query 10	12

List of Figures

1	My bucket	3
2	File of coordinates	3
3	Data file	3

4	Overview of my table	4
5	Schema of my table	4
6	Crawler for data	5
7	Query 1. SQL code	5
8	Query 1. Result	5
9	Query 2	6
10	Query 3	6
11	Query 4 in Athena	7
12	Query 4 in QuickSight	7
13	Query 5	8
14	Query 6	8
15	Query 7	10
16	Query 8 in Athena	11
17	Query 8 in QuickSight	11
18	Query 9	12
19	Query 10. SQL code	13
20	Query 10. Result	13

1 Amazon S3: directory hierarchy of my bucket

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
○	mbd-jlpuente-bucket	EU (Paris) eu-west-3	Objects can be public	June 15, 2023, 20:54:38 (UTC+02:00)

Figure 1: My bucket

Amazon S3 > Buckets > mbd-jlpuente-bucket > RIAcordoba/ > coordinates/						
coordinates/						
Copy S3 URI						
Objects Properties						
Objects (1)						
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more						
Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload						
Find objects by prefix						
< 1 > ⚙						
<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼	
<input type="checkbox"/>	ria_cordoba_localizacion_estaciones.csv	csv	June 17, 2023, 14:05:36 (UTC+02:00)	12.7 KB	Standard	

Figure 2: File of coordinates

Amazon S3 > Buckets > mbd-jlpuente-bucket > RIAcordoba/ > data/						
data/						
Copy S3 URI						
Objects Properties						
Objects (1)						
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more						
Refresh Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload						
Find objects by prefix						
< 1 > ⚙						
<input type="checkbox"/>	Name ▲	Type ▼	Last modified ▼	Size ▼	Storage class ▼	
<input type="checkbox"/>	RIA_exportacion_datos_diarios_Cordoba_20140206_punto_decimal_relleno_fecha_unix.csv	csv	June 17, 2023, 14:06:15 (UTC+02:00)	4.7 MB	Standard	

Figure 3: Data file

2 AWS Glue: Data Catalog

At this point I created a table and specified manually its schema. Then I created a crawler to automatically adjust the data type of the table fields in order to not receive an error when launching my first query in Athena.

AWS Glue > Tables

Tables

A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Tables (1)
Last updated (UTC)
June 20, 2023 at 09:40:02
Delete
Data quality New
Add tables using crawler
Add table

View and manage all available tables.

Filter tables

<input type="checkbox"/>	Name	Database	Location	Classification	Deprecated	View data
<input type="checkbox"/>	datos	riacordova-db	s3://mbd-jlpuente-bucket/f	csv	-	Table data

Figure 4: Overview of my table

The schema of the table was modified after running the crawler.

Schema (18) Edit schema as JSON Edit schema

View and manage the table schema.

Filter schemas

#	Column name	Data type	Partition key	Comment
1	idprovincia	bigint	-	-
2	sprovincia	string	-	-
3	idestacion	bigint	-	-
4	sestacion	string	-	-
5	fecha	string	-	-
6	día	bigint	-	-
7	tempmax	double	-	-
8	hormintempmax	bigint	-	-
9	tempmin	double	-	-
10	hormintempmin	bigint	-	-
11	tempmedia	double	-	-
12	humedadmax	double	-	-
13	humedadmin	double	-	-
14	humedadmedia	double	-	-
15	velviento	double	-	-
16	dirviento	double	-	-
17	radiacion	double	-	-
18	precipitacion	double	-	-

Figure 5: Schema of my table

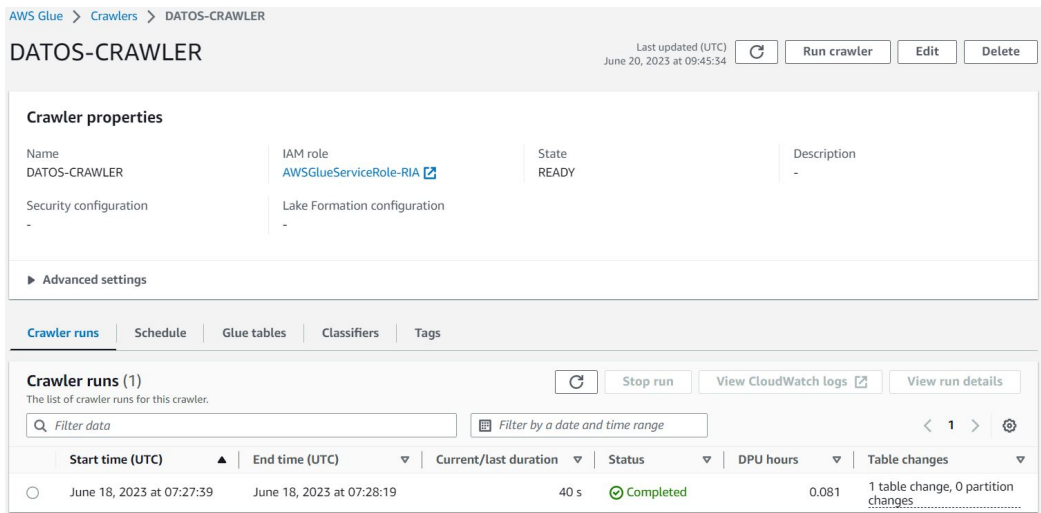


Figure 6: Crawler for data

3 Analysis of data: queries

3.1 Query 1

List of names of weather stations in Cordova.

```
1 SELECT DISTINCT sestacion
2 FROM datos;
```

Figure 7: Query 1. SQL code

The screenshot shows the results of the SQL query in the AWS Glue console. The results are displayed in a table with 12 rows, each containing a unique station name.

#	sestacion
1	IFAPA Centro de Cabra
2	Baena
3	Palma del Rio
4	SESTACION
5	Adarnuz
6	Palma del Rio
7	Hornachuelos
8	El Carpio
9	Báñez
10	IFAPA Centro de Hinojosa del Duque
11	Córdoba
12	Santaelia

Figure 8: Query 1. Result

3.2 Query 2

List of days when it rained.

Query 11

```
1 SELECT fecha
2 FROM datos
3 WHERE precipitacion > 0;
```

SQL Ln 3, Col 25

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 111 ms Run time: 524 ms Data scanned: 4.70 MB

Results (13,388) Copy Download results

Search rows

#	fecha
1	2000-04-25
2	2000-04-26
3	2000-04-27
4	2000-04-30

Figure 9: Query 2

3.3 Query 3

Name of the station, minimum temperature and date of the days when minimum temperature was below 10°C.

Query 12

```
1 SELECT sestacion, tempmin, fecha
2 FROM datos
3 WHERE tempmedia < 10;
```

SQL Ln 3, Col 22

Run Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 116 ms Run time: 582 ms Data scanned: 4.70 MB

Results (9,731) Copy Download results

Search rows

#	sestacion	tempmin	fecha
1	Bélmez	6.89	2000-04-27
2	Bélmez	4.21	2000-04-28
3	Bélmez	1.13	2000-11-01
4	Bélmez	3.2	2000-11-03
5	Bélmez	0.06	2000-11-04

Figure 10: Query 3

3.4 Query 4

Maximum of recorded rain in each station. Bar chart with that information.

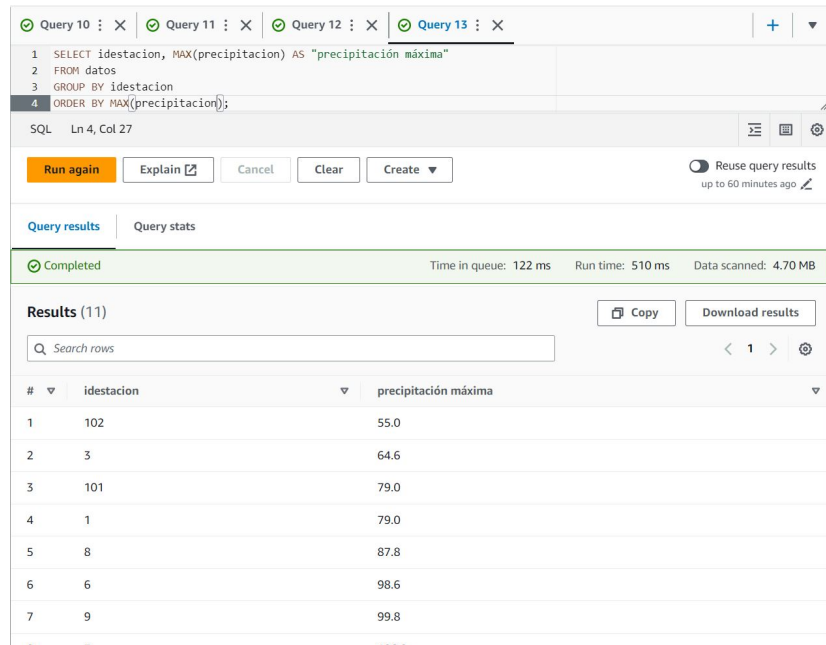


Figure 11: Query 4 in Athena

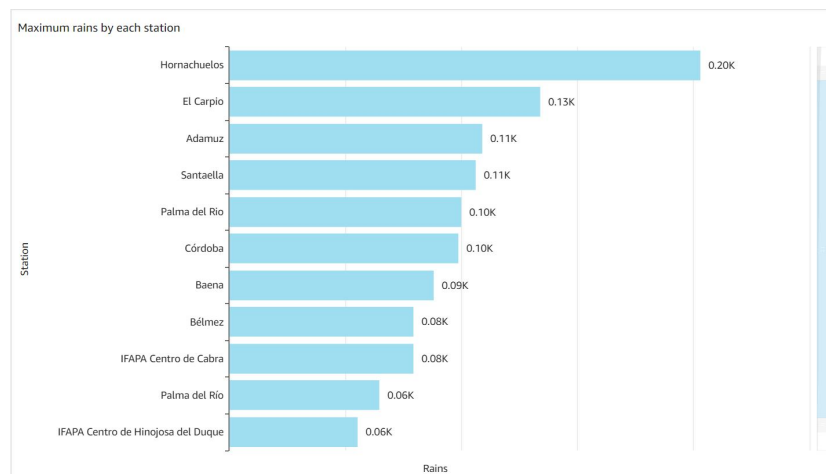


Figure 12: Query 4 in QuickSight

3.5 Query 5

Bar chart with total rain for each station.

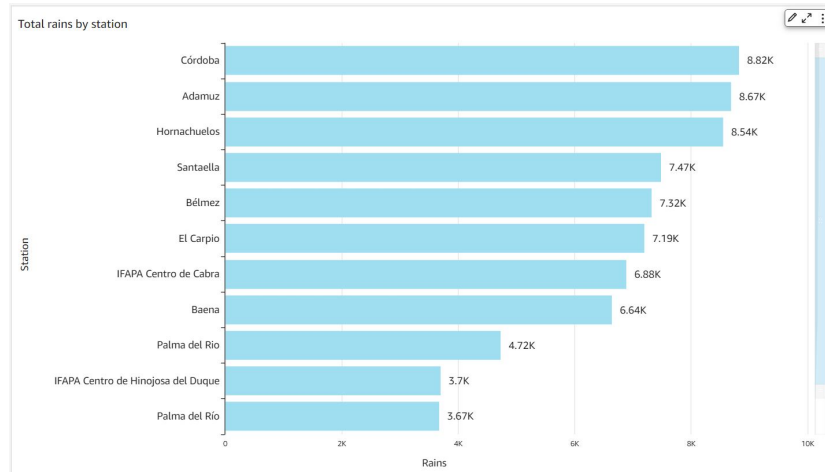


Figure 13: Query 5

3.6 Query 6

Mean maximum temperature by day.

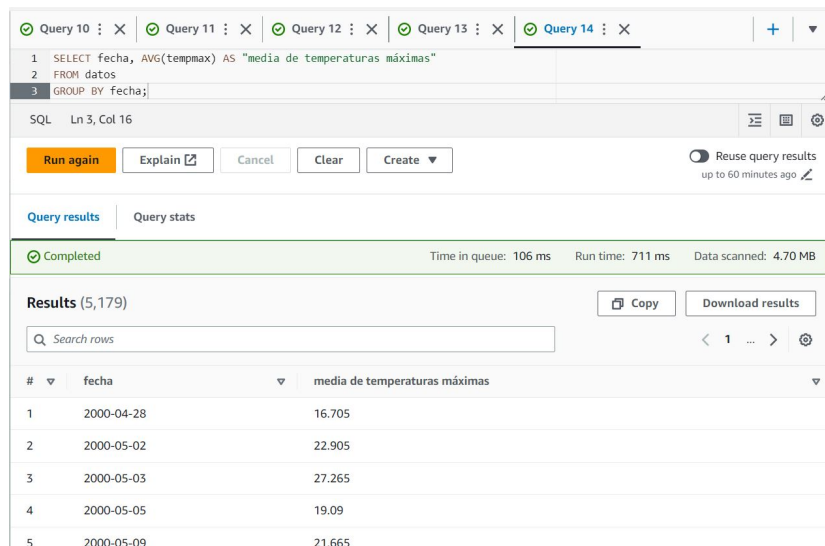


Figure 14: Query 6

3.7 Query 7

Map showing the total rain for each station. In order to convert bad degree, minutes and seconds format into decimal format, I have typed the Python code below.

```
1  import csv
2
3
4  def convert_coordinates(coordinate):
5      direction = coordinate[-1]
6      coordinate = coordinate[:-1]
7      degrees = float(coordinate[:2])
8      minutes = float(coordinate[2:4])
9      seconds = float(coordinate[4:6])
10     decimal = degrees + minutes/60 + seconds/3600
11
12     if direction in ['S', 'W']:
13         decimal *= -1
14
15     return decimal
16
17
18 def convert_csv_coordinates(input_file, output_file):
19     with open(input_file, 'r') as file:
20         reader = csv.DictReader(file, delimiter=';')
21         headers = reader.fieldnames
22
23         # Add new headers for decimal coordinates
24         headers.extend(['DEC_LATITUDE', 'DEC_LONGITUDE'])
25
26         rows = []
27         for row in reader:
28             latitude = row['SLATITUD']
29             longitude = row['SLONGITUD']
30             new_latitude = convert_coordinates(latitude)
31             new_longitude = convert_coordinates(longitude)
32             row['DEC_LATITUDE'] = new_latitude
33             row['DEC_LONGITUDE'] = new_longitude
34
35             rows.append(row)
```

```

36
37     with open(output_file, 'w', newline='') as file:
38         writer = csv.DictWriter(file, fieldnames=headers)
39         writer.writeheader()
40         writer.writerows(rows)
41
42
43 if __name__ == '__main__':
44
45     input_file = 'data/ria_exportacion_localizacion_estaciones_20140205.csv'
46     output_file = 'data/coordenadas_ria.csv'
47     convert_csv_coordinates(input_file, output_file)

```



Figure 15: Query 7

3.8 Query 8

Mean maximum temperature by month in the stations of Bélmez, Hornachuelos and Palma del Río. Line chart with that information.

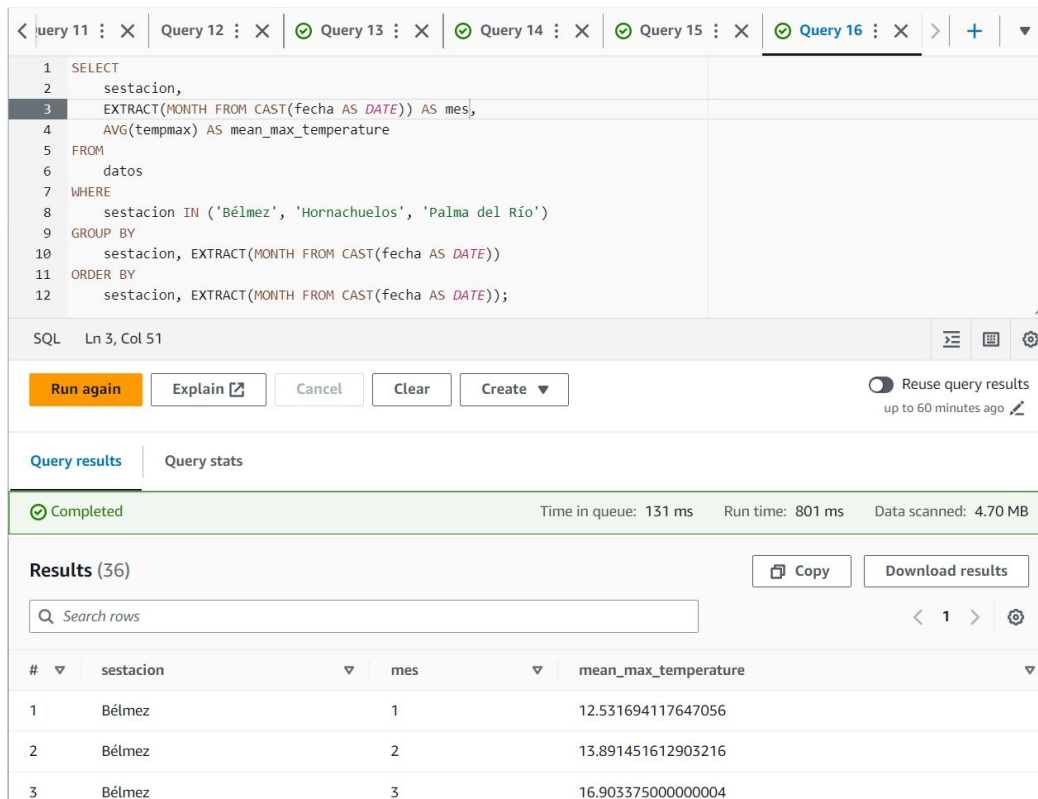


Figure 16: Query 8 in Athena

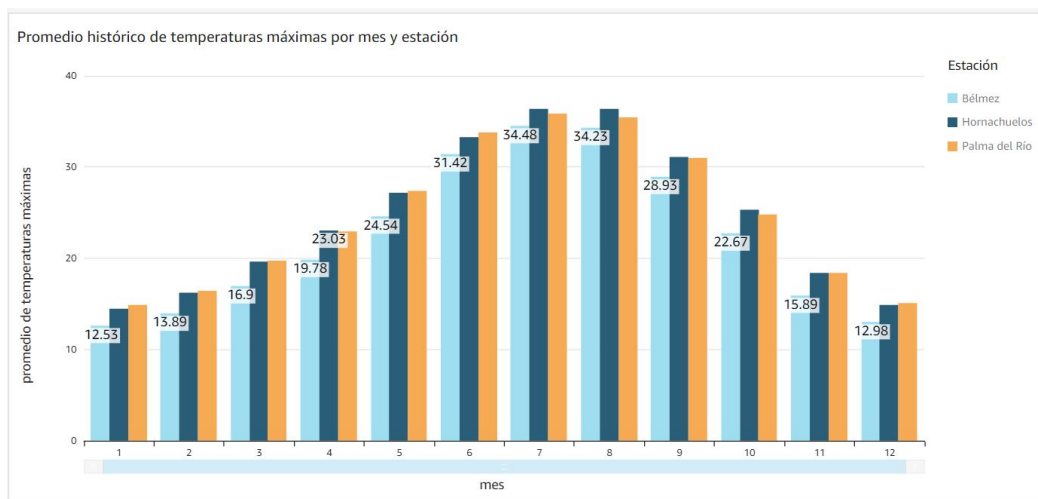


Figure 17: Query 8 in QuickSight

3.9 Query 9

Mean maximum temperature by hour for each station. Group by station and hour of the day. Keep in mind that in the data sources files, hour and minute are represented by a 4 digit number and the division (/) has to be used to extract the hour.

SQL Ln 3, Col 39

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed Time in queue: 128 ms Run time: 839 ms Data scanned: 4.70 MB

Results (250) Copy Download results

Search rows

#	idestacion	hora	media de temperaturas máximas
1	1	14	21.41619912790696
2	1	1	13.269166666666669
3	1	3	12.7025
4	1	5	10.81
5	101	12	18.5124255319149

Figure 18: Query 9

3.10 Query 10

Mean temperature depending on the wind direction. Four values are required: north wind (between 315° and 45°), east wind (between 45° and 135°), south wind (between 135° and 225°), and west wind (between 225° and 315°).

```
Query 10 : X Query 11 : X Query 12 : X Query 13 : X Query 14 : X Query 15 : X Query 16 : X Query 17 : X + ▼
1 SELECT
2 fecha,
3 estacion,
4 CASE
5     WHEN dirviento >= 315 OR dirviento < 45 THEN 'North Wind'
6     WHEN dirviento >= 45 AND dirviento < 135 THEN 'East Wind'
7     WHEN dirviento >= 135 AND dirviento < 225 THEN 'South Wind'
8     WHEN dirviento >= 225 AND dirviento < 315 THEN 'West Wind'
9     END AS "dirección del viento",
10     AVG(tempmedia) AS "temperatura media"
11 FROM datos
12 GROUP BY fecha, estacion, dirviento
13 ORDER BY fecha;
SQL Ln 1, Col 1
```

Figure 19: Query 10. SQL code

Completed Time in queue: 112 ms Run time: 859 ms Data scanned: 4.70 MB

Results (45,581)

Search rows

Copy Download results

< 1 ... > ⚙

#	fecha	estacion	dirección del viento	temperatura media
1	1999-12-02	Adamuz	North Wind	11.81
2	1999-12-03	Adamuz	North Wind	10.84
3	1999-12-04	Adamuz	North Wind	9.77
4	1999-12-05	Adamuz	North Wind	11.18
5	1999-12-06	Adamuz	North Wind	8.92
6	1999-12-07	Adamuz	North Wind	8.65
7	1999-12-08	Adamuz	North Wind	6.81
8	1999-12-09	Adamuz	West Wind	9.11
9	1999-12-10	Adamuz	North Wind	7.6
10	1999-12-11	Adamuz	West Wind	9.23
11	1999-12-12	Adamuz	South Wind	10.91
12	1999-12-13	Adamuz	South Wind	13.94
13	1999-12-14	Adamuz	West Wind	14.56
14	1999-12-15	Adamuz	West Wind	10.38
15	1999-12-16	Adamuz	North Wind	5.15

Figure 20: Query 10. Result