

Datos historicos de la calidad del aire:

Obtención, modificación e insercción en la base de datos

Por Jonatan Álvarez Jiménez

IES Punta Del Verde

Nota del autor

Trabajo final de proyecto en conjunto con el programa Erasmus+ de la union europea



Cofinanciado por el
programa Erasmus+
de la Unión Europea



Índice

Resumen.....	3
Requerimientos.	3
Obtención de los datos históricos.	4
Estaciones que miden datos de calidad de aire	4
Obtención de los datos de los contaminantes por localidades	8
Adecuación de los datos obtenidos mediante Excell.	13
Adecuación de los datos obtenidos mediante OpenOffice.	16
Insercción en la base de datos MongoDB en local.	21
Exportación desde MongoDB Compass de los datos a formato json.	22
Insercción en la base de datos Mongo Atlas en la nube.....	23

Resumen.

Informe detallado de como descargar los datos históricos, y de donde, como trabajar con ellos y llegar a la estructura deseada para facilitar la insercción y el estudio de los mismos en una base de datos y aplicación propias.

Requerimientos.

Acceso a internet para la descarga de los archivos.

Procesador de hojas de cálculo (Excel, LibreOffice o similares).

Bloc de notas.

MongoDB como gestor de bases de datos.

Obtención de los datos históricos.

Estaciones que miden datos de calidad de aire


Para la obtención de los datos recurriremos a varios sitios web oficiales de la EEA (Enviroment European Agency).

Primero entraremos en:


http://aided.apps.eea.europa.eu/?source=%7B%22query%22%3A%7B%22match_all%22%3A%7B%7D%7D%2C%22display_type%22%3A%22tabular%22%7D


En este enlace tendremos todos los datos necesarios para encontrar e identificar las estaciones que deseemos utilizar en nuestro proyecto.

En la pagina derecha de la pagina tendremos un filtro en el cual nosotros a mano pondremos los datos necesarios.


European Environment Agency

Topics
Countries
Data and maps
Indicators
Publications
Media
About us

EN
Q Search



>
aqregimes

Air quality assessment methods (data flow D)

Information on air quality assessment methods, data flow D.

Results 1 - 10 of 58574
Order: Relevance
Download TSV
Download CSV

Country or Territory	Namespace	Network	Responsible Party	Time Zone	Station Name	Eol Station...	Station Type	Station Area	Pollutant	Sampling Point Local Id	Latitude	Longitude	Altitude	Measurement Regime	Operat Activity
United Kingdom	http://environment.data.gov.uk/air-quality/so	UK Air Quality e-Reporting	The Department for Environment, Food and Rural Affairs	UTC	Billingham	GB0421A	Industrial	Urban	Nitrogen dioxide (air)	GB_SamplingPoint_42	54.60537	-1.275039	10	Continuous data collection	1987-01 01T00:00
United Kingdom	http://environment.data.gov.uk/air-quality/so	UK Air Quality e-Reporting	The Department for Environment, Food and Rural Affairs	UTC	Billingham - UKA00153	GB0421A	Industrial	Urban	Nitrogen dioxide (air)	GB_SamplingPoint_42	54.60537	-1.275039	10	Continuous data collection	1987-01 01T00:00
United Kingdom	http://environment.data.gov.uk/air-quality/so	UK Air Quality e-Reporting	The Department for Environment, Food and Rural Affairs	UTC	Aston Hill	GB0031R	Background	Rural	Ozone (air)	GB_SamplingPoint_49	52.50385	-3.034178	370	Continuous data collection	1986-06 26T00:00
United Kingdom	http://environment.data.gov.uk/air-quality/so	UK Air Quality e-Reporting	The Department for Environment, Food and Rural Affairs	UTC	Lullington Heath - UKA00152	GB0038R	Background	Rural	Nitrogen dioxide (air)	GB_SamplingPoint_51	50.7937	0.18125	125	Continuous data collection	1988-09 29T00:00
United Kingdom	http://environment.data.gov.uk/air-quality/so	UK Air Quality e-Reporting	The Department for Environment, Food and Rural Affairs	UTC	Strathvaich - UKA00162	GB0015R	Background	Rural	Ozone (air)	GB_SamplingPoint_53	57.734456	-4.776583	270	Continuous data collection	1987-03 18T00:00

Como ejemplo voy a poner los datos necesarios para buscar la estación que usaremos nosotros de Bulgaria, Sofia-Druzhba:

Country or Territory		
<input type="text"/> Search for Country or Territory		
Count	Value	Match all
6	Bulgaria	
Pollutant		
Station Type		
Station Area		
Measurement Regime		
Network		
Eol Station Code		
Station Name		
<input type="text"/> Search for Station Name		
Count	Value	Match all
6	Sofia - Druzhba	
Sampling Point Local Id		

En nuestro caso solo haremos una busqueda sencilla, modificaremos el pais señalando Bulgaria y en el nombre de estación indicaremos el que deseemos, en nuestro caso Sofia-Druzhba.

Una vez indicado los valores en el filtro, el cuadro de las estaciones cambiará mostrando así la estación que buscábamos.

Results 1 - 6 of 6

Order Relevance

Download TSV

Download CSV

Country or Ter...	Namespace	Network	Responsible Party	Time Zone	Station Name	Eol Station...	Station Type	Station Area	Pollutant	Sampling Point Local Id	Latitude	Longitude	Altitude	Measurement Regime	Operational Activity B...	Operational Activity End
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Benzene (air)	SPO-BG0052A_00020_100	42.666508	23.400164	548	Continuous data collection	2005-12-31T22:00:00Z	
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Ozone (air)	SPO-BG0052A_00007_500	42.666508	23.400164	548	Continuous data collection	1996-12-31T22:00:00Z	
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Nitrogen dioxide (air)	SPO-BG0052A_00008_500	42.666508	23.400164	548	Continuous data collection	1996-12-31T22:00:00Z	
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Nitrogen monoxide (air)	SPO-BG0052A_00038_500	42.666508	23.400164	548	Continuous data collection	1996-12-31T22:00:00Z	
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Particulate matter < 10 µm (PM10)	SPO-BG0052A_00005_100	42.666508	23.400164	548	Continuous data collection	1999-12-31T22:00:00Z	
Bulgaria	BG.BG-ExEA.AQ	National air network	Executive Environment Agency	UTC+02	Sofia - Druzhba	BG0052A	Background	Urban	Sulphur dioxide (air)	SPO-BG0052A_00001_500	42.666508	23.400164	548	Continuous data collection	1996-12-31T22:00:00Z	

.. 1 - 6 of 6 ..

Data sources

Air Quality e-Reporting provided by European Environment Agency (EEA)

En este punto podemos destacar varios factores:

- En rojo (parte superior derecha) podemos ver que podemos descargarnos la tabla completa con todos los datos de la estación.
- En amarillo (parte inferior izquierda) se nos indica que estos datos son ofrecidos por la Agencia Europea de Medioambiente, organismo oficial de la Unión Europea
- En azul (parte central izquierda) tendremos un dato importante de nuestra estación, su código de estación, este es único para cada una y es lo que nos ayudará a encontrar nuestra estación cuando busquemos los datos de los contaminantes. (en nuestro ejemplo BG0052A)
- Por ultimo en verde (parte central derecha) tendremos los contaminantes, en esta columna se nos indica que contaminantes cabe esperar que encontremos cuando busquemos los datos de los mismos.

Enlaces para obtener los datos de las estaciones:

- Bulgaria - Druzhba [enlace](#)
- Grecia – Patra-2 [enlace](#)
- España – Bermejales [enlace](#)

Obtención de los datos de los contaminantes por localidades

Una vez hemos obtenido el código de la estación entramos en la siguiente pagina web:

<http://discomap.eea.europa.eu/map/fme/AirQualityExport.htm>

En esta página rellenaremos los campos del formulario con los datos que deseemos:

Country: BG Bulgaria

For all countries, see note below.

City name: Sofia

Pollutant: PM10

Please, be aware that not every pollutant is reported by every country why the pollutant list depends on the selected country.

Year from: 2015

Year to: 2020

Source: All

E2a (UTD) data are only available for years where E1a data have not yet been delivered (this will normally be the most recent year)

Output type: HTML

Update date:

Optional, format yyyy-mm-dd hh:mm:ss. To be used when only files created or updated after a certain date is of interest.

Time coverage: Year

Files available for requested years or last 7 days.

Update request URL

Download

De arriba abajo los campos son: País, Ciudad, Contaminante, año inicio y final, fuente (ALL), tipo de salida (HTML), fecha de actualización (lo dejaremos en blanco) y cobertura de tiempo (YEAR).

Pulsando sobre el botón “Update request URL” nos saldra un link en el cuadro inferior.



Update request URL

Download

Update request URL

https://fme.discomap.eea.europa.eu/fmedatastreaming/AirQualityDownload/AQData_Extract.fmw?CountryCode=BG&CityName=Sofia&Pollutant=5&Year_from=2015&Year_to=2020&Station=&Samplingpoint=&Source=All&Output=HTML&UpdateDate=&TimeCoverage=Year

Download

Presionamos sobre “Download” y se nos abrirá una pestaña con los enlaces para descargar distintos archivos, todos son del mismo contaminante, pero de diferentes estaciones de la ciudad de Sofia y distintos años.

https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2015_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2015_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2015_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9484_2015_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2015_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2016_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2016_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2016_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2016_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2017_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2017_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2017_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2017_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_60881_2017_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2018_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2018_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2018_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2018_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_60881_2018_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2019_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2019_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2019_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2019_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_60881_2019_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9642_2020_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9572_2020_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9421_2020_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_9616_2020_timeseries.csv
https://ereporting.blob.core.windows.net/downloadservice/BG_5_60881_2020_timeseries.csv

En el nombre del archivo podemos ver cuál es el código de referencia de los archivos
eremos (9421).

Como podemos ver hay un archivo con la misma referencia por cada año, los
 amos todos y ya tendríamos todos los datos del contaminante PM10 de la estación de
 a desde 2015 hasta 2020.

Por otra parte, para hacerlo de manera mas sencilla, podemos utilizar el siguiente enlace haciendo unos pequeños cambios:

https://fme.discomap.eea.europa.eu/fmedatastreaming/AirQualityDownload/AQData_Extract.fmw?CountryCode=BG&CityName=Sofia&Pollutant=10&Year_from=2015&Year_to=2020&Station=STA-BG0052A&Samplingpoint=&Source=All&Output=HTML&UpdateDate=&TimeCoverage=Year

CountryCode=BG → Podemos cambiar BG por el código del país que busquemos

CityName=Sofia → Podemos cambiar Sofia por el nombre de la ciudad que deseemos

Pollutant=5 → Podemos cambiar 5 por el numero del contaminante que deseemos

CO: 10

O3: 7

PM10: 5

SO2: 1

NO2: 8

Year_from=2015&Year_to=2020 → Podemos cambiar los años de principio y fin

Station=STA-BG0052 → Cambiaremos el código por el de la estación que deseemos

Druzhba: STA-BG0052

Bermejales: STA-ES1638A

Patra-2: STA-GR0048A

Enlaces completos desde 2015 a 2020, por contaminante y estación:

Bulgaria, Sofía:

Druzhba PM10 [enlace](#)

Druzhba CO No hay datos

Druzhba SO2 [enlace](#)

Druzhba NO2 [enlace](#)

Druzhba O3 [enlace](#)

España, Sevilla:

Bermejales PM10 [enlace](#)

Bermejales CO [enlace](#)

Bermejales SO2 [enlace](#)

Bermejales NO2 [enlace](#)

Bermejales O3 [enlace](#)

Grecia, Patra:

Patra-2 PM10 [enlace](#)

Patra-2 CO [enlace](#)

Patra-2 SO2 [enlace](#)

Patra-2 NO2 [enlace](#)

Patra-2 O3 No hay datos

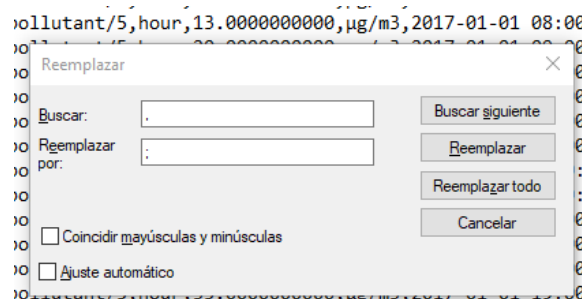
Adecuación de los datos obtenidos mediante Excell.

Una vez tenemos descargados los distintos archivos hay que hacerles distintas modificaciones.

1. Abriremos el archivo en el bloc de notas, y haremos unos pequeños cambios.

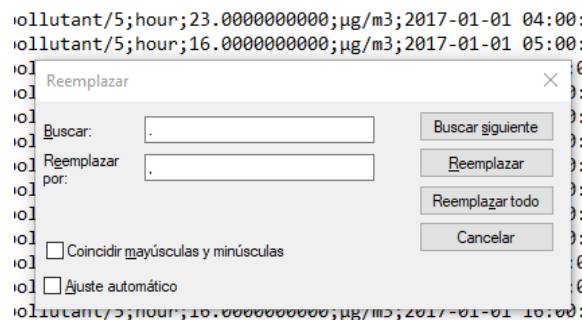
- a. Cambiaremos las comas “,” por punto y coma “;”.

Presionando CTRL+R podemos reemplazar todas las comas del archivo.



- b. A continuación cambiaremos los puntos “.” por comas “;”.

Esto solo afectará a la columna de la concentración, hará que cuando abramos el archivo en Excell dichos valores sean numéricos con decimales.



- Una vez tengamos el archivo separado en columnas podemos eliminar las que no nos hagan falta y quedarnos con las que si, que son Concentration y DateTimeBegin.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Countrycode	Namespace	AirQualityNet	AirQualitySt	AirQualitySt	SamplingPoi	SamplingPro	Sample	AirPollutant	AirPollutant	AveragingTir	Concentration	UnitOfMeas	DateTimeBegin	DateTimeEnd	Validity	Verification
2	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	268.100.000.000	Åµg/m3	2015-06-23 0	2015-06-24 0	1	1
3	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	185.800.000.000	Åµg/m3	2015-06-26 0	2015-06-27 0	1	1
4	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	279.000.000.000	Åµg/m3	2015-06-28 0	2015-06-29 0	1	1
5	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	308.700.000.000	Åµg/m3	2015-06-14 0	2015-06-15 0	1	1
6	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	290.300.000.000	Åµg/m3	2015-06-12 0	2015-06-13 0	1	1
7	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	316.500.000.000	Åµg/m3	2015-06-15 0	2015-06-16 0	1	1
8	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	262.800.000.000	Åµg/m3	2015-06-16 0	2015-06-17 0	1	1
9	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	292.300.000.000	Åµg/m3	2015-07-19 0	2015-07-20 0	1	1
10	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	300.600.000.000	Åµg/m3	2015-07-18 0	2015-07-19 0	1	1
11	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	305.000.000.000	Åµg/m3	2015-07-17 0	2015-07-18 0	1	1
12	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	269.600.000.000	Åµg/m3	2015-06-29 0	2015-06-30 0	1	1
13	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	290.500.000.000	Åµg/m3	2015-07-03 0	2015-07-04 0	1	1
14	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	239.200.000.000	Åµg/m3	2015-07-12 0	2015-07-13 0	1	1
15	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	195.600.000.000	Åµg/m3	2015-07-11 0	2015-07-12 0	1	1
16	BG	BG-BG-ExEA	NET-BG001A	STA-BG0052A		SPO-BG0052	SPP-BG_A_B	SPO_F-BG00	PM10	http://dd.eid	day	422.300.000.000	Åµg/m3	2015-07-09 0	2015-07-10 0	1	1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Concentration	DateTimeBegin														
2	92.9	2019-01-01 00:00:00														
3	38.08	2019-01-01 01:00:00														
4	35.94	2019-01-01 02:00:00														
5	35.74	2019-01-01 03:00:00														
6	31.81	2019-01-01 04:00:00														
7	26.86	2019-01-01 05:00:00														
8	22.91	2019-01-01 06:00:00														
9	37.83	2019-01-01 07:00:00														
10	45.29	2019-01-01 08:00:00														
11	21.0	2019-01-01 09:00:00														
12	9.23	2019-01-01 10:00:00														
13	6.49	2019-01-01 11:00:00														
14	1.62	2019-01-01 12:00:00														
15	5.71	2019-01-01 13:00:00														

- Una vez tengamos este resultado, en la columna de la fecha eliminaremos “+01:00” puesto que no nos es de utilidad.

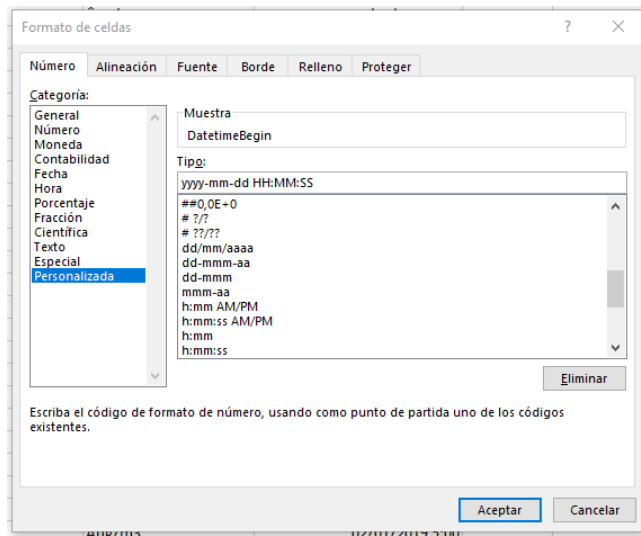
Pestaña Inicio → Buscar y reemplazar → “+01:00” a “ ” y reemplazar todos

D	E	F	G
UnitOfMeasurement	DateTimeBegin		
Åµg/m3	2019-01-01 00:00:00 +01:00		
Åµg/m3	2019-01-01 01:00:00 +01:00		
Åµg/m3	2019-01-01 02:00:00 +01:00		
Åµg/m3	2019-01-01 03:00:00 +01:00		



D	E	F	G
UnitOfMeasurement	DateTimeBegin		
Åµg/m3	01/01/2019 0:00		
Åµg/m3	01/01/2019 1:00		
Åµg/m3	01/01/2019 2:00		
Åµg/m3	01/01/2019 3:00		

4. Cambiamos el formato de la columna con las fechas, para que asi podamos insertar los archivos correctamente en la base de datos.



Seleccionamos la columna y en formato de celdas lo ponemos con tipo: “aaaa-mm-dd HH:MM:SS”

5. Añadiremos una nueva columna al principio, que tendrá como valor el código de la estación hallado anteriormente (en nuestro caso 8495).
6. A continuación le cambiamos el nombre a las columnas por:

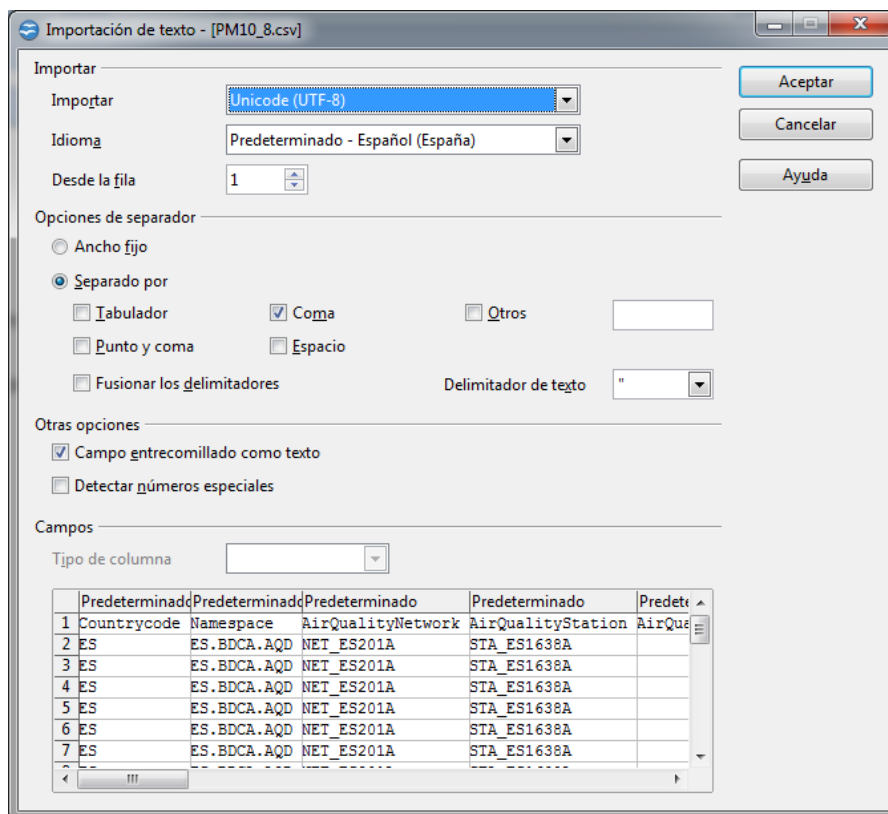
data.idx || data.iaqi.pm10.v || data.time.s

	A	B	C	D
1	data.idx	data.iaqi.pm10.v	data.time.s	
2	8495	92.9	2019-01-01 00:00:00	
3	8495	38.08	2019-01-01 01:00:00	
4	8495	35.94	2019-01-01 02:00:00	
5	8495	35.74	2019-01-01 03:00:00	
6	8495	31.81	2019-01-01 04:00:00	
7	8495	26.86	2019-01-01 05:00:00	
8	8495	22.91	2019-01-01 06:00:00	
9	8495	37.83	2019-01-01 07:00:00	
10	8495	45.29	2019-01-01 08:00:00	

Guardamos en tipo CSS delimitado por comas, en caso de que no se guarde separado por comas podemos cambiar los ; por , en el bloc de notas y ya tendremos el archivo con el formato necesario para ser introducido.

Adecuación de los datos obtenidos mediante OpenOffice.

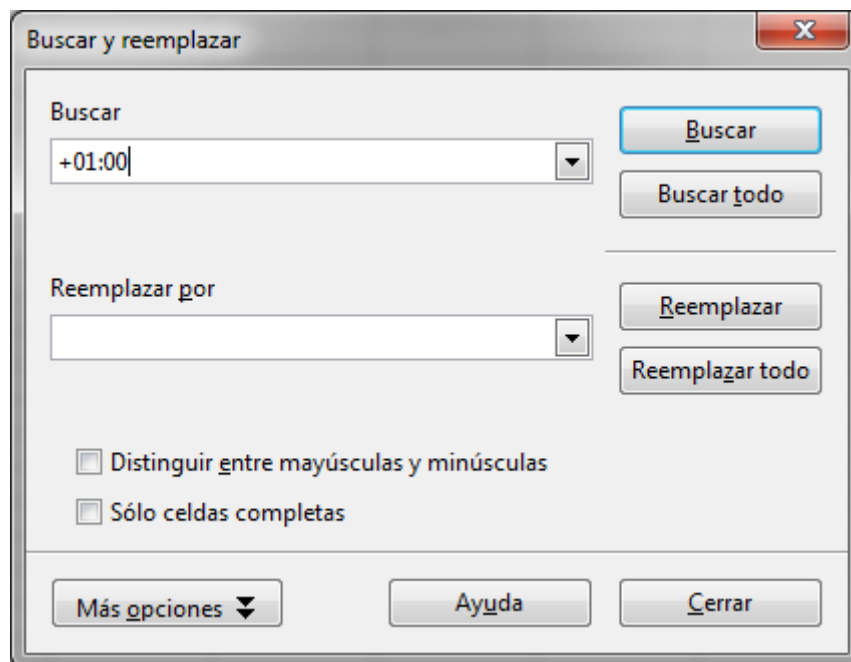
Abrir el archivo csv con OpenOffice.



Dejar solo las siguientes tres columnas:

	A	B	C
1	<u>AirQualityStation</u>	<u>Concentration</u>	<u>DatetimeBegin</u>
2	STA_ES1638A	14.2500000000	2016-12-26 09:00:00 +01:00
3	STA_ES1638A	22.8000000000	2016-12-26 12:00:00 +01:00
4	STA_ES1638A	29.4499999900	2016-12-26 03:00:00 +01:00
5	STA_ES1638A	32.2999999900	2016-12-25 23:00:00 +01:00
6	STA_ES1638A	53.2000000000	2016-12-25 01:00:00 +01:00
7	STA_ES1638A	20.8999999900	2016-12-25 15:00:00 +01:00
8	STA FS1638A	33.2500000000	2016-12-24 19:00:00 +01:00

Eliminar +01:00 del campo fecha:

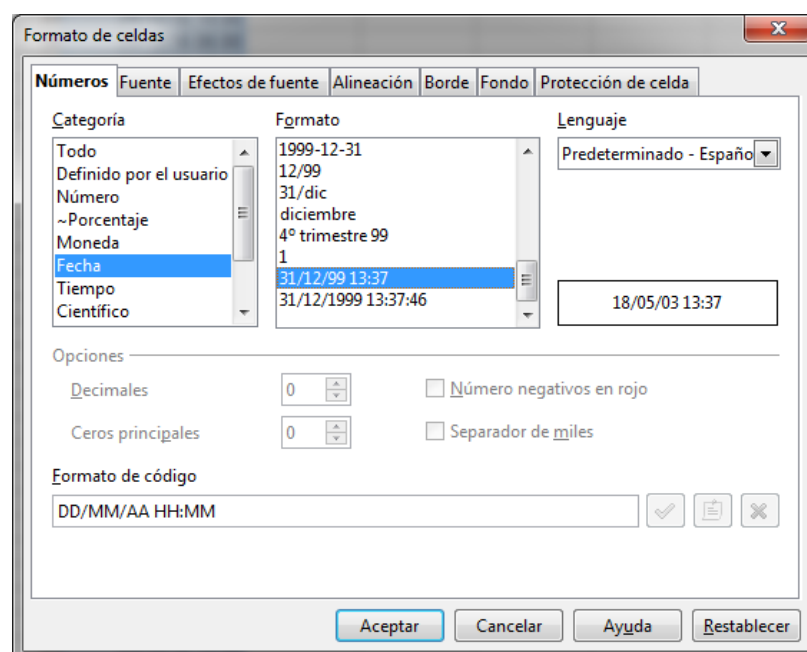


Cambiar "." por ",". Esto hará que la concentración sea numérica.

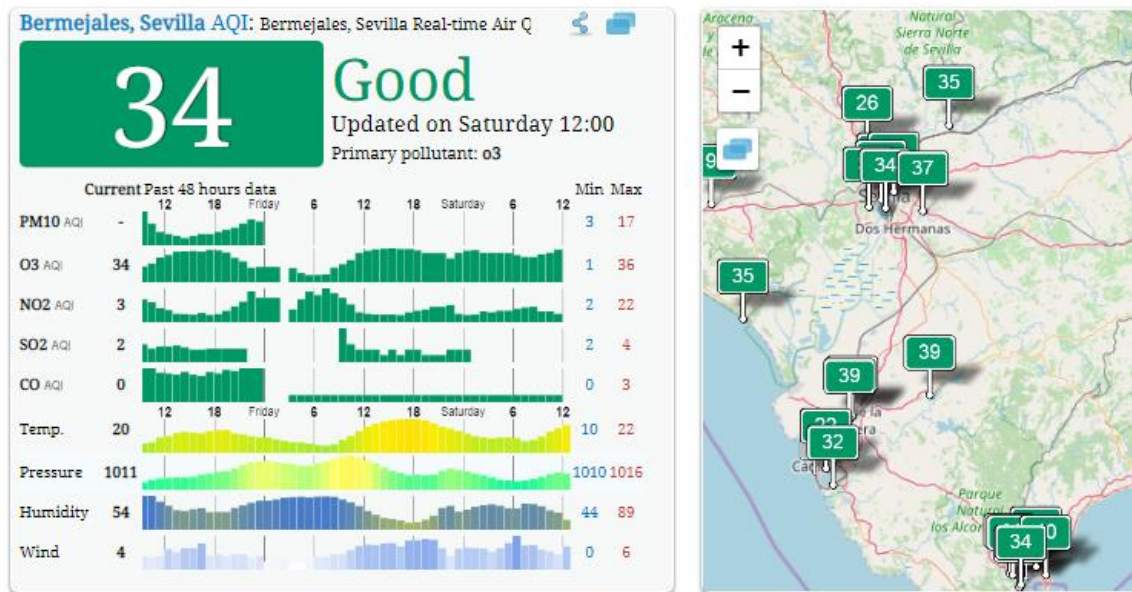
Redondear con un decimal:

Concentration
14,3
22,8
29,4

Formatear el campo DatatimeBegin a tipo Fecha:



Por compatibilidad con los datos que ofrece diariamente la web <https://waqi.info/>.
Vamos a cambiar el AirQualityStation por el idx de la estación. Por ejemplo para obtener el idx de los bermejales nos vamos a la web <https://waqi.info/> y buscamos "Bermejales":

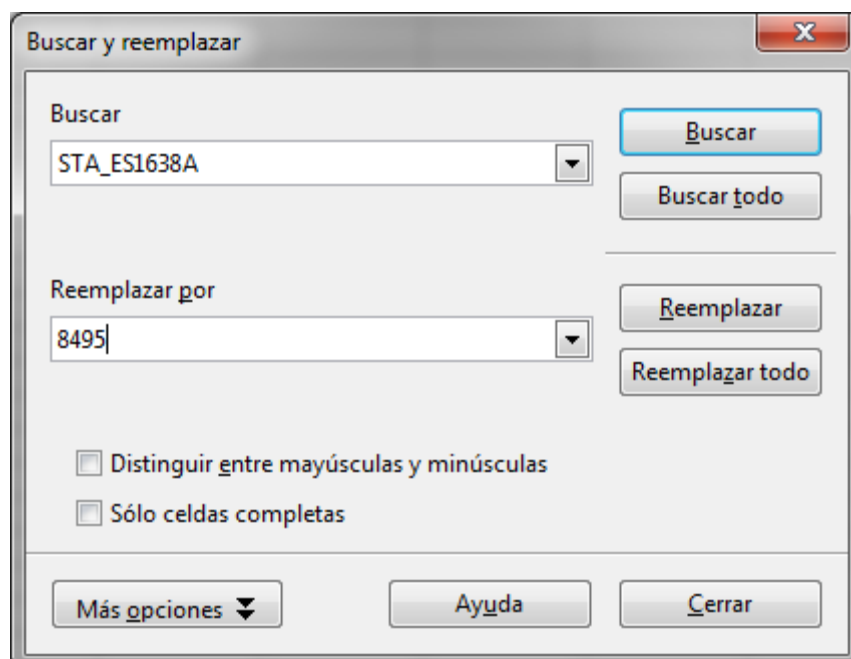


Le decimos que nos muestre el código fuente de la página, y buscamos "idx". En el código de la página nos aparecerá:

```
aqiStation =
{
  "name": "Bermejales, Sevilla",
  "idx": 8495
};
try {
  aqi18n.register("Temp", "Temp.");
} catch (e) {}
function getGetGeolocateUrl() {
  return (function() {
    var
```

Luego el idx (identificador) de la estación de los Bermejales de Sevilla es **8495**.

Cambiamos el AirQualityStation por idx, tanto en el nombre como en el valor.



Nos queda como:

8495	53,2	25/12/16 01:00
8495	20,9	25/12/16 15:00
8495	33,3	24/12/16 19:00
8495	30,4	25/12/16 06:00

Para saber los nombres de los campos vamos a usar un token solicitándolo en la web: [h https://aqicn.org/data-platform/token/#/](https://aqicn.org/data-platform/token/#/) , lo enviarán al email. Con ese token se accede a la información de los datos de las estaciones.

Nosotros usamos la siguiente url (de los Bermejales) y token:

<http://api.waqi.info/feed/@8495/?token=43b146051093d2a179f84802c061870088a7ce91>:

Los datos tal y como los suministra para los bermejales tienen la siguiente forma:

```

status:      "ok"
data:
  aqi:        34
  idx:        8495
  attributions:
    0:
      url:     "http://www.eea.europa.eu/themes/air/"
      name:    "European Environment Agency"
      logo:    "Europe-EEA.png"
    1:
      url:     "http://www.juntadeandalucia.es/medioambiente/site/portaWeb/"
      name:    "Consejería de Medio Ambiente y Ordenación del Territorio :: Junta de Andalucía"
      logo:    "Andalucia-Consejeria-de-medio-ambiente.png"
    2:
      url:     "https://waqi.info/"
      name:    "World Air Quality Index Project"
  city:
    geo:
      0:       37.3485317
      1:       -5.9877044
      name:    "Bermejales, Sevilla, Spain"
    url:       "https://aqicn.org/city/spain/andalucia/sevilla/bermejales"

dominantpol: "o3"
iaqi:
  co:
    v: 0.1
  h:
    v: 53
  no2:
    v: 3.7
  o3:
    v: 34.2
  p:
    v: 1010.9
  pm10:
    v: 11
  so2:
    v: 2.1
  t:
    v: 20
  w:
    v: 3
  wg:
    v: 9
  time:
    s: "2020-04-04 12:00:00"

```

Los nombres de los campos los ponemos como:

A	B	C
<u>data.idx</u>	<u>data.iaqui.pm10.v</u>	<u>data.time.s</u>
8495	14,3	26/12/16 09:00
8495	22,8	26/12/16 12:00

Lo guardamos.

Lo abrimos con el Bloc de notas.

Para poder subir los números con decimales (hay que cambiar de 14,3 a 14.3) es decir la "," por el punto. Para ello hacemos los siguientes reemplazos en el fichero:

- Reemplazar "," por ;
- Reemplazar "; por "
- Reemplazar " por .
- Reemplazar . por ,

El archivo CSV en el bloc de notas debe quedar Hay que comprobar que los títulos están entrecomillados:

```
"data.idx","data.iaqui.pm10.v","data.time.s"
8495,14.3,26/12/16 09:00
8495,22.8,26/12/16 12:00
8495,29.4,26/12/16 03:00
8495,32.3,25/12/16 23:00
8495,53.2,25/12/16 01:00
```

Ya está listo para importar a MongoDB.

Insercción en la base de datos MongoDB en local.

A continuación vamos a introducir los datos tratados anteriormente en una Base de datos en mongo.

Abrimos una terminal, nos posicionamos en la carpeta donde hayamos guardado el archivo con los datos.

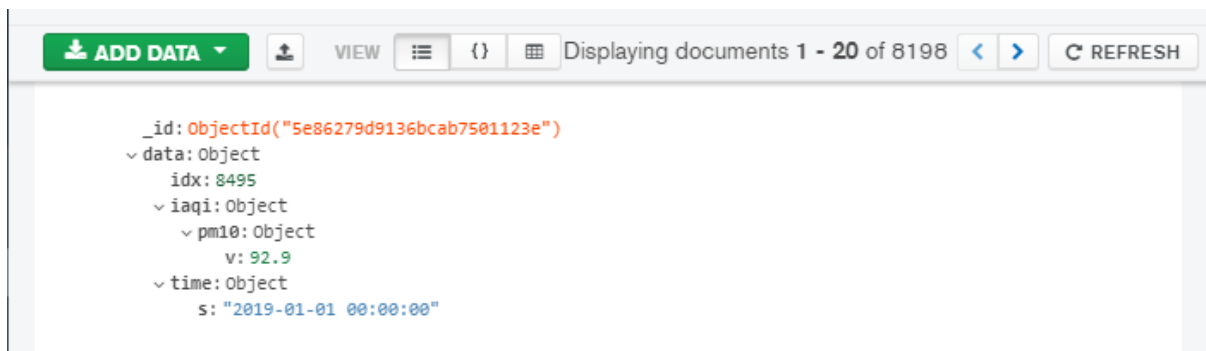
Introducimos el siguiente comando introduciendo la base de datos y la colección deseada y el archivo modificado:

```
mongoimport --db BaseDeDatos --collection Colección --headerline --type csv --file .\archivo.csv
```

El resultado sería:

```
PS C:\Users\jonatan\Desktop> mongoimport --db Historicos --collection pruebaproyecto --headerline --type csv --file .\ejemplo.csv
2020-04-02T19:57:48.922+0200    connected to: mongodb://localhost/
2020-04-02T19:57:49.184+0200    8198 document(s) imported successfully. 0 document(s) failed to import.
PS C:\Users\jonatan\Desktop>
```

Si ahora entramos en la aplicación veremos que se han insertado los datos:

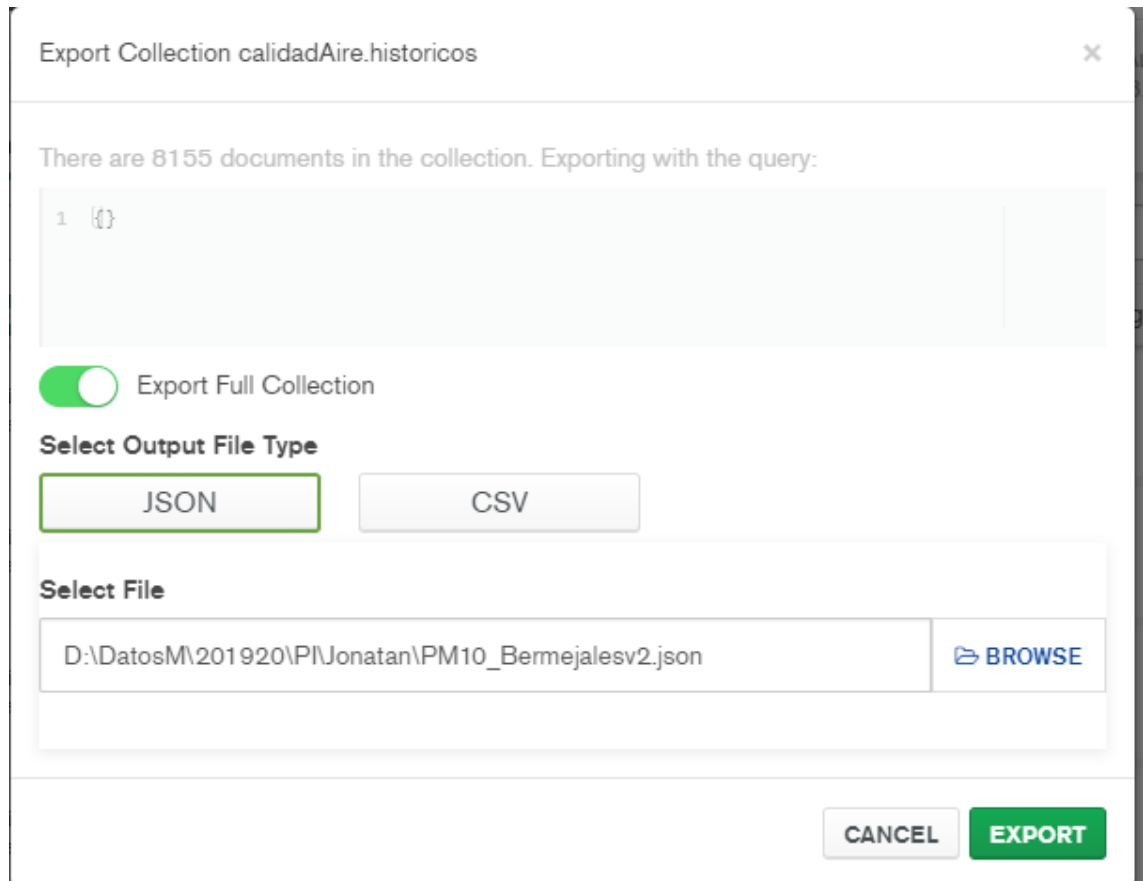
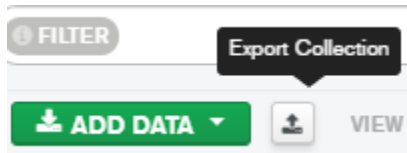


8198 Datos introducidos correctamente y con el formato y la estructura deseados.

Exportación desde MongoDB Compass de los datos a formato json.

Exportación desde MongoDB Compass de los datos a formato json.

Seleccionamos toda la colección y JSON (podíamos seleccionar csv)



Insercción en la base de datos Mongo Atlas en la nube.

En el cluster de la nube seleccionamos la colección y en la en la opción:

Command Line Tools

aparecen los comandos de exportación e importación mediante shell:

Data Import and Export Tools

Replace **PASSWORD** with the password for the admin user, **DATABASE** with the name of the database you wish to import/export to your cluster, and **COLLECTION** with the name of the collection you wish to import/export to your cluster. Replace **FILETYPE** with json or csv to specify the file type. Where applicable, replace **FILENAME** with the location and name of the output file (for export) or data source (for import).

NOTE: When exporting or importing CSV data, an additional `--fields` flag is often required. See documentation for the specific tool for additional details.

mongoimport | imports content from an Extended JSON, CSV, or TSV export

```
mongoimport --host Sandbox-shard-0/sandbox-shard-00-00-
2ghih.mongodb.net:27017,sandbox-shard-00-01-2ghih.mongodb.net:27017,sandbox-
shard-00-02-2ghih.mongodb.net:27017 --ssl --username USER --password <PASSWORD>
--authenticationDatabase admin --db <DATABASE> --collection <COLLECTION> --type
<FILETYPE> --file <FILENAME>
```

Copy

mongoexport | produces a JSON or CSV export of data stored in a MongoDB instance

```
mongoexport --host Sandbox-shard-0/sandbox-shard-00-00-
2ghih.mongodb.net:27017,sandbox-shard-00-01-2ghih.mongodb.net:27017,sandbox-
shard-00-02-2ghih.mongodb.net:27017 --ssl --username USER --password <PASSWORD>
--authenticationDatabase admin --db <DATABASE> --collection <COLLECTION> --type
<FILETYPE> --out <FILENAME>
```

Copy

A partir de aquí construimos nuestro comando de shell de importación (hay que sustituir los * por su usuario de administración y su password:

```
mongoimport --host Sandbox-shard-0/sandbox-shard-00-00-
2ghih.mongodb.net:27017,sandbox-shard-00-01-2ghih.mongodb.net:27017,sandbox-shard-
00-02-2ghih.mongodb.net:27017 --SSL --username ***** --password ***** --
authenticationDatabase admin --db CalidadAire --collection datosHistoricos --type csv --
headerline --file PM10_Bermejalesv2.csv
```

El resultado es la importación de todos los documentos:

```
D:\DatosM\201920\PI\Jonatan>mongoimport --host Sandbox-shard-0/sandbox-shard-00-00-2ghih.mongodb.net:27017,sandbox-shard-00-01-2ghih.mongodb.net:27017,sandbox-shard-00-02-2ghih.mongodb.net:27017 --ssl --username ** --password ***** --authenticationDatabase admin --db CalidadAire --collection datosHistoricos --type csv --headerline --file PM10_Bermejalesv2.csv
2020-04-05T21:40:40.943+0200 WARNING: ignoring unsupported URI parameter 'replicaset'
2020-04-05T21:40:42.427+0200 connected to: mongodb://sandbox-shard-00-00-2ghih.mongodb.net:27017,sandbox-shard-00-01-2ghih.mongodb.net:27017,sandbox-shard-00-02-2ghih.mongodb.net:27017/?replicaSet=Sandbox-shard-0
2020-04-05T21:40:44.175+0200 8155 document(s) imported successfully. 0 document(s) failed to import.
```

Y los datos ya están en la nube con el mismo formato que en local.