

Reciclaje y Limpieza de Madrid

JORGE SÁNCHEZ BERROCOSO && ALFREDO
MALDONADO



Descripción del problema propuesto

Nos ha contratado el Ayuntamiento de Madrid para que hagamos un estudio del análisis de limpieza y gestión de basuras que se hace en la ciudad.

Para ello nos han dado unos ficheros con la información deseada que debemos procesar.

Nuestro programa se debe llamar de la siguiente manera: `java -jar basuras.jar opcion [parámetros_opcion]`

Si la opción es `parser directorio_origen directorio_destino`: debe tomar los ficheros csv del directorio origen y los transformarlos en JSON y XML en el directorio destino. En dicho directorio destino deberán estar las tres versiones: CSV, JSON y XML.

Si la opción es `resumen directorio_origen directorio_destino`: debe tomar la información de los contenedores y de la recogida, independientemente de la extensión que tenga (si no corresponde a la extensión o al formato deberá indicar error) y deberá procesarla generando en directorio_destino un `resumen.html`, aplicándoles los estilos que creas oportunos, con la siguiente información:

- Título: Resumen de recogidas de basura y reciclaje de Madrid
- Fecha de generación: Fecha y hora en formato español.
- Autores: Nombre y apellidos de los dos autores.
- Número de contenedores de cada tipo que hay en cada distrito.
- Media de contenedores de cada tipo que hay en cada distrito.
- Gráfico con el total de contenedores por distrito.
- Media de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.
- Gráfico de media de toneladas mensuales de recogida de basura por distrito.
- Máximo, mínimo, media y desviación de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.
- Suma de todo lo recogido en un año por distrito.
- Por cada distrito obtener para cada tipo de residuo la cantidad recogida.
- Tiempo de generación del mismo en milisegundos.

Si la opción es `resumen distrito directorio_origen directorio_destino`: debe tomar la información de los contenedores y de la recogida, independientemente de la

extensión que tenga (si no corresponde a la extensión o al formato deberá indicar error) y deberá procesarla generando en directorio_destino un resumen_distrito.html (solo si el distrito existe, si no deberá mostrar error), aplicándoles los estilos que creas oportunos, con la siguiente información:

- Titulo: Resumen de recogidas de basura y reciclaje de “Distrito”
- Fecha de generación: Fecha y hora en formato español.
- Autores: Nombre y apellidos de los dos autores.
- Número de contenedores de cada tipo que hay en este distrito.
- Total de toneladas recogidas en ese distrito por residuo.
- Gráfico con el total de toneladas por residuo en ese distrito.
- Máximo, mínimo , media y desviación por mes por residuo en dicho distrito.
- Gráfica del máximo, mínimo y media por meses en dicho distrito.
- Tiempo de generación del mismo en milisegundos.

Por cada ejecución debemos guardar un fichero bitacora.xml donde tengamos en este XML un listado de las ejecuciones con la siguiente información:

- ID de la ejecución en base a uuid
- Instante: Instante de la ejecución en formato ISO 8601
- Tipo de opción elegida (parser, resumen global, resumen ciudad).
- Éxito: si tuvo éxito o no su procesamiento.
- Tiempo de ejecución: tiempo de ejecución si tuvo éxito en milisegundos.

Tienes los PDFs necesarios donde te explica cómo y de qué manera la información está estructurada. Debes leerlos detalladamente para saber cómo y dónde está la información relevante y cómo está expresada. Antes de programar a lo loco, te recomiendo estudiar y analizar esta información para pensar cómo y de qué manera puedes hacer lo que se te pide de la forma más efectiva y eficiente. No hay una solución buena única, pero sí muchas malas. Tú decides.

REQUISITOS FUNCIONALES

Debe realizar por cada ejecución del programa el almacenamiento de los datos en un fichero bitácora

Deberán estar las tres versiones: CSV, JSON y XML.

Se debe obtener la información de Contenedores_varios.csv y de modelo_residuos_2021-2022.

REQUISITOS NO FUNCIONALES

La opción parse deberá leer el csv y exportar a json, xml.

La opción resumen debe mostrar los siguientes:

- Título: Resumen de recogidas de basura y reciclaje de Madrid
- Fecha de generación: Fecha y hora en formato español.
- Autores: Nombre y apellidos de los dos autores.
- Número de contenedores de cada tipo que hay en cada distrito.
- Media de contenedores de cada tipo que hay en cada distrito.
- Gráfico con el total de contenedores por distrito.
- Media de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.
- Gráfico de media de toneladas mensuales de recogida de basura por distrito.
- Máximo, mínimo , media y desviación de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.
- Suma de todo lo recogido en un año por distrito.
- Por cada distrito obtener para cada tipo de residuo la cantidad recogida.
- Tiempo de generación del mismo en milisegundos

La opción resumen distrito debe mostrar la siguiente información:

El resumen por distrito debe mostrar los siguientes datos en su informe:

- Título: Resumen de recogidas de basura y reciclaje de "Distrito"
- Fecha de generación: Fecha y hora en formato español.
- Autores: Nombre y apellidos de los dos autores.
- Número de contenedores de cada tipo que hay en este distrito.
- Total de toneladas recogidas en ese distrito por residuo.

- Gráfico con el total de toneladas por residuo en ese distrito.
- Máximo, mínimo , media y desviación por mes por residuo en dicho distrito.
- Gráfica del máximo, mínimo y media por meses en dicho distrito.
- Tiempo de generación del mismo en milisegundos.

Realización de las consultas

- Número de contenedores de cada tipo que hay en cada distrito.

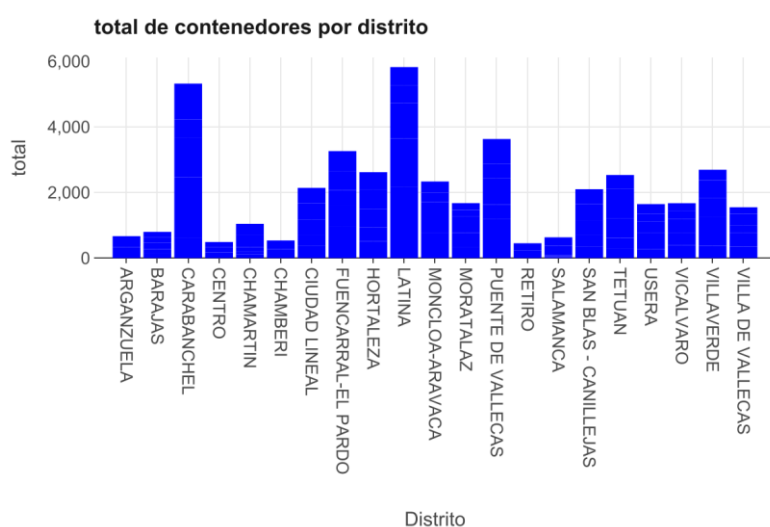
```
Número de contenedores de cada tipo que hay en cada distrito:
    distrito tipoContenedor total
0  ARGANZUELA  PAPEL-CARTON  336
1  ARGANZUELA    VIDRIO  332
2  ARGANZUELA    RESTO    1
3    BARAJAS    ENVASES  144
4    BARAJAS    ORGANICA  129
5    BARAJAS    RESTO  196
6    BARAJAS  PAPEL-CARTON  169
7    BARAJAS    VIDRIO  161
8 CARABANCHEL  PAPEL-CARTON  604
9 CARABANCHEL    RESTO 1863
10 CARABANCHEL    ENVASES 1211
11 CARABANCHEL    VIDRIO  548
12 CARABANCHEL    ORGANICA 1095
13     CENTRO  PAPEL-CARTON  175
14     CENTRO    ENVASES  154
15     CENTRO    VIDRIO  161
16 CHAMARTIN    ENVASES  109
17 CHAMARTIN    ORGANICA   91
18 CHAMARTIN    RESTO  135
19 CHAMARTIN  PAPEL-CARTON  369
```

- Media de contenedores de cada tipo que hay en cada distrito.

Media de contenedores de cada tipo que hay en cada distrito:

	distrito	tipoContenedor	media de Contenedores
0	ARGANZUELA	PAPEL-CARTON	336,0
1	ARGANZUELA	VIDRIO	332,0
2	ARGANZUELA	RESTO	1,0
3	BARAJAS	ENVASES	144,0
4	BARAJAS	ORGANICA	129,0
5	BARAJAS	RESTO	196,0
6	BARAJAS	PAPEL-CARTON	169,0
7	BARAJAS	VIDRIO	161,0
8	CARABANCHEL	PAPEL-CARTON	604,0
9	CARABANCHEL	RESTO	1863,0
10	CARABANCHEL	ENVASES	1211,0
11	CARABANCHEL	VIDRIO	548,0
12	CARABANCHEL	ORGANICA	1095,0
13	CENTRO	PAPEL-CARTON	175,0
14	CENTRO	ENVASES	154,0
15	CENTRO	VIDRIO	161,0
16	CHAMARTIN	ENVASES	109,0
17	CHAMARTIN	ORGANICA	91,0
18	CHAMARTIN	RESTO	135,0
19	CHAMARTIN	PAPEL-CARTON	369,0
...			

- Gráfico con el total de contenedores por distrito.

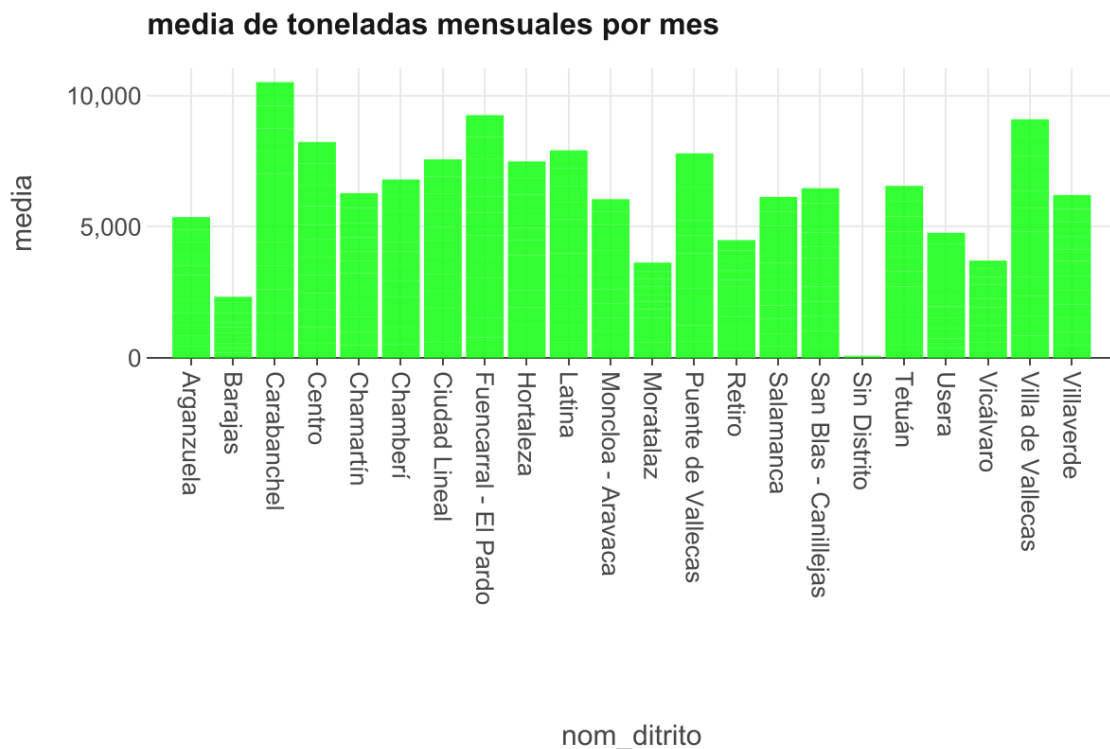


- Media de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.

Media de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito:

	nom_ditrto	anio	residuos	Media
0	Arganzuela	2021	RESTO	1894,020844
1	Arganzuela	2021	ENVASES	323,565000
2	Arganzuela	2021	VIDRIO	275,680002
3	Arganzuela	2021	ORGANICA	617,755000
4	Arganzuela	2021	PAPEL-CARTON	354,642914
5	Arganzuela	2021	CARTON COMERCIAL	25,528334
6	Arganzuela	2021	RCD	53,619092
7	Arganzuela	2021	PUNTOS LIMPIOS	72,271667
8	Arganzuela	2021	CONTENEDORES DE ROPA	0,270000
9	Barajas	2021	RESTO	818,626668
10	Barajas	2021	ENVASES	151,370000
11	Barajas	2021	VIDRIO	107,181669
12	Barajas	2021	ORGANICA	346,575002
13	Barajas	2021	PAPEL-CARTON	53,555000
14	Barajas	2021	RCD	24,888750
15	Barajas	2021	PUNTOS LIMPIOS	34,290000
16	Barajas	2021	CARTON COMERCIAL	5,610000
17	Carabanchel	2021	RESTO	4396,960815
18	Carabanchel	2021	ENVASES	731,161672
19	Carabanchel	2021	VIDRIO	281,058333
...				

- Gráfico de media de toneladas mensuales de recogida de basura por distrito.



- **Máximo, mínimo , media y desviación de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito.**

Máximo, mínimo , media y desviación de toneladas anuales de recogidas por cada tipo de basura agrupadas por distrito									
nom_ditrto	residuos	año	Maximo	Minimo	Media	Desviacion			
0	Arganzuela	RESTO	2021	2193	1569	1894,020844	147,384054		
1	Arganzuela	ENVASES	2021	370	202	323,565000	46,885439		
2	Arganzuela	VIDRIO	2021	353	202	275,680002	45,151989		
3	Arganzuela	ORGANICA	2021	694	402	617,755000	87,791320		
4	Arganzuela	PAPEL-CARTON	2021	386	288	354,642914	25,175069		
5	Arganzuela	CARTON COMERCIAL	2021	34	20	25,528334	3,829763		
6	Arganzuela	RCD	2021	142	31	53,619892	33,159174		
7	Arganzuela	PUNTOS LIMPIOS	2021	88	59	72,271667	10,631816		
8	Arganzuela	CONTENEDORES DE ROPA	2021	0	0	0,270000	0,127279		
9	Barajas	RESTO	2021	959	658	818,626668	71,678440		
10	Barajas	ENVASES	2021	194	113	151,370000	30,125478		
11	Barajas	VIDRIO	2021	152	59	107,181669	31,119079		
12	Barajas	ORGANICA	2021	448	144	346,575002	97,545376		
13	Barajas	PAPEL-CARTON	2021	65	32	53,555000	8,985836		
14	Barajas	RCD	2021	88	9	24,888750	23,262916		
15	Barajas	PUNTOS LIMPIOS	2021	43	20	34,290000	6,313676		
16	Barajas	CARTON COMERCIAL	2021	18	1	5,610000	5,128978		
17	Carabanchel	RESTO	2021	5139	3450	4396,960815	386,480146		
18	Carabanchel	ENVASES	2021	819	459	731,161672	97,392088		
19	Carabanchel	VIDRIO	2021	332	219	281,058333	35,296405		

- **Suma de todo lo recogido en un año por distrito.**

Suma de todo lo recogido en un año por distrito:

	nom_ditrto	anio	Suma	Toneladas
0	Arganzuela	2021	43351,922	
1	Barajas	2021	18499,561	
2	Carabanchel	2021	83218,828	
3	Centro	2021	74160,461	
4	Chamartín	2021	50307,875	
5	Chamberí	2021	47037,707	
6	Ciudad Lineal	2021	68199,734	
7	Fuencarral - El Pardo	2021	86267,578	
8	Hortaleza	2021	60016,137	
9	Latina	2021	72393,133	
10	Moncloa - Aravaca	2021	59491,098	
11	Moratalaz	2021	28804,938	
12	Puente de Vallecas	2021	76623,203	
13	Retiro	2021	40118,055	
14	Salamanca	2021	53657,559	
15	San Blas - Canillejas	2021	51812,082	
16	Sin Distrito	2021	153,810	
17	Tetuán	2021	53028,875	
18	Usera	2021	42565,445	
19	Vicálvaro	2021	29657,307	
...				

- Por cada distrito obtener para cada tipo de residuo la cantidad recogida.

```
Por cada distrito obtener para cada tipo de residuo la cantidad recogida:
nom_distrito      residuos Suma Toneladas
0  Arganzuela      RESTO      22728,25000
1  Arganzuela      ENVASES      3882,78027
2  Arganzuela      VIDRIO      3308,16016
3  Arganzuela      ORGANICA      7413,06006
4  Arganzuela      PAPEL-CARTON      4255,71484
5  Arganzuela      CARTON COMERCIAL      306,34000
6  Arganzuela      RCD      589,81000
7  Arganzuela      PUNTOS LIMPIOS      867,26001
8  Arganzuela      CONTENEDORES DE ROPA      0,54000
9  Barajas          RESTO      9823,52051
10 Barajas          ENVASES      1816,44006
11 Barajas          VIDRIO      1286,18005
12 Barajas          ORGANICA      4158,90039
13 Barajas          PAPEL-CARTON      642,66003
14 Barajas          RCD      298,66501
15 Barajas          PUNTOS LIMPIOS      411,48001
16 Barajas          CARTON COMERCIAL      61,71000
17 Carabanchel      RESTO      52763,52734
18 Carabanchel      ENVASES      8773,94043
19 Carabanchel      VIDRIO      3372,70020
...
```

- Tiempo de generación del mismo en milisegundos.

- Tiempo de generacion del mismo en milisegundos: 5561

Si la opción es resumen distrito directorio_origen
directorio_destino:

Hemos escogido el distrito Chamberí.

- Número de contenedores de cada tipo que hay en este distrito.

Número de contenedores de cada tipo que hay en este distrito:

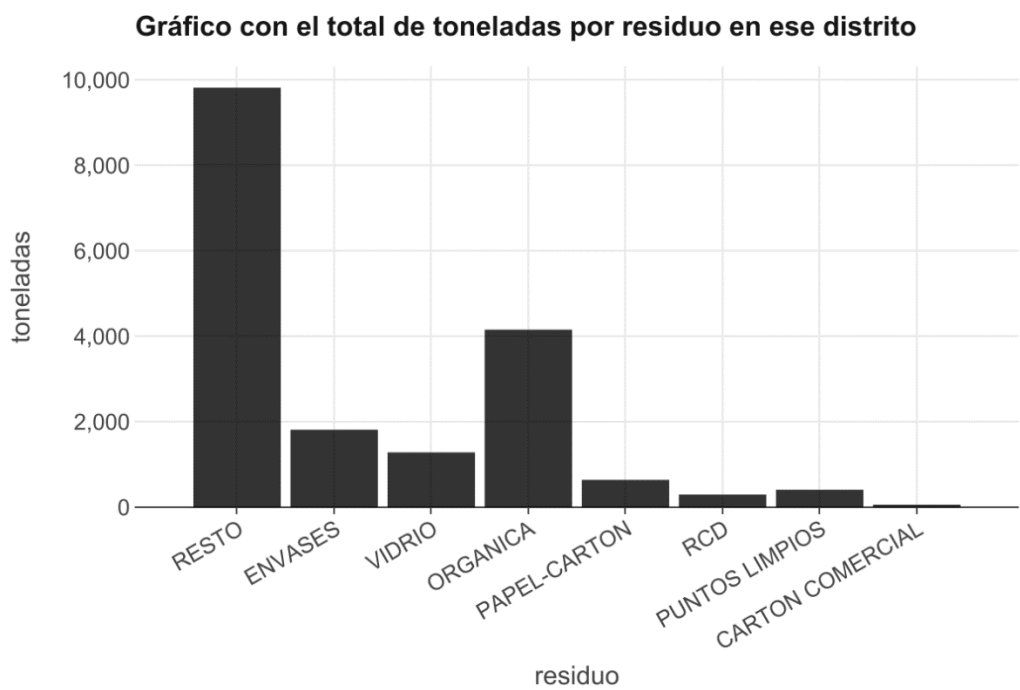
	distrito	tipoContenedor	Total Contenedores
0	CHAMBERI	PAPEL-CARTON	272
1	CHAMBERI	VIDRIO	265

- Total de toneladas recogidas en ese distrito por residuo.

Total de toneladas recogidas en ese distrito por residuo:

	nom_ditrto	residuos	Suma de Toneladas
0	Chamberí	RESTO	26352,93945
1	Chamberí	ENVASES	3823,14014
2	Chamberí	VIDRIO	2753,47998
3	Chamberí	ORGANICA	9398,88086
4	Chamberí	PAPEL-CARTON	3296,86963
5	Chamberí	CARTON COMERCIAL	760,95996
6	Chamberí	RCD	651,45496

- Gráfico con el total de toneladas por residuo en ese distrito.

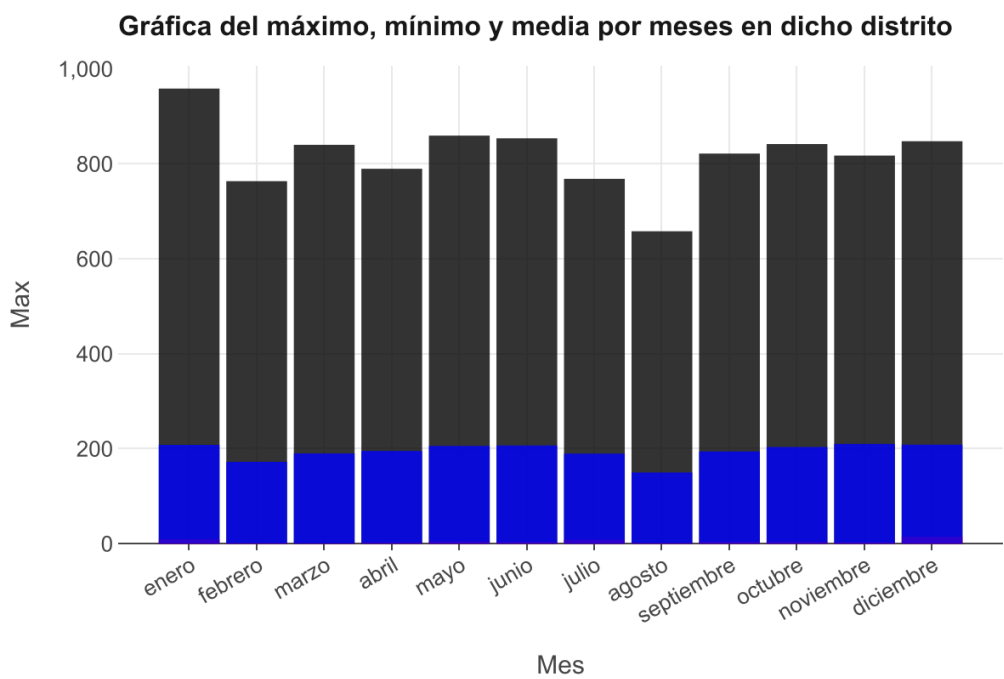


- Máximo, mínimo , media y desviación por mes por residuo en dicho distrito.

Máximo, mínimo , media y desviación por mes por residuo en dicho distrito:

	nom_ditrto	residuos	mes	Maximo	Minimo	Media	Desviacion
0	Chamberí	RESTO	enero	2403	2403	2403,340088	NaN
1	Chamberí	ENVASES	enero	213	213	212,940002	NaN
2	Chamberí	VIDRIO	enero	225	225	225,279999	NaN
3	Chamberí	ORGANICA	enero	524	524	523,760010	NaN
4	Chamberí	PAPEL-CARTON	enero	269	269	268,950012	NaN
5	Chamberí	CARTON COMERCIAL	enero	62	62	62,200001	NaN
6	Chamberí	RCD	enero	35	35	34,830002	NaN
7	Chamberí	RESTO	febrero	2166	2166	2165,620117	NaN
8	Chamberí	ENVASES	febrero	324	324	323,959991	NaN
9	Chamberí	VIDRIO	febrero	227	227	227,059998	NaN
10	Chamberí	ORGANICA	febrero	758	758	757,760010	NaN
11	Chamberí	PAPEL-CARTON	febrero	257	257	257,299988	NaN
12	Chamberí	CARTON COMERCIAL	febrero	65	65	64,800003	NaN
13	Chamberí	RESTO	marzo	2351	2351	2351,459961	NaN
14	Chamberí	ENVASES	marzo	355	355	355,179993	NaN
15	Chamberí	VIDRIO	marzo	252	252	252,139999	NaN
16	Chamberí	ORGANICA	marzo	828	828	828,119995	NaN
17	Chamberí	PAPEL-CARTON	marzo	307	307	307,250000	NaN
18	Chamberí	CARTON COMERCIAL	marzo	76	76	75,699997	NaN
19	Chamberí	RCD	marzo	86	86	85,730003	NaN
...							

- Gráfica del máximo, mínimo y media por meses en dicho distrito.



- Tiempo de generación del mismo en milisegundos.

Tiempo de generación del mismo en milisegundos:2671

Aplicación de otras técnicas que consideres interesantes

COLECCIONES

Otra técnica que me parecen muy interesante es la utilización de colecciones.

Justificación tecnológica

KOTLIN

Hemos utilizado el lenguaje Kotlin porque permite muchas de las funcionalidades que existen para Java.

Kotlin es simple pero a la vez potente, por lo que tienes un mundo abierto de posibilidades si venías de desarrollar aplicaciones en Java.

No te puedes imaginar el problema que tiene java con los null pointer exception. Por culpa de este problema se produjo el error del billón de dólares. Hoy en día tenemos herramientas como las anotaciones, o algunos patrones de diseño, que nos pueden ayudar a evitar gran parte. Pero, como siempre, la cantidad de trabajo extra necesaria es un problema en tiempo y dinero. Al utilizar Kotlin nos ahorramos este problema.



DATA FRAME

Hemos utilizado Data Frame debido a la simplicidad que ofrece y la potencia que tiene. El código es más fácil de entender cuando lo lee otra persona y, es más difícil que se introduzcan errores.