

L'execució d'aquesta pràctica es fonamenta en 2 objectius principals:

1. Generar un conjunt de scripts per extreure informació valuosa i sintetitzada d'un conjunt de dades que permetin entendre l'estat actual de desenvolupament de la xarxa de càrrega de vehicles de Catalunya.
2. Fer un seguiment de l'evolució d'aquests principals indicadors a través de l'aplicació dels mateixos scripts cada vegada que les dades siguin actualitzades.

Considerant el conjunt de dades disponible en el portal anlisi.transparenciacatalunya.cat, la resolució dels reptes plantejats per assolir els objectius esmentats anteriorment es considera òptima a través de tècniques d'*scripting* perquè aquestes permeten:

1. La gestió de les dades (descarregues, emmagatzematge, manipulació de fitxers, ...)
2. La modificació de les dades (filtrar, substituir, retallar, enganxar, ...)
3. L'anàlisi de les dades i la generació de resums ja sigui a través d'impressió en pantalla, generació d'arxius de text pla o fins i tot la creació d'informes html.

Els principals reptes que ens proposa el conjunt de dades escollit són els següents:

- L'existència de registres buits, sense informació; en diferents camps.
- El separador de camps és «,» però també tenim un camp *string* (camp=adreça) que té «,» en el seu interior.
- Registres «duplicats» (camp=municipi), escrits algunes vegades amb majúscules i altres amb minúscules, així com registres escrits algunes vegades amb accent i altres sense.
- Un camp (camp=places) que en certes ocasions representa el seu valor com a nombre enter, però altres vegades ho fa com a suma de dos o més enters generant així un *string*.

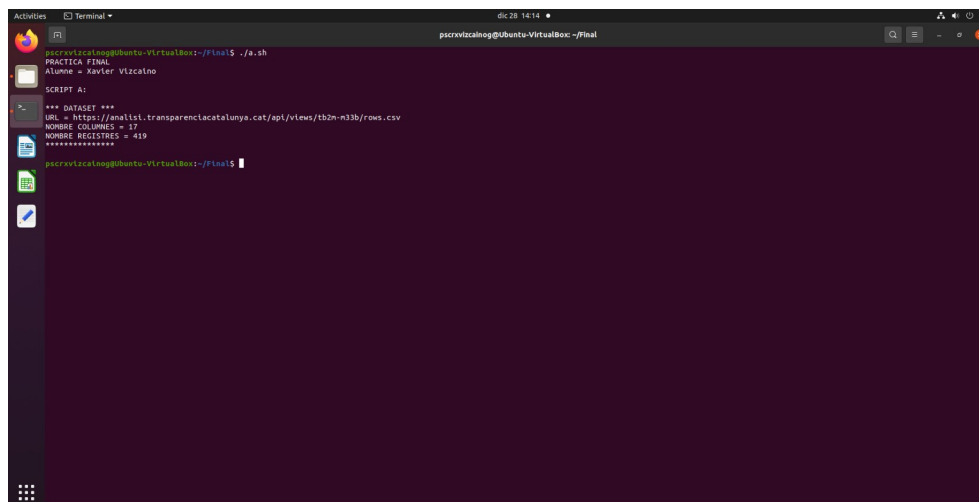
SCRIPT A:

En l'*script* A, es genera l'estructura d'informació que es mantindrà durant l'execució de tots els scripts. Aquesta estructura consta de 2 carpetes, una anomenada «dataset» on s'emmagatzemaran l'arxiu de dades inicial i les posteriors modificacions realitzades sobre aquests i una carpeta anomenada «resultats» on s'emmagatzemaran els arxius de text pla resultat de l'execució de cada un dels scripts, així com l'informe final en format html.

També es comprova si existeixen dades prèvies en l'estructura i en el cas que així sigui a aquestes se'ls canvia el nom de l'arxiu abans d'iniciar la descàrrega del nou conjunt de dades.

L'*script* A es pot executar amb opcions:

- Sense opció; retorna la *url* de descàrrega, el nombre de camps i el nombre de registres del conjunt de dades descarregat.



```
pscrvitzcalnog@ubuntu-VirtualBox: ~/Final$ ./a.sh
PRÀCTICA FINAL
Alumne = Xavier Vitzcalno

SCRIPT A:

*** DATASET ***
URL = https://onalsi.transparenciacatalunya.cat/api/views/tb2n-m33b/rows.csv
NOMBRE COLUMNES = 17
NOMBRE REGISTRES = 419
*****
pscrvitzcalnog@ubuntu-VirtualBox: ~/Final$
```

- Opció -v, addicionalment retorna el tipus de dades per cada un dels camps del *dataset* (text, decimal, enter, format desconegut). El tipus de dades s'obté, en aquest cas, comparant el valor del camp amb expressions regulars. Cal destacar que, la informació retornada prové del primer registre del *dataset* que no té cap camp buit (obtingut enllaçant comandes **sed**, **grep**, **head** i **tail**).

```

pscrxvicalnog@ubuntu-VirtualBox: ~/Final$ ./a.sh -v
PRÀCTICA FINAL
Alumne = Xavier Vicalno

SCRIPT A:

*** DATASET ***
URL = https://anlisi.transparenciacatalunya.cat/api/views/tb2n-m3b/rows.csv
NOMBRE COLUMNES = 17
NOMBRE REGISTRES = 419
FORMAT = CSV text
ESPECIFICACIO COLUMNES:
Columna 1 = text
Columna 2 = text
Columna 3 = text
Columna 4 = text
Columna 5 = decimal
Columna 6 = decimal
Columna 7 = text
Columna 8 = enter
Columna 9 = text
Columna 10 = text
Columna 11 = text
Columna 12 = text
Columna 13 = enter
Columna 14 = text
Columna 15 = enter
Columna 16 = text
Columna 17 = text
*****

pscrxvicalnog@ubuntu-VirtualBox: ~/Final$

```

- Opció -f, addicionalment retorna el tipus de dades per cada un dels camps del *dataset* (text, decimal, enter, ...). El tipus de dades s'obté, en aquest cas, emprant la comanda `csvstat --type` del paquet `csvkit`.

```

pscrxvicalnog@ubuntu-VirtualBox: ~/Final$ ./a.sh -f
PRÀCTICA FINAL
Alumne = Xavier Vicalno

SCRIPT A:

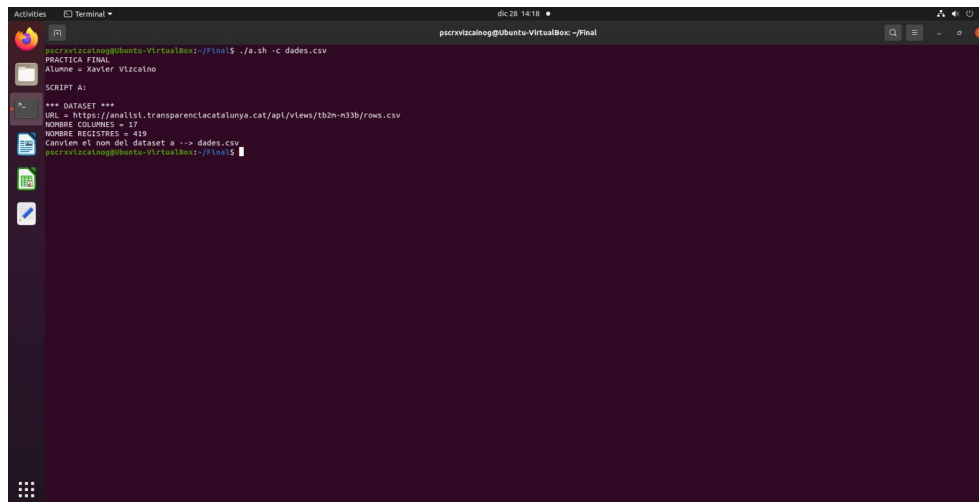
*** DATASET ***
URL = https://anlisi.transparenciacatalunya.cat/api/views/tb2n-m3b/rows.csv
NOMBRE COLUMNES = 17
NOMBRE REGISTRES = 419
FORMAT = CSV text
ESPECIFICACIO COLUMNES:
1. PROMOTOR-GESTOR: Text
2. ACCES: Text
3. TIPUS VELOCITAT: Text
4. TIPUS CONDUCTOR: Text
5. LATITUD: Number
6. LONGITUD: Number
7. DESTINACIO-DESCRIPTIVA: Text
8. POTENCIA: Number
9. TIPUS DE CORRECCIÓ: Text
10. IDENTIFICADOR: Text
11. ADRICA: Text
12. PROVINCIA: Text
13. CODIPROV: Number
14. MUNICIPI: Text
15. NPLACES ESTACIÓ: Text
16. TIPUS VEHICLES: Text
17. Columna amb georeferència: Text
*****

pscrxvicalnog@ubuntu-VirtualBox: ~/Final$

```

- Opció -c, addicionalment, es canvia el nom de l'arxiu *dataset* descarregat. Cal destacar que en aquest cas s'haurà d'afegir el nou nom de l'arxiu com a argument a la comanda:

(./a.sh [-c] <nom.ext>)

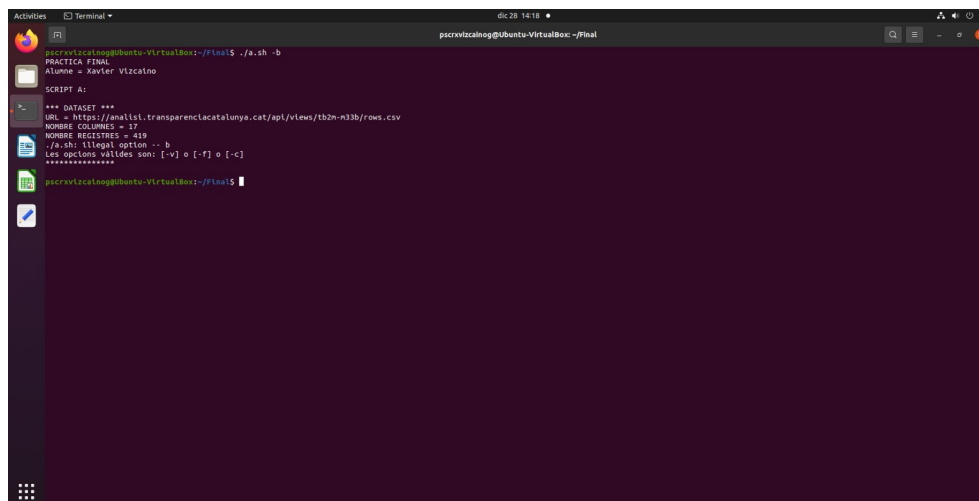


```
pscrxvicalnog@ubuntu-virtualbox:~/Final$ ./a.sh -c dades.csv
PRACTICA FINAL
Alumne = Xavier Vízcalno

SCRIPT A:

*** DATASET ***
URL = https://analis1.transparenciacatalunya.cat/api/v1em/tb2m-m3b/rows.csv
NOMBRE COLUMNES = 17
NOMBRE REGISTRES = 419
Canvien el nom del dataset a --> dades.csv
pscrxvicalnog@ubuntu-virtualbox:~/Final$
```

- Altres opcions (*), s'informa per pantalla que les opcions vàlides són [-v], [-f] o [-c]



```
pscrxvicalnog@ubuntu-virtualbox:~/Final$ ./a.sh -b
PRACTICA FINAL
Alumne = Xavier Vízcalno

SCRIPT A:

*** DATASET ***
URL = https://analis1.transparenciacatalunya.cat/api/v1em/tb2m-m3b/rows.csv
NOMBRE COLUMNES = 17
NOMBRE REGISTRES = 419
./a.sh: illegal option -- b
Les opciones vàlides son: [-v] o [-f] o [-c]
*****
pscrxvicalnog@ubuntu-virtualbox:~/Final$
```

SCRIPT B

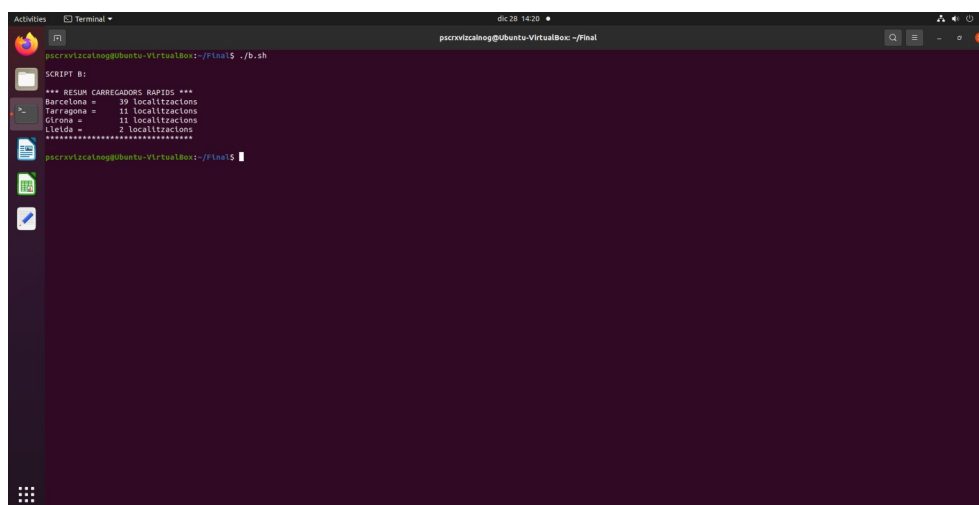
L'objectiu de l'script B és:

1. Transformar les dades per a resoldre la major part dels reptes que ens introdueix el *dataset* escollit.
2. Reportar el nombre de localitzacions on podem trobar carregadors ràpids per cada província.

Amb relació al primer objectiu, en l'*script* hi trobarem:

1. Buscar i substituir els camps buits per «no_data» a través de l'editor **sed**, per a la realització d'aquesta tasca cal destacar les següents particularitats:
 1. el camp buit pot estar a l'inici del registre.
 2. Pot existir més d'un camp buit de manera consecutiva
2. Canviar el delimitador de camps de «,» a «;» a través d'una rutina **awk**, per tal de facilitar la gestió de les dades en posteriors passos.
3. Reorganitzar els camps amb les comandes **cut** i **paste**.
4. Substituir majúscules per minúscules a través de l'editor **sed** en el primer camp (MUNICIPI), per tal d'evitar duplicats erronis.
5. Eliminar l'accentuació a través de l'editor **sed**, també per evitar falsos duplicats.
6. Finalment, substituir els « » per «_» per tal de facilitar la manipulació d'*strings* també amb l'editor **sed**.

Pel que fa al segon objectiu, es realitzarà un filtratge de les dades a través de la comanda **grep**, introduint una expressió regular com a patró de cerca i el resultat, es traslladarà a l'entrada de la comanda **wc**.



```
pscrvitzcalnog@ubuntu-VirtualBox:~/Final$ ./b.sh
SCRIPT B:
*** RESUM CARREGADORS RAPIDS ***
Barcelona = 39 localitzacions
Tarragona = 11 localitzacions
Girona = 11 localitzacions
Llíuda = 2 localitzacions
*****
```

SCRIPT C bash

L'objectiu del *script* C_bash és resoldre el darrer dels reptes que ens plantejava el *dataset* escollit i reportar el nombre de places d'aparcament aptes per a la càrrega de vehicles existents en cada municipi. Així doncs en l'*script* hi trobarem:

1. Un *loop* **while** que llegirà tots els registres, per cada registre comprovarem si el camp=places està format per la suma de dos o més enters amb un *if* i una expressió regular com a patró. Per als registres que compleixin el patró, es buscarà el nombre d'elements (3 o 5) que té la informació del camp i a través d'un *case* amb rutines **awk** definint com a limitador el símbol «+» es sumaran els enters.
2. La reorganització de les dades amb comandes **cut** i **paste**.
3. Un filtratge de dades amb la comanda **grep**, per quedar-nos només aquelles que tenen un valor enter (no alfanumèric) en el camp places.
4. L'extracció del llistat de municipis i la generació d'un arxiu de text pla amb aquest llistat, enllaçant comandes **cut**, **sort** i **uniq**.
5. Finalment, un doble *loop* **while**, on el **loop** principal llegirà cada un dels registres de l'arxiu de municipis generat. Dins del *loop* principal, trobarem:
 1. Un filtratge de dades amb la comanda **grep**, per quedar-nos només amb els camps corresponents al municipi que s'està analitzant en el *loop* principal, del *dataset* que conté les darreres modificacions.
 2. Un *loop* **while** secundari, que recorrerà tots els registres provinents del filtratge anterior, realitzant un sumatori dels valors enters trobats en el camp places.
6. Finalment, algunes tasques per a donar un format més correcte a la sortida per pantalla.

```

Activities Terminal
pscrxvicalnog@buntu-VirtualBox:~/Final$ ./c_bash.sh ./dataset/maindata.txt

SCRIPT C_bash:

***** RESUM PLACES DE CARGA *****
alcanar 1
alella 2
amposta 2
asco 2
avia 1
badalona 29
balaguer 2
benyoles 1
barbera_del_valles 4
barcelona 489
belver_de_cerdanya 1
berga 2
beuda 2
cabreria_de_mar 6
cabrits 1
caldes_de_malavella 10
caldes_de_montbui 2
caldes_d'estrach 2
calella 2
canet_de_mar 10
cardona 2
castelldefels 3
celra 1
conistants 1
cornella_de_llobregat 7
cunit 2
el_hassou 2
el_prat_de_llobregat 6
el_vendrell 3
enpuries 2
esparreguera 1
espugues_de_llobregat 1
figueras 7
gava 8
girona 11
granollers 7
guissona 1
iguada 3
la_canonja 4
la_garriga 1
la_louquera 6
l'aldea 6
l'alella_de_mar 2
l'ampolla 1

```

```

Activities Terminal
l'ampolla 1
la_rocka_del_valles 5
la_selva_del_camp 2
la_sen_d'urgell 4
la_sen_d'urgell 10
l'estarrit 2
l'hospitalet_de_llobregat 12
lleida 2
llicia_d'amunt 2
llinars_del_valles 2
lloret_de_mar 7
manresa 7
martinet 2
martorell 2
mataro 24
molins_de_rei 2
mollet_del_valles 2
monclús 2
moncada_i_reixac 2
monmeló 3
montorres_del_valles 2
mont-roig_del_camp 2
olot 2
pachs_del_penedes 2
pallás_de_piganans_l_solta 2
pallás 2
perallada 2
pinada_de_mar 2
plaja_d'aro 1
pulperda 24
reus 2
ripoll 2
ripollit 2
roses 4
rubí 13
sabadell 22
s'agaro 2
salou 2
santa_coloma_de_farners 2
santa_coloma_de_gramenet 9
santa_cristina_d'aro 1
sant_adria_de_bages 2
santa_llogaia_d'alguema 2
santa_perpetua_de_la_mogoda 1
sant_clement_sescibes 1
sant_cugat_del_valles 18
sant_feliu_de_guixols 4
sant_fruitos_de_bages 2
sant_joan_de_sop 4
sant_joan_de_vilatorrada 2
santpedor 1
sant_pere_de_ribes 8
sant_pere_pescador 4
sant_pere_vilatorrada 2
sant_quirze_del_valles 2
sant_vicenç_de_castellet 2
sitges 11
subirats 4
terragona 6
terradeselles 1
terrasa 10
tona 1
torrefrera 6
torrent 1
tortosa 2
tremo 2
vallromanes 1
valls 2
vandellos_i_l'hospitalet_i_l'infant 2
vic 0
vidreres 2
viladecans 4
viladecavalls 7
viladrau 2
vilafraanca_del_penedes 4
vilanova_l'la_geltru 2
vileca 1
*****
pscrxvicalnog@buntu-VirtualBox:~/Final$

```

```

Activities Terminal
ripoll 2
ripollit 2
roses 4
rubí 13
sabadell 22
s'agaro 2
salou 2
santa_coloma_de_farners 2
santa_coloma_de_gramenet 9
santa_cristina_d'aro 1
sant_adria_de_bages 2
santa_llogaia_d'alguema 2
santa_perpetua_de_la_mogoda 1
sant_clement_sescibes 1
sant_cugat_del_valles 18
sant_feliu_de_guixols 4
sant_fruitos_de_bages 2
sant_joan_de_sop 4
sant_joan_de_vilatorrada 2
santpedor 1
sant_pere_de_ribes 8
sant_pere_pescador 4
sant_pere_vilatorrada 2
sant_quirze_del_valles 2
sant_vicenç_de_castellet 2
sitges 11
subirats 4
terragona 6
terradeselles 1
terrasa 10
tona 1
torrefrera 6
torrent 1
tortosa 2
tremo 2
vallromanes 1
valls 2
vandellos_i_l'hospitalet_i_l'infant 2
vic 0
vidreres 2
viladecans 4
viladecavalls 7
viladrau 2
vilafraanca_del_penedes 4
vilanova_l'la_geltru 2
vileca 1
*****
pscrxvicalnog@buntu-VirtualBox:~/Final$

```

SCRIPT C_awk

L'objectiu del *script* C_awk és proporcionar informació sobre les potències de càrrega disponibles per províncies. Llavors, per a cada província es busca el valor màxim i mínim del camp potència. Tanmateix, es realitza el càlcul de la potència mitjana considerant la següent suposició:

Totes les places d'aparcament definides en el camp places, d'una localització donada (registre) tenen un carregador i aquest és de la potència definida en el camp potència

I també es reporta el número de places d'aparcament per a càrrega de vehicles totals a la província.

Cal destacar que en aquest cas, es decideix emmagatzemar la informació generada durant l'execució del script en *arrays*, per tal de poder complir amb el requisit d'aparició d'estructures iteratives.

```

pscrxvicalnog@Ubuntu-VirtualBox:~/Final$ gawk -f C_awk.awk ./dataset/maindata_2.txt

SCRIPT C_awk:

*** RESUM POTENCIES DE CÀRREGA ***
PROVINCIA  POTENCIA MITJANA  MIN  MAX  PLACES
Barcelona  29.75kW            0kW  120kW  240
Tarragona  59.21kW            7kW  120kW  40
Girona     59.68kW            11kW 120kW  60
Lleida     76.30kW            7kW  120kW  30
*****
pscrxvicalnog@Ubuntu-VirtualBox:~/Final$

```

SCRIPT D

L'objectiu de l'*script* D és generar un informe en format html que contingui la informació generada en cada un dels *scripts* anteriors.

Així doncs en l'*script* D hi trobarem, principalment 2 tipus d'estructures:

1. Concatenació d'informació fins a un indicador amb la comanda **cat**, seguint l'estructura:

```

cat > arxiu.html < < «indicador»
text
text
«indicador»

```

2. Crides a una funció anomenada creataula, que necessita l'arxiu de dades a partir del qual es vol crear la taula com a argument.

Al finalitzar la creació del document html, obrim l'informe amb *firefox*, generant un nou procés aïllat (incloent-hi & al final de la línia d'execució de *firefox*) que li permeti a *bash* continuar amb l'execució de l'*script*.

Activities Firefox Web Browser dic 28 14:26

Programació Scripting_PP file:///home/jsccvitzcalnog/Final/resultats/Report.html

Programació en Scripting

Practica Final

Nom: Xavier Vitzcalnog
Usuari: xvitzcalnog

Exercici A:

A continuació es mostra la informació bàsica obtinguda en l'exercici A: URL, número de columnes del dataset i número de registres d'aquest

Taula resultat

URL	NOMBRE COLUMNES	NOMBRE REGISTRES
https://analisi.transparenciacatalunya.cat/api/views/b2m-m3b/rows.csv	17	419

Exercici B:

A continuació es mostra el nombre d'ubicacions per cada província que tinguin com a mínim 1 carregador RAPID

Taula resultat

PROVINCIA	CARREGADORS RAPIDS
Barcelona	39
Tarragona	11
Girona	11
Lleida	2

Exercici C_bash:

A continuació es mostra el nombre de places d'aparcament designades a la càrrega de vehicles per cada municipi de catalunya

Activities Firefox Web Browser dic 28 14:28

Programació Scripting_PP file:///home/jsccvitzcalnog/Final/resultats/Report.html

sant pere de ribes	8
sant pere pescador	4
sant pere vilamajor	2
sant quirze del valles	2
sant vicenç de castellet	2
salpes	11
sallers	4
tarragona	6
terredellas	1
terrossa	18
tona	1
torrelatreia	6
torrent	1
tortosa	2
trepç	2
vallromanes	1
valls	2
vandellòs i l'hostralet i l'infant	2
vc	6
vidreres	2
viladecans	4
viladecavalls	7
viladrau	2
vilafraua del panades	4
vilanova i la geltrú	2
vilaseca	1

Exercici C_awk:

A continuació es mostra la potència mitjana, mínima i màxima dels carregadors agrupats per províncies

Taula resultat

PROVINCIA	POT	AV	MIN	MAX	PLACES
Barcelona	29.75	6	120	240	
Tarragona	50.21	7	120	48	
Girona	50.86	11	120	66	
Lleida	76.39	7	120	36	

* No s'han inclòs imatges mostrant tota l'extensió de l'informe en html per considerar que les seleccionades ja permeten tenir una visió clara del contingut d'aquest.

SCRIPT RUN

L'*script run* s'encarrega de l'execució de tots els scripts presentats anteriorment considerant les invocacions esmentades en l'enunciat de la pràctica. Addicionalment, i per tal de complir amb els requisits de la pràctica realitza un càlcul del temps d'execució per a comprovar que aquest sigui inferior a 10 seg.

The first screenshot shows a terminal window with the command `pscrvfinal@ubuntu-VirtualBox:~/Final$./run.sh`. The script output includes dataset information, a list of 17 locations, and a summary of charging stations by province.

The second screenshot shows the output of the script, which includes a list of 17 locations and a summary of charging stations by province.

The third screenshot shows a web browser displaying the final report, which includes the title "Programació en Scripting", the author "Xavier Vizcaino", and the results of the exercises.

COMENTARIS FINALS

Com a tasca més important apresada en l'elaboració del projecte destacaria: 1. la integració de totes les eines presentades al llarg de l'assignatura buscant els punts forts de cada una. Com es pot observar en els scripts preparats, en nombroses ocasions apareixen comandes **grep** i/o **sed** en *loops* de *bash*; on potser també tenim *if* amb expressions regulars i alguna invocació a **awk** i 2. la creació d'un informe en format html, força més visual que en format de text pla.

Per a l'extensió del projecte, es podria considerar: 1. la inclusió de gràfics generats amb R (ggplot2) en l'informe html i 2. una gestió millorada i més robusta de les dades; incloent-hi la data d'actualització del *dataset* en els noms dels fitxers *rows.csv*, *maindata.txt*, *maindata_2.txt* i *outputs_x.txt* (exemple *YYMMDD_rows.csv*), per tal de poder mantenir un històric de les dades i analitzar-ne l'evolució.