

Actividad 3 del Proyecto final GNU-Linux

Liliana Aguilar

Las siguientes líneas son utilizadas para obtener el total de ingresos para cada una de las estaciones de la **Línea 1** del metro para la fecha **2021-01-01** a partir del archivo **afluenciastc_desglosado_02_2025.csv**:

```
grep 2021-01-01 afluenciastc_desglosado_02_2025.csv | grep -w "Línea 1" |  
cut -d, -f 5 |  
uniq > estaciones
```

```
grep 2021-01-01 afluenciastc_desglosado_02_2025.csv | grep -w "Línea 1" > ingresos
```

```
cat estaciones | while read EST; do echo -n "$EST: "  
grep "$EST" ingresos | cut -d, -f7 | awk '{ VAR += $1 } ; END {print VAR}' done
```

a) ¿Qué resultado se obtiene al eliminar **grep -w "Línea 1"** de la secuencia de comandos? explique

b) A partir del razonamiento anterior, ¿qué modificación debe realizarse para obtener el total de ingresos de todas las estaciones (todas las líneas del metro) para el mes de enero de 2021?

c) A partir del resultado anterior, ¿qué modificaciones deben realizarse para obtener el total de ingresos de todas las estaciones (todas las líneas del metro) para el año 2021?

d) Reportar la estación con más número de ingresos para los años 2021, 2022, 2023, 2024

e) A partir de los resultados anteriores, ¿qué modificaciones deben realizarse para obtener el total de ingresos de todas las estaciones para todos los registros en el archivo, en otras palabras para todos los años registrados 2021 .. 2025?

** Recuerde que una estación puede formar parte de diferentes líneas, por ejemplo la estación "Chabacano" pertenece a las líneas 2, 8 y 9.*

Para el inciso a)

Si se elimina esta línea: `grep -w "Línea 1"`, lo que hará el comando `grep 2021-01-01` con el archivo `afluenciastc_desglosado_02_2025.csv` es que filtrará todas las líneas que se encuentren en la fecha 2021-01-01, sin importar la línea del metro a la que pertenezcan, por lo que el archivo denominado "estaciones" contendrá todas las estaciones de todas las líneas del metro para la fecha señalada y no sólo las de la Línea 1 como al principio sin el filtro. También en esta parte: `grep 2021-01-01 afluenciastc_desglosado_02_2025.csv`

| grep -w "Línea 1" > **ingresos**. El archivo que se genera denominado "ingresos" incluirá los datos de todas las estaciones de todas las líneas para esa fecha.

Para el último caso con el comando cat, la parte del bucle "while" calculará el total de ingresos para todas las estaciones de todas las líneas y no solo las de la Línea 1.

Por lo tanto, el resultado será el total de ingresos para todas las estaciones de todas las líneas del metro para la fecha del 01 de enero de 2021.

Para el inciso b)

Para poder obtener el filtro para el mes de enero de 2021, se debe modificar el comando "grep" para que busque las líneas que contengan la fecha de "Enero,2021" y no la de todas las fechas. De esta manera:

- 1) `grep "Enero,2021" afluenciastc_desglosado_02_2025.csv | cut -d, -f 5 | uniq > estaciones`
- 2) `grep "Enero,2021" afluenciastc_desglosado_02_2025.csv > ingresos`
- 3) `cat estaciones | while read EST; do echo -n "$EST: "; grep "$EST" ingresos | cut -d, -f7 | awk '{ VAR += $1 } ; END {print VAR}'; done --- No se modifica, solo las anteriores`

Con la modificación anterior se filtrarán todas las estaciones únicamente para el mes de enero para todas las líneas del metro.

Para el inciso c)

De nuevo se tiene que modificar el comando que nos permite filtrar, es decir el "grep" para que busque las líneas que contengan el 2021 como año en la columna correspondiente. Las líneas de comandos quedarían de la siguiente manera:

- 1) `grep ",2021" afluenciastc_desglosado_02_2025.csv | cut -d, -f 5 | uniq > estaciones`
- 2) `grep ",2021" afluenciastc_desglosado_02_2025.csv > ingresos`
- 3) `cat estaciones | while read EST; do echo -n "$EST: "; grep "$EST" ingresos | cut -d, -f7 | awk '{ VAR += $1 } ; END {print VAR}'; done. ----- Esta no se modifica, las otras dos sí para poder visualizarse en bash. Con los cambios anteriores esto calcula el total de ingresos para todas las estaciones durante todo el año.`

En el inciso d)

Para este inciso es necesario agregar un bucle para poder iterar sobre cada año y así ir sumando los ingresos por estación, también hay que tener en cuenta que hay estaciones

que tienen espacios en su nombre como: Gómez Farias, por lo que se agregan más filtros, por lo tanto, se modifican las líneas de comandos de la siguiente manera y se muestra los resultados de pantalla.

```
for ANIO in {2021..2024}; do echo "AÑO $ANIO"; grep "^$ANIO-"
afuenciastc_desglosado_02_2025.csv | cut -d, -f5,7 | awk -F, '{ingresos[$1]+=$2} END {for(e in ingresos) print e, ingresos[e]}' | sort -k2 -nr | head -n1; echo "-----"; done
```

Resultado de pantalla:

```
a.8133@tlaloc ~/gnu $ for ANIO in {2021..2024}; do
echo "AÑO $ANIO"
grep "^$ANIO-" afuenciastc_desglosado_02_2025.csv | cut -d, -f5,7 |
awk -F, '{ingresos[$1] += $2} END {for (e in ingresos) print e, ingresos[e]}' |
sort -k2 -nr | head -n 1
echo "-----"
done
AÑO 2021
Pantitlán 45993260
-----
AÑO 2022
Pantitlán 51542799
-----
AÑO 2023
Pantitlán 48523294
-----
AÑO 2024
Pantitlán 37141821
-----
a.8133@tlaloc ~/gnu $
```

Para el inciso e):

```
cut -d, -f5,7 afuenciastc_desglosado_02_2025.csv | awk -F, '{ingresos[$1]+=$2} END
{suma=0; for(e in ingresos) suma+=ingresos[e]; print "Total ingresos de todas las
estaciones:", suma}'
```

Se modifica:

Por lo tanto, se debe eliminar el filtro de grep para el año y el bucle for y hacer una suma general como en el inciso d), con estos cambios te hace la suma los ingresos (tomando las columnas de estación en 5 y columna 7 de ingresos) de todas las estaciones del archivo.

Salida:

```
a.8133@tlaloc ~/gnu $ echo "Total ingresos de todas las estaciones: $(cut -d, -f5,7 afuenciastc_desglosado_0
2_2025.csv | awk -F, '{ingresos[$1]+=$2} END {suma=0; for(e in ingresos) suma+=ingresos[e]; print suma}')"
Total ingresos de todas las estaciones: 4302420754
```