Practicando o aprendido: Conda, Jupyter Lab e Git

Sistemas de Big Data

12/12/2024 – IES Fernando Wirtz

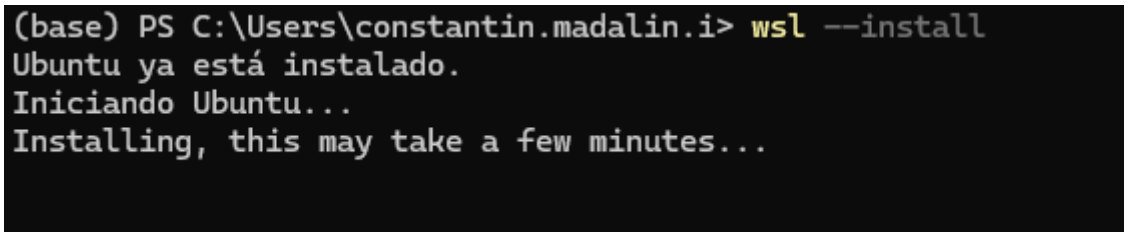
Constantin Madalin Ismana

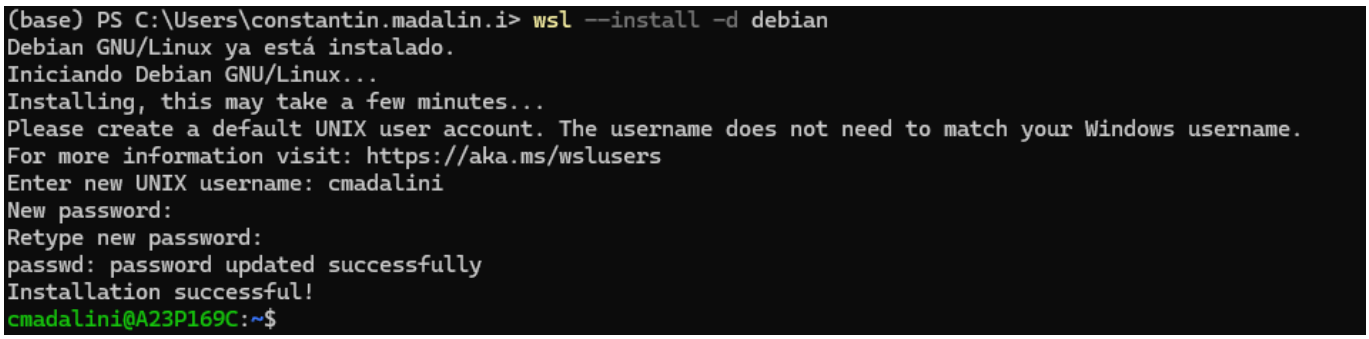
Fecha	Motivo del cambio
12/12/2024	Versión inicial
12/12/2024	Modificación
12/12/2024	Añadido del ejercicio 4

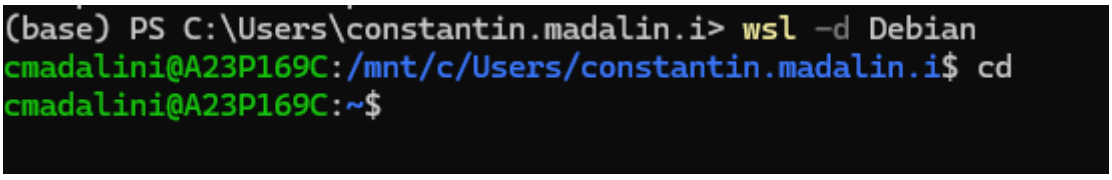
Índice

Instalacion de Wsl

Primer paso vamos a instalar el Wsl con la distrubucion de Debian con los siguientes comandos:







Instalacion del wget

Tenemos que instalar el wget para poder descargar repositorios de internet

```
cmadalini@A23P169C:~$ sudo apt-get install wget
Reading package lists... Done
Building dependency tree... Done
The following additional packages will be installed:
  ca-certificates libpsl5 openssl publicsuffix
The following NEW packages will be installed:
  ca-certificates libpsl5 openssl publicsuffix wget
0 upgraded, 5 newly installed, 0 to remove and 0 not upgraded.
Need to get 2,750 kB of archives.
After this operation, 6,881 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://deb.debian.org/debian bookworm/main amd64 openssl amd64 3.0.15-1~deb12u1 [1,427 kB]
Get:2 http://deb.debian.org/debian bookworm/main amd64 ca-certificates all 20230311 [153 kB]
Get:3 http://deb.debian.org/debian bookworm/main amd64 libpsl5 amd64 0.21.2-1 [58.7 kB]
Get:4 http://deb.debian.org/debian bookworm/main amd64 wget amd64 1.21.3-1+b2 [984 kB]
Get:5 http://deb.debian.org/debian bookworm/main amd64 publicsuffix all 20230209.2326-1 [126 kB]
Fetched 2,750 kB in 1s (4,422 kB/s)
Preconfiguring packages ...
Selecting previously unselected package openssl.
(Reading database ... 9538 files and directories currently installed.)
Preparing to unpack .../openssl_3.0.15-1~deb12u1_amd64.deb ...
```

Instalacion de Conda

Vamos a la informacion de instalación de conda en su pagina e copiamos el seguinte codigo.

```
cmadalini@A23P169C:~$ mkdir -p ~/miniconda3
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh -O ~/miniconda3/miniconda.sh
bash ~/miniconda3/miniconda.sh -b -u -p ~/miniconda3
rm ~/miniconda3/miniconda.sh
--2024-12-12 20:54:12--  https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
Resolving repo.anaconda.com (repo.anaconda.com)... 104.16.191.158, 104.16.32.241, 2606:4700::6810:20f1, ...
Connecting to repo.anaconda.com (repo.anaconda.com)|104.16.191.158|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 148337011 (141M) [application/octet-stream]
Saving to: '/home/cmadalini/miniconda3/miniconda.sh'

/home/cmadalini/miniconda3/mi 100%[=================================================>] 141.46M  58.3MB/s    in 2.4s

2024-12-12 20:54:15 (58.3 MB/s) - '/home/cmadalini/miniconda3/miniconda.sh' saved [148337011/148337011]

PREFIX=/home/cmadalini/miniconda3
Unpacking payload ...

Installing base environment...
```

Creamos el contorno de Bigdata e activarlo

Creamos el entorno bigdata y lo activamos para crear todo dentro de bigdata con los seguientes comandos:

```
cmadalini@A23P169C:~$ source ~/miniconda3/bin/activate
(base) cmadalini@A23P169C:~$
```

```
(base) cmadalini@A23P169C:~$ conda init --all
no change     /home/cmadalini/miniconda3/condabin/conda
no change     /home/cmadalini/miniconda3/bin/conda
no change     /home/cmadalini/miniconda3/bin/conda-env
no change     /home/cmadalini/miniconda3/bin/activate
no change     /home/cmadalini/miniconda3/bin/deactivate
no change     /home/cmadalini/miniconda3/etc/profile.d/conda.sh
no change     /home/cmadalini/miniconda3/etc/fish/conf.d/conda.fish
no change     /home/cmadalini/miniconda3/shell/condabin/Conda.psm1
no change     /home/cmadalini/miniconda3/shell/condabin/conda-hook.ps1
no change     /home/cmadalini/miniconda3/lib/python3.12/site-packages/xontrib/conda.xsh
no change     /home/cmadalini/miniconda3/etc/profile.d/conda.csh
modified      /home/cmadalini/.bashrc
modified      /home/cmadalini/.zshrc
modified      /home/cmadalini/.config/fish/config.fish
modified      /home/cmadalini/.xonshrc
modified      /home/cmadalini/.tcshrc

==> For changes to take effect, close and re-open your current shell. <==

(base) cmadalini@A23P169C:~$
```

```
(base) cmadalini@A23P169C:~$ conda update --all
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): /
```

```
(base) cmadalini@A23P169C:~$ conda --version
conda 24.11.1
(base) cmadalini@A23P169C:~$
```

```
(base) cmadalini@A23P169C:~$ conda create -n bigdata python=3.11
Channels:
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/cmadalini/miniconda3/envs/bigdata

  added / updated specs:
    - python=3.11


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    pip-24.2                   |  py311h06a4308_0         2.8 MB
    python-3.11.11             |       he870216_0        32.9 MB
    setuptools-75.1.0          |  py311h06a4308_0         2.2 MB
    wheel-0.44.0               |  py311h06a4308_0         145 KB
    ------------------------------------------------------------
                                           Total:        38.1 MB
```

```
(base) cmadalini@A23P169C:~$ conda activate bigdata
(bigdata) cmadalini@A23P169C:~$
```

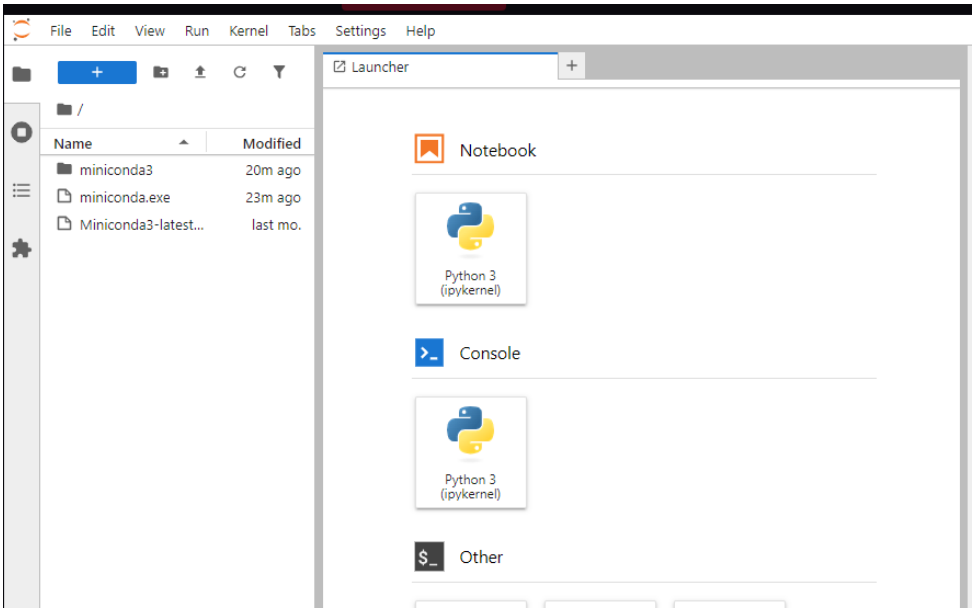
Instalamos los paquetes de Jupyterlab

```
(bigdata) cmadalini@A23P169C:~$ conda install -c conda-forge jupyterlab
Channels:
 - conda-forge
 - defaults
Platform: linux-64
Collecting package metadata (repodata.json): | |
```

```
(bigdata) cmadalini@A23P169C:~$ conda install ipykernell
Channels:
 - defaults
 - conda-forge
Platform: linux-64
Collecting package metadata (repodata.json): \ |
```

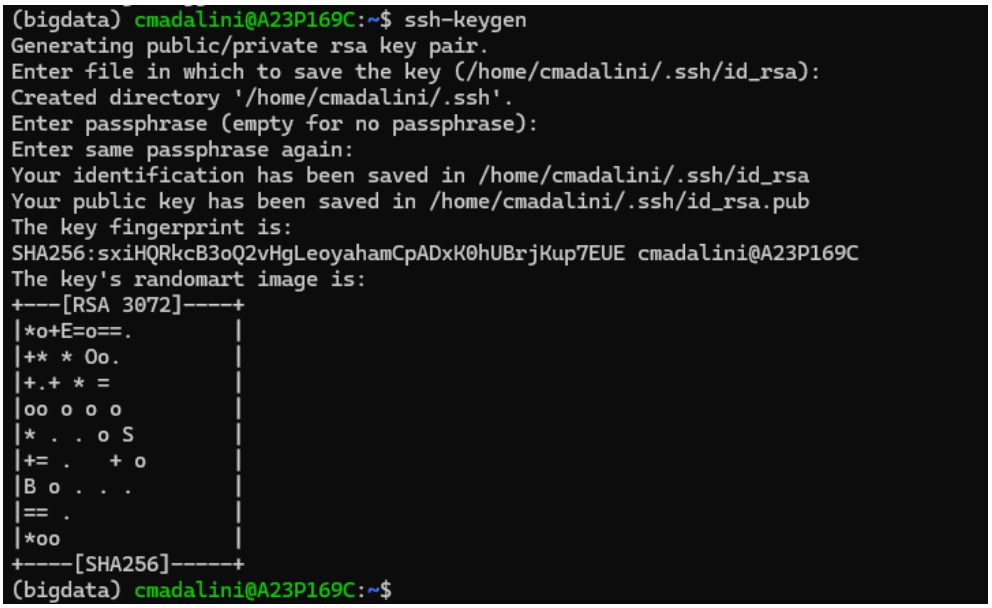
```
        Miniconda3-latest-Windows-x86_64.exe
        miniconda.exe
        miniconda3/

nothing added to commit but untracked files present (use "git add" to track)
(bigdata) cmadalini@A23P169C:~$ jupyter-lab
[I 2024-12-12 21:16:08.495 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-12-12 21:16:08.499 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-12-12 21:16:08.502 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-12-12 21:16:08.503 ServerApp] Writing Jupyter server cookie secret to /home/cmadalini/.local/share/jupyter/runt
me/jupyter_cookie_secret
[I 2024-12-12 21:16:08.660 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-12-12 21:16:08.671 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-12-12 21:16:08.673 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-12-12 21:16:08.673 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-12-12 21:16:08.678 LabApp] JupyterLab extension loaded from /home/cmadalini/miniconda3/envs/bigdata/lib/python3
11/site-packages/jupyterlab
[I 2024-12-12 21:16:08.678 LabApp] JupyterLab application directory is /home/cmadalini/miniconda3/envs/bigdata/share/ju
yter/lab
[I 2024-12-12 21:16:08.678 LabApp] Extension Manager is 'pypi'.
[I 2024-12-12 21:16:08.737 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-12-12 21:16:08.737 ServerApp] Serving notebooks from local directory: /home/cmadalini
[I 2024-12-12 21:16:08.737 ServerApp] Jupyter Server 2.14.2 is running at:
[I 2024-12-12 21:16:08.737 ServerApp] http://localhost:8888/lab?token=236bd89975acea1858c5b5ee345aa1d24c3136cb59c29f6e
[I 2024-12-12 21:16:08.737 ServerApp]     http://127.0.0.1:8888/lab?token=236bd89975acea1858c5b5ee345aa1d24c3136cb59c29
6e
[I 2024-12-12 21:16:08.738 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confir
ation).
[W 2024-12-12 21:16:10.213 ServerApp] No web browser found: Error('could not locate runnable browser').
[C 2024-12-12 21:16:10.213 ServerApp]
```


File
Edit
View
Run
Kernel
Tabs
Settings
Help
Launcher
/
Name
Modified
miniconda3
20m ago
miniconda.exe
23m ago
Miniconda3-latest...
last mo.
Notebook
Python 3
(ipykernel)
Console
Python 3
(ipykernel)
Other

Añadimos miniconda ao Path

Instalamos ssh y configuramos la llave

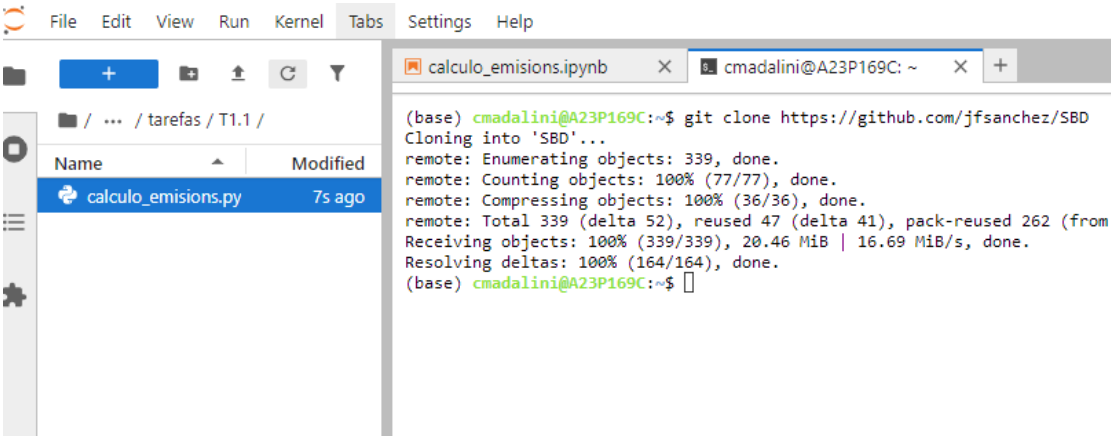


Instalacion de Git

```
(bigdata) cmadalini@A23P169C:~$ sudo apt-get install git
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  git-man libbrotli1 libcurl3-gnutls liberror-perl libgdbm-compat4 libgdbm6 libldap-2.5-0 libldap-common libnghttp2-14
  libperl5.36 librtmp1 libsasl2-2 libsasl2-modules libsasl2-modules-db libssh2-1 patch perl perl-modules-5.36
Suggested packages:
  gettext-base git-daemon-run | git-daemon-sysvinit git-doc git-email git-gui gitk gitweb git-cvs git-mediawiki
  git-svn gdbm-l10n libsasl2-modules-gssapi-mit | libsasl2-modules-gssapi-heimdal libsasl2-modules-ldap
  libsasl2-modules-otp libsasl2-modules-sql ed diffutils-doc perl-doc libterm-readline-gnu-perl
  | libterm-readline-perl-perl make libtap-harness-archive-perl
The following NEW packages will be installed:
  git git-man libbrotli1 libcurl3-gnutls liberror-perl libgdbm-compat4 libgdbm6 libldap-2.5-0 libldap-common
  libnghttp2-14 libperl5.36 librtmp1 libsasl2-2 libsasl2-modules libsasl2-modules-db libssh2-1 patch perl
  perl-modules-5.36
0 upgraded, 19 newly installed, 0 to remove and 0 not upgraded.
Need to get 18.2 MB of archives.
After this operation, 101 MB of additional disk space will be used.
Do you want to continue? [Y/n]
```

```
(bigdata) cmadalini@A23P169C:~$ git config --global user.name ""ConstantinMadalin""
(bigdata) cmadalini@A23P169C:~$ git config --global user.email ""190657180+ConstantinMadalin@users.noreply.github.com.""
(bigdata) cmadalini@A23P169C:~$
```

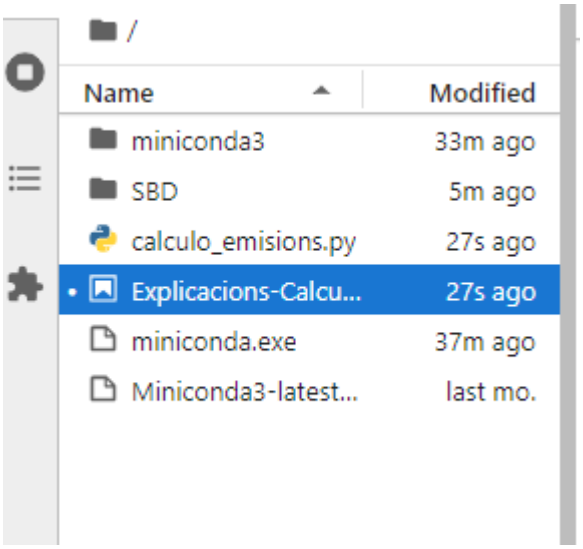
Clonamos el github en el jupyterlab

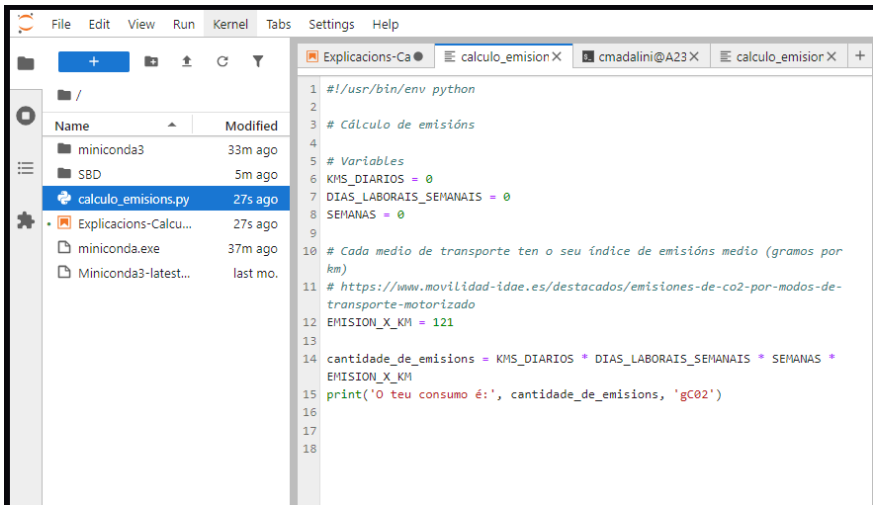


Git clone: creas una copia de un repositorio remoto en tu repositorio local para poder trabajar en él. Esto normalmente sucede solo una vez.

Git pull: pides una revisión y actualización de tu repositorio local con respecto al repositorio remoto, las actualizaciones se realizan solas a menos que existan conflictos.

Creamos un notebook en jupyter





Ejercicio 4

4.4) Modifica o script para realizar o cálculo de emisións anuais dunha persoa que viaxa en coche 100kms diarios, 5 días á semana durante 48 semanas do ano. Resultado? (captura de pantalla+texto)

```
##4.4-Modifica o script para realizar o cálculo de emisións anuais dunha persoa que viaxa en coche 100kms diarios, 5 días á semana durante 48 semanas do ano. Resultado? (captu

# Variables
KMS_DIARIOS = 100  # Distancia diaria en km
DIAS_LABORAIS_SEMANAIS = 5  # Días laborais por semana
SEMANAS = 48  # Número de semanas ao ano

# Emisións de CO2 por quilómetro (121 gCO2 por quilómetro)
EMISION_X_KM = 121  # gCO2 por quilómetro

# Cálculo da emisión de CO2 anual
cantidade_de_emisions = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_X_KM

# Mostrar o resultado
print('A emisión anual de CO2 é:', cantidade_de_emisions, 'gCO2')
```

```
A emisión anual de CO2 é: 2904000 gCO2
```

4.5) Modifica o script para realizar o cálculo de emisións anuais dunha persoa que viaxa en moto 20kms diarios, 3 días á semana, durante 40 semanas do ano. Resultado? (captura de pantalla+texto)

```
    ##4.5-Modifica o script para realizar o cálculo de emisións anuais dunha persoa que viaxa en moto 20kms diarios, 3 días á semana, durante 40 semanas do ano. Resultado? (cap

# Variables
KMS_DIARIOS = 20  # Distancia diaria en km
DIAS_LABORAIS_SEMANAIS = 3  # Días laborais por semana
SEMANAS = 40  # Número de semanas ao ano

# Emisións de CO2 por quilómetro para unha moto (60 gCO2 por quilómetro, estimado)
EMISION_X_KM = 60  # gCO2 por quilómetro

# Cálculo da emisión de CO2 anual
cantidade_de_emisions = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_X_KM

# Mostrar o resultado
print('A emisión anual de CO2 é:', cantidade_de_emisions, 'gCO2')
```

```
A emisión anual de CO2 é: 144000 gCO2
```

4.6) Calcula aproximadamente as "túas emisións" para desprazarte ao Wirtz durante todo o curso (aproxima a 24 semanas). Compara os resultados con diferentes medios de transporte. (captura de pantalla+texto)

```
##4.6-Calcula aproximadamente as "túas emisións" para desprazarte ao Wirtz durante todo o curso (aproxima a 24 semanas). Compara os resultados con diferentes medios de transpor


# Variables
KMS_DIARIOS = 5  # Distancia diaria en km (andando, coche ou moto)
DIAS_LABORAIS_SEMANAIS = 5  # Días laborais por semana
SEMANAS = 24  # Número de semanas durante o curso

# Emisións de CO2 por quilómetro para diferentes medios de transporte
EMISION_COCHE_X_KM = 121  # gCO2 por quilómetro para coche
EMISION_MOTO_X_KM = 60  # gCO2 por quilómetro para moto

# Cálculo das emisións anuais para cada medio de transporte

# Andar non emite CO2
emisiones_andando = 0

# Emisións para o coche
emisiones_coche = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_COCHE_X_KM

# Emisións para a moto
emisiones_moto = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_MOTO_X_KM

# Mostrar os resultados
print(f'Emisións anuais andando: {emisiones_andando} gCO2')
print(f'Emisións anuais en coche: {emisiones_coche} gCO2')
print(f'Emisións anuais en moto: {emisiones_moto} gCO2')
```

```
Emisións anuais andando: 0 gCO2
Emisións anuais en coche: 72600 gCO2
Emisións anuais en moto: 36000 gCO2
```

4.7) Modifica o código do notebook para que calcule o aforro en emisións para un caso definido ao mudar de medio de transporte. É dicir, dados kms, dias e semanas, indica a diferencia entre emisións usando dous medios diferentes.

```
##4.7-Modifica o código do notebook para que calcule o aforro en emisións para un caso definido ao mudar de medio de transporte. É dicir, dados kms, dias e semanas, indica a di


# Variables de entrada
KMS_DIARIOS = 5  # Distancia diaria en km
DIAS_LABORAIS_SEMANAIS = 5  # Días laborais por semana
SEMANAS = 24  # Número de semanas durante o curso

# Emisións de CO2 por quilómetro para diferentes medios de transporte
EMISION_COCHE_X_KM = 121  # gCO2 por quilómetro para coche
EMISION_MOTO_X_KM = 60  # gCO2 por quilómetro para moto

# Cálculo das emisións anuais para cada medio de transporte
emisiones_coche = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_COCHE_X_KM
emisiones_moto = KMS_DIARIOS * DIAS_LABORAIS_SEMANAIS * SEMANAS * EMISION_MOTO_X_KM

# Calcular o aforro de emisións ao cambiar de coche a moto
aforro_emisiones = emisiones_coche - emisiones_moto

# Mostrar os resultados
print(f'Emisións anuais en coche: {emisiones_coche} gCO2')
print(f'Emisións anuais en moto: {emisiones_moto} gCO2')
print(f'Aforro en emisións ao cambiar de coche a moto: {aforro_emisiones} gCO2')
```

```
Emisións anuais en coche: 72600 gCO2
Emisións anuais en moto: 36000 gCO2
Aforro en emisións ao cambiar de coche a moto: 36600 gCO2
```

4.9) Contesta: En canto os arquivos: *.ipynb e *.py do exercicio. Podes executar o código de ambos? Cal é a diferencia entre un arquivo e outro?

Solo puedo ejecutar el archivo .py porque es de python.

La diferencia que uno pertenece al python y el otro no

```
(bigdata) cmadalini@A23P169C:~$ python calculo_emisions.py
) teu consumo é: 0 gC02
(bigdata) cmadalini@A23P169C:~$ python Explicacions-Calculo-Emisions.ipynb
Traceback (most recent call last):
  File "/home/cmadalini/Explicacions-Calculo-Emisions.ipynb", line 211, in <module>
    "execution_count": null,
                       ^^^^
NameError: name 'null' is not defined
(bigdata) cmadalini@A23P169C:~$
```

5) Copia o notebook a un cartafol dun repositorio privado teu e súbeo (fai captura da consola).

```
(bigdata) cmadalini@A23P169C:~/Tarefa1$ cd ..
(bigdata) cmadalini@A23P169C:~$ sudo cp /home/cmadalini/Explicacions-Calculo-Emisions.ipynb Tarefa1/
[sudo] password for cmadalini:
Sorry, try again.
[sudo] password for cmadalini:
(bigdata) cmadalini@A23P169C:~$ ls
calculo_emisions.py                Miniconda3-latest-Windows-x86_64.exe  SBD
Explicacions-Calculo-Emisions.ipynb  miniconda.exe                        Tarea1
miniconda3                         PruebaMade                            Tarefa1
(bigdata) cmadalini@A23P169C:~$ cd Tarefa1/
(bigdata) cmadalini@A23P169C:~/Tarefa1$ ls
Explicacions-Calculo-Emisions.ipynb  README.md
(bigdata) cmadalini@A23P169C:~/Tarefa1$ git add Explicacions-Calculo-Emisions.ipynb
(bigdata) cmadalini@A23P169C:~/Tarefa1$ git commit -m "Tarefa1"
[main 29e546f] Tarefa1
 1 file changed, 239 insertions(+)
 create mode 100644 Explicacions-Calculo-Emisions.ipynb
(bigdata) cmadalini@A23P169C:~/Tarefa1$ git push origin main
Username for 'https://github.com': ConstantinMadalin
Password for 'https://ConstantinMadalin@github.com':
remote: Support for password authentication was removed on August 13, 2021.
remote: Please see https://docs.github.com/get-started/getting-started-with-git/about-remote-repositor
ies#cloning-with-https-urls for information on currently recommended modes of authentication.
fatal: Authentication failed for 'https://github.com/ConstantinMadalin/Tarefa1/'
(bigdata) cmadalini@A23P169C:~/Tarefa1$ git push origin main
Username for 'https://github.com': ConstantinMadalin
Password for 'https://ConstantinMadalin@github.com':
Enumerating objects: 4, done.
Counting objects: 100% (4/4), done.
Delta compression using up to 6 threads
Compressing objects: 100% (3/3), done.
Writing objects: 100% (3/3), 2.20 KiB | 2.20 MiB/s, done.
Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
To https://github.com/ConstantinMadalin/Tarefa1
   d7b7645..29e546f  main -> main
(bigdata) cmadalini@A23P169C:~/Tarefa1$ ls
Explicacions-Calculo-Emisions.ipynb  README.md
(bigdata) cmadalini@A23P169C:~/Tarefa1$
```