

# CPDSAI\_assignment3

August 24, 2024

```
[54]: # For converting into pdf at the end
!apt-get install -y pandoc
!pip install nbconvert
!apt-get install texlive-xetex texlive-fonts-recommended texlive-plain-generic
```

```
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3 pandoc-data
Suggested packages:
  texlive-luatex pandoc-citeproc context wkhtmltopdf librsvg2-bin groff ghc
nodejs php python
  libjs-mathjax libjs-katex citation-style-language-styles
The following NEW packages will be installed:
  libcmark-gfm-extensions0.29.0.gfm.3 libcmark-gfm0.29.0.gfm.3 pandoc pandoc-
data
0 upgraded, 4 newly installed, 0 to remove and 57 not upgraded.
Need to get 20.6 MB of archives.
After this operation, 156 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcmark-
gfm0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [115 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcmark-gfm-
extensions0.29.0.gfm.3 amd64 0.29.0.gfm.3-3 [25.1 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/universe amd64 pandoc-data all
2.9.2.1-3ubuntu2 [81.8 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 pandoc amd64
2.9.2.1-3ubuntu2 [20.3 MB]
Fetched 20.6 MB in 2s (9,302 kB/s)
Selecting previously unselected package libcmark-gfm0.29.0.gfm.3:amd64.
(Reading database ... 160890 files and directories currently installed.)
Preparing to unpack .../libcmark-gfm0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Selecting previously unselected package libcmark-gfm-
extensions0.29.0.gfm.3:amd64.
Preparing to unpack .../libcmark-gfm-
extensions0.29.0.gfm.3_0.29.0.gfm.3-3_amd64.deb ...
Unpacking libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
```

```

Selecting previously unselected package pandoc-data.
Preparing to unpack .../pandoc-data_2.9.2.1-3ubuntu2_all.deb ...
Unpacking pandoc-data (2.9.2.1-3ubuntu2) ...
Selecting previously unselected package pandoc.
Preparing to unpack .../pandoc_2.9.2.1-3ubuntu2_amd64.deb ...
Unpacking pandoc (2.9.2.1-3ubuntu2) ...
Setting up libcmark-gfm0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up libcmark-gfm-extensions0.29.0.gfm.3:amd64 (0.29.0.gfm.3-3) ...
Setting up pandoc-data (2.9.2.1-3ubuntu2) ...
Setting up pandoc (2.9.2.1-3ubuntu2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

Requirement already satisfied: nbconvert in /usr/local/lib/python3.10/dist-
packages (6.5.4)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages
(from nbconvert) (4.9.4)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (4.12.3)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages
(from nbconvert) (6.1.0)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (0.7.1)
Requirement already satisfied: entrypoints>=0.2.2 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.4)
Requirement already satisfied: jinja2>=3.0 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (3.1.4)
Requirement already satisfied: jupyter-core>=4.7 in

```

/usr/local/lib/python3.10/dist-packages (from nbconvert) (5.7.2)  
 Requirement already satisfied: jupyterlab-pygments in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.3.0)  
 Requirement already satisfied: MarkupSafe>=2.0 in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (2.1.5)  
 Requirement already satisfied: mistune<2,>=0.8.1 in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.8.4)  
 Requirement already satisfied: nbclient>=0.5.0 in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (0.10.0)  
 Requirement already satisfied: nbformat>=5.1 in /usr/local/lib/python3.10/dist-  
 packages (from nbconvert) (5.10.4)  
 Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-  
 packages (from nbconvert) (24.1)  
 Requirement already satisfied: pandocfilters>=1.4.1 in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (1.5.1)  
 Requirement already satisfied: pygments>=2.4.1 in  
 /usr/local/lib/python3.10/dist-packages (from nbconvert) (2.16.1)  
 Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-  
 packages (from nbconvert) (1.3.0)  
 Requirement already satisfied: traitlets>=5.0 in /usr/local/lib/python3.10/dist-  
 packages (from nbconvert) (5.7.1)  
 Requirement already satisfied: platformdirs>=2.5 in  
 /usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.7->nbconvert)  
 (4.2.2)  
 Requirement already satisfied: jupyter-client>=6.1.12 in  
 /usr/local/lib/python3.10/dist-packages (from nbclient>=0.5.0->nbconvert)  
 (6.1.12)  
 Requirement already satisfied: fastjsonschema>=2.15 in  
 /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert) (2.20.0)  
 Requirement already satisfied: jsonschema>=2.6 in  
 /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert) (4.23.0)  
 Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-  
 packages (from beautifulsoup4->nbconvert) (2.6)  
 Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.10/dist-  
 packages (from bleach->nbconvert) (1.16.0)  
 Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-  
 packages (from bleach->nbconvert) (0.5.1)  
 Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-  
 packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (24.2.0)  
 Requirement already satisfied: jsonschema-specifications>=2023.03.6 in  
 /usr/local/lib/python3.10/dist-packages (from  
 jsonschema>=2.6->nbformat>=5.1->nbconvert) (2023.12.1)  
 Requirement already satisfied: referencing>=0.28.4 in  
 /usr/local/lib/python3.10/dist-packages (from  
 jsonschema>=2.6->nbformat>=5.1->nbconvert) (0.35.1)  
 Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-  
 packages (from jsonschema>=2.6->nbformat>=5.1->nbconvert) (0.20.0)  
 Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-

```
packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (24.0.1)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert) (2.8.2)
Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (6.3.3)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
texlive-fonts-recommended is already the newest version (2021.20220204-1).
texlive-plain-generic is already the newest version (2021.20220204-1).
texlive-xetex is already the newest version (2021.20220204-1).
0 upgraded, 0 newly installed, 0 to remove and 57 not upgraded.
```

```
[ ]: !apt update
!apt install chromium-chromedriver
!pip install selenium
!apt-get install -y chromium-browser
```

```
Get:1 https://cloud.r-project.org/bin/linux/ubuntu jammy-cran40/ InRelease
[3,626 B]
Get:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
InRelease [1,581 B]
Get:3 http://security.ubuntu.com/ubuntu jammy-security InRelease [129 kB]
Ign:4 https://r2u.stat.illinois.edu/ubuntu jammy InRelease
Hit:5 http://archive.ubuntu.com/ubuntu jammy InRelease
Get:6 https://r2u.stat.illinois.edu/ubuntu jammy Release [5,713 B]
Get:7 https://r2u.stat.illinois.edu/ubuntu jammy Release.gpg [793 B]
Get:8 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [128 kB]
Hit:9 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu jammy InRelease
Get:10
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64
Packages [945 kB]
Get:11 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy
InRelease [24.3 kB]
Hit:12 https://ppa.launchpadcontent.net/ubuntugis/ppa/ubuntu jammy InRelease
Get:13 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [127 kB]
Get:14 https://r2u.stat.illinois.edu/ubuntu jammy/main amd64 Packages [2,560 kB]
Get:15 https://r2u.stat.illinois.edu/ubuntu jammy/main all Packages [8,232 kB]
Get:16 https://ppa.launchpadcontent.net/graphics-drivers/ppa/ubuntu jammy/main
amd64 Packages [53.3 kB]
Get:17 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 Packages [2,498
kB]
Get:18 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 Packages
[1,425 kB]
Fetched 16.1 MB in 3s (5,176 kB/s)
Reading package lists... Done
Building dependency tree... Done
```

```

Reading state information... Done
58 packages can be upgraded. Run 'apt list --upgradable' to see them.
W: Skipping acquire of configured file 'main/source/Sources' as
repository 'https://r2u.stat.illinois.edu/ubuntu jammy InRelease' does not seem
to provide it (sources.list entry misspelt?)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  apparmor chromium-browser libfuse3-3 liblzo2-2 libudev1 snapd squashfs-tools
  systemd-hwe-hwdb
  udev
Suggested packages:
  apparmor-profiles-extra apparmor-utils fuse3 zenity | kdialog
The following NEW packages will be installed:
  apparmor chromium-browser chromium-chromedriver libfuse3-3 liblzo2-2 snapd
  squashfs-tools
  systemd-hwe-hwdb udev
The following packages will be upgraded:
  libudev1
1 upgraded, 9 newly installed, 0 to remove and 57 not upgraded.
Need to get 28.5 MB of archives.
After this operation, 118 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 apparmor amd64
3.0.4-2ubuntu2.3 [595 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 liblzo2-2 amd64
2.10-2build3 [53.7 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 squashfs-tools amd64
1:4.5-3build1 [159 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libudev1 amd64
249.11-0ubuntu3.12 [78.2 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 udev amd64
249.11-0ubuntu3.12 [1,557 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfuse3-3 amd64
3.10.5-1build1 [81.2 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 snapd amd64
2.63+22.04ubuntu0.1 [25.9 MB]
Get:8 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 chromium-
browser amd64 1:85.0.4183.83-0ubuntu2.22.04.1 [49.2 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 chromium-
chromedriver amd64 1:85.0.4183.83-0ubuntu2.22.04.1 [2,308 B]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 systemd-hwe-
hwdb all 249.11.5 [3,228 B]
Fetched 28.5 MB in 2s (11.8 MB/s)
Preconfiguring packages ...
Selecting previously unselected package apparmor.
(Reading database ... 123595 files and directories currently installed.)
Preparing to unpack .../apparmor_3.0.4-2ubuntu2.3_amd64.deb ...

```

```

Unpacking apparmor (3.0.4-2ubuntu2.3) ...
Selecting previously unselected package liblzo2-2:amd64.
Preparing to unpack .../liblzo2-2_2.10-2build3_amd64.deb ...
Unpacking liblzo2-2:amd64 (2.10-2build3) ...
Selecting previously unselected package squashfs-tools.
Preparing to unpack .../squashfs-tools_1%3a4.5-3build1_amd64.deb ...
Unpacking squashfs-tools (1:4.5-3build1) ...
Preparing to unpack .../libudev1_249.11-0ubuntu3.12_amd64.deb ...
Unpacking libudev1:amd64 (249.11-0ubuntu3.12) over (249.11-0ubuntu3.10) ...
Setting up libudev1:amd64 (249.11-0ubuntu3.12) ...
Selecting previously unselected package udev.
(Reading database ... 123803 files and directories currently installed.)
Preparing to unpack .../udev_249.11-0ubuntu3.12_amd64.deb ...
Unpacking udev (249.11-0ubuntu3.12) ...
Selecting previously unselected package libfuse3-3:amd64.
Preparing to unpack .../libfuse3-3_3.10.5-1build1_amd64.deb ...
Unpacking libfuse3-3:amd64 (3.10.5-1build1) ...
Selecting previously unselected package snapd.
Preparing to unpack .../snapd_2.63+22.04ubuntu0.1_amd64.deb ...
Unpacking snapd (2.63+22.04ubuntu0.1) ...
Setting up apparmor (3.0.4-2ubuntu2.3) ...
Created symlink /etc/systemd/system/sysinit.target.wants/apparmor.service →
/lib/systemd/system/apparmor.service.
Setting up liblzo2-2:amd64 (2.10-2build3) ...
Setting up squashfs-tools (1:4.5-3build1) ...
Setting up udev (249.11-0ubuntu3.12) ...
invoke-rc.d: could not determine current runlevel
invoke-rc.d: policy-rc.d denied execution of start.
Setting up libfuse3-3:amd64 (3.10.5-1build1) ...
Setting up snapd (2.63+22.04ubuntu0.1) ...
Created symlink /etc/systemd/system/multi-
user.target.wants/snapd.apparmor.service →
/lib/systemd/system/snapd.apparmor.service.
Created symlink /etc/systemd/system/multi-
user.target.wants/snapd.autoimport.service →
/lib/systemd/system/snapd.autoimport.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.core-
fixup.service → /lib/systemd/system/snapd.core-fixup.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.recovery-
chooser-trigger.service → /lib/systemd/system/snapd.recovery-chooser-
trigger.service.
Created symlink /etc/systemd/system/multi-user.target.wants/snapd.seeded.service
→ /lib/systemd/system/snapd.seeded.service.
Created symlink /etc/systemd/system/cloud-
final.service.wants/snapd.seeded.service →
/lib/systemd/system/snapd.seeded.service.
Unit /lib/systemd/system/snapd.seeded.service is added as a dependency to a non-
existent unit cloud-final.service.

```

```

Created symlink /etc/systemd/system/multi-user.target.wants/snapd.service →
/lib/systemd/system/snapd.service.
Created symlink /etc/systemd/system/timers.target.wants/snapd.snap-repair.timer
→ /lib/systemd/system/snapd.snap-repair.timer.
Created symlink /etc/systemd/system/sockets.target.wants/snapd.socket →
/lib/systemd/system/snapd.socket.
Created symlink /etc/systemd/system/final.target.wants/snapd.system-
shutdown.service → /lib/systemd/system/snapd.system-shutdown.service.
Selecting previously unselected package chromium-browser.
(Reading database ... 124033 files and directories currently installed.)
Preparing to unpack .../chromium-
browser_1%3a85.0.4183.83-0ubuntu2.22.04.1_amd64.deb ...
=> Installing the chromium snap
==> Checking connectivity with the snap store
==> System doesn't have a working snapd, skipping
Unpacking chromium-browser (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Selecting previously unselected package chromium-chromedriver.
Preparing to unpack .../chromium-
chromedriver_1%3a85.0.4183.83-0ubuntu2.22.04.1_amd64.deb ...
Unpacking chromium-chromedriver (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Selecting previously unselected package systemd-hwe-hwdb.
Preparing to unpack .../systemd-hwe-hwdb_249.11.5_all.deb ...
Unpacking systemd-hwe-hwdb (249.11.5) ...
Setting up systemd-hwe-hwdb (249.11.5) ...
Setting up chromium-browser (1:85.0.4183.83-0ubuntu2.22.04.1) ...
update-alternatives: using /usr/bin/chromium-browser to provide /usr/bin/x-www-
browser (x-www-browser) in auto mode
update-alternatives: using /usr/bin/chromium-browser to provide /usr/bin/gnome-
www-browser (gnome-www-browser) in auto mode
Setting up chromium-chromedriver (1:85.0.4183.83-0ubuntu2.22.04.1) ...
Processing triggers for udev (249.11-0ubuntu3.12) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

```

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur\_adapter\_level\_zero.so.0 is not a symbolic link

Processing triggers for man-db (2.10.2-1) ...

Processing triggers for dbus (1.12.20-2ubuntu4.1) ...

Collecting selenium

Downloading selenium-4.23.1-py3-none-any.whl.metadata (7.1 kB)

Requirement already satisfied: urllib3<3,>=1.26 in

/usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (2.0.7)

Collecting trio~=0.17 (from selenium)

Downloading trio-0.26.2-py3-none-any.whl.metadata (8.6 kB)

Collecting trio-websocket~=0.9 (from selenium)

Downloading trio\_websocket-0.11.1-py3-none-any.whl.metadata (4.7 kB)

Requirement already satisfied: certifi>=2021.10.8 in

/usr/local/lib/python3.10/dist-packages (from selenium) (2024.7.4)

Requirement already satisfied: typing\_extensions~=4.9 in

/usr/local/lib/python3.10/dist-packages (from selenium) (4.12.2)

Requirement already satisfied: websocket-client~=1.8 in

/usr/local/lib/python3.10/dist-packages (from selenium) (1.8.0)

Requirement already satisfied: attrs>=23.2.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (24.2.0)

Requirement already satisfied: sortedcontainers in

/usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (2.4.0)

Requirement already satisfied: idna in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (3.7)

Collecting outcome (from trio~=0.17->selenium)

Downloading outcome-1.3.0.post0-py2.py3-none-any.whl.metadata (2.6 kB)

Requirement already satisfied: sniffio>=1.3.0 in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.3.1)

Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from trio~=0.17->selenium) (1.2.2)

Collecting wsproto>=0.14 (from trio-websocket~=0.9->selenium)

Downloading wsproto-1.2.0-py3-none-any.whl.metadata (5.6 kB)

Requirement already satisfied: pysocks!=1.5.7,<2.0,>=1.5.6 in

/usr/local/lib/python3.10/dist-packages (from urllib3[socks]<3,>=1.26->selenium) (1.7.1)

Collecting h11<1,>=0.9.0 (from wsproto>=0.14->trio-websocket~=0.9->selenium)

Downloading h11-0.14.0-py3-none-any.whl.metadata (8.2 kB)

Downloading selenium-4.23.1-py3-none-any.whl (9.4 MB)

9.4/9.4 MB

13.8 MB/s eta 0:00:00

Downloading trio-0.26.2-py3-none-any.whl (475 kB)

476.0/476.0 kB

20.0 MB/s eta 0:00:00



```

Downloading trio_websocket-0.11.1-py3-none-any.whl (17 kB)
Downloading wsproto-1.2.0-py3-none-any.whl (24 kB)
Downloading outcome-1.3.0.post0-py2.py3-none-any.whl (10 kB)
Downloading h11-0.14.0-py3-none-any.whl (58 kB)
58.3/58.3 kB
3.8 MB/s eta 0:00:00
Installing collected packages: outcome, h11, wsproto, trio, trio-
websocket, selenium
Successfully installed h11-0.14.0 outcome-1.3.0.post0 selenium-4.23.1
trio-0.26.2 trio-websocket-0.11.1 wsproto-1.2.0
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
chromium-browser is already the newest version
(1:85.0.4183.83-0ubuntu2.22.04.1).
chromium-browser set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 57 not upgraded.

```

## 1 Imports

```

[ ]: from selenium.webdriver.support.select import Select
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium import webdriver

import re
import pandas as pd
import time

```

## 2 Scrapping

```

[ ]: options = webdriver.ChromeOptions()
options.add_argument('--headless')
options.add_argument('--no-sandbox')
options.add_argument('--disable-dev-shm-usage')
options.add_argument('user-agent=Mozilla/5.0 (Windows NT 6.3; Win64; x64)␣
↳AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.162 Safari/537.36')

```

```

[ ]: wd = webdriver.Chrome(options=options)

wd.get("https://www.espn.in/nba/table")

table_filter1 = wd.find_element(By.XPATH, '//*[@main/div[2]/div[2]/div/div/section/
↳div/section/div[2]/div/section/div[1]/div')

```

```
table_filter2 = wd.find_element(By.XPATH, '///main/div[2]/div[2]/div/div/section/
↳div/section/div[2]/div/section/div[2]/div')
```

```
table_rows1 = table_filter1.find_elements(By.TAG_NAME, "tr")
```

```
table_rows2 = table_filter2.find_elements(By.TAG_NAME, "tr")
```

```
[ ]: team_names1 = [team.text.split('\n')[1] for team in table_rows1[1:16]]
column_names1 = [column for column in table_rows1[16].text.split(' ')[:8]]
per_team_data1 = [data.text.split('\n')[:8] for data in table_rows1[17:]]

team_names2 = [team.text.split('\n')[1] for team in table_rows2[1:16]]
column_names2 = [column for column in table_rows2[16].text.split(' ')[:8]]
per_team_data2 = [data.text.split('\n')[:8] for data in table_rows2[17:]]

df1 = pd.DataFrame(per_team_data1, columns=column_names1, index=team_names1)
df2 = pd.DataFrame(per_team_data2, columns=column_names2, index=team_names2)

scrapped_pandas = pd.concat([df1, df2])
scrapped_pandas
```

```
[ ]:
```

	W	L	PCT	GB	HOME	AWAY	DIV	CONF
Boston Celtics	64	18	.780	-	37-4	27-14	15-2	41-11
New York Knicks	50	32	.610	14	27-14	23-18	12-5	35-17
Milwaukee Bucks	49	33	.598	15	31-11	18-22	10-7	34-18
Cleveland Cavaliers	48	34	.585	16	26-15	22-19	11-5	31-21
Orlando Magic	47	35	.573	17	29-12	18-23	9-7	32-20
Indiana Pacers	47	35	.573	17	26-15	21-20	11-6	32-20
Philadelphia 76ers	47	35	.573	17	25-16	22-19	8-8	31-21
Miami Heat	46	36	.561	18	22-19	24-17	13-3	32-20
Chicago Bulls	39	43	.476	25	20-21	19-22	7-9	22-29
Atlanta Hawks	36	46	.439	28	21-20	15-26	8-8	22-30
Brooklyn Nets	32	50	.390	32	20-21	12-29	5-11	24-28
Toronto Raptors	25	57	.305	39	14-27	11-30	1-15	18-34
Charlotte Hornets	21	61	.256	43	11-30	10-31	6-10	14-38
Washington Wizards	15	67	.183	49	7-34	8-33	4-12	11-41
Detroit Pistons	14	68	.171	50	7-33	7-35	2-14	10-41
Oklahoma City Thunder	57	25	.695	-	33-8	24-17	12-4	36-16
Denver Nuggets	57	25	.695	-	33-8	24-17	10-6	33-19
Minnesota Timberwolves	56	26	.683	1	30-11	26-15	12-4	37-15
LA Clippers	51	31	.622	6	25-16	26-15	9-7	30-22
Dallas Mavericks	50	32	.610	7	25-16	25-16	11-5	31-21
Phoenix Suns	49	33	.598	8	25-16	24-17	9-9	29-23
Los Angeles Lakers	47	35	.573	10	28-14	19-21	7-10	27-25
New Orleans Pelicans	49	33	.598	8	21-19	28-14	9-7	30-22
Sacramento Kings	46	36	.561	11	24-17	22-19	10-7	30-22
Golden State Warriors	46	36	.561	11	21-20	25-16	7-9	26-26
Houston Rockets	41	41	.500	16	27-14	14-27	9-7	28-24

Utah Jazz	31	51	.378	26	21-20	10-31	5-11	16-36
Memphis Grizzlies	27	55	.329	30	9-32	18-23	8-8	14-37
San Antonio Spurs	22	60	.268	35	12-29	10-31	3-13	14-37
Portland Trail Blazers	21	61	.256	36	11-30	10-31	1-15	8-44

```
[ ]: scrapped_pandas.to_csv('ulugbek_st125457_assignment3.csv', index=False)
```

### 3 Aggregation

```
[ ]: # Find mean, median and mode wins for the combined table
```

```
scrapped_pandas.W = scrapped_pandas.W.astype(int)
print("MEAN: ", scrapped_pandas.W.mean())
print("MEDIAN: ", scrapped_pandas.W.median())
print("MODE: ", scrapped_pandas.W.mode().values[0])
```

```
MEAN: 41.0
MEDIAN: 46.5
MODE: 47
```

```
[ ]: # Segregate the home wins and home loss and away wins and away loss in separate
      ↪ columns
```

```
scrapped_pandas[["HOME_WINS", "HOME_LOSS"]] = scrapped_pandas.HOME.str.
      ↪split('-', expand=True).astype(int)
scrapped_pandas[["AWAY_WINS", "AWAY_LOSS"]] = scrapped_pandas.AWAY.str.
      ↪split('-', expand=True).astype(int)
scrapped_pandas.drop(["HOME", "AWAY"], axis=1, inplace=True)
scrapped_pandas
```

```
[ ]:
```

	W	L	PCT	GB	DIV	CONF	HOME_WINS	HOME_LOSS	\
Boston Celtics	64	18	.780	-	15-2	41-11	37	4	
New York Knicks	50	32	.610	14	12-5	35-17	27	14	
Milwaukee Bucks	49	33	.598	15	10-7	34-18	31	11	
Cleveland Cavaliers	48	34	.585	16	11-5	31-21	26	15	
Orlando Magic	47	35	.573	17	9-7	32-20	29	12	
Indiana Pacers	47	35	.573	17	11-6	32-20	26	15	
Philadelphia 76ers	47	35	.573	17	8-8	31-21	25	16	
Miami Heat	46	36	.561	18	13-3	32-20	22	19	
Chicago Bulls	39	43	.476	25	7-9	22-29	20	21	
Atlanta Hawks	36	46	.439	28	8-8	22-30	21	20	
Brooklyn Nets	32	50	.390	32	5-11	24-28	20	21	
Toronto Raptors	25	57	.305	39	1-15	18-34	14	27	
Charlotte Hornets	21	61	.256	43	6-10	14-38	11	30	
Washington Wizards	15	67	.183	49	4-12	11-41	7	34	
Detroit Pistons	14	68	.171	50	2-14	10-41	7	33	
Oklahoma City Thunder	57	25	.695	-	12-4	36-16	33	8	

Denver Nuggets	57	25	.695	-	10-6	33-19	33	8
Minnesota Timberwolves	56	26	.683	1	12-4	37-15	30	11
LA Clippers	51	31	.622	6	9-7	30-22	25	16
Dallas Mavericks	50	32	.610	7	11-5	31-21	25	16
Phoenix Suns	49	33	.598	8	9-9	29-23	25	16
Los Angeles Lakers	47	35	.573	10	7-10	27-25	28	14
New Orleans Pelicans	49	33	.598	8	9-7	30-22	21	19
Sacramento Kings	46	36	.561	11	10-7	30-22	24	17
Golden State Warriors	46	36	.561	11	7-9	26-26	21	20
Houston Rockets	41	41	.500	16	9-7	28-24	27	14
Utah Jazz	31	51	.378	26	5-11	16-36	21	20
Memphis Grizzlies	27	55	.329	30	8-8	14-37	9	32
San Antonio Spurs	22	60	.268	35	3-13	14-37	12	29
Portland Trail Blazers	21	61	.256	36	1-15	8-44	11	30

	AWAY_WINS	AWAY_LOSS
Boston Celtics	27	14
New York Knicks	23	18
Milwaukee Bucks	18	22
Cleveland Cavaliers	22	19
Orlando Magic	18	23
Indiana Pacers	21	20
Philadelphia 76ers	22	19
Miami Heat	24	17
Chicago Bulls	19	22
Atlanta Hawks	15	26
Brooklyn Nets	12	29
Toronto Raptors	11	30
Charlotte Hornets	10	31
Washington Wizards	8	33
Detroit Pistons	7	35
Oklahoma City Thunder	24	17
Denver Nuggets	24	17
Minnesota Timberwolves	26	15
LA Clippers	26	15
Dallas Mavericks	25	16
Phoenix Suns	24	17
Los Angeles Lakers	19	21
New Orleans Pelicans	28	14
Sacramento Kings	22	19
Golden State Warriors	25	16
Houston Rockets	14	27
Utah Jazz	10	31
Memphis Grizzlies	18	23
San Antonio Spurs	10	31
Portland Trail Blazers	10	31

```
[ ]: # Sort the combined table based on home wins and home loss
scrapped_pandas.sort_values(by=["HOME_WINS", "HOME_LOSS"], ascending=False,
    ↪inplace=True)
scrapped_pandas
```

```
[ ]:
```

	W	L	PCT	GB	DIV	CONF	HOME_WINS	HOME_LOSS	\
Boston Celtics	64	18	.780	-	15-2	41-11	37	4	
Oklahoma City Thunder	57	25	.695	-	12-4	36-16	33	8	
Denver Nuggets	57	25	.695	-	10-6	33-19	33	8	
Milwaukee Bucks	49	33	.598	15	10-7	34-18	31	11	
Minnesota Timberwolves	56	26	.683	1	12-4	37-15	30	11	
Orlando Magic	47	35	.573	17	9-7	32-20	29	12	
Los Angeles Lakers	47	35	.573	10	7-10	27-25	28	14	
New York Knicks	50	32	.610	14	12-5	35-17	27	14	
Houston Rockets	41	41	.500	16	9-7	28-24	27	14	
Cleveland Cavaliers	48	34	.585	16	11-5	31-21	26	15	
Indiana Pacers	47	35	.573	17	11-6	32-20	26	15	
Philadelphia 76ers	47	35	.573	17	8-8	31-21	25	16	
LA Clippers	51	31	.622	6	9-7	30-22	25	16	
Dallas Mavericks	50	32	.610	7	11-5	31-21	25	16	
Phoenix Suns	49	33	.598	8	9-9	29-23	25	16	
Sacramento Kings	46	36	.561	11	10-7	30-22	24	17	
Miami Heat	46	36	.561	18	13-3	32-20	22	19	
Atlanta Hawks	36	46	.439	28	8-8	22-30	21	20	
Golden State Warriors	46	36	.561	11	7-9	26-26	21	20	
Utah Jazz	31	51	.378	26	5-11	16-36	21	20	
New Orleans Pelicans	49	33	.598	8	9-7	30-22	21	19	
Chicago Bulls	39	43	.476	25	7-9	22-29	20	21	
Brooklyn Nets	32	50	.390	32	5-11	24-28	20	21	
Toronto Raptors	25	57	.305	39	1-15	18-34	14	27	
San Antonio Spurs	22	60	.268	35	3-13	14-37	12	29	
Charlotte Hornets	21	61	.256	43	6-10	14-38	11	30	
Portland Trail Blazers	21	61	.256	36	1-15	8-44	11	30	
Memphis Grizzlies	27	55	.329	30	8-8	14-37	9	32	
Washington Wizards	15	67	.183	49	4-12	11-41	7	34	
Detroit Pistons	14	68	.171	50	2-14	10-41	7	33	

	AWAY_WINS	AWAY_LOSS
Boston Celtics	27	14
Oklahoma City Thunder	24	17
Denver Nuggets	24	17
Milwaukee Bucks	18	22
Minnesota Timberwolves	26	15
Orlando Magic	18	23
Los Angeles Lakers	19	21
New York Knicks	23	18
Houston Rockets	14	27

Cleveland Cavaliers	22	19
Indiana Pacers	21	20
Philadelphia 76ers	22	19
LA Clippers	26	15
Dallas Mavericks	25	16
Phoenix Suns	24	17
Sacramento Kings	22	19
Miami Heat	24	17
Atlanta Hawks	15	26
Golden State Warriors	25	16
Utah Jazz	10	31
New Orleans Pelicans	28	14
Chicago Bulls	19	22
Brooklyn Nets	12	29
Toronto Raptors	11	30
San Antonio Spurs	10	31
Charlotte Hornets	10	31
Portland Trail Blazers	10	31
Memphis Grizzlies	18	23
Washington Wizards	8	33
Detroit Pistons	7	35

```
[ ]: scrapped_pandas.to_csv('modified.csv', index=False)
```

```
[45]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[53]: from google.colab import files

# !jupyter nbconvert --to pdf './drive/MyDrive/Colab Notebooks/
↳ CPDSAI_assignment3.ipynb'
!pwd
!ls
!ls drive/MyDrive/Colab\ Notebooks
!jupyter nbconvert --to pdf './drive/MyDrive/Colab Notebooks/CPDSAI_assignment3.
↳ ipynb'
# files.download('./drive/MyDrive/Colab Notebooks/CPDSAI_assignment3.pdf')
```

/content

```
drive modified.csv sample_data ulugbek_st125457_assignment3.csv
01_Intro_Pandas.ipynb
02_Data_Ingestion.ipynb
'Another copy of 03_Basic_BS4.ipynb'
'Another copy of 04_Reading_Data_from_API.ipynb'
'Another copy of selenium_ptt_oilprice_classwork.ipynb'
```

```

carPrice_3rd_attempt.ipynb
carPrice.ipynb
'Copy of 01 - Life_Expectancy.ipynb'
'Copy of ML_Assignment1_carPrice'
CPDSAI_assignment3.ipynb
DLCV_lecture3.ipynb
dsai_hw2.ipynb
Lecture3.ipynb
miniquiz_classifier.ipynb
ML_Assignment1_carPrice
numpy_intro_activity.ipynb
'Satellite Image Analysis using numpy.ipynb'
[NbConvertApp] Converting notebook ./drive/MyDrive/Colab
Notebooks/CPDSAI_assignment3.ipynb to pdf
[NbConvertApp] ERROR | Error while converting './drive/MyDrive/Colab
Notebooks/CPDSAI_assignment3.ipynb'
Traceback (most recent call last):
  File "/usr/local/lib/python3.10/dist-packages/nbconvert/nbconvertapp.py", line
488, in export_single_notebook
    output, resources = self.exporter.from_filename(
  File "/usr/local/lib/python3.10/dist-
packages/nbconvert/exporters/exporter.py", line 189, in from_filename
    return self.from_file(f, resources=resources, **kw)
  File "/usr/local/lib/python3.10/dist-
packages/nbconvert/exporters/exporter.py", line 206, in from_file
    return self.from_notebook_node(
  File "/usr/local/lib/python3.10/dist-packages/nbconvert/exporters/pdf.py",
line 181, in from_notebook_node
    latex, resources = super().from_notebook_node(nb, resources=resources, **kw)
  File "/usr/local/lib/python3.10/dist-packages/nbconvert/exporters/latex.py",
line 74, in from_notebook_node
    return super().from_notebook_node(nb, resources, **kw)
  File "/usr/local/lib/python3.10/dist-
packages/nbconvert/exporters/templateexporter.py", line 413, in
from_notebook_node
    output = self.template.render(nb=nb_copy, resources=resources)
  File "/usr/local/lib/python3.10/dist-packages/jinja2/environment.py", line
1304, in render
    self.environment.handle_exception()
  File "/usr/local/lib/python3.10/dist-packages/jinja2/environment.py", line
939, in handle_exception
    raise rewrite_traceback_stack(source=source)
  File "/usr/local/share/jupyter/nbconvert/templates/latex/index.tex.j2", line
8, in top-level template code
    ((* extends cell_style *))
  File
"/usr/local/share/jupyter/nbconvert/templates/latex/style_jupyter.tex.j2", line
176, in top-level template code

```

```

\prompt{(((prompt)))}{(((prompt_color)))}{(((execution_count)))}{(((extra_sp
ace)))}
File "/usr/local/share/jupyter/nbconvert/templates/latex/base.tex.j2", line 7,
in top-level template code
    ((* extends 'document_contents.tex.j2' *))
File
"/usr/local/share/jupyter/nbconvert/templates/latex/document_contents.tex.j2",
line 51, in top-level template code
    ((* block figure scoped *))
File "/usr/local/share/jupyter/nbconvert/templates/latex/display_priority.j2",
line 5, in top-level template code
    ((* extends 'null.j2' *))
File "/usr/local/share/jupyter/nbconvert/templates/latex/null.j2", line 30, in
top-level template code
    ((* block body *))
File "/usr/local/share/jupyter/nbconvert/templates/latex/base.tex.j2", line
215, in block 'body'
    ((( super() )))
File "/usr/local/share/jupyter/nbconvert/templates/latex/null.j2", line 32, in
block 'body'
    ((* block any_cell scoped *))
File "/usr/local/share/jupyter/nbconvert/templates/latex/null.j2", line 85, in
block 'any_cell'
    ((* block markdowncell scoped-*)) ((* endblock markdowncell -*))
File
"/usr/local/share/jupyter/nbconvert/templates/latex/document_contents.tex.j2",
line 68, in block 'markdowncell'
    ((( cell.source | citation2latex | strip_files_prefix |
convert_pandoc('markdown+tex_math_double_backslash', 'json',extra_args=[]) |
resolve_references | convert_pandoc('json','latex'))))
File "/usr/local/lib/python3.10/dist-packages/nbconvert/filters/pandoc.py",
line 24, in convert_pandoc
    return pandoc(source, from_format, to_format, extra_args=extra_args)
File "/usr/local/lib/python3.10/dist-packages/nbconvert/utils/pandoc.py", line
51, in pandoc
    check_pandoc_version()
File "/usr/local/lib/python3.10/dist-packages/nbconvert/utils/pandoc.py", line
99, in check_pandoc_version
    v = get_pandoc_version()
File "/usr/local/lib/python3.10/dist-packages/nbconvert/utils/pandoc.py", line
76, in get_pandoc_version
    raise PandocMissing()
nbconvert.utils.pandoc.PandocMissing: Pandoc wasn't found.
Please check that pandoc is installed:
https://pandoc.org/installing.html

```