

In [1]: `%conda install pandas`

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
    current version: 4.10.3
    latest version: 22.9.0
```

```
Please update conda by running
```

```
$ conda update -n base conda
```

```
## Package Plan ##
```

```
environment location: /home/studio-lab-user/.conda/envs/default
```

```
added / updated specs:
- pandas
```

```
The following packages will be downloaded:
```

package	build		
libblas-3.9.0	16_llinux64_openblas	13 KB	conda-forge
libcblas-3.9.0	16_llinux64_openblas	13 KB	conda-forge
libgfortran-ng-12.2.0	h69a702a_19	22 KB	conda-forge
libgfortran5-12.2.0	h337968e_19	1.8 MB	conda-forge
liblapack-3.9.0	16_llinux64_openblas	13 KB	conda-forge
libopenblas-0.3.21	pthreads_h78a6416_3	10.1 MB	conda-forge
numpy-1.23.5	py39h3d75532_0	5.6 MB	conda-forge
pandas-1.5.1	py39h4661b88_1	13.0 MB	conda-forge
pytz-2022.6	pyhd8ed1ab_0	235 KB	conda-forge
-----		Total:	30.7 MB

```
The following NEW packages will be INSTALLED:
```

libblas	conda-forge/linux-64::libblas-3.9.0-16_llinux64_openblas
libcblas	conda-forge/linux-64::libcblas-3.9.0-16_llinux64_openblas
libgfortran-ng	conda-forge/linux-64::libgfortran-ng-12.2.0-h69a702a_19
libgfortran5	conda-forge/linux-64::libgfortran5-12.2.0-h337968e_19
liblapack	conda-forge/linux-64::liblapack-3.9.0-16_llinux64_openblas
libopenblas	conda-forge/linux-64::libopenblas-0.3.21-pthreads_h78a6416_3
numpy	conda-forge/linux-64::numpy-1.23.5-py39h3d75532_0
pandas	conda-forge/linux-64::pandas-1.5.1-py39h4661b88_1
pytz	conda-forge/noarch::pytz-2022.6-pyhd8ed1ab_0

```
Downloading and Extracting Packages
```

pandas-1.5.1	13.0 MB	#####	100%
pytz-2022.6	235 KB	#####	100%
libblas-3.9.0	13 KB	#####	100%

```
libopenblas-0.3.21 | 10.1 MB | #####| 100%
liblapack-3.9.0 | 13 KB | #####| 100%
numpy-1.23.5 | 5.6 MB | #####| 100%
libcblas-3.9.0 | 13 KB | #####| 100%
libgfortran5-12.2.0 | 1.8 MB | #####| 100%
libgfortran-ng-12.2. | 22 KB | #####| 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

Note: you may need to restart the kernel to use updated packages.

```
In [2]: conda update -n base conda
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
## Package Plan ##
```

```
environment location: /opt/conda
```

```
added / updated specs:
```

```
- conda
```

```
The following packages will be downloaded:
```

package	build		
_openmp_mutex-4.5	2_gnu	23 KB	conda-forge
brotlipy-0.7.0	py39hb9d737c_1005	342 KB	conda-forge
bzip2-1.0.8	h7f98852_4	484 KB	conda-forge
ca-certificates-2022.9.24	ha878542_0	150 KB	conda-forge
certifi-2022.9.24	pyhd8ed1ab_0	155 KB	conda-forge
cffi-1.15.1	py39he91dace_2	230 KB	conda-forge
charset-normalizer-2.1.1	pyhd8ed1ab_0	36 KB	conda-forge
colorama-0.4.6	pyhd8ed1ab_0	25 KB	conda-forge
conda-22.9.0	py39hf3d152e_2	963 KB	conda-forge
conda-package-handling-1.9.0	py39hb9d737c_1	1015 KB	conda-forge
cryptography-38.0.3	py39h3ccb8fc_0	1.5 MB	conda-forge
idna-3.4	pyhd8ed1ab_0	55 KB	conda-forge
ld_impl_linux-64-2.39	hcc3a1bd_1	675 KB	conda-forge
libffi-3.4.2	h7f98852_5	57 KB	conda-forge
libgcc-ng-12.2.0	h65d4601_19	931 KB	conda-forge
libgomp-12.2.0	h65d4601_19	455 KB	conda-forge
libnsl-2.0.0	h7f98852_0	31 KB	conda-forge
libssqlite-3.40.0	h753d276_0	791 KB	conda-forge
libuuid-2.32.1	h7f98852_1000	28 KB	conda-forge
libzlib-1.2.13	h166bdaf_4	64 KB	conda-forge
ncurses-6.3	h27087fc_1	1002 KB	conda-forge
openssl-3.0.7	h166bdaf_0	2.8 MB	conda-forge
pycosat-0.6.4	py39hb9d737c_1	108 KB	conda-forge
pycparser-2.21	pyhd8ed1ab_0	100 KB	conda-forge
pyopenssl-22.1.0	pyhd8ed1ab_0	122 KB	conda-forge
pysocks-1.7.1	pyha2e5f31_6	19 KB	conda-forge
python-3.9.14	hba424b6_0_cpython	20.9 MB	conda-forge
readline-8.1.2	h0f457ee_0	291 KB	conda-forge
requests-2.28.1	pyhd8ed1ab_1	53 KB	conda-forge
ruamel_yaml-0.15.80	py39hb9d737c_1008	255 KB	conda-forge
setuptools-65.5.1	pyhd8ed1ab_0	731 KB	conda-forge
tk-8.6.12	h27826a3_0	3.3 MB	conda-forge
toolz-0.12.0	pyhd8ed1ab_0	48 KB	conda-forge
tqdm-4.64.1	pyhd8ed1ab_0	82 KB	conda-forge
tzdata-2022f	h191b570_0	118 KB	conda-forge
urllib3-1.26.11	pyhd8ed1ab_0	102 KB	conda-forge
wheel-0.38.4	pyhd8ed1ab_0	32 KB	conda-forge
xz-5.2.6	h166bdaf_0	409 KB	conda-forge
yaml-0.2.5	h7f98852_2	87 KB	conda-forge

```
Total: 38.4 MB
```

The following NEW packages will be INSTALLED:

bzip2	conda-forge/linux-64::bzip2-1.0.8-h7f98852_4
libnsl	conda-forge/linux-64::libnsl-2.0.0-h7f98852_0
libssqlite	conda-forge/linux-64::libssqlite-3.40.0-h753d276_0
libuuid	conda-forge/linux-64::libuuid-2.32.1-h7f98852_1000
libzlib	conda-forge/linux-64::libzlib-1.2.13-h166bdaf_4
toolz	conda-forge/noarch::toolz-0.12.0-pyhd8ed1ab_0

The following packages will be REMOVED:

chardet-4.0.0-py39hf3d152e_1
libstdcxx-ng-11.2.0-he4da1e4_9
six-1.16.0-pyh6c4a22f_0
sqlite-3.36.0-h9cd32fc_2
zlib-1.2.11-h516909a_1010

The following packages will be UPDATED:

_openmp_mutex	4.5-1_gnu --> 4.5-2_gnu
brotlipy	0.7.0-py39h3811e60_1001 --> 0.7.0-py39hb9d737c
_1005	
ca-certificates	2021.5.30-ha878542_0 --> 2022.9.24-ha878542
_0	
certifi	conda-forge/linux-64::certifi-2021.5.~ --> conda-forge/noarc
h::certifi-2022.9.24-pyhd8ed1ab_0	
cffi	1.14.6-py39h4bc2ebd_1 --> 1.15.1-py39he91dac
e_2	
charset-normalizer	2.0.0-pyhd8ed1ab_0 --> 2.1.1-pyhd8ed1ab_0
colorama	0.4.4-pyh9f0ad1d_0 --> 0.4.6-pyhd8ed1ab_0
conda	4.10.3-py39hf3d152e_2 --> 22.9.0-py39hf3d152
e_2	
conda-package-handling	1.7.3-py39h3811e60_0 --> 1.9.0-py39hb9d737c
_1	
cryptography	3.4.7-py39hbca0aa6_0 --> 38.0.3-py39h3ccb8f
c_0	
idna	3.1-pyhd3deb0d_0 --> 3.4-pyhd8ed1ab_0
ld_impl_linux-64	2.36.1-hea4e1c9_2 --> 2.39-hcc3a1bd_1
libffi	3.4.2-h9c3ff4c_4 --> 3.4.2-h7f98852_5
libgcc-ng	11.2.0-h1d223b6_9 --> 12.2.0-h65d4601_19
libgomp	11.2.0-h1d223b6_9 --> 12.2.0-h65d4601_19
ncurses	6.2-h58526e2_4 --> 6.3-h27087fc_1
openssl	1.1.11-h7f98852_0 --> 3.0.7-h166bdaf_0
pycosat	0.6.3-py39h3811e60_1006 --> 0.6.4-py39hb9d737c
_1	
pycparser	2.20-pyh9f0ad1d_2 --> 2.21-pyhd8ed1ab_0
pyopenssl	20.0.1-pyhd8ed1ab_0 --> 22.1.0-pyhd8ed1ab_0
0	
pysocks	conda-forge/linux-64::pysocks-1.7.1-p~ --> conda-forge/noarc
h::pysocks-1.7.1-pyha2e5f31_6	
python	3.9.7-hb7a2778_2_cpython --> 3.9.14-hba424b6_0
cpython	
readline	8.1-h46c0cb4_0 --> 8.1.2-h0f457ee_0
requests	2.26.0-pyhd8ed1ab_0 --> 2.28.1-pyhd8ed1ab_1
1	

```

ruamel_yaml           0.15.80-py39h3811e60_1004 --> 0.15.80-py39hb9d73
7c_1008
    setuptools      conda-forge/linux-64::setuptools-58.0~ --> conda-forge/noarc
h::setuptools-65.5.1-pyhd8ed1ab_0
        tk                         8.6.11-h27826a3_1 --> 8.6.12-h27826a3_0
        tqdm                        4.62.3-pyhd8ed1ab_0 --> 4.64.1-pyhd8ed1ab_
        0
        tzdata                      2021a-he74cb21_1 --> 2022f-h191b570_0
        urllib3                     1.26.7-pyhd8ed1ab_0 --> 1.26.11-pyhd8ed1ab_
        _0
        wheel                      0.37.0-pyhd8ed1ab_1 --> 0.38.4-pyhd8ed1ab_
        0
        xz                          5.2.5-h516909a_1 --> 5.2.6-h166bdaf_0
        yaml                        0.2.5-h516909a_0 --> 0.2.5-h7f98852_2

```

Downloading and Extracting Packages

libgcc-ng-12.2.0	931 KB	#####	100%
ld_impl_linux-64-2.3	675 KB	#####	100%
cffi-1.15.1	230 KB	#####	100%
bzip2-1.0.8	484 KB	#####	100%
pycosat-0.6.4	108 KB	#####	100%
wheel-0.38.4	32 KB	#####	100%
tk-8.6.12	3.3 MB	#####	100%
libssqlite-3.40.0	791 KB	#####	100%
charset-normalizer-2	36 KB	#####	100%
toolz-0.12.0	48 KB	#####	100%
tqdm-4.64.1	82 KB	#####	100%
urllib3-1.26.11	102 KB	#####	100%
ca-certificates-2022	150 KB	#####	100%
readline-8.1.2	291 KB	#####	100%
pysocks-1.7.1	19 KB	#####	100%
libzlib-1.2.13	64 KB	#####	100%
_openmp_mutex-4.5	23 KB	#####	100%
conda-22.9.0	963 KB	#####	100%
python-3.9.14	20.9 MB	#####	100%
brotlipy-0.7.0	342 KB	#####	100%
ruamel_yaml-0.15.80	255 KB	#####	100%
tzdata-2022f	118 KB	#####	100%
yaml-0.2.5	87 KB	#####	100%
cryptography-38.0.3	1.5 MB	#####	100%
requests-2.28.1	53 KB	#####	100%
idna-3.4	55 KB	#####	100%
libffi-3.4.2	57 KB	#####	100%
certifi-2022.9.24	155 KB	#####	100%
pycparser-2.21	100 KB	#####	100%
pyopenssl-22.1.0	122 KB	#####	100%
openssl-3.0.7	2.8 MB	#####	100%
conda-package-handli	1015 KB	#####	100%
libnsl-2.0.0	31 KB	#####	100%
colorama-0.4.6	25 KB	#####	100%
libgomp-12.2.0	455 KB	#####	100%
libuuid-2.32.1	28 KB	#####	100%
setuptools-65.5.1	731 KB	#####	100%
xz-5.2.6	409 KB	#####	100%

```
n curses-6.3 | 1002 KB | ##### | 100%
Preparing transaction: done
Verifying transaction: failed

EnvironmentNotWritableError: The current user does not have write permissions to t
he target environment.
    environment location: /opt/conda
    uid: 1000
    gid: 100
```

Note: you may need to restart the kernel to use updated packages.

In [3]: `%conda install numpy`

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
--> WARNING: A newer version of conda exists. <=-
    current version: 4.10.3
    latest version: 22.9.0
```

Please update conda by running

```
$ conda update -n base conda
```

```
# All requested packages already installed.
```

Note: you may need to restart the kernel to use updated packages.

In [4]: `%conda install matplotlib`

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
    current version: 4.10.3
    latest version: 22.9.0
```

```
Please update conda by running
```

```
$ conda update -n base conda
```

```
## Package Plan ##
```

```
environment location: /home/studio-lab-user/.conda/envs/default
```

```
added / updated specs:
- matplotlib
```

```
The following packages will be downloaded:
```

package	build	size	source
alsa-lib-1.2.8	h166bdaf_0	578 KB	conda-forge
attr-2.5.1	h166bdaf_1	69 KB	conda-forge
brotli-1.0.9	h166bdaf_8	18 KB	conda-forge
brotli-bin-1.0.9	h166bdaf_8	20 KB	conda-forge
contourpy-1.0.6	py39hf939315_0	234 KB	conda-forge
cycler-0.11.0	pyhd8ed1ab_0	10 KB	conda-forge
dbus-1.13.6	h5008d03_3	604 KB	conda-forge
expat-2.5.0	h27087fc_0	189 KB	conda-forge
fftw-3.3.10	nompi_hf0379b8_105	2.2 MB	conda-forge
font-ttf-dejavu-sans-mono-2.37	hab24e00_0	388 KB	conda-forge
font-ttf-inconsolata-3.000	h77eed37_0	94 KB	conda-forge
font-ttf-source-code-pro-2.038	h77eed37_0	684 KB	conda-forge
font-ttf-ubuntu-0.83	hab24e00_0	1.9 MB	conda-forge
fontconfig-2.14.1	hc2a2eb6_0	365 KB	conda-forge
fonts-conda-ecosystem-1	0	4 KB	conda-forge
fonts-conda-forge-1	0	4 KB	conda-forge
fonttools-4.38.0	py39hb9d737c_1	1.9 MB	conda-forge
freetype-2.12.1	hca18f0e_0	884 KB	conda-forge
gettext-0.21.1	h27087fc_0	4.1 MB	conda-forge
glib-2.74.1	h6239696_1	474 KB	conda-forge
glib-tools-2.74.1	h6239696_1	107 KB	conda-forge
gst-plugins-base-1.21.2	h3e40eee_0	2.5 MB	conda-forge
gstreamer-1.21.2	hd4edc92_0	1.9 MB	conda-forge
gstreamer-orc-0.4.33	h166bdaf_0	299 KB	conda-forge
icu-70.1	h27087fc_0	13.5 MB	conda-forge
jack-1.9.21	he978b8e_1	530 KB	conda-forge
jpeg-9e	h166bdaf_2	269 KB	conda-forge
keyutils-1.6.1	h166bdaf_0	115 KB	conda-forge
kiwisolver-1.4.4	py39hf939315_1	76 KB	conda-forge
krb5-1.19.3	h3790be6_0	1.4 MB	conda-forge

lame-3.100	h166bdaf_1003	496 KB	conda-forge
lcms2-2.14	h6ed2654_0	256 KB	conda-forge
lerc-4.0.0	h27087fc_0	275 KB	conda-forge
libbrotlicommon-1.0.9	h166bdaf_8	66 KB	conda-forge
libbrotlidec-1.0.9	h166bdaf_8	33 KB	conda-forge
libbrotlienc-1.0.9	h166bdaf_8	288 KB	conda-forge
libcap-2.66	ha37c62d_0	97 KB	conda-forge
libclang-15.0.5	default_h2e3cab8_0	127 KB	conda-forge
libclang13-15.0.5	default_h3a83d3e_0	11.0 MB	conda-forge
libcups-2.3.3	h3e49a29_2	4.5 MB	conda-forge
libdb-6.2.32	h9c3ff4c_0	23.3 MB	conda-forge
libdeflate-1.14	h166bdaf_0	81 KB	conda-forge
libedit-3.1.20191231	he28a2e2_2	121 KB	conda-forge
libevent-2.1.10	h9b69904_4	1.1 MB	conda-forge
libflac-1.4.2	h27087fc_0	411 KB	conda-forge
libgcrypt-1.10.1	h166bdaf_0	703 KB	conda-forge
libglib-2.74.1	h606061b_1	3.1 MB	conda-forge
libgpg-error-1.45	hc0c96e0_0	286 KB	conda-forge
libiconv-1.17	h166bdaf_0	1.4 MB	conda-forge
libllvm15-15.0.5	h63197d8_0	37.4 MB	conda-forge
libogg-1.3.4	h7f98852_1	206 KB	conda-forge
libopus-1.3.1	h7f98852_1	255 KB	conda-forge
libpng-1.6.38	h753d276_0	371 KB	conda-forge
libpq-14.5	hd77ab85_1	3.0 MB	conda-forge
libsndfile-1.1.0	h27087fc_0	370 KB	conda-forge
libsystemd0-252	h2a991cd_0	384 KB	conda-forge
libtiff-4.4.0	h55922b4_4	651 KB	conda-forge
libtool-2.4.6	h9c3ff4c_1008	511 KB	conda-forge
libudev1-252	h166bdaf_0	118 KB	conda-forge
libvorbis-1.3.7	h9c3ff4c_0	280 KB	conda-forge
libwebp-base-1.2.4	h166bdaf_0	404 KB	conda-forge
libxcb-1.13	h7f98852_1004	391 KB	conda-forge
libxkbcommon-1.0.3	he3ba5ed_0	581 KB	conda-forge
libxml2-2.10.3	h7463322_0	754 KB	conda-forge
lz4-c-1.9.3	h9c3ff4c_1	179 KB	conda-forge
matplotlib-3.6.2	py39hf3d152e_0	7 KB	conda-forge
matplotlib-base-3.6.2	py39hf9fd14e_0	7.5 MB	conda-forge
mpg123-1.30.2	h27087fc_1	512 KB	conda-forge
munkres-1.1.4	pyh9f0ad1d_0	12 KB	conda-forge
mysql-common-8.0.31	haf5c9bc_0	1.9 MB	conda-forge
mysql-libs-8.0.31	h28c427c_0	1.9 MB	conda-forge
nspr-4.32	h9c3ff4c_1	233 KB	conda-forge
nss-3.78	h2350873_0	2.1 MB	conda-forge
openjpeg-2.5.0	h7d73246_1	533 KB	conda-forge
openssl-1.1.1s	h166bdaf_0	2.1 MB	conda-forge
pcre2-10.40	hc3806b6_0	2.3 MB	conda-forge
pillow-9.2.0	py39hf3a2cdf_3	45.3 MB	conda-forge
ply-3.11	py_1	44 KB	conda-forge
pthread-stubs-0.4	h36c2ea0_1001	5 KB	conda-forge
pulseaudio-16.1	h4a94279_0	1.7 MB	conda-forge
pyqt-5.15.7	py39h18e9c17_2	6.2 MB	conda-forge
pyqt5-sip-12.11.0	py39h5a03fae_2	86 KB	conda-forge
python-3.9.14	h47a2c10_0_cpython	21.0 MB	conda-forge
qt-main-5.15.6	h7acdfe8_2	46.1 MB	conda-forge
sip-6.7.5	py39h5a03fae_0	464 KB	conda-forge
sqlite-3.40.0	h4ff8645_0	801 KB	conda-forge

toml-0.10.2	pyhd8ed1ab_0	18 KB	conda-forge
unicodedata2-15.0.0	py39hb9d737c_0	500 KB	conda-forge
xcb-util-0.4.0	h166bdaf_0	20 KB	conda-forge
xcb-util-image-0.4.0	h166bdaf_0	24 KB	conda-forge
xcb-util-keysyms-0.4.0	h166bdaf_0	12 KB	conda-forge
xcb-util-renderutil-0.3.9	h166bdaf_0	15 KB	conda-forge
xcb-util-wm-0.4.1	h166bdaf_0	55 KB	conda-forge
xorg-libxau-1.0.9	h7f98852_0	13 KB	conda-forge
xorg-libxdmcp-1.1.3	h7f98852_0	19 KB	conda-forge
zstd-1.5.2	h6239696_4	448 KB	conda-forge

Total: 270.5 MB

The following NEW packages will be INSTALLED:

alsa-lib	conda-forge/linux-64::alsa-lib-1.2.8-h166bdaf_0
attr	conda-forge/linux-64::attr-2.5.1-h166bdaf_1
brotli	conda-forge/linux-64::brotli-1.0.9-h166bdaf_8
brotli-bin	conda-forge/linux-64::brotli-bin-1.0.9-h166bdaf_8
certifi	conda-forge/noarch::certifi-2022.9.24-pyhd8ed1ab_0
contourpy	conda-forge/linux-64::contourpy-1.0.6-py39hf939315_0
cycler	conda-forge/noarch::cycler-0.11.0-pyhd8ed1ab_0
dbus	conda-forge/linux-64::dbus-1.13.6-h5008d03_3
expat	conda-forge/linux-64::expat-2.5.0-h27087fc_0
fftw	conda-forge/linux-64::fftw-3.3.10-nompi_hf0379b8_105
font-ttf-dejavu-s~	conda-forge/noarch::font-ttf-dejavu-sans-mono-2.37-hab24e00_0
font-ttf-inconsol~	conda-forge/noarch::font-ttf-inconsolata-3.000-h77eed37_0
font-ttf-source-c~	conda-forge/noarch::font-ttf-source-code-pro-2.038-h77eed37_0
font-ttf-ubuntu	conda-forge/noarch::font-ttf-ubuntu-0.83-hab24e00_0
fontconfig	conda-forge/linux-64::fontconfig-2.14.1-hc2a2eb6_0
fonts-conda-ecosy~	conda-forge/noarch::fonts-conda-ecosystem-1-0
fonts-conda-forge	conda-forge/noarch::fonts-conda-forge-1-0
fonttools	conda-forge/linux-64::fonttools-4.38.0-py39hb9d737c_1
freetype	conda-forge/linux-64::freetype-2.12.1-hca18f0e_0
gettext	conda-forge/linux-64::gettext-0.21.1-h27087fc_0
glib	conda-forge/linux-64::glib-2.74.1-h6239696_1
glib-tools	conda-forge/linux-64::glib-tools-2.74.1-h6239696_1
gst-plugins-base	conda-forge/linux-64::gst-plugins-base-1.21.2-h3e40eee_0
gstreamer	conda-forge/linux-64::gstreamer-1.21.2-hd4edc92_0
gstreamer-orc	conda-forge/linux-64::gstreamer-orc-0.4.33-h166bdaf_0
icu	conda-forge/linux-64::icu-70.1-h27087fc_0
jack	conda-forge/linux-64::jack-1.9.21-he978b8e_1
jpeg	conda-forge/linux-64::jpeg-9e-h166bdaf_2
keyutils	conda-forge/linux-64::keyutils-1.6.1-h166bdaf_0
kiwisolver	conda-forge/linux-64::kiwisolver-1.4.4-py39hf939315_1
krb5	conda-forge/linux-64::krb5-1.19.3-h3790be6_0
lame	conda-forge/linux-64::lame-3.100-h166bdaf_1003
lcms2	conda-forge/linux-64::lcms2-2.14-h6ed2654_0
lerc	conda-forge/linux-64::lerc-4.0.0-h27087fc_0
libbrotlicommon	conda-forge/linux-64::libbrotlicommon-1.0.9-h166bdaf_8
libbrotlidec	conda-forge/linux-64::libbrotlidec-1.0.9-h166bdaf_8
libbrotlienc	conda-forge/linux-64::libbrotlienc-1.0.9-h166bdaf_8
libcap	conda-forge/linux-64::libcap-2.66-ha37c62d_0
libclang	conda-forge/linux-64::libclang-15.0.5-default_h2e3cab8_0
libclang13	conda-forge/linux-64::libclang13-15.0.5-default_h3a83d3e_0
libcups	conda-forge/linux-64::libcups-2.3.3-h3e49a29_2

libdb	conda-forge/linux-64::libdb-6.2.32-h9c3ff4c_0
libdeflate	conda-forge/linux-64::libdeflate-1.14-h166bdaf_0
libedit	conda-forge/linux-64::libedit-3.1.20191231-he28a2e2_2
libevent	conda-forge/linux-64::libevent-2.1.10-h9b69904_4
libflac	conda-forge/linux-64::libflac-1.4.2-h27087fc_0
libgcrypt	conda-forge/linux-64::libgcrypt-1.10.1-h166bdaf_0
libglib	conda-forge/linux-64::libglib-2.74.1-h606061b_1
libgpg-error	conda-forge/linux-64::libgpg-error-1.45-hc0c96e0_0
libiconv	conda-forge/linux-64::libiconv-1.17-h166bdaf_0
libllvm15	conda-forge/linux-64::libllvm15-15.0.5-h63197d8_0
libogg	conda-forge/linux-64::libogg-1.3.4-h7f98852_1
libopus	conda-forge/linux-64::libopus-1.3.1-h7f98852_1
libpng	conda-forge/linux-64::libpng-1.6.38-h753d276_0
libpq	conda-forge/linux-64::libpq-14.5-hd77ab85_1
libsndfile	conda-forge/linux-64::libsndfile-1.1.0-h27087fc_0
libsystemd0	conda-forge/linux-64::libsystemd0-252-h2a991cd_0
libtiff	conda-forge/linux-64::libtiff-4.4.0-h55922b4_4
libtool	conda-forge/linux-64::libtool-2.4.6-h9c3ff4c_1008
libudev1	conda-forge/linux-64::libudev1-252-h166bdaf_0
libvorbis	conda-forge/linux-64::libvorbis-1.3.7-h9c3ff4c_0
libwebp-base	conda-forge/linux-64::libwebp-base-1.2.4-h166bdaf_0
libxcb	conda-forge/linux-64::libxcb-1.13-h7f98852_1004
libxkbcommon	conda-forge/linux-64::libxkbcommon-1.0.3-he3ba5ed_0
libxml2	conda-forge/linux-64::libxml2-2.10.3-h7463322_0
lz4-c	conda-forge/linux-64::lz4-c-1.9.3-h9c3ff4c_1
matplotlib	conda-forge/linux-64::matplotlib-3.6.2-py39hf3d152e_0
matplotlib-base	conda-forge/linux-64::matplotlib-base-3.6.2-py39hf9fd14e_0
mpg123	conda-forge/linux-64::mpg123-1.30.2-h27087fc_1
munkres	conda-forge/noarch::munkres-1.1.4-pyh9f0ad1d_0
mysql-common	conda-forge/linux-64::mysql-common-8.0.31-haf5c9bc_0
mysql-libs	conda-forge/linux-64::mysql-libs-8.0.31-h28c427c_0
nspr	conda-forge/linux-64::nspr-4.32-h9c3ff4c_1
nss	conda-forge/linux-64::nss-3.78-h2350873_0
openjpeg	conda-forge/linux-64::openjpeg-2.5.0-h7d73246_1
pcre2	conda-forge/linux-64::pcre2-10.40-hc3806b6_0
pillow	conda-forge/linux-64::pillow-9.2.0-py39hf3a2cdf_3
ply	conda-forge/noarch::ply-3.11-py_1
pthread-stubs	conda-forge/linux-64::pthread-stubs-0.4-h36c2ea0_1001
pulseaudio	conda-forge/linux-64::pulseaudio-16.1-h4a94279_0
pyqt	conda-forge/linux-64::pyqt-5.15.7-py39h18e9c17_2
pyqt5-sip	conda-forge/linux-64::pyqt5-sip-12.11.0-py39h5a03fae_2
qt-main	conda-forge/linux-64::qt-main-5.15.6-h7acdfc8_2
sip	conda-forge/linux-64::sip-6.7.5-py39h5a03fae_0
toml	conda-forge/noarch::toml-0.10.2-pyhd8ed1ab_0
unicodedata2	conda-forge/linux-64::unicodedata2-15.0.0-py39hb9d737c_0
xcb-util	conda-forge/linux-64::xcb-util-0.4.0-h166bdaf_0
xcb-util-image	conda-forge/linux-64::xcb-util-image-0.4.0-h166bdaf_0
xcb-util-keysyms	conda-forge/linux-64::xcb-util-keysyms-0.4.0-h166bdaf_0
xcb-util-renderutil~	conda-forge/linux-64::xcb-util-renderutil-0.3.9-h166bdaf_0
xcb-util-wm	conda-forge/linux-64::xcb-util-wm-0.4.1-h166bdaf_0
xorg-libxau	conda-forge/linux-64::xorg-libxau-1.0.9-h7f98852_0
xorg-libxdmcp	conda-forge/linux-64::xorg-libxdmcp-1.1.3-h7f98852_0
zstd	conda-forge/linux-64::zstd-1.5.2-h6239696_4

The following packages will be UPDATED:

libssqlite	3.39.4-h753d276_0 --> 3.40.0-h753d276_0
python	3.9.13-h2660328_0_cpython --> 3.9.14-h47a2c10_0
cpython	
sqlite	3.39.4-h4ff8645_0 --> 3.40.0-h4ff8645_0

The following packages will be DOWNGRADED:

openssl	3.0.7-h166bdaf_0 --> 1.1.1s-h166bdaf_0
---------	--

Downloading and Extracting Packages

mysql-common-8.0.31	1.9 MB	##### #####	##### #####	100%
fontconfig-2.14.1	365 KB	##### #####	##### #####	100%
contourpy-1.0.6	234 KB	##### #####	##### #####	100%
openjpeg-2.5.0	533 KB	##### #####	##### #####	100%
cycler-0.11.0	10 KB	##### #####	##### #####	100%
libbrotlidel-1.0.9	33 KB	##### #####	##### #####	100%
libbrotlicommon-1.0.	66 KB	##### #####	##### #####	100%
xorg-libxdmcp-1.1.3	19 KB	##### #####	##### #####	100%
libevent-2.1.10	1.1 MB	##### #####	##### #####	100%
libbrotlienc-1.0.9	288 KB	##### #####	##### #####	100%
nspr-4.32	233 KB	##### #####	##### #####	100%
libllvm15-15.0.5	37.4 MB	##### #####	##### #####	100%
libclang-15.0.5	127 KB	##### #####	##### #####	100%
libpq-14.5	3.0 MB	##### #####	##### #####	100%
libdeflate-1.14	81 KB	##### #####	##### #####	100%
unicodedata2-15.0.0	500 KB	##### #####	##### #####	100%
glib-2.74.1	474 KB	##### #####	##### #####	100%
font-ttf-ubuntu-0.83	1.9 MB	##### #####	##### #####	100%
pcre2-10.40	2.3 MB	##### #####	##### #####	100%
libxkbcommon-1.0.3	581 KB	##### #####	##### #####	100%
fftw-3.3.10	2.2 MB	##### #####	##### #####	100%
python-3.9.14	21.0 MB	##### #####	##### #####	100%
libopus-1.3.1	255 KB	##### #####	##### #####	100%
pulseaudio-16.1	1.7 MB	##### #####	##### #####	100%
lcms2-2.14	256 KB	##### #####	##### #####	100%
matplotlib-base-3.6.	7.5 MB	##### #####	##### #####	100%
libsndfile-1.1.0	370 KB	##### #####	##### #####	100%
xcb-util-image-0.4.0	24 KB	##### #####	##### #####	100%
xcb-util-renderutil-	15 KB	##### #####	##### #####	100%
glib-tools-2.74.1	107 KB	##### #####	##### #####	100%
matplotlib-3.6.2	7 KB	##### #####	##### #####	100%
gstreamer-1.21.2	1.9 MB	##### #####	##### #####	100%
freetype-2.12.1	884 KB	##### #####	##### #####	100%
libpng-1.6.38	371 KB	##### #####	##### #####	100%
kiwisolver-1.4.4	76 KB	##### #####	##### #####	100%
icu-70.1	13.5 MB	##### #####	##### #####	100%
xorg-libxau-1.0.9	13 KB	##### #####	##### #####	100%
zstd-1.5.2	448 KB	##### #####	##### #####	100%
sip-6.7.5	464 KB	##### #####	##### #####	100%
libxml2-2.10.3	754 KB	##### #####	##### #####	100%
lerc-4.0.0	275 KB	##### #####	##### #####	100%
pthread-stubs-0.4	5 KB	##### #####	##### #####	100%
libglib-2.74.1	3.1 MB	##### #####	##### #####	100%
pillow-9.2.0	45.3 MB	##### #####	##### #####	100%

libwebp-base-1.2.4	404 KB	##### #####	100%
openssl-1.1.1s	2.1 MB	##### #####	100%
xcb-util-0.4.0	20 KB	##### #####	100%
jack-1.9.21	530 KB	##### #####	100%
libogg-1.3.4	206 KB	##### #####	100%
sqlite-3.40.0	801 KB	##### #####	100%
pyqt5-sip-12.11.0	86 KB	##### #####	100%
libflac-1.4.2	411 KB	##### #####	100%
munkres-1.1.4	12 KB	##### #####	100%
attr-2.5.1	69 KB	##### #####	100%
libcups-2.3.3	4.5 MB	##### #####	100%
jpeg-9e	269 KB	##### #####	100%
xcb-util-wm-0.4.1	55 KB	##### #####	100%
brotli-1.0.9	18 KB	##### #####	100%
mpg123-1.30.2	512 KB	##### #####	100%
libdb-6.2.32	23.3 MB	##### #####	100%
pyqt-5.15.7	6.2 MB	##### #####	100%
libtool-2.4.6	511 KB	##### #####	100%
gst-plugins-base-1.2	2.5 MB	##### #####	100%
nss-3.78	2.1 MB	##### #####	100%
gettext-0.21.1	4.1 MB	##### #####	100%
libclang13-15.0.5	11.0 MB	##### #####	100%
libudev1-252	118 KB	##### #####	100%
libxcb-1.13	391 KB	##### #####	100%
fonts-conda-ecosyste	4 KB	##### #####	100%
lame-3.100	496 KB	##### #####	100%
libgpg-error-1.45	286 KB	##### #####	100%
xcb-util-keysyms-0.4	12 KB	##### #####	100%
alsa-lib-1.2.8	578 KB	##### #####	100%
libsystemd0-252	384 KB	##### #####	100%
mysql-libs-8.0.31	1.9 MB	##### #####	100%
keyutils-1.6.1	115 KB	##### #####	100%
krb5-1.19.3	1.4 MB	##### #####	100%
expat-2.5.0	189 KB	##### #####	100%
libedit-3.1.20191231	121 KB	##### #####	100%
font-ttf-source-code	684 KB	##### #####	100%
libcap-2.66	97 KB	##### #####	100%
lz4-c-1.9.3	179 KB	##### #####	100%
gstreamer-orc-0.4.33	299 KB	##### #####	100%
dbus-1.13.6	604 KB	##### #####	100%
font-ttf-inconsolata	94 KB	##### #####	100%
font-ttf-dejavu-sans	388 KB	##### #####	100%
fonts-conda-forge-1	4 KB	##### #####	100%
qt-main-5.15.6	46.1 MB	##### #####	100%
ply-3.11	44 KB	##### #####	100%
libiconv-1.17	1.4 MB	##### #####	100%
libvorbis-1.3.7	280 KB	##### #####	100%
libtiff-4.4.0	651 KB	##### #####	100%
libgcrypt-1.10.1	703 KB	##### #####	100%
fonttools-4.38.0	1.9 MB	##### #####	100%
toml-0.10.2	18 KB	##### #####	100%
brotli-bin-1.0.9	20 KB	##### #####	100%

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

Note: you may need to restart the kernel to use updated packages.

In [5]: `%conda install seaborn`

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <==
    current version: 4.10.3
    latest version: 22.9.0
```

```
Please update conda by running
```

```
$ conda update -n base conda
```

```
## Package Plan ##
```

```
environment location: /home/studio-lab-user/.conda/envs/default
```

```
added / updated specs:
- seaborn
```

```
The following packages will be downloaded:
```

package	build		
patsy-0.5.3	pyhd8ed1ab_0	189 KB	conda-forge
scipy-1.9.3	py39hddc5342_2	26.2 MB	conda-forge
seaborn-0.12.1	hd8ed1ab_0	5 KB	conda-forge
seaborn-base-0.12.1	pyhd8ed1ab_0	217 KB	conda-forge
statsmodels-0.13.5	py39h2ae25f5_2	11.2 MB	conda-forge
-----		Total:	37.8 MB

```
The following NEW packages will be INSTALLED:
```

```
patsy           conda-forge/noarch::patsy-0.5.3-pyhd8ed1ab_0
scipy           conda-forge/linux-64::scipy-1.9.3-py39hddc5342_2
seaborn          conda-forge/noarch::seaborn-0.12.1-hd8ed1ab_0
seaborn-base     conda-forge/noarch::seaborn-base-0.12.1-pyhd8ed1ab_0
statsmodels      conda-forge/linux-64::statsmodels-0.13.5-py39h2ae25f5_2
```

Downloading and Extracting Packages

```
seaborn-base-0.12.1 | 217 KB    | #####| 100%
seaborn-0.12.1    | 5 KB     | #####| 100%
statsmodels-0.13.5 | 11.2 MB   | #####| 100%
scipy-1.9.3        | 26.2 MB   | #####| 100%
patsy-0.5.3        | 189 KB    | #####| 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
In [6]: %conda install scikit-learn
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done
```

```
==> WARNING: A newer version of conda exists. <=>
  current version: 4.10.3
  latest version: 22.9.0
```

```
Please update conda by running
```

```
$ conda update -n base conda
```

```
## Package Plan ##
```

```
environment location: /home/studio-lab-user/.conda/envs/default
```

```
added / updated specs:
- scikit-learn
```

```
The following packages will be downloaded:
```

package	build		
joblib-1.2.0	pyhd8ed1ab_0	205 KB	conda-forge
scikit-learn-1.1.3	py39hd5c8da3_1	8.3 MB	conda-forge
threadpoolctl-3.1.0	pyh8a188c0_0	18 KB	conda-forge
Total:		8.5 MB	

```
The following NEW packages will be INSTALLED:
```

```
joblib           conda-forge/noarch::joblib-1.2.0-pyhd8ed1ab_0
scikit-learn     conda-forge/linux-64::scikit-learn-1.1.3-py39hd5c8da3_1
threadpoolctl    conda-forge/noarch::threadpoolctl-3.1.0-pyh8a188c0_0
```

```
Downloading and Extracting Packages
```

```
joblib-1.2.0      | 205 KB      | #####| 100%
scikit-learn-1.1.3 | 8.3 MB      | #####| 100%
threadpoolctl-3.1.0 | 18 KB      | #####| 100%
```

```
Preparing transaction: done
```

```
Verifying transaction: done
```

```
Executing transaction: done
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
In [1]: import numpy as np
from math import log10 as log10
import random
```

```
from time import time
print("Libraries imported")
```

Libraries imported

```
In [2]: print(np.log10(100))
```

2.0

```
In [3]: n = 10000000 #set the variable n to 10,000,000

l1 = list(np.random.uniform(low=1.0, high=100.0, size=n))

l2=[ ]

a1 = np.array(l1)

print("Variables are all set.")
```

Variables are all set.

```
In [4]: t0 = time()
```

```
for i in l1:
    l2.append(log10(i))

loopTime = time()-t0

print('Loop time: ' + str(round(loopTime, 3)) + ' seconds')
```

Loop time: 1.613 seconds

```
In [5]: t0 = time()
```

```
a2 = np.log10(a1)

vectorTime = time() - t0

print('Vectorized time: ' + str(round(loopTime, 3)) + ' seconds')
```

Vectorized time: 1.613 seconds

```
In [6]: ratio = (loopTime/vectorTime)
print("Ratio = " + str(round(ratio,3)))
```

Ratio = 45.166

```
In [7]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [8]: #Load training and test datasets into pandas dataframes
```

```
train = pd.read_csv('titanic_train.csv')
test = pd.read_csv('titanic_test.csv')
```

```
In [9]: #Inspect your dataset, the first 5 rows

#Some column/field definitions/explanations:
#Pclass = passenger class
#Sibsp = # of siblings/spouses on board
#parch = # of parents/children on board

train.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [10]: test.head()

Out[10]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

In [11]: `#Use pandas to display basic descriptive statistics
train.describe(include='all')`

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	891.000000
unique	Nan	Nan	Nan	891	2	Nan	Nan	Nan
top	Nan	Nan	Nan	Braund, Mr. Owen Harris	male	Nan	Nan	Nan
freq	Nan	Nan	Nan	1	577	Nan	Nan	Nan
mean	446.000000	0.383838	2.308642	Nan	Nan	29.699118	0.523008	0.381594
std	257.353842	0.486592	0.836071	Nan	Nan	14.526497	1.102743	0.806057
min	1.000000	0.000000	1.000000	Nan	Nan	0.420000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	Nan	Nan	20.125000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	Nan	Nan	28.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	Nan	Nan	38.000000	1.000000	0.000000
max	891.000000	1.000000	3.000000	Nan	Nan	80.000000	8.000000	6.000000

In [12]: `#Do the same with the test set`

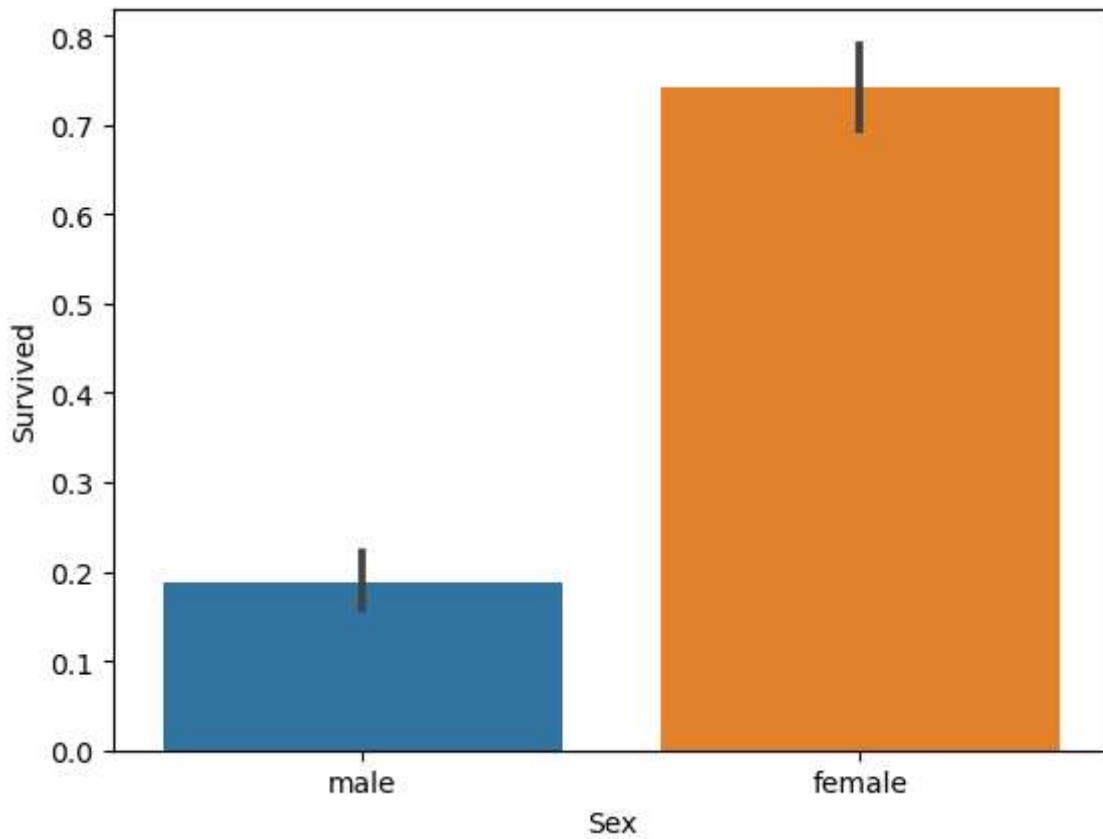
```
test.describe(include='all')
```

Out[12]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
count	418.000000	418.000000	418	418	332.000000	418.000000	418.000000	418	417.000000
unique	NaN	NaN	418	2	NaN	NaN	NaN	363	NaN
top	NaN	NaN	Kelly, Mr. James	male	NaN	NaN	NaN	PC 17608	NaN
freq	NaN	NaN	1	266	NaN	NaN	NaN	5	NaN
mean	1100.500000	2.265550	NaN	NaN	30.272590	0.447368	0.392344	NaN	35.627000
std	120.810458	0.841838	NaN	NaN	14.181209	0.896760	0.981429	NaN	55.907000
min	892.000000	1.000000	NaN	NaN	0.170000	0.000000	0.000000	NaN	0.000000
25%	996.250000	1.000000	NaN	NaN	21.000000	0.000000	0.000000	NaN	7.895000
50%	1100.500000	3.000000	NaN	NaN	27.000000	0.000000	0.000000	NaN	14.454000
75%	1204.750000	3.000000	NaN	NaN	39.000000	1.000000	0.000000	NaN	31.500000
max	1309.000000	3.000000	NaN	NaN	76.000000	8.000000	9.000000	NaN	512.329000

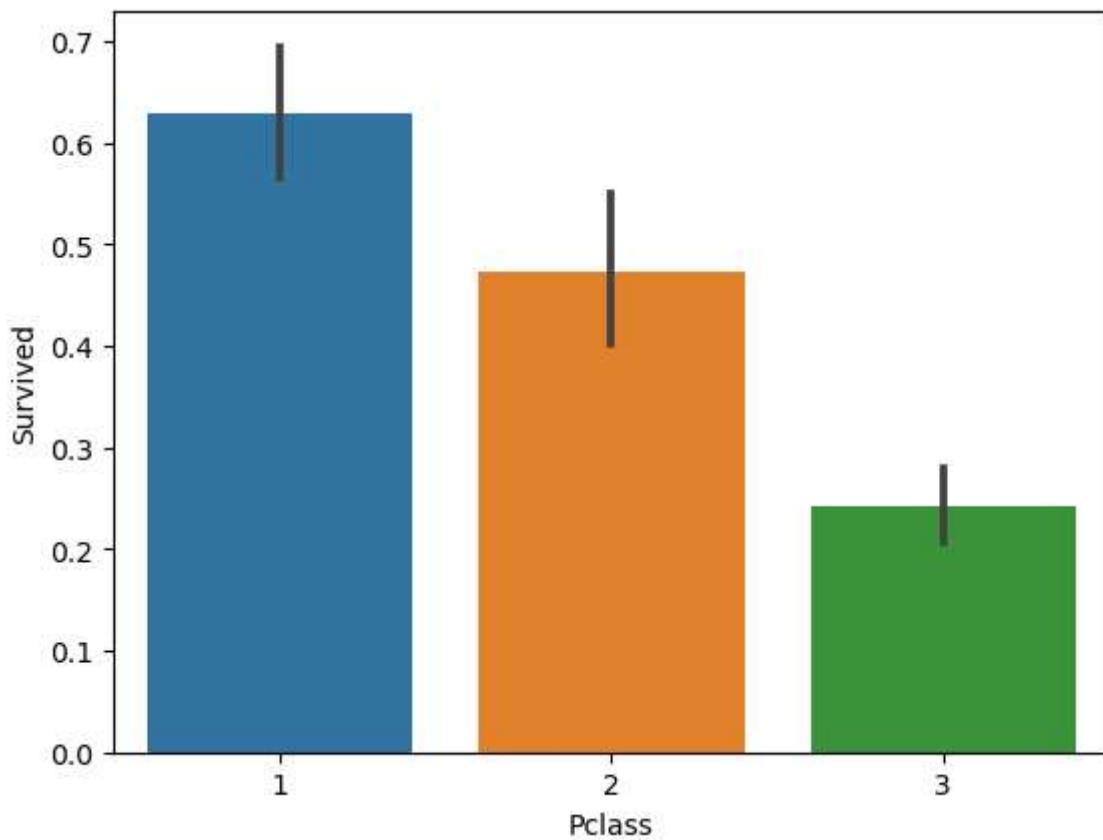
In [13]:

```
#Time for some visualizations  
#Who was more likely to survive, males or females?  
#Use a seaborn bar graph  
  
sns.barplot(x='Sex', y='Survived', data=train);
```



```
In [14]: #Which socia-economic class was more likely to survive?
```

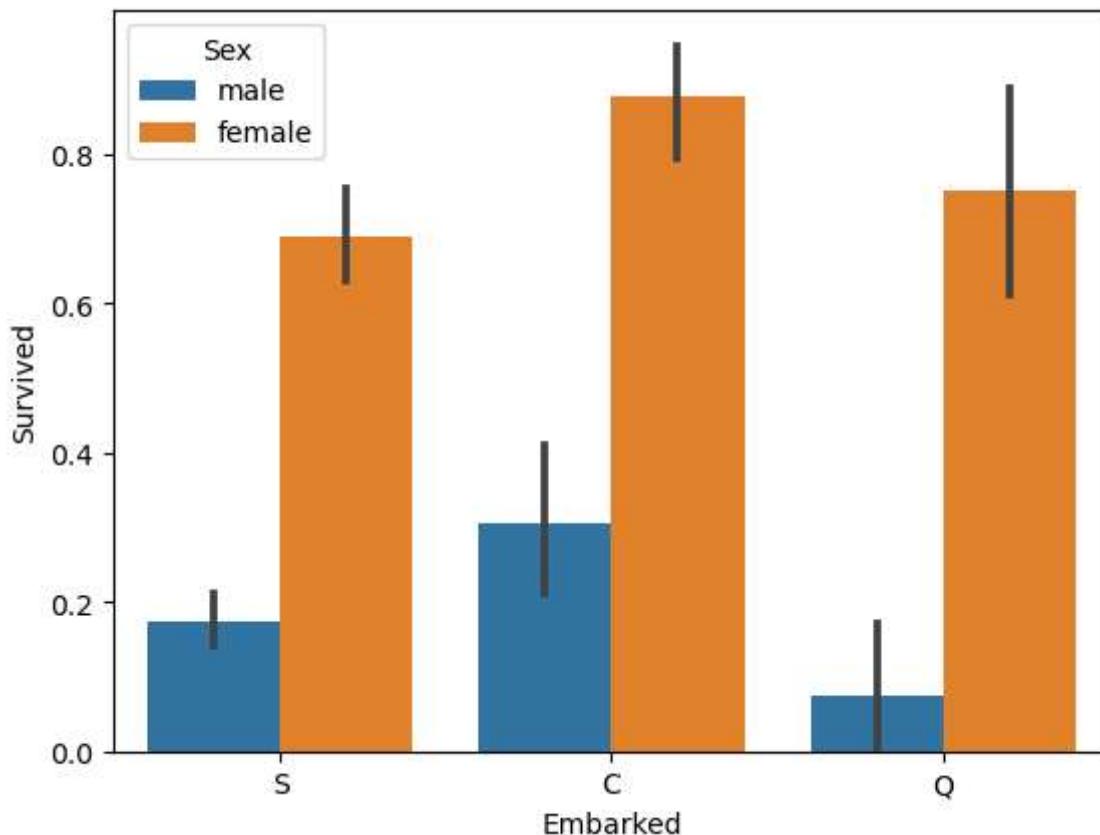
```
sns.barplot(x='Pclass', y='Survived', data=train);
```



```
In [15]: #Did the point of embarkation make a difference?
```

```
#C= Cherbourg (France), Q = Queenstown, S = Southampton
```

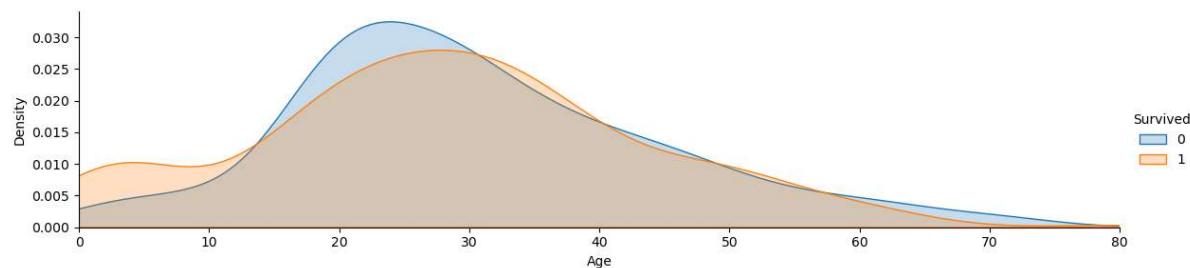
```
sns.barplot(x='Embarked', y='Survived', hue='Sex', data=train);
```



```
In [16]: #What about age distribution?
```

```
a = sns.FacetGrid(train, hue='Survived', aspect=4)
a.map(sns.kdeplot, 'Age', fill=True)
a.set(xlim=(0,train['Age'].max()))
a.add_legend()
```

```
Out[16]: <seaborn.axisgrid.FacetGrid at 0x7fe680efa1f0>
```



What are we solving for? What's our dependent y variable?

Survived (=1) or Died(=0)

```
In [17]: y = train.Survived #This y variable will store the 'survived' data
```

```
In [18]: y
```

```
Out[18]: 0      0  
1      1  
2      1  
3      1  
4      0  
..  
886    0  
887    1  
888    0  
889    1  
890    0  
Name: Survived, Length: 891, dtype: int64
```

Now let's get the data ready

```
In [19]: #Reformat the data into pandas dataframes to get ready for Machine Learning algorithm  
#We have to know the shape of these datasets later when we join and split them
```

```
train_shape = train.shape #Get the columns and rows of the training data  
train_rows = train.shape[0] #Get number of rows from index 0  
train_cols = train.shape[1] #Get number of columns from index 1  
  
print("The shape of train is" + str(train_shape))  
print("Our training set has " + str(train_rows) + ' rows')  
print ("Our training set has " + str(train_cols) + ' columns')  
  
train.head() #Displays the first few rows
```

```
The shape of train is(891, 12)  
Our training set has 891 rows  
Our training set has 12 columns
```

Out[19]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [20]: #Do the same thing for the test set

```
test_shape = test.shape #Get the columns and rows of the training data
test_rows = test.shape[0] #Get number of rows from index 0
test_cols = test.shape[1] #Get number of columns from index 1

print("The shape of test is" + str(test_shape))
print("Our test set has " + str(test_rows) + ' rows')
print ("Our test set has " + str(test_cols) + ' columns')

test.head() #Displays the first few rows
```

The shape of test is(418, 11)

Our test set has 418 rows

Our test set has 11 columns

Out[20]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindavist)	female	22.0	1	1	3101298	12.2875	NaN	S

In [21]: #Now concatenate the test and training sets together to make sure that they transform correctly

```
combined = pd.concat((train,test)) # Combine the train and test dataframes together

combined_shape = combined.shape
combined_rows = combined.shape[0]
combined_cols = combined.shape[1]
print("Our concatenated set has " + str(combined_rows) + " rows")
print("Our concatenated set has " +str(combined_cols) + " columns")
combined.head()
```

Our concatenated set has 1309 rows

Our concatenated set has 12 columns

Out[21]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1.0	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [22]: #But what about the survived column? It is in the train set but not the test set.
combined.sample(100)

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
774	775	1.0	2	Hocking, Mrs. Elizabeth (Eliza Needs)	female	54.0	1	3	29105	23.0000	Na
276	277	0.0	3	Lindblom, Miss. Augusta Charlotta	female	45.0	0	0	347073	7.7500	Na
237	1129	NaN	3	Baccos, Mr. Raffull	male	20.0	0	0	2679	7.2250	Na
658	659	0.0	2	Eitemiller, Mr. George Floyd	male	23.0	0	0	29751	13.0000	Na
63	955	NaN	3	Bradley, Miss. Bridget Delia	female	22.0	0	0	334914	7.7250	Na
...											
600	601	1.0	2	Jacobsohn, Mrs. Sidney Samuel (Amy Frances Chr...)	female	24.0	2	1	243847	27.0000	Na
325	326	1.0	1	Young, Miss. Marie Grice	female	36.0	0	0	PC 17760	135.6333	C3
153	154	0.0	3	van Billiard, Mr. Austin Blyler	male	40.5	0	2	A/5. 851	14.5000	Na
574	575	0.0	3	Rush, Mr. Alfred George John	male	16.0	0	0	A/4. 20589	8.0500	Na
450	451	0.0	2	West, Mr. Edwy Arthur	male	36.0	1	2	C.A. 34651	27.7500	Na

100 rows × 12 columns

In [34]: # Now, we transform the data, take care of the nulls, and simplify it by arranging
and drop irrelevant or difficult columns/fields.

```

# We will eliminate the survived column, because we already set the survived data at 1 or 0

def simplify_ages(df):
    df.Age = df.Age.fillna(-0.5)
    bins = (-1, 0, 5, 12, 18, 25, 35, 60, 120)
    group_names = ['Unknown', 'Baby', 'Child', 'Teenage', 'Student', 'Young Adult',
    categories = pd.cut(df.Age, bins, labels=group_names)
    df.Age = categories
    return df

def simplify_cabins(df):
    df.Cabin = df.Cabin.fillna('N')
    df.Cabin = df.Cabin.apply(lambda x: x[0])
    return df

def simplify_fares(df):
    df.Fare = df.Fare.fillna(-0.5)
    bins = (-1, 0, 8, 15, 31, 1000)
    group_names = ['Unknown', '1_quartile', '2_quartile', '3_quartile', '4_quartile']
    categories = pd.cut(df.Fare,bins,labels=group_names)
    df.Fare = categories
    return df

def drop_features(df):
    return df.drop(['Cabin', 'Name', 'Ticket', 'PassengerID', 'Survived'], axis=1)

```

In [35]: #Define a function to run those above, and run them

```

def transform_features(df):
    df = simplify_ages(df)
    df = simplify_cabins(df)
    df = simplify_fares(df)
    df = drop_features(df)
    return df

combined = transform_features(combined)
combined.head()

```

```

-----
AttributeError                                     Traceback (most recent call last)
/tmp/ipykernel_77/2402272950.py in <cell line: 10>()
      8     return df
      9
----> 10 combined = transform_features(combined)
     11 combined.head()

/tmp/ipykernel_77/2402272950.py in transform_features(df)
     2
     3 def transform_features(df):
----> 4     df = simplify_ages(df)
     5     df = simplify_cabins(df)
     6     df = simplify_fares(df)

/tmp/ipykernel_77/3169066759.py in simplify_ages(df)
     4
     5 def simplify_ages(df):
----> 6     df.Age = df.Age.fillna(-0.5)
     7     bins = (-1, 0, 5, 12, 18, 25, 35, 60, 120)
     8     group_names = ['Unknown', 'Baby', 'Child', 'Teenage', 'Student', 'Young Adult', 'Adult', 'Senior']

~/anaconda3/envs/default/lib/python3.9/site-packages/pandas/core/generic.py in __getattribute__(self, name)
  5900         ):
  5901             return self[name]
-> 5902         return object.__getattribute__(self, name)
  5903
  5904     def __setattr__(self, name: str, value) -> None:

AttributeError: 'DataFrame' object has no attribute 'Age'

```

In [36]: # Now we will do one-hot encoding - essentially pivot the binned fields into columns
#for each bin value

```

combined = pd.get_dummies(combined)
combined.head()

```

Out[36]:

	PassengerId	Survived	Pclass	SibSp	Parch	Name_Abbing, Mr. Anthony	Name_Abbott, Master. Eugene Joseph	Name_Abbott, Mr. Rossmore Edward	Name_Abbott, Mr. (
0	1	0.0	3	1	0	0	0	0	0
1	2	1.0	1	1	0	0	0	0	0
2	3	1.0	3	0	0	0	0	0	0
3	4	1.0	1	1	0	0	0	0	0
4	5	0.0	3	0	0	0	0	0	0

5 rows × 2268 columns

```
In [37]: #Load up the matrices, changes the pandas dataframes into numpy arrays, check fitting  
# and split the data back into training and test sets  
  
#Create an array from 'combined' that goes from the start to 'train_rows'  
X_train = combined[:train_rows]  
print('X_train: ' + str(X_train.shape))  
  
#Create an array from 'combined' that goes from 'train_rows' to the end  
X_test = combined[train_rows:]  
print('X_test: ' + str(X_test.shape))  
  
X_train: (891, 2268)  
X_test: (418, 2268)
```

```
In [38]: # Load up a pile of classification models and processing tools from the Scikit-learn library  
  
from sklearn.linear_model import LogisticRegression  
from sklearn.model_selection import train_test_split  
from sklearn.preprocessing import StandardScaler  
from sklearn.datasets import make_moons, make_circles, make_classification  
from sklearn.neural_network import MLPClassifier  
from sklearn.neighbors import KNeighborsClassifier  
from sklearn.svm import SVC  
from sklearn.gaussian_process import GaussianProcessClassifier  
from sklearn.gaussian_process.kernels import RBF  
from sklearn.tree import DecisionTreeClassifier  
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier  
from sklearn.naive_bayes import GaussianNB  
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
```

```
In [39]: #Create a list of names for these ML Algos  
  
names = ['Logistic regression', 'k-Nearest Neighbors', 'Linear SVM', 'RBH SVM',  
         'Gaussian Process', 'Decision Tree', 'Random Forest', 'Neural Network', 'AdaBoost Classifier']  
  
#Create a list of classifiers  
  
classifiers = [  
    LogisticRegression(),  
    KNeighborsClassifier(3),  
    SVC(kernel="linear", C=0.025),  
    SVC(gamma=2, C=1),  
    GaussianProcessClassifier(1.0 * RBF(1.0)),  
    DecisionTreeClassifier(max_depth=5),  
    RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1),  
    AdaBoostClassifier(),  
    GaussianNB(),]
```

Loop through all the classifiers to see how they perform on the dataset

```
In [40]: #Use the zip function in a for-loop to run the ML algos with their names  
#Use the fit function ot match the data to the y dependent variable in each ML also
```

```
#Measure their scores and print the results.

for name, clf in zip(names, classifiers):
    clf.fit(X_train, y)
    accuracy = round(clf.score(X_train,y) * 100, 2)
    print(name, accuracy)
```

```
/home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages/sklearn/linear_model/_logistic.py:444: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
n_iter_i = _check_optimize_result()
Logistic regression 100.0
k-Nearest Neighbors 83.61
Linear SVM 100.0
RBH SVM 100.0

/home/studio-lab-user/.conda/envs/default/lib/python3.9/site-packages/sklearn/gaussian_process/kernels.py:430: ConvergenceWarning: The optimal value found for dimension 0 of parameter k2_length_scale is close to the specified upper bound 100000.0. Increasing the bound and calling fit again may find a better value.
    warnings.warn(
Gaussian Process 61.62
Decision Tree 100.0
Random Forest 61.73
Neural Network 100.0
Adaboost 100.0
```

```
In [43]: # Let's pick the winner, the Support Vector Machine w/ RBF kernel function
#Store the model's predictions for each input (feature vector)

clf = SVC(gamma=2, C=1)
clf.fit(X_train, y)
accuracy = round(clf.score(X_train, y) * 100, 2)
print("Our accuracy score is: " + str(accuracy))
predictions = clf.predict(X_train)
```

Our accuracy score is: 100.0

```
In [46]: # Export our prediction into a file for use by the outside world

solution = pd.DataFrame({'PassengerId':test.PassengerId, 'Survived':predictions})
solution.to_csv('best_fit.csv', index=False)

print('Prediction file has been created')
```

```
-----  
ValueError Traceback (most recent call last)  
/tmp/ipykernel_77/1712183508.py in <cell line: 3>()  
      1 # Export our prediction into a file for use by the outside world  
      2  
----> 3 solution = pd.DataFrame({'PassengerId':test.PassengerId, 'Survived':predictions})  
      4 solution.to_csv('best_fit.csv', index=False)  
      5  
  
~/conda/envs/default/lib/python3.9/site-packages/pandas/core/frame.py in __init__  
(self, data, index, columns, dtype, copy)  
    660         elif isinstance(data, dict):  
    661             # GH#38939 de facto copy defaults to False only in non-dict ca  
ses  
--> 662                 mgr = dict_to_mgr(data, index, columns, dtype=dtype, copy=copy,  
typ=manager)  
    663                 elif isinstance(data, ma.MaskedArray):  
    664                     import numpy.ma.mrecords as mrecords  
  
~/conda/envs/default/lib/python3.9/site-packages/pandas/core/internals/constructi  
on.py in dict_to_mgr(data, index, columns, dtype, typ, copy)  
    491                     arrays = [x.copy() if hasattr(x, "dtype") else x for x in arr  
ays]  
    492  
--> 493             return arrays_to_mgr(arrays, columns, index, dtype=dtype, typ=typ, con  
solidate=copy)  
    494  
    495  
  
~/conda/envs/default/lib/python3.9/site-packages/pandas/core/internals/constructi  
on.py in arrays_to_mgr(arrays, columns, index, dtype, verify_integrity, typ, con  
solidate)  
    116                 # figure out the index, if necessary  
    117                 if index is None:  
--> 118                     index = _extract_index(arrays)  
    119                 else:  
    120                     index = ensure_index(index)  
  
~/conda/envs/default/lib/python3.9/site-packages/pandas/core/internals/constructi  
on.py in _extract_index(data)  
    678                     f"length {len(index)}"  
    679                 )  
--> 680             raise ValueError(msg)  
    681         else:  
    682             index = default_index(lengths[0])  
  
ValueError: array length 891 does not match index length 418
```

In [47]: best_fit = pd.read_csv('best_fit.csv')
best_fit.sample(20)

```
-----  
FileNotFoundException                                Traceback (most recent call last)  
/tmp/ipykernel_77/884015407.py in <cell line: 1>()  
----> 1 best_fit = pd.read_csv('best_fit.csv')  
      2 best_fit.sample(20)  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/util/_decorators.py in wr  
apper(*args, **kwargs)  
    209             else:  
    210                 kwargs[new_arg_name] = new_arg_value  
--> 211             return func(*args, **kwargs)  
    212  
    213         return cast(F, wrapper)  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/util/_decorators.py in wr  
apper(*args, **kwargs)  
    329                     stacklevel=find_stack_level(),  
    330                 )  
--> 331             return func(*args, **kwargs)  
    332  
    333             # error: "Callable[[VarArg(Any), KwArg(Any)], Any]" has no  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/io/parsers/readers.py in  
read_csv(filepath_or_buffer, sep, delimiter, header, names, index_col, usecols, sq  
ueeze, prefix, mangle_dupe_cols, dtype, engine, converters, true_values, false_val  
ues, skipinitialspace, skiprows, skipfooter, nrows, na_values, keep_default_na, na  
_filter, verbose, skip_blank_lines, parse_dates, infer_datetime_format, keep_date_  
col, date_parser, dayfirst, cache_dates, iterator, chunksize, compression, thousan  
ds, decimal, lineterminator, quotechar, quoting, doublequote, escapechar, comment,  
encoding, encoding_errors, dialect, error_bad_lines, warn_bad_lines, on_bad_lines,  
delim_whitespace, low_memory, memory_map, float_precision, storage_options)  
    948     kwds.update(kwds_defaults)  
    949  
--> 950     return _read(filepath_or_buffer, kwds)  
    951  
    952  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/io/parsers/readers.py in  
_read(filepath_or_buffer, kwds)  
    603  
    604     # Create the parser.  
--> 605     parser = TextFileReader(filepath_or_buffer, **kwds)  
    606  
    607     if chunksize or iterator:  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/io/parsers/readers.py in  
__init__(self, f, engine, **kwds)  
    1440  
    1441         self.handles: IOHandles | None = None  
-> 1442         self._engine = self._make_engine(f, self.engine)  
    1443  
    1444     def close(self) -> None:  
  
~/_.conda/envs/default/lib/python3.9/site-packages/pandas/io/parsers/readers.py in  
_make_engine(self, f, engine)  
    1733             if "b" not in mode:
```

```
1734             mode += "b"
-> 1735         self.handles = get_handle(
1736             f,
1737             mode,
1738
1739         ~/.conda/envs/default/lib/python3.9/site-packages/pandas/io/common.py in get_handle(path_or_buf, mode, encoding, compression, memory_map, is_text, errors, storage_options)
1740             if ioargs.encoding and "b" not in ioargs.mode:
1741                 # Encoding
1742             handle = open(
1743                 handle,
1744                 ioargs.mode,
1745
1746             FileNotFoundError: [Errno 2] No such file or directory: 'best_fit.csv'
```

In []: