out of core ML concept

In [1]:
```python
#importing important libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```python
!pip install vaex
```

```
Requirement already satisfied: vaex in c:\ana\anaconda3\lib\site-packages (4.16.0)
Requirement already satisfied: vaex-core<4.17,>=4.16.0 in c:\ana\anaconda3\lib\site-packag
es (from vaex) (4.16.1)
Requirement already satisfied: vaex-astro<0.10,>=0.9.3 in c:\ana\anaconda3\lib\site-packag
es (from vaex) (0.9.3)
Requirement already satisfied: vaex-jupyter<0.9,>=0.8.1 in c:\ana\anaconda3\lib\site-packa
ges (from vaex) (0.8.1)
Requirement already satisfied: vaex-ml<0.19,>=0.18.1 in c:\ana\anaconda3\lib\site-packages
(from vaex) (0.18.1)
Requirement already satisfied: vaex-server<0.9,>=0.8.1 in c:\ana\anaconda3\lib\site-packag
es (from vaex) (0.8.1)
Requirement already satisfied: vaex-hdf5<0.15,>=0.13.0 in c:\ana\anaconda3\lib\site-packag
es (from vaex) (0.14.1)
Requirement already satisfied: vaex-viz<0.6,>=0.5.4 in c:\ana\anaconda3\lib\site-packages
(from vaex) (0.5.4)
Requirement already satisfied: astropy in c:\ana\anaconda3\lib\site-packages (from vaex-as
tro<0.10,>=0.9.3->vaex) (4.3.1)
Requirement already satisfied: six in c:\ana\anaconda3\lib\site-packages (from vaex-core<
4.17,>=4.16.0->vaex) (1.16.0)
Requirement already satisfied: blake3 in c:\ana\anaconda3\lib\site-packages (from vaex-cor
e<4.17,>=4.16.0->vaex) (0.3.3)
Requirement already satisfied: frozendict!=2.2.0 in c:\ana\anaconda3\lib\site-packages (fr
om vaex-core<4.17,>=4.16.0->vaex) (2.3.6)
Requirement already satisfied: rich in c:\ana\anaconda3\lib\site-packages (from vaex-core<
4.17,>=4.16.0->vaex) (13.3.3)
Requirement already satisfied: dask!=2022.4.0 in c:\ana\anaconda3\lib\site-packages (from
vaex-core<4.17,>=4.16.0->vaex) (2021.10.0)
Requirement already satisfied: future>=0.15.2 in c:\ana\anaconda3\lib\site-packages (from
vaex-core<4.17,>=4.16.0->vaex) (0.18.2)
Requirement already satisfied: pandas in c:\ana\anaconda3\lib\site-packages (from vaex-cor
e<4.17,>=4.16.0->vaex) (1.3.4)
Requirement already satisfied: nest-asyncio>=1.3.3 in c:\ana\anaconda3\lib\site-packages
(from vaex-core<4.17,>=4.16.0->vaex) (1.5.1)
Requirement already satisfied: pyarrow>=5.0.0 in c:\ana\anaconda3\lib\site-packages (from
vaex-core<4.17,>=4.16.0->vaex) (11.0.0)
Requirement already satisfied: pydantic>=1.8.0 in c:\ana\anaconda3\lib\site-packages (from
vaex-core<4.17,>=4.16.0->vaex) (1.10.7)
Requirement already satisfied: tabulate>=0.8.3 in c:\ana\anaconda3\lib\site-packages (from
vaex-core<4.17,>=4.16.0->vaex) (0.9.0)
Requirement already satisfied: cloudpickle in c:\ana\anaconda3\lib\site-packages (from vae
x-core<4.17,>=4.16.0->vaex) (2.0.0)
Requirement already satisfied: requests in c:\ana\anaconda3\lib\site-packages (from vaex-c
ore<4.17,>=4.16.0->vaex) (2.26.0)
Requirement already satisfied: filelock in c:\ana\anaconda3\lib\site-packages (from vaex-c
ore<4.17,>=4.16.0->vaex) (3.3.1)
Requirement already satisfied: aplus in c:\ana\anaconda3\lib\site-packages (from vaex-core
```

```
      <4.17,>=4.16.0->vaex) (0.11.0)
Requirement already satisfied: progressbar2 in c:\ana\anaconda3\lib\site-packages (from va
ex-core<4.17,>=4.16.0->vaex) (4.2.0)
Requirement already satisfied: numpy>=1.16 in c:\ana\anaconda3\lib\site-packages (from vae
x-core<4.17,>=4.16.0->vaex) (1.20.3)
Requirement already satisfied: pyyaml in c:\ana\anaconda3\lib\site-packages (from vaex-cor
e<4.17,>=4.16.0->vaex) (6.0)
Requirement already satisfied: h5py>=2.9 in c:\ana\anaconda3\lib\site-packages (from vaex-
hdf5<0.15,>=0.13.0->vaex) (3.2.1)
Requirement already satisfied: ipyleaflet in c:\ana\anaconda3\lib\site-packages (from vaex
-jupyter<0.9,>=0.8.1->vaex) (0.17.2)
Requirement already satisfied: bqplot>=0.10.1 in c:\ana\anaconda3\lib\site-packages (from
vaex-jupyter<0.9,>=0.8.1->vaex) (0.12.38)
Requirement already satisfied: ipyvolume>=0.4 in c:\ana\anaconda3\lib\site-packages (from
vaex-jupyter<0.9,>=0.8.1->vaex) (0.6.1)
Requirement already satisfied: xarray in c:\ana\anaconda3\lib\site-packages (from vaex-jup
yter<0.9,>=0.8.1->vaex) (2023.1.0)
Requirement already satisfied: ipympl in c:\ana\anaconda3\lib\site-packages (from vaex-jup
yter<0.9,>=0.8.1->vaex) (0.9.3)
Requirement already satisfied: ipyvuetify<2,>=1.2.2 in c:\ana\anaconda3\lib\site-packages
(from vaex-jupyter<0.9,>=0.8.1->vaex) (1.8.5)
Requirement already satisfied: jinja2 in c:\ana\anaconda3\lib\site-packages (from vaex-ml<
0.19,>=0.18.1->vaex) (2.11.3)
Requirement already satisfied: traitlets in c:\ana\anaconda3\lib\site-packages (from vaex-
ml<0.19,>=0.18.1->vaex) (5.1.0)
Requirement already satisfied: numba in c:\ana\anaconda3\lib\site-packages (from vaex-ml<
0.19,>=0.18.1->vaex) (0.54.1)
Requirement already satisfied: tornado>4.1 in c:\ana\anaconda3\lib\site-packages (from vae
x-server<0.9,>=0.8.1->vaex) (6.1)
Requirement already satisfied: cachetools in c:\ana\anaconda3\lib\site-packages (from vaex
-server<0.9,>=0.8.1->vaex) (5.2.0)
Requirement already satisfied: uvicorn[standard] in c:\ana\anaconda3\lib\site-packages (fr
om vaex-server<0.9,>=0.8.1->vaex) (0.21.1)
Requirement already satisfied: fastapi in c:\ana\anaconda3\lib\site-packages (from vaex-se
rver<0.9,>=0.8.1->vaex) (0.95.0)
Requirement already satisfied: matplotlib>=1.3.1 in c:\ana\anaconda3\lib\site-packages (fr
om vaex-viz<0.6,>=0.5.4->vaex) (3.4.3)
Requirement already satisfied: pillow in c:\ana\anaconda3\lib\site-packages (from vaex-viz
<0.6,>=0.5.4->vaex) (8.4.0)
Requirement already satisfied: traittypes>=0.0.6 in c:\ana\anaconda3\lib\site-packages (fr
om bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.2.1)
Requirement already satisfied: ipywidgets<9,>=7.5.0 in c:\ana\anaconda3\lib\site-packages
(from bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (7.6.5)
Requirement already satisfied: fsspec>=0.6.0 in c:\ana\anaconda3\lib\site-packages (from d
ask!=2022.4.0->vaex-core<4.17,>=4.16.0->vaex) (2021.10.1)
Requirement already satisfied: packaging>=20.0 in c:\ana\anaconda3\lib\site-packages (from
dask!=2022.4.0->vaex-core<4.17,>=4.16.0->vaex) (23.0)
Requirement already satisfied: partd>=0.3.10 in c:\ana\anaconda3\lib\site-packages (from d
ask!=2022.4.0->vaex-core<4.17,>=4.16.0->vaex) (1.2.0)
Requirement already satisfied: toolz>=0.8.2 in c:\ana\anaconda3\lib\site-packages (from da
sk!=2022.4.0->vaex-core<4.17,>=4.16.0->vaex) (0.11.1)
Requirement already satisfied: ipyvue>=1.7.0 in c:\ana\anaconda3\lib\site-packages (from i
pyvolume>=0.4->vaex-jupyter<0.9,>=0.8.1->vaex) (1.9.0)
Requirement already satisfied: pythreejs>=2.4.0 in c:\ana\anaconda3\lib\site-packages (fro
m ipyvolume>=0.4->vaex-jupyter<0.9,>=0.8.1->vaex) (2.4.2)
Requirement already satisfied: ipywebrtc in c:\ana\anaconda3\lib\site-packages (from ipyvo
lume>=0.4->vaex-jupyter<0.9,>=0.8.1->vaex) (0.6.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\ana\anaconda3\lib\site-packages (fro
m matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.4->vaex) (3.0.4)
Requirement already satisfied: cycler>=0.10 in c:\ana\anaconda3\lib\site-packages (from ma
tplotlib>=1.3.1->vaex-viz<0.6,>=0.5.4->vaex) (0.10.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\ana\anaconda3\lib\site-packages
(from matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.4->vaex) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\ana\anaconda3\lib\site-packages (fr
om matplotlib>=1.3.1->vaex-viz<0.6,>=0.5.4->vaex) (1.3.1)
Requirement already satisfied: pytz>=2017.3 in c:\ana\anaconda3\lib\site-packages (from pa
```

```
ndas->vaex-core<4.17,>=4.16.0->vaex) (2021.3)
Requirement already satisfied: typing-extensions>=4.2.0 in c:\ana\anaconda3\lib\site-packa
ges (from pydantic>=1.8.0->vaex-core<4.17,>=4.16.0->vaex) (4.5.0)
Requirement already satisfied: pyerfa>=1.7.3 in c:\ana\anaconda3\lib\site-packages (from a
stropy->vaex-astro<0.10,>=0.9.3->vaex) (2.0.0)
Requirement already satisfied: starlette<0.27.0,>=0.26.1 in c:\ana\anaconda3\lib\site-pack
ages (from fastapi->vaex-server<0.9,>=0.8.1->vaex) (0.26.1)
Requirement already satisfied: xyzservices>=2021.8.1 in c:\ana\anaconda3\lib\site-packages
(from ipyleaflet->vaex-jupyter<0.9,>=0.8.1->vaex) (2023.2.0)
Requirement already satisfied: branca>=0.5.0 in c:\ana\anaconda3\lib\site-packages (from i
pyleaflet->vaex-jupyter<0.9,>=0.8.1->vaex) (0.6.0)
Requirement already satisfied: ipython<9 in c:\ana\anaconda3\lib\site-packages (from ipymp
l->vaex-jupyter<0.9,>=0.8.1->vaex) (7.29.0)
Requirement already satisfied: ipython-genutils in c:\ana\anaconda3\lib\site-packages (fro
m ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.2.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\ana\anaconda3\lib\site-packages (fro
m jinja2->vaex-ml<0.19,>=0.18.1->vaex) (1.1.1)
Requirement already satisfied: setuptools in c:\ana\anaconda3\lib\site-packages (from numb
a->vaex-ml<0.19,>=0.18.1->vaex) (58.0.4)
Requirement already satisfied: llvmlite<0.38,>=0.37.0rc1 in c:\ana\anaconda3\lib\site-pack
ages (from numba->vaex-ml<0.19,>=0.18.1->vaex) (0.37.0)
Requirement already satisfied: python-utils>=3.0.0 in c:\ana\anaconda3\lib\site-packages
(from progressbar2->vaex-core<4.17,>=4.16.0->vaex) (3.5.2)
Requirement already satisfied: certifi>=2017.4.17 in c:\ana\anaconda3\lib\site-packages (f
rom requests->vaex-core<4.17,>=4.16.0->vaex) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\ana\anaconda3\lib\site-packages
(from requests->vaex-core<4.17,>=4.16.0->vaex) (1.26.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\ana\anaconda3\lib\site-pack
ages (from requests->vaex-core<4.17,>=4.16.0->vaex) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\ana\anaconda3\lib\site-packages (from re
quests->vaex-core<4.17,>=4.16.0->vaex) (3.2)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in c:\ana\anaconda3\lib\site-packag
es (from rich->vaex-core<4.17,>=4.16.0->vaex) (2.14.0)
Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in c:\ana\anaconda3\lib\site-p
ackages (from rich->vaex-core<4.17,>=4.16.0->vaex) (2.2.0)
Requirement already satisfied: click>=7.0 in c:\ana\anaconda3\lib\site-packages (from uvic
orn[standard]->vaex-server<0.9,>=0.8.1->vaex) (8.0.3)
Requirement already satisfied: h11>=0.8 in c:\ana\anaconda3\lib\site-packages (from uvicor
n[standard]->vaex-server<0.9,>=0.8.1->vaex) (0.13.0)
Requirement already satisfied: python-dotenv>=0.13 in c:\ana\anaconda3\lib\site-packages
(from uvicorn[standard]->vaex-server<0.9,>=0.8.1->vaex) (1.0.0)
Requirement already satisfied: watchfiles>=0.13 in c:\ana\anaconda3\lib\site-packages (fro
m uvicorn[standard]->vaex-server<0.9,>=0.8.1->vaex) (0.19.0)
Requirement already satisfied: colorama>=0.4 in c:\ana\anaconda3\lib\site-packages (from u
vicorn[standard]->vaex-server<0.9,>=0.8.1->vaex) (0.4.4)
Requirement already satisfied: httptools>=0.5.0 in c:\ana\anaconda3\lib\site-packages (fro
m uvicorn[standard]->vaex-server<0.9,>=0.8.1->vaex) (0.5.0)
Requirement already satisfied: websockets>=10.4 in c:\ana\anaconda3\lib\site-packages (fro
m uvicorn[standard]->vaex-server<0.9,>=0.8.1->vaex) (10.4)
Requirement already satisfied: pickleshare in c:\ana\anaconda3\lib\site-packages (from ipy
thon<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.7.5)
Requirement already satisfied: matplotlib-inline in c:\ana\anaconda3\lib\site-packages (fr
om ipython<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.1.2)
Requirement already satisfied: jedi>=0.16 in c:\ana\anaconda3\lib\site-packages (from ipyt
hon<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.18.0)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in c:\ana\anac
onda3\lib\site-packages (from ipython<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (3.0.20)
Requirement already satisfied: decorator in c:\ana\anaconda3\lib\site-packages (from ipyth
on<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (5.1.0)
Requirement already satisfied: backcall in c:\ana\anaconda3\lib\site-packages (from ipytho
n<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.2.0)
Requirement already satisfied: nbformat>=4.2.0 in c:\ana\anaconda3\lib\site-packages (from
ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (5.1.3)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\ana\anaconda3\lib\site-pack
ages (from ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (1.0.0)
Requirement already satisfied: ipykernel>=4.5.1 in c:\ana\anaconda3\lib\site-packages (fro
```

```
m ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (6.4.1)
Requirement already satisfied: widgetsnbextension~=3.5.0 in c:\ana\anaconda3\lib\site-pack
ages (from ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (3.5.1)
Requirement already satisfied: mdurl~=0.1 in c:\ana\anaconda3\lib\site-packages (from mark
down-it-py<3.0.0,>=2.2.0->rich->vaex-core<4.17,>=4.16.0->vaex) (0.1.2)
Requirement already satisfied: locket in c:\ana\anaconda3\lib\site-packages\locket-0.2.1-p
y3.9.egg (from partd>=0.3.10->dask!=2022.4.0->vaex-core<4.17,>=4.16.0->vaex) (0.2.1)
Requirement already satisfied: ipydatawidgets>=1.1.1 in c:\ana\anaconda3\lib\site-packages
(from pythreejs>=2.4.0->ipyvolume>=0.4->vaex-jupyter<0.9,>=0.8.1->vaex) (4.3.3)
Requirement already satisfied: anyio<5,>=3.4.0 in c:\ana\anaconda3\lib\site-packages (from
starlette<0.27.0,>=0.26.1->fastapi->vaex-server<0.9,>=0.8.1->vaex) (3.6.2)
Requirement already satisfied: sniffio>=1.1 in c:\ana\anaconda3\lib\site-packages (from an
yio<5,>=3.4.0->starlette<0.27.0,>=0.26.1->fastapi->vaex-server<0.9,>=0.8.1->vaex) (1.2.0)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\ana\anaconda3\lib\site-packages
(from ipykernel>=4.5.1->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->va
ex) (1.4.1)
Requirement already satisfied: jupyter-client<8.0 in c:\ana\anaconda3\lib\site-packages (f
rom ipykernel>=4.5.1->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vae
x) (6.1.12)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\ana\anaconda3\lib\site-packages
(from jedi>=0.16->ipython<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex) (0.8.2)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\ana\anaconda3\lib\site-packag
es (from nbformat>=4.2.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->
vaex) (3.2.0)
Requirement already satisfied: jupyter-core in c:\ana\anaconda3\lib\site-packages (from nb
format>=4.2.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (4.8.
1)
Requirement already satisfied: wcwidth in c:\ana\anaconda3\lib\site-packages (from prompt-
toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython<9->ipympl->vaex-jupyter<0.9,>=0.8.1->vaex)
(0.2.5)
Requirement already satisfied: notebook>=4.4.1 in c:\ana\anaconda3\lib\site-packages (from
widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1-
>vaex) (6.4.5)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\ana\anaconda3\lib\site-packages (f
rom jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-j
upyter<0.9,>=0.8.1->vaex) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in c:\ana\anaconda3\lib\site-packages (from j
sonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyte
r<0.9,>=0.8.1->vaex) (21.2.0)
Requirement already satisfied: pyzmq>=13 in c:\ana\anaconda3\lib\site-packages (from jupyt
er-client<8.0->ipykernel>=4.5.1->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=
0.8.1->vaex) (22.2.1)
Requirement already satisfied: pywin32>=1.0 in c:\ana\anaconda3\lib\site-packages (from ju
pyter-core->nbformat>=4.2.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.
1->vaex) (228)
Requirement already satisfied: Send2Trash>=1.5.0 in c:\ana\anaconda3\lib\site-packages (fr
om notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.9,>=0.8.1->vaex) (1.8.0)
Requirement already satisfied: argon2-cffi in c:\ana\anaconda3\lib\site-packages (from not
ebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyte
r<0.9,>=0.8.1->vaex) (20.1.0)
Requirement already satisfied: nbconvert in c:\ana\anaconda3\lib\site-packages (from noteb
ook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<
0.9,>=0.8.1->vaex) (6.1.0)
Requirement already satisfied: terminado>=0.8.3 in c:\ana\anaconda3\lib\site-packages (fro
m notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-j
upyter<0.9,>=0.8.1->vaex) (0.9.4)
Requirement already satisfied: prometheus-client in c:\ana\anaconda3\lib\site-packages (fr
om notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.9,>=0.8.1->vaex) (0.11.0)
Requirement already satisfied: pywinpty>=0.5 in c:\ana\anaconda3\lib\site-packages (from t
erminado>=0.8.3->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>
=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.5.7)
Requirement already satisfied: cffi>=1.0.0 in c:\ana\anaconda3\lib\site-packages (from arg
on2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1
->vaex-jupyter<0.9,>=0.8.1->vaex) (1.14.6)
```

```
Requirement already satisfied: mistune<2,>=0.8.1 in c:\ana\anaconda3\lib\site-packages (fr
om nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.
10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.8.4)
Requirement already satisfied: defusedxml in c:\ana\anaconda3\lib\site-packages (from nbco
nvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->v
aex-jupyter<0.9,>=0.8.1->vaex) (0.7.1)
Requirement already satisfied: bleach in c:\ana\anaconda3\lib\site-packages (from nbconver
t->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vaex-
jupyter<0.9,>=0.8.1->vaex) (4.0.0)
Requirement already satisfied: testpath in c:\ana\anaconda3\lib\site-packages (from nbconv
ert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=0.10.1->vae
x-jupyter<0.9,>=0.8.1->vaex) (0.5.0)
Requirement already satisfied: entrypoints>=0.2.2 in c:\ana\anaconda3\lib\site-packages (f
rom nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>=
0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.3)
Requirement already satisfied: jupyterlab-pygments in c:\ana\anaconda3\lib\site-packages
(from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>
=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.1.2)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\ana\anaconda3\lib\site-package
s (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplo
t>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.5.3)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\ana\anaconda3\lib\site-packages
(from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>
=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (1.4.3)
Requirement already satisfied: pycparser in c:\ana\anaconda3\lib\site-packages (from cffi>
=1.0.0->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqp
lot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (2.20)
Requirement already satisfied: async-generator in c:\ana\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<
9,>=7.5.0->bqplot>=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (1.10)
Requirement already satisfied: webencodings in c:\ana\anaconda3\lib\site-packages (from bl
each->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets<9,>=7.5.0->bqplot>
=0.10.1->vaex-jupyter<0.9,>=0.8.1->vaex) (0.5.1)
```

In [3]:
```python
import vaex as vx
```

In [4]:
```python
df = pd.read_csv("Summary of Weather.csv")
```

In [5]:
```python
df.head()
```

Out[5]:

	STA	Date	Precip	WindGustSpd	MaxTemp	MinTemp	MeanTemp	Snowfall	PoorWeather	YR	...	FB	F
0	10001	1942-7-1	1.016	NaN	25.555556	22.222222	23.888889	0.0	NaN	42	...	NaN	Na
1	10001	1942-7-2	0	NaN	28.888889	21.666667	25.555556	0.0	NaN	42	...	NaN	Na
2	10001	1942-7-3	2.54	NaN	26.111111	22.222222	24.444444	0.0	NaN	42	...	NaN	Na
3	10001	1942-7-4	2.54	NaN	26.666667	22.222222	24.444444	0.0	NaN	42	...	NaN	Na
4	10001	1942-7-5	0	NaN	26.666667	21.666667	24.444444	0.0	NaN	42	...	NaN	Na

5 rows × 31 columns

In [6]:
```python
df.shape
```

```
Out[6]:   (119040, 31)
```

```
In [7]:   df.describe()
```

Out[7]:

	STA	WindGustSpd	MaxTemp	MinTemp	MeanTemp	YR	MO
count	119040.000000	532.000000	119040.000000	119040.000000	119040.000000	119040.000000	119040.000000	11
mean	29659.435795	37.774534	27.045111	17.789511	22.411631	43.805284	6.726016
std	20953.209402	10.297808	8.717817	8.334572	8.297982	1.136718	3.425561
min	10001.000000	18.520000	-33.333333	-38.333333	-35.555556	40.000000	1.000000
25%	11801.000000	29.632000	25.555556	15.000000	20.555556	43.000000	4.000000
50%	22508.000000	37.040000	29.444444	21.111111	25.555556	44.000000	7.000000
75%	33501.000000	43.059000	31.666667	23.333333	27.222222	45.000000	10.000000
max	82506.000000	75.932000	50.000000	34.444444	40.000000	45.000000	12.000000

8 rows × 24 columns

```
In [8]:   df.isnull().sum()
```

```
Out[8]:   STA                  0
          Date                 0
          Precip               0
          WindGustSpd     118508
          MaxTemp              0
          MinTemp              0
          MeanTemp             0
          Snowfall          1163
          PoorWeather      84803
          YR                   0
          MO                   0
          DA                   0
          PRCP              1932
          DR              118507
          SPD             118508
          MAX                474
          MIN                468
          MEA                498
          SNF               1163
          SND             113477
          FT              119040
          FB              119040
          FTI             119040
          ITH             119040
          PGT             118515
          TSHDSBRSGF       84803
          SD3             119040
          RHX             119040
          RHN             119040
          RVG             119040
          WTE             119040
          dtype: int64
```

```
In [9]:   drop_columns = df.columns[df.isnull().sum()>=80000]
```

```
In [10]:  df.drop(drop_columns,axis=1,inplace=True)
```

```
In [11]:   df.isnull().sum()
```

Out[11]:
```
STA           0
Date          0
Precip        0
MaxTemp       0
MinTemp       0
MeanTemp      0
Snowfall   1163
YR            0
MO            0
DA            0
PRCP       1932
MAX         474
MIN         468
MEA         498
SNF        1163
dtype: int64
```

```
In [12]:   #Using interpolation for filling Nan values
           df.fillna(df.mean(numeric_only=True).round(1), inplace=True)
```

```
In [13]:   df.isnull().sum()
```

Out[13]:
```
STA           0
Date          0
Precip        0
MaxTemp       0
MinTemp       0
MeanTemp      0
Snowfall   1163
YR            0
MO            0
DA            0
PRCP       1932
MAX           0
MIN           0
MEA           0
SNF        1163
dtype: int64
```

```
In [14]:   df.fillna(method='ffill', inplace=True)
```

```
In [15]:   df.isnull().sum()
```

Out[15]:
```
STA           0
Date          0
Precip        0
MaxTemp       0
MinTemp       0
MeanTemp      0
Snowfall      0
YR            0
MO            0
DA            0
PRCP          0
MAX           0
MIN           0
MEA           0
```

```
           SNF               0
           dtype: int64
```

In [16]:
```
df.columns[df.dtypes=='O']
```

Out[16]:
```
Index(['Date', 'Precip', 'Snowfall', 'PRCP', 'SNF'], dtype='object')
```

In [17]:
```
df.drop('Date',axis=1,inplace=True)
```

In [18]:
```
col = df.columns[df.dtypes=='O']
```

In [19]:
```
col
```

Out[19]:
```
Index(['Precip', 'Snowfall', 'PRCP', 'SNF'], dtype='object')
```

In [20]:
```
df = df.replace(['T'], 0)
```

In [21]:
```
df = df.replace(["#VALUE!"], 0)
```

In [22]:
```
df[col] = df[col].apply(pd.to_numeric)
```

In [23]:
```
df['Snowfall'].unique()
```

Out[23]:
```
array([ 0.  , 22.86, 33.02, 30.48,  2.54,  5.08, 10.16,  7.62, 20.32,
       15.24, 17.78, 12.7 , 25.4 , 27.94, 43.18, 38.1 , 45.72, 53.34,
       58.42, 66.04, 76.2 , 81.28, 73.66, 63.5 , 50.8 , 48.26, 60.96,
       55.88, 35.56, 78.74, 40.64, 86.36, 83.82, 68.58])
```

Using Vaex library to handle large dataset

In [24]:
```
vx_df=vx.from_pandas(df)
```

In [25]:
```
type(vx_df)
```

Out[25]:
```
vaex.dataframe.DataFrameLocal
```

In [26]:
```
vx_df.export_hdf5('df.hdf5')
```

In [27]:
```
vx_df = vx.open("df.hdf5")
```

In [28]:
```
vx_df.head()
```

Out[28]:

#	STA	Precip	MaxTemp	MinTemp	MeanTemp	Snowfall	YR	MO	DA	PRCP	MAX	MIN	MEA	SNF
0	10001	1.016	25.5556	22.2222	23.8889	0	42	7	1	0.04	78	72	75	0
1	10001	0	28.8889	21.6667	25.5556	0	42	7	2	0	84	71	78	0

#	STA	Precip	MaxTemp	MinTemp	MeanTemp	Snowfall	YR	MO	DA	PRCP	MAX	MIN	MEA	SNF
2	10001	2.54	26.1111	22.2222	24.4444	0	42	7	3	0.1	79	72	76	0
3	10001	2.54	26.6667	22.2222	24.4444	0	42	7	4	0.1	80	72	76	0
4	10001	0	26.6667	21.6667	24.4444	0	42	7	5	0	80	71	76	0
5	10001	0	26.6667	21.6667	24.4444	0	42	7	6	0	80	71	76	0
6	10001	0	28.3333	22.7778	25.5556	0	42	7	7	0	83	73	78	0
7	10001	3.556	26.6667	22.2222	24.4444	0	42	7	8	0.14	80	72	76	0
8	10001	0	27.2222	22.7778	25	0	42	7	9	0	81	73	77	0
9	10001	3.556	25.5556	21.6667	23.3333	0	42	7	10	0.14	78	71	74	0

In [29]:
```python
vx_df = vx_df.shuffle()
```

In [30]:
```python
df_train,df_test=vx_df.ml.train_test_split(test_size=0.2)
```

In [31]:
```python
df_train.shape
```

Out[31]:
```
(95232, 14)
```

In [32]:
```python
df_test.shape
```

Out[32]:
```
(23808, 14)
```

Model preparation

In [33]:
```python
from vaex.ml.sklearn import IncrementalPredictor
```

In [34]:
```python
from sklearn.linear_model import SGDRegressor
```

In [35]:
```python
features = ['STA', 'Precip', 'MaxTemp', 'MinTemp', 'MeanTemp', 'YR','MO', 'DA', 'PRCP', 'N
```

In [36]:
```python
target = 'Snowfall'
```

In [37]:
```python
model = SGDRegressor()
```

In [38]:
```python
vx_model = IncrementalPredictor(features=features, target=target, model=model, batch_size=
```

In [39]:
```python
vx_model.fit(df=df_train,progress='widget')
```

In [40]:
```python
df_test = vx_model.transform(df_test)
```

```
In [41]:   df_test.head()
```

Out[41]:

#	STA	Precip	MaxTemp	MinTemp	MeanTemp	Snowfall	YR	MO	DA	PRCP	MAX	MIN	MEA	SNF	pred
0	32701	2.54	32.7778	15	23.8889	0	44	3	20	0.1	91	59	75	0	8.25893
1	22504	8.382	27.7778	21.6667	24.4444	0	42	12	7	0.33	82	71	76	0	5.68851
2	80411	16.51	26.6667	22.2222	24.4444	0	45	5	24	0.65	80	72	76	0	2.02868
3	10704	38.862	29.4444	22.7778	26.1111	0	43	11	19	1.53	85	73	79	0	2.71454
4	33003	0	20	1.66667	11.1111	0	45	2	1	0	68	35	52	0	8.33033
5	12801	5.588	32.7778	22.7778	27.7778	0	45	8	29	0.22	91	73	82	0	3.2438
6	16407	0	3.33333	-0.555556	1.11111	2.54	45	11	3	0	38	31	34	0.1	4.14337
7	32805	0	27.2222	9.44444	18.3333	0	44	12	14	0	81	49	65	0	8.28313
8	13201	0	20.5556	15	17.7778	0	45	9	17	0	69	59	64	0	3.34102
9	16201	6.858	6.11111	1.11111	3.33333	2.54	44	12	16	0.27	43	34	38	0.1	4.09227

```
In [42]:   from sklearn.metrics import r2_score,mean_absolute_error
```

```
In [43]:   print(r2_score(df_test['Snowfall'].values,df_test['prediction'].values))
```

```
-1.3397696395944782e+35
```

```
In [44]:   print(mean_absolute_error(df_test['Snowfall'].values,df_test['prediction'].values))
```

```
7.543213651628429e+17
```

This method can handle large datasets, by using your local machine. We can use Out of core ML methods using Vaex libraray.

```
In [45]:   !pip install -U notebook-as-pdf
```

```
Requirement already satisfied: notebook-as-pdf in c:\ana\anaconda3\lib\site-packages (0.5.
0)
Requirement already satisfied: PyPDF2 in c:\ana\anaconda3\lib\site-packages (from notebook
-as-pdf) (3.0.1)
Requirement already satisfied: pyppeteer in c:\ana\anaconda3\lib\site-packages (from noteb
ook-as-pdf) (1.0.2)
Requirement already satisfied: nbconvert in c:\ana\anaconda3\lib\site-packages (from noteb
ook-as-pdf) (6.1.0)
Requirement already satisfied: bleach in c:\ana\anaconda3\lib\site-packages (from nbconver
t->notebook-as-pdf) (4.0.0)
Requirement already satisfied: pygments>=2.4.1 in c:\ana\anaconda3\lib\site-packages (from
nbconvert->notebook-as-pdf) (2.14.0)
Requirement already satisfied: testpath in c:\ana\anaconda3\lib\site-packages (from nbconv
ert->notebook-as-pdf) (0.5.0)
Requirement already satisfied: nbformat>=4.4 in c:\ana\anaconda3\lib\site-packages (from n
bconvert->notebook-as-pdf) (5.1.3)
Requirement already satisfied: jinja2>=2.4 in c:\ana\anaconda3\lib\site-packages (from nbc
onvert->notebook-as-pdf) (2.11.3)
Requirement already satisfied: jupyterlab-pygments in c:\ana\anaconda3\lib\site-packages
(from nbconvert->notebook-as-pdf) (0.1.2)
Requirement already satisfied: defusedxml in c:\ana\anaconda3\lib\site-packages (from nbco
nvert->notebook-as-pdf) (0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\ana\anaconda3\lib\site-package
```

s (from nbconvert->notebook-as-pdf) (0.5.3)
Requirement already satisfied: entrypoints>=0.2.2 in c:\ana\anaconda3\lib\site-packages (f
rom nbconvert->notebook-as-pdf) (0.3)
Requirement already satisfied: traitlets>=5.0 in c:\ana\anaconda3\lib\site-packages (from
nbconvert->notebook-as-pdf) (5.1.0)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\ana\anaconda3\lib\site-packages (fr
om nbconvert->notebook-as-pdf) (0.8.4)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\ana\anaconda3\lib\site-packages
(from nbconvert->notebook-as-pdf) (1.4.3)
Requirement already satisfied: jupyter-core in c:\ana\anaconda3\lib\site-packages (from nb
convert->notebook-as-pdf) (4.8.1)
Requirement already satisfied: typing_extensions>=3.10.0.0 in c:\ana\anaconda3\lib\site-pa
ckages (from PyPDF2->notebook-as-pdf) (4.5.0)
Requirement already satisfied: websockets<11.0,>=10.0 in c:\ana\anaconda3\lib\site-package
s (from pyppeteer->notebook-as-pdf) (10.4)
Requirement already satisfied: urllib3<2.0.0,>=1.25.8 in c:\ana\anaconda3\lib\site-package
s (from pyppeteer->notebook-as-pdf) (1.26.7)
Requirement already satisfied: pyee<9.0.0,>=8.1.0 in c:\ana\anaconda3\lib\site-packages (f
rom pyppeteer->notebook-as-pdf) (8.2.2)
Requirement already satisfied: importlib-metadata>=1.4 in c:\ana\anaconda3\lib\site-packag
es (from pyppeteer->notebook-as-pdf) (4.8.1)
Requirement already satisfied: appdirs<2.0.0,>=1.4.3 in c:\ana\anaconda3\lib\site-packages
(from pyppeteer->notebook-as-pdf) (1.4.4)
Requirement already satisfied: tqdm<5.0.0,>=4.42.1 in c:\ana\anaconda3\lib\site-packages
(from pyppeteer->notebook-as-pdf) (4.62.3)
Requirement already satisfied: certifi>=2021 in c:\ana\anaconda3\lib\site-packages (from p
yppeteer->notebook-as-pdf) (2021.10.8)
Requirement already satisfied: zipp>=0.5 in c:\ana\anaconda3\lib\site-packages (from impor
tlib-metadata>=1.4->pyppeteer->notebook-as-pdf) (3.6.0)
Requirement already satisfied: MarkupSafe>=0.23 in c:\ana\anaconda3\lib\site-packages (fro
m jinja2>=2.4->nbconvert->notebook-as-pdf) (1.1.1)
Requirement already satisfied: async-generator in c:\ana\anaconda3\lib\site-packages (from
nbclient<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (1.10)
Requirement already satisfied: jupyter-client>=6.1.5 in c:\ana\anaconda3\lib\site-packages
(from nbclient<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (6.1.12)
Requirement already satisfied: nest-asyncio in c:\ana\anaconda3\lib\site-packages (from nb
client<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (1.5.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\ana\anaconda3\lib\site-packag
es (from nbformat>=4.4->nbconvert->notebook-as-pdf) (3.2.0)
Requirement already satisfied: ipython-genutils in c:\ana\anaconda3\lib\site-packages (fro
m nbformat>=4.4->nbconvert->notebook-as-pdf) (0.2.0)
Requirement already satisfied: colorama in c:\ana\anaconda3\lib\site-packages (from tqdm<
5.0.0,>=4.42.1->pyppeteer->notebook-as-pdf) (0.4.4)
Requirement already satisfied: six>=1.9.0 in c:\ana\anaconda3\lib\site-packages (from blea
ch->nbconvert->notebook-as-pdf) (1.16.0)
Requirement already satisfied: packaging in c:\ana\anaconda3\lib\site-packages (from bleac
h->nbconvert->notebook-as-pdf) (23.0)
Requirement already satisfied: webencodings in c:\ana\anaconda3\lib\site-packages (from bl
each->nbconvert->notebook-as-pdf) (0.5.1)
Requirement already satisfied: pywin32>=1.0 in c:\ana\anaconda3\lib\site-packages (from ju
pyter-core->nbconvert->notebook-as-pdf) (228)
Requirement already satisfied: setuptools in c:\ana\anaconda3\lib\site-packages (from json
schema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert->notebook-as-pdf) (58.0.4)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\ana\anaconda3\lib\site-packages (f
rom jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert->notebook-as-pdf) (0.18.0)
Requirement already satisfied: attrs>=17.4.0 in c:\ana\anaconda3\lib\site-packages (from j
sonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert->notebook-as-pdf) (21.2.0)
Requirement already satisfied: tornado>=4.1 in c:\ana\anaconda3\lib\site-packages (from ju
pyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (6.1)
Requirement already satisfied: pyzmq>=13 in c:\ana\anaconda3\lib\site-packages (from jupyt
er-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (22.2.1)
Requirement already satisfied: python-dateutil>=2.1 in c:\ana\anaconda3\lib\site-packages
(from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert->notebook-as-pdf) (2.8.2)
```

In [47]:
```python
!pyppeteer-install
```

chromium is already installed.

In [ ]:

chromium is already installed.