unswDataImportSingle

Data import test on singular TEBT file, using the 01/01/2018 dated file. We are forced to skip the first 6 rows (which is when the actual data starts), and to ignore the headers. The original headers in the file are partially uncomaptable due to an encoding issue (pandas doesn't like the degree symbol in the encoding it chose) and despite a true solution being avaialble, the workaround is to specify all the headings hardcoded. Once we have all the data we can rename the headers to the original headers by enumerating the header list.

Beware, the headers string is very long and is left in for reproducibility purposes. It shall be truncated later.

In [2]:

```
import pandas as pd
file_name = r"C:\Users\Clairvoyant Cabbage\Documents\PythonProject\Thesis\UNSWData\2018
-Array\2018-01-01.csv"
#file_name = r"C:\Users\Clairvoyant Cabbage\Documents\PythonProject\Thesis\84-Site_12-B
P-Solar.csv"
df1 = pd.read_csv(file_name, delimiter=";", header=None, skiprows=6)
headers = "TimeStamp;ExlSolIrr;IntSolIrr;SMA-h-On;TmpAmb C;TmpMdul C;WindVel km/h;A.Ms.
Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;G
M.TotWhOut;GridMs.A.phsA;GridMs.A.phsB;GridMs.A.phsC;GridMs.Hz;GridMs.PhV.phsA;GridMs.P
hV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;GridMs.TotVA;GridMs.TotVAr;GridMs.VA.phsA;GridM
s.VA.phsB;GridMs.VA.phsC;GridMs.VAr.phsA;GridMs.VAr.phsB;GridMs.VAr.phsC;GridMs.W.phsA;
GridMs.W.phsB;GridMs.W.phsC;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtC
ntUsr;Op.EvtNo;Op.GriSwStt;Op.Health;Op.Prio;Op.TmsRmg;Pac;PCM-DigInStt;PlntCtl.Stt;Ser
ial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Am
p;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.
PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.T
otTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Am
p;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vo
l;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.Ph
V.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUs
r;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.W
att;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.M
s.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.To
tPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.Gri
SwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.
Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Tota
l;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimS
tt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pa
c;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A
4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridM
s.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mod
e;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Ser
ial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Am
p;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.
PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.T
otTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Am
p;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vo
l;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.Ph
V.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUs
r;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.W
att;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.M
s.Amp;Error;E-Total;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.To
tPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.Gri
SwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.
Ms.Amp;A3.Ms.Amp;A4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Tota
l;GridMs.Hz;GridMs.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimS
tt;InvCtl.Stt;Mode;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pa
c;PlntCtl.Stt;Serial Number;A.Ms.Amp;A.Ms.Vol;A.Ms.Watt;A1.Ms.Amp;A2.Ms.Amp;A3.Ms.Amp;A
4.Ms.Amp;A5.Ms.Amp;B.Ms.Amp;B.Ms.Vol;B.Ms.Watt;B1.Ms.Amp;Error;E-Total;GridMs.Hz;GridM
s.PhV.phsA;GridMs.PhV.phsB;GridMs.PhV.phsC;GridMs.TotPFPrc;Inv.TmpLimStt;InvCtl.Stt;Mod
e;Mt.TotOpTmh;Mt.TotTmh;Op.EvtCntUsr;Op.EvtNo;Op.GriSwStt;Op.TmsRmg;Pac;PlntCtl.Stt;Ser
ial Number"
headers = dict(enumerate(headers.split(';')))
df1 = df1.rename(columns = headers)
print(df1)
```

```
    TimeStamp  ExlSolIrr  IntSolIrr  SMA-h-On  TmpAmb C  TmpMdul C  \
0       00:00          0        0.0  34654.71     22.33      22.23
1       00:05          0        0.0  34654.79     22.33      22.23
2       00:10          0        0.0  34654.88     22.31      22.23
3       00:15          0        0.0  34654.96     22.33      22.23
4       00:20          0        0.0  34655.05     22.27      22.13
..        ...        ...        ...       ...       ...        ...
278     23:10          0        0.0  34677.69     22.43      22.13
279     23:15          0        0.0  34677.78     22.39      22.15
280     23:20          0        0.0  34677.86     22.29      21.85
281     23:25          0        0.0  34677.94     22.09      21.45
282     23:30          0        0.0  34678.03     22.05      21.15

     WindVel km/h  A.Ms.Amp  A.Ms.Vol  A.Ms.Watt  ...  Mode  Mt.TotOpTmh
\
0           11.86       NaN       NaN        NaN  ...   NaN          NaN
1            4.38       NaN       NaN        NaN  ...   NaN          NaN
2            5.08       NaN       NaN        NaN  ...   NaN          NaN
3            6.46       NaN       NaN        NaN  ...   NaN          NaN
4            4.34       NaN       NaN        NaN  ...   NaN          NaN
..            ...       ...       ...        ...  ...   ...          ...
278          6.04       NaN       NaN        NaN  ...   NaN          NaN
279          5.22       NaN       NaN        NaN  ...   NaN          NaN
280          4.38       NaN       NaN        NaN  ...   NaN          NaN
281          6.24       NaN       NaN        NaN  ...   NaN          NaN
282          3.68       NaN       NaN        NaN  ...   NaN          NaN

     Mt.TotTmh  Op.EvtCntUsr  Op.EvtNo  Op.GriSwStt  Op.TmsRmg  Pac  \
0          NaN           NaN       NaN          NaN        NaN  NaN
1          NaN           NaN       NaN          NaN        NaN  NaN
2          NaN           NaN       NaN          NaN        NaN  NaN
3          NaN           NaN       NaN          NaN        NaN  NaN
4          NaN           NaN       NaN          NaN        NaN  NaN
..         ...           ...       ...          ...        ...  ...
278        NaN           NaN       NaN          NaN        NaN  NaN
279        NaN           NaN       NaN          NaN        NaN  NaN
280        NaN           NaN       NaN          NaN        NaN  NaN
281        NaN           NaN       NaN          NaN        NaN  NaN
282        NaN           NaN       NaN          NaN        NaN  NaN

     PlntCtl.Stt  Serial Number
0            NaN            NaN
1            NaN            NaN
2            NaN            NaN
3            NaN            NaN
4            NaN            NaN
..           ...            ...
278          NaN            NaN
279          NaN            NaN
280          NaN            NaN
281          NaN            NaN
282          NaN            NaN

[283 rows x 362 columns]
```

Result

We can see from the initial result that we have roughly 24 hours of data available, with 362 columns of data points to choose from. Deciding which columns to keep may prove challenging.