seattle_jithin

August 6, 2018

```
In [1]: import pandas as pd
        import seaborn as sns
        from matplotlib import pyplot as plt

In [2]: df=pd.read_excel("Seattle_dataset.xlsx")

In [4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 365 entries, 0 to 364
Data columns (total 17 columns):
STATION         365 non-null object
STATION_NAME    365 non-null object
DATE            365 non-null int64
PRCP            365 non-null int64
SNWD            365 non-null int64
SNOW            365 non-null int64
TMAX            365 non-null int64
TMIN            365 non-null int64
AWND            365 non-null int64
WDF2            365 non-null int64
WDF5            365 non-null int64
WSF2            365 non-null int64
WSF5            365 non-null int64
WT01            365 non-null int64
WT05            365 non-null int64
WT02            365 non-null int64
WT03            365 non-null int64
dtypes: int64(15), object(2)
memory usage: 48.6+ KB


In [5]: df.isnull().sum()

Out[5]: STATION         0
        STATION_NAME    0
        DATE            0
        PRCP            0
```

```
SNWD            0
SNOW            0
TMAX            0
TMIN            0
AWND            0
WDF2            0
WDF5            0
WSF2            0
WSF5            0
WT01            0
WT05            0
WT02            0
WT03            0
dtype: int64
```

In [6]:
```
df.describe()
```

Out[6]:

	DATE	PRCP	SNWD	SNOW	TMAX
count	3.650000e+02	365.000000	365.000000	365.000000	365.000000
mean	2.014067e+07	33.775342	0.219178	-54.531507	169.958904
std	3.454755e+02	67.815757	4.187391	739.171802	72.687242
min	2.014010e+07	0.000000	0.000000	-9999.000000	-16.000000
25%	2.014040e+07	0.000000	0.000000	0.000000	111.000000
50%	2.014070e+07	0.000000	0.000000	0.000000	161.000000
75%	2.014100e+07	36.000000	0.000000	0.000000	222.000000
max	2.014123e+07	467.000000	80.000000	74.000000	356.000000

	TMIN	AWND	WDF2	WDF5	WSF2
count	365.000000	365.000000	365.000000	365.000000	365.000000
mean	86.624658	33.876712	181.657534	-227.232877	75.463014
std	49.746433	14.398895	101.995505	2027.917682	27.043146
min	-60.000000	6.000000	10.000000	-9999.000000	31.000000
25%	56.000000	24.000000	100.000000	110.000000	58.000000
50%	94.000000	31.000000	200.000000	210.000000	72.000000
75%	128.000000	42.000000	230.000000	240.000000	89.000000
max	178.000000	88.000000	360.000000	360.000000	183.000000

	WSF5	WT01	WT05	WT02	WT03
count	365.000000	365.000000	365.0	365.000000	365.000000
mean	-314.090411	-5834.616438	-9999.0	-9642.835616	-9916.808219
std	2008.055912	4936.446921	0.0	1855.862874	904.102747
min	-9999.000000	-9999.000000	-9999.0	-9999.000000	-9999.000000
25%	72.000000	-9999.000000	-9999.0	-9999.000000	-9999.000000
50%	94.000000	-9999.000000	-9999.0	-9999.000000	-9999.000000
75%	116.000000	1.000000	-9999.0	-9999.000000	-9999.000000
max	250.000000	1.000000	-9999.0	1.000000	1.000000

In [17]:
```
sns.pairplot(df[['PRCP','SNWD','SNOW']])
```

Out[17]: <seaborn.axisgrid.PairGrid at 0x7fbad997b278>