MEASURE ENERGY CONSUMPTION

PHASE 3

import matplotlib.pyplot as plt

import numpy as np

import os

import pandas as pd

import seaborn as sns

plt.style.use('ggplot')

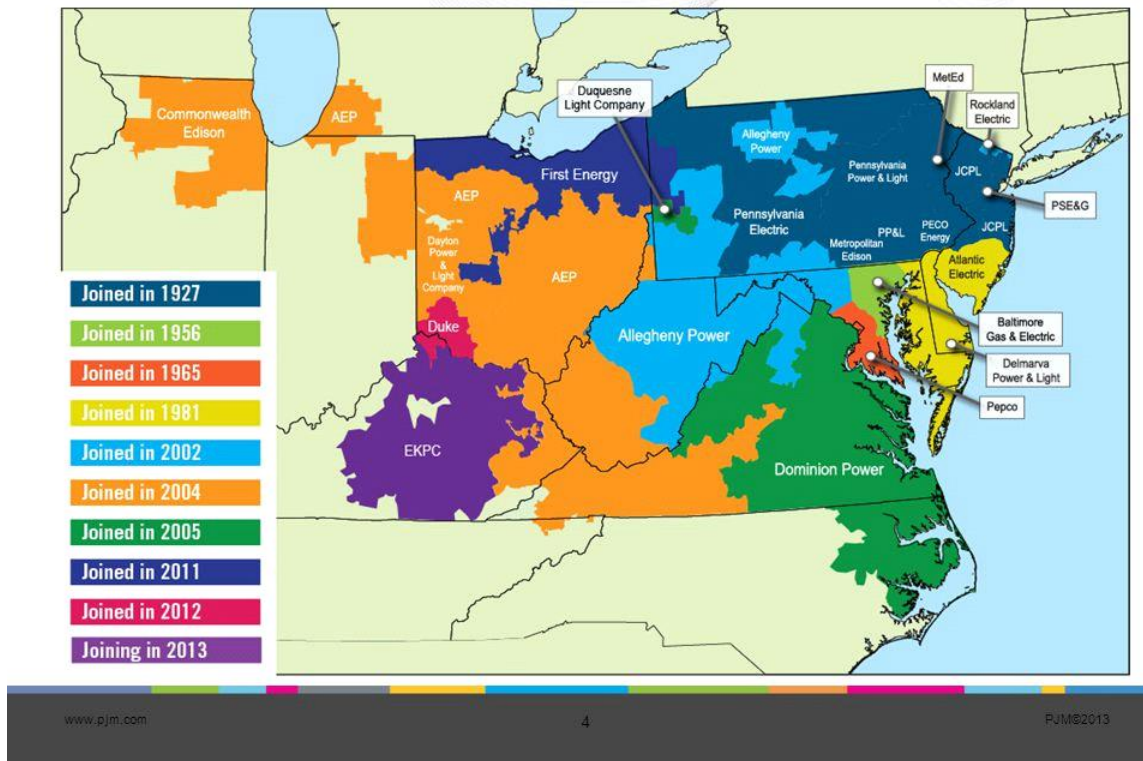df = pd.read_parquet('../input/est_hourly.paruqet')

Data index is the date/hour, columns are for different regions within PJM.

Regions joined at different times, so not all have data for all dates. Regions also split (PJM_Load split to East and West)

from IPython.display import Image

Image(url= "http://slideplayer.com/4238181/14/images/4/PJM+Evolution.jpg")

<u>OUTPUT:</u>

df.head()

OUTPUT:

	AEP	COMED	DAYTON	DEOK	DOM	DUQ	EKPC	FE	NI	PJME	PJMW	PJM_Load
Datetime												
1998-12-31 01:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	29309.0
1998-12-31 02:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28236.0
1998-12-31 03:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	27692.0

1998-12-31 04:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	27596.0

1998-12-31 05:00:00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	27888.0

df.describe().T

OUTPUT:

	count	mean	std	min	25%	50%	75%	max
AEP	121273.0	15499.513717	2591.399065	9581.0	13630.0	15310.0	17200.00	25695.0
COMED	66497.0	11420.152112	2304.139517	7237.0	9780.0	11152.0	12510.00	23753.0
DAYTON	121275.0	2037.851140	393.403153	982.0	1749.0	2009.0	2279.00	3746.0
DEOK	57739.0	3105.096486	599.859026	907.0	2687.0	3013.0	3449.00	5445.0
DOM	116189.0	10949.203625	2413.946569	1253.0	9322.0	10501.0	12378.00	21651.0
DUQ	119068.0	1658.820296	301.740640	1014.0	1444.0	1630.0	1819.00	3054.0
EKPC	45334.0	1464.218423	378.868404	514.0	1185.0	1386.0	1699.00	3490.0
FE	62874.0	7792.159064	1331.268006	0.0	6807.0	7700.0	8556.00	14032.0
NI	58450.0	11701.682943	2371.498701	7003.0	9954.0	11521.0	12896.75	23631.0
PJME	145366.0	32080.222831	6464.012166	14544.0	27573.0	31421.0	35650.00	62009.0
PJMW	143206.0	5602.375089	979.142872	487.0	4907.0	5530.0	6252.00	9594.0
PJM_Load	32896.0	29766.427408	5849.769954	17461.0	25473.0	29655.0	33073.25	54030.0

_ = df['PJME'].plot.hist(figsize=(15, 5), bins=200, title='Distribution of PJME Load')
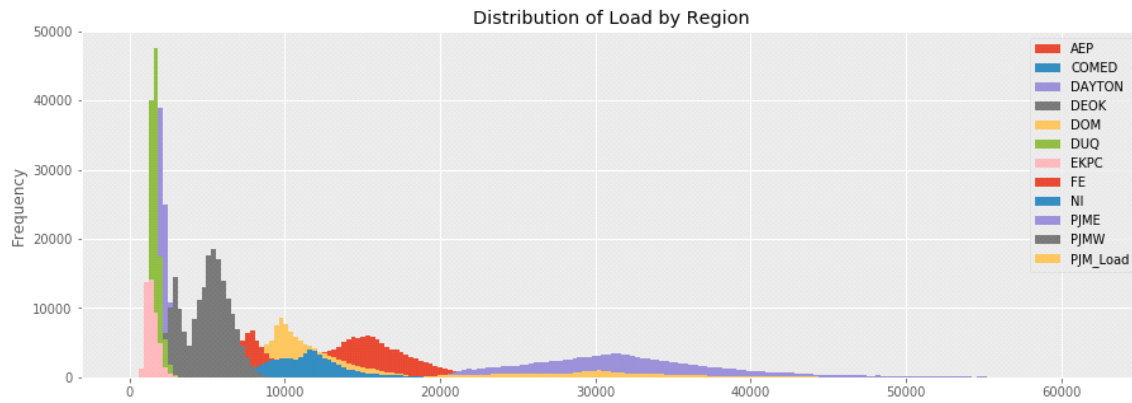
OUTPUT:

Distribution of PJME Load
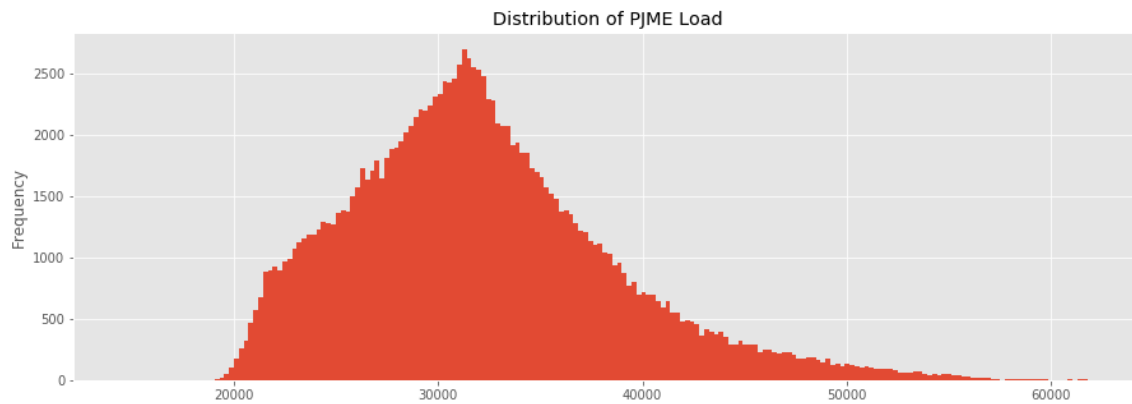
_ = df['DOM'].plot.hist(figsize=(15, 5), bins=200, title='Distribution of DOMINION Load')

OUTPUT:

Distribution of DOMINION Load
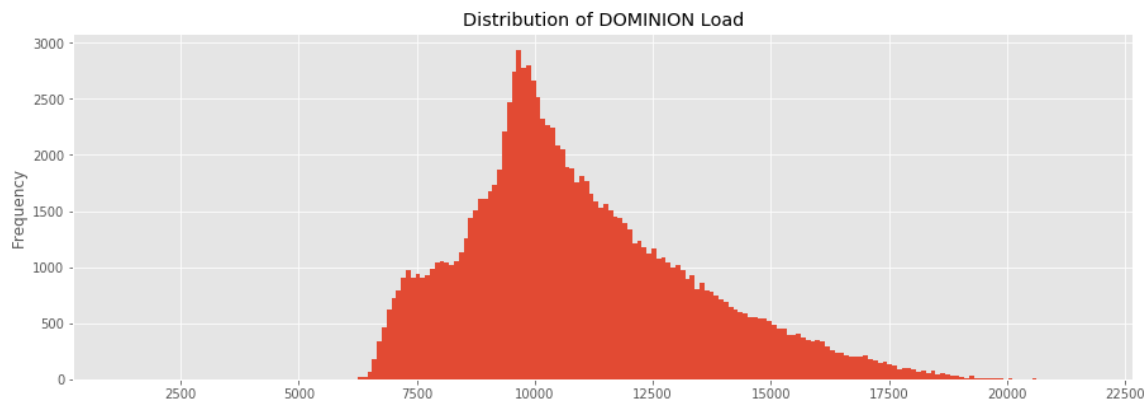
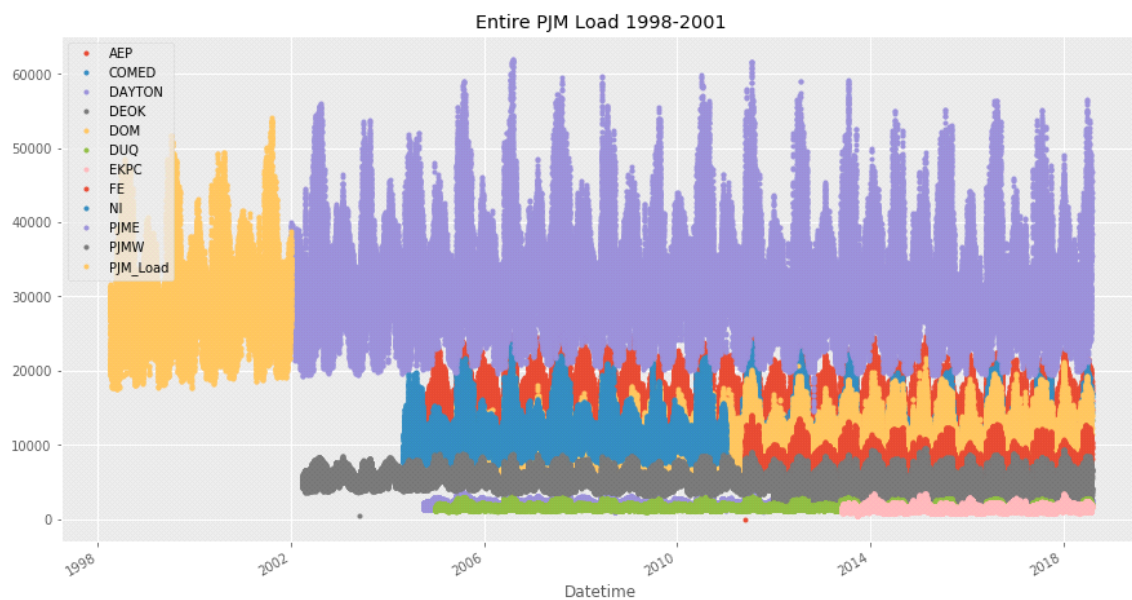_ = df.plot.hist(figsize=(15, 5), bins=200, title='Distribution of Load by Region')

OUTPUT:

Plot Time Series

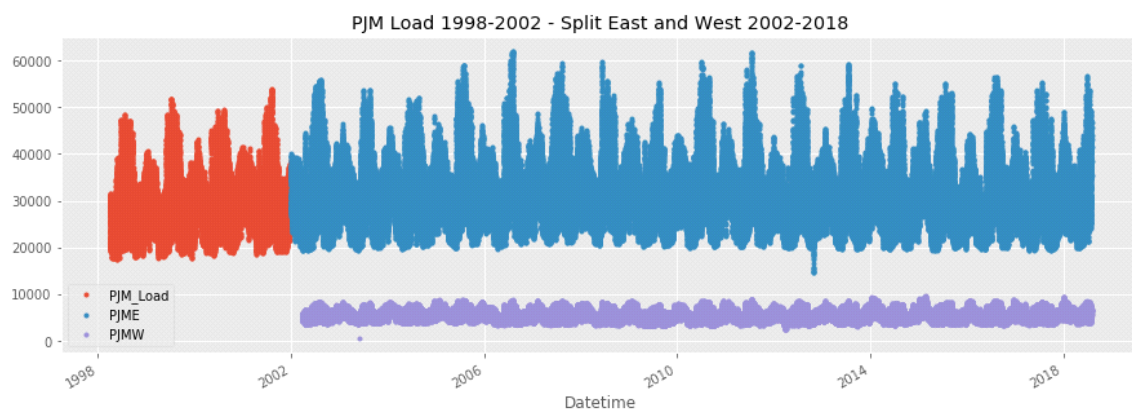plot = df.plot(style='.', figsize=(15, 8), title='Entire PJM Load 1998-2001')

OUTPUT:

Plotting Regions

```
_ = df[['PJM_Load','PJME','PJMW']] ¥

    .plot(style='.', figsize=(15, 5), title='PJM Load 1998-2002 - Split East and West 2002-2018')
```

OUTPUT:

Summer Demand vs Winter Demand

Note the dips mid-day in the winter months. Conversely in summer months the daily load is more bell shaped. This is due to high mid-day energy consumtion by air conditioning. In winter months people tend to use less energy mid-day.