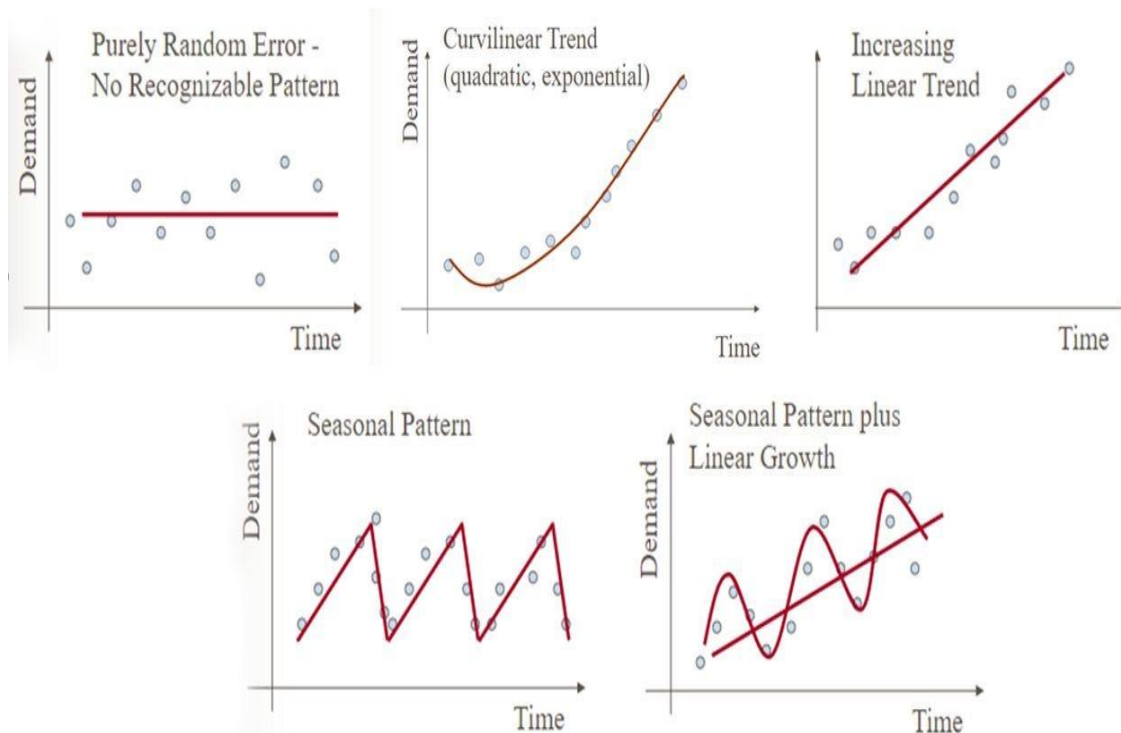


PHASE 4: DEVELOPMENT PART 2

Types of Time Series Data



Data:

The data we will be using is hourly power consumption data from PJM. Energy consumption has some unique characteristics. It will be interesting to see how prophet picks them up.

Pulling the PJM East which has data from 2002-2018 for the entire east region.

INPUT:

```
pjme = pd.read_csv('../input/PJME_hourly.csv',  
                    index_col=[0],  
                    parse_dates=[0])  
pjme.head()
```

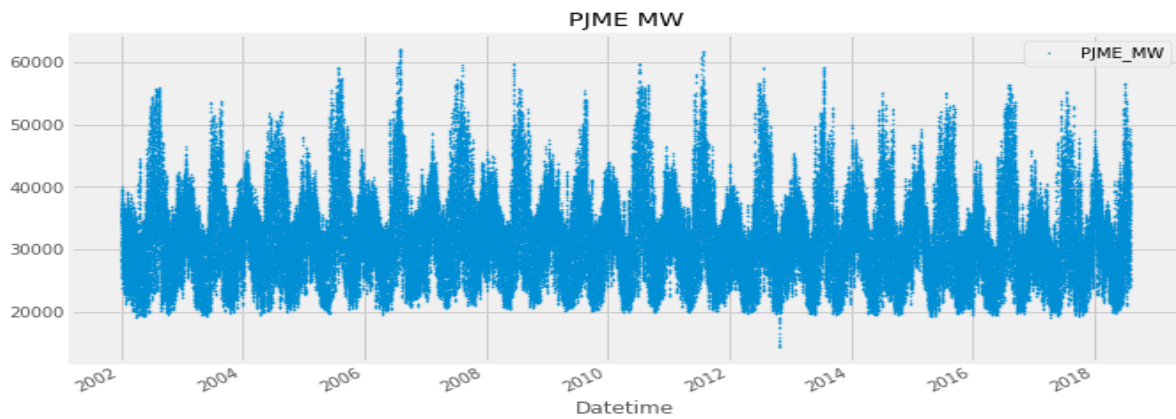
Output:

	PJME_MW
Datetime	
2002-12-31 01:00:00	26498.0
2002-12-31 02:00:00	25147.0
2002-12-31 03:00:00	24574.0
2002-12-31 04:00:00	24393.0
2002-12-31 05:00:00	24860.0

INPUT:

```
color_pal = sns.color_palette()
pjme.plot(style='.',
          figsize=(10, 5),
          ms=1,
          color=color_pal[0],
          title='PJME MW')
plt.show()
```

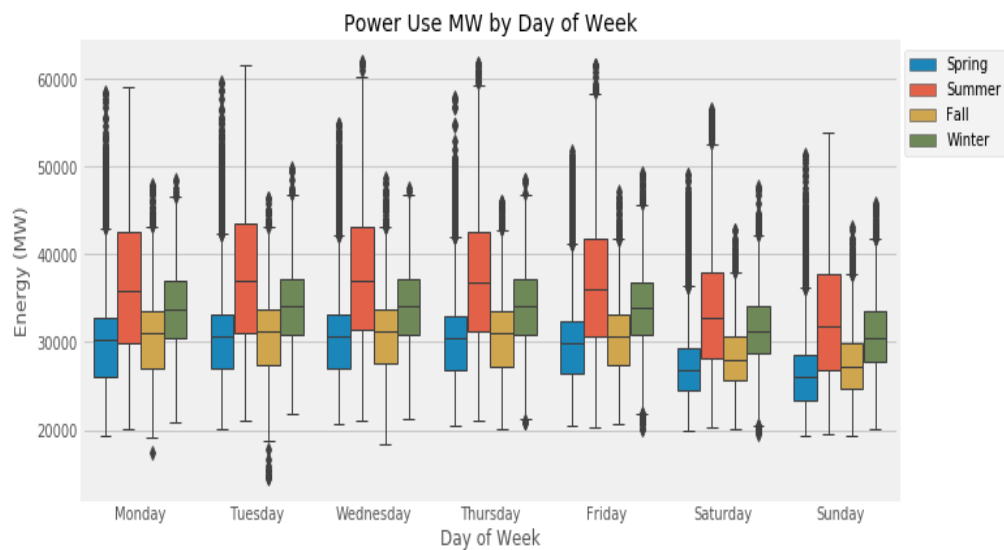
OUTPUT:



INPUT:

```
fig, ax = plt.subplots(figsize=(10, 5))
sns.boxplot(data=features_and_target.dropna(),
            x='weekday',
            y='PJME_MW',
            hue='season',
            ax=ax,
            linewidth=1)
ax.set_title('Power Use MW by Day of Week')
ax.set_xlabel('Day of Week')
ax.set_ylabel('Energy (MW)')
ax.legend(bbox_to_anchor=(1, 1))
plt.show()
```

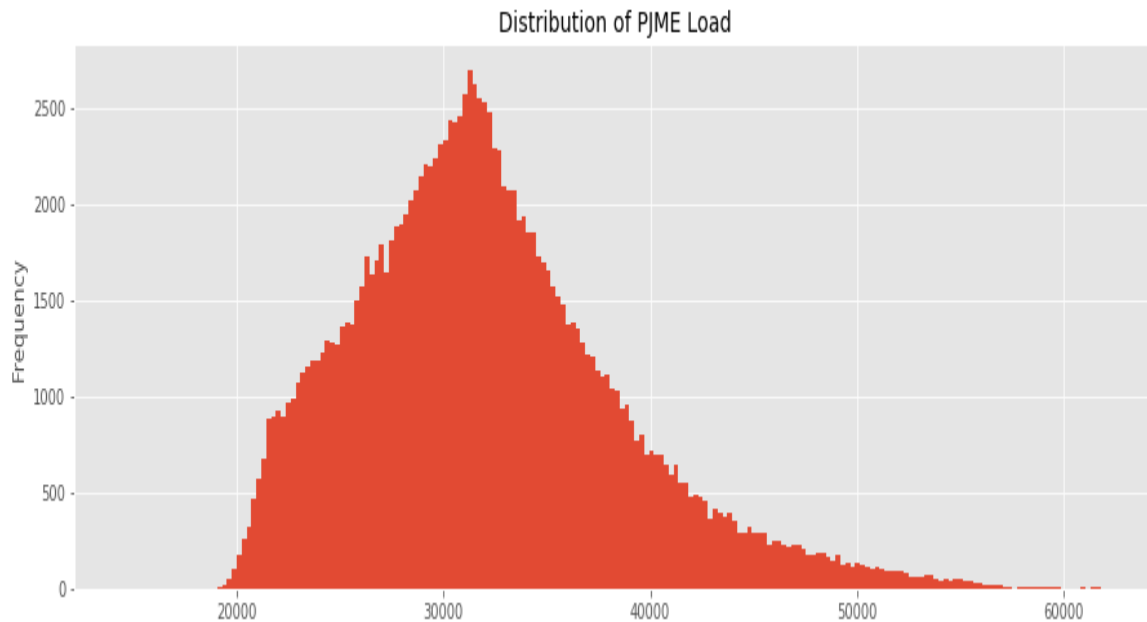
OUTPUT:



INPUT:

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

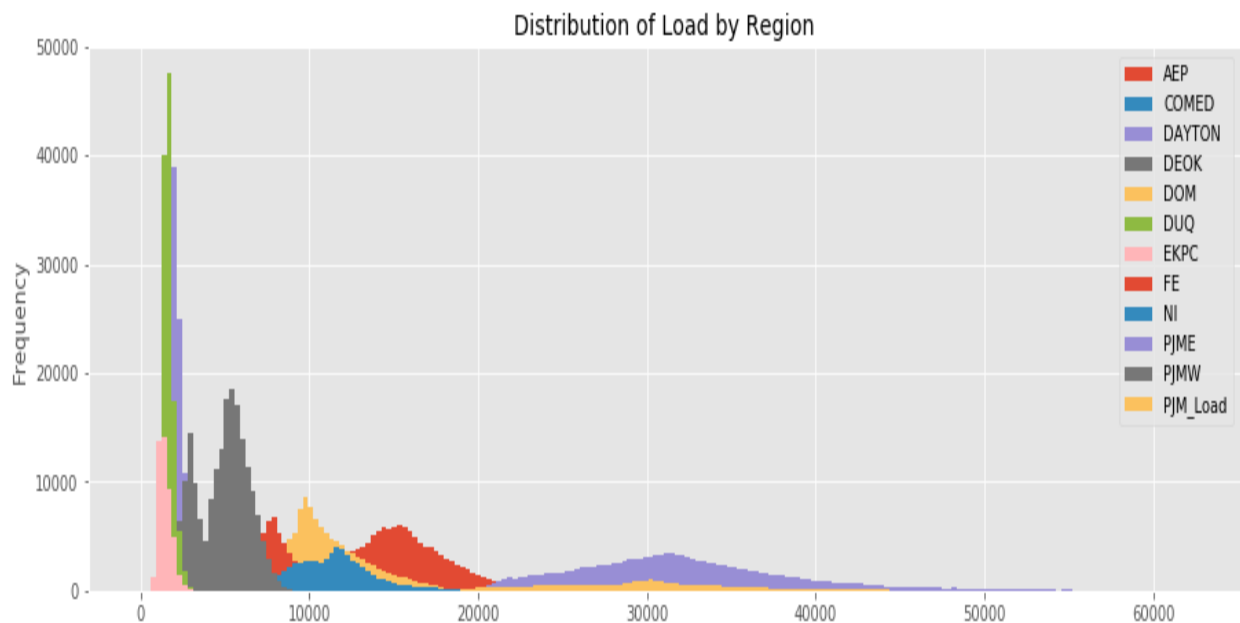
OUTPUT:



INPUT:

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

OUTPUT:

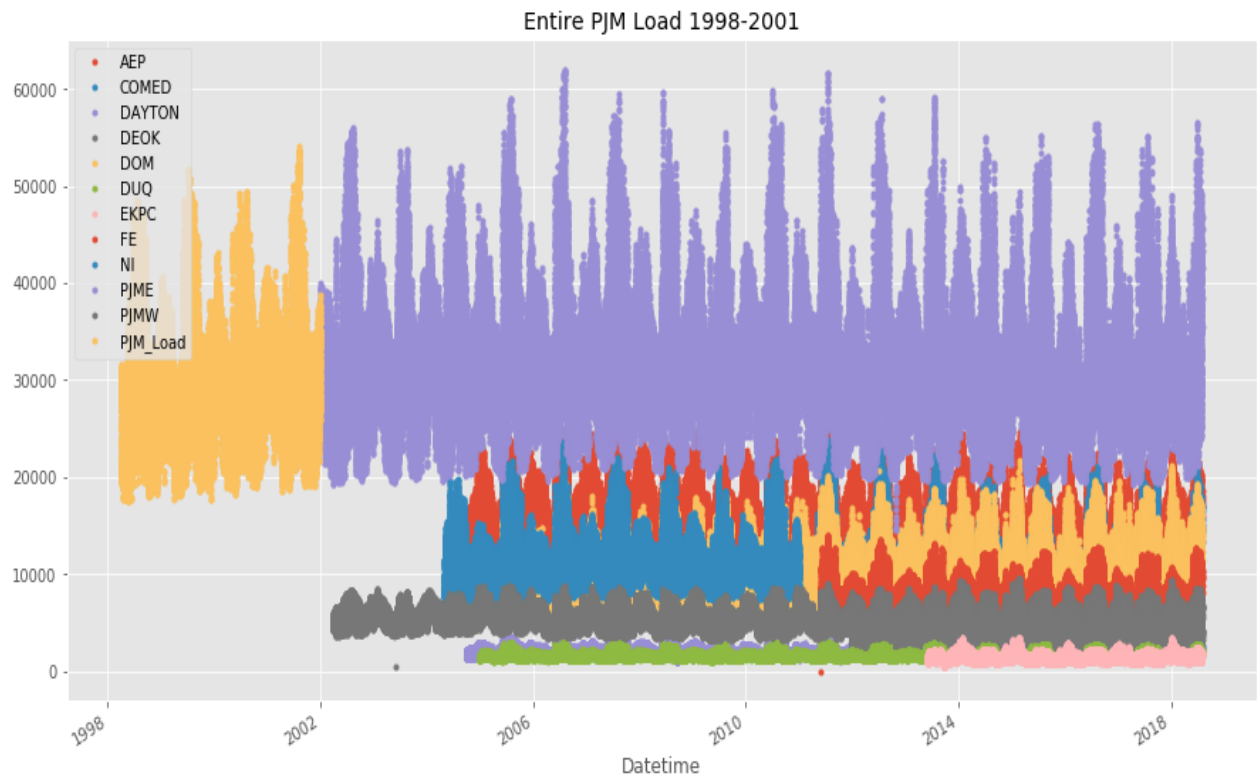


Plot Time Series

INPUT:

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

OUTPUT:



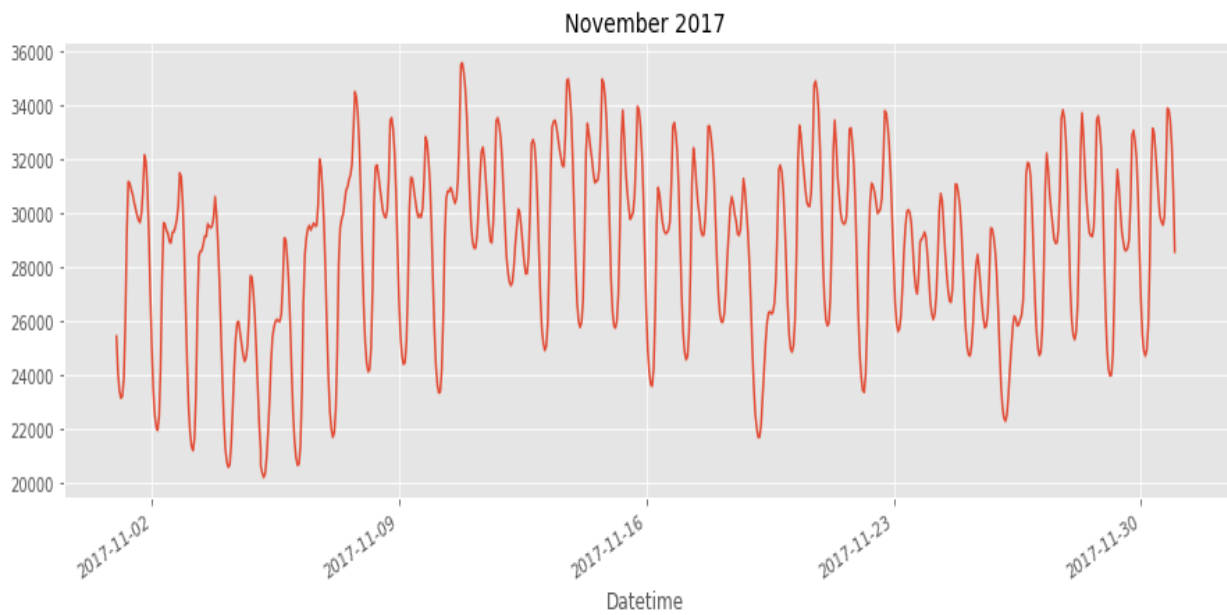
Summer Demand vs Winter Demand

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 consumption by air conditioning. In winter months people tend to use less energy mid-day.

INPUT:

```
_ = df['PJME'].loc[(df['PJME'].index >= '2017-11-01') &  
                  (df['PJME'].index < '2017-12-01')] \  
    .plot(figsize=(15, 5), title = 'November 2017')
```

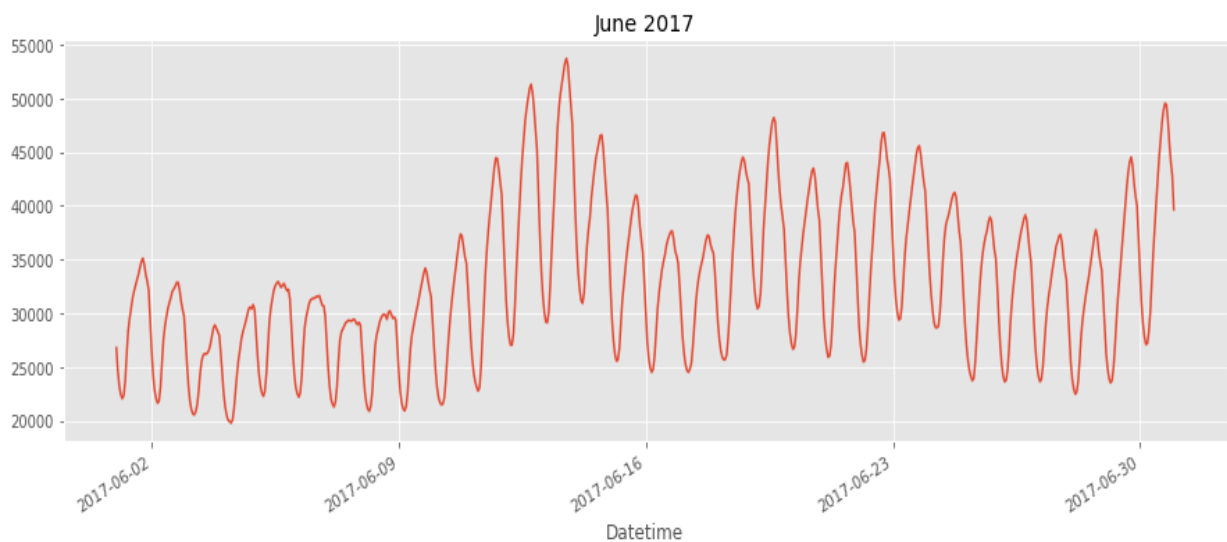
OUTPUT:



INPUT:

```
= df['PJME'].loc[(df['PJME'].index >= '2017-06-01') &  
                 (df['PJME'].index < '2017-07-01')] \  
   .plot(figsize=(15, 5), title = 'June 2017')
```

OUTPUT:



Creating Time Series Features

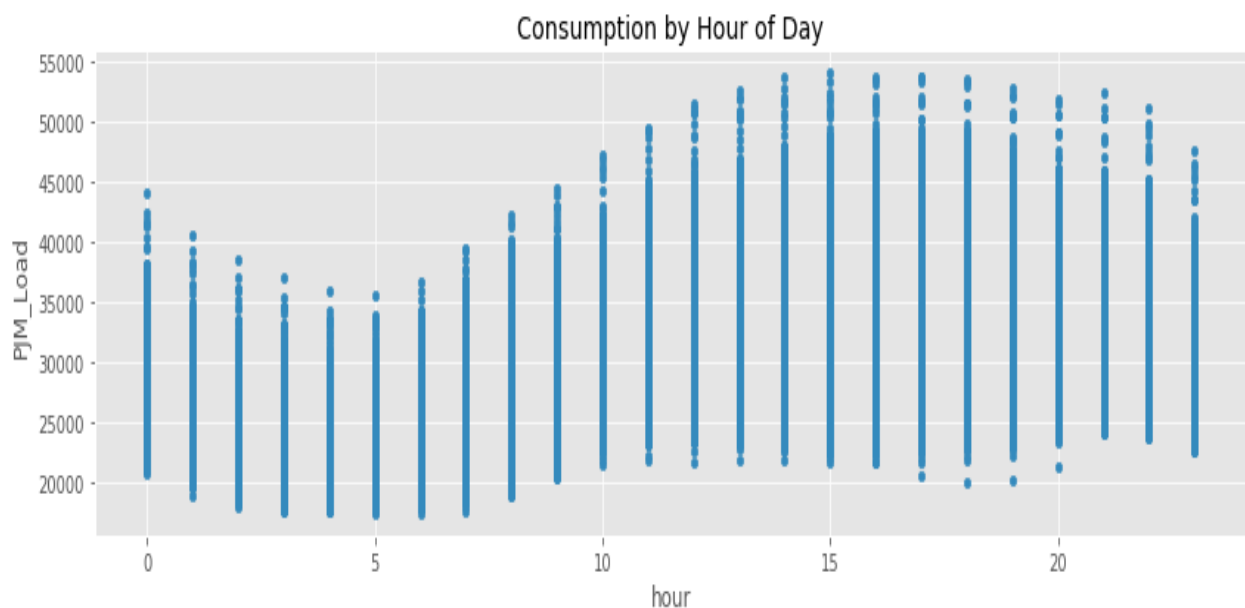
INPUT:

```
df['dow'] = df.index.dayofweek
df['doy'] = df.index.dayofyear
df['year'] = df.index.year
df['month'] = df.index.month
df['quarter'] = df.index.quarter
df['hour'] = df.index.hour
df['weekday'] = df.index.weekday_name
df['woy'] = df.index.weekofyear
df['dom'] = df.index.day # Day of Month
df['date'] = df.index.date
```

INPUT:

```
_ = df[['PJM_Load', 'hour']].plot(x='hour',
                                  y='PJM_Load',
                                  kind='scatter',
                                  figsize=(14,4),
                                  title='Consumption by Hour of Day
ay')
```

OUTPUT:



Note Saturday and Sunday demand is much less than during a work week. This is also true for holidays.

INPUT:

```
df.pivot_table(index=df['hour'],
                columns='weekday',
                values='PJME',
                aggfunc='sum').plot(figsize=(15,4),
                title='PJM East - Daily Trends')
```

OUTPUT:

