In [42]:

```python
#import libraries
import xgboost as xgb
import itertools
from prophet import Prophet
import pandas as pd
import numpy as np
```

In [2]:

```python
data = pd.read_csv("D:/Forecasting Models and Time Series for Business in Python/Dai
data.head()
```

Out[2]:

|  | instant | dteday | season | yr | mnth | holiday | weekday | workingday | weathersit | temp |  |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 2011-01-01 | 1 | 0 | 1 | 0 | 6 | 0 | 2 | 0.344167 | ( |
| **1** | 2 | 2011-01-02 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0.363478 | ( |
| **2** | 3 | 2011-01-03 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0.196364 | ( |
| **3** | 4 | 2011-01-04 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 0.200000 | ( |
| **4** | 5 | 2011-01-05 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 0.226957 | ( |

In [32]:

```python
#select variables
dataset = data.loc[:, ["dteday","cnt", "holiday", "workingday", "weathersit",
                       "temp", "atemp", "hum", "windspeed"]]
dataset.head()
```

Out[32]:

|  | dteday | cnt | holiday | workingday | weathersit | temp | atemp | hum | windspeed |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 2011-01-01 | 985 | 0 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 |
| **1** | 2011-01-02 | 801 | 0 | 0 | 2 | 0.363478 | 0.353739 | 0.696087 | 0.248539 |
| **2** | 2011-01-03 | 1349 | 0 | 1 | 1 | 0.196364 | 0.189405 | 0.437273 | 0.248309 |
| **3** | 2011-01-04 | 1562 | 0 | 1 | 1 | 0.200000 | 0.212122 | 0.590435 | 0.160296 |
| **4** | 2011-01-05 | 1600 | 0 | 1 | 1 | 0.226957 | 0.229270 | 0.436957 | 0.186900 |

In [33]:

```
#Date variable
dataset.dteday = pd.to_datetime(dataset.dteday,
                                 format = "%Y-%m-%d")
dataset.dteday
```

Out[33]:

```
0      2011-01-01
1      2011-01-02
2      2011-01-03
3      2011-01-04
4      2011-01-05
          ...
726    2012-12-27
727    2012-12-28
728    2012-12-29
729    2012-12-30
730    2012-12-31
Name: dteday, Length: 731, dtype: datetime64[ns]
```

In [34]:

```
#renaming variable
dataset = dataset.rename(columns = {'cnt' : 'y'})
dataset = dataset.rename(columns = {'dteday' : 'ds'})
dataset.head()
```

Out[34]:

| | ds | y | holiday | workingday | weathersit | temp | atemp | hum | windspeed |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 2011-01-01 | 985 | 0 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 |
| **1** | 2011-01-02 | 801 | 0 | 0 | 2 | 0.363478 | 0.353739 | 0.696087 | 0.248539 |
| **2** | 2011-01-03 | 1349 | 0 | 1 | 1 | 0.196364 | 0.189405 | 0.437273 | 0.248309 |
| **3** | 2011-01-04 | 1562 | 0 | 1 | 1 | 0.200000 | 0.212122 | 0.590435 | 0.160296 |
| **4** | 2011-01-05 | 1600 | 0 | 1 | 1 | 0.226957 | 0.229270 | 0.436957 | 0.186900 |

In [35]:

```
holiday_dates = dataset[dataset.holiday == 1].ds
holidays = pd.DataFrame({'holiday' : 'holi',
                         'ds': pd.to_datetime(holiday_dates),
                         'lower_window': -3,
                         'upper_window': 1})
holidays.head()
```

Out[35]:

| | holiday | ds | lower_window | upper_window |
|---|---|---|---|---|
| **16** | holi | 2011-01-17 | -3 | 1 |
| **51** | holi | 2011-02-21 | -3 | 1 |
| **104** | holi | 2011-04-15 | -3 | 1 |
| **149** | holi | 2011-05-30 | -3 | 1 |
| **184** | holi | 2011-07-04 | -3 | 1 |

In [36]:

```
#removing holiday column
dataset = dataset.drop(columns = "holiday")
dataset.head(1)
```

Out[36]:

| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed |
|---|---|---|---|---|---|---|---|---|
| **0** | 2011-01-01 | 985 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 |

In [37]:

```
#Training and test set
test_days = 31
training_set = dataset.iloc[:-test_days, :]
test_set = dataset.iloc[-test_days:, :]
test_set.tail()
```

Out[37]:

| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed |
|---|---|---|---|---|---|---|---|---|
| **726** | 2012-12-27 | 2114 | 1 | 2 | 0.254167 | 0.226642 | 0.652917 | 0.350133 |
| **727** | 2012-12-28 | 3095 | 1 | 2 | 0.253333 | 0.255046 | 0.590000 | 0.155471 |
| **728** | 2012-12-29 | 1341 | 0 | 2 | 0.253333 | 0.242400 | 0.752917 | 0.124383 |
| **729** | 2012-12-30 | 1796 | 0 | 1 | 0.255833 | 0.231700 | 0.483333 | 0.350754 |
| **730** | 2012-12-31 | 2729 | 1 | 2 | 0.215833 | 0.223487 | 0.577500 | 0.154846 |

In [38]:

```
#Facebook Prophet model
m = Prophet(growth = "linear",
            yearly_seasonality = True,
            weekly_seasonality = True,
            daily_seasonality = False,
            holidays = holidays,
            seasonality_mode = "multiplicative",
            seasonality_prior_scale = 20,
            holidays_prior_scale = 20,
            changepoint_prior_scale = 0.01)
m.add_regressor('workingday')
m.add_regressor('weathersit')
m.add_regressor('temp')
m.add_regressor('atemp')
m.add_regressor('hum')
m.add_regressor('windspeed')
m.fit(training_set)
```

```
18:12:20 - cmdstanpy - INFO - Chain [1] start processing
18:12:20 - cmdstanpy - INFO - Chain [1] done processing
```

Out[38]:

```
<prophet.forecaster.Prophet at 0x2a7f2ef3640>
```

In [39]:

```
#Create Future Dataframe
future = m.make_future_dataframe(periods = len(test_set),
                                 freq = "D")
future.tail()
```

Out[39]:

| | ds |
|---|---|
| **726** | 2012-12-27 |
| **727** | 2012-12-28 |
| **728** | 2012-12-29 |
| **729** | 2012-12-30 |
| **730** | 2012-12-31 |

In [40]:

```python
#merge regressors
future = pd.concat([future, dataset.iloc[:,2:]],
                   axis = 1)
future.tail()
```

Out[40]:

| | ds | workingday | weathersit | temp | atemp | hum | windspeed |
|---|---|---|---|---|---|---|---|
| **726** | 2012-12-27 | 1 | 2 | 0.254167 | 0.226642 | 0.652917 | 0.350133 |
| **727** | 2012-12-28 | 1 | 2 | 0.253333 | 0.255046 | 0.590000 | 0.155471 |
| **728** | 2012-12-29 | 0 | 2 | 0.253333 | 0.242400 | 0.752917 | 0.124383 |
| **729** | 2012-12-30 | 0 | 1 | 0.255833 | 0.231700 | 0.483333 | 0.350754 |
| **730** | 2012-12-31 | 1 | 2 | 0.215833 | 0.223487 | 0.577500 | 0.154846 |

In [41]:

```python
#forecast
forecast = m.predict(future)
forecast.head()
```

Out[41]:

| | ds | trend | yhat_lower | yhat_upper | trend_lower | trend_upper | atemp | atemp_ |
|---|---|---|---|---|---|---|---|---|
| **0** | 2011-01-01 | 2429.776904 | 274.639879 | 2164.623680 | 2429.776904 | 2429.776904 | -0.055503 | -0.0 |
| **1** | 2011-01-02 | 2435.398244 | 52.781019 | 1982.814006 | 2435.398244 | 2435.398244 | -0.060175 | -0.0 |
| **2** | 2011-01-03 | 2441.019585 | 370.166907 | 2254.347723 | 2441.019585 | 2441.019585 | -0.137830 | -0.1 |
| **3** | 2011-01-04 | 2446.640925 | 450.453198 | 2305.364375 | 2446.640925 | 2446.640925 | -0.127095 | -0.1 |
| **4** | 2011-01-05 | 2452.262265 | 639.018538 | 2532.923198 | 2452.262265 | 2452.262265 | -0.118992 | -0.1 |

5 rows × 46 columns

In [43]:

```
#XGBoost

prophet_variables = forecast.loc[:, ["trend", "holi", "weekly", "yearly"]]
df_xgb = pd.concat([dataset, prophet_variables], axis = 1)
df_xgb.head()
```

Out[43]:

| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed | tren |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 2011-01-01 | 985 | 0 | 2 | 0.344167 | 0.363625 | 0.805833 | 0.160446 | 2429.77690 |
| **1** | 2011-01-02 | 801 | 0 | 2 | 0.363478 | 0.353739 | 0.696087 | 0.248539 | 2435.39824 |
| **2** | 2011-01-03 | 1349 | 1 | 1 | 0.196364 | 0.189405 | 0.437273 | 0.248309 | 2441.01958 |
| **3** | 2011-01-04 | 1562 | 1 | 1 | 0.200000 | 0.212122 | 0.590435 | 0.160296 | 2446.64092 |
| **4** | 2011-01-05 | 1600 | 1 | 1 | 0.226957 | 0.229270 | 0.436957 | 0.186900 | 2452.26226 |

In [44]:

```
#Training and test set
test_days = 31
training_set = df_xgb.iloc[:-test_days, :]
test_set = df_xgb.iloc[-test_days:, :]
test_set.tail()
```

Out[44]:

| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed | tr |
|---|---|---|---|---|---|---|---|---|---|
| **726** | 2012-12-27 | 2114 | 1 | 2 | 0.254167 | 0.226642 | 0.652917 | 0.350133 | 6403.273 |
| **727** | 2012-12-28 | 3095 | 1 | 2 | 0.253333 | 0.255046 | 0.590000 | 0.155471 | 6408.712 |
| **728** | 2012-12-29 | 1341 | 0 | 2 | 0.253333 | 0.242400 | 0.752917 | 0.124383 | 6414.152 |
| **729** | 2012-12-30 | 1796 | 0 | 1 | 0.255833 | 0.231700 | 0.483333 | 0.350754 | 6419.591 |
| **730** | 2012-12-31 | 2729 | 1 | 2 | 0.215833 | 0.223487 | 0.577500 | 0.154846 | 6425.031 |

In [45]:

```
#isolate X and Y
y_train = training_set.y
y_test = test_set.y
X_train = training_set.iloc[:, 2:]
X_test = test_set.iloc[:, 2:]
```

In [46]:

```
#create XGBoost Matrices
Train = xgb.DMatrix(data = X_train, label = y_train)
Test = xgb.DMatrix(data = X_test, label = y_test)
```

In [50]:

```
#Set the parameters
parameters = {'learning_rate': 0.1,
              'max_depth': 3,
              'colsample_bytree': 1,
              'subsample': 1,
              'min_child_weight': 1,
              'gamma': 1,
              'random_state': 1502,
              'eval_metric': "rmse",
              'objective': "reg:squarederror"}
```

In [52]:

```
#XGBoost Model
model = xgb.train(params = parameters,
                  dtrain = Train,
                  num_boost_round = 100,
                  evals = [(Test, "y")])
```

```
[0]	y-rmse:3891.58894
[1]	y-rmse:3563.69059
[2]	y-rmse:3274.75734
[3]	y-rmse:2950.62380
[4]	y-rmse:2716.06266
[5]	y-rmse:2522.39072
[6]	y-rmse:2317.96280
[7]	y-rmse:2156.40260
[8]	y-rmse:2024.56200
[9]	y-rmse:1934.05555
[10]	y-rmse:1866.22580
[11]	y-rmse:1791.30150
[12]	y-rmse:1740.37022
[13]	y-rmse:1686.21001
[14]	y-rmse:1645.00477
[15]	y-rmse:1607.49185
[16]	y-rmse:1550.18027
[17]	y-rmse:1505.37005
[18]	y-rmse:1497.57975
[19]	y-rmse:1468.93298
[20]	y-rmse:1447.49851
[21]	y-rmse:1470.55166
[22]	y-rmse:1458.53921
[23]	y-rmse:1450.04251
[24]	y-rmse:1436.79017
[25]	y-rmse:1437.45231
[26]	y-rmse:1434.34955
[27]	y-rmse:1426.79344
[28]	y-rmse:1405.47873
[29]	y-rmse:1382.48373
[30]	y-rmse:1368.36629
[31]	y-rmse:1372.42708
[32]	y-rmse:1364.42167
[33]	y-rmse:1354.23317
[34]	y-rmse:1351.62577
[35]	y-rmse:1345.77694
[36]	y-rmse:1345.84280
[37]	y-rmse:1341.13245
[38]	y-rmse:1338.39152
[39]	y-rmse:1326.23727
[40]	y-rmse:1326.54284
[41]	y-rmse:1325.68060
[42]	y-rmse:1313.05545
[43]	y-rmse:1304.07389
[44]	y-rmse:1299.94952
[45]	y-rmse:1290.71035
[46]	y-rmse:1293.13980
[47]	y-rmse:1293.57000
[48]	y-rmse:1279.87623
[49]	y-rmse:1279.88202
[50]	y-rmse:1282.58339
[51]	y-rmse:1282.58685
[52]	y-rmse:1279.46683
[53]	y-rmse:1268.14083
[54]	y-rmse:1262.12109
[55]	y-rmse:1257.84226
[56]	y-rmse:1258.36123
[57]	y-rmse:1261.15983
[58]	y-rmse:1261.33869
[59]	y-rmse:1260.90076
[60]	y-rmse:1257.12425
```

```
[61]	y-rmse:1237.10839
[62]	y-rmse:1236.71193
[63]	y-rmse:1235.84508
[64]	y-rmse:1238.61690
[65]	y-rmse:1235.14405
[66]	y-rmse:1235.04261
[67]	y-rmse:1223.69949
[68]	y-rmse:1221.92818
[69]	y-rmse:1216.91122
[70]	y-rmse:1216.01084
[71]	y-rmse:1218.76650
[72]	y-rmse:1218.01949
[73]	y-rmse:1217.88212
[74]	y-rmse:1207.76128
[75]	y-rmse:1205.51613
[76]	y-rmse:1202.63937
[77]	y-rmse:1202.52231
[78]	y-rmse:1203.09201
[79]	y-rmse:1202.82743
[80]	y-rmse:1201.42101
[81]	y-rmse:1192.39748
[82]	y-rmse:1187.60552
[83]	y-rmse:1187.49280
[84]	y-rmse:1188.94596
[85]	y-rmse:1188.47885
[86]	y-rmse:1186.20477
[87]	y-rmse:1186.19375
[88]	y-rmse:1186.19311
[89]	y-rmse:1184.95346
[90]	y-rmse:1184.79903
[91]	y-rmse:1183.38123
[92]	y-rmse:1183.37107
[93]	y-rmse:1178.39280
[94]	y-rmse:1178.29195
[95]	y-rmse:1181.48770
[96]	y-rmse:1182.13371
[97]	y-rmse:1181.01612
[98]	y-rmse:1180.26977
[99]	y-rmse:1174.05661
```

In [53]:

```python
#Forecasting
predictions_xgb = pd.Series(model.predict(Test), name = "XGBoost")
predictions_xgb.index = test_set.ds
predictions_xgb[:2]
```

Out[53]:

```
ds
2012-12-01    3996.192383
2012-12-02    3068.570312
Name: XGBoost, dtype: float32
```

In [56]:

```
#set up index
training_set.index = training_set.ds
test_set.index = test_set.ds
```

Out[56]:

| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed | |
|---|---|---|---|---|---|---|---|---|---|
| **ds** | | | | | | | | | |
| **2012-12-01** | 2012-12-01 | 5191 | 0 | 2 | 0.298333 | 0.316904 | 0.806667 | 0.059704 | 6261.8 |
| **2012-12-02** | 2012-12-02 | 4649 | 0 | 2 | 0.347500 | 0.359208 | 0.823333 | 0.124379 | 6267.2 |
| **2012-12-03** | 2012-12-03 | 6234 | 1 | 1 | 0.452500 | 0.455796 | 0.767500 | 0.082721 | 6272.7 |
| **2012-12-04** | 2012-12-04 | 6606 | 1 | 1 | 0.475833 | 0.469054 | 0.733750 | 0.174129 | 6278.1 |
| **2012-12-05** | 2012-12-05 | 5729 | 1 | 1 | 0.438333 | 0.428012 | 0.485000 | 0.324021 | 6283.6 |
| **2012-12-06** | 2012-12-06 | 5375 | 1 | 1 | 0.255833 | 0.258204 | 0.508750 | 0.174754 | 6289.0 |
| **2012-12-07** | 2012-12-07 | 5008 | 1 | 2 | 0.320833 | 0.321958 | 0.764167 | 0.130600 | 6294.4 |
| **2012-12-08** | 2012-12-08 | 5582 | 0 | 2 | 0.381667 | 0.389508 | 0.911250 | 0.101379 | 6299.9 |
| **2012-12-09** | 2012-12-09 | 3228 | 0 | 2 | 0.384167 | 0.390146 | 0.905417 | 0.157975 | 6305.3 |
| **2012-12-10** | 2012-12-10 | 5170 | 1 | 2 | 0.435833 | 0.435575 | 0.925000 | 0.190308 | 6310.8 |
| **2012-12-11** | 2012-12-11 | 5501 | 1 | 2 | 0.353333 | 0.338363 | 0.596667 | 0.296037 | 6316.2 |
| **2012-12-12** | 2012-12-12 | 5319 | 1 | 2 | 0.297500 | 0.297338 | 0.538333 | 0.162937 | 6321.6 |
| **2012-12-13** | 2012-12-13 | 5532 | 1 | 1 | 0.295833 | 0.294188 | 0.485833 | 0.174129 | 6327.1 |
| **2012-12-14** | 2012-12-14 | 5611 | 1 | 1 | 0.281667 | 0.294192 | 0.642917 | 0.131229 | 6332.5 |
| **2012-12-15** | 2012-12-15 | 5047 | 0 | 1 | 0.324167 | 0.338383 | 0.650417 | 0.106350 | 6337.9 |
| **2012-12-16** | 2012-12-16 | 3786 | 0 | 2 | 0.362500 | 0.369938 | 0.838750 | 0.100742 | 6343.4 |
| **2012-12-17** | 2012-12-17 | 4585 | 1 | 2 | 0.393333 | 0.401500 | 0.907083 | 0.098258 | 6348.8 |
| **2012-12-18** | 2012-12-18 | 5557 | 1 | 1 | 0.410833 | 0.409708 | 0.666250 | 0.221404 | 6354.3 |
| **2012-12-19** | 2012-12-19 | 5267 | 1 | 1 | 0.332500 | 0.342162 | 0.625417 | 0.184092 | 6359.7 |
| **2012-12-20** | 2012-12-20 | 4128 | 1 | 2 | 0.330000 | 0.335217 | 0.667917 | 0.132463 | 6365.1 |
| **2012-12-21** | 2012-12-21 | 3623 | 1 | 2 | 0.326667 | 0.301767 | 0.556667 | 0.374383 | 6370.6 |
| **2012-12-22** | 2012-12-22 | 1749 | 0 | 1 | 0.265833 | 0.236113 | 0.441250 | 0.407346 | 6376.0 |
| **2012-12-23** | 2012-12-23 | 1787 | 0 | 1 | 0.245833 | 0.259471 | 0.515417 | 0.133083 | 6381.5 |

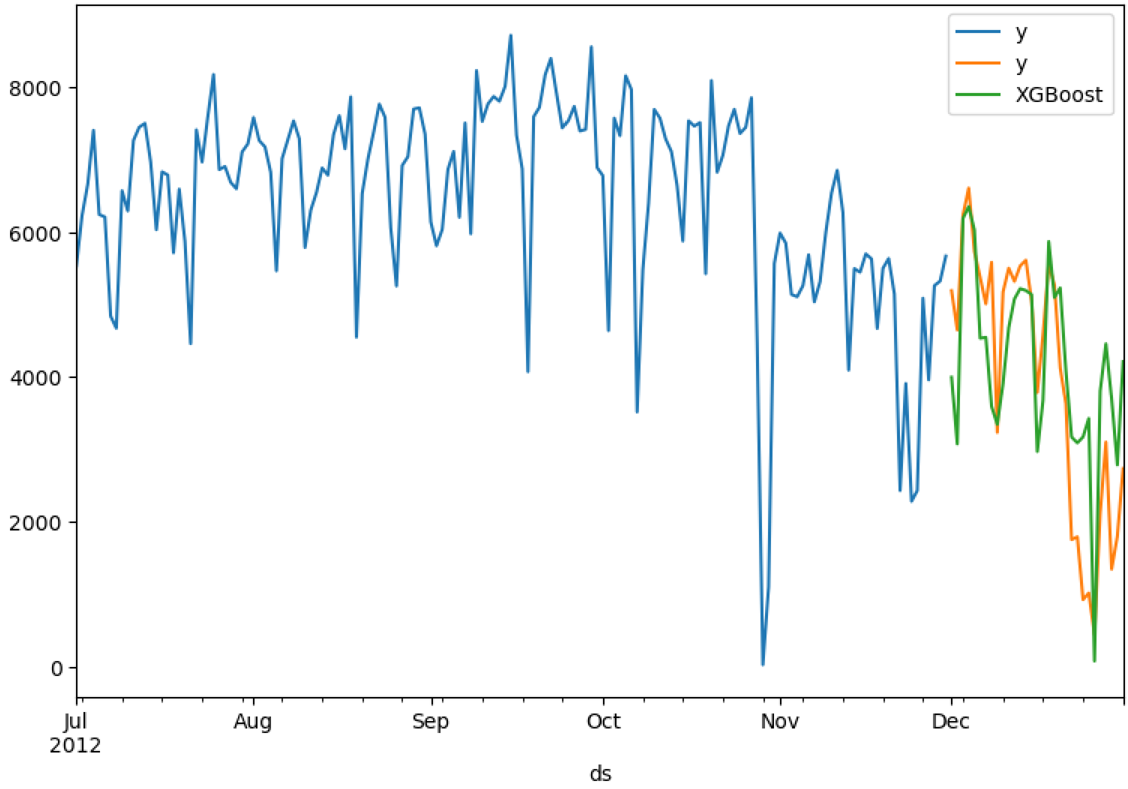
| | ds | y | workingday | weathersit | temp | atemp | hum | windspeed | |
|---|---|---|---|---|---|---|---|---|---|
| **ds** | | | | | | | | | |
| **2012-12-24** | 2012-12-24 | 920 | 1 | 2 | 0.231304 | 0.258900 | 0.791304 | 0.077230 | 6386.9 |
| **2012-12-25** | 2012-12-25 | 1013 | 0 | 2 | 0.291304 | 0.294465 | 0.734783 | 0.168726 | 6392.3 |
| **2012-12-26** | 2012-12-26 | 441 | 1 | 3 | 0.243333 | 0.220333 | 0.823333 | 0.316546 | 6397.8 |
| **2012-12-27** | 2012-12-27 | 2114 | 1 | 2 | 0.254167 | 0.226642 | 0.652917 | 0.350133 | 6403.2 |
| **2012-12-28** | 2012-12-28 | 3095 | 1 | 2 | 0.253333 | 0.255046 | 0.590000 | 0.155471 | 6408.7 |
| **2012-12-29** | 2012-12-29 | 1341 | 0 | 2 | 0.253333 | 0.242400 | 0.752917 | 0.124383 | 6414.1 |
| **2012-12-30** | 2012-12-30 | 1796 | 0 | 1 | 0.255833 | 0.231700 | 0.483333 | 0.350754 | 6419.5 |
| **2012-12-31** | 2012-12-31 | 2729 | 1 | 2 | 0.215833 | 0.223487 | 0.577500 | 0.154846 | 6425.0 |

In [57]:

```
#Viz
training_set.y['2012-07-01':].plot(figsize = (9,6), legend = True)
test_set.y.plot(legend = True)
predictions_xgb.plot(legend = True)
```

Out[57]:

```
<AxesSubplot: xlabel='ds'>
```



In [58]:

```
#MAE and RMSE
from sklearn.metrics import mean_squared_error, mean_absolute_error
print(round(mean_absolute_error(test_set['y'], predictions_xgb),0))
print(round(np.sqrt(mean_squared_error(test_set['y'], predictions_xgb)), 0))
```

```
946.0
1174.0
```

In [59]:

```
#MAPE function
def MAPE(y_true, y_pred):
  y_true, y_pred = np.array(y_true), np.array(y_pred)
  return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
MAPE(test_set['y'], predictions_xgb)
```

Out[59]:

```
45.377004733275925
```

In [ ]:

```

```