Amazon Forecast ハンズオン

2019/10/31

エバンジェリスト 亀田

はじめに：Amazon ForecastはAWSが提供している時系列データをもとに予測を行う機械学習サービスです。

1. リージョンの選択

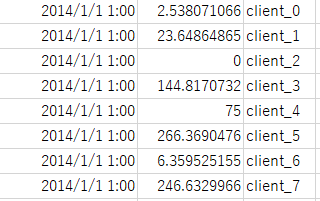
本ハンズオンで作業を行うリージョンは講師に確認をしてください。画面右上の地名を選択すると変更できます。



1. 学習用データのダウンロード

http://bit.ly/22Z4QCj

以下のようなクライアントごとの消費電力データが大量に含まれています。



このハンズオンでは、この時系列データをもとに将来の電力消費予測を行います。Forecastでは、時系列データであれば処理が可能ですので、売り上げデータ、在庫消費データなどを用いることもできます。

3. S3バケットの作成

Forecastは学習データを任意のS3バケットよりインポートして利用するため、先ほどダウンロードしたCSVファイルを格納するS3バケットを作成します。

3-1．以下のように[s3]と入力してS3のマネージメントコンソールにアクセスします。



3-2.　【バケットを作成する】を押します。



3-3．　【バケット名】に適当な名前を入れます。

S3バケットはAWS全ユーザーの全バケットで一意である必要があるため、ある程度長い名前が必要です。

リージョンはForecastの作業を行うリージョンと同じものを指定してください。



3-4.　すべてデフォルトでS3バケットを作りますので画面右下の【次へ】を2回押し、最後に【バケットを作成】を押してください。

3-5.　作成されたバケットをダブルクリックしてください。



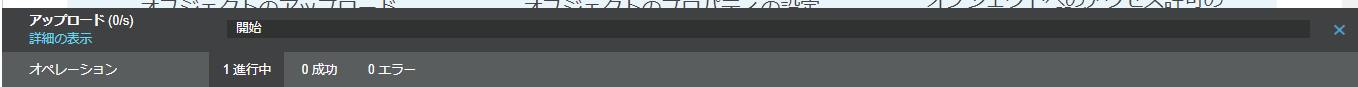
3-6．　【アップロード】を押します。



3-7.　先ほどダウンロードした[electricityusagedata.csv]をドラッグ＆ドロップで選択します。



3-8.画面左下の【アップロード】ボタンをおすとアップロード作業が開始されます。



3-9．アップロードが完了すると以下の画面となります。



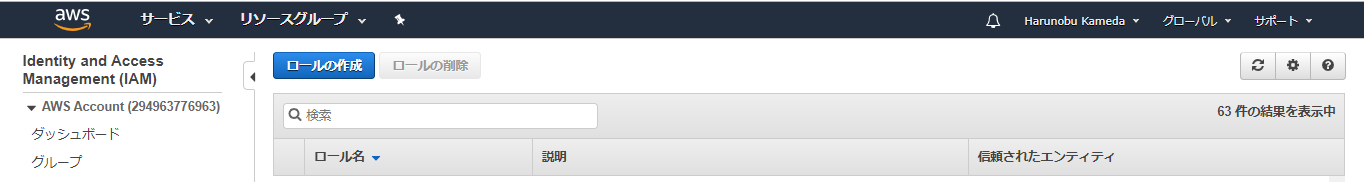
1. IAMロールの作成

ForecastがS3へアクセスするために必要な権限を作成します。ここで作成された権限は、後ほどForecastにアタッチされ、S3へアクセスできるようになります。

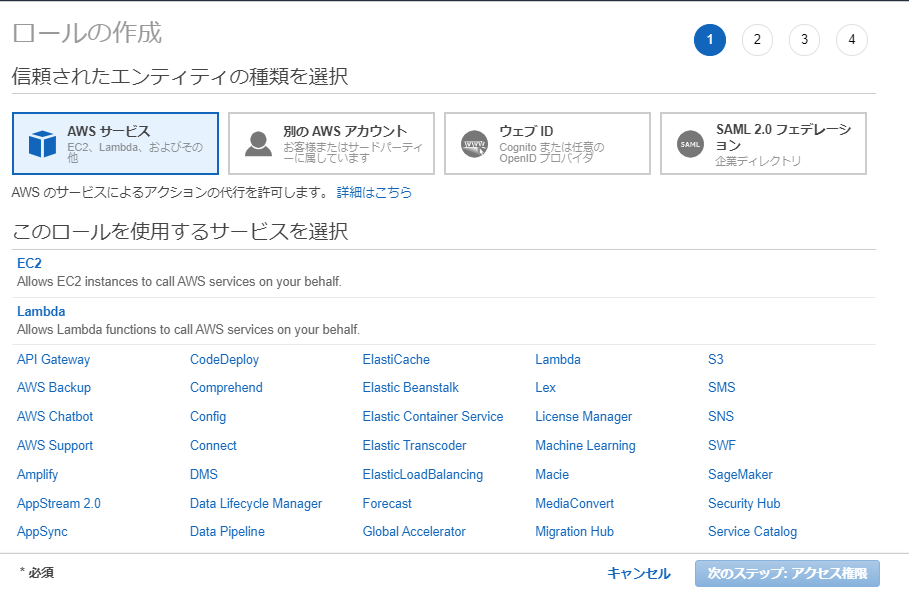
4-1.　S3と同じ要領でIAMのマネージメントコンソールへアクセスします。



4-2．画面左のナビゲーションペインでロールを選択します。



4-3．【ロールの作成】をおします。



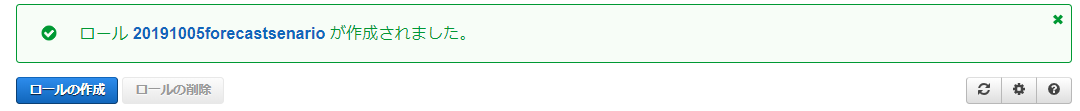
4-4．　【このロールを使用するサービスを選択】から【Forecast】選び、【次のステップ】を押します。



4-5.　画面右下の【次のステップ】を押します。その次に表示されるタグの画面はそのまま何も入力せずもう一度【次のステップ】を押します。



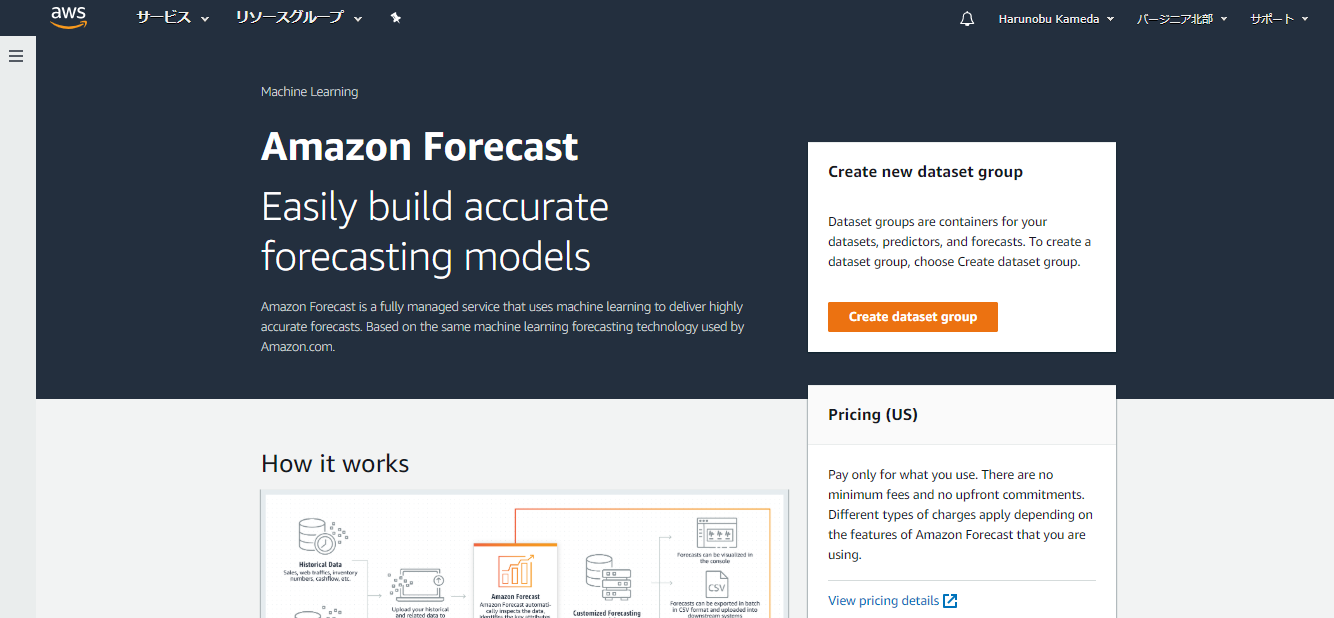
4-6.　【ロール名】に適当な名前を入力し【ロールの作成】を押します。



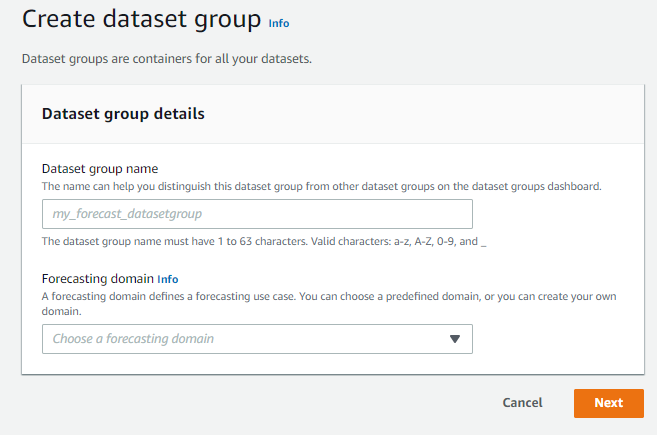
4-7.　作成が完了しました。作成されたロールをクリックして詳細画面に行きます。後ほどこのロールをForecaseに付与して、Forecastに対するS3へのアクセス権限を付与しますので【ロールARN】をメモっておいてください。



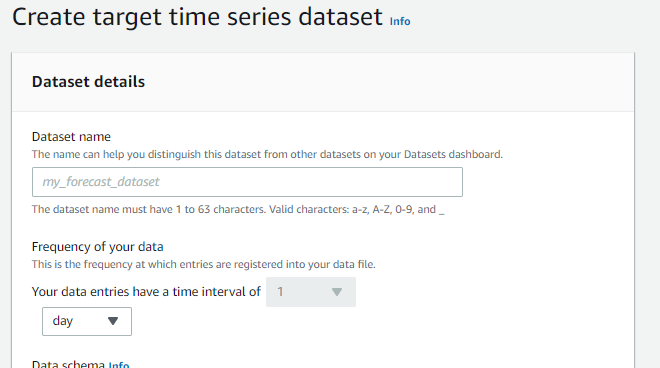
5．　S3,IAMと同じようにForecastの画面にいきます。



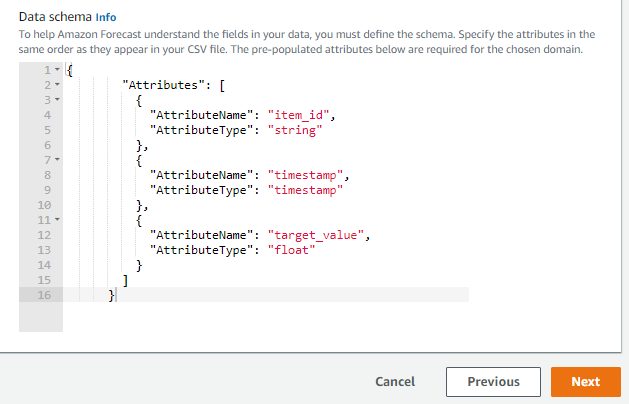
5-1．　【Create dataset group】を押します。



5-2.　適当な名前を【Dataset group name】に入力し、【Forecasting domain】 を【Custom】に設定し【Next】を押します。名前は、数字から始まる文字列は指定できませんので注意してください。



5-3.　【Dataset name】に適当な名前をいれます。【Frequency of your data】は【hour】を選びます。（この値は学習させる元データの時系列間隔と合わせる必要があります。）



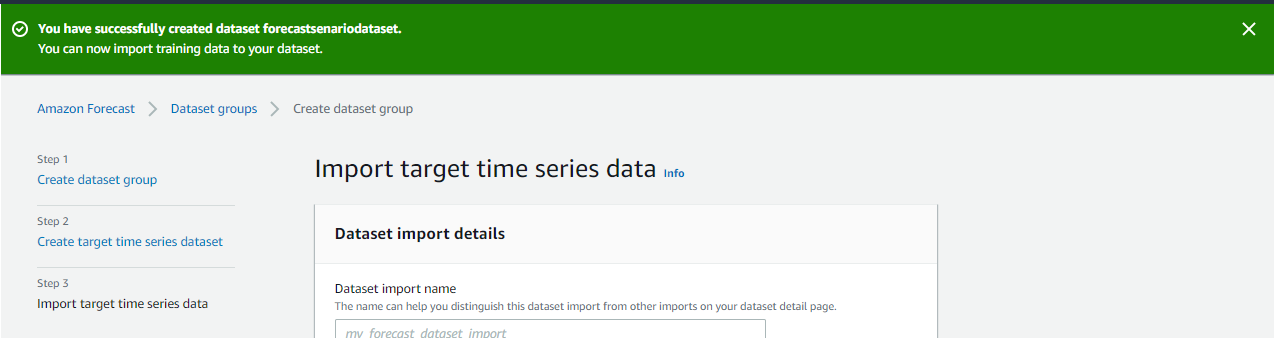
5-4. 【Data Schema】の値を学習させる元データの並び順に合わせるため以下のように入れ替えます。



5-5.　【Next】を押します。

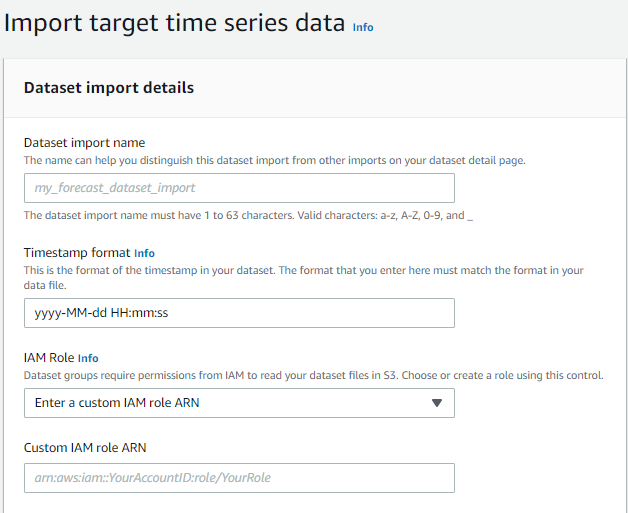


5-6．以下の緑のバーが表示されれば、正しくDatasetが設定されています。



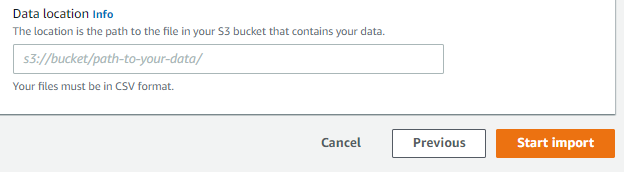
続いてImportの設定を行います。

5-7.　【Dataset import name】に適当な名前を入力します。【Timestamp format】はインポートする元データと合致した表記でなければならないですが、このハンズオンでは元データとデフォルト値が同じなのでそのままにします。【Custom IAM role ARN】に作ほどコピーした新しく作成したIAM RoleのARNを貼り付けます。

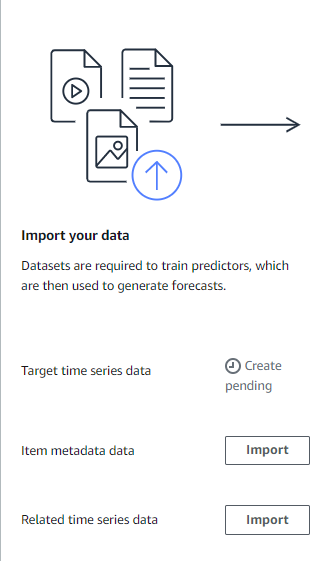


5-8.　 先ほどアップロードした学習用元データを以下の形式で【Data location】に入力し、【Start Import】を押します。S3://<バケット名>/<ファイル名>

(外部からアクセスできるURIとは異なりますので注意してください)

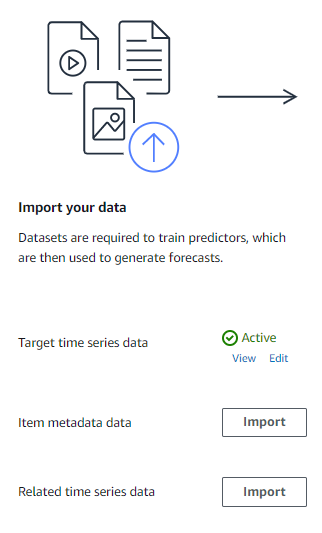


5-9．　Importが開始され以下の画面になります。



インポート作業は時間が10分弱かかりますので、その間にPersonalizeハンズオンシナリオを行います。

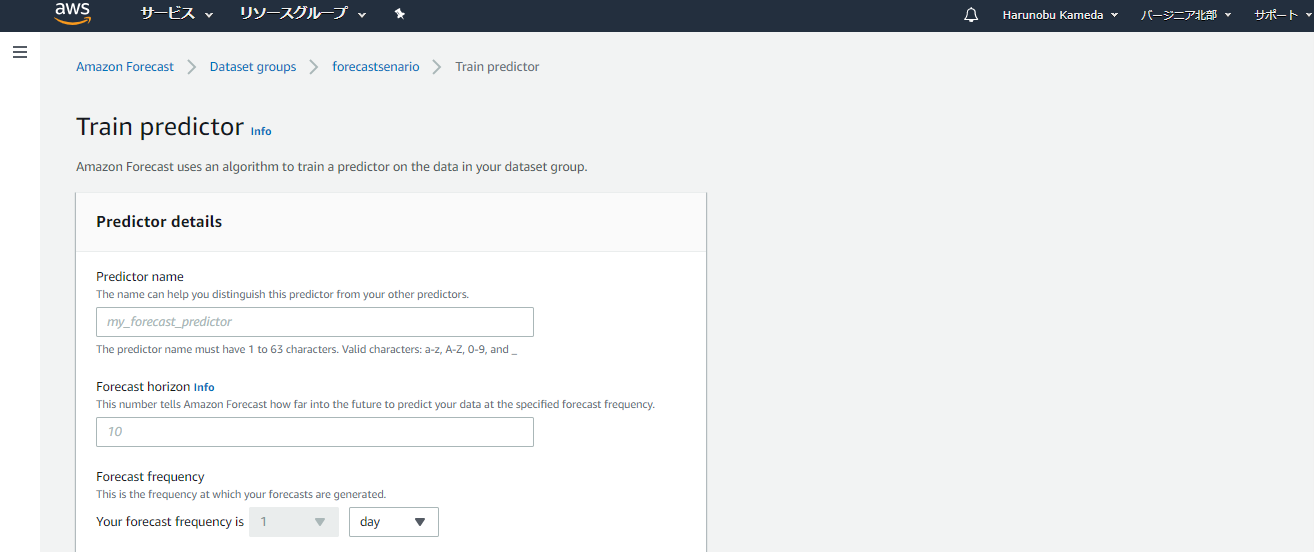
5-10.　以下のように【Active】と緑で表示されればインポート完了です。



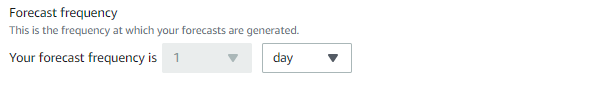
1. 学習

データのインポートが完了したので、学習を行います。

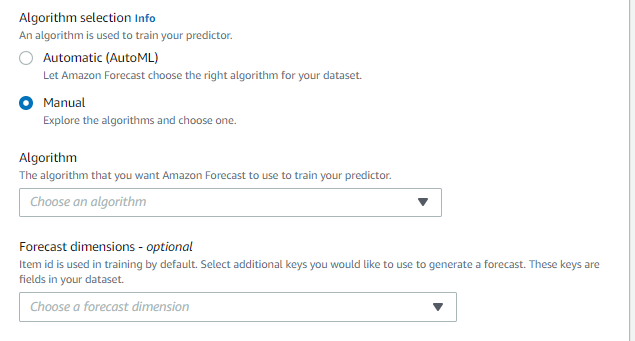
6-1. 【Train a predictor】の【Start】を押します。



6-2.　【Predictor name】に適当な名前を入れます。【Forecast horizon】は予測を行う時間間隔をしています。このハンズオンでは　36（36時間分）の予測を作成するため、【36】と入力します。



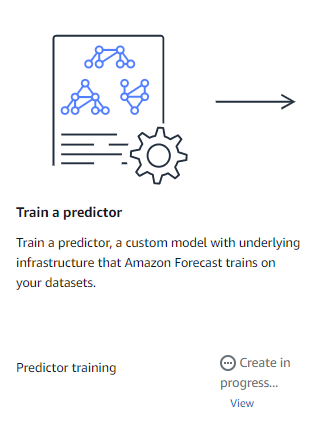
6-3.　【Forecast frequency】は元データと同じである必要があるため【hour】を選びます。【Algorithm selection】は【Manual】のままとし【Algorithm】は【ETS】を選びます。



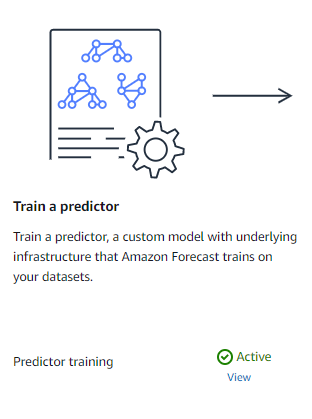
6-4．残りの値は全てそのままで【Train predictor】を押します。



6-5.　以下が表示されたら学習開始となっているので、学習が終わるまで約25分程度かかりますので、その間にPersonalizeのハンズオンに戻り、続きを行います。



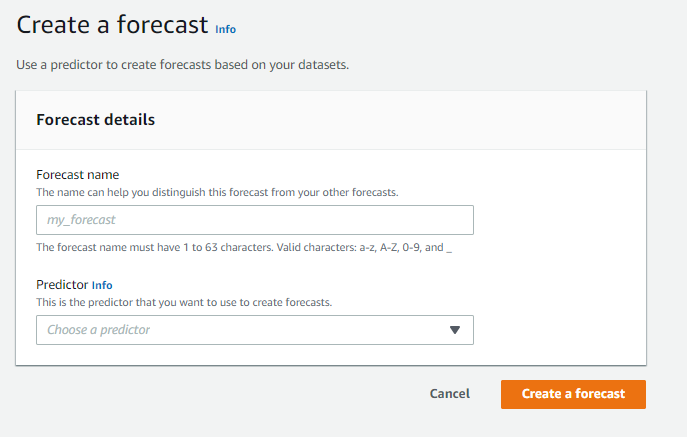
6-6.　以下のように【Active】と表示されれば学習が完了です。



1. 予測エンドポイント作成

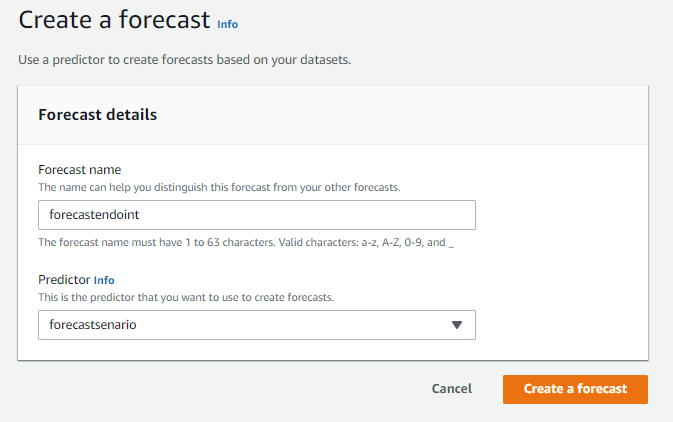
学習が開始したので予測を行うエンドポイントを作成します。

7-1． 画面右の【Generate forecasts】のところで【Start】ボタンを押します。

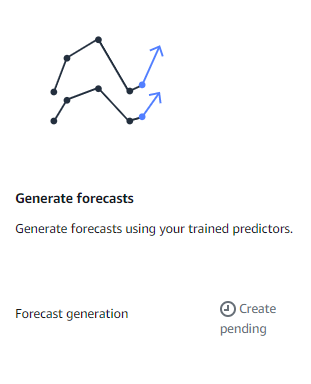


7-2．　【Forecast name】に適当な名前を入れます。【Predictor】のドロップダウンから先ほど学習環境につけた名前を選びます。

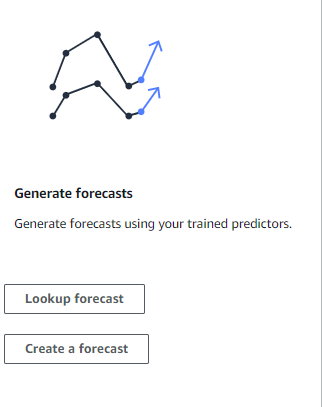
＃何も表示されない場合は、cancelを押して再度やり直すと表示されます。



7-3.　【Create a forecast】のボタンを押します。



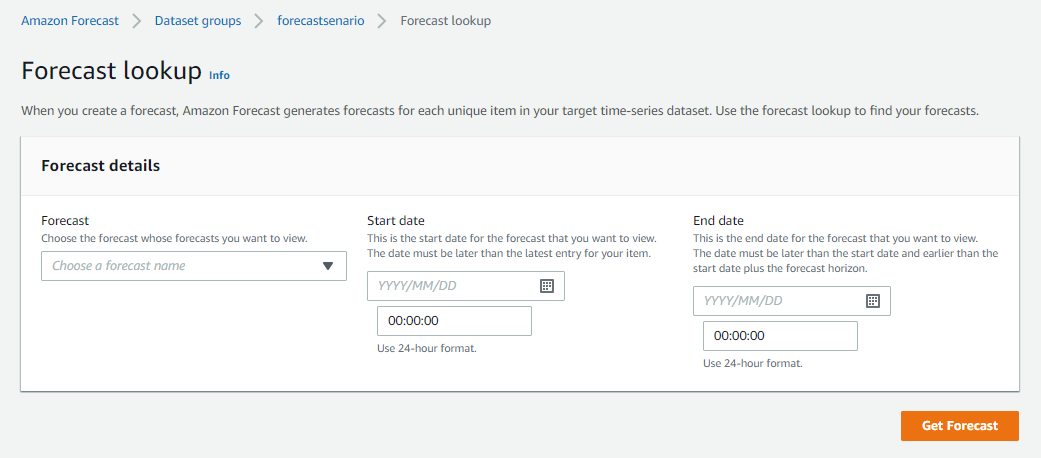
7-4. 予測エンドポイントの生成が開始されます。完了するまで20分程度かかりますので。その間Personalizeのハンズオンを続行します。以下の表示になれば完了です。予測ができるようになりました。



1. 予測を行う

以上の手順で学習済データから予測を行うことができます。

【Look up Forecast】を押します。

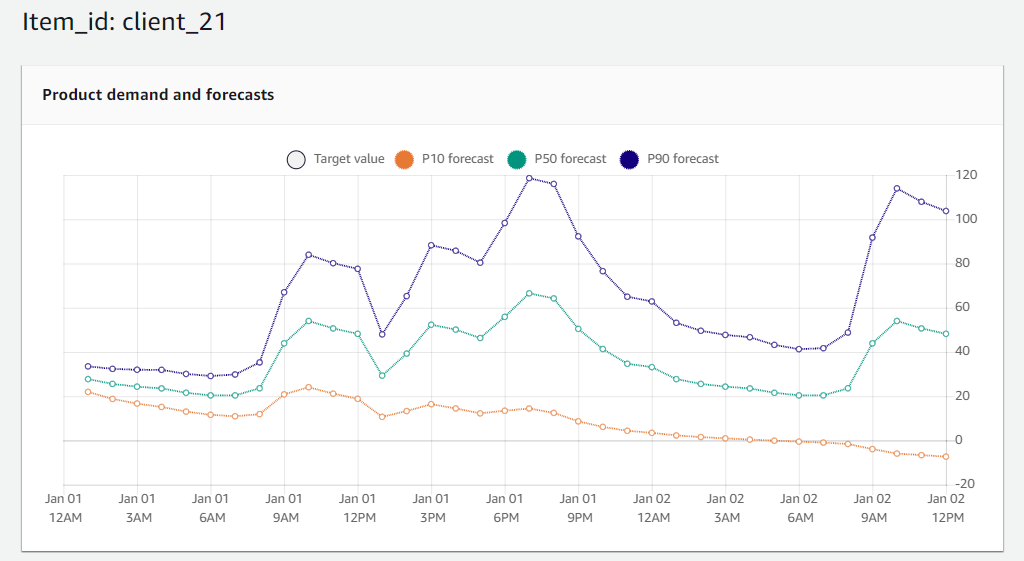


8-1.　【Forecast】に先ほど作成したエンドポイントの名前をドロップダウンから選びます。

* [**開始日**] – **2015/01/01** と入力します。デフォルト時間である 00:00:00 のままにします。
* [**終了日**] – **2015/01/02** と入力します。時間を 12:00:00 に変更します。

【Value】には学習元データに入っているクライアントのIDを入力します。(例：client\_2)

8-2.　【Get Forecast】 を押すと以下のように予測がグラフで出力されます。



**P10**、**P50**、**P90**の数値は、それぞれ 10%、50%、90% の確立で、実際の需要量に合致します。これらを利用する方法は、使っているユースケースと、需要に対する過小および過大な予測がどう影響するかにより変わってきます。P50 の予測値が、実際の需要にもっとも近いと考えられます。