***Amazon Forecast***

* It is a fully managed service that uses statistical and machine learning algorithms to deliver highly accurate time-series forecasts.
* Time-series forecasting is useful in multiple fields, including retail, finance, logistics, and healthcare. It can also be used to predict domain-specific metrics for your inventory, workforce, web traffic, server capacity, and finances.
* Most common use cases for Amazon Forecast :
  + **Retail Demand Planning:**
    - Predict product demand, allowing you to more accurately vary inventory and pricing at different store locations.
  + **Supply Chain Planning:**
    - Predict the quantity of raw goods, services, or other inputs required by manufacturing.
  + **Resource Planning:**
    - Predict requirements for staffing, advertising, energy consumption, and server capacity.
  + **Operational Planning:**
    - Predict levels of web traffic, AWS usage, and IoT sensor usage.
* Amazon Forecast automates much of the time-series forecasting process, providing the following features:
  + **Automated Machine Learning:**
    - Forecast automates complex machine learning tasks by finding the optimal combination of machine learning algorithms for your datasets.
  + **State-of-the-art algorithms:**
    - Forecast offers a wide range of training algorithms, from commonly used statistical methods to complex neural networks.
  + **Missing values support:**
    - Forecast provides several filling methods to automatically handle missing values in your datasets.
  + **Additional built-in datasets:**
    - Forecast can automatically incorporate built-in datasets to improve your model.
* When working with Forecast we deal with:
  + Importing Datasets
  + Training Predictors
  + Generating Forecast

## **Importing Datasets:**

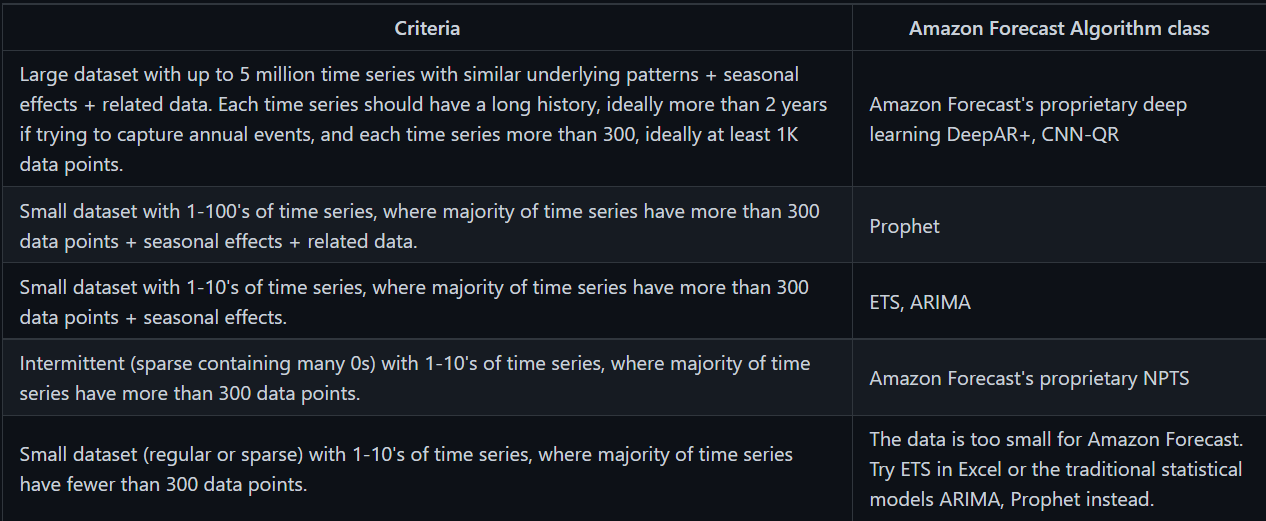
* + A *dataset group* is a collection of complementary datasets that detail a set of changing parameters over a series of time. After creating a dataset group, you use it to train a predictor.
  + When creating datasets we provide the following data:
    - The frequency/interval of the record in the dataset
    - The predicting format(Domain) and the dataset type
    - The Dataset Schema
    - Geolocation and Timezone information
  + Types of dataset:
    - **Target Time Series dataset:**
      * The main dataset that include the timestamp, unique-id and the demand value of a record, i.e., it includes the historical time-series demand data for the retail items(*item\_id*, *timestamp* and *demand* target field)
    - **Related Time Series dataset (Numerical values):**
      * The helping dataset that includes historical time-series data other than the target field like the price or revenue. It is mappable to the target dataset.
    - **Metadata dataset (Categorical value):**
      * Includes the metadata related to the item like color, genre, brand, category,etc.
  + The ***Data Schema*** contains the headers for all the columns in the corresponding order and the type of the data in each column.
  + A group of datasets including the Target time series dataset, Related time series dataset and Metadata dataset or a combination of them is called the ***Dataset Group.***
  + Types of Domain supported by Forecast:
    - **Retail Domain:** 
      * For retail demand forecasting
    - **Inventory Planning Domain:**
      * For supply chain and inventory planning
    - **EC2 Capacity Domain:** 
      * For forecasting Amazon Elastic Compute Cloud (Amazon EC2) capacity
    - **WorkForce Domain:**
      * For workforce planning
    - **Web Traffic Domain:**
      * For estimating future web traffic
    - **Metric Domain:**
      * For forecasting metrics, such as revenue and cash flow
    - **Custom Domain:**
      * For all other types of time-series forecasting
  + Handling Missing Values:
    - Forecast support following filling methods:
      * Front filling:
        + Fills any missing value between the Global Start date and the first recorded datapoint
      * Back filling:
        + Similarly, fills any missing values between the last recorded datapoint and the Global End date
      * MIddle filling:
        + Fills any missing value between time start and item end date of a dataset
      * Future filling:
        + Fills any missing value between the Global end date and the end of Forecast horizon
    - All the following filling methods fill the missing value with
      * Zero, Nan
      * Min, Max
      * Mean, Median
      * Any value

## **Training Predictors**

* + A predictor is an Amazon Forecast model that is trained using your target time series, related time series, item metadata, and any additional datasets you include.
  + Amazon Forecast requires the following inputs to train a predictor:
    - Dataset group
    - Forecast horizon
    - Forecast frequency
    - Few others can also be used:
      * Time alignment boundary
      * Forecast dimension
      * Forecast types
      * Optimization metrics
      * Additional datasets(Holidays and Weather)
  + We can customize our own time boundaries such as first day of the week as Wednesday for weekly frequency, or first day of the month for monthly frequency as 15th of the month,etc using *Time alignment boundary*
  + Optimization Metrics:
    - Amazon Forecast provides the following metrics to evaluate the predictors:
      * ***Root Mean Square Error*** (RMSE),
      * ***Weighted Quantile Loss*** (wQL), P
      * ***Average Weighted Quantile Loss*** (Average wQL),
      * ***Mean Absolute Scaled Error*** (MASE),
      * ***Mean Absolute Percentage Error*** (MAPE), and
      * ***Weighted Absolute Percentage Error*** (WAPE).
  + We can also **retain our predictors** to further add new data and try to increase the accuracy of the model. This is faster than training a new model using the same dataset.
  + Forecast also allows us to enter two additional datasets to further help building our model
    - Weather- based on the given location or country, it takes the available data for the weather to help in our forecast
    - Holiday - similar to weather, it is also based on the location.
  + Once the predictors are working and we forecast the values, we can also use **Predictor Explainability** to help us understand how the attributes in your datasets impact your target variable. Forecast uses a metric called ***Impact scores*** to quantify the relative impact of each attribute and determine whether they increase or decrease forecast values.

## **Generating Forecast**

* + After you create an Amazon Forecast predictor, you are ready to create a forecast. By default, a forecast includes predictions for every item (item\_id) in the dataset group that was used to train the predictor.
  + However, you can specify a subset of items that are used to generate a forecast.
  + Forecast Explainability
    - Forecast Explainability helps you better understand how the attributes in your datasets impact forecasts for specific time series (item and dimension combinations) and time points.
    - Forecast uses a metric called Impact scores to quantify the relative impact of each attribute and determine whether they increase or decrease forecast values.



# **References:**

* <https://docs.aws.amazon.com/forecast/latest/dg/what-is-forecast.html>
* <https://github.com/aws-samples/amazon-forecast-samples>
* <https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/forecast.html>