AWS re:Invent

AIM312

Predict future business outcomes using Amazon Forecast

Rohit Menon
Sr. Product Manager
Amazon Forecast
Amazon Web Services

Chris King
Sr. Solutions Architect
Vertical AI Services
Amazon Web Services

Sean Patrick McCrary, CPIM
Director, Supply Chain Data Analytics, LSG
Thermo Fisher Scientific

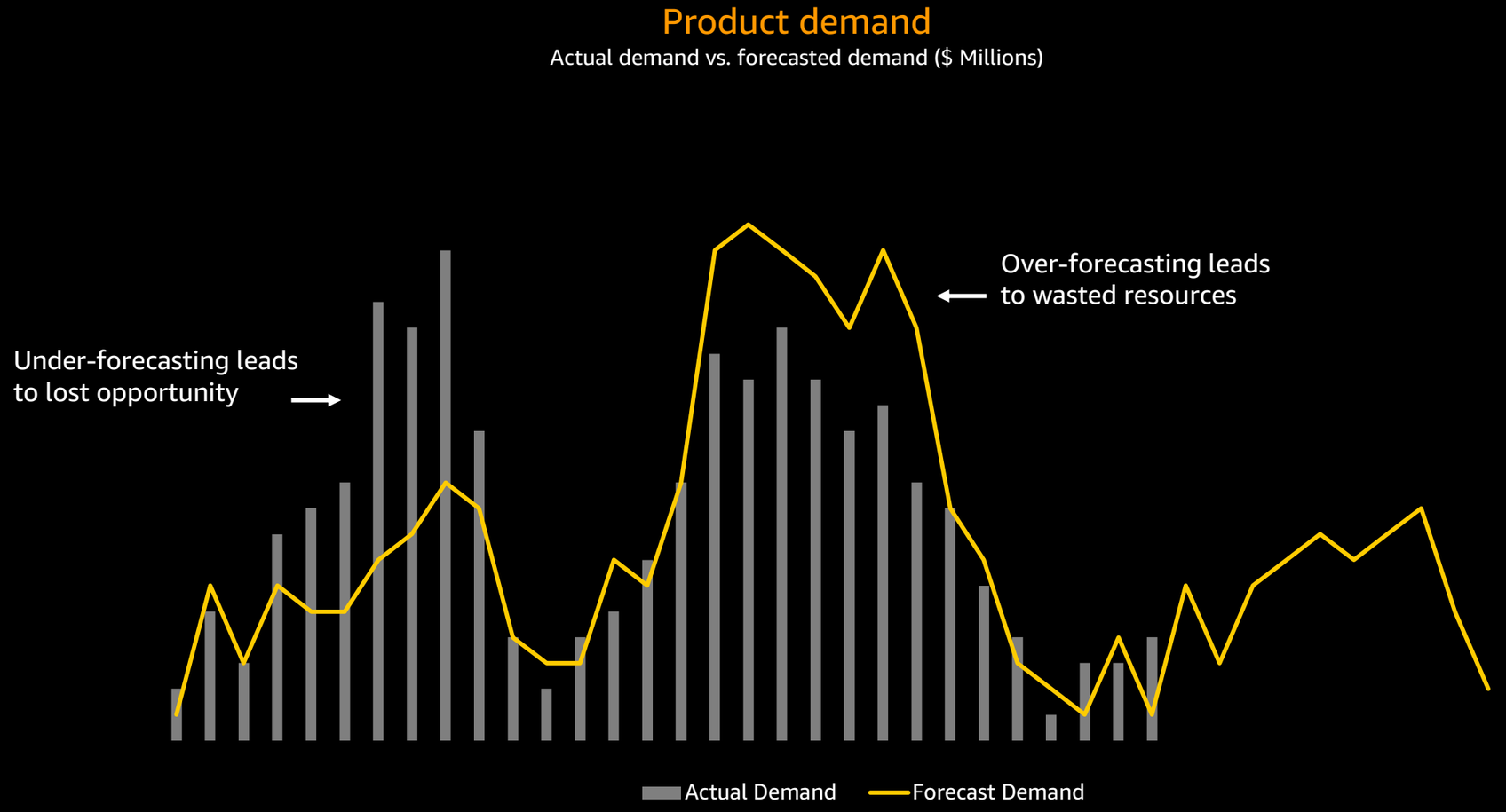
Agenda

- The case for forecasting
- Overview of Amazon Forecast
- Behind the scenes: The algorithms powering Amazon Forecast
- Evaluating your Forecast
- Case study: Reducing excess inventory at Thermo Fisher Scientific
- Wrap up and Q&A

The case for forecasting

The case for forecasting

Forecasting is the science of predicting the future



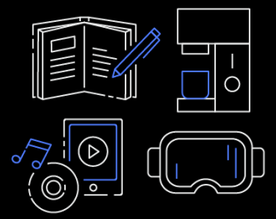
The case for Forecasting

Impact of under and over forecasting

	Inventory planning	Workforce planning	Capacity planning	Financial planning
Over-forecasting	Excess inventory	Unutilized labor	Uncapitalized infrastructure	Depleted cash reserves
Under-forecasting	Lost sale	Overtime costs	Unmet demand	Undercutting

Forecasting at Amazon.com

Accurate forecasting is critical for delivering on customer promises

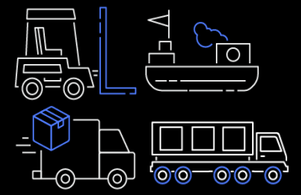


Product availability

Demand forecasting for over 400 million products every day

Lower price

Inventory and fulfillment cost reduction to provide low prices to customers



Fast delivery

12 shipping options with free same-day delivery

Forecasting process

Three standard steps involved in forecasting

- Looking backward
  - We always begin with historical data; ideally, that contains a timestamp, an item, a value
  - Larger datasets contain potentially millions of rows over thousands of items, but the construct is the same
  - This provides the baseline
- Identifying trends
  - Using statistical, deep learning, or other approaches, you look over the historical data to determine trends within your data—trends that hopefully continue into the future
- Projecting forward
  - Given the trends identified, take each item and predict in increments the expected future values

Amazon Forecast: An overview

The AWS ML Stack

Broadest and deepest set of capabilities

AI Services

VISION			SPEECH		LANGUAGE		CHATBOTS	FORECASTING	RECOMMENDATIONS
Amazon Rekognition Image	Amazon Rekognition Video	Amazon Textract	Amazon Polly	Amazon Transcribe	Amazon Translate	Amazon Comprehend & Amazon Comprehend Medical	Amazon Lex	Amazon Forecast	Amazon Personalize

ML Services

Amazon SageMaker

Ground Truth	Notebooks	Algorithms + Marketplace	Reinforcement Learning	Training	Optimization	Deployment	Hosting

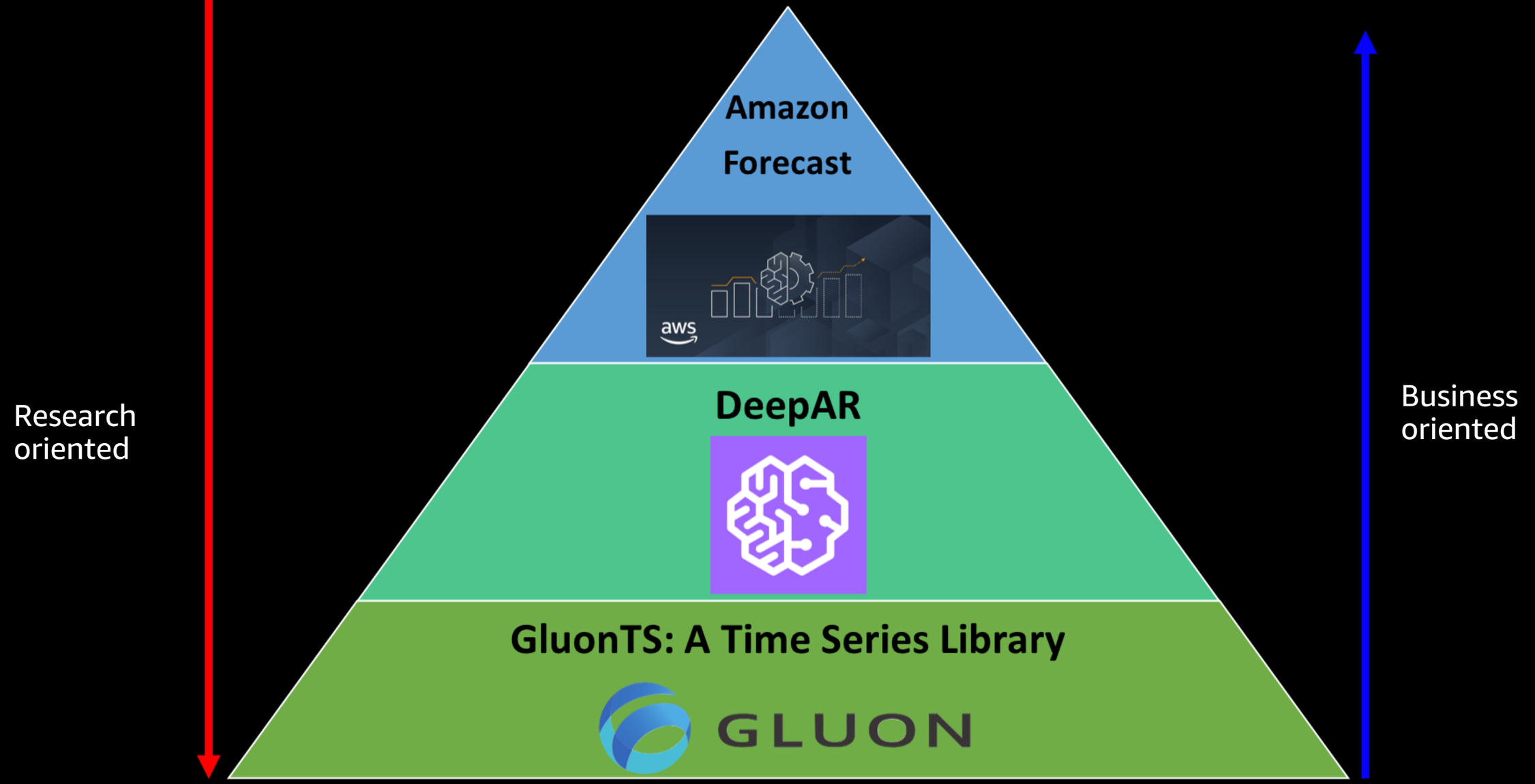
ML frameworks + infrastructure

FRAMEWORKS	INTERFACES
TensorFlow, mxnet, PYTORCH	GLUON, Keras

INFRASTRUCTURE							
Amazon EC2 P3 & P3DN	Amazon EC2 G4	Amazon EC2 C5	FPGAS	DL Containers & AMIs	AWS IoT Greengrass	Amazon Elastic Inference	AWS Inferentia

The AWS ML Stack

The forecast view



Amazon Forecast

Automated machine learning service for accurate forecasting

Fully managed service

Automatically sets up data pipeline, training, and prediction

Highly accurate

50% improvement in accuracy over traditional methods

Easy to use

No deep learning experience required

Your data, your models

Encrypted with customer keys through Amazon Key Management Service (AWS KMS)

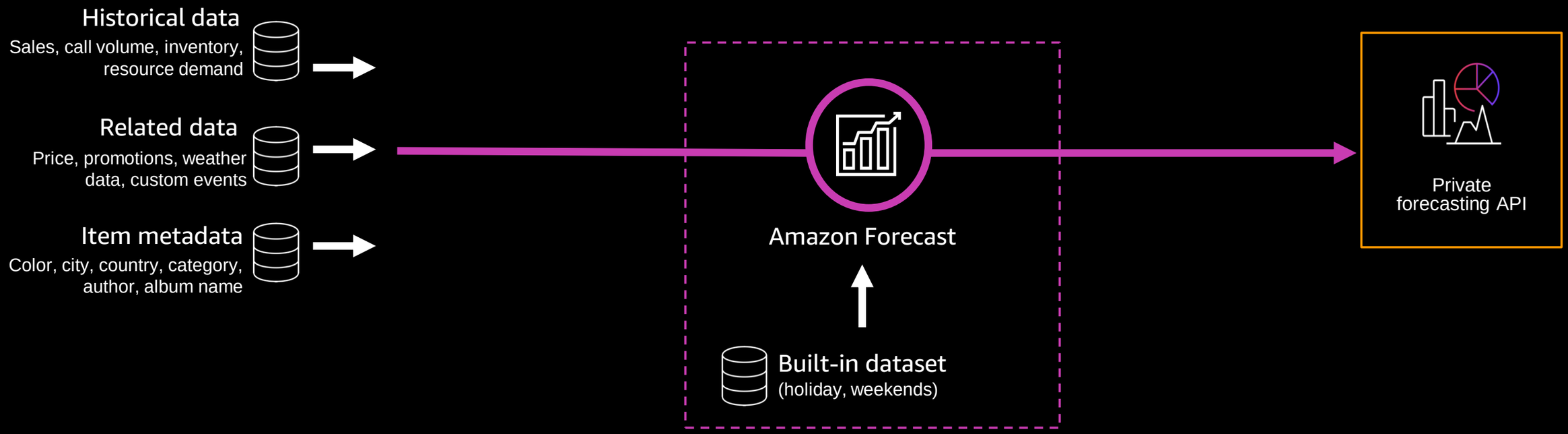
Amazon Forecast

Datasets used for forecasting

Use of historical data to predict future values	Target time-series dataset	The primary variable to predict with its historical values (demand, sales)
Use of known time-varying data specific to your business	Related time-series dataset	Time-varying related features that may impact the target value (price, promotion, weather)
Use of related attributes and categorical data	Item metadata (non-time-varying)	Categorical data that provide more context about items (color, city, channel)

Amazon Forecast

The technology that powers the world's largest ecommerce business



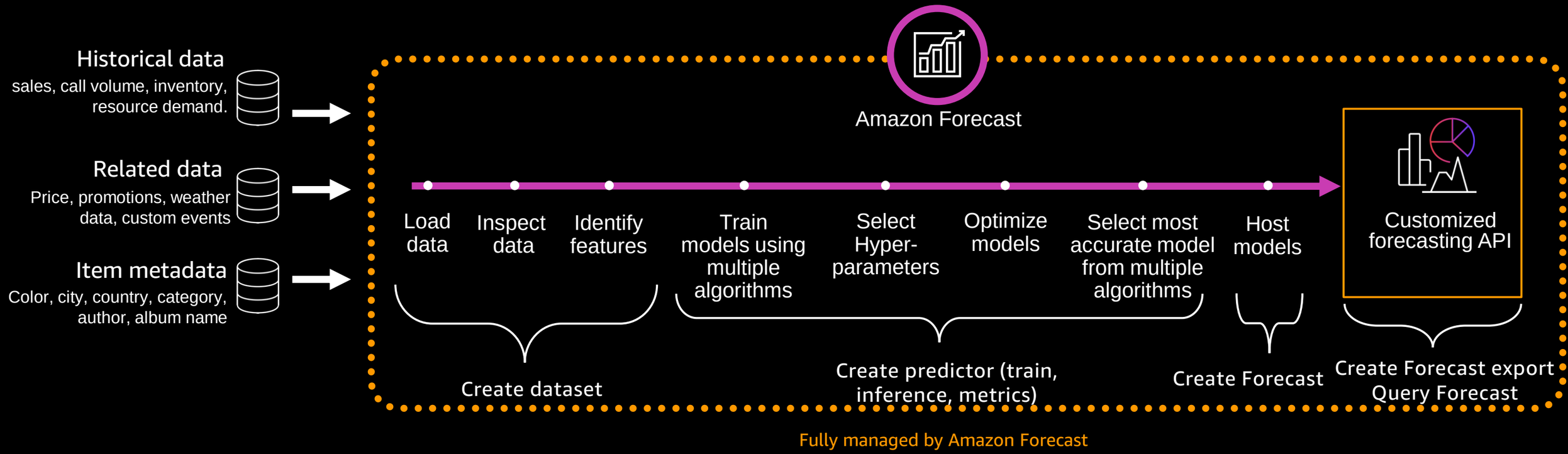
Get started with the console or API
Point Amazon Forecast to your data stored in Amazon Simple Storage Service (Amazon S3)

Automatically train your custom ML model
Let Amazon Forecast auto select the best one for your data through AutoML

Generate accurate forecasts
Retrieve forecasts through the console or private API

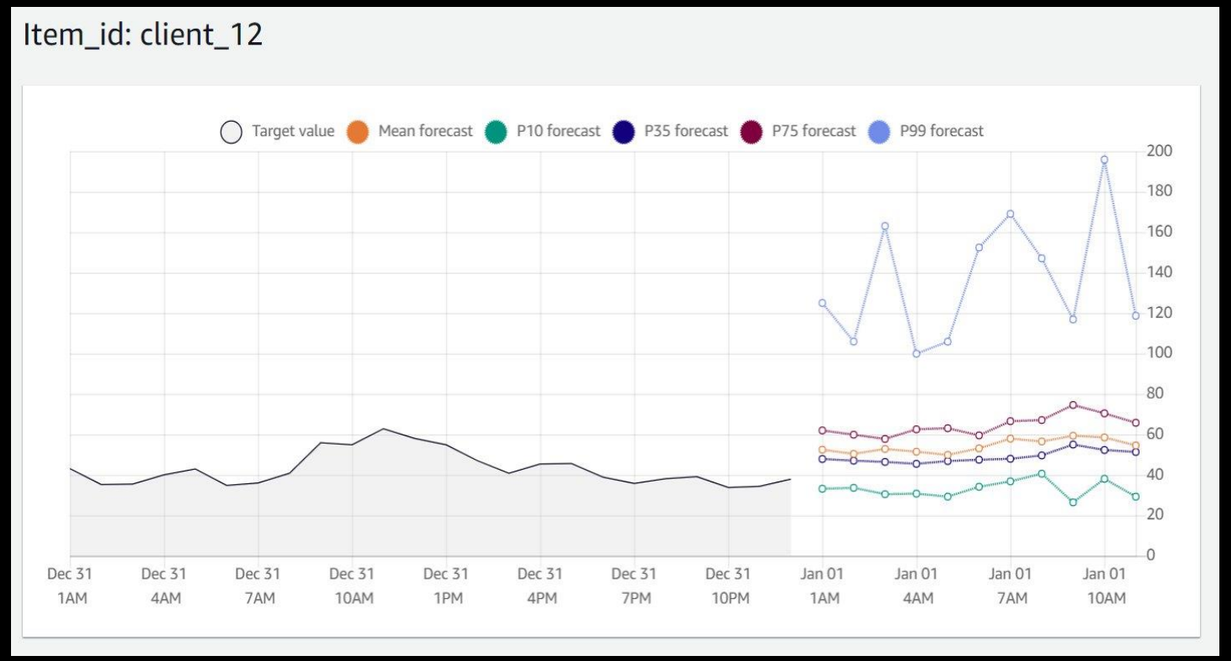
Amazon Forecast

Behind the scenes



Amazon Forecast

Visualize the distribution of forecasted values



View probabilistic forecasts at any quantile in the console

Retrieve forecasts through your private API

Export forecasts to .csv

Amazon Forecast

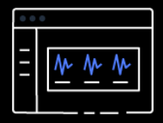
Handles tricky forecasting scenarios

Missing values

Product discontinuation

Cold start
(new product introduction)

Highly spiky data



Irregular seasonality

Sensitivity analysis
(future price change)

Pricing for Amazon Forecast

Pay-as-you-go pricing model

Cost type	**Pricing**
Generated forecasts	$0.60 per 1,000 forecasts (rounded to the nearest 1,000)
Data storage	$0.088 per GB
Training hours	$0.24 per hour

Deep dive: The algorithms powering Amazon Forecast

AWS re:Invent

aws

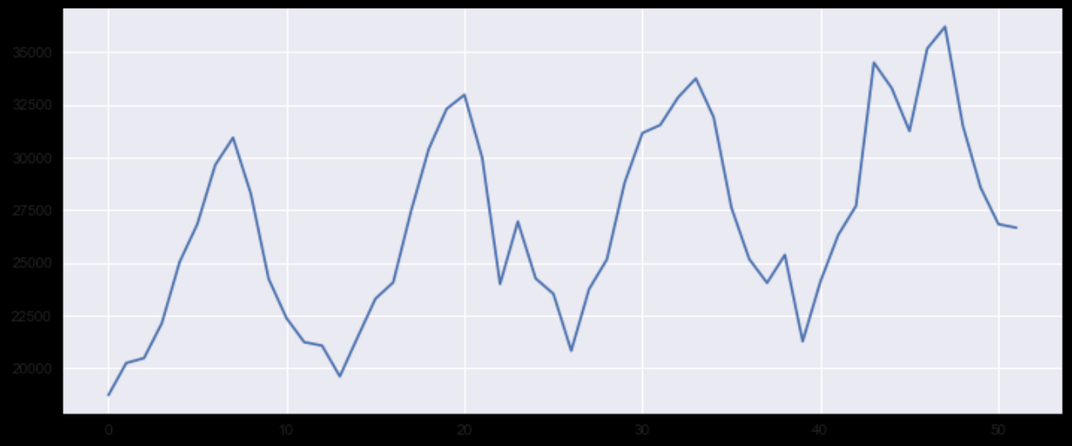
Algorithms powering Amazon Forecast

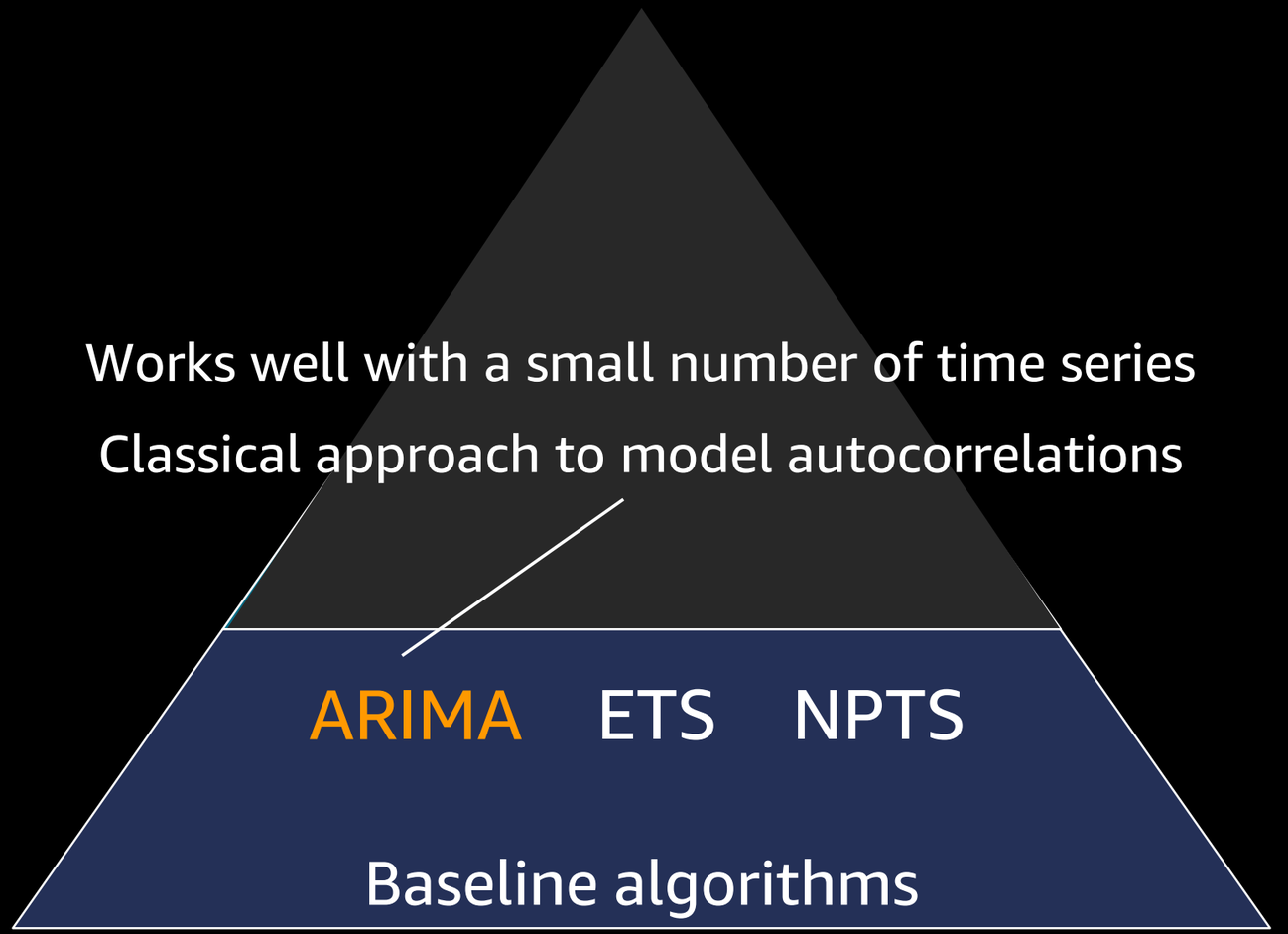
Amazon Forecast includes a gallery of statistical and machine learning forecasting algorithms

ARIMA

Auto-**r**egressive **i**ntegrated **m**oving **a**verage

De facto statistical method





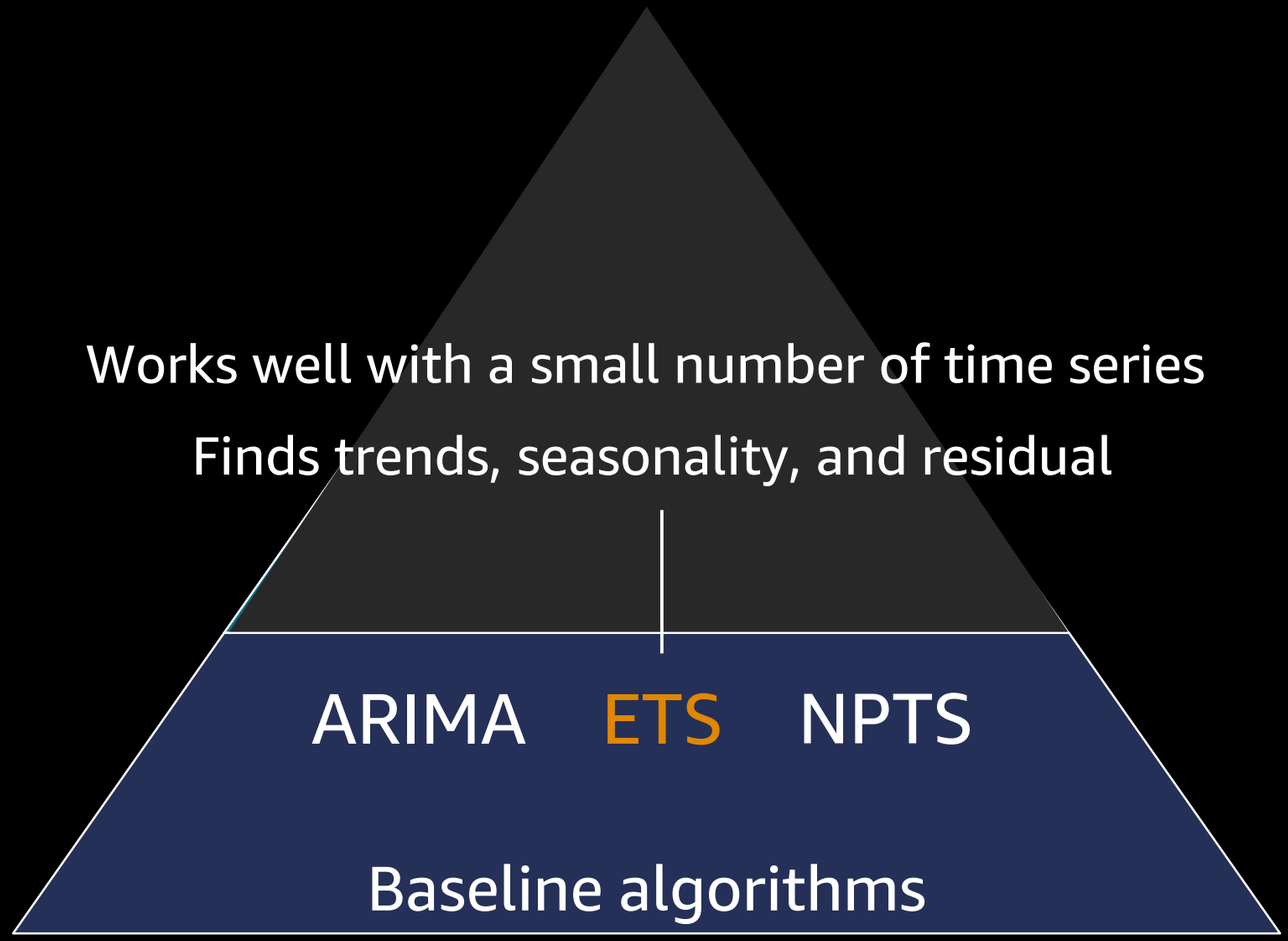
Algorithms powering Amazon Forecast

Amazon Forecast includes a gallery of statistical and machine learning forecasting algorithms

ETS

Error trend seasonality

Statistical algorithm that uses exponential smoothing

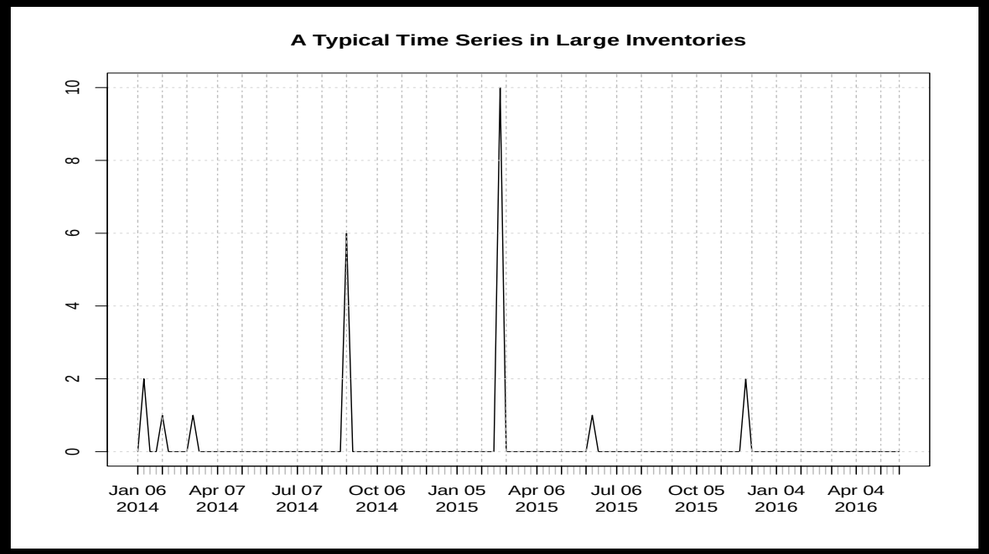


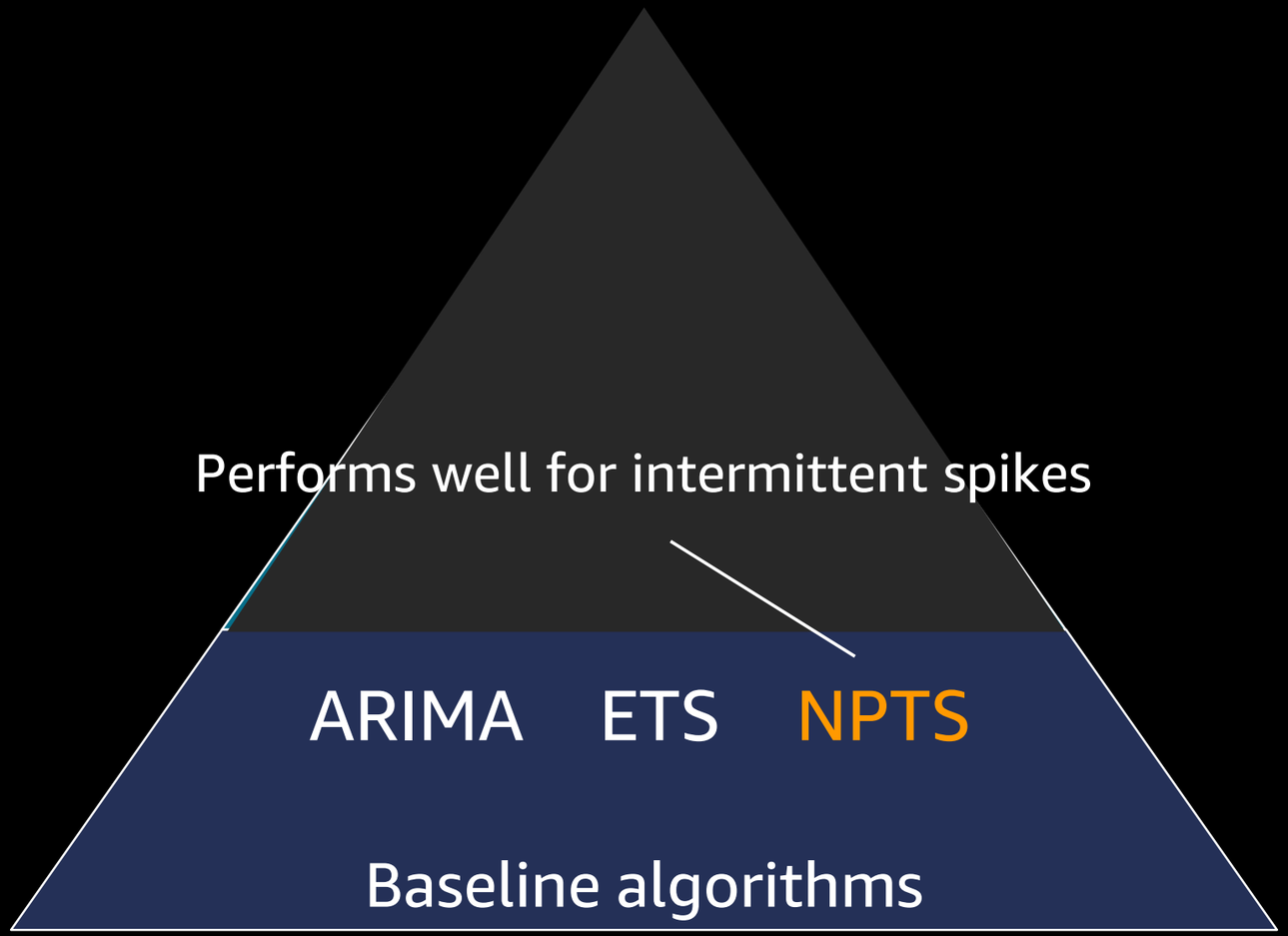
Algorithms powering Amazon Forecast

Amazon Forecast includes a gallery of statistical and machine learning forecasting algorithms

NPTS

Non-parametric time series



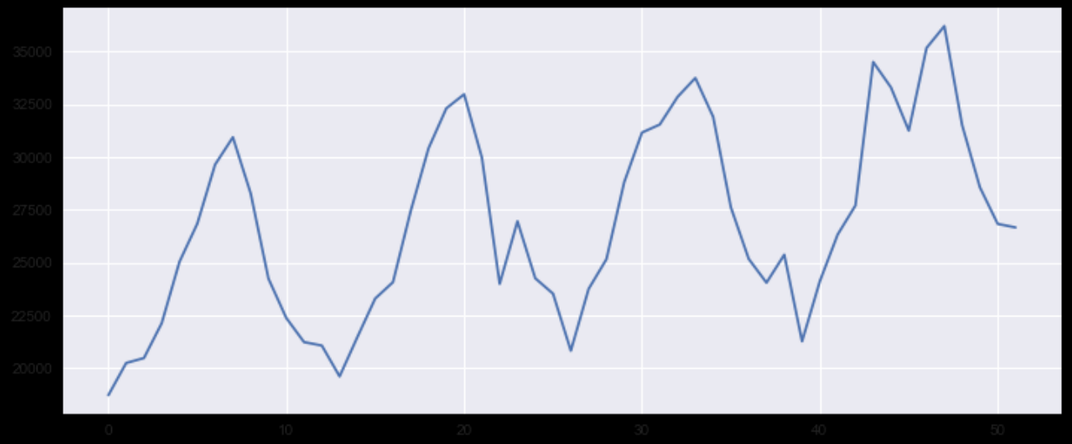


Algorithms powering Amazon Forecast

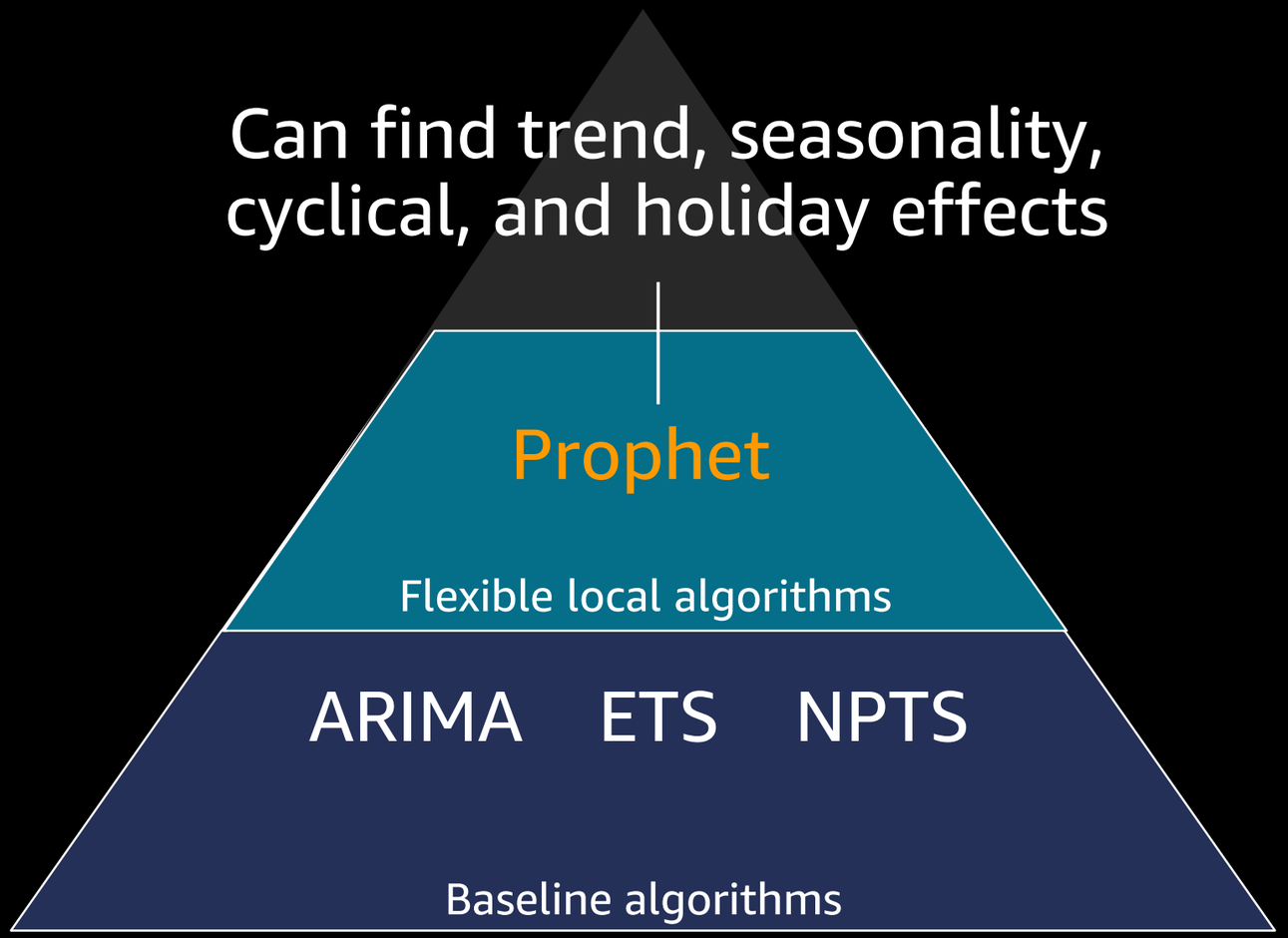
Amazon Forecast includes a gallery of statistical and machine learning forecasting algorithms

Prophet

Additive model with Gaussian likelihood



Can find trend, seasonality, cyclical, and holiday effects

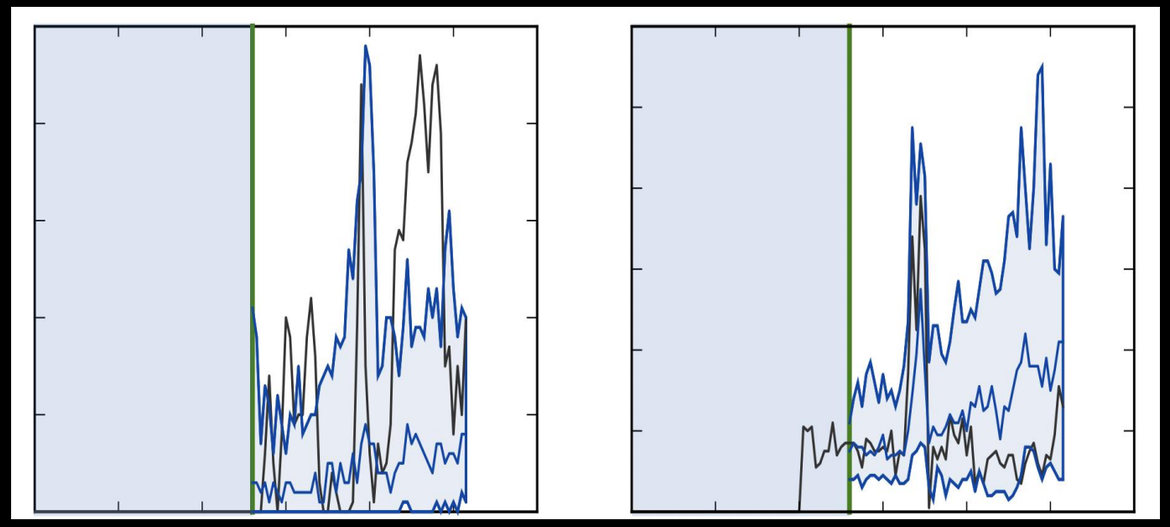


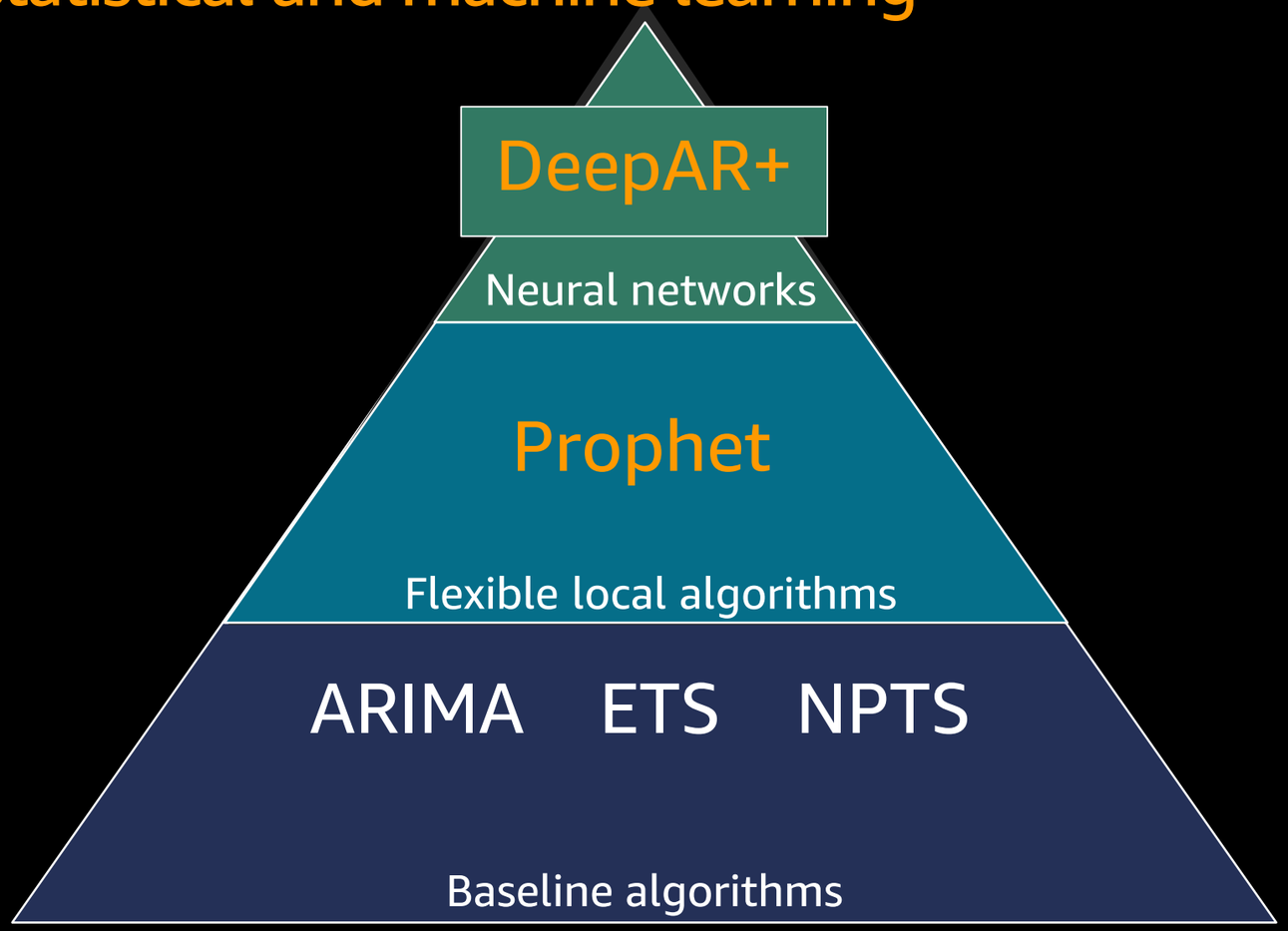
Algorithms powering Amazon Forecast

Amazon Forecast includes a gallery of statistical and machine learning forecasting algorithms

DeepAR+

Global model that can use related time series and attributes





How is DeepAR+ different

- The DeepAR+ forecasting algorithm has been used internally in Amazon for mission-critical decisions

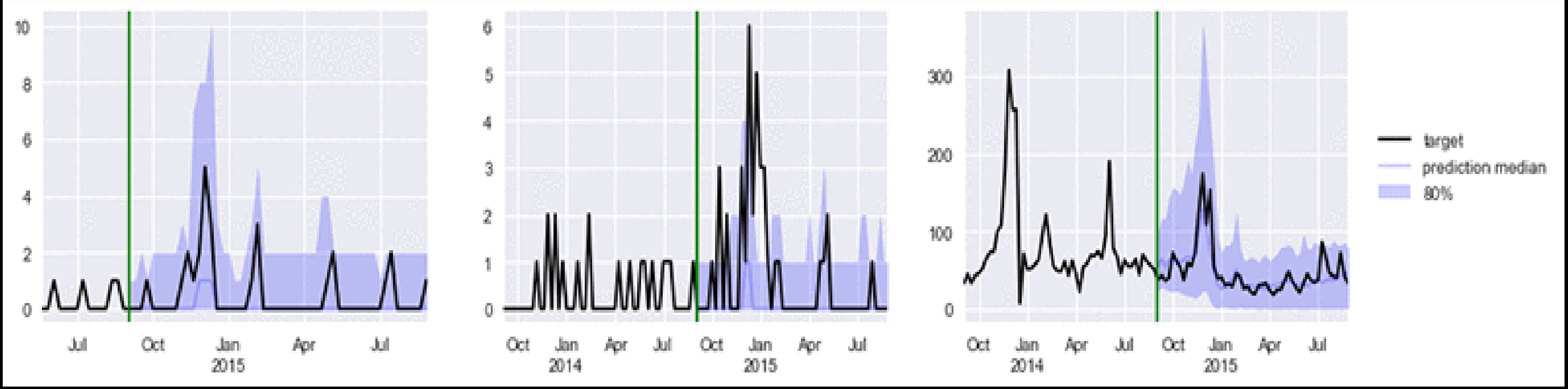
- Classical forecasting techniques such as ARIMA and ETS fit one model to an individual time series. However, in many situations, a set of related time series have been or can be collected.

- DeepAR+ can train a model over such a set of related time series for additional insights and increased predictive power

- Requires minimal feature engineering and can produce forecasts that are either point (amount sold was X) or probabilistic (amount sold was between X and Y with Z probability).

DeepAR+ for time series forecasting

- Modeling many related times series: Over 100,000 at one time
- Cold-start forecasting: Based on data from related time series

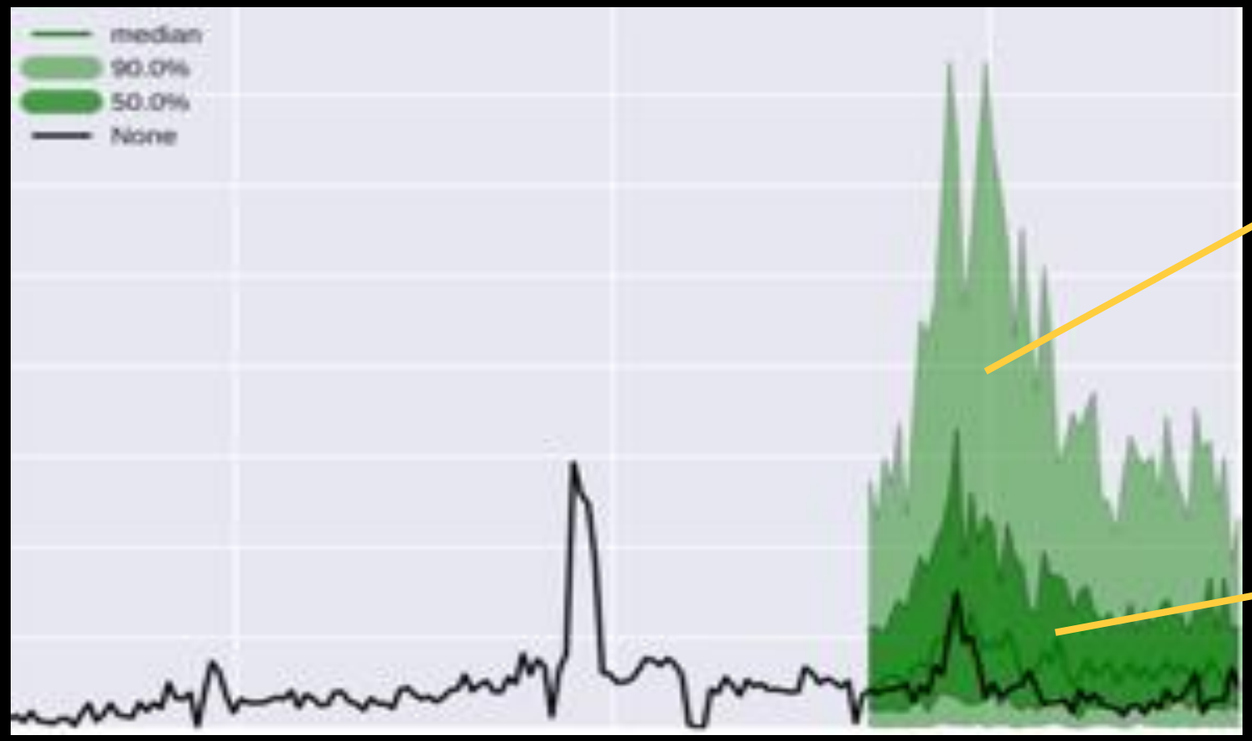


Evaluating your Forecast

Evaluating your forecast

Select the quantile most appropriate for your use case

Quantiles are used in probabilistic forecasts to provide ranges of predicted values



Examples

A P90 quantile predicts that, 90% of the time, the true value will be less than the predicted value

A P50 quantile predicts that, 50% of the time, the true value will be less than the predicted value

Evaluating your forecast

Understanding error and loss metrics

Error/loss functions are used to evaluate machine learning models by calculating the error between true and predicted values. Error functions are chosen on the basis of the algorithms and type of data.

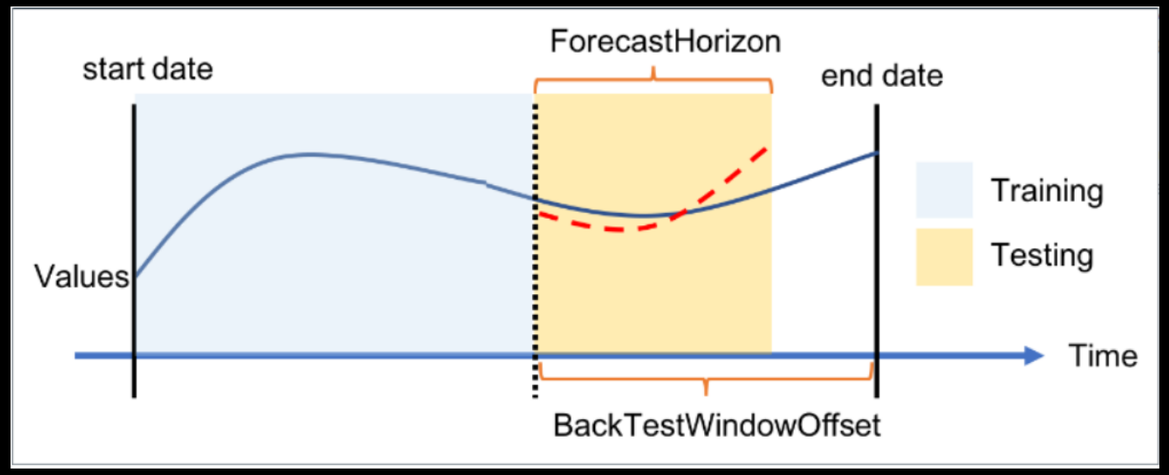
Weighted quantile loss (wQuantileLoss) calculates how far off the forecast a certain quantile is from actual demand

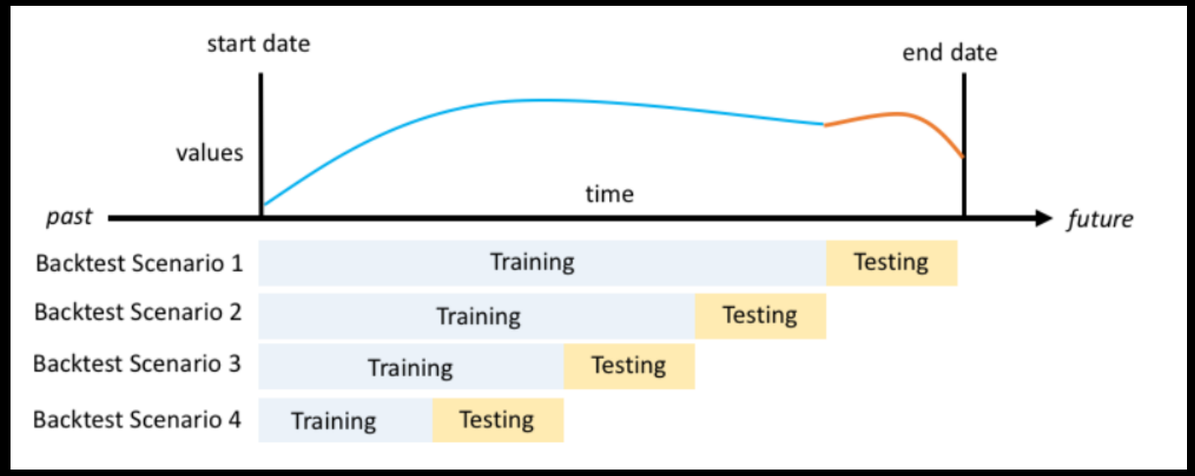
RMSE (root mean square error) calculates the difference between the actual target value and the predicted (forecasted) mean value

Many more: MAPE, MASE, log loss

Evaluating your forecast: Back testing

How would your model perform when deployed at different points in time





Reduce excess inventory with better forecast accuracy

AWS re:Invent

aws

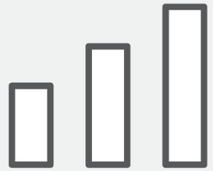
ThermoFisher SCIENTIFIC

thermo scientific | applied biosystems | invitrogen | fisher scientific | unity lab services

70,000
employees

5,000
R&D scientists/engineers

$1B
invested in R&D



$24B
in revenues

The world leader in serving science

ThermoFisher SCIENTIFIC

Organizational structure

ThermoFisher SCIENTIFIC

The world leader in serving science

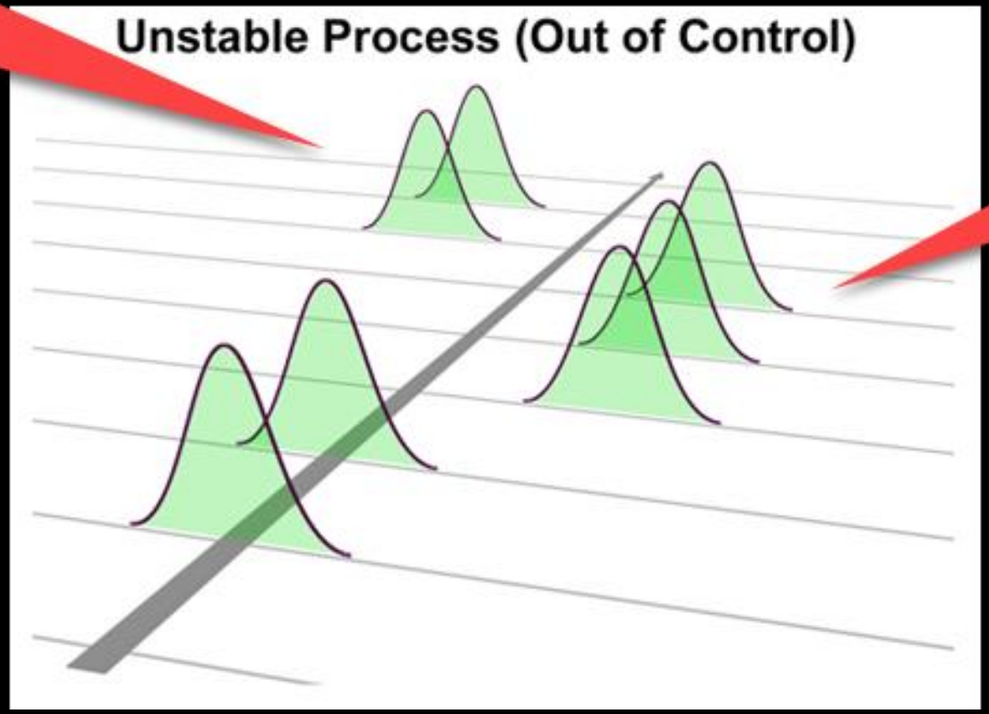
thermo scientific	applied biosystems	invitrogen	fisher scientific	unity lab services
Analytical precision and diagnostics excellence	Inspiring meaningful genetic analysis	Accelerating discovery research	One-stop access for scientific products	Instrument and enterprise services

Highly complex and diverse inventory to manage

158,000 SKUs from highly valuable devices to perishable consumables

Manufacturing Forecast inaccuracy is costly

Underforecasting creates customer availability issues and increased expedite costs



Overforecasting creates excess inventory and scrap.

Traditional forecast models

75006580	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Statistical	15	18	13	15	9	10	11	11	11	13	12	10
Override 1												
Override 2												
Override 3												
Forecast	15	18	13	15	9	10	11	11	11	13	12	10
ASP	15,468	15,468	15,468	15,468	15,468	15,468	15,468	15,468	15,468	15,468	15,468	15,468
Extended ASP	232,122	277,964	196,849	230,193	139,692	161,055	174,699	173,973	165,020	203,867	182,189	155,135

Formulas | Percent | 10 | Increment | 1 | Value | 0 | Override 1 | Commit | Help

Our approach to experimenting with and adopting Amazon Forecast

- Input file: 3 years of sales data for 158,000 SKUs for all sites
- Inspected and preprocessed data feed to remove:
  - Financial adjustments and internal transfers
  - SKUs with less than 3 transactions per year
  - SKUs with aggregate demand or price equal to or less than 0
- Used sales price as related time series and other item metadata information
- Trained and optimized models using various algorithms and parameters, including DeepAR+
- Focus group: 20,000 SKUs where we generated forecasts
- Conducted back-testing to evaluate accuracy and compared with our legacy system by looking at MAPE, WAPE, and mean bias metrics for various SKU groups

Amazon Forecast vs. legacy system results

Forecast MAPE accuracy results – Set A

Categories	Other (legacy system)	Amazon Forecast	Difference
W-Runner-High * GB01	78.78%	89.21%	10.43%
W-Runner-High * US02	61.25%	82.08%	20.83%
W-Runner-High * US03	68.55%	80.81%	12.26%
W-Runner-High * MTS Manuf	70.74%	79.87%	9.13%
W-Runner-High * MTS Purch	69.26%	84.55%	15.29%
W-Runner-High * Prod Line FCR	72.75%	83.44%	10.69%

Amazon Forecast Deep AR+ showed better results in all the SKU groups evaluated

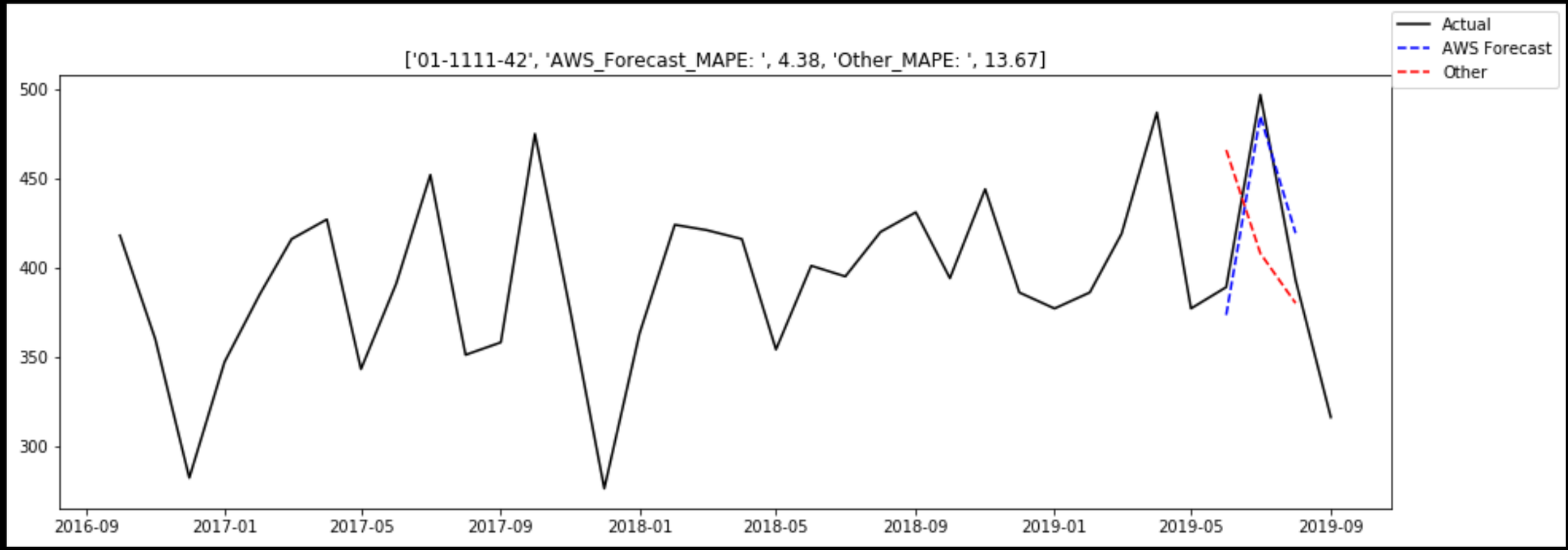
Amazon Forecast vs. legacy system results

Forecast MAPE Accuracy Results – Set B

Categories	Other (legacy system)	Amazon Forecast	Difference
Primary Site: U50	16.40%	43.49%	27.09% ↑
Primary Site: UM	-69.78%	20.15%	89.93% ↑
Prod Line Group: 1C7	5.06%	36.99%	31.93% ↑
Prod Line Group: ANT	40.60%	53.20%	12.60% ↑
Prod Line Group: ELI	17.90%	42.94%	25.04% ↑
Item Group Code: INV250	19.45%	40.45%	21.00% ↑
Stocking Type: MTS Purch Catalog-Vendor	14.32%	48.45%	34.13% ↑

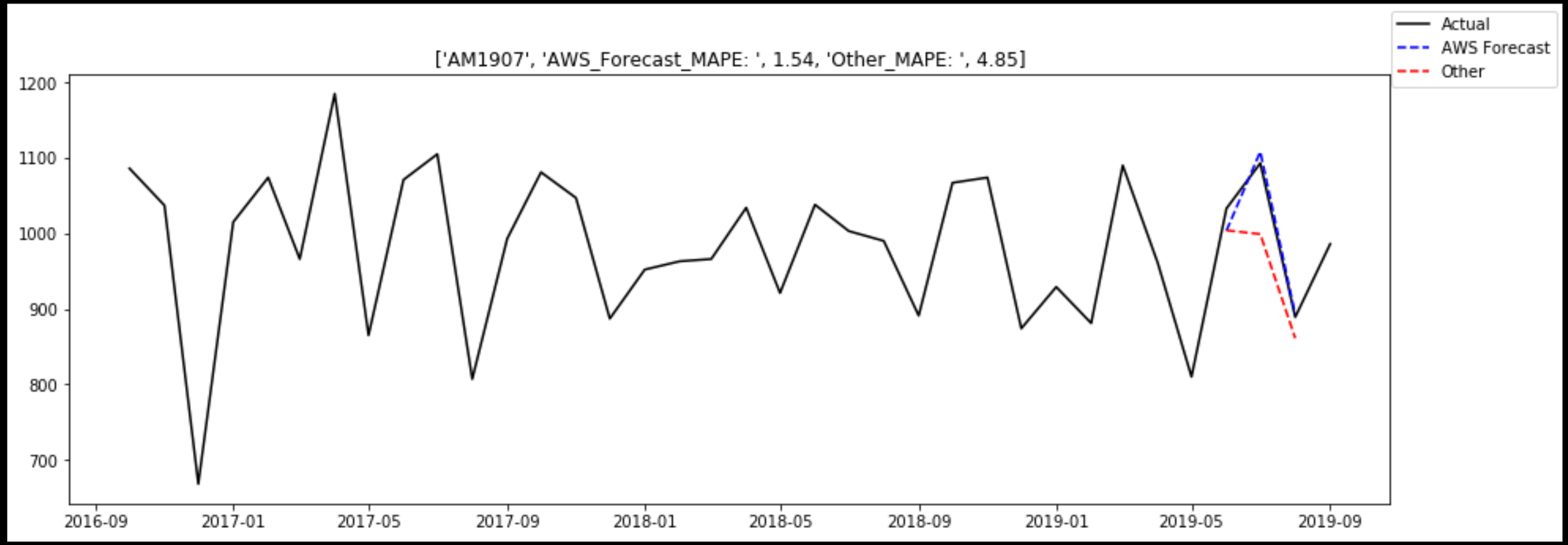
Amazon Forecast inventory accuracy visualization

Visualization - Example 1



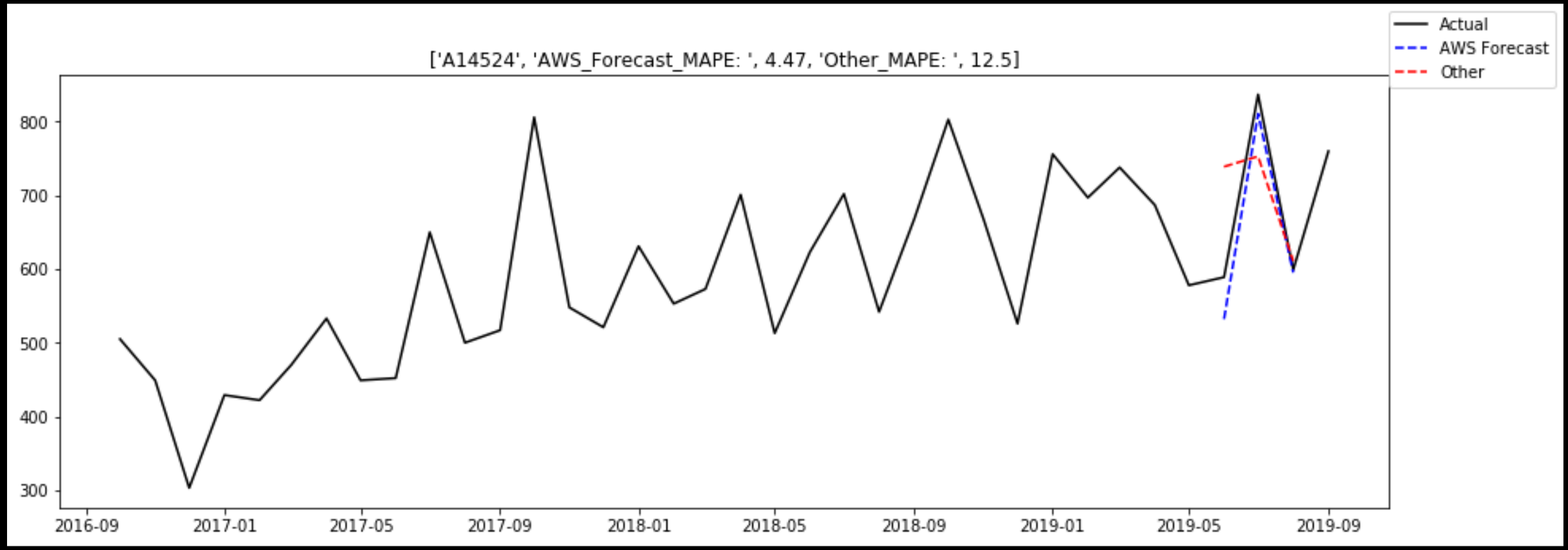
Amazon Forecast inventory accuracy visualization

Visualization - Example 2



Amazon Forecast inventory accuracy visualization

Visualization - Example 3



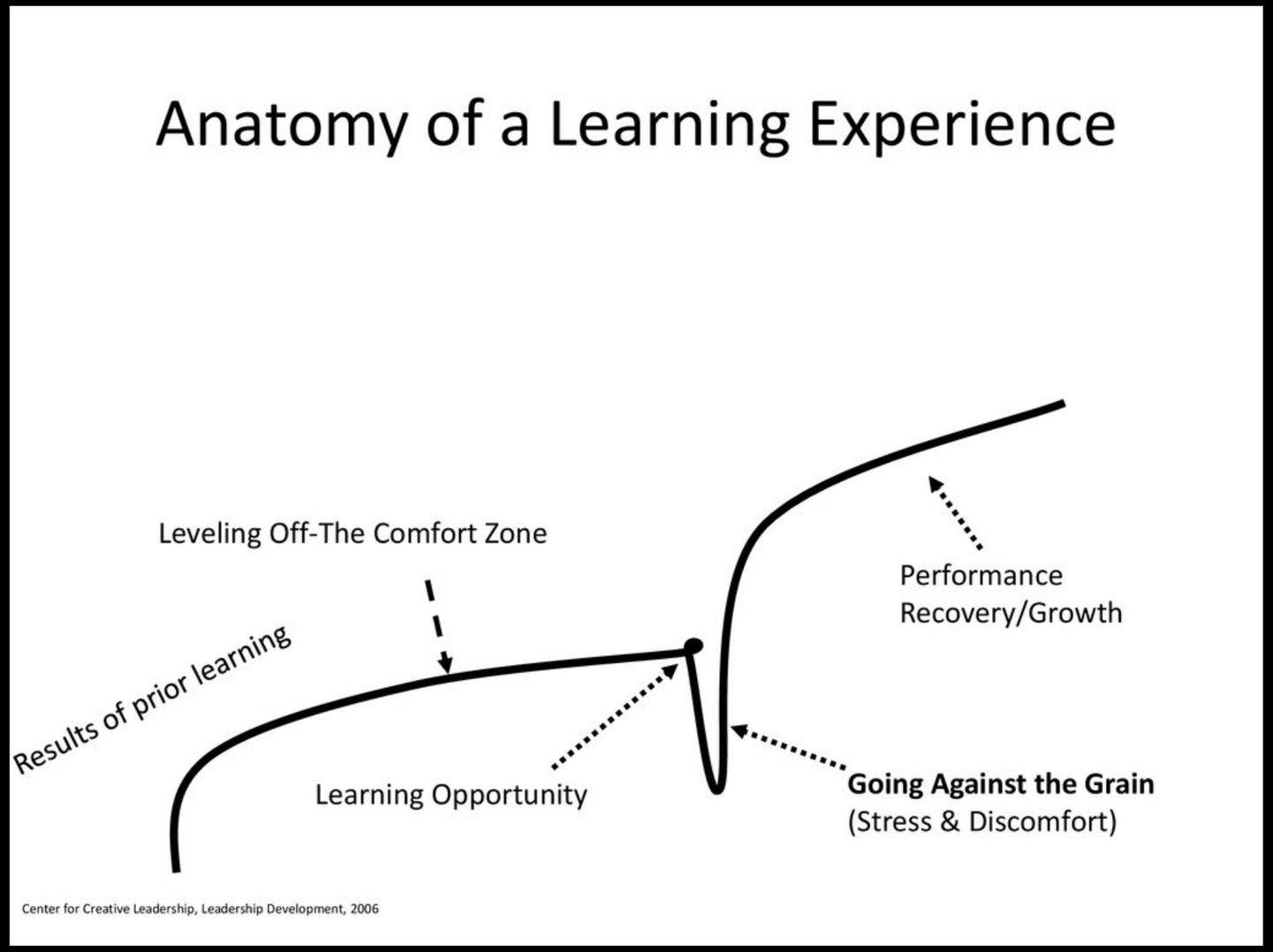
Increase in forecast accuracy translates to millions of dollars of reduction of inventory and/or improvement in managing backorders

Amazon Forecast vs. legacy system TCO

	Other (legacy system)	Amazon Forecast
Installation time	2–8 months	6–8 weeks
Pricing model	License fee	Pay-as-you-go model
Maintenance costs (3 years)	Up to a $1,000,000	$6,000*
Professional services	Up to $150K	As needed

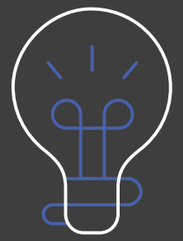
*Includes training, model maintenance for monthly forecast for 35,000 SKUs, and two quantiles

Our AWS AI/ML Journey



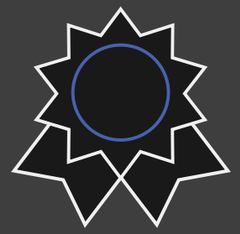
Using Amazon Forecast will allow us to **leap-frog to better results** after overcoming initial fears and discomfort of learning a new technology

Thermo Fisher Scientific and AWS collaboration



AWS Partner & ML Solutions Lab

- Brainstorming
- Custom modeling
- Training
- Work side-by-side with Amazon experts



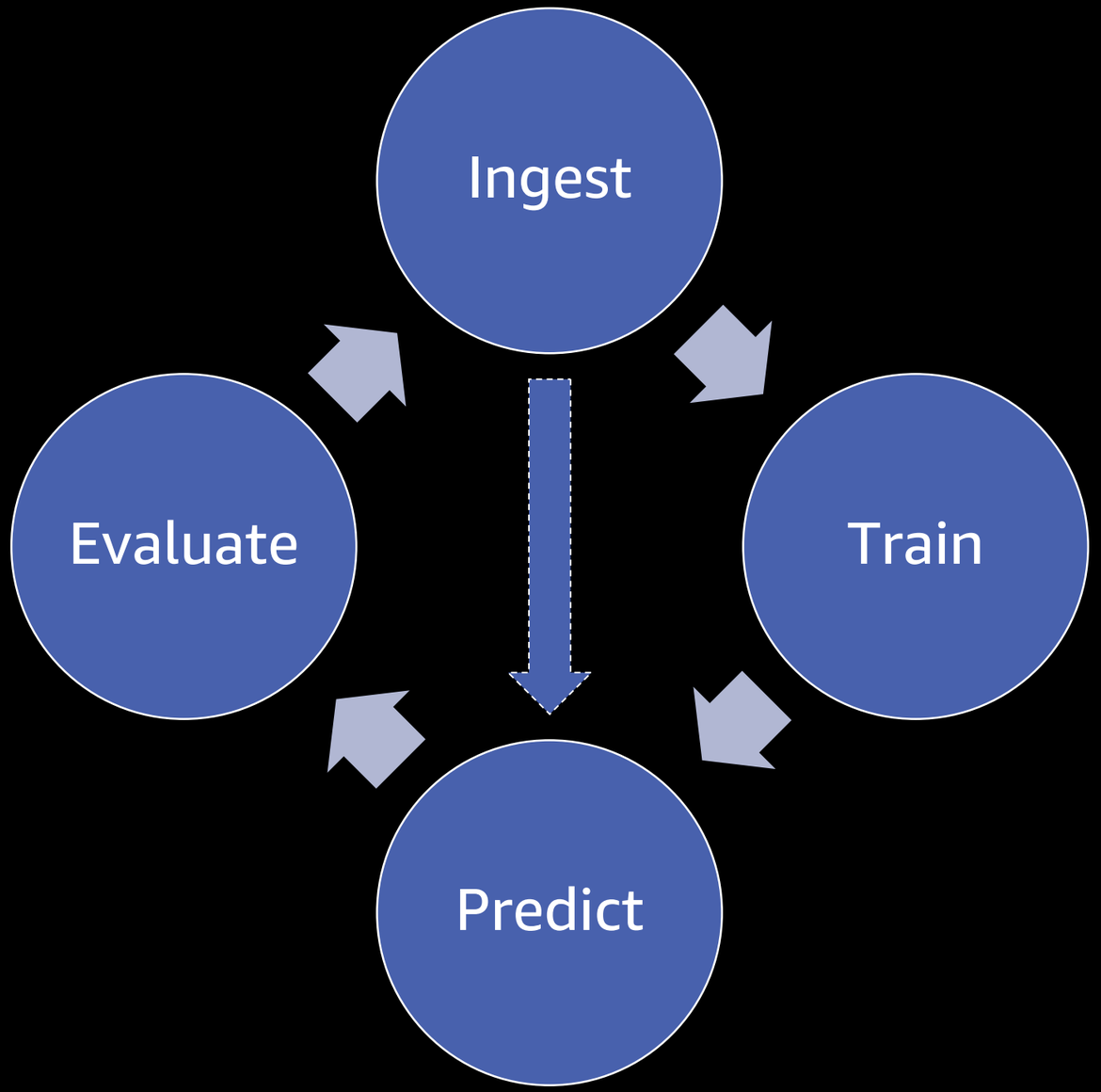
Machine learning training and certification

- Practical education on ML for new and experienced practitioners
- Based on the same material used to train Amazon developers

Amazon Forecast in production

Amazon Forecast in production

- Automate into a repeating process
- Define your integration points and create data pipelines for input data sources
- Continually infer new predictions
- Train models on an as need basis
- Automate evaluation and trigger based on error thresholds

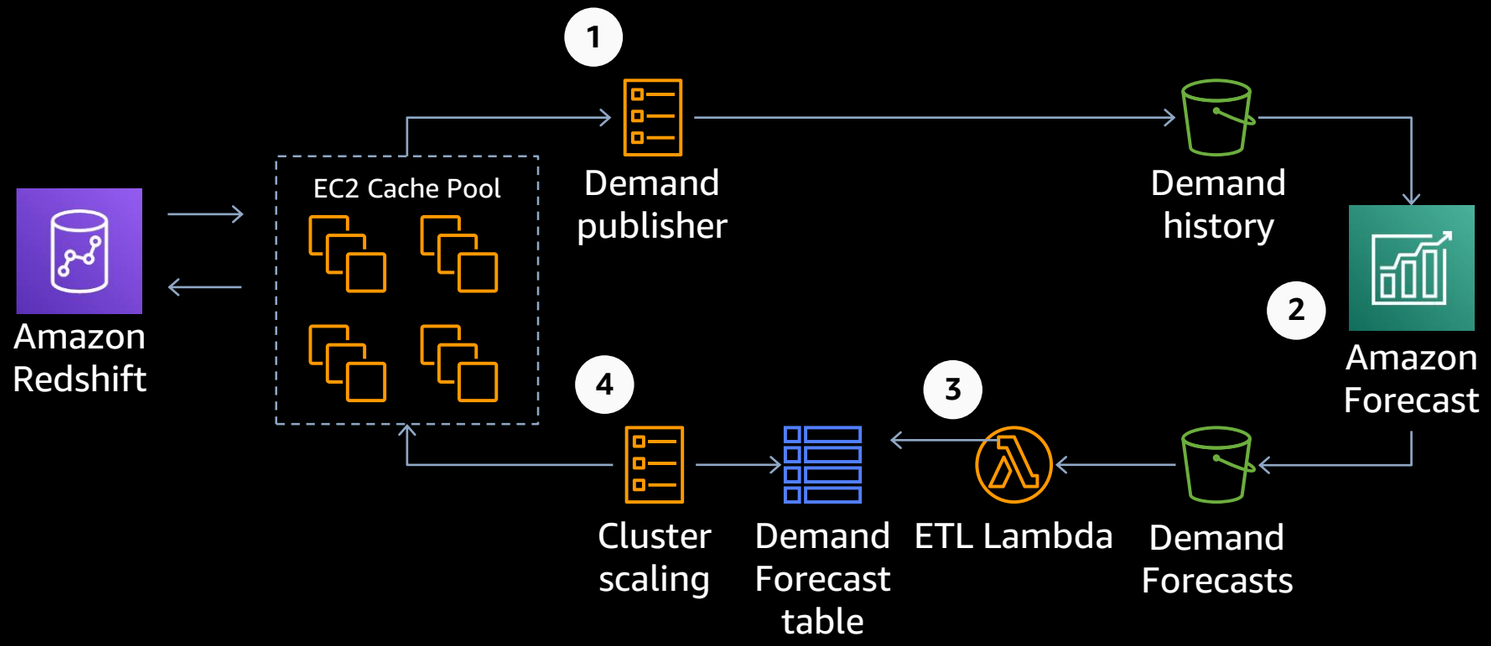


Amazon Forecast in production

Amazon Redshift Cluster Management

Manages a set of EC2 instances
Maintains cache pools for EC2 instances
Needs to forecast cache pool size

1. EC2 Cache pool demand changes are published to S3 bucket
2. New data is ingested into Amazon Forecast and new forecast predictions are stored in S3 bucket
3. A Lambda function copies new forecasts to an Amazon DynamoDB table
4. The cluster scaling logic reads forecasts and adjust the cache pool size based on projected demand



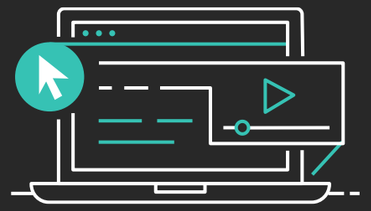
Working with the AWS Partner Network

Partners with capabilities in Amazon Forecast

Learn ML with AWS Training and Certification

The same training that our own developers use, now available on demand

Role-based ML learning paths for developers, data scientists, data platform engineers, and business decision makers



70+ free digital ML courses from AWS experts let you learn from real-world challenges tackled at AWS

Validate expertise with the
AWS Certified Machine Learning - Specialty exam

Visit https://aws.training/machinelearning

aws training and certification

Thank you!

AWS re:Invent

aws

Please complete the session survey in the mobile app.