AWS re:Invent

AIM307

Amazon SageMaker: A deep dive

Shyam Srinivasan
Product Marketing Lead, ML
Amazon Web Services

Julien Simon
Principal Evangelist, AI/ML
Amazon Web Services

Nils Mohr
Flight Data Recorder Programming Analyst
British Airways


“Great things are never done by one person, they’re done by a team of people.”
—Steve Jobs

Agenda

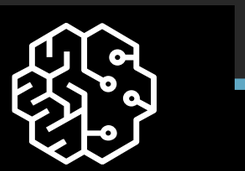
What is Amazon SageMaker?

Why did we build Amazon SageMaker?

Ongoing challenges with machine learning and how to overcome them

Demo

Seeing is believing – Amazon SageMaker at British Airways

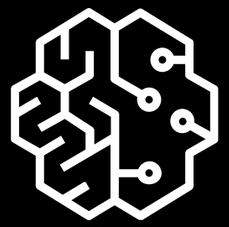


Amazon SageMaker

Machine learning for every developer & data scientist
Build, train, and deploy ML models at scale

Amazon SageMaker
Getting models from concept to production

Amazon SageMaker
Build, train, and deploy machine learning models quickly at scale



Amazon SageMaker

Ground Truth

ML Marketplace

Algorithms & Frameworks

Notebooks

Training & Tuning

Reinforcement Learning

Neo

Deployment & Hosting

But, what are we seeing as the ongoing challenges to machine learning?

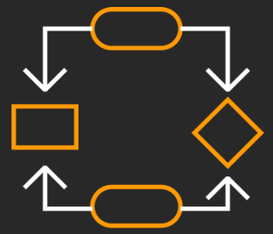
Amazon SageMaker

Addressing challenges to machine learning

First fully integrated development environment (IDE) for machine learning
Amazon SageMaker Studio

Enhanced notebook experience with quick-start & easy collaboration
Amazon SageMaker Notebooks (Preview)

Experiment management system to organize, track & compare thousands of experiments
Amazon SageMaker Experiments

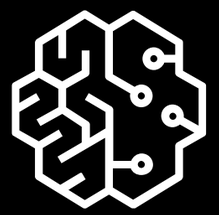


Automatic debugging, analysis, and alerting
Amazon SageMaker Debugger

Model monitoring to detect deviation in quality & take corrective actions
Amazon SageMaker Model Monitor

Automatic generation of ML models with full visibility & control
Amazon SageMaker Autopilot

Build, train, and deploy machine learning models quickly at scale



Amazon SageMaker

NEW! Amazon SageMaker Studio IDE

Ground Truth | ML Marketplace | Algorithms and Frameworks | NEW! SageMaker Notebooks | Training and Tuning | Reinforcement Learning | NEW! SageMaker Experiments | NEW! SageMaker Debugger | NEW! SageMaker Autopilot | Neo | Deployment and Hosting | NEW! SageMaker Model Monitor

Machine learning is iterative, involving dozens of tools and hundreds of iterations

Multiple tools needed for different phases of the ML workflow

+

Lack of an integrated experience

+

Large number of iterations

=

Cumbersome, lengthy processes, resulting in loss of productivity

NEW

Introducing Amazon SageMaker Studio

The first fully integrated development environment (IDE) for machine learning

Collaboration at scale

Share scalable notebooks without tracking code dependencies

Easy experiment management

Organize, track, and compare thousands of experiments

Automatic model generation

Get accurate models for with full visibility & control without writing code

Higher quality ML models

Automatically debug errors, monitor models & maintain high quality

Increased productivity

Code, build, train, deploy & monitor in a unified visual interface

Data science and collaboration needs to be easy

Set up and manage resources

+

Collaboration across multiple data scientists

+

Different data science projects have different resource needs

=

Managing notebooks and collaborating across multiple data scientists is highly complicated

NEW

Introducing Amazon SageMaker Notebooks

(Available in Preview)

Fast-start shareable notebooks

Easy access with Single Sign-On (SSO)

Access your notebooks in seconds with your corporate credentials

Fully managed and secure

Administrators manage access and permissions

No explicit setup

Start your notebooks without spinning up compute resources

Easy collaboration

Share your notebooks as a URL with a single click

Flexibility

Dial up or down compute resources

Managing trials and experiments is cumbersome

Thousands of experiments

+

Hundreds of parameters per experiment

+

Compare and evaluate

=

Very cumbersome and error-prone

NEW

Introducing Amazon SageMaker Experiments

Organize, track, and compare training experiments

Tracking at scale

Track parameters & metrics across experiments & users

Custom organization

Organize experiments by teams, goals & hypotheses

Visualization

Easily visualize experiments and compare

Metrics and logging

Log custom metrics using the Python SDK & APIs

Fast Iteration

Quickly go back & forth & identify the best performing experiments

Analysis & debugging while training ML models is painful

Large neural networks with many layers

+

Data capture with many connections

+

Additional tooling for analysis and debug

=

Extraordinarily difficult to inspect, debug, and profile the "black box"

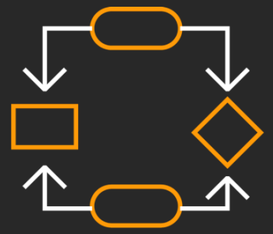
NEW

Introducing Amazon SageMaker Debugger

Analysis & debugging, explainability, and alert generation

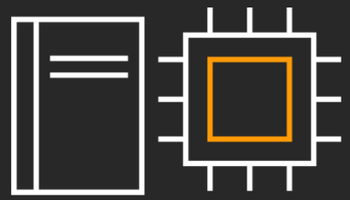
Relevant data capture

Data is automatically captured for analysis



Data analysis & debugging

Analyze & debug data with no code changes



Automatic error detection

Errors are automatically detected based on rules

Improved productivity with alerts

Take corrective action based on alerts

Visual analysis and debugging

Visually analyze & debug from Amazon SageMaker Studio

Deploying a model is not the end

You need to continuously monitor models in production and iterate

Concept drift due to divergence of data

+

Model performance can change due to unknown factors

+

Continuous monitoring involves a lot of tooling and expense

=

Model monitoring is cumbersome but critical

NEW

Introducing Amazon SageMaker Model Monitor
Continuous monitoring of models in production

Automatic data collection

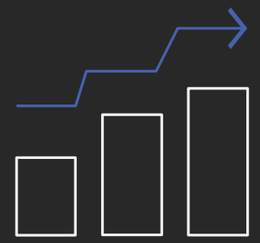
Data is automatically collected from your endpoints

Continuous monitoring

Define a monitoring schedule and detect changes in quality against a pre-defined baseline

Flexibility with rules

Use built-in rules to detect data drift or write your own rules for custom analysis



Visual data analysis

See monitoring results, data statistics, and violation reports in Amazon SageMaker Studio

Amazon CloudWatch integration

Automate corrective actions based on Amazon CloudWatch alerts

Successful ML requires complex, hard to discover combinations

of algorithms, data, parameters

Largely explorative & iterative

+

Requires broad and complete knowledge of ML domain

+

Lack of visibility

=

Time-consuming, Error-prone process, even for ML experts

NEW

Introducing Amazon SageMaker Autopilot
Automatic model creation with full visibility & control

Quick to start

Provide your data in a tabular form & specify target prediction

Automatic model creation

Get ML models with feature engineering & automatic model tuning automatically done

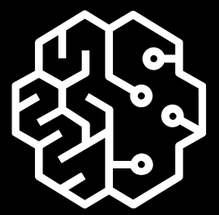
Visibility & control

Get notebooks for your models with source code

Recommendations & optimization

Get a leaderboard & continue to improve your model

Build, train, and deploy machine learning models quickly at scale



Amazon SageMaker

NEW! Amazon SageMaker Studio IDE

Ground Truth | ML Marketplace | Algorithms and Frameworks | NEW! SageMaker Notebooks | Training and Tuning | Reinforcement Learning | NEW! SageMaker Experiments | NEW! SageMaker Debugger | NEW! SageMaker Autopilot | Neo | Deployment and Hosting | NEW! SageMaker Model Monitor

Demo

Julien Simon
Principal Evangelist, AWS AI/ML

AWS re:Invent

aws

Machine learning at British Airways Engineering

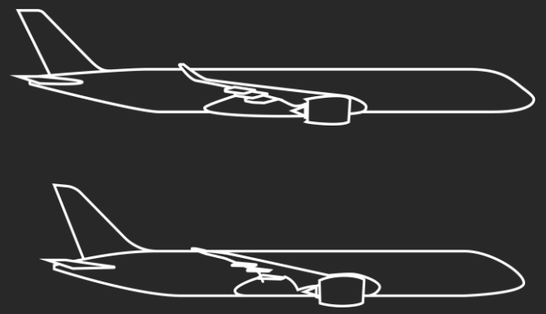
Nils Mohr
Flight Data Recorder Programming Analyst
British Airways

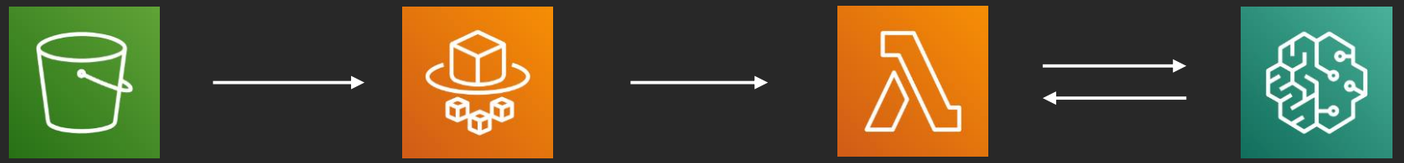
AWS re:Invent

aws


BRITISH AIRWAYS
Ai2
Aircraft Information Intelligence

Architecture & intentions





Aeroplane	Date	Flight No	Route	Condition a	Condition b	Condition c
G-ZBWW	12th Nov	BA 012	LHR → LAS	●	●	●
G-ZBWW	13th Nov	BA 013	LAS → LHR	●	●	●
G-ZBWW	13th Nov	BA 132	LHR → JFK	●	●	●
G-ZBWW	14th Nov	BA 133	JFK → LHR	●	●	●

BRITISH AIRWAYS

Using built-in algorithms in Amazon SageMaker

AWS re:Invent

aws

Using built-in algorithms

Unsupervised

- K-Means

  Clustering of discrete data points

- Random Cut Forest

  Anomaly detection in continuous data

Supervised

- XGBoost

  Further classification/root cause detection

- DeepAR

  Time-series forecasting

Our dataset

ACARS

QX QUKMEBA LONKFBA

.LONAXBA 221134

OFF01

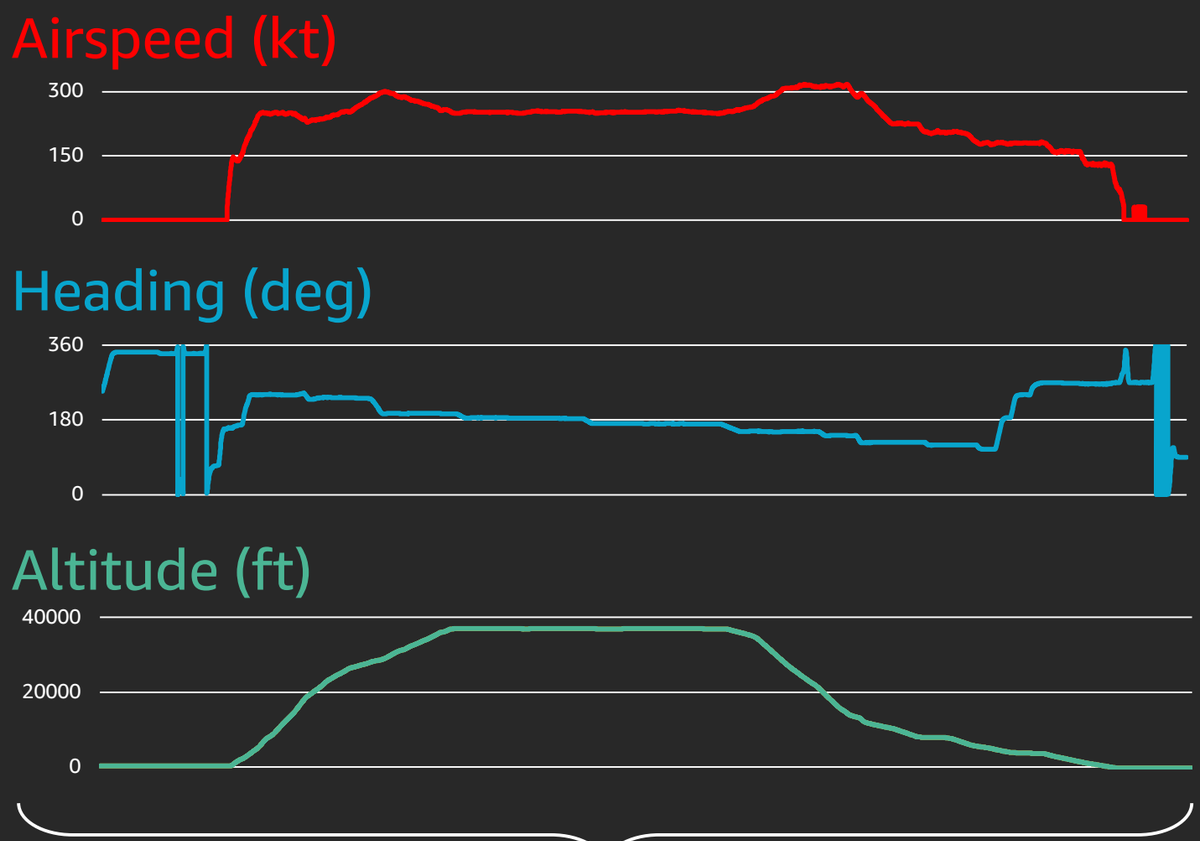
BA414 G-NEOV 221134

1134DIA02EGLLEDDT154853

15.4t fuel 85.3t take-off weight

Discrete values: K-Means

Timeseries data



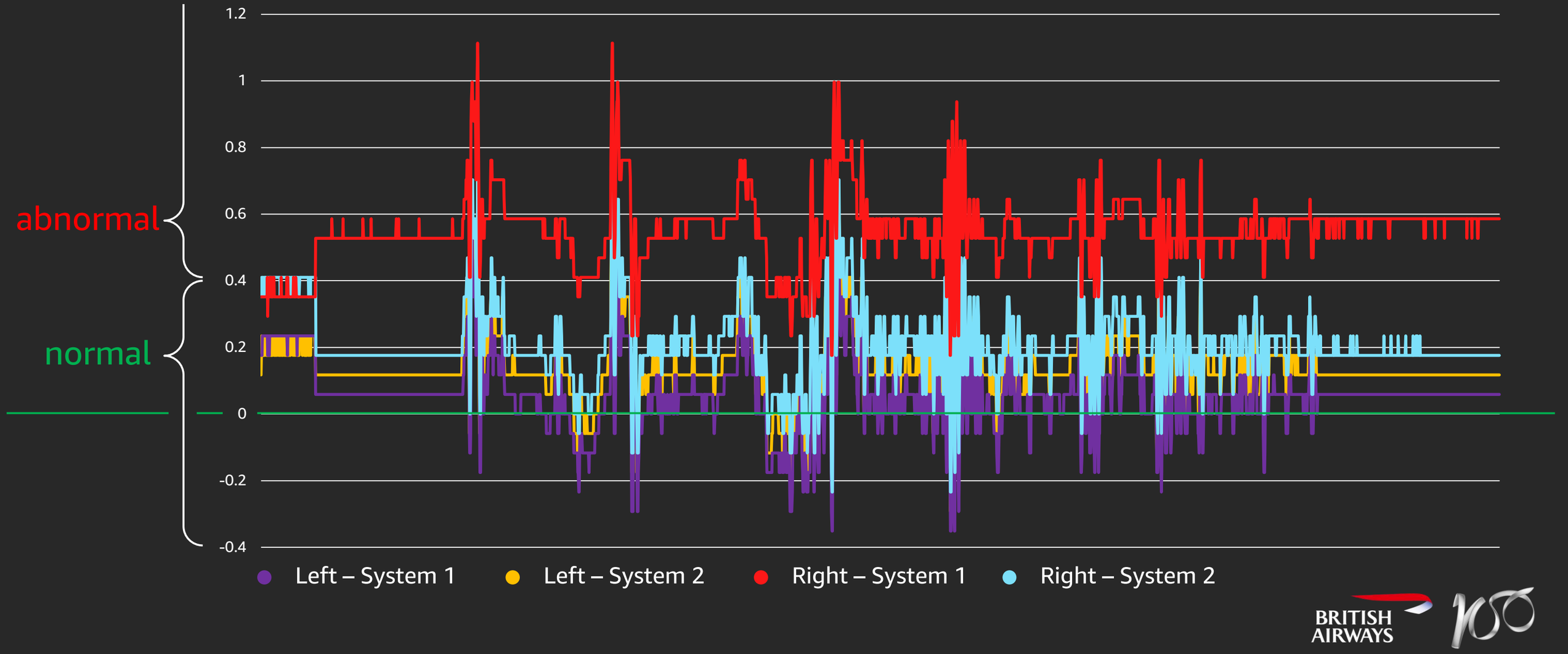
Continues data: RCF

BRITISH AIRWAYS

Example use case

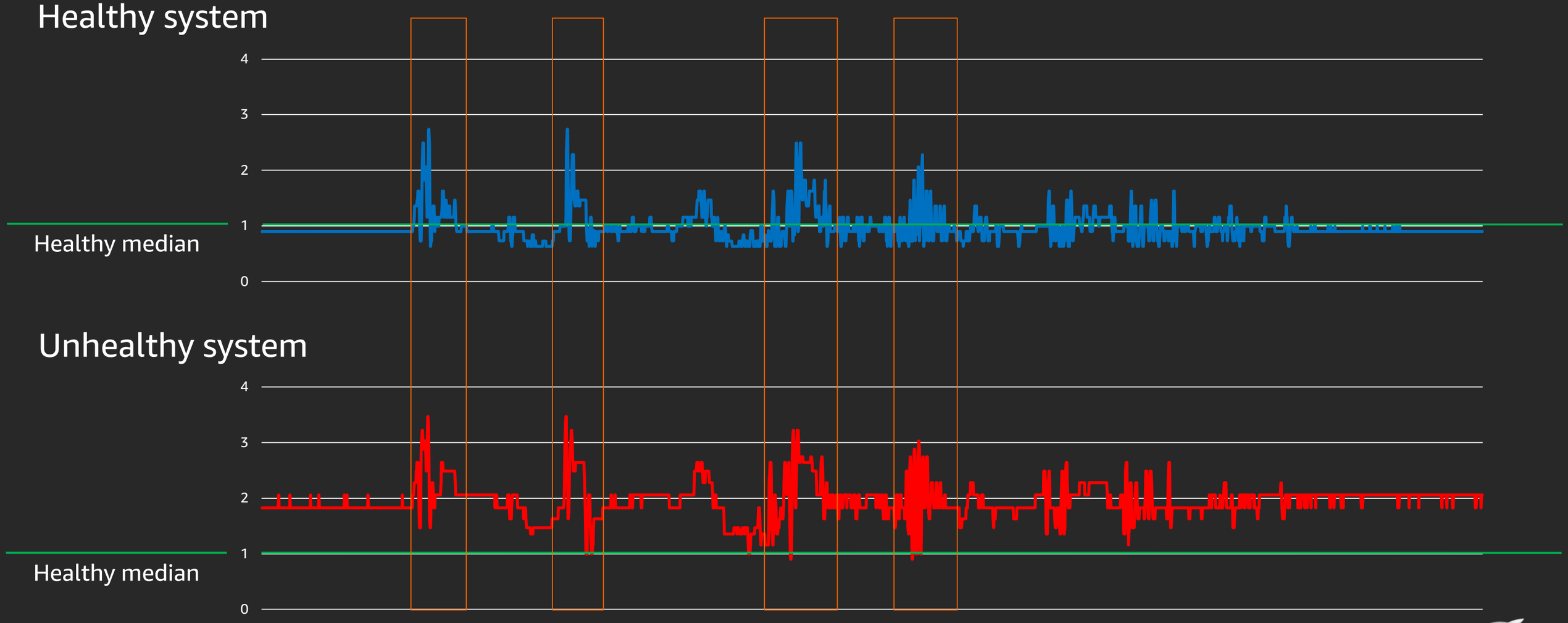
Example dataset

Decoded flight data, filtered to show only the relevant data



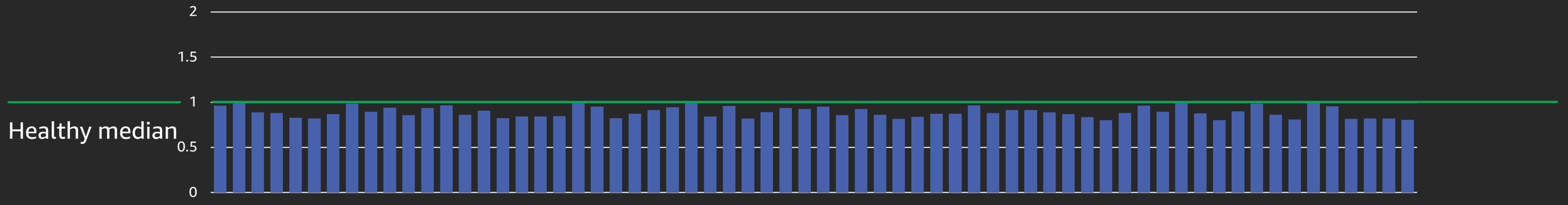
BRITISH AIRWAYS

Resulting Amazon SageMaker scores

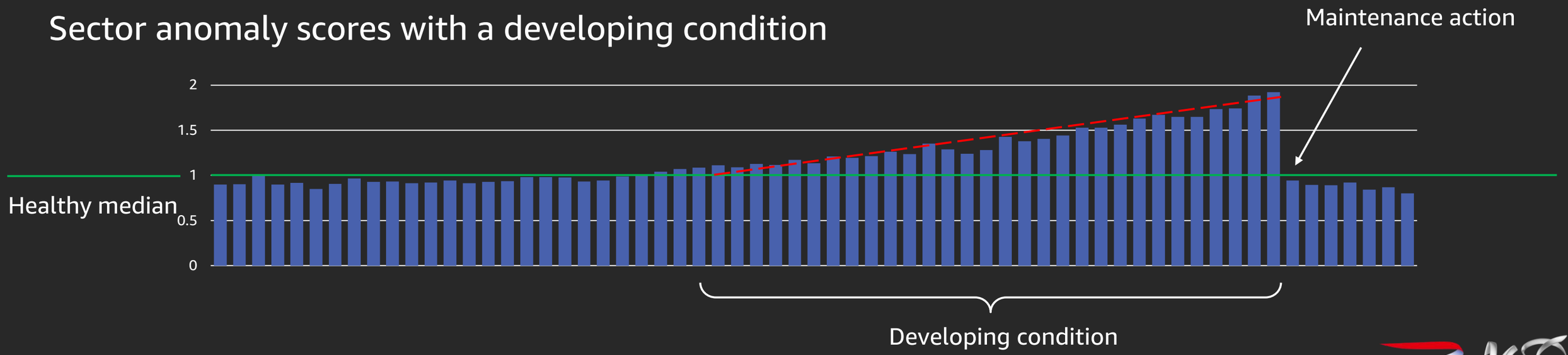


Condition monitoring over multiple flights

Sector anomaly scores on a healthy system



Sector anomaly scores with a developing condition



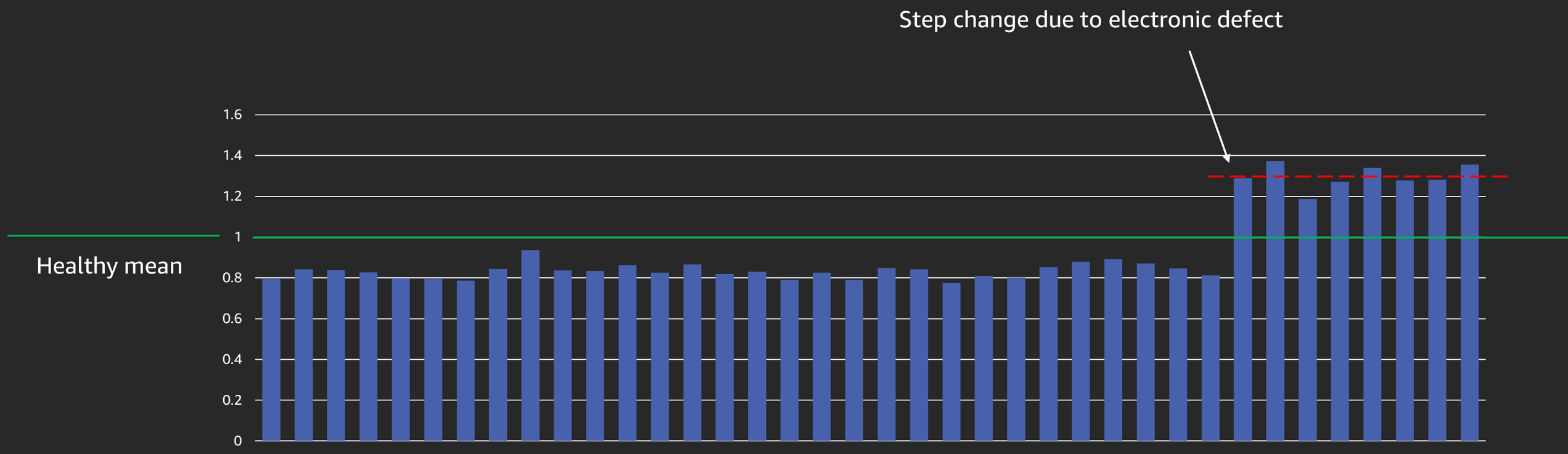
BRITISH AIRWAYS

A year of using Amazon SageMaker

- About 4 flights earlier detection than a “traditional” detection mechanism
- Amazon SageMaker built-in algorithms allow us to focus on our domain expertise – handling flight data
- BUT: We found some “false” positives, which turned out to be genuine events of a different type

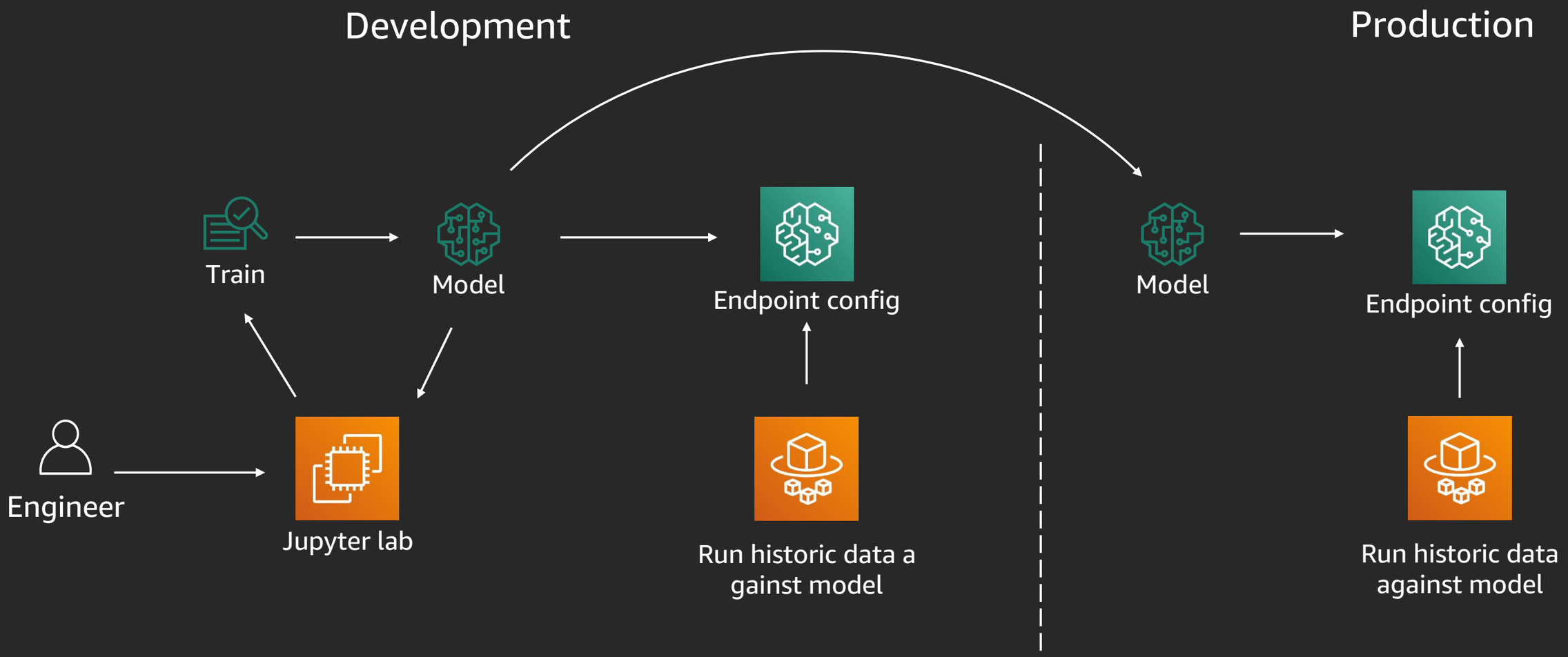
BRITISH AIRWAYS

Expect the unexpected

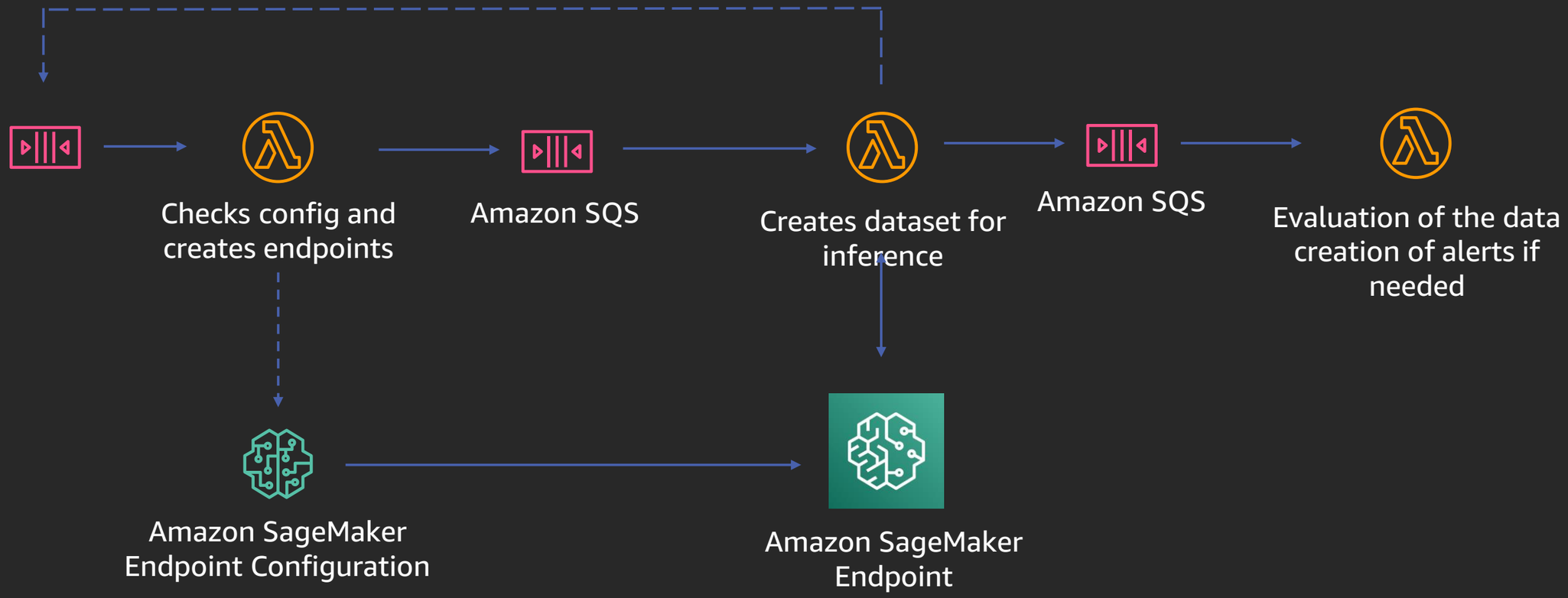


Train, test, and deploy

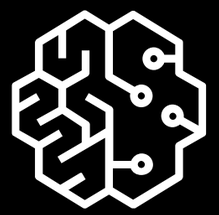
Build, train, and test with Amazon SageMaker



Interfacing with Amazon SageMaker



Build, train, and deploy machine learning models quickly at scale



Amazon SageMaker

NEW! Amazon SageMaker Studio IDE

Ground Truth | ML Marketplace | Algorithms and Frameworks | NEW! SageMaker Notebooks | Training and Tuning | Reinforcement Learning | NEW! SageMaker Experiments | NEW! SageMaker Debugger | NEW! SageMaker Autopilot | Neo | Deployment and Hosting | NEW! SageMaker Model Monitor

Related sessions

Session ID	Title	When	Where
AIM361-R1	Optimizing your ML models with Amazon SageMaker Autopilot	Wed, Dec. 4, 3:15 PM	Venetian, Level 4, Lando 4304
AIM214-R1	Introducing Amazon SageMaker Studio	Wed, Dec. 4, 6:15 PM	Venetian, Level 2, Titian 2202
AIM362-R1	Build, train, tune & debug, and deploy & monitor with Amazon SageMaker	Thu, Dec. 5, 12:15 PM	Aria, Level 1 East, Joshua 2
AIM213-R1	Introducing Amazon SageMaker Model Monitor	Thu, Dec. 5, 1 PM	Venetian, Level 4, Delfino 405
AIM215-R1	Introducing Amazon SageMaker Autopilot	Thu, Dec. 5, 1:45 PM	Mirage, Events Center C3

Thank you!

AWS re:Invent

aws

Please complete the session survey in the mobile app.