



아마존 퀵사이트 와 아마존 세이지메이커 오토 파일럿 연동 Amazon QuickSight with Amazon SageMaker AutoPilot

문곤수

AI / ML 스페셜리스트, 솔루션즈 아키텍트

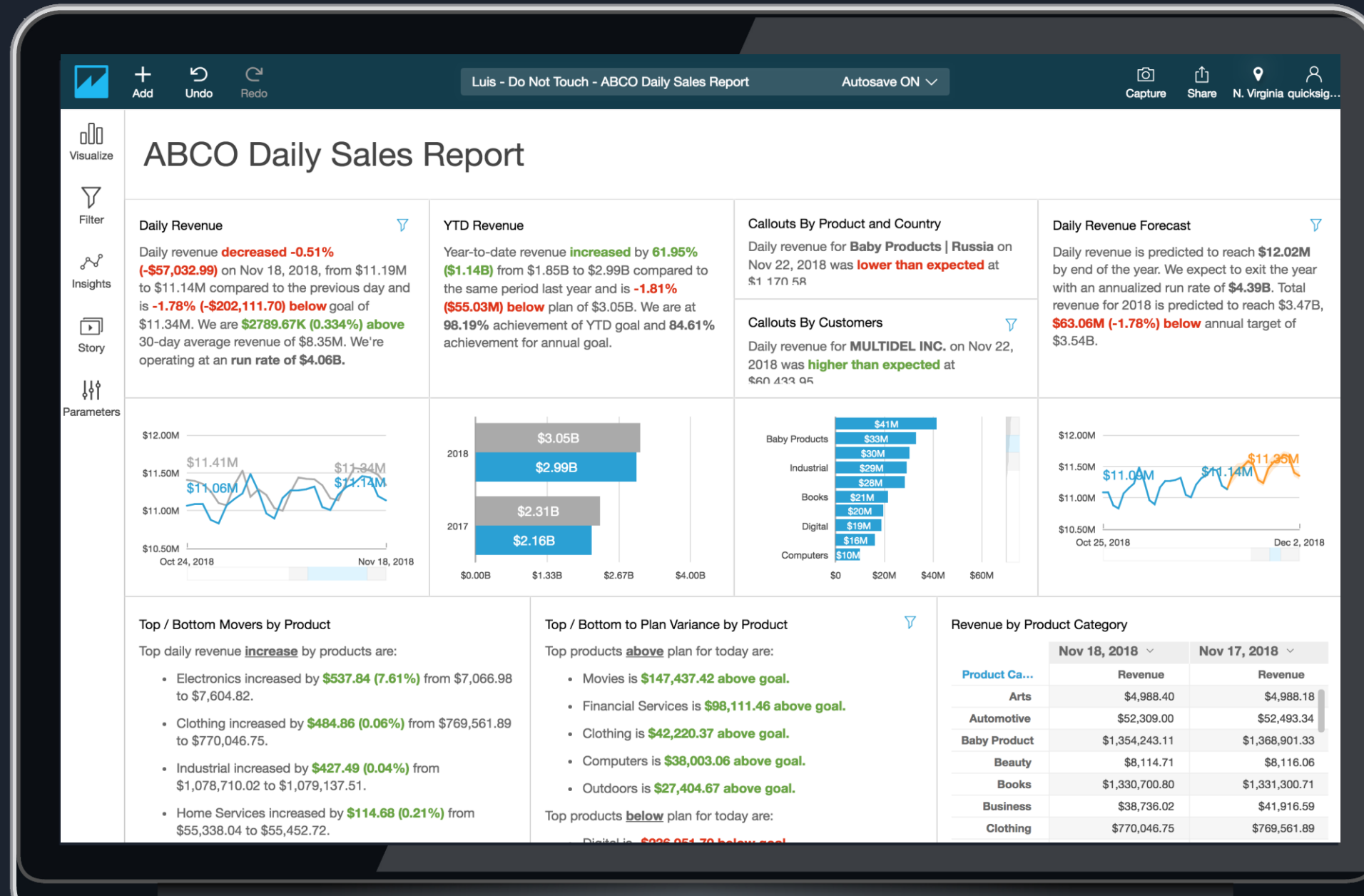
목차

- 소개: Amazon QuickSight
- 소개: Amazon SageMaker AutoPilot
- 데모

Amazon QuickSight

Introduction

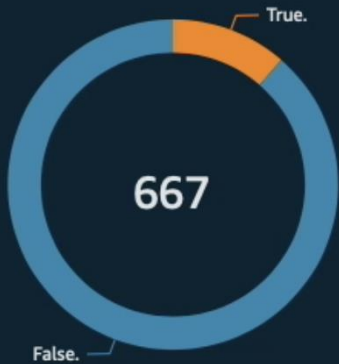
Amazon QuickSight



Amazon QuickSight – 고객 이탈 예측 대시보드

Customer Churn Analysis

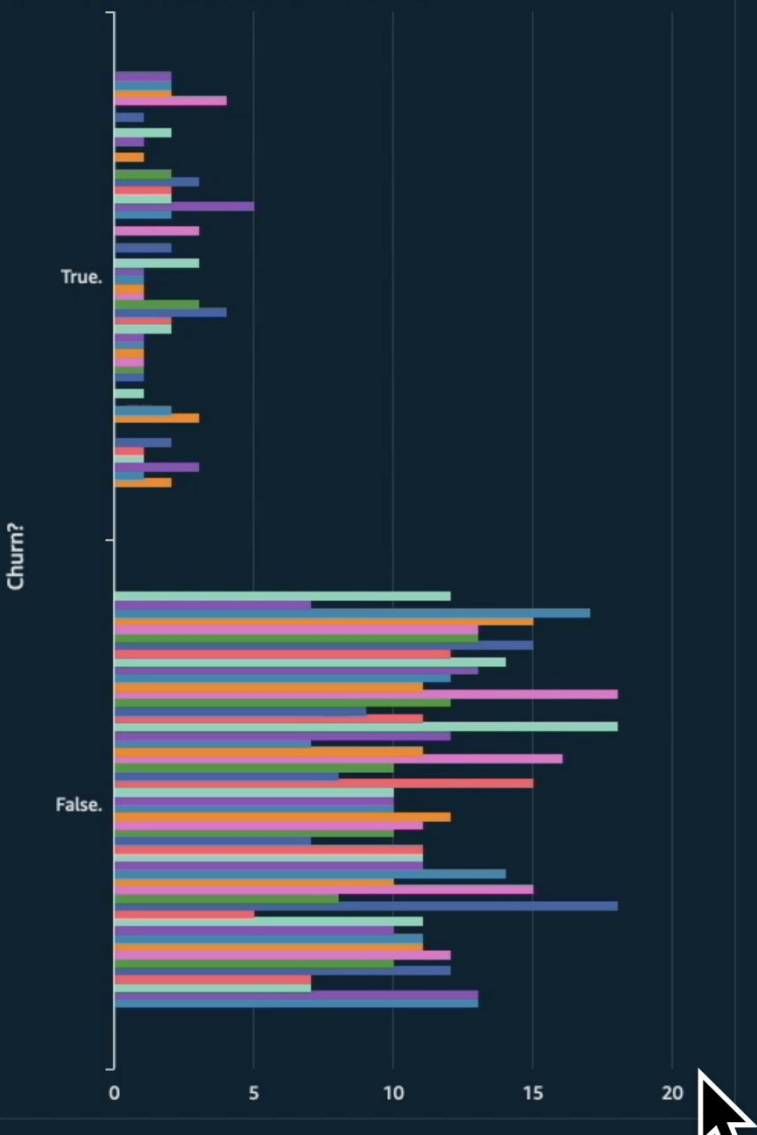
Customers likely to churn



Predicted Loss for Churn Customers

4,990.19

Churn by State
SHOWING TOP 2 IN CHURN? AND TOP 50 IN STATE



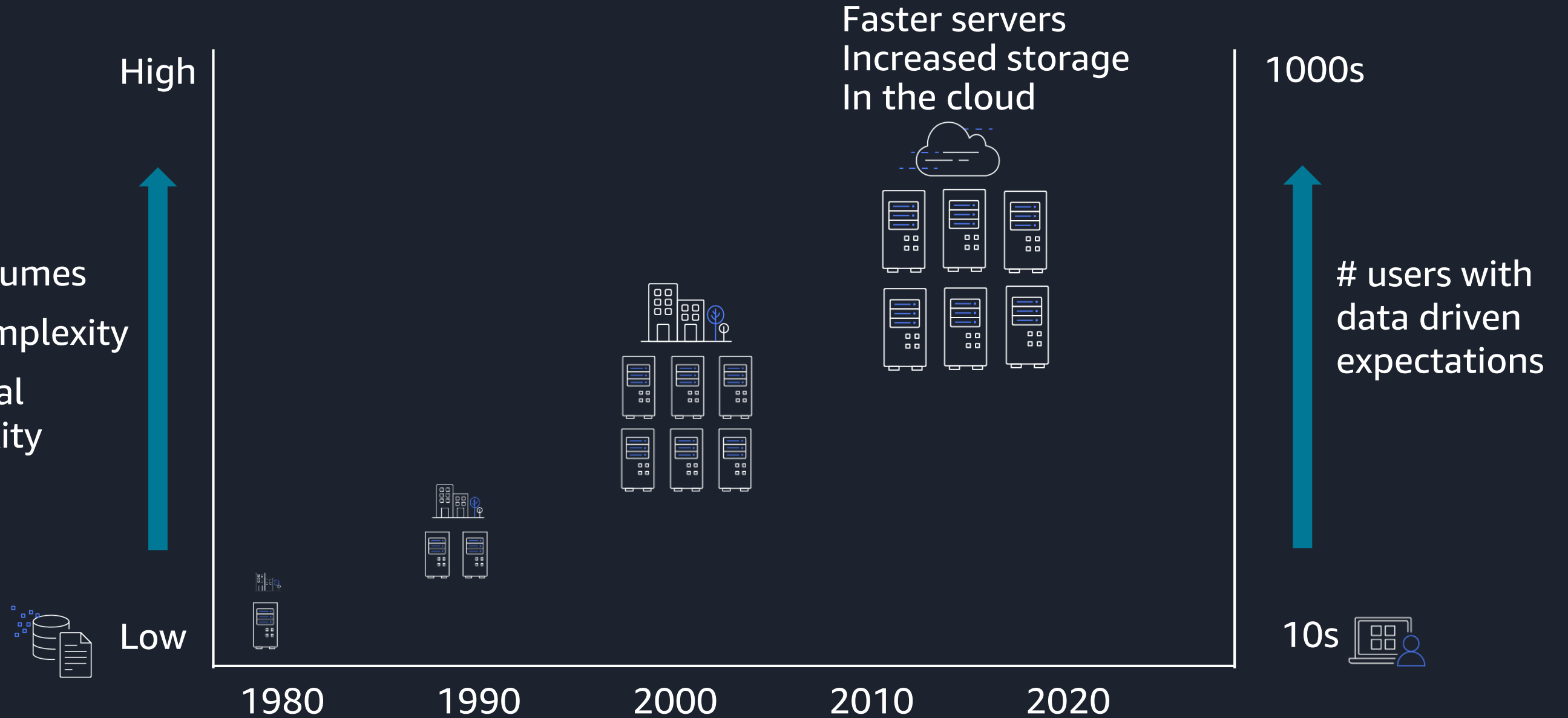
Churn Deailed List

Phone	Churn?	State	Account Length	CustServ Calls	Call Charges
333-5295	True.	AK	126	0	81.17
334-4506	False.	AK	96	1	51.61
341-9764	False.	AK	36	0	48.42
345-9153	False.	AK	132	2	58.25
364-7719	False.	AK	111	4	54.22
366-4467	False.	AK	104	1	64.21
382-5743	False.	AK	121	0	67.18
384-6132	True.	AK	177	1	68.62
394-4548	False.	AK	110	0	52.83
396-2335	False.	AK	110	3	49.05
402-1381	False.	AK	136	3	58.03
404-1931	False.	AK	76	1	51.47
336-5920	False.	AL	122	0	69.83
341-7296	False.	AL	72	0	65.91



데이터 양은 증가, 하지만 BI 기술은?

- Data volumes
- Data complexity
- Analytical complexity



Cloud-Powered Amazon QuickSight -

빠르고, 일관된 성능에 기반한 오토 스케일링

	Server-based BI 아키텍처	Serverless QuickSight 아키텍처
추가 동시 사용자	추가 서버 필요	유저수에 기반하여 자동적으로 확장
더욱 깊이 있는 분석	추가 서버 필요	성능에 기반한 자동적 확장
지리적 분산	추가 서버 + 관리	필요한 만큼 배치, 인프라 관리 없음
고 가용성	추가 서버 + 관리	내재적인 고가용성, 인프라 관리 없음



Amazon QuickSight - Cloud 기반 BI (Business Intelligence) 와 ML Insights

Cloud native = No servers = Auto-Scale
서버, 소프트웨어 관리 필요 없이 확장

Fully Integrated w/AWS
AWS 서비스와 결합 (보안, 머신러닝)

Secure and global
End-to-End 보안, 글로벌 확장

Easy to develop and maintain
UI Design, API 결합

Fast, consistent performance
언제나 빠르고 일관된 성능(속도)

ML-insights
이상감지, 예측, 데이터 기반 알람

Insights for everyone
사용에 따른 과금

Customize and embed
애플리케이션에 삽입

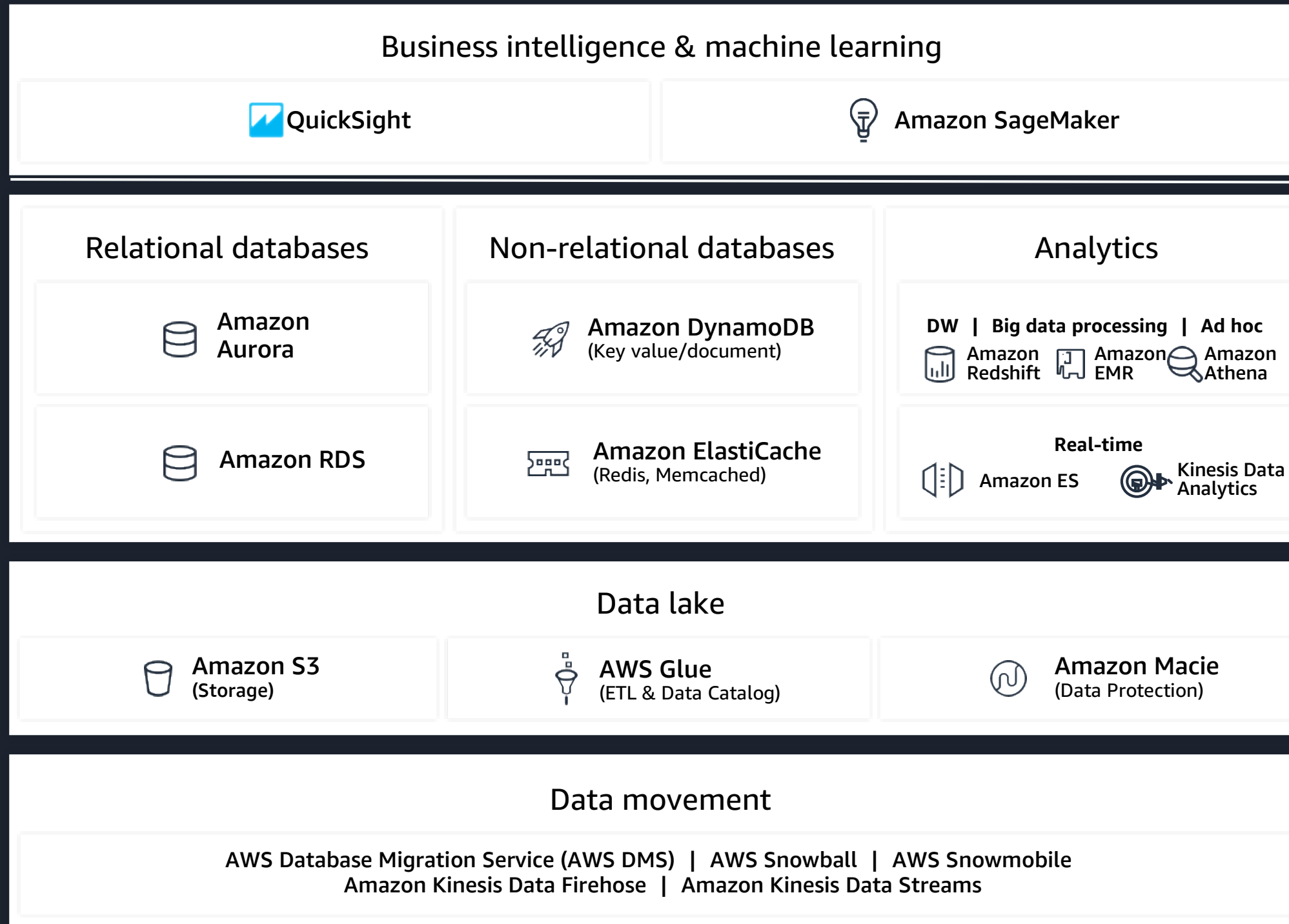
Administrator



End-user

AWS database and analytics stack

가장 넓고 깊은
포트폴리오



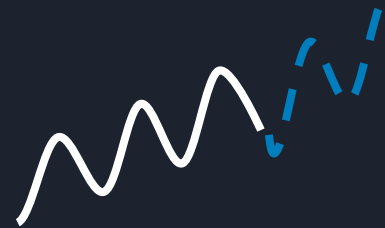
What's New?

Machine learning in Amazon QuickSight



Anomaly detection

이상 트렌드 및
이상 지표 데이터 발견



Forecasting

간단한 조작으로
머신러닝 예측



Auto narratives

자동 요약문 작성



ML predictions

아마존 세이지메이커
모델을 통한
예측 대시보드 작성

일반적인 비즈니스 과제

Predicting price

Employee attrition prediction

Scoring sales leads

Credit scoring

Text analytics

고객 이탈 분석

Detecting fraudulent patterns

Demand forecasting

Assessing loan default risk

Today: How does the BI analyst incorporate ML



Step 1: 데이터 과학자 협업

머신러닝 모델 생성

Step 2: 데이터 및 머신 러닝
엔지니어 협업

데이터 파이프 라인, 머신 러닝
서비스 인프라 생성

Step 3: 리포트 생성

예측 데이터 사용하여
리포트 생성

Takes weeks to months and multiple teams to complete

Amazon SageMaker AutoPilot

소개



**성공적인 ML은
복잡하고 발견하기 어려운 알고리즘,
데이터,
파라미터 조합이 필요합니다.**

방대한 탐색 & 반복

+

ML 도메인에 대한
광범위하고 완전한
지식 필요

+

가시성 부족

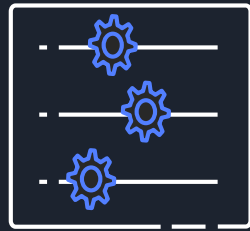
=

ML 전문가도 시간이
많이 걸리고
에러가 발생하기 쉬운
프로세스입니다.

Automated machine learning with Amazon SageMaker Autopilot

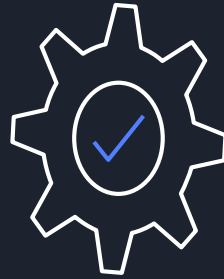


예측 대상(타겟)
기술

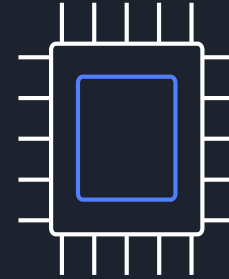


알고리즘 지원
: 회귀 및 분류

(Regression &
classification)



자동화된
피쳐 생성
Automated
feature
engineering



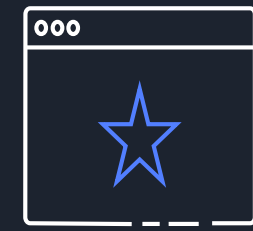
자동화된
알고리즘 선택 및
파라미터 튜닝

Automated
algorithm
selection & HPO



가시성 및 액션
제어

Commented
notebook
describing actions



아마존
세이지메이커
스튜디오와
통합

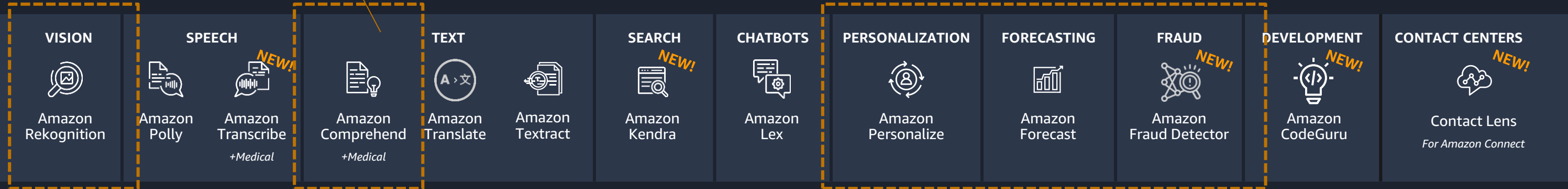
이미지인식, 텍스트분석

The AWS ML Stack

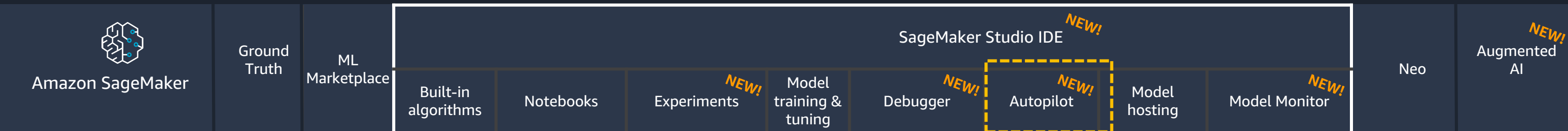
시계열 예측, 추천,
사기탐지 문제

가장 광범위하고 가장 완벽한 머신 러닝 기능 제공

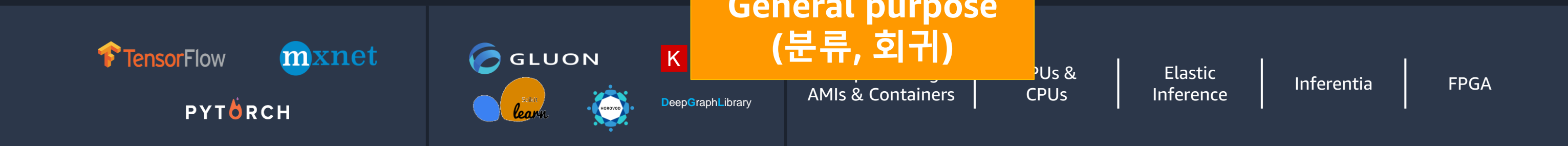
AI SERVICES



ML SERVICES

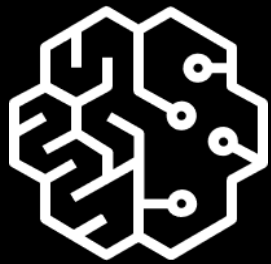


ML FRAMEWORKS & INFRASTRUCTURE

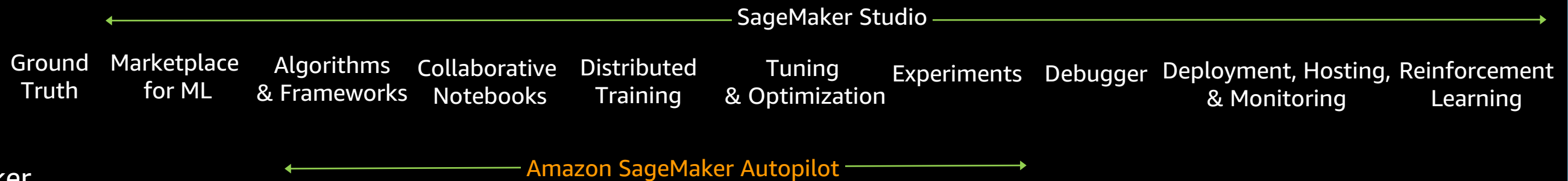


Our goal is to help you scale

빠르고 확장 가능한 머신 러닝 모델 생성, 학습, 배포
Build, Train, Deploy Machine Learning Models Quickly at Scale



Amazon SageMaker



Simply to Get Started

Is your S3 input a manifest file?

For more information on the format of a manifest file, please see the [AWS Docs](#)

Target attribute name

The target attribute is the attribute in your dataset that you want Amazon SageMaker Autopilot to make predictions for.

frst_mv_md1_trgt_yn

The attribute name is case-sensitive and must match exactly the name in your input dataset

Output data location (S3 bucket)

Enter the location in S3 where you want to store the output.

☐ Find S3 bucket ☒ Enter S3 bucket location

Note: The S3 bucket must be in the same AWS Region where you're running SageMaker Studio because SageMaker doesn't allow cross-region requests.

S3 bucket address

s3://dongseok-sagemaker-studio/out/

Select the machine learning problem type

- ☐ Auto
- ☒ Binary classification
- ☐ Regression
- ☐ Multiclass classification

Objective metric

F1

문제 유형 선택

Do you want to run a complete experiment?

- ☒ Yes
- ☐ No, run a pilot to create a notebook with candidate definitions

완전 자동 실행 여부

EXPERIMENT: LG-AUTOPILOT-FROM-STUDIO

Analyzing Data

Feature Engineering

Model Tuning

Completed

If experiment is taking too long to run, you can [stop the experiment](#)

You can always return to this page later by choosing this experiment on the Experiments tab in the navigation panel.



Autopilot 작동 원리

1. 후보 생성: Autopilot은 학습 데이터셋을 분석하고 통계를 계산하며 모델을 평가할 인스턴스 유형, 전처리 및 알고리즘의 조합을 결정합니다.

1-1. (Optional) 후보 노트북: Autopilot은 선택 사항을 설명하는 주석이 달린 노트를 제공하여 데이터 과학자가 프로세스에 입력을 제공할 수 있도록 합니다.

2. 학습 및 튜닝: Autopilot은 병렬로 최대 7개의 모델 파이프라인을 튜닝하여 전처리, 알고리즘 및 하이퍼파라미터의 가장 정확한 조합을 결정합니다.

Autopilot에서 자동 생성된 노트북

1.Data Exploration notebook

Input 데이터 통계 분석

Amazon SageMaker Autopilot Data Exploration ¶

This report provides insights about the dataset you provided as input to the AutoML job. It was automatically generated by the AutoML training job: **automl-banking-26-09-07-31**.

As part of the AutoML job, the input dataset was randomly split into two pieces, one for **training** and one for **validation**. The training dataset was randomly sampled, and metrics were computed for each of the columns. This notebook provides these metrics so that you can:

1. Understand how the job analyzed features to select the candidate pipelines.
2. Modify and improve the generated AutoML pipelines using knowledge that you have about the dataset.

We read **32950** rows from the training dataset. The dataset has **21** columns and the column named **y** is used as the target column. This is identified as a **BinaryClassification** problem. Here are **2** examples of labels: `['no', 'yes']`.

💡 Suggested Action Items

- Look for sections like this for recommended actions that you can take.

Contents

1. [Dataset Sample](#)
2. [Column Analysis](#)

Dataset Sample

The following table is a random sample of **10** rows from the training dataset. For ease of presentation, we are only showing **20 of the 21** columns of the dataset.

💡 Suggested Action Items

	% of Numerical Values	Mean	Median	Min	Max
age	100.0%	40.0005	38.0	17.0	98.0
duration	100.0%	257.23	178.0	0.0	4918.0
campaign	100.0%	2.57517	2.0	1.0	56.0
pdays	100.0%	963.225	999.0	0.0	999.0
previous	100.0%	0.171533	0.0	0.0	7.0
emp.var.rate	100.0%	0.0815781	1.1	-3.4	1.4
cons.price.idx	100.0%	93.5736	93.444	92.201	94.767
cons.conf.idx	100.0%	-40.48	-41.8	-50.8	-26.9
euribor3m	100.0%	3.62236	4.857	0.634	5.045
nr.employed	100.0%	5167.15	5191.0	4963.6	5228.1

Autopilot에서 자동 생성된 노트북

2. Candidate Generation Notebook-

Candidate Pipelines 생성 (Feature Transformer + Algorithm)

```
automl_interactive_runner.display_candidates()
```

Candidate Name	Algorithm	Feature Transformer
dpp0-xgboost	xgboost	dpp0.py
dpp2-linear-learner	linear-learner	dpp2.py
dpp3-xgboost	xgboost	dpp3.py
dpp4-xgboost	xgboost	dpp4.py
dpp5-xgboost	xgboost	dpp5.py
dpp6-xgboost	xgboost	dpp6.py
dpp7-xgboost	xgboost	dpp7.py
dpp8-xgboost	xgboost	dpp8.py
dpp9-xgboost	xgboost	dpp9.py

Transformer 학습 인스턴스, 피쳐 변환 인스턴스
및 Alogrithm 학습 인스턴스로 구성

dpp0-xgboost: This data transformation strategy first transforms 'numeric' features using [RobustImputer](#) (converts missing values to nan), 'categorical' features using [ThresholdOneHotEncoder](#). It merges all the generated features and applies [RobustStandardScaler](#). The transformed data will be used to tune a *xgboost* model. Here is the definition:

```
automl_interactive_runner.select_candidate({
  "data_transformer": {
    "name": "dpp0",
    "training_resource_config": {
      "instance_type": "ml.m5.4xlarge",
      "instance_count": 1,
      "volume_size_in_gb": 50
    },
    "transform_resource_config": {
      "instance_type": "ml.m5.4xlarge",
      "instance_count": 1,
    },
    "transforms_label": True,
    "transformed_data_format": "text/csv",
    "sparse_encoding": False
  },
  "algorithm": {
    "name": "xgboost",
    "training_resource_config": {
      "instance_type": "ml.m5.4xlarge",
      "instance_count": 1,
    }
  }
})
```



튜닝에 따른 결과 비교 (이탈 예측 정확도)

aws

File Edit View Run Kernel Git Tabs Settings Help

+
/ automl-preview /

Name	Last Modified
bank-additional	12 hours ago
model	12 hours ago
sagemaker_auto...	3 hours ago

EXPERIMENT: MY-SAGEMAKER-AUTOPILOT

Trials Job profile

TRIALS

0 row selected

Trial name	Status	Start time	End time	Objective
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9206119775772095
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9202479720115662
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9200050234794617
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9195190072059631
my-sagemaker-tuning-job-...	Completed	9 hours ago		0.9191550016403198
my-sagemaker-tuning-job-...	Completed	7 hours ago		0.9190340042114258
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9189119935035706
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9186699986457825
my-sagemaker-tuning-job-...	Completed	8 hours ago		0.9186699986457825

0 \$ 1

my-sagemaker-autopilot

Best 모델 확인 및 선택 배포

TRIAL COMPONENTS

1 row selected

Add chart

Deploy model

Status	Experiment	Type	Trial	Trial component	Created on	Last modified	ObjectiveMe
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.90823
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.910658
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.899005
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.91612
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.913693
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.919641
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	130129
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	13013
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	07745
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	9949
✓ Completed	automl-20200124-ful...	Training job	tuning-job-1-42f63c9...	tuning-job-1-42f63c9...	22 hours ago	22 hours ago	0.907138

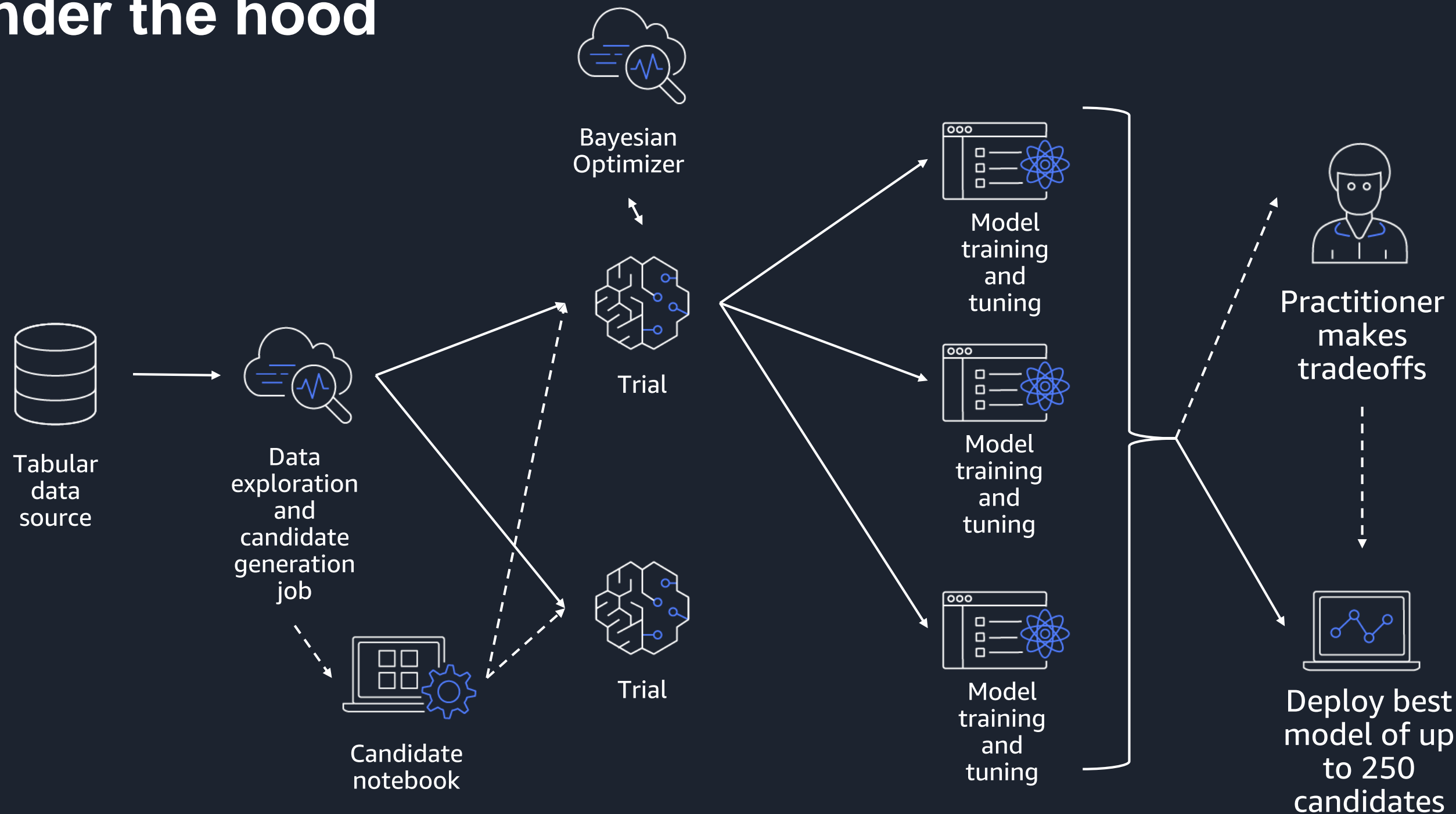
Deploy model

Open in trial details

Copy cell contents

Shift+Right Click for Browser Menu

Under the hood



1

7 in parallel

250

데모

고객 이탈 예측 대시보드 개발 (Amazon QuickSight + SageMaker AutoPilot)

QuickSight with SageMaker 통합 자료:

Amazon QuickSight와 Amazon SageMaker 모델 통합

- https://docs.aws.amazon.com/ko_kr/quicksight/latest/user/sagemaker-integration.html?icmpid=icmpid%3Ddocs_quicksight_console_whatsnew

블로그: Visualizing Amazon SageMaker machine learning predictions with Amazon QuickSight

- <https://aws.amazon.com/blogs/machine-learning/making-machine-learning-predictions-in-amazon-quicksight-and-amazon-sagemaker/>

AWS 교육 및 자격증



조직을 위한 맞춤 교육

고객 및 파트너를 위해
준비된 데이터 및
데이터베이스 관련
맞춤형 교육 여정을
확인해 보세요.



원하는 방법으로 - 유연한 학습 형태

"The elements of Data
Science" 과정을 포함한 무료
디지털 교육 또는 강의실
교육을 통해 클라우드 역량을
향상시키세요.



AWS 자격증을 통한 기술 역량 입증

업계에서 인정받는
데이터 분석 또는
데이터베이스 - 전문분야
AWS 자격증을 통해 전문성을
입증할 수 있습니다.

aws.amazon.com/training

AWS 데이터분석 관련 자료를 원하시면 ...

데이터 분석 관련 기술 백서 및 전자책을 자세히 살펴보면
데이터에서 새로운 통찰력과 가치를 발견 할 수 있습니다!

- 클라우드 기반 데이터 분석 서비스
- 데이터의 분석 활용 사례
- 최신 분석 아키텍처 생성 방법
- 데이터 중심 기업 전환
- 동영상, 기술 백서 등

지금 방문하세요! »



<https://tinyurl.com/data-databases-analytics-kr>

AWS 데이터 분석 특집 웨비나에 참석해주셔서 대단히 감사합니다.

저희가 준비한 내용, 어떻게 보셨나요?
더 나은 세미나를 위하여 **설문을 꼭 작성해 주시기 바랍니다.**



aws-korea-marketing@amazon.com



twitter.com/AWSKorea



facebook.com/amazonwebservices.ko



youtube.com/user/AWSKorea



slideshare.net/awskorea



twitch.tv/aws

Thank you!

소중한 시간 감사 드립니다.