

Getting Started with AI/ML: With No Code or Low Code

Michael Lin
Solutions Architect, AWS



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

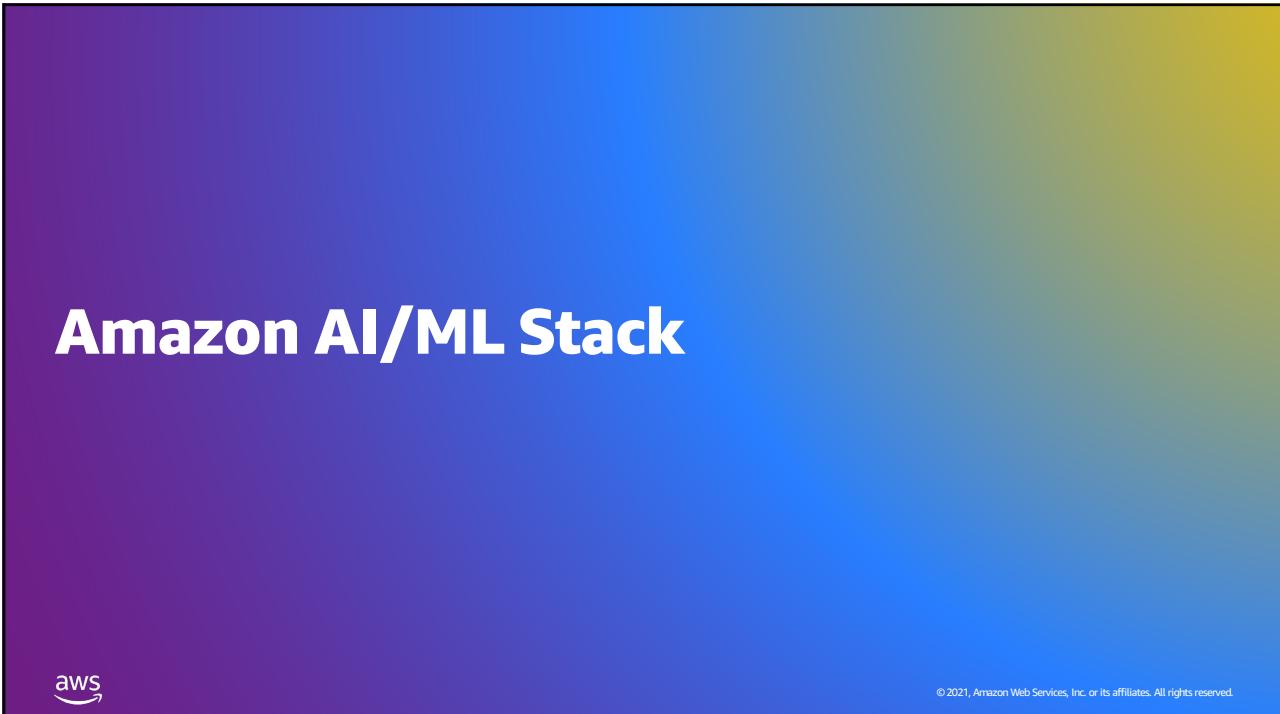
1

Agenda

- Amazon AI/ML Stack
- Amazon SageMaker Canvas
- Amazon SageMaker Data Wrangler
- Q & A

2

1



3

A slide titled "AWS AI and machine learning services" with the subtitle "BROADEST AND DEEPEST SET OF CAPABILITIES". It features three main sections: "AI services", "ML services", and "ML frameworks + infrastructure".

- AI services:** A grid of icons and names for various AI services: VISION (Amazon Rekognition Image, Amazon Rekognition Video, Amazon Textract), SPEECH (Amazon Polly, Amazon Transcribe), LANGUAGE (Amazon Translate, Amazon Comprehend & Comprehend Medical), CHATBOTS (Amazon Lex), FORECASTING (Amazon Forecast), and RECOMMENDATIONS (Amazon Personalize).
- ML services:** A horizontal row of icons and names for Amazon SageMaker and its associated services: Ground Truth, Notebooks, Algorithms + Marketplace, Reinforcement learning, Training, Optimization, Deployment, and Hosting.
- ML frameworks + infrastructure:** A grid of icons and names for ML frameworks (TensorFlow, MXNet, PyTorch, Keras) and infrastructure services (EC2 P3 & P3dn, EC2 G4, EC2 G5, FPGAs, DL containers & AMIs, Amazon ECS, Amazon EKS, AWS IoT Greengrass, Amazon Elastic Inference, AWS Inferentia).

4

2

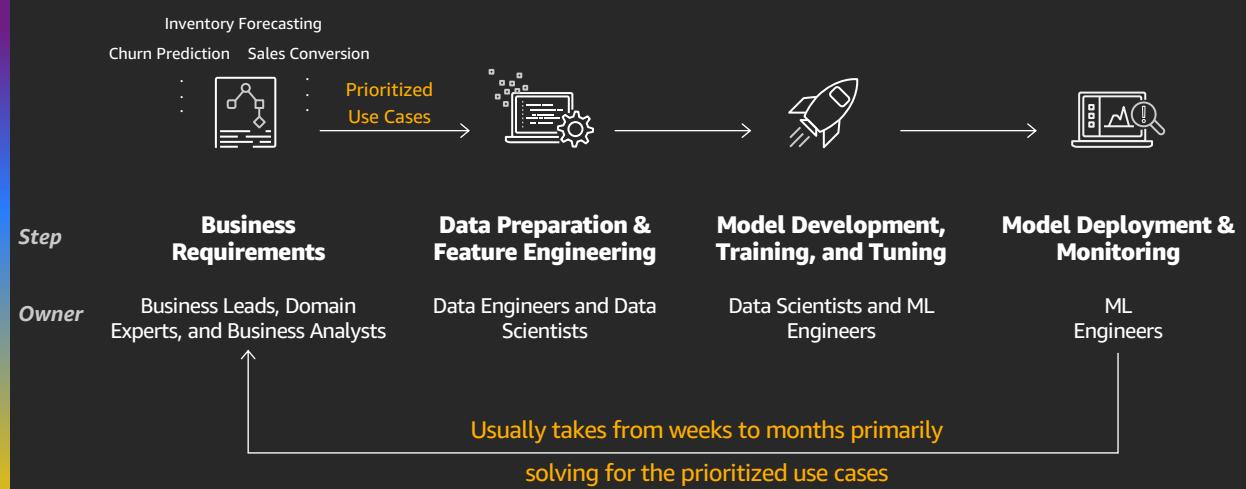
Amazon SageMaker Canvas



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

5

How ML drives value creation today



6

3

Challenges analysts face in building ML



Analysts lack deep ML expertise, and learning curve is steep



Business needs explainability and validation from experts



Available no-code ML tools tend to lack transparency and have upfront fees

7

Amazon SageMaker Canvas

Build ML models and generate accurate predictions—no code required



Quickly access and prepare data for Machine Learning



Built-in AutoML to build models and generate accurate predictions



Share ML models and collaborate with data science teams

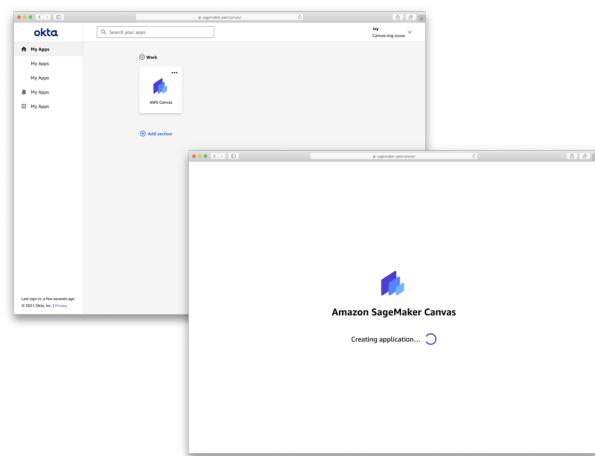


Usage-based pricing to avoid licensing fees and reduce TCO

8

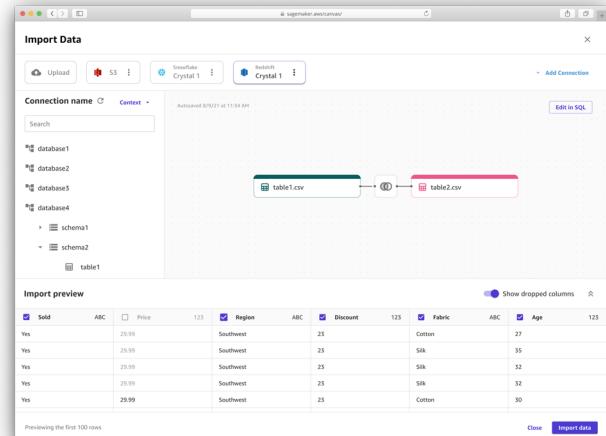
4

Self-service access to a business-friendly tool for Machine Learning, outside of the AWS console



9

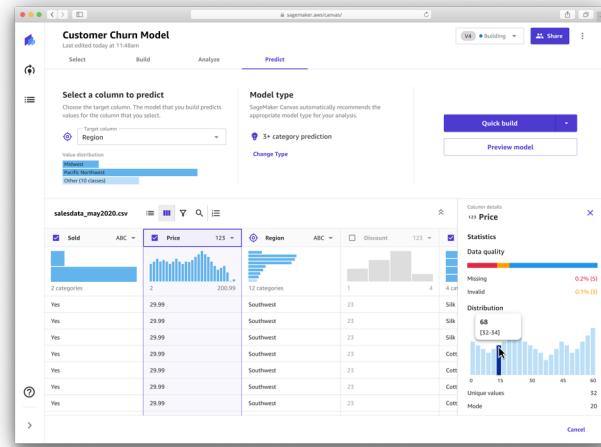
Combine datasets from various sources like local disk, Amazon S3, Amazon RedShift, and Snowflake



10

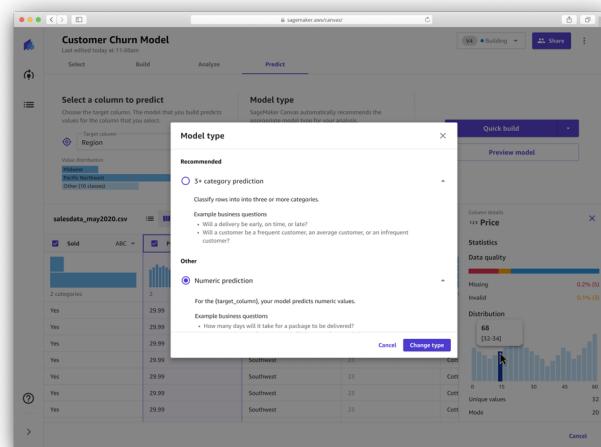
5

Quickly understand and prepare your data via a visual interface



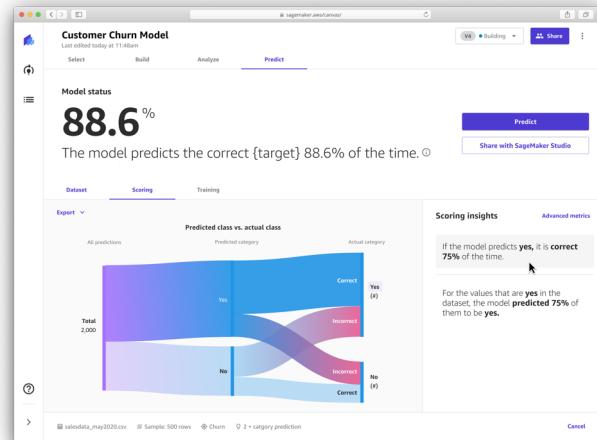
11

Automatically build an accurate ML model for your dataset



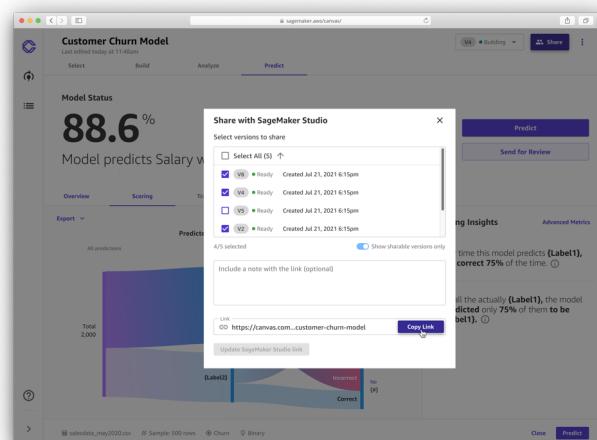
12

Get the first ML model in minutes. Review advanced metrics and feature importance to understand and explain predictions.



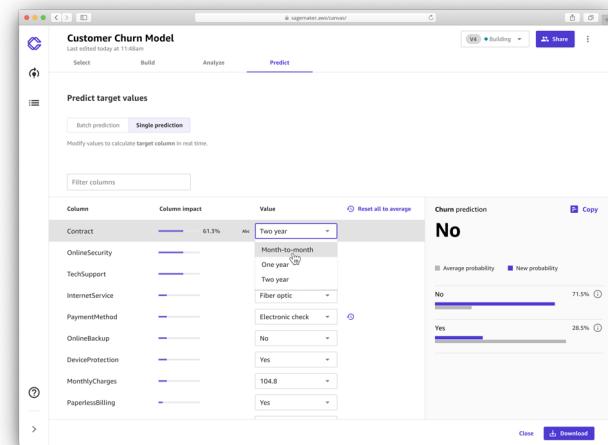
13

Easily share your models with data scientists to get feedback



14

Run what-if scenarios, or get predictions on an entire dataset



15

Lab 1 SageMaker Canvas



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

16

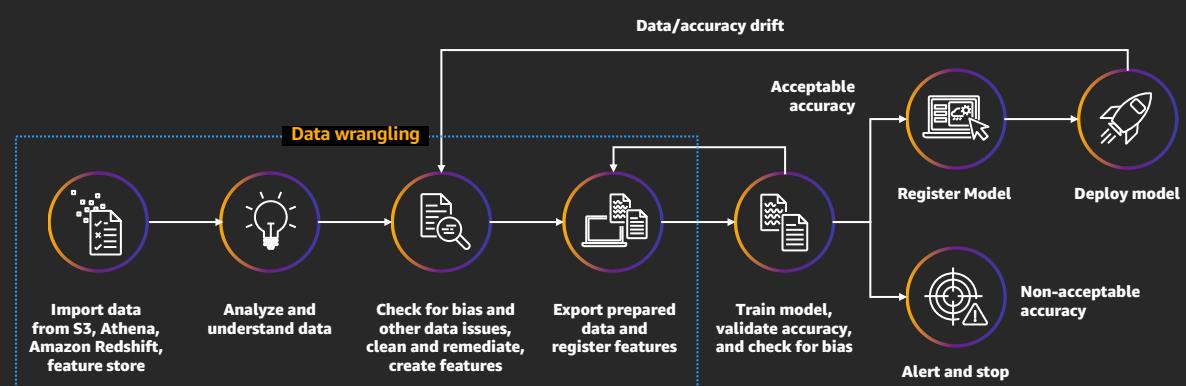
Amazon SageMaker Data Wrangler



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

17

Data preparation is a critical part of the end-to-end ML workflow



18

Data preparation challenges

- ⌚ Data preparation is time consuming and requires multiple tools and tasks
- ⌚ Simple tasks require a lot of code
- ⌚ Deployment can require a code rewrite, and productionizing can take months

19

Introducing Amazon SageMaker Data Wrangler



20

10

Quickly select and query data

SELECT

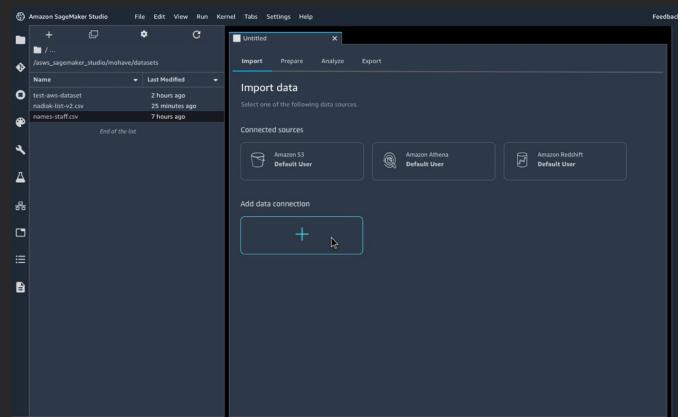
Quickly select data from Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon S3, and features from SageMaker Feature Store

WRITE

Write queries for data sources before importing data over to SageMaker Data Wrangler

IMPORT

Easily import data in various file formats, such as CSV files, parquet files, as well as database tables, directly into Amazon SageMaker



21

Easily transform data

TRANSFORM

Transform your data without writing a single line of code using 300+ built-in data transforms

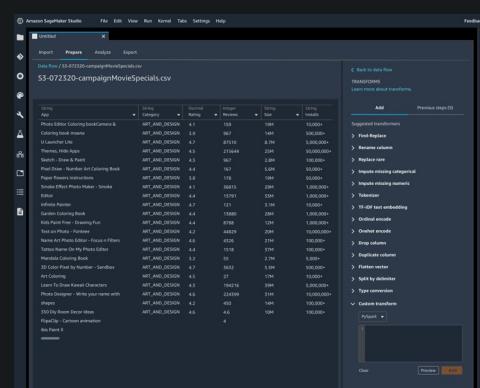
PREPARE

Preconfigured data transforms include:

- Missing value detection and imputation
- Outlier detection and handling
- Featurizers for string and date-time columns
- Column manipulation
- String cleaning and processing tools
- Categorical encoding

AUTHOR

Author custom transforms in PySpark, SQL, and Pandas



22

11

Understand your data visually

UNDERSTAND

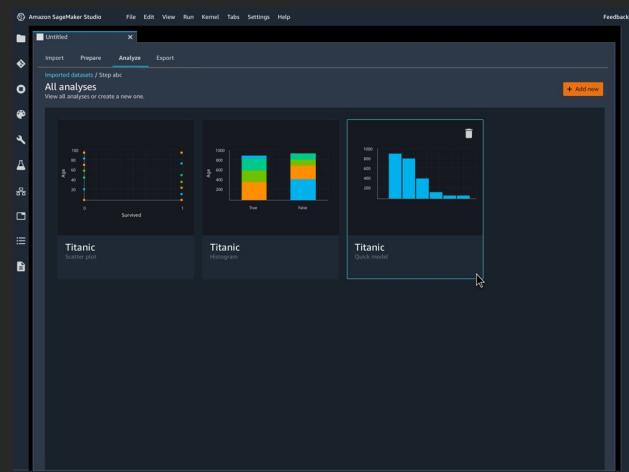
Intuitively understand your data with a set of preconfigured visualization templates which include histograms and scatter plots

CREATE

Create your own custom templates using Altair for data visualization

SAVE

Save and organize your visualizations easily for future reference and reuse



23

Quickly estimate ML model accuracy

IDENTIFY

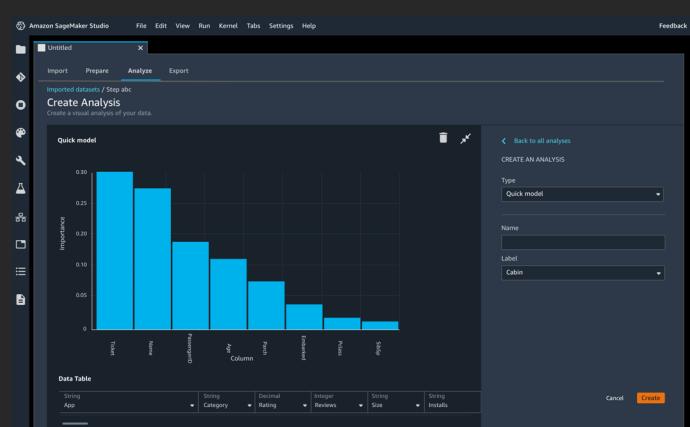
Quickly identify inconsistencies in your data preparation workflow and diagnose issues before ML models are deployed into production

DETECT

Detect which features are contributing to model performance relative to others

DETERMINE

Determine if additional feature engineering is needed to improve model performance



24

12

Deploy data preparation workflows into production with a single click

EXPORT

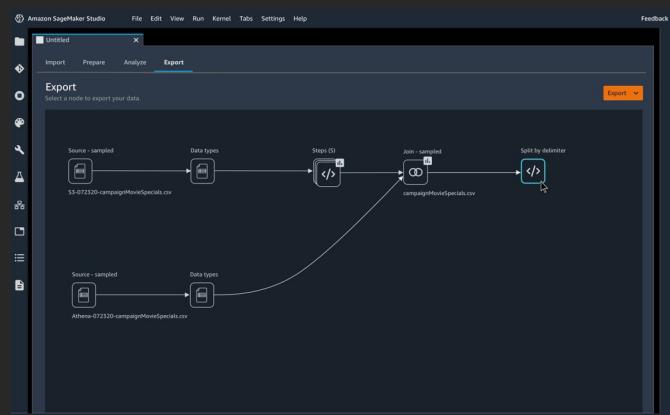
Export data preparation workflow as a processing job notebook or Python code with a single click

INTEGRATE

Integrate your workflow with SageMaker Pipelines to automate model deployment and management

PUBLISH

Publish created features to SageMaker Feature Store for reuse and syndication across teams and projects



25

Lab 2 SageMaker Data Wrangler



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

26

13

Q & A



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

27

Thank you!

Michael Lin

linmicht@amazon.com



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

28

14