Getting Started with AI/ML: Low Code and No Code Style

Michael Lin

Senior Solutions Architect
Amazon Web Services

Agenda

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

Amazon AI/ML Stack

AWS AI and machine learning services

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 & 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								
EC2 P3 & P3dn	EC2 G4 EC2 C5	FPGAs	DL containers & AMIs	Amazon ECS	Amazon EKS	AWS IoT Greengrass	Amazon Elastic Inference	AWS Inferentia

Amazon SageMaker Canvas

How ML drives value creation today

Inventory Forecasting

Churn Prediction Sales Conversion

Step	**Business Requirements**	**Data Preparation & Feature Engineering**	**Model Development, Training, and Tuning**	**Model Deployment & Monitoring**
Owner	Business Leads, Domain Experts, and Business Analysts	Data Engineers and Data Scientists	Data Scientists and ML Engineers	ML Engineers

Usually takes from weeks to months primarily solving for the prioritized use cases

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

Amazon SageMaker Canvas

Build ML models and generate accurate predictions—no code required

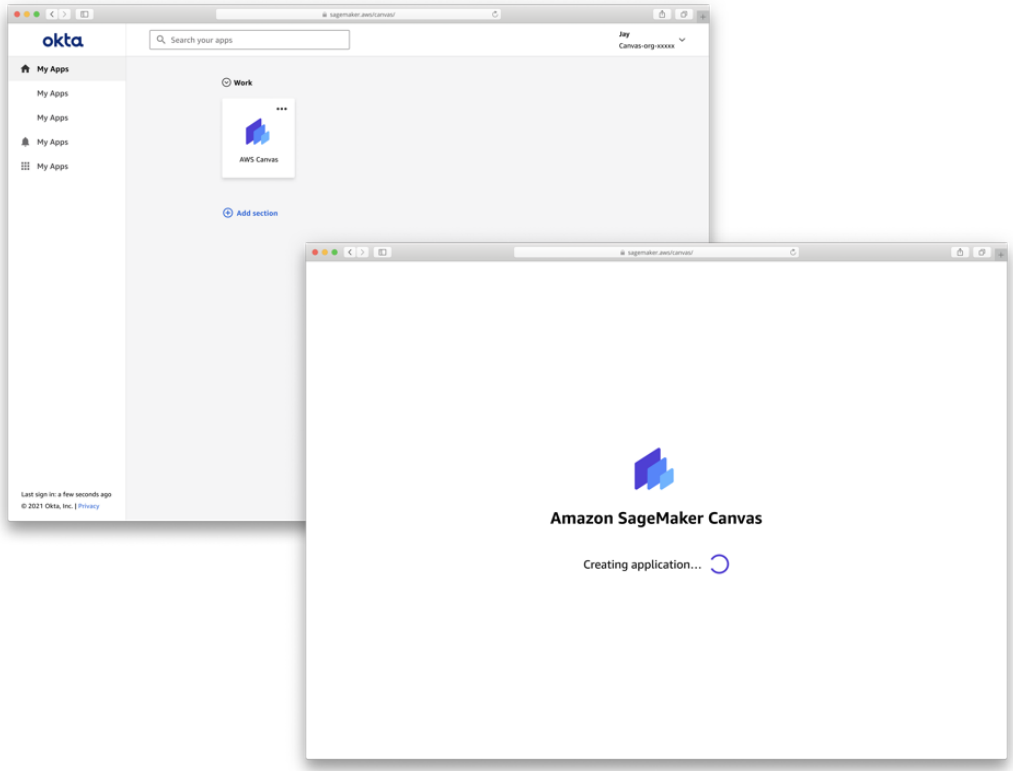
Quickly access and prepare data for Machine Learning

Built-in AutoML to build models and generate accurate predictions

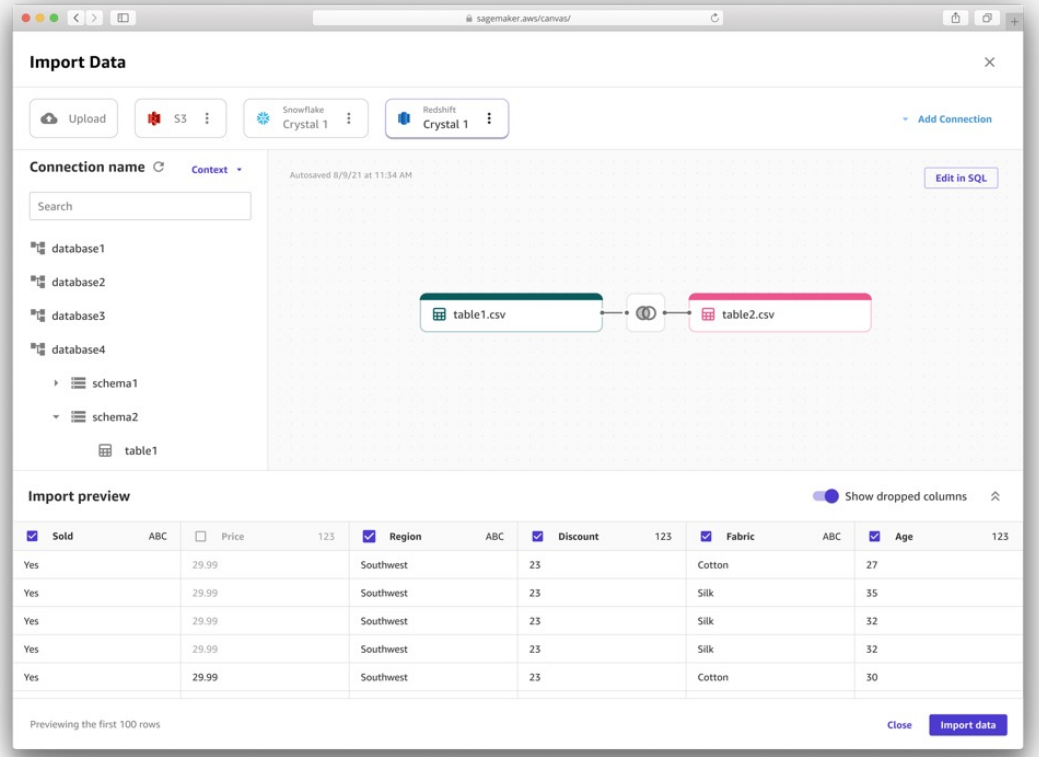
Share ML models and collaborate with data science teams

Usage-based pricing to avoid licensing fees and reduce TCO

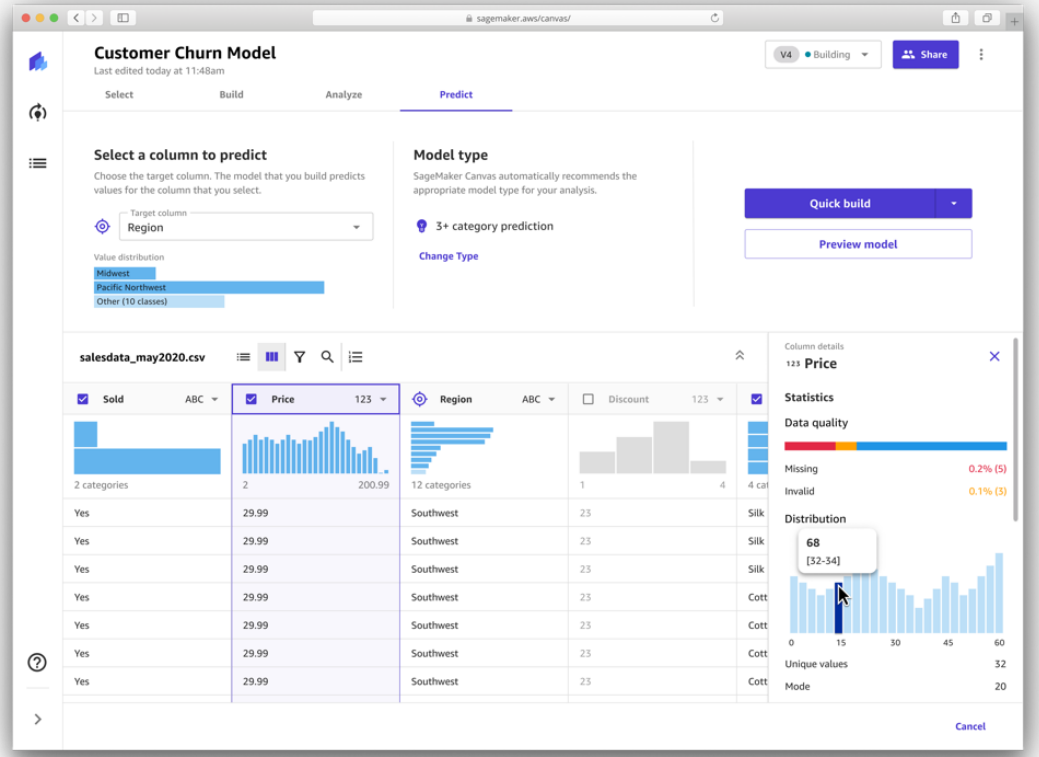
Self-service access to a business-friendly tool for Machine Learning, outside of the AWS console



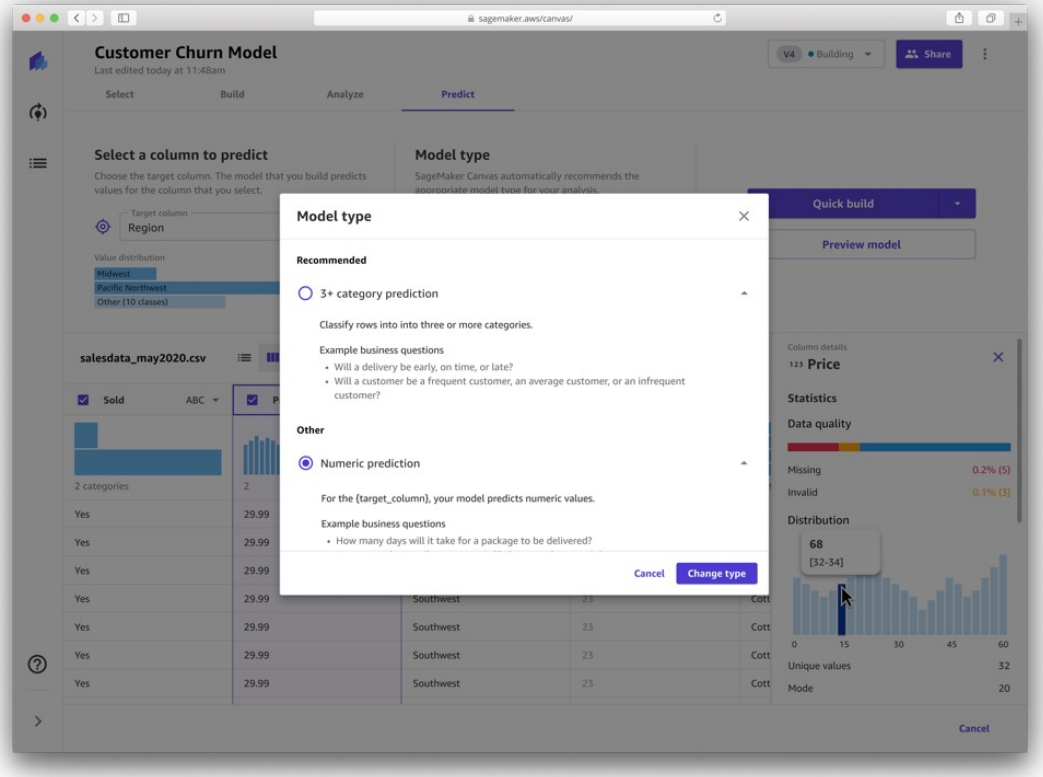
Combine datasets from various sources like local disk, Amazon S3, Amazon RedShift, and Snowflake



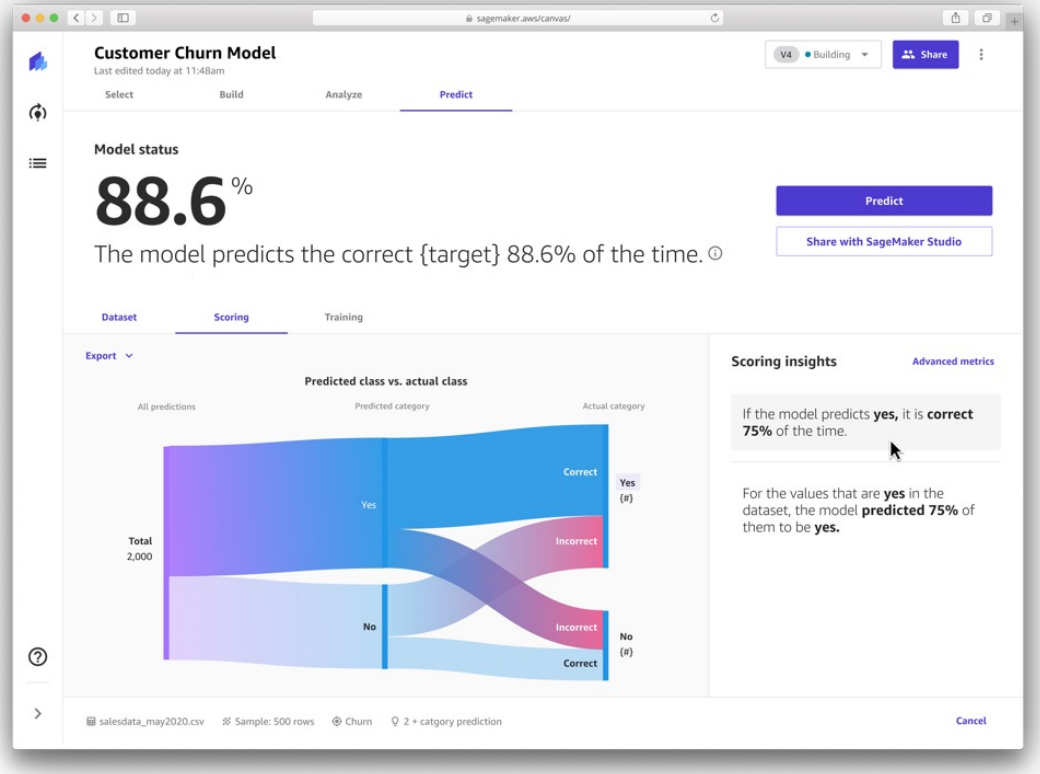
Quickly understand and prepare your data via a visual interface



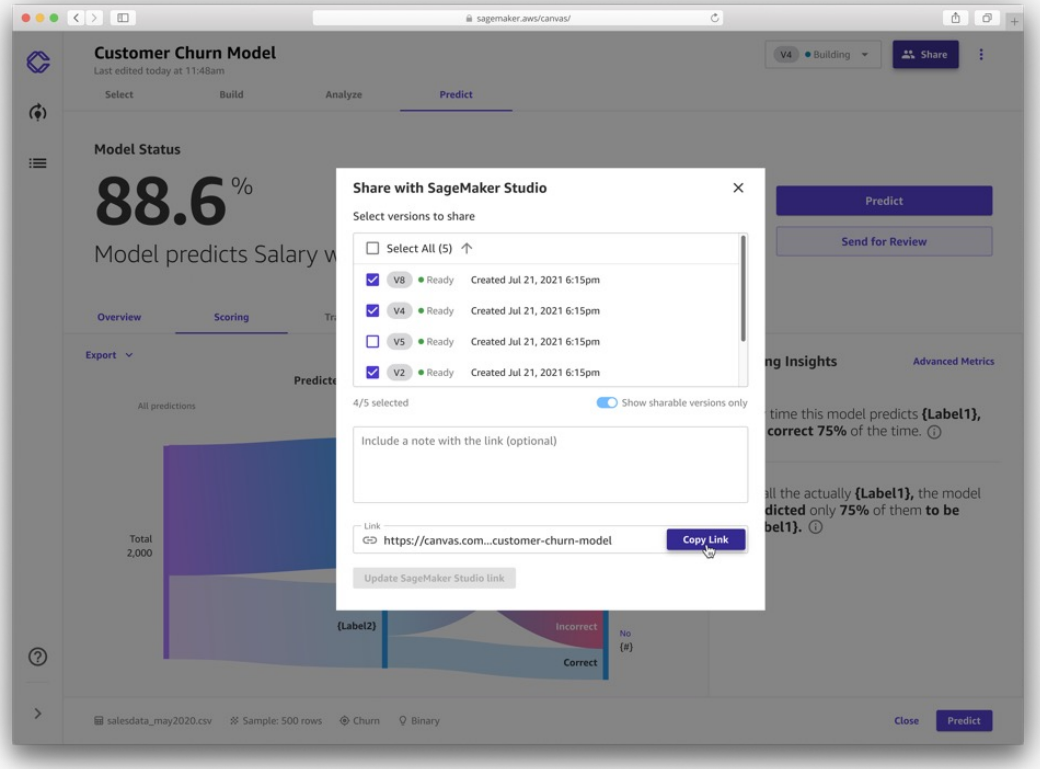
Automatically build an accurate ML model for your dataset



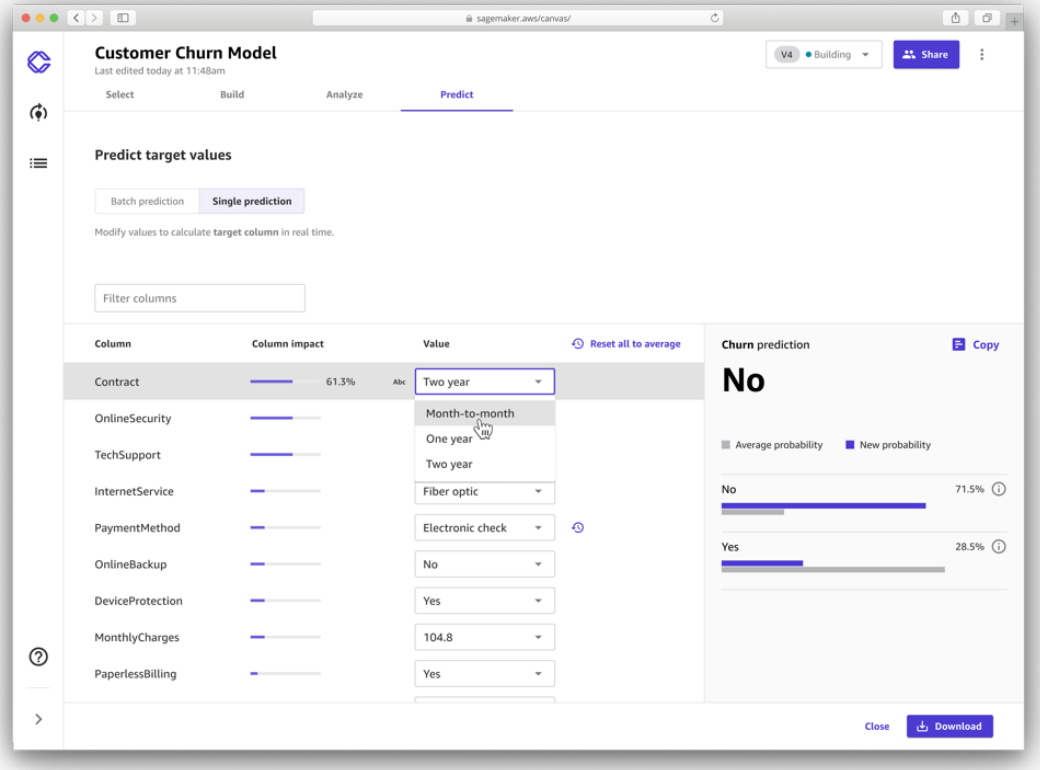
Get the first ML model in minutes. Review advanced metrics and feature importance to understand and explain predictions.



Easily share your models with data scientists to get feedback



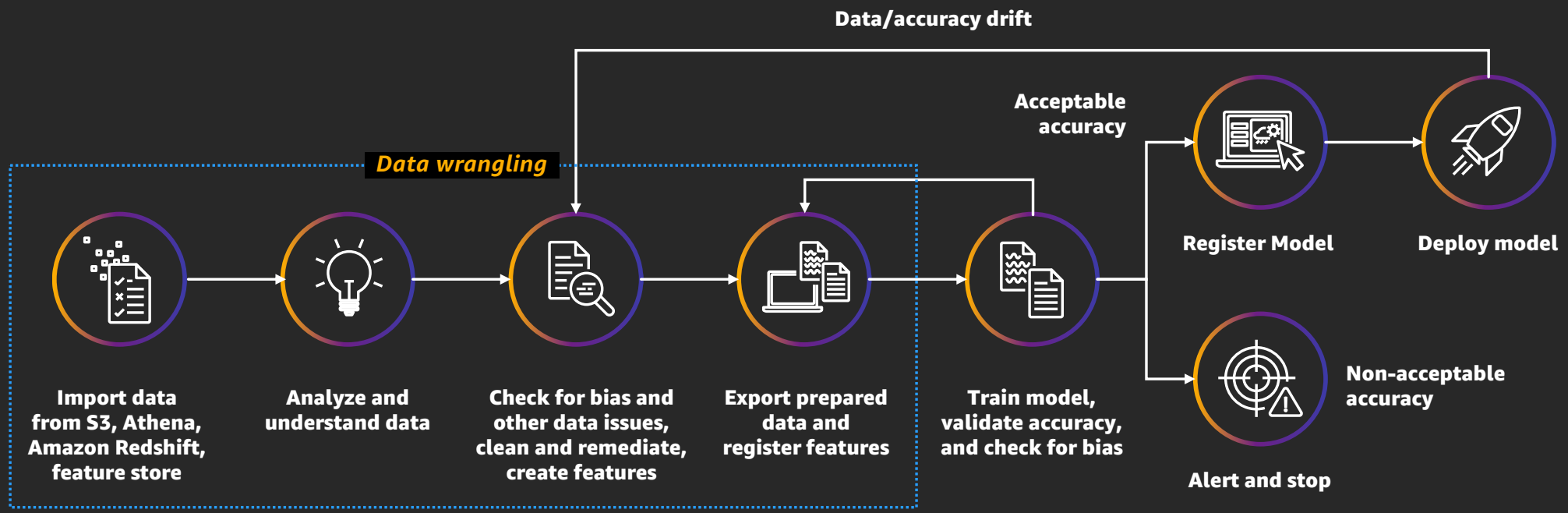
Run what-if scenarios, or get predictions on an entire dataset



Lab 1 SageMaker Canvas

Amazon SageMaker Data Wrangler

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



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*

Introducing Amazon SageMaker Data Wrangler

Quickly select and query data

SELECT

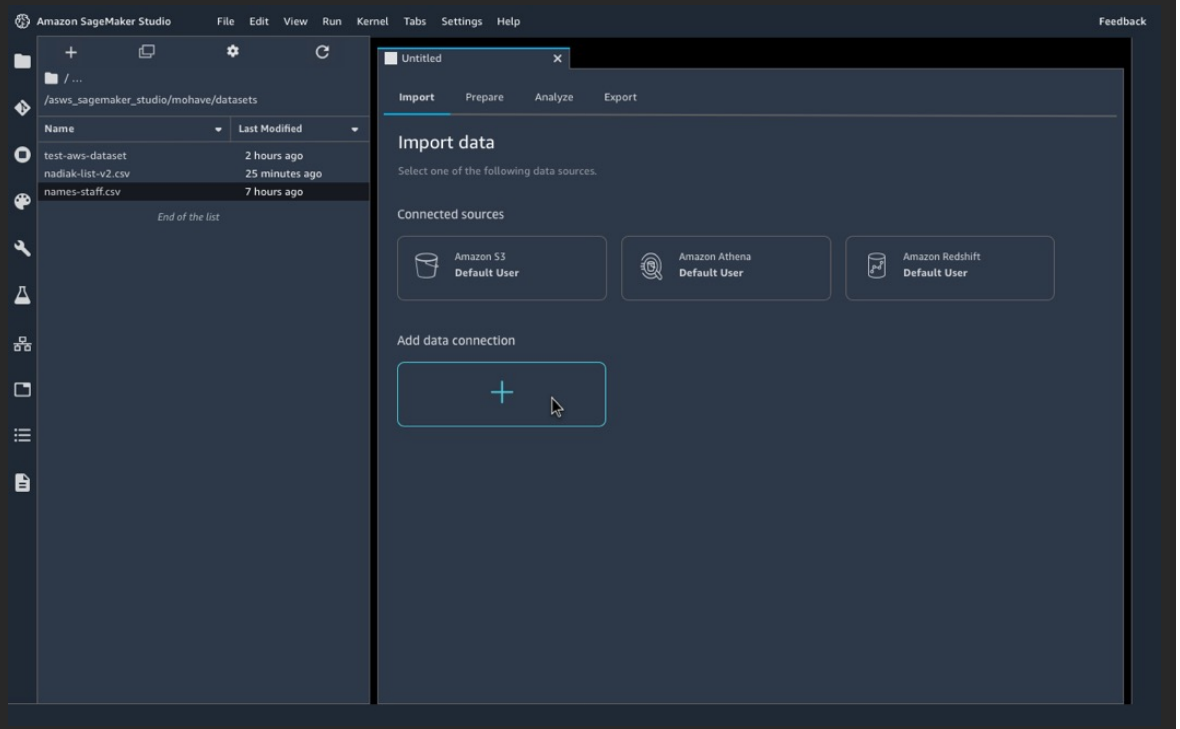
Quickly select data from Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon S3, and features from SageMaker Feature Store

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



Easily transform data

TRANSFORM

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

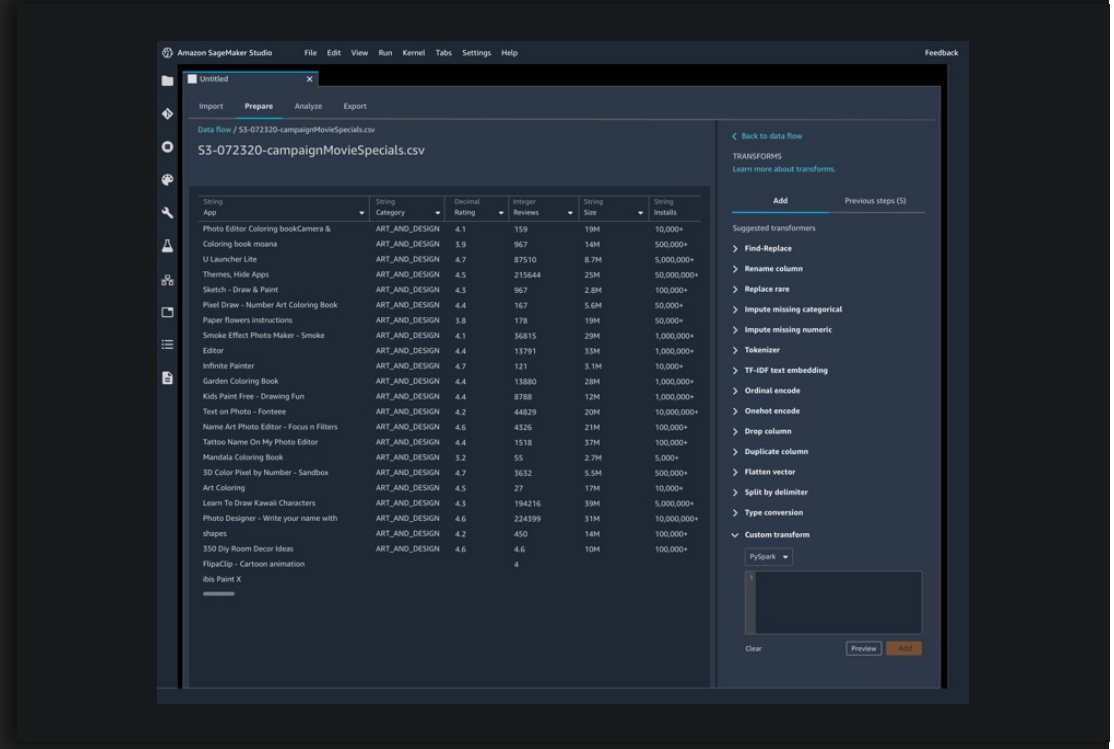
PREPARE

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



Understand your data visually

UNDERSTAND

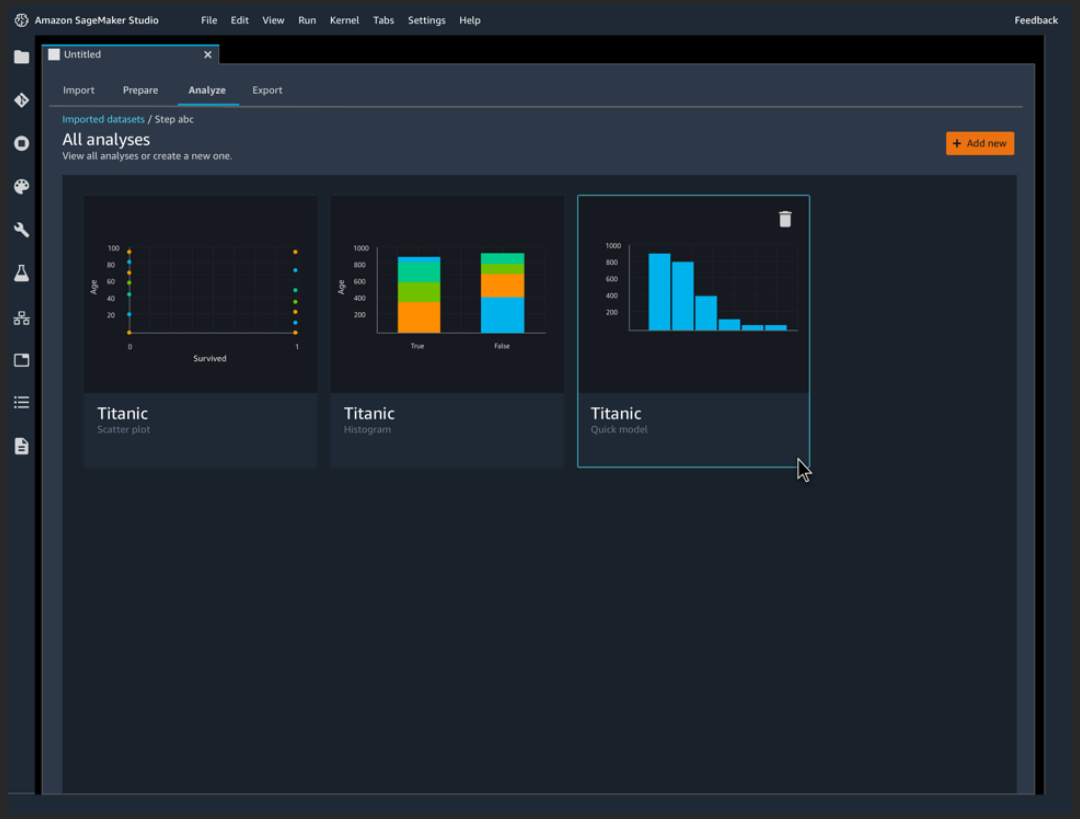
Intuitively understand your data with a set of preconfigured visualization templates which include histograms and scatter plots

CREATE

Create your own custom templates using Altair for data visualization

SAVE

Save and organize your visualizations easily for future reference and reuse



Quickly estimate ML model accuracy

IDENTIFY

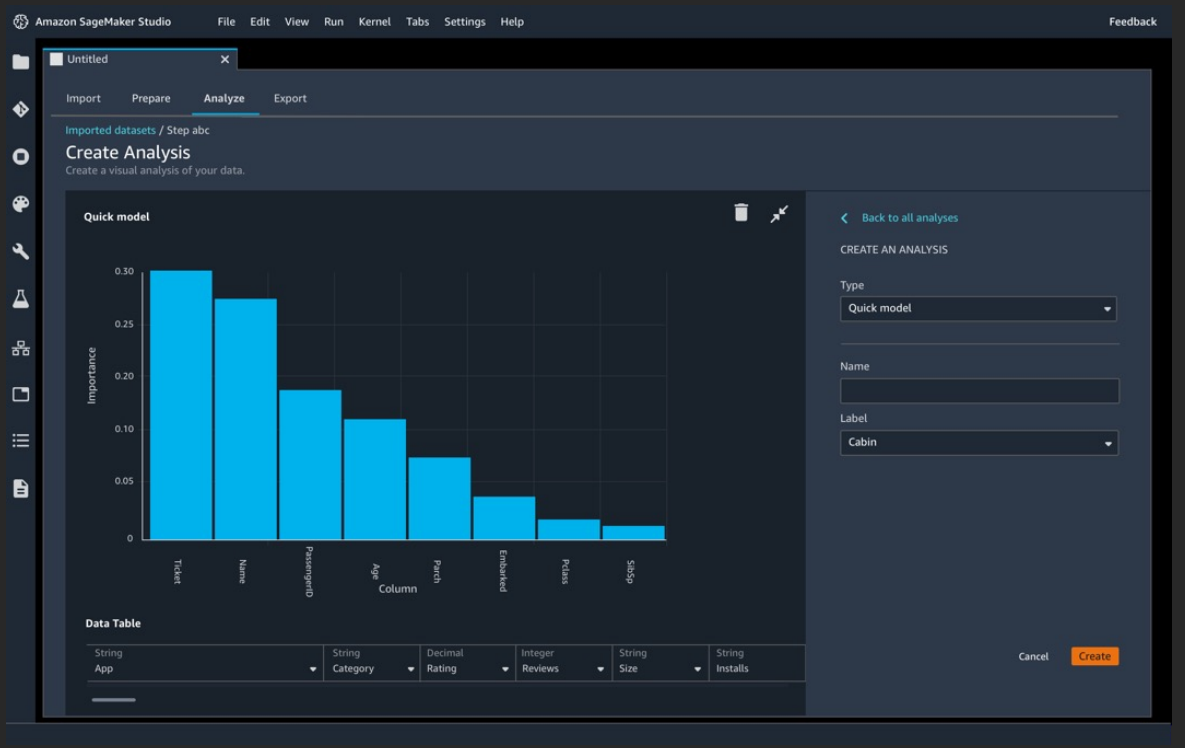
Quickly identify inconsistencies in your data preparation workflow and diagnose issues before ML models are deployed into production

DETECT

Detect which features are contributing to model performance relative to others

DETERMINE

Determine if additional feature engineering is needed to improve model performance



Deploy data preparation workflows into production with a single click

EXPORT

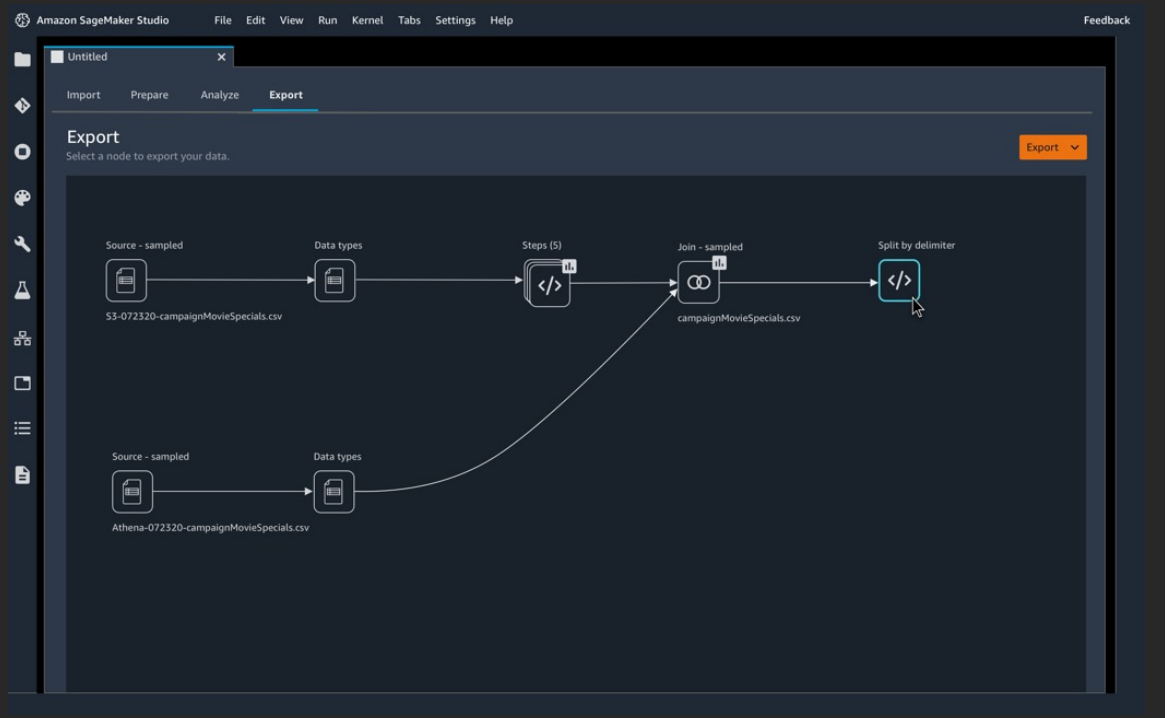
Export data preparation workflow as a processing job notebook or Python code with a single click

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



Lab 2 SageMaker Data Wrangler

Q & A

Thank you!

Michael Lin

linmicht@amazon.com