

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

Michael Lin

Solutions Architect, AWS



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
---	--------------	-----------	--------------------------	------------------------	----------	--------------	------------	---------

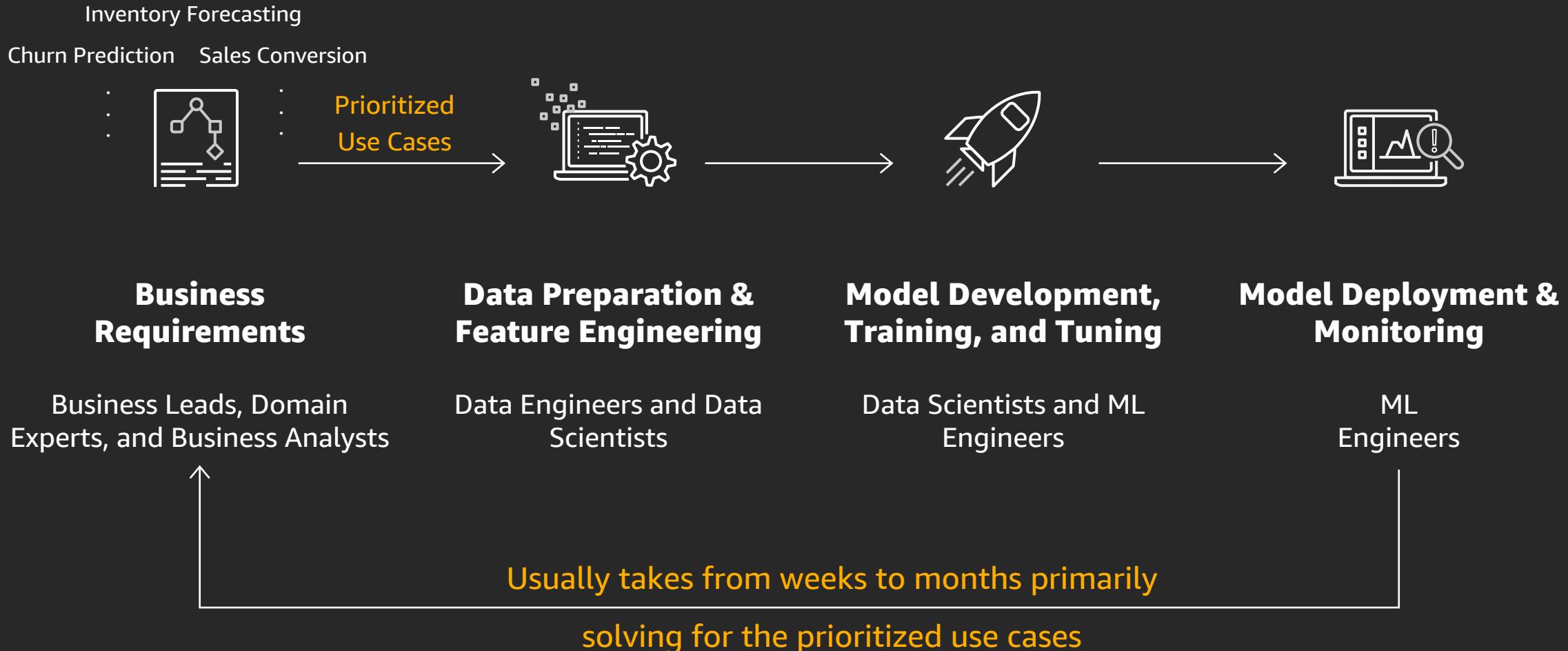
ML frameworks + infrastructure

FRAMEWORKS	INTERFACES	INFRASTRUCTURE								
 TensorFlow  PyTorch	 MXNet  Keras	 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



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

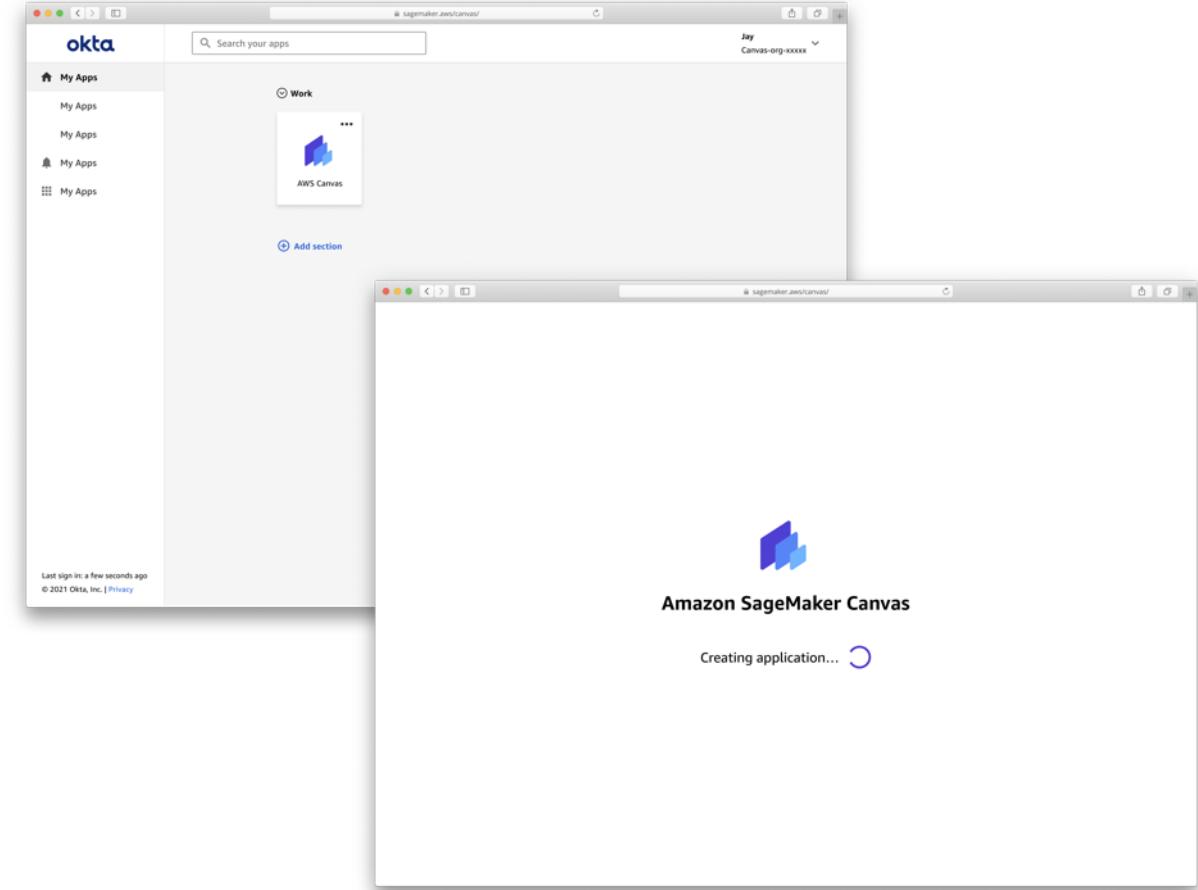


Share ML models and collaborate with data science teams



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**



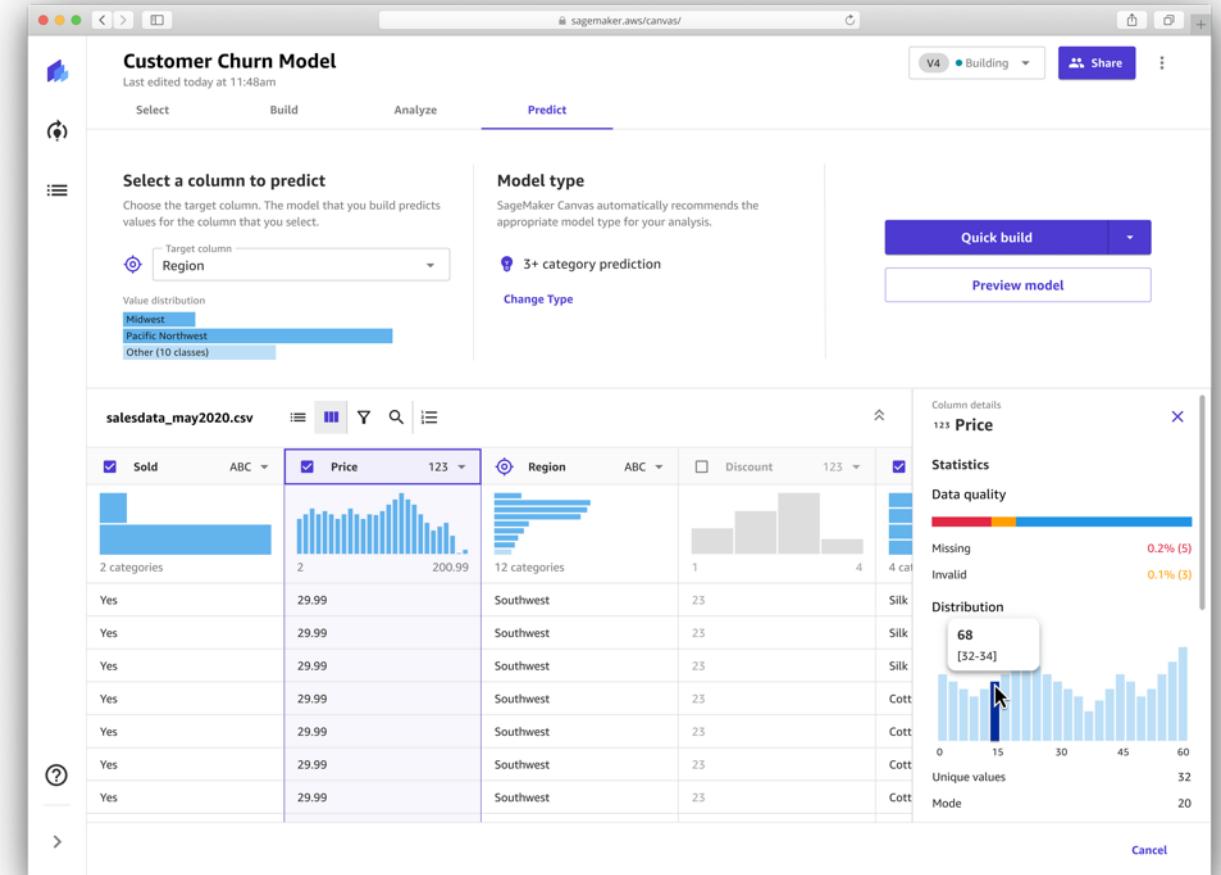
Combine datasets from various sources like local disk, Amazon S3, Amazon RedShift, and Snowflake

The screenshot shows the AWS SageMaker Canvas interface titled "Import Data". At the top, there are connection buttons for "Upload", "S3", "Snowflake Crystal 1", and "Redshift Crystal 1". A "Connection name" dropdown is set to "database1", and a "Context" dropdown is set to "Autosaved 8/9/21 at 11:34 AM". A search bar is present. On the left, a tree view shows "database1" expanded to include "schema1" and "schema2", with "table1" under "schema2". On the right, a diagram shows two boxes labeled "table1.csv" and "table2.csv" connected by a central node, indicating a join operation. Below the diagram, the "Import preview" section displays a table of data with columns: Sold, ABC, Price, 123, Region, ABC, Discount, 123, Fabric, ABC, Age, and 123. The first few rows of data are:

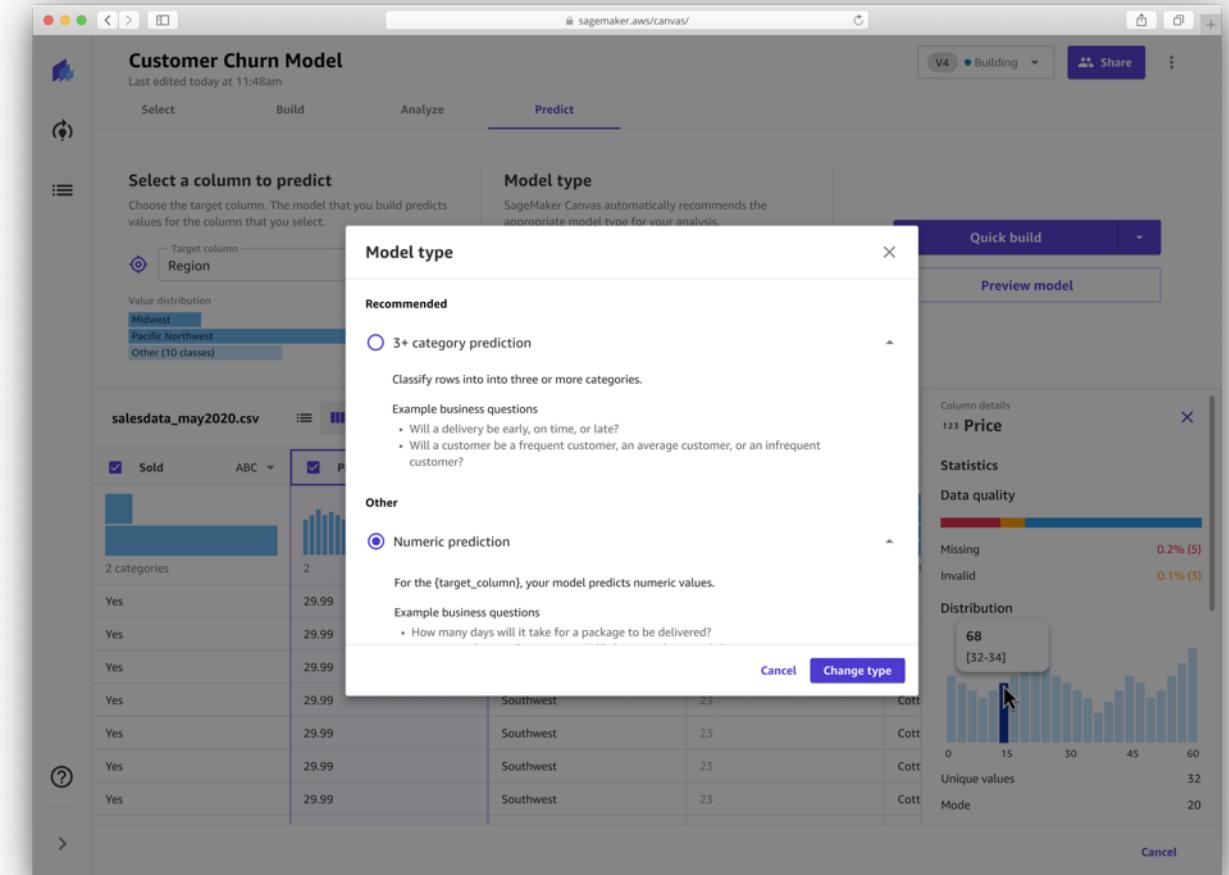
Sold	ABC	Price	123	Region	ABC	Discount	123	Fabric	ABC	Age	123
Yes		29.99		Southwest		23		Cotton		27	
Yes		29.99		Southwest		23		Silk		35	
Yes		29.99		Southwest		23		Silk		32	
Yes		29.99		Southwest		23		Silk		32	
Yes		29.99		Southwest		23		Cotton		30	

At the bottom, a message says "Previewing the first 100 rows" and there are "Close" and "Import data" buttons.

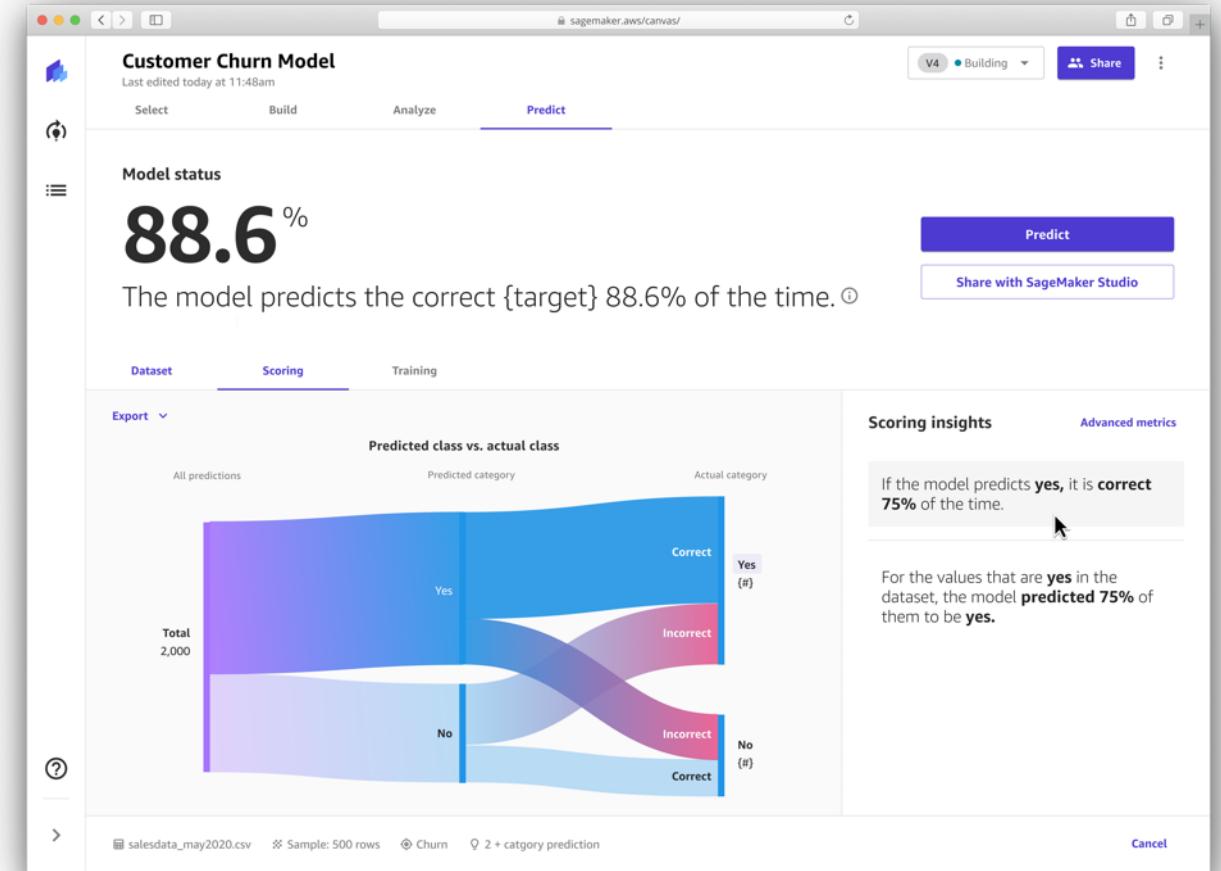
Quickly understand and prepare your data via a visual interface



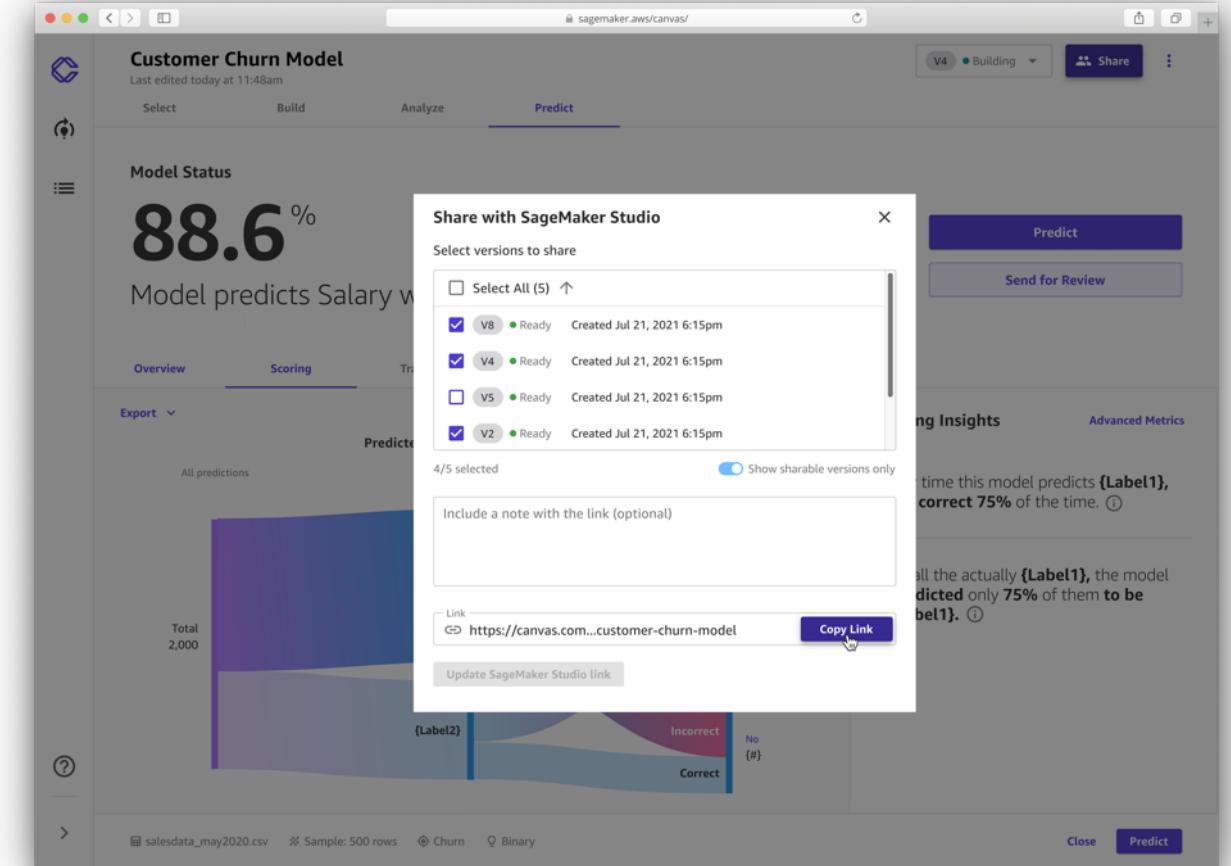
Automatically build an accurate ML model for your dataset



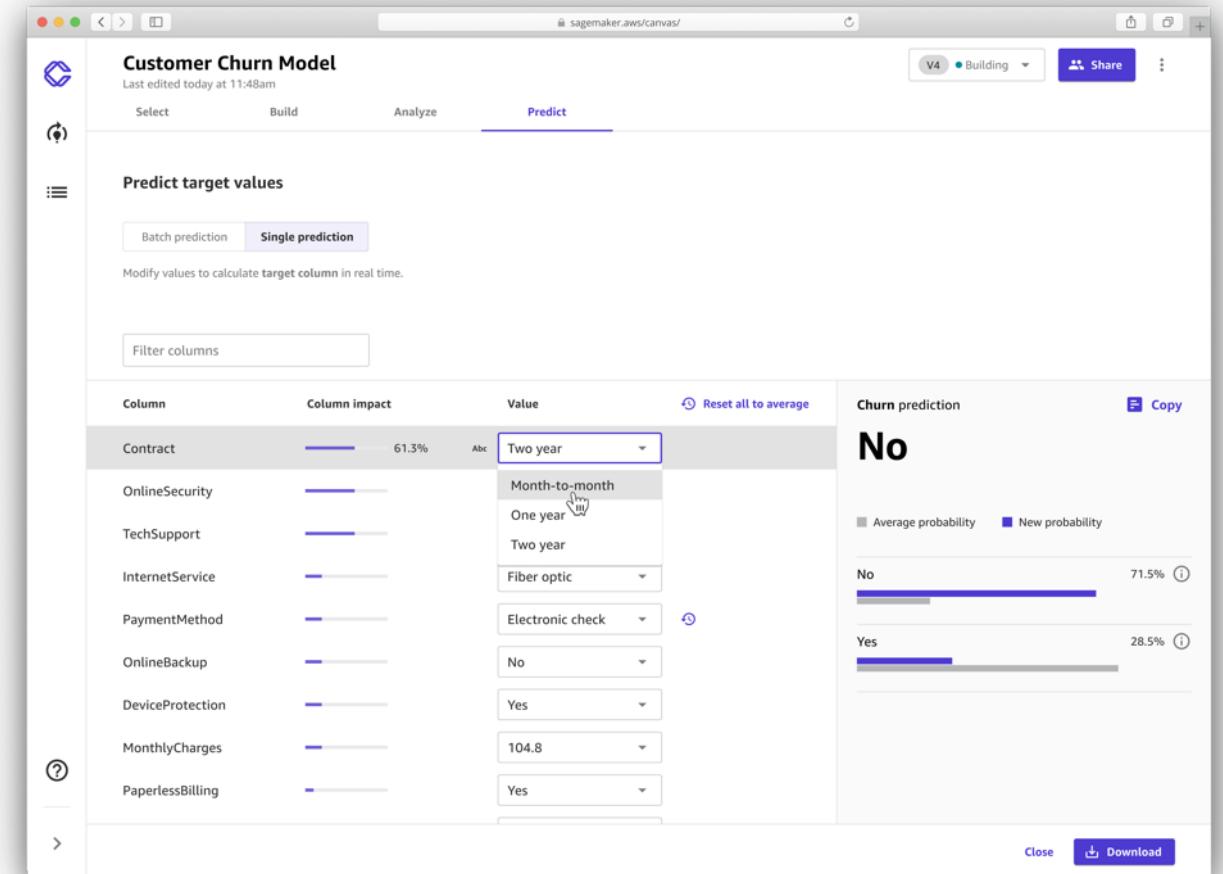
Get the first ML model in minutes. Review advanced metrics and feature importance to understand and explain predictions.



Easily share your models with data scientists to get feedback



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



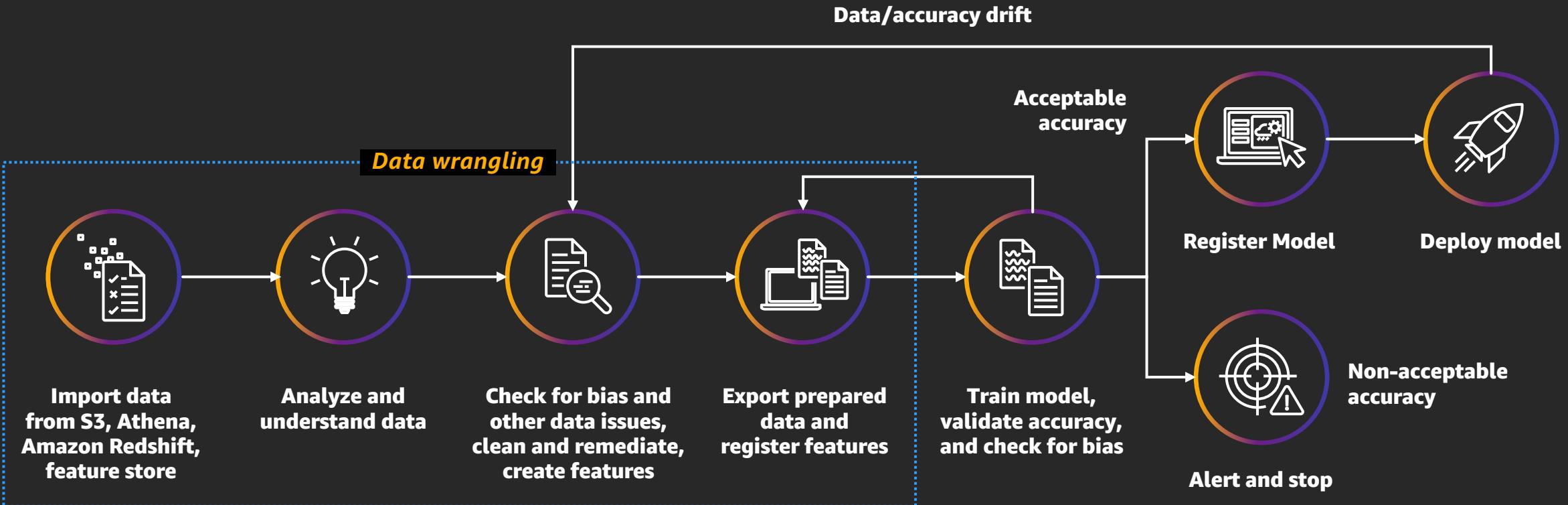
Lab 1 SageMaker Canvas



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

Amazon SageMaker Data Wrangler

The fastest and easiest way to
prepare data for machine learning

Easily transform data
with 300+ built-in
data transformations

Customize data
transformations in PySpark,
SQL, or Pandas

Quickly select
and query data

Deploy data preparation
workflows into production
with a single click

Quickly estimate ML
model accuracy

Understand
data visually

Quickly select and query data

SELECT

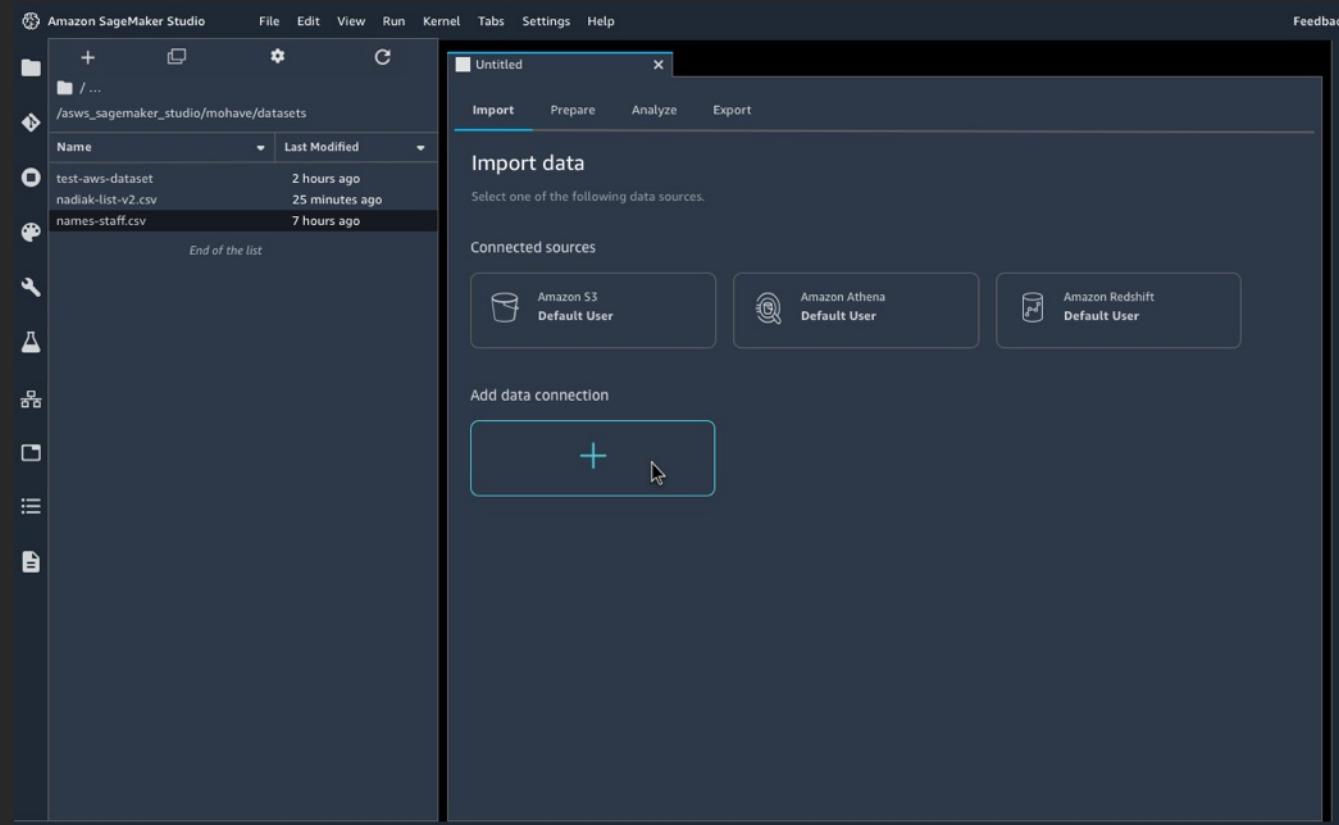
Quickly select data from Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon S3, and features from SageMaker Feature Store

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

The screenshot shows the Amazon SageMaker Studio Data Flow interface. The main area displays a table of data transforms, each with a name, category, rating, reviews, size, and installs. The categories listed are ART_AND DESIGN, and the ratings range from 3.9 to 4.7. The sidebar on the right lists various transform suggestions under sections like 'Suggested transformers' and 'Custom transform'. The 'Custom transform' section is expanded, showing options for PySpark, including 'Preview' and 'Add' buttons.

Name	Category	Rating	Reviews	String Size	Installs
Photo Editor Coloring book Camera &	ART_AND DESIGN	4.1	159	19M	10,000+
Coloring book moana	ART_AND DESIGN	3.9	967	14M	500,000+
U Launcher Lite	ART_AND DESIGN	4.7	87510	8.7M	5,000,000+
Themes, Hide Apps	ART_AND DESIGN	4.5	215644	25M	50,000,000+
Sketch - Draw & Paint	ART_AND DESIGN	4.5	967	2.8M	100,000+
Pixel Draw - Number Art Coloring Book	ART_AND DESIGN	4.4	167	5.6M	50,000+
Paper flowers instructions	ART_AND DESIGN	3.8	178	19M	50,000+
Smoke Effect Photo Maker - Smoke	ART_AND DESIGN	4.1	36815	29M	1,000,000+
Editor	ART_AND DESIGN	4.4	13791	35M	1,000,000+
Infinite Painter	ART_AND DESIGN	4.7	121	3.1M	10,000+
Garden Coloring Book	ART_AND DESIGN	4.4	13880	28M	1,000,000+
Kids Paint Free - Drawing Fun	ART_AND DESIGN	4.4	8788	12M	1,000,000+
Text on Photo - Fontee	ART_AND DESIGN	4.2	44829	20M	10,000,000+
Name Art Photo Editor - Focus n Filters	ART_AND DESIGN	4.6	4326	21M	100,000+
Tattoo Name On My Photo Editor	ART_AND DESIGN	4.4	1518	37M	100,000+
Mandala Coloring Book	ART_AND DESIGN	3.2	55	2.7M	5,000+
3D Color Pixel by Number - Sandbox	ART_AND DESIGN	4.7	3632	5.5M	500,000+
Art Coloring	ART_AND DESIGN	4.5	27	17M	10,000+
Learn To Draw Kawaii Characters	ART_AND DESIGN	4.3	194216	39M	5,000,000+
Photo Designer - Write your name with shapes	ART_AND DESIGN	4.6	224399	31M	10,000,000+
350 DIY Room Decor Ideas	ART_AND DESIGN	4.2	450	14M	100,000+
FlipaClip - Cartoon animation	ART_AND DESIGN	4.6	4,6	10M	100,000+
ibis Paint X					

Understand your data visually

UNDERSTAND

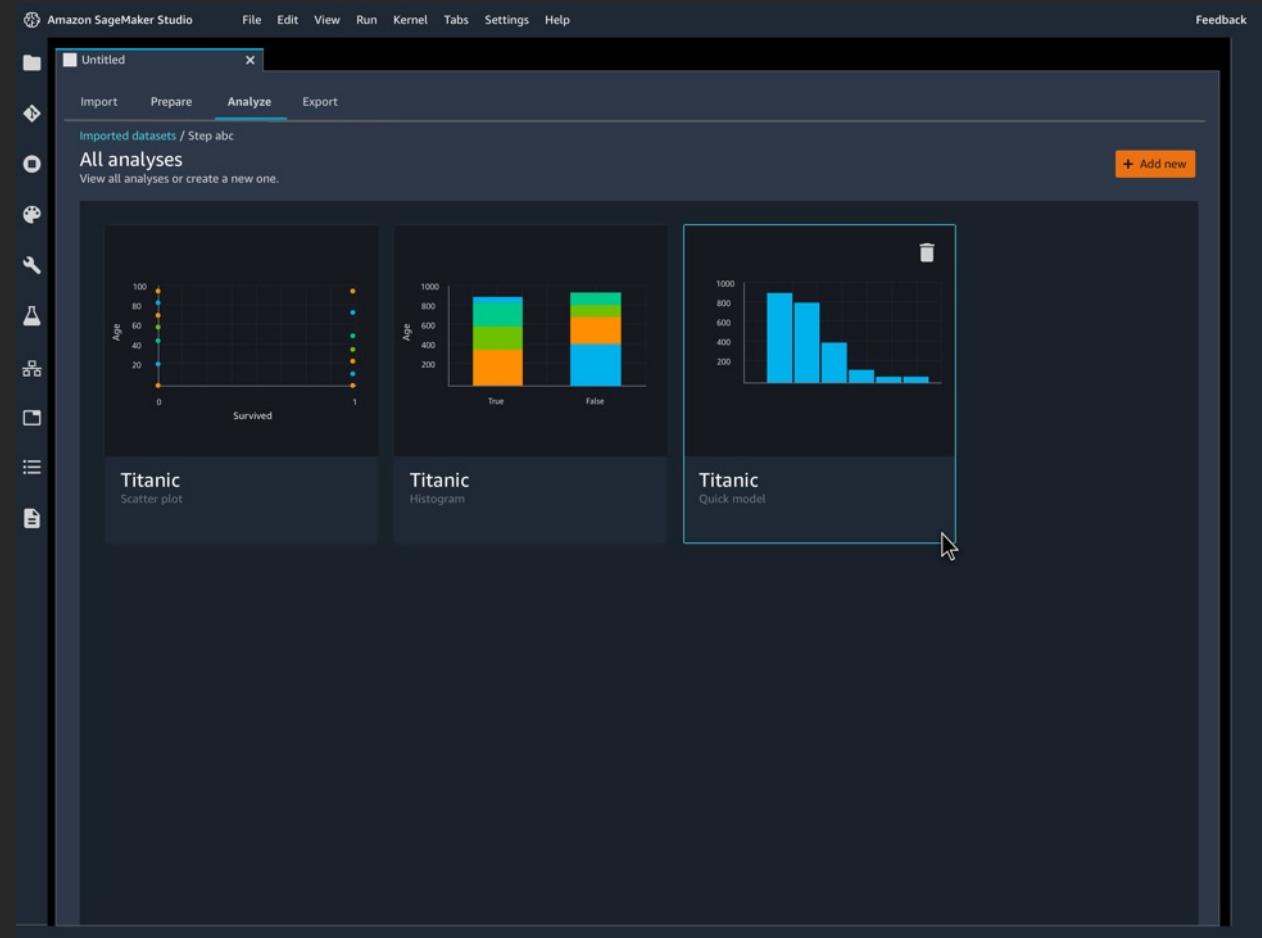
Intuitively understand your data with a set of preconfigured visualization templates which include histograms and scatter plots

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

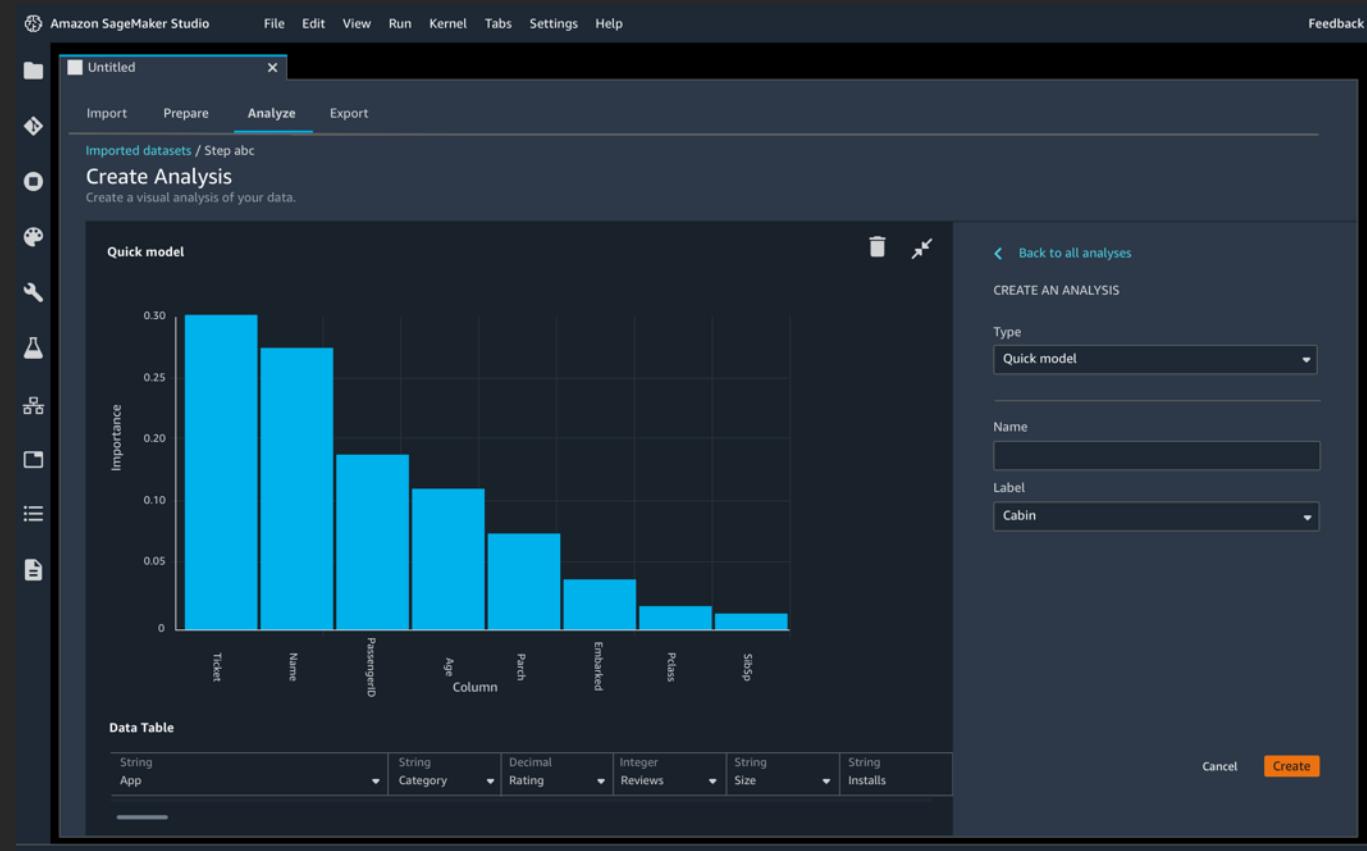
Quickly identify inconsistencies in your data preparation workflow and diagnose issues before ML models are deployed into production

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

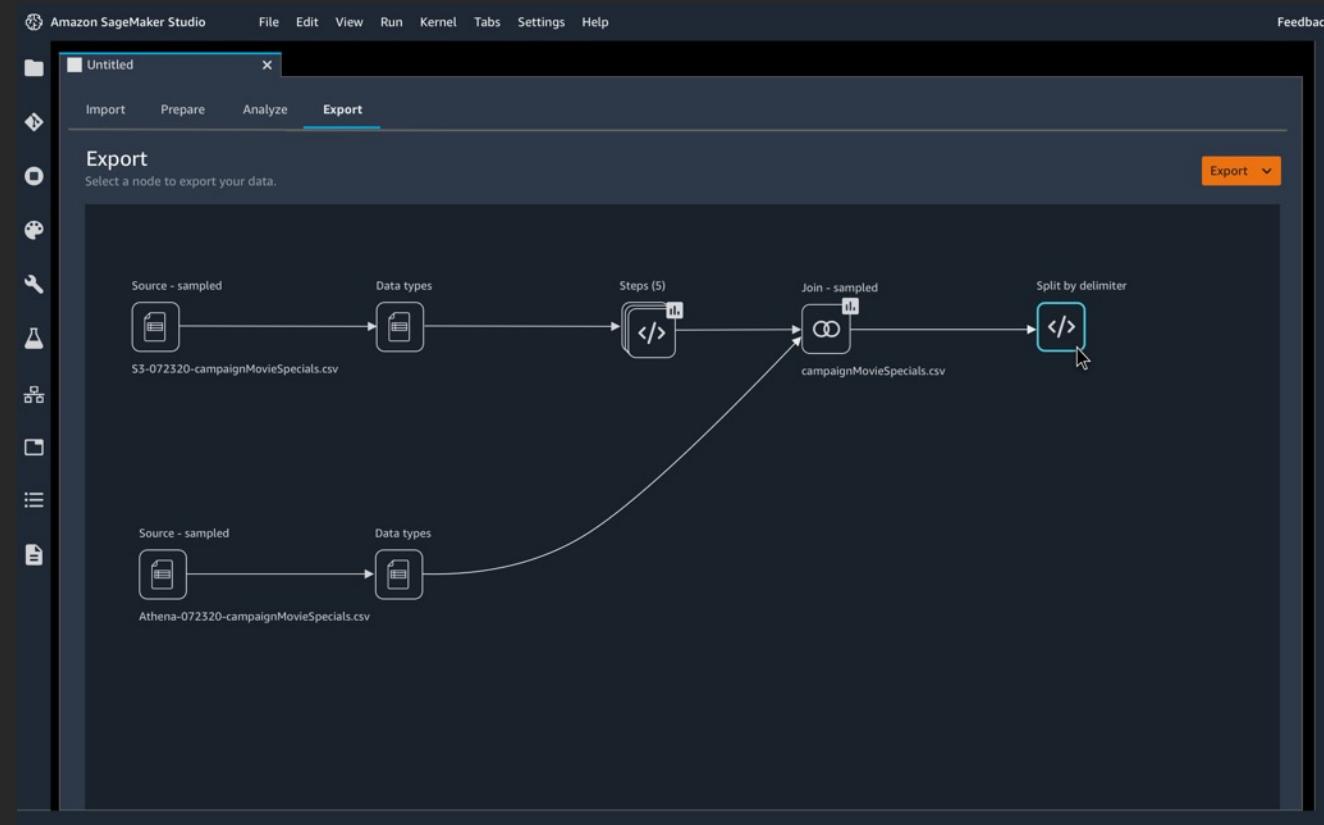
Export data preparation workflow as a processing job notebook or Python code with a single click

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

