

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

Michael Lin

Senior Solutions Architect
Amazon Web Services



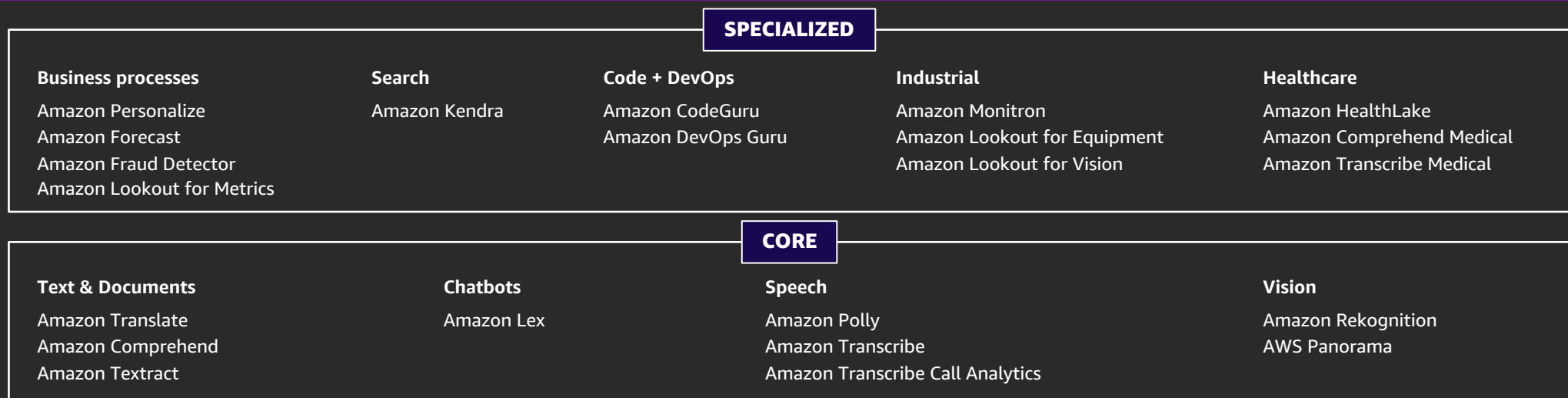
Agenda

- Amazon AI/ML Stack
- Amazon SageMaker Canvas
- Amazon SageMaker Data Wrangler
- Next Step

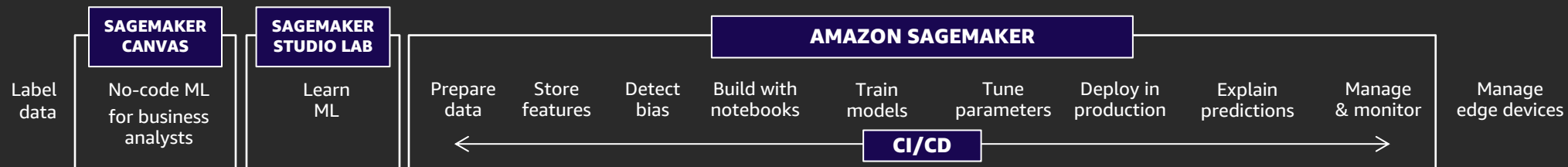
Amazon AI/ML Stack

The AWS AI/ML stack

AI SERVICES



ML SERVICES



ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, Apache MXNet, TensorFlow Amazon EC2 CPUs GPUs AWS Inferentia AWS Trainium Habana Gaudi FPGA Elastic inference

Lab 1 Setup

<https://ppt.cc/fi4Exx>

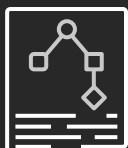
Amazon SageMaker Canvas

How ML drives value creation today

Inventory Forecasting

Churn Prediction Sales Conversion

⋮



⋮

Prioritized
Use Cases



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

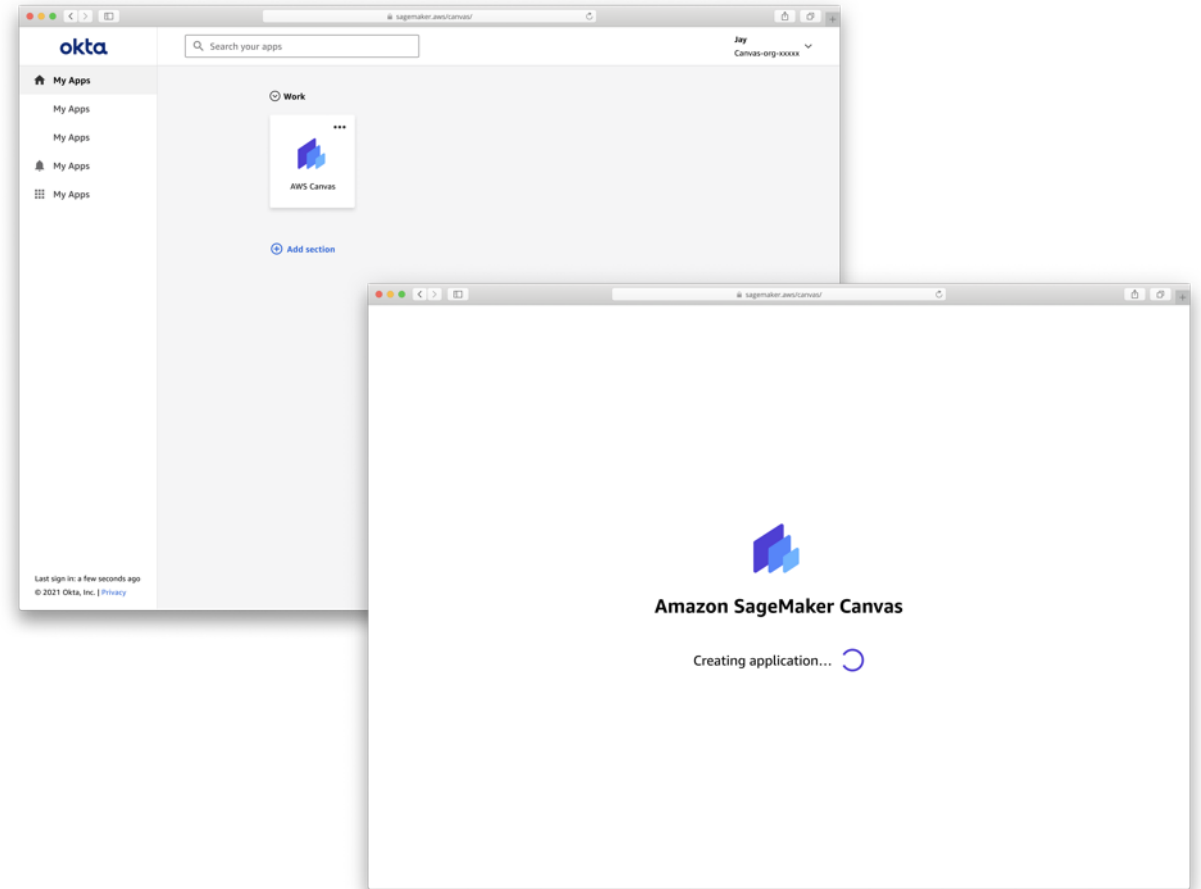


Share ML models and collaborate with data science teams



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

Upload

S3

Snowflake
Crystal 1

Redshift
Crystal 1

Add Connection

Connection name

Context

Autosaved 8/9/21 at 11:34 AM

Edit in SQL

Search

database1

database2

database3

database4

schema1

schema2

table1

table1.csv

table2.csv

Import preview

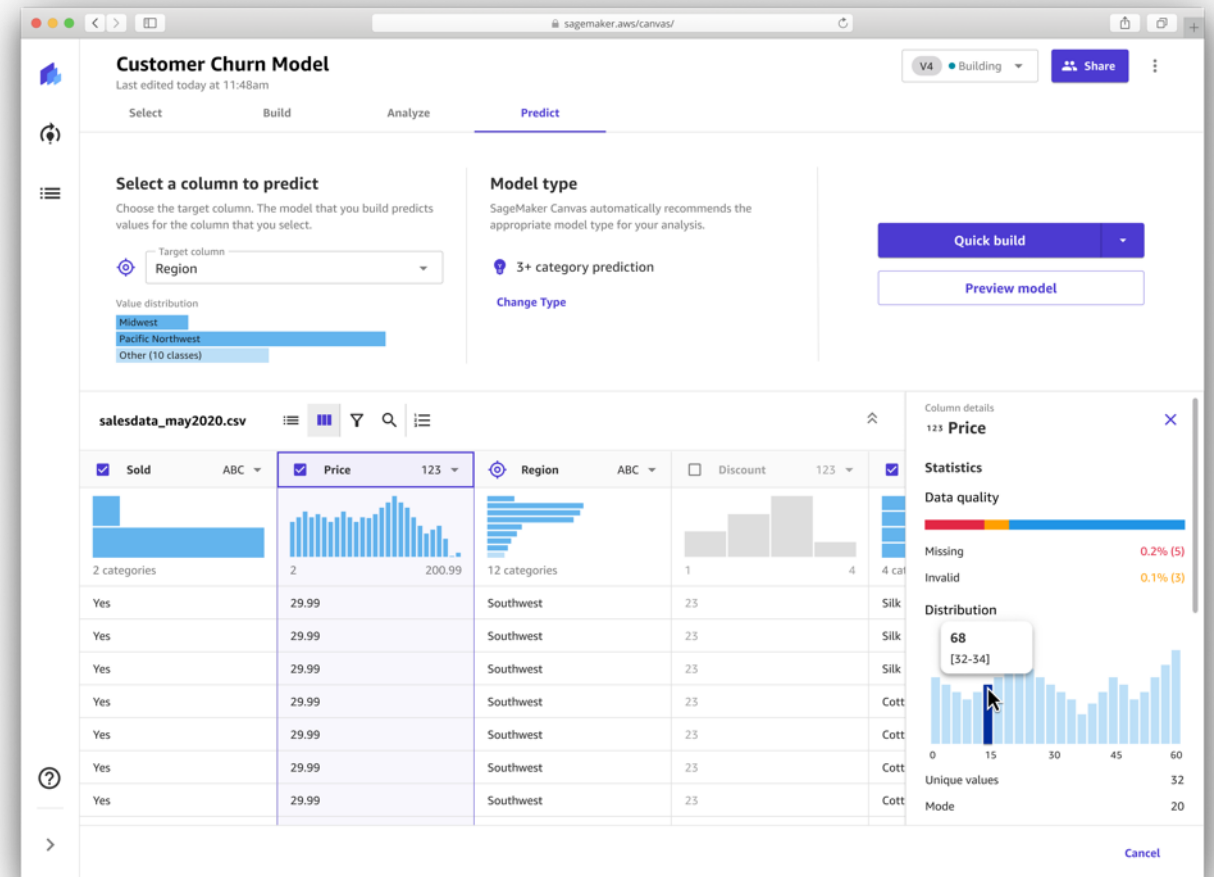
Show dropped columns

<input checked="" type="checkbox"/> Sold	ABC	<input type="checkbox"/> Price	123	<input checked="" type="checkbox"/> Region	ABC	<input checked="" type="checkbox"/> Discount	123	<input checked="" type="checkbox"/> Fabric	ABC	<input checked="" type="checkbox"/> 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	

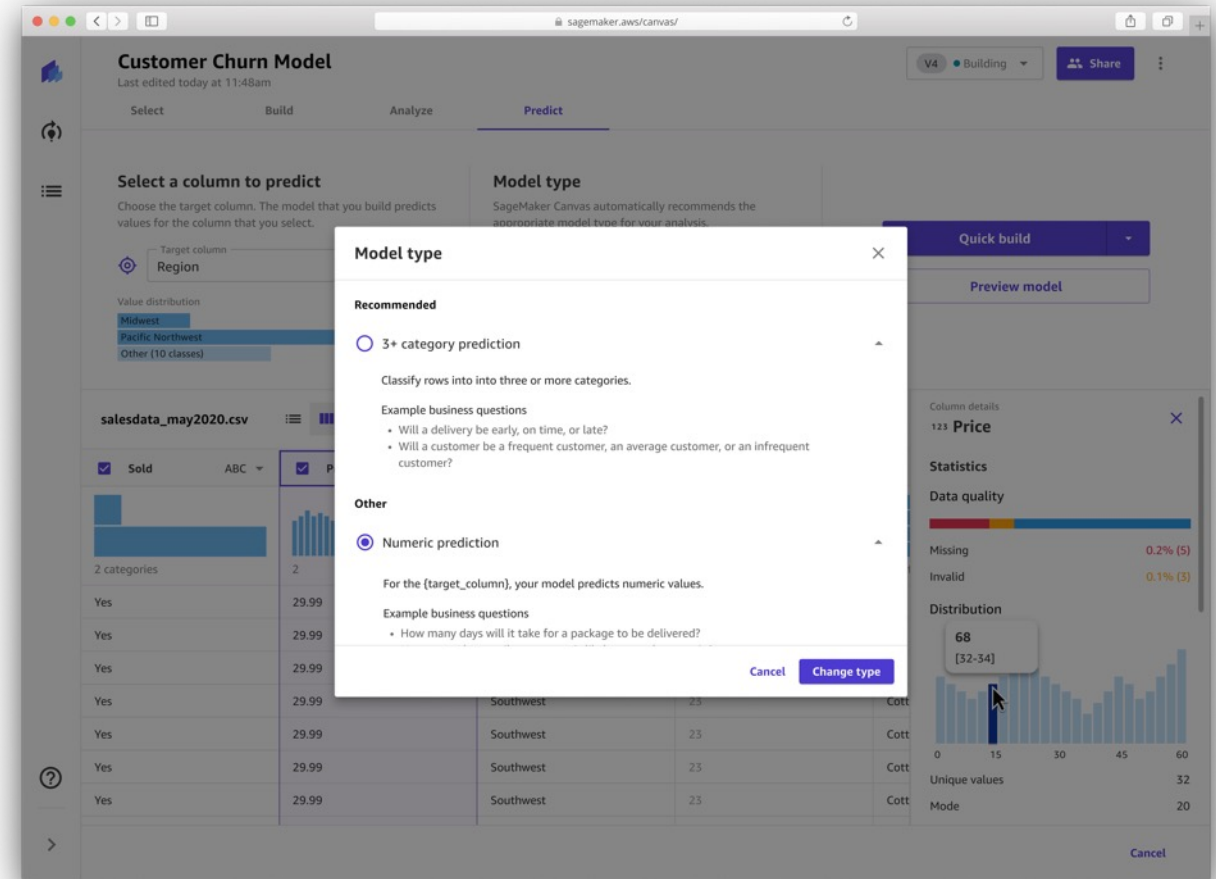
Previewing the first 100 rows

CloseImport data

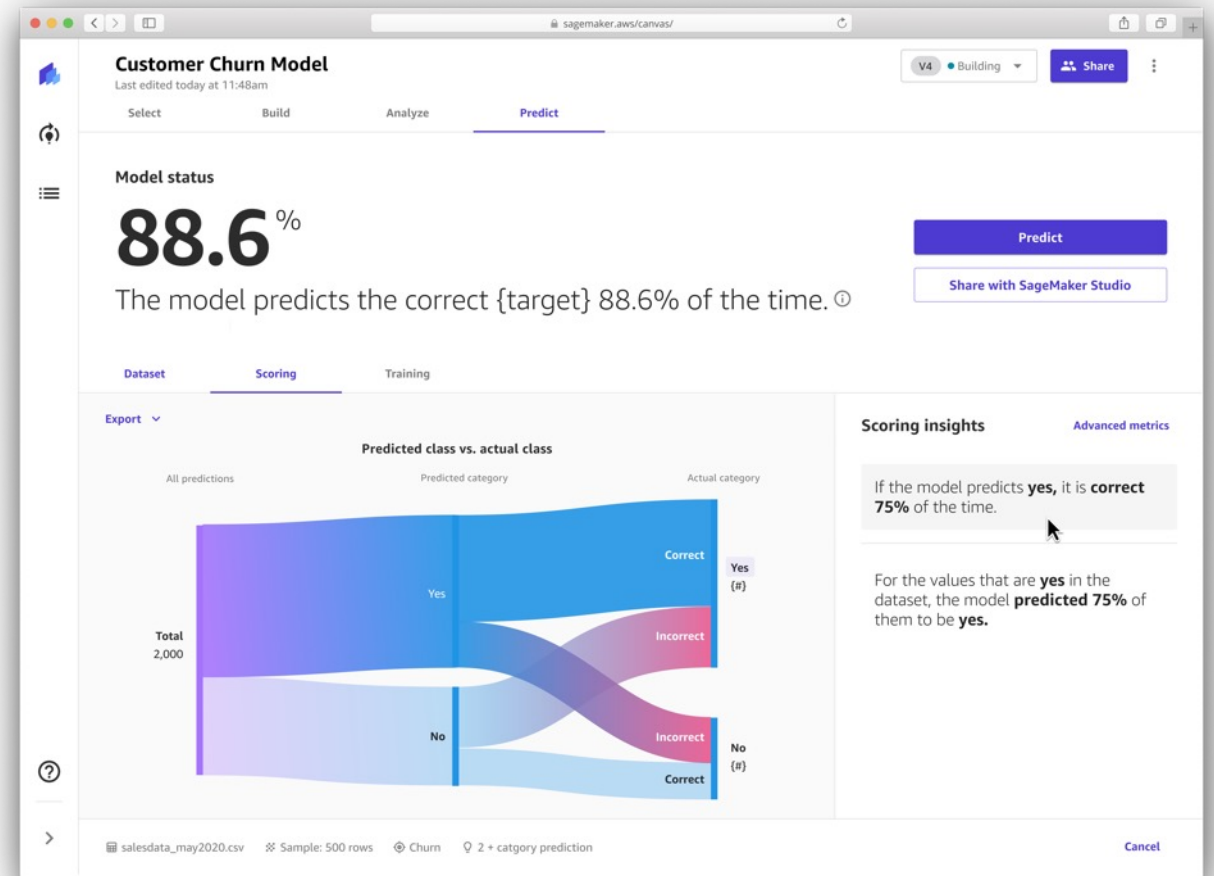
Quickly understand
and prepare your data via
a visual interface



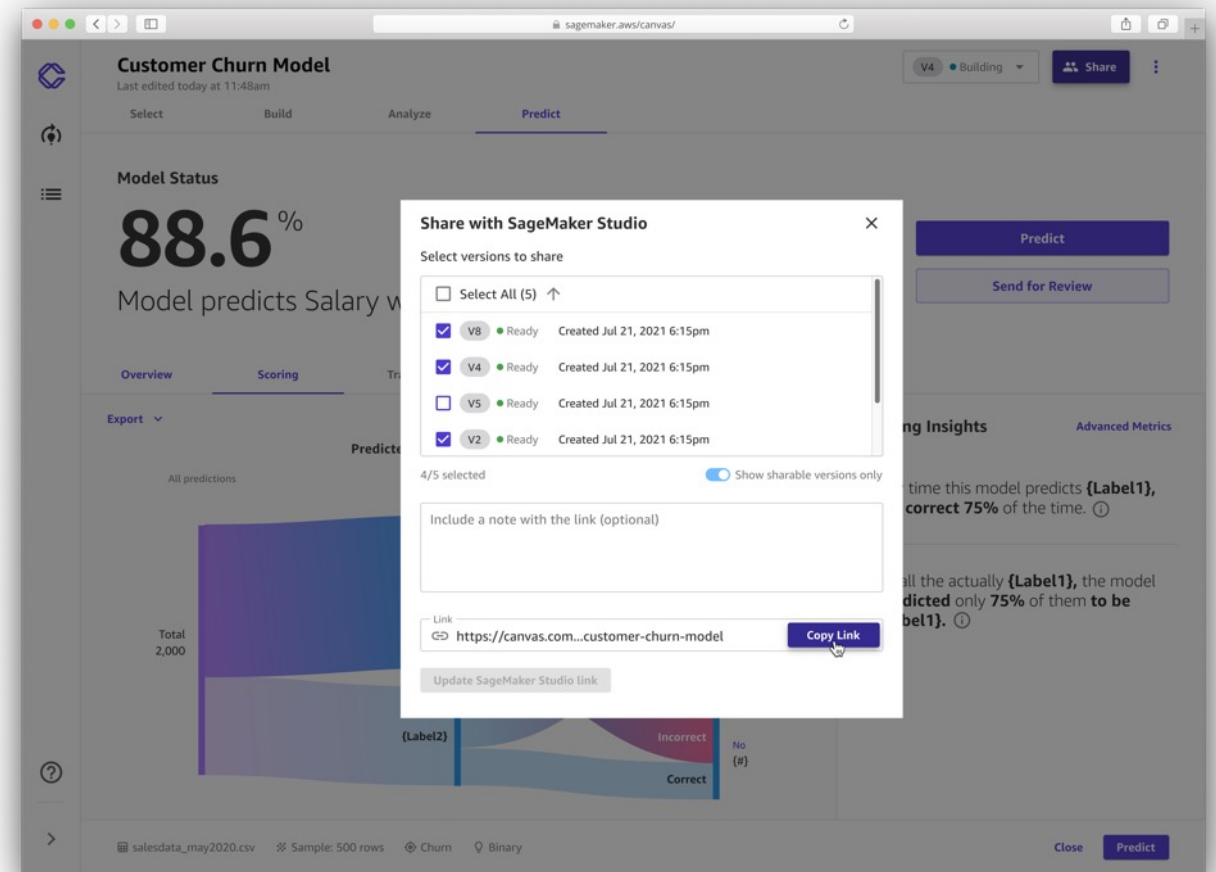
Automatically build an accurate ML model for your dataset



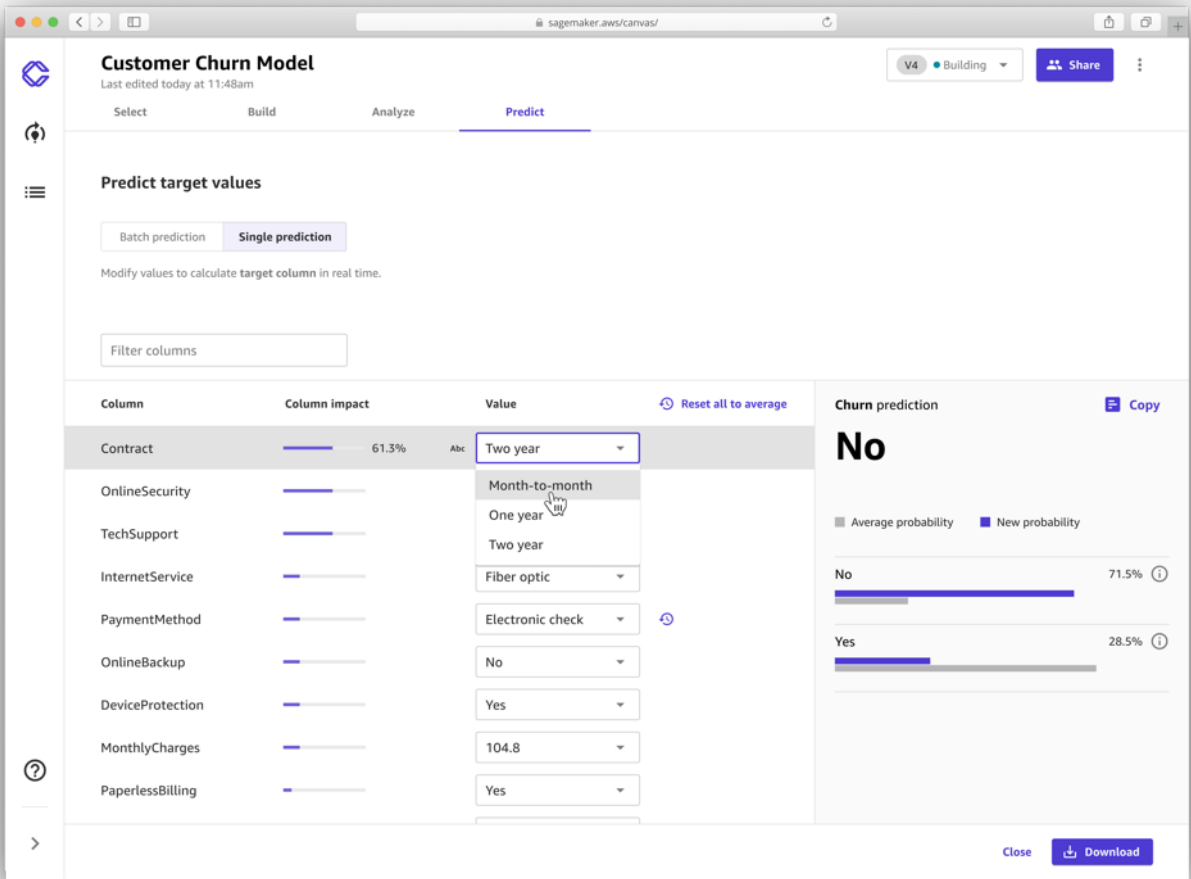
Get the first ML model in minutes. Review advanced metrics and feature importance to understand and explain predictions.



**Easily share your models
with data scientists to
get feedback**



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



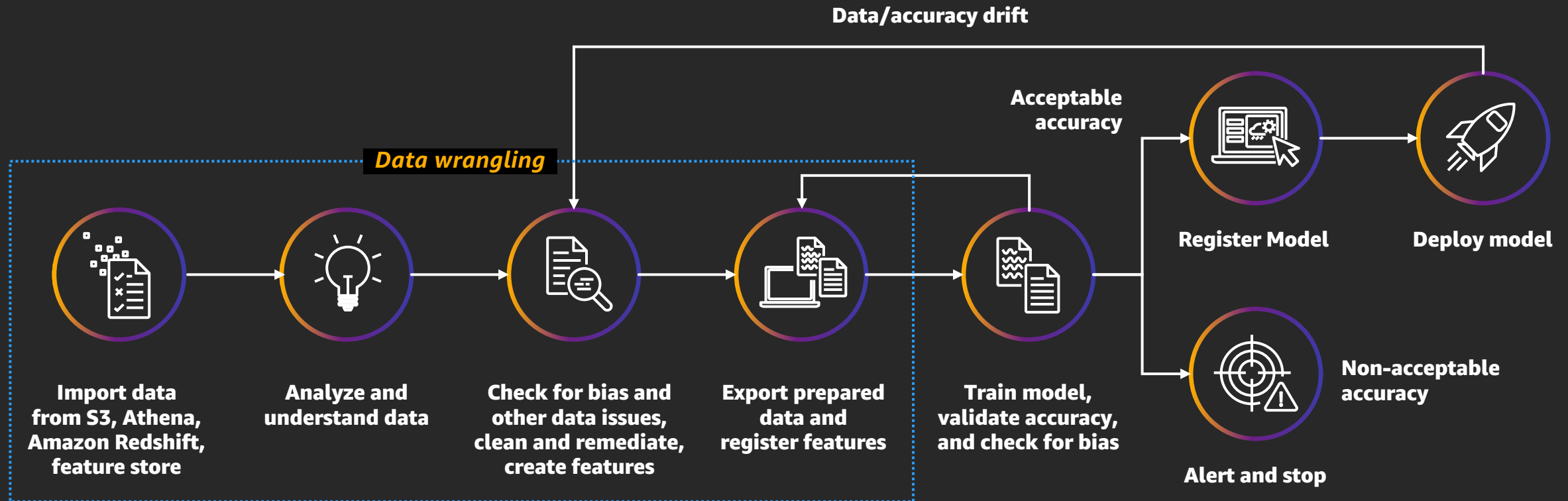
Lab 1 SageMaker Canvas

Amazon SageMaker Data Wrangler

Lab 2 Setup

<https://ppt.cc/fi4Exx>

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

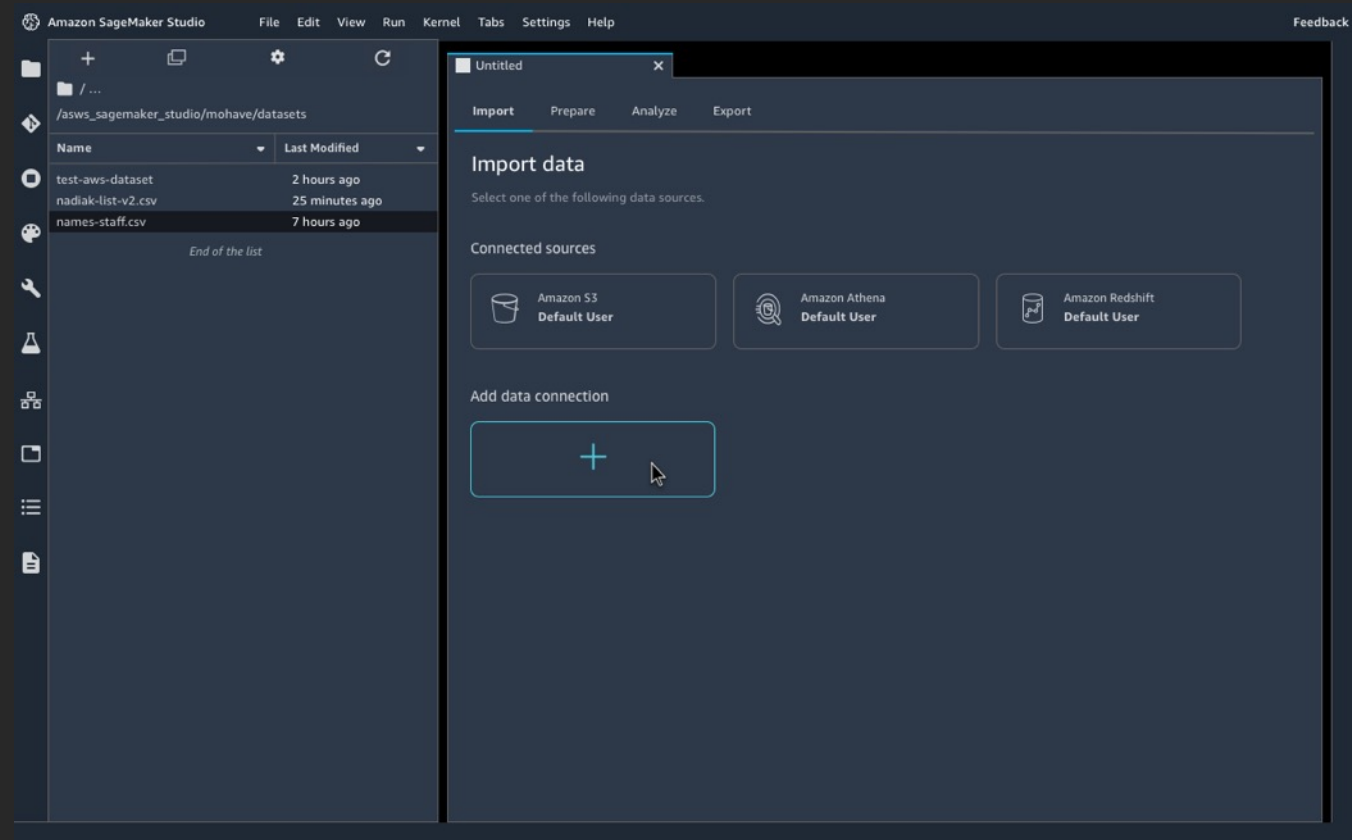
Quickly select data from Amazon Athena, Amazon Redshift, AWS Lake Formation, Amazon S3, and features from SageMaker Feature Store

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

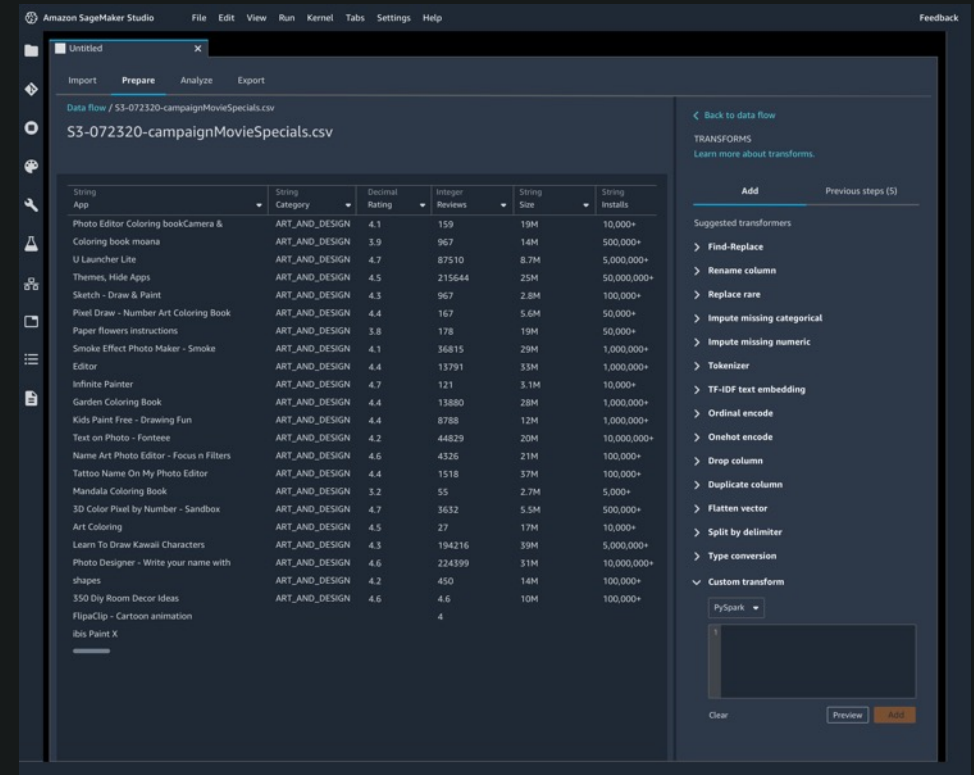
PREPARE

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

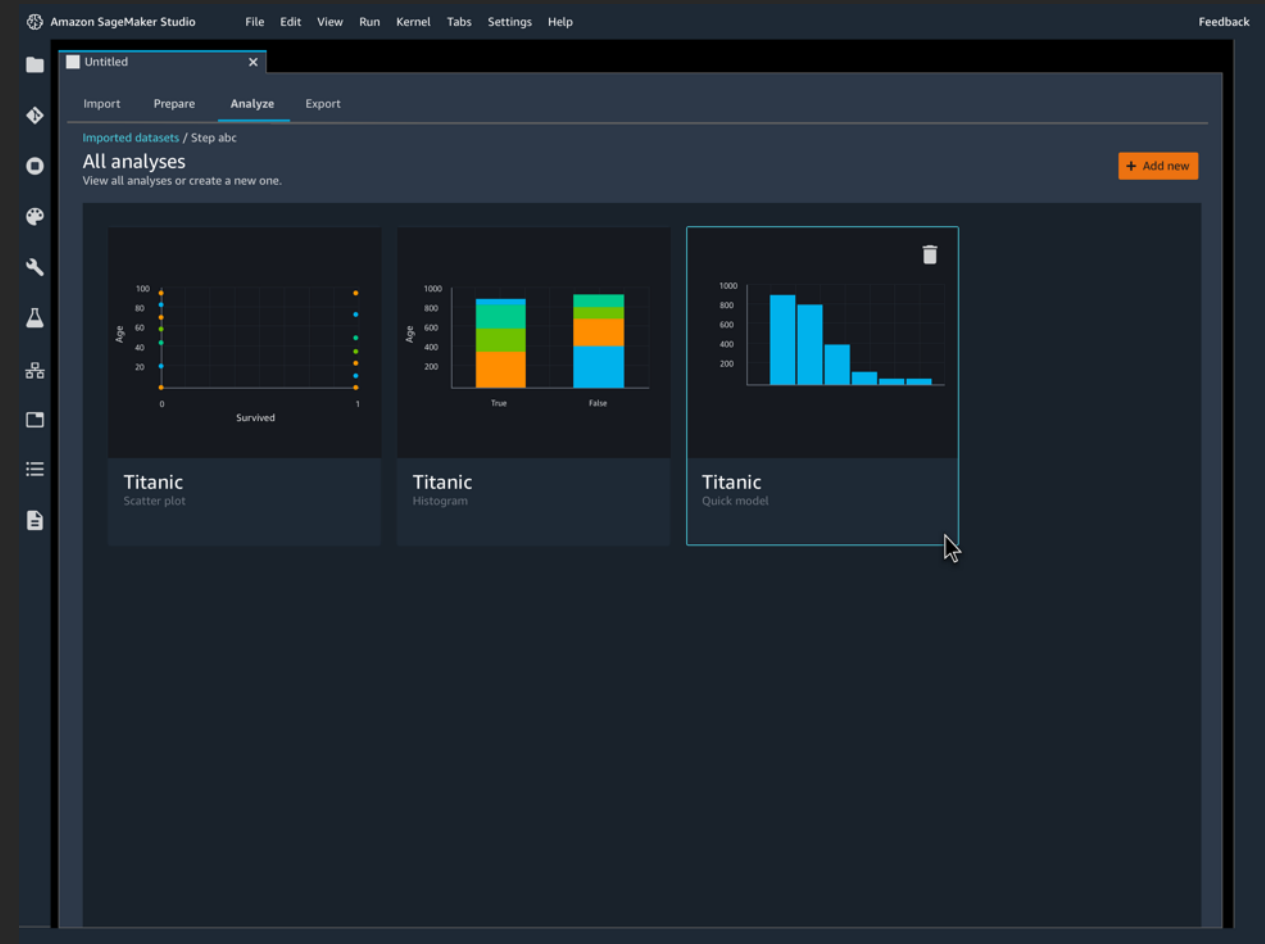
Intuitively understand your data with a set of preconfigured visualization templates which include histograms and scatter plots

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

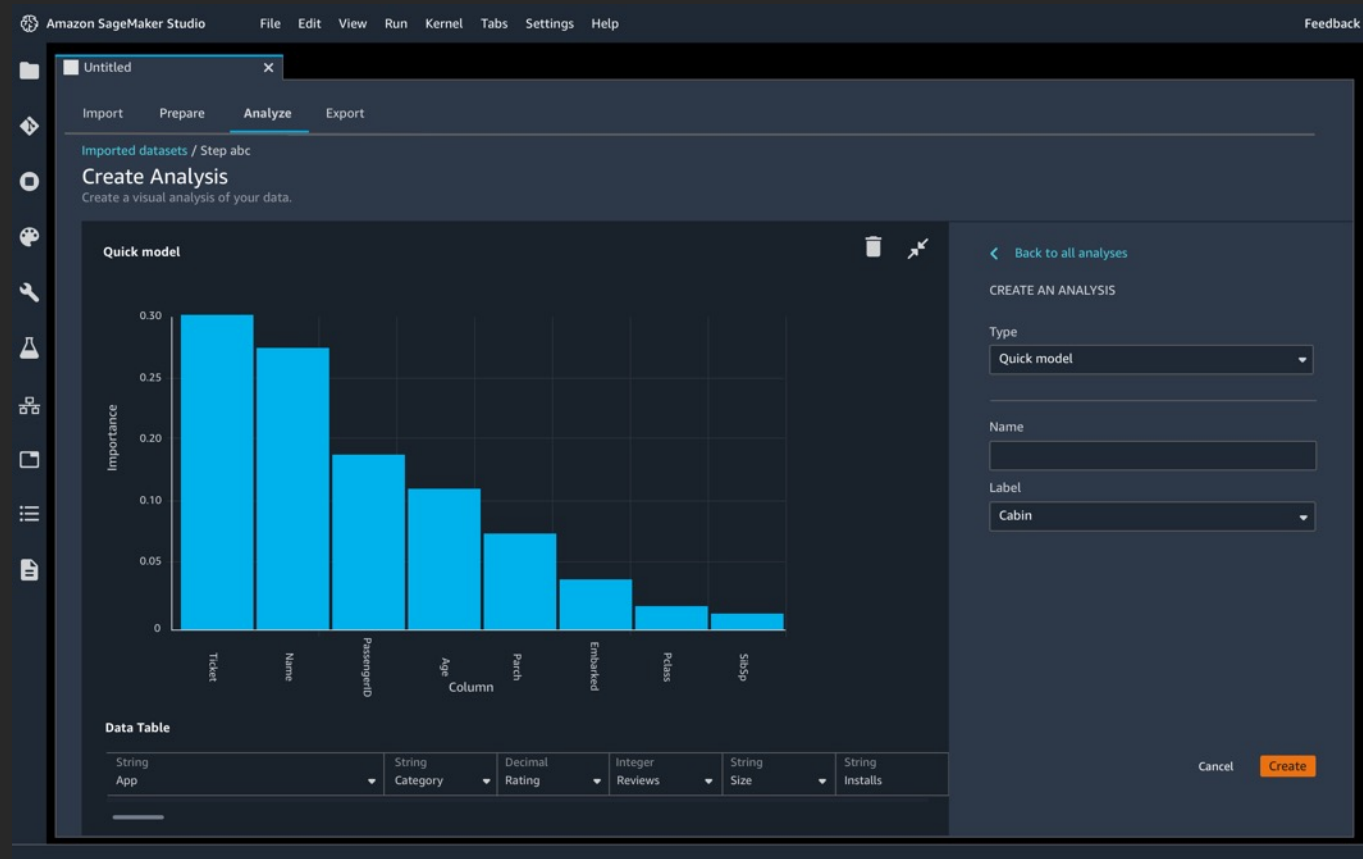
Quickly identify inconsistencies in your data preparation workflow and diagnose issues before ML models are deployed into production

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

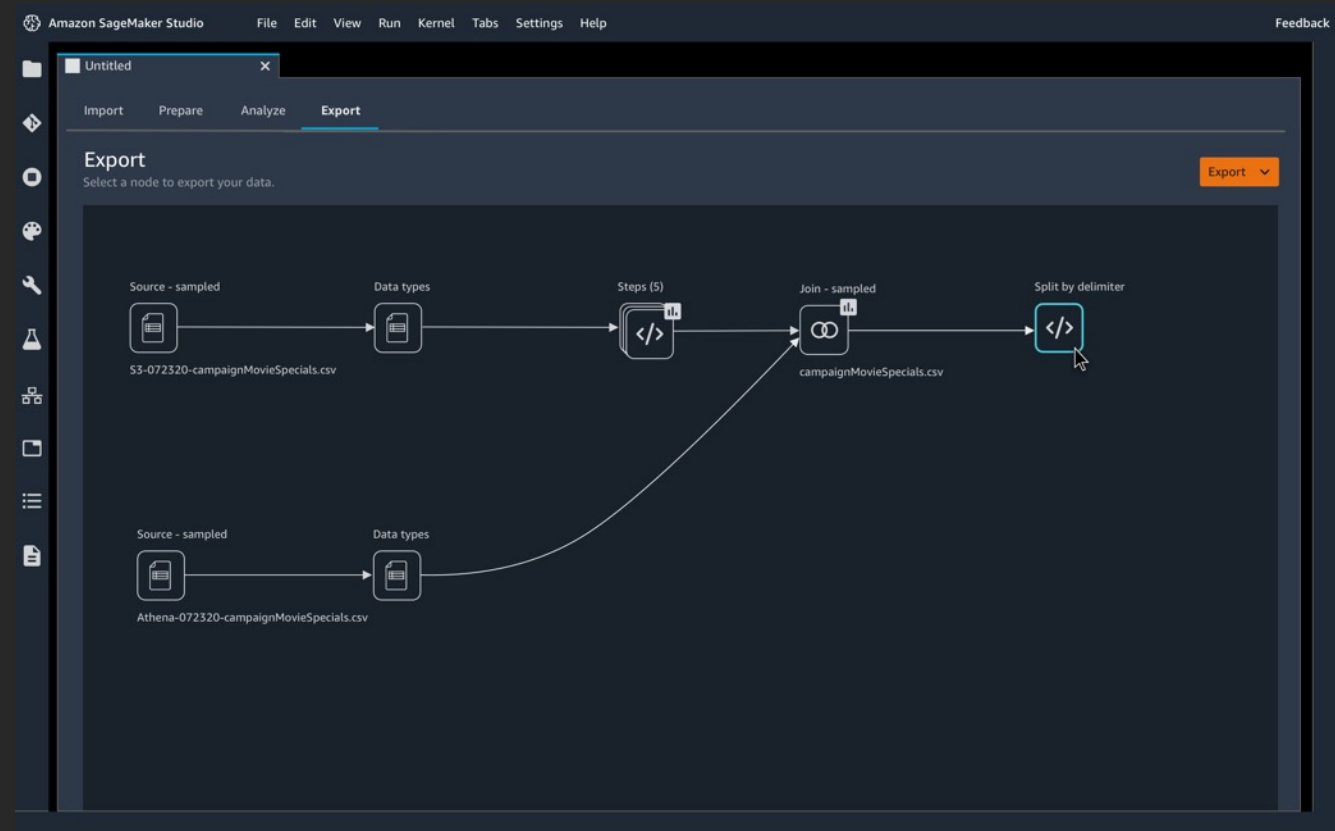
Export data preparation workflow as a processing job notebook or Python code with a single click

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

Data Prep in the Lab

- Ingest Data
- Review Data Types
- Generate Summary Statistics
- Visualize Data
- Handle Missing Data
- Feature Selection
- Feature Engineering
- One-Hot Encoding

Next Step

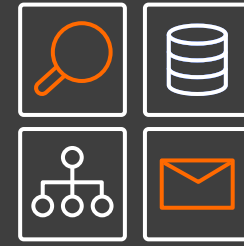
Embarking on the Journey



Training



Proof of concept



Production

Thank you!

Michael Lin

linmicht@amazon.com

