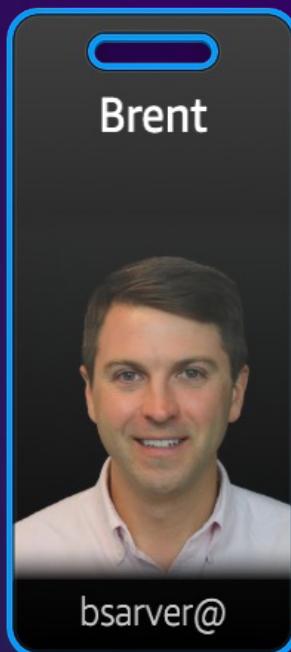
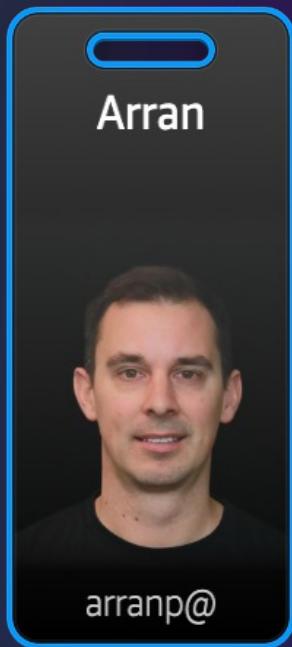
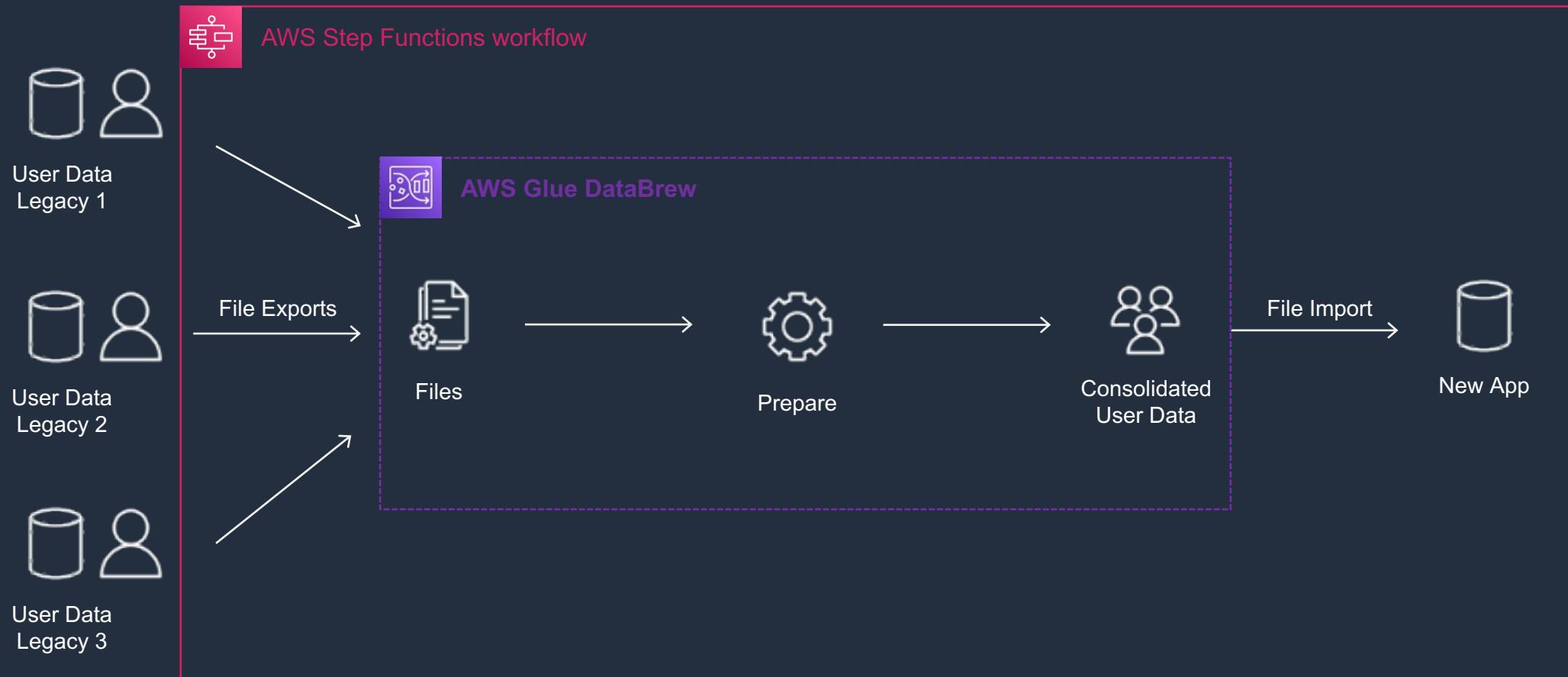


Todays AWS Team



Reflection on Learnings

My Real World Example with DataBrew





DAY 4

Working with Data

Predictions & Forecast

Arran Peterson (He/Him)

Solutions Architect
Amazon Web Services (ANZ)



Business impact of scaling predictive analytics



Data value lags behind data growth

90%

of today's data has
been created in the
last **2 years** alone

Statistic provided by
[Worldwide IDC Global DataSphere Forecast, 2022–2026](#)

32%

of companies reported
that they are **able** to
realize value from data

Statistic provided by
[Accenture study](#)





**Enhance
customer
experience**



**Better and faster
decision-making**



**Improve business
operations**



**New products
and services**



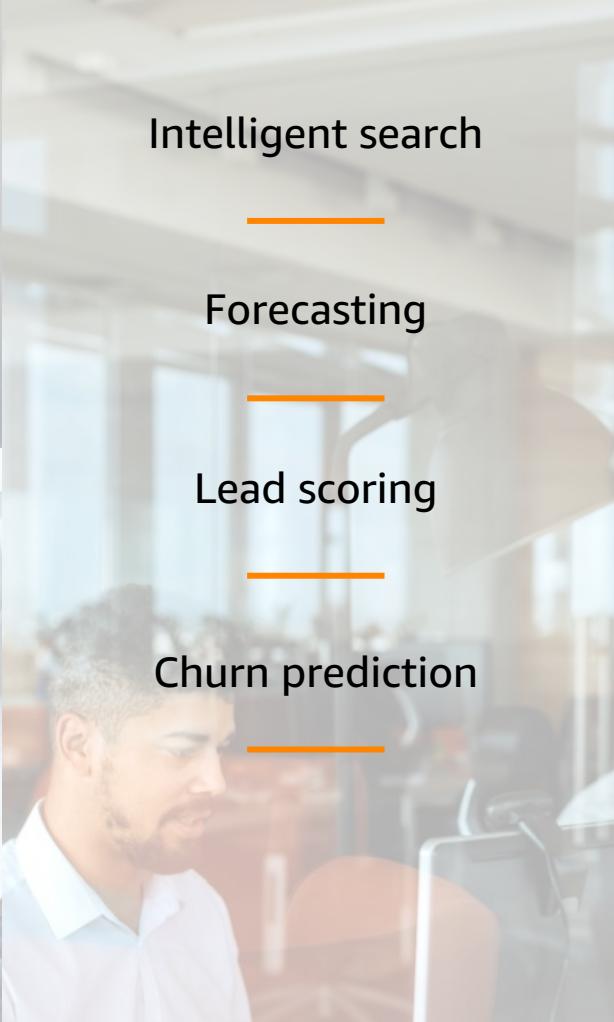
Contact center
intelligence

Chatbots and
virtual assistants

Personalization

Identity verification

Fraud prevention



Intelligent search

Forecasting

Lead scoring

Churn prediction



Intelligent document
processing

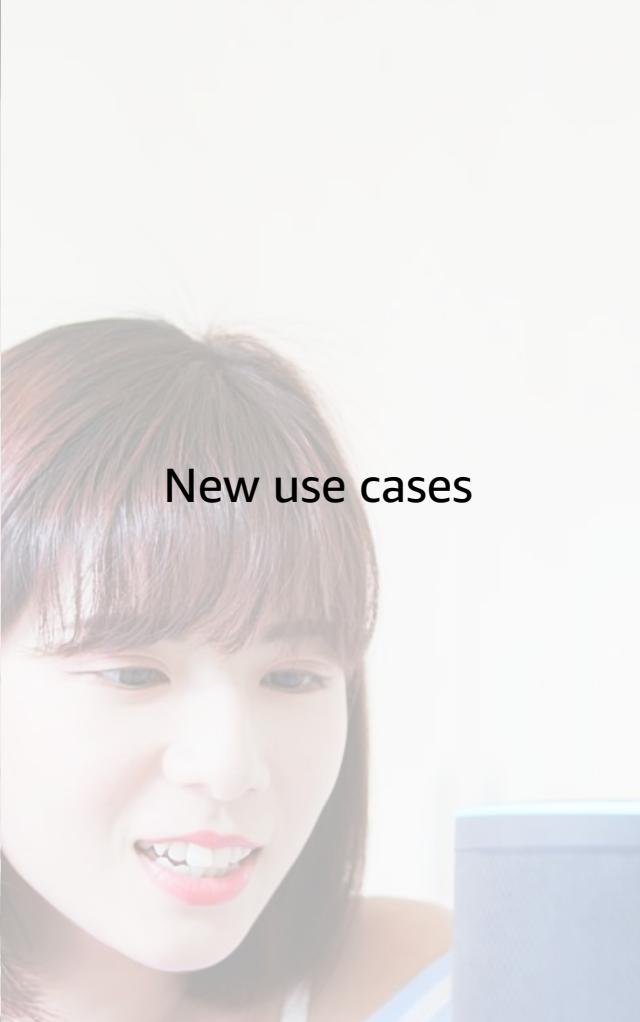
Content moderation

Predictive Maintenance

Machine translation

MLOps

Coding & DevOps



New use cases

**Enhance
customer
experience**

**Better and faster
decision-making**

**Improve business
operations**

**New products
and services**



What is Machine Learning

What is it?



Artificial intelligence (AI)

Any technique that enables computers to mimic human intelligence using logic, if-then statements, and machine learning (ML), including deep learning



Machine learning

Subset of AI that uses machines to search for patterns in data to build logic models automatically



Deep learning

Subset of ML composed of deeply multilayered neural networks that perform tasks like speech and image recognition

Use Cases

COMMON PATTERNS WITH AWS CUSTOMERS



Autonomous
driving



Predictive
maintenance



Churn
prediction



Computer
vision



Personalized
recommendations



Fraud
detection



Extract data
from documents

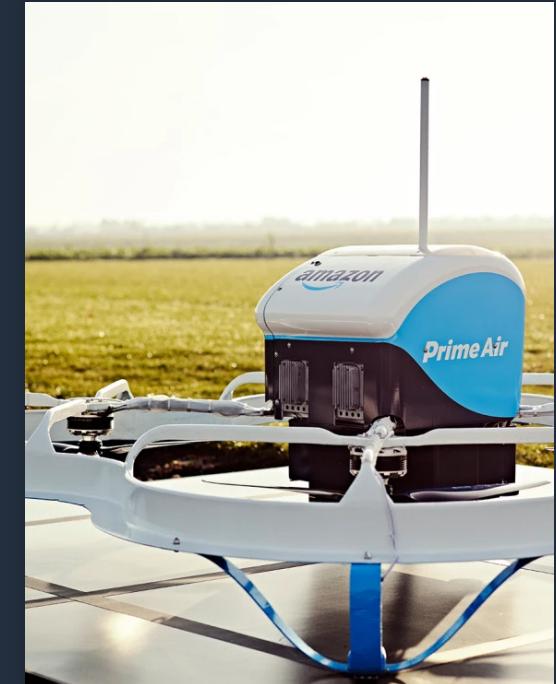
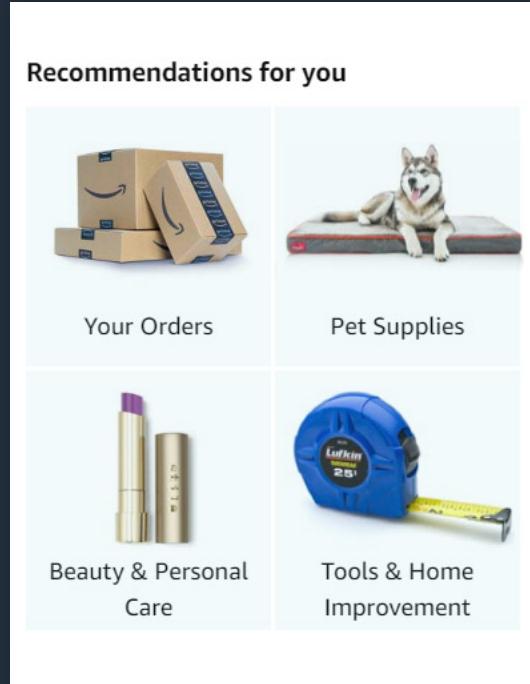


Demand
forecasting



Credit risk
prediction

Amazon machine learning innovation at scale



**4,000 products
per minute** sold
on Amazon.com

**1.6 million packages
every day**

Billions of Alexa
interactions
each week

**First Prime Air
delivery on
December 7, 2016**

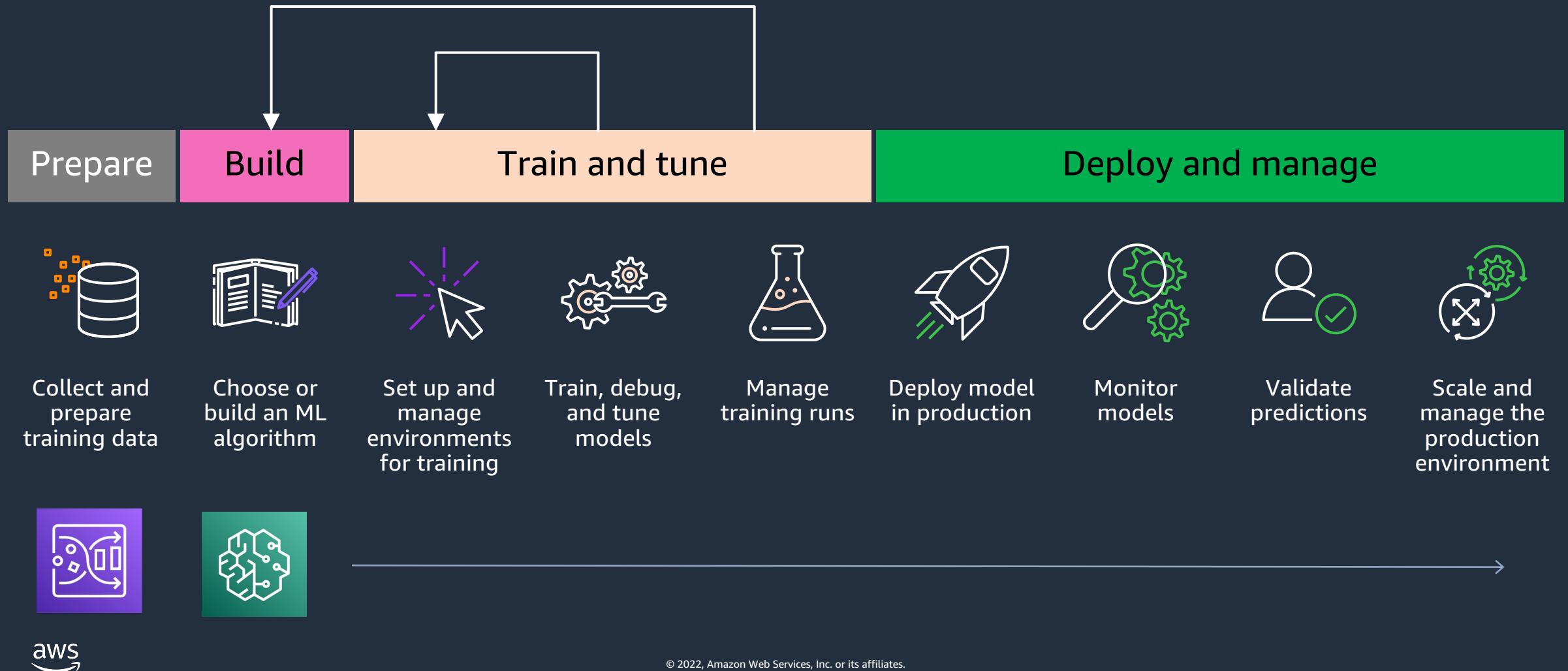


Our Use Case

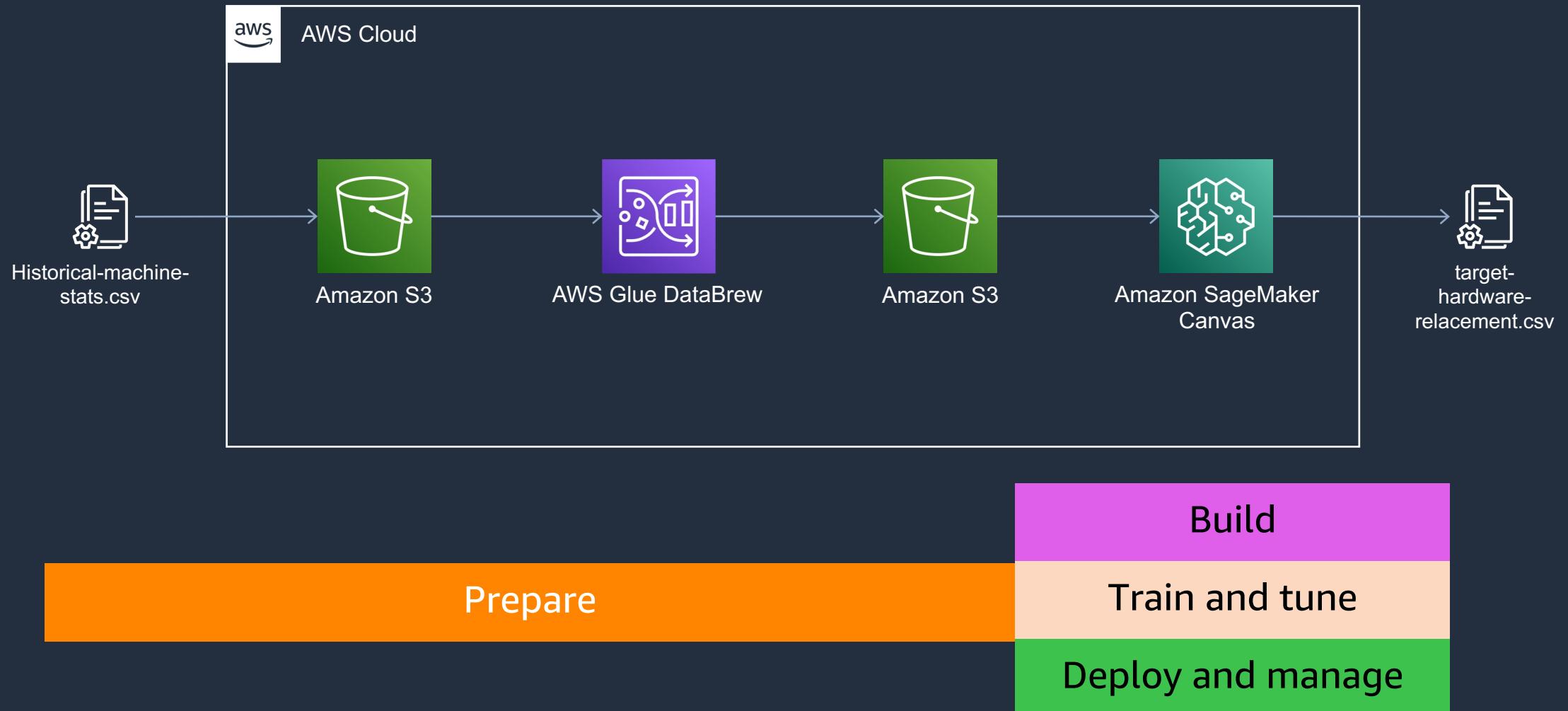
Immersion Week Architecture Pattern



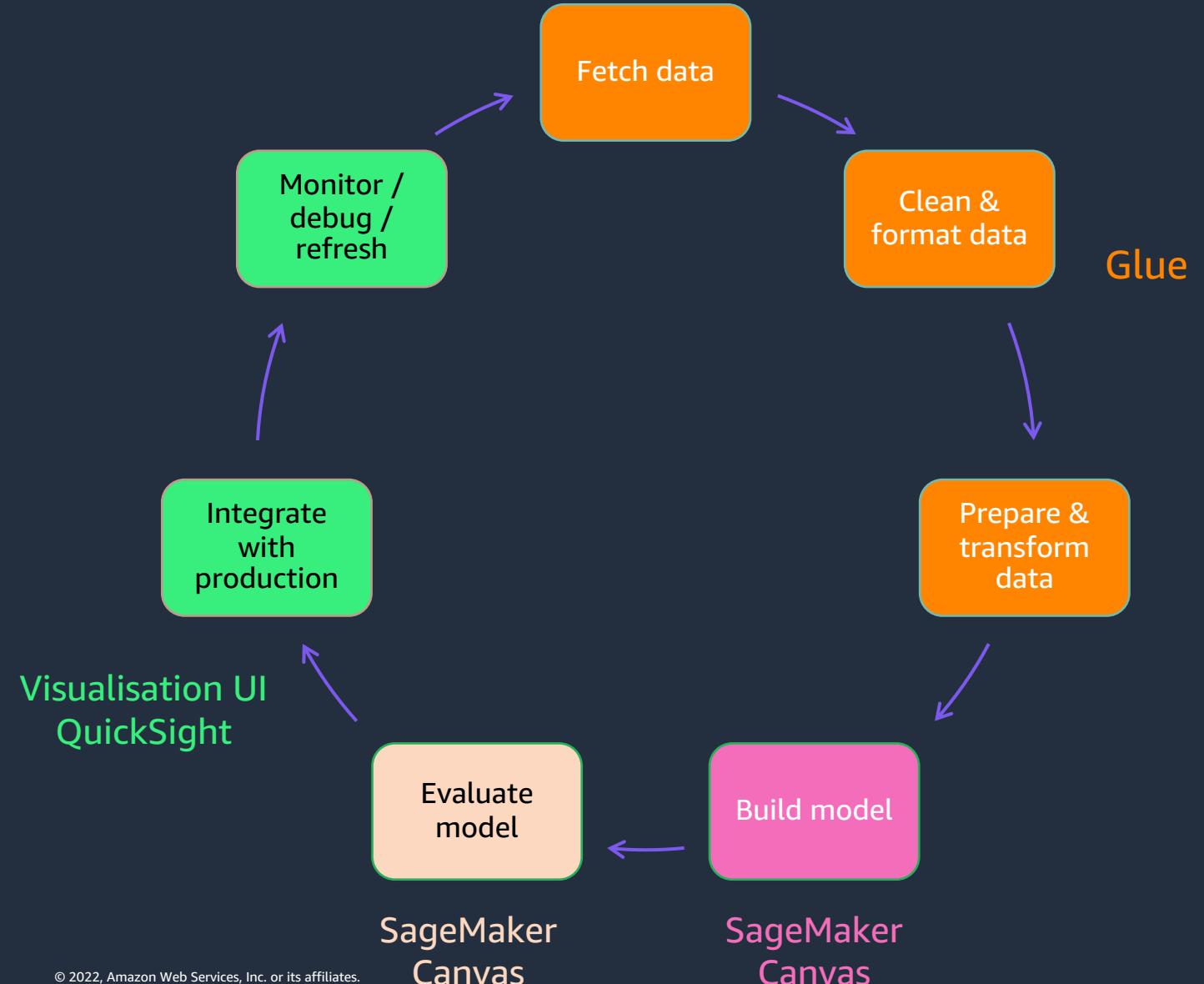
ML workflows can be complex and iterative



Immersion Week Outcome



The Machine Learning Lifecycle





Amazon SageMaker Canvas

A Visual, No Code Machine Learning
Capability

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



Import ML models from any tool within or outside Amazon SageMaker and generate predictions directly in SageMaker Canvas



Amazon SageMaker



Demo



Labs



HackMD

<https://tinyurl.com/42aws23>

Access Token: 6cc8-0e317c-f1





Thank you!

Arran Peterson

Brent Sarver

Karan Sethi

Andy Pettica

Monica Moorfield

Dylan Thompson

Justin Brien

Nicole Scott



<https://www.pulse.aws/survey/IXBEZK2I>

No/Low Code Predictive Analytics with AWS



Import the dataset from S3 to SageMaker Canvas to train the ML model

Generate batch inferences, download the data or export to S3

Import data from S3/local into SPICE, and create dashboards with QuickSight

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

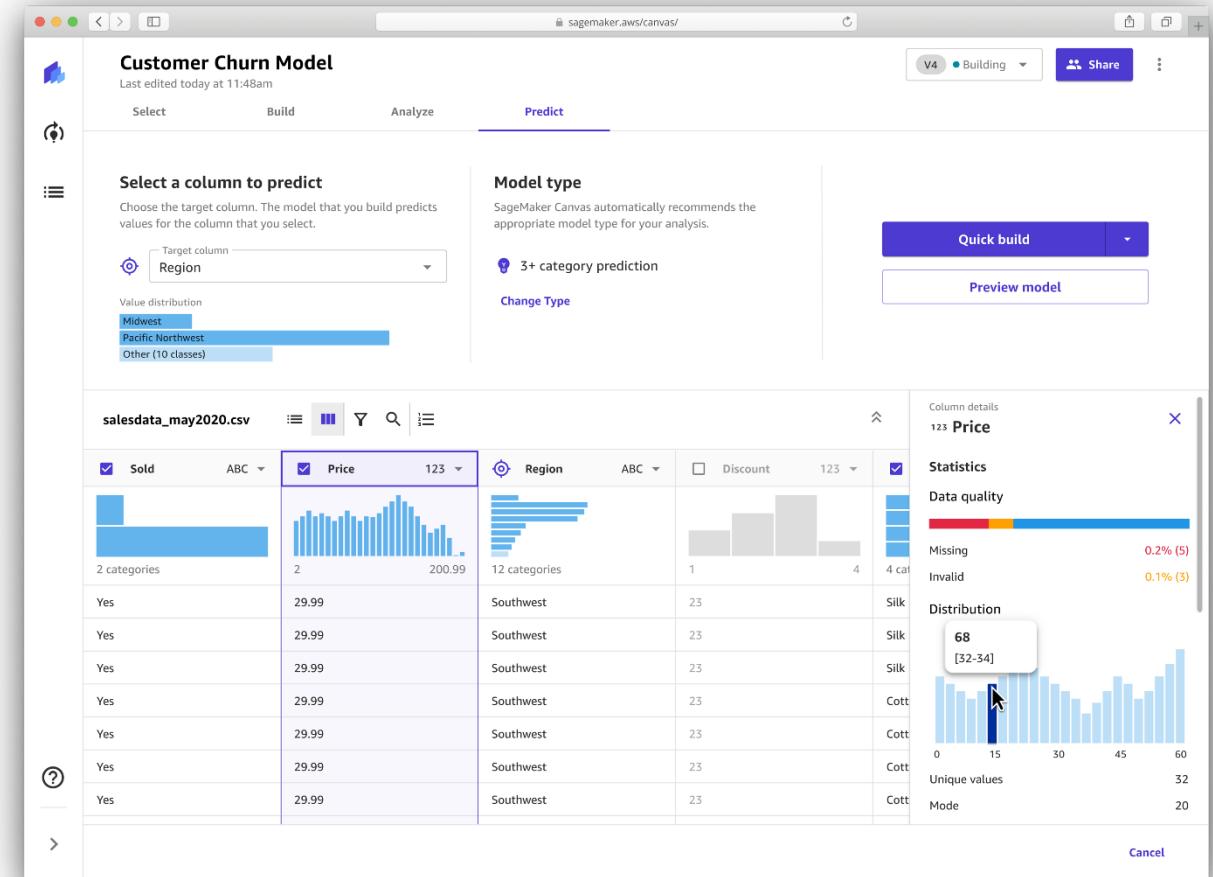
The screenshot shows the AWS SageMaker Canvas interface with the title "Import Data". At the top, there are connection options: "Upload", "S3", "Snowflake Crystal 1", and "Redshift Crystal 1". Below this is a sidebar titled "Connection name" with a search bar and a list of connections: "database1", "database2", "database3", "database4", "schema1", and "schema2". Under "schema2", there is a "table1" entry. In the main area, there is a diagram showing a green box labeled "table1.csv" connected to a white circle node, which is then connected to a pink box labeled "table2.csv". Below the diagram is an "Import preview" table with the following data:

Sold	ABC	Price	Region	ABC	Discount	Fabric	ABC	Age
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

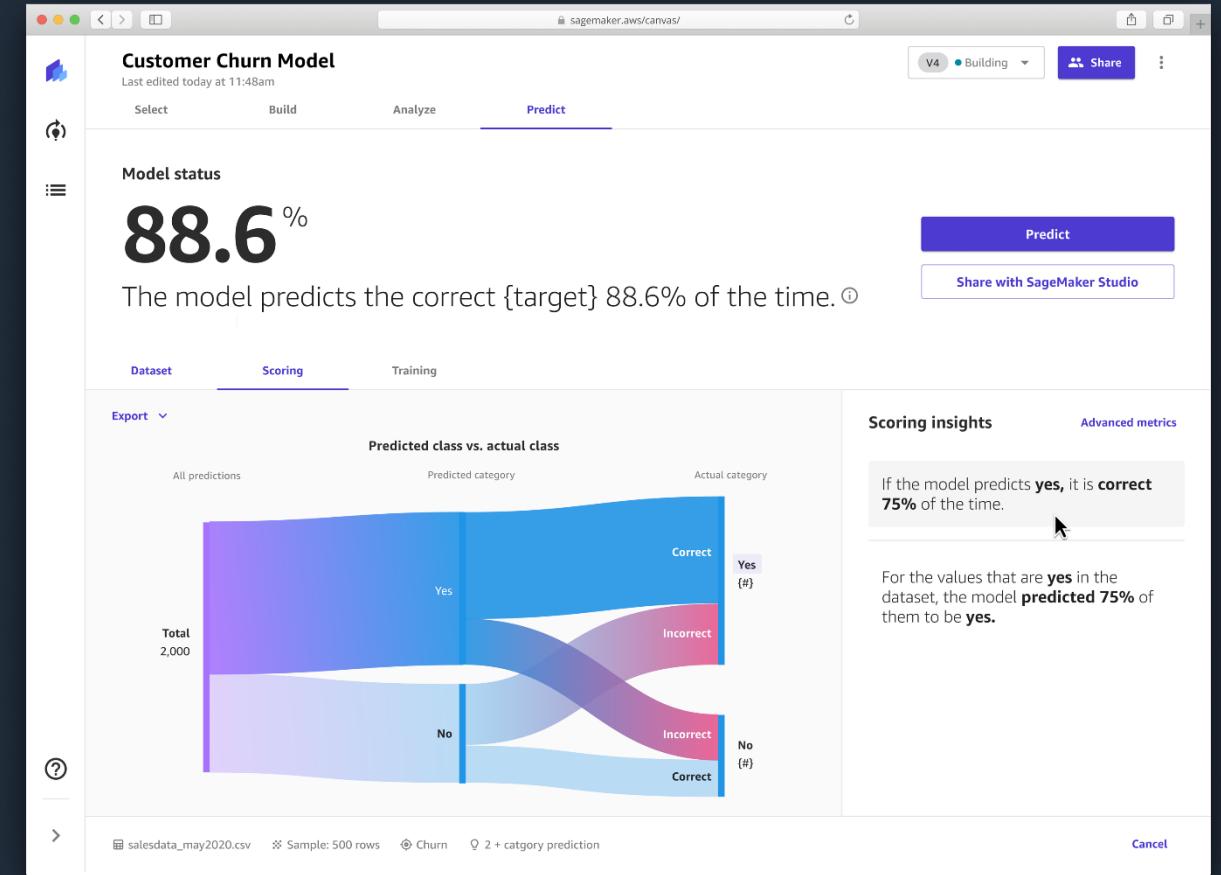
At the bottom right of the preview area are buttons for "Close" and "Import data".



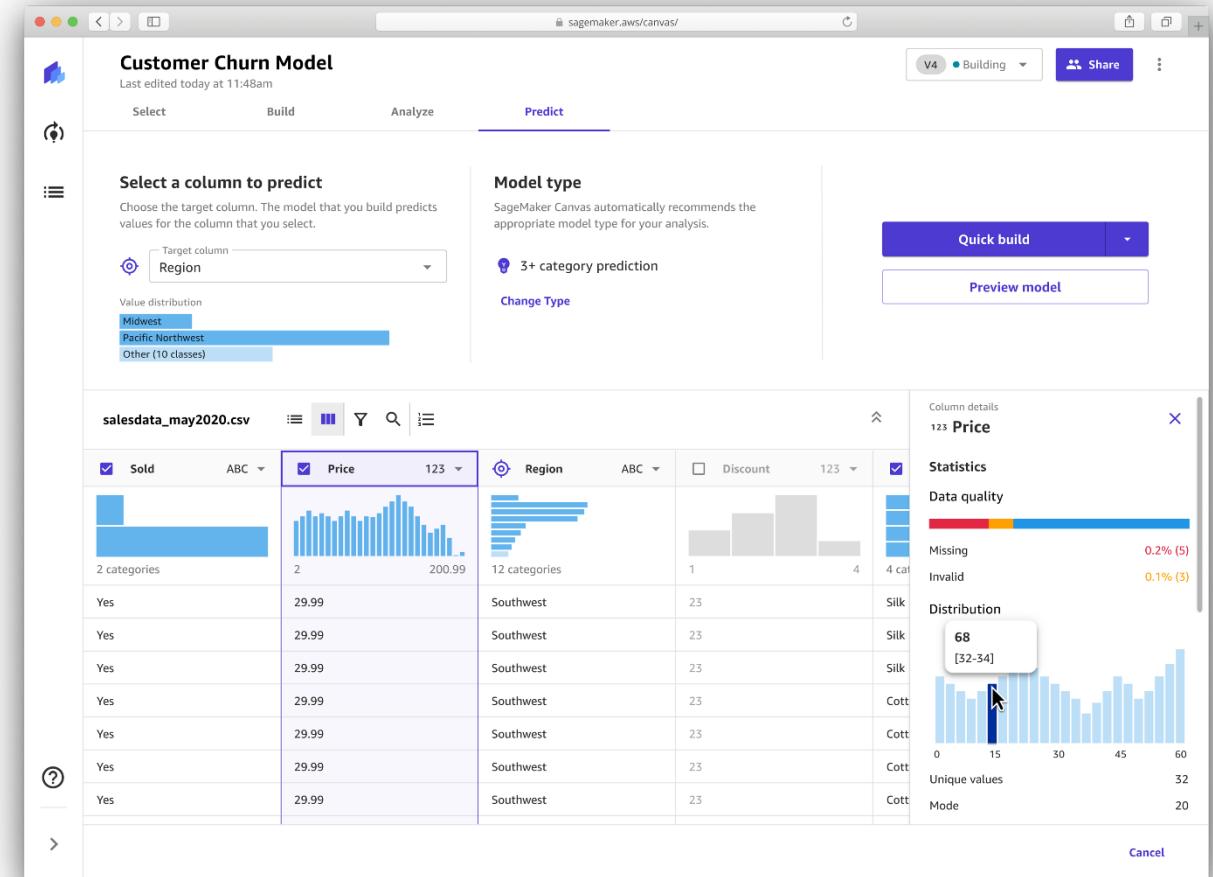
Quickly understand and prepare your data via a visual interface



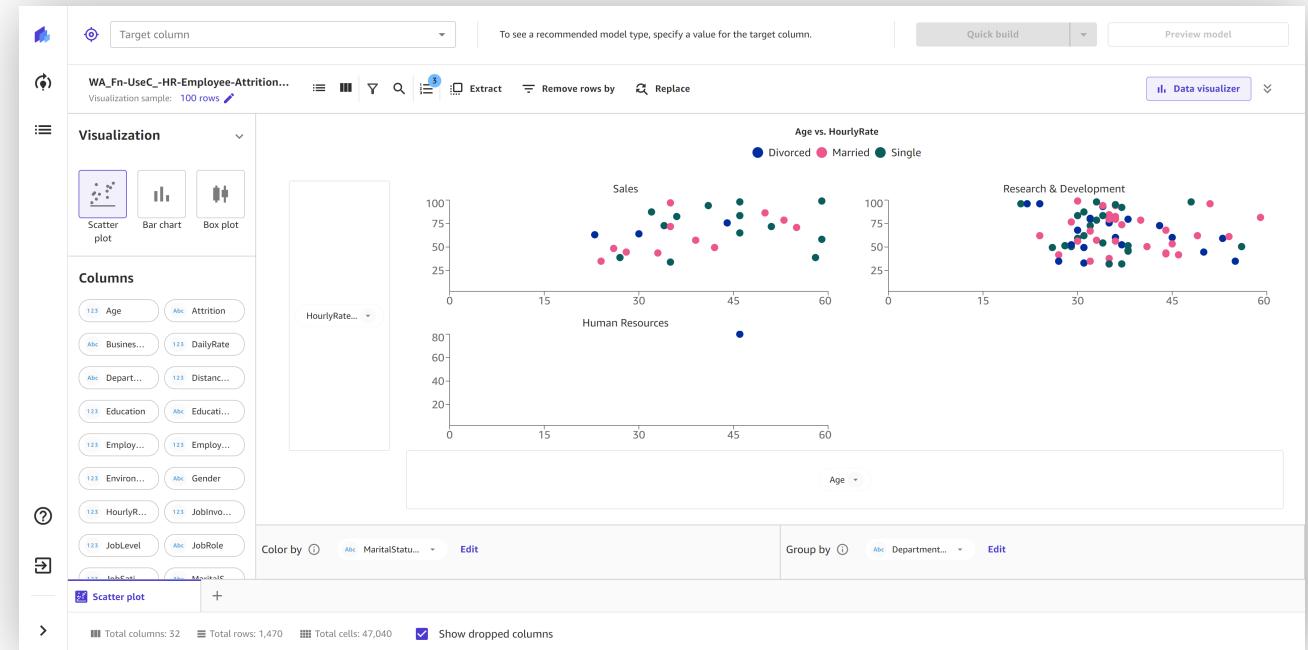
Get the first ML model in minutes. Review advanced metrics and feature importance to understand and explain predictions.



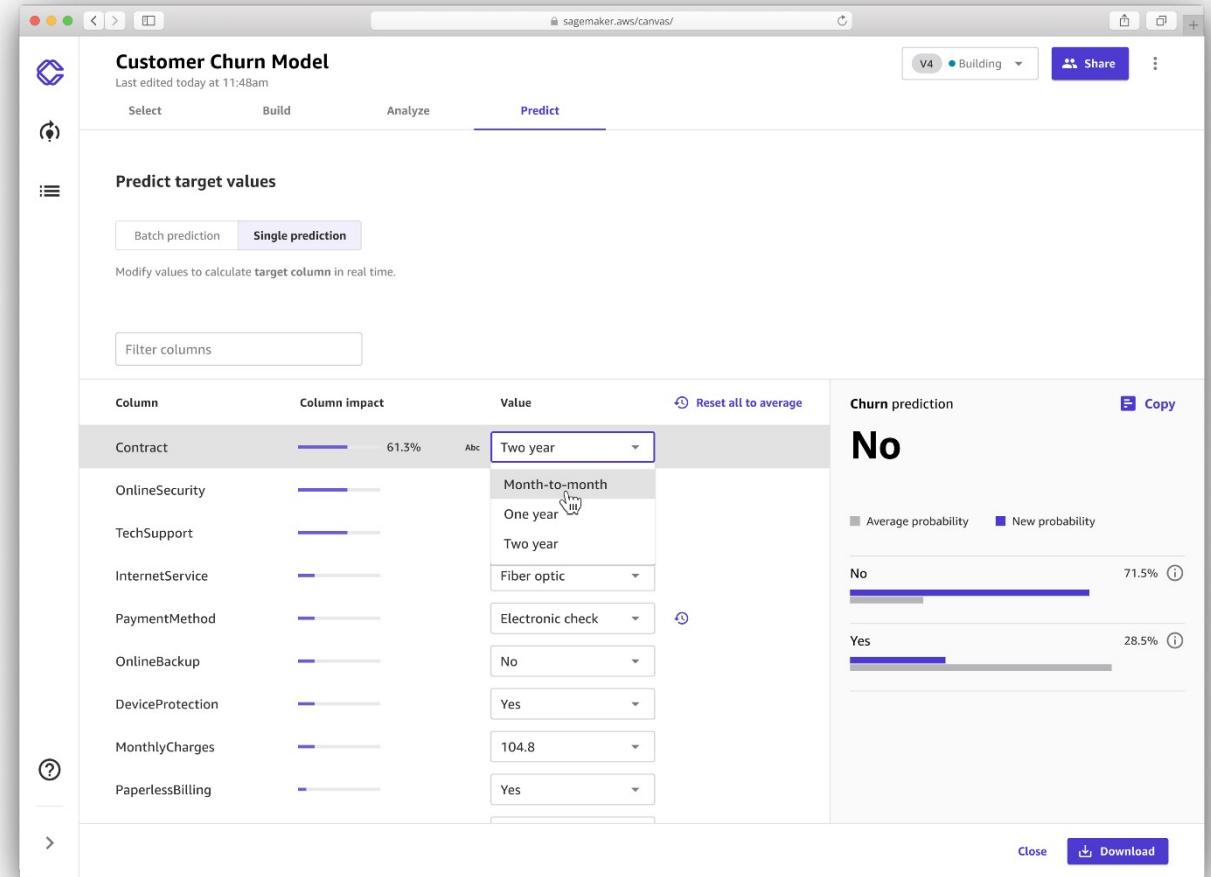
Quickly understand and prepare your data via a visual interface



Explore and visualize your data to gain insights into your data before building ML models



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



SageMaker Canvas Use Cases

VAST ARRAY OF USE CASES ACROSS DIFFERENT BUSINESS FUNCTIONS, OR VERTICALS



Sales and Marketing

1. Sales conversion
2. Sales forecasting
3. Propensity to churn
4. Customer lifetime value prediction
5. Marketing mix modeling



Finance and Accounting

1. Credit risk scoring
2. Delayed payments prediction
3. Fraud detection
4. Portfolio optimization
5. Account payables automation



Operations and Logistics

1. Demand forecasting
2. Inventory planning and scheduling
3. Delivery time forecasting
4. Predictive Maintenance

and many more...



Tips and tricks

Canvas operates on tabular data. Form your question for your data accordingly.

Var 1	Var 2	Var 3	Var 4	Var 5	Target	Var 4	Var 5	Target
11.0656	7.7798	12.9536	9.4292	3/1/2018	9.6			
8.5304	1.2543	11.3047	5.1858	3/2/2018	10.5			
5.4827	-10.3581	10.1407	7.0479	3/3/2018	9.4	292	3/1/2018	Bad
8.5374	-1.3222	12.022	6.5749	3/4/2018	9.8	858	3/2/2018	Bad
11.7058	-0.1327	14.1295	7.7506	3/5/2018	4.5	479	3/3/2018	Bad
5.9862	-2.2913	8.6058	7.0685	3/6/2018	4.3	749	3/4/2018	Bad
8.4624	-6.1065	7.3603	8.2627	3/7/2018	10.1	506	3/5/2018	Good
			5.9862	-2.2913	8.6058	7.0685	3/6/2018	Good
			8.4624	-6.1065	7.3603	8.2627	3/7/2018	Bad

“Can the target column value be explained by/predicted from the other column values?”

Manufacturing quality engineer wants to predict end-of-line quality using mid-line tests and equipment sensor readings

	Up-/mid-stream tests			Sensor readings				Digitize optical inspection data		EOL quality
N21	A	B	C	D	E	F	G	H	I	J
1	Test1	Gate1	Gate2	Reading1	Reading5	Reading6	Reading7	Xoffset	Yoffset	EOLTest
2	L	NORMAL	HIGH	80	7998.46	72.03854	0.032772	21.22587	17.6728	Fail
3	L	LOW	HIGH	77	7730.573	67.08473	0.920558	23.53545	16.012	Fail
4	L	LOW	HIGH	77	7702.598	77.68998	0.008467	22.05945	16.07217	Fail
5	L	LOW	HIGH	77	7701.867	77.61067	0.752178	22.02701	15.92174	Fail
6	L	LOW	HIGH	77	7699.826	67.69932	0.00831	23.89724	15.69795	Fail
7	S	HIGH	NORMAL	76	7603.656	136.0996	0.606646	14.11135	9.656831	Fail
8	L	LOW	HIGH	76	7600.318	78.37614	0.008194	20.48083	15.68505	Fail
9	L	LOW	HIGH	76	7600.298	78.83789	0.352657	20.59402	15.47902	Fail
10	S	HIGH	NORMAL	76	7599.836	140.6925	0.598734	13.60846	10.16249	Fail
11	S	HIGH	NORMAL	76	7596.798	140.0011	0.537771	15.18518	9.044426	Fail
12	S	LOW	NORMAL	76	7497.453	152.9493	-0.02966	11.69514	10.23416	Pass



Thank you!



Thank you!