Use no-code ML to solve real-world challenges

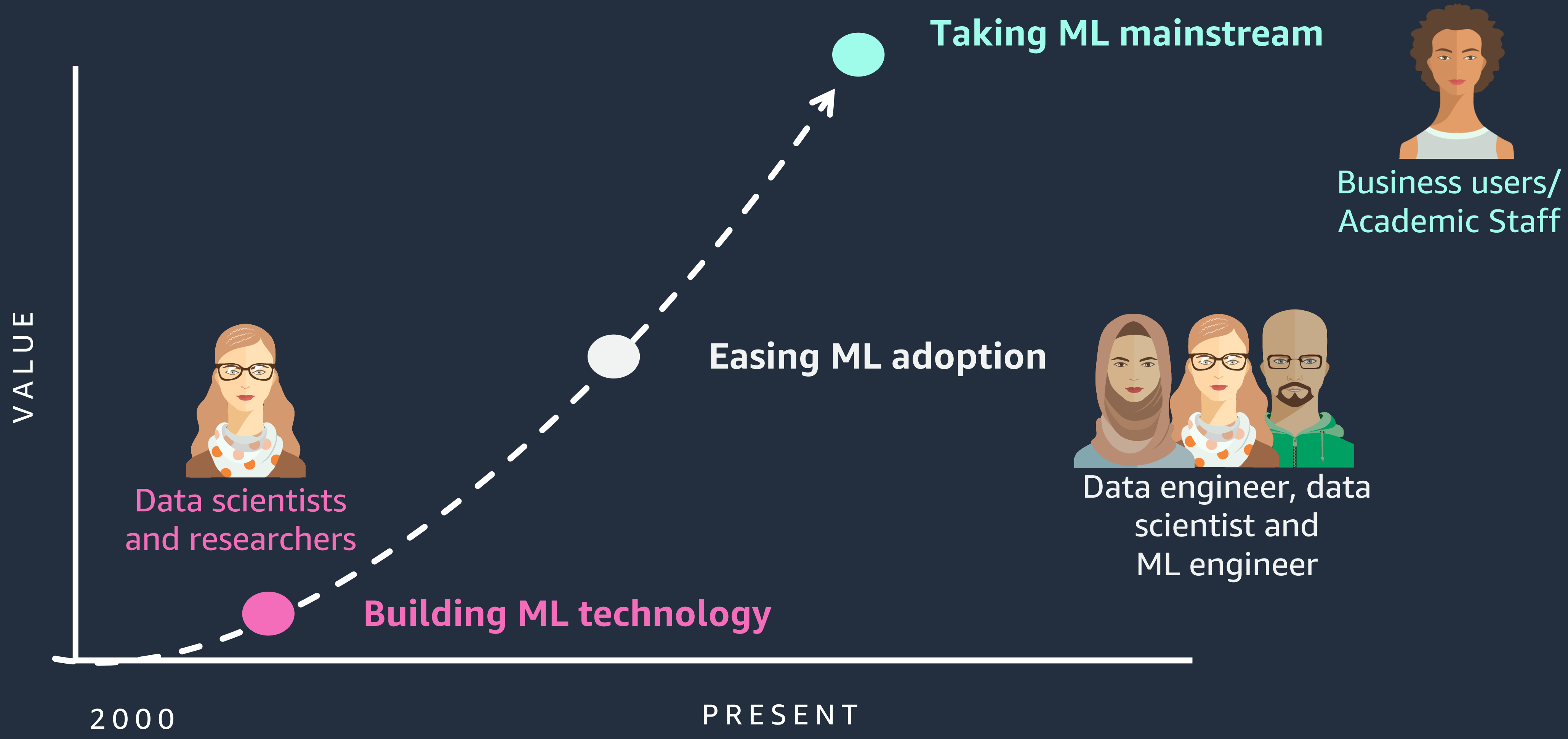
Rifkhan Anoor

Solutions Architect
Amazon Web Services
rifkhata@amazon.com

Agenda

- Why no-code ML?
- Using ready-to-use models
- Preparing and analyzing data
- Creating custom models
- Using generative AI
- Demo

Why no-code machine learning?

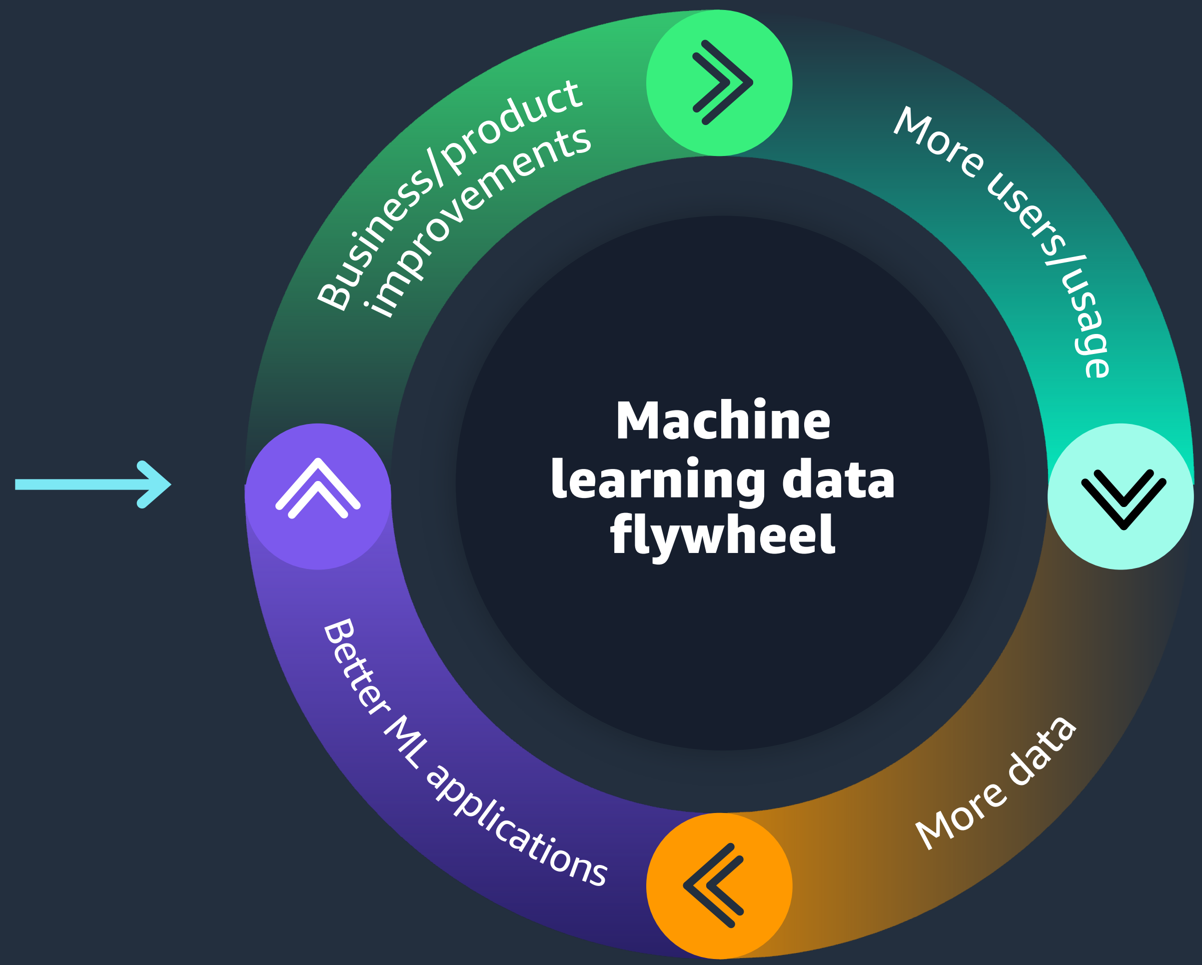


Building practical ML applications is hard

The bottleneck

BUILDING BETTER ML APPLICATIONS CAN BE THE BOTTLENECK BECAUSE:

- Line of business teams depend on data science teams to make machine learning (ML) powered decisions
- Data science teams are oversubscribed, while ML needs are only increasing
- High learning curve for technical users to learn how to code
- Friction points in the machine learning journey and prioritizing use cases



Business requirements → **Data preparation and feature engineering** → **Model development, training, and tuning** → **Model deployment inference and monitoring**

Business Teams

Amazon SageMaker Canvas

A **dedicated no-code workspace for business analysts** to generate ML-powered predictions

Canvas

A visual point-and-click interface that allows analysts to generate accurate ML predictions on their own — without requiring any machine learning experience or having to write a single line of code.

2-Way Collaboration

\+

Data Science Teams

Amazon SageMaker Studio

A **dedicated workspace for data engineers, data scientists and ML Ops teams** to collaborate and bring ML to market faster

Data Wrangler

A faster, visual way to aggregate and prepare data for machine learning

Autopilot

AutoML capability that automatically prepares your data, as well as builds, trains, and tunes the best machine learning models for your tabular and time-series datasets

JumpStart

Pre-built solutions and a model zoo of pre-trained and easily tunable state-of-the-art models for computer vision, and natural language processing

Many deployment options

Amazon QuickSight

Data Stores – 40+ Connectors

Work with BI data, enhance dashboards with predictions, export models and artifacts

Amazon SageMaker Canvas

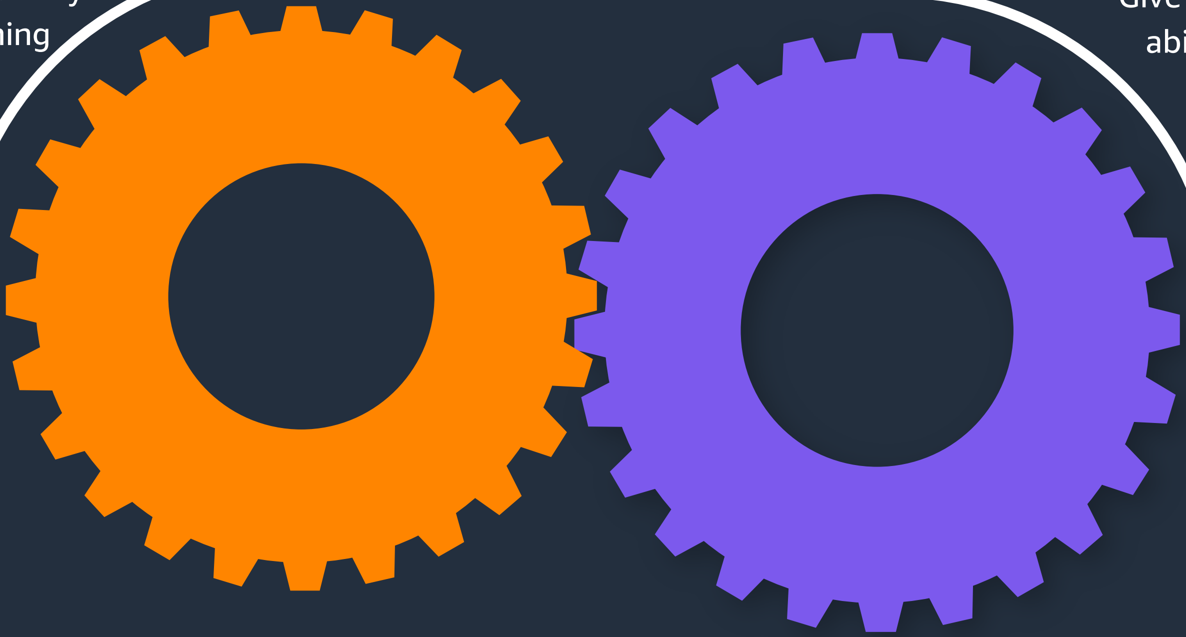
Build ML models and generate accurate predictions — no code required

- **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**

Ways SageMaker low-code/no-code helps

Accelerate data science teams

Do more with your current team by using low-code machine learning tools in order to get to the desired outcomes faster.



Enable business users

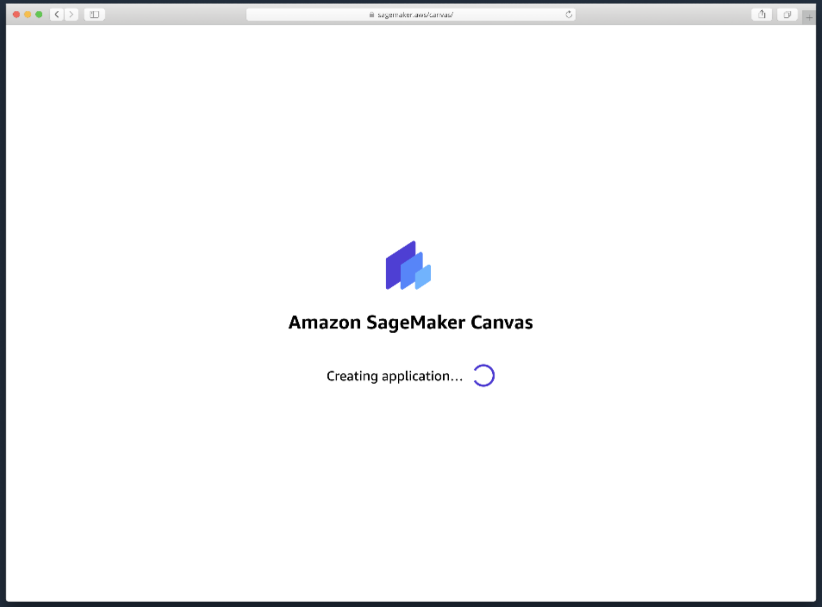
Give business users and analysts the ability to do ML without any code, scaling the number of people who can create ML powered insights, forecasts, and predictions

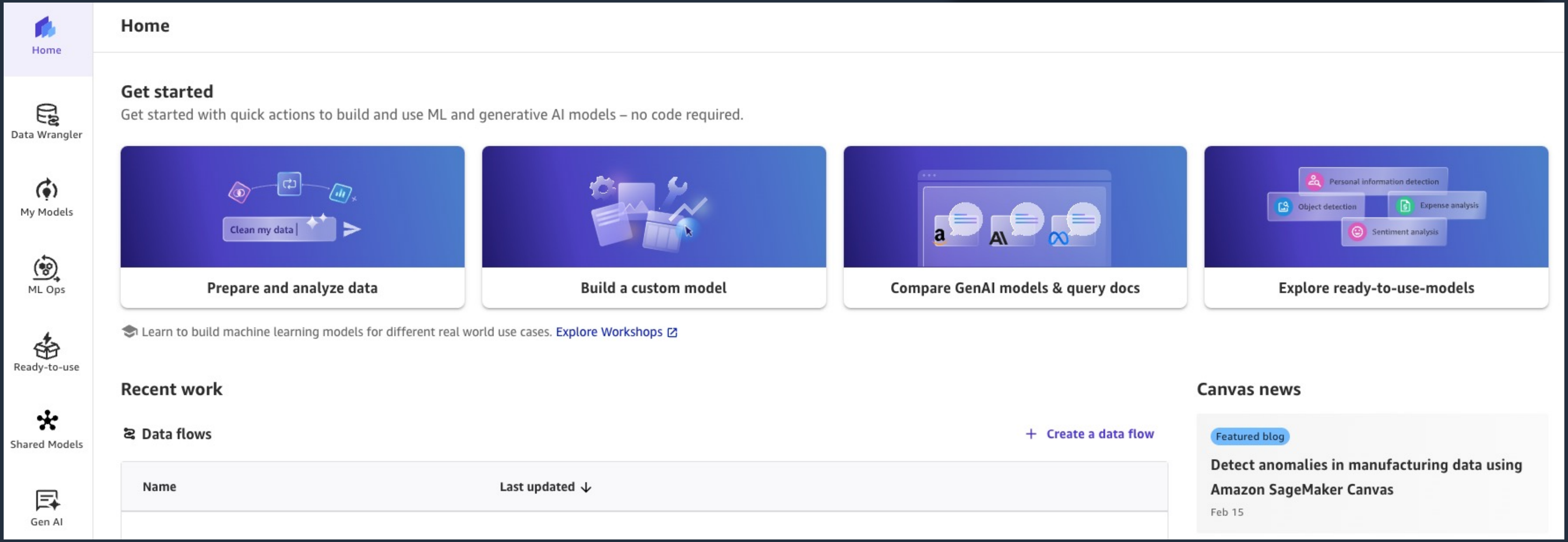
Collaborate together

SageMaker LCNC ML has several points of collaboration making it efficent for Business users to use data scientist models or for data scientists to make changes on the models analysts build and creating one place for all the analytics and machine learning in a team or organization

Amazon SageMaker Canvas

Self-service access to a business-friendly tool for machine learning





Ready-to-use models

Home

Data Wrangler

My Models

ML Ops

Ready-to-use

Shared Models

Gen AI

Help

Log out

Ready-to-use models

Document queries
Extract information from structured documents such as paystubs, bank statements, W-2s, and mortgage application forms by asking questions using natural language.

Powered by Amazon Textract

Sentiment analysis
Detect sentiment in lines of text, which can be positive, negative, neutral, or mixed.

Powered by Amazon Comprehend

Entities extraction
Extract entities, which are real-world objects such as people, places, and commercial items, or units such as dates and quantities, from text.

Powered by Amazon Comprehend

Language detection
Determine the dominant language in text such as English, French or German.

Powered by Amazon Comprehend

Personal information detection
Detect personal information that could be used to identify an individual, such as addresses, bank account numbers, and phone numbers, from text.

Powered by Amazon Comprehend

Object detection in images
Detect objects, concepts, scenes, and actions in your images.

Powered by Amazon Rekognition

Text detection in images
Detect text in your images.

Powered by Amazon Rekognition

Expense analysis
Extract information from invoices and receipts, such as date, number, item prices, total amount, and payment terms.

Powered by Amazon Textract

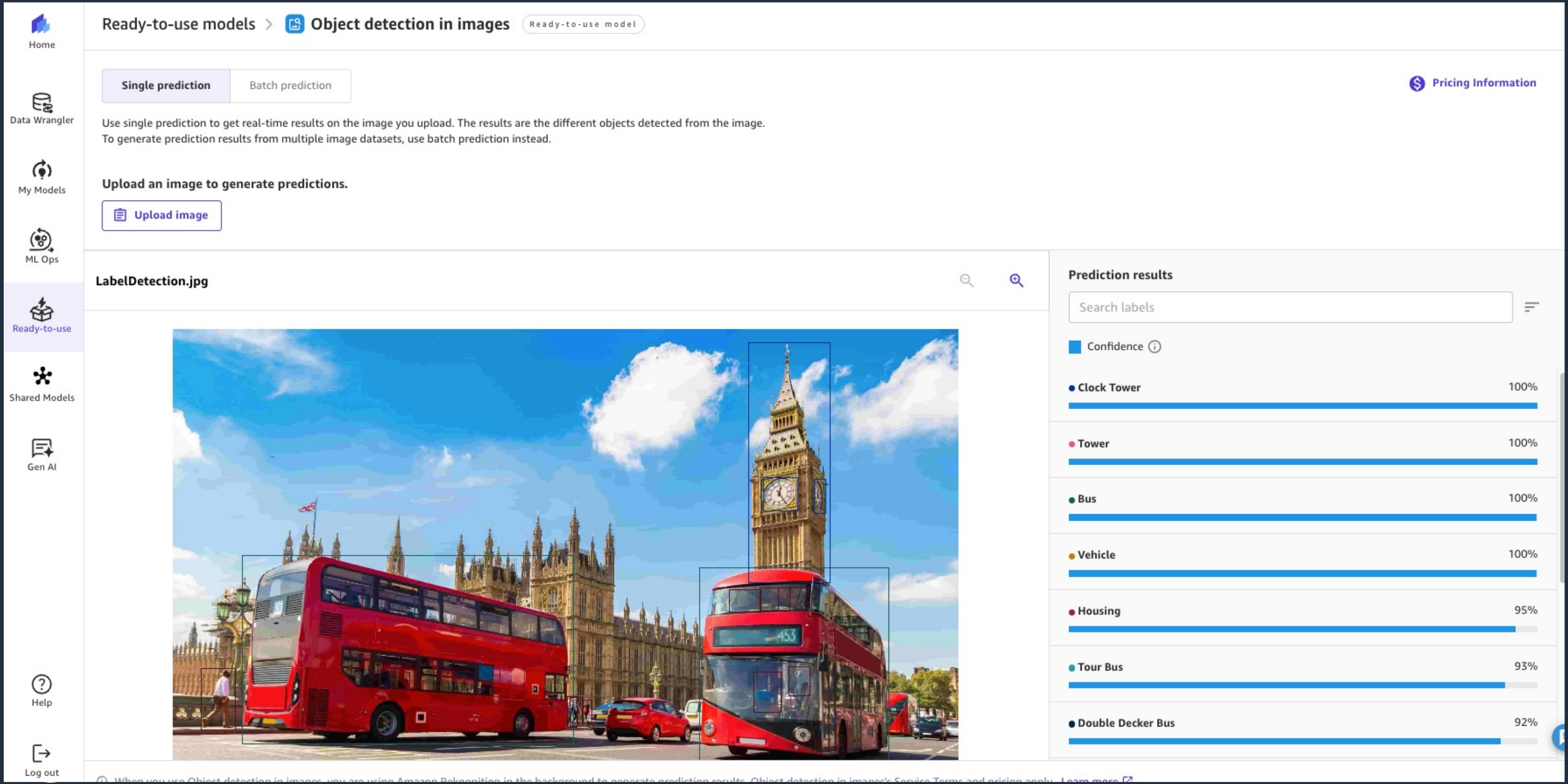
Identity document analysis
Extract information from passports, driver licenses, and other identity documentation issued by the US Government.

Powered by Amazon Textract

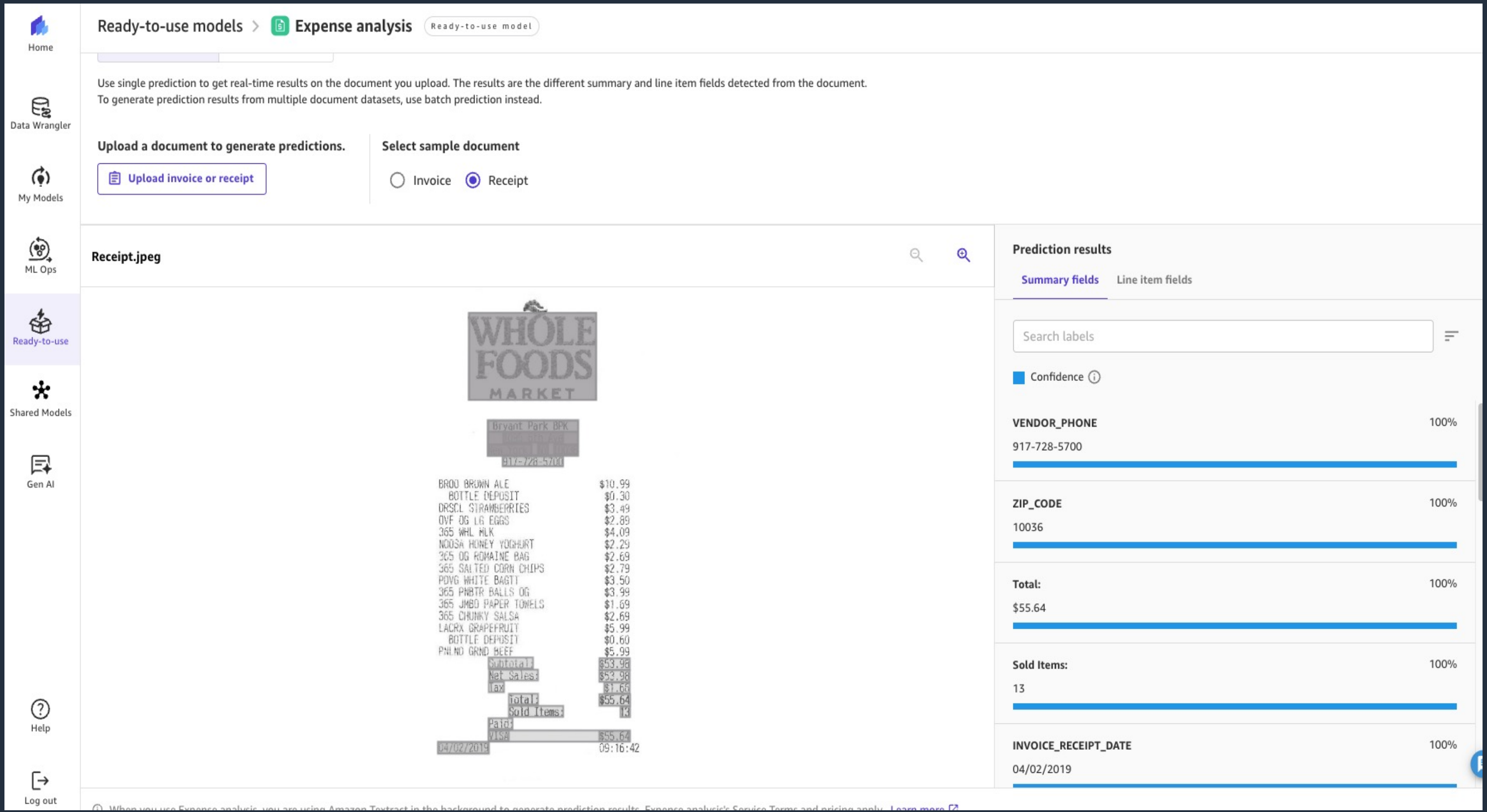
Document analysis
Analyze documents and forms for relationships among detected text.

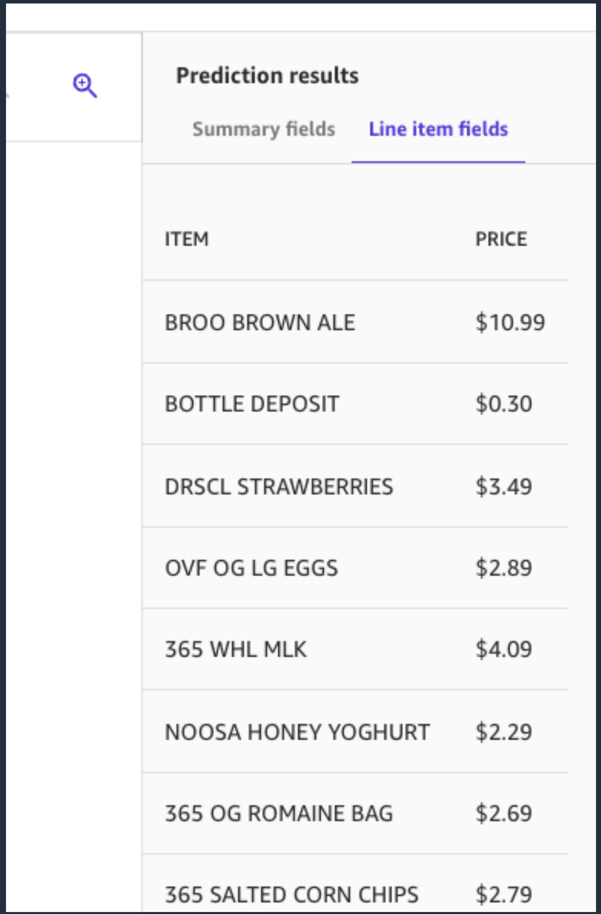
Powered by Amazon Textract

Object detection of images

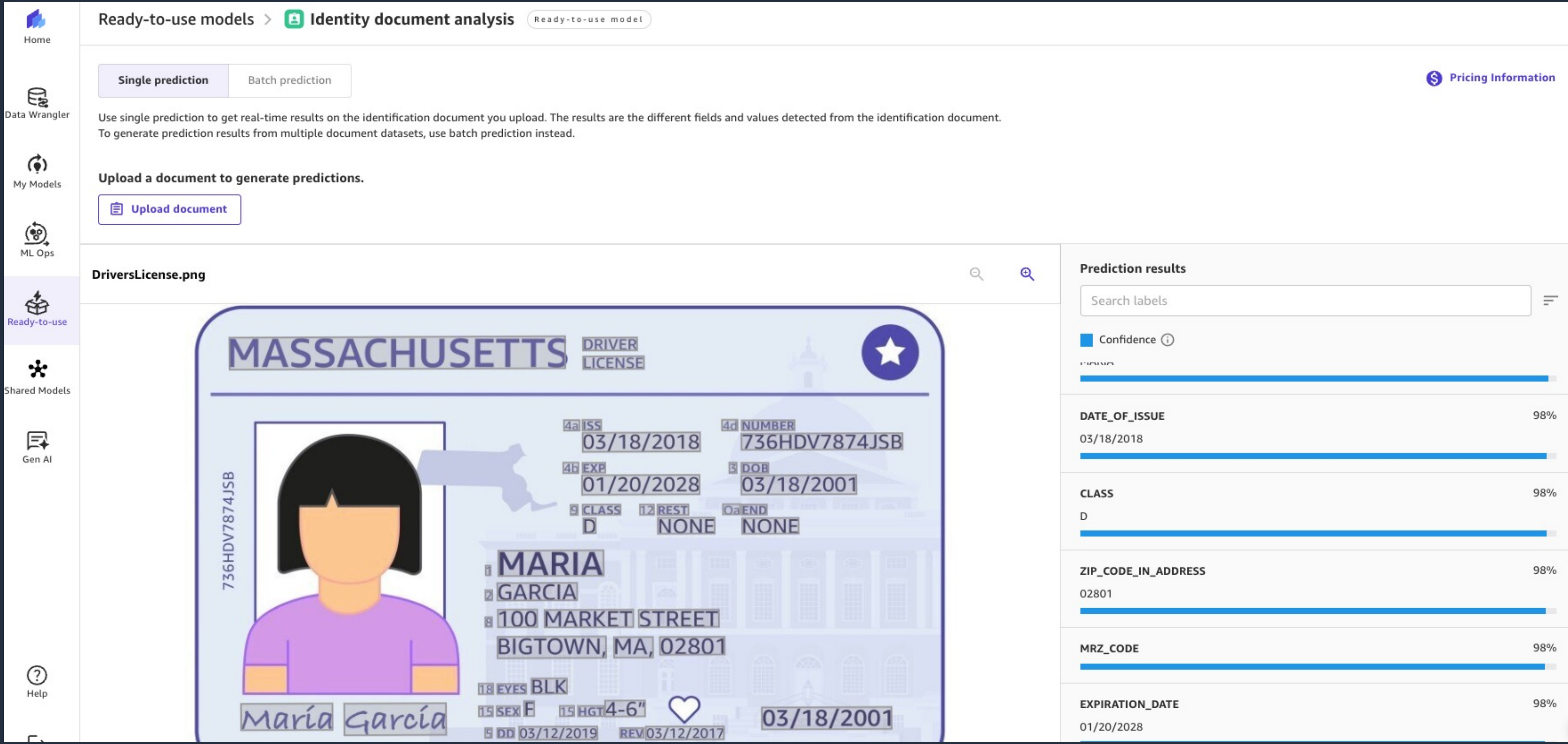


Expense analysis

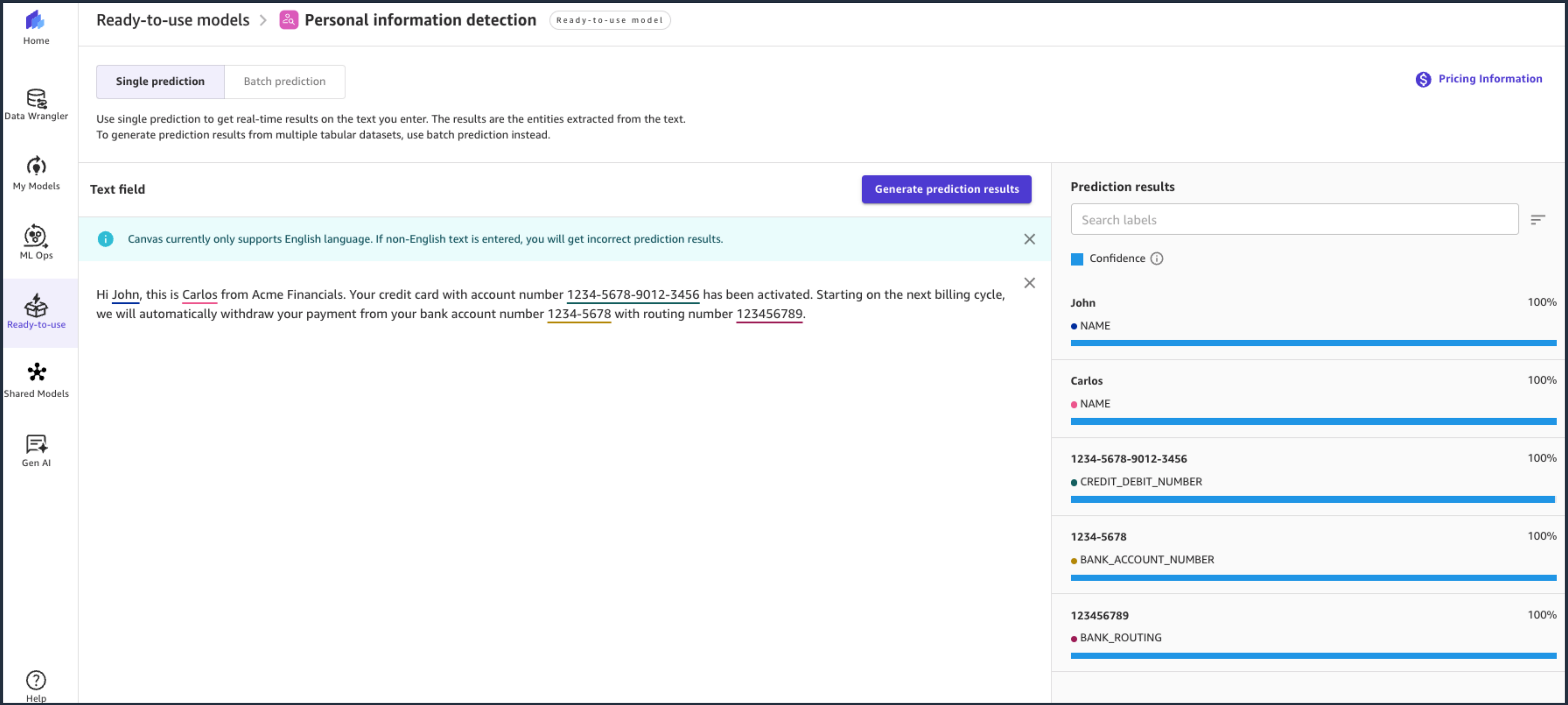




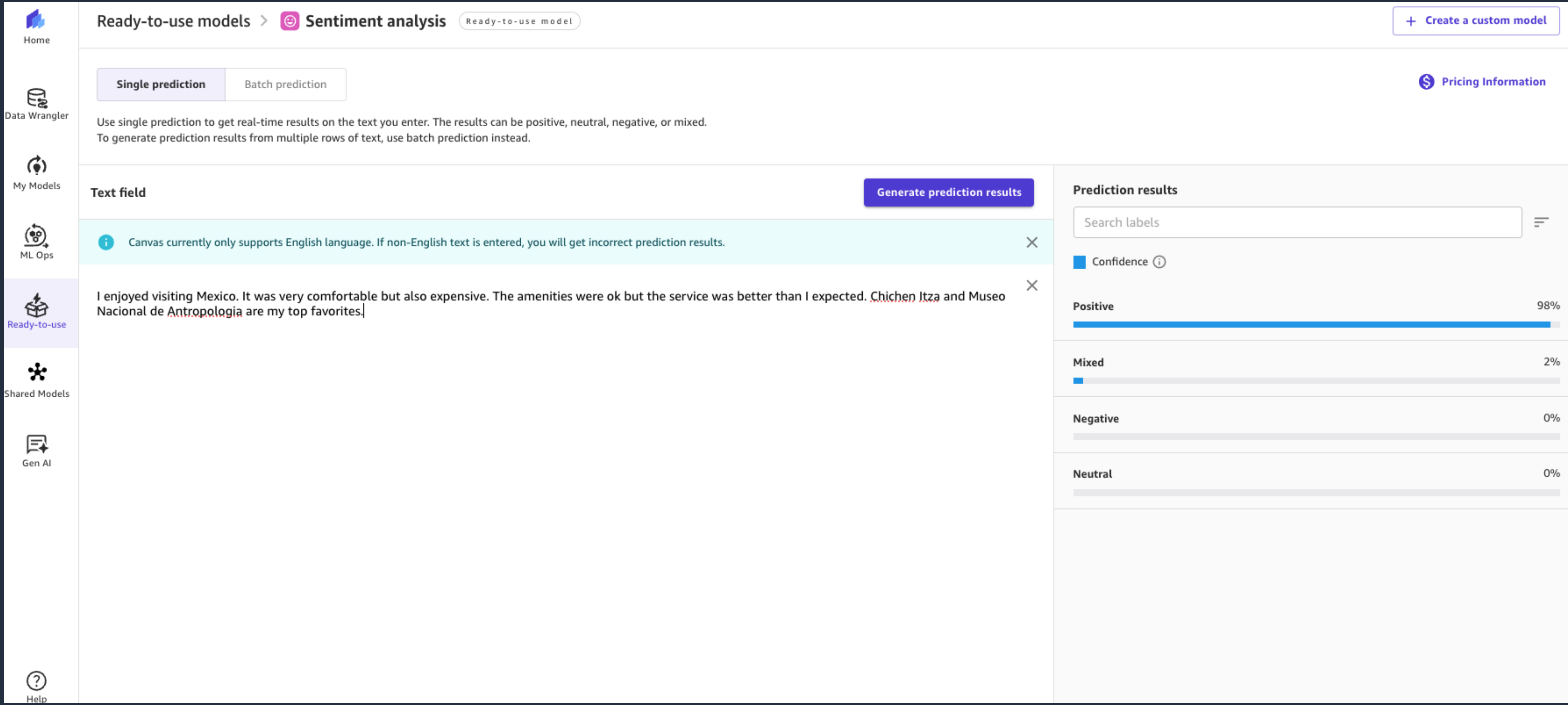
Identity document analysis



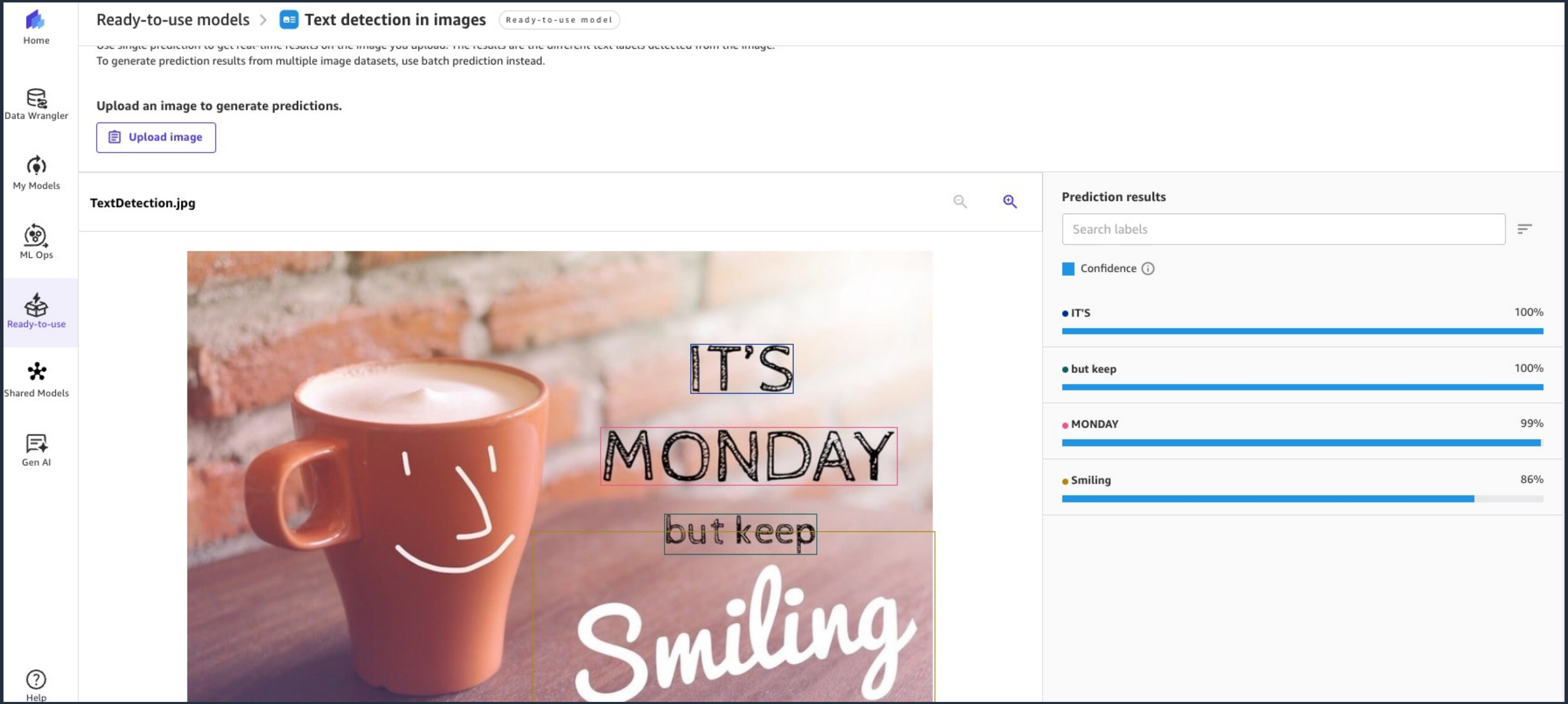
Personal information detection



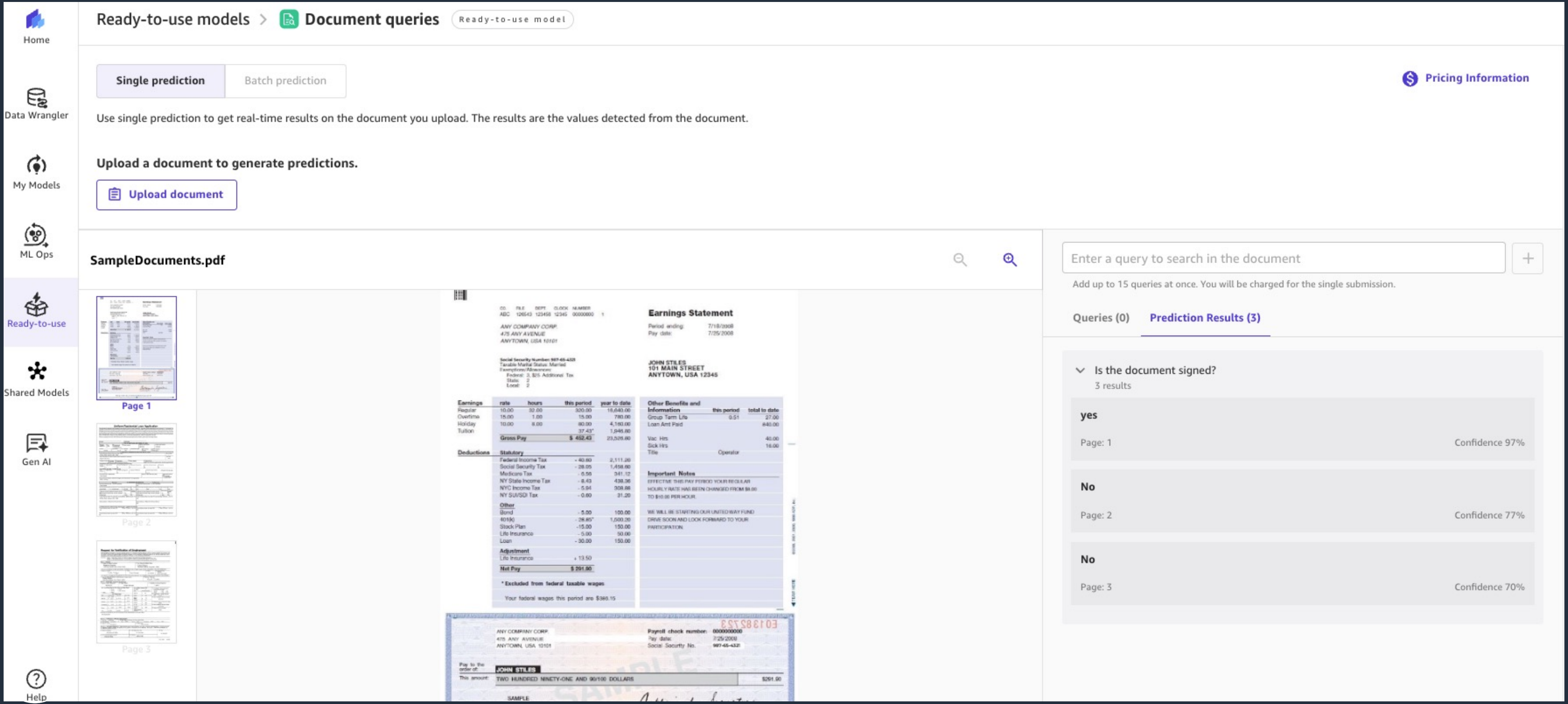
Sentiment analysis



Text detection in images

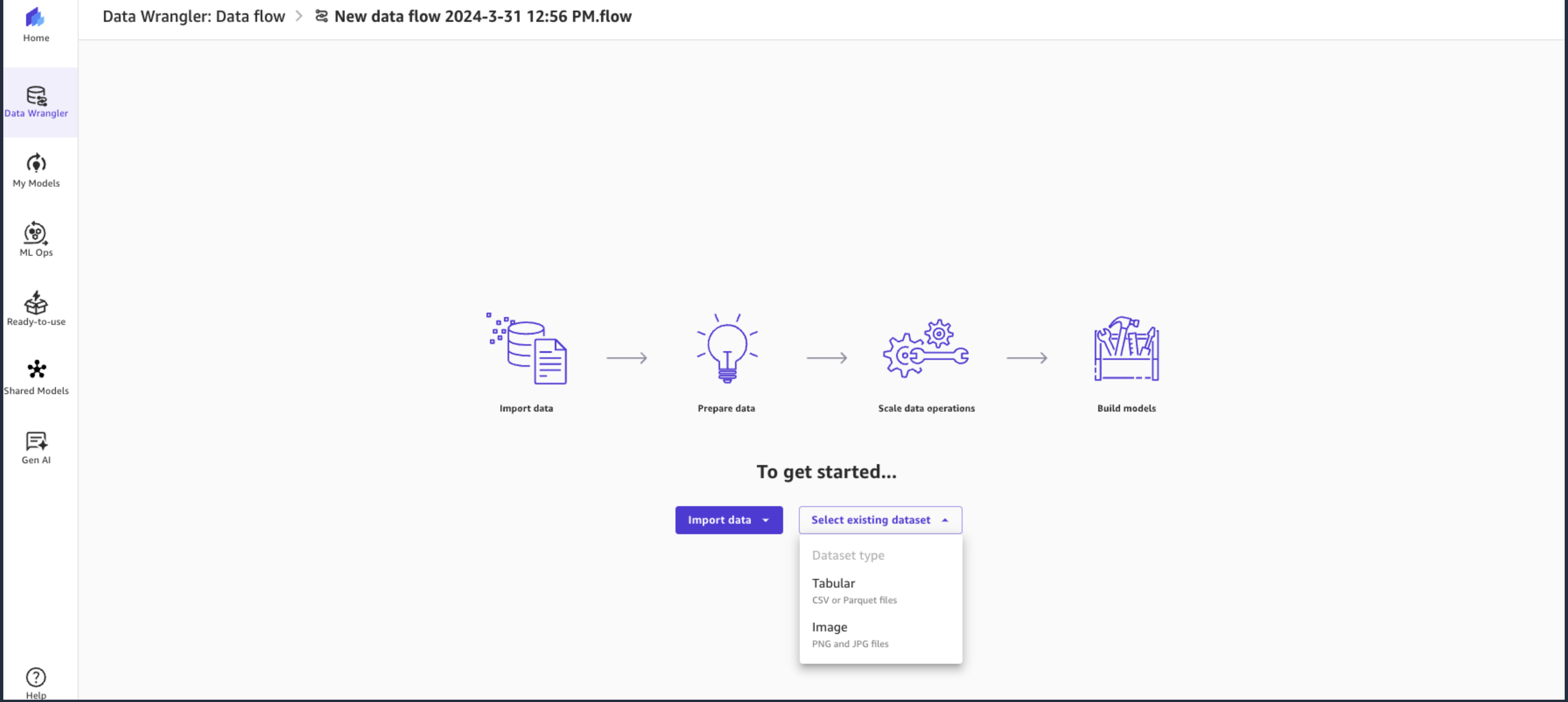


Document analysis

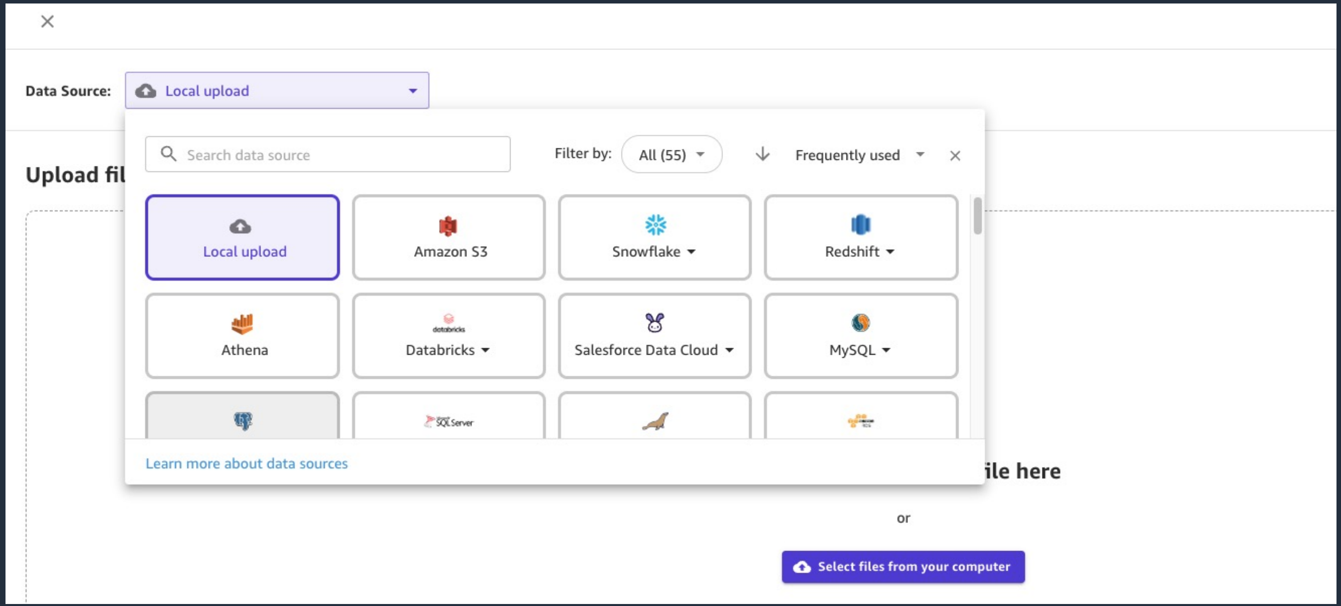


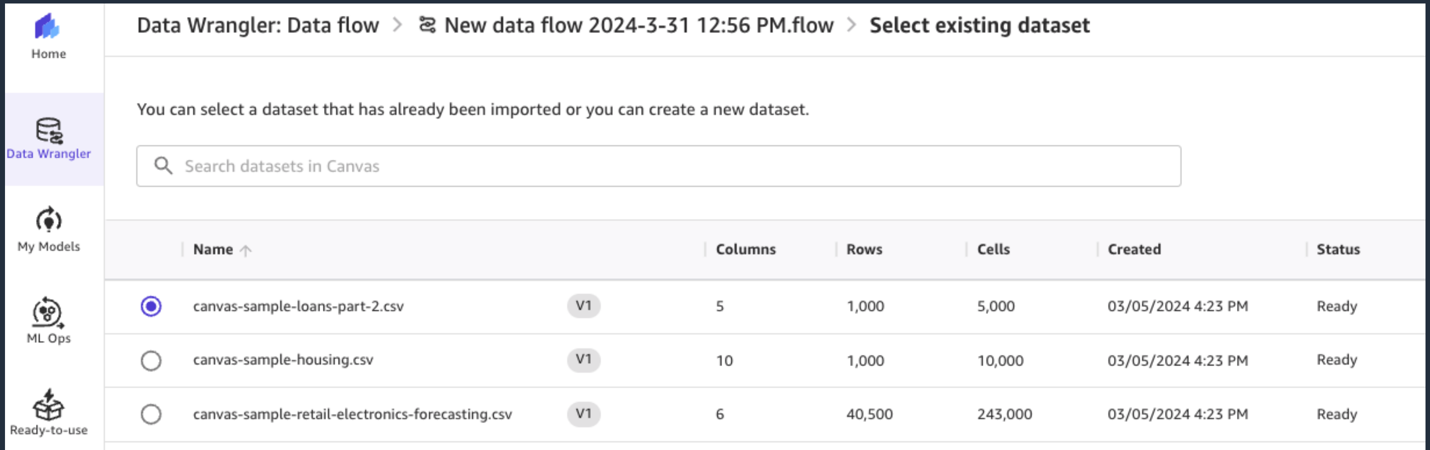
Data preparation

Creating data flow

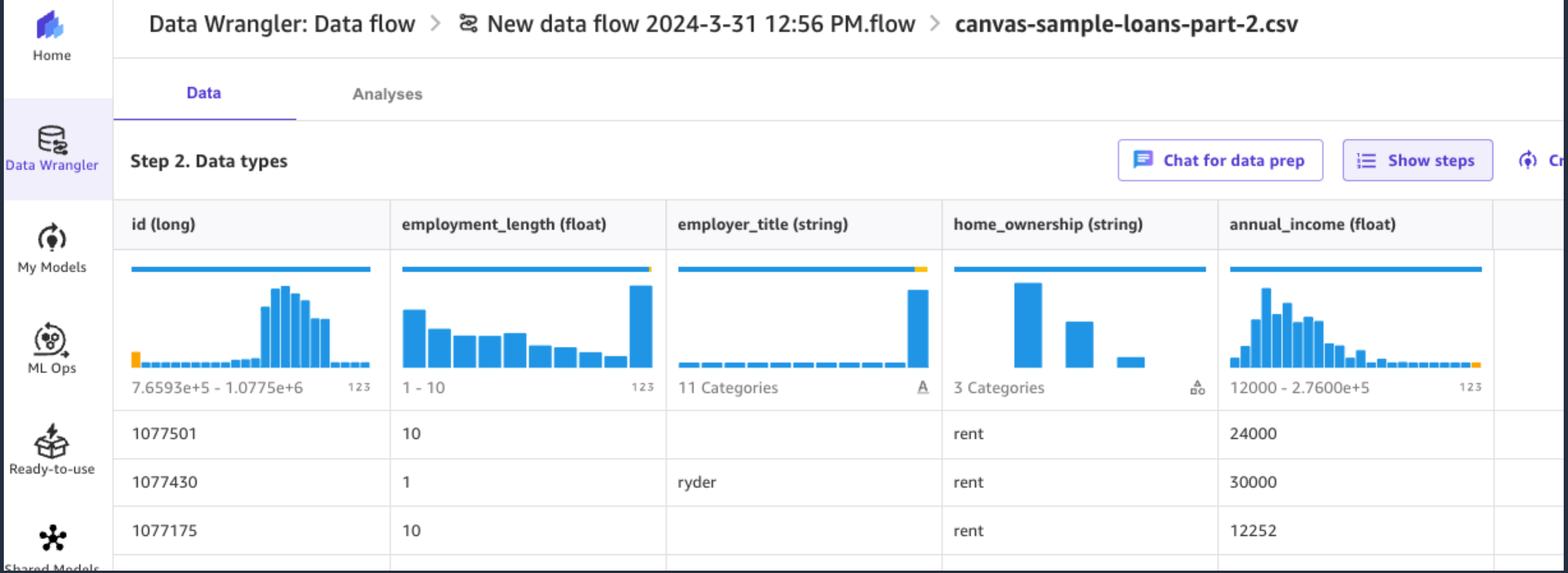


Selecting data sets

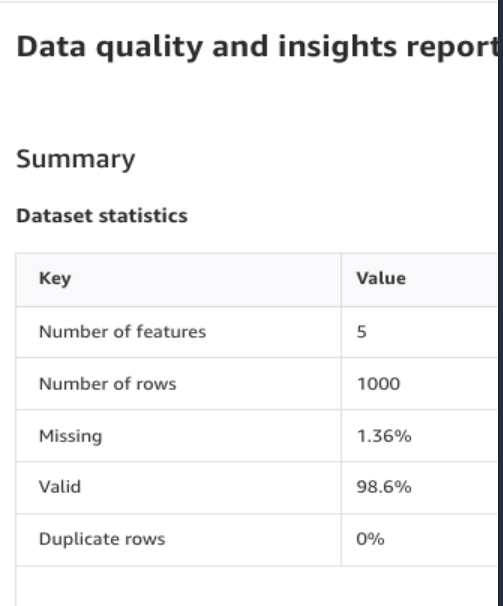




Amazon Sagemaker Data Wrangler



id (long)	employment_length (float)	employer_title (string)	home_ownership (string)	annual_income (float)
1077501	10		rent	24000
1077430	1	ryder	rent	30000
1077175	10		rent	12252

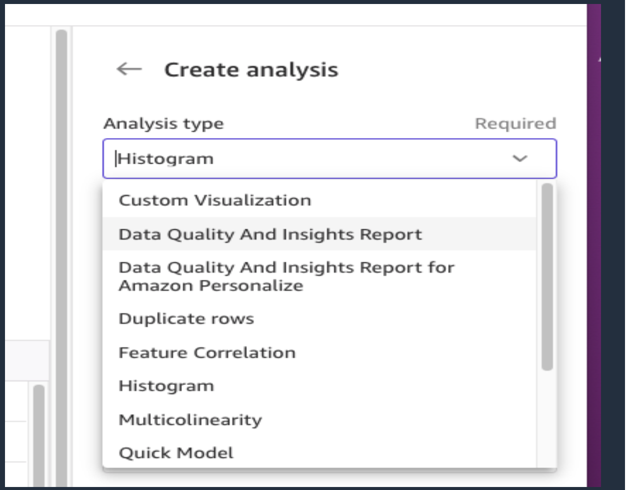


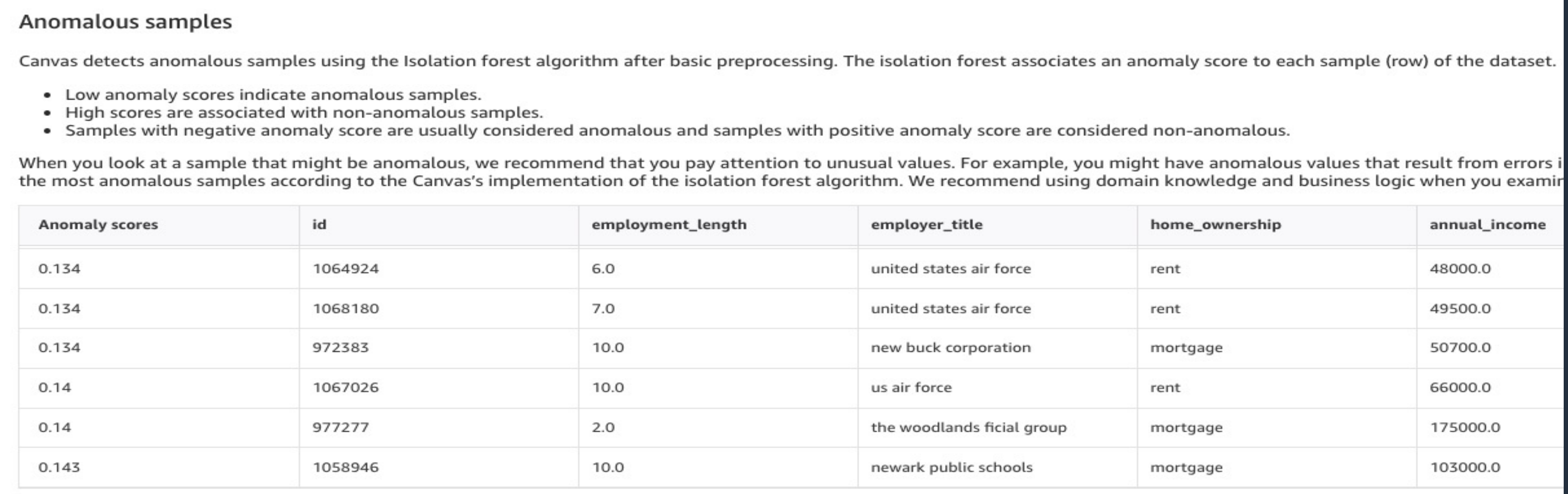
Data quality and insights report

Summary

Dataset statistics

Key	Value
Number of features	5
Number of rows	1000
Missing	1.36%
Valid	98.6%
Duplicate rows	0%





Anomalous samples

Canvas detects anomalous samples using the Isolation forest algorithm after basic preprocessing. The isolation forest associates an anomaly score to each sample (row) of the dataset.

- Low anomaly scores indicate anomalous samples.
- High scores are associated with non-anomalous samples.
- Samples with negative anomaly score are usually considered anomalous and samples with positive anomaly score are considered non-anomalous.

When you look at a sample that might be anomalous, we recommend that you pay attention to unusual values. For example, you might have anomalous values that result from errors i the most anomalous samples according to the Canvas's implementation of the isolation forest algorithm. We recommend using domain knowledge and business logic when you examin

Anomaly scores	id	employment_length	employer_title	home_ownership	annual_income
0.134	1064924	6.0	united states air force	rent	48000.0
0.134	1068180	7.0	united states air force	rent	49500.0
0.134	972383	10.0	new buck corporation	mortgage	50700.0
0.14	1067026	10.0	us air force	rent	66000.0
0.14	977277	2.0	the woodlands ficial group	mortgage	175000.0
0.143	1058946	10.0	newark public schools	mortgage	103000.0

Chat for data prep

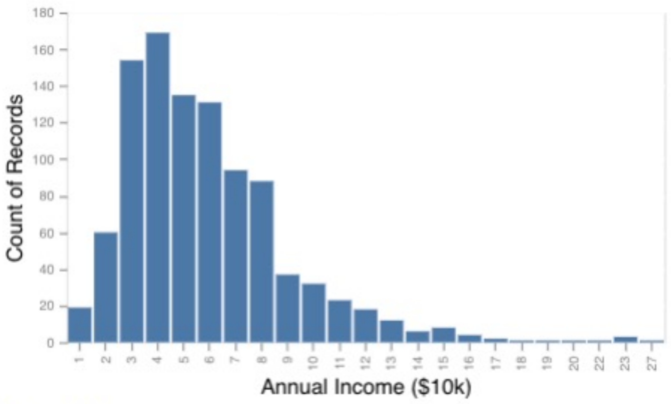
Welcome to chat for data prep!
From here, you can explore, visualize, and transform your data using natural language. To get started, we have some guided prompts for you.

- Plot histogram of the column id
- What is the maximum value of the column annual_income
- What is the mean value of the column annual_income

Plot a chart for annual income in $10000 buckets

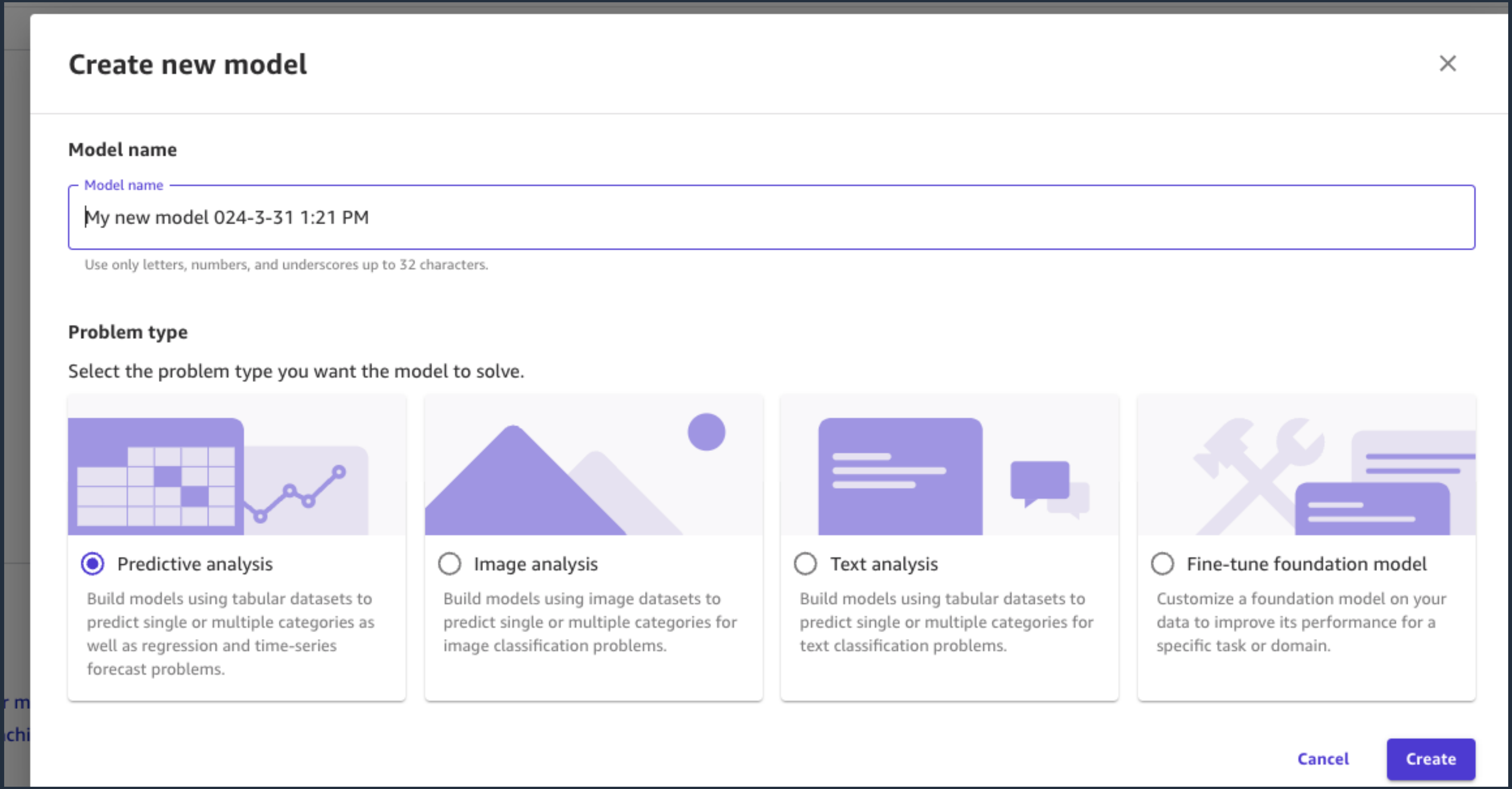
Plot a chart for annual income in $10000 buckets

The code first converts the annual_income column to $10,000 buckets by floor dividing the values by 10,000. It then creates a bar chart using altair, with the bucketed annual_income column on the x-axis and the count of records in each bucket on the y-axis. This generates a histogram showing the distribution of annual_income in $10k buckets.

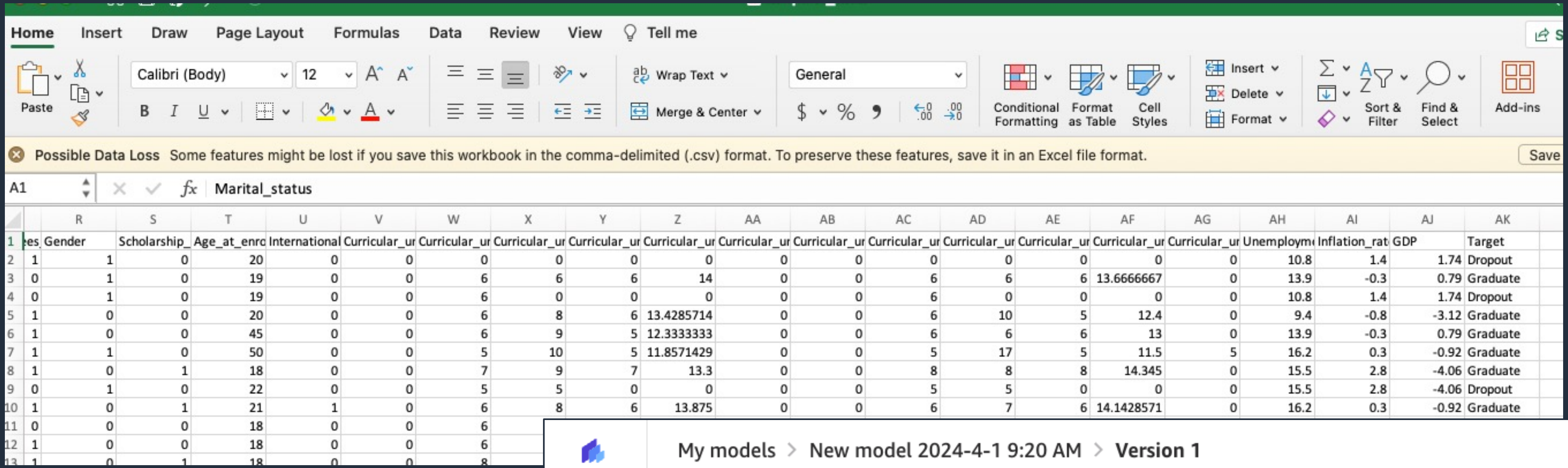


Let’s build a ML model to predict college student dropout

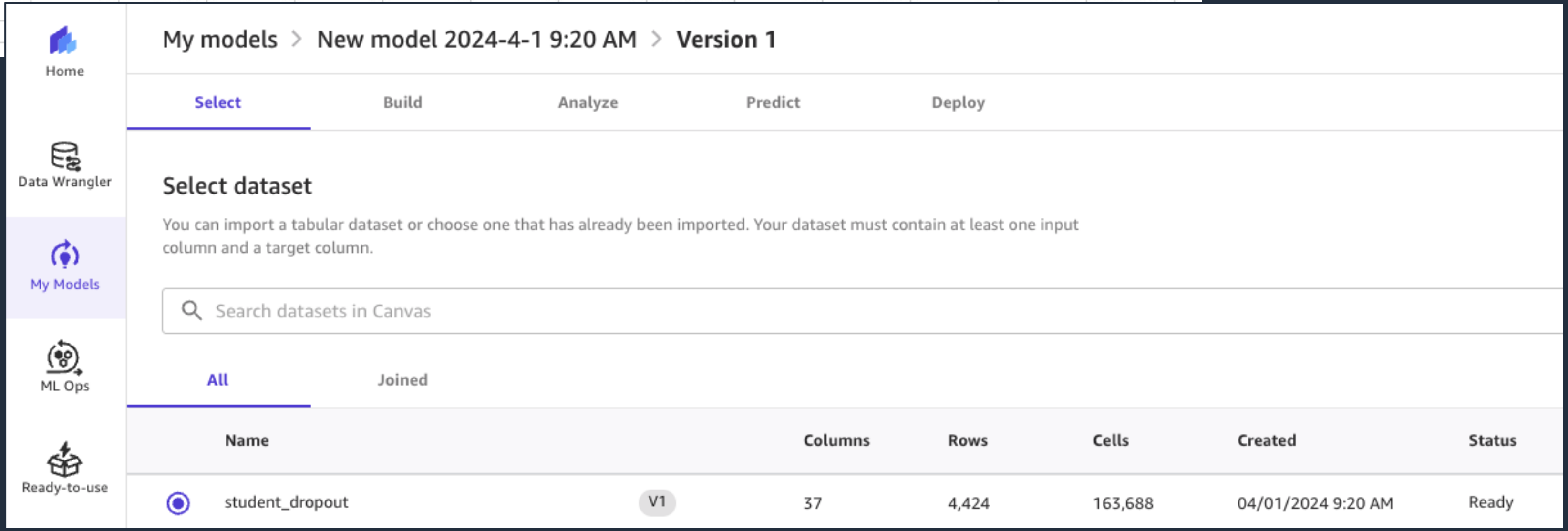
Creating a new model



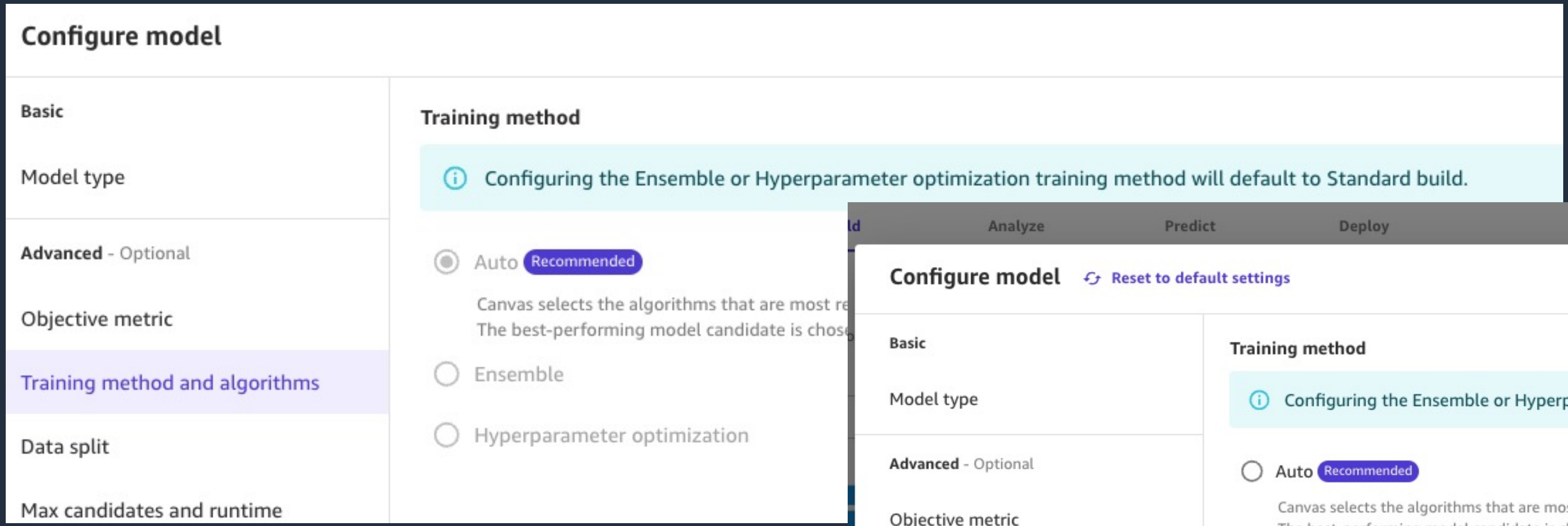
Training data set

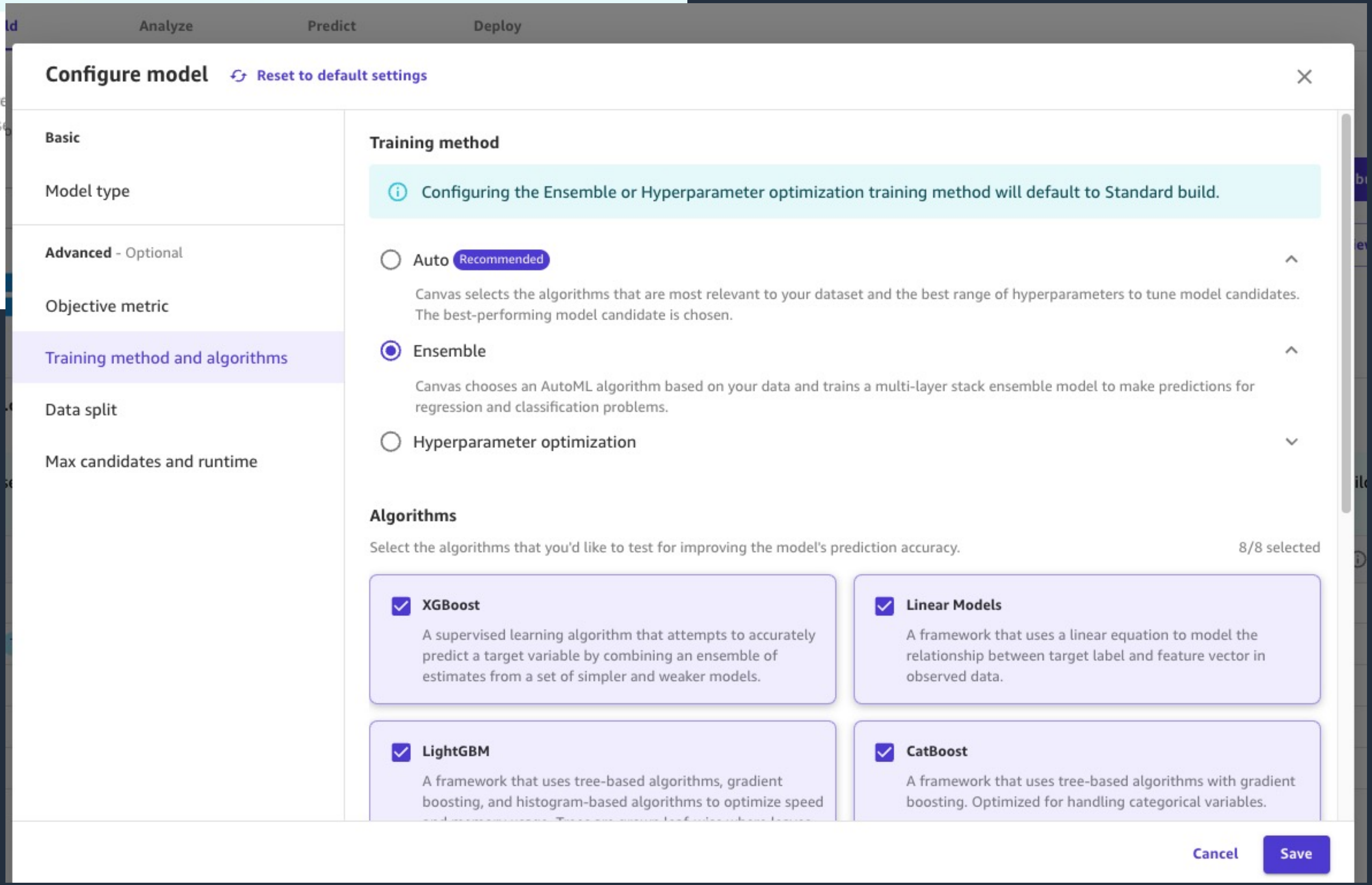


		R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	es_	Gender	Scholarship_	Age_at_enro	International	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Unemploym	Inflation_rat	GDP	Target
2	1	1	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	10.8	1.4	1.74	Dropout
3	0	1	0	19	0	0	6	6	6	14	0	0	6	6	6	13.6666667	0	13.9	-0.3	0.79	Graduate
4	0	1	0	19	0	0	6	0	0	0	0	0	6	0	0	0	0	10.8	1.4	1.74	Dropout
5	1	0	0	20	0	0	6	8	6	13.4285714	0	0	6	10	5	12.4	0	9.4	-0.8	-3.12	Graduate
6	1	0	0	45	0	0	6	9	5	12.3333333	0	0	6	6	6	13	0	13.9	-0.3	0.79	Graduate
7	1	1	0	50	0	0	5	10	5	11.8571429	0	0	5	17	5	11.5	5	16.2	0.3	-0.92	Graduate
8	1	0	1	18	0	0	7	9	7	13.3	0	0	8	8	8	14.345	0	15.5	2.8	-4.06	Graduate
9	0	1	0	22	0	0	5	5	0	0	0	0	5	5	0	0	0	15.5	2.8	-4.06	Dropout
10	1	0	1	21	1	0	6	8	6	13.875	0	0	6	7	6	14.1428571	0	16.2	0.3	-0.92	Graduate
11	0	0	0	18	0	0	6														
12	1	0	0	18	0	0	6														
13	1	0	1	18	0	0	8														

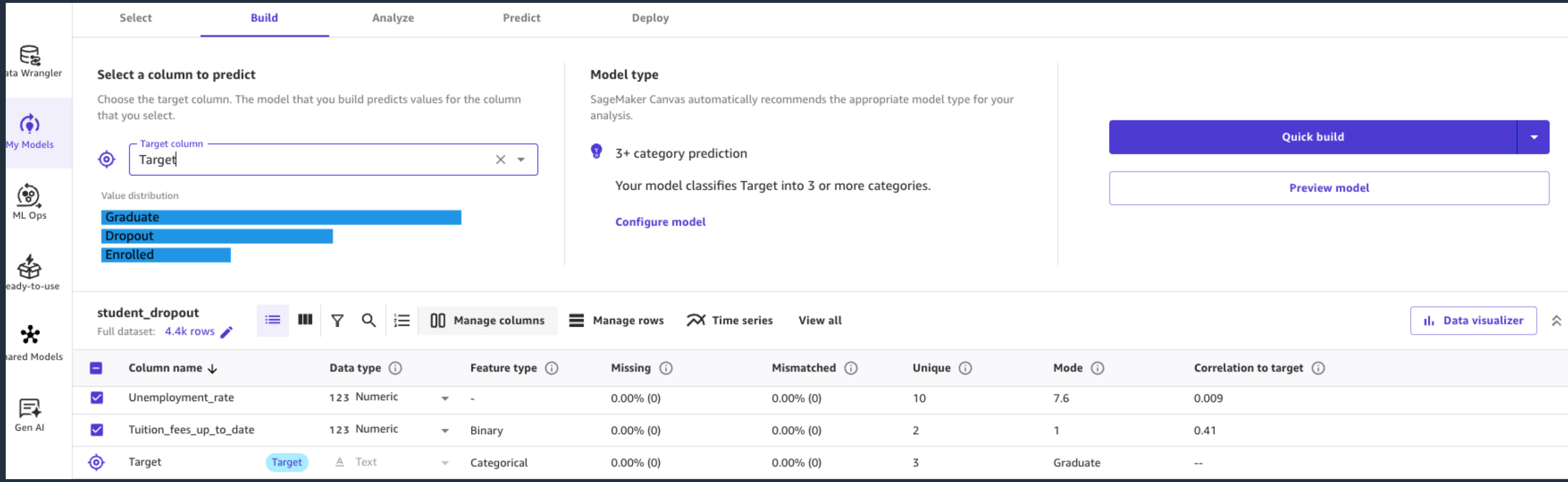


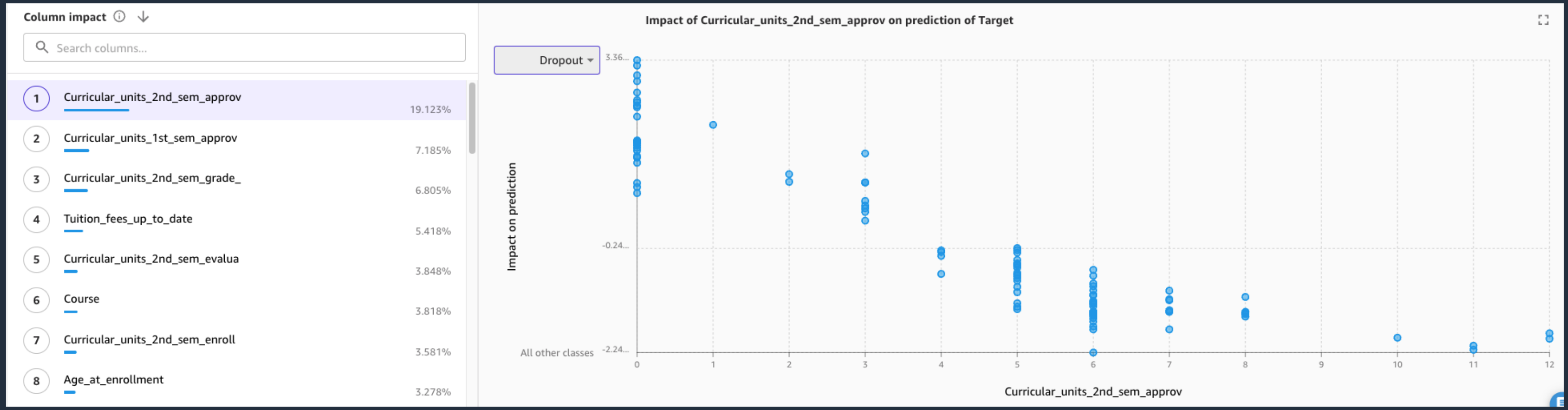
Model configurations

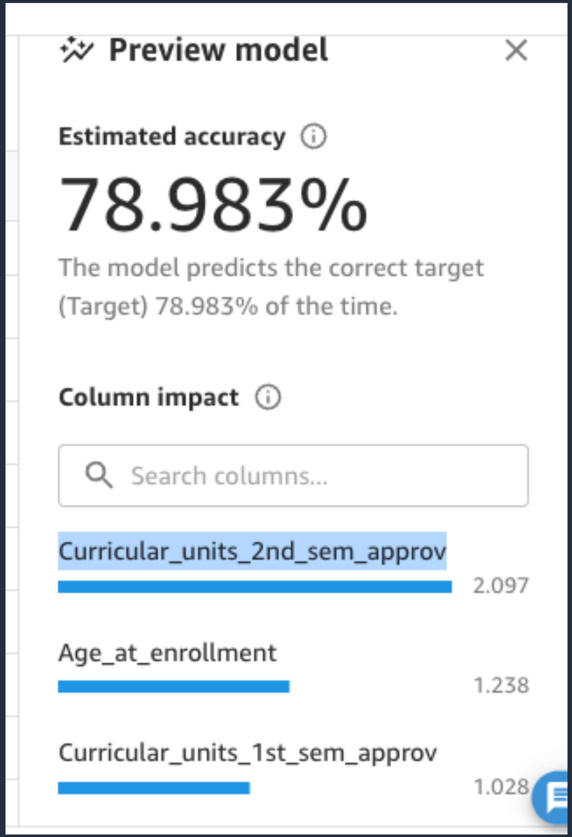




Preview model







Prediction using the model

AC21

	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	tional	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Curricular_ur	Unemployme	Inflation_rat	GDP	Target	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	10.8	1.4	1.74		
3	0	0	6	6	6	14	0	0	6	6	6	13.6666667	0	13.9	-0.3	0.79		
4	0	0	6	0	0	0	0	0	6	0	0	0	0	10.8	1.4	1.74		
5	0	0	6	8	6	13.4285714	0	0	6	10	5	12.4	0	9.4	-0.8	-3.12		
6	0	0	6	9	5	12.3333333	0	0	6	6	6	13	0	13.9	-0.3	0.79		
7	0	0	5	10	5	11.8571429	0	0	5	17	5	11.5	5	16.2	0.3	-0.92		
8	0	0	7	9	7	13.3	0	0	8	8	8	14.345	0	15.5	2.8	-4.06		
9	0	0	5	5	0	0	0	0	5	5	0	0	0	15.5	2.8	-4.06		
10	1	0	6	8														
11	0	0	6	9														
12	0	0	6	6														
13	0	0	8	8														
14	0	0	6	6														
15																		

To make predictions on a dataset, select it or import it. The dataset that you select must have the same number of feature columns as the training dataset.

Search datasets in Canvas

Name		Columns	Rows	Cells	Created	Status
student_dropout_prediction	V1	37	13	481	04/01/2024 10:00 AM	Ready
student_dropout						

batchInfer-New model 2024-4-1 9:20 AM-student_dropout_prediction-1711980102

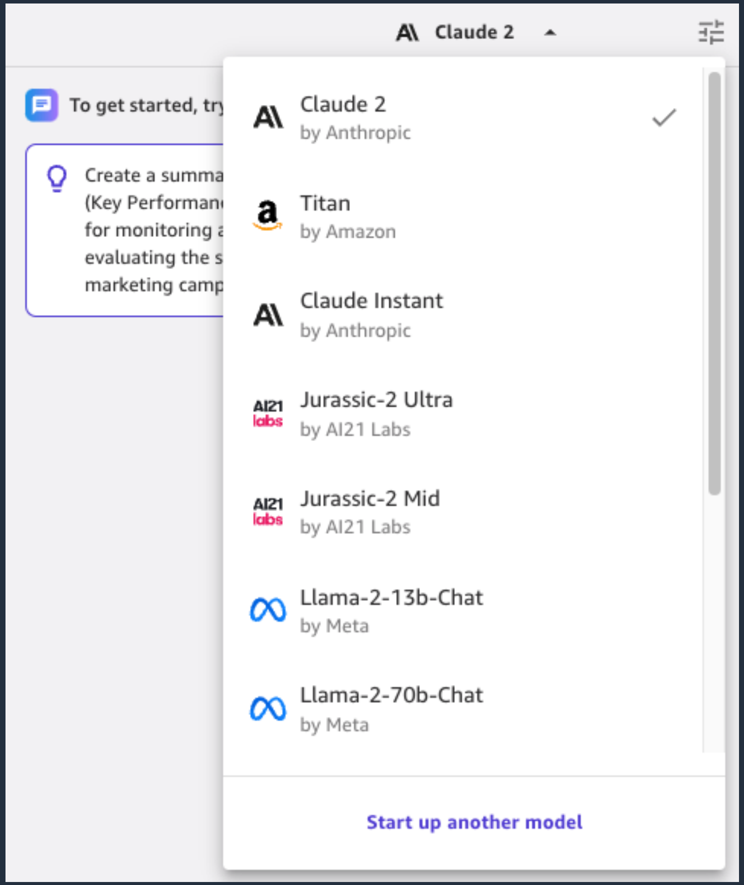
Prediction (Target)	Probability	Marital_status	Application_...	Application_...	Course	Daytime_eve...	Previous_
Dropout	77.5%	1	17	5	171	1	1
Graduate	95.9%	1	15	1	9254	1	1
Dropout	99.5%	1	1	5	9070	1	1
Graduate	76.4%	1	17	2	9773	1	1
Graduate	96.7%	2	39	1	8014	0	1
Dropout	74.1%	2	39	1	9991	0	19

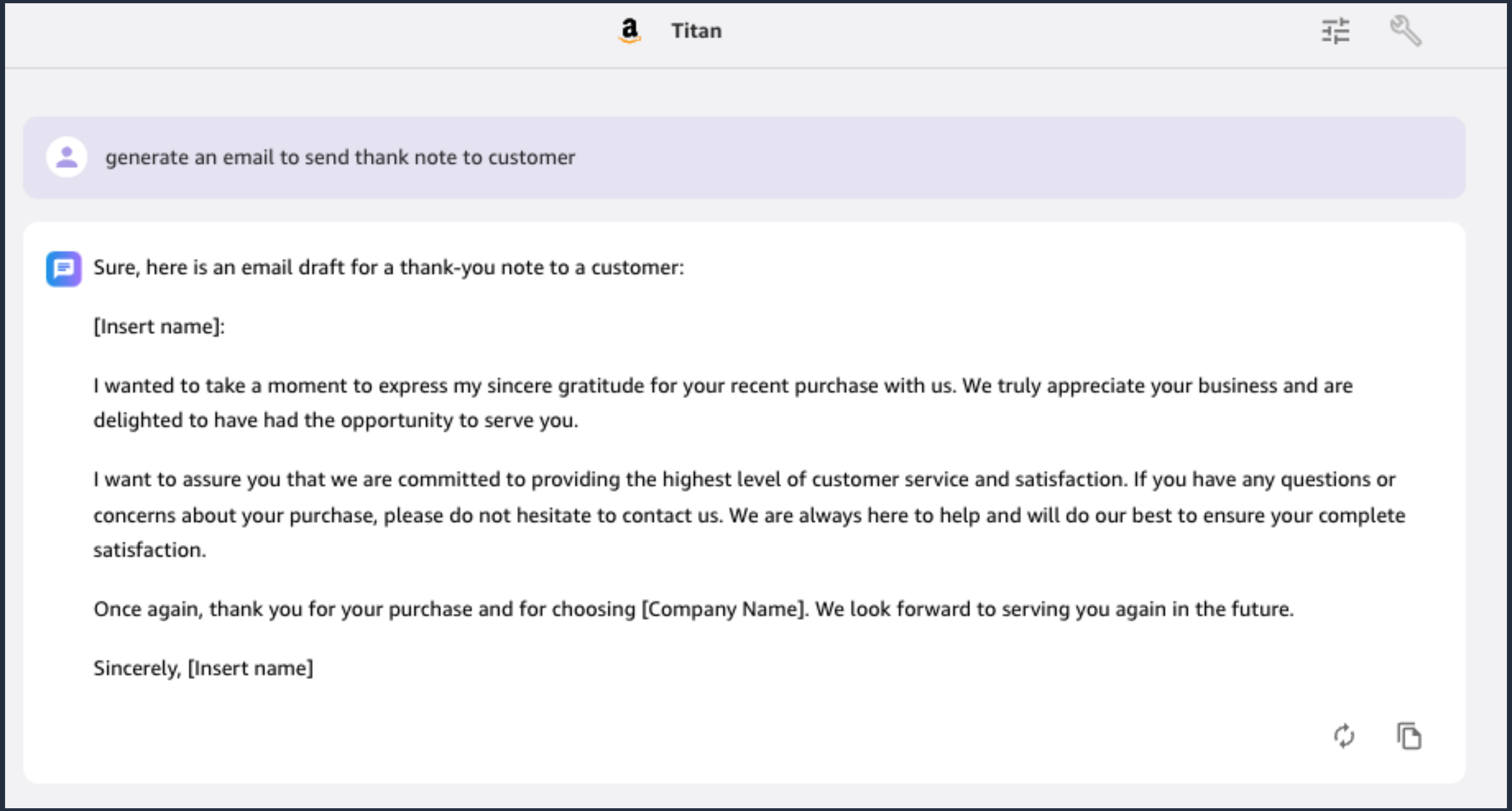
What is generative AI?

- Generative artificial intelligence is artificial intelligence capable of generating text, images, videos, or other data using generative models, often in response to prompts. Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics.

- Large language models (LLM) are very large deep learning models that are pre-trained on vast amounts of data

No code generative AI





Customer stories

What universities are saying

"One of the hardest parts about programming with machine learning is configuring the environment to build. Students usually have to choose the compute instances, security polices, and provide a credit card. My students needed Amazon SageMaker to abstract away all of the complexity of setup and provide a free powerful sandbox to experiment. This lets them **write code immediately without needing to spend time configuring the ML environment**."

Dan Roth

Distinguished Professor of Computer and Information Science at the University of Pennsylvania

Penn
UNIVERSITY of PENNSYLVANIA

What universities are saying

"Amazon SageMaker will help my students learn the building blocks of machine learning by removing the cloud configuration steps required to get started. Now, in my natural language processing classes, students have more time to enhance their skills."

Sanjiv Das

Professor of Finance and Data Science at Santa Clara University

How to get started with Amazon SageMaker Canvas

1. Immersion Day
2. MOOC via Coursera
3. Getting started tutorial

aws.amazon.com/sagemaker/canvas

Thank you!

Please complete the session survey by scanning the QR code