



Denver Learning Days

Amazon SageMaker Canvas

Grant Joslyn
Senior Solutions Architect
gajoslyn@amazon.com

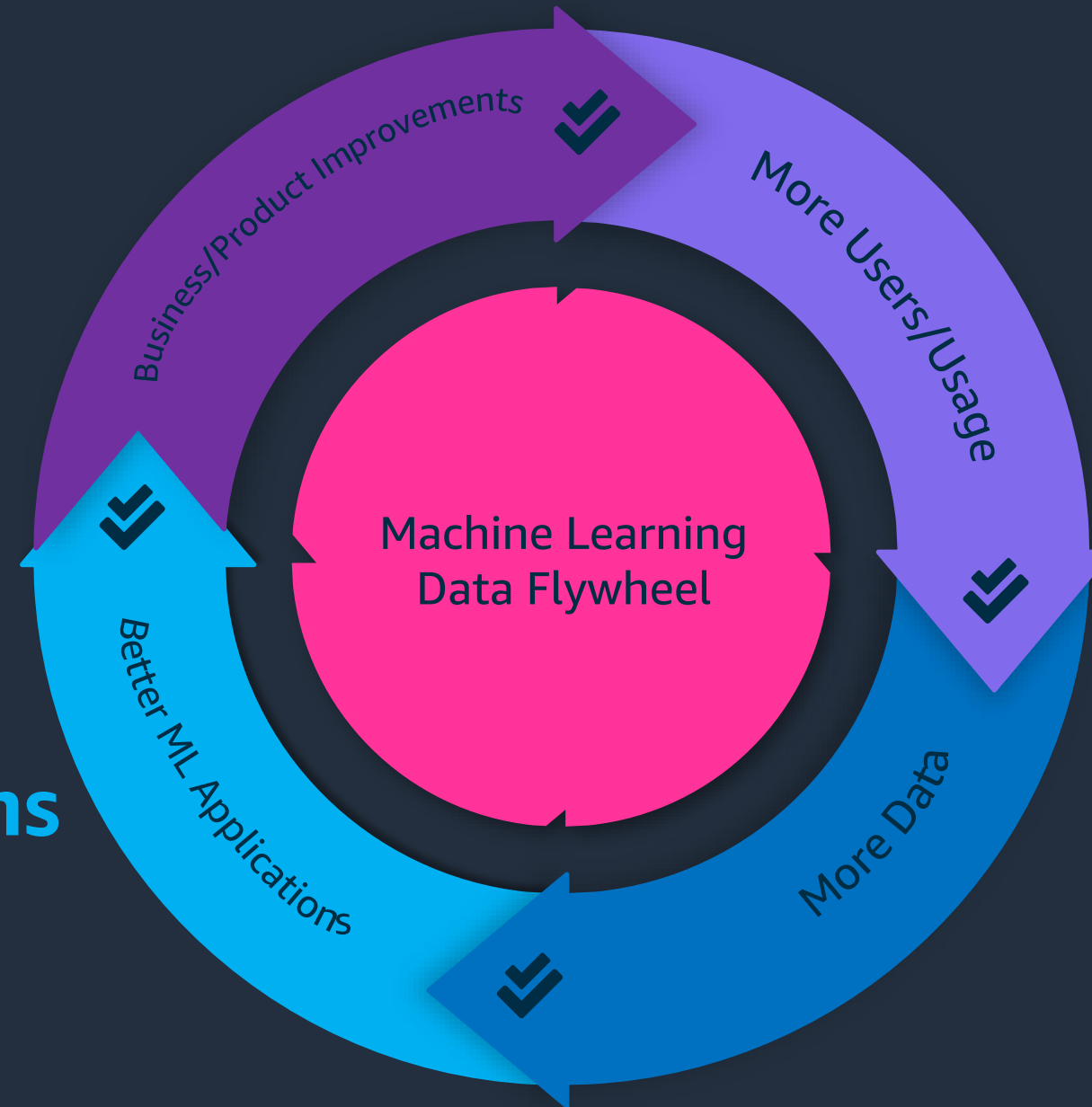
ML Provides Repeatable Value

Business/Product Improvements

ML powered insights and applications improve business decisions and lead to better products/services

Better ML Applications

Building more models to solve more use cases and improving the existing models accuracy.



More Users/ Usage

Having more users or more for your current users to consume.

More Data

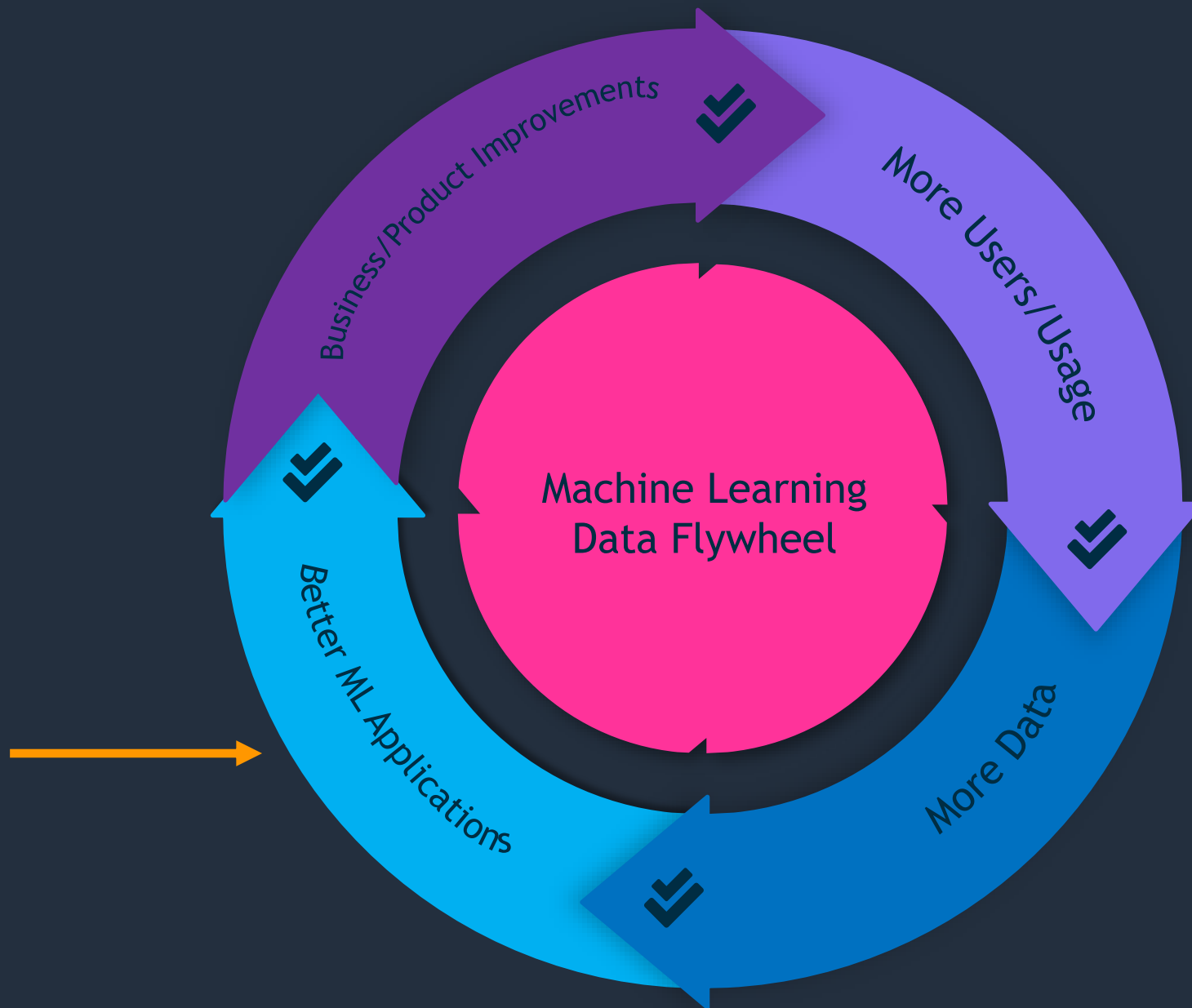
With more data available, there are more use cases and insights that ML can provide valuable impact

Data Science Resources Are A Limiting Factor

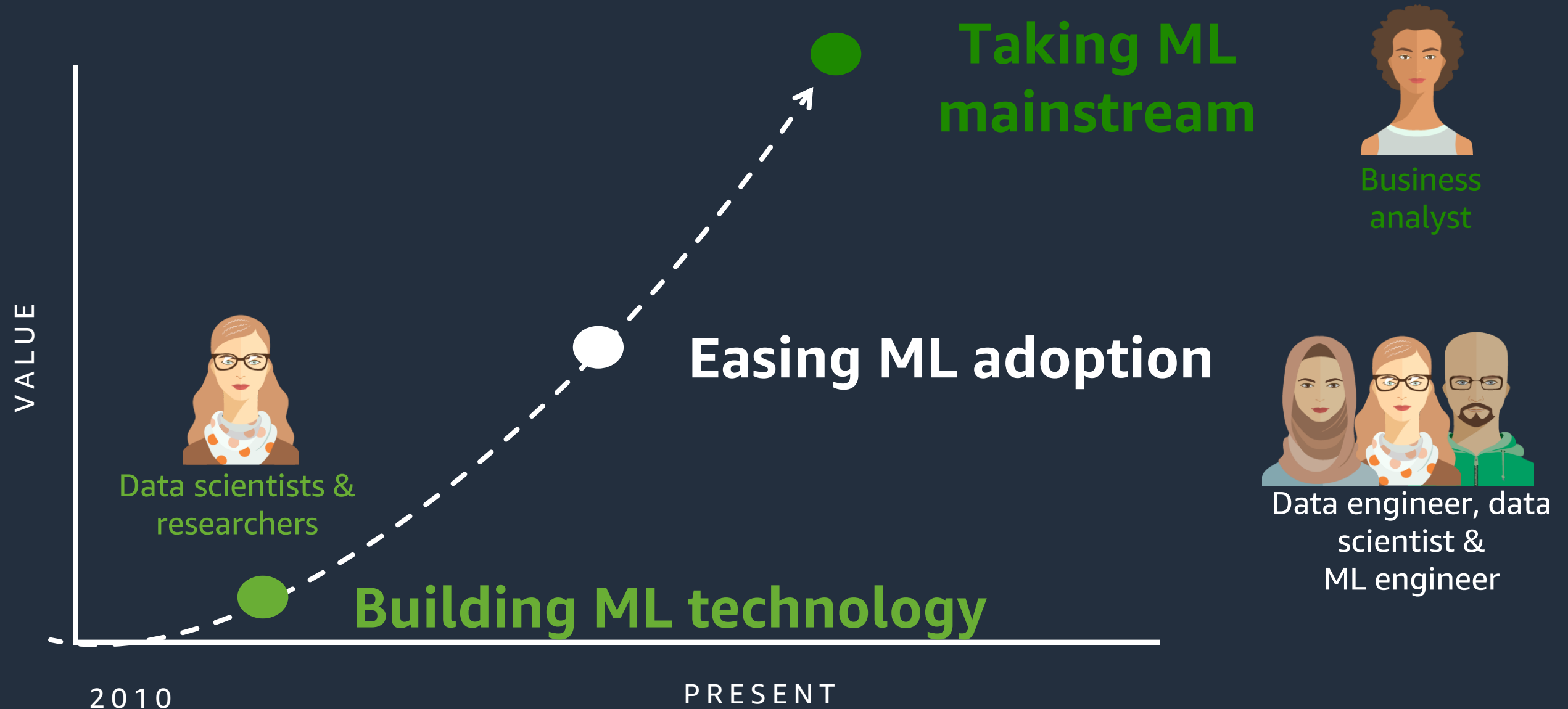
The Bottleneck

Building better ML applications can be the bottleneck because:

- Data Science talent is expensive to scale
- Friction points in the Machine Learning journey



ML technology maturity and democratizing ML to accelerate ML-driven value creation





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.

+
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

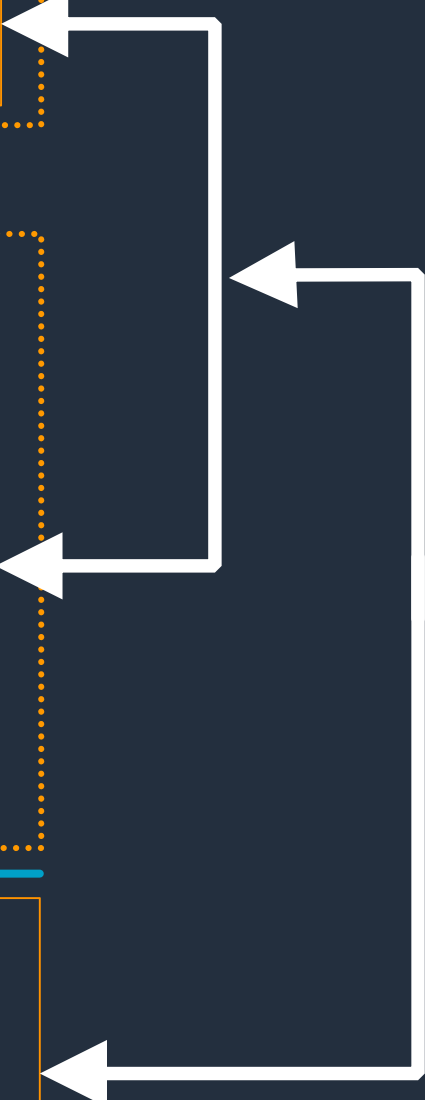
Data Wrangler
A faster, visual way to aggregate and prepare data for machine learning

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

2-Way Collaboration




Amazon Quicksight


Data Stores – 40+ Connectors

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



Amazon SageMaker Canvas

No-code workspace for **business teams** to build, customize, and deploy **ML and Generative AI models**



Ready-to-Use Foundation Models

Customize and fine-tune multiple LLMs using your own documents with a few clicks



Custom Models with AutoML

Build highly accurate tabular, CV, and NLP models specific to your data



End-to-End Model-Building

Prepare data, train, and deploy models



Collaboration with Experts

Share models between business analysts and data scientists for validation and governance

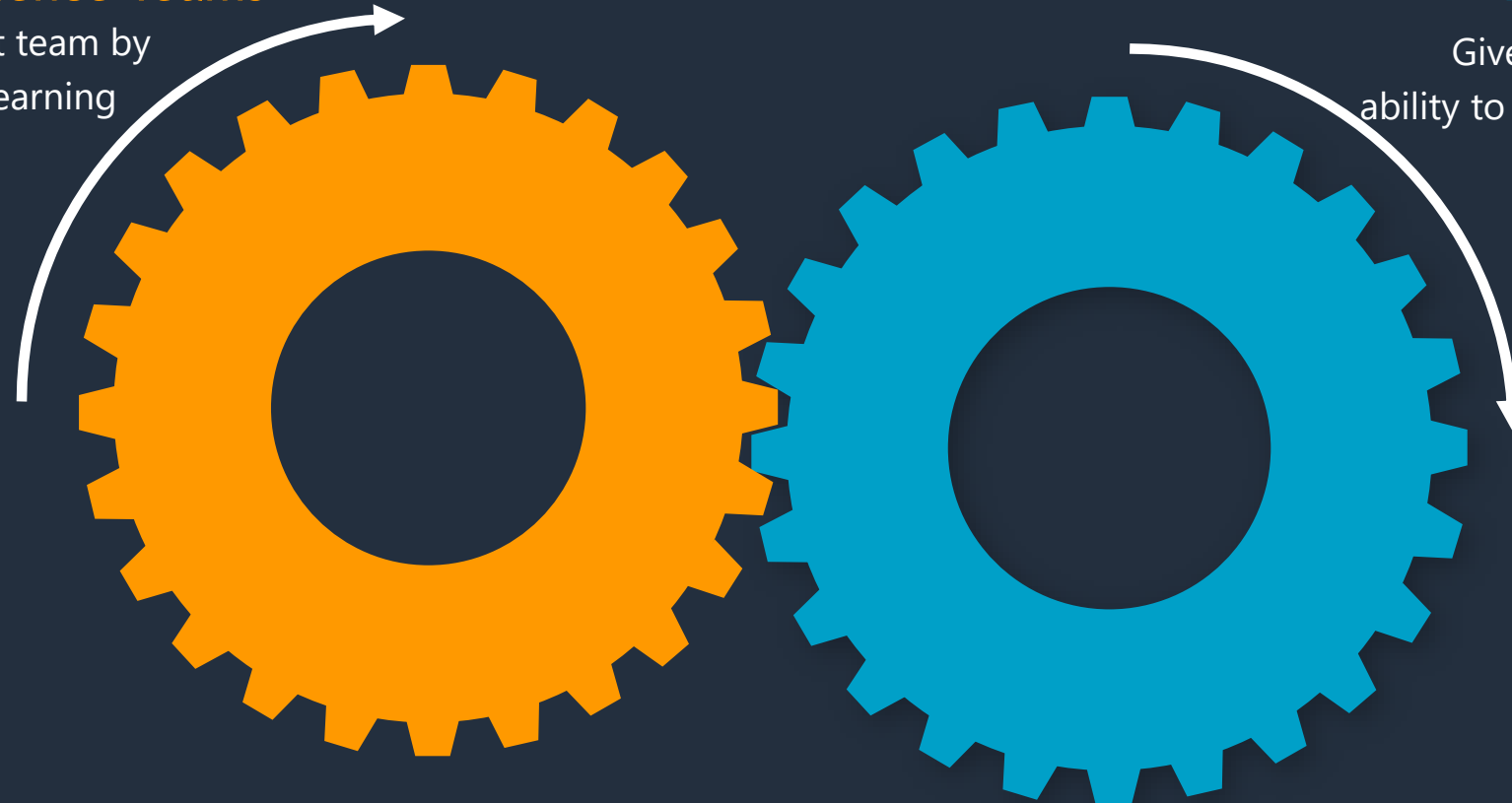
Ways SageMaker Low-code/No-code Helps

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

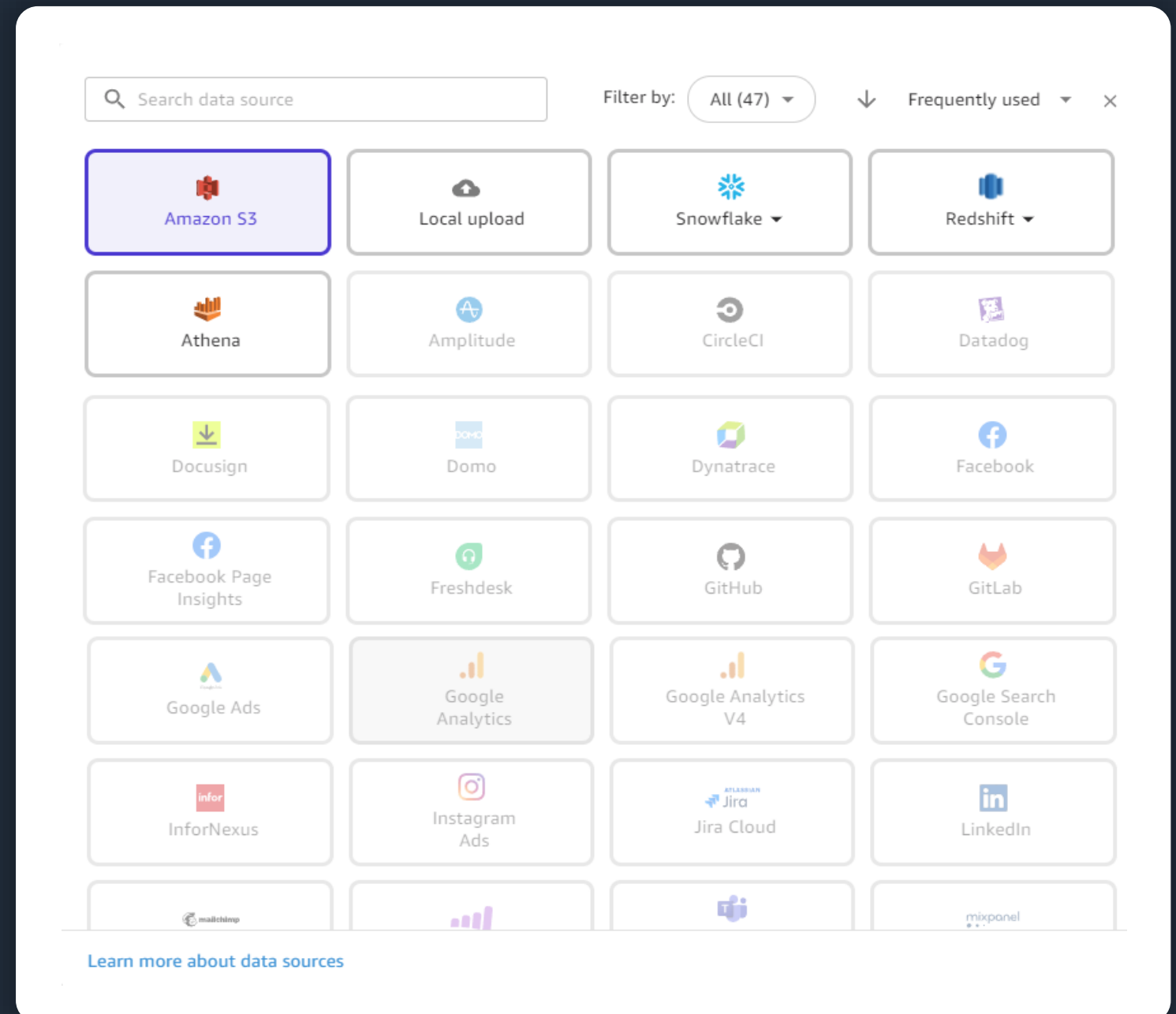


Collaborate together

SageMaker Low-Code/No-Code ML has several points of collaboration making it easy 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

Quickly connect and join data to build your dataset

- ✓ Import from 50+ data sources
- ✓ Support for native Snowflake and Salesforce connectors using SSO
- ✓ **[NEW]** Support for Amazon DocumentDB



[NEW] Point and click UI to explore and prepare data

- ✓ SageMaker Data Wrangler capabilities now available in Canvas
- ✓ 300+ ML-specific transformations
- ✓ Export data flow as SageMaker notebook, SageMaker pipelines, S3 or as Canvas dataset
- ✓ Accessible via SageMaker Studio

Data: Data flow / Titanic_exploration / Dataset name

Step 2. Data types Train model Export snapshot

The following is a visualization of the first 2,000 rows of your dataset. To analyze the entire dataset, [Check Data Quality](#)

Sold	Price	Region	Discount	Fabric
Yes		Southwest	-	yes
Yes	null	-	23	yes
Yes		Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes
Yes	null	Southwest	23	yes

Dataset type: Tabular Total dataset cells (columns x rows): 390972 (12 x 32581) Data source: S3

[NEW] Quickly explore and prepare data powered using natural language

- ✓ Use natural language to generate visualizations and data preparation steps
- ✓ Powered by Amazon Bedrock
- ✓ Data remains in your environment
- ✓ Managed via separate policy

Data Wrangler: Data flow > canvas-data-prep.flow > canvas-sample-housing.csv

Data Analyses

Previewing: Chat: 2023-11-9 2:08 PM

Chat for data prep Show steps Create model Export data

remove rows where population is less than 1000

The code filters the dataframe to only keep rows where the population column is greater than or equal to 1000.

View code






PySpark

```
1 from pyspark.sql.functions import col
2
3 df = df.filter(col("population") >= 1000)
```

Cancel Add to steps

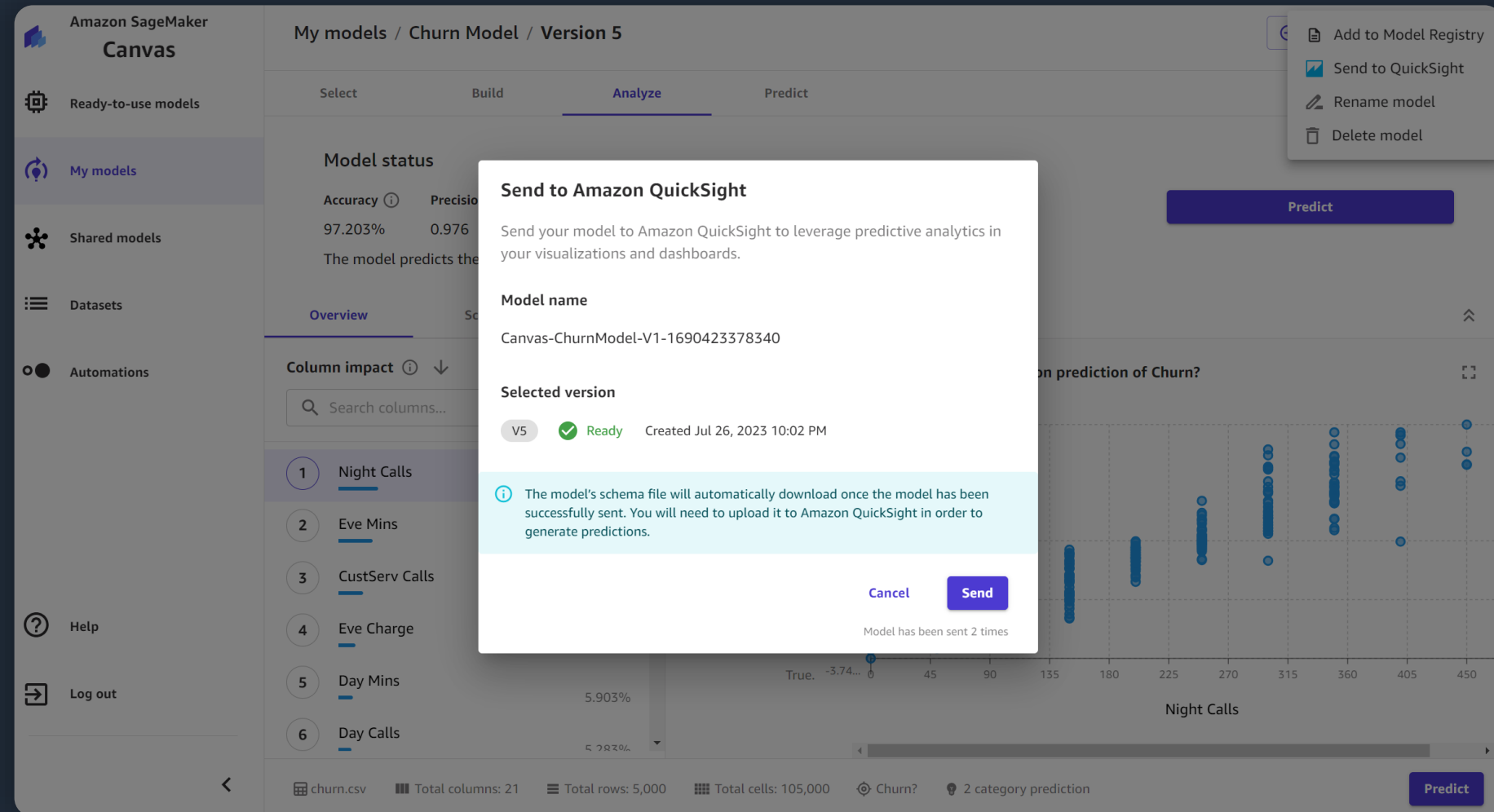
e.g. Help me understand my data with a summary

Preview

longitude (float)	latitude (float)	housing_median_age (float)	total_rooms (float)	total_bedrooms (float)
				
-122.34 - -121.73 123	37.47 - 37.9 123	3 - 52 123	609 - 28258 123	236 - 3864
-122.22	37.86	21	7099	1106

Consume predictions visually or programmatically


- ✓ Interact with predictions in Canvas UI
- ✓ Click-to-deploy to SageMaker endpoint
- ✓ Visualize predictions in QuickSight



Use ready to use models

OR

Build your own



Ready-to-use models

Here are some ready-to-use models we've prepared for you to use.


You can start generating predictions with pre-built models without writing a single line of code. To get started, bring your data such as text, images, or documents and select a model to extract information and insights.


JetBrains Toolbox


Can't find the right model? [Create a custom model](#)


Filter by data type: Text Image Document


↓ Last used ▾ Grid List


**Object detection in images**
Detect objects, concepts, scenes, and actions in your images.
Powered by Amazon Rekognition


**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


**Document analysis**
Analyze documents and forms for relationships among detected text.
Powered by Amazon Textract


**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


**Sentiment analysis**
Detect sentiment in lines of text, which can be positive, negative, neutral, or mixed.
Powered by Amazon Comprehend

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

**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

**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

My Custom Model


Use only letters, numbers, and underscores up to 32 characters.

Problem type


Select the problem type you want the model to solve.




☒ **Predictive analysis**
Build models using tabular datasets to predict single or multiple categories or regression problems.



☐ **Forecast**
Build models for time-series forecast problems.




☐ **Anomaly Detection**
Build models to predict anomalies in your tabular time-series datasets.



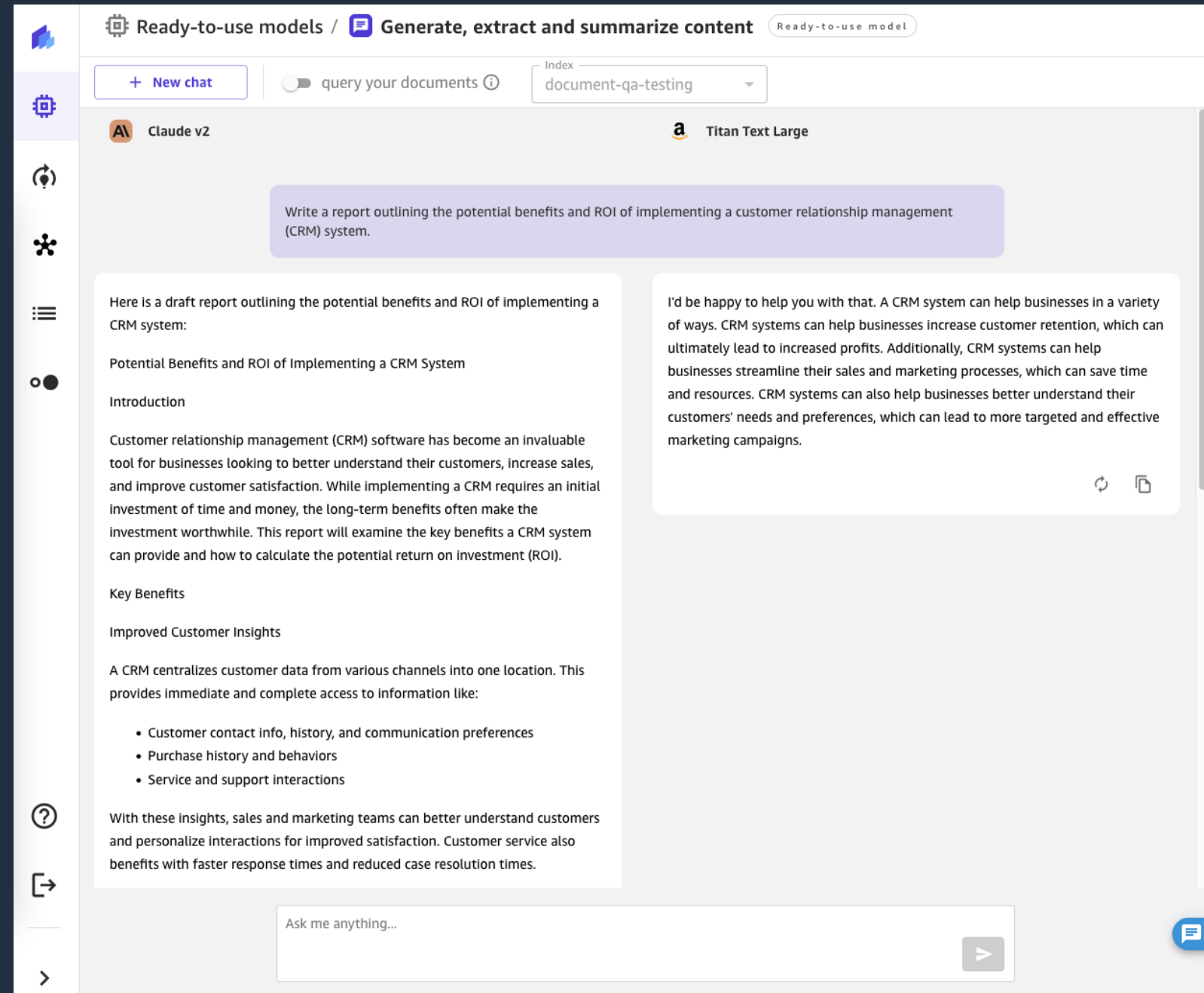
☐ **Text analysis**
Build models using tabular datasets to predict single or multiple categories for text classification problems.

© 2024, Amazon Web Services, Inc. or its Affiliates.



Safely and securely use Foundation Models within seconds

- ✓ Compare model output from different models
- ✓ Content is not used to improve base models or shared with 3P model developers
- ✓ Data encrypted in transit and at rest



[NEW] Customize foundation models using RAG

- ✓ Improve relevance and accuracy of model responses
- ✓ Native support for Kendra as document store

The screenshot displays the Amazon SageMaker Canvas interface, which is used for building and deploying machine learning applications. The interface is divided into several sections:

- Header:** Shows the SageMaker Canvas logo and the current application name, "Ready-to-use models".
- Navigation:** A sidebar on the left contains icons for various functions, including a search icon, a document icon, and a chat icon.
- Main Content Area:** Displays a technical document titled "DCN00669(1)_QTRCO_IOM_LDA_typeC.pdf". The document includes a table of models and torque specifications, and diagrams of bolt patterns.
- Chat Window:** A chat window on the right shows a query: "What do I do if my L Series type C double acting actuator is leaking?". The response provides a list of steps to take if the actuator is leaking, based on the excerpts provided.

Table of Models and Torque Specifications:

MODEL	LB* FT	Nm
L04	10	14
L06	20	27
L08	35	48
L10	35	48
L12	35	48
L14	56	76
L16	56	76
L20	58	79
L24	87	118

Bolt Pattern Diagrams:

- 4 Bolt Pattern
- 6 Bolt Pattern
- 8 Bolt Pattern
- 12 Bolt Pattern

Chat Window:

Query: What do I do if my L Series type C double acting actuator is leaking?

Response: Based on the excerpts provided, here are some steps to take if your L Series type C double acting actuator is leaking:

- Inspect the piston o-ring and wiper ring. Replace them if needed, using the proper QTRCO approved lubricant.
- Make sure the internal surface of the cylinder is lubricated before reinstalling the cylinder over the piston.
- Check that the cylinder is properly seated in the seal groove when reassembling.
- Inspect the piston for corrosion or wear. Replace the piston if excessive corrosion or wear is found.
- Check the rod seals and bushings. Replace them if needed.
- If the leaks persist after replacing seals and lubricating, you may need a new actuator or actuator components.

The excerpts mention inspecting and replacing seals, lubricating properly during reassembly, and replacing damaged components as ways to address leaks in this type of actuator. Let me know if you need any clarification on these steps!

Source: 1. DCN00669(1)_QTRCO_IOM_L... 2. DCN00677(1)_QTRCO_IOM_L... 3. DCN00669(1)_QTRCO_IOM_L... 4. DCN00669(1)_QTRCO_IOM_L... 5. DCN04081(1)_QTRCO_IOM_D...

Index: manufacturing-demo

[NEW] Customize foundation models with fine-tuning

- ✓ Adapt models to specific domains that could be under-represented in original training
- ✓ Data used in fine-tuning is encrypted and does not leave your VPC
- ✓ Fine-tune models in parallel to accelerate workflow and select from model leaderboard

My models / { Name of model } / Version 1

+ Add version

Select Build Analyze Deploy

Fine-tune model configurations

Select up to 3 base model

The foundation models you choose will be the base models that are customized to perform your desired language task. [Learn more about our foundation models.](#)

Titan Text Large Jurassic 2 Ultra MPT 7B

Amazon Bedrock models (2/2)

- ☒ Titan Text Large
- ☒ Jurassic 2 Ultra
- ☐ Jurassic 2 Mid

Amazon SageMaker JumpStart models

- ☐ Falcon 40b
- ☐ Falcon 7b
- ☒ MPT 7b

Advanced settings

Select Input column

The input column contains the text that are used to train the model.

Input column

Select Output column

The output column contains the text that you want the model to generate.

Output column

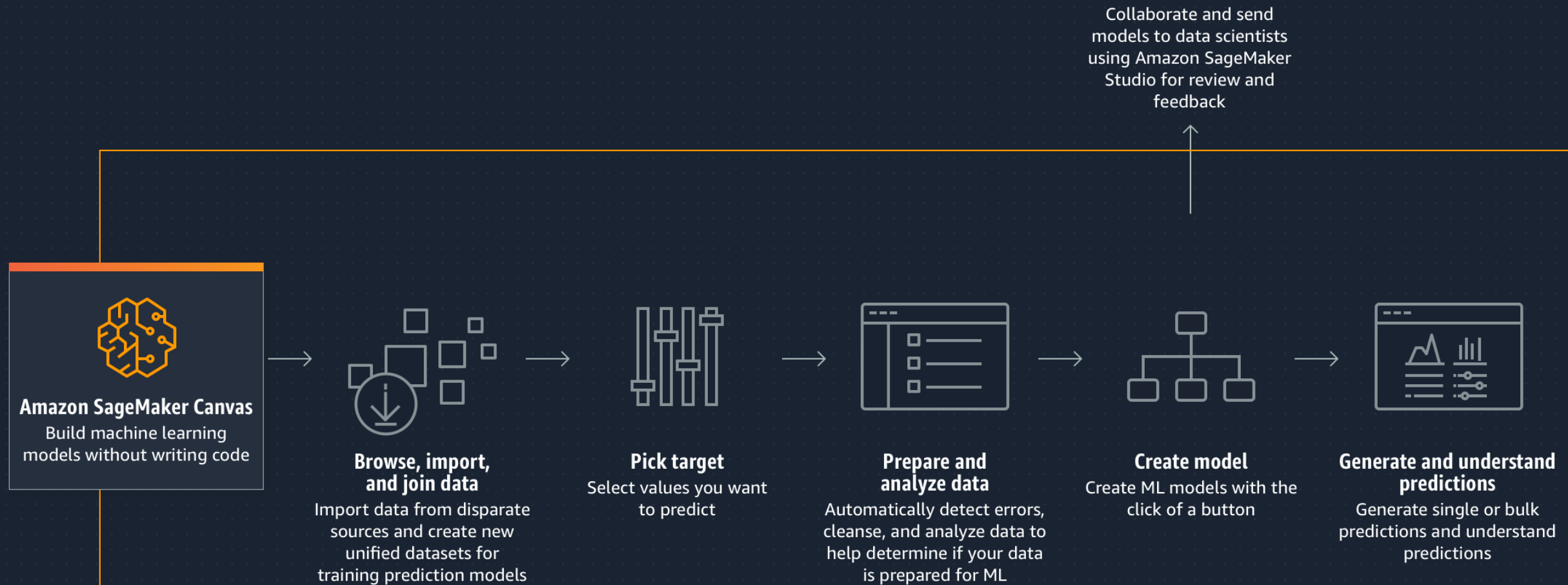
Build

Prepare data

Feature type	Missing	Mismatched	Mode	Unique
-	0% (0)	0% (0)	Northeast	22
Completion	0% (0)	0% (0)	Middle	22
Column 3	5.6% (1836)	0% (0)	Middle	22
Column 4	0% (0)	0% (0)	Middle	22
Column 5	0% (0)	0% (0)	Middle	22
Column 6	0% (0)	0% (0)	Middle	22

Total columns: 5 Total rows: 5,110 Total cells: 5,110

Amazon SageMaker Canvas: no-code ML



How to get started with
Amazon SageMaker Canvas

SageMaker Canvas

Demo / Hands-on

Thank you!

Feedback is a
gift!
Survey QR Code



Grant Joslyn
Senior Solutions Architect
gajoslyn@amazon.com