

Lab 2 SageMaker Data Wrangler





Services

Q s3



N. Virginia ▾

TeamRole/MasterKey @ 8341-2056-7544 ▾

Amazon SageMa

[Dashboard](#)[Search](#)

SageMaker Domain

[Studio](#)[RStudio](#)[Canvas](#)

Images

[▶ Ground Truth](#)[▶ Notebook](#)[▶ Processing](#)[▶ Training](#)[▶ Inference](#)[▶ Edge Manager](#)[▶ Augmented AI](#)[▶ AWS Marketplace](#)[Services \(7\)](#)[Features \(10\)](#)[Blogs \(1,036\)](#)[Documentation \(96,682\)](#)[Knowledge Articles \(30\)](#)[Tutorials \(4\)](#)[Events \(14\)](#)[Marketplace \(720\)](#)

Search results for 's3'

Services

[See all 7 results ▶](#)**S3**

Scalable Storage in the Cloud

**S3 Glacier**

Archive Storage in the Cloud

**Athena**

Query Data in S3 using SQL

**AWS Snow Family**

Large Scale Data Transport

Features

[See all 10 results ▶](#)**Amazon S3 File Gateway**

Storage Gateway feature

Datasets



IoT Analytics feature

Amazon SageMaker project templates enabled for Studio users

[Add user](#)

< 1 >

[Launch app ▾](#)[Delete Domain](#)[Edit Settings](#)

Authentication method

AWS Identity and Access Management (IAM)

Amazon S3 X**Buckets**

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 3

▶ AWS Marketplace for S3

Amazon S3 > Buckets

▶ Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)[View Storage Lens dashboard](#)**Buckets (3) Info**[Copy ARN](#)

Empty

Delete

[Create bucket](#)Buckets are containers for data stored in S3. [Learn more](#) Find buckets by name< 1 >

Name	AWS Region	Access	Creation date
sagemaker-studio-834120567544-5zbey54q71a	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 20:55:19 (UTC+08:00)
sagemaker-us-east-1-834120567544	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 21:06:52 (UTC+08:00)

Amazon S3 > Create bucket

Create bucket Info

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

Bucket name must be unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

Copy settings from existing bucket - optional

Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Object Ownership Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

ACLs disabled (recommended)

All objects in this bucket are owned by this account.
Access to this bucket and its objects is specified using

ACLs enabled

Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be

Bucket Versioning

- Disable
 Enable

Tags (0) - optional

Track storage cost or other criteria by tagging your bucket. [Learn more](#)

No tags associated with this bucket.

[Add tag](#)**Default encryption**

Automatically encrypt new objects stored in this bucket. [Learn more](#)

Server-side encryption

- Disable
 Enable

► Advanced settings

After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.

[Cancel](#)[Create bucket](#)

s3.console.aws.amazon.com/s3/home?region=us-east-1

aws Services Search for services, features, blogs, docs, and more [Option+S]

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight 3

AWS Marketplace for S3

Success message: Successfully created bucket "titanic-michlin-20220218". To upload files and folders, or to configure additional bucket settings choose View details.

Read the S3 resources page for documentation and technical content.

Amazon S3

▶ Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens dashboard](#)

Buckets (3) Info

Buckets are containers for data stored in S3. [Learn more](#)

Name	AWS Region	Access	Creation date
sagemaker-studio-834120567544-5zbey54q71a	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 20:55:19 (UTC+08:00)
sagemaker-us-east-1-834120567544	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 21:06:52 (UTC+08:00)
titanic-michlin-20220218	US East (N. Virginia) us-east-1	Bucket and objects not public	February 16, 2022, 22:15:58 (UTC+08:00)

Event Engine - Team Dashboard X | Amazon SageMaker X | titanic-michlin-20220218 - S3 X | JupyterLab X | Introducing Amazon SageMake X | +

s3.console.aws.amazon.com/s3/buckets/titanic-michlin-20220218?region=us-east-1&tab=objects

Paused

AWS Services Search for services, features, blogs, docs, and more [Option+S]

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight 3

▶ AWS Marketplace for S3

Amazon S3 > titanic-michlin-20220218

titanic-michlin-20220218 Info

Objects Properties Permissions Metrics Management Access Points

Objects (0)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI Copy URL Download Open Delete Actions Create folder

Upload

Find objects by prefix

< 1 >

	Name	Type	Last modified	Size	Storage class	
No objects						
You don't have any objects in this bucket.						
Upload						

Event Engine - Team Dashboard X | Amazon SageMaker X | S3 Management Console X | JupyterLab X | Introducing Amazon SageMake X | +

← → C ⌂ 🔒 s3.console.aws.amazon.com/s3/upload/titanic-michlin-20220218?region=us-east-1

aws Services Search for services, features, blogs, docs, and more [Option+S]

Amazon S3 > titanic-michlin-20220218 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files**, or **Add folders**.

Files and folders (0)

All files and folders in this table will be uploaded.

	Name	Folder	Type	Size
No files or folders				

You have not chosen any files or folders to upload.

Destination

Destination

Event Engine - Team Dashboard X | Amazon SageMaker X | S3 Management Console X | JupyterLab X | Introducing Amazon SageMake X | +

s3.console.aws.amazon.com/s3/upload/titanic-michlin-20220218?region=us-east-1

AWS Services Search for services, features, blogs, docs, and more [Option+S]

Amazon S3 > titanic-michlin-20220218 > Upload

Upload Info

Favorites

- mba
- Desktop
- Documents
- Downloads

iCloud

- Shared

Locations

- Network

Media

- Music
- Photos
- Movies

Data

test.csv

train.csv

winequality-red.csv

train.csv

Comma Separated Spreadsheet (.csv) - 61 KB

Information

Created

Show More

Yesterday, 10:23 PM

Cancel Open

Destination

Destination

Amazon S3 > titanic-michlin-20220218 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files**, or **Add folders**.

Files and folders (1 Total, 59.8 KB)

All files and folders in this table will be uploaded.

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	train.csv	-	text/csv	59.8 KB

Remove **Add files** **Add folder**

Find by name < 1 >

Destination

Destination
[s3://titanic-michlin-20220218](#)

▶ **Destination details**
Bucket settings that impact new objects stored in the specified destination.

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files**, or **Add folders**.

Files and folders (1 Total, 59.8 KB)

All files and folders in this table will be uploaded.

[Remove](#)[Add files](#)[Add folder](#) Find by name

< 1 >

<input type="checkbox"/>	Name	Folder	Type	Size
<input type="checkbox"/>	train.csv	-	text/csv	59.8 KB

Destination

Destination

<s3://titanic-michlin-20220218>

► Destination details

Bucket settings that impact new objects stored in the specified destination.

► Permissions

Grant public access and access to other AWS accounts.

► Properties

Specify storage class, encryption settings, tags, and more.

[Cancel](#)[Upload](#)



Services

Search for services, features, blogs, docs, and more

[Option+S]



Global ▾

TeamRole/MasterKey @ 8341-2056-7544 ▾

✓ Upload succeeded
View details below.

Upload: status

[Close](#)

The information below will no longer be available after you navigate away from this page.

Summary

Destination

[s3://titanic-michlin-20220218](#)

Succeeded

1 file, 59.8 KB (100.00%)

Failed

0 files, 0 B (0%)

[Files and folders](#)[Configuration](#)

Files and folders (1 Total, 59.8 KB)

Find by name

< 1 >

Name	▲	Folder	▼	Type	▼	Size	▼	Status	▼	Error	▼
train.csv	-			text/csv		59.8 KB		Succeeded		-	

Event Engine - Team Dashboard X Amazon SageMaker X Amazon SageMaker Immersion X +

console.aws.amazon.com/sagemaker/home?region=us-east-1#/landing

Paused

aws Services Q sagemaker|Canvas X

N. Virginia ▾ TeamRole/MasterKey @ 8341-2056-7544 ▾

Amazon SageMa

Search results for 'sagemaker'

Services (2)

Amazon SageMaker Build, Train, and Deploy Machine Learning Models

AWS Glue DataBrew Visual data preparation tool to clean and normalize data for analytics and machine learn...

Features

SageMaker Studio Amazon SageMaker feature

SageMaker Canvas Amazon SageMaker feature

Notebooks IoT Analytics feature

Autopilot

Studio, a integrated environment (IDE) for building and debugging experiments, and monitoring

With SageMaker, you pay per use. Authoring, training, and inference are billed by the minute, with minimum fees and discounts.

dio

Dashboard

Search

SageMaker Domain

Studio

RStudio

Canvas

Images

▶ Ground Truth

▶ Notebook

▶ Processing

▶ Training

▶ Inference

▶ Edge Manager

▶ Augmented AI

▶ AWS Model Registry

Event Engine - Team Dashboard X Amazon SageMaker X +

console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/d-iceiqcsfx9pm

Services Search for services, features, blogs, docs, and more [Option+S]

N. Virginia TeamRole/MasterKey @ 8341-2056-7544

Amazon SageMaker

SageMaker Domain

Users

Add user

Search users

Name Modified on Created on

default-1645016110854	Feb 16, 2022 13:01 UTC	Feb 16, 2022 13:00 UTC	Launch app
-----------------------	------------------------	------------------------	------------

Domain

Status: Ready
The status of the SageMaker Domain, and is not the status of the compute resources such as EC2 instances to execute notebook.

Domain ID: d-iceiqcsfx9pm
Use the SageMaker Domain ID for troubleshooting and tracking usage.

Execution role: arn:aws:iam::834120567544:role/TteamRole

Authentication method: AWS Identity and Access Management (IAM)

How to delete the domain Delete Domain Edit Settings

Use Domain for troubleshooting and tracking usage.
The status shown is for the SageMaker Studio service, and is not the status of compute resources such as EC2 instances to execute notebooks.

Projects

Amazon SageMaker project templates enabled for this account

Dashboard Search SageMaker Domain Studio RStudio Canvas Images Ground Truth Notebook Processing Training Inference Edge Manager Augmented AI AWS Machine Learning

Amazon SageMaker X

Dashboard

Search

SageMaker Domain

Studio

RStudio

Canvas

Images

▶ Ground Truth

▶ Notebook

▶ Processing

▶ Training

▶ Inference

▶ Edge Manager

▶ Augmented AI

▶ AWS Marketplace

Amazon SageMaker > SageMaker Domain

SageMaker Domain

Users

Add user

Search users

< 1 >



Name

Modified on

Created on

default-1645016110854

Feb 16, 2022 13:01 UTC

Feb 16, 2022 13:00 UTC

Launch app ▲

Studio

▼ Domain

How to delete the domain

Delete

Status

Ready

The status of the SageMaker Domain, and is not the status of the compute resources such as EC2 instances to execute notebook.

Domain ID

d-iceiqcsfx9pm

Use the SageMaker Domain ID for troubleshooting and tracking usage.

Execution role

arn:aws:iam::834120567544:role/T
eamRole

A

A

M

M

- Open Link in New Tab
- Open Link in New Window
- Open Link in Incognito Window
- Save Link As...
- Copy Link Address
- Inspect

Use Domain for troubleshooting and tracking usage.

The status shown is for the SageMaker Studio service, and is not the status of compute resources such as EC2 instances to execute notebooks.

Projects

✓ Amazon SageMaker project templates enabled for this account

Launch constraint role: arn:aws:iam::834120567544:role/service-role/AmazonSageMakerServiceCatalogProductsLaunchRole

Product use role: arn:aws:iam::834120567544:role/service-role/AmazonSageMakerServiceCatalogProductsUseRole

✓ Amazon SageMaker project templates enabled for Studio users



Amazon SageMaker Studio

Creating the JupyterServer application default...

Event Engine - Team Dashboard X | Amazon SageMaker X JupyterLab X +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help Feedback

Launcher

Get started

JumpStart models, algorithms, and solutions

SageMaker JumpStart

Solution: Detect malicious users and transactions →

Solution: Demand forecasting →

Go to SageMaker JumpStart →

Build models automatically

SageMaker Autopilot

Video: Get started with Autopilot →

Blog: Getting started with Autopilot →

New autopilot experiment →

Instantly prepare your data

SageMaker Data Wrangler

Blog: Getting started with Data Wrangler →

Blog: Predicting credit risk with SageMaker →

Start now →

ML tasks and components

New compilation job +

Create a new compilation job. [View compilation jobs](#)

New feature group +

Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

New data flow +

Prepare and visualize your data with SageMaker Data Wrangler. [View data flows](#)

New project +

Organize ML components and automate MLOps with built-in or custom project templates. [View projects](#)

New Autopilot experiment +

Create prediction models from your data and start making predictions in a few clicks. [View current experiments](#)

Event Engine - Team Dashboard X | Amazon SageMaker X JupyterLab X +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help Feedback

SageMaker resources Select the resource to view.

Projects Create project

PROJECT 0 rows selected 1/20 filters Search column name to start

Status: Default Clear all

ML tasks and components

New compilation job Create a new compilation job. [View compilation jobs](#)

New data flow Prepare and visualize your data with SageMaker Data Wrangler. [View data flows](#)

New Autopilot experiment Create prediction models from your data and start making predictions in a few clicks. [View current experiments](#)

New feature group Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

New project Organize ML components and automate MLOps with built-in or custom project templates. [View projects](#)

Launcher

Get started

JumpStart models, algorithms, and solutions

SageMaker JumpStart

Solution: Detect malicious users and transactions →

Solution: Demand forecasting →

Go to SageMaker JumpStart →

Build models automatically

SageMaker Autopilot

Video: Get started with Autopilot →

Blog: Getting started with Autopilot →

New autopilot experiment →

Instantly prepare

SageMaker Data Wrangler

Blog: Getting started with Data Wrangler →

Blog: Predicting credit risk with SageMaker →

Start now →

Event Engine - Team Dashboard X | Amazon SageMaker X | JupyterLab X +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help Feedback

SageMaker resources Select the resource to view.

Projects Data Wrangler Feature Store Pipelines Experiments and trials Model registry Endpoints Compilation Jobs Inference Recommender Jobs

There are no Projects yet. Create a Project using the SageMaker SDK and track your work automatically.

Launcher

Get started

JumpStart models, algorithms, and solutions

SageMaker JumpStart

Solution: Detect malicious users and transactions →

Solution: Demand forecasting →

Go to SageMaker JumpStart →

Build models automatically

SageMaker Autopilot

Video: Get started with Autopilot →

Blog: Getting started with Autopilot →

New autopilot experiment →

Instantly prepare

SageMaker Data Wrangler

Blog: Getting started with Data Wrangler →

Blog: Predicting credit risk with SageMaker →

Start now →

ML tasks and components

New compilation job +

Create a new compilation job. [View compilation jobs](#)

New data flow +

Prepare and visualize your data with SageMaker Data Wrangler. [View data flows](#)

New Autopilot experiment +

Create prediction models from your data and start making predictions in a few clicks. [View current experiments](#)

New feature group +

Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

New project +

Organize ML components and automate MLOps with built-in or custom project templates. [View projects](#)

Event Engine - Team Dashboard X | Amazon SageMaker X JupyterLab X +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help Feedback

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

Launcher

Get started

JumpStart models, algorithms, and solutions

SageMaker JumpStart

Solution: Detect malicious users and transactions →

Solution: Demand forecasting →

Go to SageMaker JumpStart →

Build models automatically

SageMaker Autopilot

Video: Get started with Autopilot →

Blog: Getting started with Autopilot →

New autopilot experiment →

Instantly prepare

SageMaker Data Wrangler

Blog: Getting started with Data Wrangler →

Blog: Predicting credit risk with SageMaker →

Start now →

ML tasks and components

New compilation job +

Create a new compilation job. [View compilation jobs](#)

New data flow +

Prepare and visualize your data with SageMaker Data Wrangler. [View data flows](#)

New Autopilot experiment +

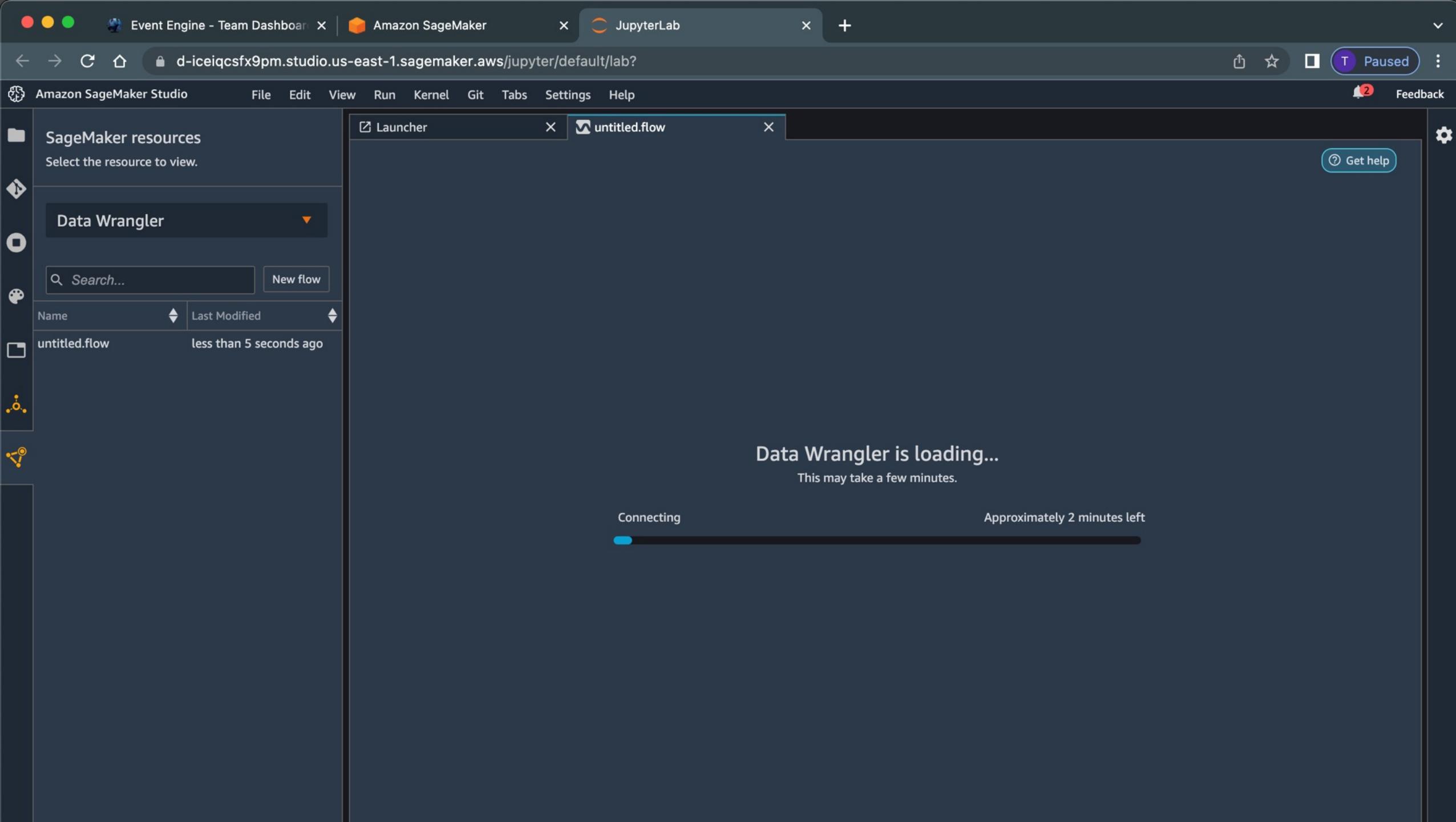
Create prediction models from your data and start making predictions in a few clicks. [View current experiments](#)

New feature group +

Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

New project +

Organize ML components and automate MLOps with built-in or custom project templates. [View projects](#)



Event Engine - Team Dashboard X | Amazon SageMaker X JupyterLab X +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Launcher untitled.flow

Import Data Flow

Get help

Add data source ▾

Import data

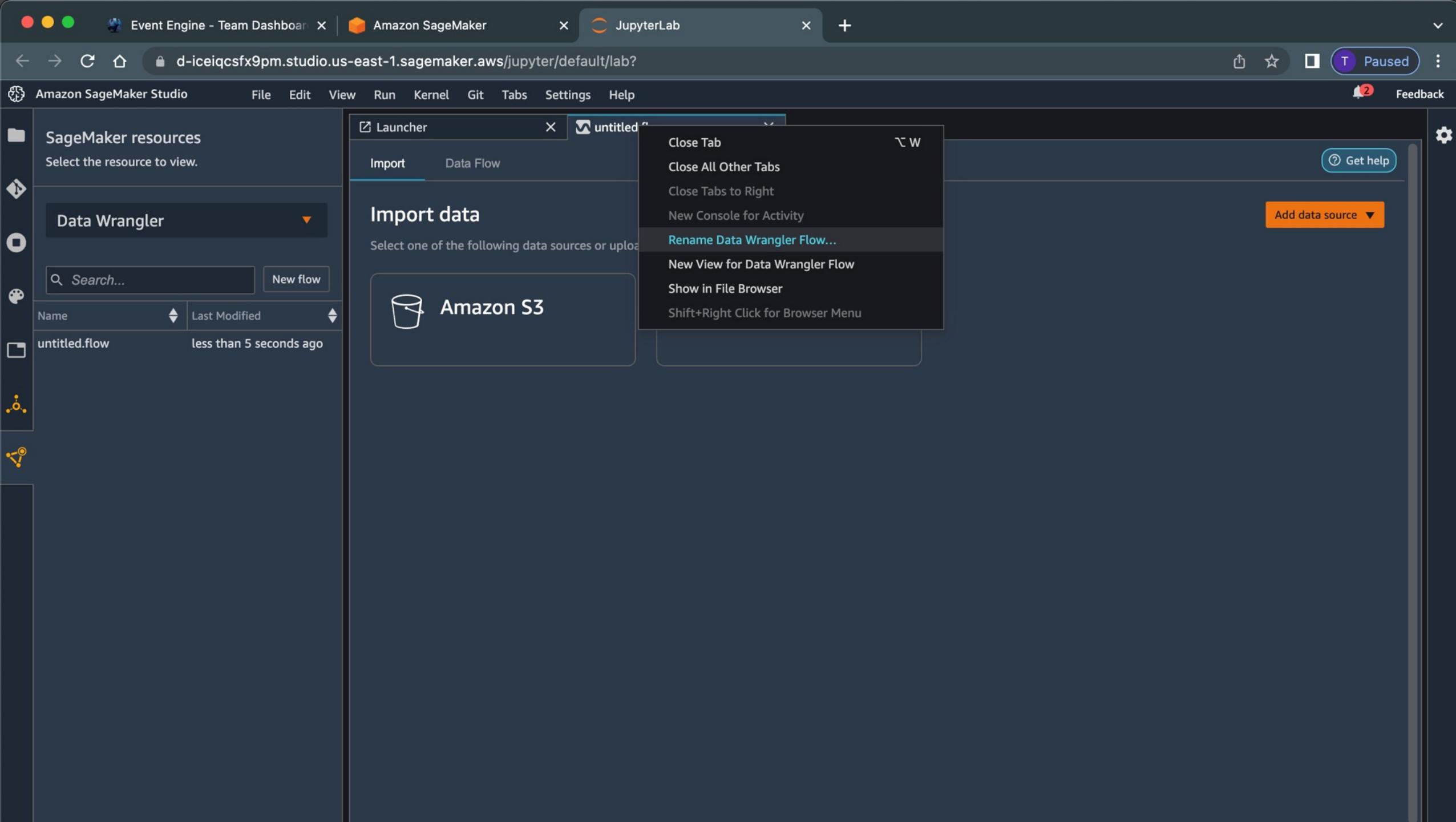
Select one of the following data sources or upload a file to import a dataset.

Amazon S3

Amazon Athena

Feedback

Paused



Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Feedback

SageMaker resources
Select the resource to view.

Data Wrangler

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import Data Flow

Get help

Add data source ▾

Launcher untitled.flow

Import Data Flow

Import data
Select one of the following data sources or upload a file to import a dataset.

Amazon S3 Amazon Athena

Rename File

File Path
untitled.flow

New Name
TitanicPrep.flow

Cancel Rename

3 minutes ago

A small modal dialog box titled "Rename File" is displayed in the center. It contains a "File Path" field with the value "untitled.flow" and a "New Name" field with the value "TitanicPrep.flow". At the bottom are two buttons: "Cancel" and "Rename".

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher  TitanicPrep.flow

Import Data Flow

Get help

Import data

Select one of the following data sources or upload a file to import a dataset.

Add data source ▾

 Amazon S3

 Amazon Athena

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher X TitanicPrep.flow X

Back to Import Get help

Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 /

Bucket name	Region	Creation date
sagemaker-studio-834120567544-5zbey54q71a	us-east-1	2022-02-16 12:55:19+00:00
sagemaker-us-east-1-834120567544	us-east-1	2022-02-16 13:06:52+00:00
titanic-michlin-20220218	us-east-1	2022-02-16 14:15:58+00:00

Event Engine - Team Dashboard X | Amazon SageMaker X | titanic-michlin-20220218 - S3 X | JupyterLab X | Introducing Amazon SageMake X | +

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher X TitanicPrep.flow X

Back to Import Get help

Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 / titanic-michlin-20220218

Object name	Size	Last modified
train.csv	59.76KB	2022-02-16 14:16:59+00:00

Previous Displaying 1 - 1 Next

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher ✎ TitanicPrep.flow

Back to Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 / titanic-michlin-20220218

Object name	Size	Last modified
train.csv	59.76KB	2022-02-16 14:16:59+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (first 100 rows shown)

PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

DETAILS

Name Preview on
train.csv

File type csv

First row is header

Import nested directories

Advanced configuration

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher ✎ TitanicPrep.flow

Back to Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 / titanic-michlin-20220218

Object name	Size	Last modified
train.csv	59.76KB	2022-02-16 14:16:59+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (first 100 rows shown)

PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

DETAILS

Name Preview on

File type

First row is header

Import nested directories

Advanced configuration

Delimiter

Enable sampling

Filename as separate column

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher X TitanicPrep.flow X

Back to Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 / titanic-michlin-20220218

Object name	Size	Last modified
train.csv	59.76KB	2022-02-16 14:16:59+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (first 100 rows shown)

PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

DETAILS Get help

Name Preview on train.csv

File type csv

First row is header

Import nested directories

Advanced configuration

Delimiter COMMA

Enable sampling

Filename as separate column

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

TitanicPrep.flow 9 minutes ago

Launcher ✎ TitanicPrep.flow

Back to Import

Import a dataset from S3

Enter the S3 URL of a file or prefix (folder) in the text box, or use the following table to browse S3

Enter an S3 URL Go

S3 / titanic-michlin-20220218

Object name	Size	Last modified
train.csv	59.76KB	2022-02-16 14:16:59+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (first 100 rows shown)

PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

DETAILS

Name Preview on

File type

First row is header

Import nested directories

Advanced configuration

Delimiter

Enable sampling

Filename as separate column

Get help



SageMaker resources
Select the resource to view.

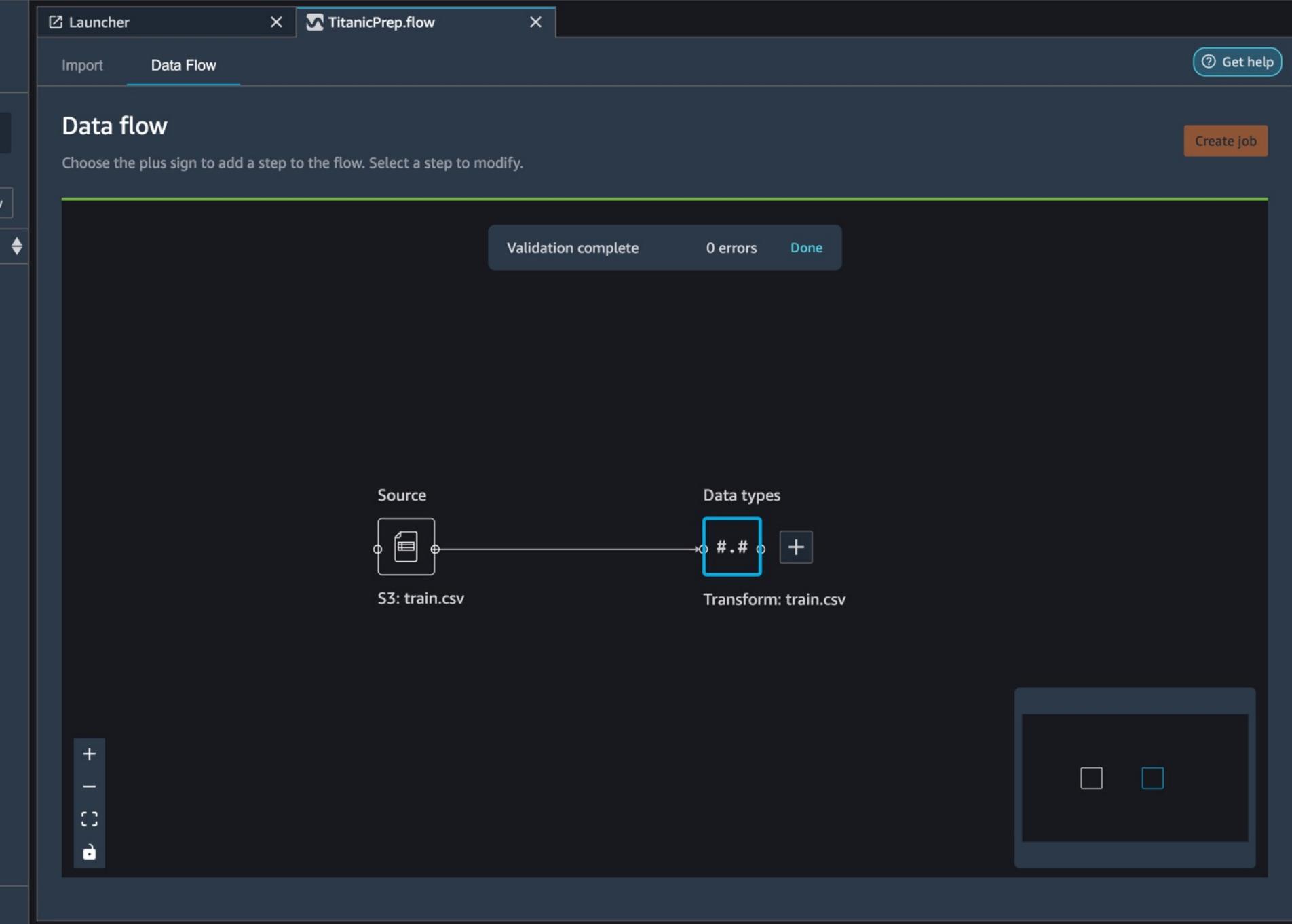
Data Wrangler ▾

Search... New flow

Name	Last Modified
TitanicPrep.flow	9 minutes ago

+

7 minutes ago



Launcher



TitanicPrep.flow



Import Data Flow

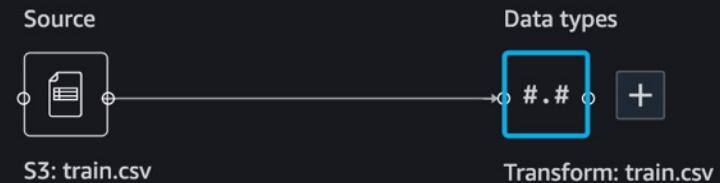
Get help

Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete 0 errors Done



Launcher X TitanicPrep.flow X

Import

Data Flow

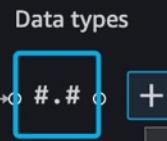
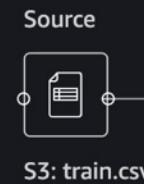
Get help

Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete 0 errors Done



Edit data types

Add transform

Add analysis

Add destination >

Export to >

Join

Concatenate



Launcher

TitanicPrep.flow

Get help

[Back to data flow](#)

Data types · Transform: train.csv

Data

Analysis

Step 2. Data types

Export data

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

CONFIGURE TYPES

Column name	Type
PassengerId	Long
Survived	Long
Pclass	Long
Name	String
Sex	String
Age	Long
SibSp	Long
Parch	Long
Ticket	String
Fare	Float
Cabin	String
Embarked	String
Clear	
Preview	
Apply	

Launcher



TitanicPrep.flow



Import

Data Flow

Get help



Data flow

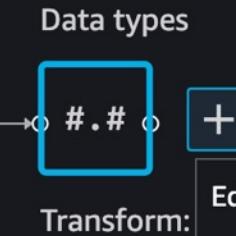
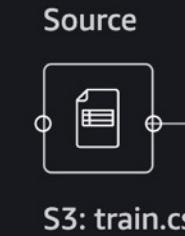
Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete

0 errors

Done



- [Edit data types](#)
- [Add transform](#)
- [Add analysis](#)
- [Add destination >](#)
- [Export to >](#)
- [Join](#)
- [Concatenate](#)

+
-
[]
[]

← → ⌂ ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Feedback

Launcher X TitanicPrep.flow X

Back to data flow Get help

Data types · Transform: train.csv

Data Analysis

Histogram: Untitled

No Preview available

Use Configure for built-in analyses

Use Code to create a custom analysis

Back to all analyses Create analysis

Analysis type

Histogram

Bias Report

Custom Visualization

Duplicate rows

Feature Correlation

Histogram

Multicollinearity

Quick Model

Scatter Plot

Table Summary

Select...

Optional

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	1	Palsson, Master. Gosta	male	

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Data Analysis

Table Summary: Summary

summary	PassengerId	Survived	Pclass	Name	Sex
count	891	891	891	891	891
mean	446.0	0.3838383838383838	2.308641975308642	None	None
stddev	257.3538420152301	0.48659245426485753	0.8360712409770491	None	None
min	1	0	1	Abbing, Mr. Anthony	female
max	891	1	3	van Melkebeke, Mr. Phil...	male

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Lundstrom, Mr. Gustaf ...	male	57

Back to all analyses

Create analysis

Analysis type

Table Summary

A limit of 100,000 rows is used for this analysis.

Analysis name

Summary

Optional

Launcher



TitanicPrep.flow



Data

Analysis

Table Summary: Summary

summary	PassengerId	Survived	Pclass	Name	Sex
count	891	891	891	891	891
mean	446.0	0.3838383838383838	2.308641975308642	None	None
stddev	257.3538420152301	0.48659245426485753	0.8360712409770491	None	None
min	1	0	1	Abbing, Mr. Anthony	female
max	891	1	3	van Melkebeke, Mr. Phil...	male

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14

[Back to all analyses](#)

Create analysis

Analysis type

Table Summary

A limit of 100,000 rows is used for this analysis.

Analysis name

Summary

Optional

Clear

Preview

Save

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow ●

Back to data flow Get help

Data types · Transform: train.csv

Data Analysis Create new analysis

Summary Table Summary

The screenshot shows the Amazon SageMaker Studio interface. At the top, there are several tabs: Event Engine - Team Dashboard, Amazon SageMaker, titanic-michlin-20220218 - S3, JupyterLab, and Introducing Amazon SageMake. Below the tabs is a navigation bar with icons for back, forward, search, and other functions. The main workspace is titled 'Launcher' and contains a tab for 'TitanicPrep.flow'. The current view is 'Data types · Transform: train.csv'. There are two tabs: 'Data' (disabled) and 'Analysis' (selected). A large icon of a bar chart is centered in the workspace. In the bottom left corner, there is a summary section labeled 'Summary' and 'Table Summary'. On the far left, there is a vertical sidebar with various icons representing different data types and analysis tools. The overall theme is dark blue.

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow

Data types · Transform: train.csv

Data Analysis

Histogram: Sex

No Preview available

Use Configure for built-in analyses

Use Code to create a custom analysis

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta	male	

Back to all analyses

Create analysis

Analysis type: Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name: Sex

X axis: Sex

Color by: Survived

Facet by: Select...

Launcher

TitanicPrep.flow

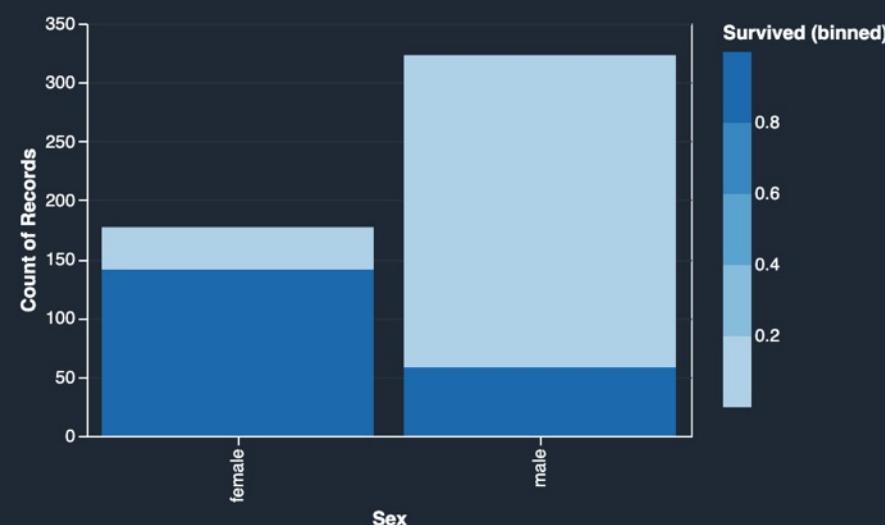
X

Data types · Transform: train.csv

Data

Analysis

Histogram: Sex



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	1	Moran, Mr. James	male	30

Back to all analyses

Create analysis

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Sex

Optional

X axis

Sex

X | ▾

Color by

Survived

X | ▾

Optional

Facet by

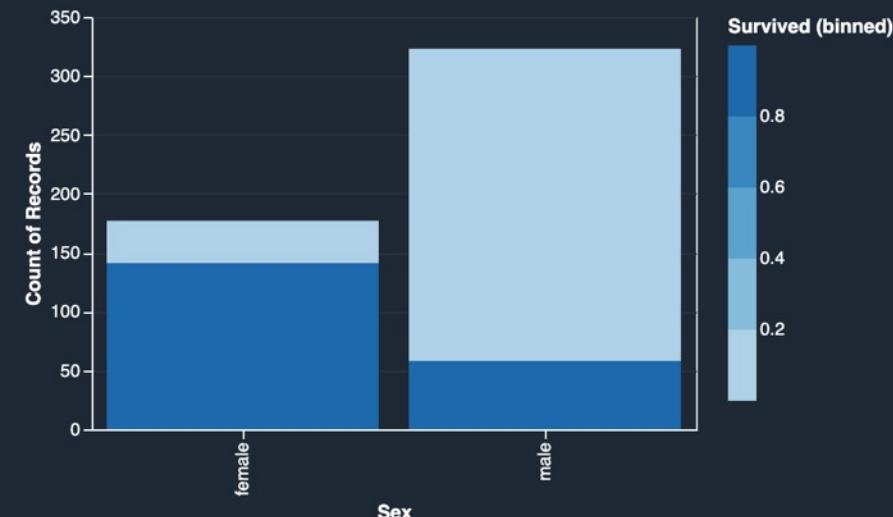
Select...

X | ▾

Optional

Launcher

TitanicPrep.flow



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Sex

Optional

X axis

Sex

Color by

Survived

Optional

Facet by

Select...

Optional

Preview

Save

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow ●

Back to data flow Get help

Data types · Transform: train.csv

Data Analysis Create new analysis

 Summary
Table Summary

 Sex
Histogram
Sex · Survived

Launcher

X

TitanicPrep.flow

X



Data types · Transform: train.csv

Data

Analysis

Histogram: Pclass



No Preview available

Use Configure for built-in analyses

Use Code to create a custom analysis

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27

Back to all analyses

Create analysis

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Pclass

Optional

X axis

Pclass

Color by

Survived

Optional

Facet by

Select...

Optional

Clear

Preview

Save

Launcher

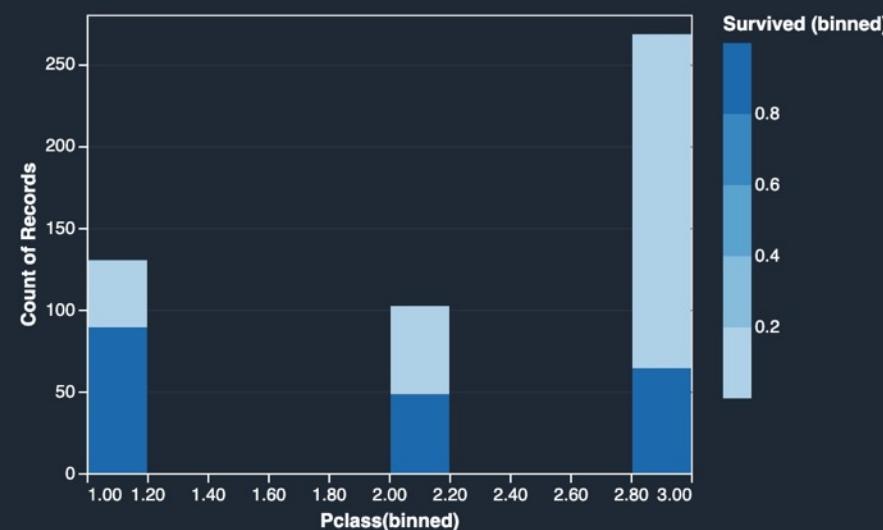
TitanicPrep.flow

Data types · Transform: train.csv

Data

Analysis

Histogram: Pclass



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	29

[Back to all analyses](#)

Create analysis

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Pclass

Optional

X axis

Pclass

Color by

Survived

Optional

Facet by

Select...

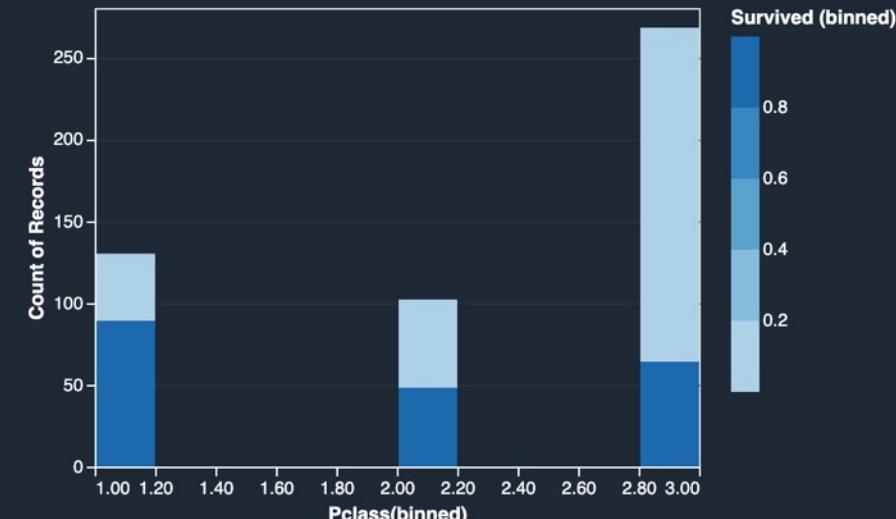
Optional

Launcher

X

TitanicPrep.flow

X



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Pclass

Optional

X axis

Pclass



Color by

Survived



Optional

Facet by

Select...



Optional

Clear

Preview

Save

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help Feedback

Launcher **TitanicPrep.flow** Back to data flow

Data types · Transform: train.csv

Data Analysis

Histogram: Parch

No Preview available

Use Configure for built-in analyses

Use Code to create a custom analysis

Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gusta...	male	2

Back to all analyses Create analysis

Analysis type: Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name: Parch

X axis: Parch

Color by: Survived

Facet by: Select...

Launcher

TitanicPrep.flow

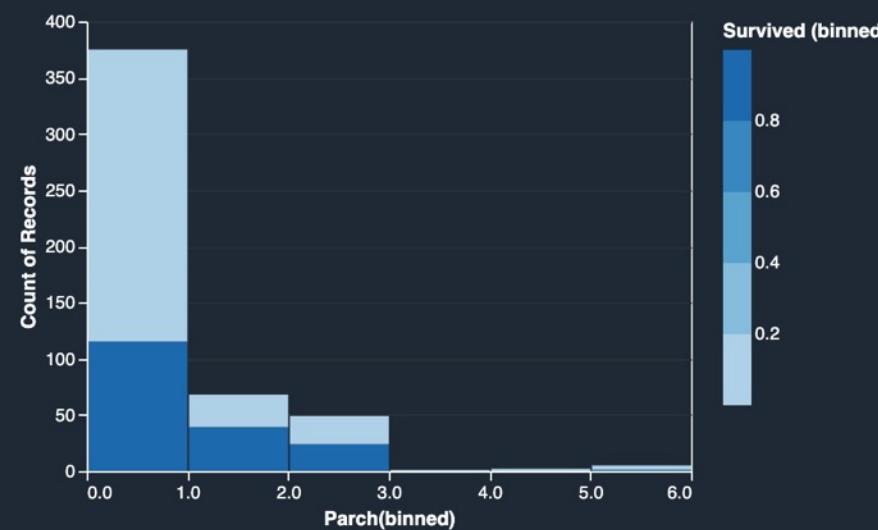
[Back to data flow](#)

Data types · Transform: train.csv

Data

Analysis

Histogram: Parch



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	1	Moran, Mr. James	male	34

[Back to all analyses](#)

Create analysis

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Parch

Optional

X axis

Parch



Color by

Survived



Optional

Facet by

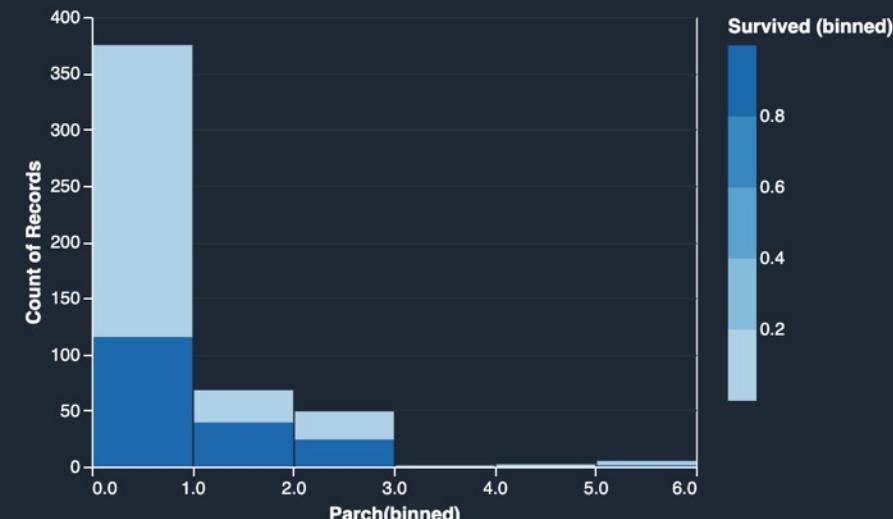
Select...



Optional

Launcher

TitanicPrep.flow



Data table

PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20

Analysis type

Histogram

A limit of 100,000 rows is used for this analysis.

Analysis name

Parch

Optional

X axis

Parch

Color by

Survived

Optional

Facet by

Select...

Optional

Preview

Save

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Feedback

Launcher X TitanicPrep.flow X

Import Data Flow

Get help

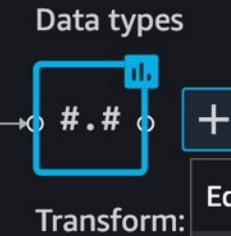
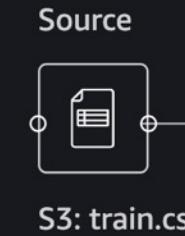


Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete 0 errors Done



- + Edit data types
- + Add transform
- Add analysis
- Add destination >
- Export to >
- Join
- Concatenate

← → ⌂ ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher ✎ TitanicPrep.flow X

Back to data flow Get help

Data types · Transform: train.csv

Data Analysis

Step 2. Data types

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	

Export data

TRANSFORMS X

+ Add step

▶ 1. S3 Source

▶ 2. Data types

← → ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab? ⌃ ⌂ ⌂ Paused ⌂

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X Back to data flow Get help

Data types · Transform: train.csv

Data Analysis

Step 2. Data types

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

Export data

ADD TRANSFORM

- Featurize date/time
- Featurize text
- Format string
- Group by
- Handle missing
- Handle outliers
- Handle structured column
- Manage columns
- Manage rows

← → ⌂ ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow

Data types · Transform: train.csv

Data Analysis

Step 2. Data types

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

Export data

HANDLE MISSING

Replace, drop, or add indicators for missing values. [Learn more.](#)

Transform i

Impute

Column type i

Numeric

Input column

Age

Imputing strategy i

Approximate Median

Output column i

Age_Imputed

Optional

Clear

Preview Add

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X Back to data flow Get help

Data types · Transform: train.csv

Data Analysis

Previewing: Handle missing

h (long)	Ticket (string)	Fare (float)	Cabin (string)	Embarked (string)	Age_Imputed (float)
A/5 21171	7.25			S	22
PC 17599	71.2833		C85	C	38
STON/O2. 3101282	7.925			S	26
113803	53.1		C123	S	35
373450	8.05			S	35
330877	8.4583			Q	28
17463	51.8625	E46		S	54
349909	21.075			S	2
347742	11.1333			S	27
237736	30.0708			C	14
PP 9549	16.7	G6		S	4
113783	26.55	C103		S	58
A/5. 2151	8.05			S	20
347082	31.275			S	39
350406	7.8542			S	14
248706	16			S	55
382652	29.125			Q	2
244373	13			S	28
345763	18			S	31
2649	7.225			C	28

Export data

HANDLE MISSING

Replace, drop, or add indicators for missing values. [Learn more.](#)

Transform *i*

Impute

Column type *i*

Numeric

Input column

Age

Imputing strategy *i*

Approximate Median

Output column *i*

Age_Imputed

Optional

Clear Preview Add

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Feedback

Launcher X TitanicPrep.flow

Back to data flow Get help

Impute · Transform: train.csv

Data Analysis

Step 3. Impute

Export data

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	

TRANSFORMS

+ Add step

- ▶ 1. S3 Source
- ▶ 2. Data types
- ▶ 3. Impute

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow ●

[Back to data flow](#) (?) Get help ⚙️

Impute · Transform: train.csv

Data Analysis

Step 3. Impute

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (...	female	27
10	1	2	Nasser, Mrs. Nicholas (A...	female	14
11	1	3	Sandstrom, Miss. Margu...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35
22	1	2	Browley, Mr. Lawrence	male	

Export data

ADD TRANSFORM X

- Featurize date/time**
Encode date/time values to numeric and vector representations. [Learn more...](#)
- Featurize text**
Generate vector representations from natural language text. [Learn more...](#)
- Format string**
Clean and prepare strings using standard string formatting operations. [Learn more...](#)
- Group by**
Add an aggregated column after group by as a new column.
- Handle missing**
Replace, drop, or add indicators for missing values. [Learn more.](#)
- Handle outliers**
Remove or replace outlier numeric and categorical values. [Learn more.](#)
- Handle structured column**
Flatten JSON and perform other operations on structured data
- Manage columns**
Move, drop, duplicate or rename columns in the dataset. [Learn more.](#)
- Manage rows**
Sort, shuffle or drop duplicate rows.
- Manage vectors**
Expand or create vector columns. [Learn more.](#)

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow

Impute · Transform: train.csv

Data Analysis

Previewing: Manage columns

(long)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)
1	0		7.25	S	22
1	0		71.2833	C	38
0	0		7.925	S	26
1	0		53.1	S	35
0	0		8.05	S	35
0	0		8.4583	Q	28
0	0		51.8625	S	54
3	1		21.075	S	2
0	2		11.1333	S	27
1	0		30.0708	C	14
1	1		16.7	S	4
0	0		26.55	S	58
0	0		8.05	S	20
1	5		31.275	S	39
0	0		7.8542	S	14
0	0		16	S	55
4	1		29.125	Q	2
0	0		13	S	28
1	0		18	S	31
0	0		7.225	C	28

Export data

MANAGE COLUMNS

Move, drop, duplicate or rename columns in the dataset. [Learn more.](#)

Transform *i*

Drop column

Columns to drop

PassengerId X Name X Cabin X Ticket X

Clear Preview Add

Amazon SageMaker Studio File Edit View Run Kernel Tabs Settings Help

Launcher X TitanicPrep.flow ●

Back to data flow Get help

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

Survived (long)	Pclass (long)	Sex (string)	Age (long)	SibSp (long)	Parch (long)
0	3	male	22	1	0
1	1	female	38	1	0
1	3	female	26	0	0
1	1	female	35	1	0
0	3	male	35	0	0
0	3	male		0	0
0	1	male	54	0	0
0	3	male	2	3	1
1	3	female	27	0	2
1	2	female	14	1	0
1	3	female	4	1	1
1	1	female	58	0	0
0	3	male	20	0	0
0	3	male	39	1	5
0	3	female	14	0	0
1	2	female	55	0	0
0	3	male	2	4	1
1	2	male		0	0
0	3	female	31	1	0
1	3	female		0	0

Export data

TRANSFORMS

+ Add step

- ▶ 1. S3 Source
- ▶ 2. Data types
- ▶ 3. Impute
- ▶ 4. Drop column

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow Get help

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

Survived (long)	Pclass (long)	Sex (string)	Age (long)	SibSp (long)	Parch (long)
0	3	male	22	1	0
1	1	female	38	1	0
1	3	female	26	0	0
1	1	female	35	1	0
0	3	male	35	0	0
0	3	male		0	0
0	1	male	54	0	0
0	3	male	2	3	1
1	3	female	27	0	2
1	2	female	14	1	0
1	3	female	4	1	1
1	1	female	58	0	0
0	3	male	20	0	0
0	3	male	39	1	5
0	3	female	14	0	0
1	2	female	55	0	0
0	3	male	2	4	1
1	2	male		0	0
0	3	female	31	1	0

Export data

ADD TRANSFORM

- Custom transform Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations...
- Balance data Balance the data for binary classification problems using random oversampling...
- Custom formula Define a new column using a Spark SQL expression to query data in the ...
- Encode categorical Convert categorical variables to numeric or vector representations. Learn more...
- Featurize date/time Encode date/time values to numeric and vector representations. Learn more...
- Featurize text Generate vector representations from natural language text. Learn more...
- Format string Clean and prepare strings using standard string formatting operations. Learn more...
- Group by Add an aggregated column after group by as a new column.
- Handle missing Replace, drop, or add indicators for missing values. Learn more...

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow Get help

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

Survived (long)	Pclass (long)	Sex (string)	Age (long)	SibSp (long)	Parch (long)
0	3	male	22	1	0
1	1	female	38	1	0
1	3	female	26	0	0
1	1	female	35	1	0
0	3	male	35	0	0
0	3	male		0	0
0	1	male	54	0	0
0	3	male	2	3	1
1	3	female	27	0	2
1	2	female	14	1	0
1	3	female	4	1	1
1	1	female	58	0	0
0	3	male	20	0	0
0	3	male	39	1	5
0	3	female	14	0	0
1	2	female	55	0	0
0	3	male	2	4	1
1	2	male		0	0
0	3	female	31	1	0

Export data

CUSTOM PYSPARK X

Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations. [Learn more](#).

|Python (PySpark)

Python (PySpark)

Python (Pandas)

SQL (PySpark SQL)

← → C ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow Get help

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

	Survived (long)	Pclass (long)	Sex (string)	Age (long)	SibSp (long)	Parch (long)	Export data
0	3	male	22	1	0		
1	1	female	38	1	0		
1	3	female	26	0	0		
1	1	female	35	1	0		
0	3	male	35	0	0		
0	3	male		0	0		
0	1	male	54	0	0		
0	3	male	2	3	1		
1	3	female	27	0	2		
1	2	female	14	1	0		
1	3	female	4	1	1		
1	1	female	58	0	0		
0	3	male	20	0	0		
0	3	male	39	1	5		
0	3	female	14	0	0		
1	2	female	55	0	0		
0	3	male	2	4	1		
1	2	male		0	0		
0	3	female	31	1	0		

CUSTOM PANDAS

Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations. [Learn more.](#)

Using Python (Pandas) requires your dataset to fit in memory and only uses a single instance in batch computation. It is ideal for smaller datasets less than 2GB and experimentation but we recommend Python (PySpark) for production use-cases

Python (Pandas)

```
1 df["With_Family"] = ( (df["SibSp"] != 0) | (df["Parch"] != 0) ).astype(int)
```

Clear Preview Add

d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow

Drop column · Transform: train.csv

Data Analysis

Previewing: Python (Pandas)

P (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)	With_Family (long)
0	7.25	S	22	1	
0	71.2833	C	38	1	
0	7.925	S	26	0	
0	53.1	S	35	1	
0	8.05	S	35	0	
0	8.4583	Q	28	0	
0	51.8625	S	54	0	
1	21.075	S	2	1	
2	11.1333	S	27	1	
0	30.0708	C	14	1	
1	16.7	S	4	1	
0	26.55	S	58	0	
0	8.05	S	20	0	
5	31.275	S	39	1	
0	7.8542	S	14	0	
0	16	S	55	0	
1	29.125	Q	2	1	
0	13	S	28	0	
0	18	S	31	1	
0	7.225	C	28	0	

Export data

CUSTOM PANDAS

Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations. [Learn more.](#)

Using Python (Pandas) requires your dataset to fit in memory and only uses a single instance in batch computation. It is ideal for smaller datasets less than 2GB and experimentation but we recommend Python (PySpark) for production use-cases

Python (Pandas)

```
1 df["With_Family"] = ( (df["SibSp"] !=0 ) | (df["Parch"] !=0 ) ).astype(int)
```

Clear Preview Add

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Feedback

Launcher TitanicPrep.flow

Back to data flow Get help

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

p (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)	With_Family (long)
0	7.25	S	22	1	
0	71.2833	C	38	1	
0	7.925	S	26	0	
0	53.1	S	35	1	
0	8.05	S	35	0	
0	8.4583	Q	28	0	
0	51.8625	S	54	0	
1	21.075	S	2	1	
2	11.1333	S	27	1	
0	30.0708	C	14	1	
1	16.7	S	4	1	
0	26.55	S	58	0	
0	8.05	S	20	0	
5	31.275	S	39	1	
0	7.8542	S	14	0	
0	16	S	55	0	
1	29.125	Q	2	1	
0	13	S	28	0	
0	18	S	31	1	
0	7.225	C	28	0	
0	26	S	35	0	

Export data

TRANSFORMS

+ Add step

- ▶ 1. S3 Source
- ▶ 2. Data types
- ▶ 3. Impute
- ▶ 4. Drop column
- ▶ 5. Custom Pandas

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher TitanicPrep.flow

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

p (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)	With_Family (long)
0		7.25	S	22	1
0		71.2833	C	38	1
0		7.925	S	26	0
0		53.1	S	35	1
0		8.05	S	35	0
0		8.4583	Q	28	0
0		51.8625	S	54	0
1		21.075	S	2	1
2		11.1333	S	27	1
0		30.0708	C	14	1
1		16.7	S	4	1
0		26.55	S	58	0
0		8.05	S	20	0
5		31.275	S	39	1
0		7.8542	S	14	0
0		16	S	55	0
1		29.125	Q	2	1
0		13	S	28	0
0		18	S	31	1
0		7.225	C	28	0

Export data

ADD TRANSFORM

- Custom transform
- Balance data
- Custom formula
- Encode categorical
- Featurize date/time
- Featurize text
- Format string
- Group by
- Handle missing

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow (?) Get help

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

P (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)	With_Family (long)
0	7.25	S	22	1	
0	71.2833	C	38	1	
0	7.925	S	26	0	
0	53.1	S	35	1	
0	8.05	S	35	0	
0	8.4583	Q	28	0	
0	51.8625	S	54	0	
1	21.075	S	2	1	
2	11.1333	S	27	1	
0	30.0708	C	14	1	
1	16.7	S	4	1	
0	26.55	S	58	0	
0	8.05	S	20	0	
5	31.275	S	39	1	
0	7.8542	S	14	0	
0	16	S	55	0	
1	29.125	Q	2	1	
0	13	S	28	0	
0	18	S	31	1	
0	7.225	C	28	0	
0	26	S	35	0	

Export data

ENCODE CATEGORICAL X

Convert categorical variables to numeric or vector representations. [Learn more.](#)

Transform i

|Ordinal encode X | ▾

Ordinal encode

One-hot encode

Similarity encode

Optional

Invalid handling strategy i

Replace with NaN X | ▾

Clear Preview Add

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X Get help Feedback

Back to data flow

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

Survived (long)	Pclass (long)	Sex (string)	Age (float)	SibSp (long)	Parch (long)
0	3	male	22	1	0
1	1	female	38	1	0
1	3	female	26	0	0
1	1	female	35	1	0
0	3	male	35	0	0
0	3	male		0	0
0	1	male	54	0	0
0	3	male	2	3	1
1	3	female	27	0	2
1	2	female	14	1	0
1	3	female	4	1	1
1	1	female	58	0	0
0	3	male	20	0	0
0	3	male	39	1	5
0	3	female	14	0	0
1	2	female	55	0	0
0	3	male	2	4	1
1	2	male		0	0
0	3	female	31	1	0
1	3	female		0	0
0	2	male	35	0	0

Export data

ENCODE CATEGORICAL

Convert categorical variables to numeric or vector representations. [Learn more](#)

Transform i

One-hot encode

Input columns i

Pclass X

Input already ordinal encoded i

Invalid handling strategy i

Skip

Drop last i

Output style i

Columns

Output column i

Optional

Clear Preview Add

← → ⌂ ⌂ 🔒 d-iceiqcsfx9pm.studio.us-east-1.sagemaker.aws/jupyter/default/lab?

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Custom Pandas · Transform: train.csv

Data Analysis

Previewing: Encode categorical

Survived (string)	Age_Imputed (float)	With_Family (long)	Pclass_3 (float)	Pclass_1 (float)	Pclass_2 (float)
22	1	1	0	0	1
38	1	0	1	0	0
26	0	1	0	1	0
35	1	0	1	0	0
35	0	1	0	0	1
28	0	1	0	1	0
54	0	0	1	0	0
2	1	1	0	0	1
27	1	1	0	1	0
14	1	0	0	0	1
4	1	1	0	1	0
58	0	0	1	0	0
20	0	1	0	0	1
39	1	1	0	1	0
14	0	1	0	0	1
55	0	0	1	0	0
2	1	1	0	0	1
28	0	0	1	0	0
31	1	1	0	0	1
28	0	1	0	0	1
35	0	0	1	0	0

Export data

ENCODE CATEGORICAL

Convert categorical variables to numeric or vector representations. [Learn more.](#)

Transform [i](#)

One-hot encode

Input columns [i](#)

Pclass X

Input already ordinal encoded [i](#)

Invalid handling strategy [i](#)

Skip

Drop last [i](#)

Output style [i](#)

Columns

Output column [i](#)

Optional

Clear Preview Add

Back to data flow

One-hot encode · Transform: train.csv

Data Analysis

Step 6. One-hot encode

Survived (long)	Sex (string)	Age (float)	SibSp (long)	Parch (long)	Fare (float)
0	male	22	1	0	7.25
1	female	38	1	0	71.2833
1	female	26	0	0	7.925
1	female	35	1	0	53.1
0	male	35	0	0	8.05
0	male		0	0	8.4583
0	male	54	0	0	51.8625
0	male	2	3	1	21.075
1	female	27	0	2	11.1333
1	female	14	1	0	30.0708
1	female	4	1	1	16.7
1	female	58	0	0	26.55
0	male	20	0	0	8.05
0	male	39	1	5	31.275
0	female	14	0	0	7.8542
1	female	55	0	0	16
0	male	2	4	1	29.125
1	male		0	0	13
0	female	31	1	0	18
1	female		0	0	7.225
0	male	35	0	0	26

Export data

ENCODE CATEGORICAL

Convert categorical variables to numeric or vector representations. [Learn more.](#)

Transform [i](#)

One-hot encode

Input columns [i](#)

Sex [x](#)

Input already ordinal encoded [i](#)

Invalid handling strategy [i](#)

Skip

Drop last [i](#)

Output style [i](#)

Columns

Output column [i](#)

Optional

Clear Preview Add

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow •

Back to data flow Get help

One-hot encode · Transform: train.csv

Data Analysis

Previewing: Encode categorical

1_Family (long)	Pclass_3 (float)	Pclass_1 (float)	Pclass_2 (float)	Sex_male (float)	Sex_female (float)
1	0			1	
0		1		0	
1		0		0	
0		1		0	
1		0		1	
1		0		1	
0		1		1	
1		0		1	
1		0		0	
0		0		0	
1		0		0	
0		1		0	
1		0		1	
1		0		1	
1		0		0	
0		0		0	
1		0		1	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0		1	
0		0		0	
1		0		1	
1		0		0	
0		0		1	
1		0		0	
1		0			

Launcher X TitanicPrep.flow ●

[Back to data flow](#)[Get help](#)

One-hot encode · Transform: train.csv

Data

Analysis

Step 7. One-hot encode

[Export data](#)

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)
0	22	1	0	7.25	S
1	38	1	0	71.2833	C
1	26	0	0	7.925	S
1	35	1	0	53.1	S
0	35	0	0	8.05	S
0		0	0	8.4583	Q
0	54	0	0	51.8625	S
0	2	3	1	21.075	S
1	27	0	2	11.1333	S
1	14	1	0	30.0708	C
1	4	1	1	16.7	S
1	58	0	0	26.55	S
0	20	0	0	8.05	S
0	39	1	5	31.275	S
0	14	0	0	7.8542	S
1	55	0	0	16	S
0	2	4	1	29.125	Q
1		0	0	13	S
0	31	1	0	18	S
1		0	0	7.225	C
0	35	0	0	26	S

TRANSFORMS X

[+ Add step](#)

▶ 1. S3 Source

▶ 2. Data types

▶ 3. Impute

▶ 4. Drop column

▶ 5. Custom Pandas

▶ 6. One-hot encode

▶ 7. One-hot encode

Amazon SageMaker Studio File Edit View Run Kernel Tabs Settings Help Feedback

Launcher TitanicPrep.flow

Back to data flow

One-hot encode · Transform: train.csv

Data Analysis

Step 7. One-hot encode

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)
0	22	1	0	7.25	S
1	38	1	0	71.2833	C
1	26	0	0	7.925	S
1	35	1	0	53.1	S
0	35	0	0	8.05	S
0		0	0	8.4583	Q
0	54	0	0	51.8625	S
0	2	3	1	21.075	S
1	27	0	2	11.1333	S
1	14	1	0	30.0708	C
1	4	1	1	16.7	S
1	58	0	0	26.55	S
0	20	0	0	8.05	S
0	39	1	5	31.275	S
0	14	0	0	7.8542	S
1	55	0	0	16	S
0	2	4	1	29.125	Q
1		0	0	13	S
0	31	1	0	18	S
1		0	0	7.225	C

Export data

ADD TRANSFORM

- Custom transform
- Balance data
- Custom formula
- Encode categorical
- Featurize date/time
- Featurize text
- Format string
- Group by
- Handle missing

Amazon SageMaker Studio File Edit View Run Kernel Git Tabs Settings Help

Launcher X TitanicPrep.flow X

Back to data flow (?) Get help

One-hot encode · Transform: train.csv

Data Analysis

Step 7. One-hot encode

ived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)
22	1	0	0	7.25	S
38	1	0	0	71.2833	C
26	0	0	0	7.925	S
35	1	0	0	53.1	S
35	0	0	0	8.05	S
	0	0	0	8.4583	Q
54	0	0	0	51.8625	S
2	3	1	0	21.075	S
27	0	2	0	11.1333	S
14	1	0	0	30.0708	C
4	1	1	0	16.7	S
58	0	0	0	26.55	S
20	0	0	0	8.05	S
39	1	5	0	31.275	S
14	0	0	0	7.8542	S
55	0	0	0	16	S
2	4	1	0	29.125	Q
	0	0	0	13	S
31	1	0	0	18	S
	0	0	0	7.225	C
35	0	0	0	26	S

Export data

ENCODE CATEGORICAL

Convert categorical variables to numeric or vector representations. [Learn more.](#)

Transform i

One-hot encode

Input columns i

Embarked x

Input already ordinal encoded i

Invalid handling strategy i

Skip

Drop last i

Output style i

Columns

Output column i

Optional

Clear Preview Add

 Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

-auncher

TitanicPrep.flow

| (?) Get help

Feedback

One-hot encode · Transform: train.csv

Data Analysis

Previewing: Encode categorical

◀ ENCODE CATEGORICAL X

Convert categorical variables to numeric or vector representations. [Learn more.](#) 

Transform 

One-hot encode

Input columns ⓘ

Embarked ✕

Input already ordinal encoded. 

Invalid handling strategy

skip

Drop last

Output style ⓘ

Columns

Output column

Optional

Clear

Launcher X TitanicPrep.flow ●

[Back to data flow](#)[Get help](#)

One-hot encode · Transform: train.csv

Data

Analysis

Step 8. One-hot encode

[Export data](#)

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Age_Imputed (float)
0	22	1	0	7.25	22
1	38	1	0	71.2833	38
1	26	0	0	7.925	26
1	35	1	0	53.1	35
0	35	0	0	8.05	35
0		0	0	8.4583	28
0	54	0	0	51.8625	54
0	2	3	1	21.075	2
1	27	0	2	11.1333	27
1	14	1	0	30.0708	14
1	4	1	1	16.7	4
1	58	0	0	26.55	58
0	20	0	0	8.05	20
0	39	1	5	31.275	39
0	14	0	0	7.8542	14
1	55	0	0	16	55
0	2	4	1	29.125	2
1		0	0	13	28
0	31	1	0	18	31
1		0	0	7.225	28
0	35	0	0	26	35

TRANSFORMS X

[+ Add step](#)

- ▶ 1. S3 Source
- ▶ 2. Data types
- ▶ 3. Impute
- ▶ 4. Drop column
- ▶ 5. Custom Pandas
- ▶ 6. One-hot encode
- ▶ 7. One-hot encode
- ▶ 8. One-hot encode

X

Amazon SageMaker Studio

File Edit View Run Kernel Tabs Settings Help

Launcher X TitanicPrep.flow

Back to data flow

One-hot encode · Transform: train.csv

Data Analysis

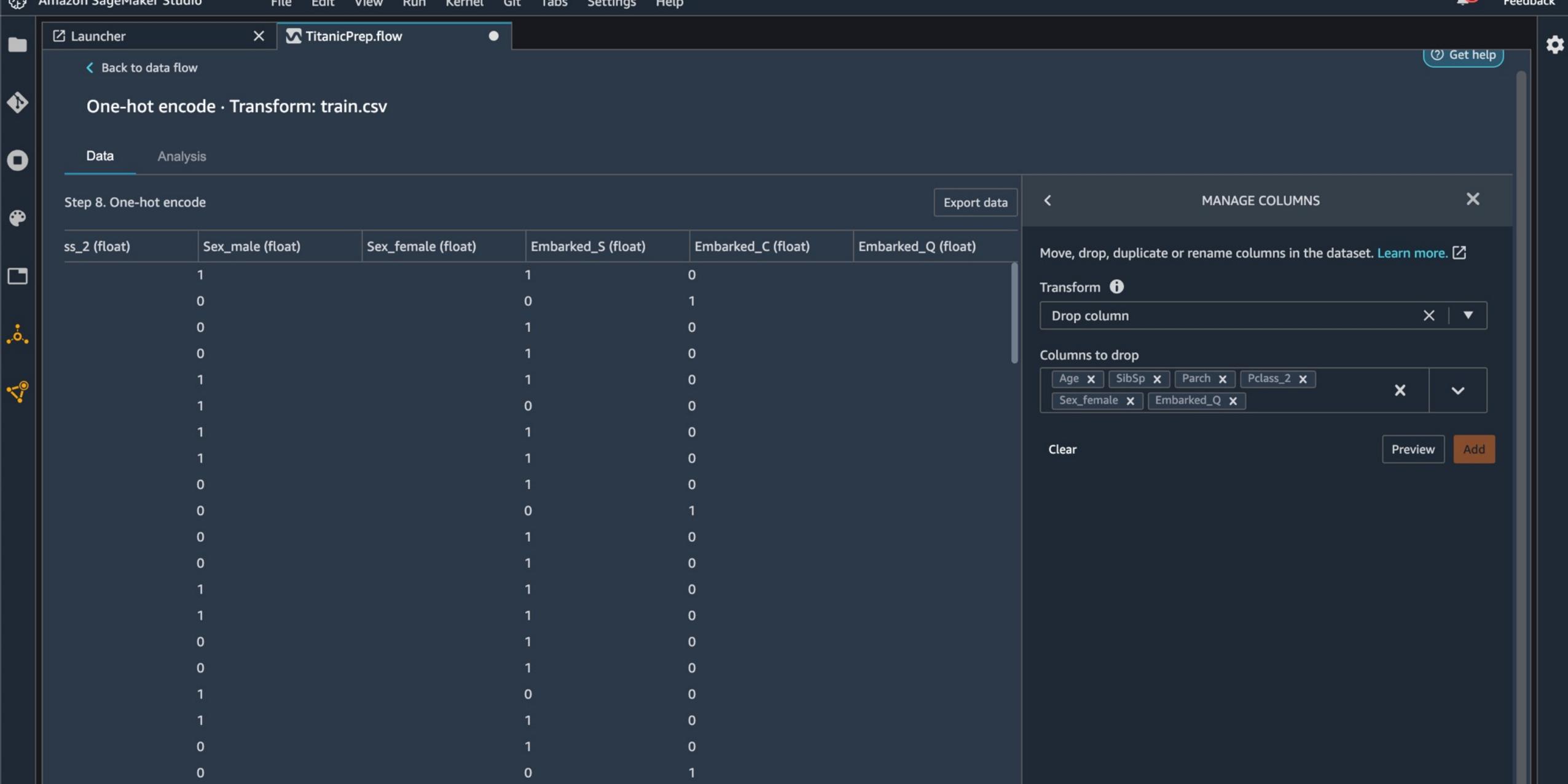
Step 8. One-hot encode

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Age_Imputed (float)
0	22	1	0	7.25	22
1	38	1	0	71.2833	38
1	26	0	0	7.925	26
1	35	1	0	53.1	35
0	35	0	0	8.05	35
0		0	0	8.4583	28
0	54	0	0	51.8625	54
0	2	3	1	21.075	2
1	27	0	2	11.1333	27
1	14	1	0	30.0708	14
1	4	1	1	16.7	4
1	58	0	0	26.55	58
0	20	0	0	8.05	20
0	39	1	5	31.275	39
0	14	0	0	7.8542	14
1	55	0	0	16	55
0	2	4	1	29.125	2
1		0	0	13	28
0	31	1	0	18	31
1		0	0	7.225	28

Export data

ADD TRANSFORM

- Featurize text
- Format string
- Group by
- Handle missing
- Handle outliers
- Handle structured column
- Manage columns
- Manage rows
- Manage vectors



Launcher

TitanicPrep.flow

Back to data flow

(?) Get help

One-hot encode · Transform: train.csv

Data

Analysis

Previewing: Manage columns

Survived (long)	Fare (float)	Age_Imputed (float)	With_Family (long)	Pclass_3 (float)	Pclass_1 (float)
0	7.25	22	1	1	0
1	71.2833	38	1	0	1
1	7.925	26	0	1	0
1	53.1	35	1	0	1
0	8.05	35	0	1	0
0	8.4583	28	0	1	0
0	51.8625	54	0	0	1
0	21.075	2	1	1	0
1	11.1333	27	1	1	0
1	30.0708	14	1	0	0
1	16.7	4	1	1	0
1	26.55	58	0	0	1
0	8.05	20	0	1	0
0	31.275	39	1	1	0
0	7.8542	14	0	1	0
1	16	55	0	0	0
0	29.125	2	1	1	0
1	13	28	0	0	0
0	18	31	1	1	0
1	7.225	28	0	1	0
0	26	35	0	0	0

Export data

< MANAGE COLUMNS X

Move, drop, duplicate or rename columns in the dataset. [Learn more.](#) ↗

Transform ⓘ

Drop column

Columns to drop

Age X SibSp X Parch X Pclass_2 X
Sex_female X Embarked_Q X

Clear

Preview Add

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Feedback

Launcher X TitanicPrep.flow X

Import Data Flow [Get help](#)

Create job

Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Validation complete 0 errors Done

Source S3: train.csv

Data types Transform: train.csv

Steps (7) </> +

The screenshot shows the Amazon SageMaker Studio Data Flow interface. At the top, there are tabs for 'Import' and 'Data Flow', with 'Data Flow' being the active tab. A validation message 'Validation complete' with '0 errors' and a 'Done' button is displayed. Below this, a data flow diagram is shown with three main components: 'Source' (S3: train.csv), 'Data types' (Transform: train.csv), and 'Steps (7)' (represented by a stack of seven blue rounded rectangles). A '+' icon is located next to the steps component, indicating where new steps can be added. On the left side, there is a vertical toolbar with various icons for file operations like copy, paste, and lock. On the right side, there is a large empty workspace area with a progress bar at the bottom.

Launcher

X

TitanicPrep.flow

X

Import

Data Flow

Get help



Data flow

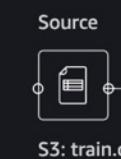
Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete

0 errors

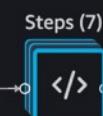
Done



S3: train.csv



Transform: train.csv



Steps (7)

1. Impute
2. Drop column
3. Custom Pandas
4. One-hot encode
5. One-hot encode
6. One-hot encode
7. Drop column



Launcher X TitanicPrep.flow X[Import](#) [Data Flow](#)[Get help](#)[Create job](#)

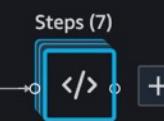
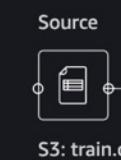
Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Validation complete

0 errors

Done



1. [Impute](#)
2. [Drop column](#)
3. [Custom Pandas](#)
4. [One-hot encode](#)
5. [One-hot encode](#)
6. [One-hot encode](#)
7. [Drop column](#)

[Add transform](#)
[Add analysis](#)

[Add destination >](#)

[Export to >](#)

[Delete step](#)

Launcher



TitanicPrep.flow

Import

Data Flow

Get help



Data flow

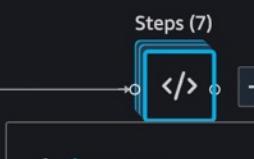
Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete

0 errors

Done



⋮

1. Impute

2. Drop column

3. Custom Pandas

- Amazon S3 (via Jupyter Notebook)
- SageMaker Pipelines (via Jupyter Notebook)
- Python Code
- SageMaker Feature Store (via Jupyter Notebook)

Add transform

Add analysis

Add destination >

Export to >

Delete step

Launcher



TitanicPrep.flow

Import

Data Flow

Get help



Data flow

Choose the plus sign to add a step to the flow. Select a step to modify.

Create job

Validation complete

0 errors

Done



1. Impute
2. Drop column
3. Custom Pandas

Add transform

Add analysis

Add destination >

Export to >

Delete step

Amazon S3 (via Jupyter Notebook)

SageMaker Pipelines (via Jupyter Notebook)

Python Code

SageMaker Feature Store (via Jupyter Notebook)

Save to S3 with a SageMaker Processing Job

Quick Start To save your processed data to S3, select the Run menu above and click **Run all cells**. [View the status of the export job and the output S3 location.](#)

This notebook executes your Data Wrangler Flow `TitanicPrep.flow` on the entire dataset using a SageMaker Processing Job and will save the processed data to S3.

This notebook saves data from the step `Manage Columns` from `Source: Train.Csv`. To save from a different step, go to Data Wrangler to select a new step to export.

Contents

1. [Inputs and Outputs](#)
2. [Run Processing Job](#)
 - A. [Job Configurations](#)
 - B. [Create Processing Job](#)
 - C. [Job Status & S3 Output Location](#)
3. [Optional Next Steps](#)
 - A. [Load Processed Data into Pandas](#)
 - B. [Train a model with SageMaker](#)

Inputs and Outputs

The below settings configure the inputs and outputs for the flow export.

Configurable Settings

In `Input - Source` you can configure the data sources that will be used as input by Data Wrangler

1. For S3 sources, configure the source attribute that points to the input S3 prefixes



Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 3

▶ AWS Marketplace for S3

Read the S3 resources page for documentation and technical content.

Learn more



Amazon S3

▶ Account snapshot

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

Buckets (3) [Info](#)

Copy ARN

Empty

Delete

Create bucket

Buckets are containers for data stored in S3. [Learn more](#)

Find buckets by name

< 1 >



Name	AWS Region	Access	Creation date
sagemaker-studio-834120567544-5zbey54q71a	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 20:55:19 (UTC+08:00)
sagemaker-us-east-1-834120567544	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 21:06:52 (UTC+08:00)
titanic-michlin-20220218	US East (N. Virginia) us-east-1	Bucket and objects not public	February 16, 2022, 22:15:58 (UTC+08:00)

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight (3)

AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-834120567544

sagemaker-us-east-1-834120567544 Info

Objects Properties Permissions Metrics Management Access Points

Objects (3)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)



Find objects by prefix < 1 > 

-	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 Canvas/	Folder	-	-	-
<input type="checkbox"/>	 data_wrangler_flows/	Folder	-	-	-
<input checked="" type="checkbox"/>	 export-flow-16-15-04-15-fd0b1758/	Folder	-	-	-

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight

▶ AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/

export-flow-16-15-04-15-fd0b1758/

Copy S3 URI

Objects Properties

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI Copy URL Download Open Delete Actions Create folder

Upload

Find objects by prefix

< 1 >

Name	Type	Last modified	Size	Storage class
output/	Folder	-	-	-

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight (3)

▶ AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/ > output/

output/

Objects (1) **Properties**

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

[Upload](#)

Find objects by prefix < 1 >

<input checked="" type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input checked="" type="checkbox"/>	data-wrangler-flow-processing-16-15-04-15-fd0b1758/	Folder	-	-	-

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight 3

▶ AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/ > output/ > data-wrangler-flow-processing-16-15-04-15-fd0b1758/

[Copy S3 URI](#) [Objects](#) [Properties](#) Objects (1) Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#) [Actions ▾](#) [Create folder](#) Find objects by prefix < 1 > | <input checked="" type="checkbox"/> | Name | Type | Last modified | Size | Storage class | |-------------------------------------|---|--------|---------------|------|---------------| | <input checked="" type="checkbox"/> | 7f06826f-b540-4996-b6d5-e1a8eb70f110/ | Folder | - | - | - |

Amazon S3

Buckets

- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- Access analyzer for S3

Block Public Access settings for this account

Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight (3)

AWS Marketplace for S3

Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/ > output/ > data-wrangler-flow-processing-16-15-04-15-fd0b1758/ > 7f06826f-b540-4996-b6d5-e1a8eb70f110/ > default/

default/

Objects **Properties**

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input checked="" type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input checked="" type="checkbox"/>	part-00000-b8bd0b1c-acdc-439d-bf2c-837fa0f2faee-c000.csv	csv	February 16, 2022, 23:10:52 (UTC+08:00)	30.5 KB	Standard

Possible Data Loss Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

Save As...

A1 f_x Survived

1	Survived	Fare	Age_Imputed	With_Family	Pclass_3	Pclass_1	Sex_male	Embarked_S	Embarked_C	J	K	L	M	N	O	P	Q	R	S	T	U
2	0	7.25	22	1	1	0	1	1	0												
3	1	71.2833	38	1	0	1	0	0	1												
4	1	7.925	26	0	1	0	0	0	1												
5	1	53.1	35	1	0	1	0	0	1												
6	0	8.05	35	0	1	0	0	1	1												
7	0	8.4583	28	0	1	0	0	1	0												
8	0	51.8625	54	0	0	1	1	1	1												
9	0	21.075	2	1	1	0	0	1	1												
10	1	11.1333	27	1	1	0	0	0	1												
11	1	30.0708	14	1	0	0	0	0	0												
12	1	16.7	4	1	1	0	0	0	1												
13	1	26.55	58	0	0	1	0	0	1												
14	0	8.05	20	0	1	0	0	1	1												
15	0	31.275	39	1	1	0	0	1	1												
16	0	7.8542	14	0	1	0	0	0	1												
17	1	16	55	0	0	0	0	0	1												
18	0	29.125	2	1	1	0	0	1	0												
19	1	13	28	0	0	0	0	1	1												
20	0	18	31	1	1	0	0	0	1												
21	1	7.225	28	0	1	0	0	0	0												
22	0	26	35	0	0	0	0	1	1												
23	1	13	34	0	0	0	0	1	1												
24	1	8.0292	15	0	1	0	0	0	0												
25	1	35.5	28	0	0	1	1	1	1												
26	0	21.075	8	1	1	0	0	0	1												
27	1	31.3875	38	1	1	0	0	0	1												
28	0	7.225	28	0	1	0	0	1	0												
29	0	263	19	1	0	1	1	1	1												
30	1	7.8792	28	0	1	0	0	0	0												
31	0	7.8958	28	0	1	0	0	1	1												
32	0	27.7208	40	0	0	1	1	1	0												
33	1	146.5208	28	1	0	1	0	0	0												
34	1	7.75	28	0	1	0	0	0	0												
35	0	10.5	66	0	0	0	0	1	1												
36	0	82.1708	28	1	0	1	1	1	0												
37	0	52	42	1	0	1	1	1	1												
38	1	7.2292	28	0	1	0	1	0	1												

part-00000-b8bd0b1c-acdc-439d-b

+

Ready

part-00000-b8bd0b1c-acdc-439d-b.csv

Grid View + 100%

Thank you!

Michael Lin

linmicht@amazon.com

