

Lab 2 SageMaker Data Wrangler



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data Prep in the Lab

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

AWS Services Search results for 's3' N. Virginia TeamRole/MasterKey @ 8341-2056-7544

Amazon SageMaker

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
- Top features** Buckets Access points Batch Operations
- 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

Add user < 1 > 🔍 Launch app ▾ Delete Domain Edit Settings Authentication method AWS Identity and Access Management (IAM)

Amazon SageMaker project templates enabled for Studio users

Amazon S3



Amazon S3 > Buckets

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

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](#)



[Copy ARN](#)

Empty

Delete

[Create bucket](#)

[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)

Event Engine - Team Dashboard | Amazon SageMaker | S3 bucket | JupyterLab | Introducing Amazon SageMake | + | Paused

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

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

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

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global ▾ TeamRole/MasterKey @ 8341-2056-7544 ▾

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

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

Cancel [Create bucket](#)

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

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

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. [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](#)

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 | Amazon SageMaker | titanic-michlin-20220218 - S3 | JupyterLab | Introducing Amazon SageMake | +

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

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

Global TeamRole/MasterKey @ 8341-2056-7544

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](#)

C 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 | Amazon SageMaker | S3 Management Console | JupyterLab | Introducing Amazon SageMake | + | Paused

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.

Find by name

	Name	Folder	Type	Size
No files or folders				

You have not chosen any files or folders to upload.

Destination

Destination

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

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

File Selection:

- test.csv
- train.csv**
- winequality-red.csv

train.csv
Comma Separated Spreadsheet (.csv) - 61 KB

Information

Created Yesterday, 10:23 PM

Show More

Cancel Open

Destination

Destination

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

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 (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

Destination

Destination
<s3://titanic-michlin-20220218>

▶ **Destination details**

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

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 8341-2056-7544

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

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**

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

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]

Global TeamRole/MasterKey @ 8341-2056-7544

Upload succeeded View details below.

Upload: status

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

Summary

Destination	Succeeded	Failed
s3://titanic-michlin-20220218	1 file, 59.8 KB (100.00%)	0 files, 0 B (0%)

Files and folders Configuration

Files and folders (1 Total, 59.8 KB)

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 sagemaker|Canvas X

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

Amazon SageMa

Search results for 'sagemaker'

Services (2)

Features (4)

Blogs (614)

Documentation (58,491)

Knowledge Articles (24)

Tutorials (9)

Events (3)

Marketplace (127)

Services

 **Amazon SageMaker** ☆
Build, Train, and Deploy Machine Learning Models

Top features
SageMaker Studio Autopilot SageMaker Canvas

 **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

dio

Maker, you pay per use. Authoring, publishing and inference are billed by the number of requests and inferences.

es

Amazon SageMaker



Getting started

Control panel

Studio

Studio Lab

Canvas

RStudio

SageMaker dashboard

Images

Lifecycle configurations

Search

► **Ground Truth**

► **Notebook**

► **Processing**

► **Training**

► **Inference**

► **Edge Manager**

Amazon SageMaker

SageMaker Studio

The first fully integrated development environment (IDE) for machine learning.

Launch Studio

Choose a user to launch Studio

Users

Add user

Search users

< 1 >

Name

Modified on

Created on

Amazon SageMaker

Getting started

Control panel

Studio

Studio Lab

Canvas

RStudio

SageMaker dashboard

Images

Lifecycle configurations

Search

► Ground Truth

► Notebook

► Processing

► Training

► Inference

► Edge Manager

Users

Add user

Search users

< 1 >



Name

Modified on

Created on

sagemakeruser

Oct 09, 2022 05:02 UTC

Oct 09, 2022 05:02 UTC

Open Studio

How it works

What is Studio?

Amazon SageMaker Studio provides a single, web-based visual interface where you can perform all ML development steps, improving data science team productivity by up to 10x. SageMaker Studio gives you complete access, control, and visibility into each step required to build, train, and deploy models.

[Get Started with SageMaker](#)

Introduction video



Pricing (US)

With Amazon SageMaker Studio, you pay only for what you use. Authoring, training and hosting is billed by the second, with no minimum fees and no upfront commitments.

[Learn more](#)

Documentation

[Tutorials](#)



Amazon SageMaker Studio

Creating the JupyterServer application default...

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

SageMaker resources
Select the resource to view.

Data Wrangler Data Wrangler

Search... New flow

Name Last Modified

SageMaker JumpStart

Accelerate your machine learning workflows with one-click access to popular model collections, example Jupyter notebooks, and to end-to-end solutions that solve common use cases.

Solutions Data types ML Tasks Notebooks Frameworks Resources

Solutions

Launch end-to-end machine learning solutions that tie SageMaker to other AWS services with one click.

[Explore All Solutions \(19\)](#)

Fraud Detection in Financial Transactions 
Financial Services

Use Deep Graph Library (DGL) to train a graph neural network model to detect fraud in financial transactions.

Corporate Credit Rating Prediction 
Financial Services

Use company's financial data (numeric + text) to predict corporate credit rating.

Demand Forecasting 
Supply Decision Making

Demand forecasting for multi-variate time series data using deep learning models.

Product Defect 
Product Defect

Identify defective images.

Vision Models

Fine-tune and deploy pretrained vision models with one click.

[Explore All Vision Models \(204\)](#)

Image Classification 
Featured Mobilenet V2 1.0...

Dataset: ImageNet
Fine-tunable: Yes

Object Detection 
Featured Frcnn Resnet 50 F...

Dataset: COCO 2017
Fine-tunable: Yes

Semantic Segmentation 
Featured Fcn Resnet 101 C...

Dataset: COCO 2017
Fine-tunable: Yes

Instance Seg 
Featured Ma...

Dataset: COCO 2017
Fine-tunable: No

less than 10 seconds ago

0 \$ 0 Git: idle SageMaker JumpStart

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

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 Create job

Data flow

Import your data to prepare or analyze it.

```
graph LR; A[Import Data] --> B[Prepare]; B --> C[Process];
```

Import Data Prepare Process

Import data Loading sample data

less than 10 seconds ago

0 \$ 1 Git: idle untitled.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾
Search... New flow
Name Last Modified
untitled.flow less than 5 seconds ago

untitled.flow 16 vCPU + 64 GiB Get help

Close Tab Close All Other Tabs Close Tabs to Right New Console for Activity Rename Data Wrangler Flow... New View for Data Wrangler Flow Show in File Browser Shift+Right Click for Browser Menu

Data flow
Import your data to prepare or analyze it.

Import Data Prepare Process

Import data Use sample dataset

+

less than 20 seconds ago

Git: refreshing... untitled.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

. SageMaker JumpStart X untitled.flow X Import Data Flow 16 vCPU + 64 GiB Get help

Data flow

Import your data to prepare or analyze it.

Create job

Rename File

File Path
untitled.flow

New Name
TitanicPrep.flow

Cancel Rename

Import data Use sample dataset

Process

+

-

↻

⌚ half a minute ago

0 \$ 1 Git: idle untitled.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾
Search... New flow
Name Last Modified
untitled.flow less than 5 seconds ago

.SageMaker JumpStart X TitanicPrep.flow X
Import Data Flow 16 vCPU + 64 GiB Get help

Data flow

Import your data to prepare or analyze it.

Import Data Prepare Process

Import data Use sample dataset

less than a minute ago

0 \$ 1 Git: idle TitanicPrep.flow

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

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 16 vCPU + 64 GiB Get help

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

Amazon S3  Amazon Athena

less than a minute ago

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import data 16 vCPU + 64 GiB Get help

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

S3 URI path

Enter an S3 URI Go

S3 /

Bucket name	Region	Creation date
sagemaker-studio-0f2b2f90	us-east-1	2022-10-09 04:52:06+00:00
sagemaker-us-east-1-200409934317	us-east-1	2022-10-09 04:52:06+00:00
titanic-dataset-20221009	us-east-1	2022-10-09 05:26:05+00:00

1 minute ago

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import data 16 vCPU + 64 GiB Get help

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

S3 URI path

Enter an S3 URI Go

S3 /

Bucket name	Region	Creation date
sagemaker-studio-0f2b2f90	us-east-1	2022-10-09 04:52:06+00:00
sagemaker-us-east-1-200409934317	us-east-1	2022-10-09 04:52:06+00:00
titanic-dataset-20221009	us-east-1	2022-10-09 05:26:05+00:00

1 minute ago

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import data

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

S3 URI path

Enter an S3 URI Go

S3 / titanic-dataset-20221009

Object name	Size	Last modified
train.csv	59.76KB	2022-10-09 05:26:22+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (First 100 rows shown. The preview doesn't reflect your sampling configuration.)

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
7	0	1	McCarthy, Mr. Timothy J	male
8	0	3	Palsson, Master. Gosta ...	male

1 minute ago

16 vCPU + 64 GiB Get help

DETAILS

Name Preview on

File type

First row is header

Import nested directories

Delimiter

Sampling 

Sample size

Filename as separate column

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import data

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

S3 URI path

Enter an S3 URI Go

S3 / titanic-dataset-20221009

Object name	Size	Last modified
train.csv	59.76KB	2022-10-09 05:26:22+00:00

PREVIEW • train.csv (First 100 rows shown. The preview doesn't reflect your sampling configuration.)

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
7	0	1	McCarthy, Mr. Timothy J	male
8	0	3	Palsson, Master. Gosta ...	male

1 minute ago

16 vCPU + 64 GiB Get help

DETAILS

Name Preview on

File type

First row is header

Import nested directories

Delimiter

Sampling

First K ▾

None (highlighted)

First K

Random

Stratified

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

Import data

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

S3 URI path

Enter an S3 URI Go

S3 / titanic-dataset-20221009 C

Object name	Size	Last modified
train.csv	59.76KB	2022-10-09 05:26:22+00:00

Previous Displaying 1 - 1 Next

PREVIEW • train.csv (First 100 rows shown. The preview doesn't reflect your sampling configuration.)

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
7	0	1	McCarthy, Mr. Timothy J	male
8	0	3	Palsson, Master. Gosta ...	male

2 minutes ago

0 \$ 1 Git: idle TitanicPrep.flow

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾
Search... New flow
Name Last Modified
untitled.flow less than 5 seconds ago

. SageMaker JumpStart X TitanicPrep.flow

Data types · Transform: train.csv

Data Analysis Training NEW

Step 2. Data types

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

Export and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▼ 2. Data 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 Update

Report a bug

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾

Search... New flow

Name Last Modified

untitled.flow less than 5 seconds ago

2 minutes ago

. SageMaker JumpStart X TitanicPrep.flow ●

Data flow 

Data types · Transform: train.csv

Data Analysis Training NEW

Step 2. Data types

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

Export and train Export data

ALL STEPS 

+ Add step

▶ 1. S3 Source

▼ 2. Data 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 Update 

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

SageMaker resources
Select the resource to view.

Data Wrangler ▾
Search... New flow
Name Last Modified
untitled.flow less than 5 seconds ago

. S. SageMaker JumpStart X TitanicPrep.flow ● 16 vCPU + 64 GiB Get help

Import Data Flow

Data flow

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

Validation complete 0 errors Done

Source → Data types

S3: train.csv Transform: train.csv

+ - ↻ 🔒

4 minutes ago

0 \$ 1 Git: idle TitanicPrep.flow

The screenshot shows the Amazon SageMaker Studio interface with the Data Wrangler feature selected. On the left, there's a sidebar with 'SageMaker resources' and a list of flows. The main area is titled 'Data flow' and displays a flow diagram. It starts with a green 'Source' block containing 'S3: train.csv', which has an arrow pointing to a blue 'Data types' block containing 'Transform: train.csv'. There are buttons for '+' and '-' below the blocks, and a settings gear icon in the top right corner of the main area. A validation message at the top says 'Validation complete 0 errors Done'. The bottom status bar shows '4 minutes ago' and the flow name 'TitanicPrep.flow'.

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

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

16 vCPU + 64 GiB Get help

Data flow
Choose the plus sign to add a step to the flow. Select a step to modify.
Validation complete 0 errors

Add transform
Add analysis ↗
Train model NEW
Get data insights
Add destination >
Export to >
Join
Concatenate
Edit

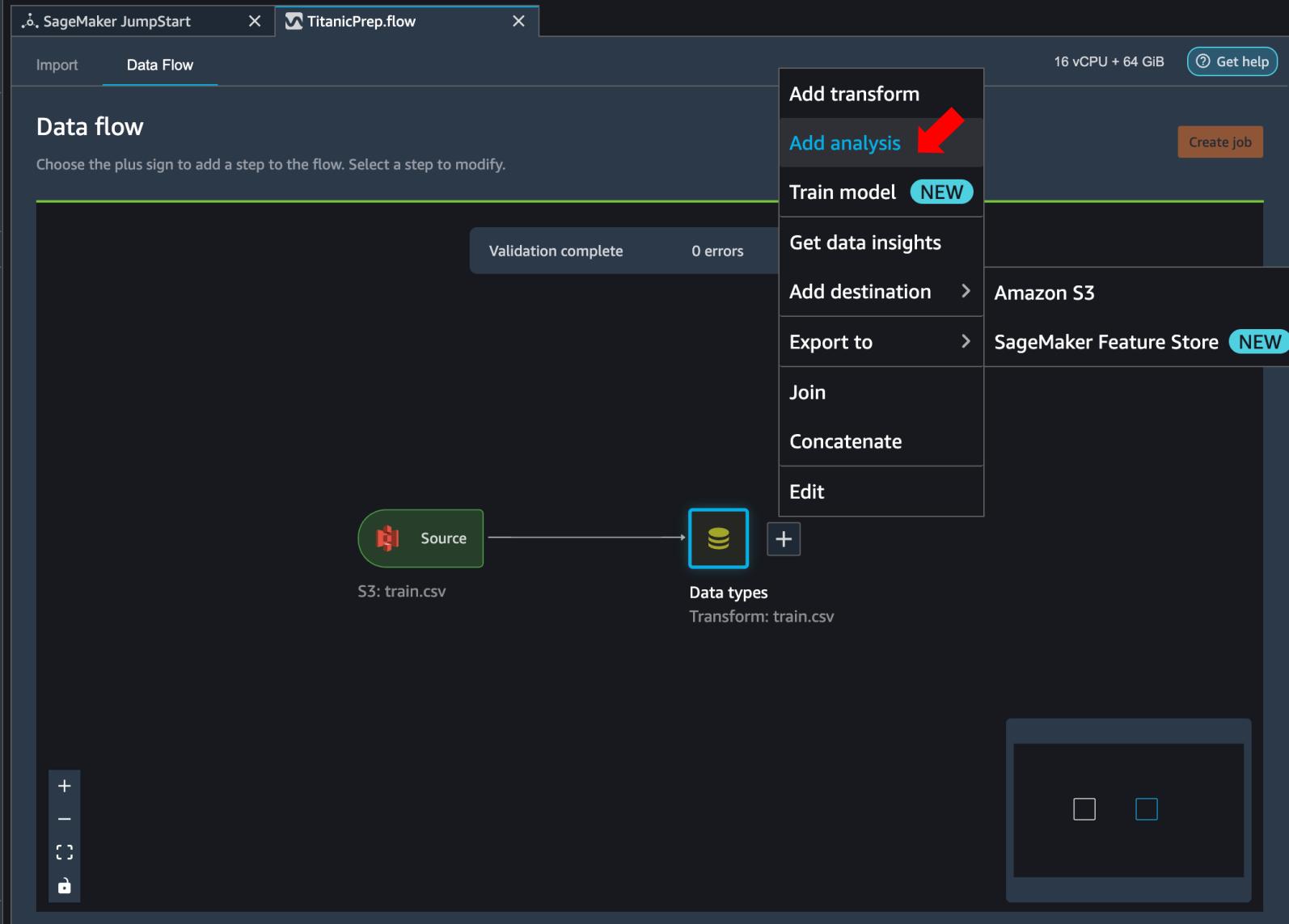
Amazon S3
SageMaker Feature Store NEW

Source → Data types
S3: train.csv Transform: train.csv

+

4 minutes ago

Git: refreshing... TitanicPrep.flow



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

.s. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	35 35
4	0	3	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ... Johnson, Mrs. Oscar W (...	male	27
7	1	3		female	

0 \$ 1 Git: idle TitanicPrep.flow

All analyses

Create analysis

Analysis type: Histogram

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

Analysis name: Untitled

X axis: Select...

Color by: Select...

Facet by: Select...

Clear Preview Save

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

.5. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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

All analyses Create analysis

Analysis type

Histogram Duplicate rows
Feature Correlation
Histogram
Multicollinearity
Quick Model
Scatter Plot
Table Summary Target Leakage
Time Series
Select...
Optional
Clear Preview Save

0 \$ 1 Git: idle TitanicPrep.flow

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

.s. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X

Data types · Transform: train.csv

Data Analysis Training NEW

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

All analyses

Create analysis

Analysis type: Table Summary

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

Analysis name: Summary

Optional

Preview Save



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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data types · Transform: train.csv

Data Analysis Training NEW

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

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



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

.δ. SageMaker JumpStart X TitanicPrep.flow

16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Create new analysis

Summary
Table Summary

0 \$ 1 Git: idle TitanicPrep.flow

The screenshot shows the Amazon SageMaker Studio interface. The main area displays the 'Data types' section for a 'Transform: train.csv' file. There are three tabs at the top: 'Data', 'Analysis' (which is selected), and 'Training'. Below the tabs, there's a large icon representing data analysis and a summary card labeled 'Summary' and 'Table Summary'. In the bottom right corner of the main area, there's an orange button labeled 'Create new analysis'. A prominent red arrow points from the left side of the image towards this button. The top navigation bar includes links for 'File', 'Edit', 'View', 'Run', 'Kernel', 'Git', 'Tabs', 'Settings', 'Help', and a 'Report a bug' link. The top right corner shows resource information ('16 vCPU + 64 GiB') and a 'Get help' button. The bottom navigation bar has icons for file operations like 'New', 'Open', 'Save', and 'Exit', along with a 'Git' status indicator and the file name 'TitanicPrep.flow'.

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

.s. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	26
4	1	1		female	35
5	0	3		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 (... Johnson, Mrs. Oscar W (... 	female	27

All analyses

Create analysis

Analysis type: Histogram

Analysis name: Untitled

X axis: Select...

Color by: Select...

Facet by: Select...

Clear Preview Save

Git: refreshing... TitanicPrep.flow

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

.s. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female male	35 35
4	0	3	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ... Johnson, Mrs. Oscar W (...	male female	27
7	1	3			
8	1	3			
9	1	3			

PassengerId
Survived
Pclass
Name
Sex →
Age
SibSp
Parch
Ticket
-

Select... ▾

Color by
Select... ▾

Optional

Facet by
Select... ▾

Optional

Clear Preview Save

0 \$ 1 Git: idle TitanicPrep.flow

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

.s. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	35 35
4	1	1	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ...	male	27
7	0	1	Johnson, Mrs. Oscar W (... Johnson. Mrs. Oscar W (...	female	27
8	0	3			
9	1	3			

All analyses

Create analysis

Analysis type: Histogram

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

Analysis name: Untitled

X axis: Sex

Color by: Survived

Facet by: Select...

Clear Preview Save

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	35 35
4	1	1	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ... Johnson, Mrs. Oscar W (...	male	27
7	0	1			
8	0	3			
9	1	3			

16 vCPU + 64 GiB (Get help)

All analyses

Create analysis

Analysis type: Histogram

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

Analysis name: Sex

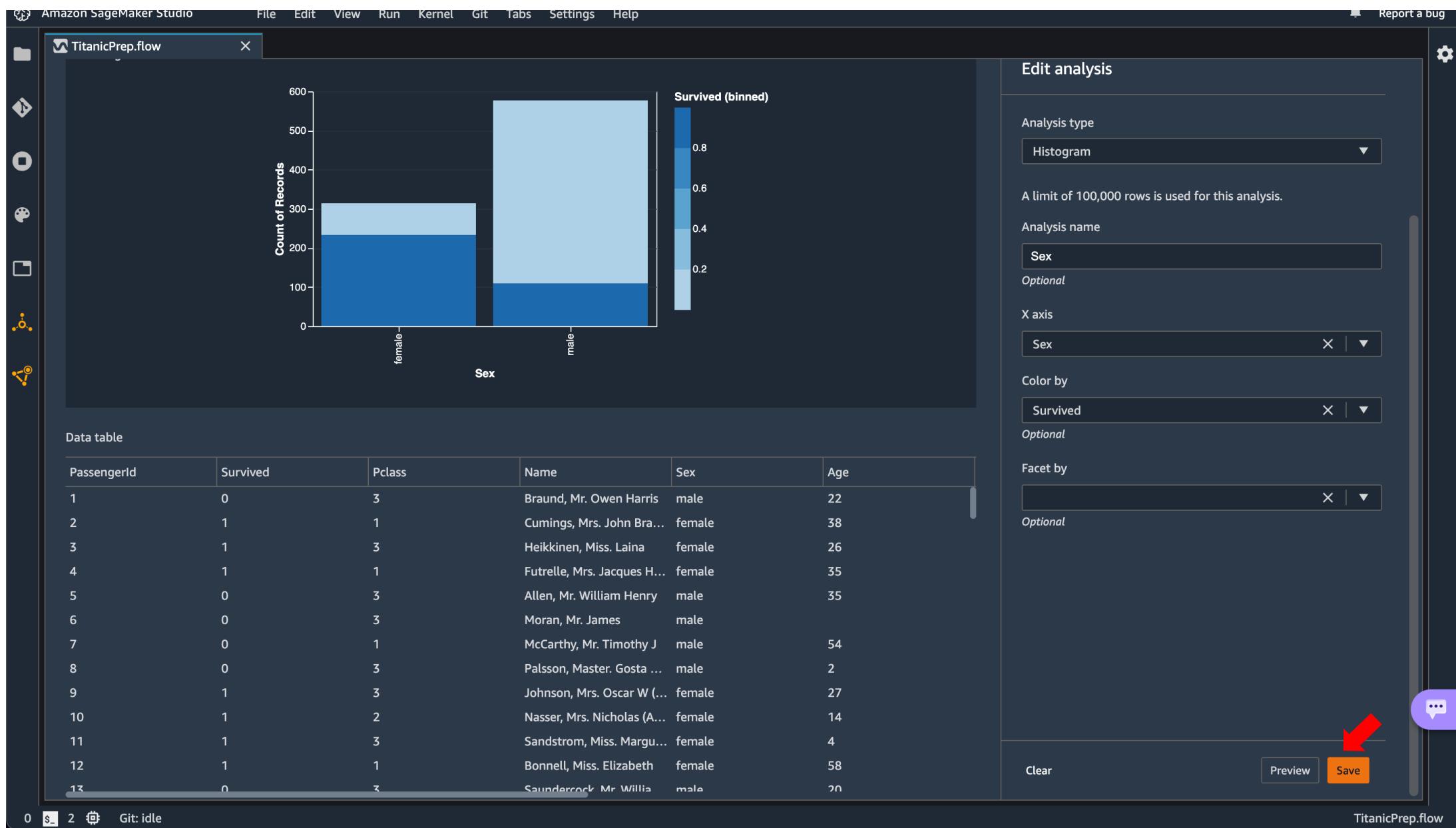
X axis: Sex

Color by: Survived

Facet by: Select...

Preview Save

0 \$ 1 Git: idle TitanicPrep.flow



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

.δ. SageMaker JumpStart X TitanicPrep.flow

16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Create new analysis

Summary
Table Summary

Sex
Histogram
Sex · Survived

0 \$ 1 Git: idle TitanicPrep.flow

A screenshot of the Amazon SageMaker Studio interface. The top navigation bar includes 'Amazon SageMaker Studio', 'File', 'Edit', 'View', 'Run', 'Kernel', 'Git', 'Tabs', 'Settings', 'Help', and 'Report a bug'. Below the tabs, there are two tabs: '.δ. SageMaker JumpStart' and 'TitanicPrep.flow', with 'TitanicPrep.flow' being active. On the right side of the header, it shows '16 vCPU + 64 GiB' and a 'Get help' button. The main content area is titled 'Data types · Transform: train.csv'. It has three tabs: 'Data', 'Analysis' (which is selected), and 'Training'. Below these tabs is a 'Create new analysis' button with a red arrow pointing to it. There are two analysis cards displayed: 'Summary' (Table Summary) and 'Sex' (Histogram Sex · Survived). The bottom of the screen shows a footer with icons for file operations and a status bar indicating 'Git: idle'.

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

.SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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

PassengerId
Survived
Pclass (highlighted)
Name
Sex
Age
SibSp
Parch
Ticket
-

Select... ▾

Color by
Select... ▾

Optional

Facet by
Select... ▾

Optional

Clear Preview Save

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

.SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	26
4	1	1		female	35
5	0	3		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 (... Johnson, Mrs. Oscar W (... 	female	27

All analyses

Create analysis

Analysis type: Histogram

Analysis name: Untitled

X axis: Pclass

Color by: Survived

Facet by: Select...

Clear Preview Save

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female male	35 35
4	0	3	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ... Johnson, Mrs. Oscar W (...	male female	27
7	1	3			
8	0	3			
9	1	3			

All analyses

Create analysis

Analysis type: Histogram

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

Analysis name: Pclass

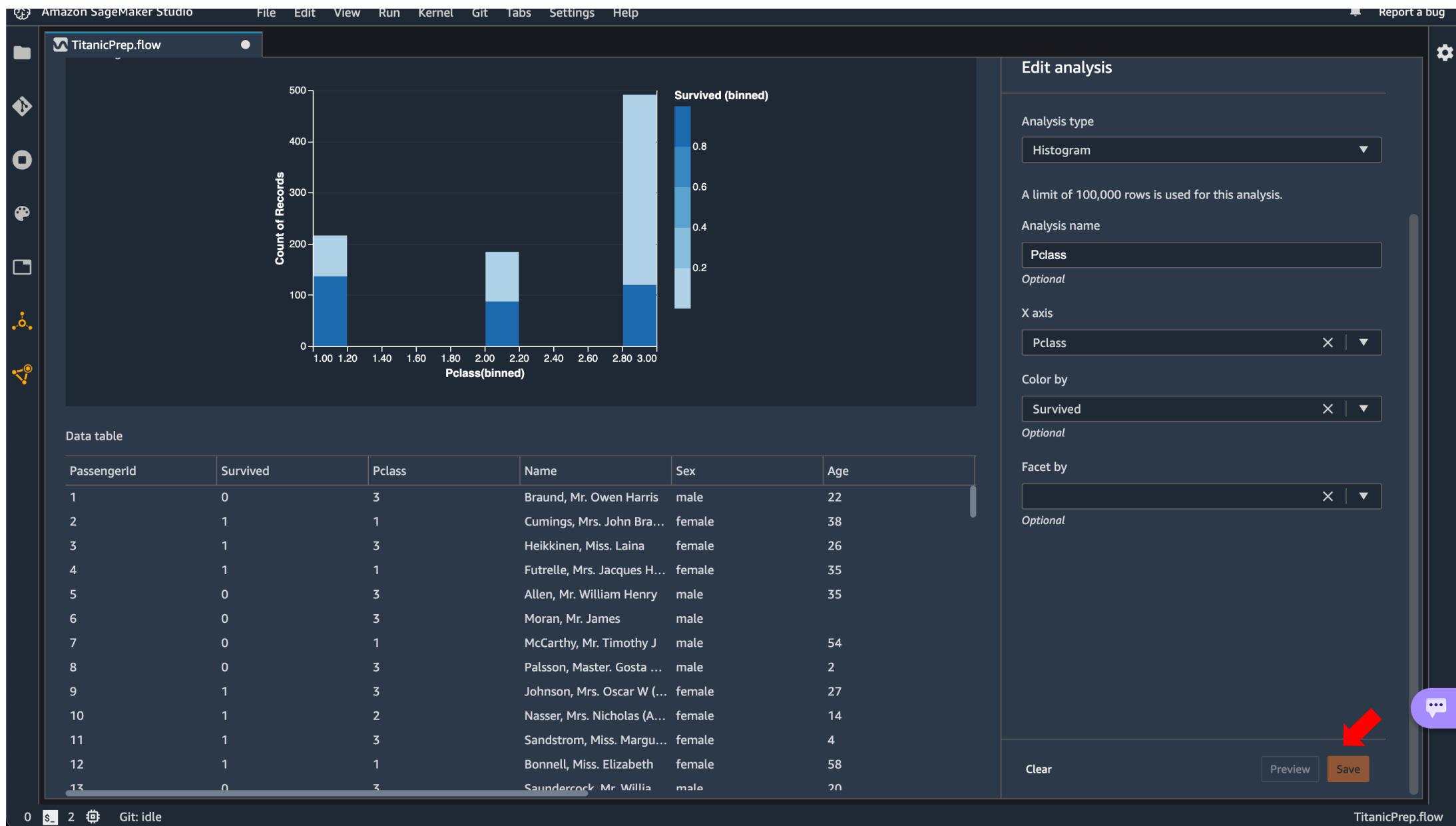
X axis: Pclass

Color by: Survived

Facet by: Select...

Preview Save

0 \$ 1 Git: idle TitanicPrep.flow



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

.δ. SageMaker JumpStart X TitanicPrep.flow ● 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

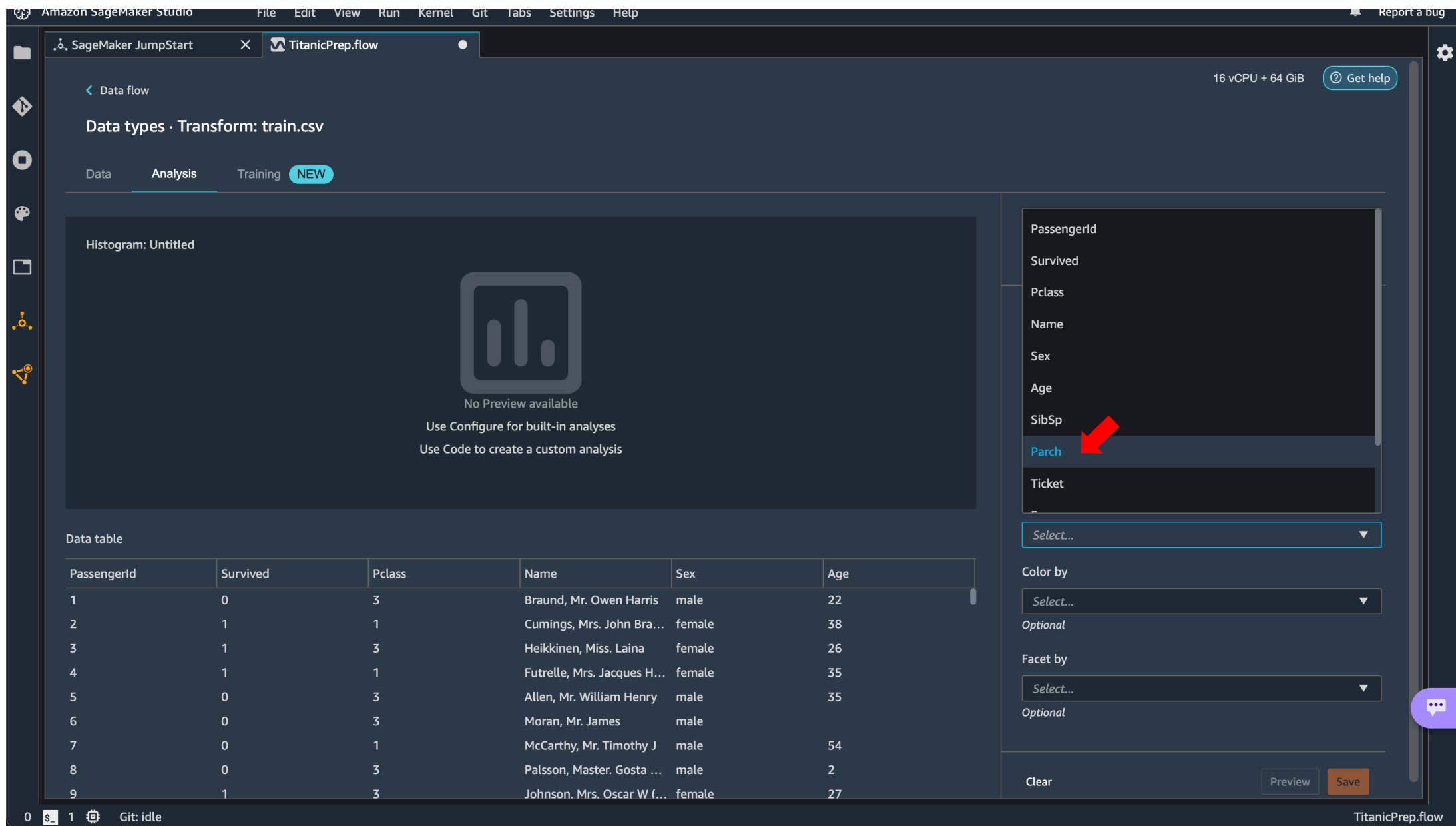
Create new analysis

Summary
Table Summary

Sex
Histogram
Sex · Survived

Pclass
Histogram
Pclass · Survived

0 \$ 1 Git: refreshing... TitanicPrep.flow



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

.s. SageMaker JumpStart X TitanicPrep.flow ● 16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

Histogram: Untitled

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... Heikkinen, Miss. Laina	female	38 26
3	1	3	Futrelle, Mrs. Jacques H... Allen, Mr. William Henry	female	35 35
4	1	1	Moran, Mr. James	male	54
5	0	3	McCarthy, Mr. Timothy J	male	2
6	0	3	Palsson, Master. Gosta ...	male	27
7	0	1	Johnson, Mrs. Oscar W (... Johnson. Mrs. Oscar W (...	female	27
8	0	3			
9	1	3			

0 \$ 1 Git: idle TitanicPrep.flow

All analyses

Create analysis

Analysis type: Histogram

Analysis name: Untitled

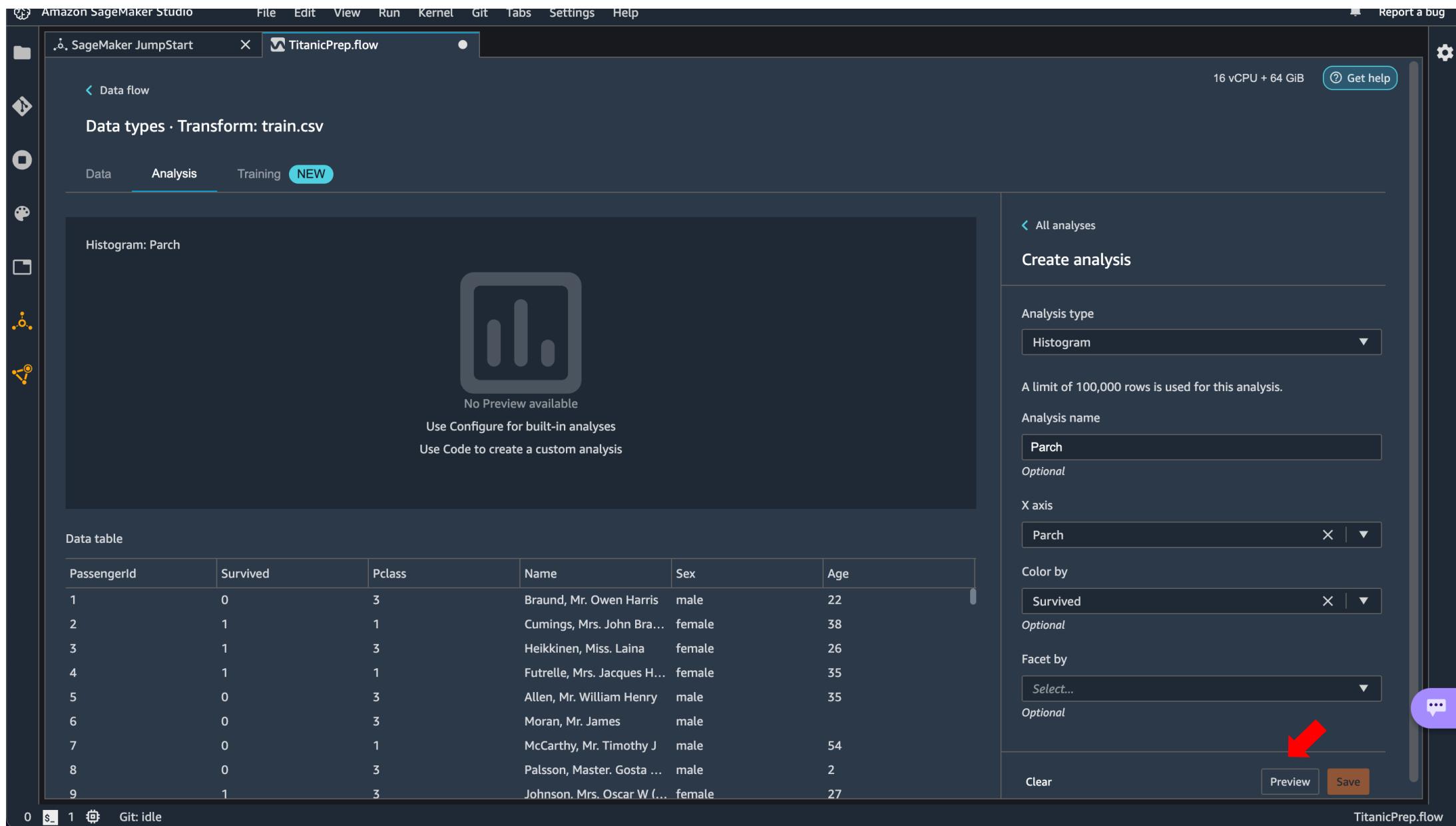
X axis: Parch

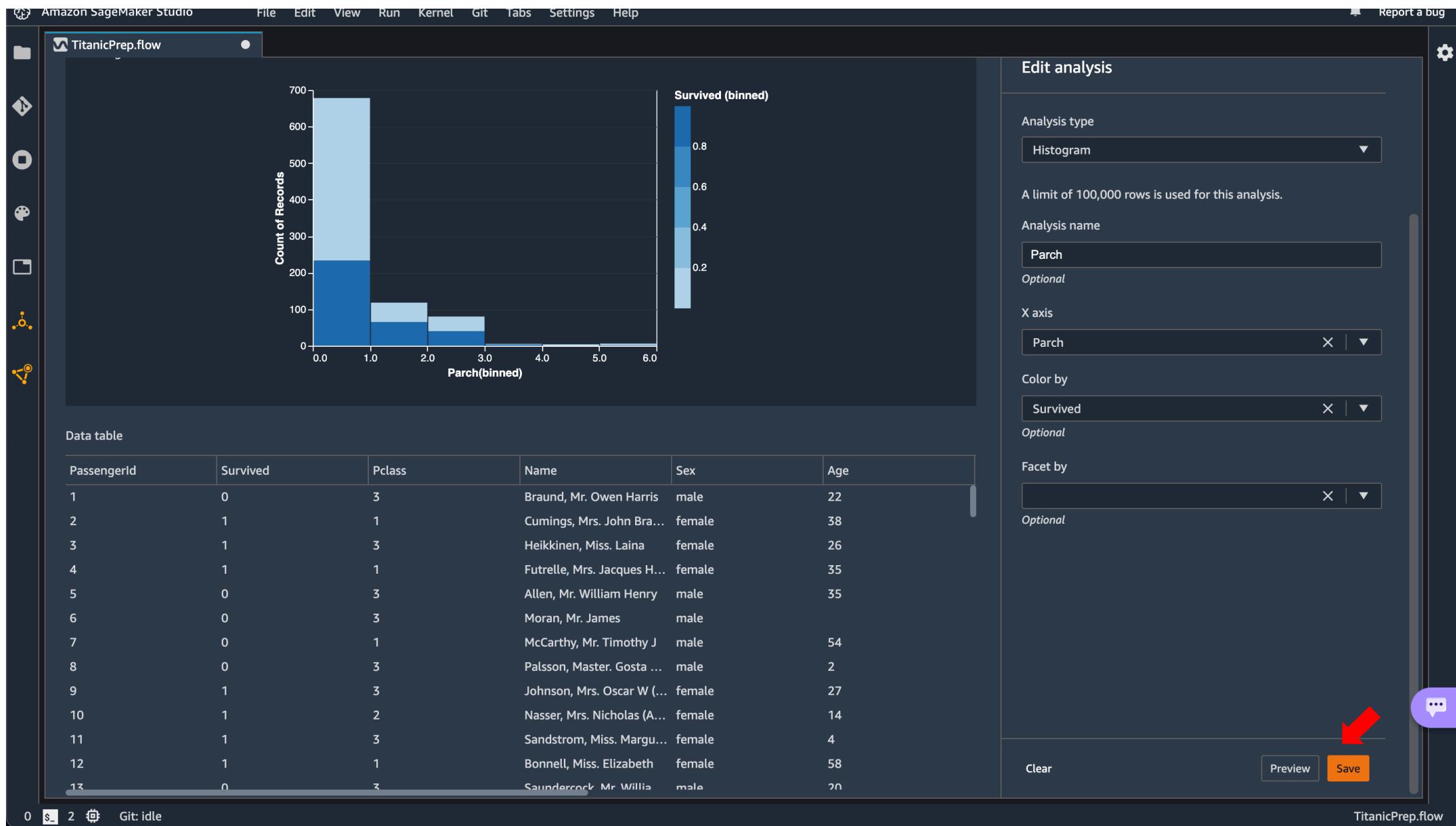
Color by: Survived

Facet by: Select...

Clear Preview Save

A red arrow points to the "Color by" dropdown menu, which contains the value "Survived".





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

.δ. SageMaker JumpStart X TitanicPrep.flow • 16 vCPU + 64 GiB Get help

Data flow 

Data types · Transform: train.csv

Data Analysis Training NEW

Create new analysis

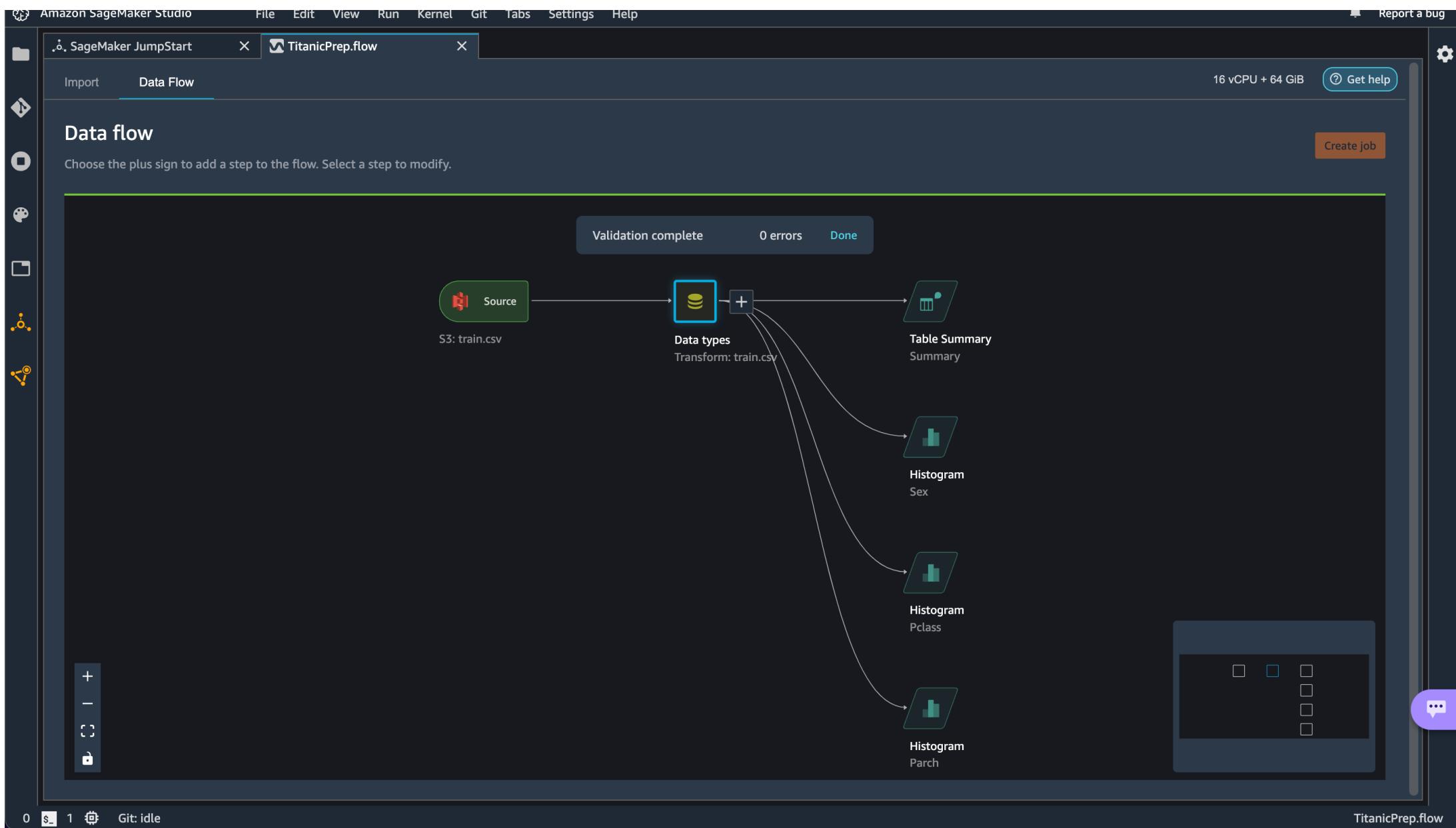
Summary Table Summary

Sex Histogram Sex · Survived

Pclass Histogram Pclass · Survived

Parch Histogram Parch · Survived

0 \$ 1 Git: idle TitanicPrep.flow



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

.SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Import Data Flow 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

S3: train.csv

Source → [Data types] → + → Table Summary

Add transform → Train model NEW

Add analysis

Get data insights

Add destination >

Export to >

Join

Concatenate

Edit

Histogram Sex

Histogram Pclass

Histogram Parch

0 \$ 1 Git: idle TitanicPrep.flow

The screenshot shows the Amazon SageMaker Studio Data Flow interface. A pipeline is defined starting from an 'S3: train.csv' source node, followed by a 'Transform' node, and finally a 'Table Summary' destination node. A context menu is open over the 'Transform' node, with the 'Add transform' option highlighted by a red arrow. Other options in the menu include 'Add analysis', 'Train model NEW', 'Get data insights', 'Add destination >', 'Export to >', 'Join', 'Concatenate', and 'Edit'. To the right of the menu, three histogram visualizations are displayed for the columns 'Sex', 'Pclass', and 'Parch'. The top navigation bar includes tabs for 'Import' and 'Data Flow', along with resource information like '16 vCPU + 64 GiB' and a 'Get help' button. The bottom navigation bar shows the current file as 'TitanicPrep.flow'.

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

.s. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▼ 2. Data 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 Update

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

[Data flow](#)

Data types · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ADD TRANSFORM

Search transforms

CUSTOM

Custom formula
Define a new column using a Spark SQL expression to query data in the current dataframe.

Custom transform
Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations.

STANDARD

Balance data
Balance the data for binary classification problems using random oversampling, random undersampling or SMOTE.

Dimensionality Reduction
For the top K principal components, trains a model to project vectors to a lower dimensional space.

Encode categorical
Convert categorical variables to numeric or vector representations.

Featurize date/time
Encode date/time values to numeric and vector representations.

Featurize text

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ADD TRANSFORM

missing

RESULTS

Handle missing

Replace, drop, or add indicators for missing values.

Encode categorical

Convert categorical variables to numeric or vector representations.

Related: missing

Search and edit

Find, replace, split, and otherwise transform input string values using search and edit functions.

Related: missing

Time Series

Transformers to preprocess and manipulate time series.

Related: missing

...

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

HANDLE MISSING

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

Transform [i](#)

Impute

Column type [i](#)

Numeric

Input columns

Select...

Imputing strategy [i](#)

Approximate Median

Output column [i](#)

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

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	Saunderscock, 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 and train Export data

HANDLE MISSING

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

Transform [i](#)

Impute

Column type [i](#)

Numeric

Input columns

Age [X](#)

Imputing strategy [i](#)

Approximate Median

Output column [i](#)

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

HANDLE MISSING

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

Transform [i](#)

Impute

Column type [i](#)

Numeric

Input columns

Age [X](#)

Imputing strategy [i](#)

Approximate Median

Output column [i](#)

Age_Imputed

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Data types · Transform: train.csv

Data Analysis Training NEW

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	Saunderscock, 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 and train Export data

HANDLE MISSING

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

Transform [i](#)
Impute

Column type [i](#)
Numeric

Input columns
Age X

Imputing strategy [i](#)
Approximate Median

Output column [i](#)
Age_Imputed

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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
239865		26		S	35

Export and train Export data

HANDLE MISSING

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

Transform [i](#)

Impute

Column type [i](#)

Numeric

Input columns

Age [X](#)

Imputing strategy [i](#)

Approximate Median

Output column [i](#)

Age_Imputed

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Data types · Transform: train.csv

Data Analysis Training NEW

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
239865		26		S	35

Export and train Export data

HANDLE MISSING

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

Transform [i](#)
Impute

Column type [i](#)
Numeric

Input columns
Age X

Imputing strategy [i](#)
Approximate Median

Output column [i](#)
Age_Imputed

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.s. SageMaker JumpStart X TitanicPrep.flow ●

Data flow 16 vCPU + 64 GiB Get help

Impute · Transform: train.csv

Data Analysis Training NEW

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

Export and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▶ 2. Data types

▼ 3. Impute ...

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

Transform i

Impute

Column type i

Numeric

Input columns

Age X

Imputing strategy i

Approximate Median

Output column i

Age_Imputed

Optional

Clear Preview Update

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow 16 vCPU + 64 GiB Get help

Impute · Transform: train.csv

Data Analysis Training NEW

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

Export and train Export data

ALL STEPS

- + Add step
- ▶ 1. S3 Source
- ▶ 2. Data types
- ▶ 3. Impute

0 \$ 1 Git: idle Saving started TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow 16 vCPU + 64 GiB Get help

Impute · Transform: train.csv

Data Analysis Training NEW

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	Saunderscock, 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 and train Export data

ADD TRANSFORM

columns

RESULTS

Manage columns 

Move, drop, duplicate or rename **columns** in the dataset.

Manage vectors

Expand or create vector **columns**.

Encode categorical

Convert categorical variables to numeric or vector representations.

Related: **columns**

Featurize date/time

Encode date/time values to numeric and vector representations.

Related: **columns**

Featurize text

Generate vector representations from natural language text.

Related: **columns**

Format string

Clean and prepare strings using standard string formatting operations.

Related: **columns**

Handle missing

Replace, drop, or add indicators for missing values.

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Impute · Transform: train.csv

Data Analysis Training NEW

Previewing: Manage columns

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
0	2	male	35	0	0

Export and train 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~~ ~~x~~ ~~v~~

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▶ 2. Data types

▶ 3. Impute

▼ 4. Drop column

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 Update

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X

Data flow

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

ALL STEPS

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

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

ADD TRANSFORM

Search transforms

CUSTOM

Custom formula
Define a new column using a Spark SQL expression to query data in the current dataframe.

Custom transform 
Use Pyspark, Pandas, or Pyspark (SQL) to define custom transformations.

STANDARD

Balance data
Balance the data for binary classification problems using random oversampling, random undersampling or SMOTE.

Dimensionality Reduction
For the top K principal components, trains a model to project vectors to a lower dimensional space.

Encode categorical
Convert categorical variables to numeric or vector representations.

Featurize date/time
Encode date/time values to numeric and vector representations.

Featurize text

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X

Data flow 16 vCPU + 64 GiB Get help

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

CUSTOM TRANSFORM

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

Name:

Optional:

Python (PySpark): X | ▾

Search example snippets

Your custom transform:

```
1 # Table is available as variable `df`  
2
```

Clear Preview Add

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

CUSTOM TRANSFORM

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

Name:

Optional:

|Python (Pandas) Python (PySpark) Python (Pandas) Python (User-Defined Function) SQL (PySpark SQL)

Python (Pandas) 

1 # Table is available as variable `df`
2

Clear Preview Add

0 \$ 1 Git: refreshing... TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Drop column · Transform: train.csv

Data Analysis Training NEW

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
0	2	male	35	0	0

Export and train Export data

CUSTOM TRANSFORM

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

Name: Custom Pandas

Optional

Python (Pandas)

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 PySpark or Python (User-Defined Function) for production use-cases

```
1 df["With_Family"] = ( (df["SibSp"] != 0) | (df["Parch"] != 0) ).astype(int)
2
```

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

Drop column · Transform: train.csv

Data Analysis Training NEW

Previewing: Custom transform

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 and train Export data

CUSTOM TRANSFORM

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

Name: Custom Pandas

Optional

Python (Pandas)

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) or Python (User-Defined Function) for production use-cases

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

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow ●

Data flow 16 vCPU + 64 GiB Get help

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ALL STEPS

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

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

Name: Custom Pandas
Optional
Python (Pandas)

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) or Python (User-Defined Function) for production use-cases

Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X

Data flow 16 vCPU + 64 GiB Get help

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ALL STEPS

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

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ADD TRANSFORM

RESULTS

Encode categorical

Convert categorical variables to numeric or vector representations.

Featurize date/time

Encode date/time values to numeric and vector representations.

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart TitanicPrep.flow

Data flow 16 vCPU + 64 GiB Get help

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ENCODE CATEGORICAL

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

Transform | One-hot encode

Ordinal encode

One-hot encode

Similarity encode

Invalid handling strategy

Keep

Drop last

Output style

Vector

Output column

Optional

Clear

Preview Add

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train 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](#)

Keep

Drop last [i](#)

Output style [i](#)

Vector

Output column [i](#)

Optional

Clear Preview Add ⚙️

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train 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](#)

Vector

Output column [i](#)

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X 16 vCPU + 64 GiB Get help

Data flow Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train 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 [x](#)

Output column [i](#)

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow X Data flow 16 vCPU + 64 GiB Get help

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ENCODE CATEGORICAL

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

Transform [i](#)

One-hot encode

Input already ordinal encoded [i](#)

Input columns [i](#)

Pclass

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 Report a bug

.δ. SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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	0
38	1	0	1	1	0
26	0	1	0	0	1
35	1	0	1	0	0
35	0	1	0	0	1
28	0	1	0	0	1
54	0	0	1	0	0
2	1	1	0	0	1
27	1	1	0	0	1
14	1	0	0	0	1
4	1	1	0	0	1
58	0	0	1	0	0
20	0	1	0	0	1
39	1	1	0	0	1
14	0	1	0	0	1
55	0	0	0	0	1
2	1	1	0	0	1
28	0	0	0	0	1
31	1	1	0	0	1
28	0	1	0	0	1
35	0	0	0	0	1

Export and train 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 [x](#)

Output column [i](#)

Optional

Clear

Preview Add

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

Custom Pandas · Transform: train.csv

Data Analysis Training NEW

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	0
38	1	0	1	0	0
26	0	1	0	0	1
35	1	0	1	0	0
35	0	1	0	0	0
28	0	1	0	0	0
54	0	0	1	0	0
2	1	1	0	0	0
27	1	1	0	0	0
14	1	0	0	0	0
4	1	1	0	0	0
58	0	0	1	0	0
20	0	1	0	0	0
39	1	1	0	0	0
14	0	1	0	0	0
55	0	0	0	0	0
2	1	1	0	0	0
28	0	0	0	0	0
31	1	1	0	0	0
28	0	1	0	0	0
35	0	0	0	0	0

Export and train 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 [x](#)

Output column [i](#)

Optional

Clear

Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow ●

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▶ 2. Data types

▶ 3. Impute

▶ 4. Drop column

▶ 5. Custom Pandas

▼ 6. One-hot encode

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

Transform *One-hot encode*

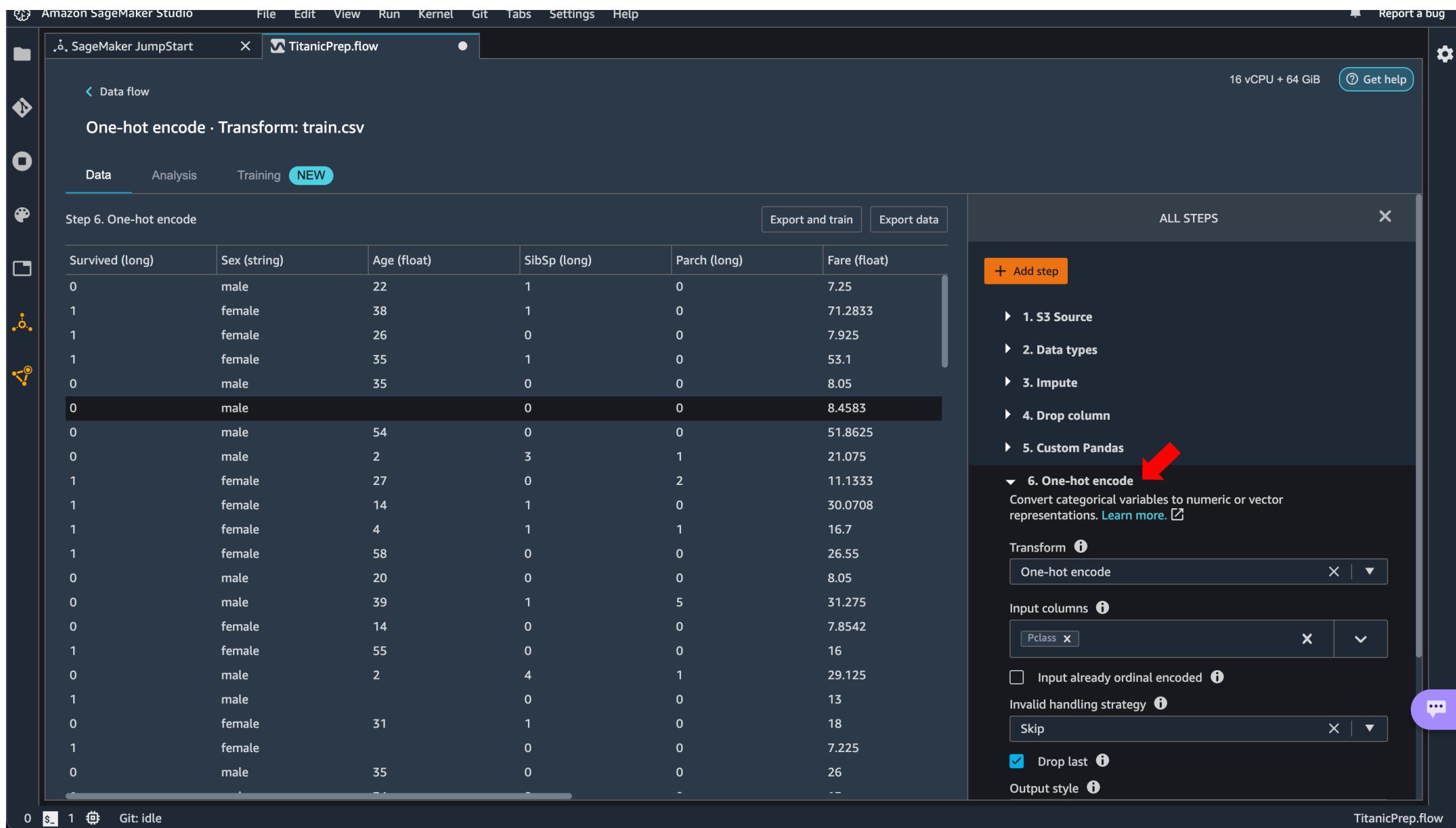
Input columns *Pclass*

Input already ordinal encoded *i*

Invalid handling strategy *Skip*

Drop last *i*

Output style *i*



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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ALL STEPS

+ Add step

▶ 1. S3 Source

▶ 2. Data types

▶ 3. Impute

▶ 4. Drop column

▶ 5. Custom Pandas

▶ 6. One-hot encode

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow 16 vCPU + 64 GiB Get help

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ADD TRANSFORM

RESULTS

Encode categorical 

Convert categorical variables to numeric or vector representations.

Featurize date/time

Encode date/time values to numeric and vector representations.

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow ●

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

ENCODE CATEGORICAL

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

Transform [i](#) One-hot encode

Input columns [i](#) Sex

Input already ordinal encoded [i](#)

Invalid handling strategy [i](#) Skip

Drop last [i](#)

Output style [i](#) Columns

Output column [i](#)

Optional

Clear Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow ●

16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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			0	
0		0		1	

Export and train 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 [...](#)

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow •

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Previewing: Encode categorical

·_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			0	
0		0		1	

Export and train 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 [...](#)

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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
0	35	0	0	26	S

Export and train Export data

ALL STEPS

+ Add step

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

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

Transform One-hot encode

Input columns Sex

Input already ordinal encoded

Invalid handling strategy Skip

Drop last

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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	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	0	13	S
0	31	1	0	18	S
1	0	0	0	7.225	C
0	35	0	0	26	S

Export and train Export data

ALL STEPS

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

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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	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	0	13	S
0	31	1	0	18	S
1	0	0	0	7.225	C
0	35	0	0	26	S

Export and train Export data

ADD TRANSFORM

encode

RESULTS

Encode categorical 

Convert categorical variables to numeric or vector representations.

Featurize date/time

Encode date/time values to numeric and vector representations.

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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	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	0	13	S
0	31	1	0	18	S
1	0	0	0	7.225	C
0	35	0	0	26	S

Export and train Export data

ENCODE CATEGORICAL

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 Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Previewing: Encode categorical

ss_2 (float)	Sex_male (float)	Sex_female (float)	Embarked_S (float)	Embarked_C (float)	Embarked_Q (float)
1			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			0	0	
1			1	0	
1			1	0	
0			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			1	0	
0			1	0	
0			1	0	
1			0	0	
1			1	0	
0			0	1	
0			0	0	
1			1	0	

Export and train Export data

ENCODE CATEGORICAL

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

Preview Add

Clear

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X

16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Previewing: Encode categorical

ss_2 (float)	Sex_male (float)	Sex_female (float)	Embarked_S (float)	Embarked_C (float)	Embarked_Q (float)
1			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			0	0	
1			1	0	
1			1	0	
0			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			1	0	
0			1	0	
0			1	0	
1			0	0	
1			1	0	
0			0	1	
0			0	0	
1			1	0	

Export and train Export data

ENCODE CATEGORICAL

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

Preview Add

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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	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	0	13	28
0	31	1	0	18	31
1	0	0	0	7.225	28
0	35	0	0	26	35

Export and train Export data

ALL STEPS

- + 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
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  

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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	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	0	13	28
0	31	1	0	18	31
1	0	0	0	7.225	28
0	35	0	0	26	35

Export and train Export data

ALL STEPS

- + 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

0 \$ 1 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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
0	35	0	0	26	35

Export and train Export data

ADD TRANSFORM

RESULTS

Handle missing
Replace, **drop**, or add indicators for missing values.

Manage columns

Move, **drop**, duplicate or rename columns in the dataset.

Manage rows
Sort, shuffle or **drop** duplicate rows.

Time Series
Transformers to preprocess and manipulate time series.

Related: **drop**

Encode categorical
Convert categorical variables to numeric or vector representations.

Related: **drop**

Featurize text
Generate vector representations from natural language text.

Related: **drop**

Manage vectors
Expand or create vector columns.

Related: **drop**

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Step 8. One-hot encode

Age (float)	SibSp (long)	Parch (long)	Fare (float)	Age_Imputed (float)	With_Family (long)
22	1	0	7.25	22	1
38	1	0	71.2833	38	1
26	0	0	7.925	26	0
35	1	0	53.1	35	1
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	1
14	0	0	7.8542	14	0
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	0
...

Export and train Export data

MANAGE COLUMNS

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

Transform Drop column

Columns to drop

Age	X
-----	---

Name Type

FLOAT (11)

<input checked="" type="checkbox"/> Age	float
<input type="checkbox"/> Age_Imputed	float
<input type="checkbox"/> Embarked_C	float
<input type="checkbox"/> Embarked_Q	float
<input type="checkbox"/> Embarked_S	float
<input type="checkbox"/> Fare	float
<input type="checkbox"/> Pclass_1	float
<input type="checkbox"/> Pclass_2	float

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Step 8. One-hot encode

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

Export and train Export data

MANAGE COLUMNS

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

Transform *i*

Drop column

Columns to drop

Age <input checked="" type="checkbox"/>	Pclass_2 <input checked="" type="checkbox"/>	X	v
---	--	---	---

<input type="checkbox"/> Age_Imputed	float
<input type="checkbox"/> Embarked_C	float
<input type="checkbox"/> Embarked_Q	float
<input type="checkbox"/> Embarked_S	float
<input type="checkbox"/> Fare	float
<input type="checkbox"/> Pclass_1	float
<input checked="" type="checkbox"/> Pclass_2	float
<input type="checkbox"/> Pclass_3	float
<input type="checkbox"/> Sex_female	float
<input type="checkbox"/> Sex_male	float

Git: refreshing... TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Step 8. One-hot encode

ass_1 (float)	Pclass_2 (float)	Sex_male (float)	Sex_female (float)	Embarked_S (float)	Embarked_C (float)
1				1	0
0				0	1
0				1	0
0				1	0
1				1	0
1				0	0
1				0	0
1				1	0
0				1	0
0				0	1
0				1	0
0				1	0
1				1	0
1				1	0
0				1	0
0				1	0
1				0	0
1				1	0
0				0	1
0				1	0
1				1	0

Export and train Export data

MANAGE COLUMNS

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

Transform Drop column

Columns to drop

- Age x Pclass_2 x Sex_female x

<input type="checkbox"/> Fare	float
<input type="checkbox"/> Pclass_1	float
<input checked="" type="checkbox"/> Pclass_2	float
<input type="checkbox"/> Pclass_3	float
<input checked="" type="checkbox"/> Sex_female	float
<input type="checkbox"/> Sex_male	float
LONG (4)	
<input type="checkbox"/> Parch	long
<input type="checkbox"/> SibSp	long
<input type="checkbox"/> Survived	long

Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X 📈 TitanicPrep.flow X 📈 TitanicPrep.ipynb X

16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

Step 8. One-hot encode

ss_2 (float)	Sex_male (float)	Sex_female (float)	Embarked_S (float)	Embarked_C (float)	Embarked_Q (float)
1			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			0	0	
1			1	0	
1			1	0	
0			1	0	
0			0	1	
0			1	0	
0			1	0	
1			1	0	
1			1	0	
0			1	0	
0			1	0	
1			0	0	
1			1	0	
0			1	0	
0			0	1	
1			1	0	

Export and train Export data

MANAGE COLUMNS

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

Transform [i](#)

Drop column

Columns to drop

Age	Pclass_2	Sex_female
		Embarked_Q

Age_Imputed float
 Embarked_C float
 Embarked_Q float
 Embarked_S float
 Fare float
 Pclass_1 float
 Pclass_2 float
 Pclass_3 float
 Sex_female float

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB

← Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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
0	35	0	0	26	35

Export and train Export data

MANAGE COLUMNS

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

Transform

Columns to drop

Age Pclass_2 Sex_female
 Embarked_Q Parch

Column	Type
Fare	float
Pclass_1	float
<input checked="" type="checkbox"/> Pclass_2	float
<input type="checkbox"/> Pclass_3	float
<input checked="" type="checkbox"/> Sex_female	float
<input type="checkbox"/> Sex_male	float
LONG (4)	
<input checked="" type="checkbox"/> Parch	long
<input type="checkbox"/> SibSp	long
<input type="checkbox"/> Survived	long
<input type="checkbox"/> With_Family	long

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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
0	35	0	0	26	35

Export and train Export data

MANAGE COLUMNS

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

Transform

Drop column

Columns to drop

Age Pclass_2 Sex_female Embarked_Q Parch + 1 item selected

Column	Type
Fare	float
Pclass_1	float
<input checked="" type="checkbox"/> Pclass_2	float
<input type="checkbox"/> Pclass_3	float
<input checked="" type="checkbox"/> Sex_female	float
<input type="checkbox"/> Sex_male	float
LONG (4)	
<input checked="" type="checkbox"/> Parch	long
<input checked="" type="checkbox"/> SibSp	long
<input type="checkbox"/> Survived	long
<input type="checkbox"/> With_Family	long

0 \$ 2 Git: idle TitanicPrep.ipynb

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

Data flow

One-hot encode · Transform: train.csv

Data Analysis Training NEW

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 and train Export data

MANAGE COLUMNS

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

Transform Drop column

Columns to drop

Age Pclass_2 Sex_female
Embarked_Q Parch + 1 item selected

Clear Preview Add

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X

16 vCPU + 64 GiB

← Data flow

One-hot encode · Transform: train.csv

Data Analysis Training **NEW**

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 and train Export data

MANAGE COLUMNS

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

Transform Drop column

Columns to drop

Age Pclass_2 Sex_female
Embarked_Q Parch + 1 item selected

Clear Preview **Add**

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow ● TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

Data flow

Drop column · Transform: train.csv

Data Analysis Training NEW

Step 9. Drop column

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 and train Export data

ALL STEPS

- + 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
- 9. Drop column

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

Transform *i*

Drop column

Columns to drop

Age X Pclass_2 X Sex_female X Embarked_Q X Parch X + 1 item selected

Clear Preview Update

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X □ TitanicPrep.flow ● □ TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

◀ Data flow

Drop column · Transform: train.csv

Data Analysis Training NEW

Step 9. Drop column

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 and train Export data

ALL STEPS

- + 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
- ▶ 9. Drop column

Git: refreshing... TitanicPrep.flow

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

.SageMaker JumpStart X TitanicPrep.flow . TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

Import Data Flow 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

```
graph LR; S[S3: train.csv] --> TS[Table Summary]; TS --> H1[Histogram Sex]; TS --> H2[Histogram Pclass]; TS --> H3[Histogram Parch]; H1 --> I[Impute Transform: train.csv]; I --> D1[Drop column Transform: train.csv]; D1 --> C[Custom Pandas Transform: train.csv]; C --> O1[One-hot encode Transform: train.csv]; O1 --> O2[One-hot encode Transform: train.csv]; O2 --> O3[One-hot encode Transform: train.csv]; O3 --> D2[Drop column Transform: train.csv];
```

0 \$ 2 Git: idle TitanicPrep.flow

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep.ipynb X 16 vCPU + 64 GiB Get help

Import Data Flow 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

S3: train.csv

Source
Transform: train.csv

Table Summary
Summary

Histogram
Sex

Histogram
Pclass

Histogram
Parch

Validation complete 0 errors Done

Impute
Transform: train.csv

Drop column
Transform: train.csv

Custom Pandas
Transform: train.csv

One-hot encode
Transform: train.csv

One-hot encode
Transform: train.csv

One-hot encode
Transform: train.csv

Drop column
Transform: train.csv

Add transform

Add analysis

Train model NEW

Get data insights

Add destination >

Export to > red arrow

Join

Concatenate

Delete

Amazon S3 (via Jupyter Notebook)

SageMaker Pipelines (via Jupyter Notebook)

Python Code

SageMaker Feature Store (via Jupyter Notebook)

0 \$ 2 Git: idle TitanicPrep.flow

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

.SageMaker JumpStart X | TitanicPrep.flow X | TitanicPrep1.ipynb X | ● | TitanicPrep1.ipynb X |

File + X D C Markdown \$ git 2 vCPU + 4 GiB Cluster Python 3 (Data Science) Share

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

0 \$ 3 Git: idle Python 3 (Data Science) | Idle Kernel: Idle | Instance MEM Mode: Command X Ln 1, Col 1 TitanicPrep1.ipynb

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

.SageMaker JumpStart TitanicPrep.flow **TitanicPrep1.ipynb** TitanicPrep.ipynb

2 vCPU + 4 GiB Cluster Python 3 (Data Science) Share

Upload Flow to S3

To use the Data Wrangler as an input to the processing job, first upload your flow file to Amazon S3.

```
[6]: import os
import json
import boto3

# name of the flow file which should exist in the current notebook working directory
flow_file_name = "TitanicPrep.flow"

# Load .flow file from current notebook working directory
!echo "Loading flow file from current notebook working directory: $PWD"

with open(flow_file_name) as f:
    flow = json.load(f)

# Upload flow to S3
s3_client = boto3.client("s3")
s3_client.upload_file(flow_file_name, bucket, f"data_wrangler_flows/{flow_export_name}.flow", ExtraArgs={"ServerSideEncryption": "aws:kms"})

flow_s3_uri = f"s3://{bucket}/data_wrangler_flows/{flow_export_name}.flow"

print(f"Data Wrangler flow {flow_file_name} uploaded to {flow_s3_uri}")

Loading flow file from current notebook working directory: /root
Data Wrangler flow TitanicPrep.flow uploaded to s3://sagemaker-us-east-1-200409934317/data_wrangler_flows/flow-09-07-49-00-8c9fe661.flow
```

The Data Wrangler Flow is also provided to the Processing Job as an input source which we configure below.

```
[7]: ## Input - Flow: TitanicPrep.flow
flow_input = ProcessingInput(
    source=flow_s3_uri,
    destination="/opt/ml/processing/flow",
    input_name="flow",
    s3_data_type="S3Prefix",
    s3_input_mode="File",
    s3_data_distribution_type="FullyReplicated"
)
```

0 \$ 3 Git: idle Python 3 (Data Science) | Busy Kernel: Busy | Instance MEM Mode: Command Ln 1, Col 1 TitanicPrep1.ipynb

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

.SageMaker JumpStart TitanicPrep.flow TitanicPrep.ipynb TitanicPrep1.ipynb

2 vCPU + 4 GiB Cluster Python 3 (Data Science) Share

Create Processing Job

To launch a Processing Job, you will use the SageMaker Python SDK to create a Processor function.

```
[9]: from sagemaker.processing import Processor
from sagemaker.network import NetworkConfig

processor = Processor(
    role=iam_role,
    image_uri=container_uri,
    instance_count=instance_count,
    instance_type=instance_type,
    volume_size_in_gb=volume_size_in_gb,
    network_config=NetworkConfig(enable_network_isolation=enable_network_isolation),
    sagemaker_session=sess,
    output_kms_key=kms_key,
    tags=user_tags
)

# Start Job
processor.run(
    inputs=[flow_input] + data_sources,
    outputs=[processing_job_output],
    arguments=[f"--output-config '{json.dumps(output_config)}'" + [f"--refit-trained-params '{json.dumps(refit_trained_params)}'"],
    wait=False,
    logs=False,
    job_name=processing_job_name
)
```

Job Name: data-wrangler-flow-processing-09-07-49-00-8c9fe661
Inputs: [{"InputName": "flow", "AppManaged": False, "S3Input": {"S3Uri": "s3://sagemaker-us-east-1-200409934317/data_wrangler_flows/flow-09-07-49-00-8c9fe661.flow", "LocalPath": "/opt/ml/processing/flow", "S3DataType": "S3Prefix", "S3InputMode": "File", "S3DataDistributionType": "FullyReplicated", "S3CompressionType": "None"}}, {"InputName": "train.csv", "AppManaged": False, "S3Input": {"S3Uri": "s3://titanic-dataset-20221009/train.csv", "LocalPath": "/opt/ml/processing/train.csv", "S3DataType": "S3Prefix", "S3InputMode": "File", "S3DataDistributionType": "FullyReplicated", "S3CompressionType": "None"}}]
Outputs: [{"OutputName": "e60b1970-d5df-494d-bcde-ac792c400813.default", "AppManaged": False, "S3Output": {"S3Uri": "s3://sagemaker-us-east-1-200409934317/export-flow-09-07-49-00-8c9fe661/output", "LocalPath": "/opt/ml/processing/output", "S3UploadMode": "EndOfJob"}}]

0 \$ 3 Git: refreshing... Python 3 (Data Science) | Busy Kernel: Busy | Instance MEM Mode: Command Report a bug Ln 1, Col 1 TitanicPrep1.ipynb

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

.δ. SageMaker JumpStart X TitanicPrep.flow X TitanicPrep1.ipynb X TitanicPrep.ipynb X

2 vCPU + 4 GiB Cluster Python 3 (Data Science) Share

Job Status & S3 Output Location

Below you wait for processing job to finish. If it finishes successfully, the raw parameters used by the Processing Job will be printed

```
[10]: s3_job_results_path = f"s3://{bucket}/{s3_output_prefix}/{processing_job_name}"
print(f"Job results are saved to S3 path: {s3_job_results_path}")

job_result = sess.wait_for_processing_job(processing_job_name)
job_result

Job results are saved to S3 path: s3://sagemaker-us-east-1-200409934317/export-flow-09-07-49-00-8c9fe661/output/data-wrangler-flow-processing-09-07-49-00-8c9fe661
.....
[10]: {'ProcessingInputs': [{'InputName': 'flow',
   'AppManaged': False,
   'S3Input': {'S3Uri': 's3://sagemaker-us-east-1-200409934317/data_wrangler_flows/flow-09-07-49-00-8c9fe661.flow',
   'LocalPath': '/opt/ml/processing/flow',
   'S3DataType': 'S3Prefix',
   'S3InputMode': 'File',
   'S3DataDistributionType': 'FullyReplicated',
   'S3CompressionType': 'None'}},
  {'InputName': 'train.csv',
   'AppManaged': False,
   'S3Input': {'S3Uri': 's3://titanic-dataset-20221009/train.csv',
   'LocalPath': '/opt/ml/processing/train.csv',
   'S3DataType': 'S3Prefix',
   'S3InputMode': 'File',
   'S3DataDistributionType': 'FullyReplicated',
   'S3CompressionType': 'None'}},
  'ProcessingOutputConfig': {'Outputs': [{'OutputName': 'e60b1970-d5df-494d-bcde-ac792c400813.default',
   'S3Output': {'S3Uri': 's3://sagemaker-us-east-1-200409934317/export-flow-09-07-49-00-8c9fe661/output',
   'LocalPath': '/opt/ml/processing/output',
   'S3UploadMode': 'EndOfJob'},
   'AppManaged': False}]},
  'ProcessingJobName': 'data-wrangler-flow-processing-09-07-49-00-8c9fe661',
  'ProcessingResources': {'ClusterConfig': {'InstanceCount': 2,
   'InstanceType': 'ml.m5.4xlarge',
   'VolumeSizeInGB': 30}},
  'StoppingCondition': {'MaxRuntimeInSeconds': 86400},
  'AppSpecification': '!Tmagentbill! 1662277280841_dkr.ecr.us-east-1.amazonaws.com/sagemaker-data-wrangler-container:1.x'}
```

0 \$ 3 Git: idle Python 3 (Data Science) | Idle Kernel: Idle | Instance MEM Mode: Command X Ln 1, Col 1 TitanicPrep1.ipynb

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3.

Learn more X i

Amazon S3 > Buckets

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](#)

[C](#) [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

Name	AWS Region	Access	Creation date
titanic-dataset-20221009	US East (N. Virginia) us-east-1	Bucket and objects not public	October 9, 2022, 13:26:05 (UTC+08:00)
sagemaker-us-east-1-200409934317	US East (N. Virginia) us-east-1	Objects can be public	October 9, 2022, 12:52:06 (UTC+08:00)
sagemaker-studio-0f2b2f90	US East (N. Virginia) us-east-1	Objects can be public	October 9, 2022, 12:52:06 (UTC+08:00)

[Find buckets by name](#) < 1 > [⚙️](#)



AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more X

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317

sagemaker-us-east-1-200409934317 Info

Objects Properties Permissions Metrics Management Access Points

Objects (5)

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix < 1 > ⚙

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	canvas-clarify-job-Canvas140959-2022-10-09-06-27-30/	Folder	-	-	-
<input type="checkbox"/>	canvas/	Folder	-	-	-
<input type="checkbox"/>	Canvas/	Folder	-	-	-
<input type="checkbox"/>	data_wrangler_flows/	Folder	-	-	-
<input type="checkbox"/>	export-flow-09-07-49-00-8c9fe661/	Folder	-	-	-



Feedback Looking for language selection? Find it in the new Unified Settings. © 2022 Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/

export-flow-09-07-49-00-8c9fe661/

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](#)

C Copy S3 URI Copy URL Download Open Delete Actions Create folder

Upload

Find objects by prefix 1 /

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	output/	Folder	-	-	-

A red arrow points to the 'output/' folder in the table.

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/ > output/

output/

Objects Properties

Copy S3 URI

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix

Name Type Last modified Size Storage class

data-wrangler-flow-processing-09-07-49-00-8c9fe661/ Folder - - -

A red arrow points to the 'data-wrangler-flow-processing-09-07-49-00-8c9fe661/' folder entry in the list.

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/ > output/ > data-wrangler-flow-processing-09-07-49-00-8c9fe661/

data-wrangler-flow-processing-09-07-49-00-8c9fe661/

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
e60b1970-d5df-494d-bcde-ac792c400813/	Folder	-	-	-

Feedback Looking for language selection? Find it in the new Unified Settings © 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences



AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/ > output/ > data-wrangler-flow-processing-09-07-49-00-8c9fe661/ > e60b1970-d5df-494d-bcde-ac792c400813/

e60b1970-d5df-494d-bcde-ac792c400813/

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix

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

Feedback Looking for language selection? Find it in the new Unified Settings

© 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more X

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/ > output/ > data-wrangler-flow-processing-09-07-49-00-8c9fe661/ > e60b1970-d5df-494d-bcde-ac792c400813/ > default/ default/

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
part-00000-85e5831f-c1f5-4131-8429-8f842a6ad693-c000.csv	csv	October 9, 2022, 15:55:03 (UTC+08:00)	30.5 KB	Standard

A red arrow points to the 'Name' column header.

AWS Services Search for services, features, blogs, docs, and more [Option+S] Global TeamRole/MasterKey @ 2004-0993-4317

EC2 VPC Support

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

Follow security best practices for S3. Learn more X

Amazon S3 > Buckets > sagemaker-us-east-1-200409934317 > export-flow-09-07-49-00-8c9fe661/ > output/ > data-wrangler-flow-processing-09-07-49-00-8c9fe661/ > e60b1970-d5df-494d-bcde-ac792c400813/ > default/ default/

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](#)

C Copy S3 URI Copy URL Download Open Actions Create folder

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
part-00000-85e5831f-c1f5-4131-8429-8f842a6ad693-c000.csv	csv	October 9, 2022, 15:55:03 (UTC+08:00)	30.5 KB	Standard

Feedback Looking for language selection? Find it in the new [Unified Settings](#) © 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

part-00000-85e....csv Show All X

Paste B I U Merge & Center \$ % , .⁰⁰ Conditional Formatting as Table Cell Styles Insert Delete Format Sort & Filter Find & Select Analyze Data

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.

A1 Survived

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Survived	Fare	Age_Impute	With_Family	Pclass_3	Pclass_1	Sex_male	Embarked_S	Embarked_C													
2	0	7.25	22	1	1	0	1	1	1													
3	1	71.2833	38	1	0	1	0	0	0													
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	1	1	1													
7	0	8.4583	28	0	1	0	1	0	0													
8	0	51.8625	54	0	0	1	1	1	1													
9	0	21.075	2	1	1	0	1	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	1	1	1													
15	0	31.275	39	1	1	0	1	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	1	0	0													
19	1	13	28	0	0	0	1	1	0													
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	0	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													
39	0	8.05	21	0	1	0	1	1	1													
40	0	10	10	1	1	0	0	0	1													

part-00000-85e5831f-c1f5-4131-8

Ready Accessibility: Unavailable

100%

Thank you!

Michael Lin

linmicht@amazon.com



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.