



1

A screenshot of the AWS SageMaker console. The left sidebar shows navigation options like Dashboard, Search, SageMaker Domain (Studio, RStudio, Canvas), Images, and various AI services. A search bar at the top is set to 's3'. The main content area displays search results for 's3' under the 'Services' category. It lists S3 (Scalable Storage in the Cloud), S3 Glacier (Archive Storage in the Cloud), Athena (Query Data in S3 using SQL), and AWS Snow Family (Large Scale Data Transport). Below this, under 'Features', it lists the Amazon S3 File Gateway (Storage Gateway feature) and Datasets (IoT Analytics feature). On the right side, there are buttons for 'Add user', 'Launch app', 'Delete Domain', and 'Edit Settings', along with information about the authentication method (AWS Identity and Access Management (IAM)).

Search results for 's3'

Services

- S3 ★ Scalable Storage in the Cloud
- S3 Glacier ★ Archive Storage in the Cloud
- Athena ★ Query Data in S3 using SQL
- AWS Snow Family ★ Large Scale Data Transport

Features

- Amazon S3 File Gateway
- Datasets

Add user

Launch app

Delete Domain

Edit Settings

Authentication method
AWS Identity and Access Management (IAM)

2

1

Event Engine - Team Dashboard | Amazon SageMaker | S3 bucket | JupyterLab | Introducing Amazon SageMak... | Paused

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

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

3

Bucket Versioning

Disable
 Enable

Tags (0) - *optional*
Track storage cost or other criteria by tagging your bucket. [Learn more](#)

No tags associated with this bucket.
[Add tag](#)

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

Server-side encryption

Disable
 Enable

▶ Advanced settings

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

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

4

2

The screenshot shows the AWS S3 Management Console interface. At the top, there is a green success message: "Successfully created bucket 'titanic-michlin-20220218'. To upload files and folders, or to configure additional bucket settings choose View details." Below this, a blue banner says "Read the S3 resources page for documentation and technical content." On the left sidebar, under the "Buckets" section, the newly created bucket is listed. The main content area displays an "Account snapshot" with a link to "View Storage Lens dashboard". A table lists three buckets, including the newly created one:

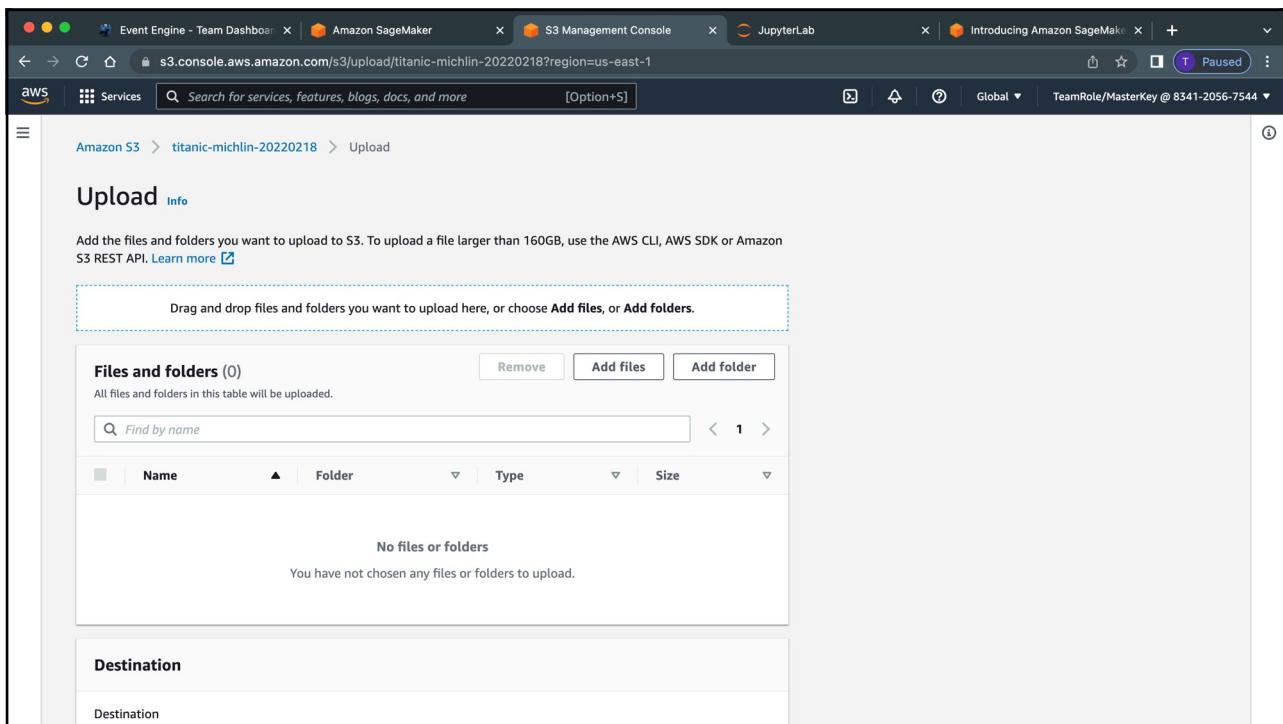
Name	AWS Region	Access	Creation date
sagemaker-studio-834120567544-5zbey5q71a	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 20:55:19 (UTC+08:00)
sagemaker-us-east-1-834120567544	US East (N. Virginia) us-east-1	Objects can be public	February 16, 2022, 21:06:52 (UTC+08:00)
titanic-michlin-20220218	US East (N. Virginia) us-east-1	Bucket and objects not public	February 16, 2022, 22:15:58 (UTC+08:00)

5

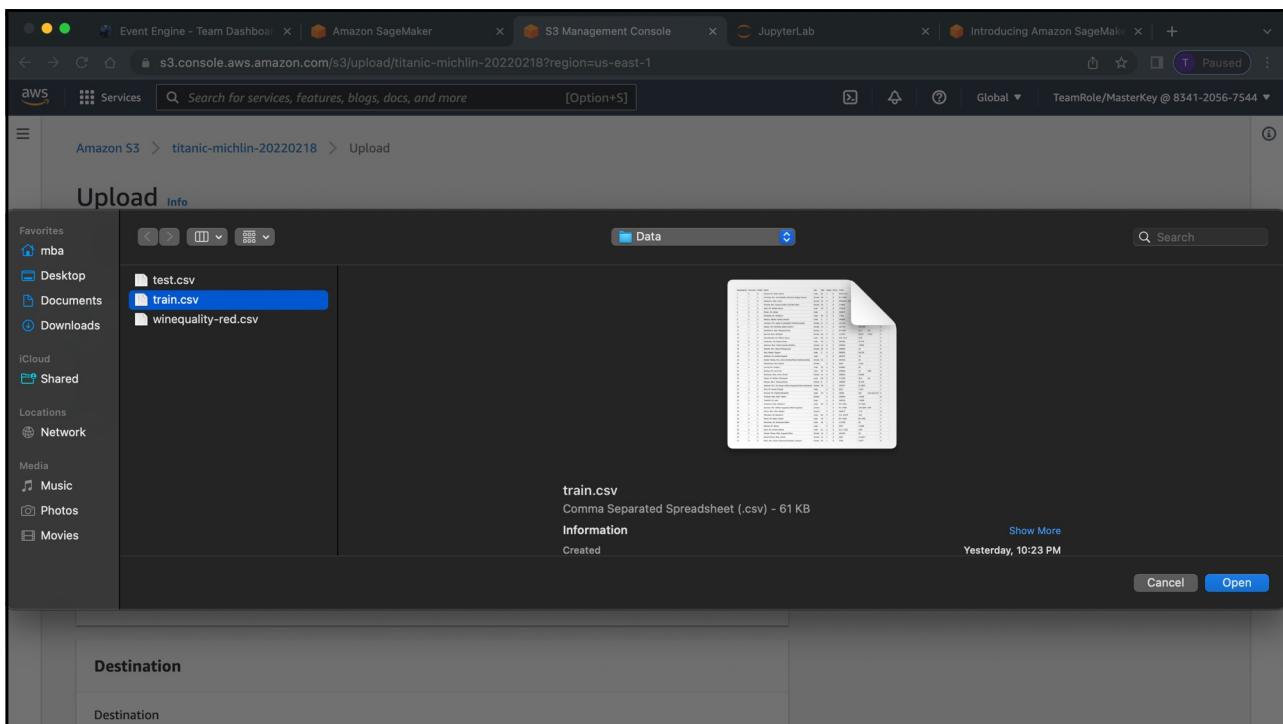
The screenshot shows the AWS S3 Management Console interface, specifically the "Objects" tab for the bucket "titanic-michlin-20220218". The left sidebar is identical to the previous screenshot. The main content area shows a table with zero rows, indicating "No objects". There is a single "Upload" button at the bottom.

6

3



7



8

4

The screenshot shows the AWS S3 Management Console interface. At the top, there are several tabs: Event Engine - Team Dashboard, Amazon SageMaker, S3 Management Console, JupyterLab, and Introducing Amazon SageMaker. The main window title is "Upload" under "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.

9

The screenshot shows the AWS S3 Management Console interface, similar to the previous one but with additional sections at the bottom.

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.

▶ Permissions

Grant public access and access to other AWS accounts.

▶ Properties

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

Cancel **Upload**

10

5

The screenshot shows a browser window with multiple tabs open, including Event Engine - Team Dashboard, Amazon SageMaker, S3 Management Console, JupyterLab, and Introducing Amazon SageMaker. The active tab is the S3 Management Console, displaying a successful upload of a file named 'train.csv' to the bucket 's3://titanic-michlin-20220218'. The upload status is shown as 'Succeeded' with a green circular icon and '1 file, 59.8 KB (100.00%)'. There are no failed files.

11

The screenshot shows a browser window with multiple tabs open, including Event Engine - Team Dashboard, Amazon SageMaker, Amazon SageMaker Immersion, and console.aws.amazon.com. The active tab is the Amazon SageMaker Immersion search results page, which displays search results for 'sagemaker'. The results are categorized under 'Services' and 'Features'. Under 'Services', there are sections for 'Amazon SageMaker' (with sub-sections for Top features: SageMaker Studio, Autopilot, SageMaker Canvas), 'AWS Glue DataBrew', and 'Amazon SageMaker Immersion'. Under 'Features', there are sections for 'SageMaker Studio', 'SageMaker Canvas', 'Notebooks', and 'Autopilot'. A sidebar on the left lists various SageMaker domains and features.

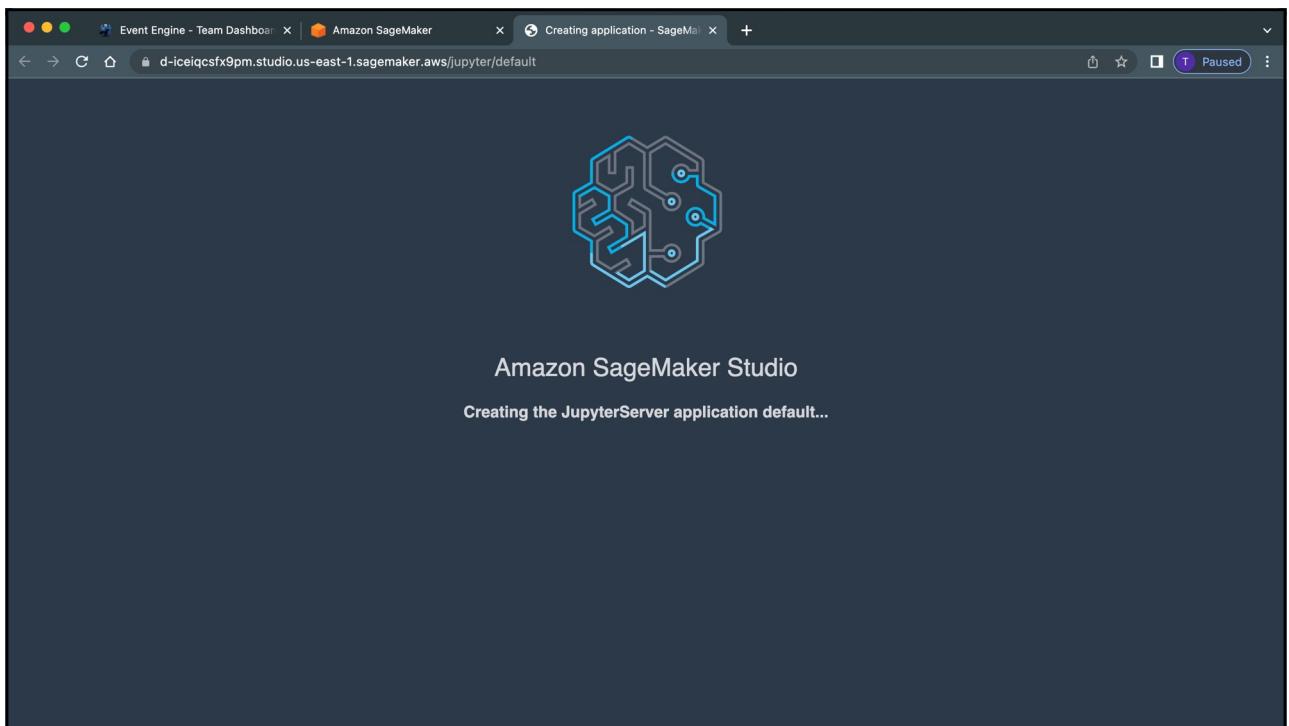
12

The screenshot shows the Amazon SageMaker Domain configuration page. On the left, there's a sidebar with navigation links for Dashboard, Search, SageMaker Domain (Studio, RStudio, Canvas), and Images (Ground Truth, Notebook, Processing, Training, Inference, Edge Manager, Augmented AI). The main content area is titled "SageMaker Domain". It shows a table with one row for a domain named "default-1645016110854". The domain is listed as "Ready". The table includes columns for Name, Modified on, Created on, Status, Domain ID, Execution role, and Authentication method. A "Launch app" button is also present. Below the table, there's a section for "Projects" which lists "Amazon SageMaker project templates enabled for this account".

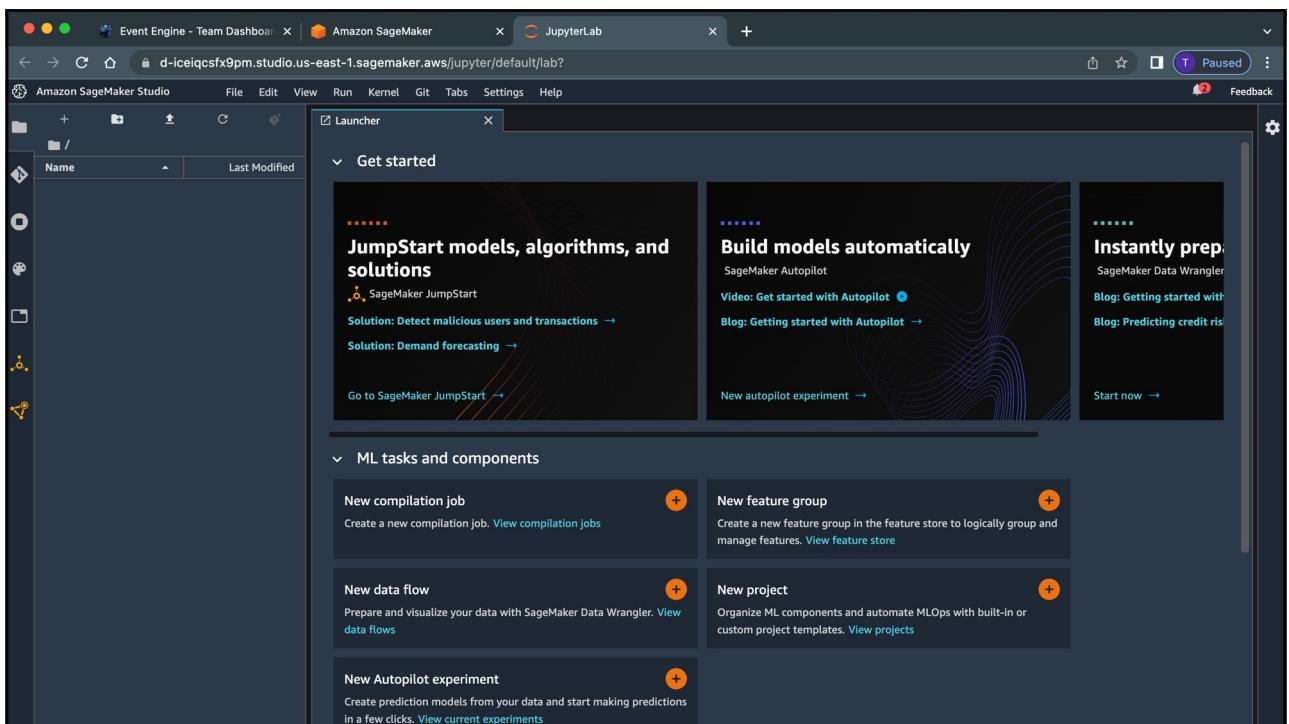
13

This screenshot is identical to the one above, showing the Amazon SageMaker Domain configuration page. However, a context menu is open over the "Launch app" button in the top right corner of the domain table. The menu options visible are "Open Link in New Tab", "Open Link in New Window", "Open Link in Incognito Window", "Save Link As...", "Copy Link Address", and "Inspect".

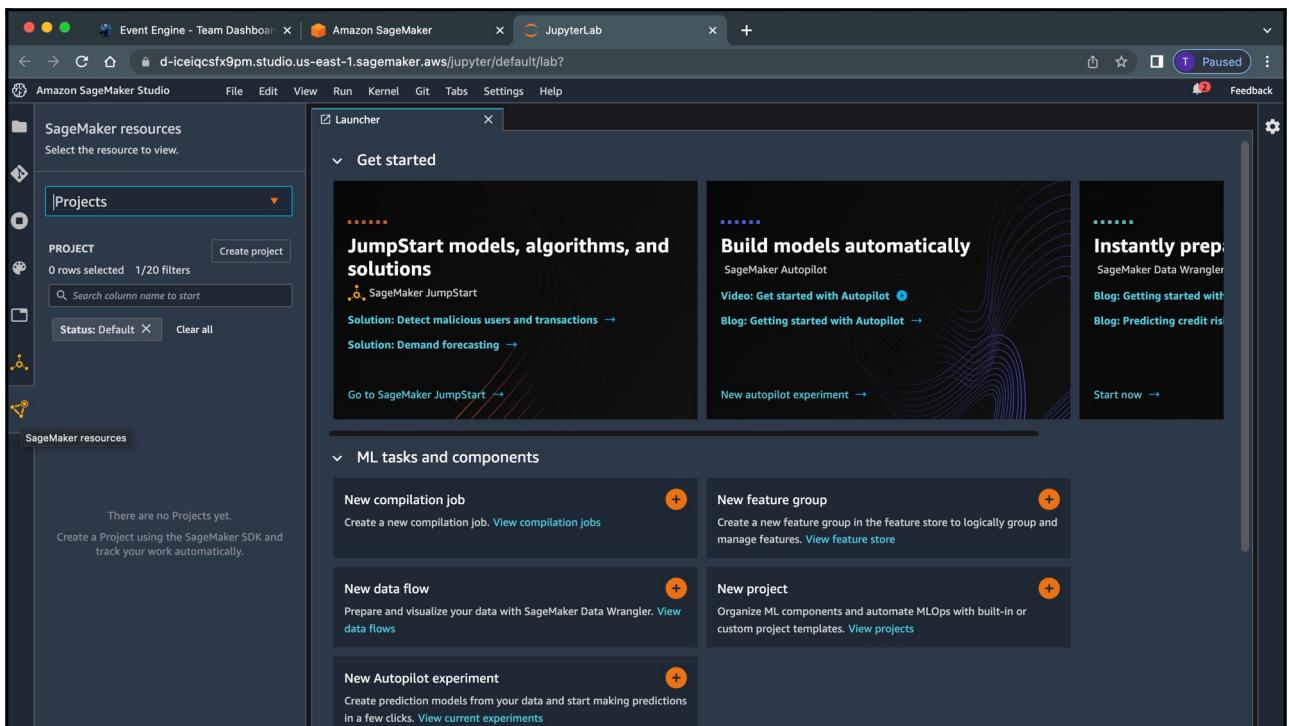
14



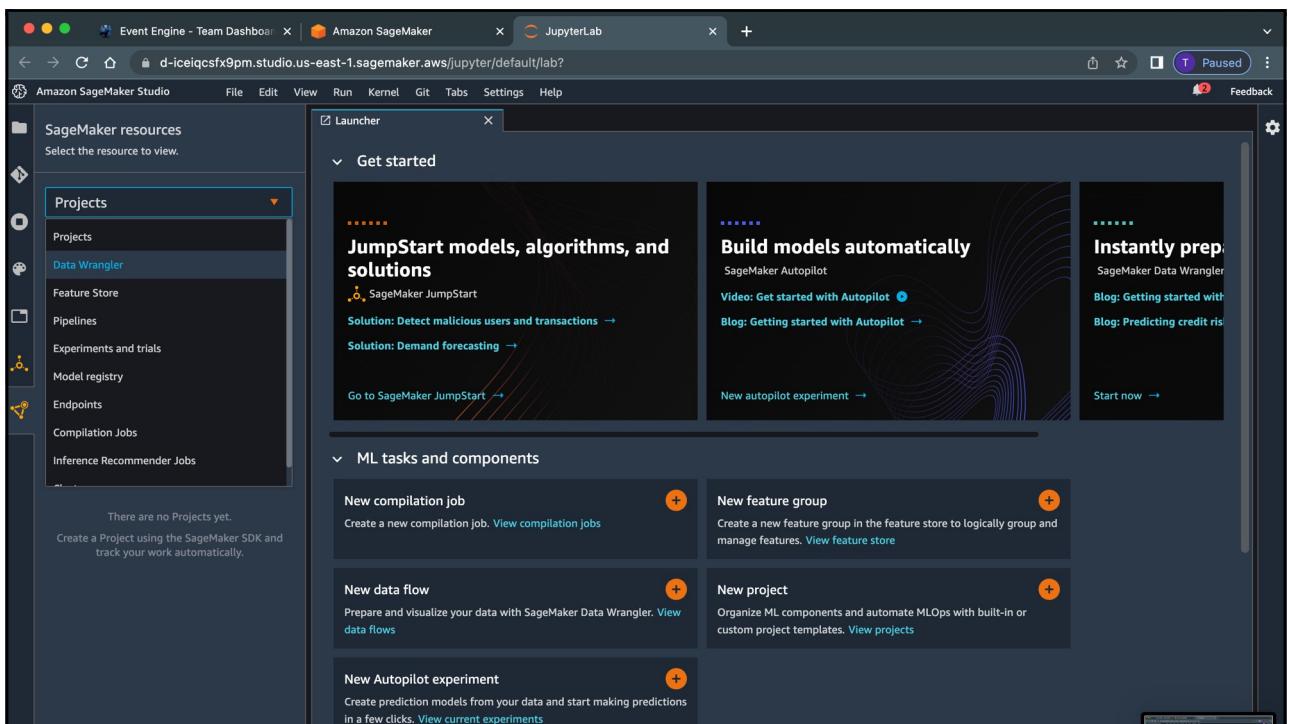
15



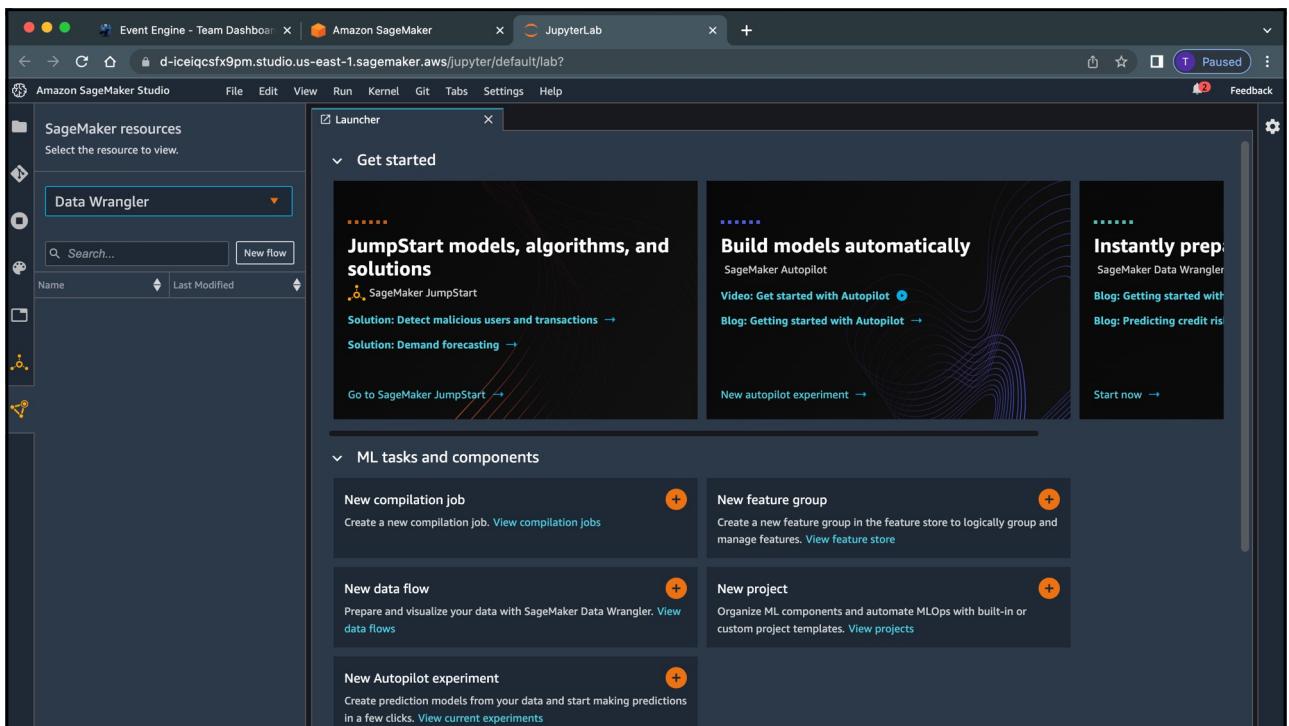
16



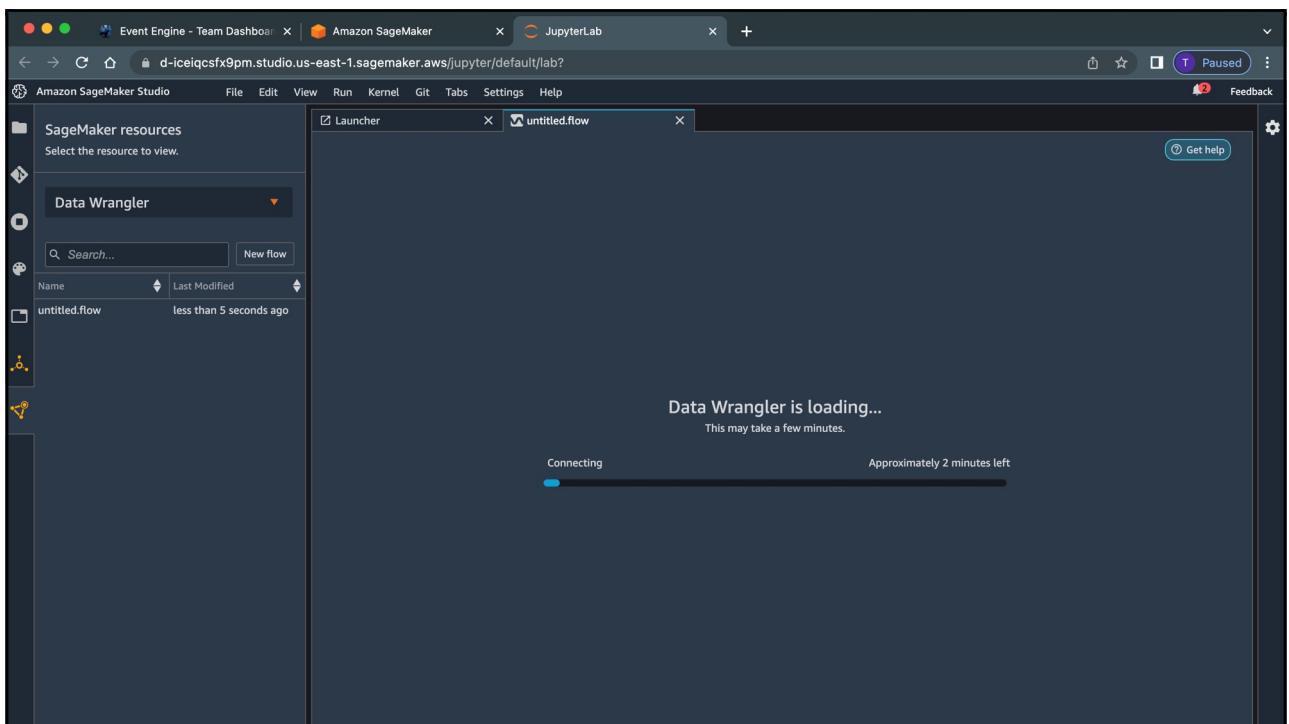
17



18

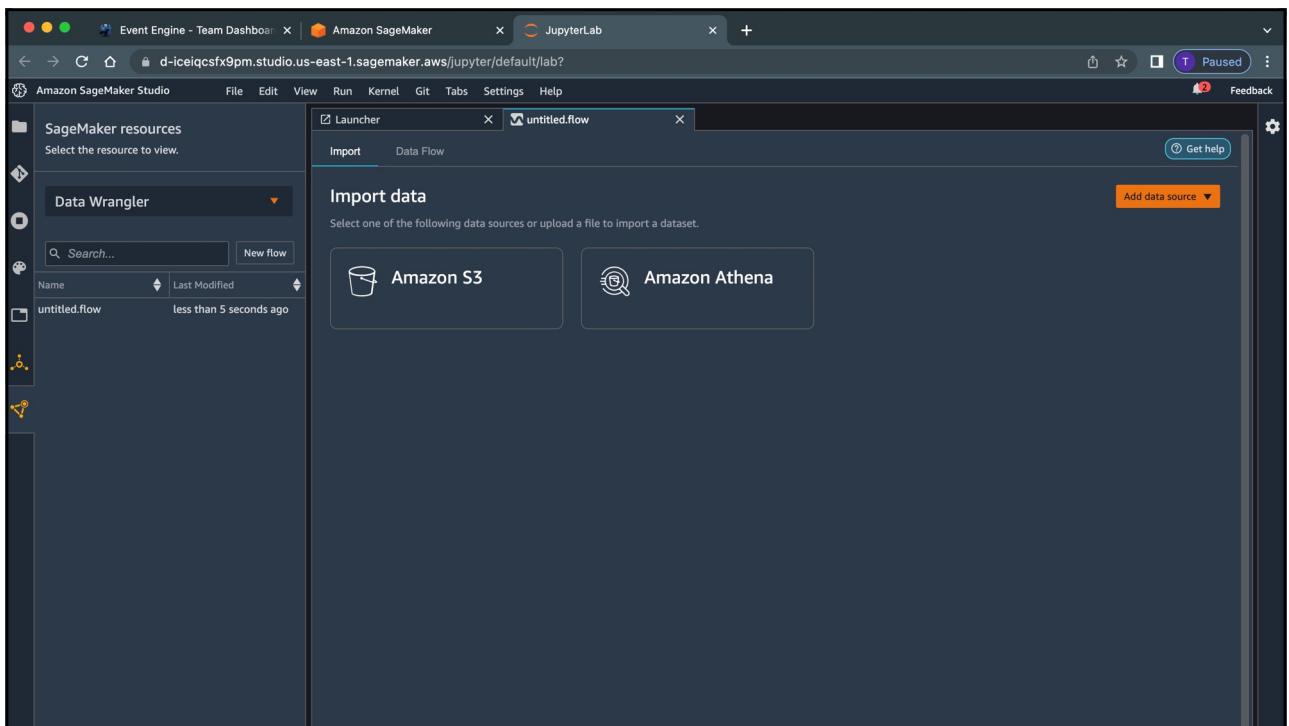


19

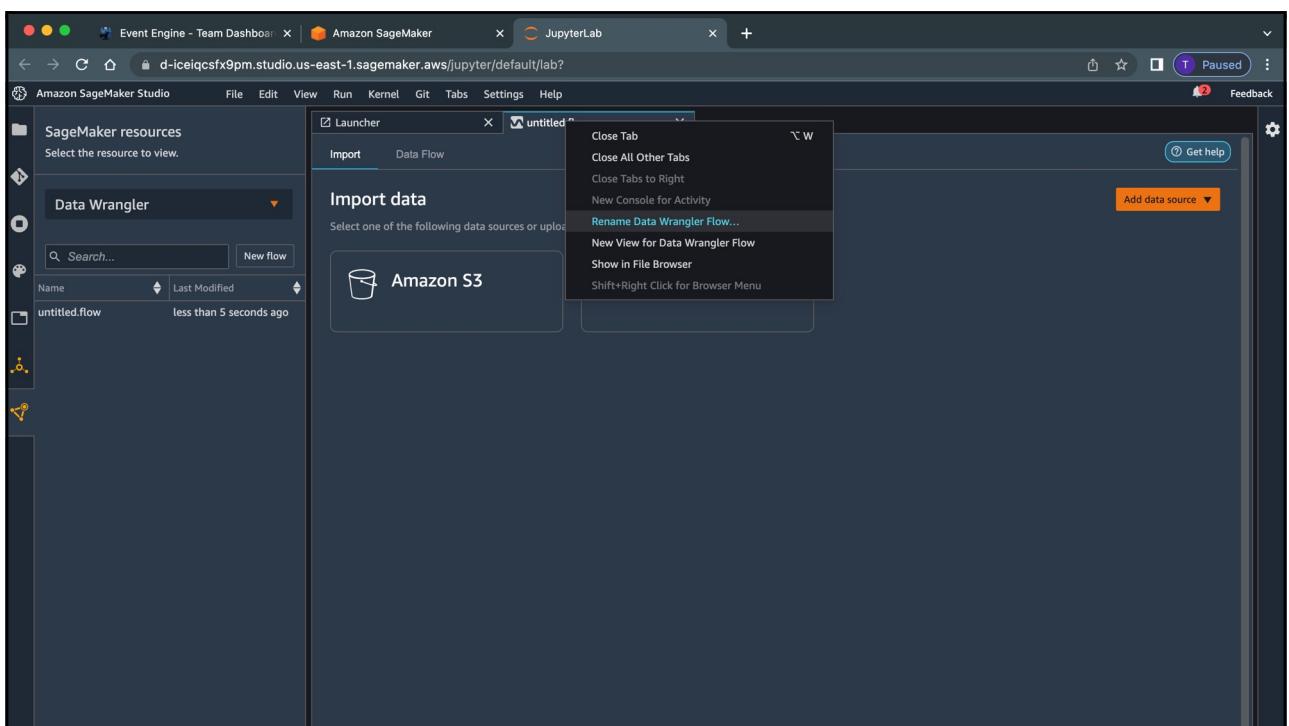


20

10

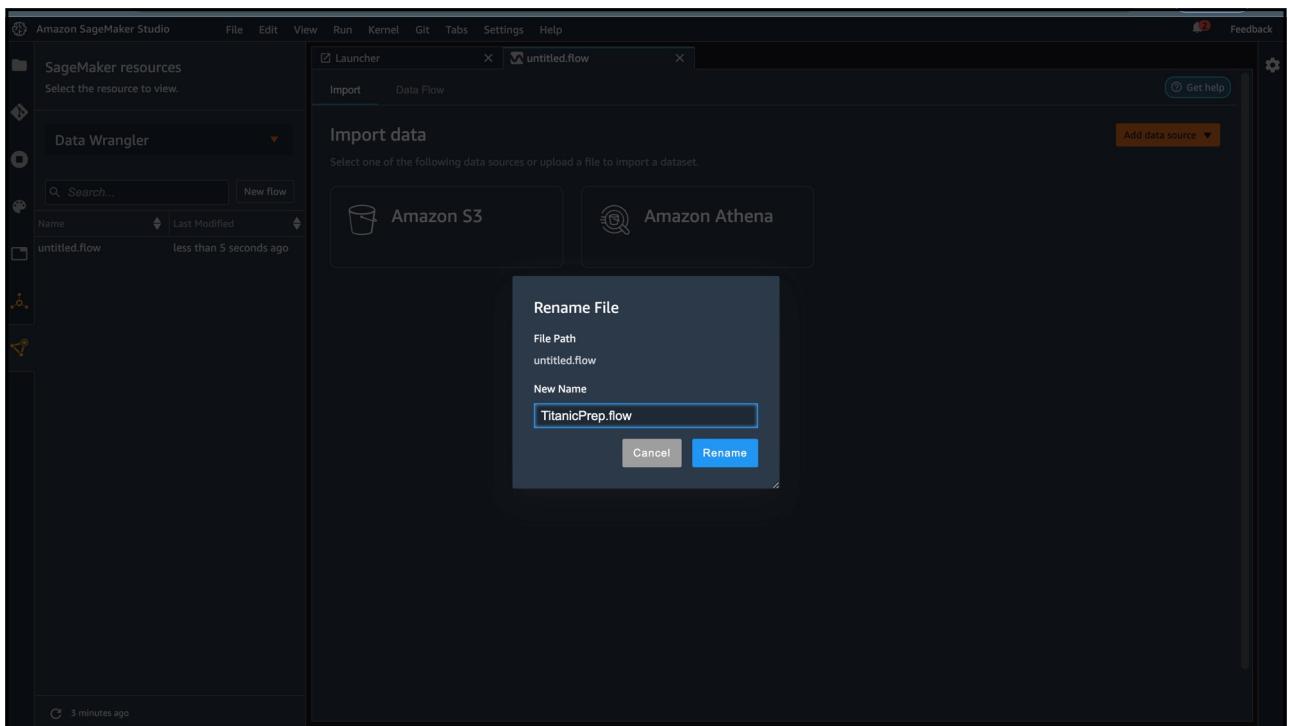


21

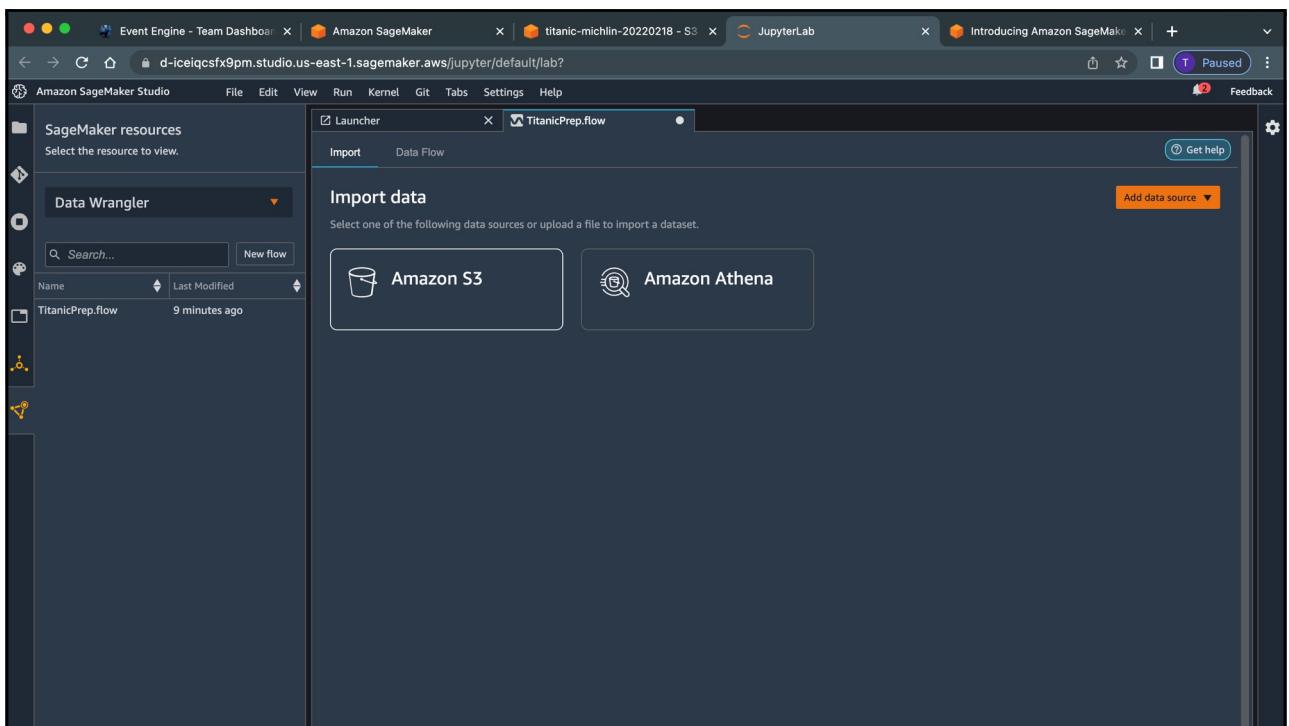


22

11

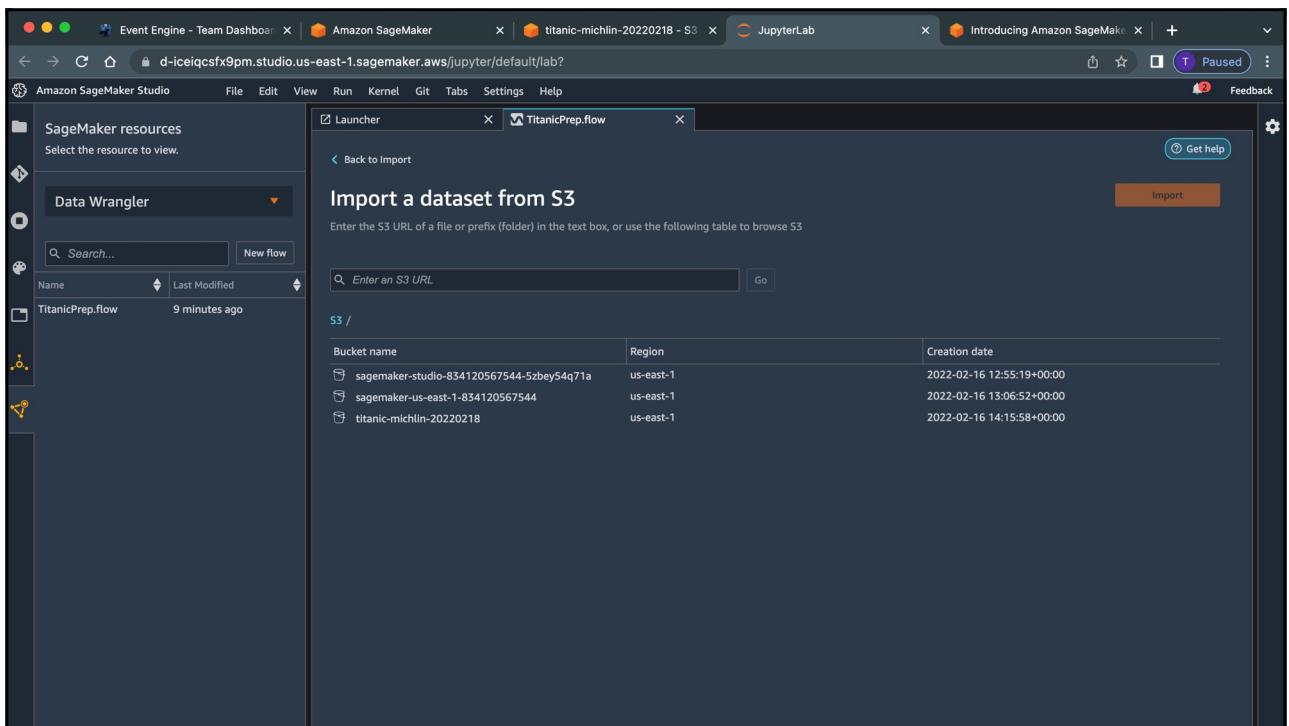


23

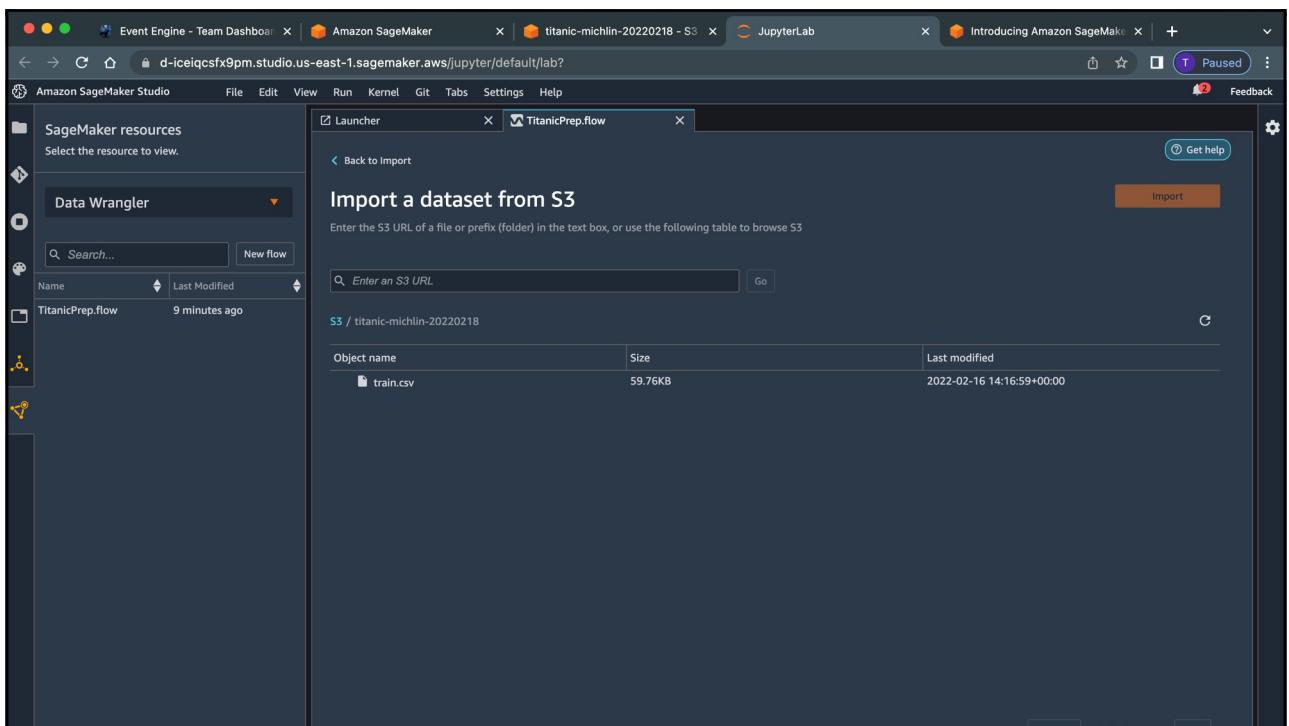


24

12



25



26

The screenshot shows the Amazon SageMaker Studio interface. On the left, there's a sidebar titled "SageMaker resources" with a "Data Wrangler" dropdown. The main area is titled "Launcher" and "TitanicPrep.flow". A sub-section titled "Import a dataset from S3" is displayed. It includes a search bar "Enter an S3 URL" and a table showing an object named "train.csv" with size 59.76KB and last modified date 2022-02-16 14:16:59+00:00. To the right, there's a "DETAILS" panel with fields for "Name" (set to "train.csv"), "File type" (set to "csv"), and checkboxes for "First row is header" (checked) and "Import nested directories". Below the import section is a preview table titled "PREVIEW • train.csv (first 100 rows shown)" showing the first six rows of the titanic dataset.

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

27

This screenshot is nearly identical to the one above, showing the "TitanicPrep.flow" launcher in Amazon SageMaker Studio. The "Import a dataset from S3" section and the preview table are the same. However, the "DETAILS" panel on the right shows different configurations under "Advanced configuration": "Delimiter" is set to "COMMA" and "Enable sampling" is checked, while "Filename as separate column" is unchecked.

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

28

The screenshot shows the Amazon SageMaker Studio interface. On the left, there's a sidebar titled "SageMaker resources" with a "Data Wrangler" dropdown. The main area is titled "Launcher" and "TitanicPrep.flow". A sub-section titled "Import a dataset from S3" is displayed. It includes a search bar "Enter an S3 URL" and a table showing an object named "train.csv" with size 59.76KB and last modified date 2022-02-16 14:16:59+00:00. To the right, there's a "DETAILS" panel with settings for the import: "Name" set to "train.csv", "File type" set to "csv", and "First row is header" checked. Below the preview table, a "PREVIEW" section shows the first 100 rows of the "train.csv" file.

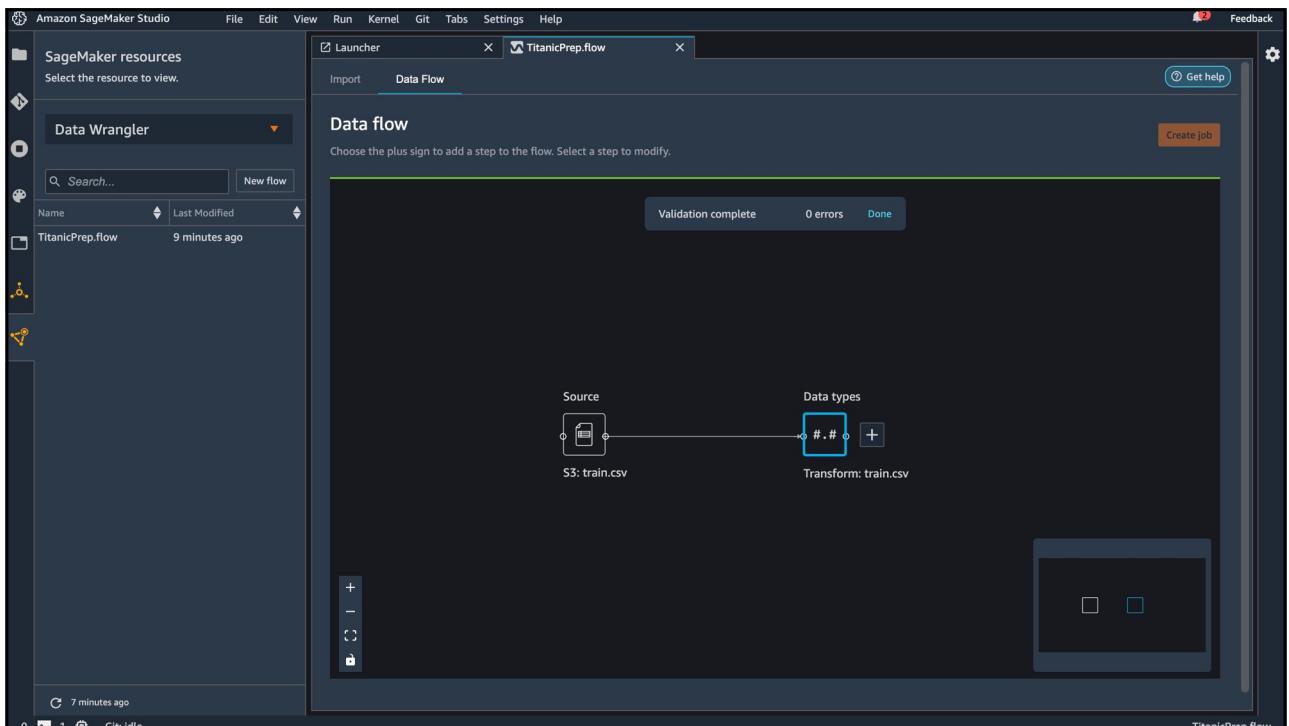
PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra... Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

29

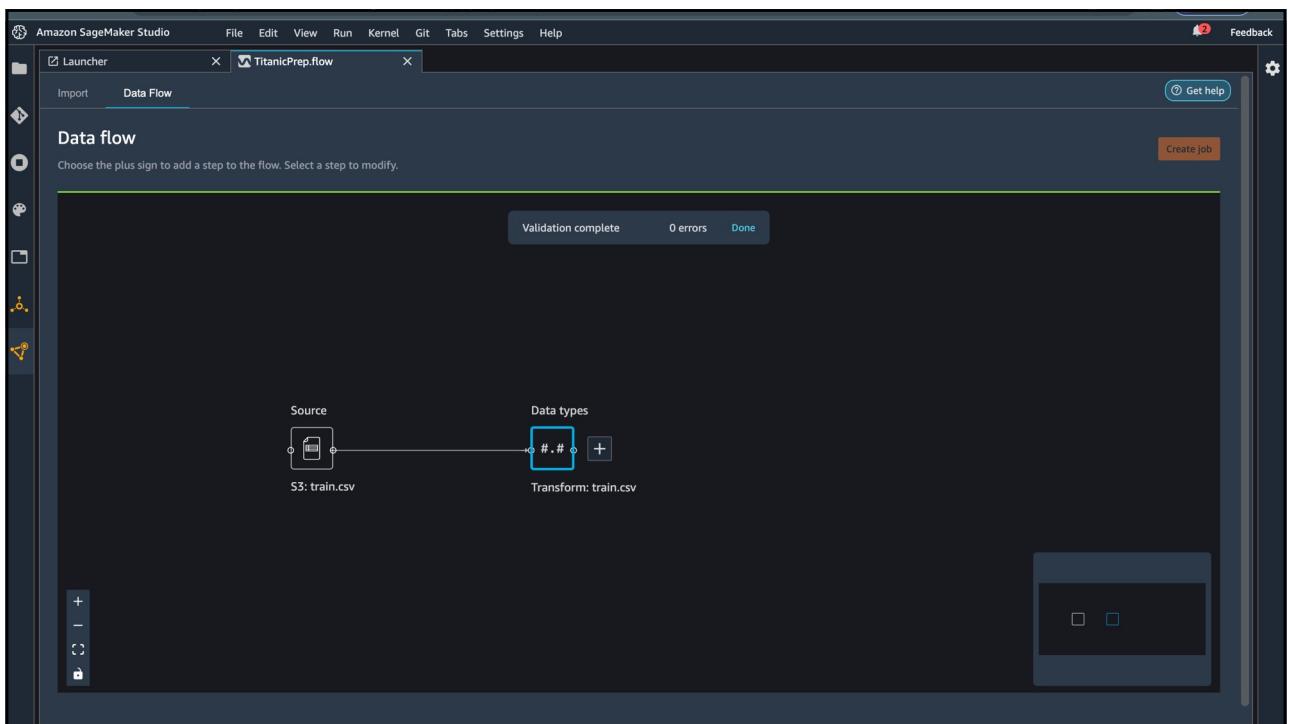
This screenshot is nearly identical to the one above, showing the same interface and data. The main difference is the timestamp in the "Last modified" column of the preview table, which has changed to 2022-02-16 14:16:59+00:00.

PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra... 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

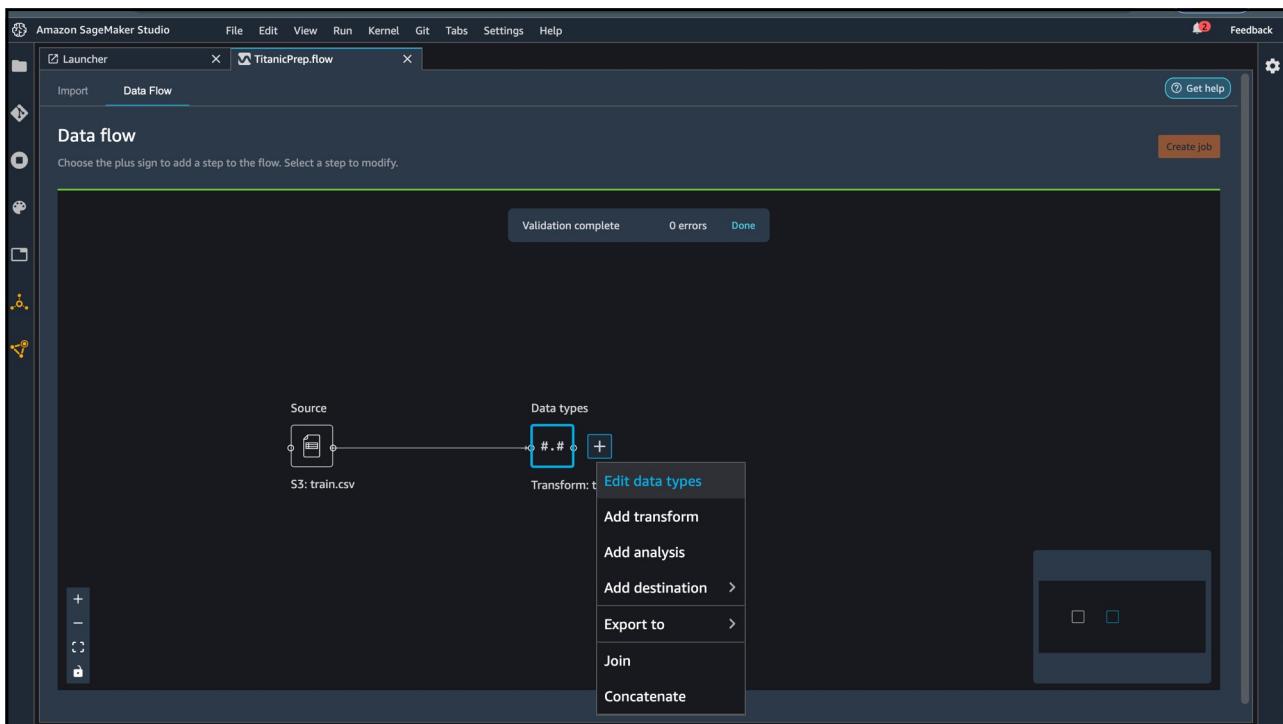
30



31



32

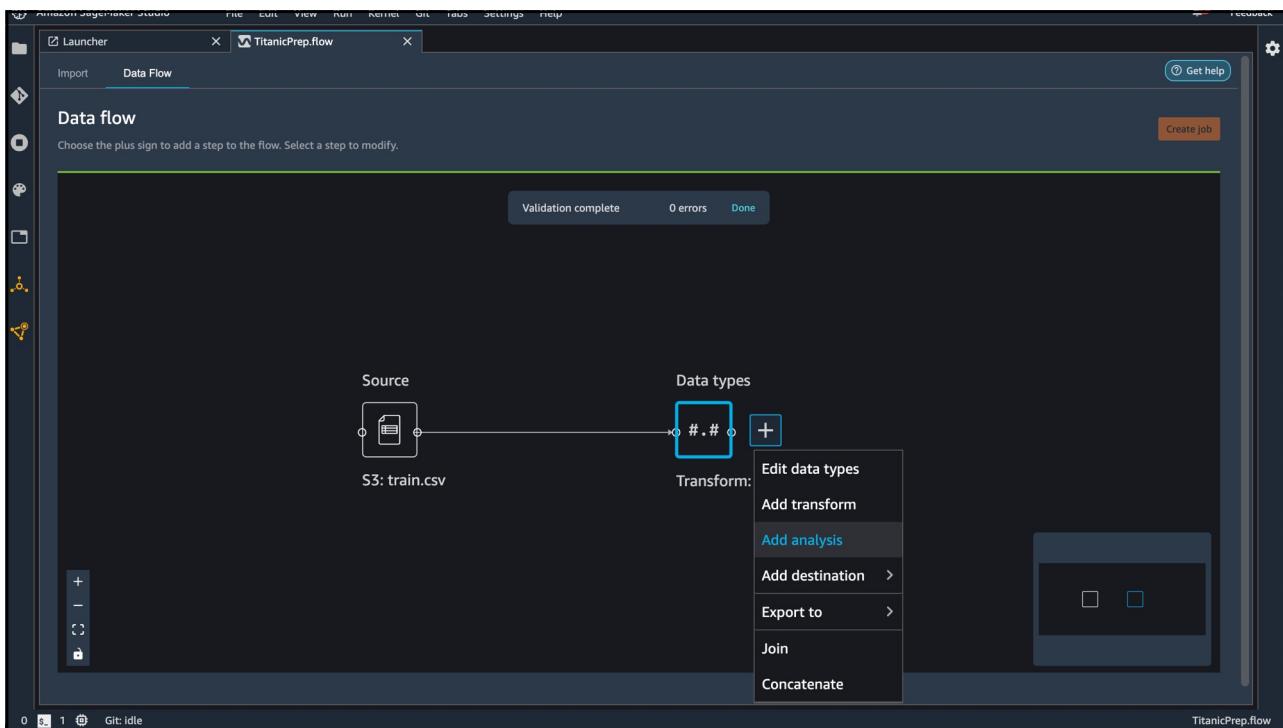


33

The screenshot shows the Amazon SageMaker Studio Data types configuration interface for 'train.csv'. It displays a preview of the data and a configuration table for mapping columns to 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

34



35

The screenshot shows the Amazon SageMaker Studio Data Flow interface. A flow named "TitanicPrep.flow" is displayed. A "Data types · Transform: train.csv" node is selected. On the left, there is a "Histogram: Untitled" visualization with a message "No Preview available" and "Use Configure for built-in analyses" and "Use Code to create a custom analysis". Below it is a "Data table" showing the first seven rows of the "train.csv" dataset:

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	54
7	0	1	McCarthy, Mr. Timothy J	male	54

On the right, a sidebar titled "Create analysis" lists various analysis types: Histogram (selected), Bias Report, Custom Visualization, Duplicate rows, Feature Correlation, Histogram, Multicollinearity, Quick Model, Scatter Plot, Table Summary (highlighted in blue), Select..., and Optional.

36

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

37

Table Summary: Summary

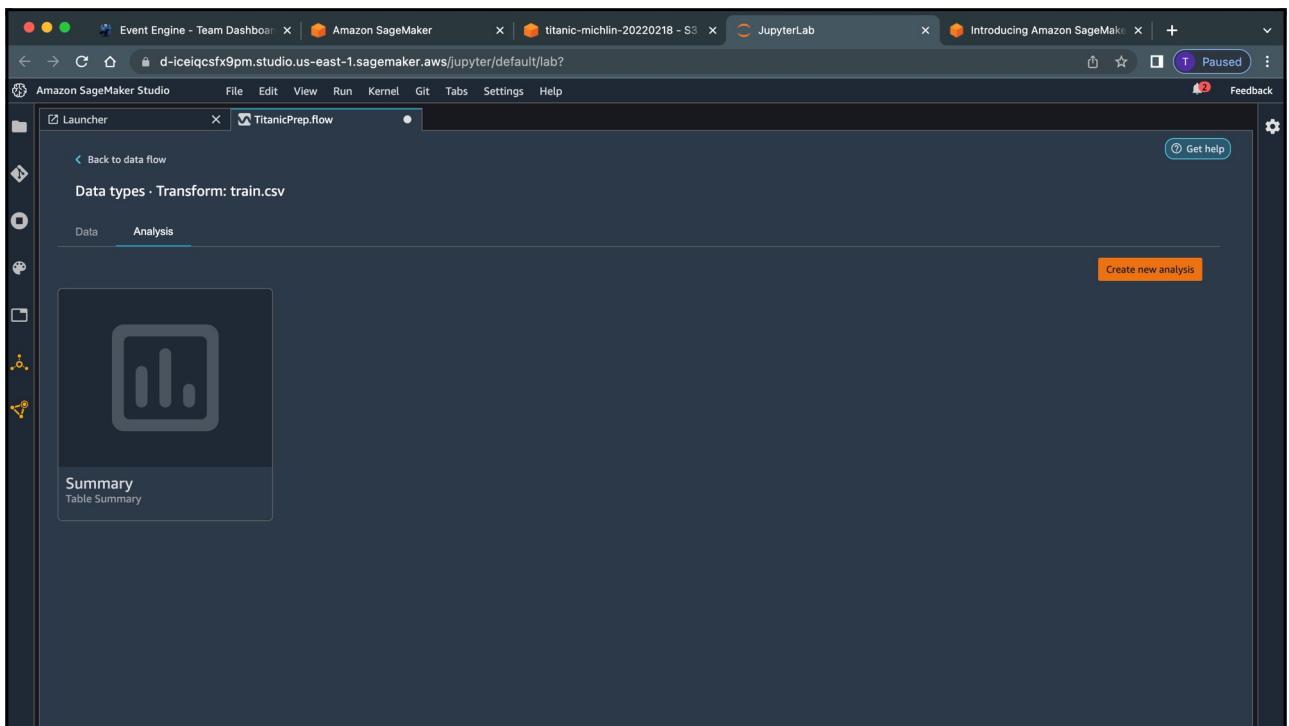
summary	PassengerId	Survived	Pclass	Name	Sex
count	891	891	891	891	891
mean	446.0	0.3838383838383838	2.308641975308642	None	None
stddev	257.3538420152301	0.48659245426485753	0.8360712409770491	None	None
min	1	0	1	Abbing, Mr. Anthony	female
max	891	1	3	van Melkebeke, Mr. Phil...	male

Data table

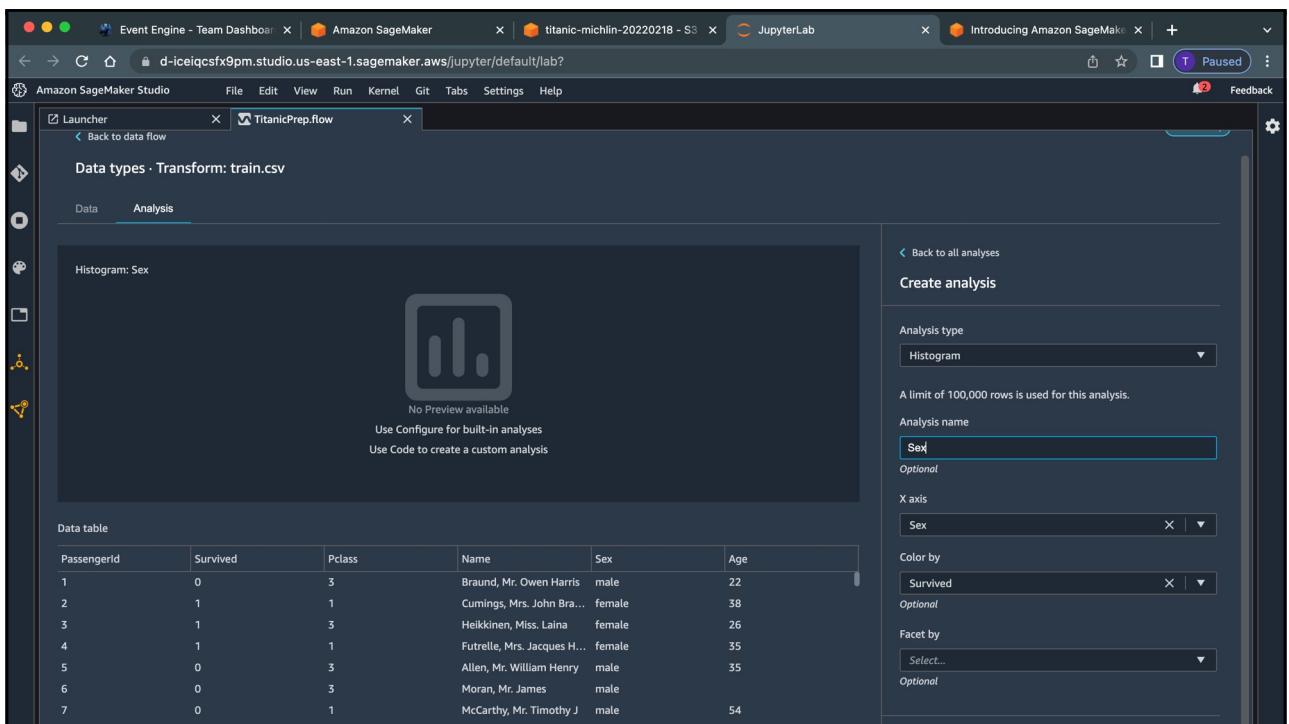
PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J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

38

19

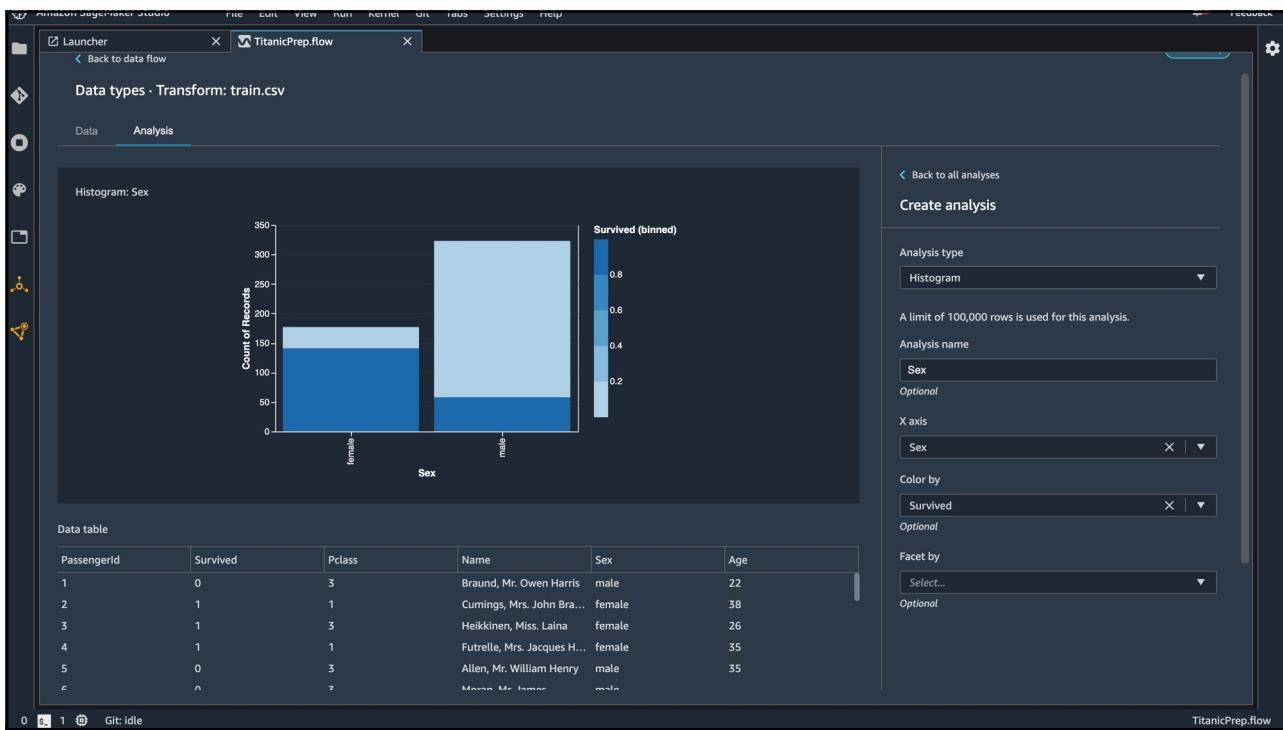


39

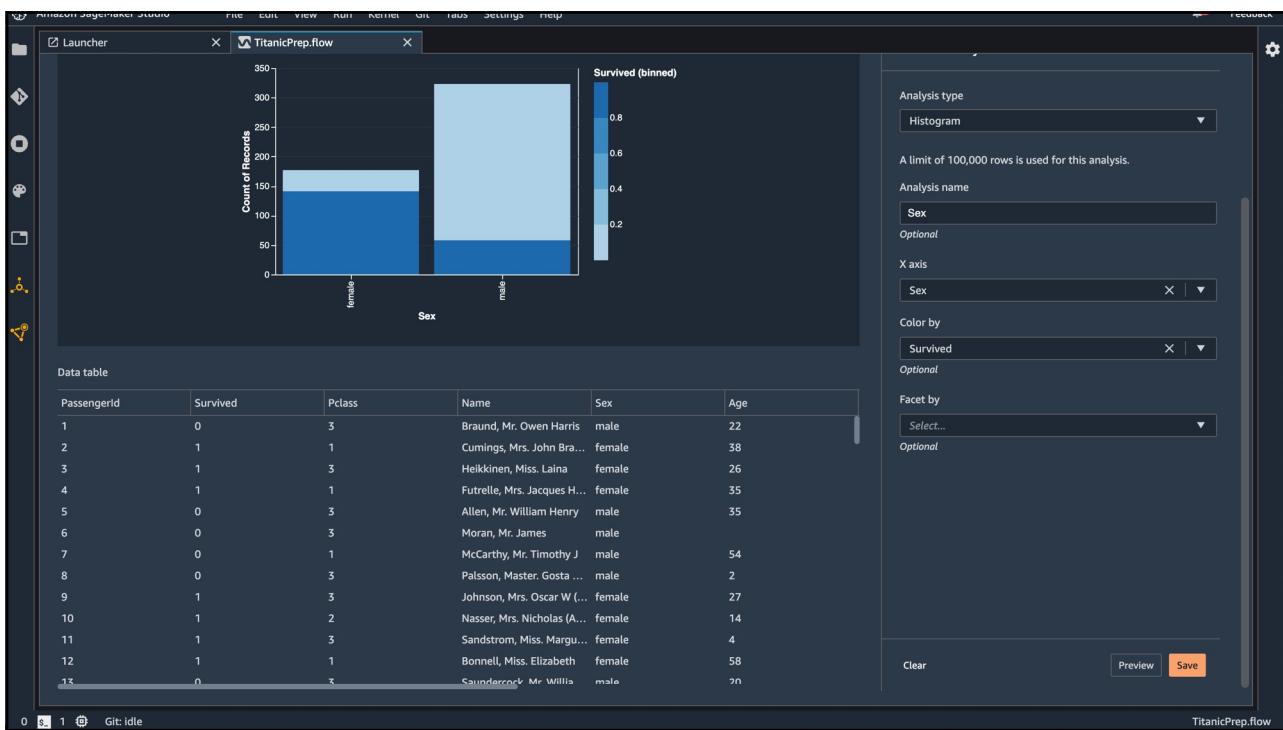


40

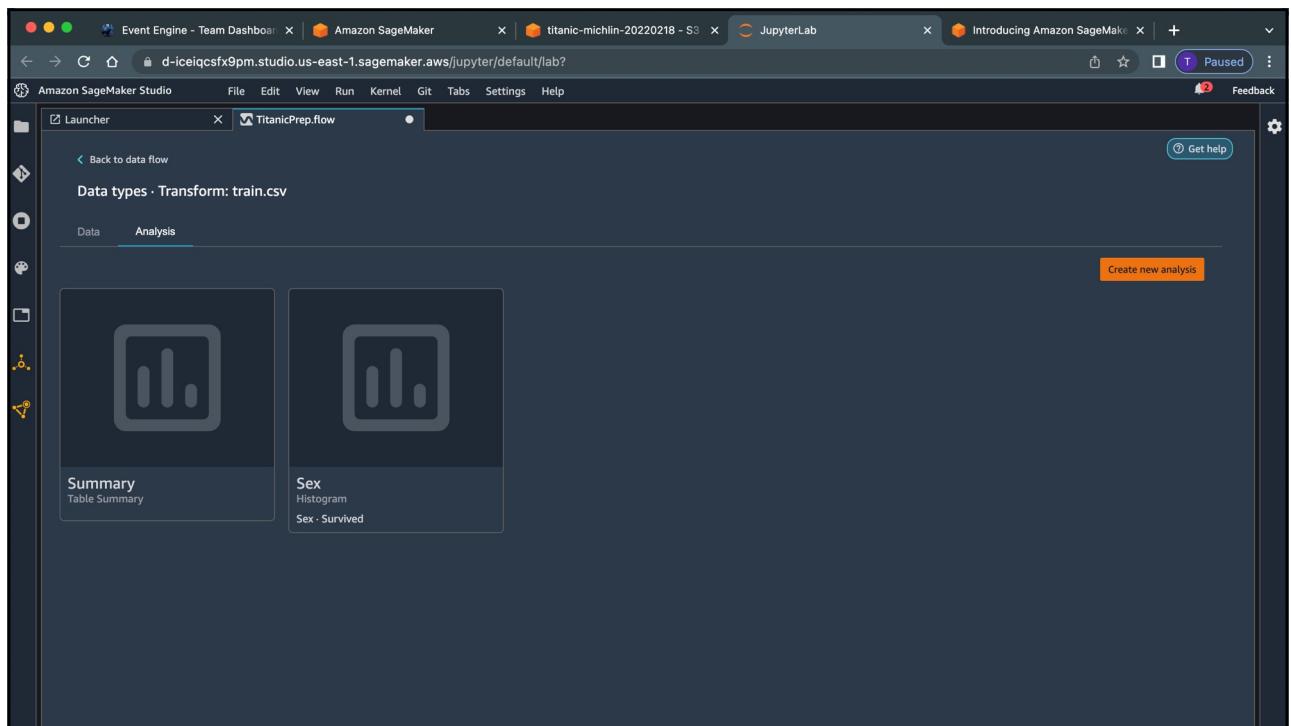
20



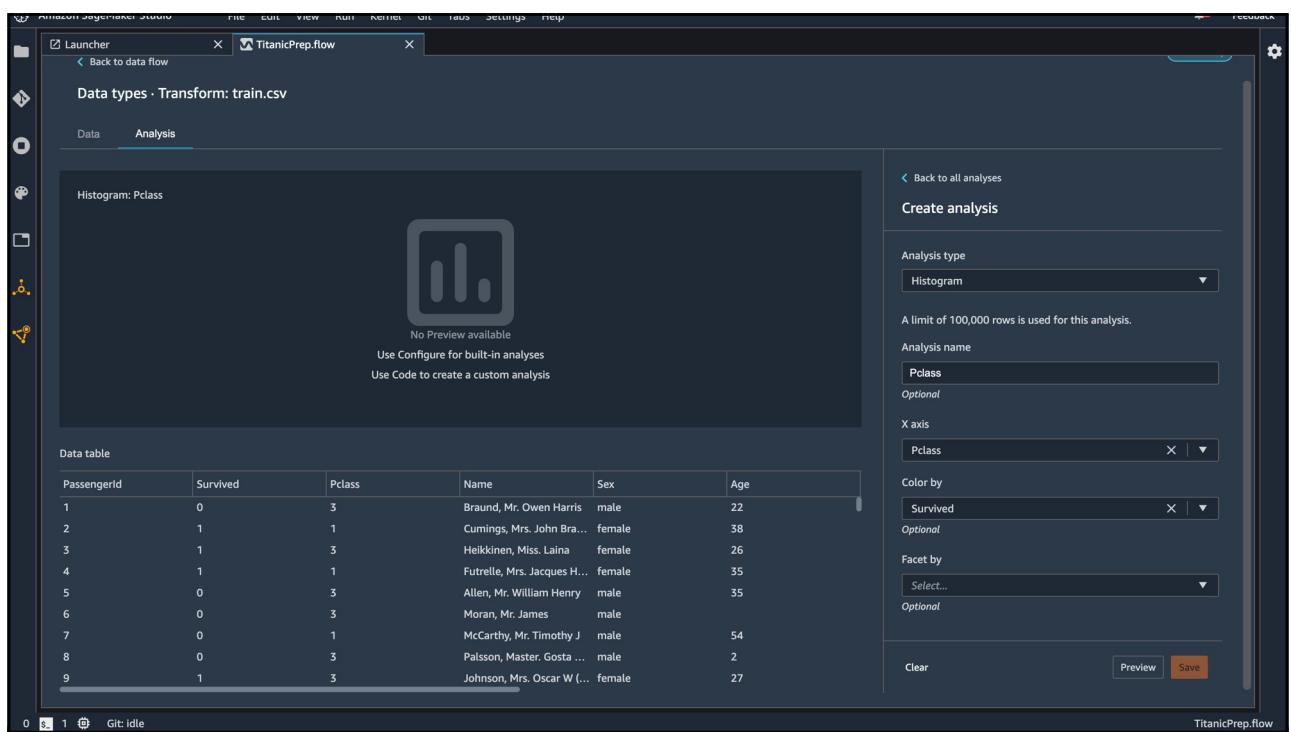
41



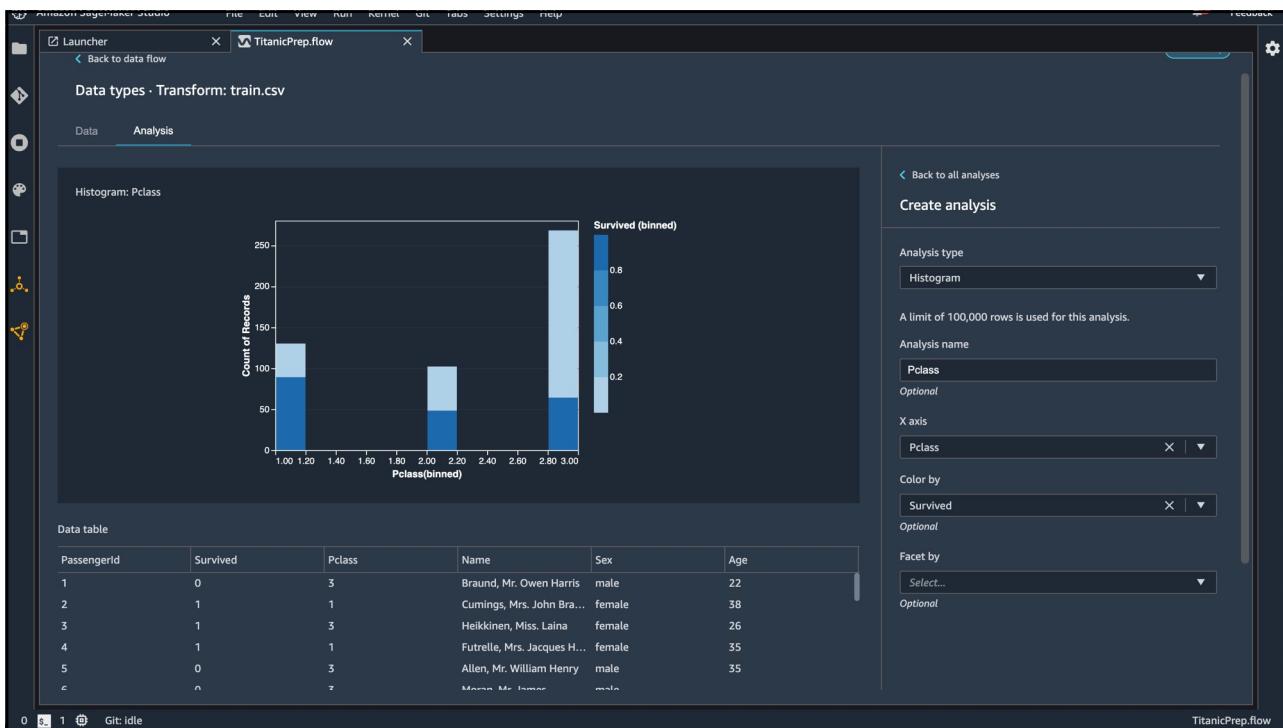
42



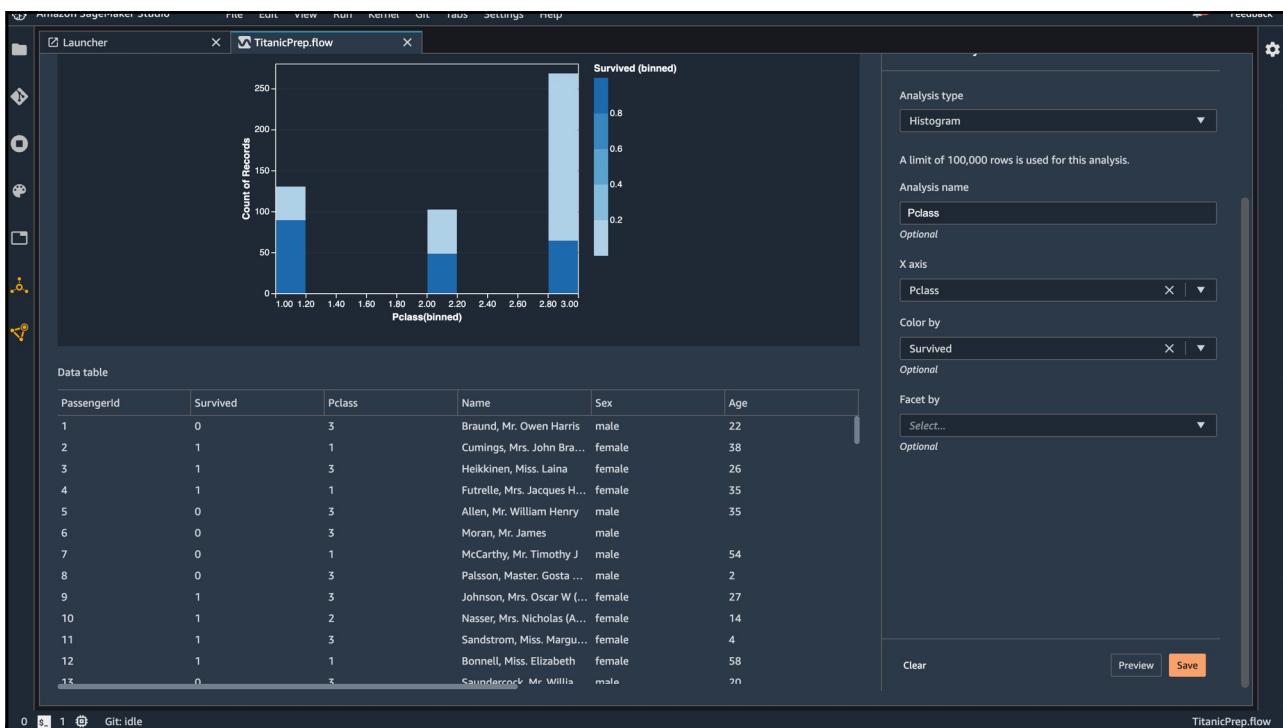
43



44



45



46

Event Engine - Team Dashboard | Amazon SageMaker | titanic-michlin-20220218 - S3 | JupyterLab | Introducing Amazon SageMaker | Paused | Feedback

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

Launcher > TitanicPrep.flow

Data types · Transform: train.csv

Data Analysis

Histogram: Parch

No Preview available

Use Configure for built-in analyses
Use Code to create a custom analysis

Data table

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

Create analysis

Analysis type: Histogram

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

Analysis name: Parch

X axis: Parch

Color by: Survived

Facet by: Select...

47

Event Engine - Team Dashboard | Amazon SageMaker | titanic-michlin-20220218 - S3 | JupyterLab | Introducing Amazon SageMaker | Paused | Feedback

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

Launcher > TitanicPrep.flow

Data types · Transform: train.csv

Data Analysis

Histogram: Parch

Data table

Passengerid	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allan, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	54

Create analysis

Analysis type: Histogram

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

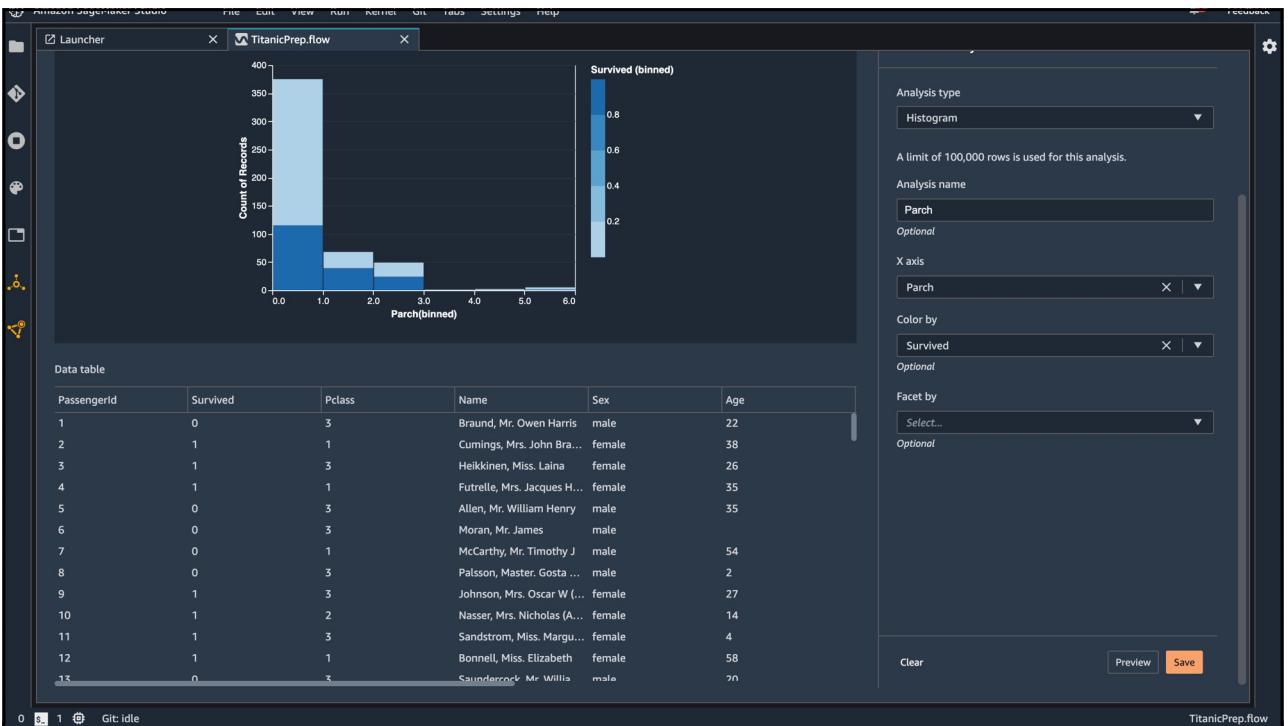
Analysis name: Parch

X axis: Parch

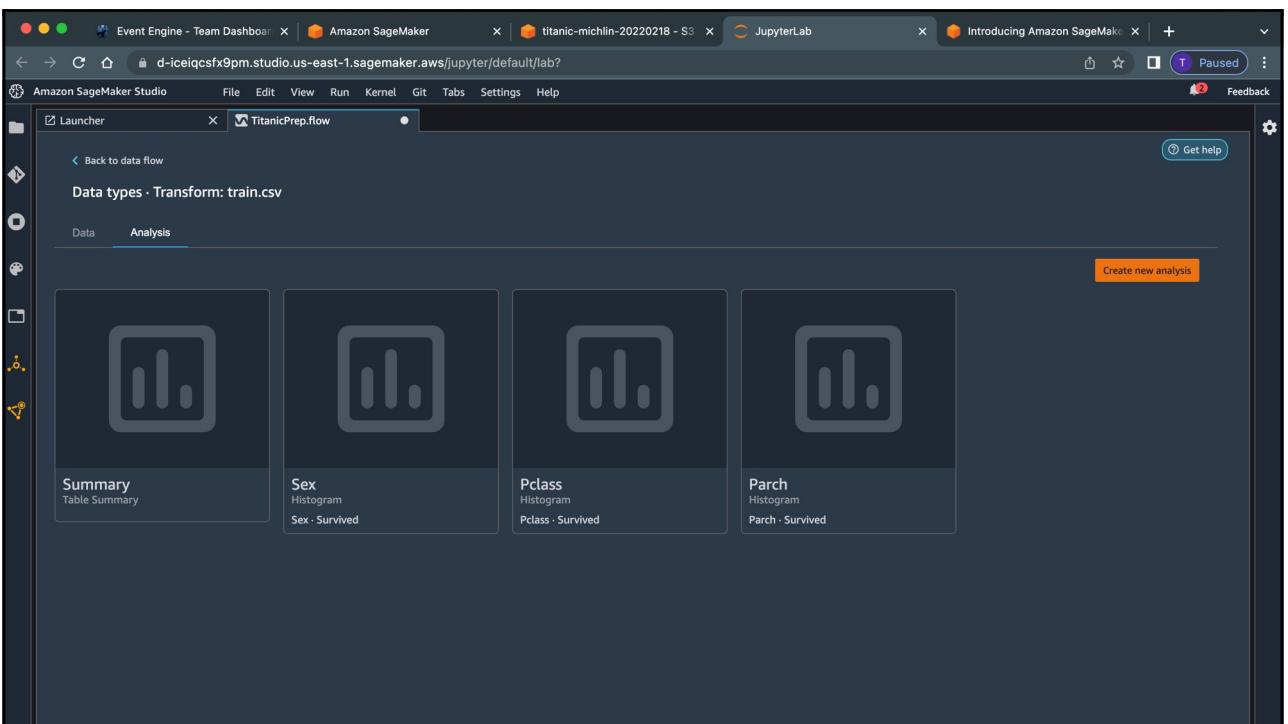
Color by: Survived

Facet by: Select...

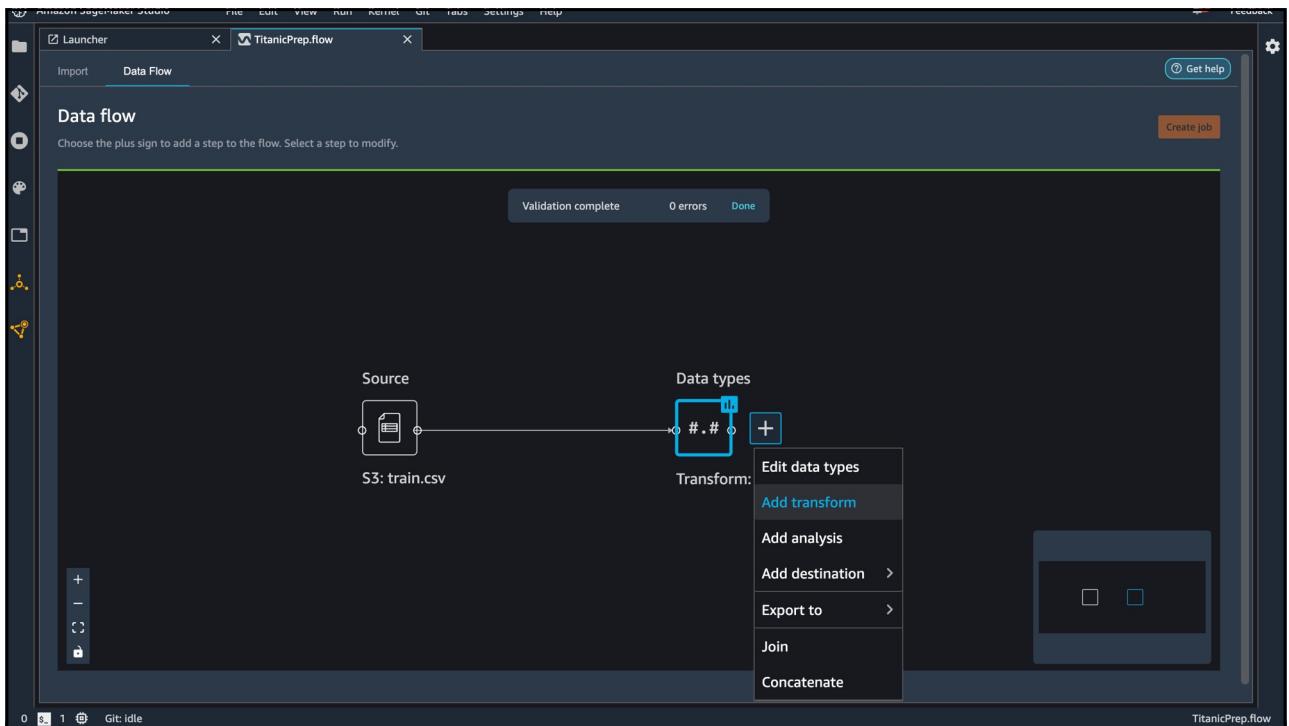
48



49



50



51

The screenshot shows the "Data types" step for the "train.csv" file within the Data Flow. The "Data" tab is selected, displaying a table of data with columns: PassengerId (long), Survived (long), Pclass (long), Name (string), Sex (string), and Age (long). The table contains 41 rows of passenger information. To the right of the table, a "TRANSFORMS" sidebar shows a list of steps: "1. S3 Source" and "2. Data types". The "Data types" step is highlighted with a red border.

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	Allan, 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 Plank, Mrs. Juli...	female	31
20	1	2	Macewani, Mrs. Estiwa...	female	

52

Amazon SageMaker Studio

Launcher TitanicPrep.flow

Data types · Transform: train.csv

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. Hulda A...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

Export data

ADD TRANSFORM

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

53

Amazon SageMaker Studio

Launcher TitanicPrep.flow

Data types · Transform: train.csv

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. Hulda A...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia...	male	20
14	0	3	Andersson, Mr. Anders J...	male	39
15	0	3	Vestrom, Miss. Hulda A...	female	14
16	1	2	Hewlett, Mrs. (Mary D K...	female	55
17	0	3	Rice, Master. Eugene	male	2
18	1	2	Williams, Mr. Charles Eu...	male	
19	0	3	Vander Planke, Mrs. Juli...	female	31
20	1	3	Masselmani, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

Export data

HANDLE MISSING

Replace, drop, or add indicators for missing values. Learn more.

Transform ?

Impute

Column type ?

Numeric

Input column ?

Age

Imputing strategy ?

Approximate Median

Output column ?

Age_Imputed

Optional

Clear

Preview Add

54

27

Data types - Transform: train.csv

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

HANDLE MISSING

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

Transform [?](#)

Impute

Column type [?](#)

Numeric

Input column [?](#)

Age

Imputing strategy [?](#)

Approximate Median

Output column [?](#)

Age_Imputed

Optional

Clear

Preview Add

55

Impute - Transform: train.csv

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. Hilda A... ...	female	14
16	1	2	Hewlett, Mrs. (Mary D K... ...	female	55
17	0	3	Rice, Master. Eugene ...	male	2
18	1	2	Williams, Mr. Charles Eu... ...	male	
19	0	3	Vander Planke, Mrs. Juli... ...	female	31
20	1	3	Masselmani, Mrs. Fatima ...	female	

TRANSFORMS

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

56

Amazon SageMaker Studio

TitanicPrep.flow

Step 3. Impute

Data Analysis

Export data

Passengerid (long) Survived (long) Pclass (long) Name (string) Sex (string) Age (long)

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

Featureize date/time
Encode date/time values to numeric and vector representations. Learn more...

Featureize text
Generate vector representations from natural language text. Learn more...

Format string
Clean and prepare strings using standard string formatting operations. Learn more...

Group by
Add an aggregated column after group by as a new column.

Handle missing
Replace, drop, or add indicators for missing values. Learn more...

Handle outliers
Remove or replace outlier numeric and categorical values. Learn more...

Handle structured column
Flatten JSON and perform other operations on structured data.

Manage columns
Move, drop, duplicate or rename columns in the dataset. Learn more...

Manage rows
Sort, shuffle or drop duplicate rows.

Manage vectors
Expand or create vector columns. Learn more...

57

Amazon SageMaker Studio

TitanicPrep.flow

Step 3. Impute

Data Analysis

Export data

Passengerid (long) Survived (long) Pclass (long) Name (string) Sex (string) Age (long)

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

Move, drop, duplicate or rename columns in the dataset. Learn more...

Transform
Drop column

Columns to drop
Passengerid X Name X Cabin X Ticket X

Clear Preview Add

58

29

The screenshot shows the Amazon SageMaker Studio interface. In the center, there is a data preview titled "TitanicPrep.flow" showing the first few rows of a CSV file named "train.csv". The columns include PassengerId, SibSp, Parch, Fare, Embarked, and Age_Imputed. On the right side, a "MANAGE COLUMNS" dialog is open, allowing users to move, drop, duplicate, or rename columns. The "Transform" section shows the current step: "Drop column". The "Columns to drop" list contains "PassengerId", "Name", "Cabin", and "Ticket".

59

The screenshot shows the Amazon SageMaker Studio interface. The data preview now displays the results of the "Drop column" step, showing the same columns as before but with the "PassengerId", "Name", "Cabin", and "Ticket" columns removed. On the right side, the "TRANSFORMS" sidebar is visible, listing the steps taken so far: 1. S3 Source, 2. Data types, 3. Impute, and 4. Drop column.

60

30

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher TitanicPrep.flow

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

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

Export data

Get help

ADD TRANSFORM

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

61

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher TitanicPrep.flow

Drop column · Transform: train.csv

Data Analysis

Step 4. Drop column

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

Export data

Get help

CUSTOM PYSPARK

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

Python (PySpark)

Python (Pandas)

SQL (PySpark SQL)

62

The screenshot shows the Amazon SageMaker Studio interface. On the left, there's a sidebar with icons for S3, Lambda, Step Functions, and ML Flow. The main area has tabs for 'Event Engine - Team Dashboard', 'Amazon SageMaker', 'titanic-michlin-20220218 - S3', 'JupyterLab', and 'Introducing Amazon SageMaker'. The 'JupyterLab' tab is active.

In the center, there's a 'TitanicPrep.flow' step titled 'Drop column · Transform: train.csv'. It has a 'Data' tab selected, showing a preview of the 'train.csv' dataset. The columns are: Survived (long), Pclass (long), Sex (string), Age (long), SibSp (long), and Parch (long). The data shows rows of passenger information.

To the right of the preview, there's a 'CUSTOM PANDAS' panel. It contains a note: 'Use Pyspark, Pandas, or PySpark (SQL) to define custom transformations. Learn more.' Below this is a code editor with Python (Pandas) code:

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

Buttons for 'Preview' and 'Add' are at the bottom of the panel.

63

This screenshot is identical to the one above, showing the same interface and data preview for the 'train.csv' file. The 'CUSTOM PANDAS' panel also contains the same Python code for creating a 'With_Family' column.

64

The screenshot shows the Amazon SageMaker Studio interface. A flow named "TitanicPrep.flow" is open, specifically Step 5: Custom Pandas. The data preview shows the first few rows of a "train.csv" file. The right panel displays the "TRANSFORMS" configuration, which includes five steps: 1. S3 Source, 2. Data types, 3. Impute, 4. Drop column, and 5. Custom Pandas.

p (long)	Parch (long)	Fare (float)	Embarked (string)	Age_Imputed (float)	With_Family (long)
0		7.25	S	22	1
0		71.2833	C	38	1
0		7.925	S	26	0
0		53.1	S	35	1
0		8.05	S	35	0
0		8.4583	Q	28	0
0		51.8625	S	54	0
1		21.075	S	2	1
2		11.3333	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

65

The screenshot shows the Amazon SageMaker Studio interface. A flow named "TitanicPrep.flow" is open, specifically Step 5: Custom Pandas. The data preview shows the first few rows of a "train.csv" file. The right panel displays the "ADD TRANSFORM" modal, which lists several transformation options: Custom transform, Balance data, Custom formula, Encode categorical, Featureize date/time, Featureize text, Format string, Group by, and Handle missing.

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		11.3333	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

66

67

68

The screenshot shows the Amazon SageMaker Studio interface with the 'TitanicPrep.flow' data flow open. On the left, there's a preview of the 'train.csv' data with columns: 'Survived (string)', 'Age_Imputed (float)', 'With_Family (long)', 'Pclass_3 (float)', 'Pclass_1 (float)', and 'Pclass_2 (float)'. On the right, the 'ENCODE CATEGORICAL' configuration panel is displayed for the 'Pclass' column. The 'Transform' dropdown is set to 'One-hot encode'. Under 'Input columns', 'Pclass' is selected. Under 'Output style', 'Columns' is selected. The 'Drop last' checkbox is checked. The 'Output column' field is empty. There are 'Preview' and 'Add' buttons at the bottom.

69

The screenshot shows the Amazon SageMaker Studio interface with the 'TitanicPrep.flow' data flow open. On the left, there's a preview of the 'train.csv' data with columns: 'Survived (long)', 'Sex (string)', 'Age (float)', 'SibSp (long)', 'Parch (long)', and 'Fare (float)'. On the right, the 'ENCODE CATEGORICAL' configuration panel is displayed for the 'Sex' column. The 'Transform' dropdown is set to 'One-hot encode'. Under 'Input columns', 'Sex' is selected. Under 'Output style', 'Columns' is selected. The 'Drop last' checkbox is checked. The 'Output column' field is empty. There are 'Preview' and 'Add' buttons at the bottom.

70

One-hot encode - Transform: train.csv

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	
0	0			0	
1	0			0	
0	0			0	
1	0			1	
0	0			1	
1	0			0	
1	0			0	

ENCODE CATEGORICAL

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

Transform [?](#)

One-hot encode

Input columns [?](#)

Sex [X](#)

Input already ordinal encoded [?](#)

Invalid handling strategy [?](#)

Skip

Drop last [?](#)

Output style [?](#)

Columns [X](#)

Output column [?](#)

Optional

Preview Add

71

One-hot encode - Transform: train.csv

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

TRANSFORMS

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

72

Amazon SageMaker Studio

TitanicPrep.flow

One-hot encode - Transform: train.csv

Data Analysis

Step 7. One-hot encode

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)
0	22	1	0	7.25	S
1	38	1	0	71.2833	C
1	26	0	0	7.925	S
1	35	1	0	53.1	S
0	35	0	0	8.05	S
0	0	0	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

Export data

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

Balance data
Balance the data for binary classification problems using random oversampling...

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

Encode categorical
Convert categorical variables to numeric or vector representations. Learn more

Featetrize date/time
Encode date/time values to numeric and vector representations. Learn more

Featetrize text
Generate vector representations from natural language text. Learn more

Format string
Clean and prepare strings using standard string formatting operations. Learn more

Group by
Add an aggregated column after group by as a new column.

Handle missing
Replace, drop, or add indicators for missing values. Learn more

73

Amazon SageMaker Studio

TitanicPrep.flow

One-hot encode - Transform: train.csv

Data Analysis

Step 7. One-hot encode

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Embarked (string)
0	22	1	0	7.25	S
1	38	1	0	71.2833	C
1	26	0	0	7.925	S
1	35	1	0	53.1	S
0	35	0	0	8.05	S
0	0	0	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
1	14	0	0	7.8542	S
0	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

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

74

37

75

76

One-hot encode - Transform: train.csv

Data Analysis

Step 8. One-hot encode

Survived (long)	Age (float)	SibSp (long)	Parch (long)	Fare (float)	Age_Imputed (float)
0	22	1	0	7.25	22
1	38	1	0	71.2833	38
1	26	0	0	7.925	26
1	35	1	0	53.1	35
0	35	0	0	8.05	35
0	0	0	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

Export data

ADD TRANSFORM

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

77

One-hot encode - Transform: train.csv

Data Analysis

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			0	1	
1			0	0	
1			1	0	
0			1	0	
0			0	1	

Export data

MANAGE COLUMNS

Move, drop, duplicate or rename columns in the dataset. Learn more.

Transform

Drop column

Columns to drop

Age x SibSp x Parch x Pclass_2 x
Sex_female x Embarked_Q x

Clear Preview Add

78

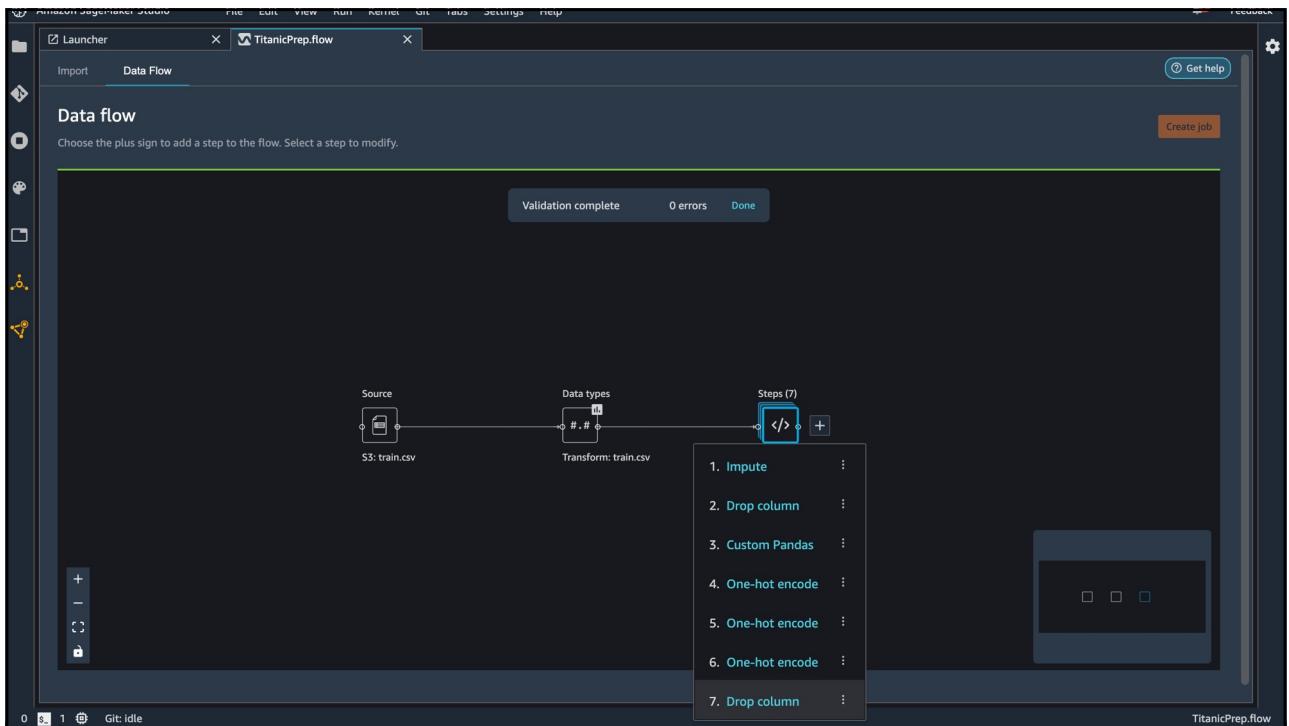
39

The screenshot shows the Amazon SageMaker Studio interface. A data preview window titled "One-hot encode - Transform: train.csv" displays a portion of a dataset with columns: Survived (long), Fare (float), Age_Imputed (float), With_Family (long), Pclass_3 (float), and Pclass_1 (float). To the right, a "MANAGE COLUMNS" panel is open, showing a list of columns to drop: Age, SibSp, Parch, Pclass_2, Sex_female, and Embarked_Q.

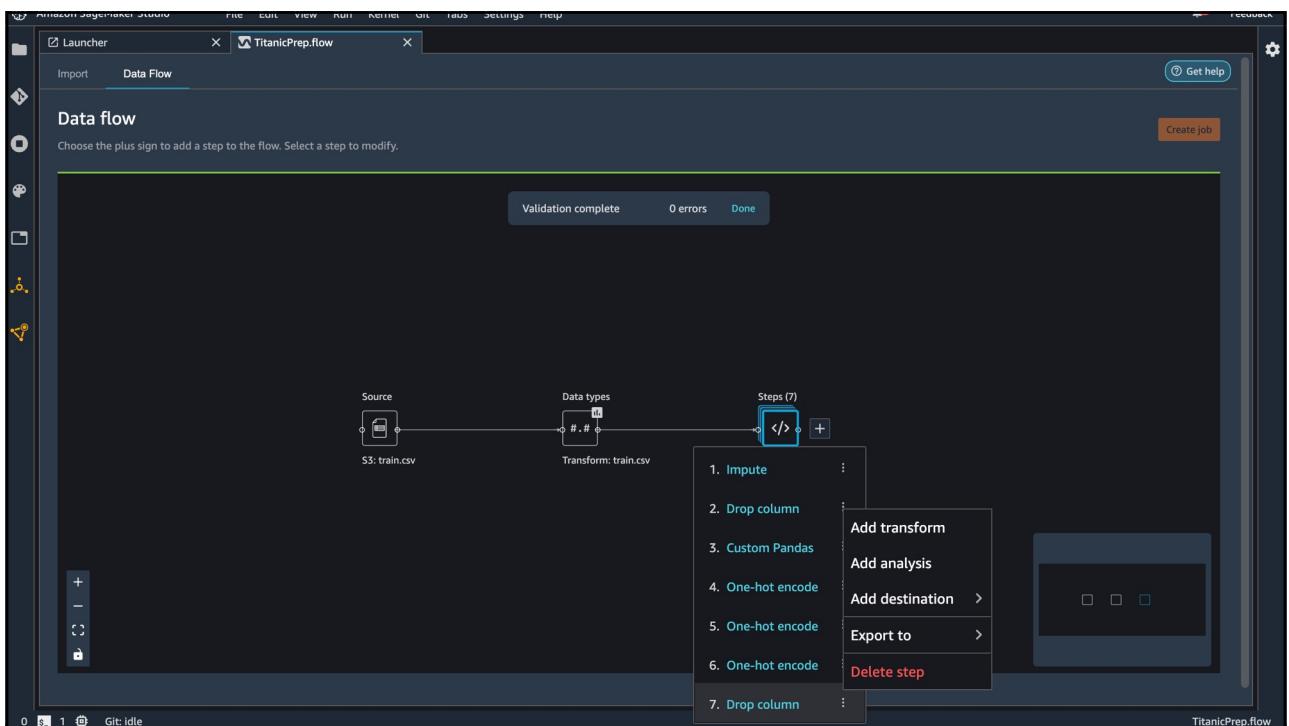
79

The screenshot shows the Amazon SageMaker Studio Data Flow interface. It displays a flowchart with three main components: "Source" (S3: train.csv), "Data types" (Transform: train.csv), and "Steps (7)". The "Data types" component has two input ports labeled "#, #". The "Steps (7)" component has one output port labeled "</>". A validation message at the top indicates "Validation complete" with "0 errors" and a "Done" button.

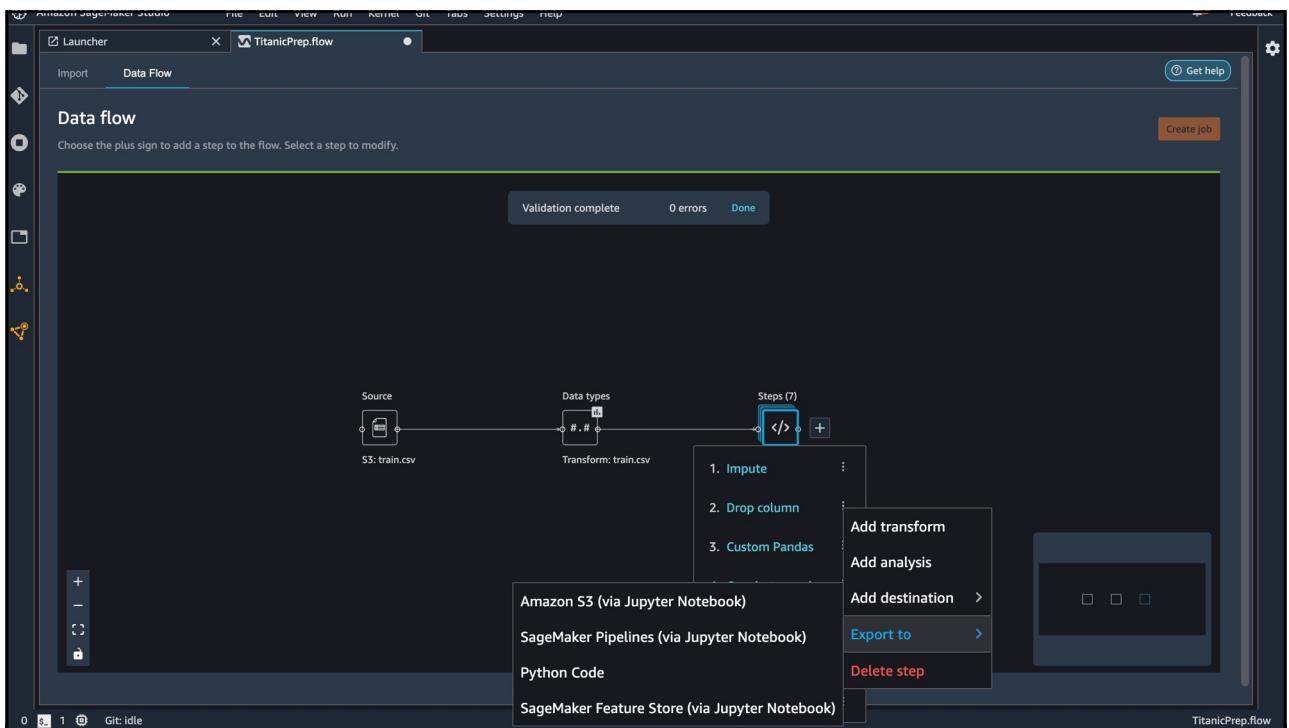
80



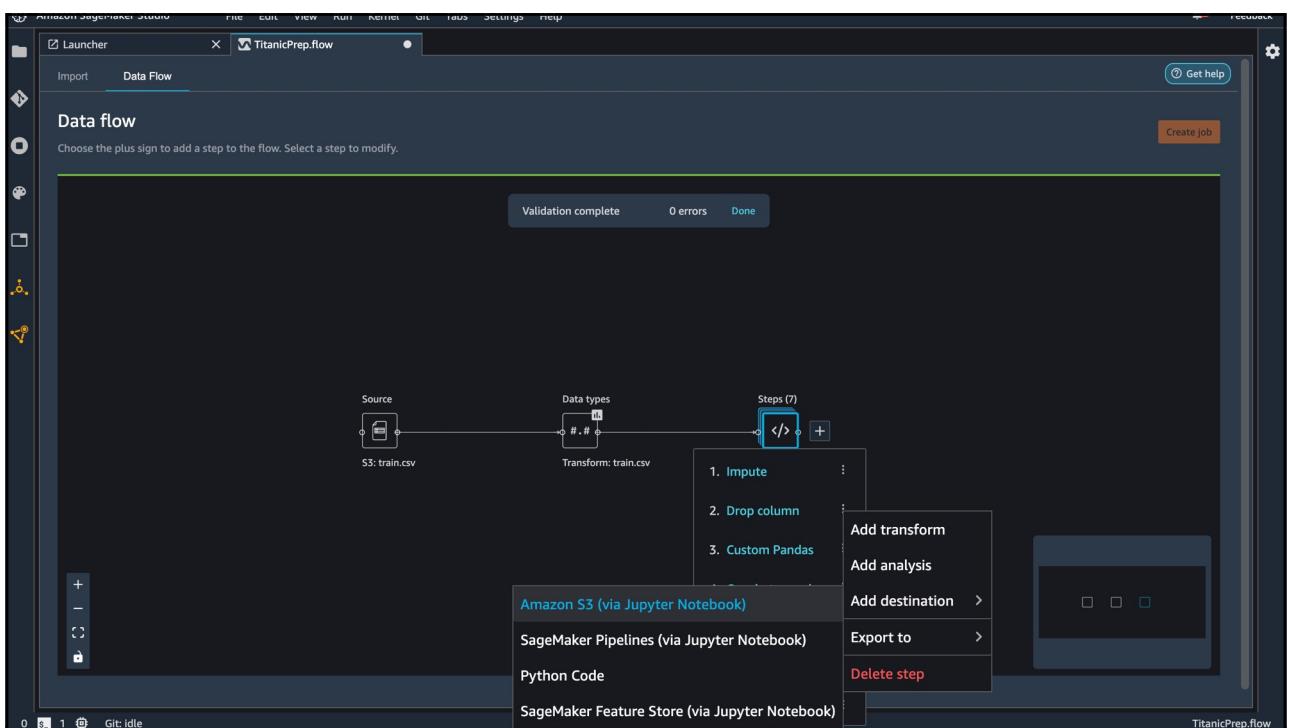
81



82



83



84

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

85

Amazon S3

Buckets

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

Block Public Access settings for this account

Storage Lens

- Dashboards
- AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

Read the S3 resources page for documentation and technical content.

Amazon S3

▶ Account snapshot

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

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-michelin-20220218	US East (N. Virginia) us-east-1	Bucket and objects not public	February 16, 2022, 22:15:58 (UTC+08:00)

86

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with 'Amazon S3' selected under 'Buckets'. The main area shows the 'sagemaker-us-east-1-834120567544' bucket with three objects listed:

Name	Type	Last modified	Size	Storage class
Canvas/	Folder	-	-	-
data_wrangler_flows/	Folder	-	-	-
export-flow-16-15-04-15-fd0b1758/	Folder	-	-	-

87

The screenshot shows the AWS S3 console interface, navigating into the 'export-flow-16-15-04-15-fd0b1758/' folder. The main area shows one object listed:

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

88

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

output/

Objects (1)

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

Actions

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
data-wrangler-flow-processing-16-15-04-15-fd0b1758/	Folder	-	-	-

89

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

data-wrangler-flow-processing-16-15-04-15-fd0b1758/

Objects (1)

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

Actions

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
7f06826f-b540-4996-b6d5-e1a8eb70f110/	Folder	-	-	-

90

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with various options like Buckets, Storage Lens, and Feature spotlight. The main area shows a breadcrumb navigation path: Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/ > output/ > data-wrangler-flow-processing-16-15-04-15-fd0b1758/ > 7f06826f-b540-4996-b6d5-e1a8eb70f110/. Below the path, the folder name '7f06826f-b540-4996-b6d5-e1a8eb70f110/' is displayed. A 'Copy S3 URI' button is located to the right. The 'Objects' tab is selected, showing a table with one item: 'default/' (Type: Folder). There are also buttons for Copy S3 URI, Copy URL, Download, Open, Delete, Actions, and Create folder.

91

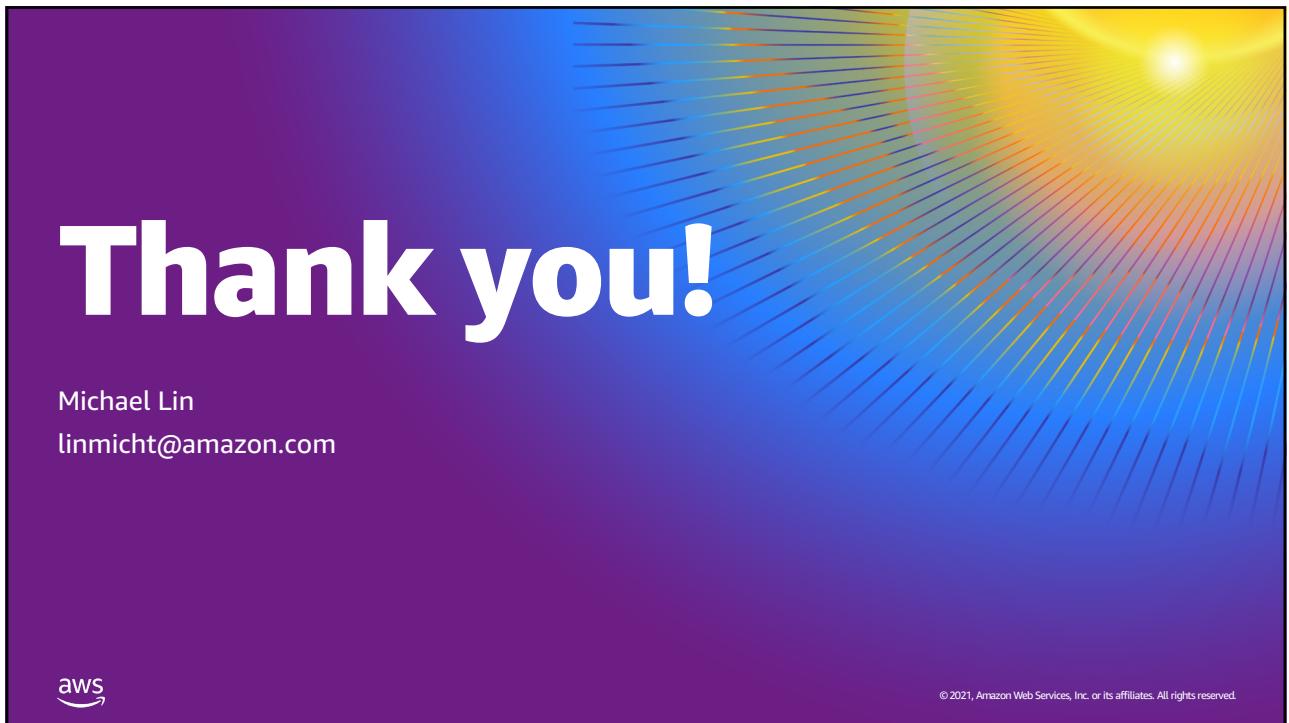
This screenshot is similar to the previous one, showing the AWS S3 console. The breadcrumb path is identical. The folder 'default/' is selected. Inside, there is a single object: 'part-00000-b8bd0b1c-acdc-439d-bf2c-837fa0f2faee-c000.csv'. This is a CSV file, as indicated by the 'csv' extension in the table. The table includes columns for Name, Type, Last modified, Size, and Storage class. The file was last modified on February 16, 2022, at 23:10:52 (UTC+08:00) and is 30.5 KB in size, stored in the Standard storage class.

92

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	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Survived	Fare	Age_Imputer_With_Family_Pclass_3	Pclass_1	Sex_male	Embarked_S	Embarked_C													
2	0	7.25	22	1	1	0	1	1	0											
3	1	71.2833	38	1	0	1	0	0	0	1										
4	1	7.925	26	0	1	0	0	0	1	0										
5	1	53.1	35	1	0	1	0	1	1	0										
6	0	8.05	35	0	1	0	1	1	1	1										
7	0	8.4583	28	0	1	0	1	0	0	0										
8	0	51.8625	54	0	0	1	1	1	1	0										
9	0	21.075	2	1	1	0	1	1	1	0										
10	1	11.1333	27	1	1	0	0	0	1	0										
11	1	30.0708	14	1	0	0	0	0	0	1										
12	1	16.7	4	1	1	0	0	0	1	0										
13	1	26.55	58	0	0	1	0	1	1	0										
14	0	8.05	20	0	1	0	1	1	1	1										
15	0	31.275	39	1	1	0	1	1	1	0										
16	0	7.8542	14	0	1	0	0	0	1	0										
17	1	16	55	0	0	0	0	0	1	0										
18	0	29.125	2	1	1	0	1	0	0	0										
19	1	13	28	0	0	0	1	1	1	0										
20	0	18	31	1	1	0	0	0	1	0										
21	1	7.225	28	0	1	0	0	0	0	1										
22	0	26	35	0	0	0	0	1	1	1										
23	1	13	34	0	0	0	1	1	1	0										
24	1	8.0292	15	0	1	0	0	0	0	0										
25	1	35.5	28	0	0	1	1	1	1	0										
26	0	21.075	8	1	1	0	0	0	1	0										
27	1	31.3875	38	1	1	0	0	0	1	0										
28	0	7.225	28	0	1	0	1	0	0	1										
29	0	263	19	1	0	1	1	1	1	0										
30	1	7.8792	28	0	1	0	0	0	0	0										
31	0	7.8958	28	0	1	0	1	1	1	0										
32	0	27.7208	40	0	0	1	1	1	0	1										
33	1	146.5208	28	1	0	1	0	0	0	1										
34	1	7.75	28	0	1	0	0	0	0	0										
35	0	10.5	66	0	0	0	1	1	1	0										
36	0	82.1708	28	1	0	1	1	1	0	1										
37	0	52	42	1	0	1	1	1	1	0										
38	1	7.2292	28	0	1	0	1	0	1	0										

93



94