



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 Storage Gateway feature
- Datasets IoT Analytics feature

Add user

Launch app

Delete Domain

Edit Settings

Authentication method
AWS Identity and Access Management (IAM)

2

1

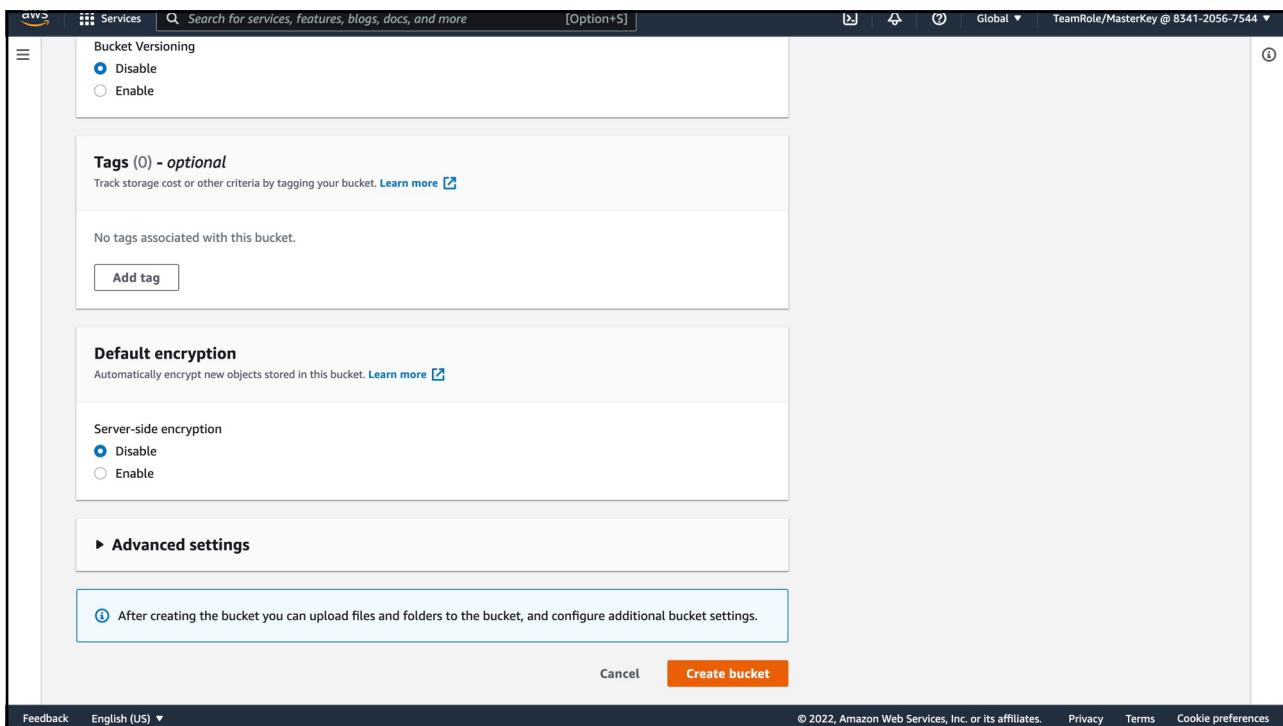
The screenshot shows the AWS S3 console. On the left, there's a sidebar with 'Buckets' selected. The main area displays an 'Account snapshot' with a link to 'View Storage Lens dashboard'. Below it is a table titled 'Buckets (3) Info' showing three buckets: 'sagemaker-studio-834120567544-5zbey54q71a' and 'sagemaker-us-east-1-834120567544'. Both buckets are located in 'US East (N. Virginia) us-east-1' and have 'Objects can be public' access. The table has columns for Name, AWS Region, Access, and Creation date.

3

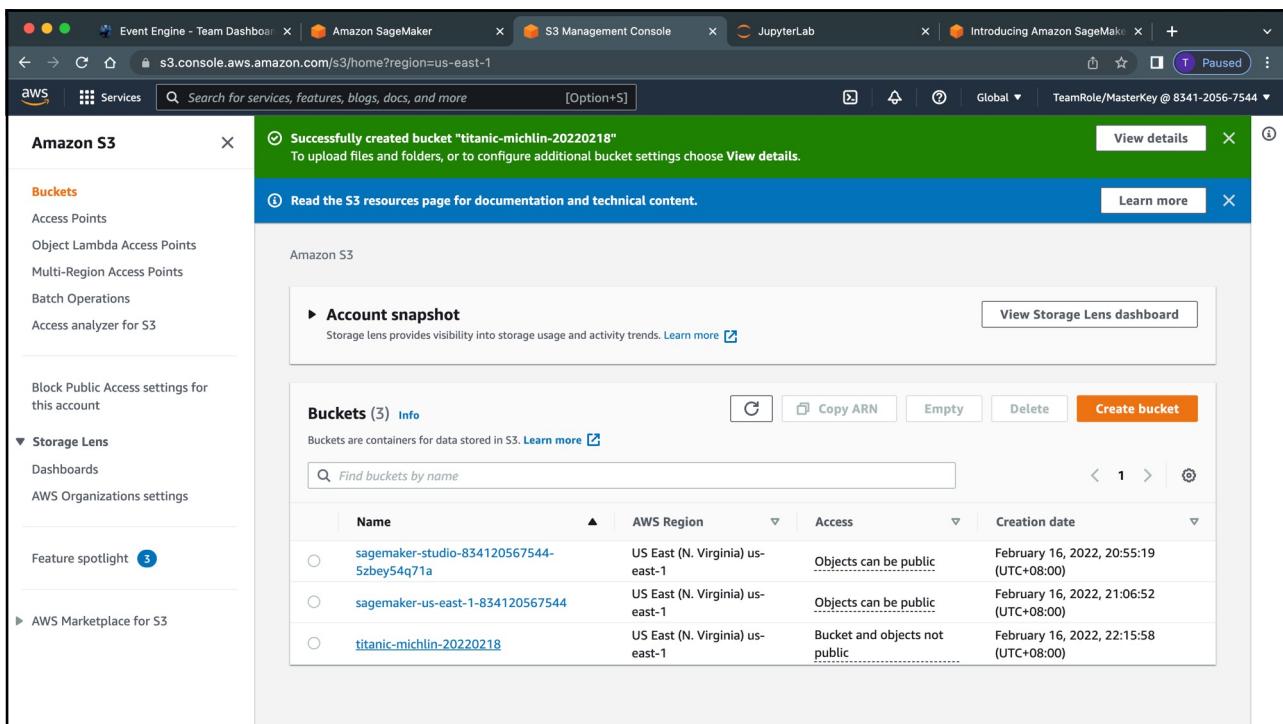
The screenshot shows the 'Create bucket' wizard. In the 'General configuration' section, the 'Bucket name' is set to 'titanic-michlin-20220218' and the 'AWS Region' is set to 'US East (N. Virginia) us-east-1'. In the 'Object Ownership' section, the 'ACLs disabled (recommended)' option is selected, indicating that all objects in the bucket are owned by the account owner.

4

2



5



6

3

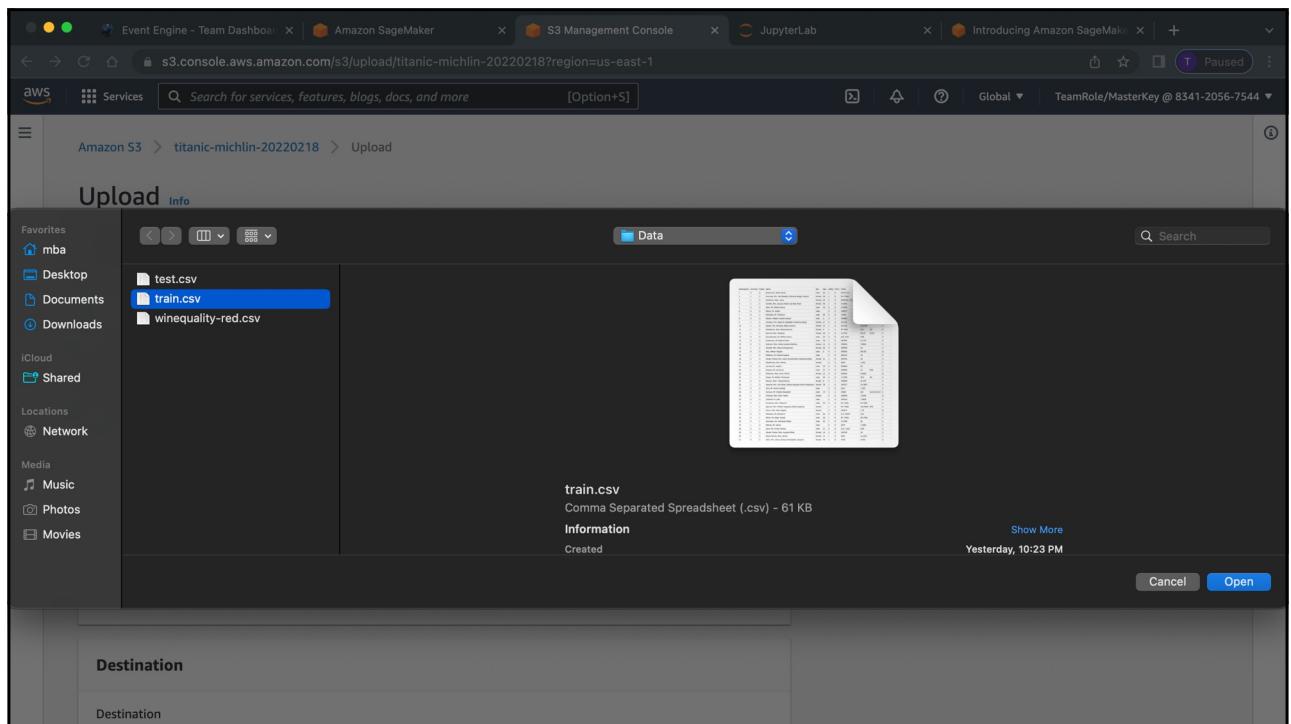
The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with various options like Buckets, Storage Lens, and Feature spotlight. The main area shows the 'titanic-michlin-20220218' bucket. At the top, there are tabs for Objects, Properties, Permissions, Metrics, Management, and Access Points. Below the tabs, a section titled 'Objects (0)' is displayed. It includes a search bar, a file upload button, and a table header for Name, Type, Last modified, Size, and Storage class. A message at the bottom states 'No objects' and 'You don't have any objects in this bucket.'

7

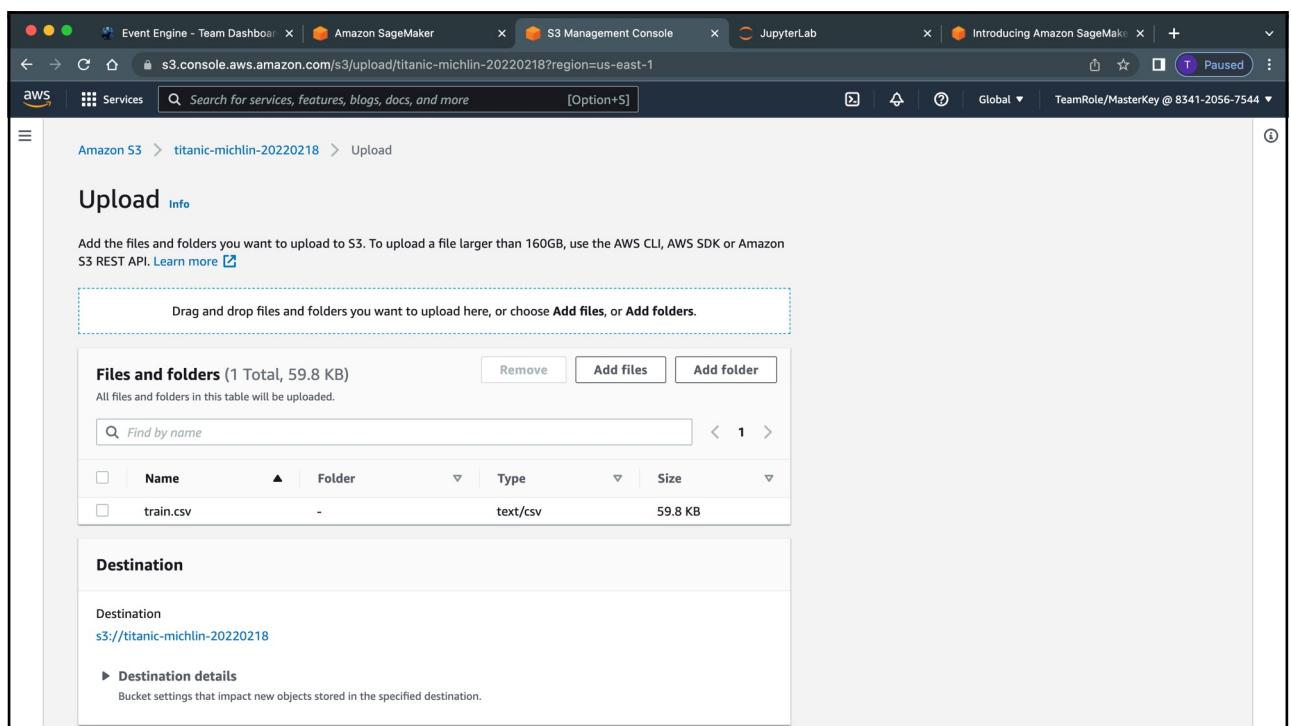
The screenshot shows the 'Upload' step in the AWS S3 Management Console. The URL in the browser is 's3.console.aws.amazon.com/s3/upload/titanic-michlin-20220218?region=us-east-1'. The interface includes a large 'Upload' button with a 'Info' link, a note about supported file sizes, and a 'Drag and drop files and folders you want to upload here, or choose Add files, or Add folders.' area. Below this is a 'Files and folders (0)' table with a 'Find by name' search bar and a message 'All files and folders in this table will be uploaded.' Further down is a 'Destination' section with a 'Destination' input field.

8

4



9



10

5

The screenshot shows the AWS S3 Management Console interface. A progress bar at the top indicates the upload of 'train.csv' to the 'titanic-michlin-20220218' bucket is 100% complete. Below the progress bar, the 'Upload' button is highlighted in orange.

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)

Name	Folder	Type	Size
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**

11

The screenshot shows the AWS S3 Management Console after the upload has completed. A green header bar indicates 'Upload succeeded'. The main area displays the upload status summary and a detailed file list.

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 (1 Total, 59.8 KB)

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

12

The screenshot shows the AWS SageMaker search results for 'sagemaker'. The left sidebar includes 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 displays 'Search results for "sagemaker"' under 'Services' and 'Features'. Under 'Services', 'Amazon SageMaker' is highlighted with a callout: 'Build, Train, and Deploy Machine Learning Models' and 'Top features: SageMaker Studio, Autopilot, SageMaker Canvas'. Other services listed are AWS Glue DataBrew and SageMaker Studio. Under 'Features', there are four items: SageMaker Studio (Amazon SageMaker feature), SageMaker Canvas (Amazon SageMaker feature), Notebooks (IoT Analytics feature), and Autopilot.

13

The screenshot shows the 'SageMaker Domain' configuration page. The left sidebar is identical to the previous screenshot. The main content area shows the 'SageMaker Domain' section with a table for 'Users'. It lists one user: 'default-1645016110854' (Status: Ready, Modified on: Feb 16, 2022 13:01 UTC, Created on: Feb 16, 2022 13:00 UTC). Below the users is a 'Domain' section with details: Status (Ready), Domain ID (d-iceiqcsfx9pm), Execution role (arn:aws:iam::834120567544:role/TteamRole), and Authentication method (AWS Identity and Access Management (IAM)). A note states: 'Use Domain for troubleshooting and tracking usage. The status shown is for the SageMaker Studio service, and is not the status of compute resources such as EC2 instances to execute notebooks.' At the bottom, it says 'Projects' and 'Amazon SageMaker project templates enabled for this account'.

14

Amazon SageMaker

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

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

Amazon SageMaker > SageMaker Domain

SageMaker Domain

Users Add user

Name	Modified on	Created on
default-1645016110854	Feb 16, 2022 13:01 UTC	Feb 16, 2022 13:00 UTC

Domain How to delete the domain Delete

Status	Domain ID	Execution role
Ready The status of the SageMaker Domain, and is not the status of the compute resources such as EC2 instances to execute notebook.	d-iceiqcsfx9pm Use the SageMaker Domain ID for troubleshooting and tracking usage.	arn:aws:iam::834120567544:role/TteamRole

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

Projects

Amazon SageMaker project templates enabled for this account
Launch constraint role: arn:aws:iam::834120567544:role/service-role/AmazonSageMakerServiceCatalogProductsLaunchRole
Product use role: arn:aws:iam::834120567544:role/service-role/AmazonSageMakerServiceCatalogProductsUserRole

Amazon SageMaker project templates enabled for Studio users

https://console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/open/d-iceiqcsfx9pm/default-1645016110854 © 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

A context menu is open over the domain entry, showing options: Open Link in New Tab, Open Link in New Window, Open Link in Incognito Window, Save Link As..., Copy Link Address, and Inspect.

15

Event Engine - Team Dashboard | Amazon SageMaker | Creating application - SageMaker

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

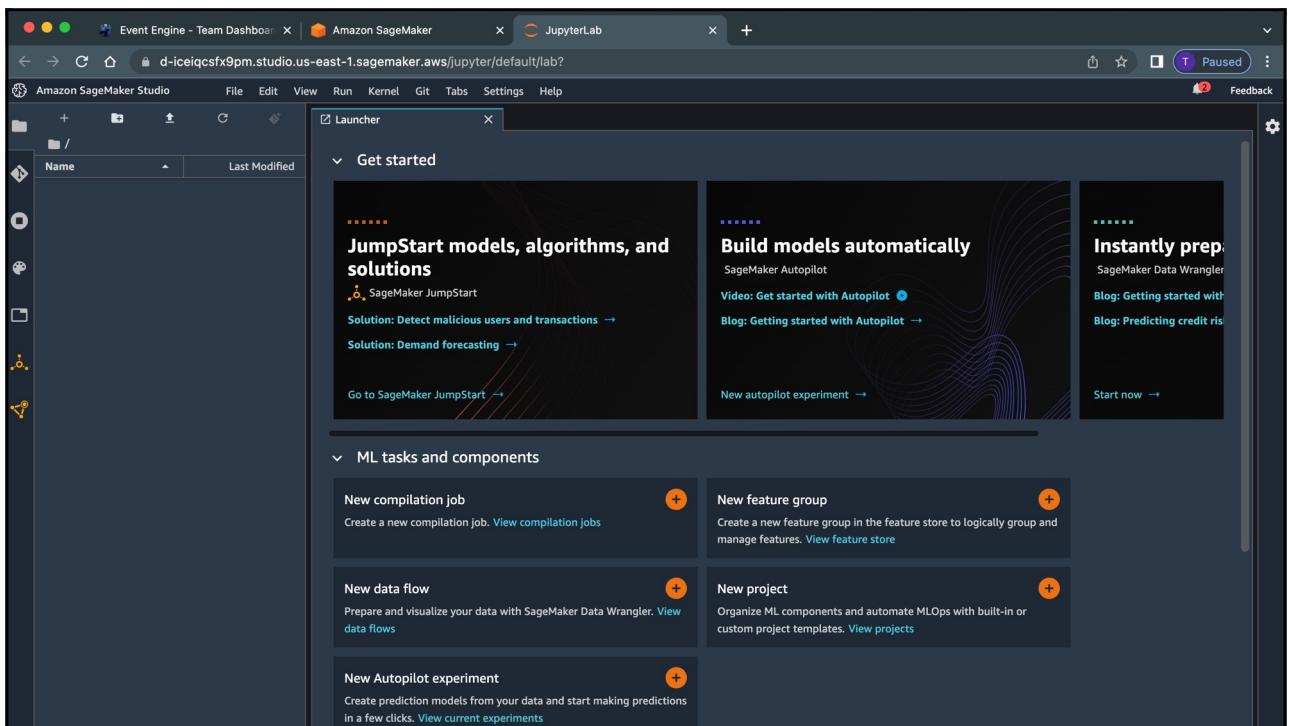
Paused



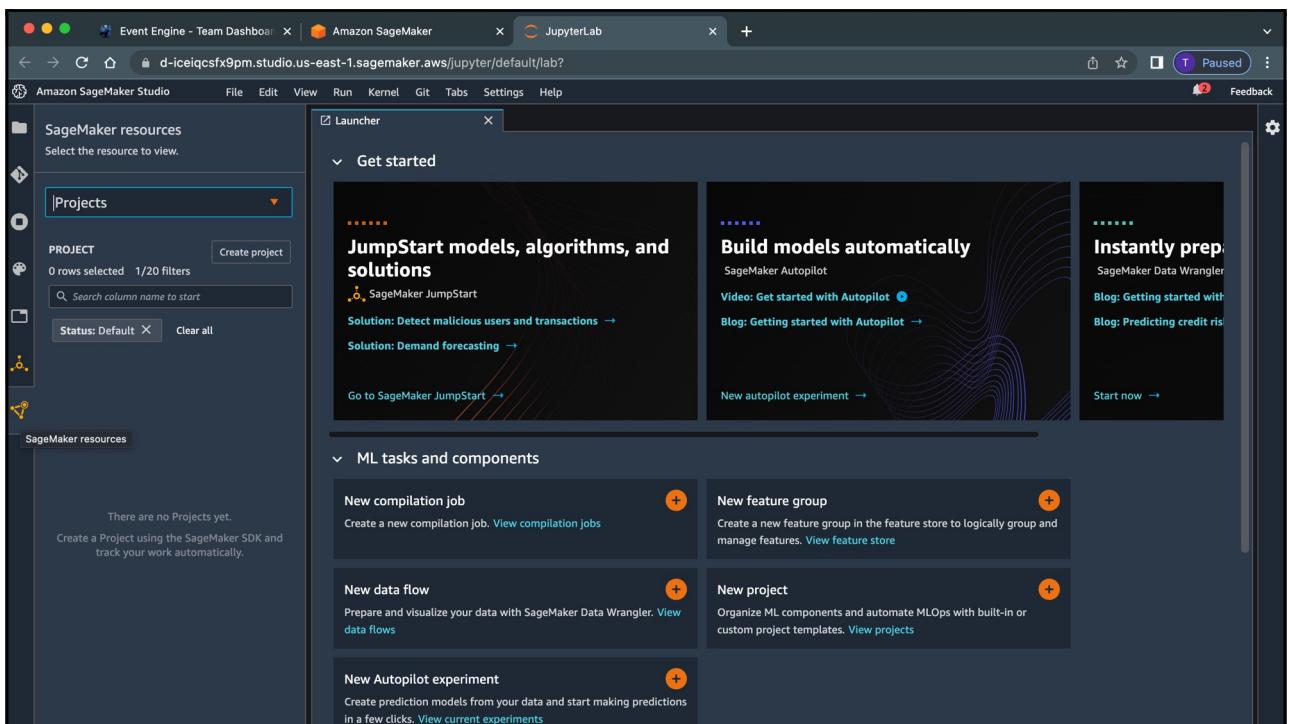
Amazon SageMaker Studio

Creating the JupyterServer application default...

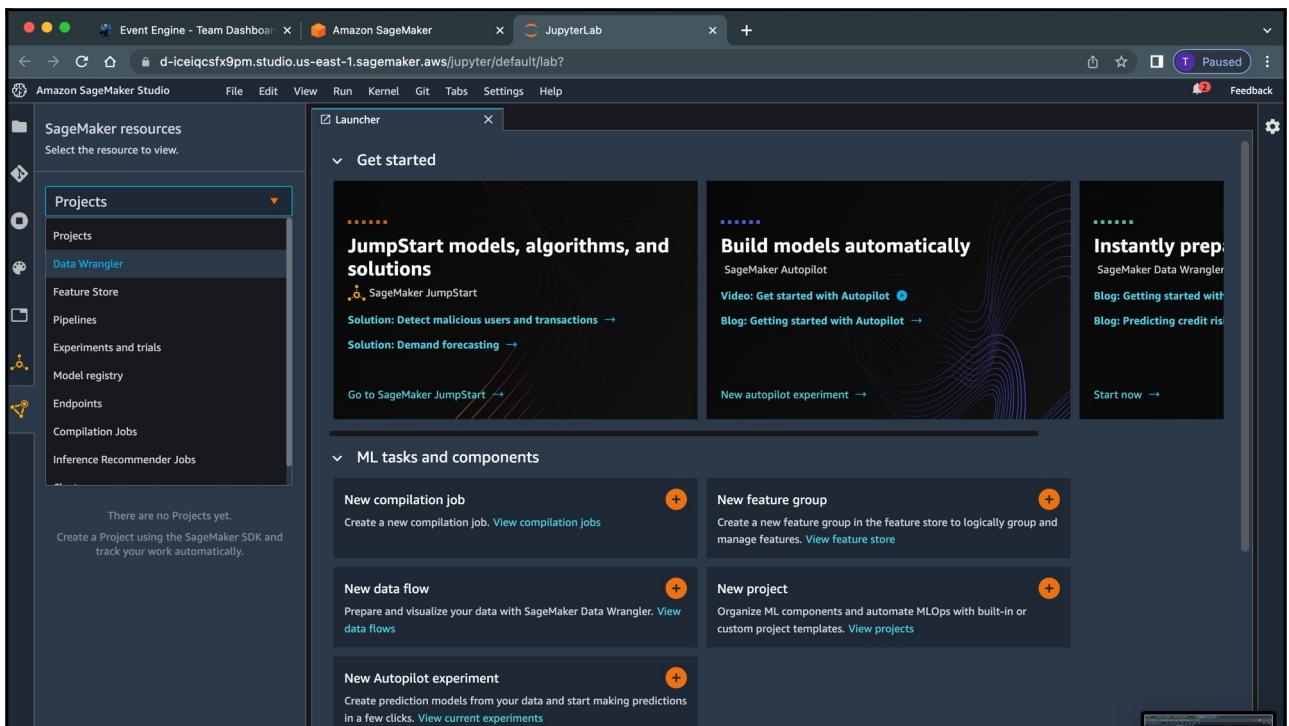
16



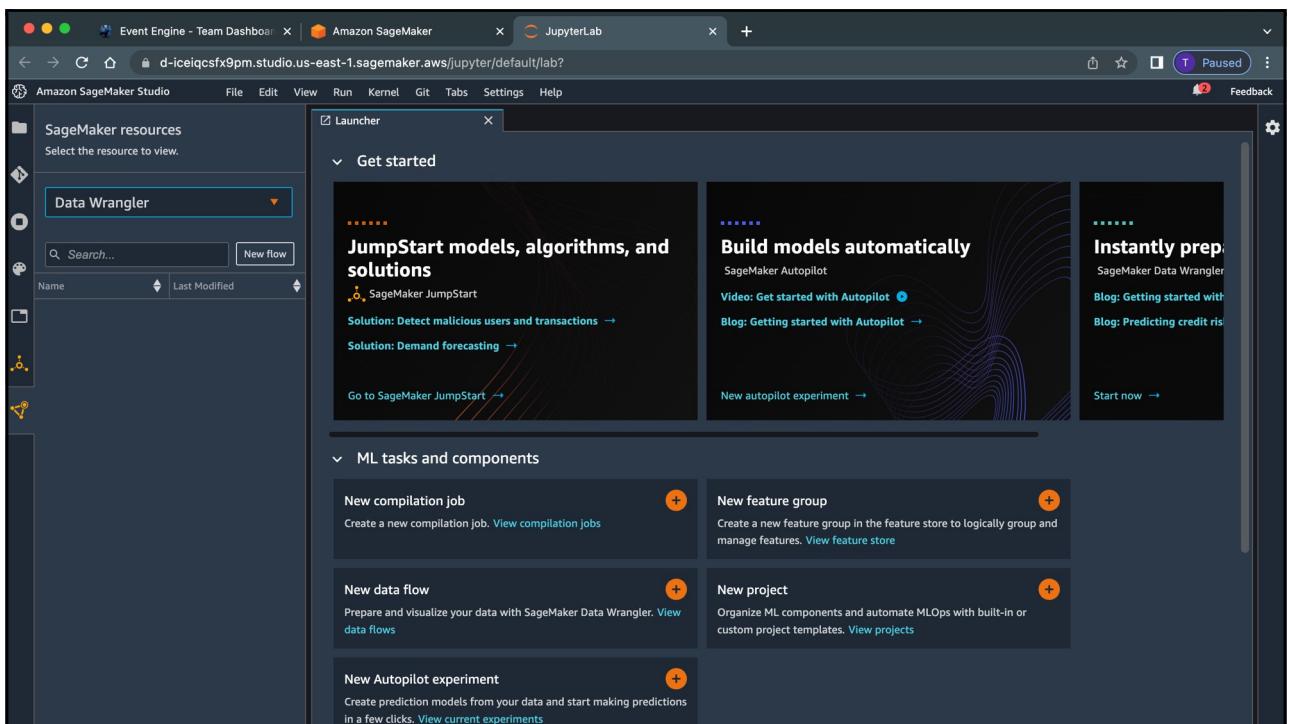
17



18

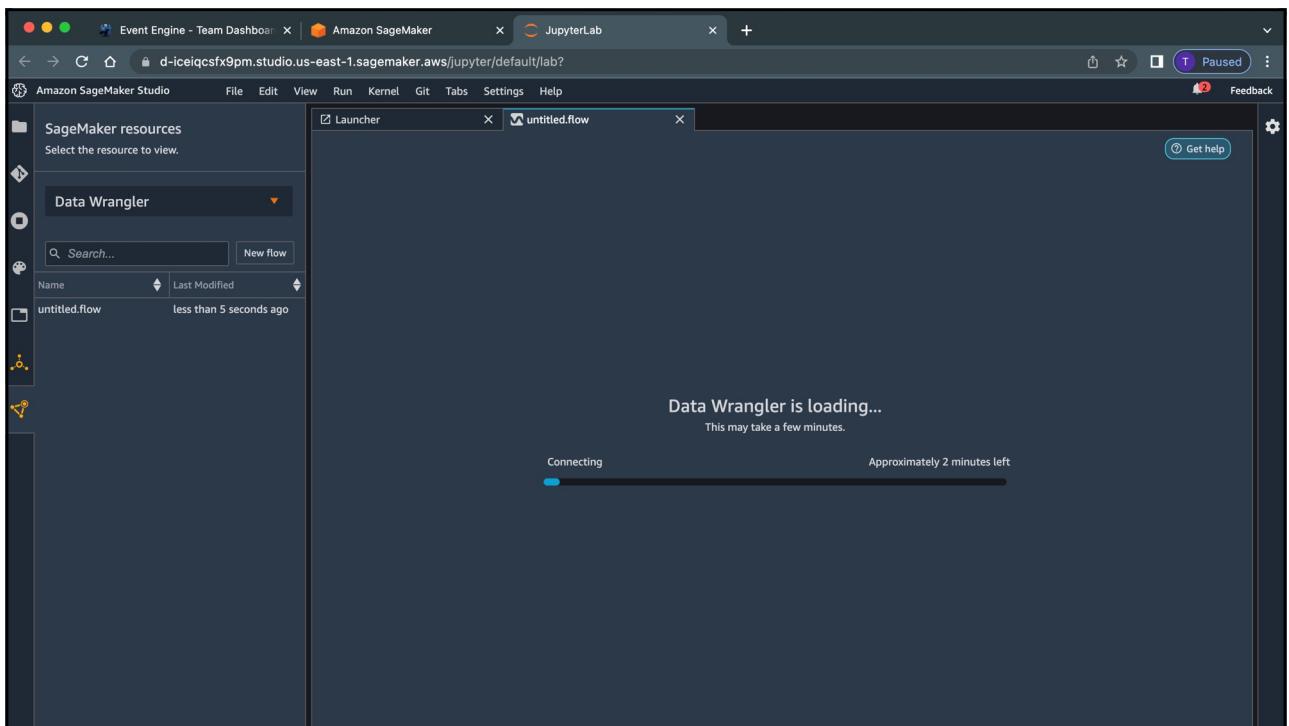


19

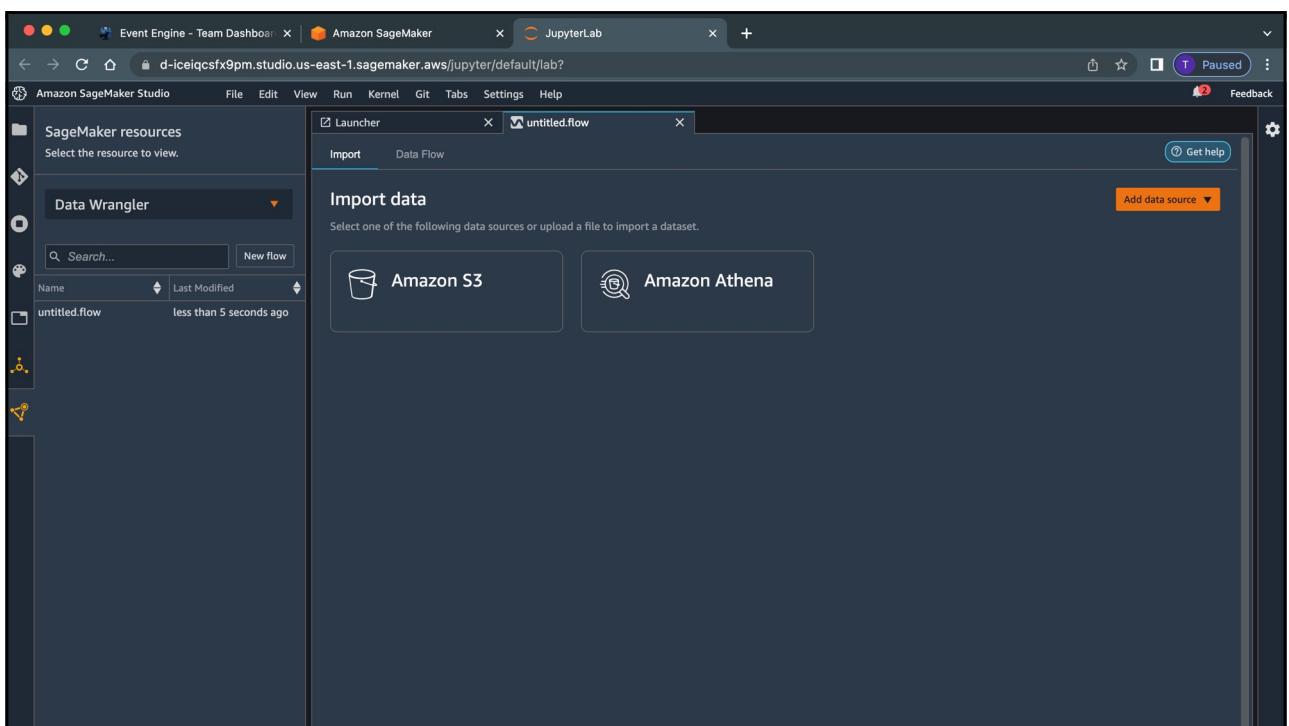


20

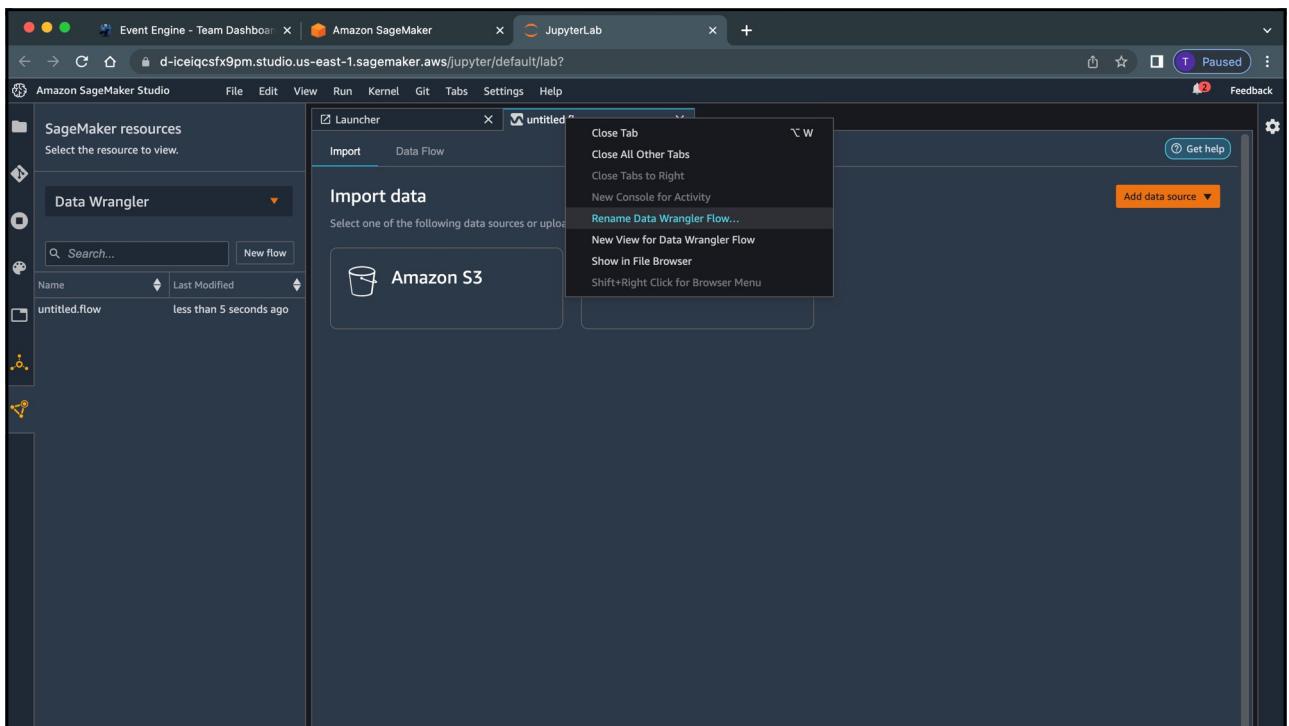
10



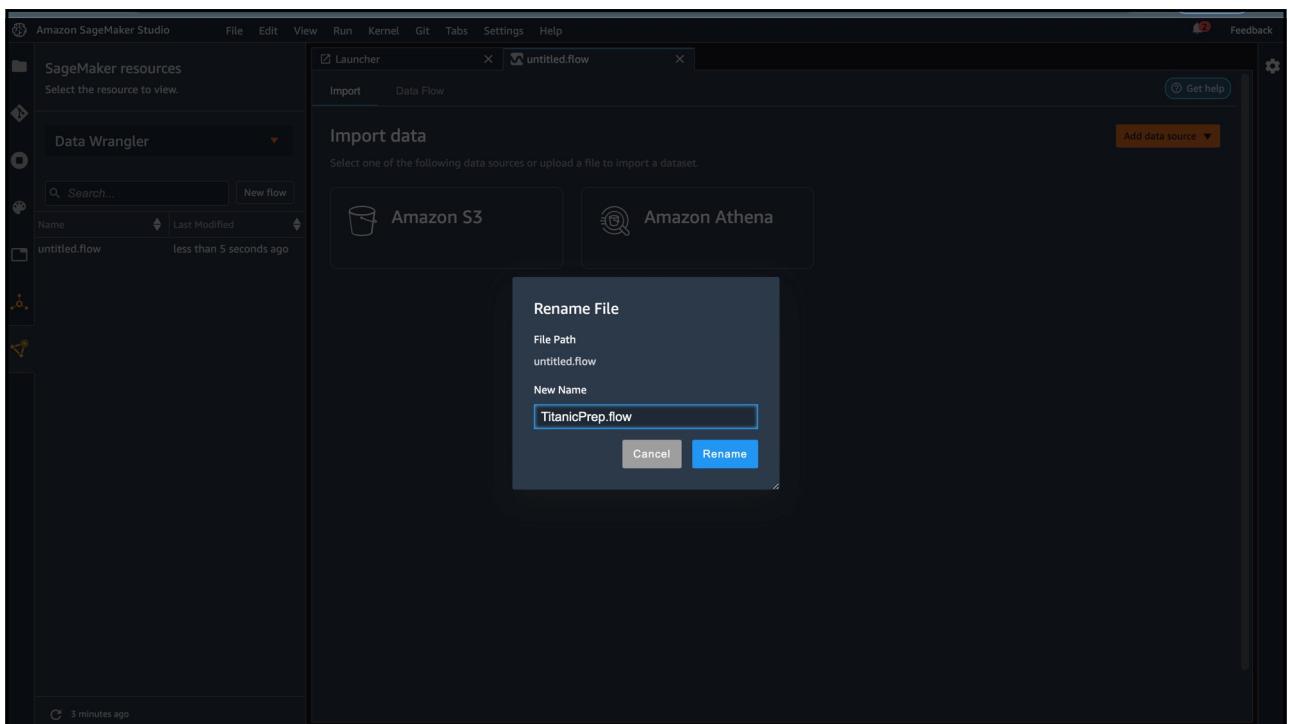
21



22

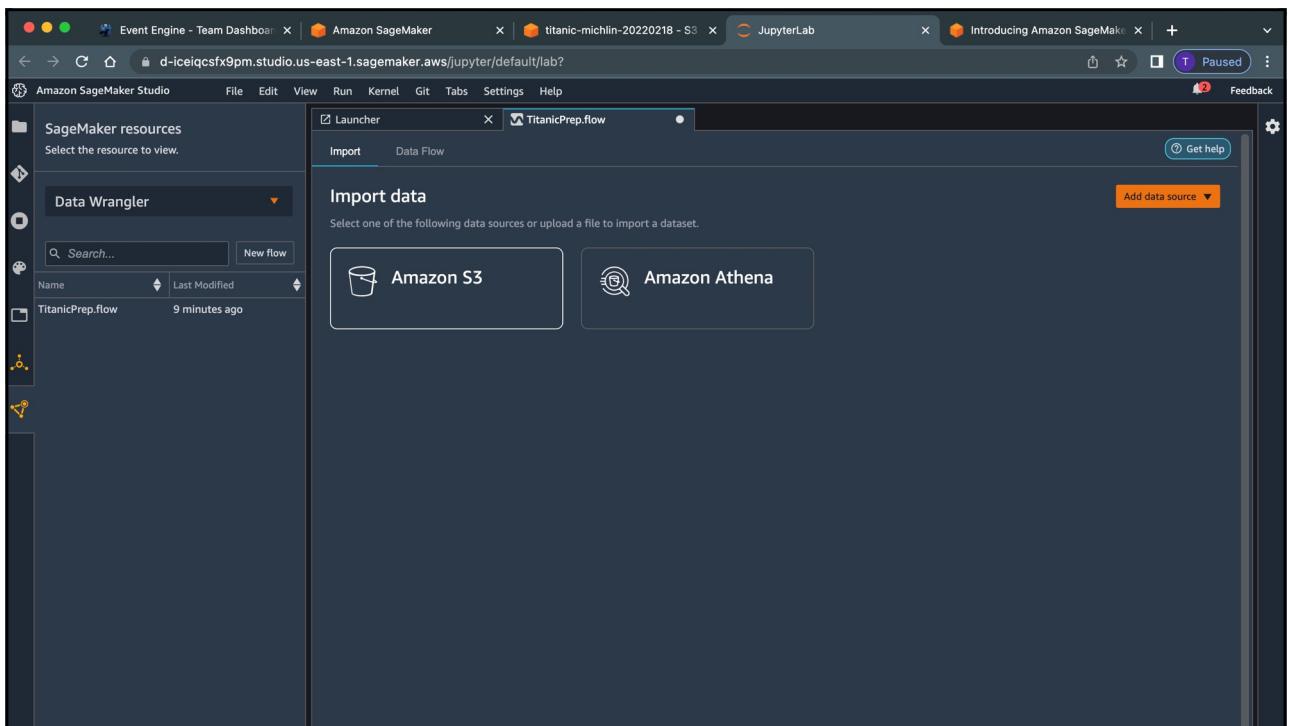


23

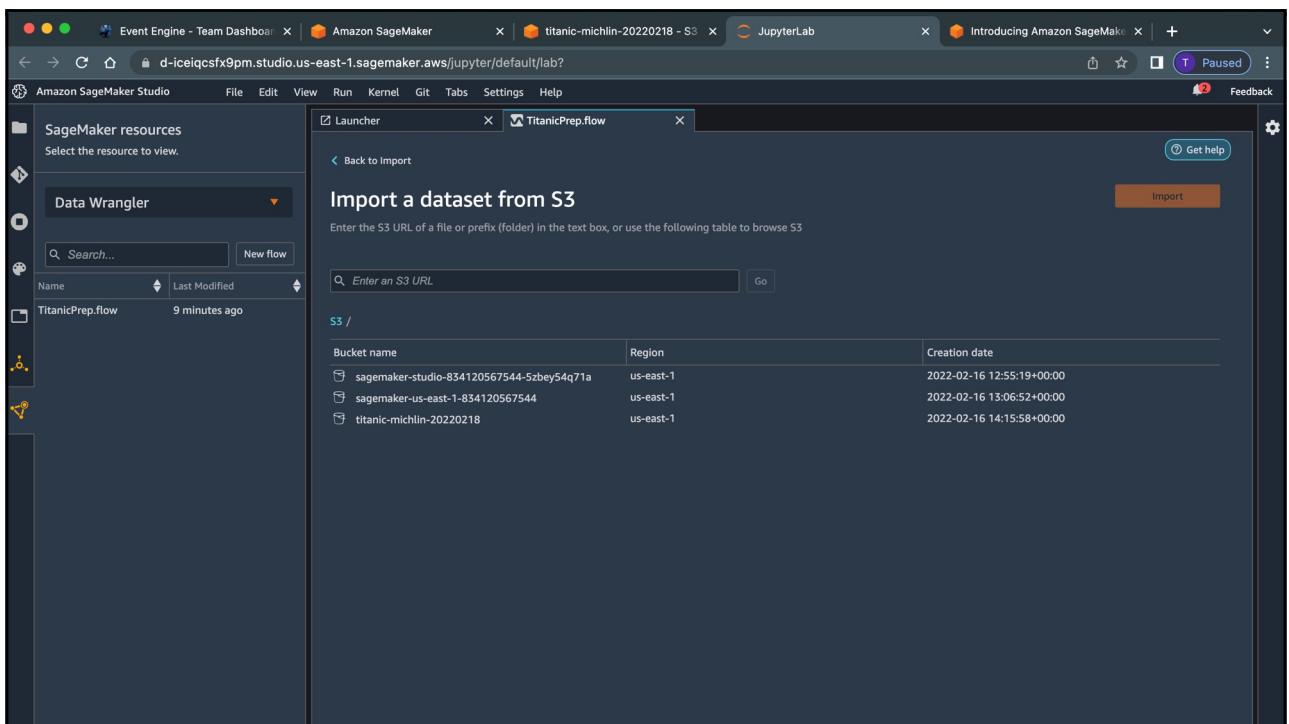


24

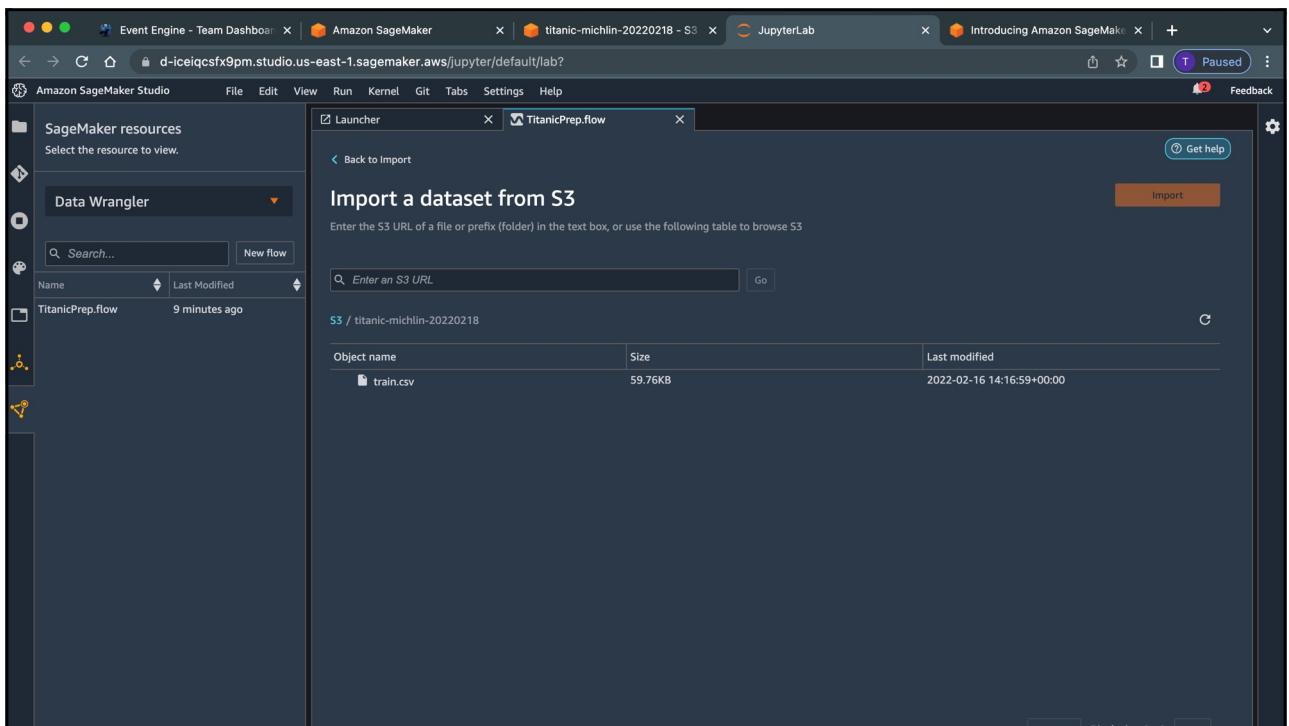
12



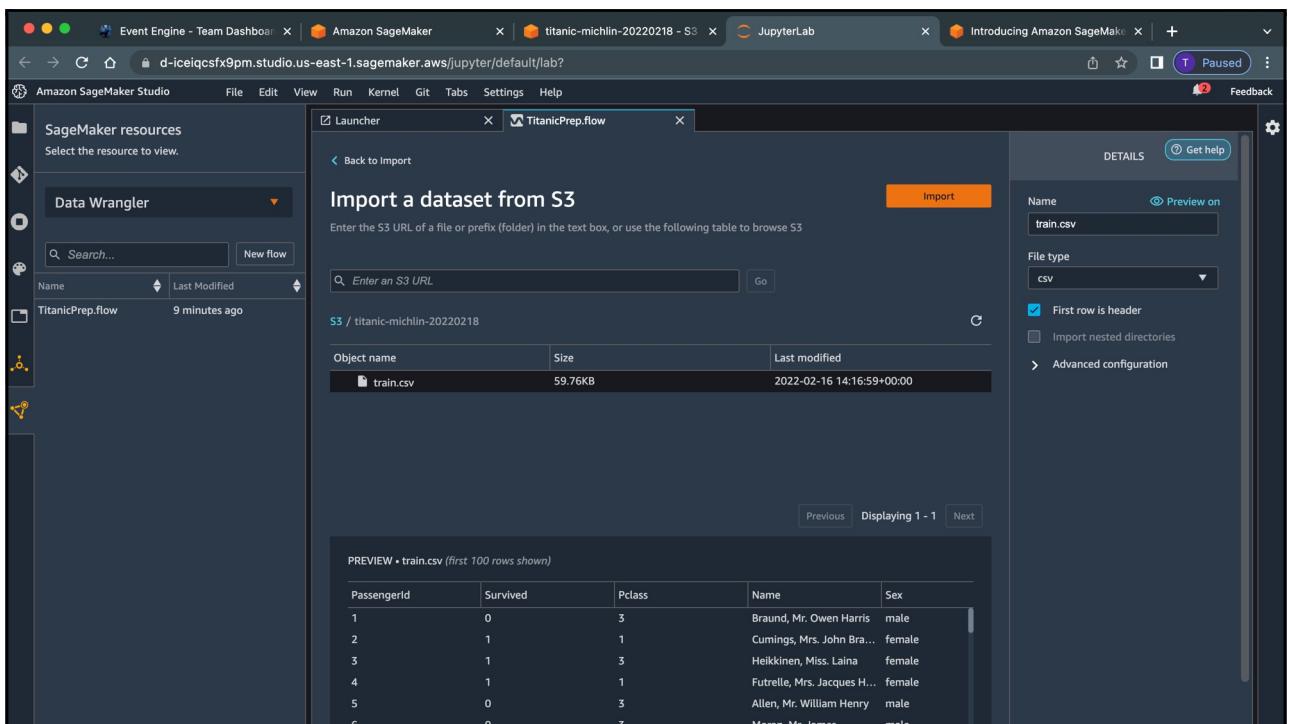
25



26



27



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, with a search bar "Enter an S3 URL" and a table showing an object named "train.csv" from the path "S3 / titanic-michelin-20220218". The "DETAILS" panel on the right shows "Name: train.csv", "File type: csv", and "First row is header" checked. Below the preview table, a "PREVIEW" section shows the first 100 rows of the "train.csv" file.

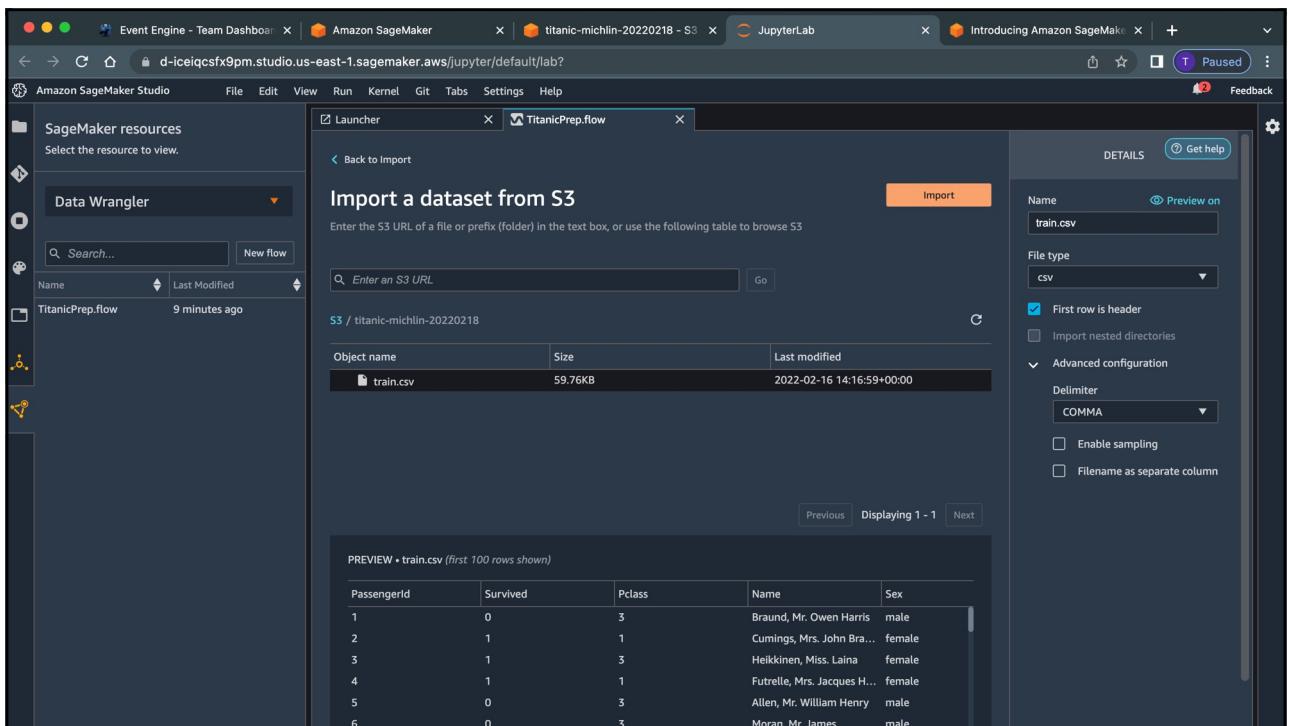
PassengerId	Survived	Pclass	Name	Sex
1	0	3	Braund, Mr. Owen Harris	male
2	1	1	Cumings, Mrs. John Bra...	female
3	1	3	Heikkinen, Miss. Laina	female
4	1	1	Futrelle, Mrs. Jacques H...	female
5	0	3	Allen, Mr. William Henry	male
6	0	3	Moran, Mr. James	male

29

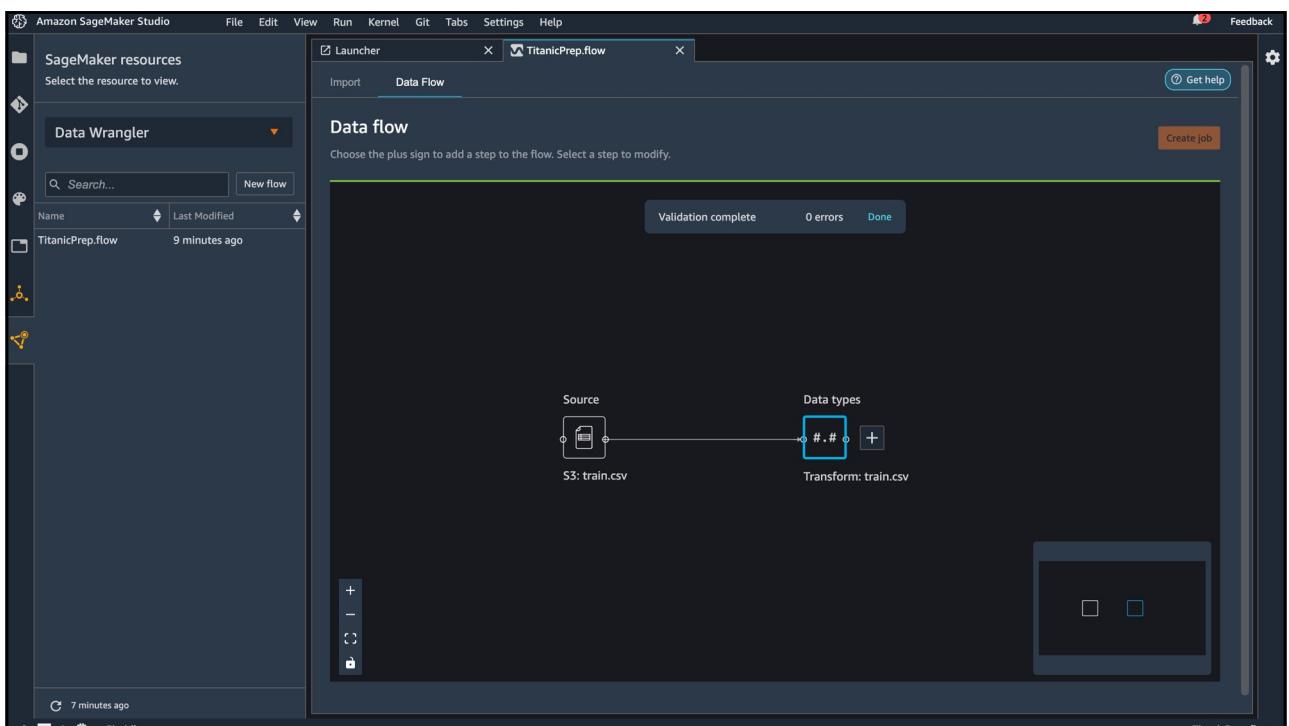
This screenshot is nearly identical to the one above, showing the same interface in Amazon SageMaker Studio. The "Data Wrangler" sidebar is open, and the "TitanicPrep.flow" launcher is active. The "Import a dataset from S3" section shows the "train.csv" file from the "titanic-michelin-20220218" bucket. The "DETAILS" panel on the right shows the same configuration: "Name: train.csv", "File type: csv", and "First row is header" checked. The "PREVIEW" section at the bottom shows the first 100 rows of the CSV file.

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

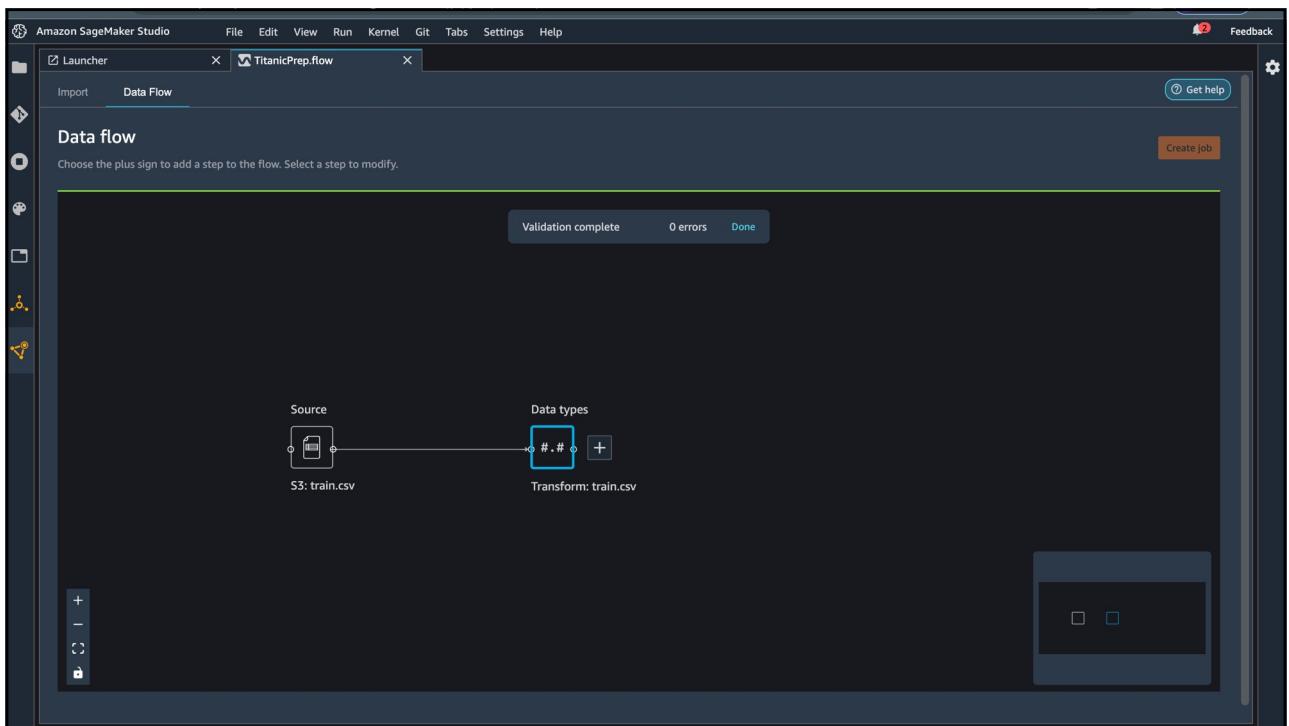
30



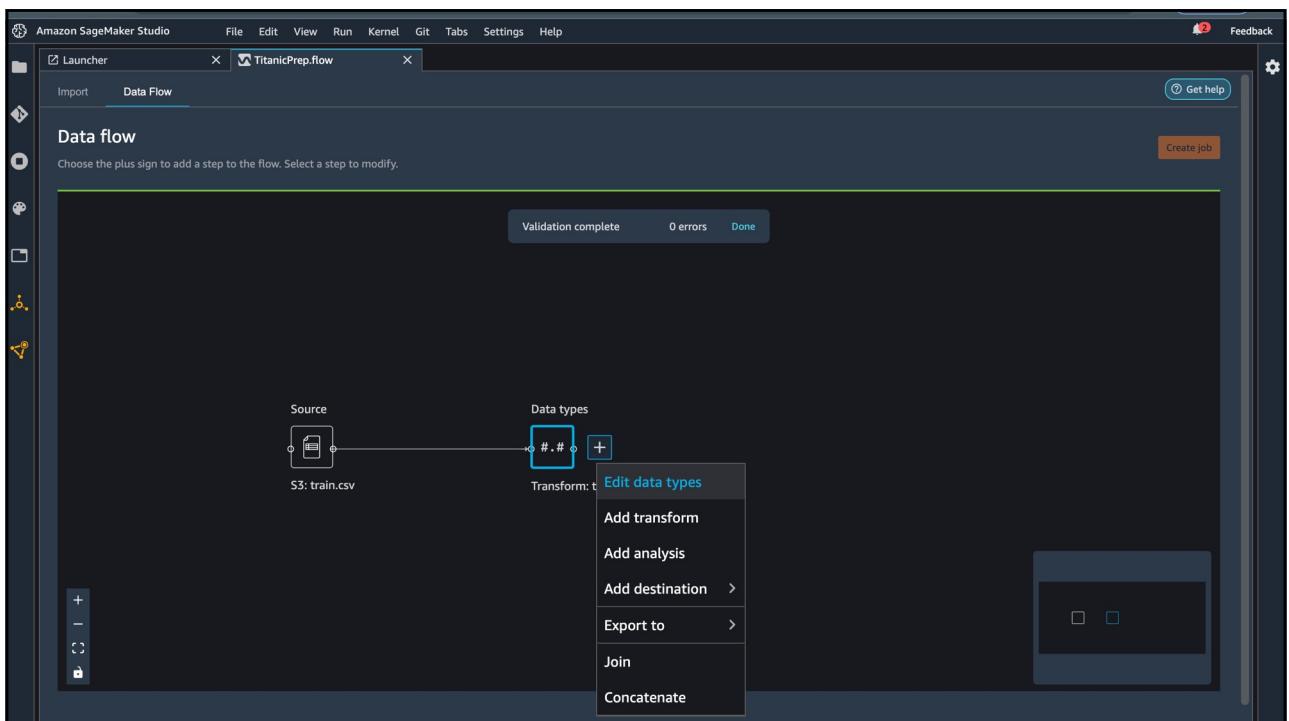
31



32



33



34

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	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	
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 Plank, Mrs. Juli...	female	31
20	1	3	Masseyman, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

CONFIGURE TYPES

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

35

Data flow

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

Validation complete 0 errors Done

Source

S3: train.csv

Transform:

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

36

Histogram: Untitled

No Preview available

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

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

Analysis type: Histogram

Analysis type: Bias Report
Analysis type: Custom Visualization
Analysis type: Duplicate rows
Analysis type: Feature Correlation
Analysis type: Histogram
Analysis type: Multicollinearity
Analysis type: Quick Model
Analysis type: Scatter Plot
Analysis type: Table Summary
Analysis type: Select...
Analysis type: Optional

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

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
8	0	3	Palsson, Master, Gosta ...	male	2

Analysis type: Table Summary

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

Analysis name: Summary

38

19

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

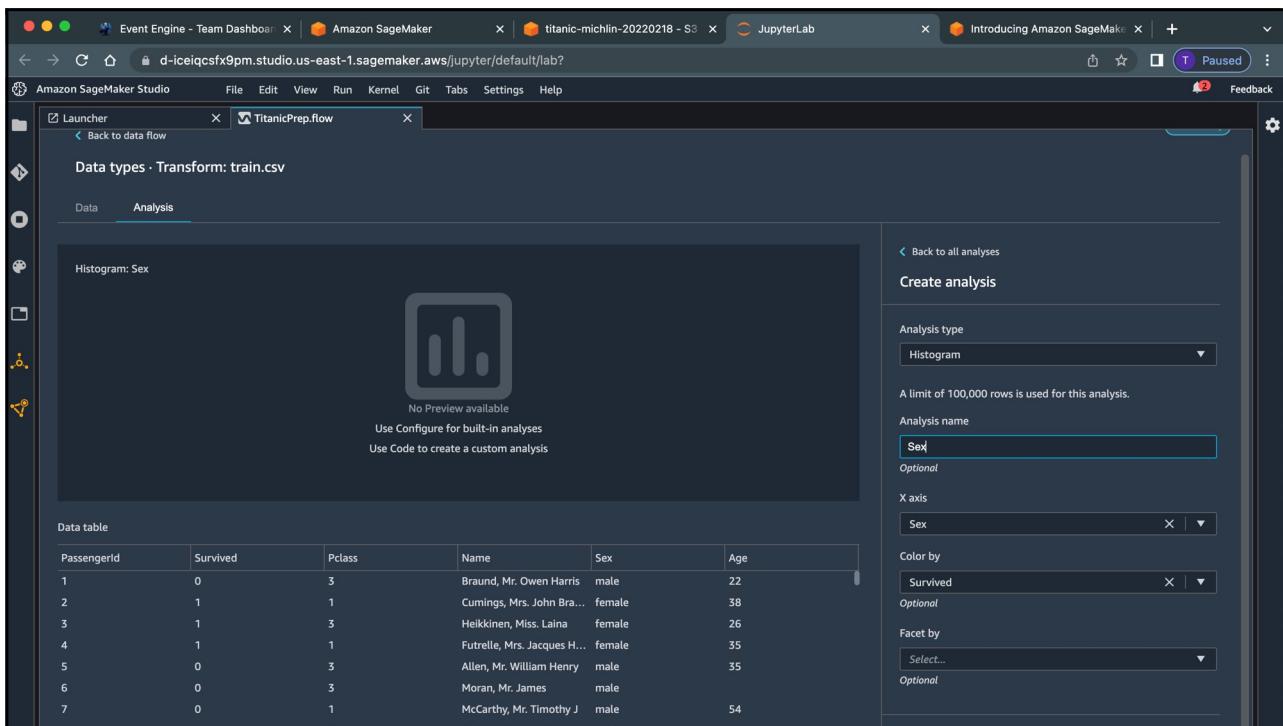
PassengerId	Survived	Pclass	Name	Sex	Age
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allan, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	2
7	0	1	McCarthy, Mr. Timothy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

39

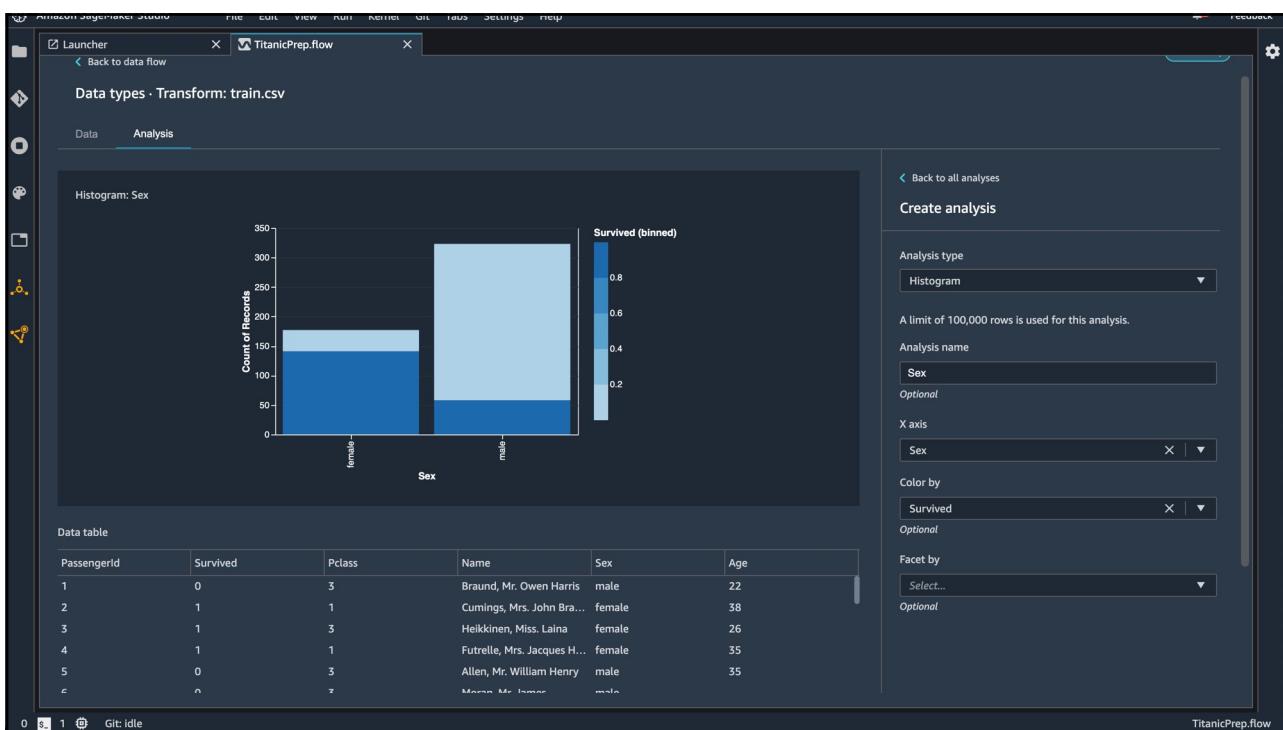
Data types · Transform: train.csv

Summary
Table Summary

40

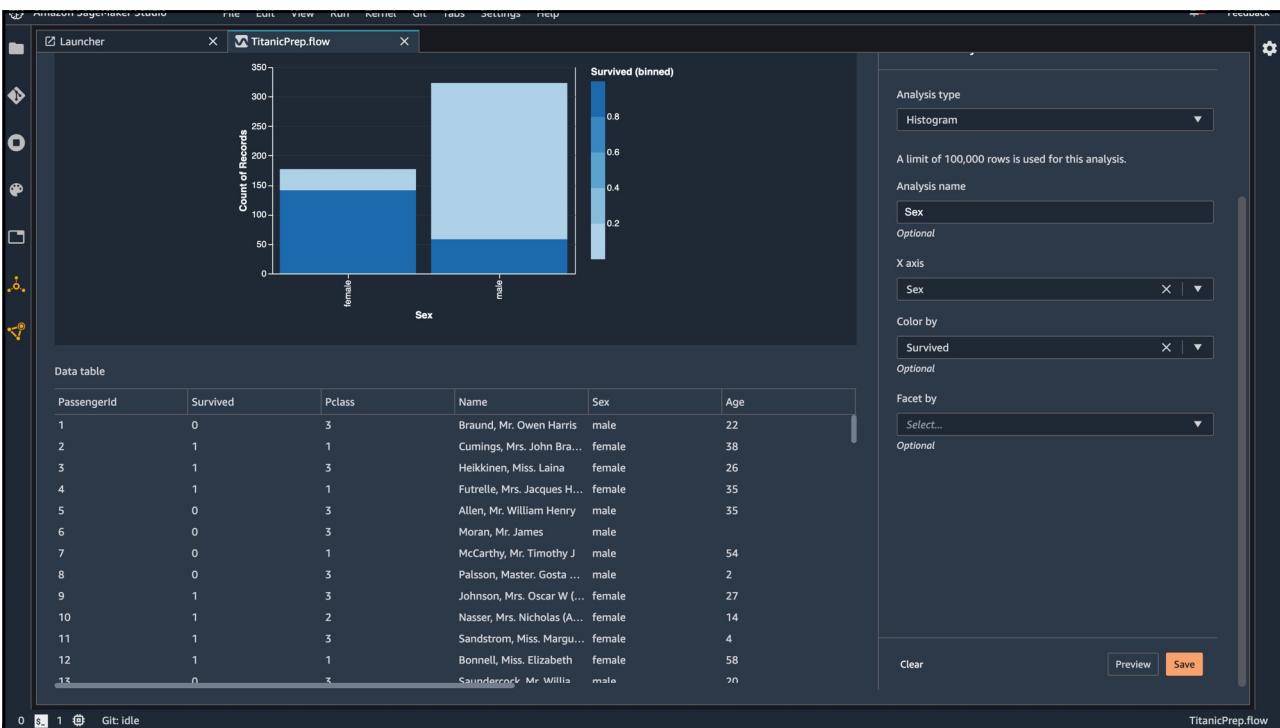


41

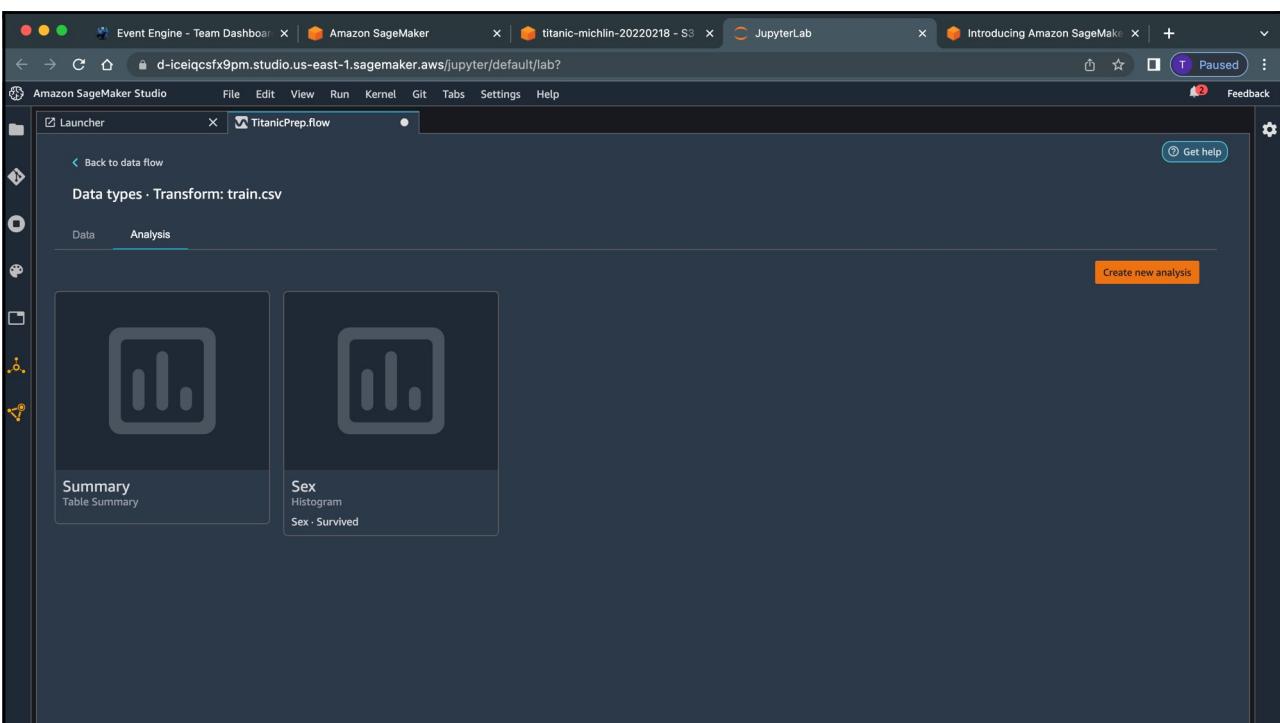


42

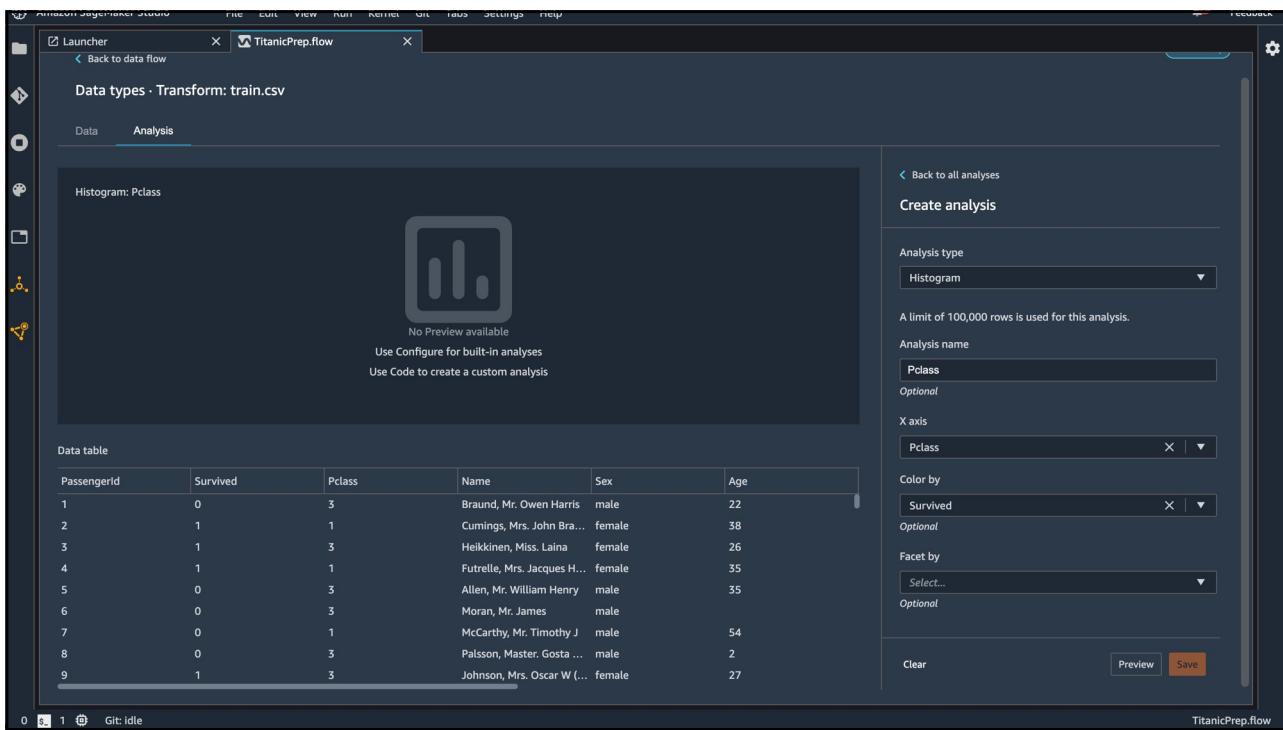
21



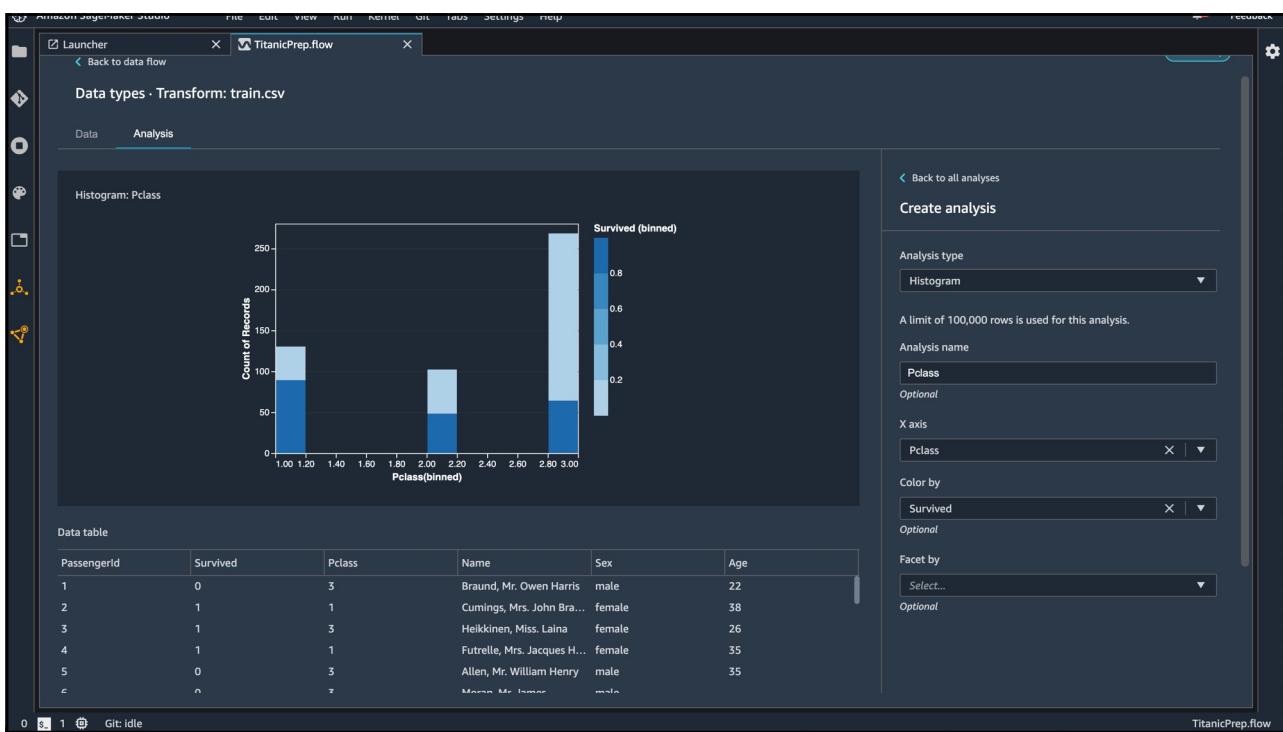
43



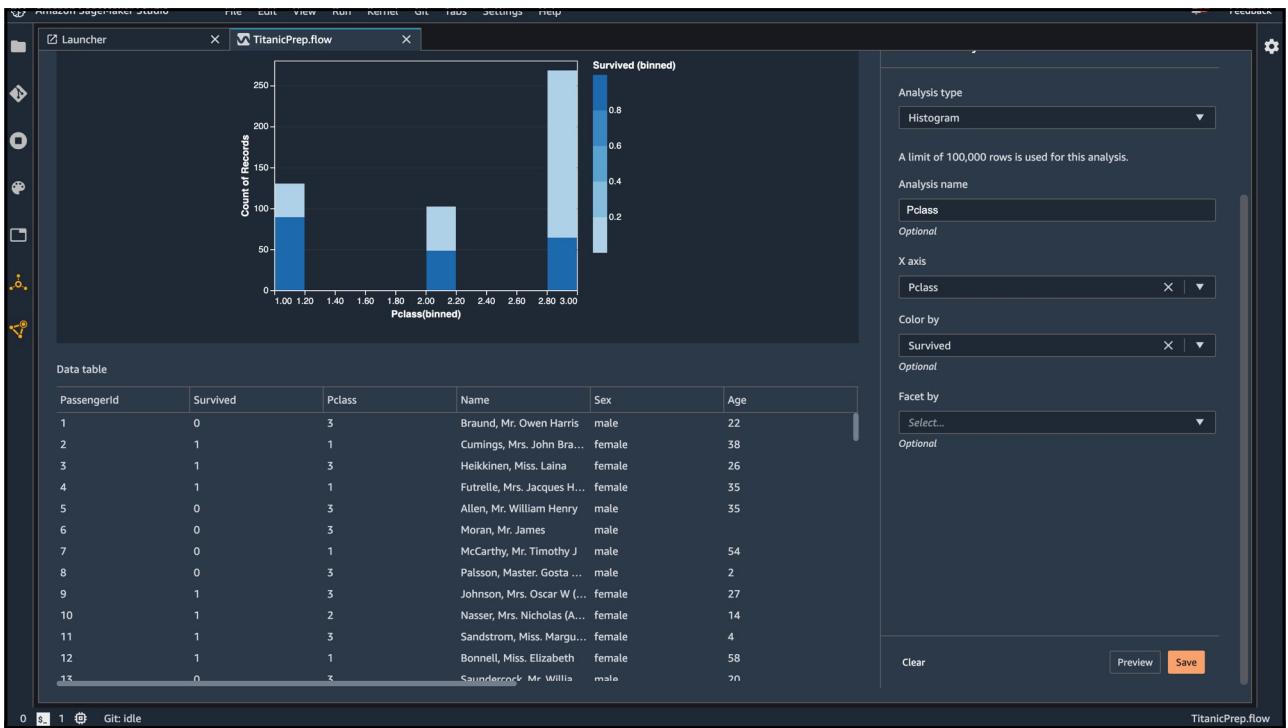
44



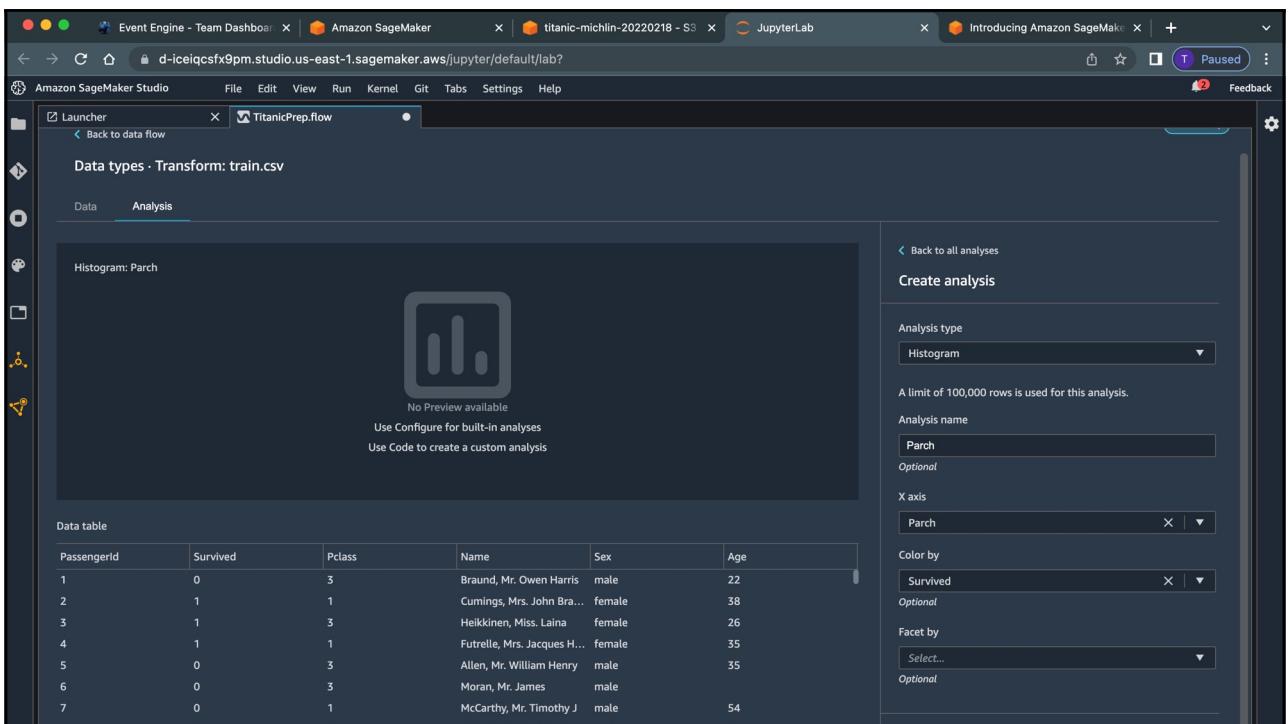
45



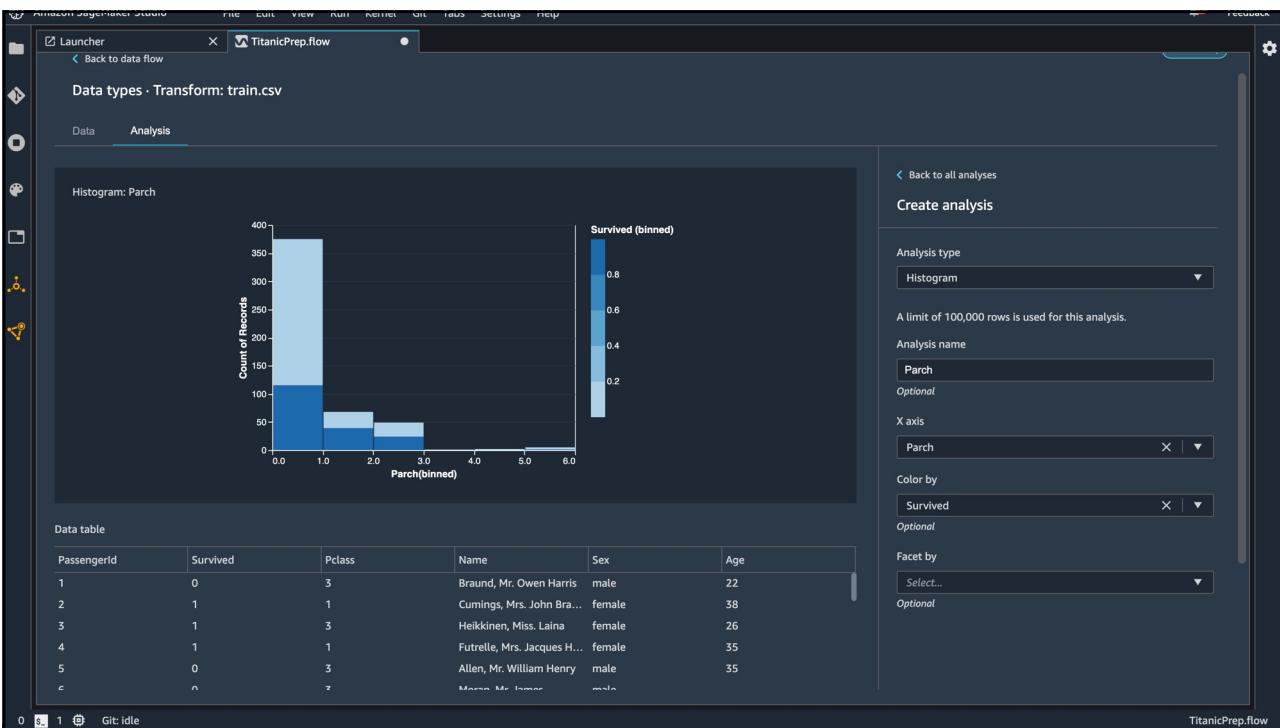
46



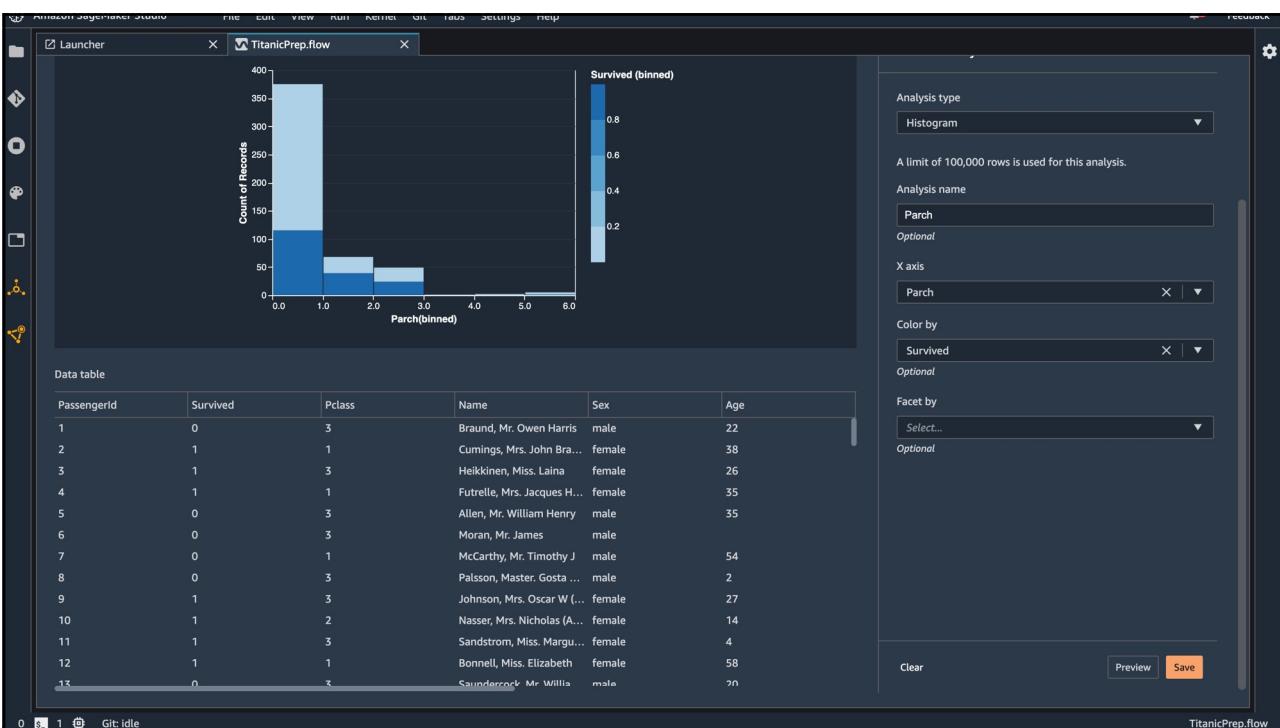
47



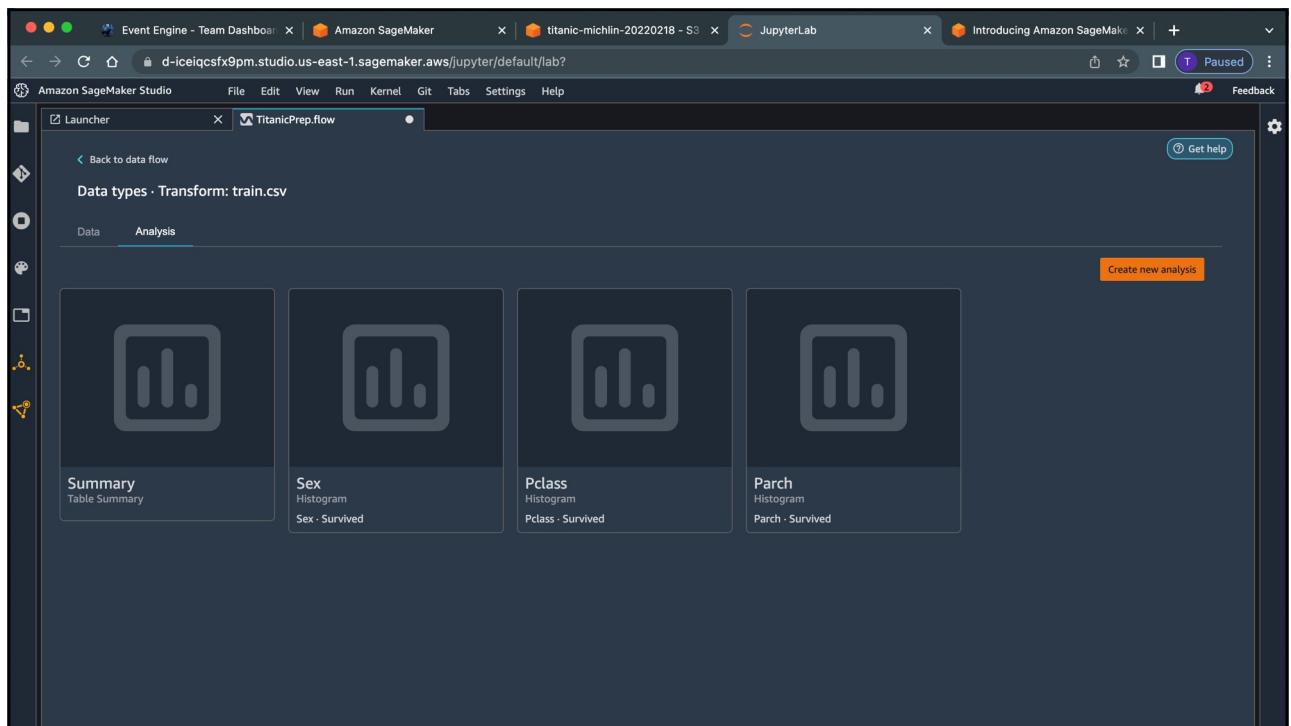
48



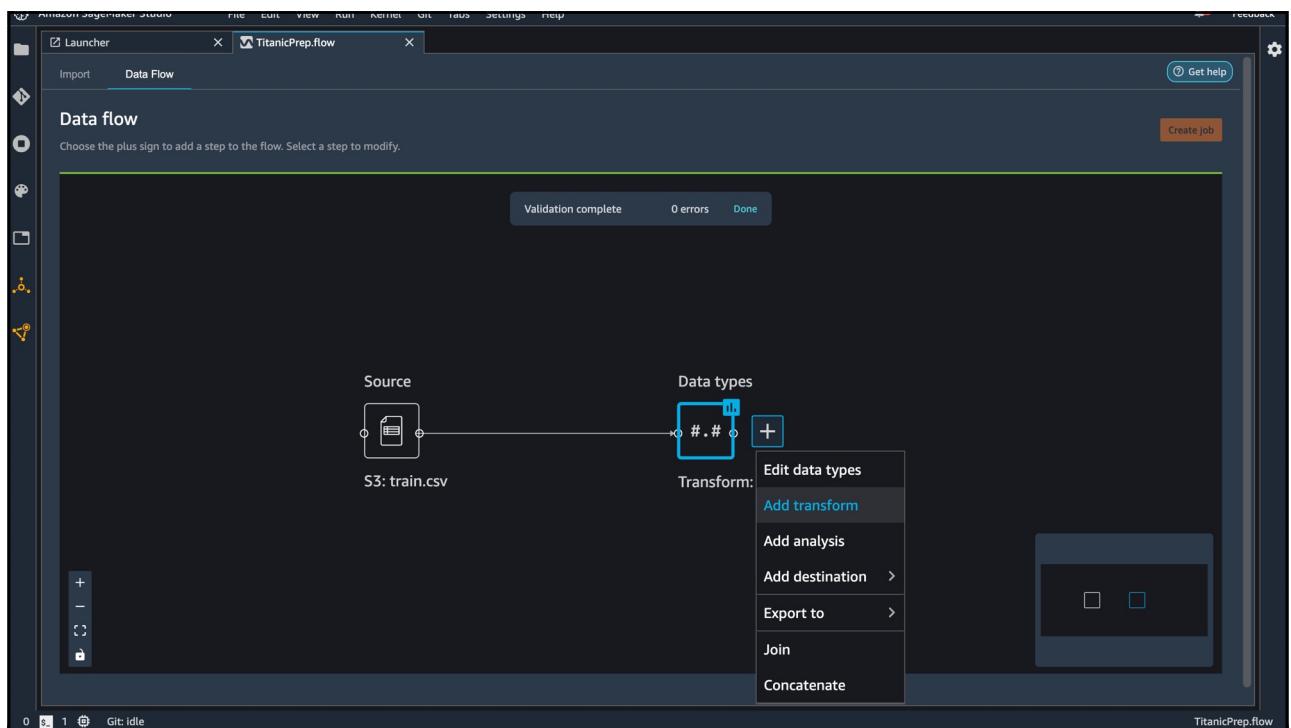
49



50



51



52

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. 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 Planké, Mrs. Juli...	female	31
20	1	2	Masseyman, Mrs. Fatima	female	
21	0	3	Funnebo, Mr. Joseph I.	male	35

TRANSFORMS

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

53

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. 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 Planké, Mrs. Juli...	female	31
20	1	2	Masseyman, Mrs. Fatima	female	
21	0	3	Funnebo, Mr. Joseph I.	male	35

ADD TRANSFORM

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

54

Amazon SageMaker Studio

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. 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	Masseyman, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

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

Preview Add

55

Amazon SageMaker Studio

TitanicPrep.flow

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	
113893	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

Preview Add

56

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher

TitanicPrep.flow

Impute - Transform: train.csv

Data Analysis

Step 3. Impute

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

EXPORT DATA

TRANSFORMS

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

57

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher

TitanicPrep.flow

Impute - Transform: train.csv

Data Analysis

Step 3. Impute

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

ADD TRANSFORM

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

58

Amazon SageMaker Studio

TitanicPrep.flow

Impute - Transform: train.csv

Data Analysis

Step 3. Impute

PassengerId (long)	Survived (long)	Pclass (long)	Name (string)	Sex (string)	Age (long)
1	0	3	Braund, Mr. Owen Harris	male	22
2	1	1	Cumings, Mrs. John Bra... ...	female	38
3	1	3	Heikkinen, Miss. Laina	female	26
4	1	1	Futrelle, Mrs. Jacques H...	female	35
5	0	3	Allen, Mr. William Henry	male	35
6	0	3	Moran, Mr. James	male	
7	0	1	McCarthy, Mr. Timothy J	male	54
8	0	3	Palsson, Master. Gosta ...	male	2
9	1	3	Johnson, Mrs. Oscar W (... ...	female	27
10	1	2	Nasser, Mrs. Nicholas (A... ...	female	14
11	1	3	Sandstrom, Miss. Margu... ...	female	4
12	1	1	Bonnell, Miss. Elizabeth	female	58
13	0	3	Saundercock, Mr. Willia... ...	male	20
14	0	3	Andersson, Mr. Anders J... ...	male	39
15	0	3	Vestrom, Miss. 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	Masseyman, Mrs. Fatima	female	
21	0	2	Fynney, Mr. Joseph J	male	35

Export data

MANAGE COLUMNS

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

Transform *Drop column*

Columns to drop

PassengerId Name Cabin Ticket

Clear Preview Add

59

Amazon SageMaker Studio

TitanicPrep.flow

Impute - Transform: train.csv

Data Analysis

Previewing: Manage columns

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

Export data

MANAGE COLUMNS

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

Transform *Drop column*

Columns to drop

PassengerId Name Cabin Ticket

Clear Preview Add

60

Amazon SageMaker Studio

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
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
0	3	female	31	1	0
1	3	female	0	0	0

TRANSFORMS

- + Add step
- 1. \$S Source
- 2. Data types
- 3. Impute
- 4. Drop column

61

Amazon SageMaker Studio

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
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
0	3	female	31	1	0

ADD TRANSFORM

- Custom transform
Use PySpark, Pandas, or PySpark (SQL) to define custom transformations...
- Balance data
Balance the data for binary classification problems using random oversa...
- 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. Lear...
- Featureize date/time
Encode date/time values to numeric and vector representations. Lear ...
- Featureize text
Generate vector representations from natural language text. Learn more...
- Format string
Clean and prepare strings using standard string formatting operations. I...
- Group by
Add an aggregated column after group by as a new column.
- Handle missing

62

Amazon SageMaker Studio

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	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
0	3	female	31	1	0

CUSTOM PYSPARK

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

Python (PySpark)

Python (Pandas)

SQL (PySpark SQL)

63

Amazon SageMaker Studio

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
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
0	3	female	31	1	0

CUSTOM PANDAS

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

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

Python (Pandas)

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

Clear Preview Add

64

Drop column · Transform: train.csv

Data Analysis

Reviewing: Python (Pandas)

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

CUSTOM PANDAS

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

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

Python (Pandas)

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

Clear Preview Add

65

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

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

TRANSFORMS

+ Add step

- 1. \$3 Source
- 2. Data types
- 3. Impute
- 4. Drop column
- 5. Custom Pandas

66

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

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

Export data

ADD TRANSFORM

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

67

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

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

ENCODE CATEGORICAL

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

Transform

- Ordinal encode
- One-hot encode
- Similarity encode

Optional

Invalid handling strategy

Replace with NaN

Clear Preview Add

68

Custom Pandas · Transform: train.csv

Data Analysis

Step 5. Custom Pandas

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

ENCODE CATEGORICAL

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

Transform One-hot encode

Input columns Pclass

Input already ordinal encoded

Invalid handling strategy Skip

Drop last

Output style Columns

Output column

Optional

Preview Add

69

Custom Pandas · Transform: train.csv

Data Analysis

Previewing: Encode categorical

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

ENCODE CATEGORICAL

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

Transform One-hot encode

Input columns Pclass

Input already ordinal encoded

Invalid handling strategy Skip

Drop last

Output style Columns

Output column

Optional

Preview Add

70

One-hot encode - Transform: train.csv

Data Analysis

Step 6. One-hot encode

Survived (long)	Sex (string)	Age (float)	SibSp (long)	Parch (long)	Fare (float)
0	male	22	1	0	7.25
1	female	38	1	0	71.2833
1	female	26	0	0	7.925
1	female	35	1	0	53.1
0	male	35	0	0	8.05
0	male	0	0	0	8.4583
0	male	54	0	0	51.8625
0	male	2	3	1	21.075
1	female	27	0	2	11.1533
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	0	13
0	female	31	1	0	18
1	female	0	0	0	7.225

ENCODE CATEGORICAL

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

Output style [?](#)

Columns

Output column [?](#)

Optional

Preview Add

71

One-hot encode - Transform: train.csv

Data Analysis

Previewing: Encode categorical

t_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	

ENCODE CATEGORICAL

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

Output style [?](#)

Columns

Output column [?](#)

Optional

Preview Add

72

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 'Launcher' and 'TitanicPrep.flow'. Below the tabs, it says 'One-hot encode - Transform: train.csv'. There are two tabs: 'Data' (selected) and 'Analysis'. Under 'Data', there's a table titled 'Step 7. One-hot encode' with columns: Survived (long), Age (float), SibSp (long), Parch (long), Fare (float), and Embarked (string). The table contains 12 rows of data. To the right of the table is a 'TRANSFORMS' panel with a list of steps: 1. S3 Source, 2. Data types, 3. Impute, 4. Drop column, 5. Custom Pandas, 6. One-hot encode, and 7. One-hot encode. A button '+ Add step' is at the top of the panel.

73

This screenshot is similar to the previous one, showing the Amazon SageMaker Studio interface with the 'TitanicPrep.flow' step selected. The 'Data' tab is active, displaying the same table of 12 rows. However, a large 'ADD TRANSFORM' dialog box is overlaid on the right side of the screen. The dialog has a header 'ADD TRANSFORM' and a list of transformation options: 'Custom transform', 'Balance data', 'Custom formula', 'Encode categorical', 'Featetrize date/time', 'Featetrize text', 'Format string', 'Group by', and 'Handle missing'. Each option has a brief description and a 'Learn more' link.

74

37

One-hot encode - Transform: train.csv

Step 7. One-hot encode

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

75

One-hot encode - Transform: train.csv

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		1	0		
0		0	1		

76

Amazon SageMaker Studio

Launcher

TitanicPrep.flow

One-hot encode - Transform: train.csv

Data Analysis

Step 8. One-hot encode

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

TRANSFORMS

- + Add step
- 1. 53 Source
- 2. Data types
- 3. Impute
- 4. Drop column
- 5. Custom Pandas
- 6. One-hot encode
- 7. One-hot encode
- 8. One-hot encode

77

Amazon SageMaker Studio

Launcher

TitanicPrep.flow

One-hot encode - Transform: train.csv

Data Analysis

Step 8. One-hot encode

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

ADD TRANSFORM

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

78

39

Event Engine - Team Dashboard | Amazon SageMaker | titanic-michelin-20220218 - S3 | JupyterLab | Introducing Amazon SageMak

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

Launcher | TitanicPrep.flow | Back to data flow

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		
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, SibSp, Parch, Pclass_2, Sex_female, Embarked_Q

Clear Preview Add

79

Event Engine - Team Dashboard | Amazon SageMaker | titanic-michelin-20220218 - S3 | JupyterLab | Introducing Amazon SageMak

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

Launcher | TitanicPrep.flow | Back to data flow

One-hot encode - Transform: train.csv

Data Analysis

Previewing: Manage columns

Survived (long)	Fare (float)	Age_Imputed (float)	With_Family (long)	Pclass_3 (float)	Pclass_1 (float)
0	7.25	22	1	1	0
1	71.2833	38	1	0	1
1	7.925	26	0	1	0
1	53.1	35	1	0	1
0	8.05	35	0	1	0
0	8.4583	28	0	1	0
0	51.8625	54	0	0	1
0	21.075	2	1	1	0
1	11.1333	27	1	1	0
1	30.0708	14	1	0	0
1	16.7	4	1	1	0
1	26.55	58	0	0	1
0	8.05	20	0	1	0
0	31.275	39	1	1	0
0	7.8542	14	0	1	0
1	16	55	0	0	0
0	29.125	2	1	1	0
1	13	28	0	0	0
0	18	31	1	1	0
1	7.225	28	0	1	0
0	26	76	0	0	0

Export data

MANAGE COLUMNS

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

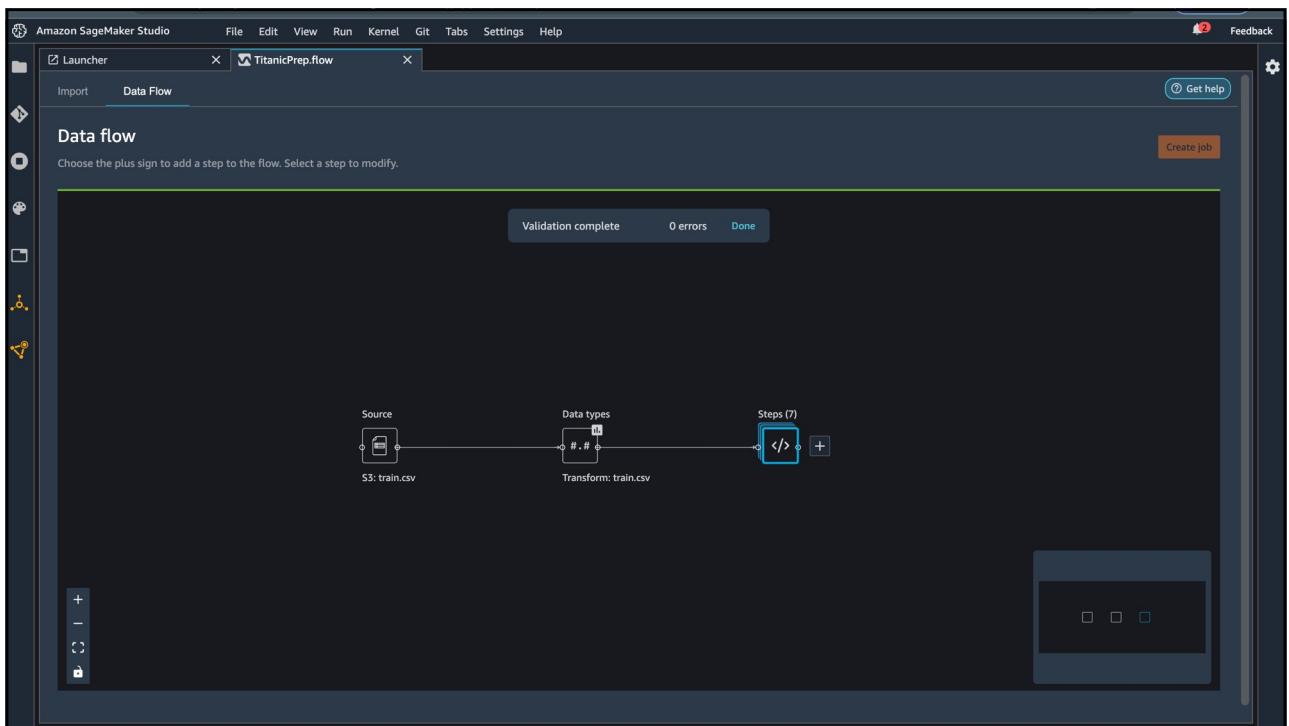
Transform *Drop column*

Columns to drop

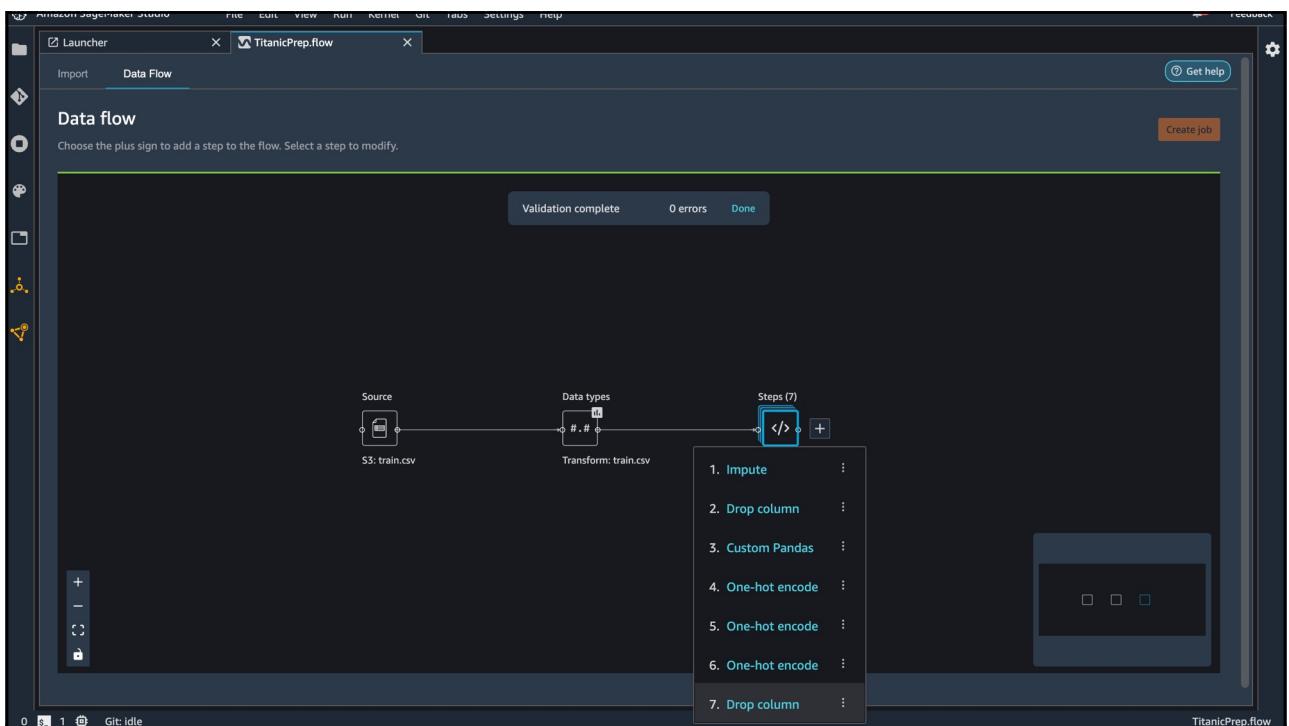
Age, SibSp, Parch, Pclass_2, Sex_female, Embarked_Q

Clear Preview Add

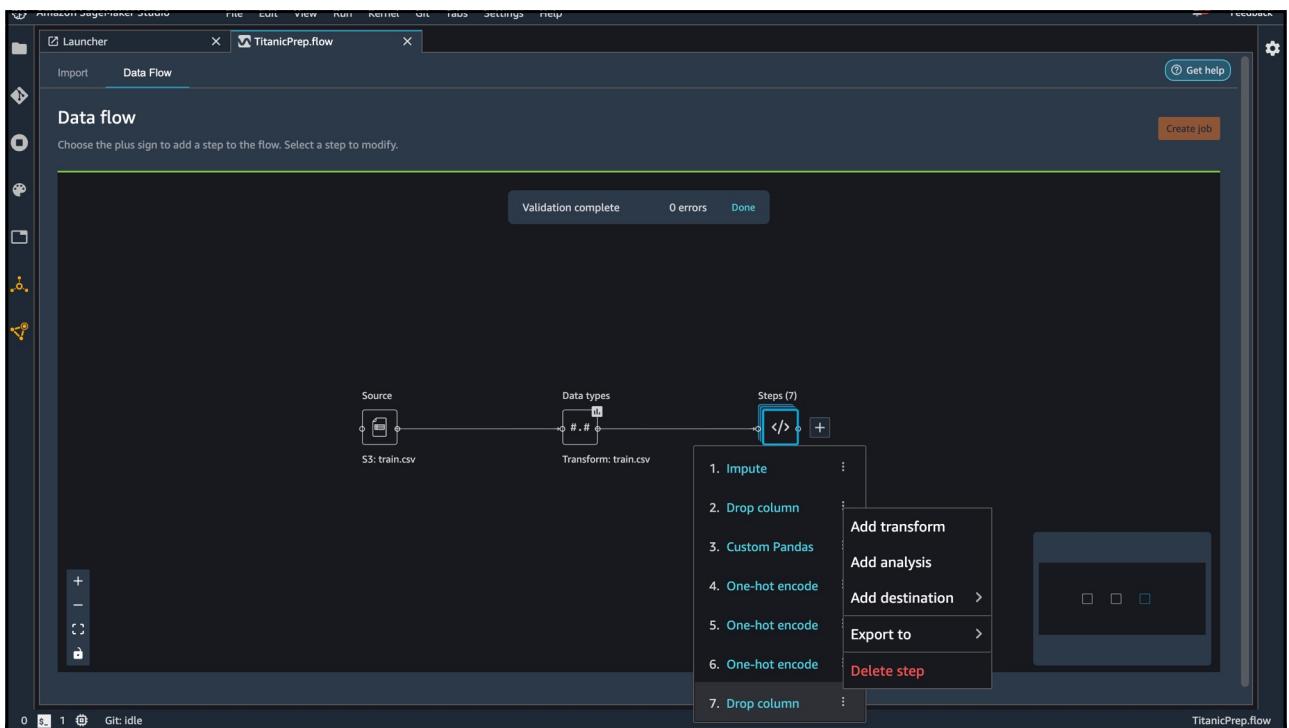
80



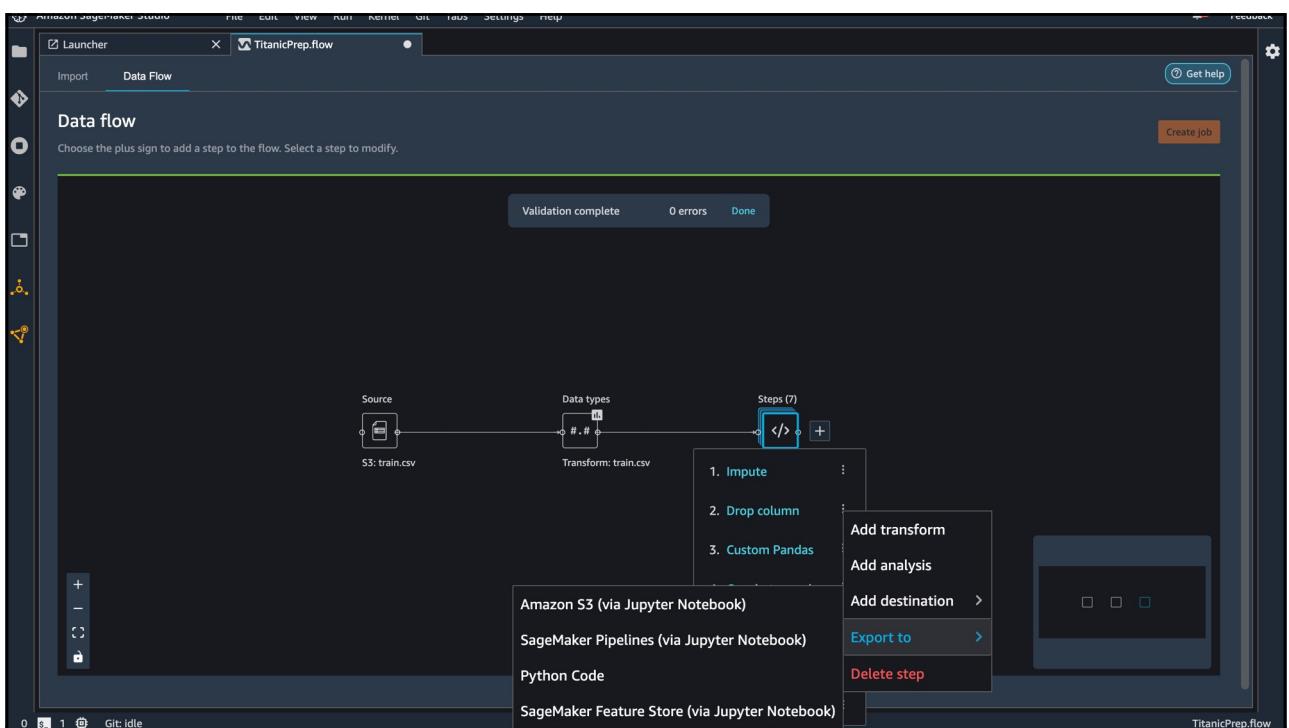
81



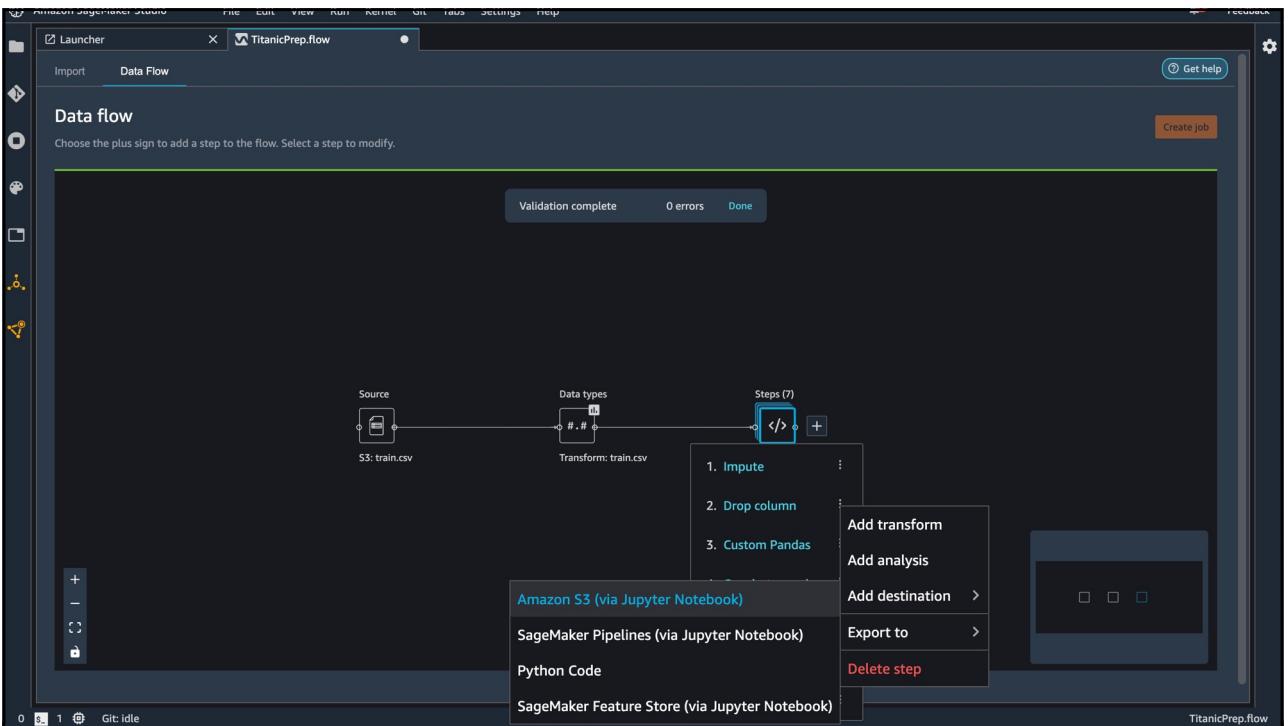
82



83



84



85

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

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

86

The screenshot shows the AWS S3 Management Console dashboard. On the left, there's a sidebar with links for Buckets, Storage Lens, Feature spotlight, and AWS Marketplace for S3. The main area has a heading "Amazon S3" and a callout "Read the S3 resources page for documentation and technical content." Below this is an "Account snapshot" section with a link to "View Storage Lens dashboard". The "Buckets" section shows three buckets: "sagemaker-studio-834120567544-", "sagemaker-us-east-1-834120567544", and "titanic-michelin-20220218".

Name	AWS Region	Access	Creation date
sagemaker-studio-834120567544-	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)

87

This screenshot shows the "Objects" tab for the bucket "sagemaker-us-east-1-834120567544". The sidebar is identical to the previous screenshot. The main area shows three objects: "Canvas/", "data_wrangler_flows/", and "export-flow-16-15-04-15-fd0b1758/".

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

88

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 path: Amazon S3 > sagemaker-us-east-1-834120567544 > export-flow-16-15-04-15-fd0b1758/. A sub-breadcrumb indicates we're in the 'output/' folder. There are two tabs: 'Objects' (selected) and 'Properties'. Below the tabs is a section titled 'Objects (1)'. It contains a table with one item:

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

At the top right of the main area, there's a 'Copy S3 URI' button.

89

This screenshot is similar to the previous one but shows the 'output/' folder expanded. The breadcrumb path now includes 'output/'. The table in the 'Objects (1)' section shows:

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

90

The screenshot shows the AWS S3 console interface. The left sidebar is titled "Amazon S3" and includes sections for Buckets, Storage Lens, Feature spotlight, and AWS Marketplace for S3. The main content 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/. Below this, a sub-path 7f06826f-b540-4996-b6d5-e1a8eb70f110/ is displayed. A "Copy S3 URI" button is located in the top right of this sub-path area. The "Objects" tab is selected, showing a table with one item:

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

91

This screenshot is identical to the one above, showing the AWS S3 console with the same navigation paths and the same single folder entry in the object list.

92

Amazon S3 > default/

Objects (1)

part-00000-b8bd0b1c-acdc-439d-bf2c-837fa0f2faee-c000.csv

CSV | Type: CSV | Size: 30.5 KB | Last modified: February 16, 2022, 23:10:52 (UTC+08:00) | Storage class: Standard

93

	A	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	0	0	1	0	1	0											
3	1	71.2833	38	1	0	1	0	0	0	0	1	0									
4	1	7.925	26	0	1	0	0	0	1	0	0	0									
5	1	53.1	35	1	0	1	0	0	1	0	1	0									
6	0	8.05	35	0	1	0	0	1	1	1	0										
7	0	8.4583	28	0	1	0	0	1	0	0	0	0									
8	0	51.8625	54	0	0	1	1	1	1	1	0										
9	0	21.075	2	1	1	0	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	0	1	0											
14	0	8.05	20	0	1	0	0	1	1	1	0										
15	0	31.275	39	1	1	0	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	0										
22	0	26	35	0	0	0	0	1	1	1	0										
23	1	13	34	0	0	0	0	1	1	1	0										
24	1	8.0292	15	0	1	0	0	0	0	0	0										
25	1	35.5	28	0	0	1	0	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	0										
29	0	263	19	1	0	1	1	1	1	0											
30	1	7.8792	28	0	1	0	0	0	0	0	0										
31	0	7.8958	28	0	1	0	0	1	1	1	0										
32	0	27.7208	40	0	0	1	1	1	0	1	0										
33	1	146.5208	28	1	0	1	0	0	0	0	1	0									
34	1	7.75	28	0	1	0	0	0	0	0	0										
35	0	10.5	66	0	0	0	0	1	1	0											
36	0	82.1708	28	1	0	1	1	1	0	1	0										
37	0	52	42	1	0	1	0	1	1	1	0										
38	1	7.2292	28	0	1	0	0	1	0	1	0										

94

Thank you!

Michael Lin

linmicht@amazon.com



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.