
AWS
re:Invent

EMB033

Zero-code data preparation with AWS Glue DataBrew

Surbhi Dangi

Senior Product Manager

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What to expect from this session

State of union

Introducing AWS Glue DataBrew

Key features

Demo

Use cases



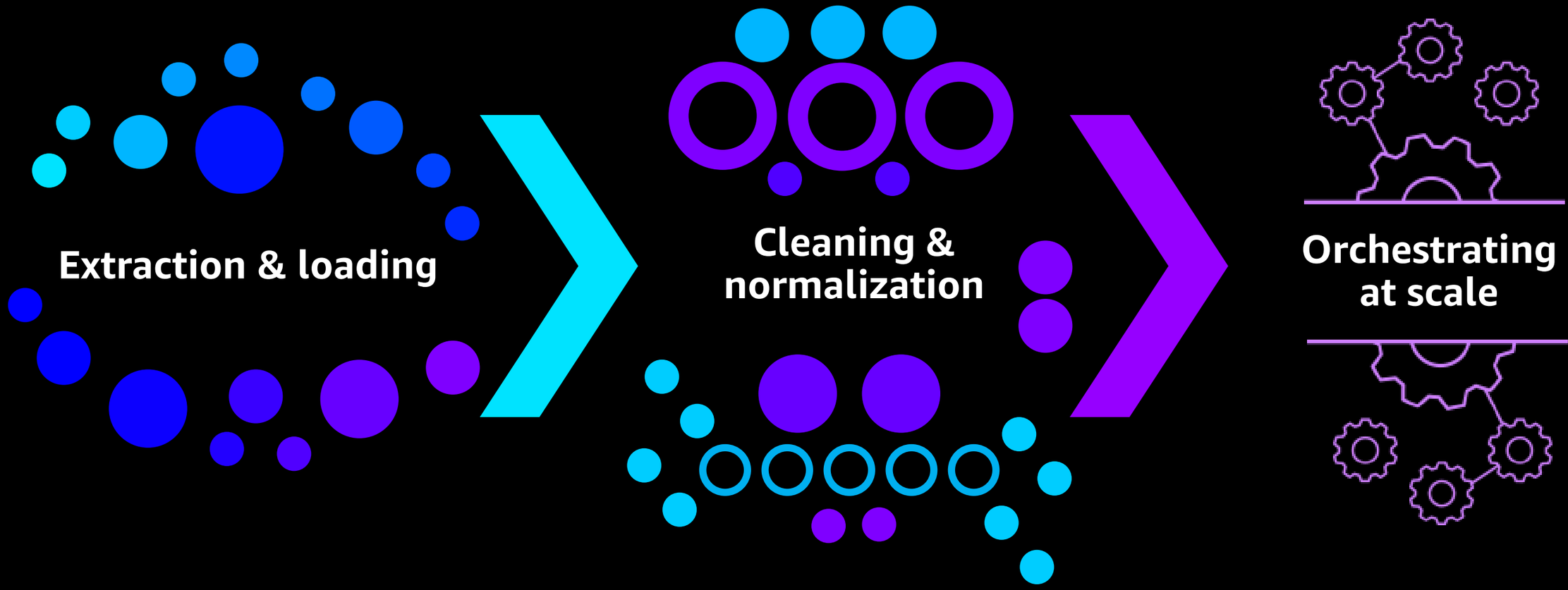
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Our teams spend too much time on the undifferentiated, repetitive, and mundane tasks associated with data preparation.”



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

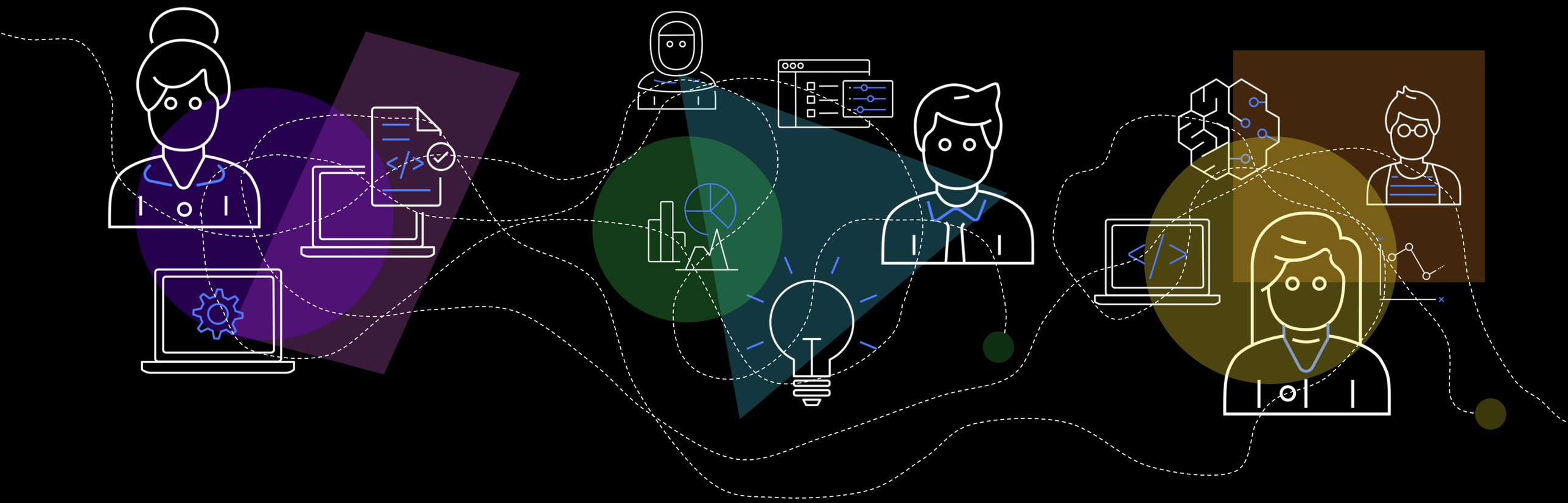
Preparing data involves several complex tasks



Needs a lot of code-based heavy-lifting to work at scale

As much as 80% of time is spent preparing data today

Data engineers | ETL developers | Data analysts | Data scientists



Needs the right tool for the right persona

Challenges with traditional data preparation

Manual

Needs a lot of code-based heavy-lifting for it to work at scale

Siloed

Often requires moving large amounts of data into silos, at times out of VPCs

Time consuming

Needs the right tools for the right persona that are integrated

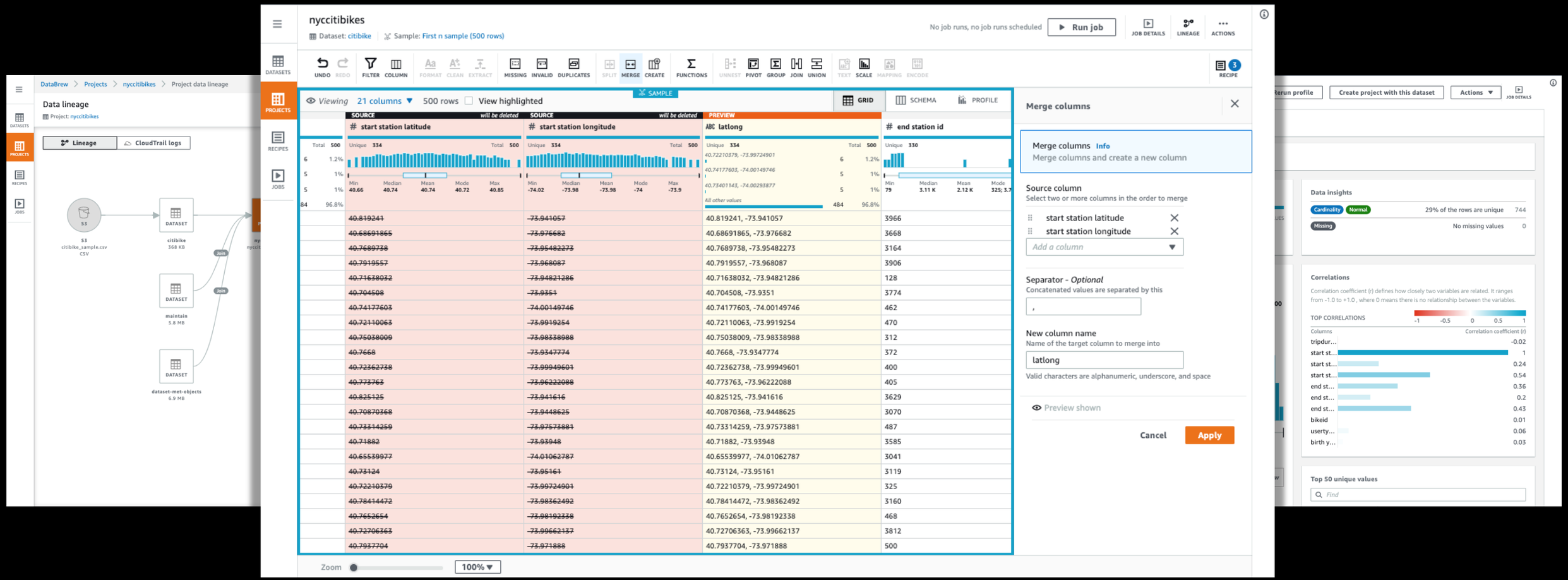
Introducing
AWS Glue DataBrew



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Clean and normalize data up to 80% faster



Built for data analysts and data scientists



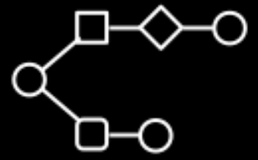
Understand data quality

Understand patterns and detect anomalies using profiles



Clean and normalize data

Over 250 built-in transformations



Visually map data lineage

Understand steps that the data has been through



Automate at scale

Save transformations and apply to new data as it comes in

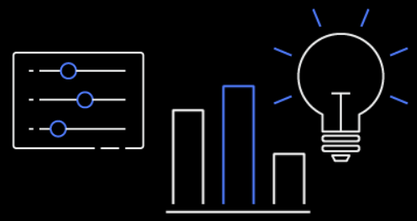
Data preparation made easy

Demo



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What we saw in the demo



Profile the data



Build a recipe



Run a job

Operationalize at scale

Schedule jobs

Use APIs/SDK

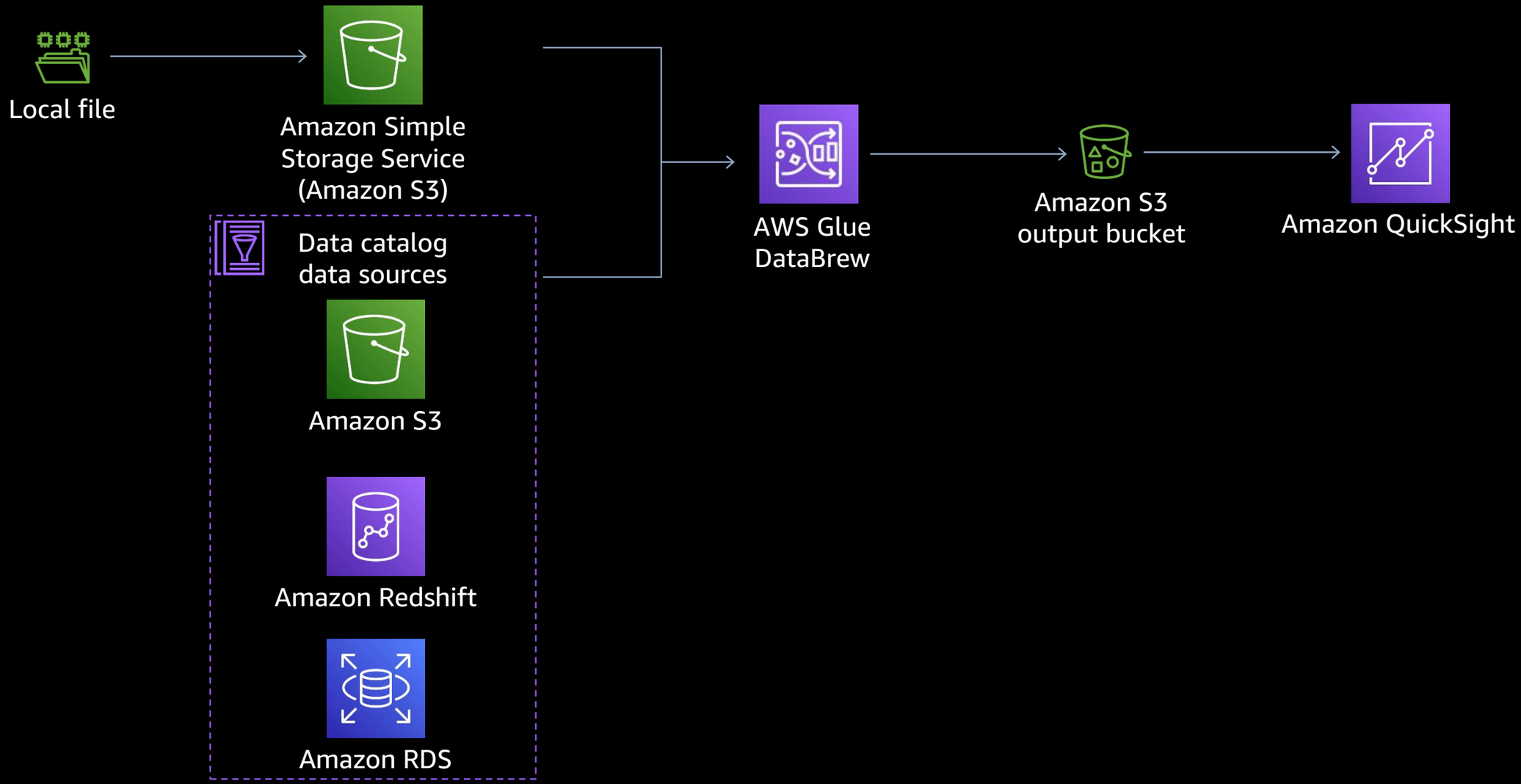
Reuse recipes

Popular use cases

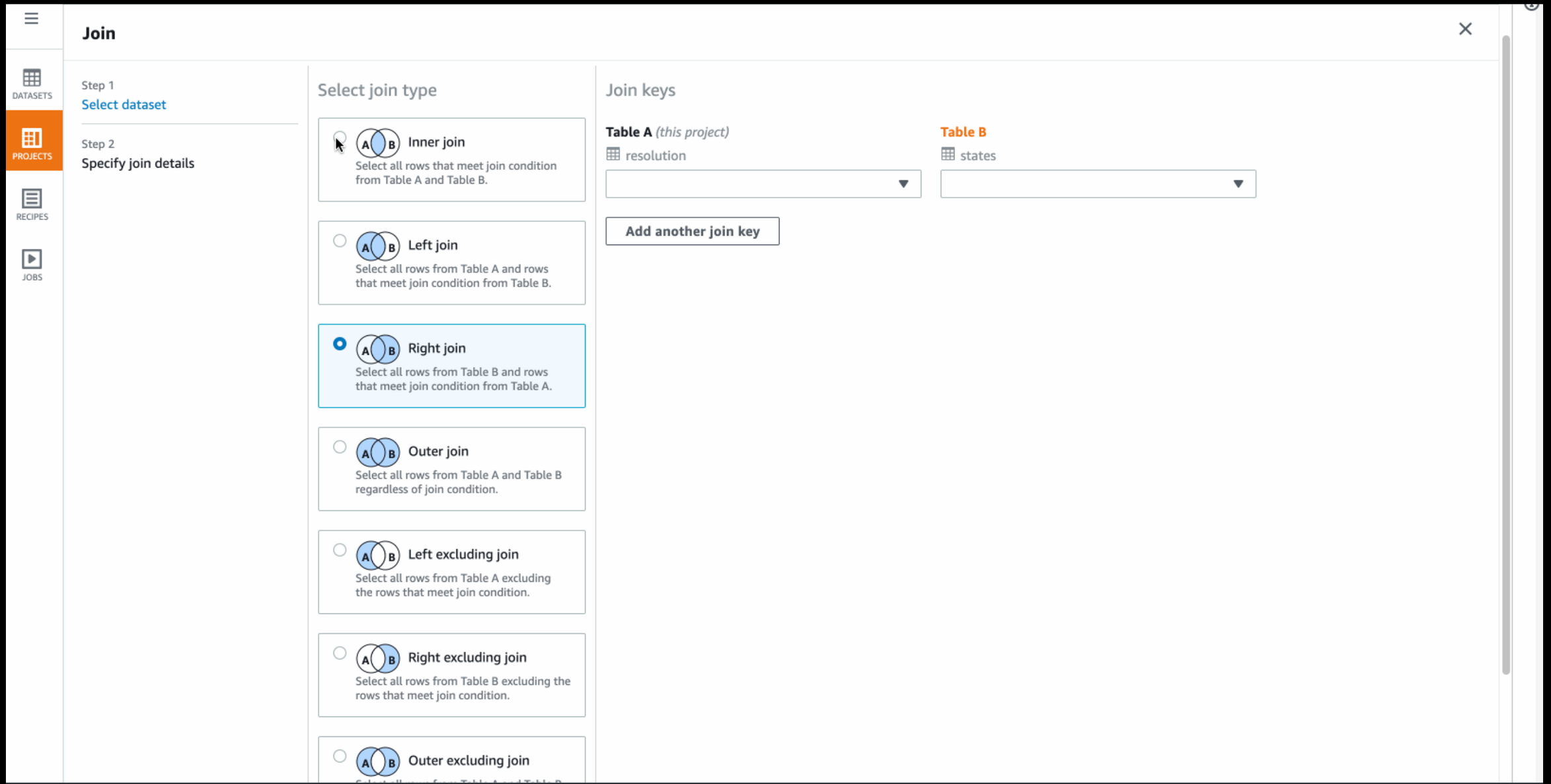


© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

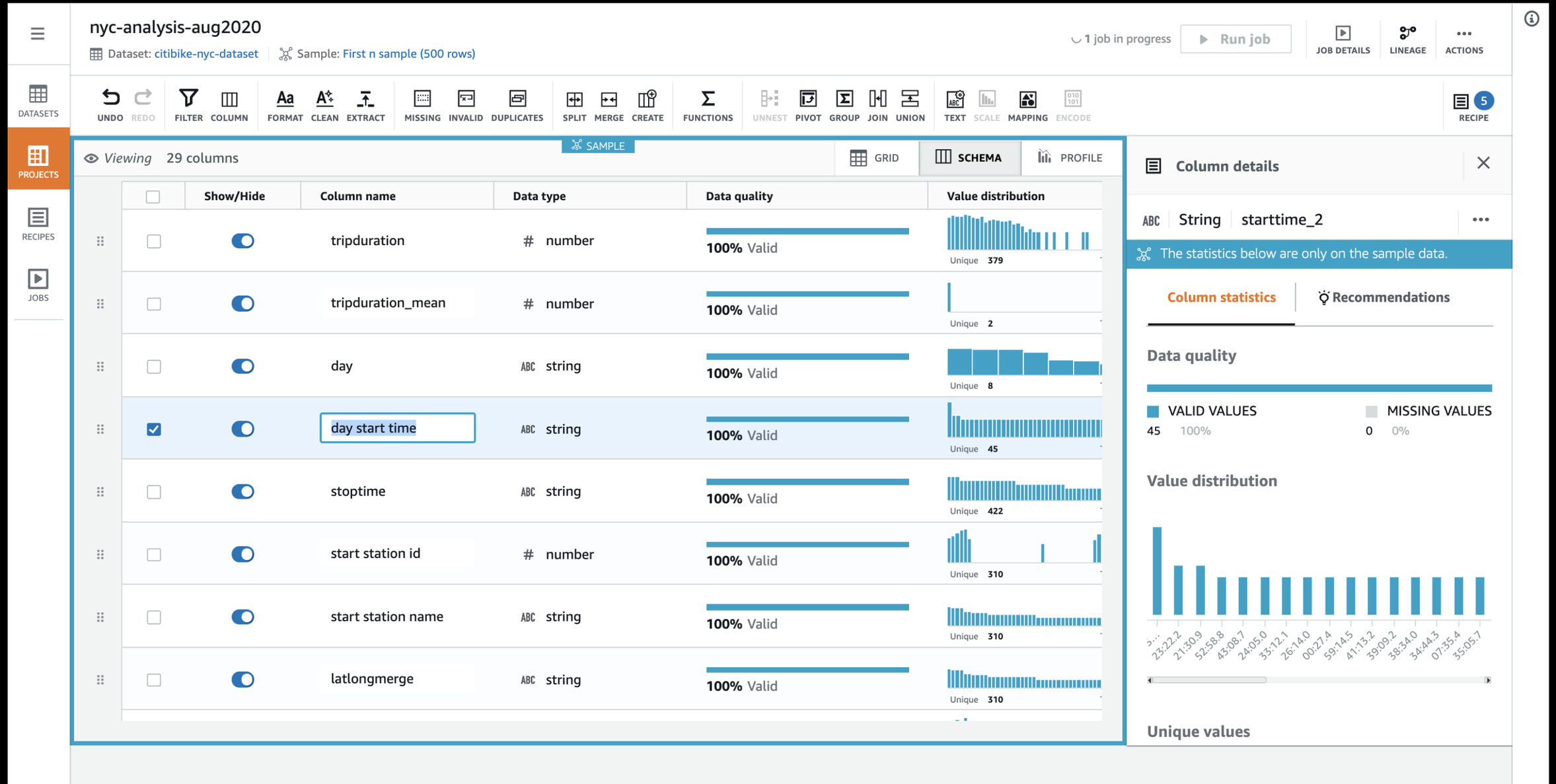
One-time data analysis for business reporting



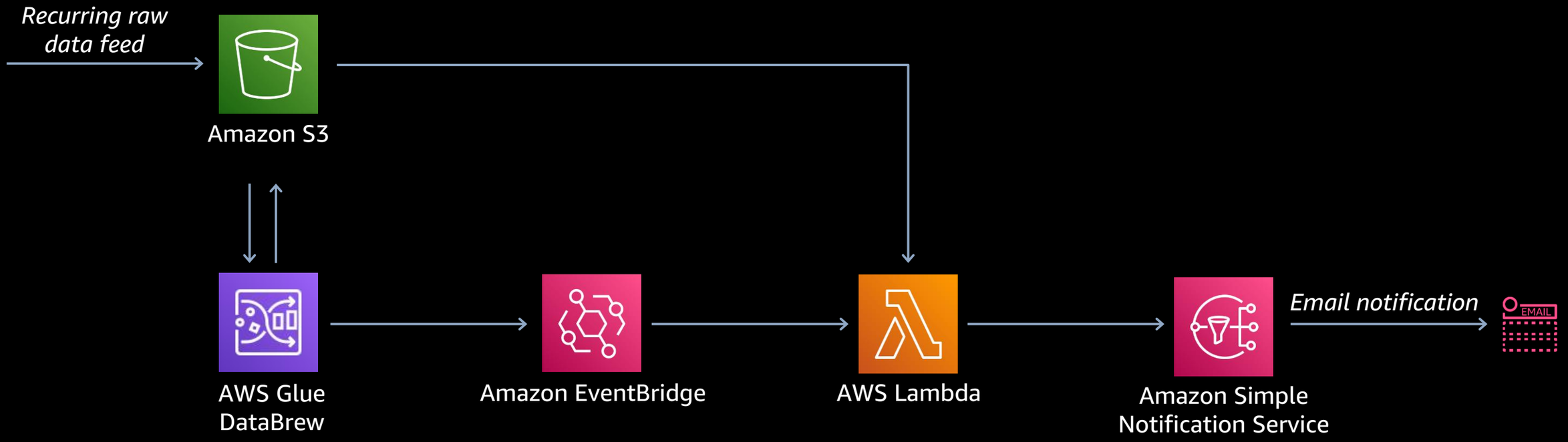
Enrich data with unions and joins



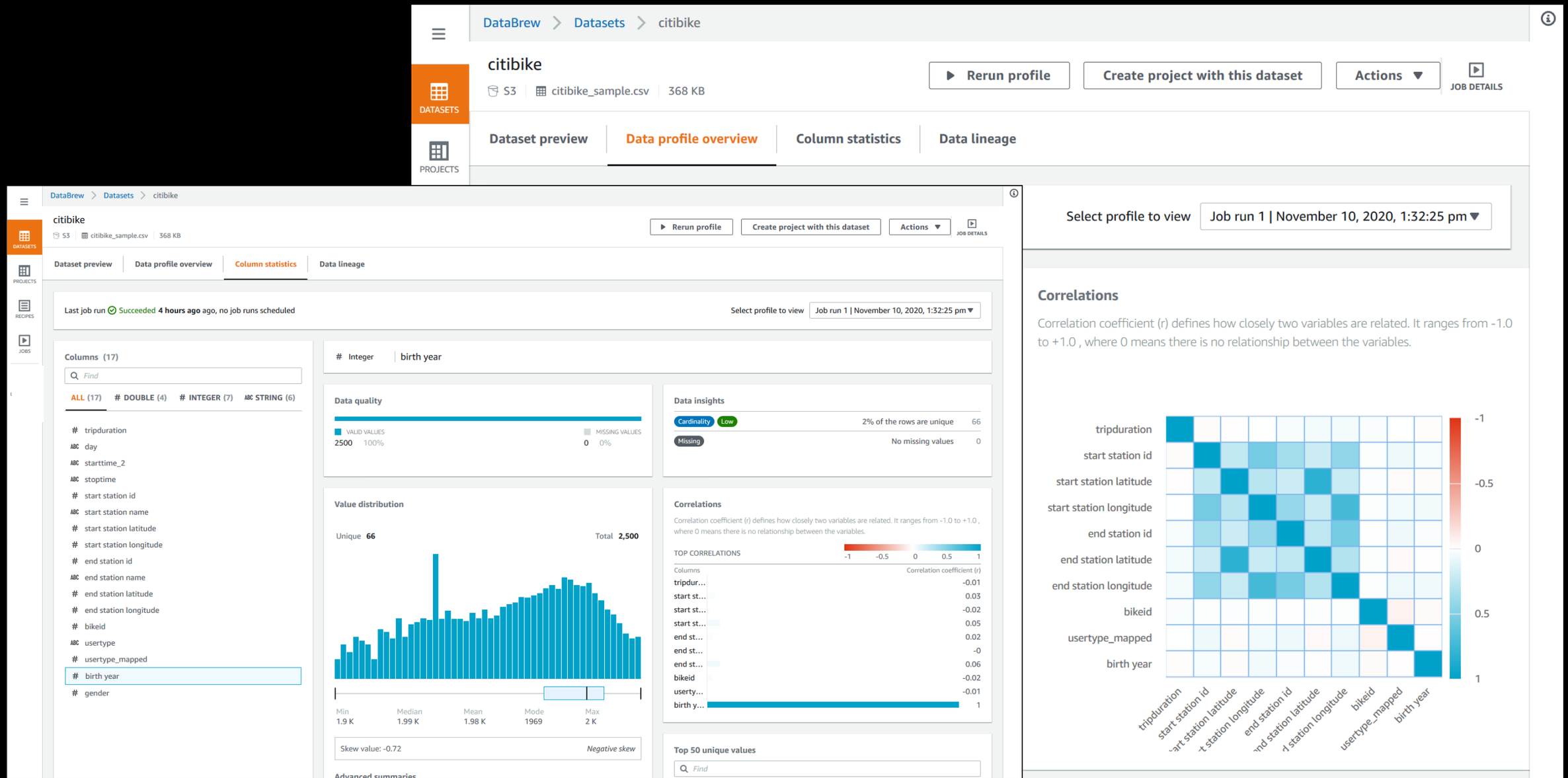
Update the schema



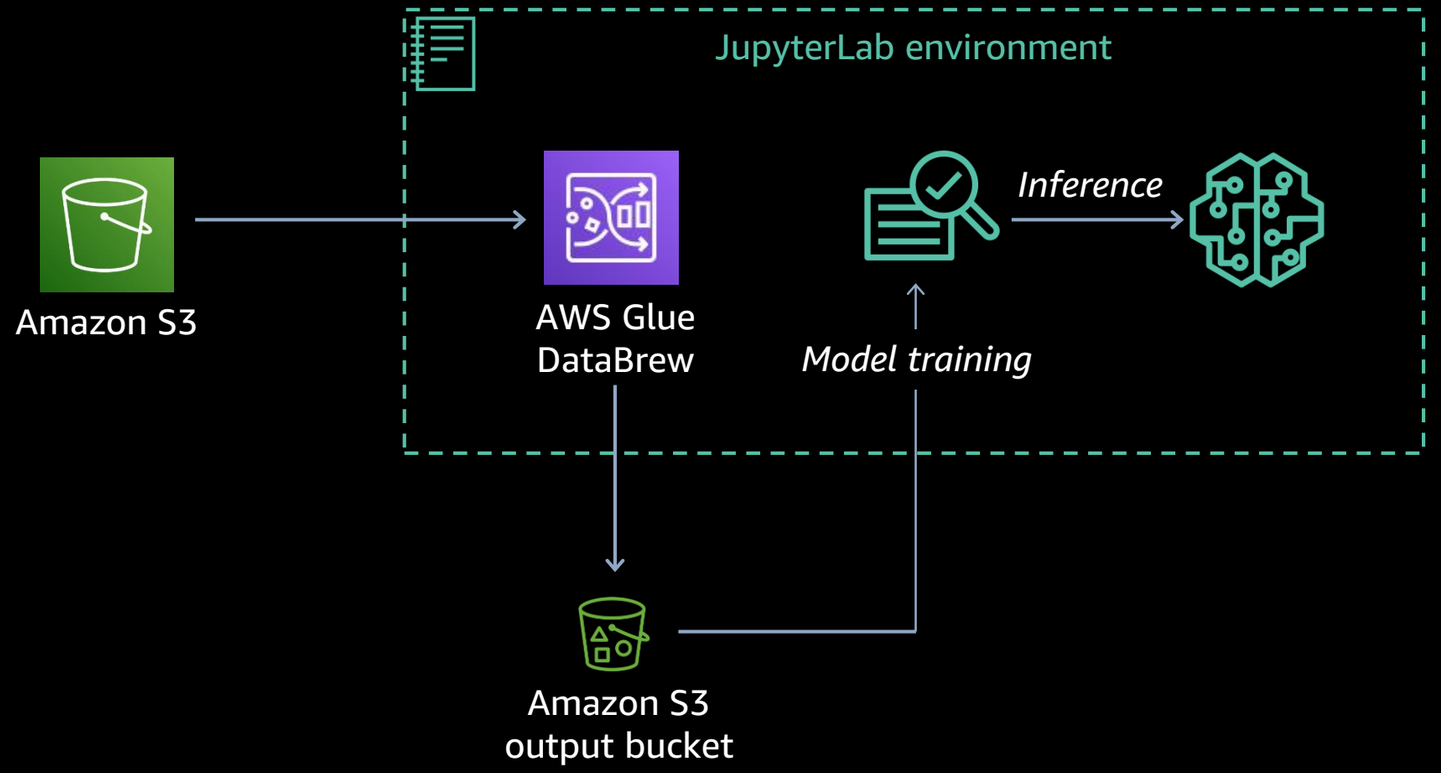
Set up data quality rules with AWS Lambda



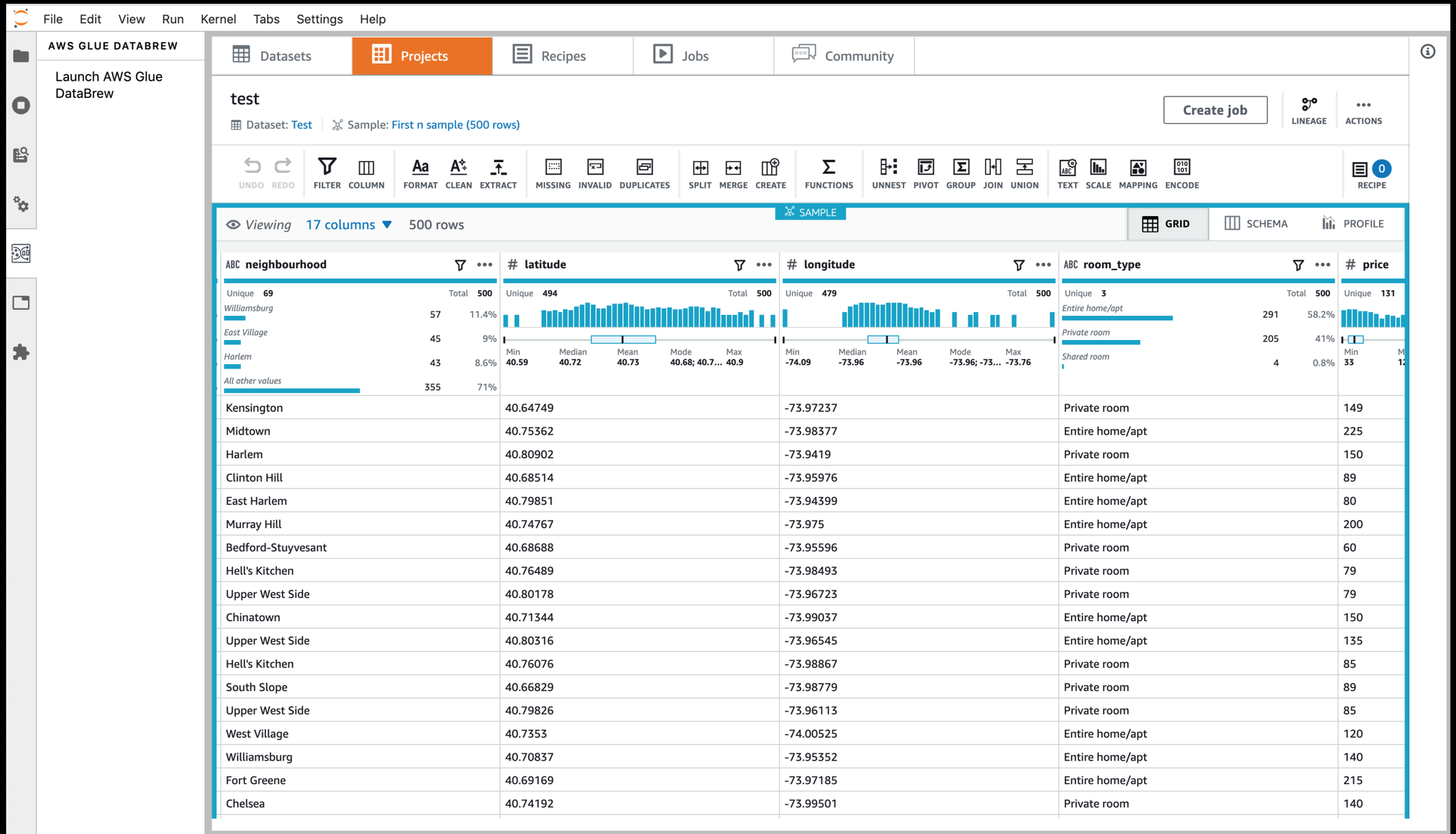
Generate data profiles



Data preprocessing for machine learning

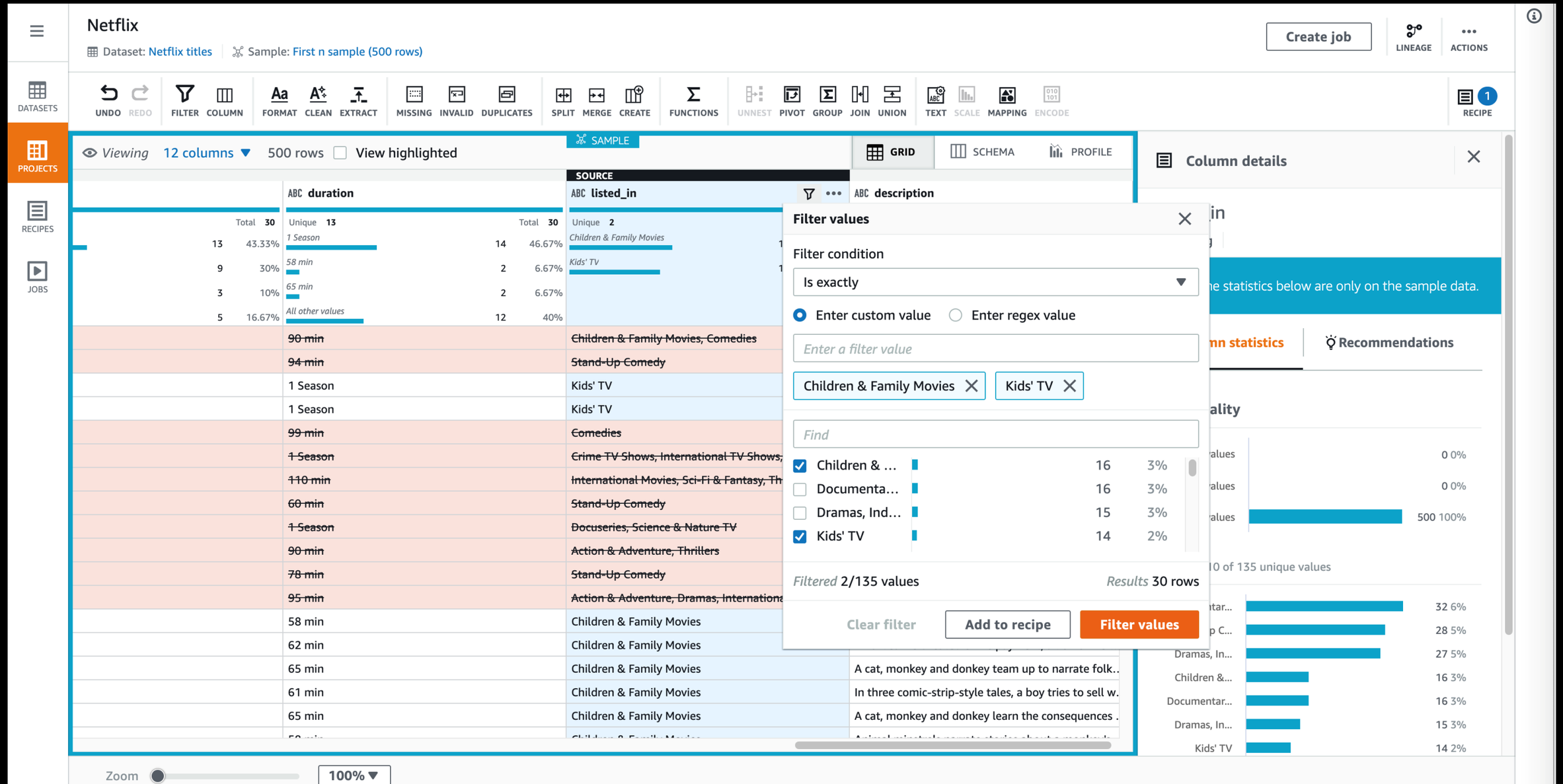


Use with Jupyter Notebooks

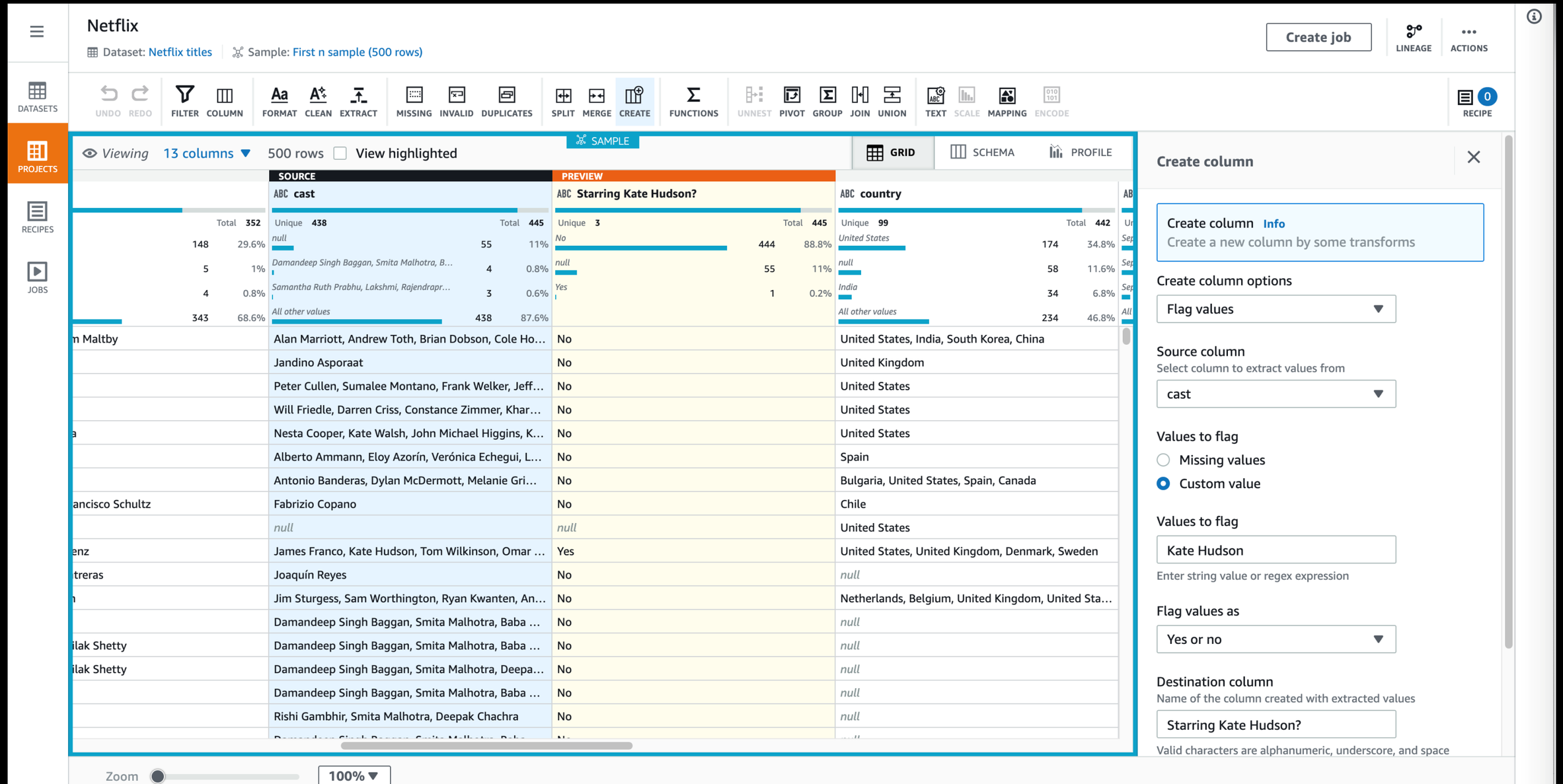


neighbourhood	latitude	longitude	room_type	price
Kensington	40.64749	-73.97237	Private room	149
Midtown	40.75362	-73.98377	Entire home/apt	225
Harlem	40.80902	-73.9419	Private room	150
Clinton Hill	40.68514	-73.95976	Entire home/apt	89
East Harlem	40.79851	-73.94399	Entire home/apt	80
Murray Hill	40.74767	-73.975	Entire home/apt	200
Bedford-Stuyvesant	40.68688	-73.95596	Private room	60
Hell's Kitchen	40.76489	-73.98493	Private room	79
Upper West Side	40.80178	-73.96723	Private room	79
Chinatown	40.71344	-73.99037	Entire home/apt	150
Upper West Side	40.80316	-73.96545	Entire home/apt	135
Hell's Kitchen	40.76076	-73.98867	Private room	85
South Slope	40.66829	-73.98779	Private room	89
Upper West Side	40.79826	-73.96113	Private room	85
West Village	40.7353	-74.00525	Entire home/apt	120
Williamsburg	40.70837	-73.95352	Entire home/apt	140
Fort Greene	40.69169	-73.97185	Entire home/apt	215
Chelsea	40.74192	-73.99501	Private room	140

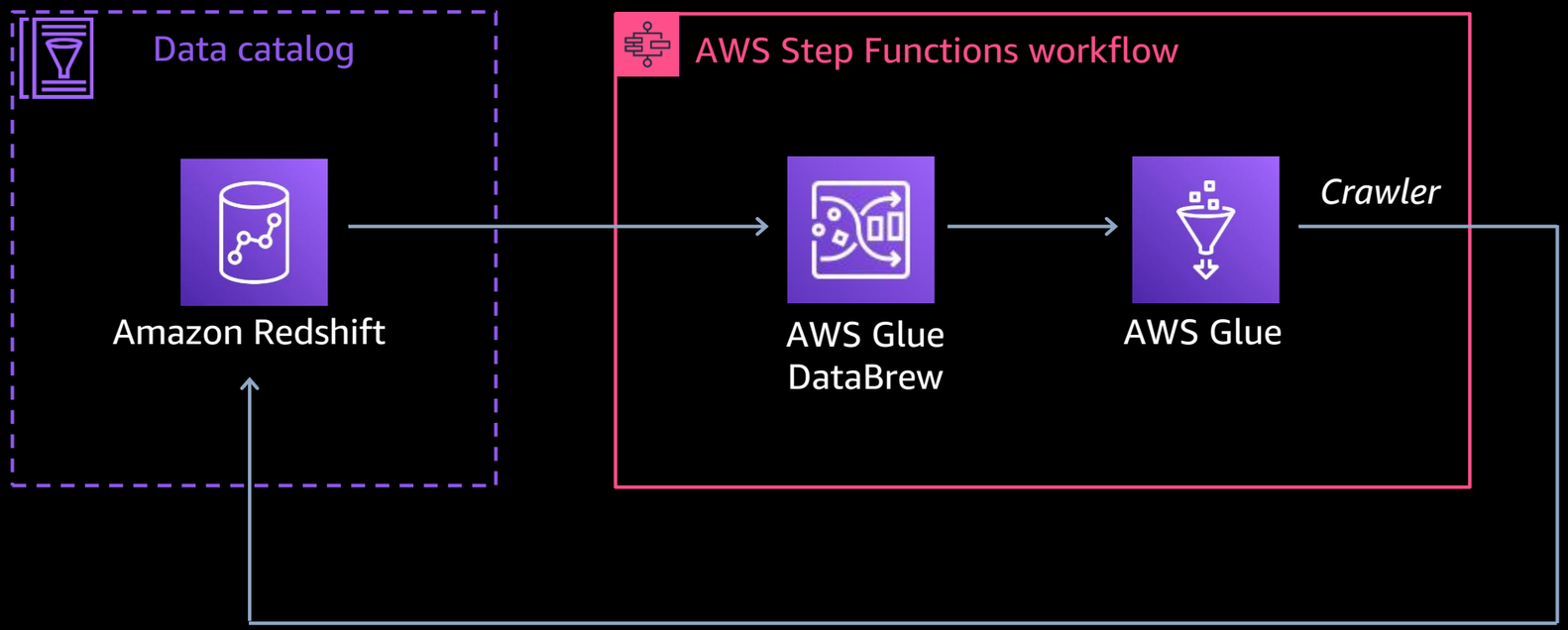
Remove outliers



Feature engineering



Orchestrating data preparation in workflows



AWS Glue DataBrew

Clean and normalize data up to 80% faster

https://aws.amazon.com/glue/features/databrew/

https://github.com/aws/aws-glue-databrew-jupyter-extension



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Surbhi Dangi

Twitter: @surbhidangi
LinkedIn: https://www.linkedin.com/in/surbhidangi/

databrew-feedback@amazon.com

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.