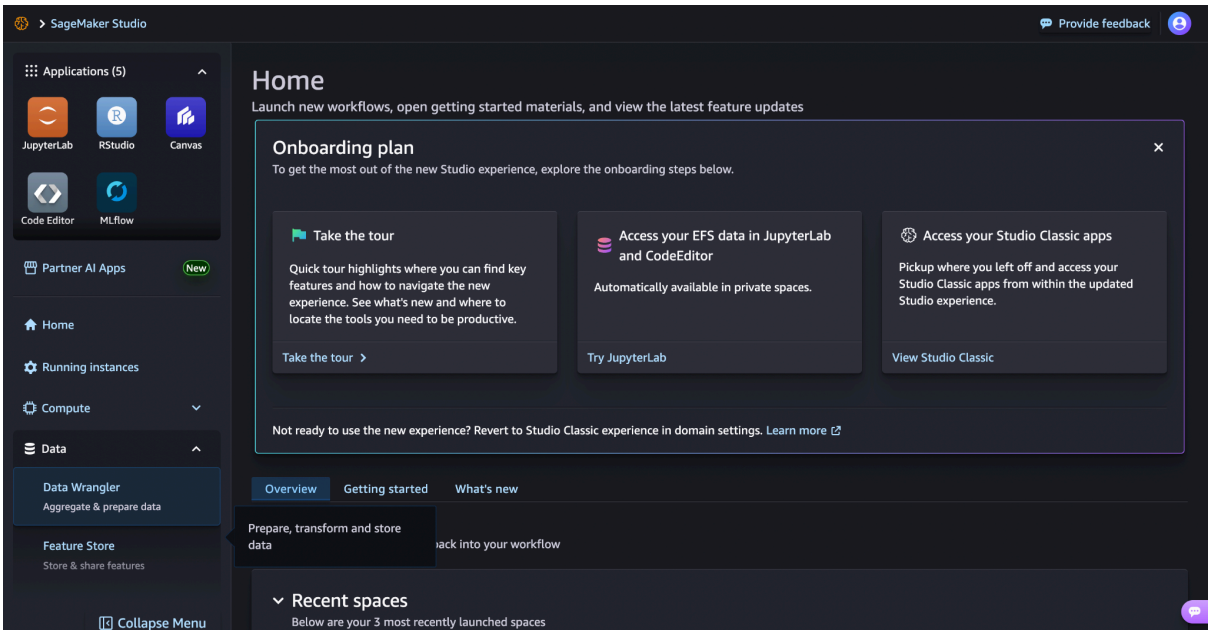
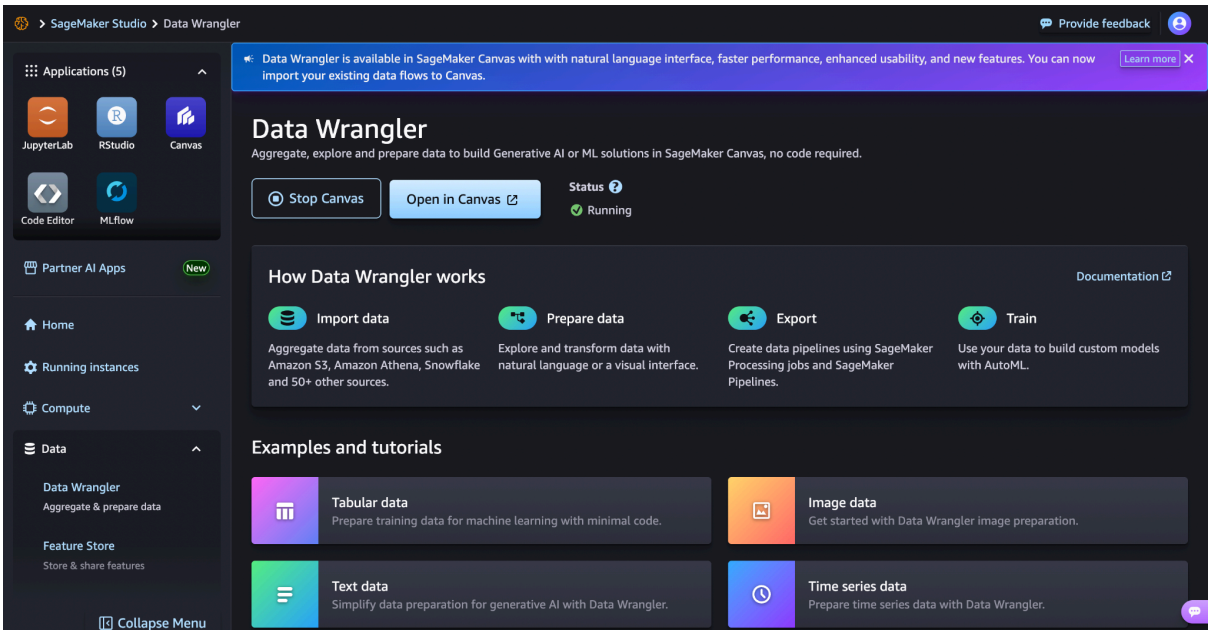
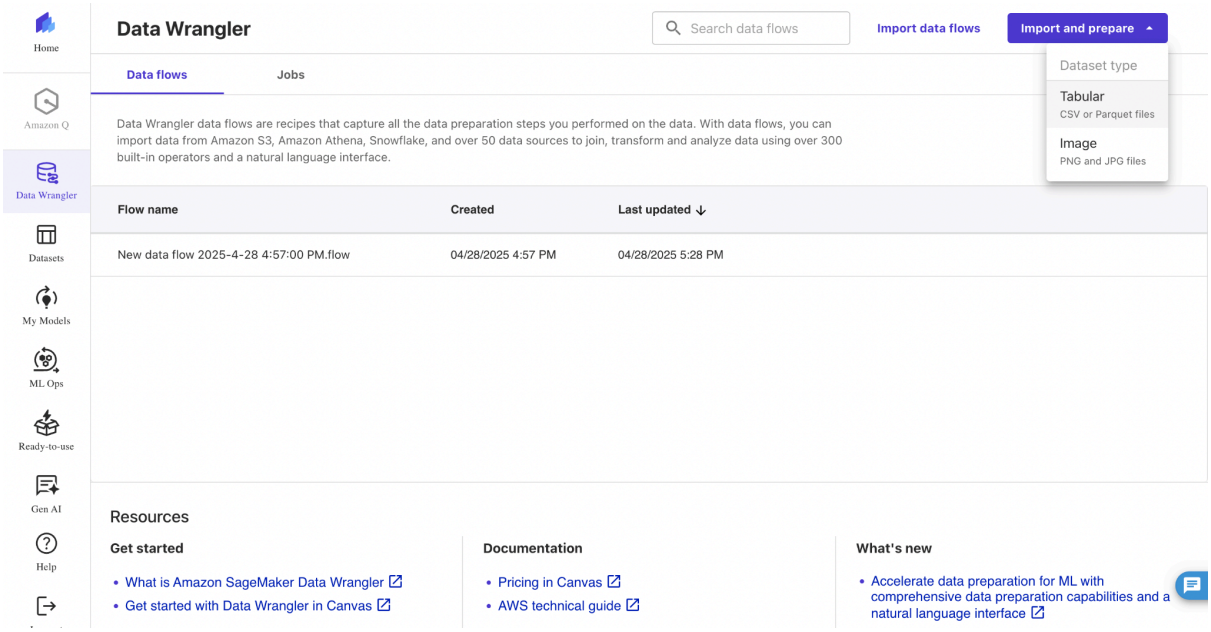
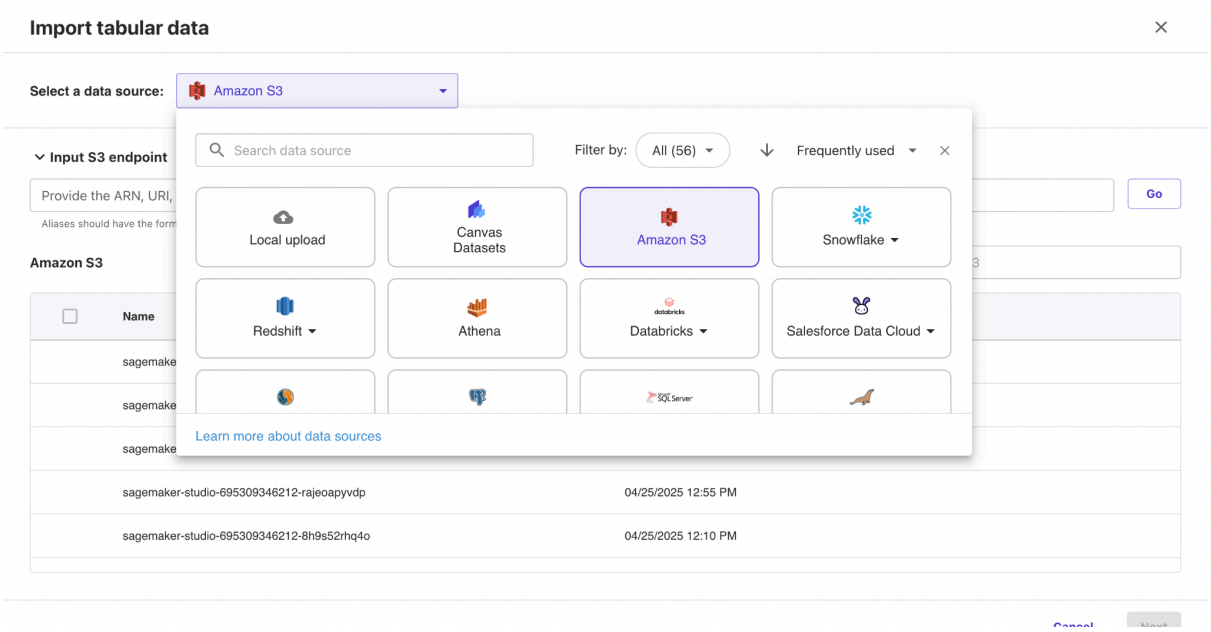
Step 0- Start Creating Flow

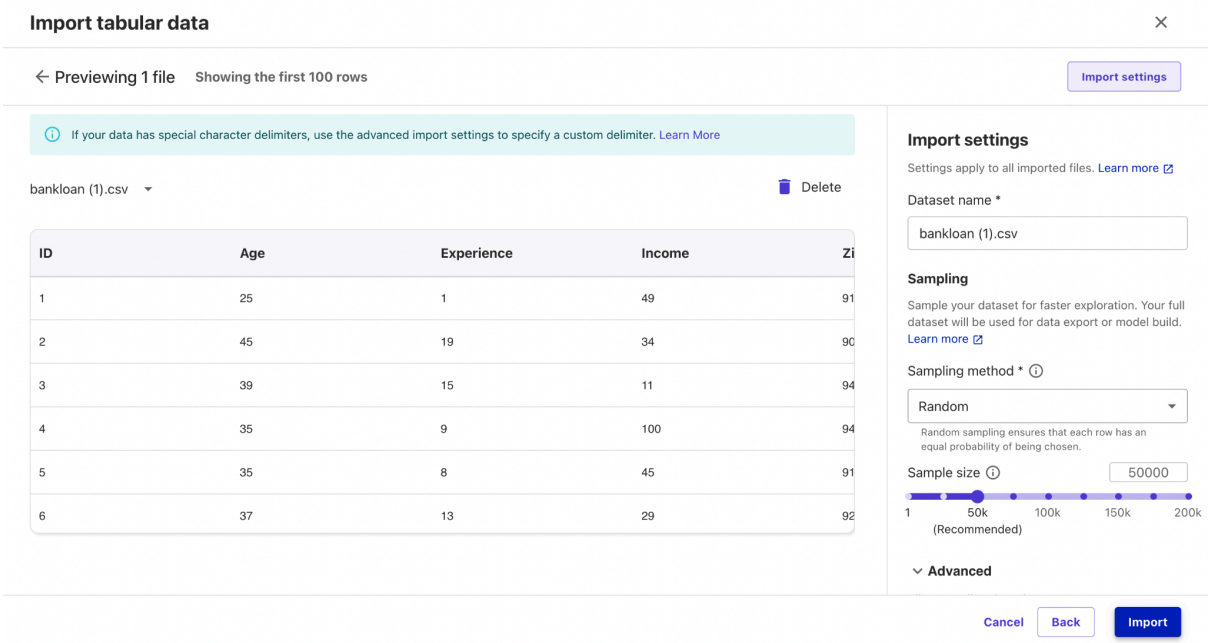




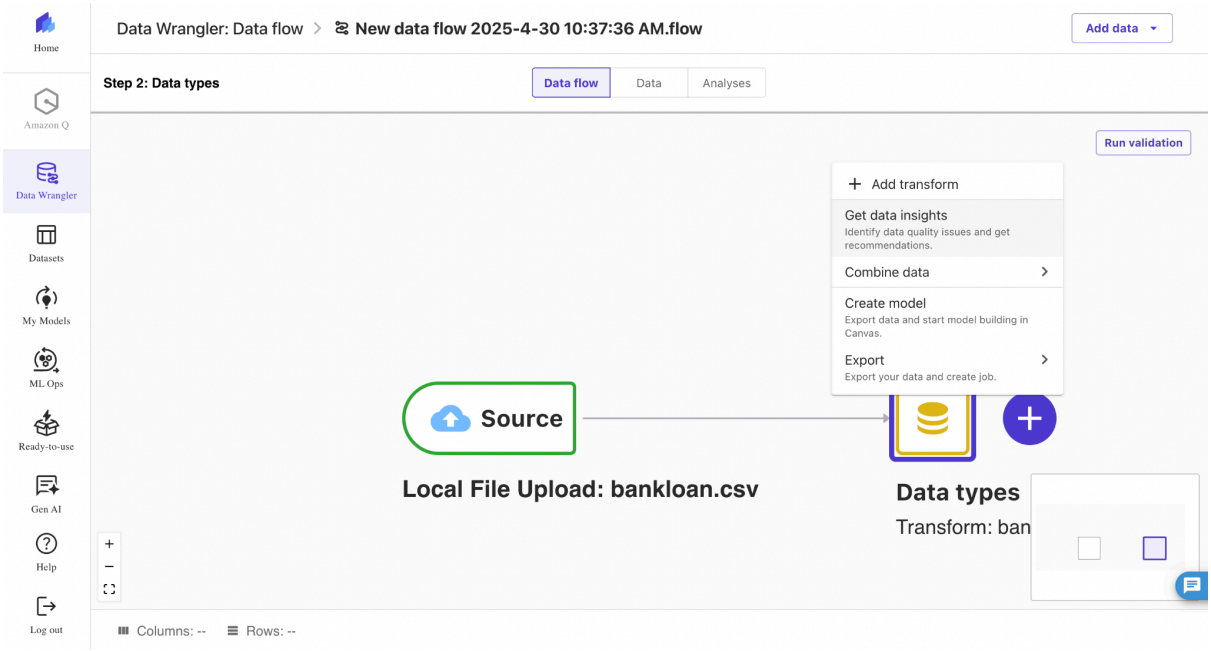
Step 1- Import Data

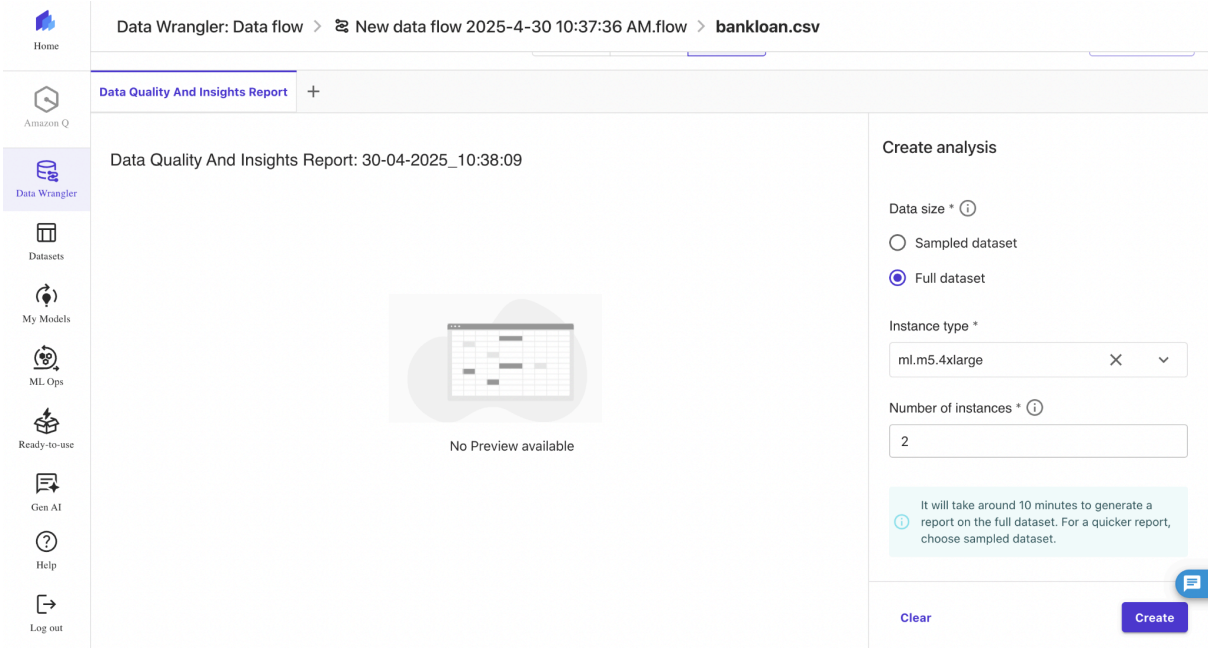

Home
Amazon Q
Data Wrangler
Datasets
My Models
ML Ops
Ready-to-use
Gen AI
Help
Data Wrangler
Search data flows
Import data flows
Import and prepare
Dataset type
Tabular
CSV or Parquet files
Image
PNG and JPG files
Data flows
Jobs
Data Wrangler data flows are recipes that capture all the data preparation steps you performed on the data. With data flows, you can import data from Amazon S3, Amazon Athena, Snowflake, and over 50 data sources to join, transform and analyze data using over 300 built-in operators and a natural language interface.
Flow name
Created
Last updated
New data flow 2025-4-28 4:57:00 PM.flow
04/28/2025 4:57 PM
04/28/2025 5:28 PM
Resources
Get started
What is Amazon SageMaker Data Wrangler
Get started with Data Wrangler in Canvas
Documentation
Pricing in Canvas
AWS technical guide
What's new
Accelerate data preparation for ML with comprehensive data preparation capabilities and a natural language interface

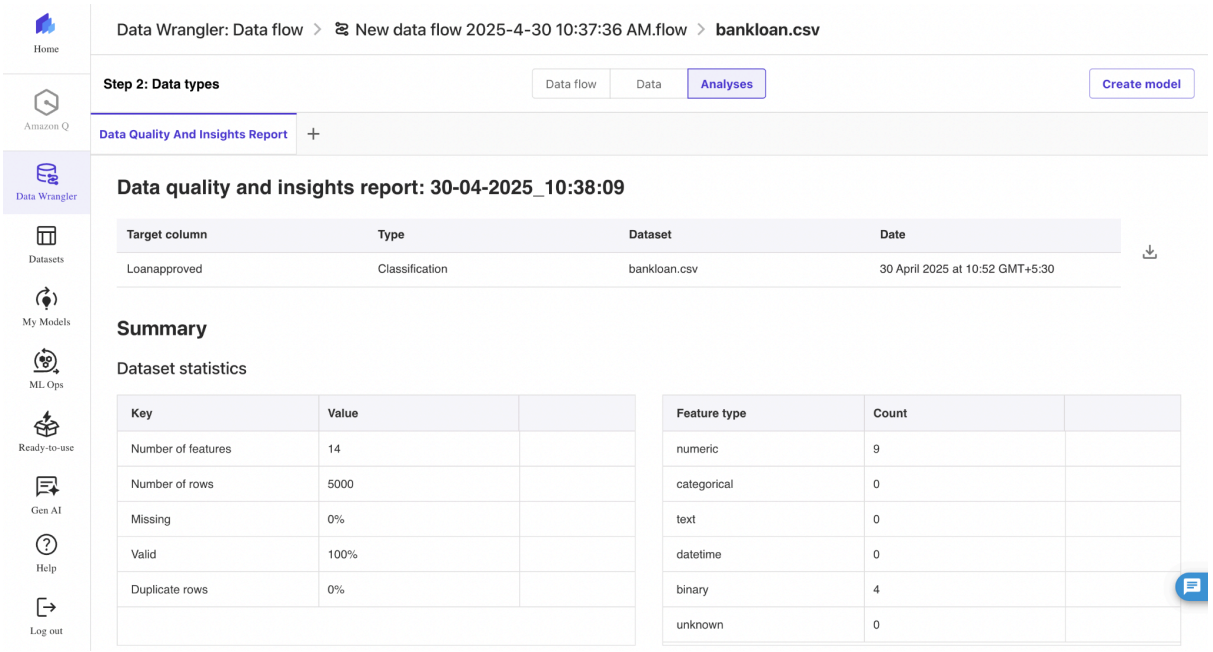

Import tabular data
Select a data source:
Amazon S3
Search data source
Filter by:
All (56)
Frequently used
Local upload
Canvas Datasets
Amazon S3
Snowflake
Redshift
Athena
Databricks
Salesforce Data Cloud
Learn more about data sources
Input S3 endpoint
Provide the ARN, URI,
Aliases should have the form
Go
Amazon S3
Name
sagemaker-studio-695309346212-rajeoapyvdp
04/25/2025 12:55 PM
sagemaker-studio-695309346212-8h9s52rhq4o
04/25/2025 12:10 PM



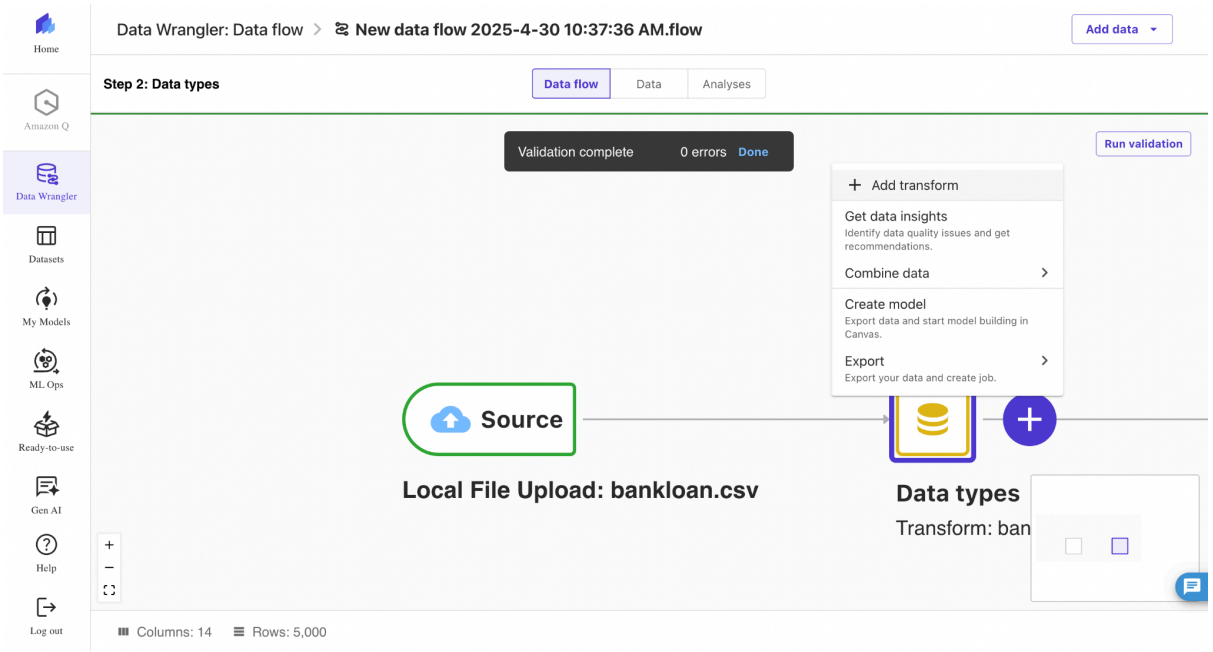
Step 2- Generate Data Quality Insights

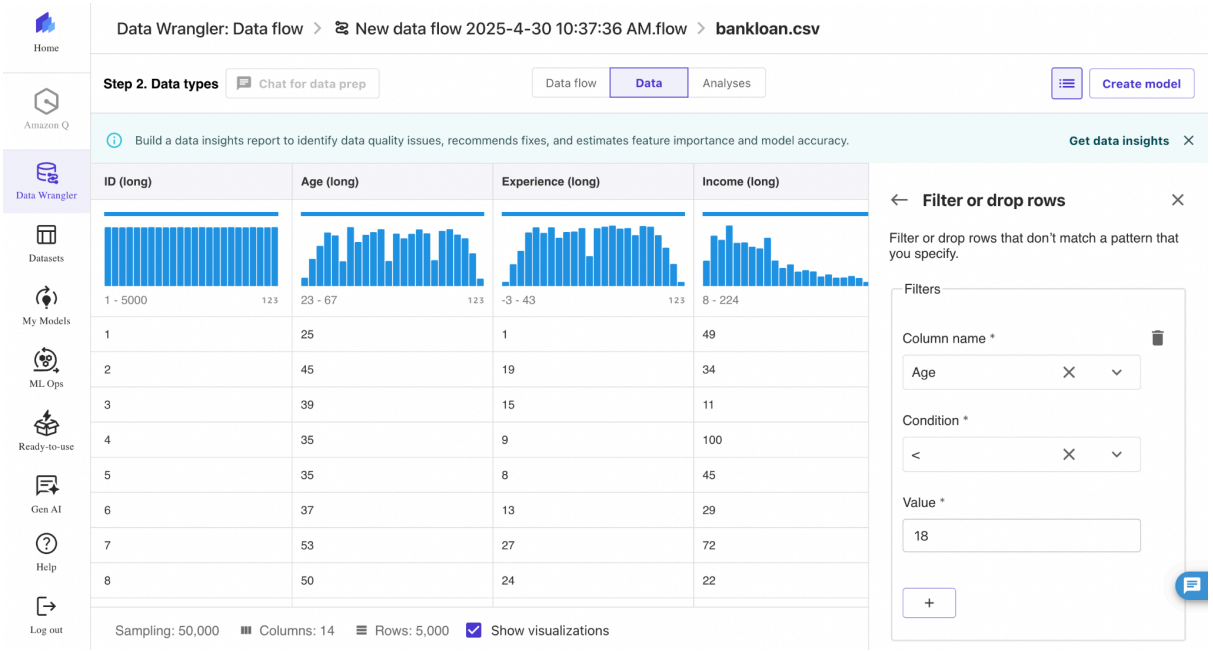


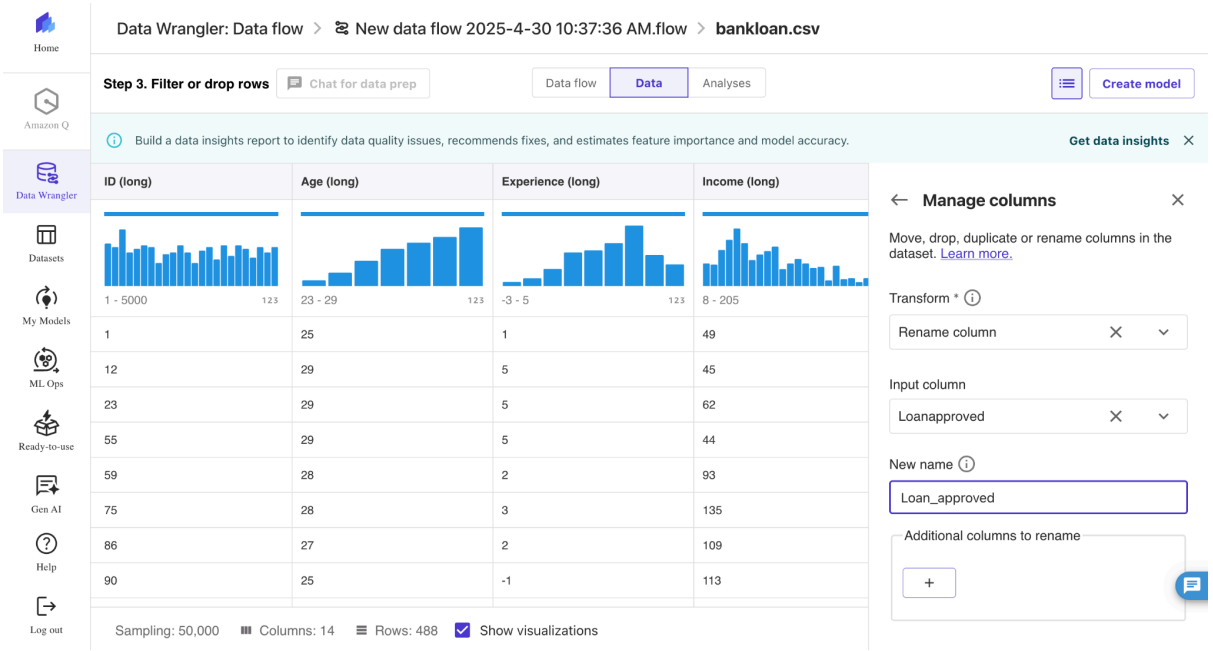

Home
Amazon Q
Data Wrangler
Datasets
My Models
ML Ops
Ready-to-use
Gen AI
Help
Log out
Data Wrangler: Data flow
New data flow 2025-4-30 10:37:36 AM.flow
bankloan.csv
Data Quality And Insights Report
Data Quality And Insights Report: 30-04-2025_10:38:09
No Preview available
Create analysis
Data size *
Sampled dataset
Full dataset
Instance type *
ml.m5.4xlarge
Number of instances *
2
It will take around 10 minutes to generate a report on the full dataset. For a quicker report, choose sampled dataset.
Clear
Create


Home
Amazon Q
Data Wrangler
Datasets
My Models
ML Ops
Ready-to-use
Gen AI
Help
Log out
Data Wrangler: Data flow
New data flow 2025-4-30 10:37:36 AM.flow
bankloan.csv
Step 2: Data types
Data flow
Data
Analyses
Create model
Data Quality And Insights Report
Data quality and insights report: 30-04-2025_10:38:09
Target column
Type
Dataset
Date
Loanapproved
Classification
bankloan.csv
30 April 2025 at 10:52 GMT+5:30
Summary
Dataset statistics
Key
Value
Number of features
14
Number of rows
5000
Missing
0%
Valid
100%
Duplicate rows
0%
Feature type
Count
numeric
9
categorical
0
text
0
datetime
0
binary
4
unknown
0

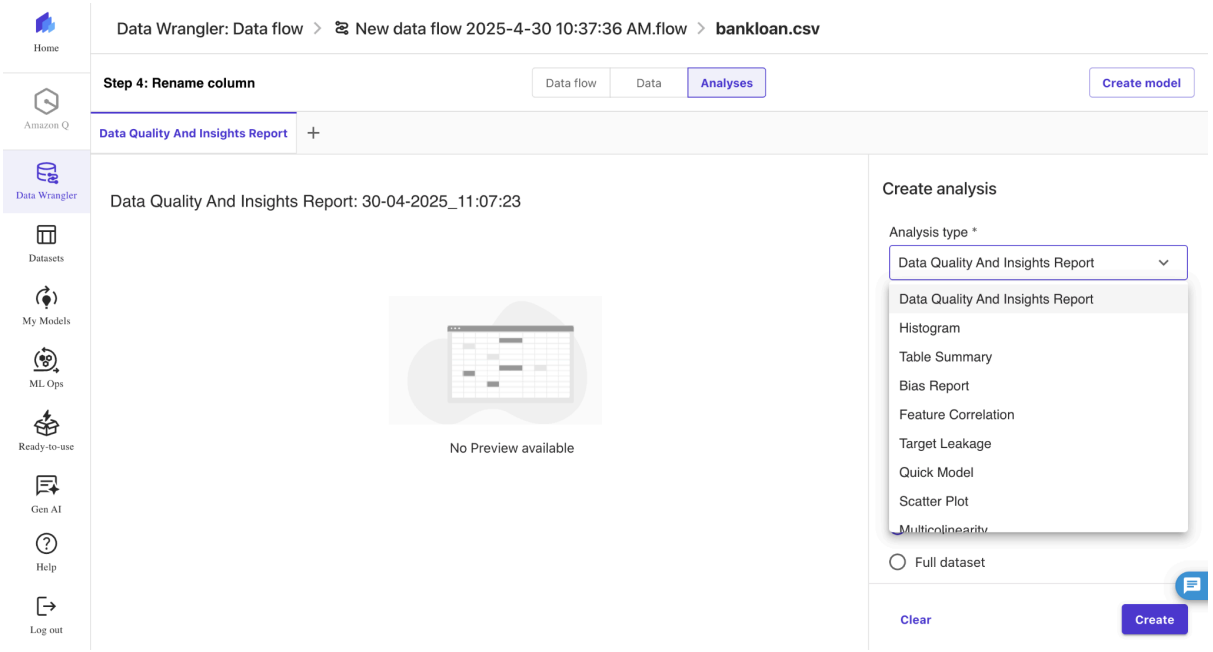
Step 3- Add Cleaning Transformations- click on data flow and then add transform







Step 4&5 - Univariate EDA, Bivariate and Correlation



Step 6- Feature Engineering