

Experiment with machine learning model training and deployment using cloud-based AI services.

Ebuka Obiakor – 11th March, 2024

Using AWS Sage Maker

Amazon SageMaker is a fully managed machine learning (ML) service. With SageMaker, data scientists and developers can quickly and confidently build, train, and deploy ML models into a production-ready hosted environment. It provides a UI experience for running ML workflows that makes SageMaker ML tools available across multiple integrated development environments (IDEs).

With SageMaker, you can store and share your data without having to build and manage your own servers. This gives you or your organizations more time to collaboratively build and develop your ML workflow, and do it sooner. SageMaker provides managed ML algorithms to run efficiently against extremely large data in a distributed environment

Typical SageMaker workflow



1. Label data

Set up and manage labeling jobs for highly accurate training datasets within Amazon SageMaker, using active learning and human labeling.



2. Build

Connect to other AWS services and transform data in Amazon SageMaker notebooks.



3. Train

Use Amazon SageMaker's algorithms and frameworks, or bring your own, for distributed training.



4. Tune

Amazon SageMaker automatically tunes your model by adjusting multiple combinations of algorithm parameters.



5. Deploy

After training is completed, models can be deployed to Amazon SageMaker endpoints, for real-time predictions.



6. Discover

Find, buy, and deploy ready-to-use model packages, algorithms, and data products in AWS Marketplace.

Start by navigating to the AWS Sage Maker.

Create a domain and roles for Data Scientist and MLOpsEngineer

Amazon SageMaker > Domains > Domain: QuickSetupDomain-20240311T203305

QuickSetupDomain-20240311T203305

Domain details

Configure and manage the domain.

User profiles

Space management

Environment

Domain settings

User profiles Info

⌂

Add user

A user profile represents a single user within a domain. It is the main way to reference a user for the purposes of sharing, reporting, and other user-oriented features.

🔍 Search users

< 1 > ⌂

Name	Modified on	Created on	
default-20240311T203305	Mar 12, 2024 02:49 UTC	Mar 12, 2024 02:49 UTC	<div>Launch ▾</div>

Open Studio labs (Canvas)

Home

Data Wrangler

My Models

ML Ops

Ready-to-use

Shared Models

Get AI

Data Wrangler

Datasets


Data flows

🔍 Search data flows


+ Create

Name	Created	Last updated ↓
New data flow 2024-3-11 9:27 PM.flow	03/11/2024 9:27 PM	03/11/2024 9:27 PM
New data flow 2024-3-11.flow	03/11/2024 9:09 PM	03/11/2024 9:09 PM


Select Create data flow




Import data



Prepare data



Scale data operations



Build models

To get started...

Import data ▾

Select existing dataset ▾

Import data or use local available data

Import data

Data Source:

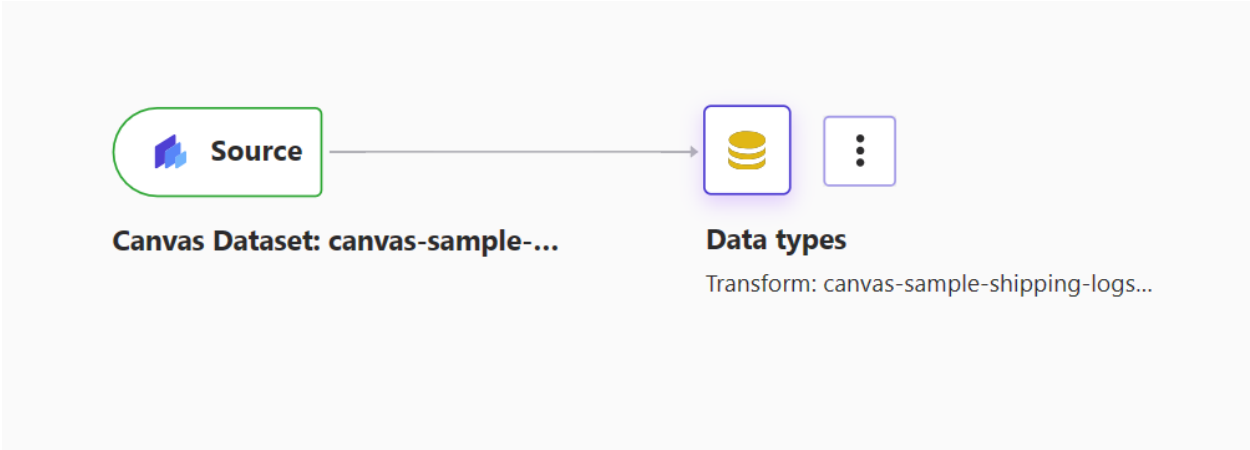
Amazon S3

> Input S3 endpoint

Amazon S3

Search Amazon S3

<input type="checkbox"/>	Name	Created on ↓
	oregon-bucket-s3	03/11/2024 8:57 PM
	sagemaker-us-west-2-058264200429	03/11/2024 8:33 PM
	sagemaker-studio-058264200429-rosigfn5gp	03/11/2024 8:33 PM
	aws-glue-assets-058264200429-us-west-2	03/11/2024 4:12 PM
	elasticbeanstalk-us-west-2-058264200429	03/07/2024 11:42 PM
	elasticbeanstalk-us-east-2-058264200429	03/07/2024 11:39 PM
	acebucket0303	03/03/2024 10:38 AM



Build a model to use with the data

My models

Search models

New model

Grid

List

Name	Versions	Problem type	Target	Updated
------	----------	--------------	--------	---------

Select Build Analyze Predict Deploy

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

Target column
median_house_value

Value distribution



Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

Numeric prediction

For the median_house_value, your model predicts numeric values.

Configure model

Quick build

Preview model

canvas-sample-housing.csv

Full dataset: 1.0k rows

Manage columns Manage rows Time series View all

Data visualizer

To build a model with this dataset, set **median_house_value** as a target column, and **Numeric prediction** as the model type. To learn more about how to build a model with this dataset, check our [workshop page](#).

Column name	Data type	Feature type	Missing	Mismatched	Unique	Mode	Correlation to target
population	123 Numeric	-	0.00% (0)	0.00% (0)	783	1,109	0.152
ocean_proximity	Δ Text	Categorical	0.00% (0)	0.00% (0)	3	NEAR BAY	N/A
median_income	123 Numeric	-	0.00% (0)	0.00% (0)	935	2.88	0.767
median_house_value	123 Numeric	-	0.00% (0)	0.00% (0)	781	500,001	--
longitude	123 Numeric	-	0.00% (0)	0.00% (0)	57	-122.27	0.197
latitude	123 Numeric	-	0.00% (0)	0.00% (0)	44	37.77	0
housing_median_age	123 Numeric	-	0.00% (0)	0.00% (0)	51	52	-0.138
households	123 Numeric	-	0.00% (0)	0.00% (0)	575	267	0.176

Total columns: 10 Total rows: 1,000 Total cells: 10,000 Show dropped columns

Sample visualizations.

canvas-sample-housing.csv

Visualization sample: 100 rows

Manage columns Manage rows Time series View all

Data visualizer

To build a model with this dataset, set **median_house_value** as a target column, and **Numeric prediction** as the model type. To learn more about how to build a model with this dataset, check our [workshop page](#).

Visualizations

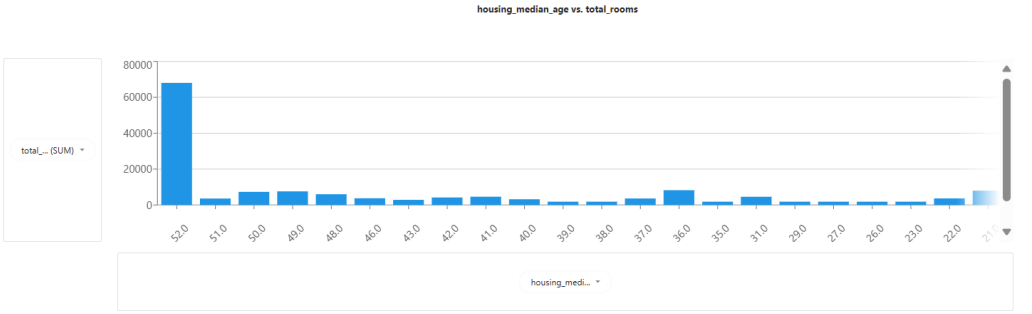
Analytics

Visualization

Scatter plot Bar chart Box plot

Columns

longitude latitude
housing_... total_roo...
total_be... populati...
househo... median_h...
median_... ocean_pr...



Group by Drop column here Edit

Stack by Drop column here Edit

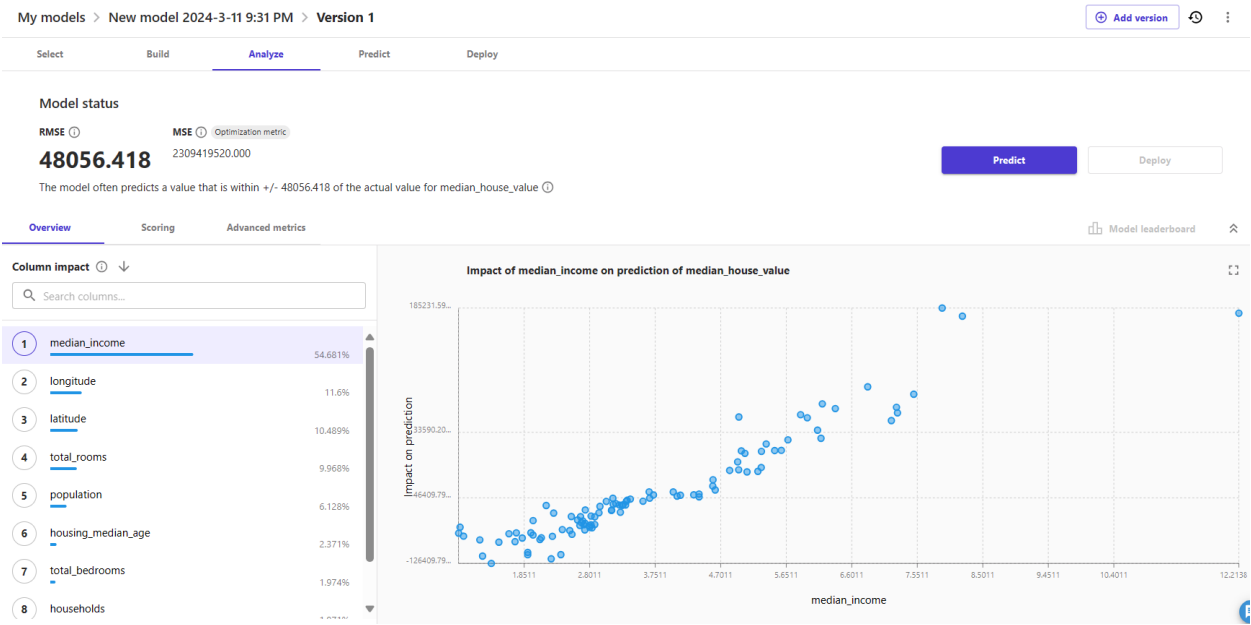
Bar chart

Sample Analytics



Wait for model to train

Once model is trained, view results



Test prediction results

My models > New model 2024-3-11 9:31 PM > Version 1

Add version

SelectBuildAnalyzePredictDeploy

Predict target values

Batch predictionSingle prediction

Modify values to predict median_house_value in real time.

Filter columns

Column	Feature importance	Value
longitude	11.795%	-122.27
latitude	3.218%	37.77
housing_median_age	2.426%	52
total_rooms	6.445%	1487
total_bedrooms	0.655%	365
population	2.535%	1109
households	3.987%	267
median_income	68.932%	1.875
ocean_proximity	0.007%	NEAR BAY

median_house_value Prediction

Copy

115208.422

New predictionAverage prediction

115208.422

115208.422

Iterate and test model.

Deploy when satisfied with results.

References

<https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>