

PROJECT OVERVIEW [SKIP IF FAMILIAR]



PROJECT CARD

GOAL:

- The objective of this case study is to predict the health insurance cost incurred by Individuals based on their age, gender, BMI, number of children, smoking habit and geo-location.

TOOL:

- *AWS SageMaker Autopilot*

PRACTICAL REAL-WORLD APPLICATION:

- *This project can be effectively used by insurance companies to predict healthcare insurance cost, increase revenues and reduce costs.*

DATA:

• **INPUTS:**

- age, gender, BMI, number of children, smoking habit and geo-location

• **OUTPUT:**

- *Insurance Charges*



DATA OVERVIEW

The available features are:

- sex: insurance contractor gender
- bmi: Body mass index (ideally 18.5 to 24.9)
- children: Number of children covered by health insurance / Number of dependents
- smoker: Smoking
- region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.

Target (output):

- charges: Individual medical costs billed by health insurance

- Data Source: <https://www.kaggle.com/mirichoi0218/insurance>

DATA OVERVIEW

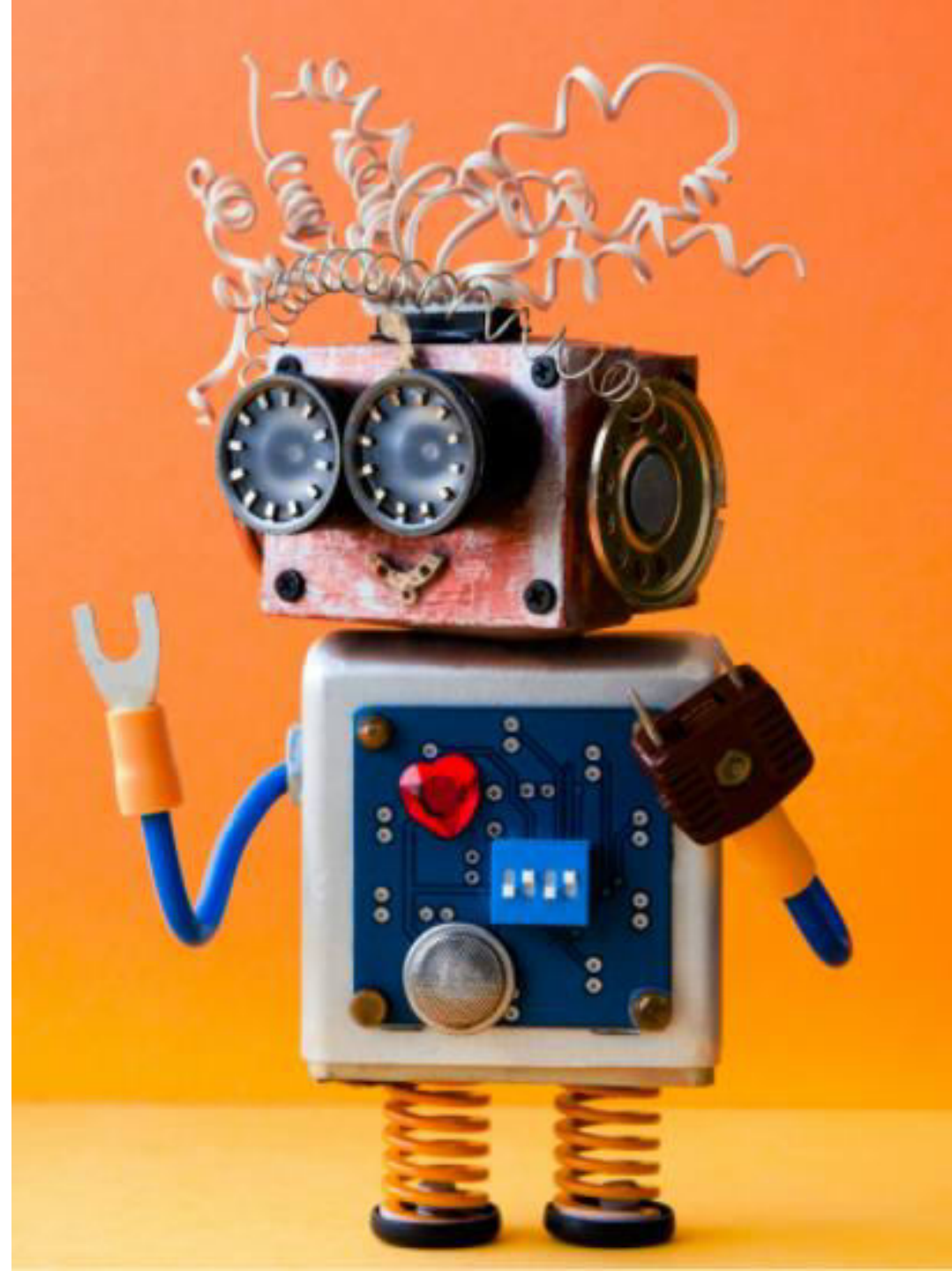
	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns



TARGET
COLUMN

AMAZON SAGEMAKER AUTOPILOT



AMAZON SAGEMAKER

- Amazon SageMaker is a fully-managed machine learning workflow platform that provides services on data labeling, model building, training, tuning and deployment.
- SageMaker allows data scientists and developers to build scalable AI/ML models easily and efficiently.
- Models could be deployed in production at a much faster rate and with a fraction of the cost.
- Let's explore SageMaker:
<https://aws.amazon.com/sagemaker/#>

BUILD

- SageMaker offers data labeling service
- Prebuilt available notebooks with state-of-the-art algorithms on AWS marketplace

TRAIN

- Train models using EC2 instances (on-demand and spot)
- Manage environments for training
- Hyperparameters optimization for model tuning

DEPLOY

- Easily deploy and scale models
- Autoscaling with 75% savings

AMAZON SAGEMAKER MODEL TRAINING AND DEPLOYMENT OVERVIEW

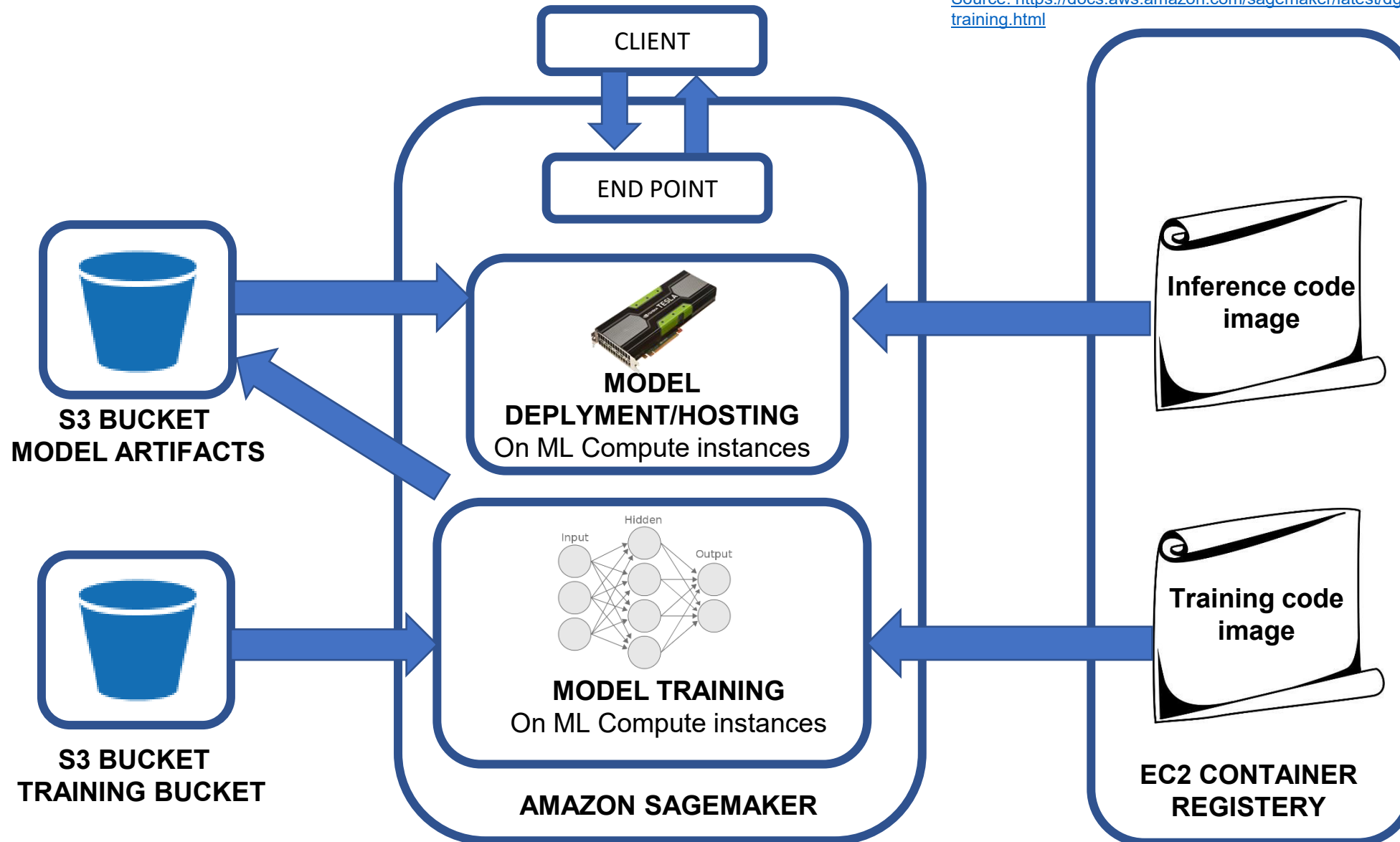
https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg

https://commons.wikimedia.org/wiki/File:AWS_Simple_Icons_Storage_Amazon_S3.svg

<https://docs.aws.amazon.com/sagemaker/latest/dg/how-it-works-training.html>

<https://www.flickr.com/photos/gbpublic/8178512552>

Source: <https://docs.aws.amazon.com/sagemaker/latest/dg/how-it-works-training.html>



AMAZON SAGEMAKER AUTOPILOT 101

Check this out: <https://aws.amazon.com/sagemaker/autopilot/>

Amazon SageMaker Autopilot

Automatically create machine learning models with full visibility

Get Started with Amazon SageMaker Autopilot

FEATURED EVENT
SageMaker Fridays Are Back With a New Season
Sign up to get started faster with machine learning using Amazon SageMaker.

Register now >

Amazon SageMaker Autopilot automatically builds, trains, and tunes the best machine learning models based on your data, while allowing you to maintain full control and visibility.

Building machine learning (ML) models requires you to manually prepare features, test multiple algorithms, and optimize hundreds of model parameters in order to find the best model for your data. However, this approach requires deep ML expertise. If you don't have that expertise, you could use an automated approach (AutoML), but AutoML approaches typically provide very little visibility into the impact of your features for model predictions. As a result, you may have less trust in it because you can't recreate it and you can't learn how it makes predictions.

Amazon SageMaker Autopilot eliminates the heavy lifting of building ML models, and helps you automatically build, train, and tune the best ML model based on your data. With SageMaker Autopilot, you simply provide a tabular dataset and select the target column to predict, which can be a number (such as a house price, called regression), or a category (such as spam/not spam, called classification). SageMaker Autopilot will automatically explore different solutions to find the best model. You then can directly deploy the model to production with just one click, or iterate on the recommended solutions with Amazon SageMaker Studio to further improve the model quality.

How it works

```
graph LR; A[Raw data] --> B[Target]; B --> C[Automatic model creation]; C --> D[Full visibility and control]; D --> E[Model leaderboard]; E --> F[Deploy and monitor the model]; F -- "Choice to optimize and retrain, to improve model quality" --> E
```

Raw data
Load tabular data from Amazon S3 to train the model

Target
Select target column for prediction

Automatic model creation
The correct algorithm is chosen, training and tuning is done automatically for the right model

Full visibility and control
Full visibility with model notebooks

Model leaderboard
Select the best model for your needs from a ranked list of recommendations

Deploy and monitor the model

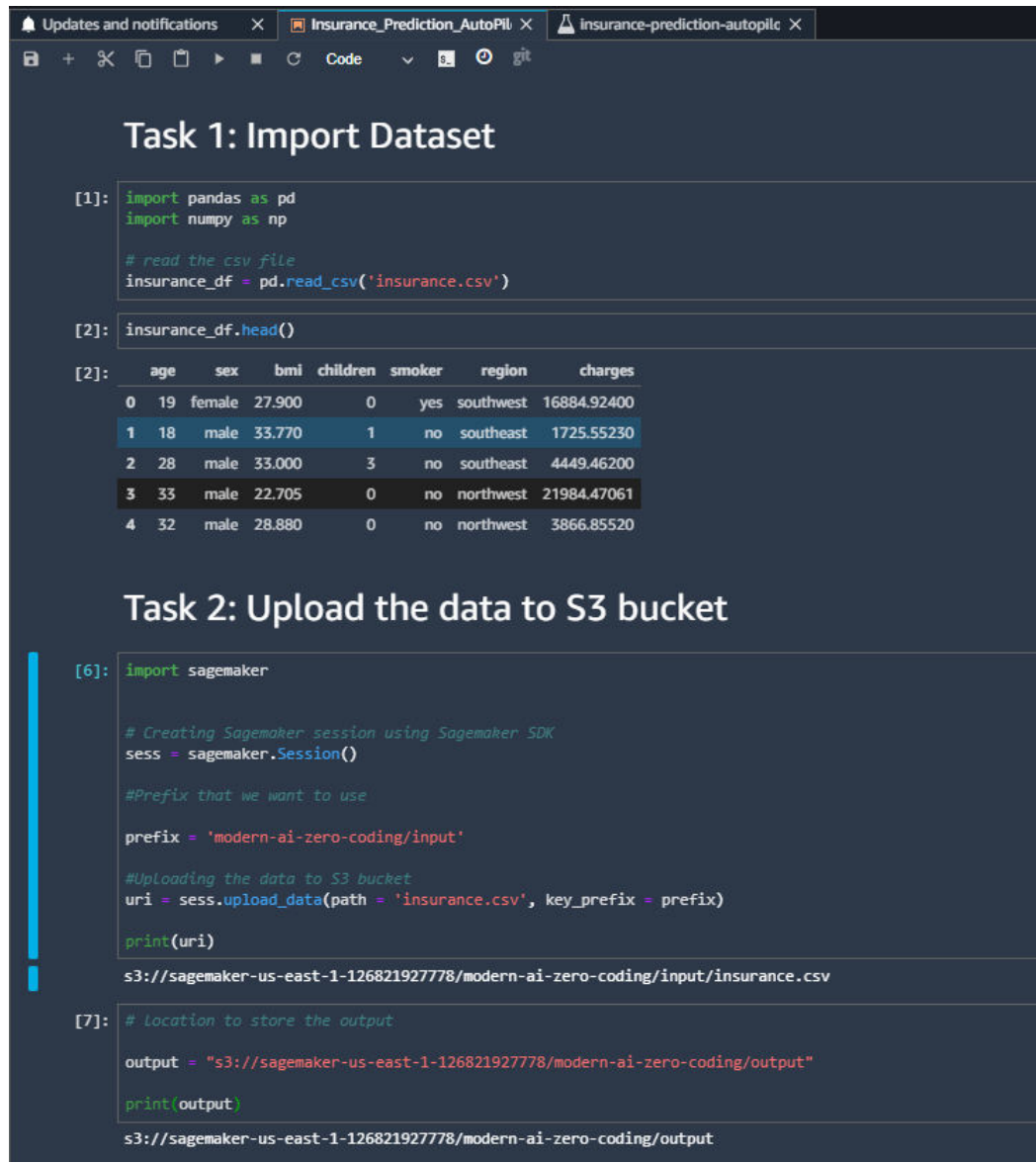
Choice to optimize and retrain, to improve model quality

AMAZON SAGEMAKER AUTOPILOT DEMO



AMAZON SAGEMAKER AUTOPILOT

RUN THIS SCRIPT



```
Updates and notifications X Insurance_Prediction_AutoPil X insurance-prediction-autopilc X
+ ✂ 📄 📁 ▶ ■ ↺ Code ▾ 📄 ⌚ git
```

Task 1: Import Dataset

```
[1]: import pandas as pd
import numpy as np

# read the csv file
insurance_df = pd.read_csv('insurance.csv')
```

```
[2]: insurance_df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Task 2: Upload the data to S3 bucket

```
[6]: import sagemaker

# Creating Sagemaker session using Sagemaker SDK
sess = sagemaker.Session()

#Prefix that we want to use
prefix = 'modern-ai-zero-coding/input'

#Uploading the data to S3 bucket
uri = sess.upload_data(path = 'insurance.csv', key_prefix = prefix)

print(uri)

s3://sagemaker-us-east-1-126821927778/modern-ai-zero-coding/input/insurance.csv
```

```
[7]: # Location to store the output

output = "s3://sagemaker-us-east-1-126821927778/modern-ai-zero-coding/output"

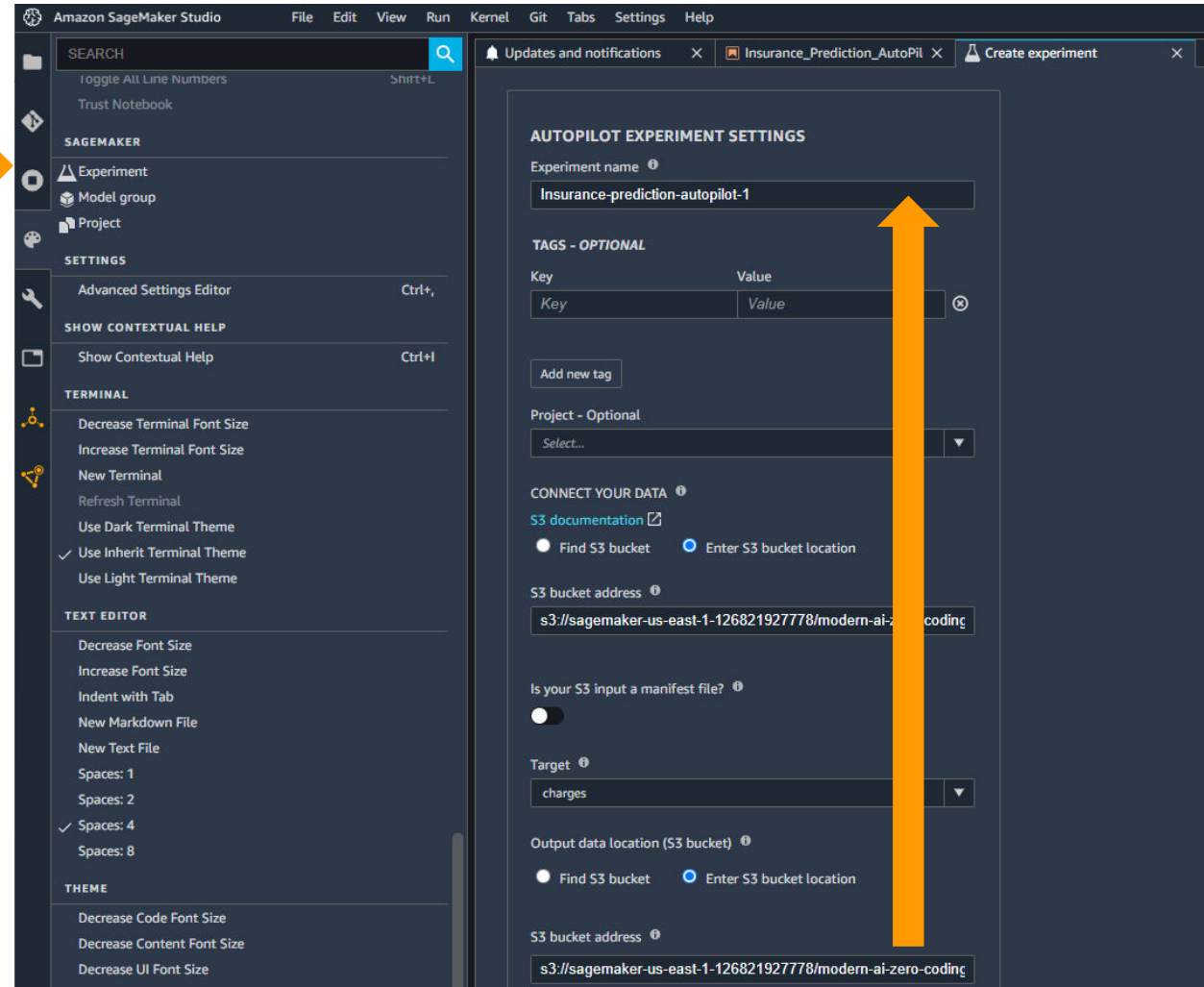
print(output)

s3://sagemaker-us-east-1-126821927778/modern-ai-zero-coding/output
```

AMAZON SAGEMAKER AUTOPILOT

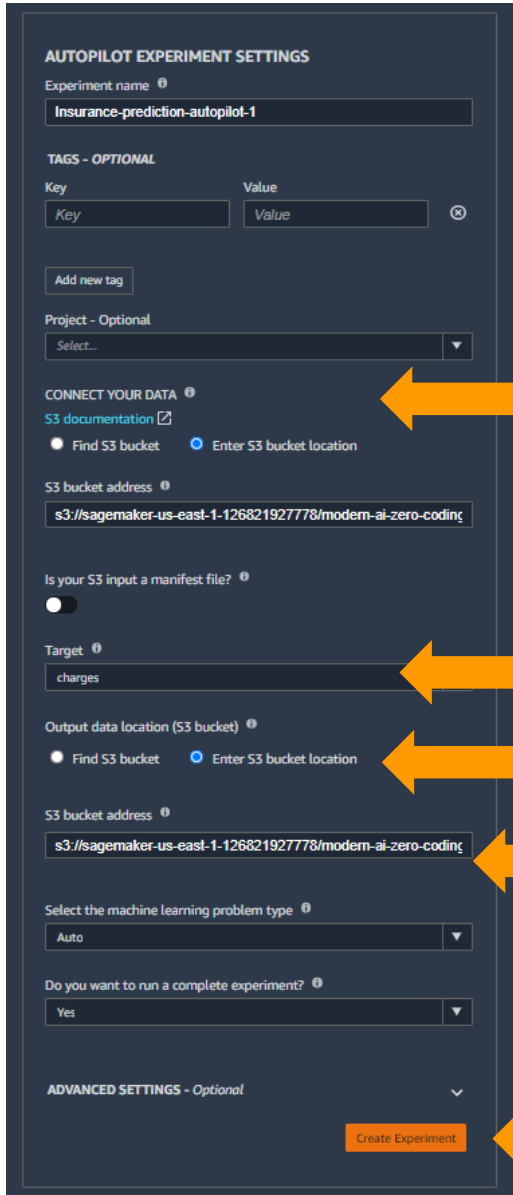
Check this out: <https://aws.amazon.com/sagemaker/autopilot/>

1. CLICK ON
EXPERIMENT



2. PROVIDE A NAME FOR THE EXPERIMENT

AMAZON SAGEMAKER AUTOPILOT



AUTOPILOT EXPERIMENT SETTINGS

Experiment name ⓘ
Insurance-prediction-autopilot-1

TAGS - OPTIONAL

Key	Value
Key	Value ⓘ

[Add new tag](#)

Project - Optional
Select...

CONNECT YOUR DATA ⓘ
[S3 documentation](#) ⓘ

☐ Find S3 bucket ☒ Enter S3 bucket location

S3 bucket address ⓘ
s3://sagemaker-us-east-1-126821927778/modern-ai-zero-coding

Is your S3 input a manifest file? ⓘ
☐

Target ⓘ
charges

Output data location (S3 bucket) ⓘ
☐ Find S3 bucket ☒ Enter S3 bucket location

S3 bucket address ⓘ
s3://sagemaker-us-east-1-126821927778/modern-ai-zero-coding

Select the machine learning problem type ⓘ
Auto

Do you want to run a complete experiment? ⓘ
Yes

ADVANCED SETTINGS - Optional ▾

Create Experiment

SELECT "ENTER S3 BUCKET LOCATION"

PROVIDE THE LOCATION OF TRAIN DATA IN S3 (CAN BE FOUND IN .IPYNB NOTEBOOK)

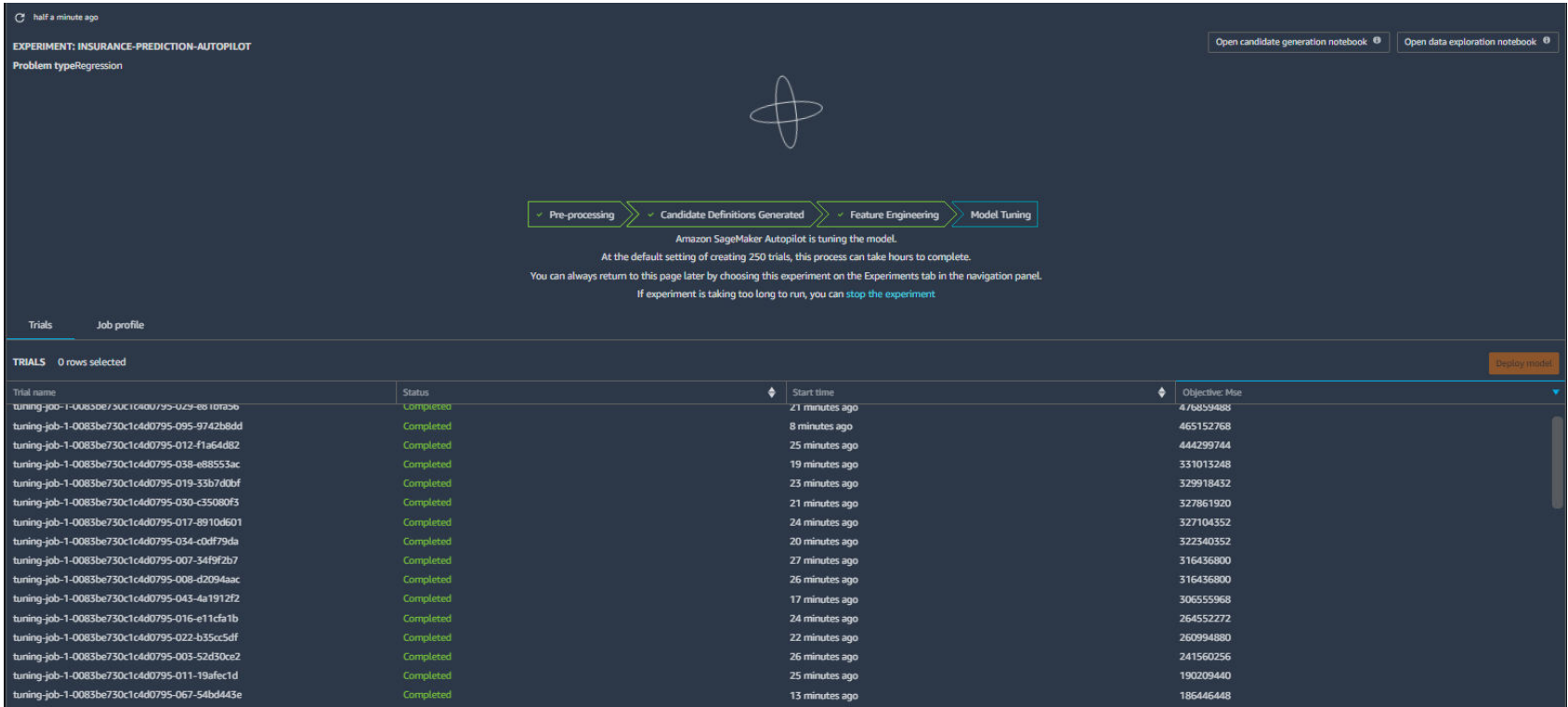
PROVIDE THE COLUMN NAME OF TARGET VARIABLE

SELECT "ENTER S3 BUCKET LOCATION"

PROVIDE THE PATH TO STORE THE OUTPUT IN S3 (CAN BE FOUND IN .IPYNB NOTEBOOK)

CLICK CREATE EXPERIMENT

AMAZON SAGEMAKER AUTOPILOT

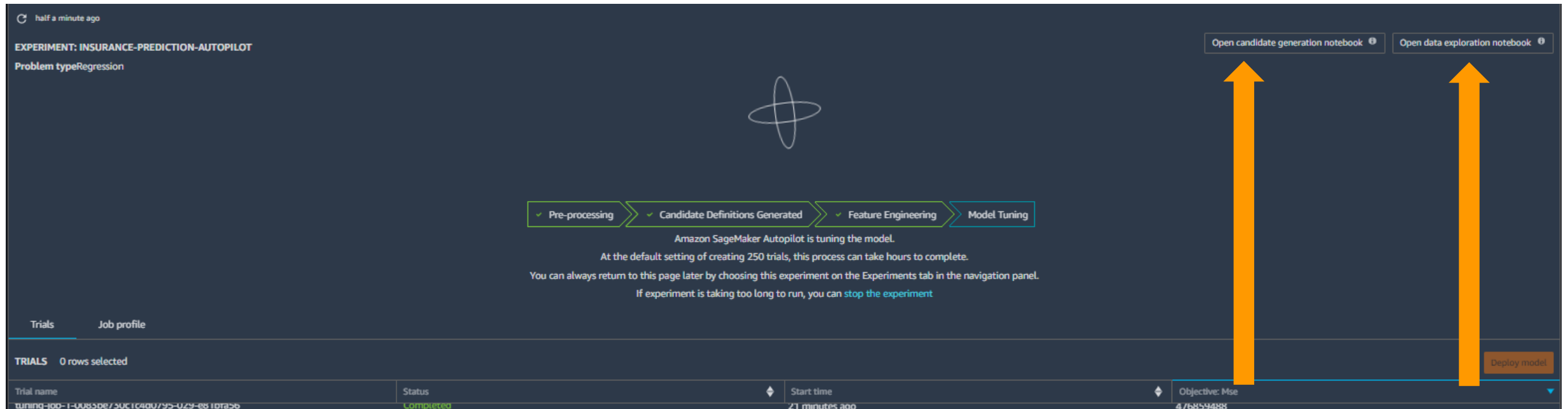


AFTER CREATING EXPERIMENT, YOU SHOULD SEE SOMETHING LIKE THIS.

IT WOULD TAKE AROUND 2 HOURS TO COMPLETE. THIS IS BECAUSE BY-DEFAULT IT WOULD RUN 250 TUNING JOBS TO FIGURE-OUT THE BEST HYPER-PARAMETER

AMAZON SAGEMAKER AUTOPILOT

Once first 2 steps are completed, you can view candidate generation notebook and data exploration notebook.



The screenshot displays the Amazon SageMaker Autopilot console for an experiment named "INSURANCE-PREDICTION-AUTOPILOT". The problem type is "Regression". The progress bar shows four steps: "Pre-processing" (completed), "Candidate Definitions Generated" (completed), "Feature Engineering" (in progress), and "Model Tuning" (pending). Below the progress bar, a message states: "Amazon SageMaker Autopilot is tuning the model. At the default setting of creating 250 trials, this process can take hours to complete. You can always return to this page later by choosing this experiment on the Experiments tab in the navigation panel. If experiment is taking too long to run, you can stop the experiment".

Two orange arrows point from the bottom of the console to the buttons "Open candidate generation notebook" and "Open data exploration notebook".

At the bottom, there is a table with the following data:

Trial name	Status	Start time	Objective: Mse
tuning-job-1-00830e750c1c4d0795-029-e810a36	Completed	21 minutes ago	476859488

A "Deploy model" button is located at the bottom right of the console.

CLICK ON THESE TO OPEN RESPECTIVE NOTEBOOKS.

NOTE: GENERATED NOTEBOOKS FROM OUR EXPERIMENT ARE PROVIDED FOR REFERENCE.

AMAZON SAGEMAKER AUTOPILOT

Amazon SageMaker Studio

half a minute ago

AI_in_Business_final.ipynb AI_in_Business_autopilot.ip... ai-in-business-autopilot Create experiment SageMakerAutopilotDataE SageMakerAutopilotCandik % Deploy model SageMakerAutopilo

about 2 hours ago

EXPERIMENT: AI-IN-BUSINESS-AUTOPILOT

Open candidate generati

Trials Job profile

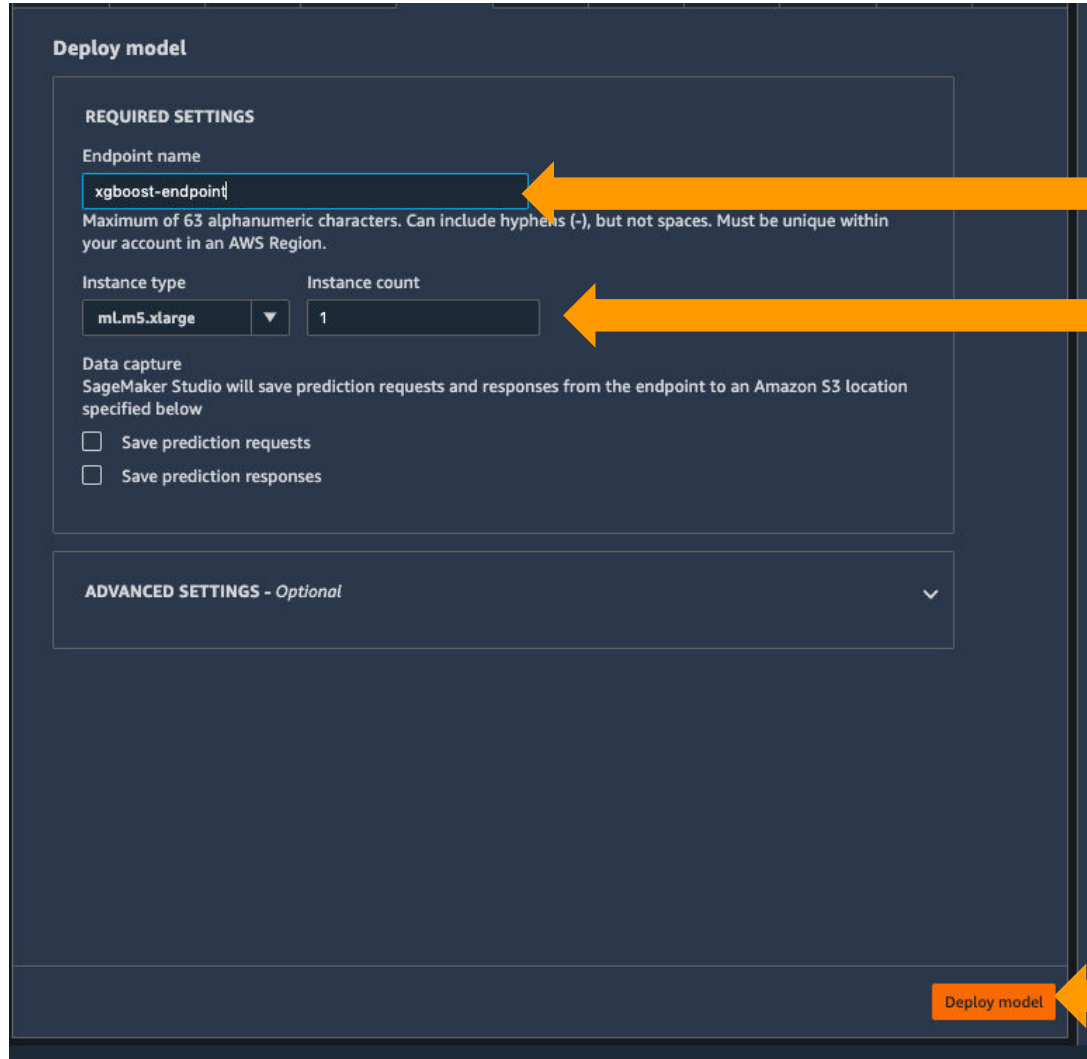
TRIALS

1 row selected

Trial name	Status	Start time	Objective
★ Best: tuning-job-1-458405b91f75482993-106-fcd537f	Completed	3 hours ago	
tuning-job-1-458405b91f75482993-204-fe99bc21	Completed	2 hours ago	0.8288300037384033
tuning-job-1-458405b91f75482993-145-a472f54d	Completed	3 hours ago	0.8286700248718262
tuning-job-1-458405b91f75482993-240-05f7f5e4	Completed	2 hours ago	0.8286700248718262
tuning-job-1-458405b91f75482993-078-1bdf6d15	Completed	3 hours ago	0.828499972820282
tuning-job-1-458405b91f75482993-015-7fc8d320	Completed	3 hours ago	0.8283299803733826
tuning-job-1-458405b91f75482993-203-8692d67d	Completed	2 hours ago	0.8283299803733826
tuning-job-1-458405b91f75482993-189-7c64de82	Completed	2 hours ago	0.8281700015068054
tuning-job-1-458405b91f75482993-227-f707b42c	Completed	2 hours ago	0.8281700015068054
tuning-job-1-458405b91f75482993-245-adc46278	Completed	2 hours ago	0.828000009059906
tuning-job-1-458405b91f75482993-103-04309b4c	Completed	3 hours ago	0.828000009059906
tuning-job-1-458405b91f75482993-012-a9449b9b	Completed	3 hours ago	0.828000009059906
tuning-job-1-458405b91f75482993-137-d154ad2d	Completed	3 hours ago	0.8278300166130066
tuning-job-1-458405b91f75482993-226-857e19f4	Completed	2 hours ago	0.8278300166130066
tuning-job-1-458405b91f75482993-188-a70384aa	Completed	2 hours ago	0.8276699781417847
tuning-job-1-458405b91f75482993-247-0f6e3f69	Completed	2 hours ago	0.8276699781417847
tuning-job-1-458405b91f75482993-191-44857475	Completed	2 hours ago	0.8276699781417847
tuning-job-1-458405b91f75482993-121-f53afb3e	Completed	3 hours ago	0.8276699781417847
tuning-job-1-458405b91f75482993-094-4eb02a00	Completed	3 hours ago	0.8274999856948853
tuning-job-1-458405b91f75482993-246-b4efc25a	Completed	2 hours ago	0.8274999856948853
tuning-job-1-458405b91f75482993-058-7282421c	Completed	3 hours ago	0.8274999856948853
tuning-job-1-458405b91f75482993-110-a64b2f27	Completed	3 hours ago	0.8274999856948853
tuning-job-1-458405b91f75482993-005-c5f3f789	Completed	3 hours ago	0.8274999856948853
tuning-job-1-458405b91f75482993-148-d035832f	Completed	3 hours ago	0.8274999856948853

ONCE THE JOB IS COMPLETED, YOU CAN SELECT THE BEST MODEL AND THEN CLICK ON DEPLOY MODEL TO DEPLOY.

AMAZON SAGEMAKER AUTOPILOT



Deploy model

REQUIRED SETTINGS

Endpoint name
xgboost-endpoint

Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.

Instance type
mL.m5.xlarge

Instance count
1

Data capture
SageMaker Studio will save prediction requests and responses from the endpoint to an Amazon S3 location specified below

☐ Save prediction requests

☐ Save prediction responses

ADVANCED SETTINGS - Optional

Deploy model

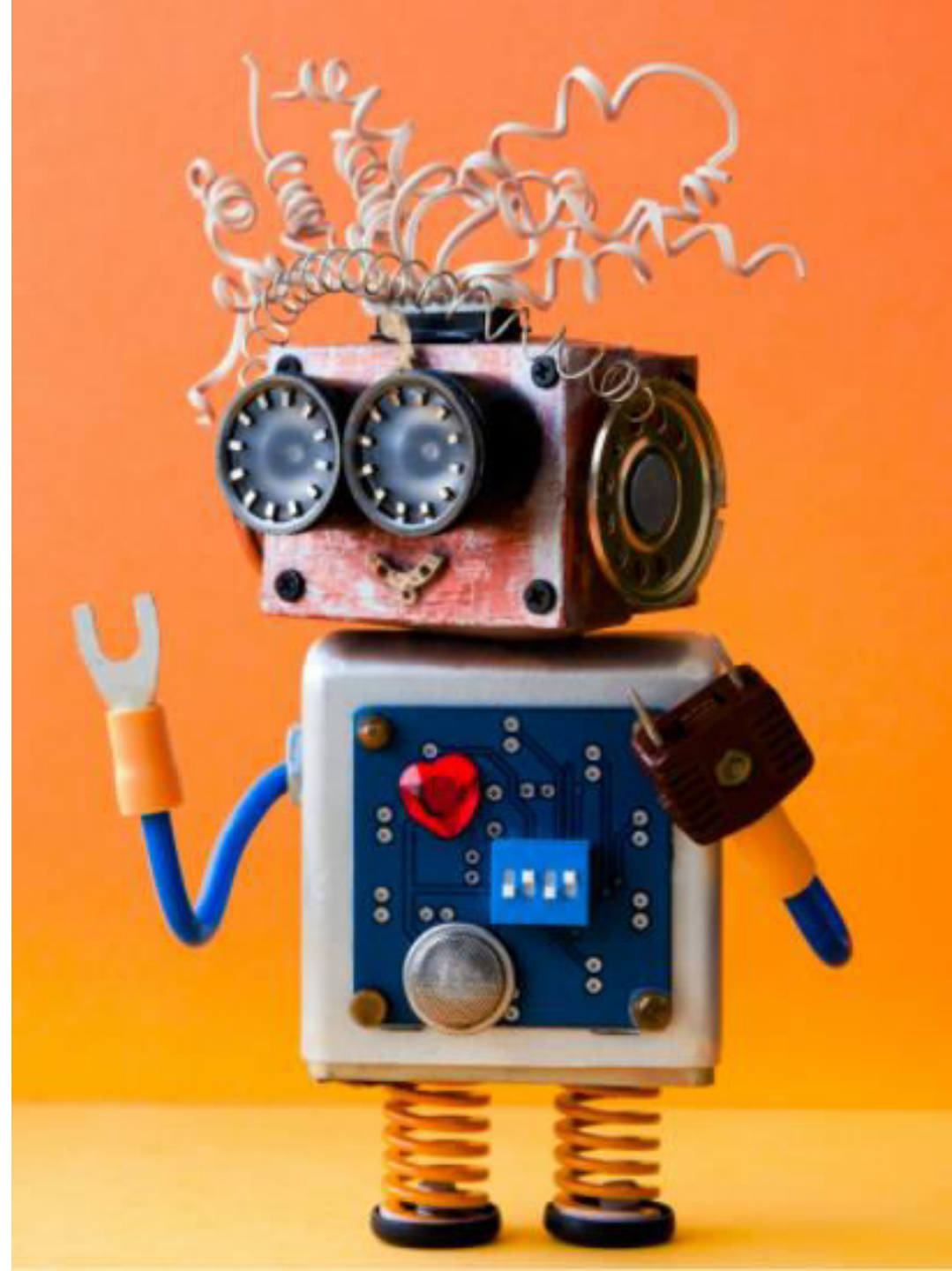
PROVIDE THE NAME FOR END-POINT

SELECT THE INSTANCE TYPE
AND THE NUMBER OF INSTANCE

CLICK TO DEPLOY THE MODEL

DELETE ENDPOINT

[IMPORTANT]

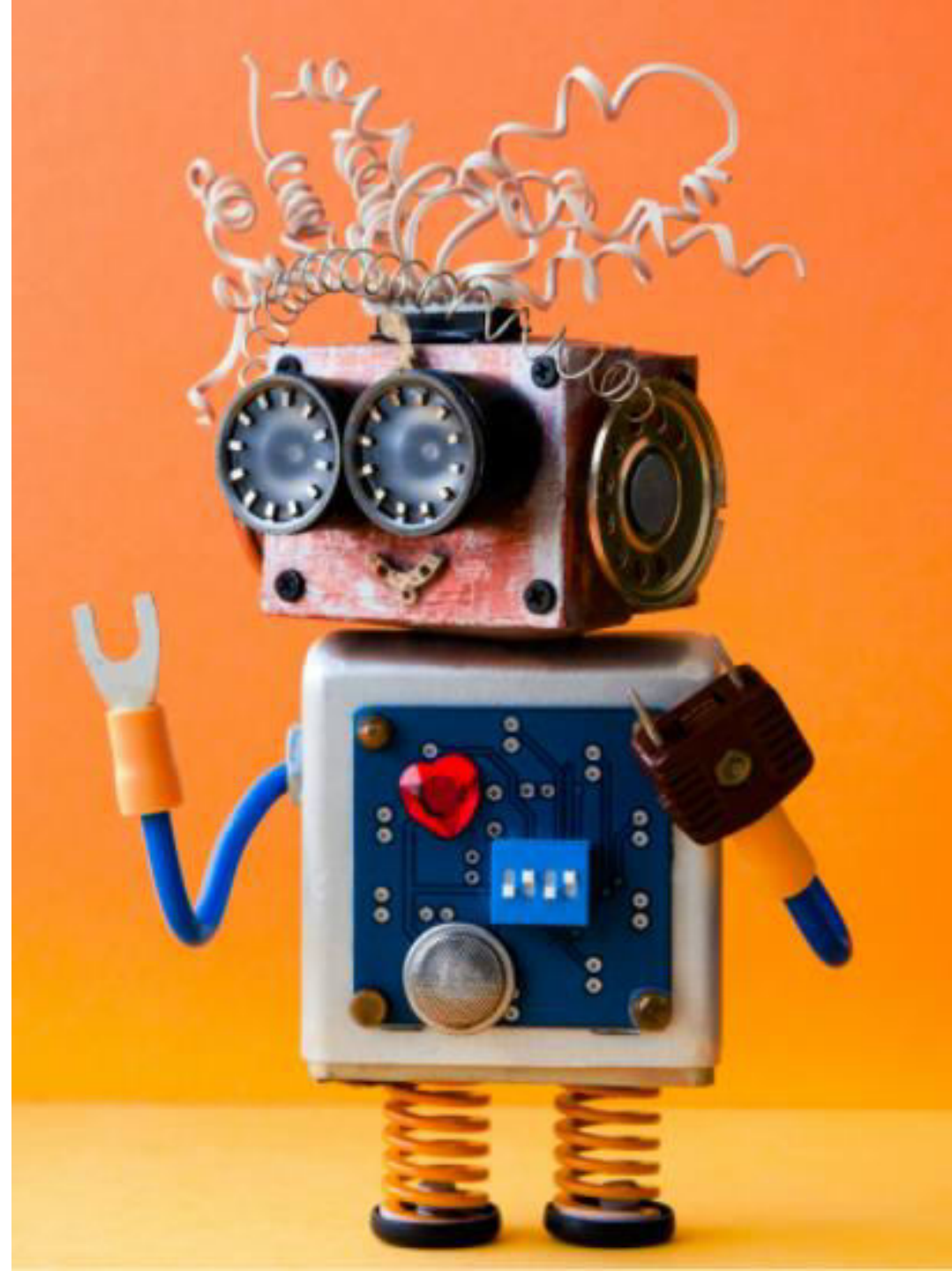


DELETE ENDPOINT [IMPORTANT]

The screenshot shows the Amazon SageMaker console interface. The left sidebar contains a navigation menu with the following items: **Notebook instances**, Lifecycle configurations, Git repositories, **Processing**, Processing jobs, **Training**, Algorithms, Training jobs, Hyperparameter tuning jobs, **Inference**, Compilation jobs, Model packages, Models, Endpoint configurations, **Endpoints** (highlighted with a red arrow), and Batch transform jobs. The main content area is titled 'Amazon SageMaker > Endpoints'. It features a search bar labeled 'Search endpoints', a refresh button, an 'Update endpoint' button, an 'Actions' dropdown menu (highlighted with a red arrow), and a 'Create endpoint' button. Below these controls is a table with the following data:

	Name	ARN	Status
<input type="radio"/>	insurance-prediction-endpoint	arn:aws:sagemaker:us-east-1:126821927778:endpoint/insurance-prediction-endpoint	✓ InService

FINAL END-OF-DAY CAPSTONE PROJECT



PROJECT OVERVIEW

- *The goal of this project is to use AWS SageMaker Autopilot to build, train, test machine learning models to predict bike rental usage using inputs such as temperature, humidity, wind speed..etc.*
- *This project can be effectively used by bike rental shops to predict demand and expected future sales and understand key factors that contribute to generating revenue.*



Image Source: <https://pixabay.com/photos/bike-rental-bikes-rent-pay-284380/> <https://www.flickr.com/photos/pasa/6757993805>

Dataset Source: Hadi Fanaee-T, Laboratory of Artificial Intelligence and Decision Support (LIAAD), University of Porto INESC
Porto, Campus da FEUP Rua Dr. Roberto Frias, 378 4200 - 465 Porto, Portugal <https://www.kaggle.com/janjughazyan/cars1>

MODEL OVERVIEW

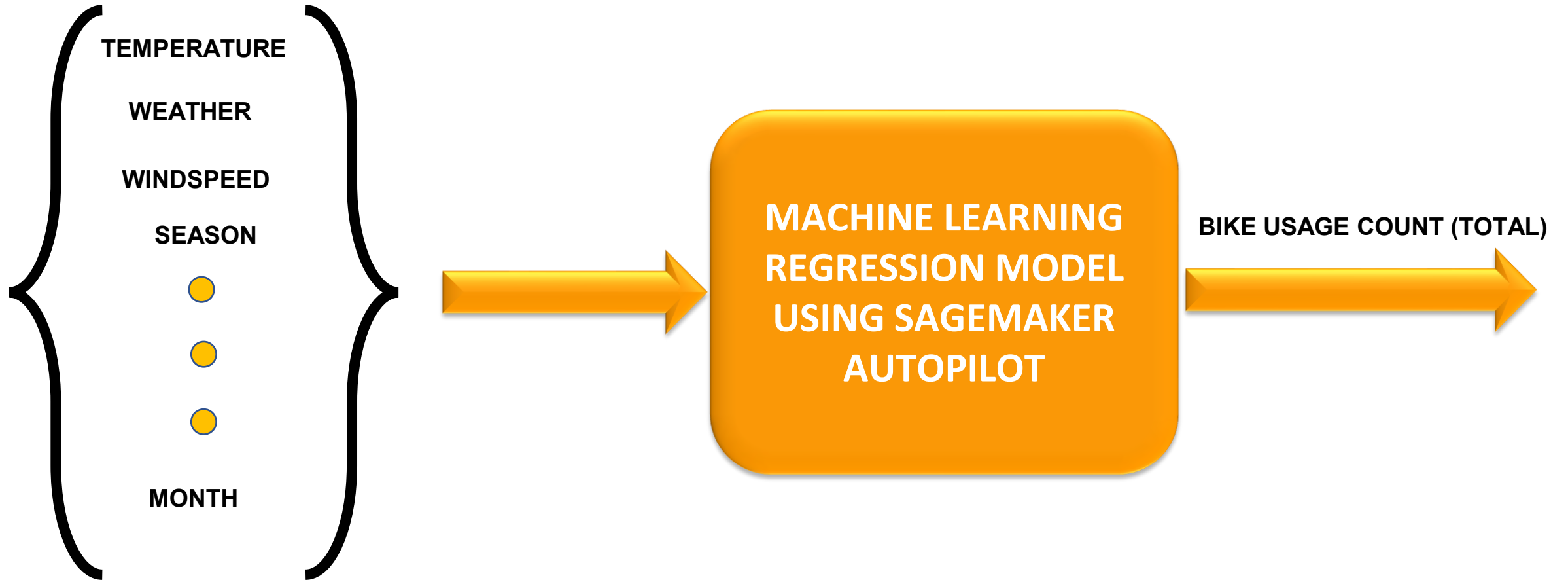


Photo Credit: https://commons.wikimedia.org/wiki/File:Neural_network.svg

PROJECT TASKS

Please complete the following:

1. Load the “*bike_sharing_daily.csv*” dataset
2. Train a machine learning models using AWS SageMaker Autopilot
3. Assess trained models' performance
4. Explore the candidate notebooks
5. Deploy the best model as an endpoint