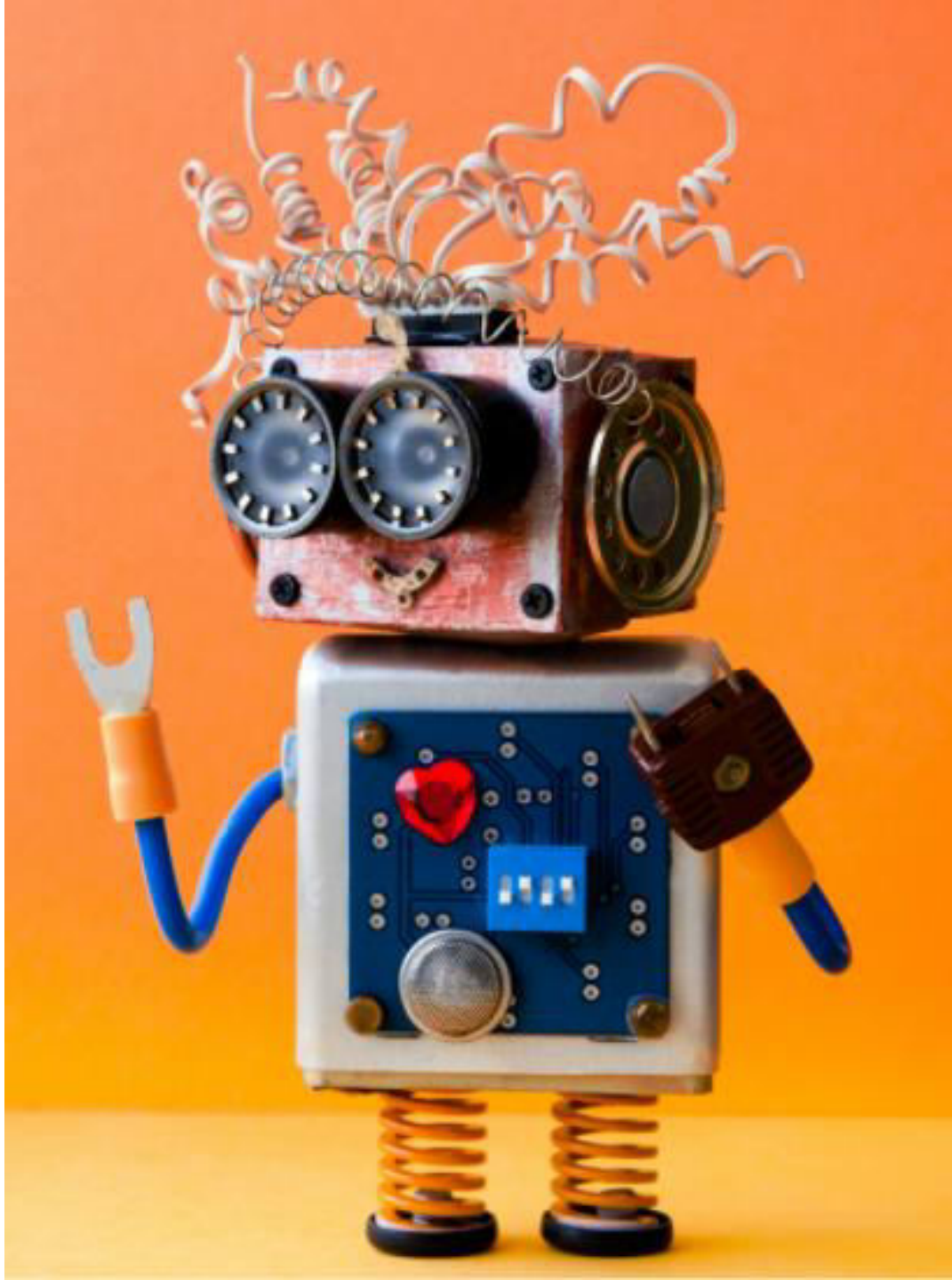
PROJECT CARD [SKIP IF FAMILIAR]

EASY ADVANCED



INTRODUCTION AND KEY LEARNING OUTCOMES

- We will an XG-Boost SageMaker Built-in algorithm using SageMaker JumpStart.
- We will learn how to:
  1. Train an XG-boost algorithm in SageMaker to predict university admission using AWS JumpStart
  2. Train an XG-boost algorithm in SageMaker to predict life expectancy (capstone project) using AWS JumpStart
  3. List XG-Boost hyperparameters
  4. Assess trained regression models performance
  5. Plot the residuals
  6. Deploy an endpoint and perform inference

PROJECT CARD

GOAL:

- *Build, train, test and deploy a machine learning model to predict chances of university admission into a particular university* given student's profile.

TOOL:

- *AWS SageMaker Studio – Jump Start*

PRACTICAL REAL-WORLD APPLICATION:

- *This project can be effectively used by university admission departments to determine top qualifying students.*

DATA:

INPUTS (FEATURES):

- *GRE Scores (out of 340)*
- *TOEFL Scores (out of 120)*
- *University Rating (out of 5)*
- *Statement of Purpose (SOP)*
- *Letter of Recommendation (LOR) Strength (out of 5)*
- *Undergraduate GPA (out of 10)*
- *Research Experience (either 0 or 1)*

OUTPUTS:

- *Chance of admission (ranging from 0 to 1)*

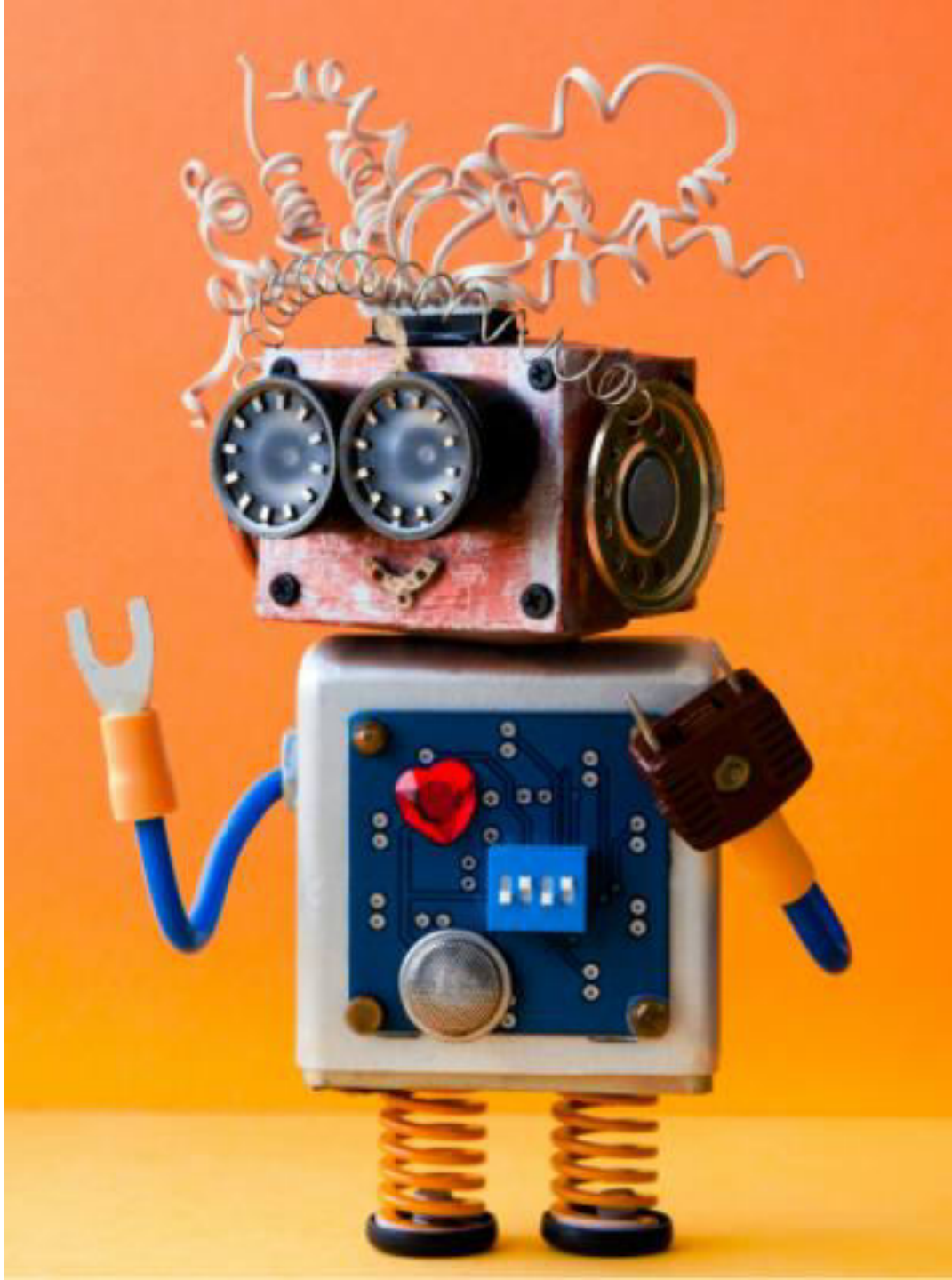
- Data Source: https://www.kaggle.com/mohansacharya/graduate-admissions
- Photo Credit: https://www.pexels.com/photo/accomplishment-ceremony-education-graduation-267885/

https://www.flickr.com/photos/pasa/6767993805

https://www.kaggle.com/ljanjughazyan/cars1

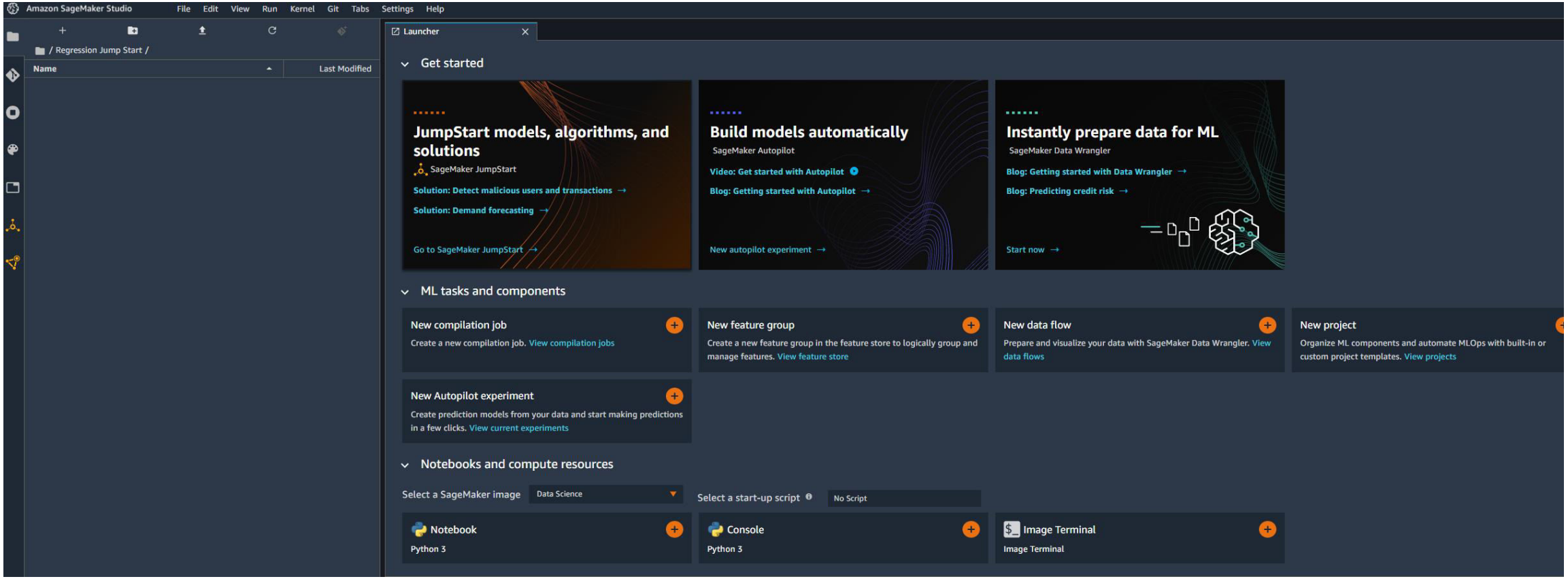
JUMPSTART DEMO: OVERVIEW

EASY ADVANCED



JUMP START DEMO

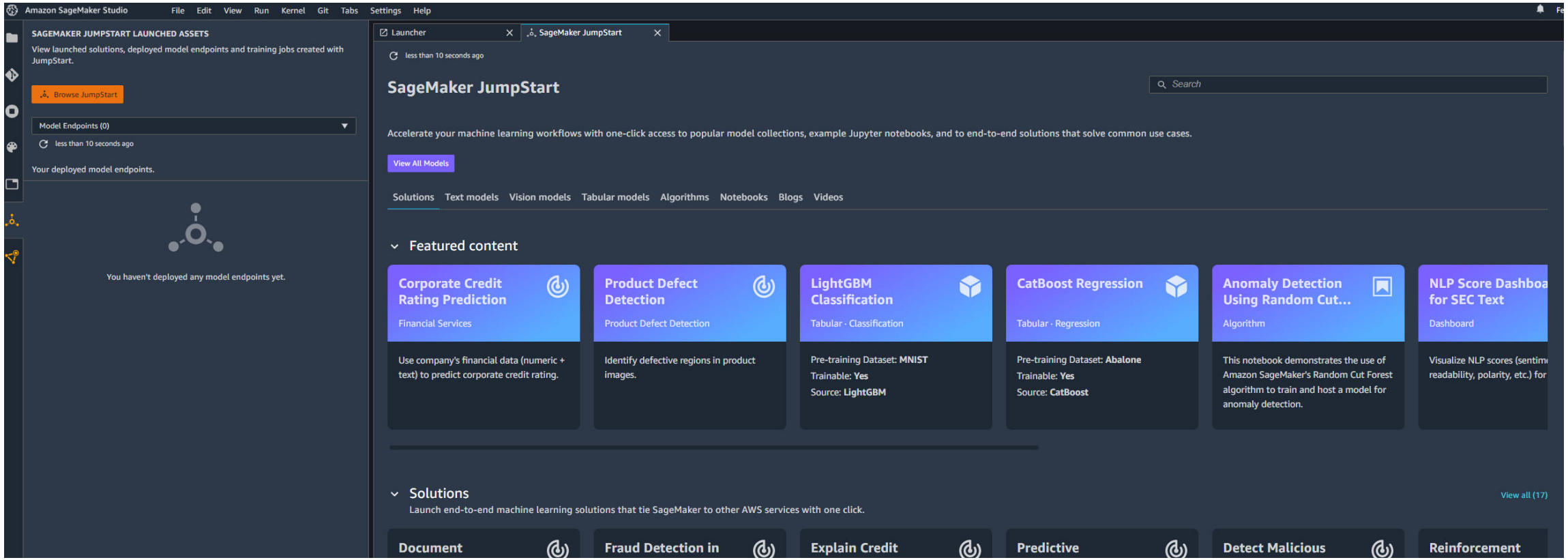
NAVIGATE TO AMAZON SAGEMAKER STUDIO HOMEPAGE AND CLICK ON SAGEMAKER JUMPSTART



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

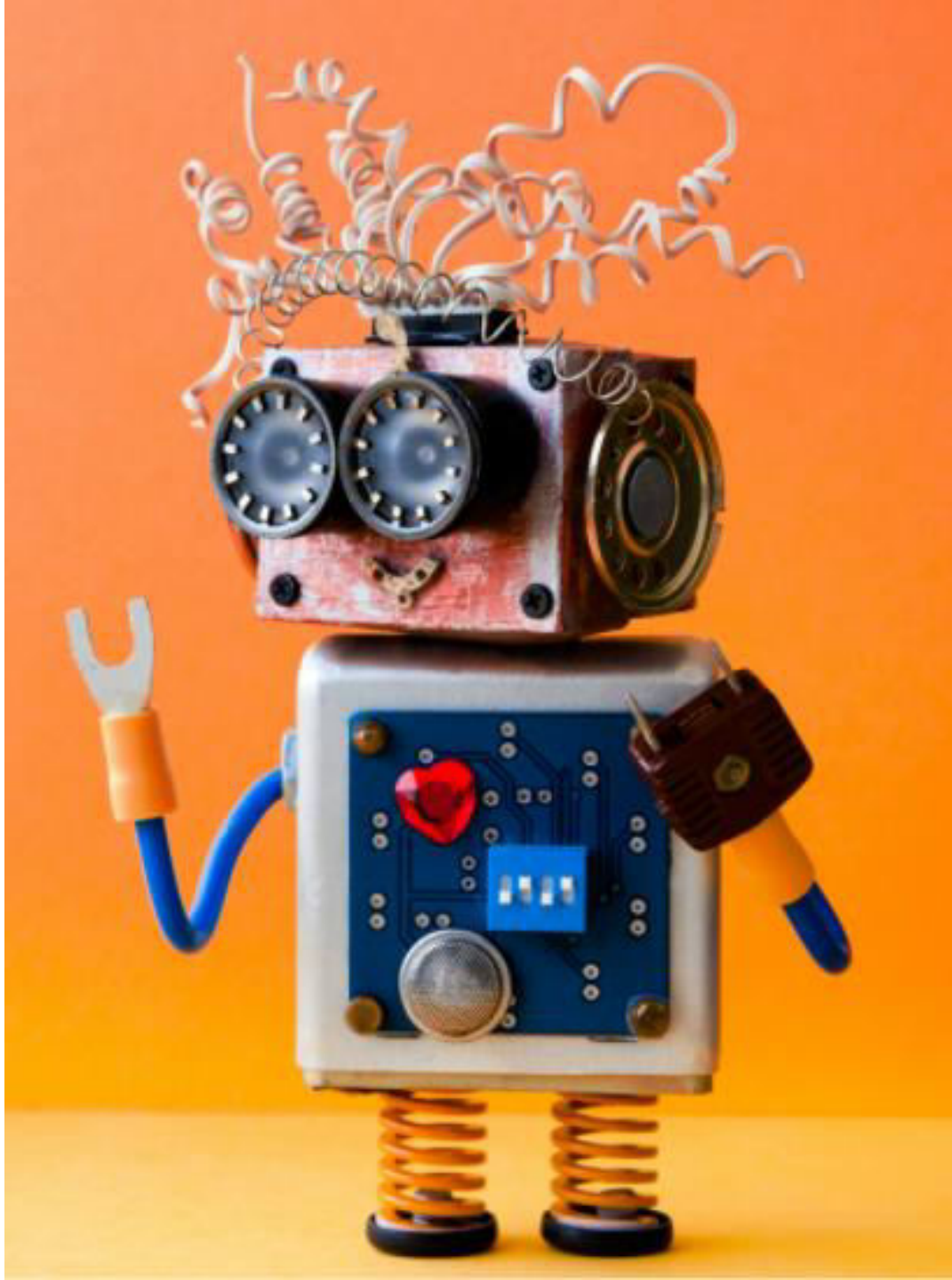
YOU SHOULD FIND SEVERAL ML WORKFLOWS THAT CAN BE DEPLOYED WITH ONE CLICK ACCESS. THERE ARE VISION, TEXT, AND TABULAR MODELS..ETC.



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

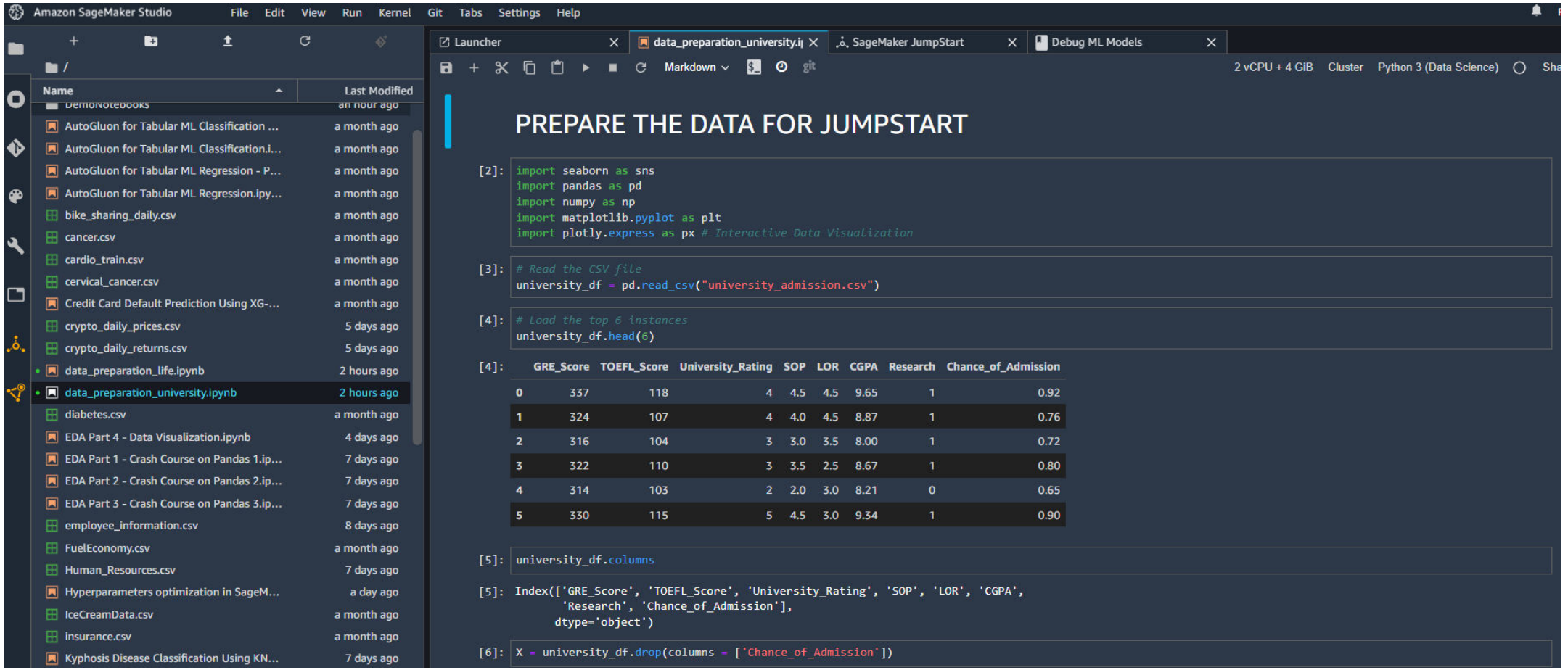
JUMPSTART DEMO: DATA PREP PRIOR TO TRAINING

EASY ADVANCED



DATA PREP

IN THIS NOTEBOOK, WE WILL SPLIT THE DATA INTO TRAINING, VALIDATION AND TESTING. WE WILL ALSO SET THE FIRST COLUMN IN THE DATA TO BE OUR TARGET (OUTPUT) COLUMN.

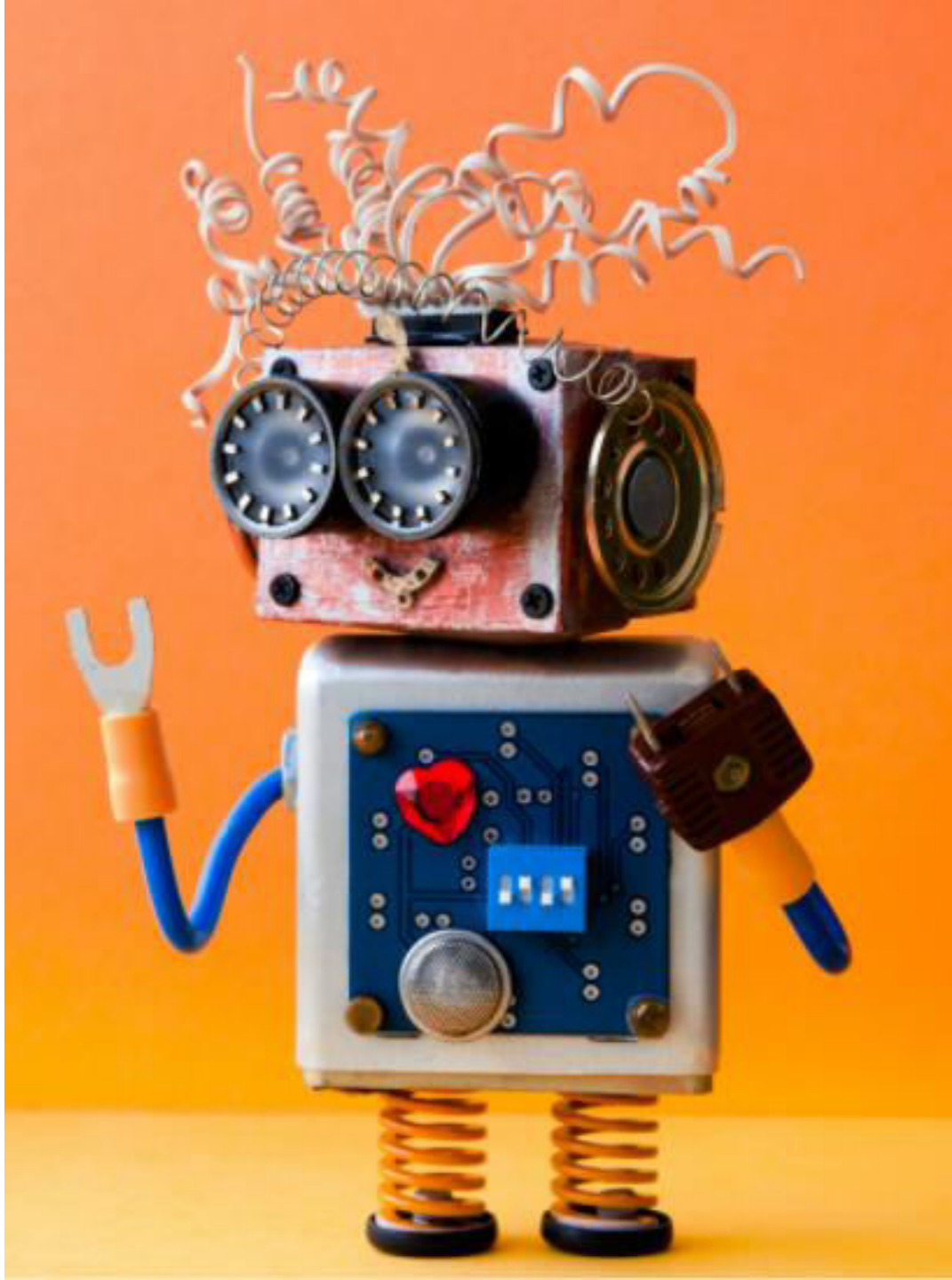


https://www.kaggle.com/ljanjughazyan/cars1

JUMPSTART DEMO: DATA UPLOAD

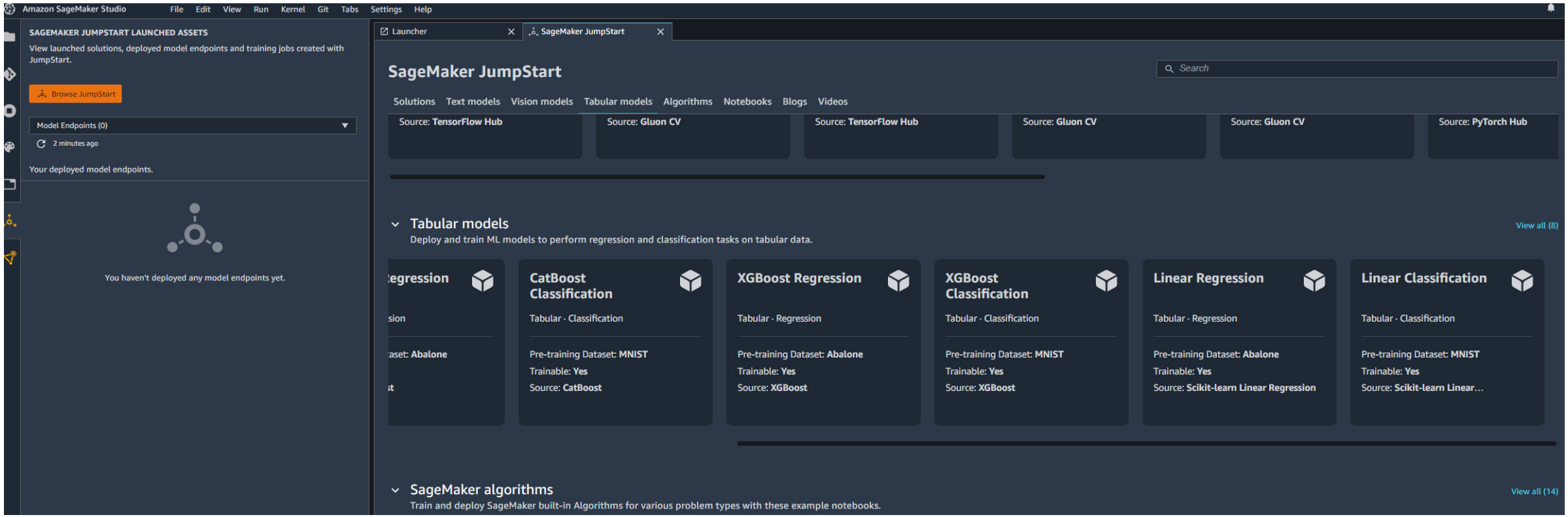
EASY

ADVANCED



JUMP START DEMO

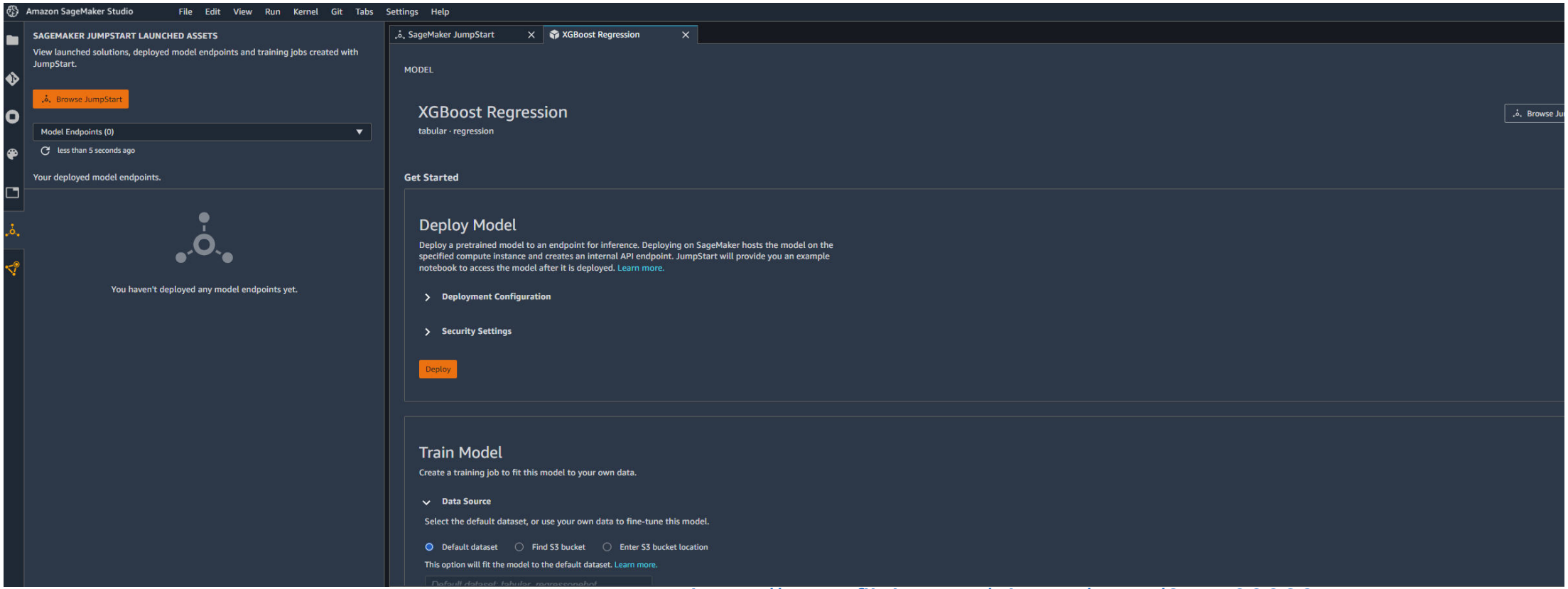
CLICK ON XGBOOST REGRESSION



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

YOU CAN DEPLOY A PRETRAINED MODEL (MODEL TRAINED ON THE ABALONE DATASET) OR YOU CAN TRAIN THE MODEL WITH YOUR OWN CUSTOM DATASET

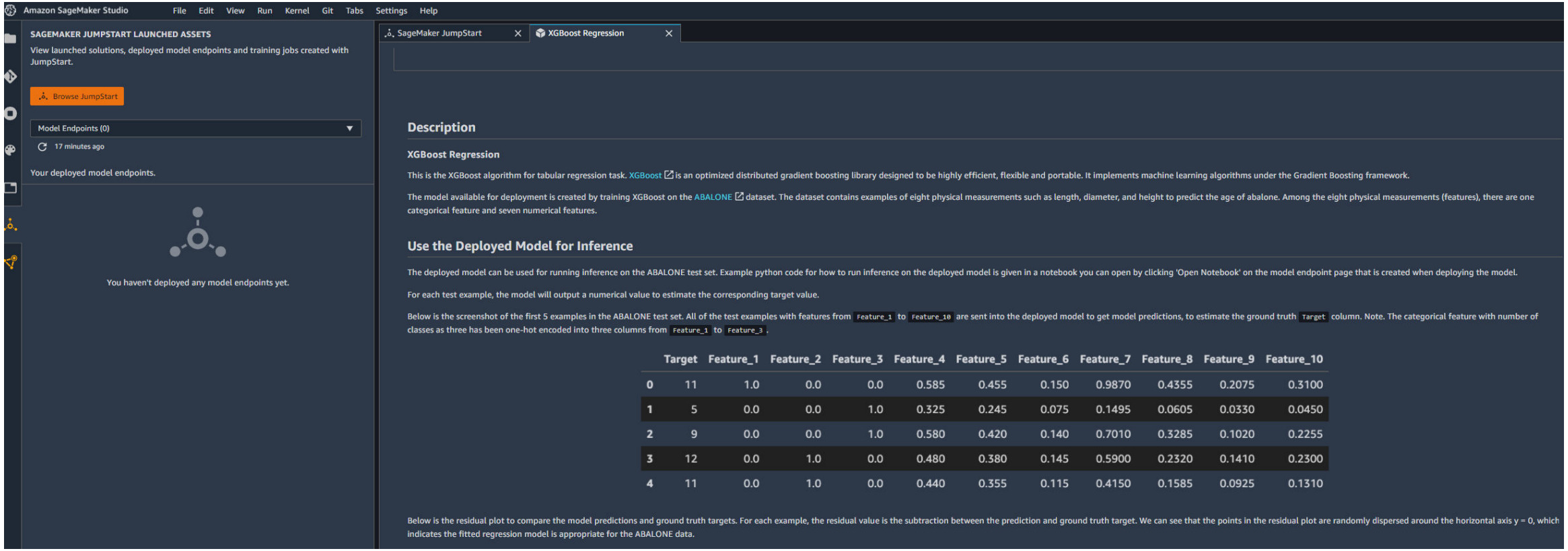


https://www.flickr.com/photos/pasa/6757993805

https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

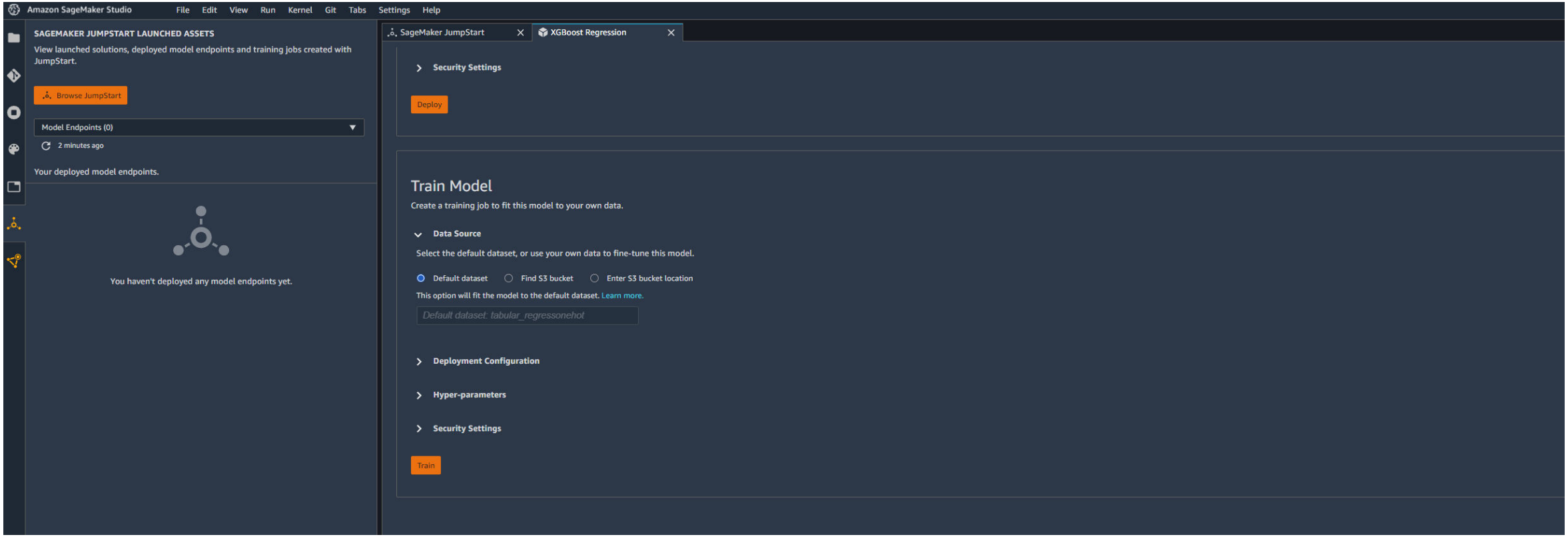
YOU CAN READ THE DESCRIPTION BELOW.



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

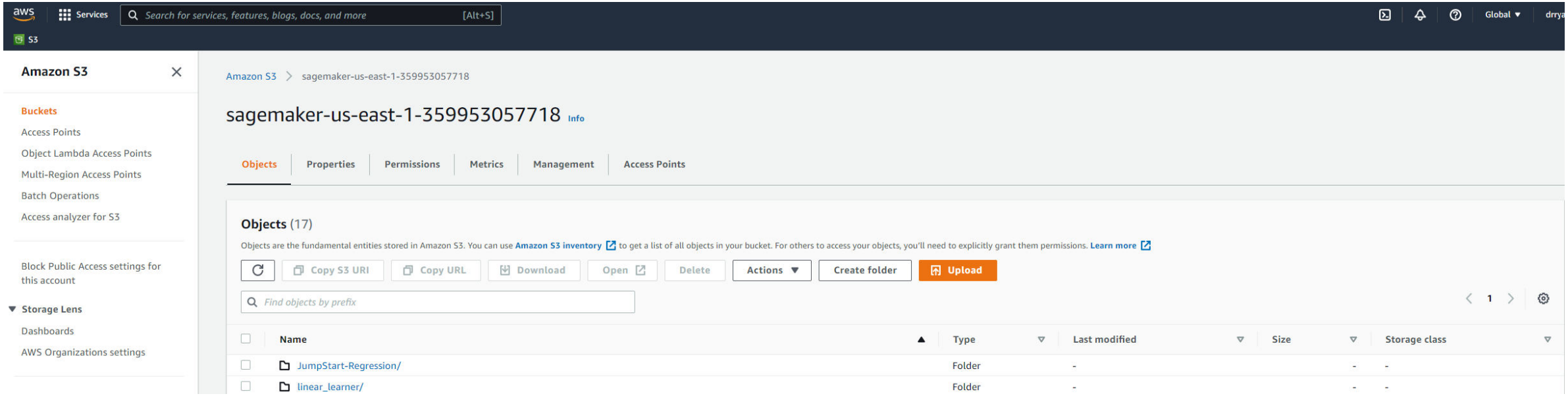
JUMP START DEMO

- INPUT: A DIRECTORY CONTAINS 2 SUB-DIRECTORY 'TRAIN/' 'VALIDATION/'.
- EACH SUB-DIRECTORY CONTAINS A 'DATA.CSV' FILE. THE 'DATA.CSV' FILES UNDER SUB-DIRECTORY 'TRAIN/' AND 'VALIDATION/' ARE FOR TRAINING/VALIDATION.
- THE FIRST COLUMN OF THE 'DATA.CSV' SHOULD HAVE THE CORRESPONDING TARGET VARIABLE AND THE REST CONTAINS THE FEATURES.
- YOU MUST ENCODE ALL CATEGORICAL FEATURES FIRST.
- OUTPUT: A TRAINED MODEL THAT CAN BE DEPLOYED FOR INFERENCE.



JUMP START DEMO

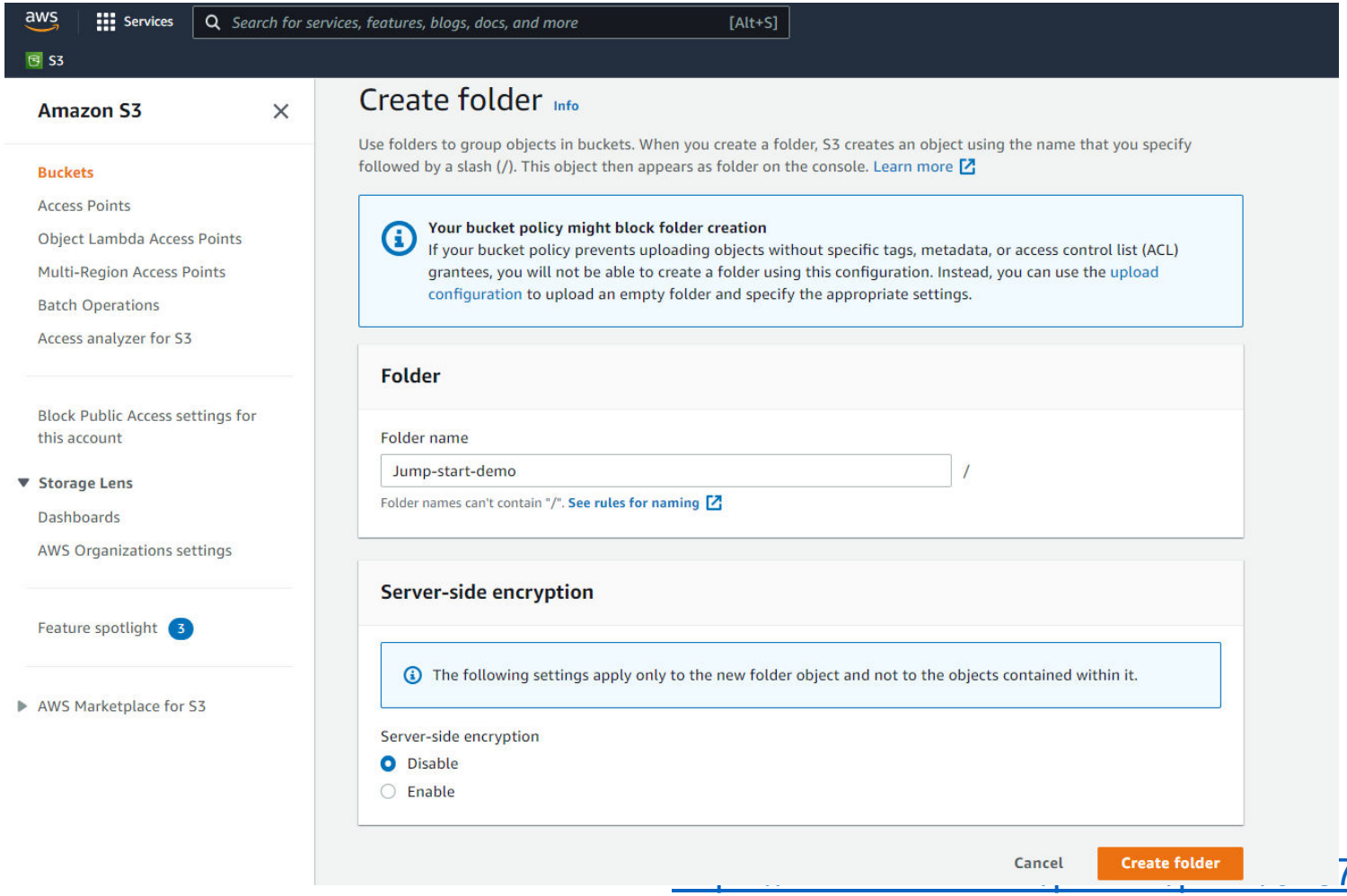
NAVIGATE TO S3, CREATE A DIRECTORY AND LIST THE TRAIN AND TEST DATASET



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

CREATE A NEW FOLDER IN THE DEFAULT SAGEMAKER BUCKET

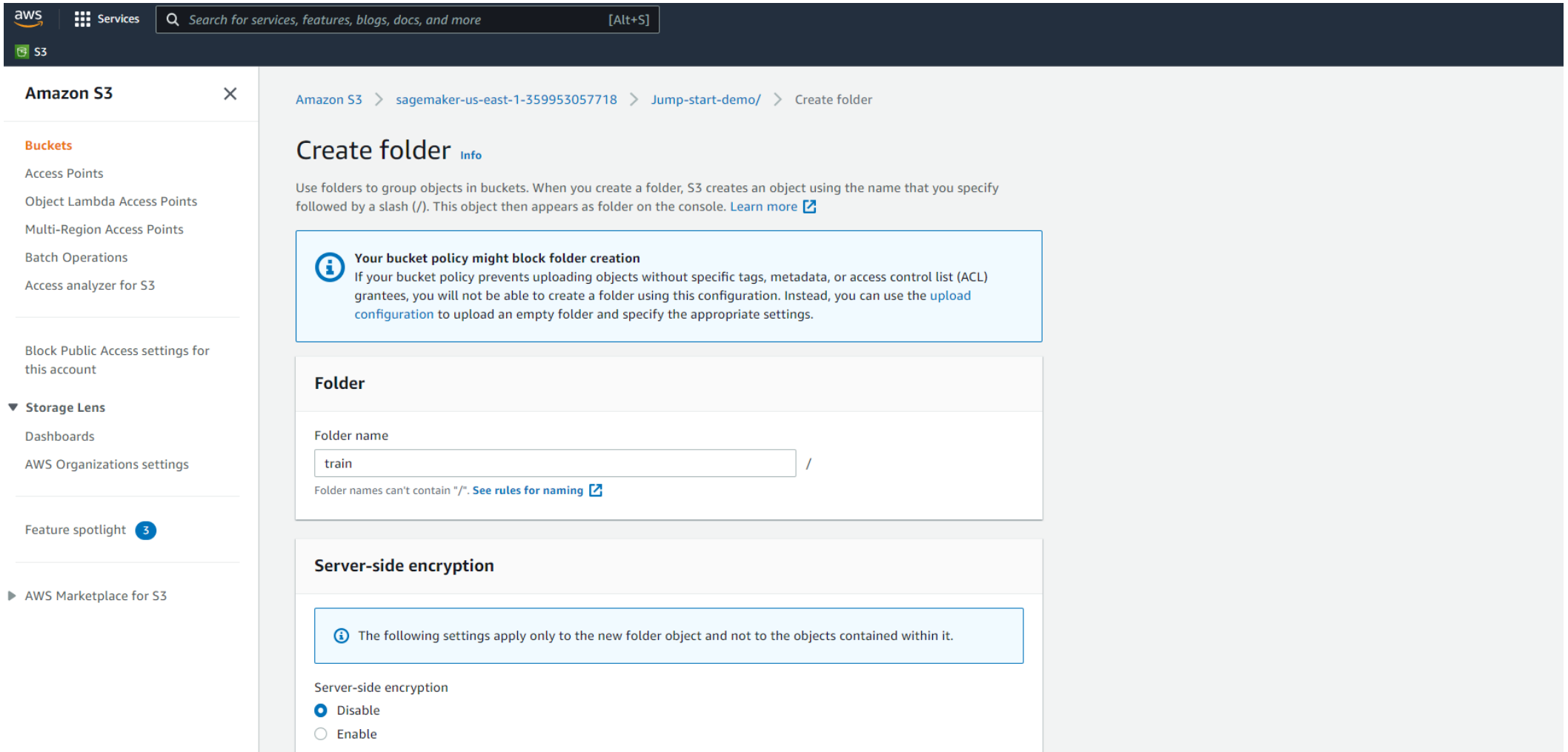


7993805

https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

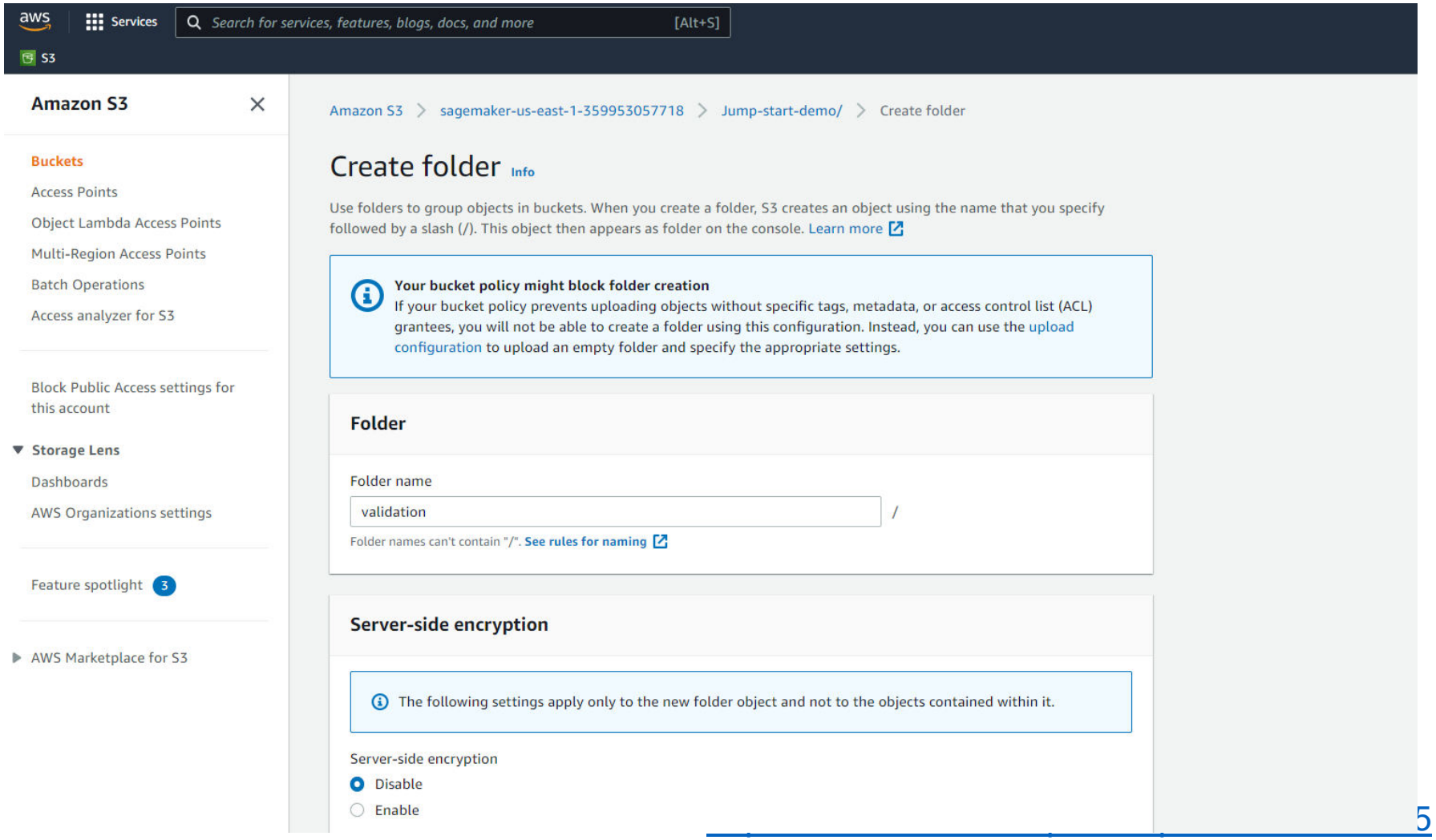
CREATE A SUB-DIRECTORY AND NAME IT TRAIN



https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

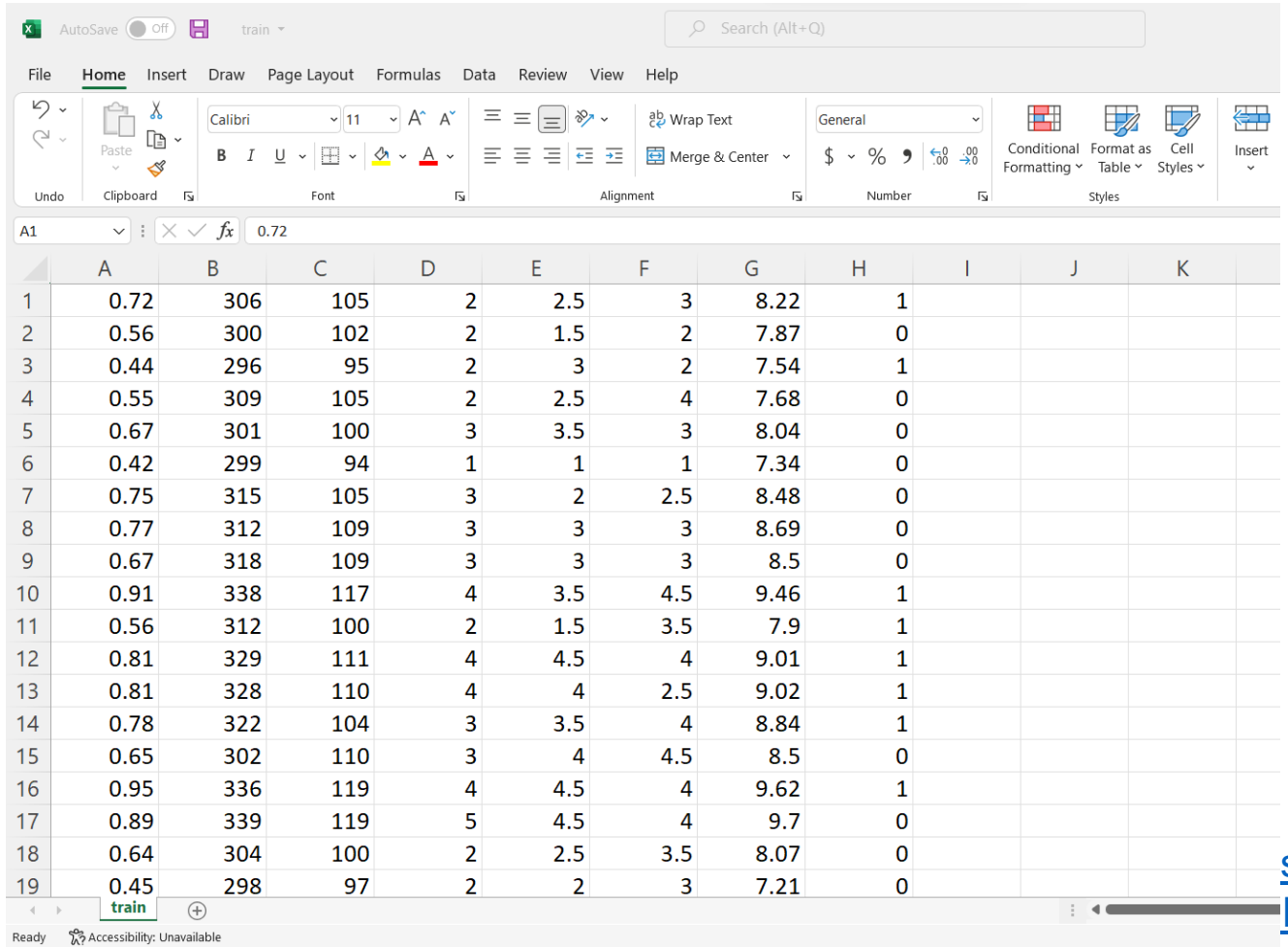
CREATE A SUB-DIRECTORY AND NAME IT VALIDATION



https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

UPLOAD THE TRAINING DATASET TO TRAIN DIRECTORY. NOTE THAT THE FOLLOWING: (1) NO HEADERS, (2) TARGET VARIABLE IS LISTED FIRST, (3) YOU WILL NEED TO RENAME THE FILE TO DATA.CSV AFTER IT'S UPLOADED TO S3

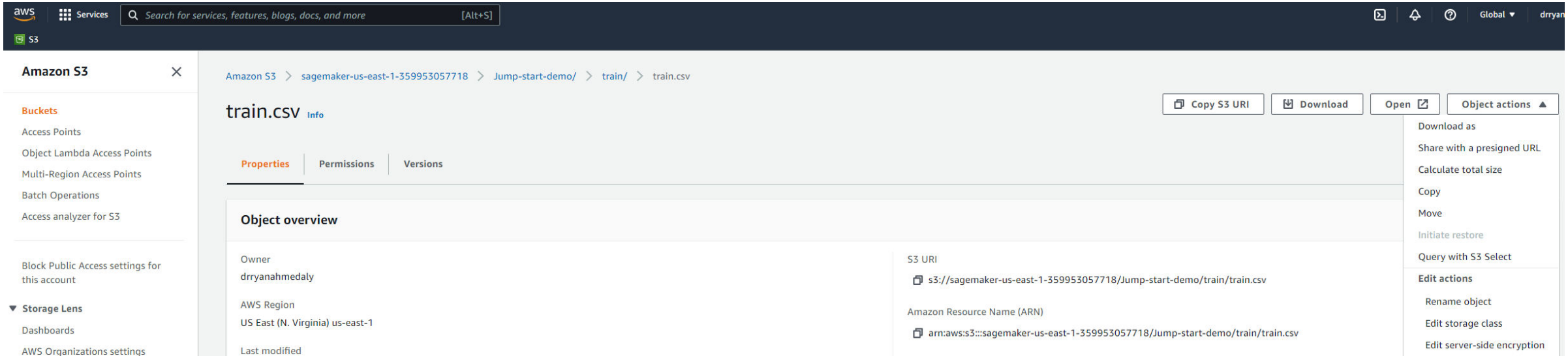


	A	B	C	D	E	F	G	H	I	J	K
1	0.72	306	105	2	2.5	3	8.22	1			
2	0.56	300	102	2	1.5	2	7.87	0			
3	0.44	296	95	2	3	2	7.54	1			
4	0.55	309	105	2	2.5	4	7.68	0			
5	0.67	301	100	3	3.5	3	8.04	0			
6	0.42	299	94	1	1	1	7.34	0			
7	0.75	315	105	3	2	2.5	8.48	0			
8	0.77	312	109	3	3	3	8.69	0			
9	0.67	318	109	3	3	3	8.5	0			
10	0.91	338	117	4	3.5	4.5	9.46	1			
11	0.56	312	100	2	1.5	3.5	7.9	1			
12	0.81	329	111	4	4.5	4	9.01	1			
13	0.81	328	110	4	4	2.5	9.02	1			
14	0.78	322	104	3	3.5	4	8.84	1			
15	0.65	302	110	3	4	4.5	8.5	0			
16	0.95	336	119	4	4.5	4	9.62	1			
17	0.89	339	119	5	4.5	4	9.7	0			
18	0.64	304	100	2	2.5	3.5	8.07	0			
19	0.45	298	97	2	2	3	7.21	0			

[sa/6757993805
hazyan/cars1]

JUMP START DEMO

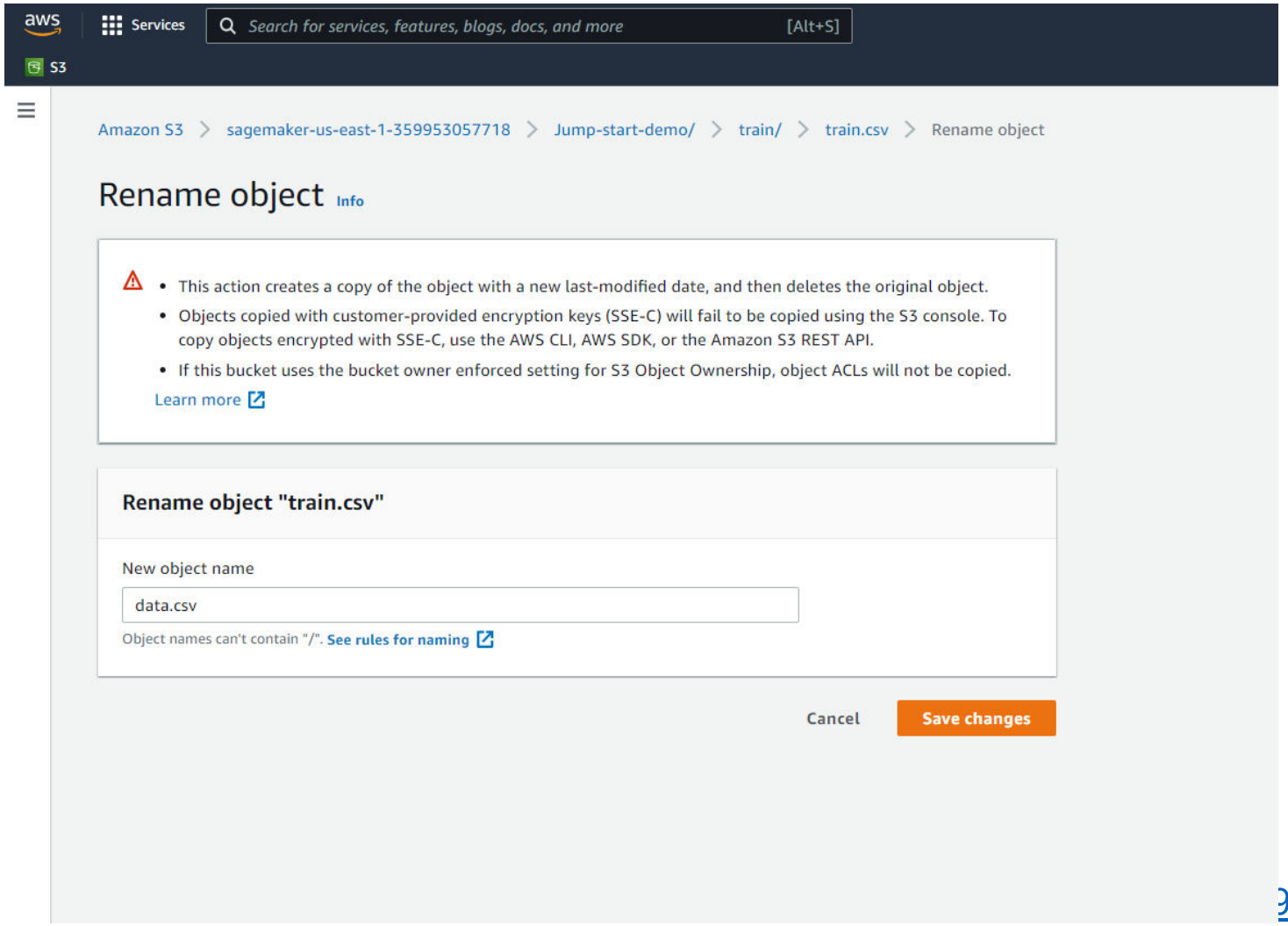
CLICK ON RENAME OBJECT



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

RENAMED IT FROM TRAIN.CSV TO DATA.CSV

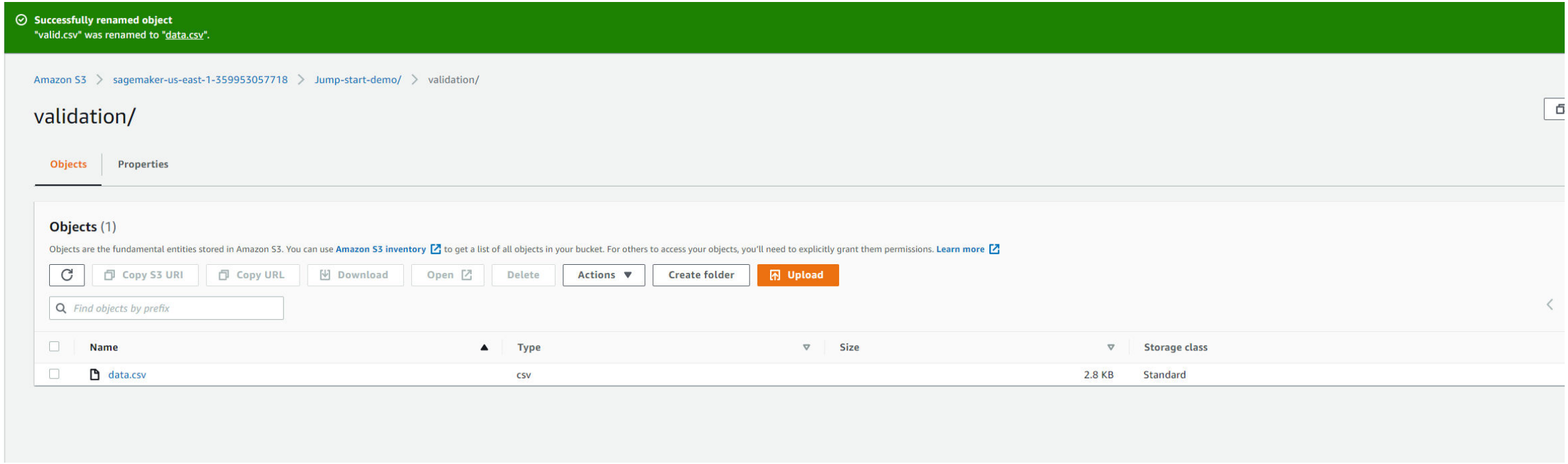


993805

https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

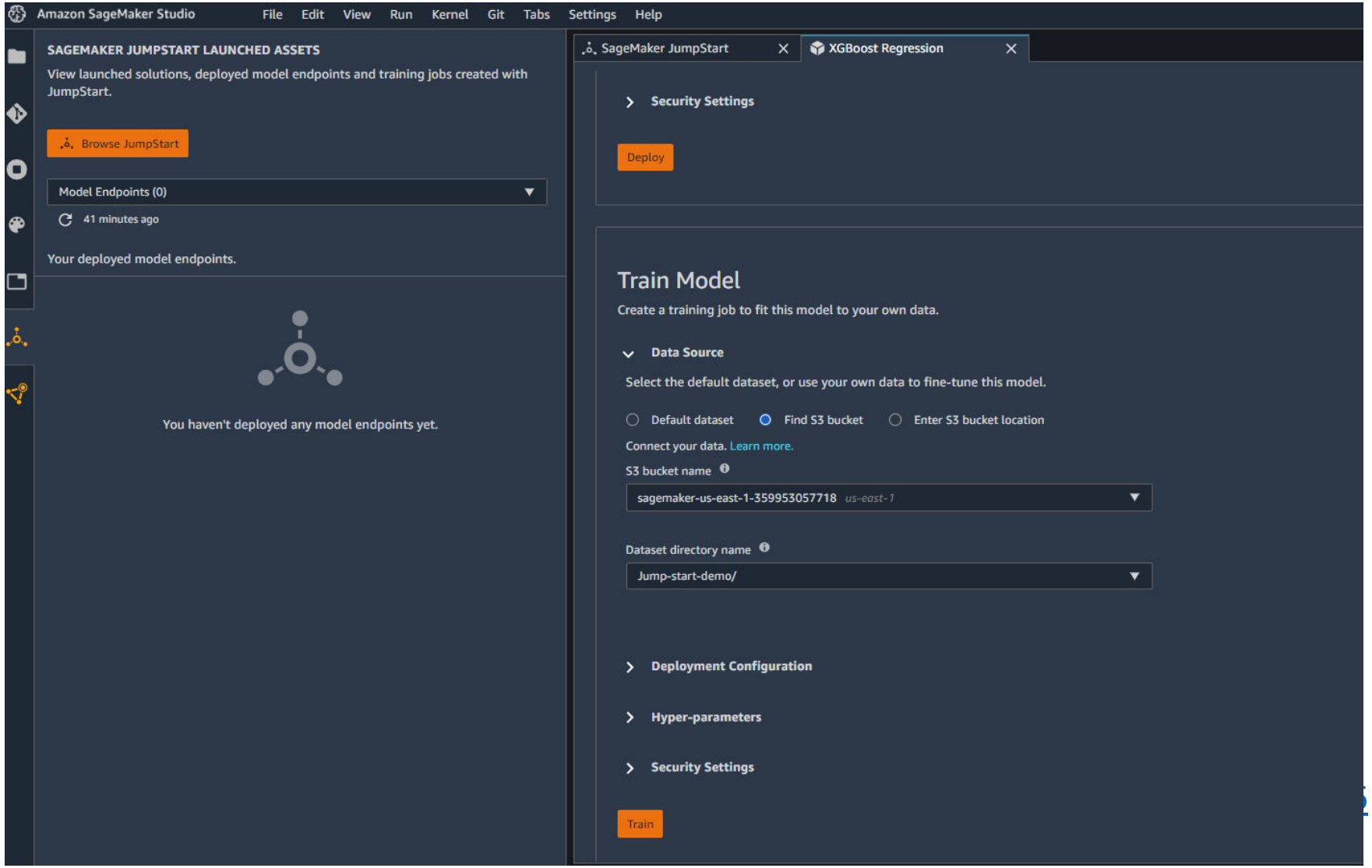
DO THE SAME FOR THE VALIDATION DATASET



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

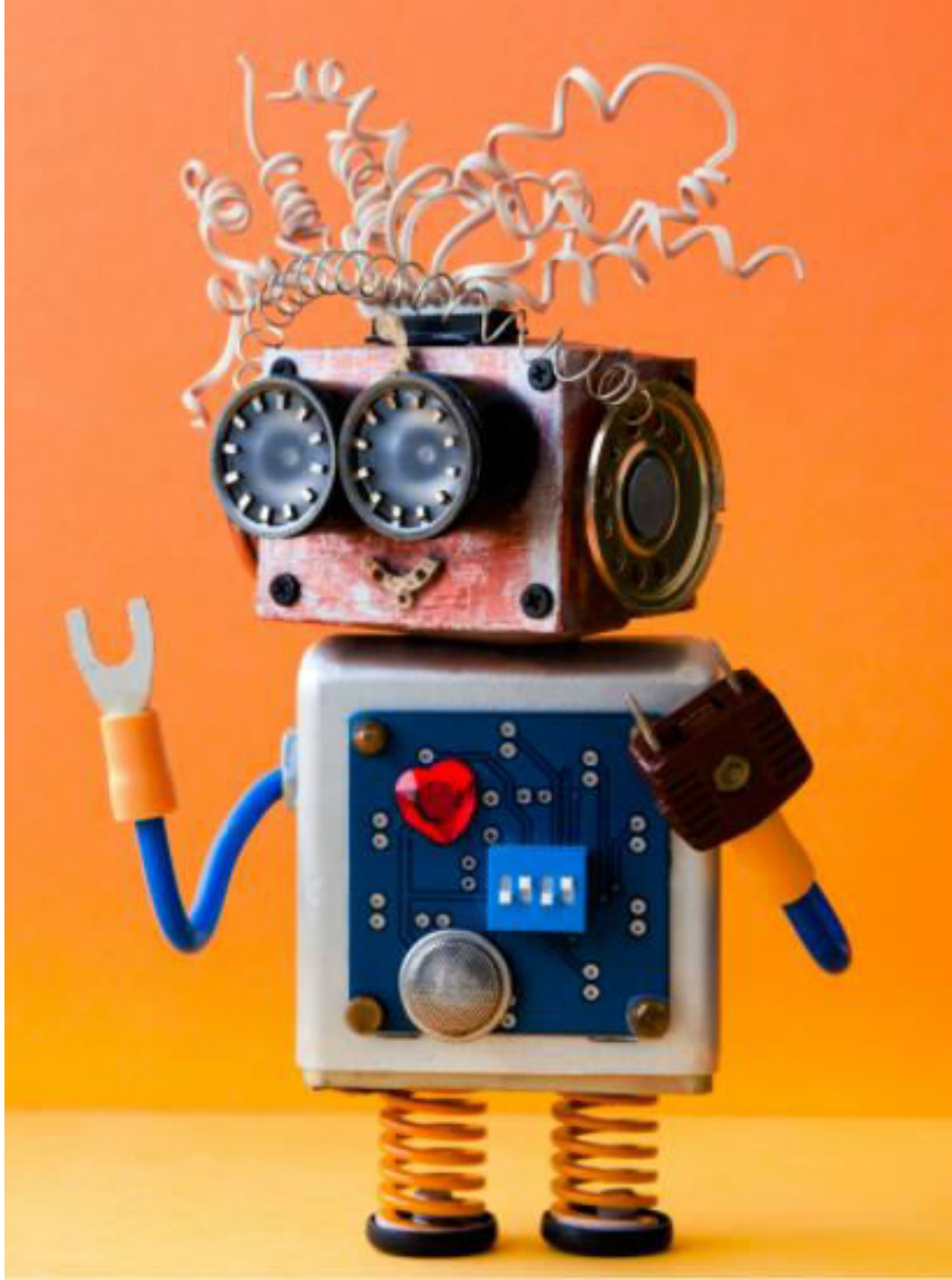
SELECT THE S3 BUCKET AND DATASET DIRECTORY NAME. GO AHEAD, CLICK TRAIN TO START THE TRAINING PROCESS



JUMPSTART DEMO: MODEL TRAINING

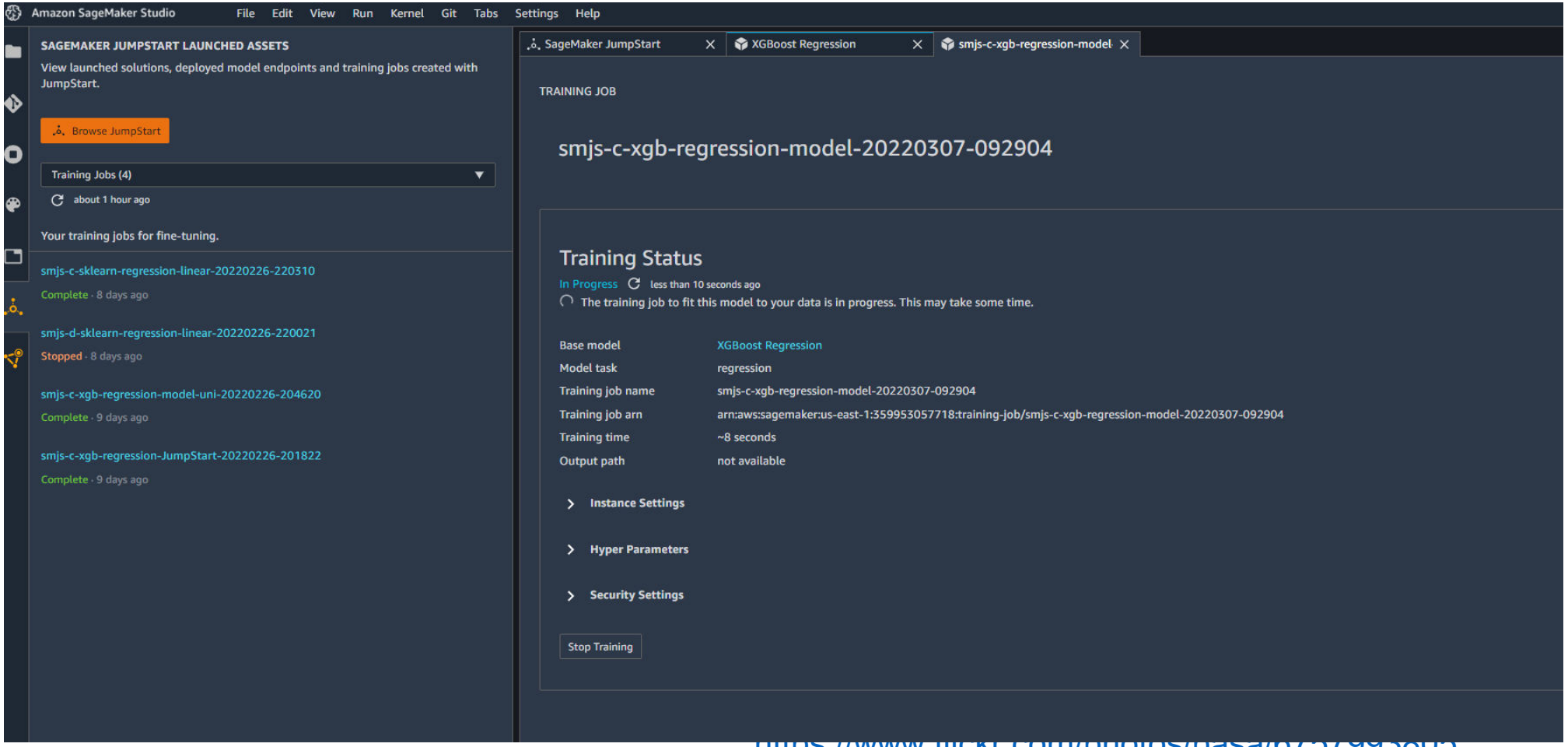
EASY

ADVANCED



JUMP START DEMO

TRAINING PROCESS HAS JUST STARTED!

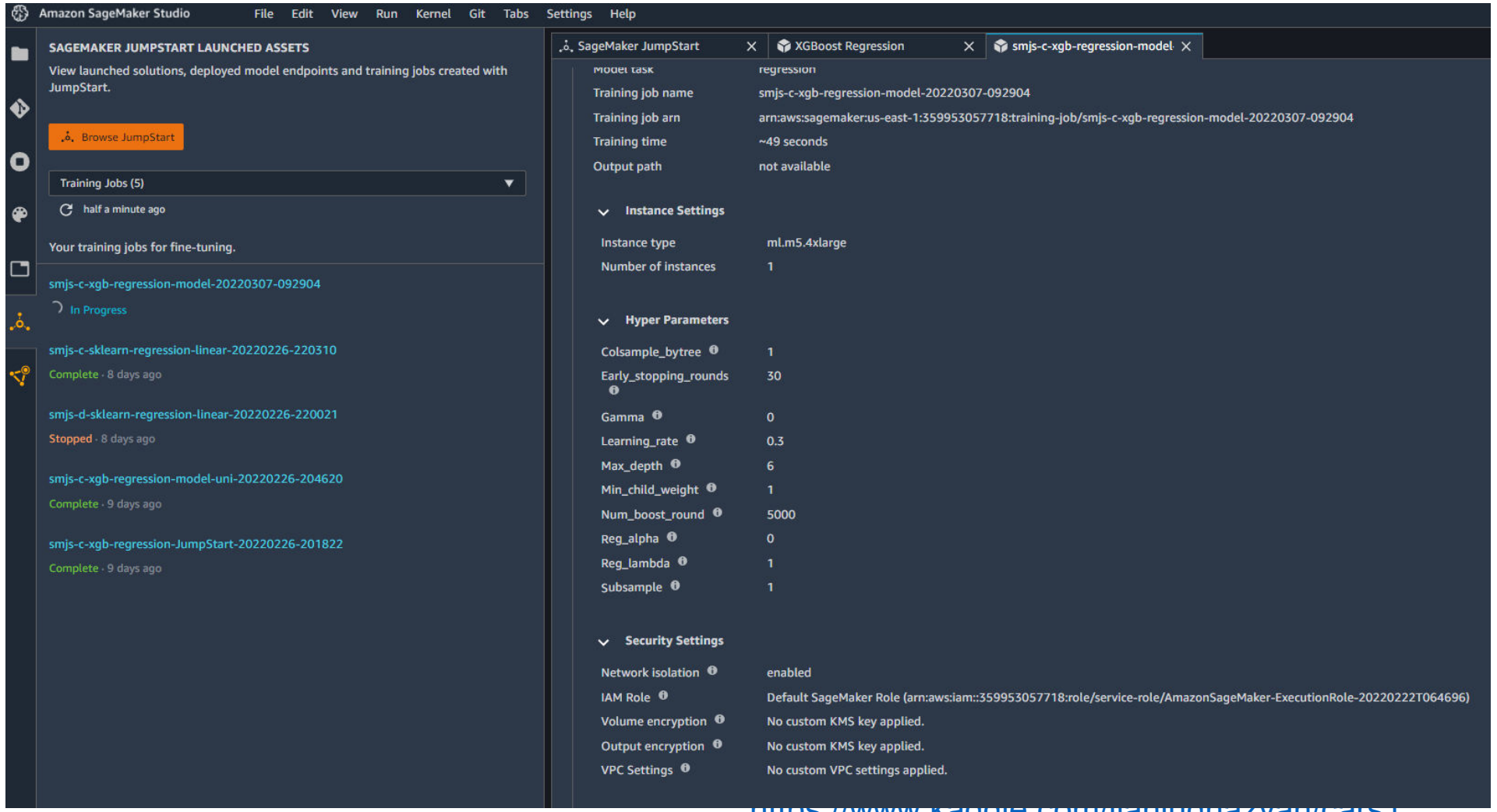


https://www.flickr.com/photos/pasa/6757993805

https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

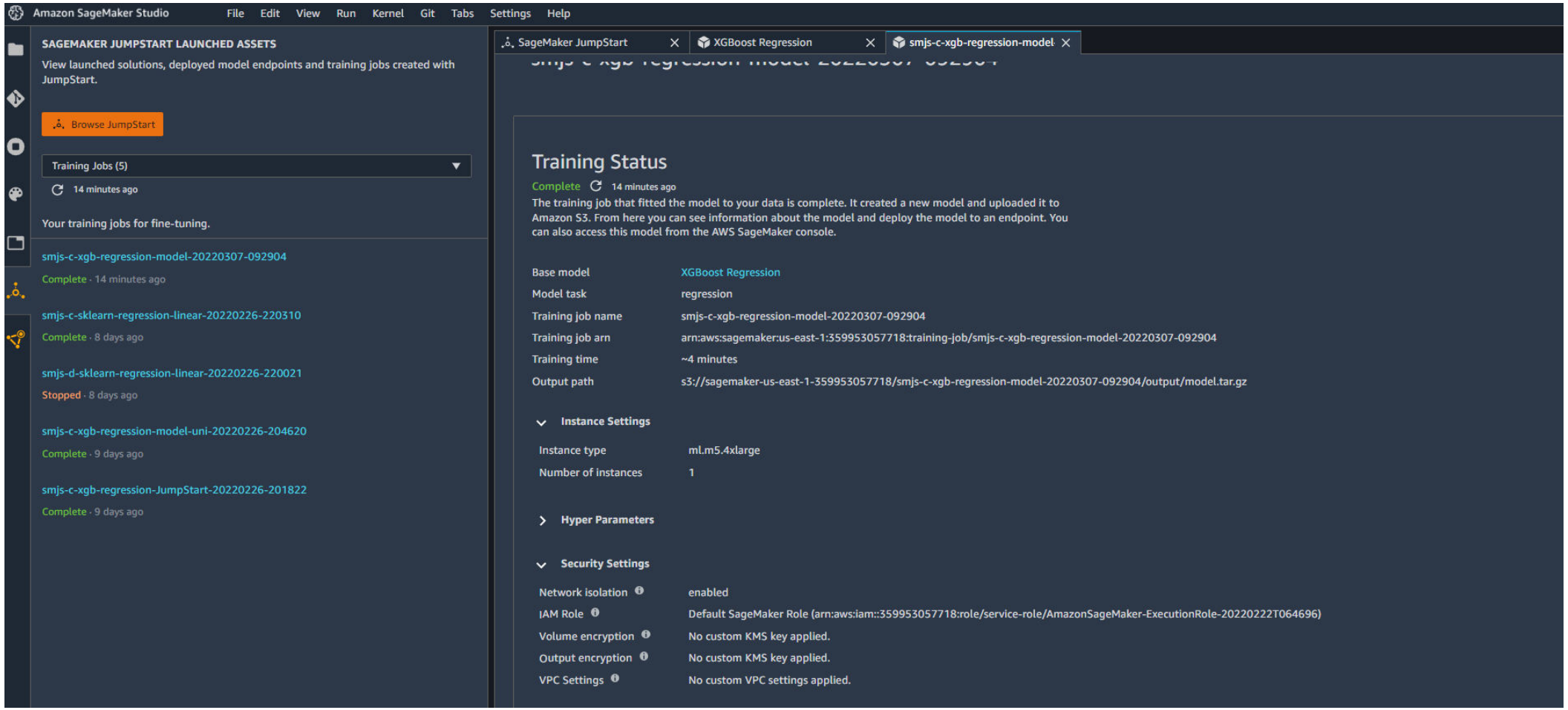
YOU CAN VIEW THE HYPERPARAMETERS AND INSTANCE SETTINGS



https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

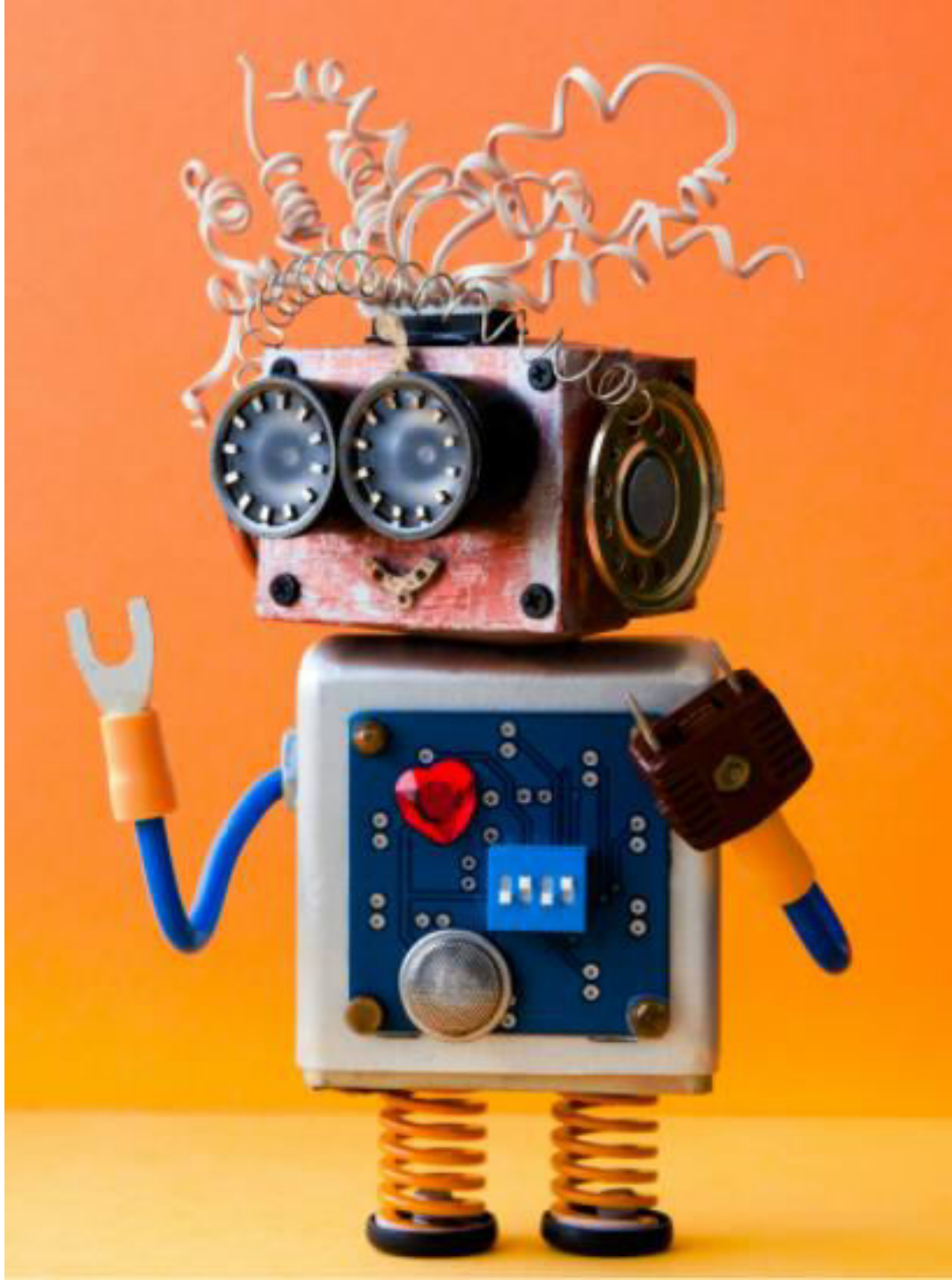
TRAINING IS NOW COMPLETE!



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

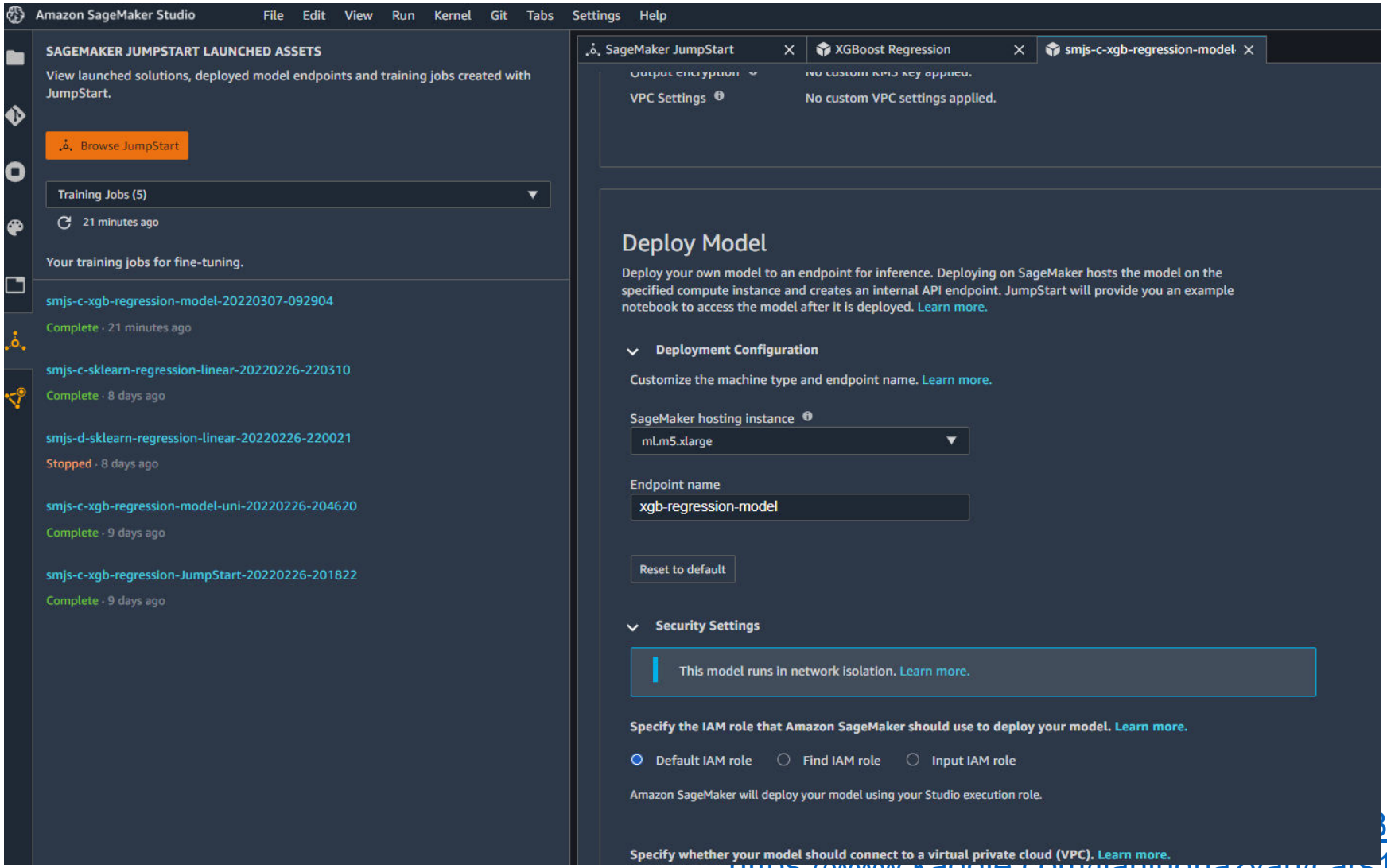
JUMPSTART DEMO: ENDPOINT DEPLOYMENT

EASY ADVANCED



JUMP START DEMO

MODEL DEPLOYMENT

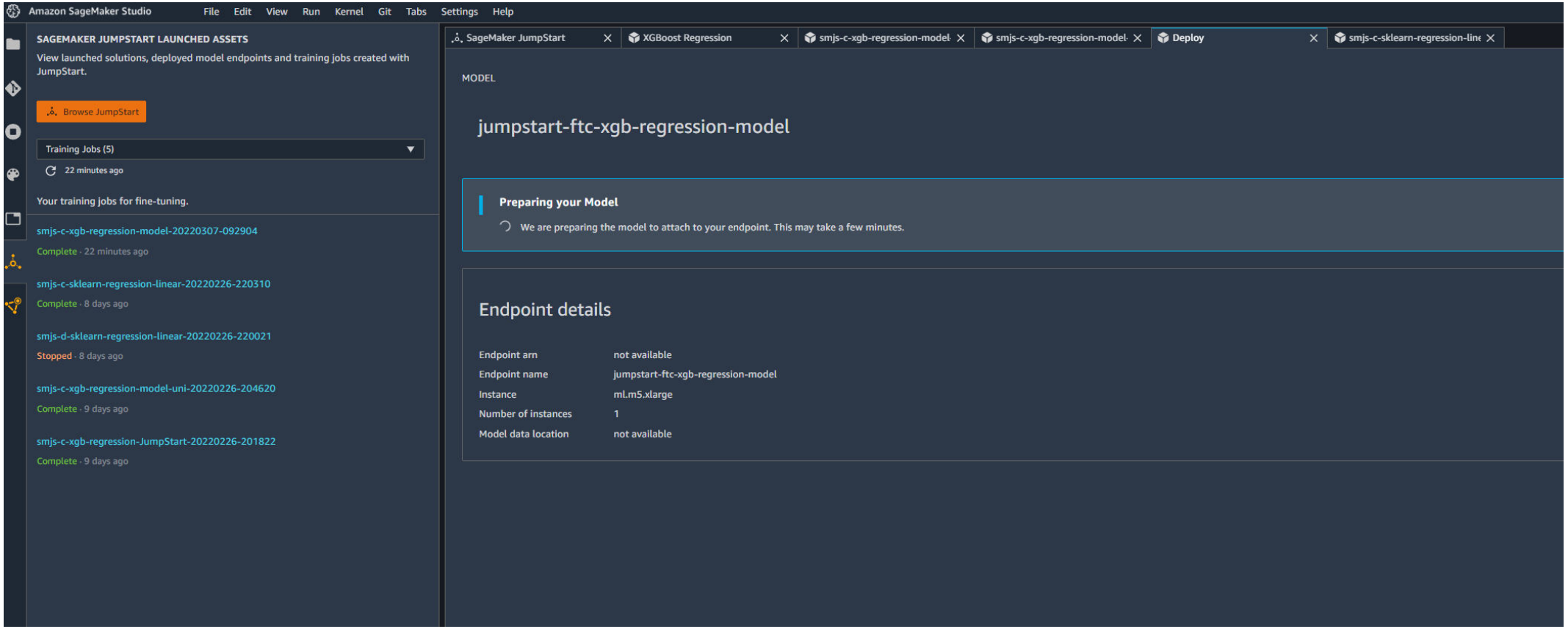


305

https://www.kaggle.com/ijanjughazyan/cars1

JUMP START DEMO

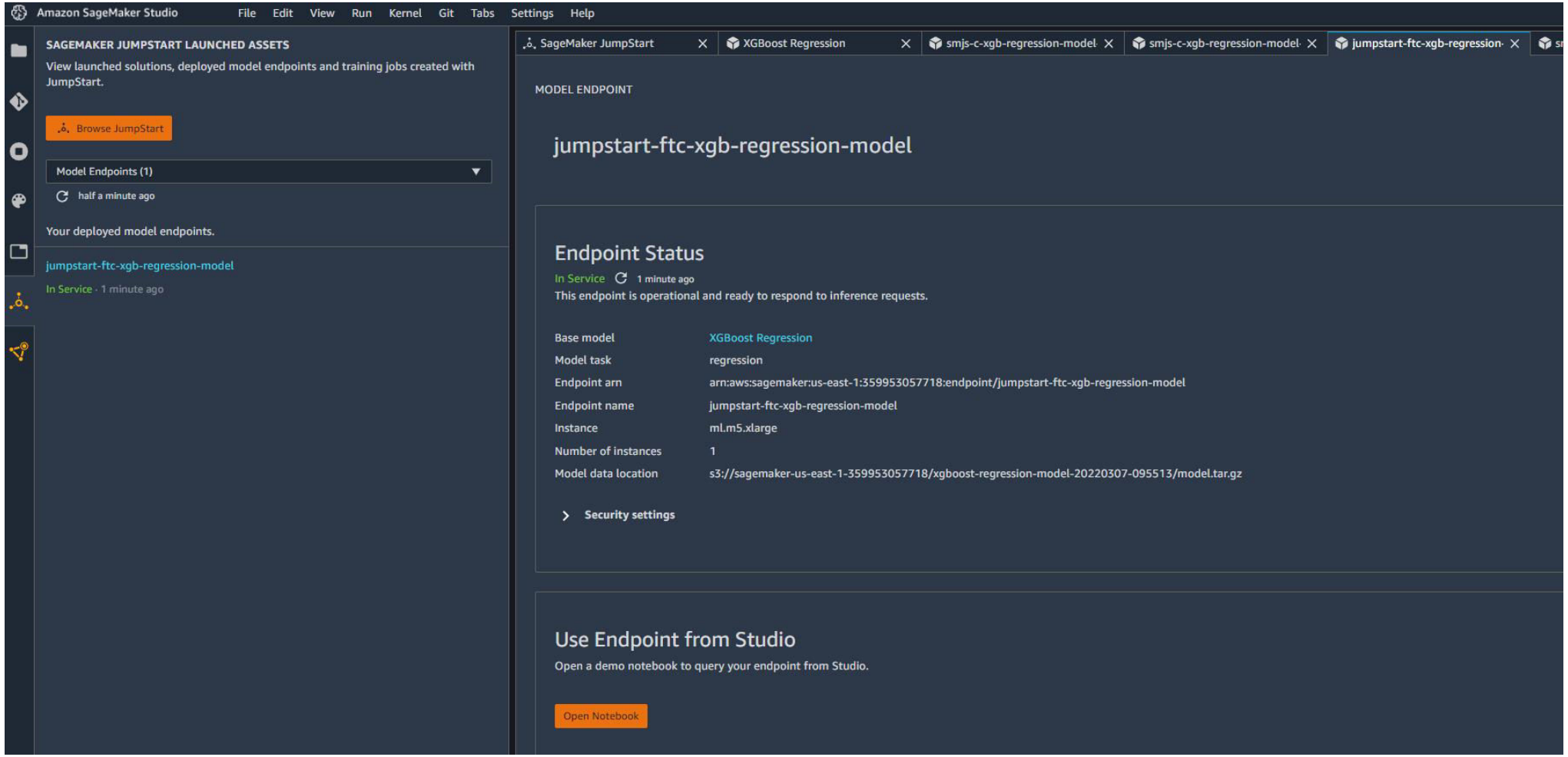
MODEL IS BEING DEPLOYED…THIS SHOULD TAKE COUPLE IF MINUTES.



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

NOW THE ENDPOINT IS IN-SERVICE

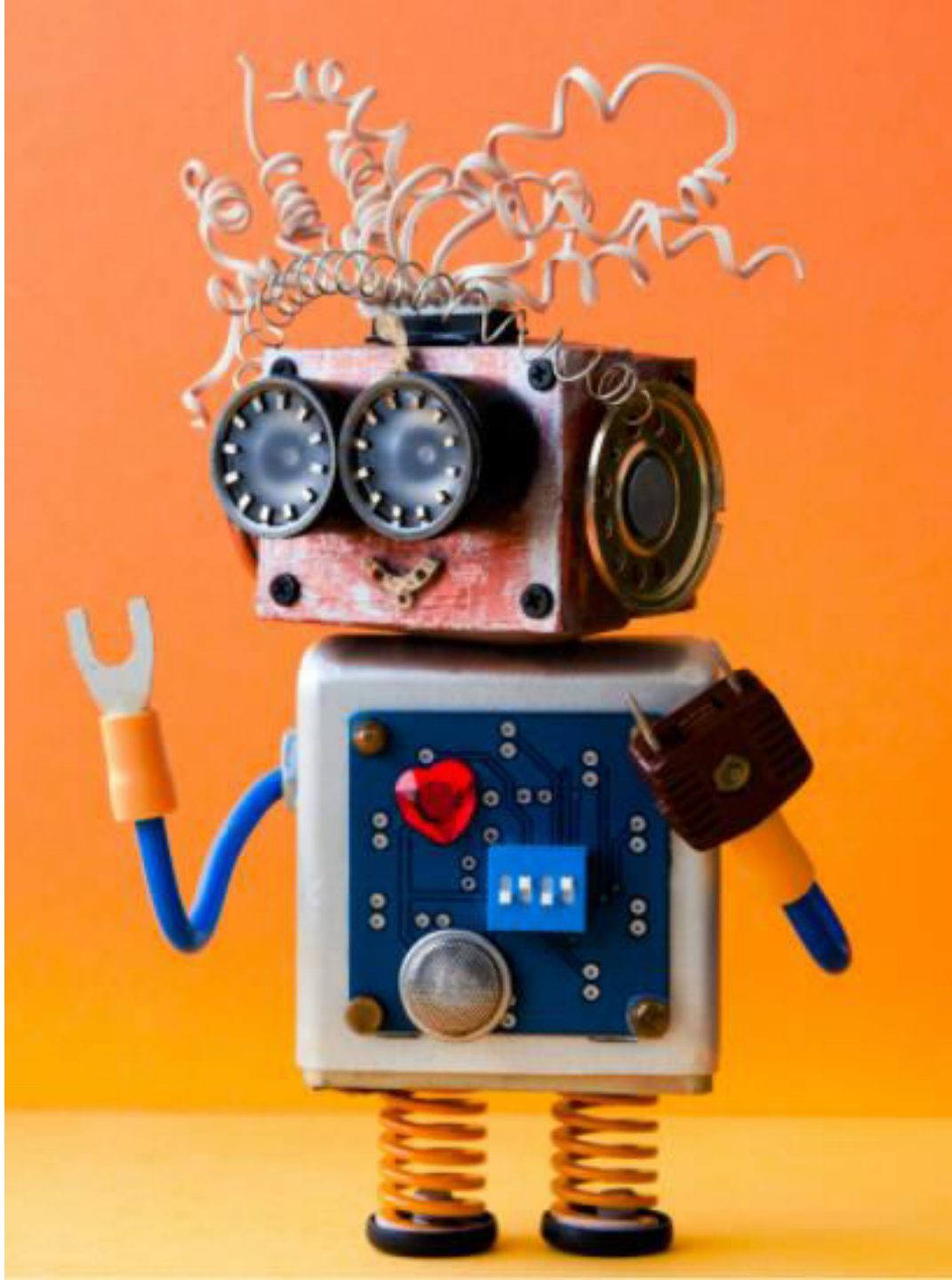


https://www.kaggle.com/ljanjughazyan/cars1

JUMPSTART DEMO: INVOKE ENDPOINT

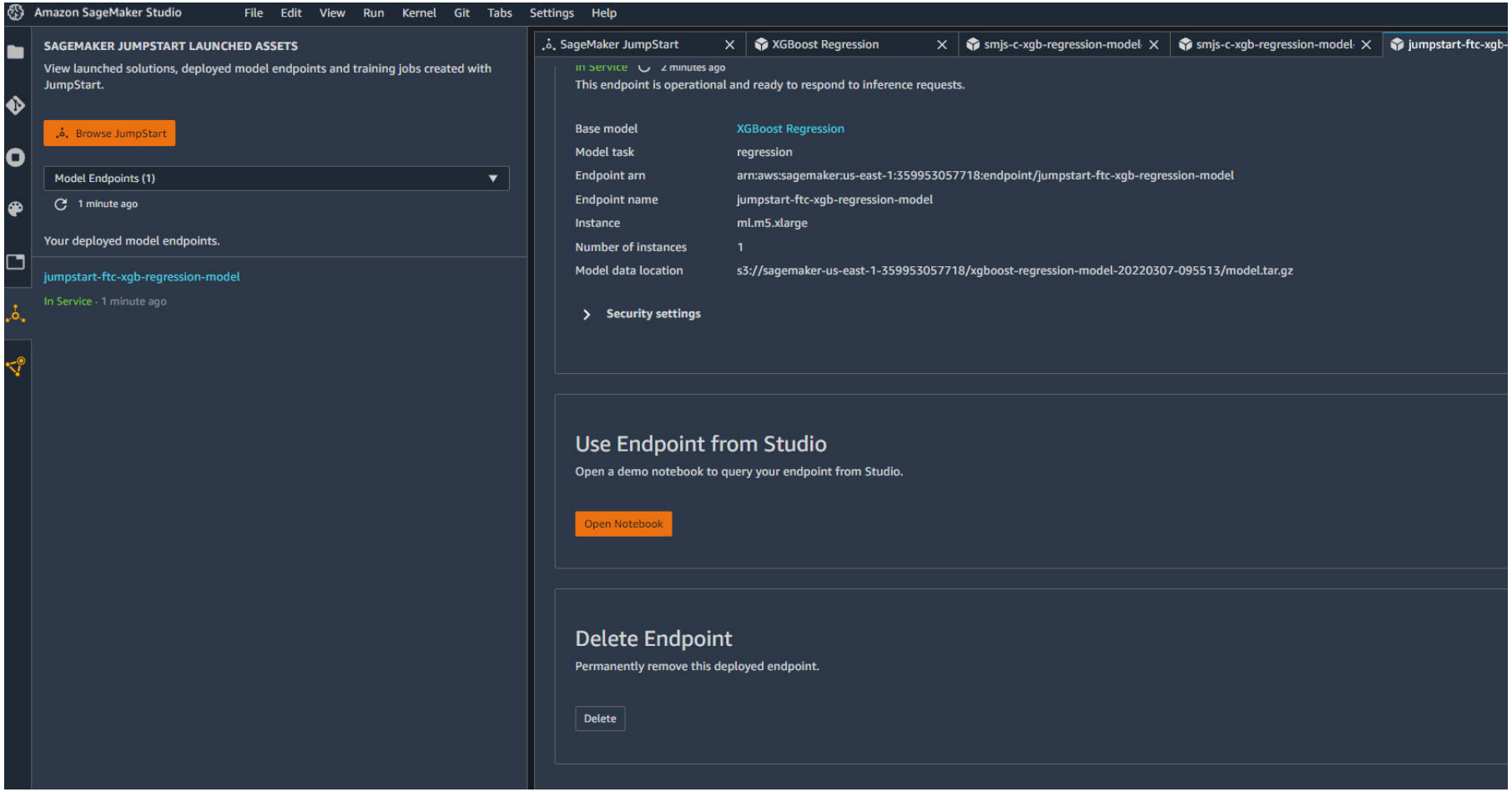
EASY

ADVANCED



JUMP START DEMO

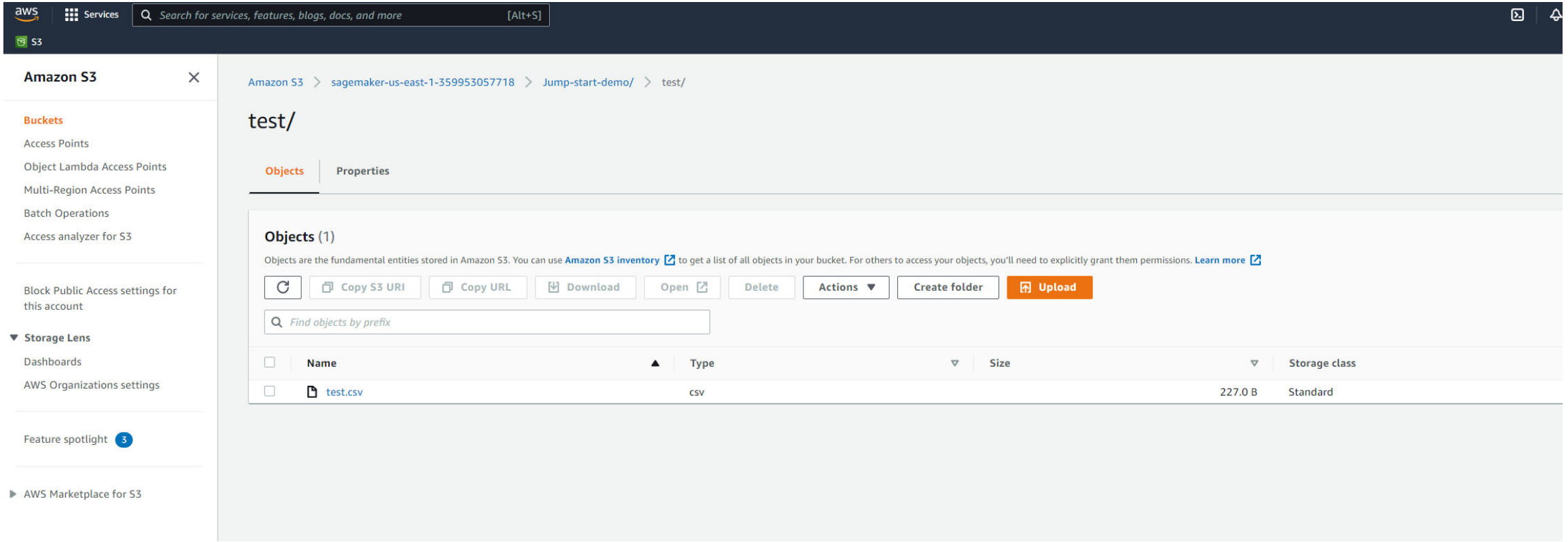
TO MAKE INFERENCE ON THE ENDPOINT, CLICK ON OPEN NOTEBOOK TO START MAKING QUERIES



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

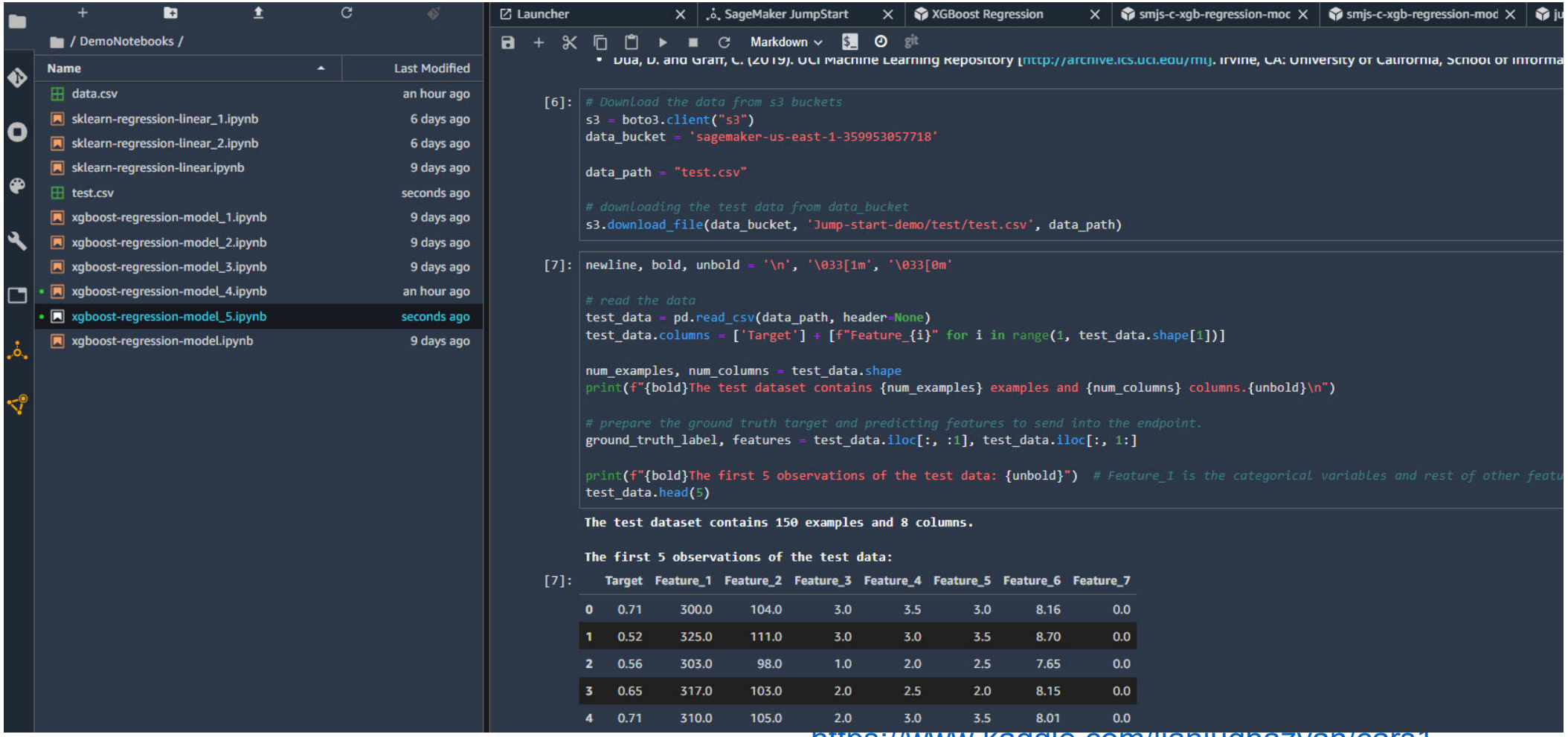
UPLOAD THE TESTING DATASETS TO THE S3 BUCKET



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

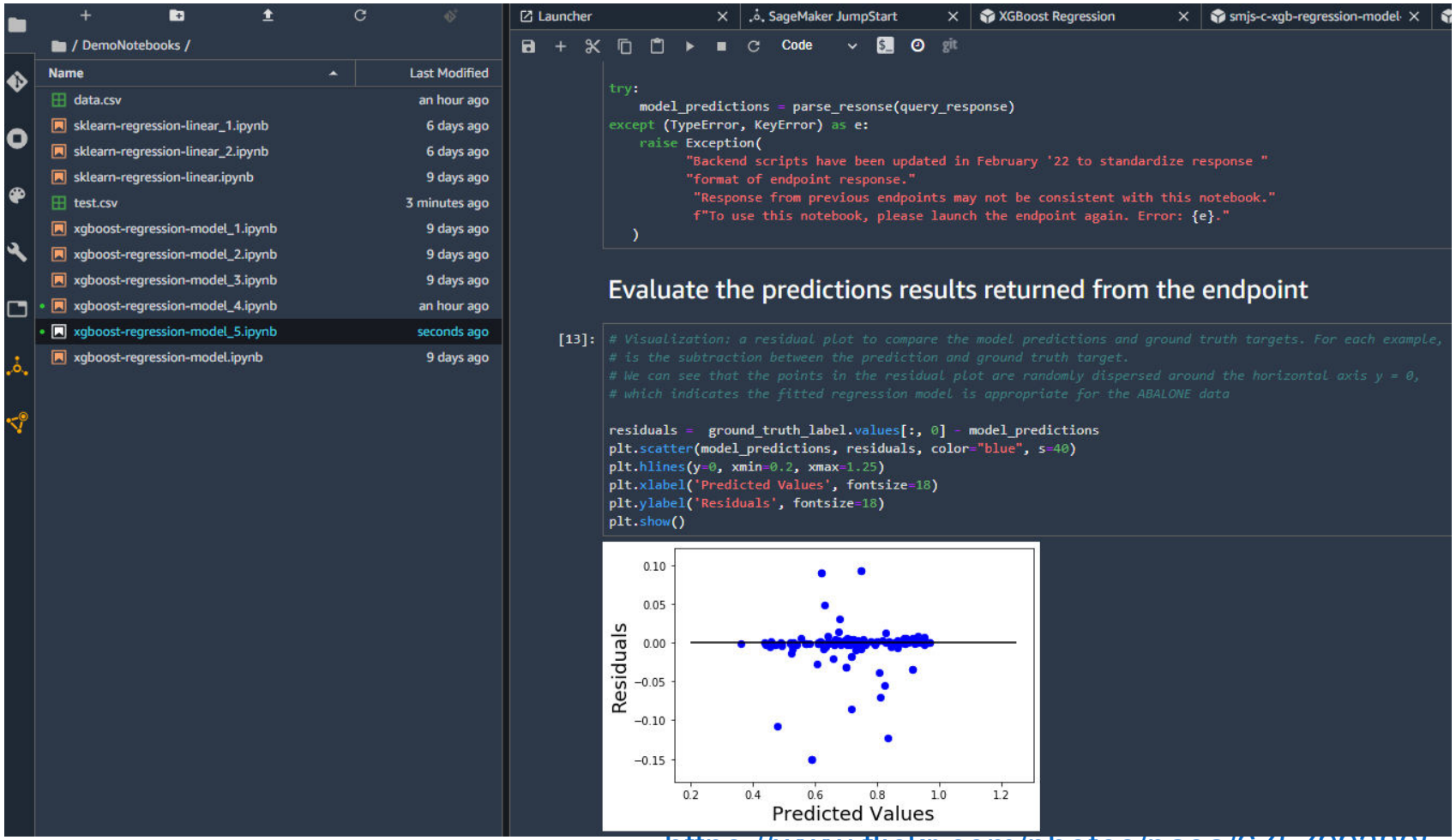
TO MAKE INFERENCE ON THE ENDPOINT, CLICK ON OPEN NOTEBOOK TO START MAKING QUERIES



https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

ASSESS TRAINED MODEL PERFORMANCE



https://www.flickr.com/photos/pasa/675799380
https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

PRINT OUT REGRESSION MODELS EVALUATION METRICS

[10]:
```
# Evaluate the model predictions quantitatively.
eval_r2_score = r2_score(ground_truth_label.values, model_predictions)
eval_mse_score = mean_squared_error(ground_truth_label.values, model_predictions)
eval_mae_score = mean_absolute_error(ground_truth_label.values, model_predictions)
print (
    f"{bold}Evaluation result on test data{unbold}:{newline}"
    f"{bold}{r2_score.__name__}{unbold}: {eval_r2_score}{newline}"
    f"{bold}{mean_squared_error.__name__}{unbold}: {eval_mse_score}{newline}"
    f"{bold}{mean_absolute_error.__name__}{unbold}: {eval_mae_score}{newline}"
)
```

```
Evaluation result on test data:
r2_score: 0.9557161949003675
mean_squared_error: 0.00089036054130454 46
mean_absolute_error: 0.011734435749053957
```

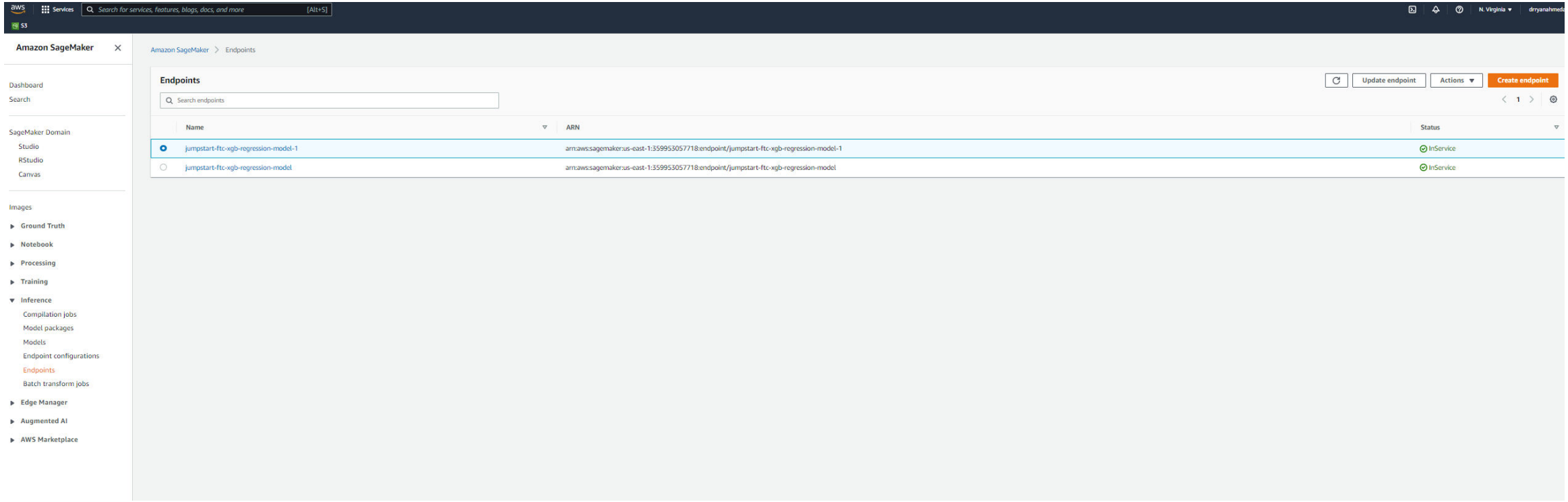
[ ]:

https://www.flickr.com/photos/pasa/6757993805

https://www.kaggle.com/ljanjughazyan/cars1

JUMP START DEMO

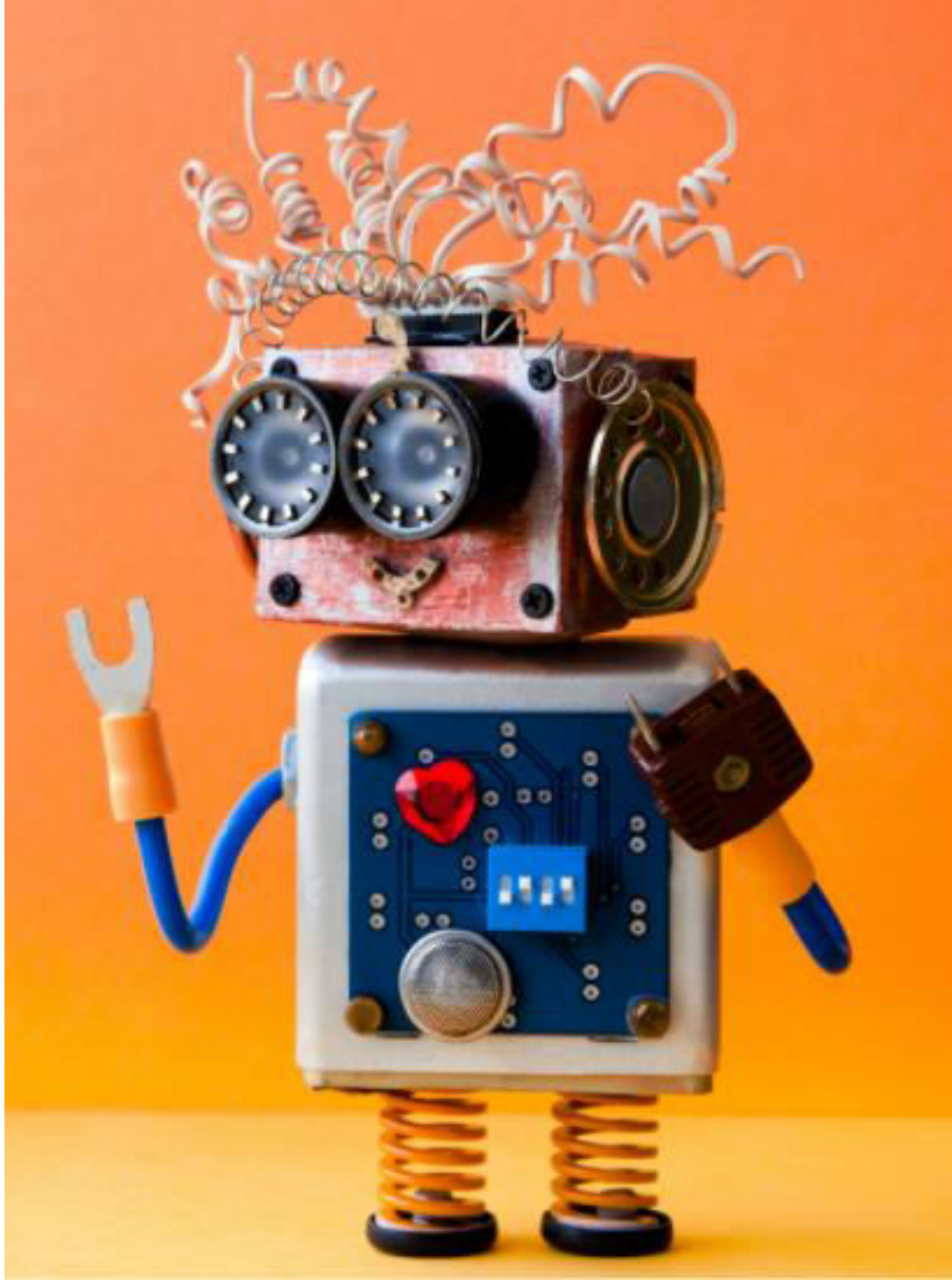
MAKE SURE TO DELETE THE DEPLOYED ENDPOINT AND SHUT DOWN ALL RUNNING INSTANCES



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1

FINAL END-OF-DAY CAPSTONE PROJECT

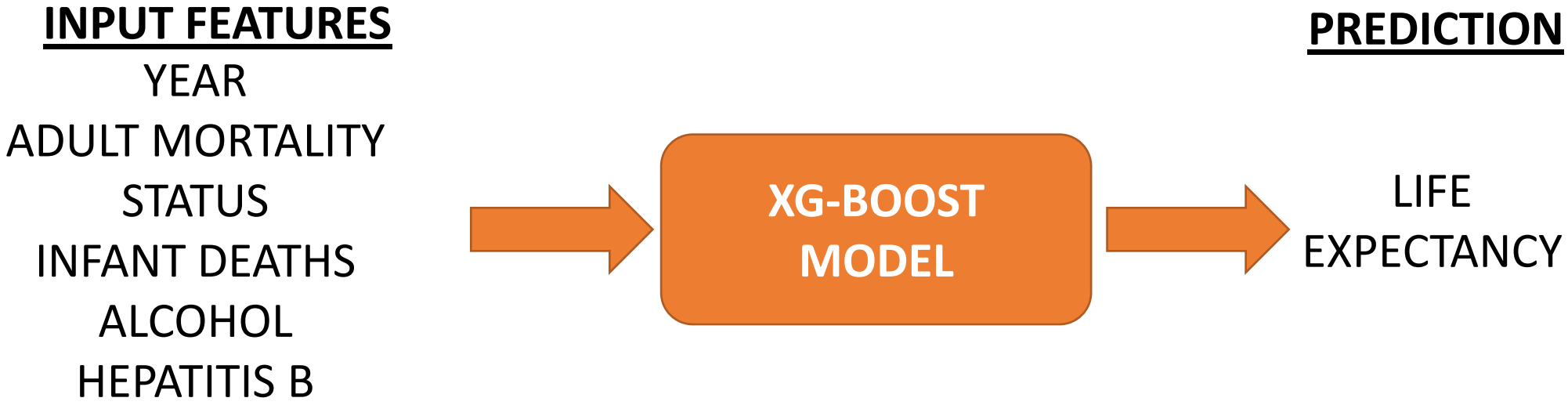
EASY ADVANCED



PROJECT OVERVIEW: SAGEMAKER JUMPSTART

- In this hands-on project, we will train an XG-Boost Regression model to predict life expectancy using SageMaker Jumpstart.
- This data was initially obtained from World Health Organization (WHO) and United Nations Website. Data contains features like year, status, life expectancy, adult mortality, infant deaths, percentage of expenditure, alcohol etc.
- **Tasks:**

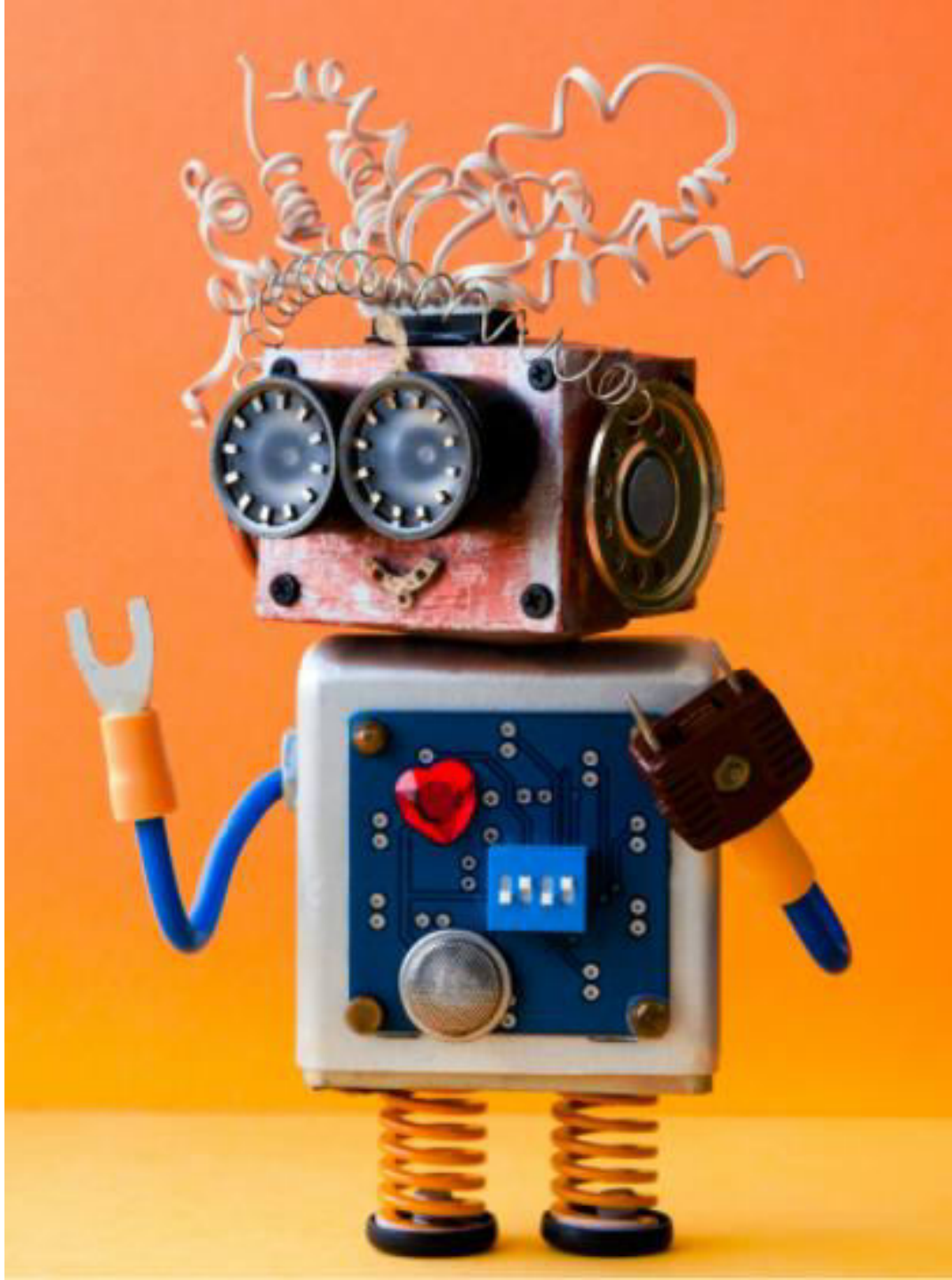
1. Upload the dataset Life_Expectancy_test_NoHeader.csv and Life_Expectancy_train_NoHeader.csv to S3
2. Using AWS SageMaker JumpStart, train an XG-Boost model to predict life expectancy.
3. Deploy the model and assess its performance. What's R2?



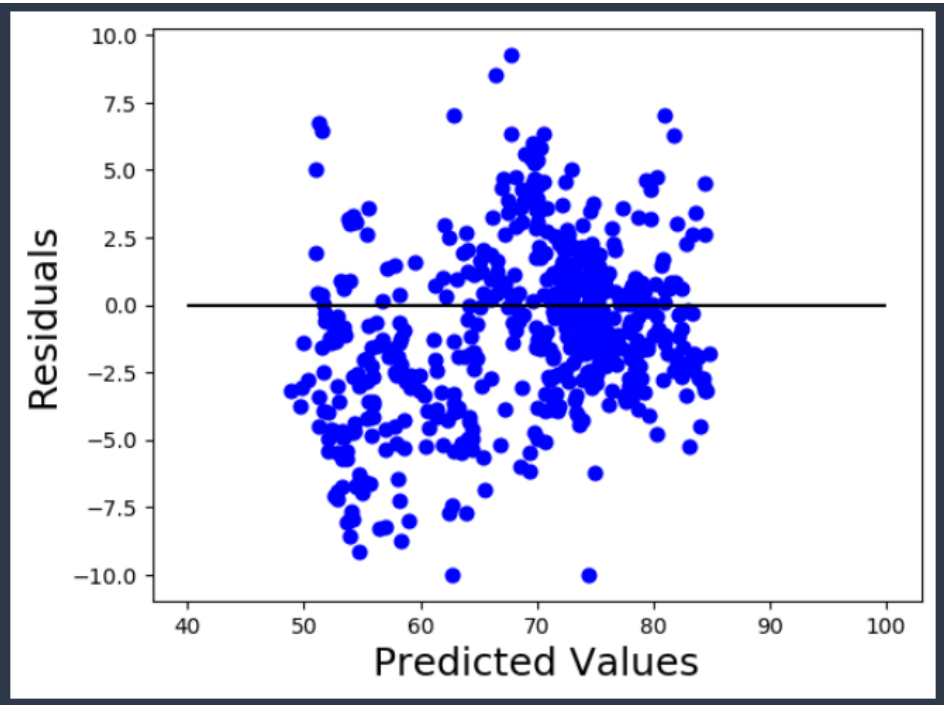
Source: https://www.kaggle.com/jkumarajarshi/life-expectancy-who

FINAL END-OF-DAY CAPSTONE PROJECT SOLUTION

EASY ADVANCED



FINAL PROJECT SOLUTION



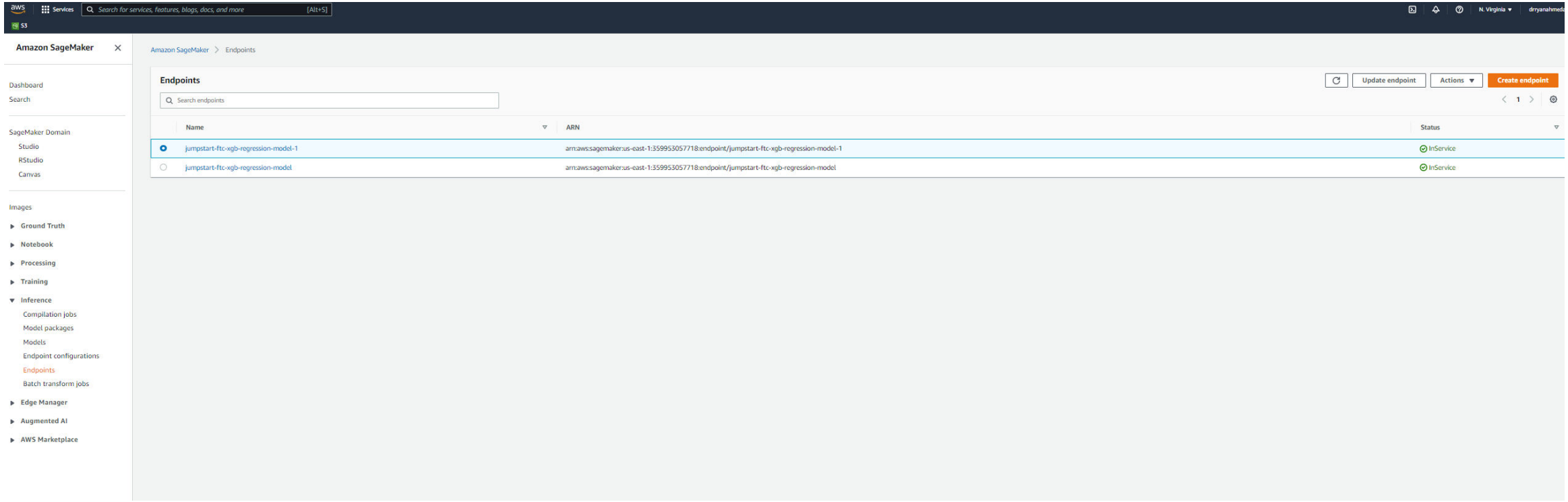
[10]:
```python
# Evaluate the model predictions quantitatively.
eval_r2_score = r2_score(ground_truth_label.values, model_predictions)
eval_mse_score = mean_squared_error(ground_truth_label.values, model_predictions)
eval_mae_score = mean_absolute_error(ground_truth_label.values, model_predictions)
print (
    f"{bold}Evaluation result on test data{unbold}:{newline}"
    f"{bold}{r2_score.__name__}{unbold}: {eval_r2_score}{newline}"
    f"{bold}{mean_squared_error.__name__}{unbold}: {eval_mse_score}{newline}"
    f"{bold}{mean_absolute_error.__name__}{unbold}: {eval_mae_score}{newline}"
)
```

```
Evaluation result on test data:
r2_score: 0.905099280527262
mean_squared_error: 9.99137901437254
mean_absolute_error: 2.451313835233201
```

https://www.kaggle.com/ijanjughazyan/cars1

FINAL PROJECT SOLUTION

MAKE SURE TO DELETE THE DEPLOYED ENDPOINT AND SHUT DOWN ALL RUNNING INSTANCES



https://www.flickr.com/photos/pasa/6757993805
https://www.kaggle.com/ljanjughazyan/cars1