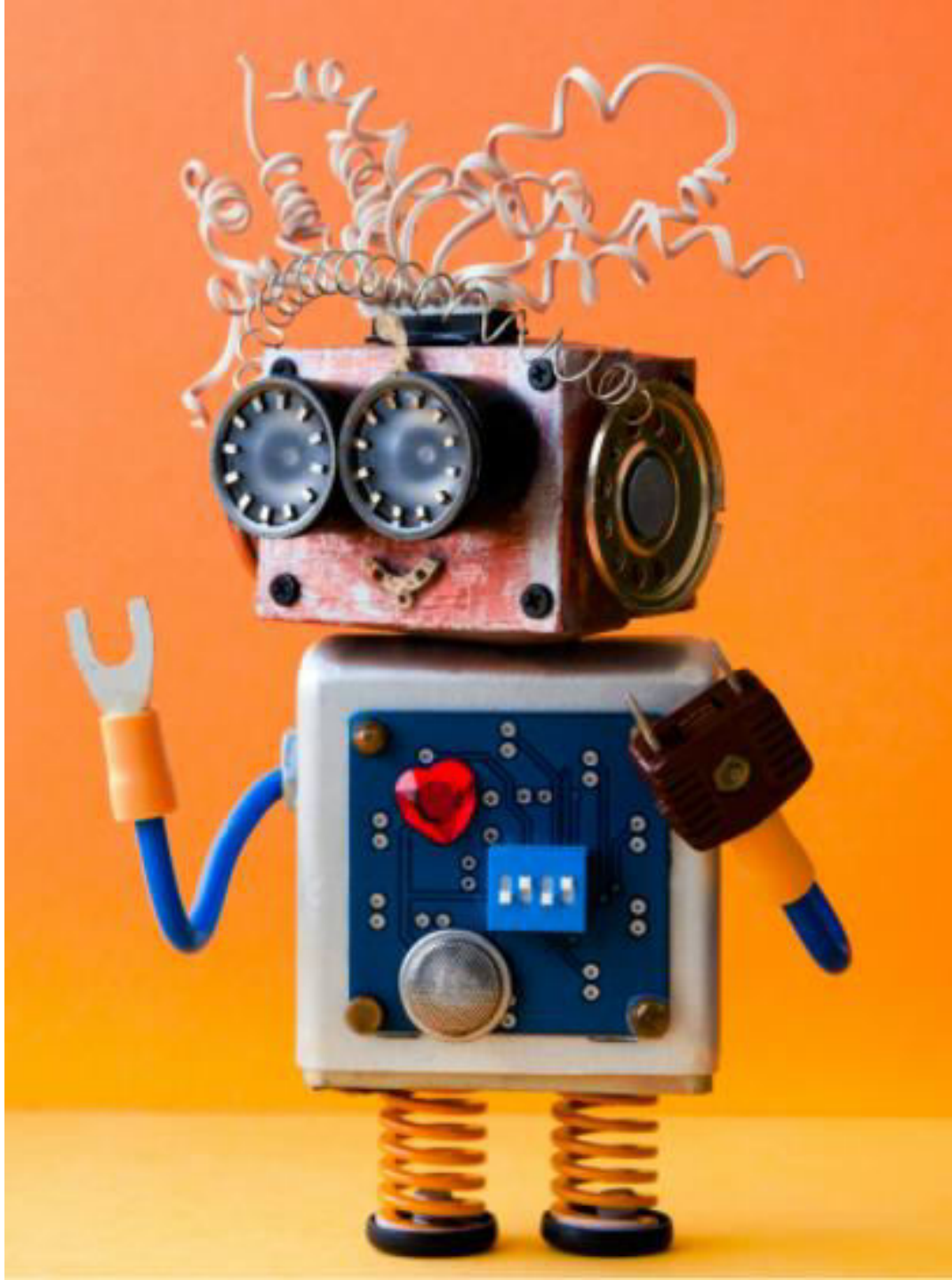
PROJECT CARD [SKIP IF FAMILIAR]

EASY ADVANCED



PROJECT CARD

GOAL:

- *Build, train, test and deploy a machine learning regression model to predict used car prices based on their features*

TOOL:

- *AWS SageMaker Studio*

PRACTICAL REAL-WORLD APPLICATION:

- *This project can be effectively used by car dealerships to predict used car prices and understand key factors that contribute to used car prices.*

DATA:

- **INPUTS:**
  - *Make, Model, Type, Origin, Drivetrain, Invoice, EngineSize, Cylinders, Horsepower, MPG_City, MPG_Highway, Weight, Wheelbase, and Length*

- **OUTPUT:**
  - *MSRP (Price)*

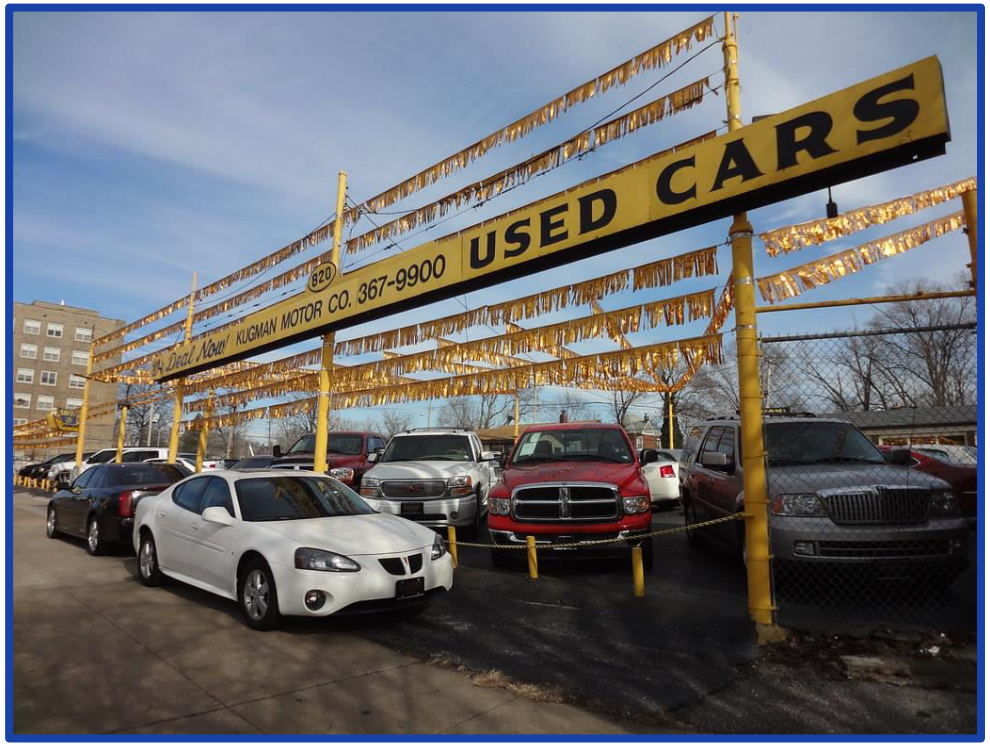
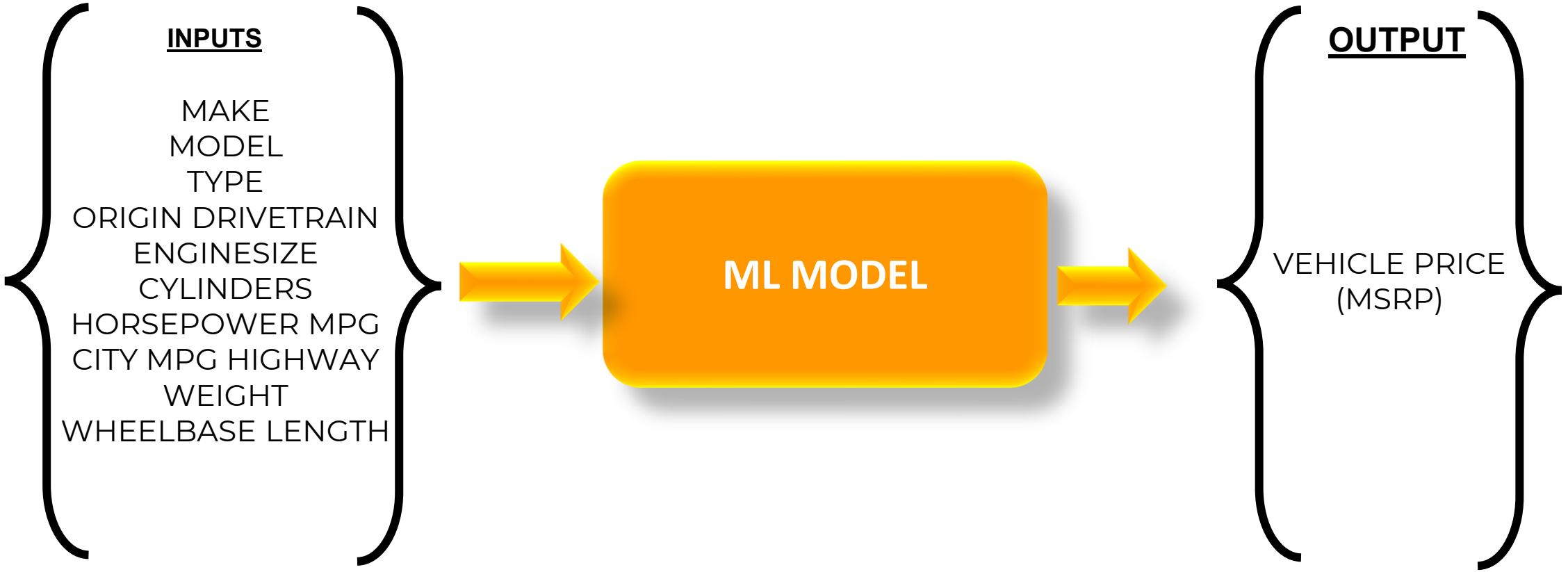


Image Source: https://www.flickr.com/photos/pasa/6757993805
Dataset Source: https://www.kaggle.com/ljanjughazyan/cars1

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

INPUTS AND OUTPUTS

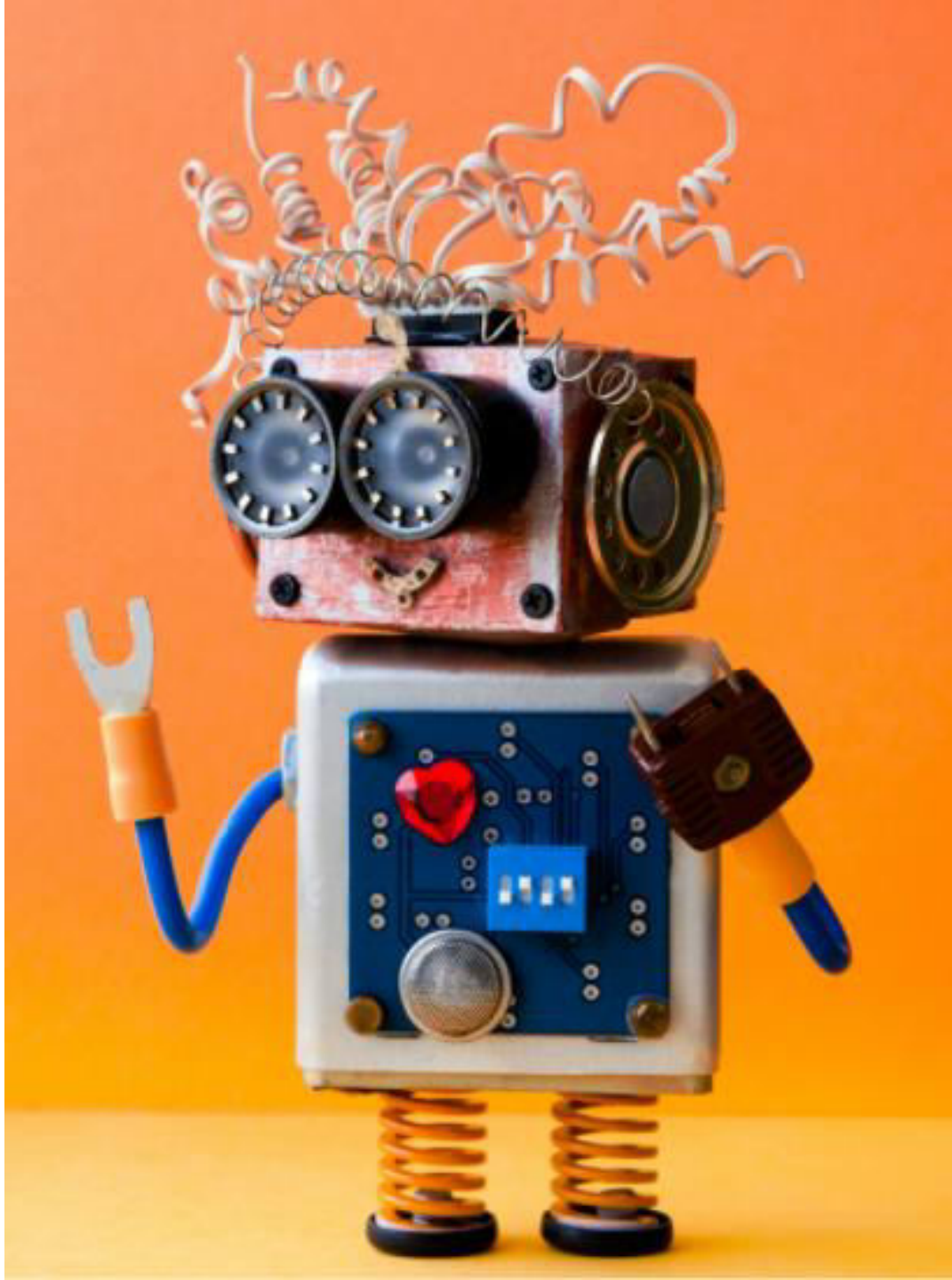


DATA OVERVIEW

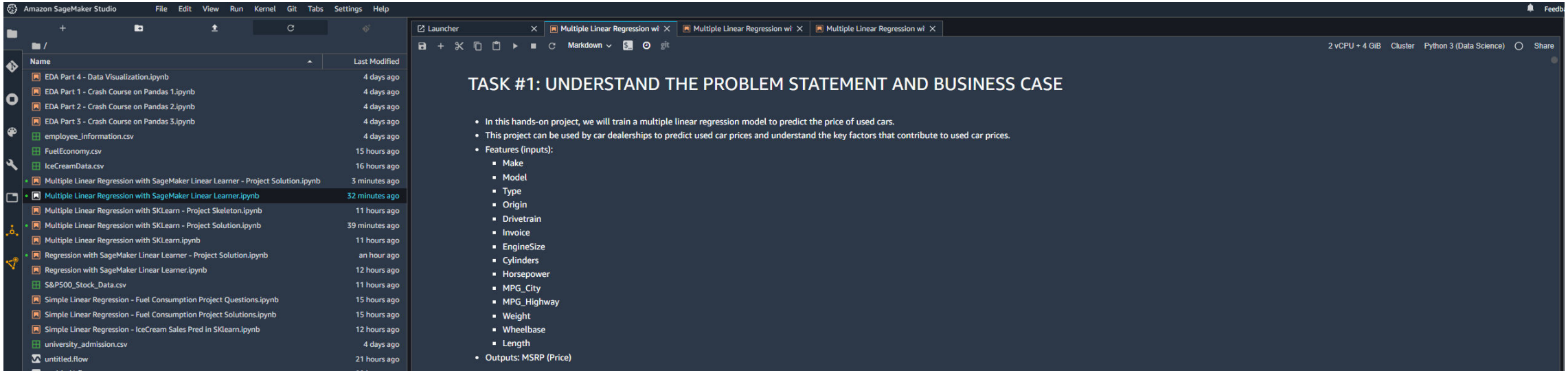
	Make	Model	Type	Origin	DriveTrain	MSRP	EngineSize	Cylinders	Horsepower	MPG_City	MPG_Highway	Weight	Wheelbase	Length
0	Acura	MDX	SUV	Asia	All	36945	3.5	6.0	265	17	23	4451	106	189
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	23820	2.0	4.0	200	24	31	2778	101	172
2	Acura	TSX 4dr	Sedan	Asia	Front	26990	2.4	4.0	200	22	29	3230	105	183
3	Acura	TL 4dr	Sedan	Asia	Front	33195	3.2	6.0	270	20	28	3575	108	186
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	43755	3.5	6.0	225	18	24	3880	115	197
5	Acura	3.5 RL w/Navigation 4dr	Sedan	Asia	Front	46100	3.5	6.0	225	18	24	3893	115	197
6	Acura	NSX coupe 2dr manual S	Sports	Asia	Rear	89765	3.2	6.0	290	17	24	3153	100	174
7	Audi	A4 1.8T 4dr	Sedan	Europe	Front	25940	1.8	4.0	170	22	31	3252	104	179
8	Audi	A41.8T convertible 2dr	Sedan	Europe	Front	35940	1.8	4.0	170	23	30	3638	105	180
9	Audi	A4 3.0 4dr	Sedan	Europe	Front	31840	3.0	6.0	220	20	28	3462	104	179

MODEL OUTPUT: MSRP
MANUFACTURER'S SUGGESTED RETAIL PRICE

CODE DEMO

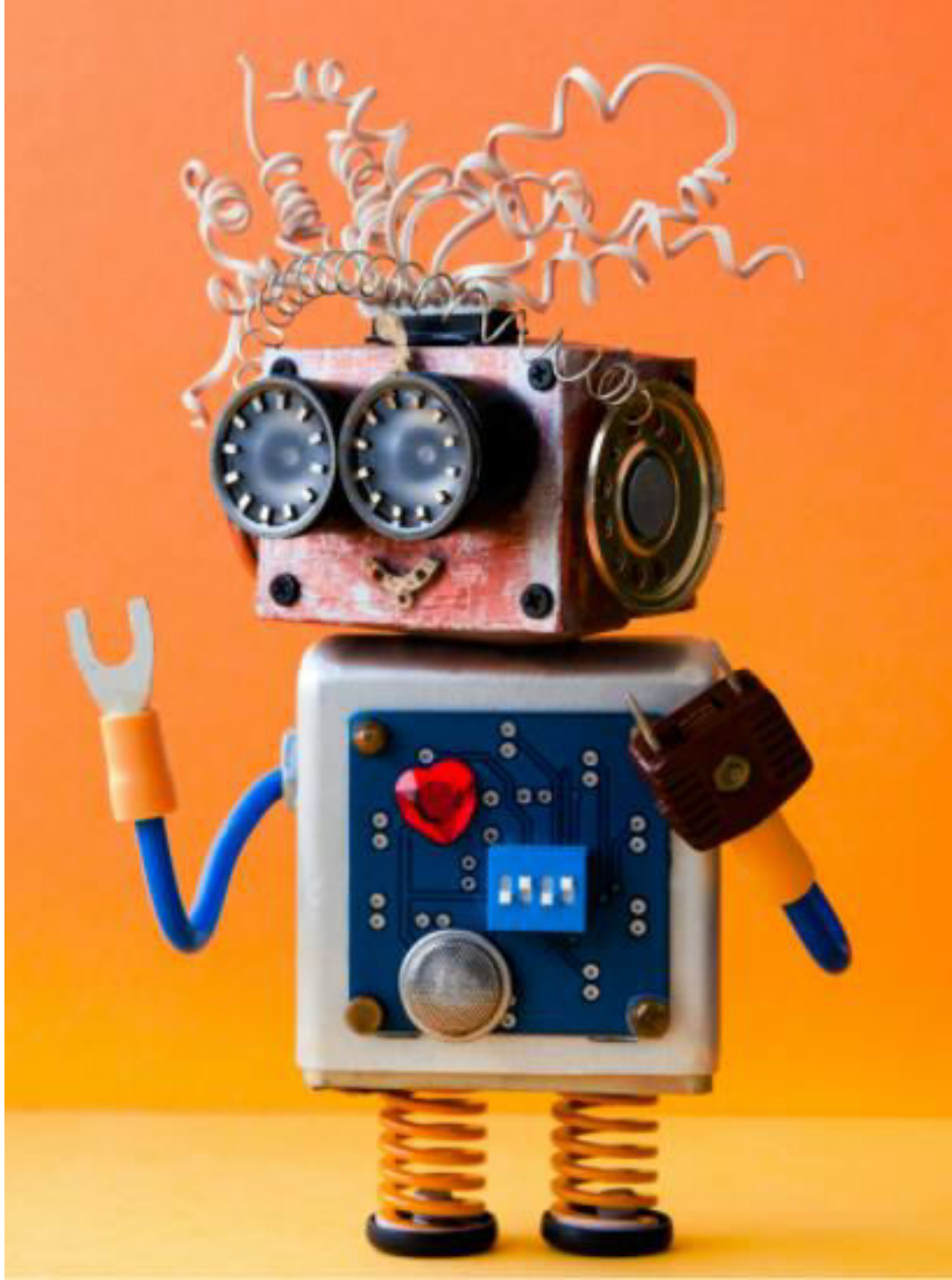


CODE DEMO


Amazon SageMaker Studio
File Edit View Run Kernel Git Tabs Settings Help
Launcher
Multiple Linear Regression wi
Multiple Linear Regression wi
Multiple Linear Regression wi
Markdown
2 vCPU + 4 GiB Cluster Python 3 (Data Science) Share
Name
Last Modified
EDA Part 4 - Data Visualization.ipynb 4 days ago
EDA Part 1 - Crash Course on Pandas 1.ipynb 4 days ago
EDA Part 2 - Crash Course on Pandas 2.ipynb 4 days ago
EDA Part 3 - Crash Course on Pandas 3.ipynb 4 days ago
employee_information.csv 4 days ago
FuelEconomy.csv 15 hours ago
IceCreamData.csv 16 hours ago
Multiple Linear Regression with SageMaker Linear Learner - Project Solution.ipynb 3 minutes ago
Multiple Linear Regression with SageMaker Linear Learner.ipynb 32 minutes ago
Multiple Linear Regression with SKLearn - Project Skeleton.ipynb 11 hours ago
Multiple Linear Regression with SKLearn - Project Solution.ipynb 39 minutes ago
Multiple Linear Regression with SKLearn.ipynb 11 hours ago
Regression with SageMaker Linear Learner - Project Solution.ipynb an hour ago
Regression with SageMaker Linear Learner.ipynb 12 hours ago
S&P500_Stock_Data.csv 11 hours ago
Simple Linear Regression - Fuel Consumption Project Questions.ipynb 15 hours ago
Simple Linear Regression - Fuel Consumption Project Solutions.ipynb 15 hours ago
Simple Linear Regression - IceCream Sales Pred in SKlearn.ipynb 12 hours ago
university_admission.csv 4 days ago
untitled.flow 21 hours ago
TASK #1: UNDERSTAND THE PROBLEM STATEMENT AND BUSINESS CASE
In this hands-on project, we will train a multiple linear regression model to predict the price of used cars.
This project can be used by car dealerships to predict used car prices and understand the key factors that contribute to used car prices.
Features (inputs):
Make
Model
Type
Origin
Drivetrain
Invoice
EngineSize
Cylinders
Horsepower
MPG_City
MPG_Highway
Weight
Wheelbase
Length
Outputs: MSRP (Price)

END-OF-DAY CAPSTONE PROJECT

EASY ADVANCED



PROJECT

- We would like to predict the S&P500 Price using interest rate and employment using Amazon SageMaker Linear Learner algorithm.
  - Independent variable X: Interest Rate and Employment
  - Dependent variable Y: S&P 500 Price

Interest Rates	Employment	S&P 500 Price
1.943859273	55.41357113	2206.680582
2.258228944	59.54630512	2486.474488
2.215862783	57.41468676	2405.868337
1.977959542	49.90835272	2140.434475
2.437722808	52.03549192	2411.275663
2.143636835	56.06059825	2187.344909
2.148646786	51.51320834	2263.049249
2.176183572	53.4759086	2281.496374
2.125351611	63.66842224	2355.163011
2.225681934	56.99339607	2326.330337
1.814687751	55.36178043	2078.553895
2.281897215	58.48475241	2337.504507
2.426737871	55.7093282	2485.774097
2.259270476	61.8872018	2478.413528
2.38801924	66.55127056	2665.00807
1.715103596	60.20251695	2057.393366
2.392425284	60.57381954	2423.590565
2.388766722	58.26132918	2605.470983
2.25666065	52.77316693	2303.851816
2.089815376	48.80721748	2095.440317
2.348535874	58.65942761	2495.24303
1.751579397	54.1482556	1871.361622
2.043664892	55.88532564	2213.4959

PROJECT TASKS

Using AWS SageMaker Linear Learner, perform the following:

- 1. Upload the "*S&P500_Stock_Data.csv*" dataset into S3
- 2. Split the data into 80% for training and 20% for testing
- 3. Train a machine linear regression model using SageMaker SDK
- 4. Deploy trained model as an endpoint
- 5. Assess trained model performance, what is the RMSE?
- 6. Delete the endpoint