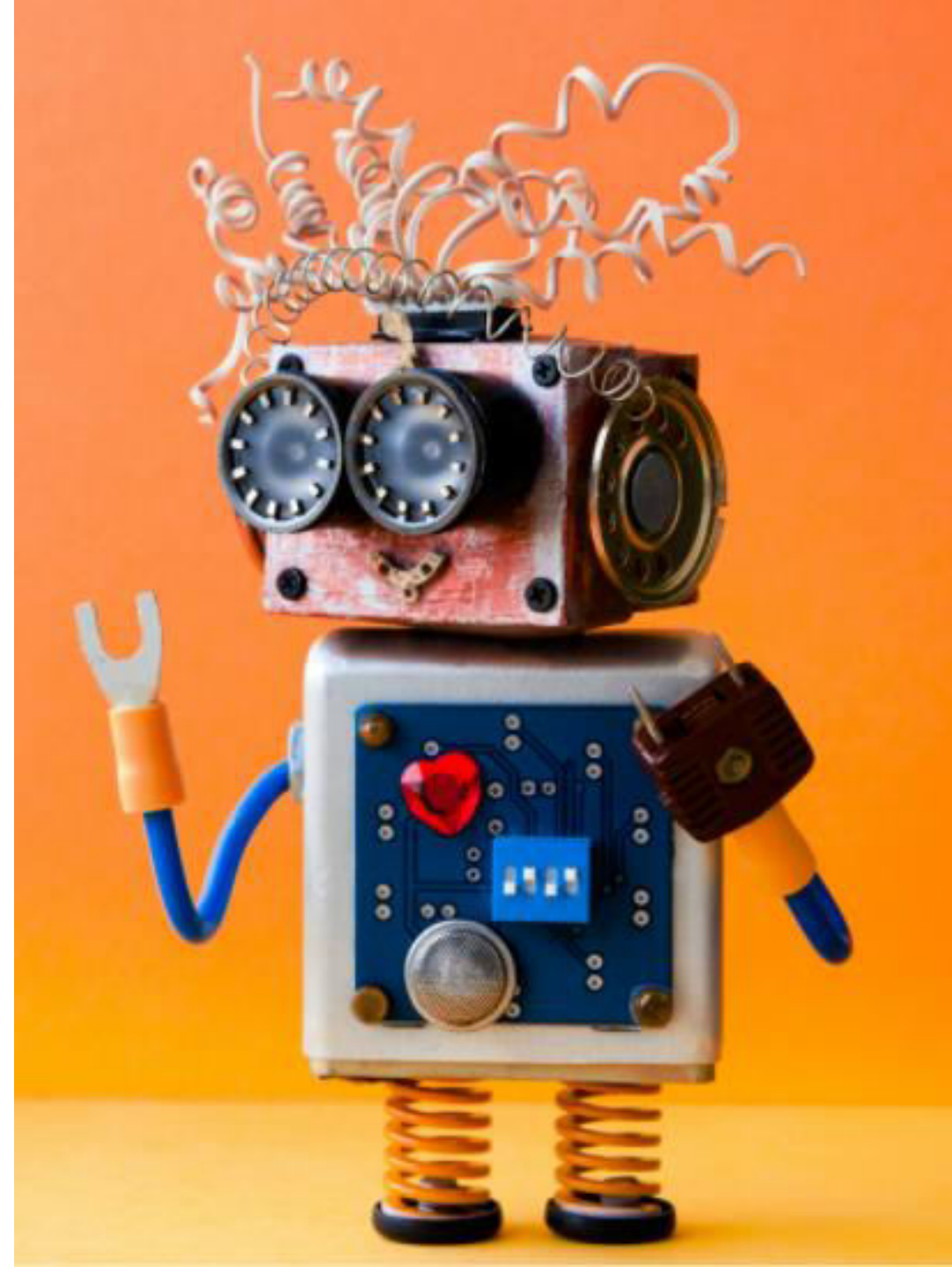


# PROJECT CARD & DATA OVERVIEW [SKIP IF FAMILIAR]



# HYPERPARAMETERS OPTIMIZATION RECAP

- Hyperparameters optimization is the problem of choosing a set of optimal hyperparameters for a given machine learning algorithm.
- A hyperparameters are used to control the learning process and are set before the training job starts.
- After training multiple models, you would like to fine tune them so that they perform better on a given dataset.

## HYPERPARAMETER

***Hyper parameter: values set prior to the training process such as number of neurons, layers, learning rate..etc***

## PARAMETER

***Parameter: values that are obtained by the training process such as network weights and biases.***

# PROJECT CARD

## GOAL:

- *Build, train, test and deploy a machine learning model to predict bike rental usage based on inputs such as temperature, humidity, wind speed..etc.*
- *We will train a model and optimize its hyperparameters in SageMaker.*

## TOOL:

- *AWS SageMaker Studio (Hyperparameters Tuning Jobs)*

## PRACTICAL REAL-WORLD APPLICATION:

- *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.*

## DATA:

### • **INPUTS:**

- *Instant, date, season, year, hour, month, holiday, weather situation, temperature, and windspeed.*

### • **OUTPUT:**

- *casual: count of casual users*
- *registered: count of registered users*
- *cnt: count of total rental bikes including both casual and registered*



**Image Source:** <https://pixabay.com/photos/bike-rental-bikes-rent-pay-2284380/>

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

# DATA EXPLORATION: INPUTS

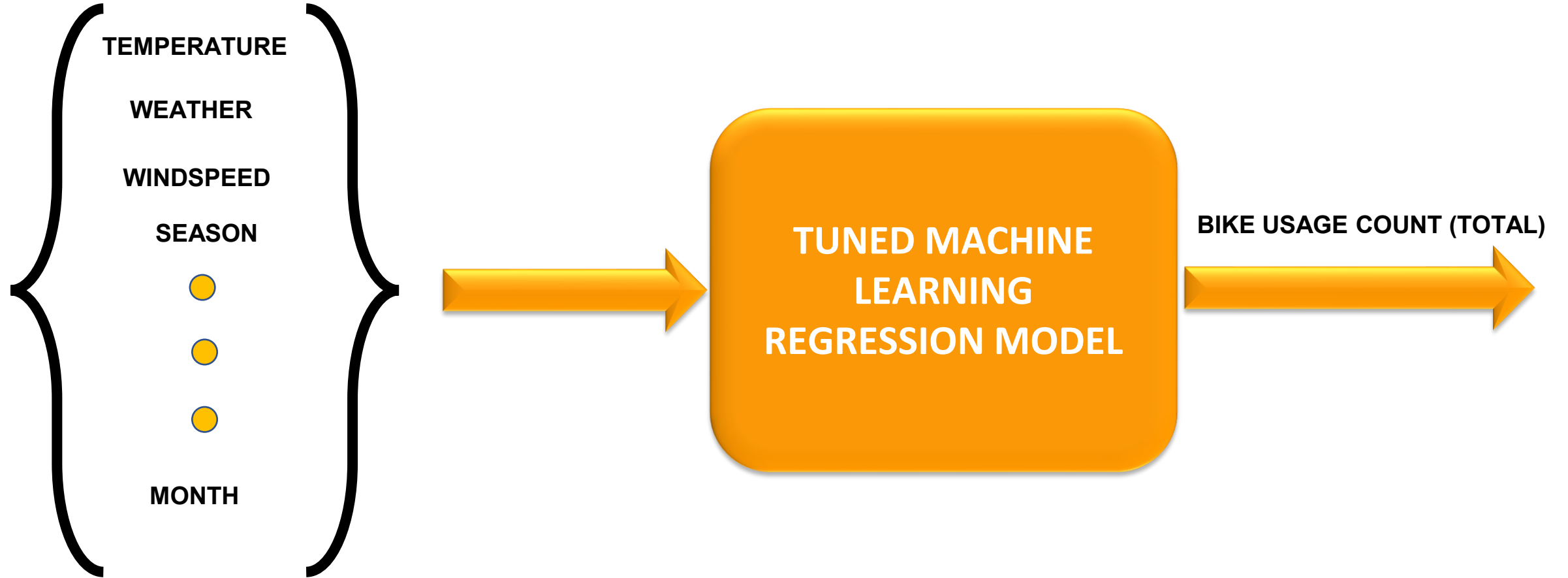
- **Inputs:**

- instant: record index
- dteday: date
- season: season (1: springer, 2: summer, 3: fall, 4: winter)
- yr: year (0: 2011, 1: 2012)
- mnth: month ( 1 to 12)
- hr: hour (0 to 23)
- holiday: whether day is holiday or not - weekday : day of the week
- Working day: if day is neither weekend nor holiday is 1, otherwise is 0.
- weathersit :
  - ❖ 1: Clear, Few clouds, Partly cloudy
  - ❖ 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
  - ❖ 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
  - ❖ 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp : Normalized temperature in Celsius. The values are divided to 41 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)

- **Outputs:**

- casual: count of casual users
- registered: count of registered users
- cnt: count of total rental bikes including both casual and registered

# MODEL OVERVIEW



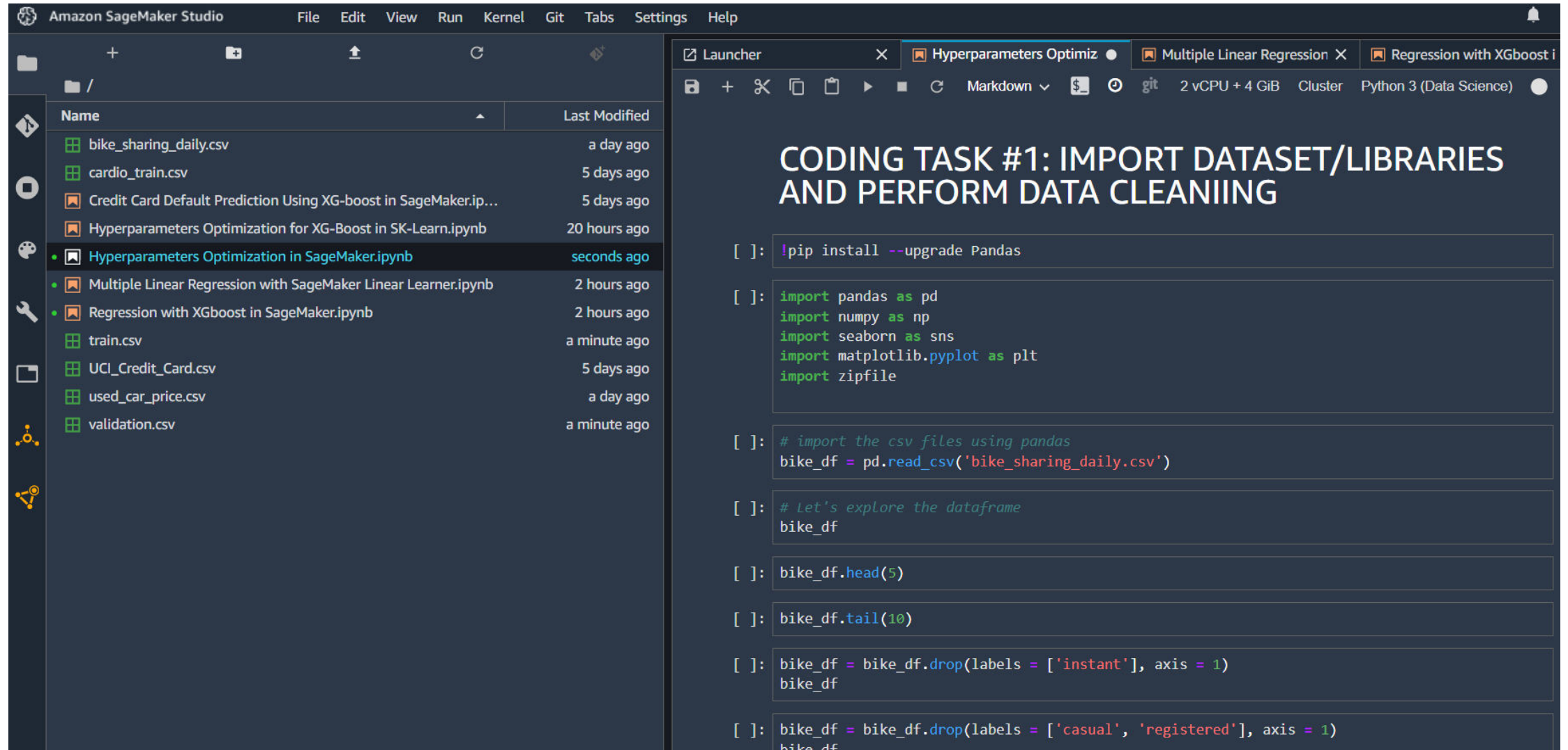


# HYPERPARAMETERS OPTIMIZATION USING SAGEMAKER DEMO

---



# HYPERPARAMETERS OPTIMIZATION DEMO IN AMAZON SAGEMAKER



The screenshot displays the Amazon SageMaker Studio interface. On the left, a file explorer shows a list of files and notebooks. The selected notebook, "Hyperparameters Optimization in SageMaker.ipynb", is highlighted. The main area on the right shows the Jupyter notebook content, which includes a title and several code cells.

**File Explorer:**

Name	Last Modified
bike_sharing_daily.csv	a day ago
cardio_train.csv	5 days ago
Credit Card Default Prediction Using XG-boost in SageMaker.ip...	5 days ago
Hyperparameters Optimization for XG-Boost in SK-Learn.ipynb	20 hours ago
<b>Hyperparameters Optimization in SageMaker.ipynb</b>	<b>seconds ago</b>
Multiple Linear Regression with SageMaker Linear Learner.ipynb	2 hours ago
Regression with XGboost in SageMaker.ipynb	2 hours ago
train.csv	a minute ago
UCI_Credit_Card.csv	5 days ago
used_car_price.csv	a day ago
validation.csv	a minute ago

**Jupyter Notebook Content:**

## CODING TASK #1: IMPORT DATASET/LIBRARIES AND PERFORM DATA CLEANIING

```
[ ]: !pip install --upgrade Pandas

[ ]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import zipfile

[ ]: # import the csv files using pandas
bike_df = pd.read_csv('bike_sharing_daily.csv')

[ ]: # Let's explore the dataframe
bike_df

[ ]: bike_df.head(5)

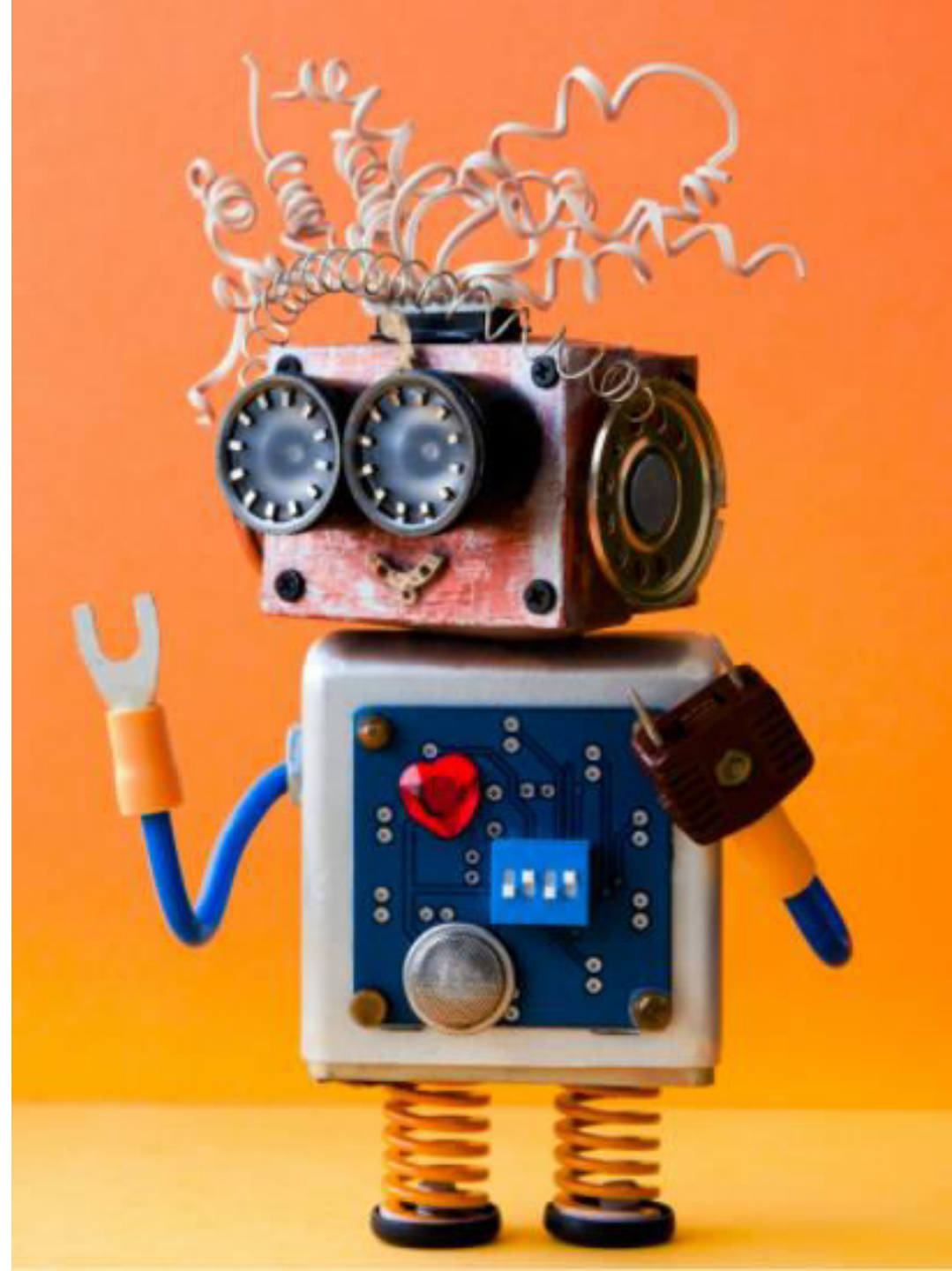
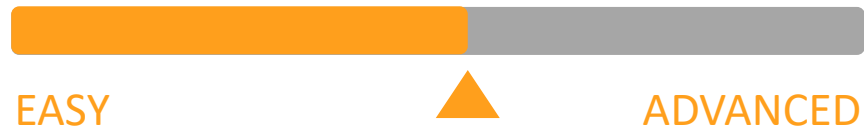
[ ]: bike_df.tail(10)

[ ]: bike_df = bike_df.drop(labels = ['instant'], axis = 1)
bike_df

[ ]: bike_df = bike_df.drop(labels = ['casual', 'registered'], axis = 1)
bike_df
```

# FINAL END-OF-DAY CAPSTONE PROJECT

---





# PROJECT

Using the used car prices dataset included in the course package, perform the following:

1. Load the “*used\_car\_price.csv*” dataset
2. Split the data into training, validation and testing
3. Train an XG-Boost model using SageMaker Built-in XG-Boost Algorithm
4. Assess trained XG-Boost model performance using various KPIs
5. Perform hyperparameters optimization using SageMaker SDK
6. Compare the performance of the tuned model using various KPIs. Comment on your results.