Problem Statement:

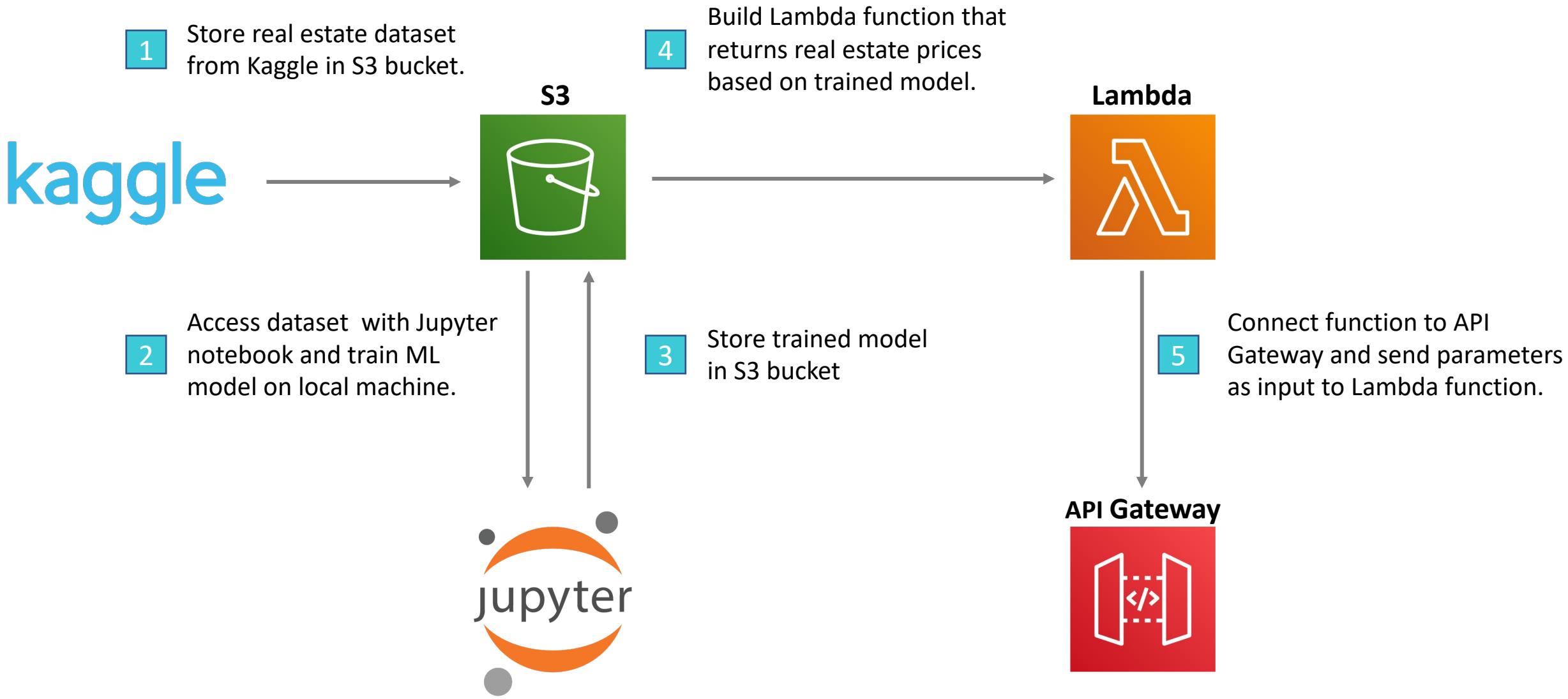
- Pricing real estate correctly can be a challenging task for sellers as there is a variety of features of the property that determine the price.
- While overpricing a property can result in an extended time on the market, reducing the likelihood of the sale, underpricing will lead to a loss of profit for the seller.

Proposed Solution:

- The goal is to create a machine learning model to help sellers easily find the optimal price for their property & sell it.
- To achieve this, we take a housing dataset from Kaggle that will be stored in an S3 bucket. (We will focus on real estate prices in Madrid using this dataset: https://www.kaggle.com/datasets/mirbektoktogaraev/madrid-real-estate-market)
- The data will be used to train a machine learning model on our local machine using Jupyter notebook.
- The final model will be uploaded to an S3 bucket and used by a Lambda function to predict real estate prices.
- The Lambda function will be connected to API Gateway to make it publicly accessible and predict real estate prices, based on the parameters received from the API call.

Architecture diagram:

See next page.


1
Store real estate dataset from Kaggle in S3 bucket.
4
Build Lambda function that returns real estate prices based on trained model.
S3
Lambda
kaggle
2
Access dataset with Jupyter notebook and train ML model on local machine.
3
Store trained model in S3 bucket
5
Connect function to API Gateway and send parameters as input to Lambda function.
jupyter
API Gateway