MLOPS

# Using Designer

We are going to build restaurant ratingsnased on the restaurants dataset. Available in samples.

## Step 1: Create Compute Instance

A screenshot of a computer

Description automatically generated

## Step 2: Data Wrangling and Data Transformation

Select "Designer" from the left menu. Choose the "Classic Prebuilt" tab and then click the + sign to create a new pipeline.

A screenshot of a computer

Description automatically generated

### Dataset

1. Add the sample datasets, **Restaurant Ratings** and **Restaurant Feature Data**, to the pipeline canvas.
2. Add Select Columns in Dataset for **Restaurant Ratings**  and **Restaurant Feature Data**.
3. Double click on Select Columns in Dataset

A screenshot of a computer

Description automatically generated

* 1. Select placeID and rating from the **Restaurant Ratings** data source.

A screenshot of a computer

Description automatically generated

* 1. Select placeID, alcohol, dress\_code, price, and Rambience from the **Restaurant Feature Data** source.

A screenshot of a computer

Description automatically generated

Output

A screenshot of a computer

Description automatically generated

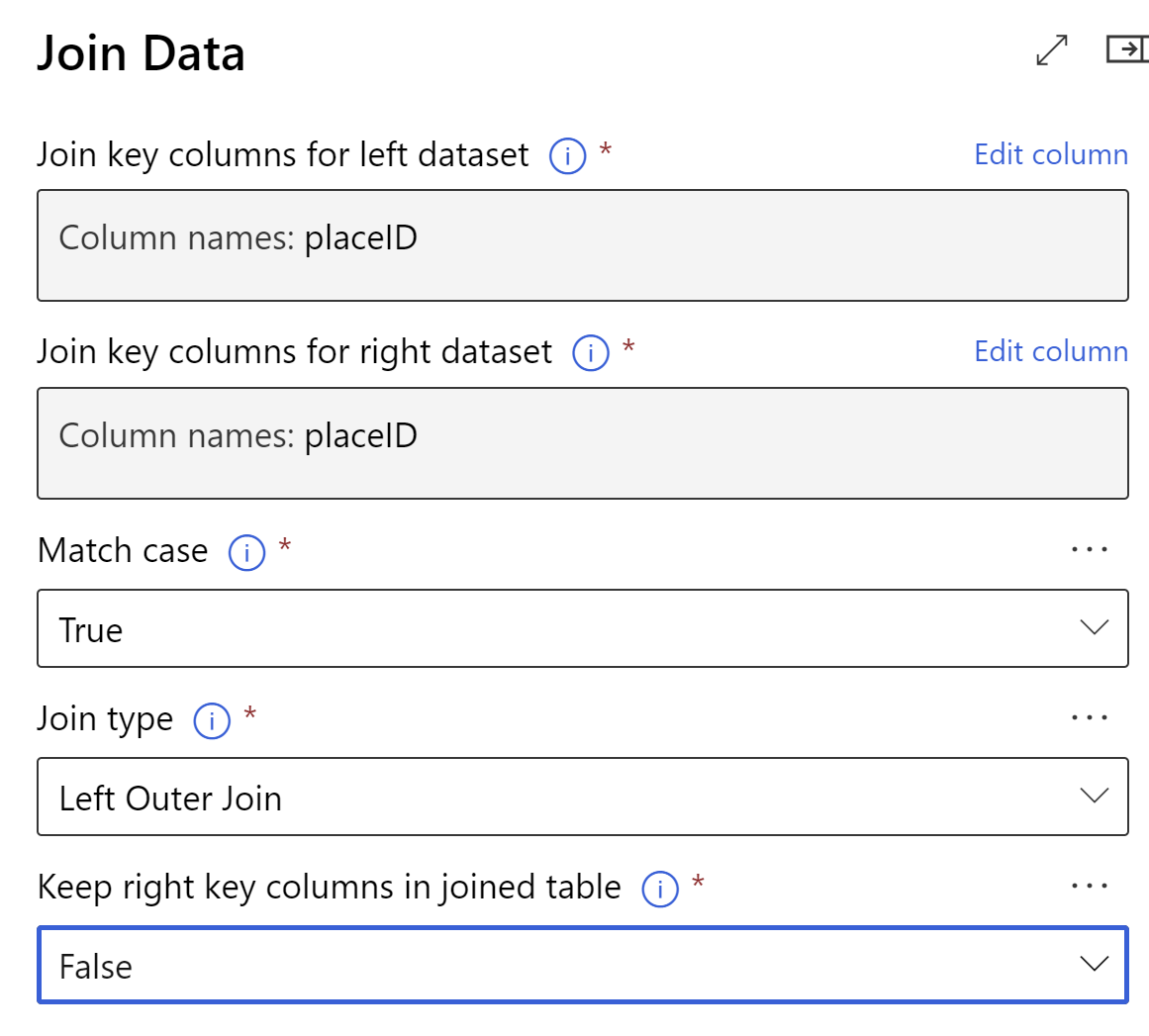
### Transformation

1. Join the data sources using placeID as key.

A screenshot of a computer

Description automatically generated

Double click on Join Data, add placeID for columns and select Left Outer Join as the Join Type. Set False for the Keep Right Key Columns in joined table



1. Replace missing data in columns (placeID, rating, alcohol, dress\_code, price, Rambience) with *0*.

### Submit the pipeline

Give the experiment name

A screenshot of a computer

Description automatically generated

Select the compute

A screenshot of a computer

Description automatically generated

Hit Submit to create a new pipeline job

A screenshot of a phone

Description automatically generated

Navigate to Jobs

A screenshot of a computer

Description automatically generated

Check for Outputs + Logs

A screenshot of a computer

Description automatically generated

You can look at the output data as well once the job is complete.

A screenshot of a computer

Description automatically generated

The results are ultimately stored in the Azure Blob Store container.

A screenshot of a computer

Description automatically generated

### Clean the data

Add all the columns to cleanup any missing data

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Step 3: Split Data & Train the Model

1. Connect the Cleaned Dataset from Clean Missing Data to Split Data

A screenshot of a data flow

Description automatically generated

1. Split data using a 60/40 split.

A screenshot of a computer

Description automatically generated

* + 60% should go to a filter using Pearson correlation
  + 40% should be used as test

### Feature Selection

1. Create a Pearson correlation Feature selection using **rating** as a target column (select columns to transform and apply transformation).

A screenshot of a computer

Description automatically generated

Select Columns to Transform

A screenshot of a computer

Description automatically generated

Apply Transformations

A screenshot of a computer

Description automatically generated

### Model Training

1. Create a Boosted Decision Tree Regression with the following settings:
   * Create trainer mode: SingleParameter
   * Maximum number of leaves per tree: *20*
   * Minimum number of leaves per tree: *10*
   * Learning rate: *0.2*
   * Total number of trees constructed: *100*

A screenshot of a computer

Description automatically generated

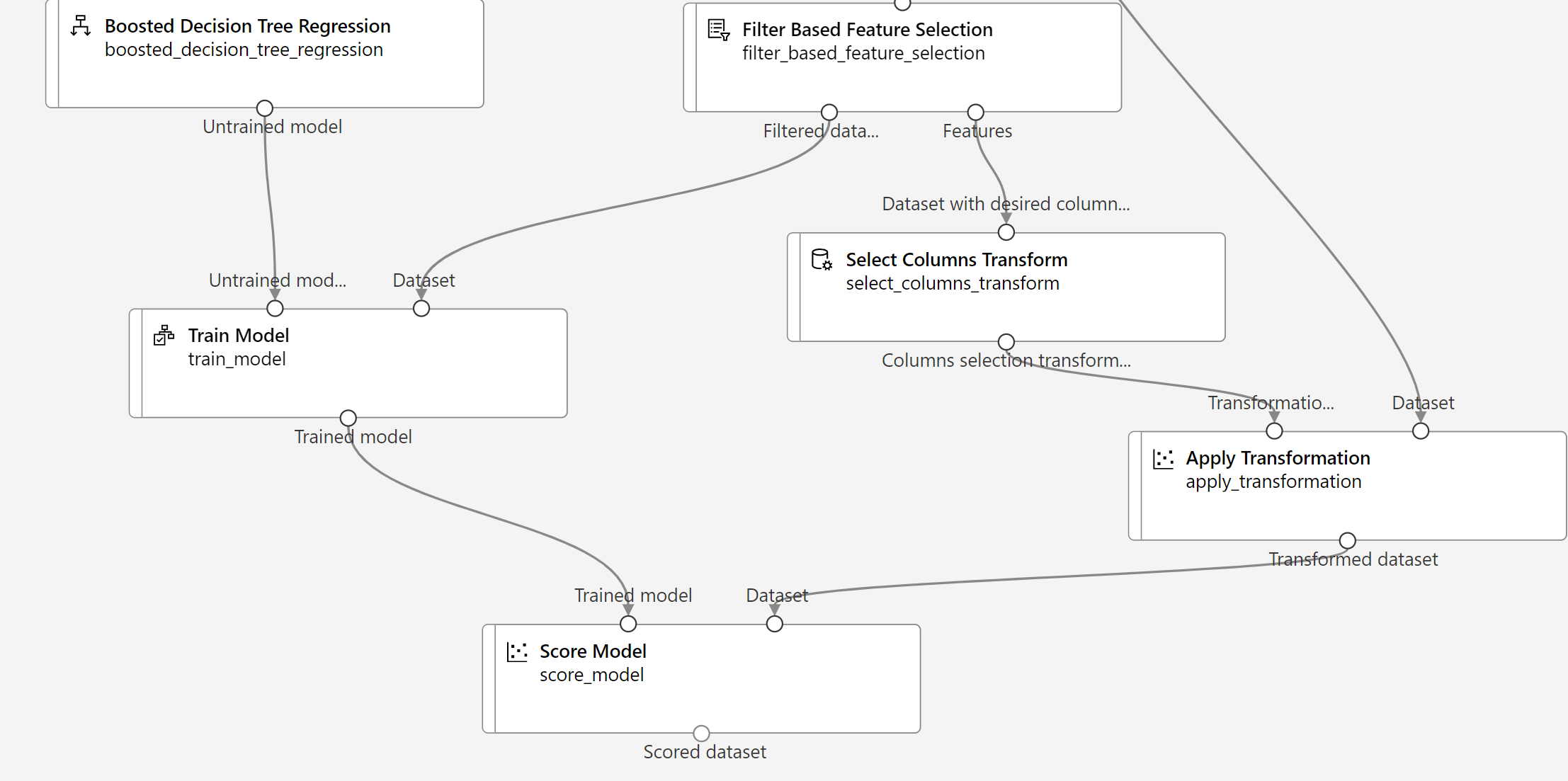
1. Create Train Model using rating as label column.

A screenshot of a computer program

Description automatically generated

## Step 4: Scoring and Evaluate

1. Create Score Model activity.



1. Create Evaluate Model activity.

A screenshot of a computer

Description automatically generated

1. Submit Model.
2. Evaluate Results.

A screenshot of a computer

Description automatically generated

Improvements can be made further.

A screenshot of a computer

Description automatically generated

## Step 5: Register the Model Registry

1. Navigate to the Jobs and get the latest experiment run.
2. Select train model
3. Look at Outputs + Logs and get the model type

A screenshot of a computer

Description automatically generated

1. Now, click on the Register model.

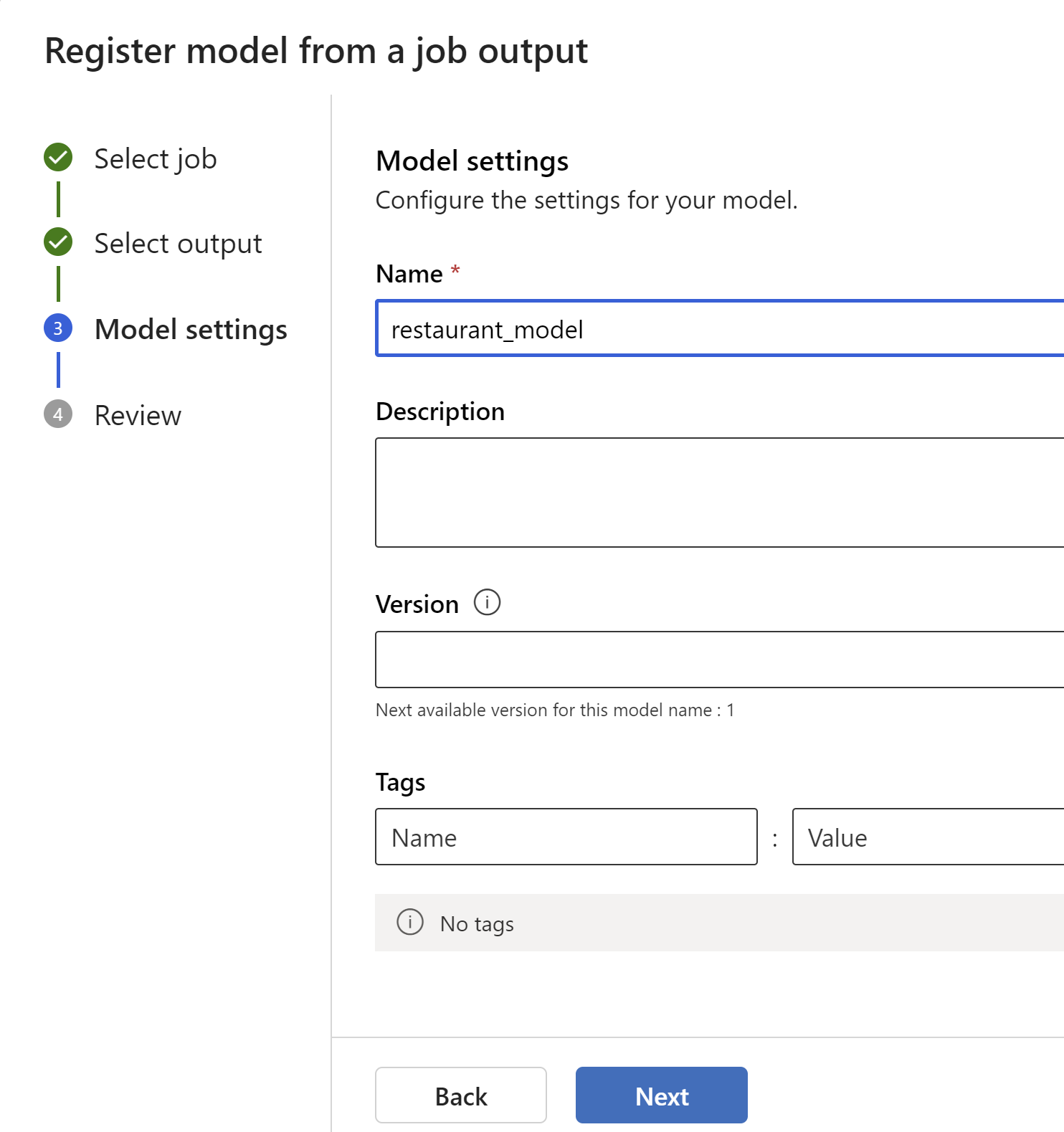
A screenshot of a computer

Description automatically generated

1. Get the trained model outputs

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

## Step 6: Deploy Endpoint

1. Create a compute cluster

A screenshot of a computer

Description automatically generated

1. Download conda and score.py files
2. Create a custom environment

A screenshot of a computer

Description automatically generated

Copy conda file contents

A screenshot of a computer

Description automatically generated

After creation,

A screenshot of a computer

Description automatically generated

Navigate to Model List, and select the model that you have recently created.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Review & then Create.

Create a batch scoring job

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Next step is to move the model from Dev to QA.

# Using ML Notebooks

References: Download templates from here: <https://github.com/Azure/mlops-templates>

Download the relevant code base from here: <https://github.com/Azure/mlops-project-template/tree/main/classical>

## Step 1: Configure your environment

1. If you do not have an existing environment, use Terraforms to create a new one. infrastructure\terraform\github-actions\tf-gha-deploy-infra.yml
2. Set your config-infra-dev.yml & config-infra-prod.yml
3. Run your mlops\github-actions\deploy-model-training-pipeline-classical.yml

