

Cloud Application Development

Phase 3: Project Development Part 1

Project title:

Machine learning model deployment with IBM cloud Watson Studio.

Problem Statement:

Become a wizard of predictive analytics with IBM Cloud Watson Studio. Train machine learning models to predict the outcomes in real time. Deploy the models as web services and integrate them into your applications. Unlock the magic of data driven insights and make informed decisions like never before.

House Price Prediction Analysis - Part 1

- Project overview
- Data Understanding
- Data Visualization
- Data Preparation
- Modeling
- Evaluation

Project Overview:

House Price Prediction Analysis aims to use Machine learning analysis algorithms to predict the price of houses based on their features like number of rooms, number of bedrooms, age of the house, population of the respective area where the house is located, location of the house and the area income with other relevant factors if available. By this Machine Learning model user can predict the price of the house that can be sold.

Data Understanding:

Data analysis is the process of inspecting, cleaning, transforming, and modeling data to uncover useful information, draw conclusions, and support decision-making.

- Data visualization
- Exploratory Data Analysis

Data Visualization:

Data visualization plays a crucial role in understanding the data, identifying trends, and communicating insights.

- Histograms and Distributions

- Scatter Plots
- Correlation Matrix
- Box Plots
- Time Series Plots
- Geospatial Visualization
- Feature Importance Plot
- Residual Plots
- Interactive Dashboards
- Comparison Charts

Data Preparation:

1. Data Cleaning:

- Handle missing data.
- Address outliers.

2. Feature Engineering:

- Create relevant new features.
- Encode categorical variables.
- Scale or normalize numerical features.

3. Data Splitting:

- Split the data into training, validation, and test sets.

4. Normalization.

5. Feature Selection.

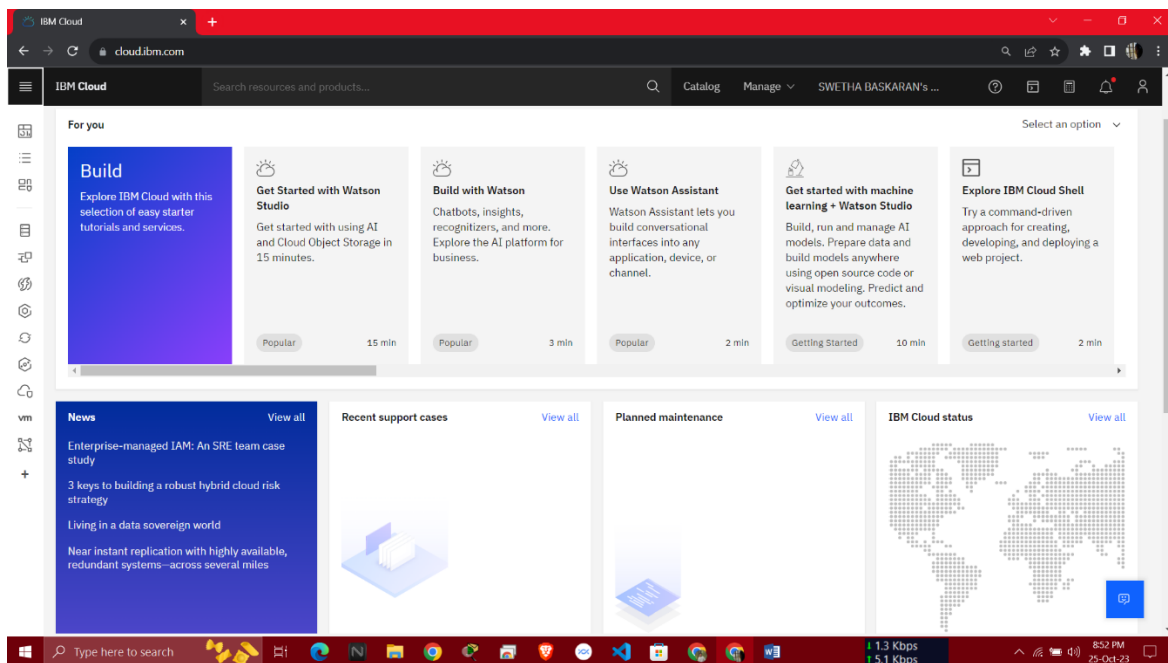
6. Documentation:

- Keep records of data preparation steps for reproducibility.

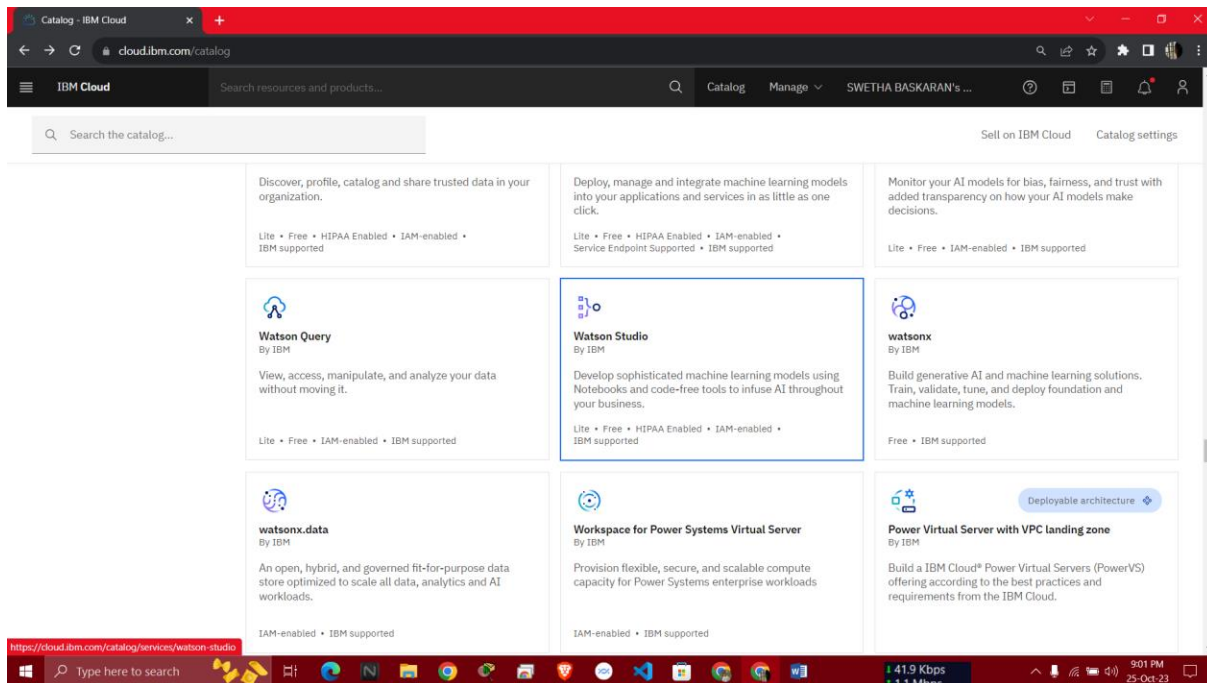
Modeling and Evaluation:

Step wise process for the House Price Prediction Analysis Machine learning model deployment.

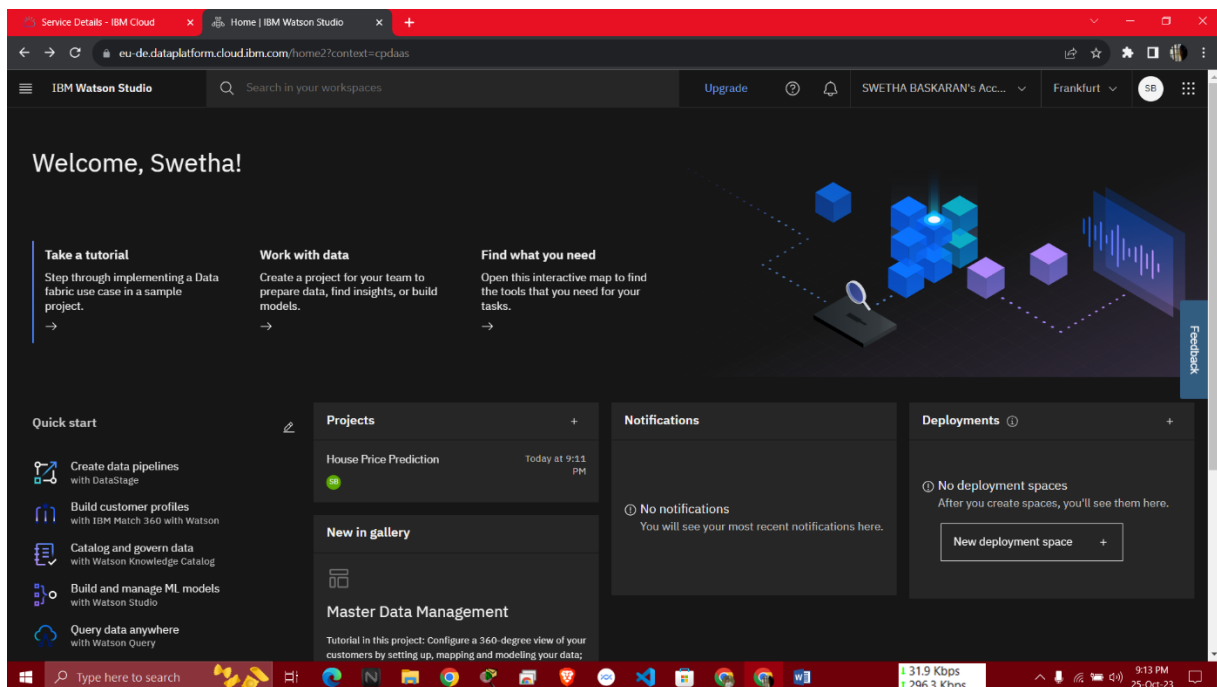
Step 1: Login to IBM cloud



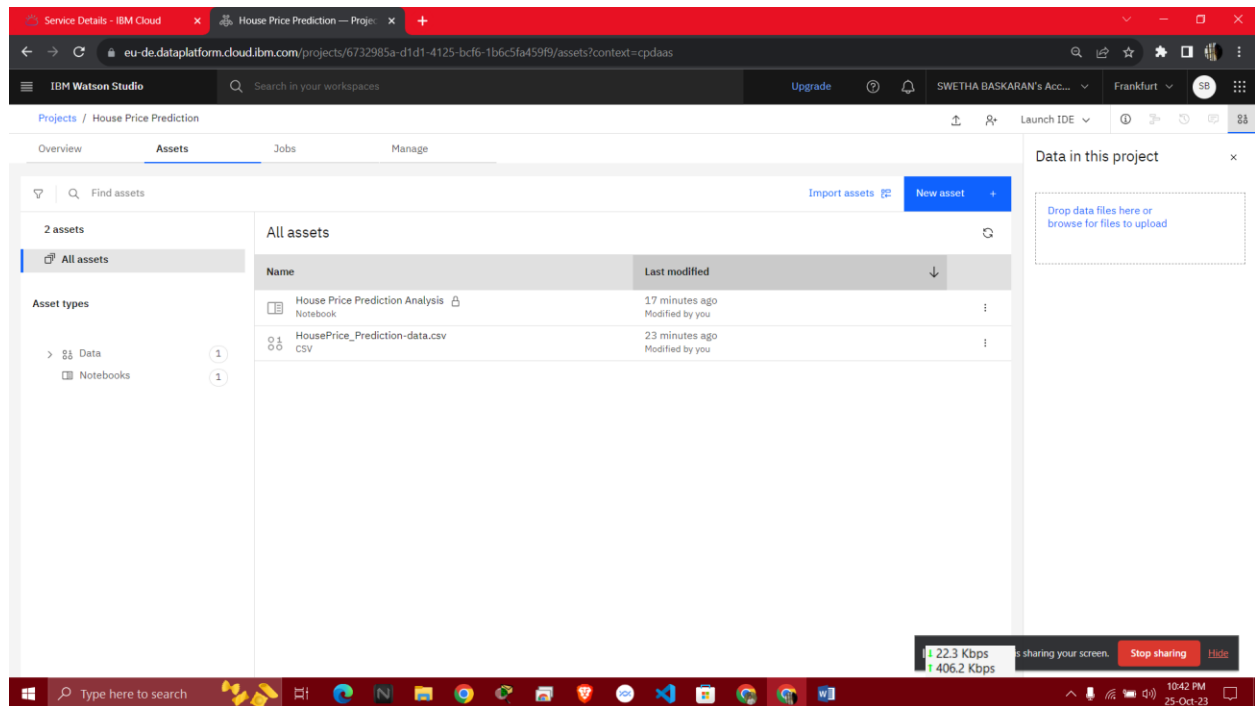
Step 2: Go to catalog and create a Watson Studio service in AI category.



Step 3: Get started to launch Watson Studio Dashboard.



Step 4: Create a project in IBM Watson Studio in IBM Cloud and assign a Cloud object storage to manage datasets.

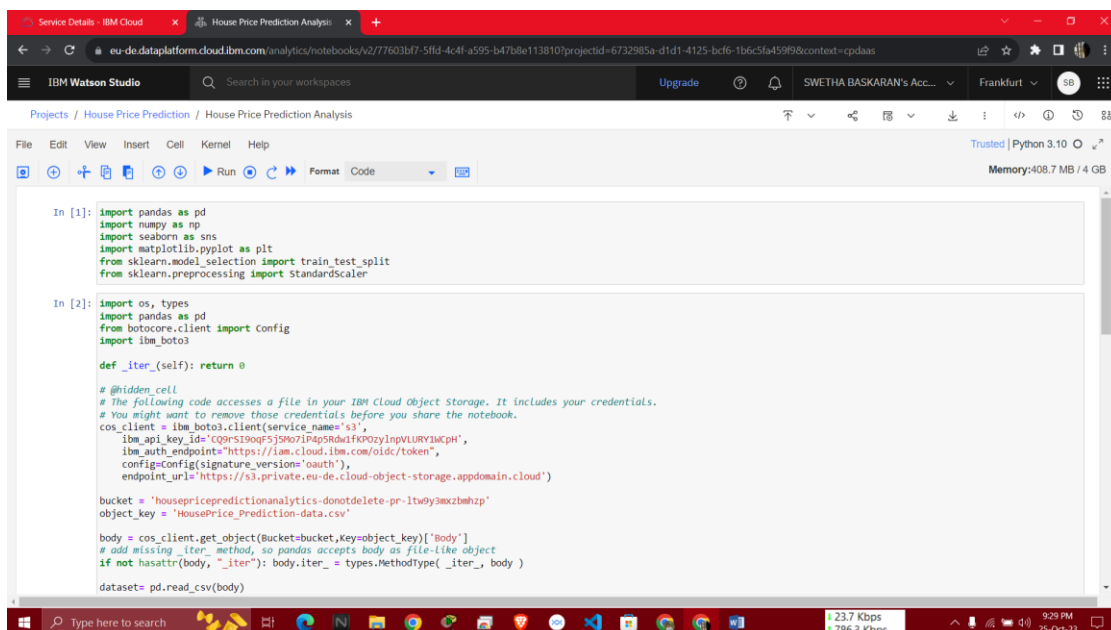


Cloud Object Storage is a storage service in IBM Cloud. We use this service to manage our datasets for training the ML Model and store required files.

Step 5: Add a jupyter notebook instance in your project to Develop and Deploy Machine Learning Model.

i. Import necessary library packages.

- import pandas as pd: For handling your dataset.
- import numpy as np: Useful for numerical operations.
- from sklearn.model_selection import train_test_split: Split your data into training and testing sets.
- import seaborn as sns: for data visualization in python.
- import matplotlib.pyplot as plt: another powerful library for creating visualizations.



```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

In [2]: import os, types
import pandas as pd
from botocore.client import Config
import boto3

def _iter_(self): return 0

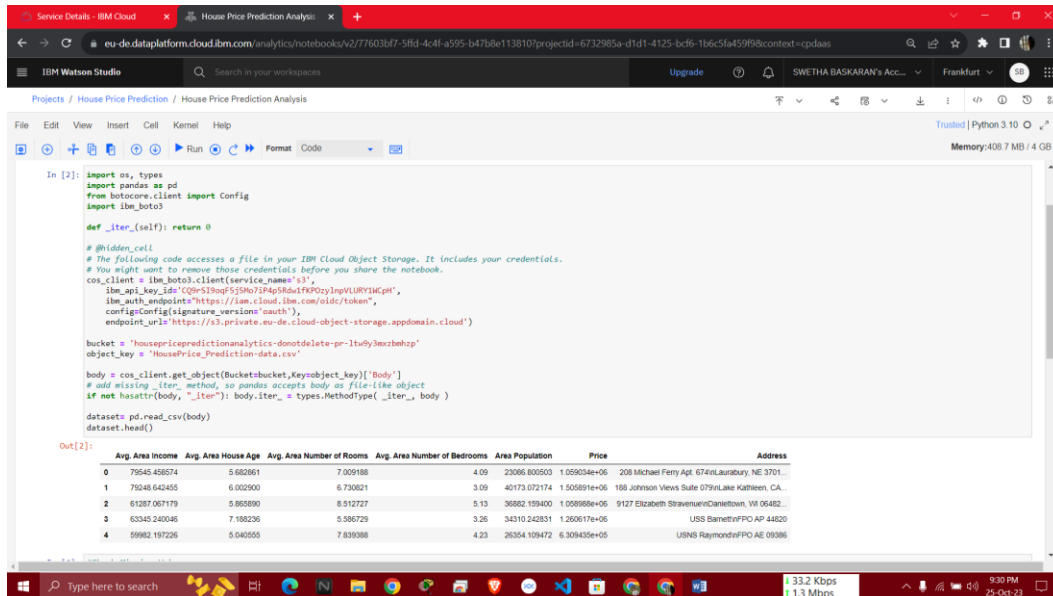
#@hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = boto3.client(service_name='s3',
    aws_access_key_id='cosStoogSj9w7IPapsdaifK0zylnpVLURYACph',
    aws_secret_access_key='ibm_auth_endpoint=https://iam.cloud.ibm.com/oidc/token',
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.private.eu-de.cloud-object-storage.appdomain.cloud')

bucket = 'housepricepredictionanalytics-donotdelete-pr-lt9w3mx2mh2p'
object_key = 'HousePrice_Prediction-data.csv'

body = cos_client.get_object(Bucket=bucket,Key=object_key)['body']
# add missing _iter_ method, so pandas accepts body as file-like object
if not hasattr(body, "_iter"): body._iter_ = types.MethodType(_iter_, body)

dataset = pd.read_csv(body)
```

- ii. Import dataset and proceed further with pre-processing steps and build the model.



The screenshot shows an IBM Watson Studio notebook titled 'House Price Prediction Analysis'. The code in the cell imports necessary libraries (os, types, pandas, boto3, botocore) and defines a function to access an S3 bucket containing the dataset. The output shows the first five rows of the dataset, which includes columns for area income, house age, number of rooms, bedrooms, population, price, and address.

```
In [2]: import os, types
import pandas as pd
from botocore.client import import Config
import boto3

def _iter_(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
cos_client = boto3.client(service_name='s3',
                           aws_access_key_id='COS3S0p55807184p4k4d4r4n4c4y4p4V4U4Y4M4p4',
                           aws_secret_access_key='COS3S0p55807184p4k4d4r4n4c4y4p4V4U4Y4M4p4',
                           config=Config(signature_version='s3v4'),
                           endpoint_url='https://s3.private.eu-de.cloud-object-storage.appdomain.cloud')

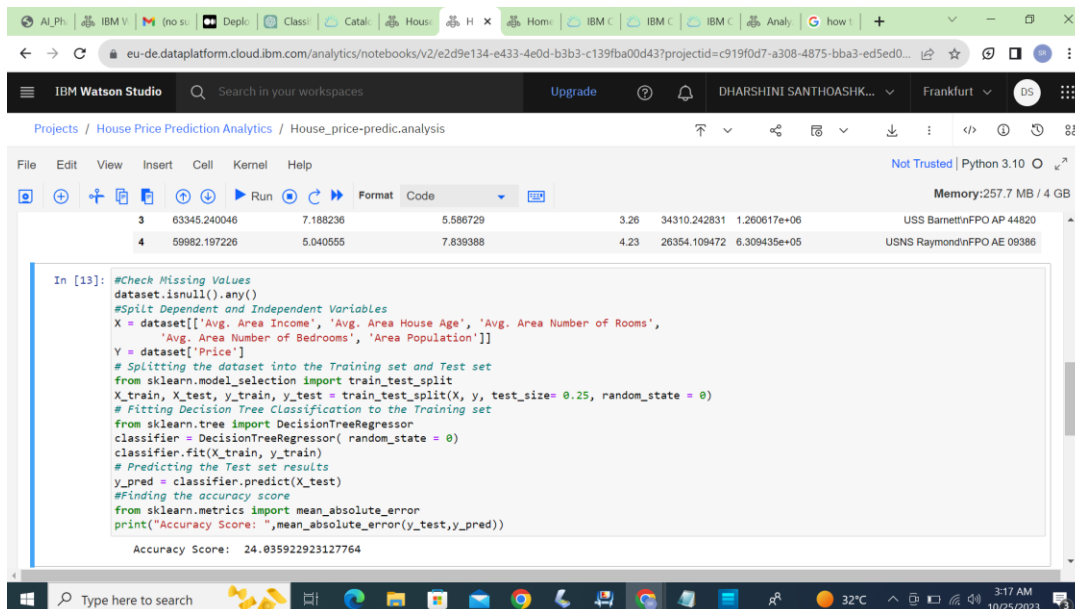
bucket = 'housepricepredictionanalytics-donotdelete-pr-1tu0y3mzbehp'
object_key = 'HousePricePrediction-data.csv'

body = cos_client.get_object(Bucket=bucket, Key=object_key)['Body']
# add missing _iter_ method, so pandas accepts body as file-like object
if not hasattr(body, '_iter_'): body._iter_ = types.MethodType(_iter_, body)

dataset = pd.read_csv(body)
dataset.head()
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674mLaunbury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.508919e+06	168 Johnson Views Suite 079mLake Kathleen, CA...
2	61287.067179	5.866890	8.512727	5.13	36882.159400	1.058986e+06	9127 Elizabeth StravenueinCamelton, WI 06482...
3	63345.240046	7.188236	5.986729	3.26	34310.242831	1.260617e+06	USS BarnettinFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS RaymondinFPO AE 09386

Step 6: Train the build model and evaluate them.



The screenshot shows the same IBM Watson Studio notebook with the next cell of code. This cell performs data preprocessing (checking for nulls, splitting into training and testing sets) and builds a Decision Tree Regressor model. The output shows the accuracy score of the model.

```
In [13]: #Check Missing Values
dataset.isnull().any()
#Split Dependent and Independent Variables
X = dataset[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
              'Avg. Area Number of Bedrooms', 'Area Population']]
Y = dataset['Price']
# Splitting the dataset into the Training set and Test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.25, random_state = 0)
# Fitting Decision Tree Classification to the Training set
from sklearn.tree import DecisionTreeRegressor
classifier = DecisionTreeRegressor( random_state = 0)
classifier.fit(X_train, y_train)
# Predicting the Test set results
y_pred = classifier.predict(X_test)
#Finding the accuracy score
from sklearn.metrics import mean_absolute_error
print("Accuracy Score: ",mean_absolute_error(y_test,y_pred))

Accuracy Score: 24.035922923127764
```

Model was build trained and tested.

Conclusion:

The housing price prediction project has successfully demonstrated the ability to forecast real estate prices with a reasonable degree of accuracy. Through the utilization of machine learning algorithms and a robust dataset, this project has provided valuable insights into factors influencing housing prices. It can serve as a valuable tool for both prospective homebuyers and real estate professionals seeking to make informed decisions in the dynamic housing market. However, it's important to acknowledge that market conditions can change over time, and ongoing model updates and data monitoring are essential to maintain predictive accuracy.