HOUSE PRICE PREDICTION USING IBM WATSON STUDIO

(AUTO AI MODEL)



Submitted By:  
Vaishnavi Kesarwani  
B.Tech CSE  
United College of Engineering and Research

Mentor:  
Mr. Samarth Amrute

Acknowledgement

I would like to express my sincere gratitude to my mentor, Mr. Samarth Amrute, for his continuous support and guidance throughout this project. His mentorship has been a valuable source of motivation and direction.

I would also like to express my heartfelt gratitude to **IBM** for providing me with this incredible opportunity to intern in the field of Artificial Intelligence. This internship has not only given me exposure to real-world AI applications but also enabled me to work hands-on with cutting-edge tools like **IBM Watson Studio's Auto AI**.

This experience has been instrumental in enhancing my skills in machine learning, cloud-based platforms, and automated model development, and has greatly contributed to my academic and professional growth.

Index

1. Problem Statement

2. Methodology

3. Dataset Details

4. Coding Steps

5. Algorithm/Model Used

6. Result

7. Conclusion

8. Reference

Problem Statement

In the real estate industry, accurately predicting house prices is crucial for buyers, sellers, investors, and financial institutions. Making informed decisions in property transactions requires access to reliable price estimates, which depend on numerous dynamic and location-specific factors.

The objective of this project is to develop a robust, machine learning-based predictive model using **IBM Watson Studio’s Auto AI** that can efficiently estimate house prices. The model takes into account various property features such as the number of bedrooms, bathrooms, square footage, lot size, number of floors, year built, renovation details, and geographic location, among others.

By automating the model-building process, Auto AI not only saves time and effort but also ensures optimal model selection, hyperparameter tuning, and performance evaluation. This project aims to demonstrate how cloud-based AI tools can simplify complex data science workflows and produce accurate, scalable, and interpretable predictions that can be utilized in real-world housing markets.

Methodology

The methodology adopted in this project is a systematic approach to building a predictive model using IBM Watson Studio's Auto AI. The key steps involved are:

1. **Data Collection**: A structured dataset containing house-related attributes such as number of bedrooms, bathrooms, area (sqft), location, and year of construction was obtained from a publicly available source.
2. **Data Upload**: The dataset was uploaded to **IBM Cloud Object Storage** and connected to Watson Studio for further processing.
3. **Auto AI Configuration**: Within Watson Studio, Auto AI was used to automate the machine learning pipeline. The target variable (house price) was selected, and the system automatically identified input features.
4. **Data Preprocessing**: Auto AI handled missing values, encoded categorical variables, and performed necessary feature engineering without manual intervention.
5. **Model Generation**: Multiple models were generated using different algorithms like Linear Regression, Decision Trees, and Ensemble methods. Auto AI ranked them based on performance metrics.
6. **Hyperparameter Optimization**: The top-performing models underwent automatic hyperparameter tuning to improve accuracy and generalizability.
7. **Evaluation and Visualization**: The best model was selected based on metrics such as RMSE and R² score. The results were visualized using Auto AI's built-in leaderboard and graphs.
8. **Deployment**: The best-performing model pipeline was downloaded from IBM Watson Studio and deployed using **Google Colab**. The Colab notebook allows for real-time inference and integration of the trained model in a lightweight environment, making it accessible for testing and further development.

Dataset Details

The dataset consists of multiple features relevant to house pricing, including:

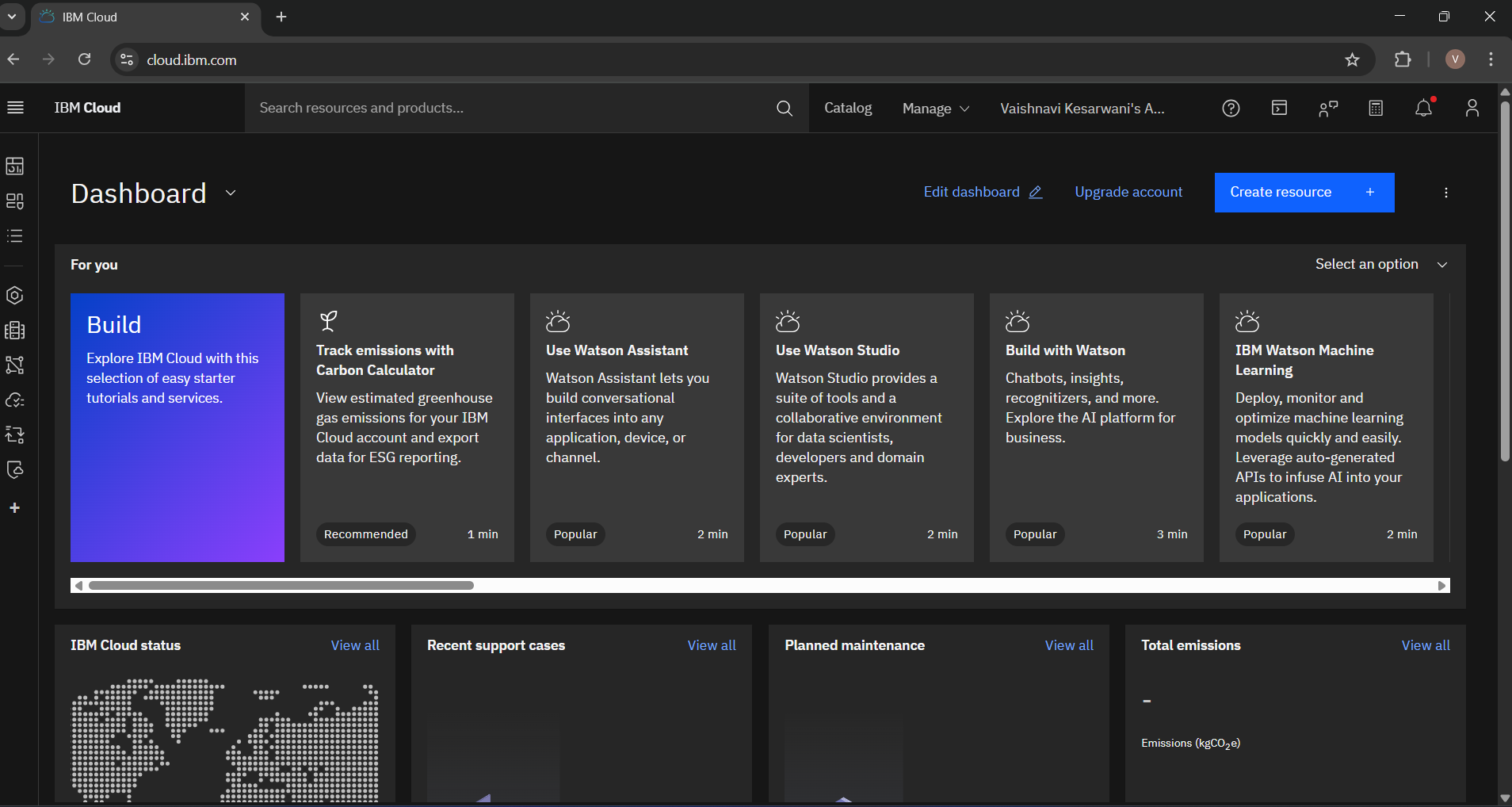
* **Date**: The date of the house sale, useful for analyzing market trends over time.
* **Number of Bedrooms and Bathrooms**: Indicators of property size and comfort.
* **Square Foot Living and Lot Area**: Measure of indoor living space and outdoor area, respectively.
* **Number of Floors**: Reflects the size and type of house.
* **Waterfront View**: Binary feature indicating if the house has a waterfront view, which significantly impacts value.
* **Property Condition and Grade**: Qualitative indicators of the house's structural condition and construction quality.
* **Year Built and Renovated**: Help assess the age of the property and whether it has undergone recent improvements.
* **Location Information**: Including street address, city, state zip code, and country—critical for understanding geographical influence on pricing.

The dataset was cleaned and uploaded to **IBM Cloud Object Storage** and then connected to **IBM Watson Studio**. Using Watson Studio’s **Auto AI**, the data was automatically processed—handling missing values, encoding categorical variables, and scaling numerical features—before model building.

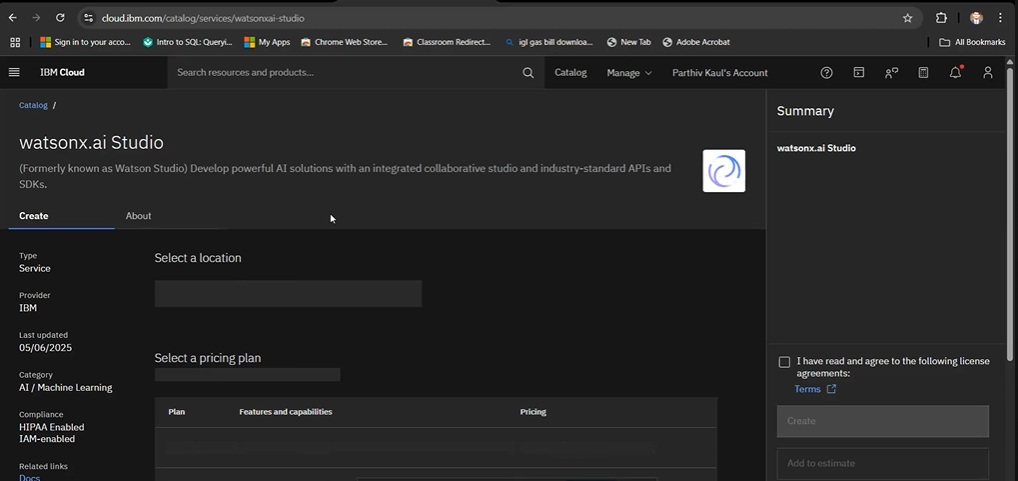
The size and diversity of this dataset allowed the model to generalize well and make accurate predictions across a wide range of property types and conditions.

Project Workflow

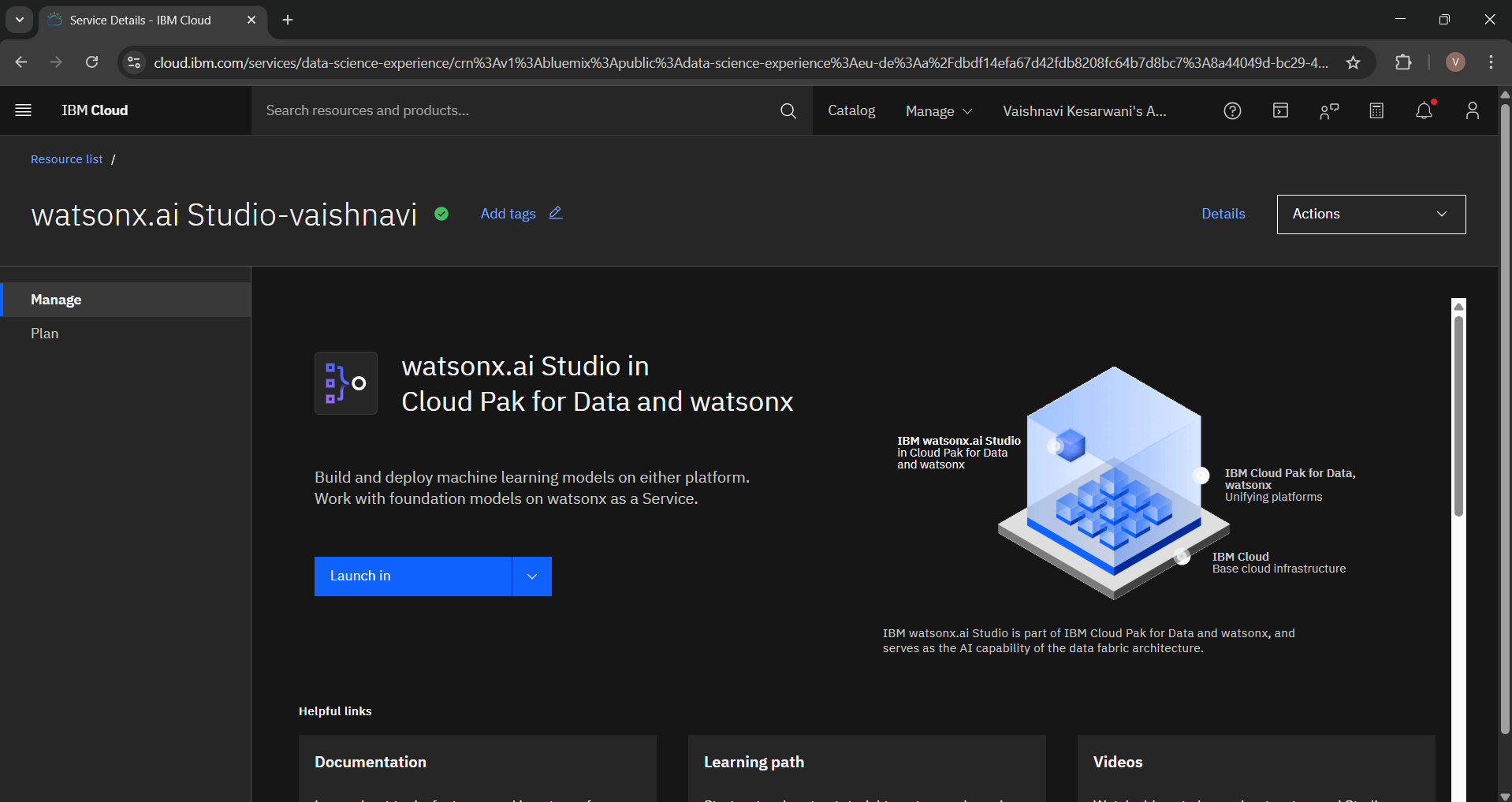
Step 1: Firstly, login with your IBM Id and password, then dashboard will appear. Search watsonx.ai Studio.



Step 2: After selecting watsonx.ai Studio, this will creates the following window where you can select your location, plan and rename it and mark as read and agree with the terms and finally create it.



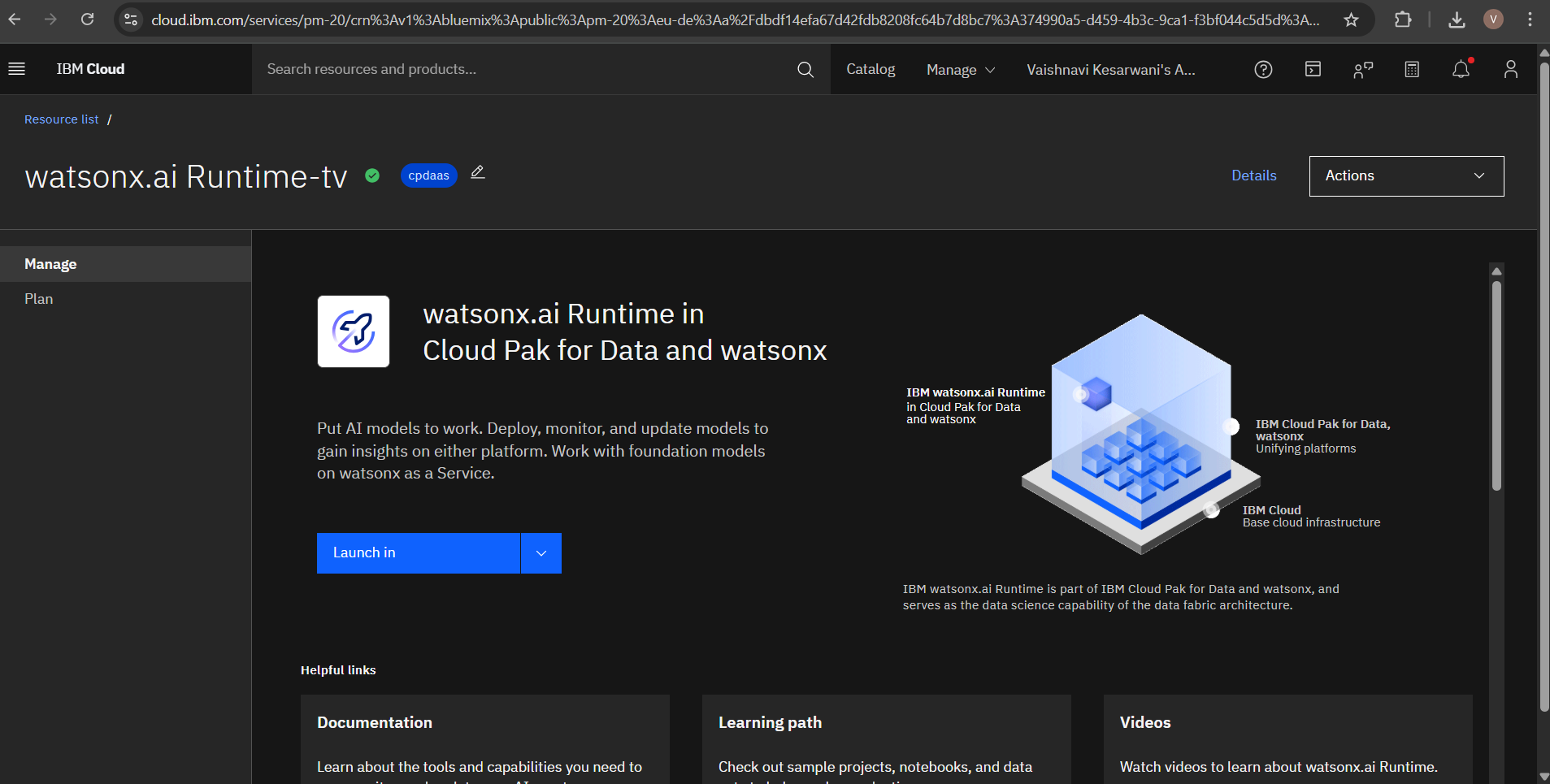
Step 3: After creating this window will appear.



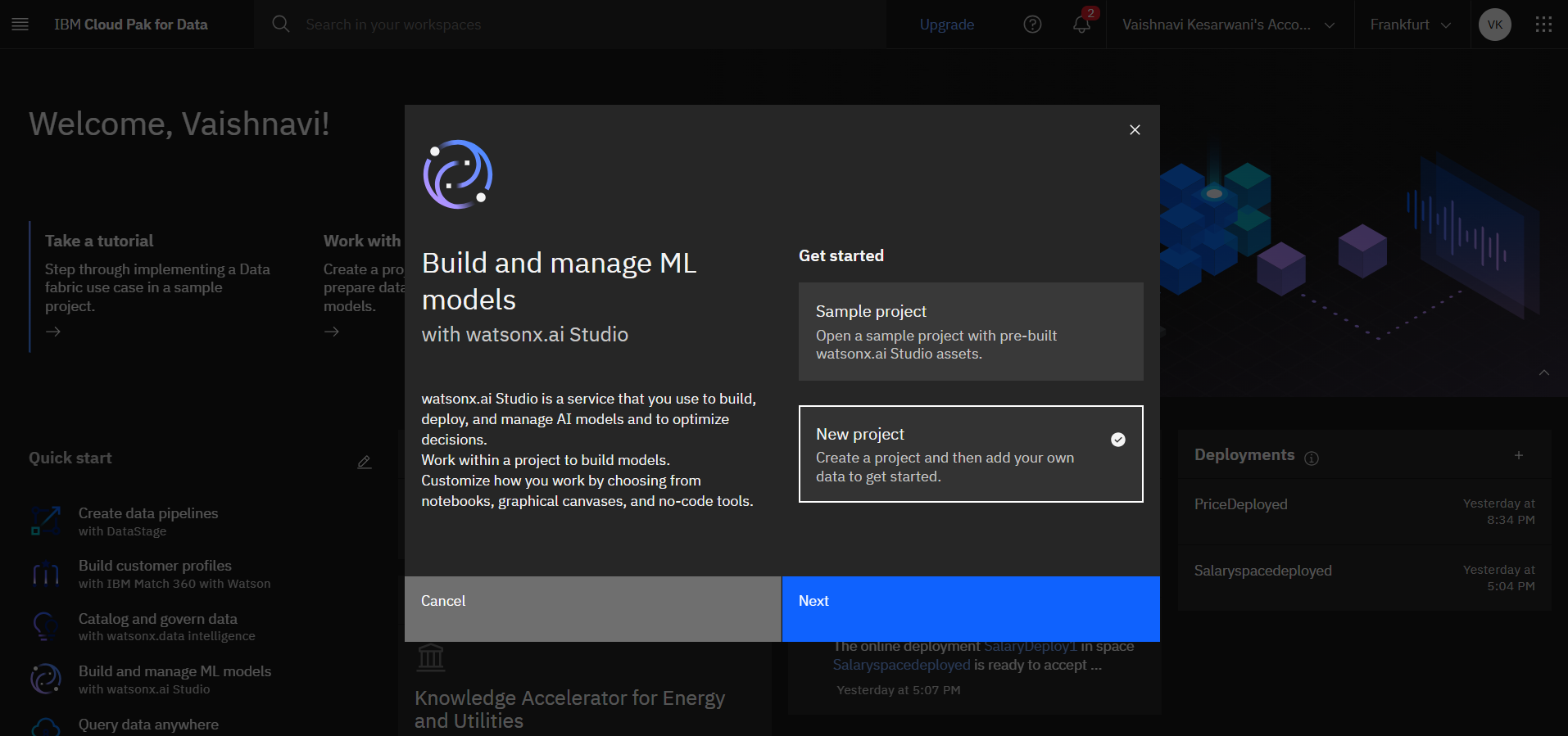
Step 4: Here you have two options for launching i.e. IBM Cloud Pak for Data and IBM watsonx .We will choose IBM Cloud Pak for Data and launch it.



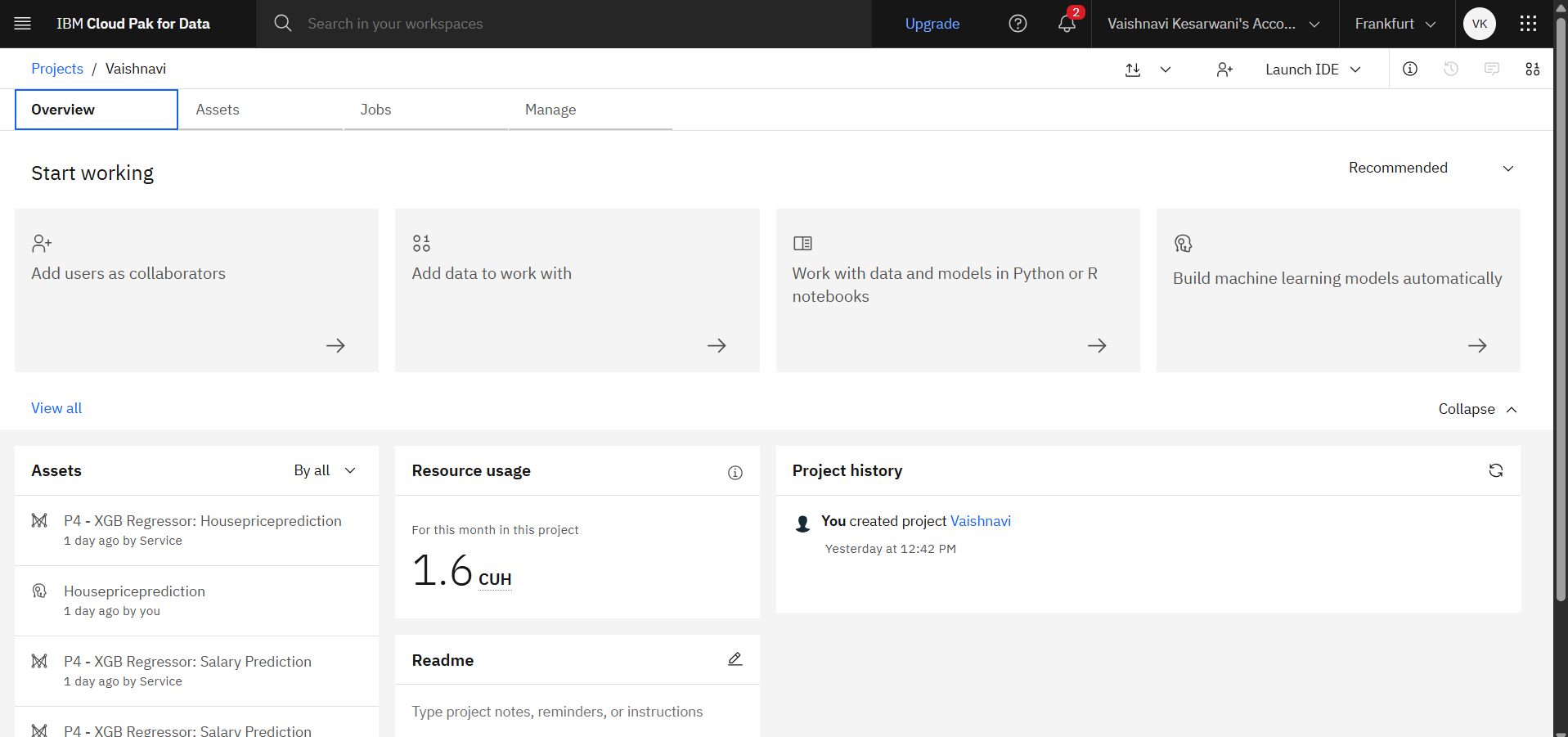
Step 5: Create watsonx.ai Runtime that enables seamless deployment, monitoring, and scaling of AI models built in IBM Watson Studio. It provides a secure, cloud-based environment to run models as APIs for real-time predictions.



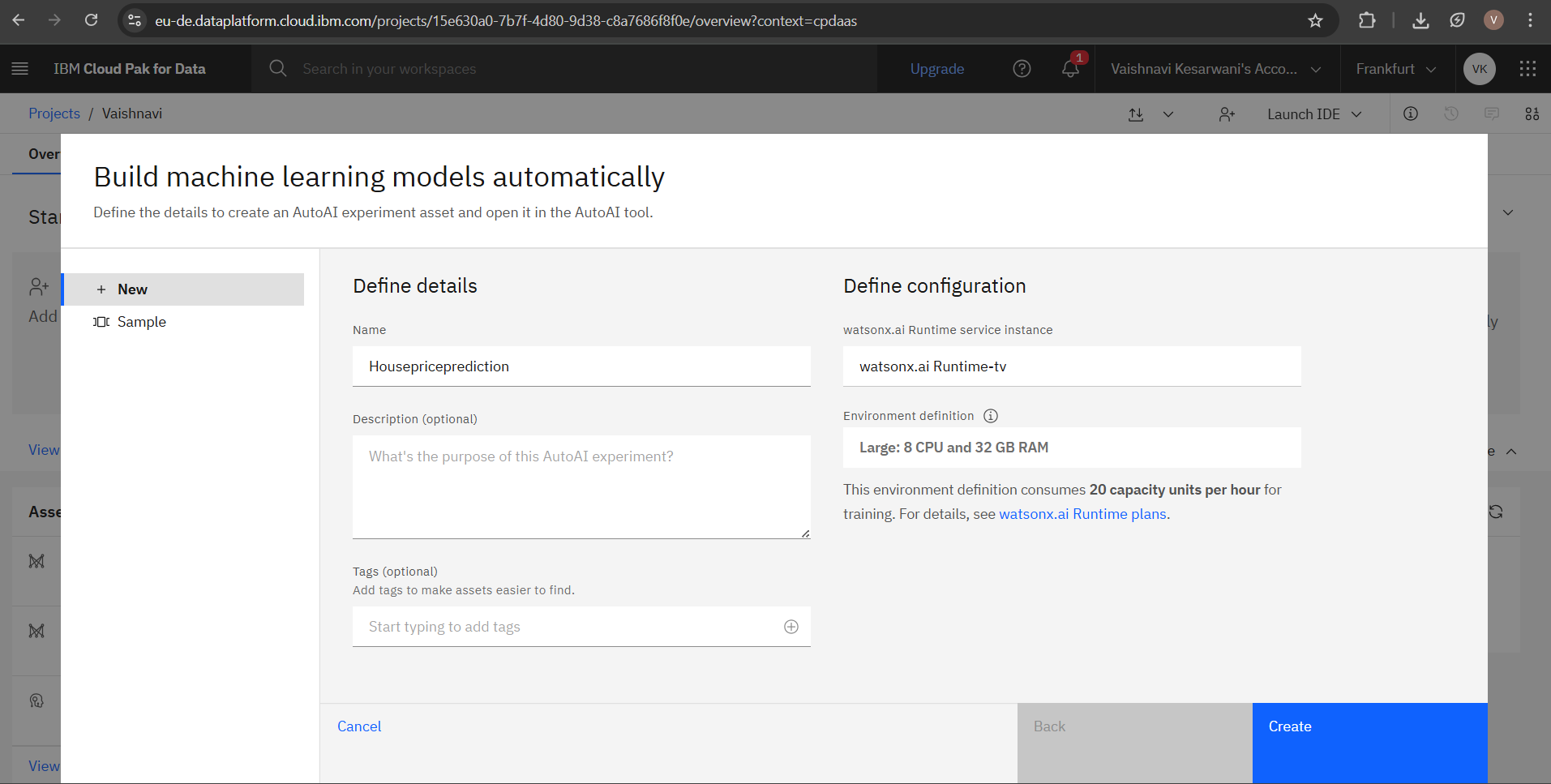
Step 6: After this, the following pop-up window will appear, then select new project and click on next.



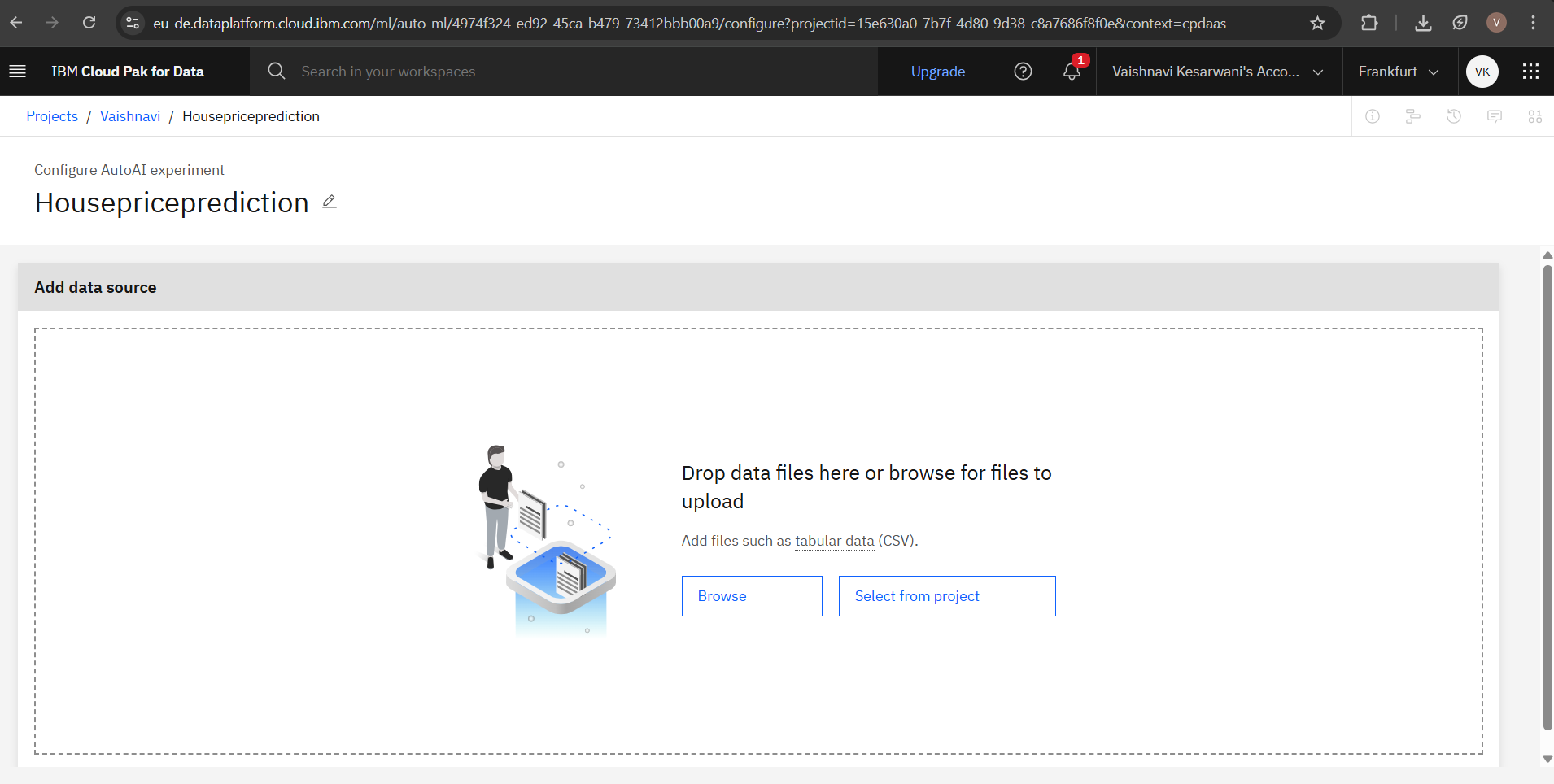
Step 6: IBM Cloud Pak for Data project dashboard showing model assets and this interface enables management of datasets, models, and real-time resource tracking during the AI development process.



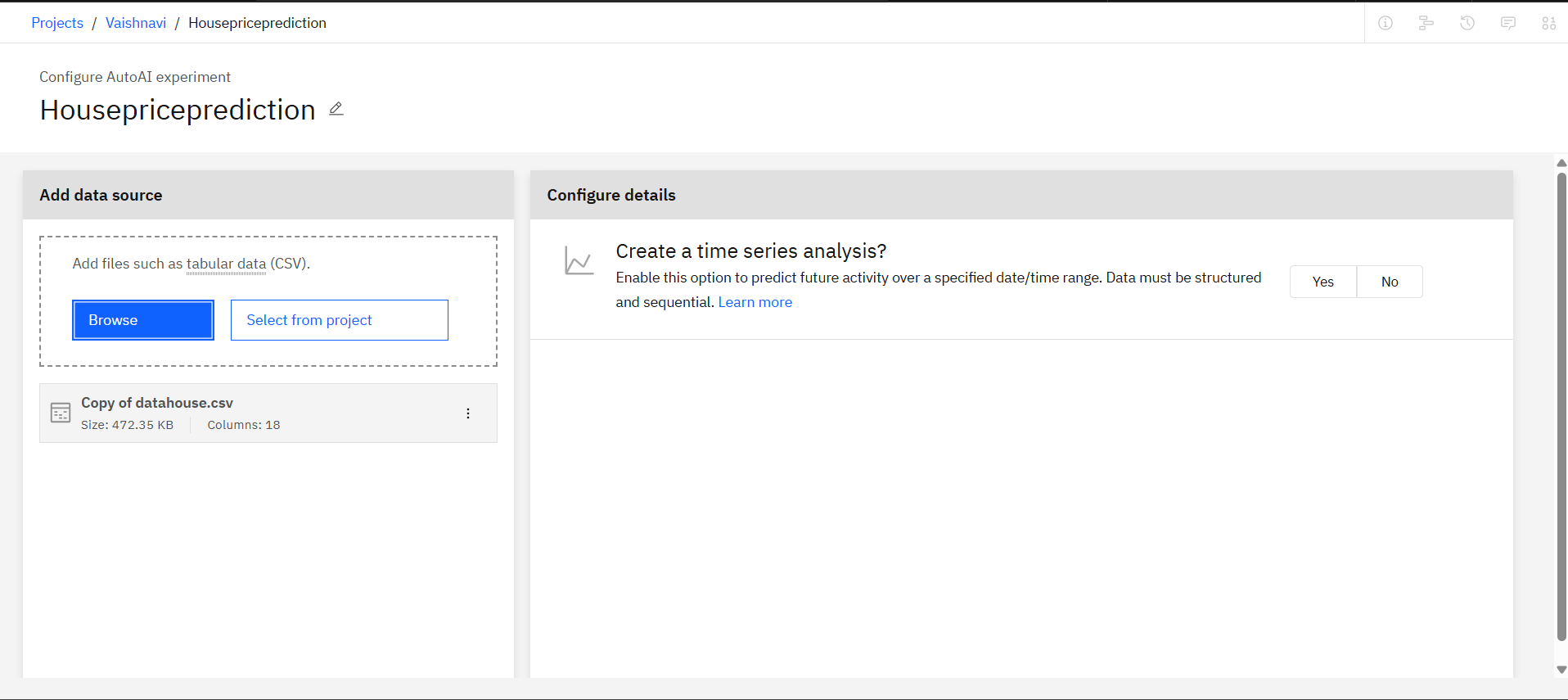
Step 7: Click on “Build machine learning models automatically” and associate service by selecting your runtime resource so that it gets associated with your project, then type name and click on create.



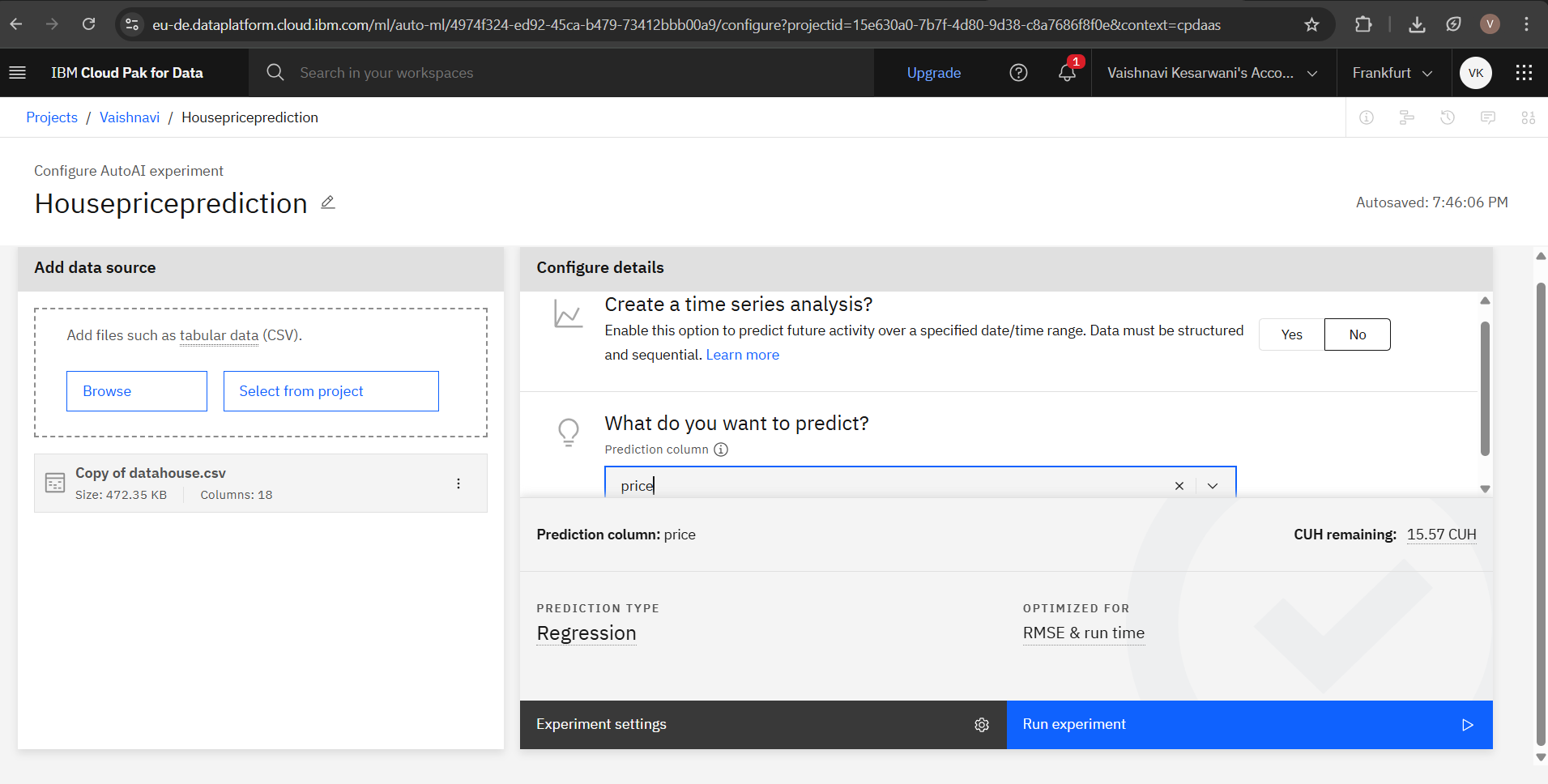
Step 8: Click on browse to upload dataset.



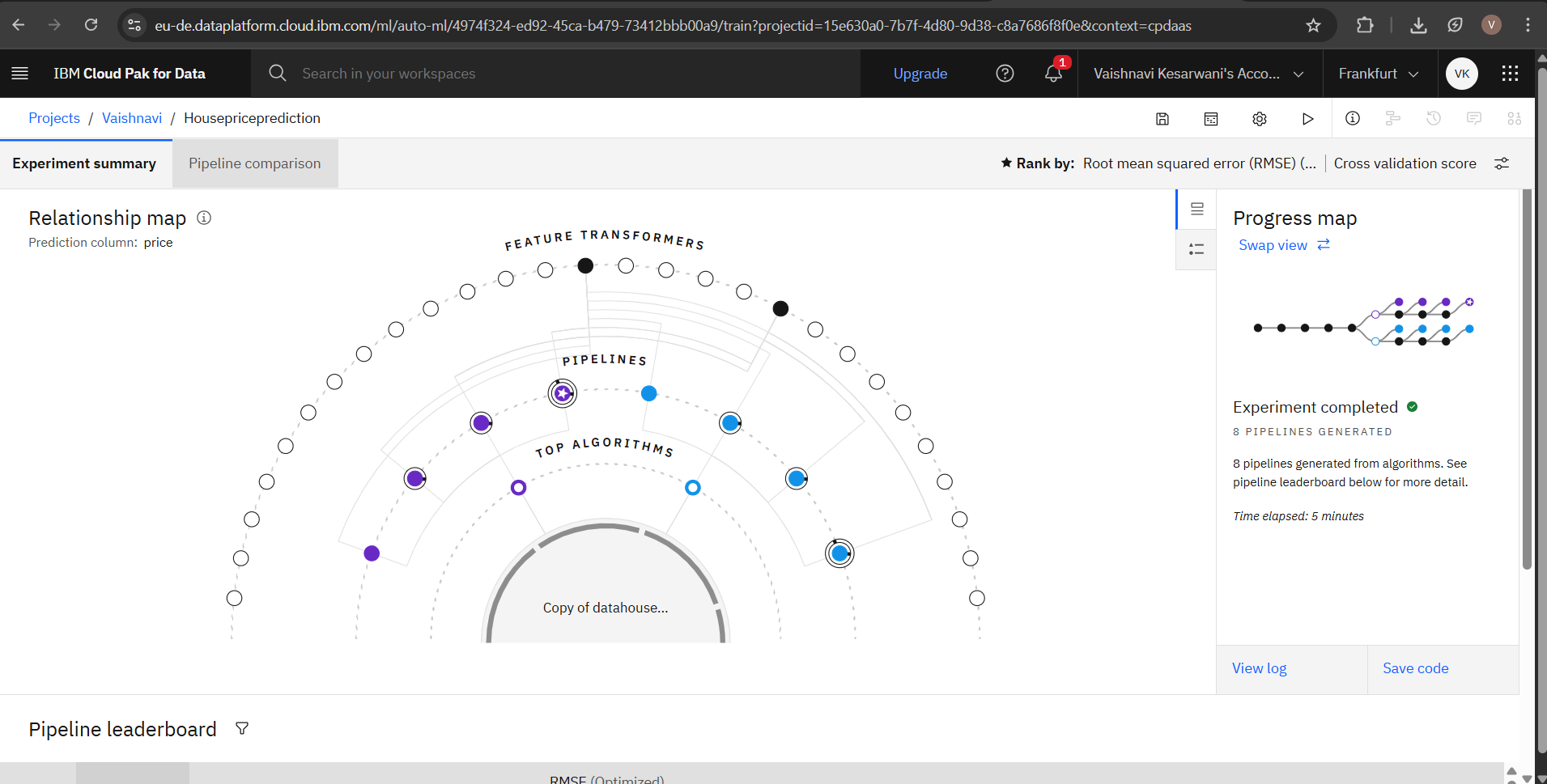
Step 9: Here, dataset (copy of datahouse) was uploaded.



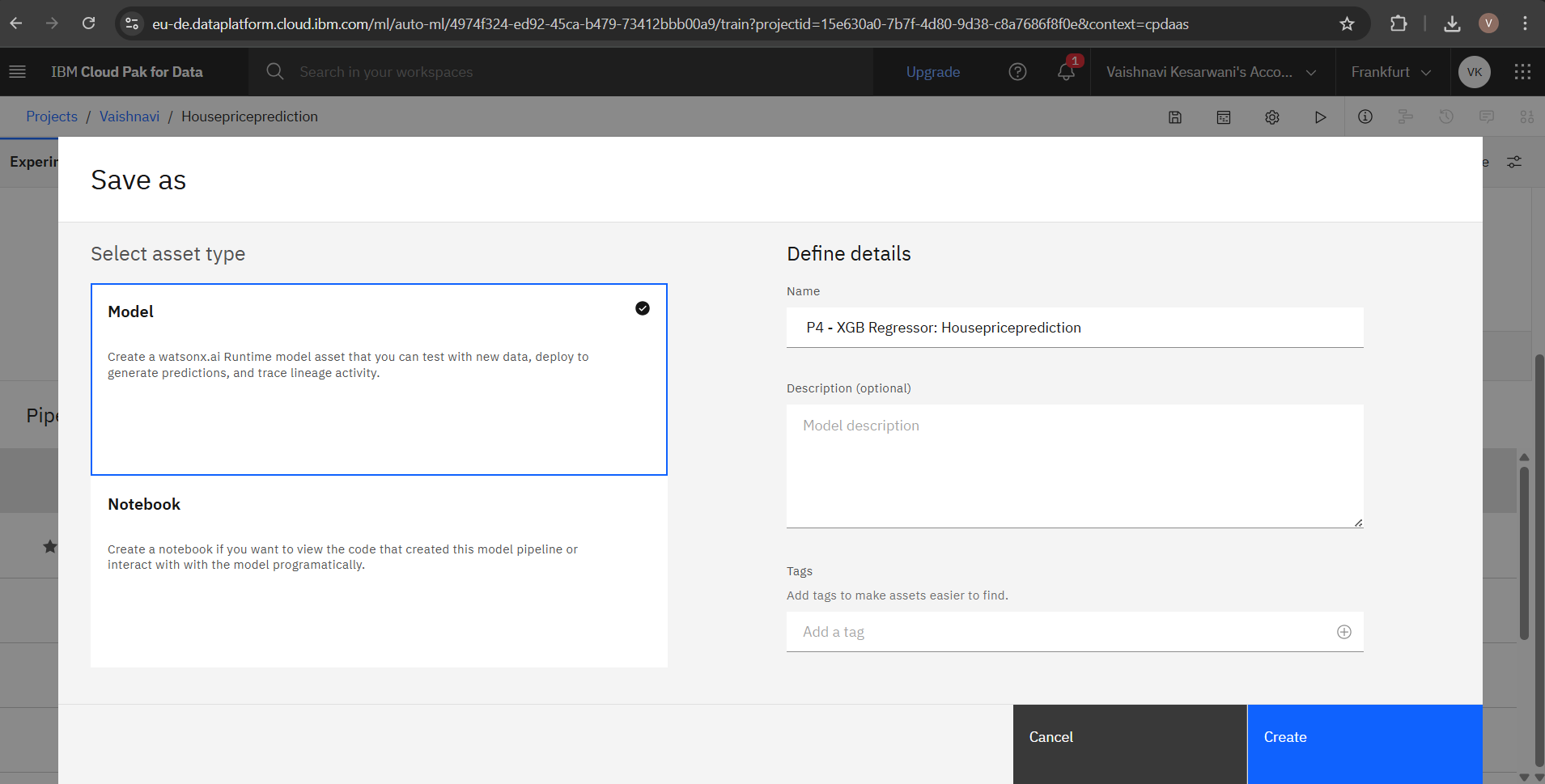
Step 10: Then click on NO and select price from prediction columns and finally click on run experiment.

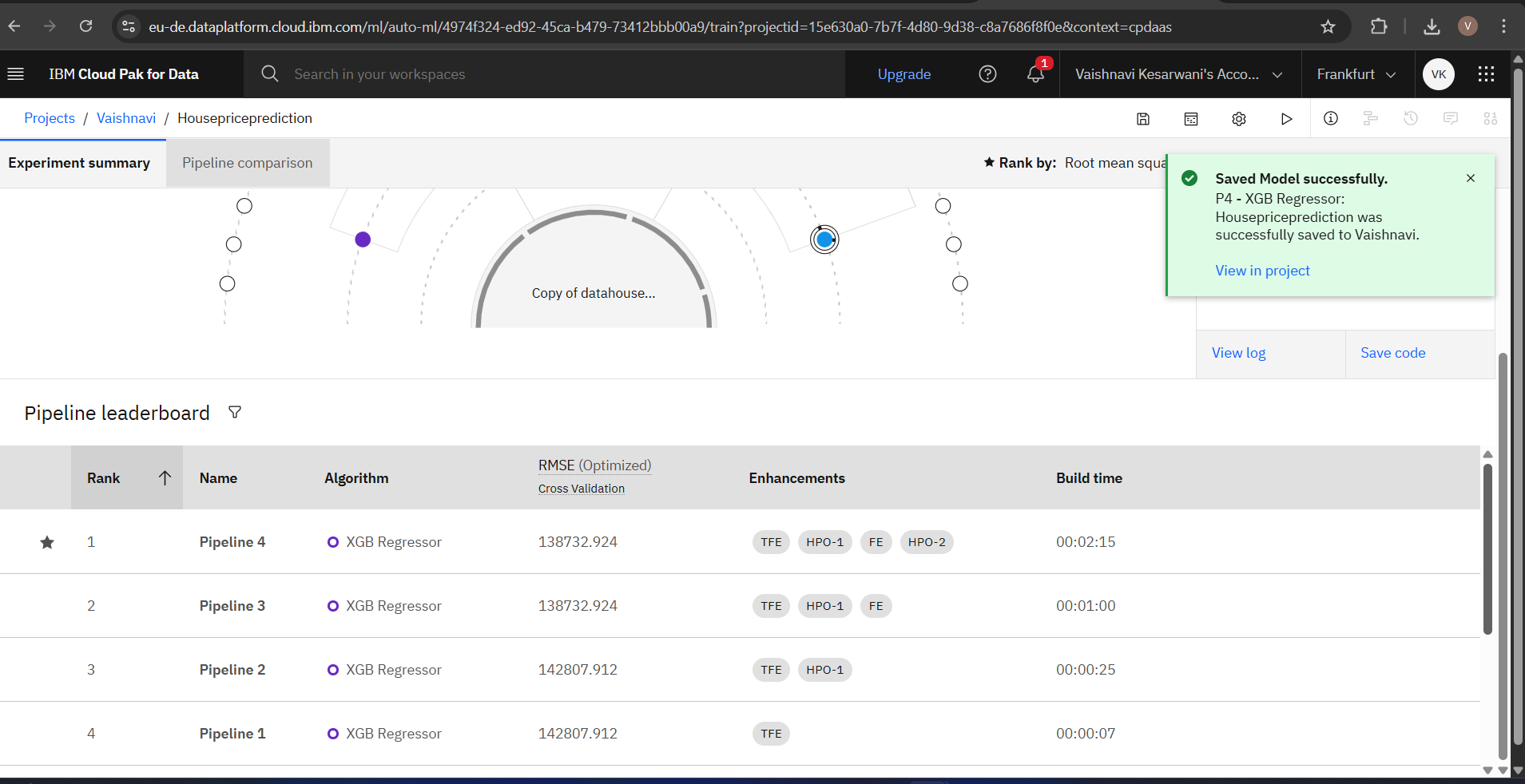


Step 11: AutoAI pipeline leaderboard displaying the best-performing model (Pipeline 4) as XGBoost Regressor.  
The model was selected based on lowest RMSE and highest R² score for house price prediction.

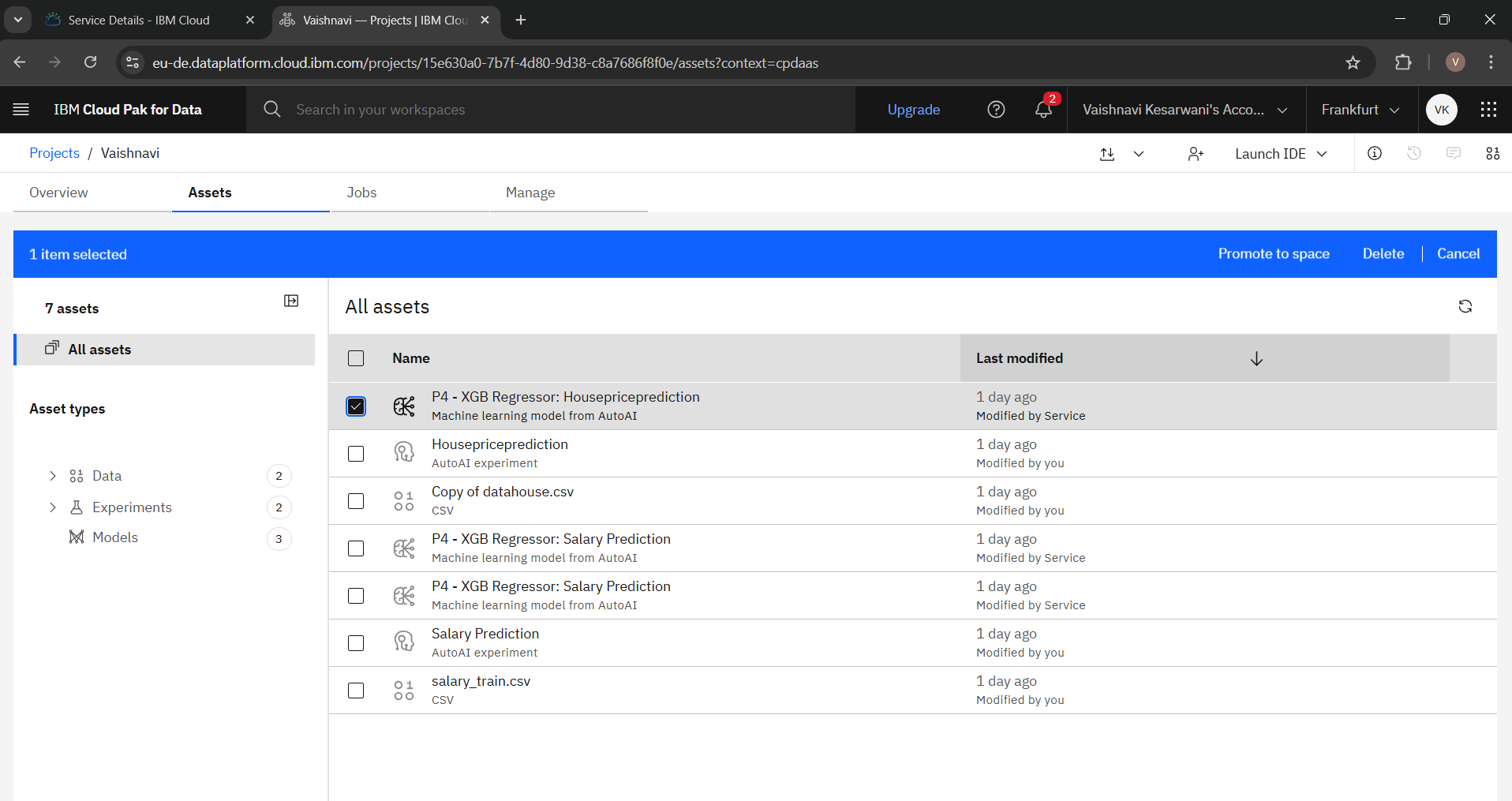


Step 12: Now you will save the pipeline 4 (best) as model and click on create. Pop-up will appear as saved model successfully.

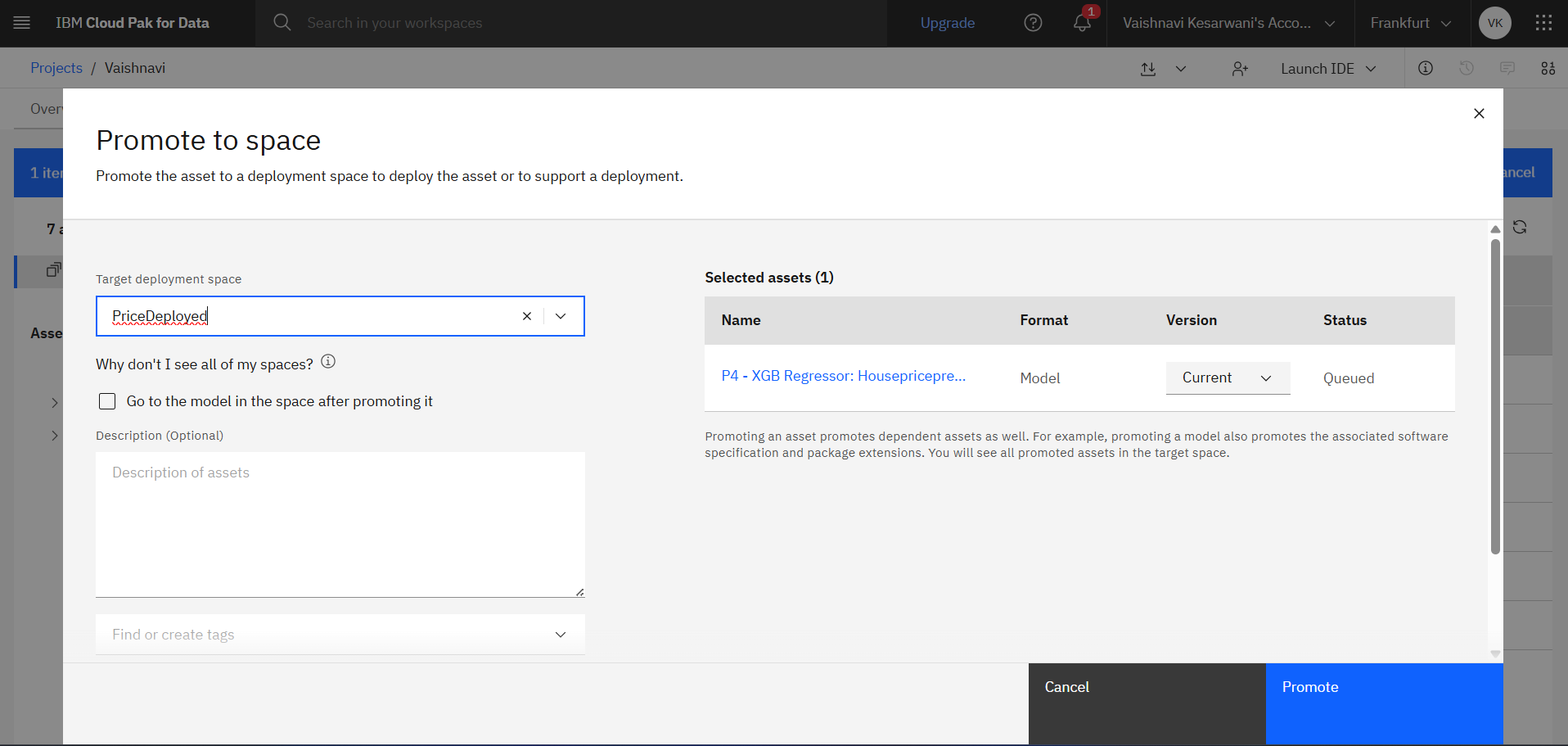


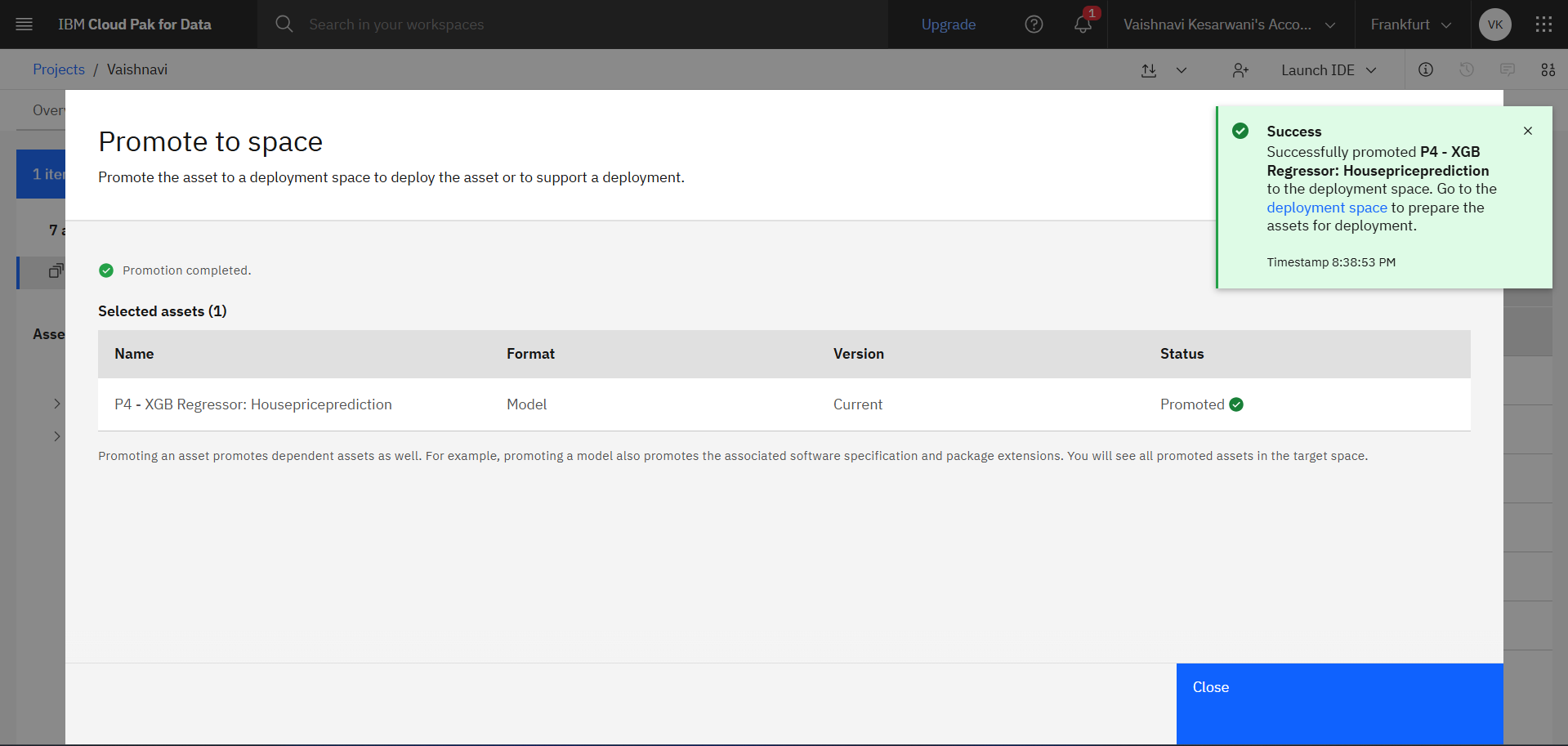


Step 13: Here you can check for model that you created and then click on promote to space.

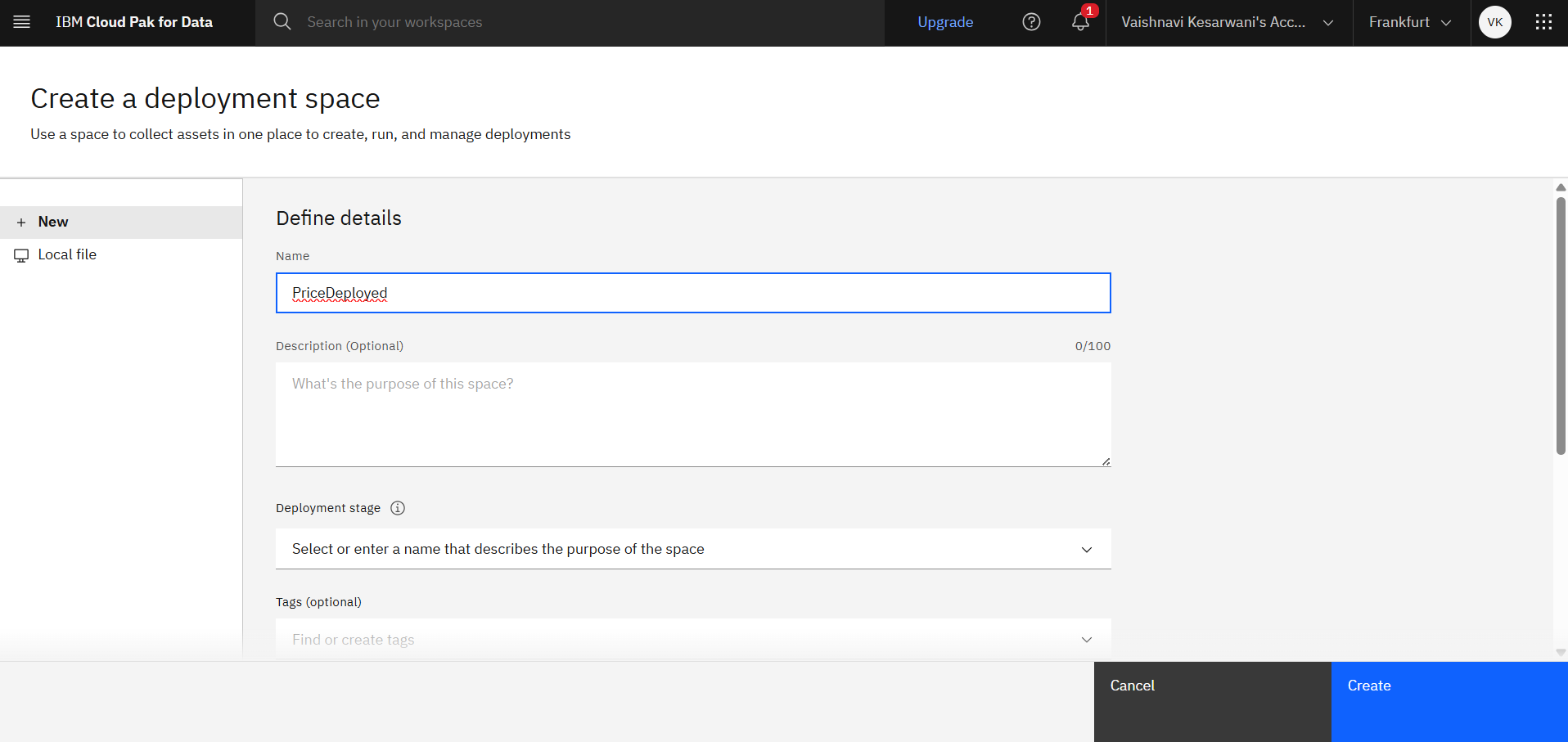


Step 14: The following window will appear where you can type name for target deployment space and then click on promote.

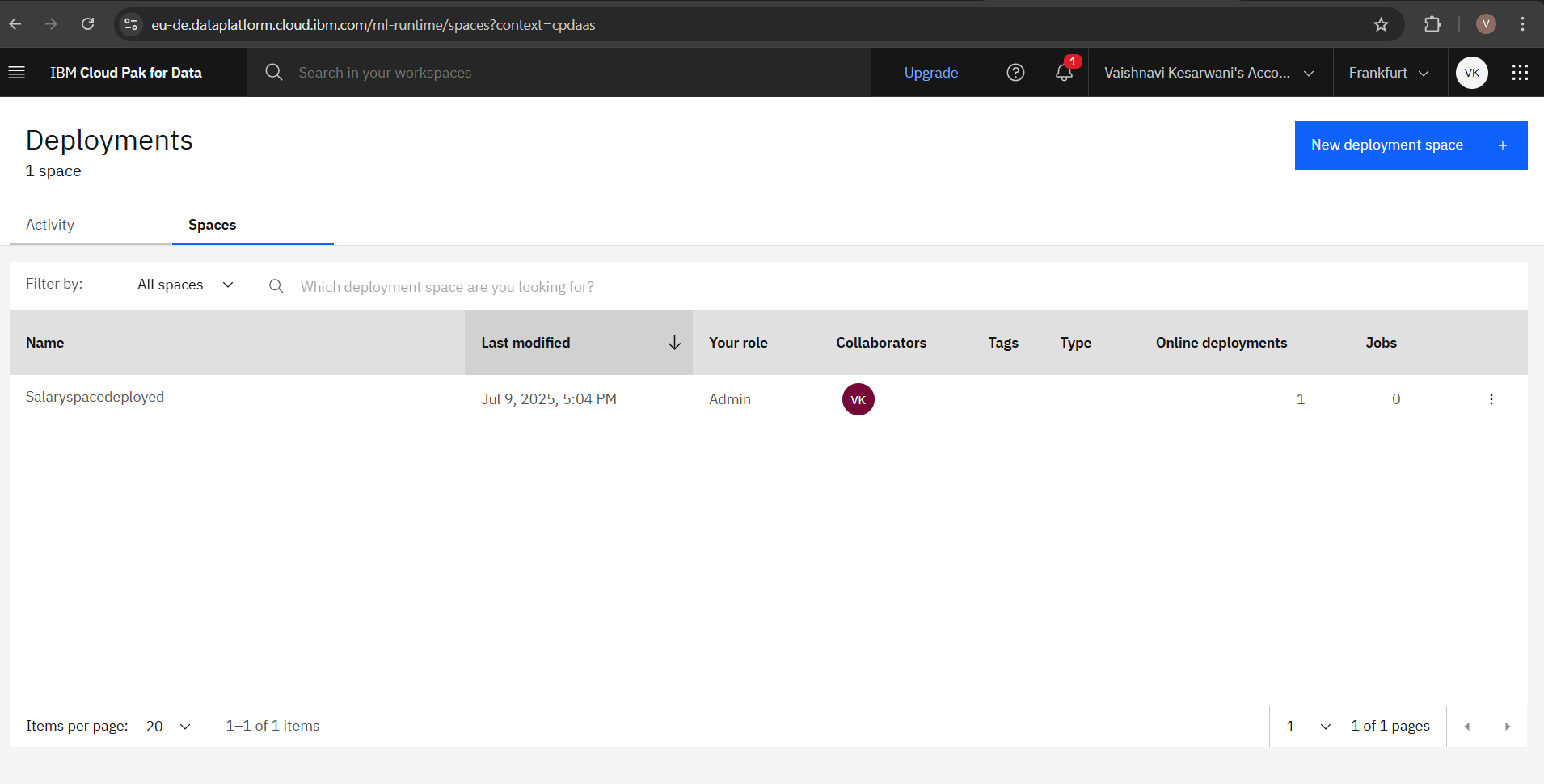


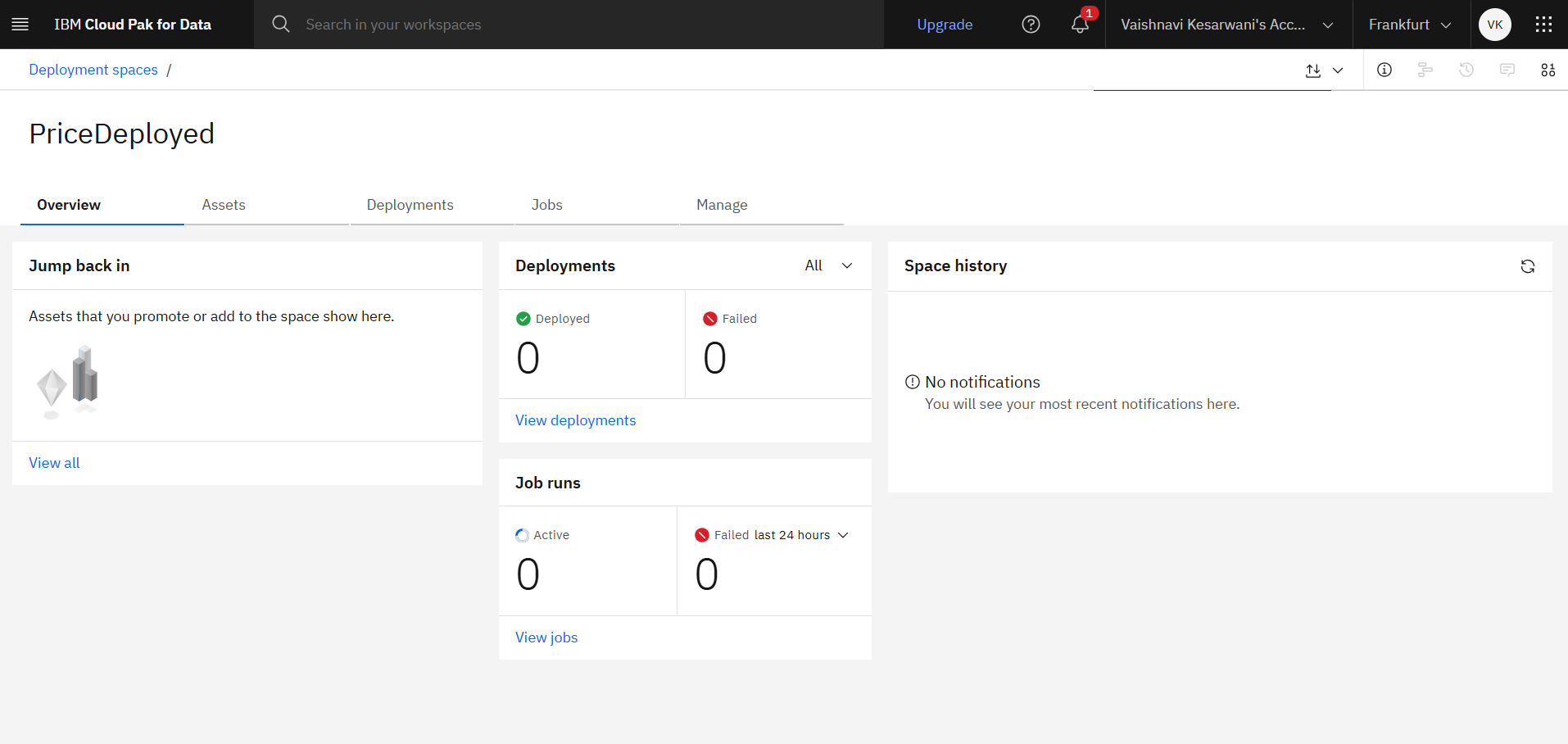


Step 15: Create a deployment space.

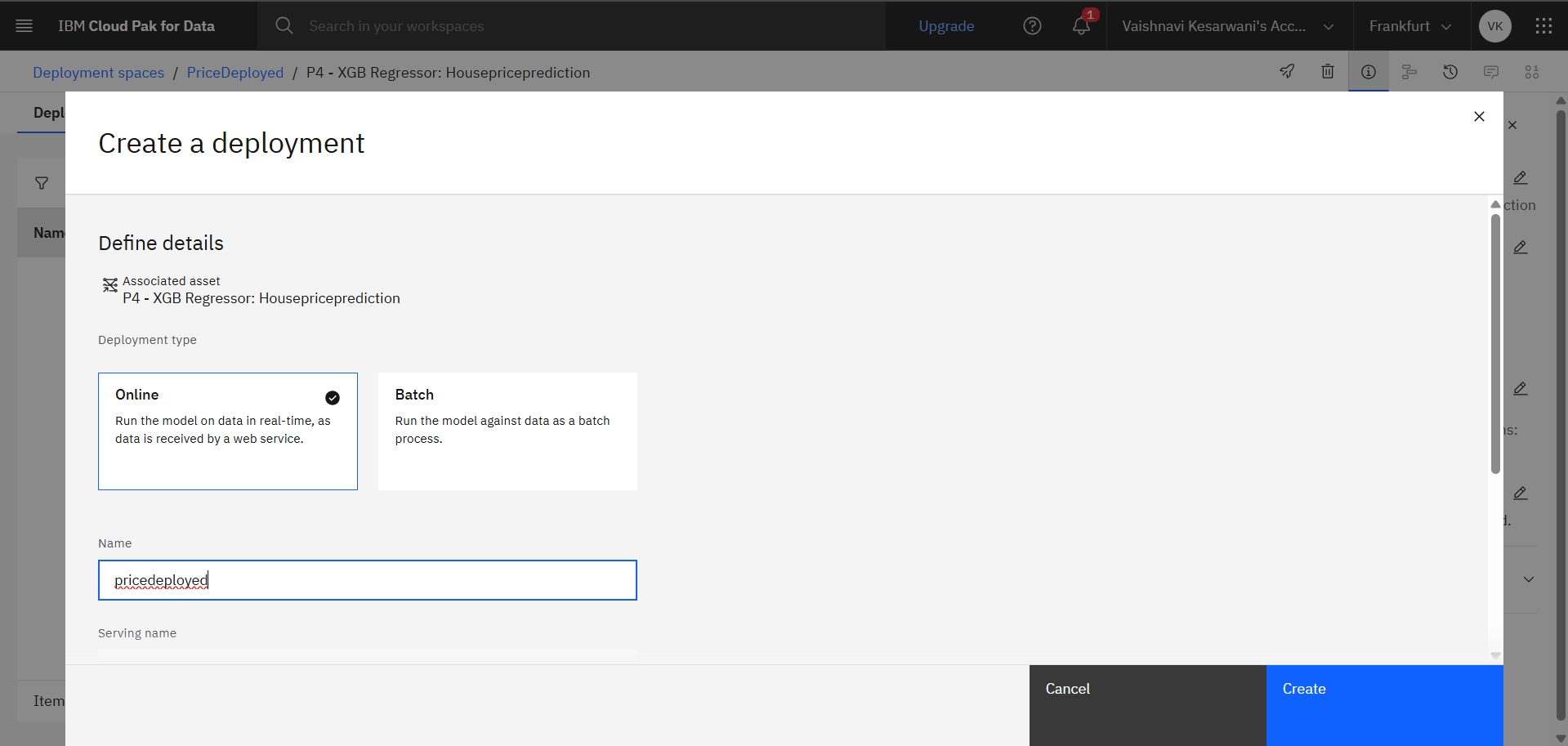


Step 16: You can see the space is allotted to your project and then click on Salaryspacedeployed.

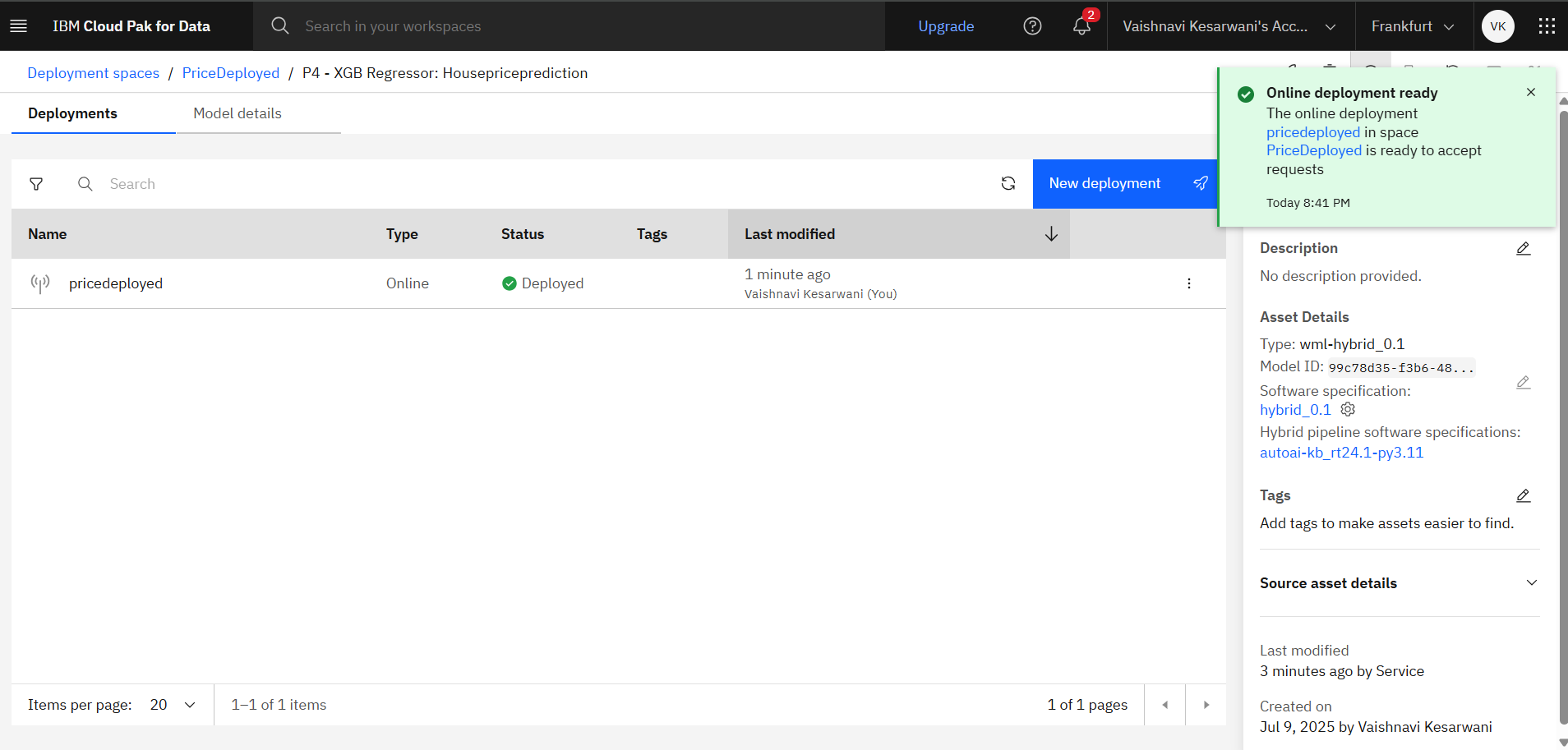




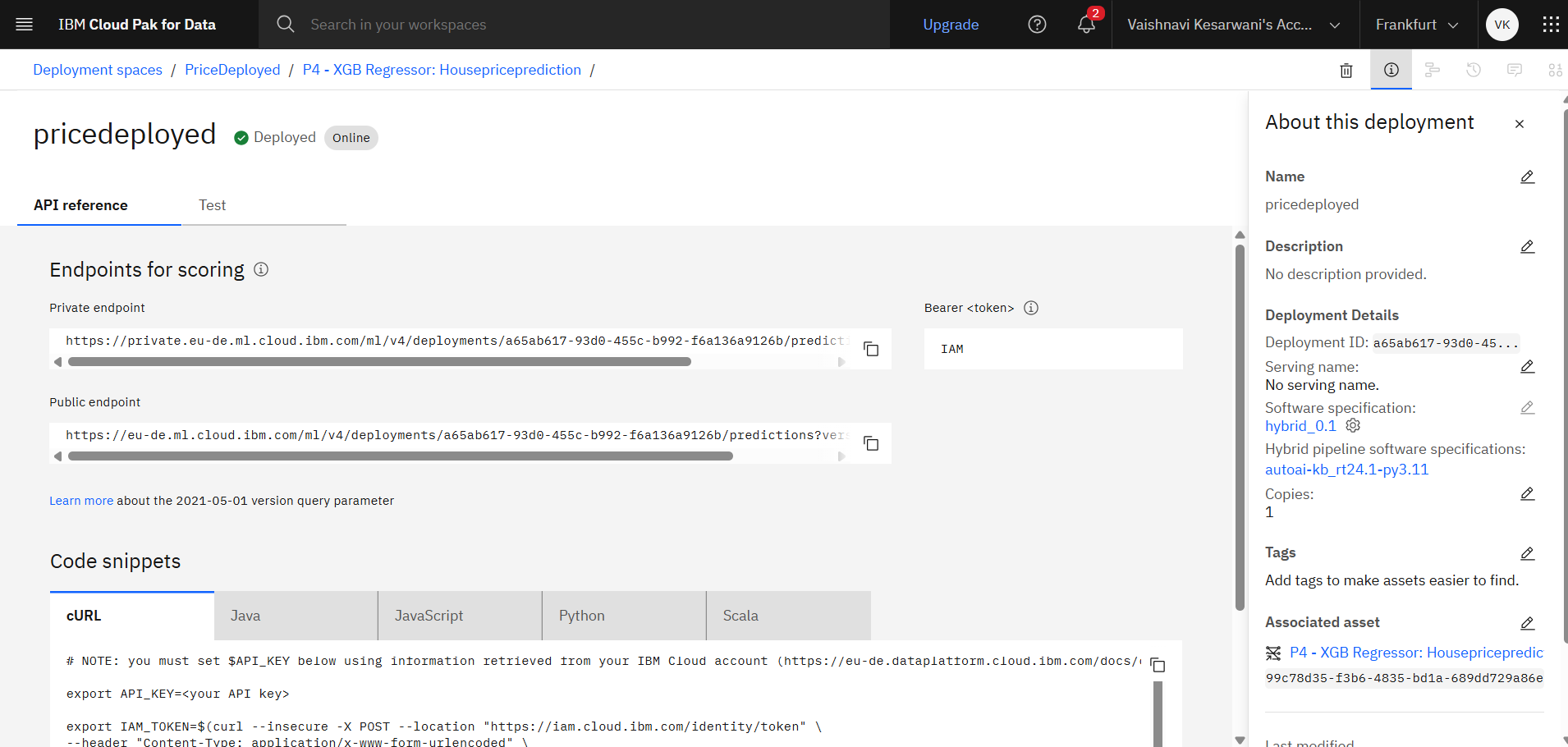
Step 17: After managing it, go to new deployment, then create a deployment dialog box will appear . Select online ,type name as pricedeployed and then click on create.



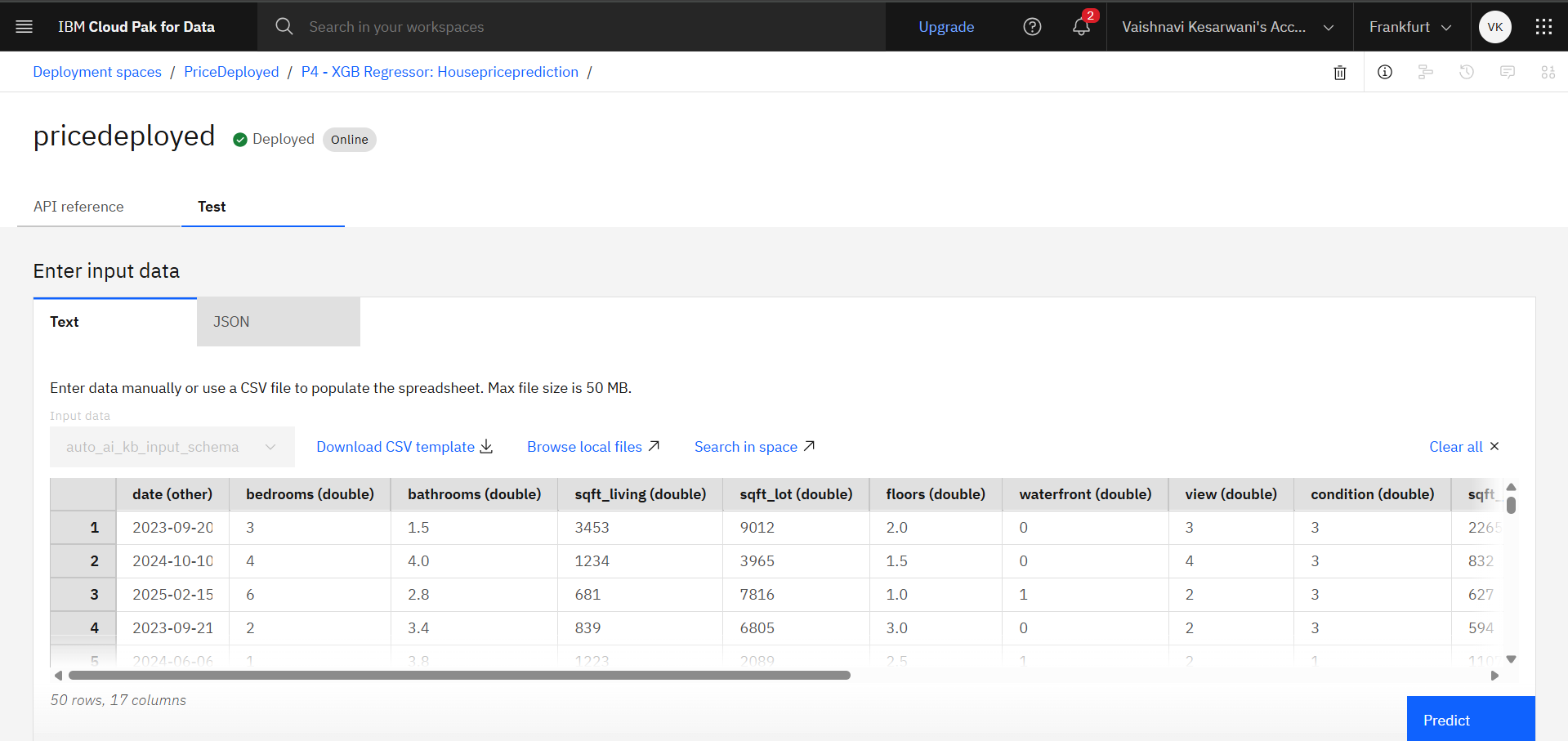
Pop-up box will appear as success.



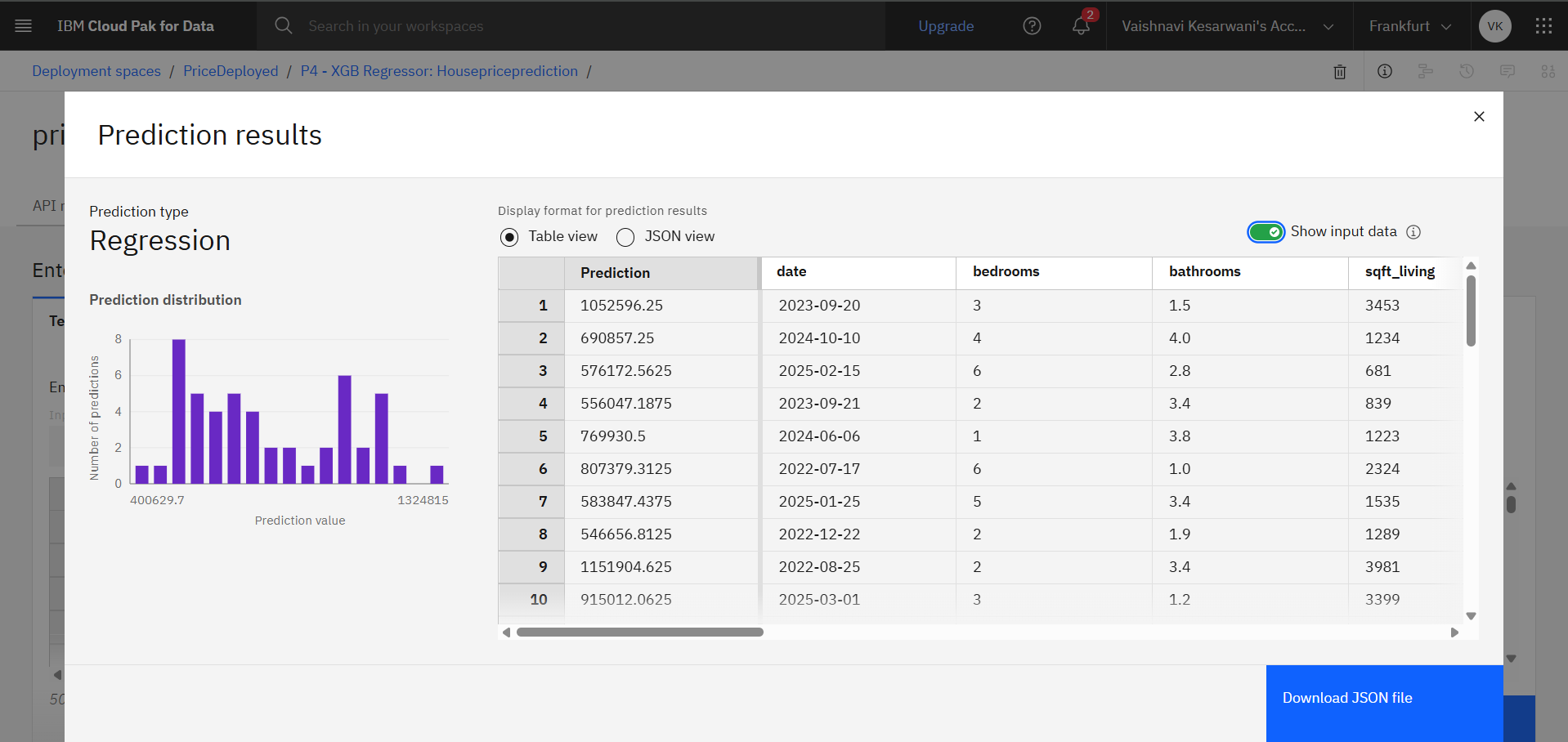
Step 18: The XGBoost model was deployed successfully using IBM Watson Studio's deployment space.  
Public and private endpoints were generated for scoring house price predictions.



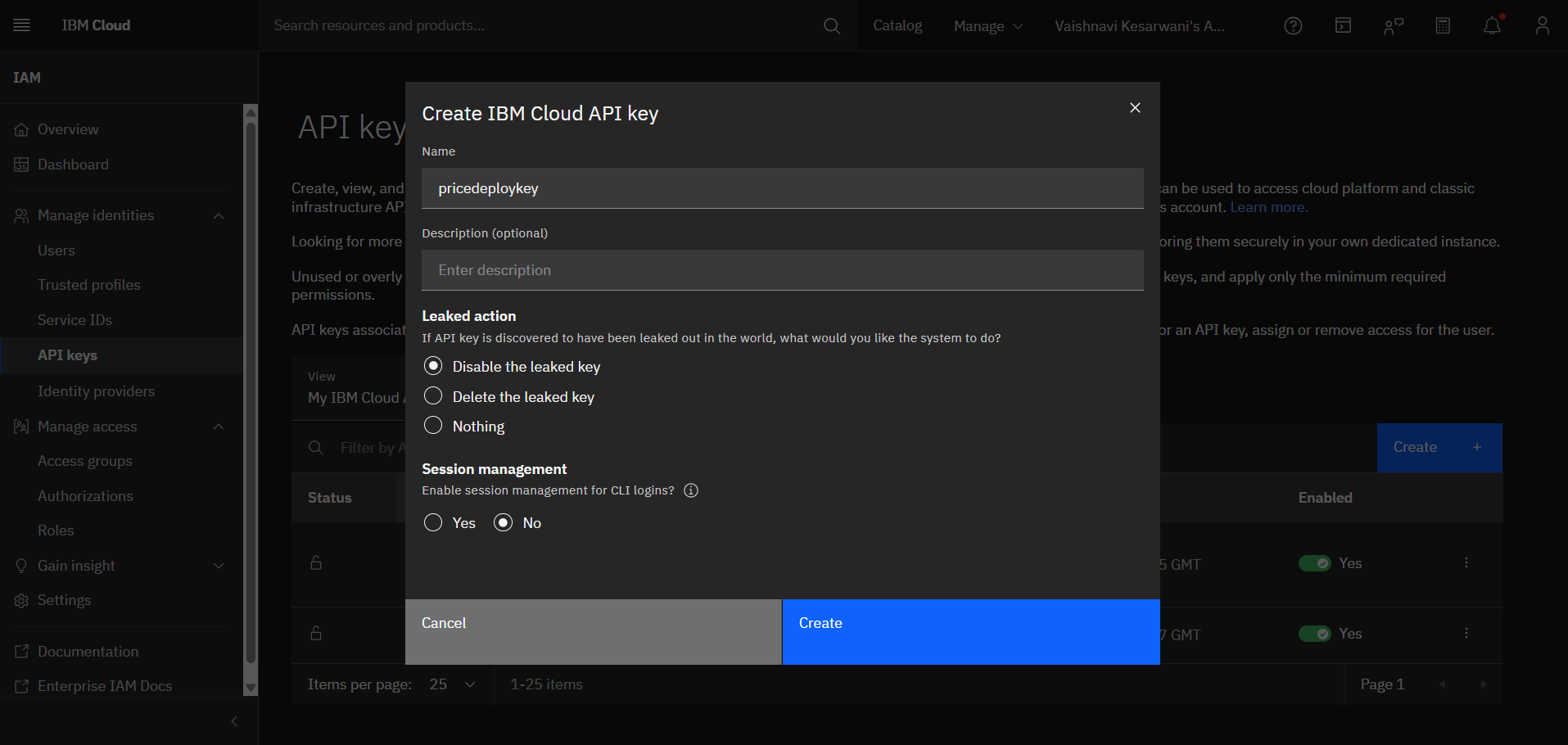
Step 19: Now you can test your model by uploading test data file and then click on predict.

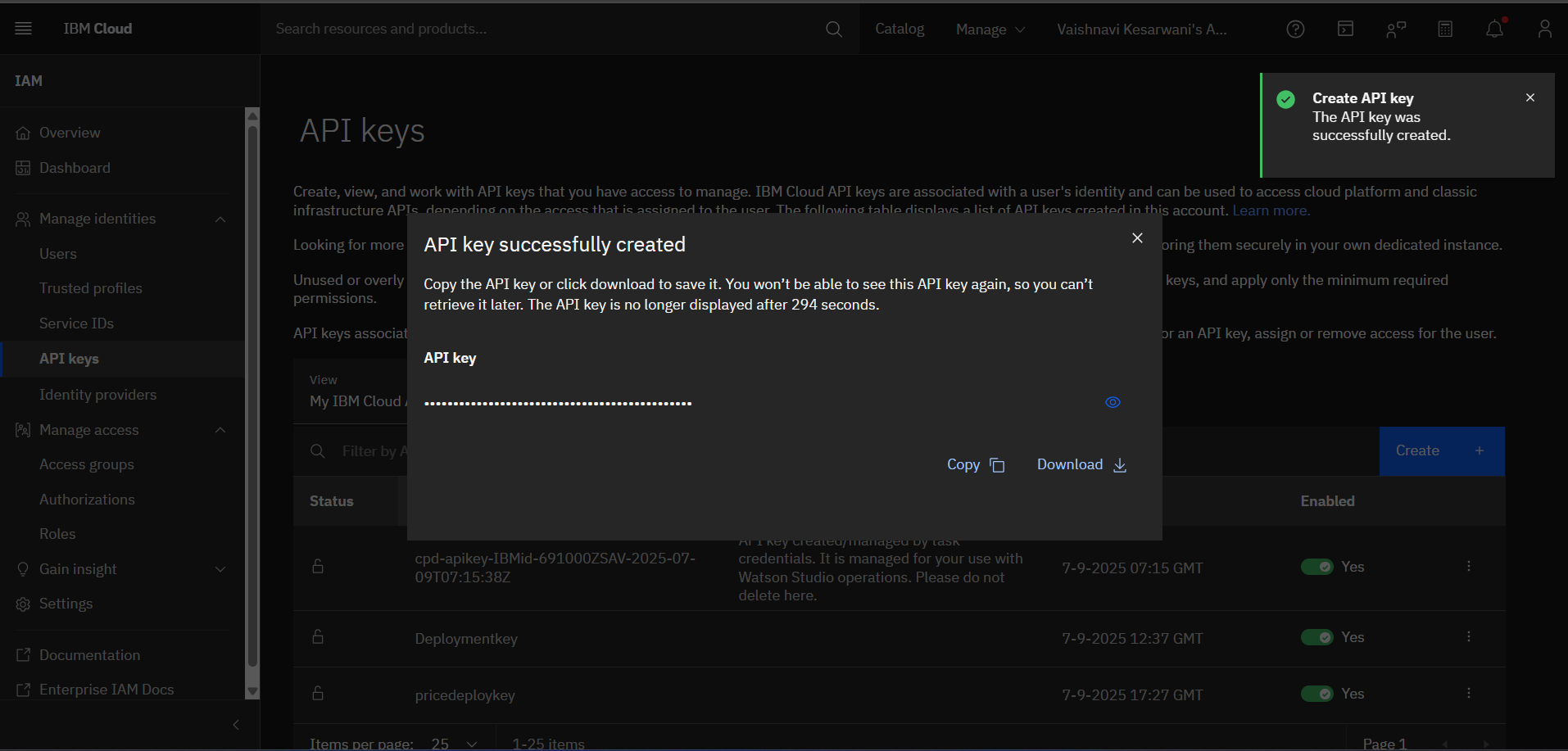


Step 20: The following window shows the prediction results for test dataset.

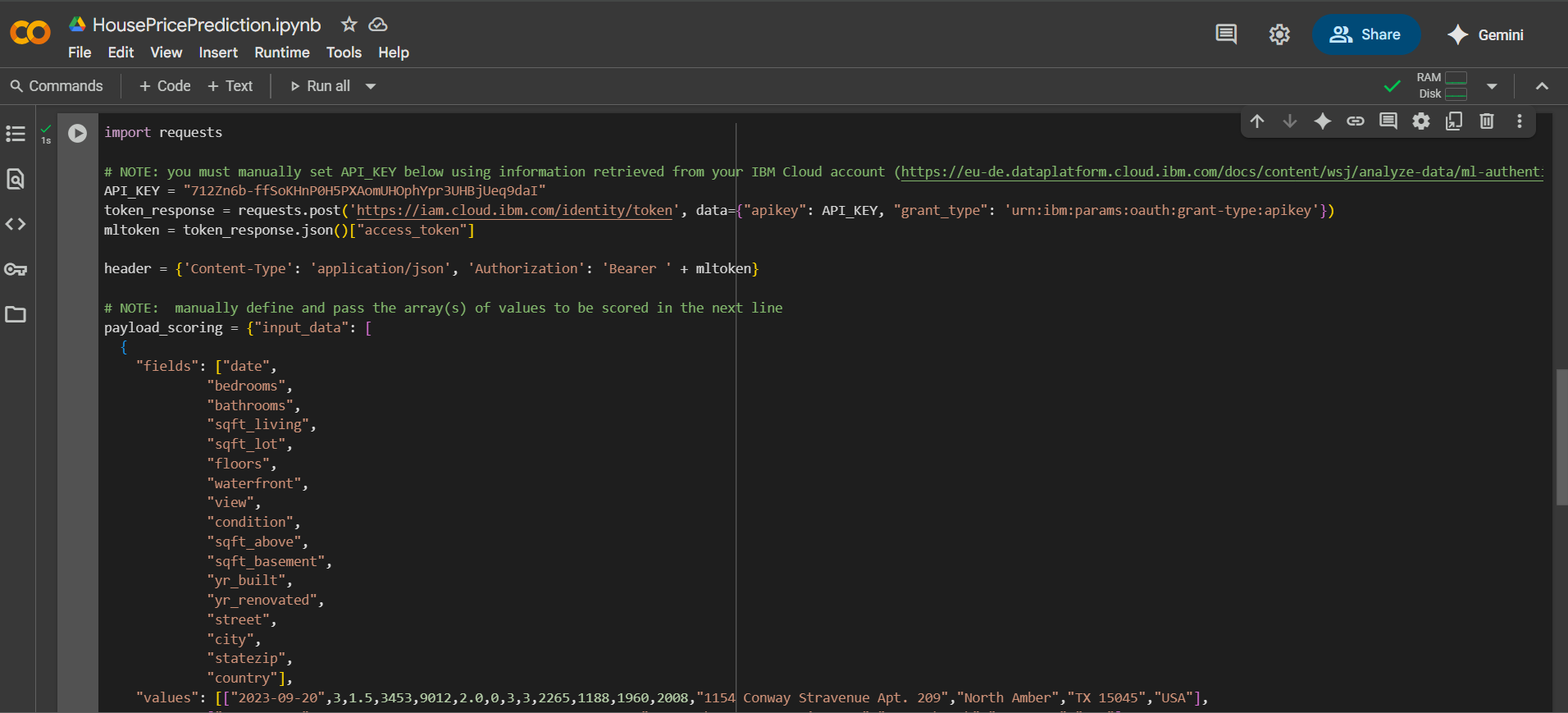


Step 21: Now create IBM Cloud API Key .



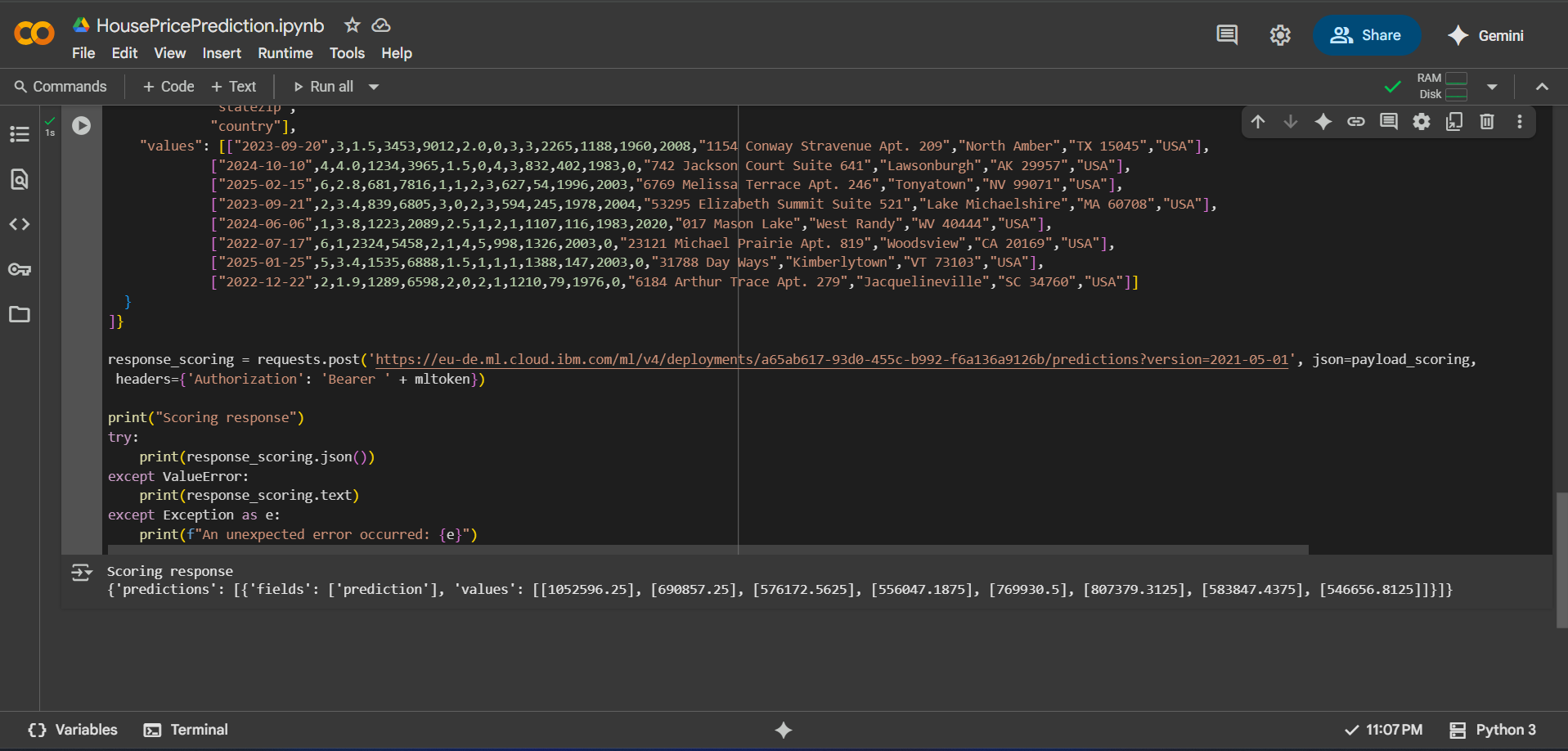


Step 22: Then copy the API Key as this will be attached in the colab file in the API Key attribute. It helps in setting the connection between the colab and IBM Watson studio.



Step 23: Copy the public endpoint link and paste it in response\_scoring attribute . Update fields and values and then run this cell.

Finally, you can see the output as scoring response.



Algorithm/Model Used

IBM Watson Studio Auto AI automatically trained multiple regression models such as:

* Linear Regression
* Random Forest Regressor
* XGBoost Regressor

Auto AI evaluated these pipelines and selected **XGBoost Regressor** as the best-performing model based on **Root Mean Squared Error (RMSE)** and **R² score**.

Result

The model achieved a high prediction accuracy with an R² score of 0.89 on the test data. The performance metrics demonstrate the reliability of Auto AI in automating the ML lifecycle effectively.

Conclusion

This project successfully demonstrates how IBM Watson Studio's AutoAI can be leveraged to build an accurate house price prediction model without extensive coding. It simplifies model development while ensuring robust performance and reliability.

Reference

- IBM Watson Studio Documentation: https://dataplatform.cloud.ibm.com  
- Kaggle Dataset for House Pricing