TERRATREND:HOUSE PRICE PREDICTION

Group No. 1

Siddhant Annadate (250243025004)

Sakshee Kandarkar (250243025028)

Santosh Kriplani (250243025033)

Siddhesh Suryavanshi (250243025043)

Guided By:

Milind Kapse sir

TerraTrend – Navigating the Real Estate

The Problem we solve : Navigating the Real Estate “Black Box”

For the average Indian, buying or selling property is the most significant financial decision of their lifetime. Yet the real estate market often operates like a “Black Box”. Pricing can seem arbitrary, information is scattered, and a persistent lack of transparency creates a stressful and unequal environment.

This leads to critical problems for everyone involved:

- **For Homebuyers:** The constant fear of overpaying.
- **For Sellers:** The challenge of setting a competitive price.
- **For the Market:** This lack of clear, data-driven valuation contributes to market inefficiency, slows down transactions, and erodes trust

Brief Introduction of TerraTrend – House Price Prediction

The Impact We Make: Empowering Decisions with Data

TerraTrend is designed to dismantle this black box. By leveraging the power of machine learning, we transform complex market data into a clear, actionable, and reliable price prediction. Our goal is not just to build a tool, but to create a more transparent and equitable real estate ecosystem.

- **Empowering Homebuyers:** We provide homebuyers with a data-driven estimate of a property's fair market value. This gives them the confidence to make informed offers
- **Enabling Sellers:** Our tool helps sellers price their properties strategically from day one. By providing an unbiased valuation based on thousands of similar properties
- **Creating a More Efficient Market:** By bringing transparency to the forefront, TerraTrend helps streamline the entire transaction process.

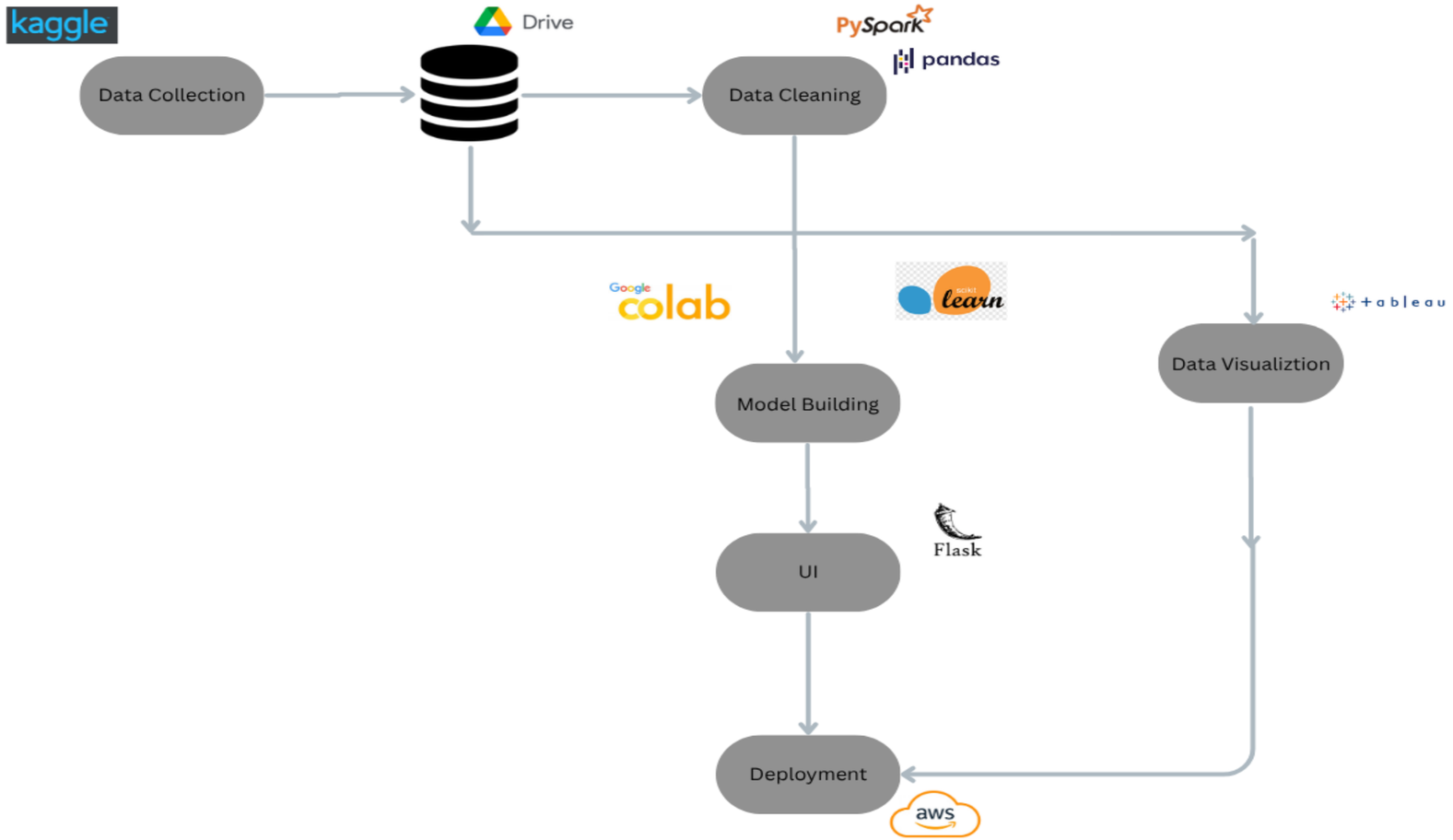
METHODOLOGY

Data collection and preparation

Exploratory Data Analysis (EDA)

Data Modeling

Deployment



FLOW DIAGRAM

Data Preparation and Preprocessing

Sourcing the dataset –

Our dataset were taken from publicly available and reliable sources which were further sourced through various listings across India. All the details were brought together across several locations and fetched through publicly available datasets.

Key Features and attributes -

The dataset includes critical features such as location, size, number of bedrooms, and bathrooms. It allows for robust and predictive Modelling.

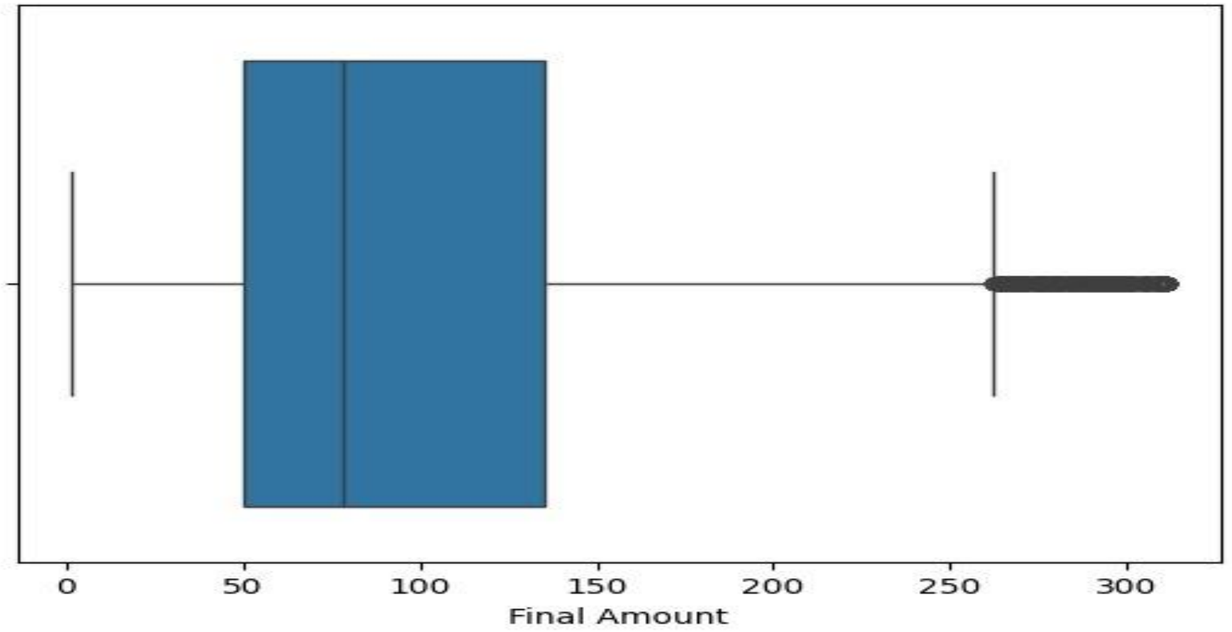
Data Import -

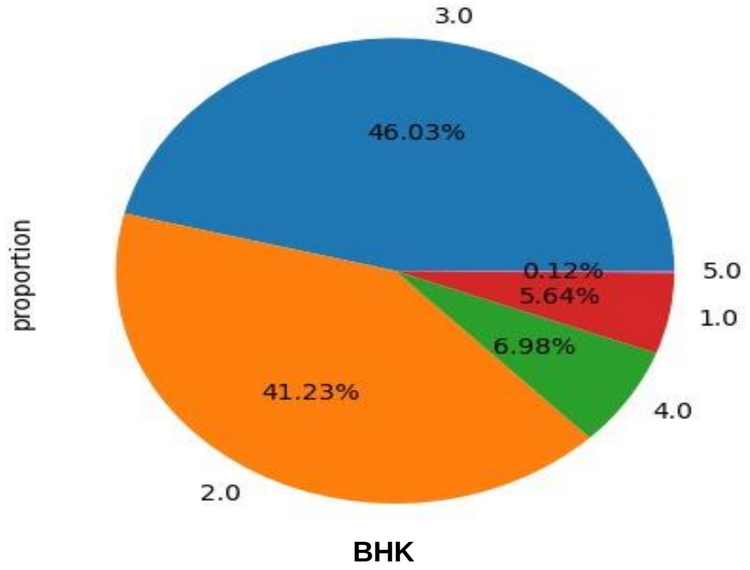
Data is imported from the datasets which were in comma separated values into a data frame for easier pre processing and cleaning to ensure a strong foundation for the machine learning model.

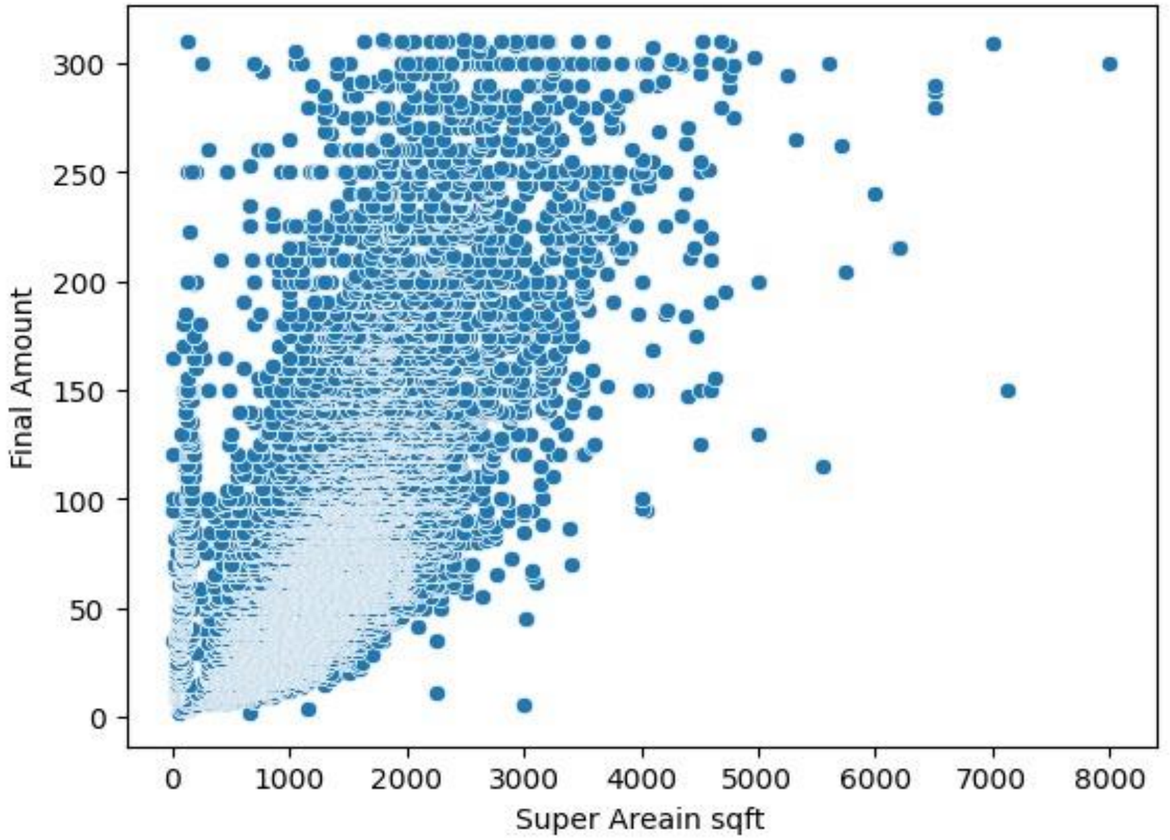
Data cleaning –

Data is duly cleaned by working on various aspects. Null values, Blanks are handled depending upon the significance of the said rows and columns, they were either imputed, or eliminated. Duplicates are removed, and data was properly put together. The columns not contributing to the prediction were also removed ensuring a strong foundation for ML models

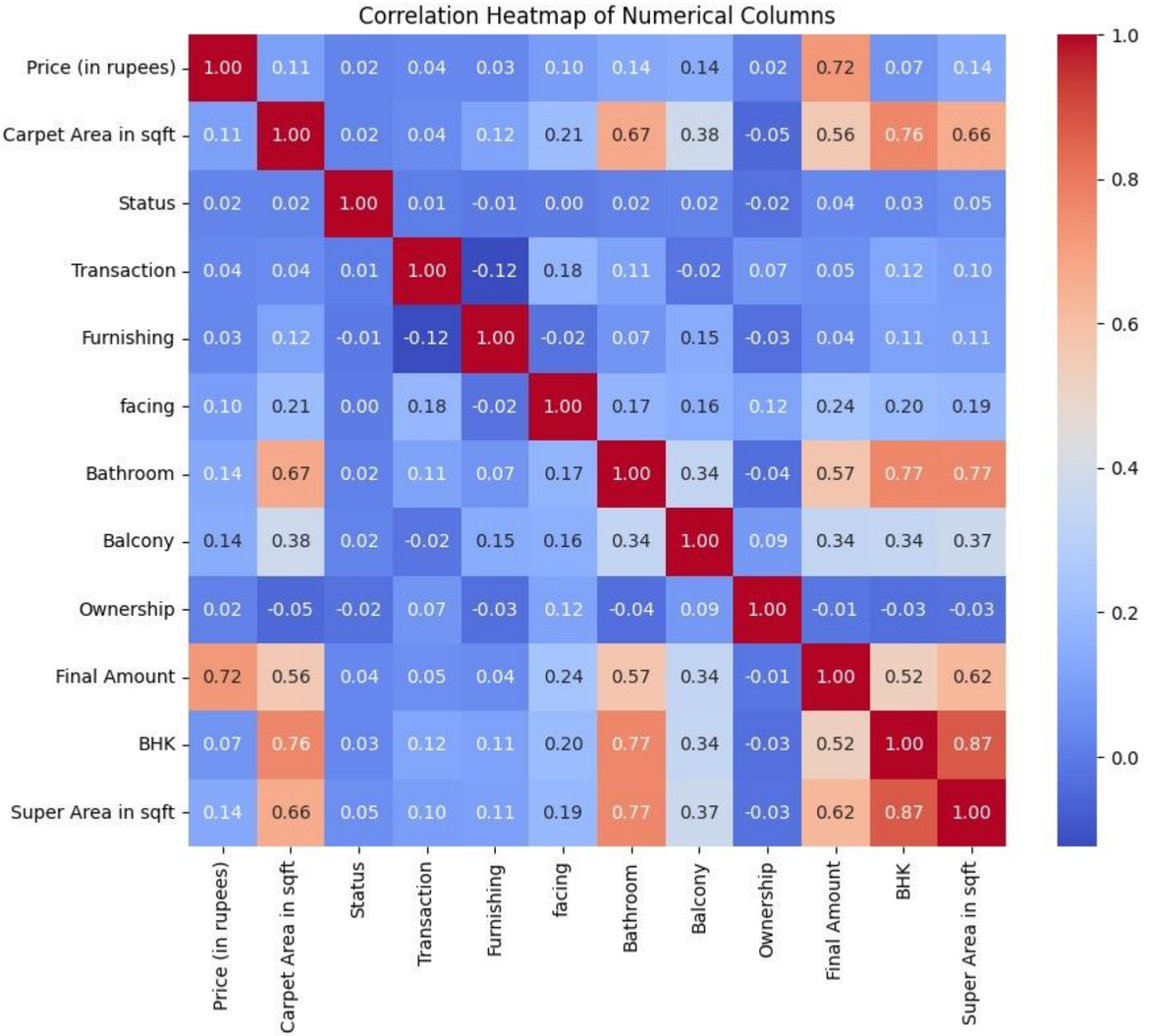
Exploratory Data Analysis

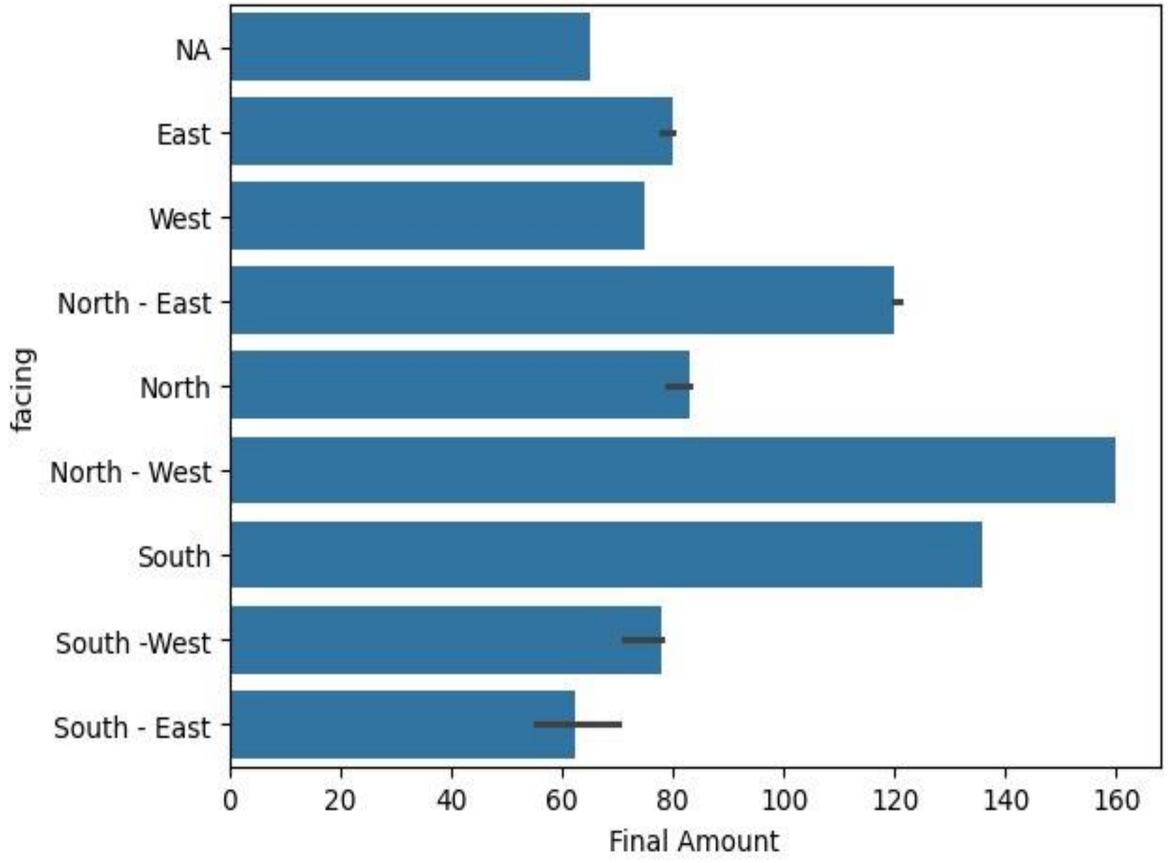






Exploratory Data Analysis





Data Modeling

Linear Regression

Set as a benchmark.

.

Random Forest

Used for its robustness and ability to reduce overfitting through ensemble learning.

XGBoost

Employed for its high performance and speed in handling large datasets with boosting techniques.

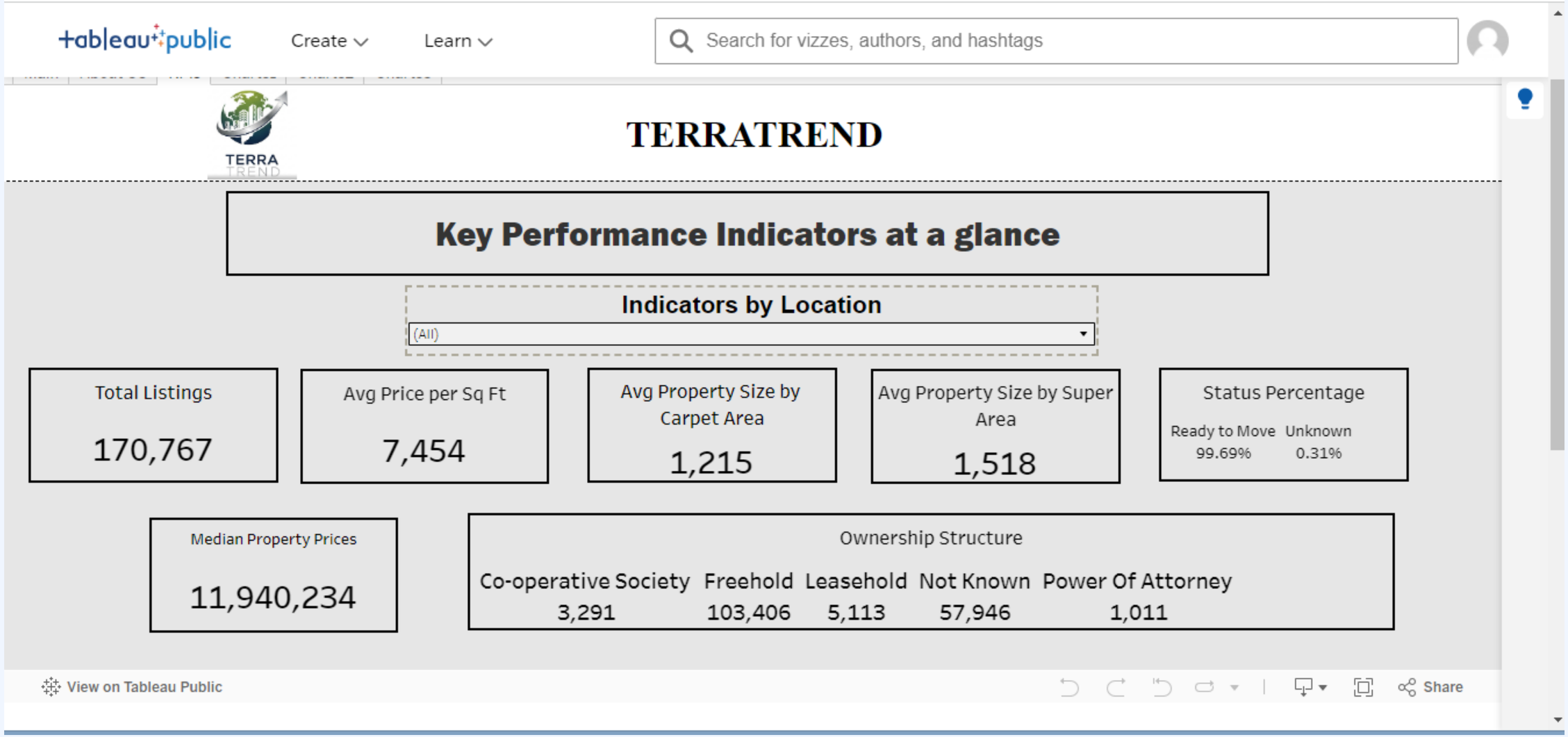
CatBoost

Leveraged for its ability to handle categorical features without extensive preprocessing.

Model Performance & Hyperparameter Tuning

Model	R^2 Score	MAE	RMSE
Linear Regression	0.8699	14.85	21.29
Random Forest	0.9914	1.531	5.46
XGBoost	0.9922	2.172	5.66
CatBoost	0.989	2.668	6.13

TABLEAU DASHBOARD

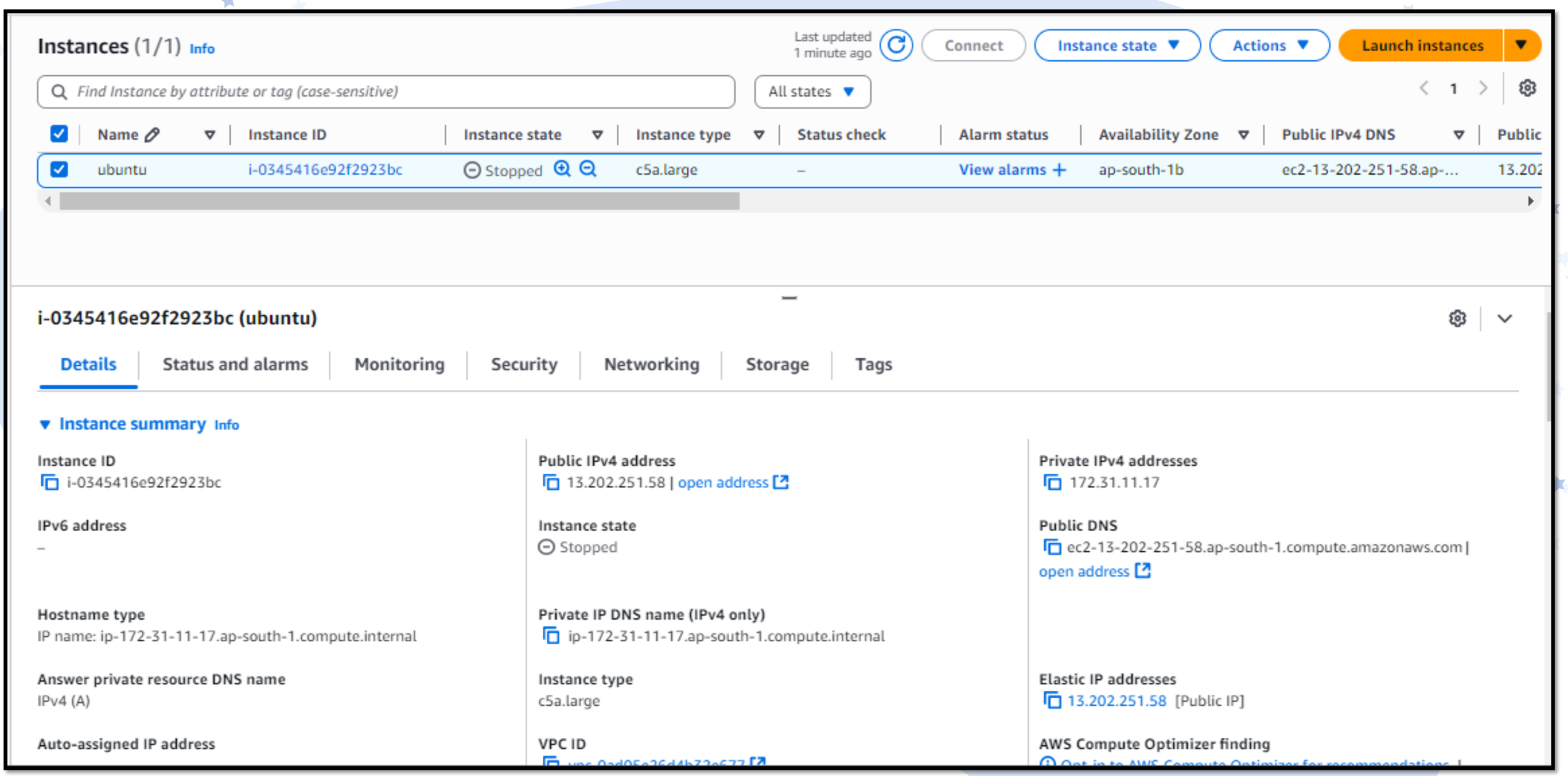


https://public.tableau.com/app/profile/santosh.kriplani/viz/Project2_17544349064900/Main

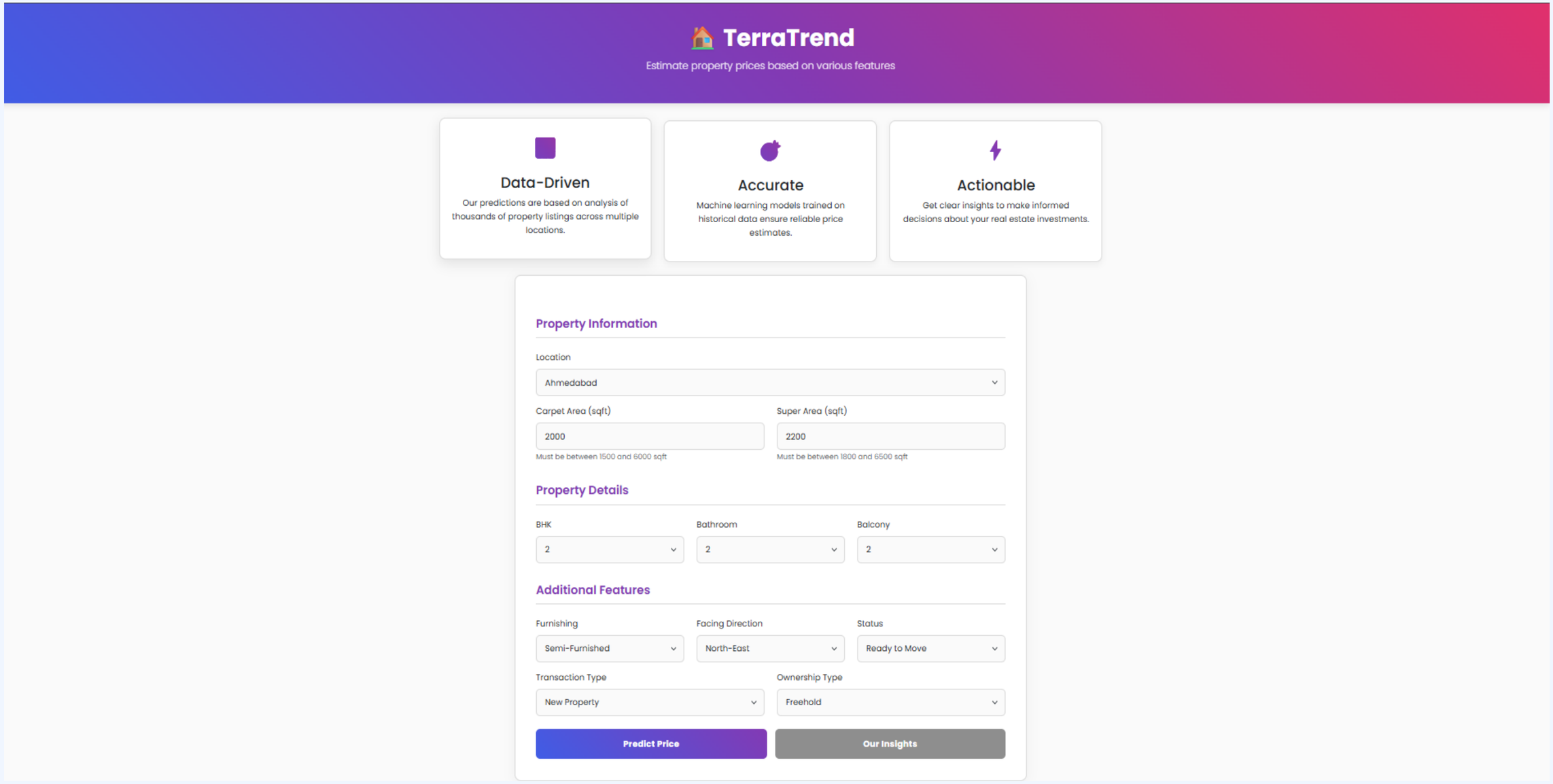
Deployment

1.Launch EC2 Instance
2.Configure Security Groups
3.Connect to Instance and Install Python & pip
4.Instal, Create & Activate Virtual Environment Tool
6. Install Project Dependencies and Upload Project files
7. Run Flask Application
8. Access Application via Public IP

Deployment


Instances (1/1) Info
Last updated 1 minute ago
Connect
Instance state
Actions
Launch instances
Find Instance by attribute or tag (case-sensitive)
All states
1
Name
Instance ID
Instance state
Instance type
Status check
Alarm status
Availability Zone
Public IPv4 DNS
Public
ubuntu
i-0345416e92f2923bc
Stopped
c5a.large
–
View alarms +
ap-south-1b
ec2-13-202-251-58.ap-...
13.202
i-0345416e92f2923bc (ubuntu)
Details
Status and alarms
Monitoring
Security
Networking
Storage
Tags
Instance summary Info
Instance ID
i-0345416e92f2923bc
IPv6 address
–
Hostname type
IP name: ip-172-31-11-17.ap-south-1.compute.internal
Answer private resource DNS name
IPv4 (A)
Auto-assigned IP address
Public IPv4 address
13.202.251.58 | open address
Instance state
Stopped
Private IP DNS name (IPv4 only)
ip-172-31-11-17.ap-south-1.compute.internal
Instance type
c5a.large
VPC ID
Private IPv4 addresses
172.31.11.17
Public DNS
ec2-13-202-251-58.ap-south-1.compute.amazonaws.com |
open address
Elastic IP addresses
13.202.251.58 [Public IP]
AWS Compute Optimizer finding

PREDICTION PAGE



Future Scope

- **Advanced Algorithms** : Deep Learning (Example Neural Networks) for model improvement and optimization
- **Integration with real time data** : incorporate real time market trends.
- **Explainable AI** : Integrate tools like SHAP or LIME to explain model predictions.
- **Natural Language description Generation using LLMs** : Fine tune LLMs like T5, LLaMA, on your dataset to automatically generate detailed and well written property descriptions from structured data.

Thank you