|  |  |
| --- | --- |
| Internship Project Title | Intelligent Property Analyser |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Debashis Roy |
| Name of the Institute | MIT ADT UNIVERSITY |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Start Date | End Date | | Total Effort (hrs.) | | Project Environment | Tools used |
| 05/05/2023 | 09/05/2023 | | 25.5 | | Windows 11 | Python,Jupyter notebook,Pycharm,  Flask |
| Milestone # | 1 | Milestone: | | Predict price of housing property based on multiple factors | | | |

**TABLE OF CONTENTS**

[**1. Acknowledgements** 2](#_Toc127215298)

[**2. Objectives** 2](#_Toc127215299)

[**3. Introduction** 2](#_Toc127215300)

[**4. Internship Activities** 2](#_Toc127215301)

[**5. Methods and Algorithms** 3](#_Toc127215302)

[**6. Assumptions** 3](#_Toc127215303)

[**7. Project Analysis** 3](#_Toc127215304)

[**Model Evaluation** 6](#_Toc127215305)

[**Model Prediction** 6](#_Toc127215306)

[**10. Conclusion** 7](#_Toc127215307)

[**11. Enhancement Scope** 7](#_Toc127215308)

[**12. Link to Code** 7](#_Toc127215309)

[**13. References** 7](#_Toc127215310)

# **1. Acknowledgements**

1. I am extremely grateful to TCS iON for providing me with this amazing internship opportunity. I am deeply thankful to my industry mentor, Mr. Debashis Roy, for his invaluable guidance and support throughout the project. I could not have completed this project without his time and effort. Thank you from the bottom of my heart!

# **2. Objectives**

We are building a website that can tell you how much a house is worth based on different factors, like its size, location, and age. To do this, we are using a computer program that has been taught to understand the relationships between these factors and the prices of houses. The website will have a user-friendly interface that lets you enter the details of a property and get an estimated value. Our goal is to create a reliable tool that can help people make informed decisions when buying or selling a house.

# **3. Introduction**

Demand for houses grew rapidly over the years as people's living standards improved. House price prediction can be done using multiple prediction models (Machine Learning Model) such as linear regression, Decision tree, etc. There are many benefits that home buyers, property investors, and housebuilders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors, and housebuilders, such as the valuation of house prices in the present market, which will help them determine house prices. The target feature in this proposed model is the price of the housing property and the independent features are: are type, availability, location, size and balcony. The whole implementation is done using the python programming language, flask framework and MySQL database.

# **4. Internship Activities**

Internship activities are divided in three stages which are as follows:

**Stage 1: Develop ML model using python**

1. Data Collection
2. Data Preprocessing
3. Model Selection
4. Training the model
5. Evaluating model
6. Parameter Tuning
7. Making Predictions

**Stage 2: Setting up MySQL database**

1. Setup MySQL database
2. Database would be used by users/admins to populate records and predicted prices of properties

**Stage 3: Develop Django based web project**

1. Develop Web-UI
2. Users can view properties and get predicted prices
3. Designing report that displays predicted property prices

# **5. Methods and Algorithms**

Steps Involved:

1. Import the necessary Python packages into our environment.
2. Import house price data and perform exploratory data analysis (EDA) on it.
3. Visualize statistics on house prices using visualization techniques.
4. Split the data into training and testing sets and select the most relevant features.
5. Use various machine learning algorithms to model the data, such as regression and decision tree models.
6. Evaluate the performance of each algorithm using appropriate assessment metrics.
7. Use the best machine learning model to predict house prices based on new data inputs.
8. Develop a web application that integrates our machine learning model into a user-friendly interface that allows users to input their data and receive accurate predictions.

We have used these models for regression: Linear Regression,Ridge Regression.

Linear Regression:

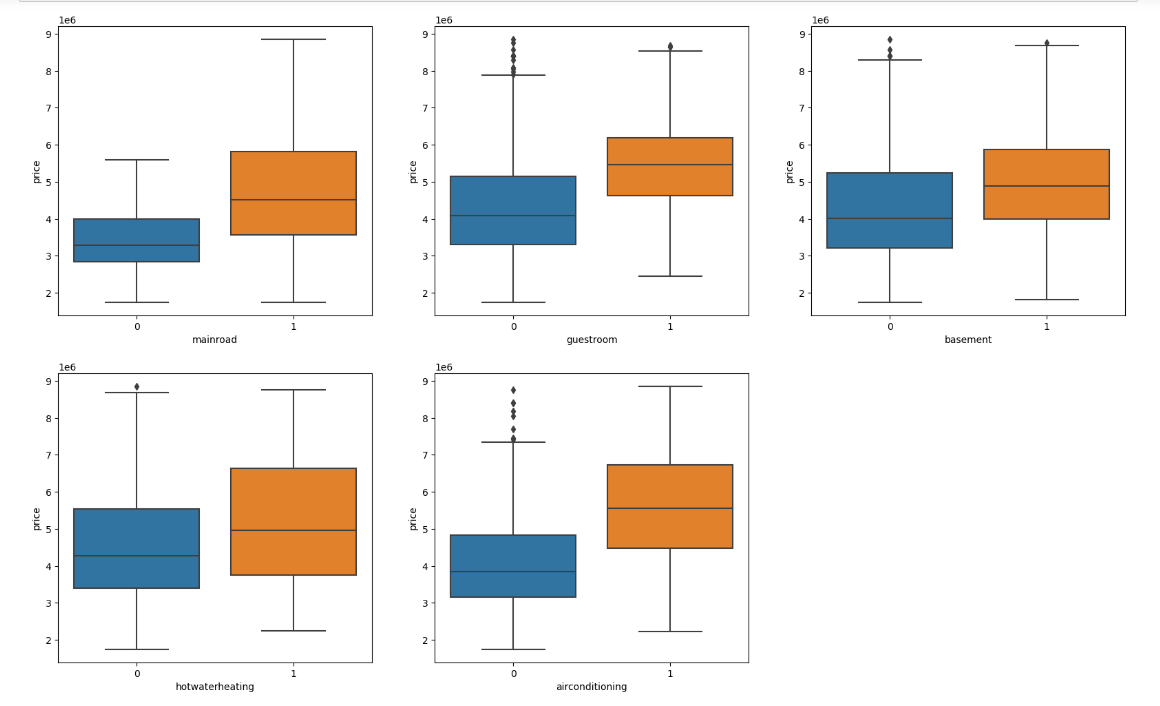
Linear regression is a statistical technique for examining the relationship between two or more variables. It can be used to predict values based on past data and to explain variation in the response variable caused by variation in the explanatory variables. Linear regression models can be simple, with only one independent variable (simple linear regression), or complex, with multiple independent variables (complex linear regression) (multiple linear regression). In both cases, the model is a linear function of the independent variables, which means that for each unit of change in the independent variable, a fixed amount is added to the final result. When using multiple linear regression, the model is expressed as an equation with the independent variables on the right and the response variable on the left. The coefficients of the equation show how much a change in the independent variables is predicted to affect the response variable.

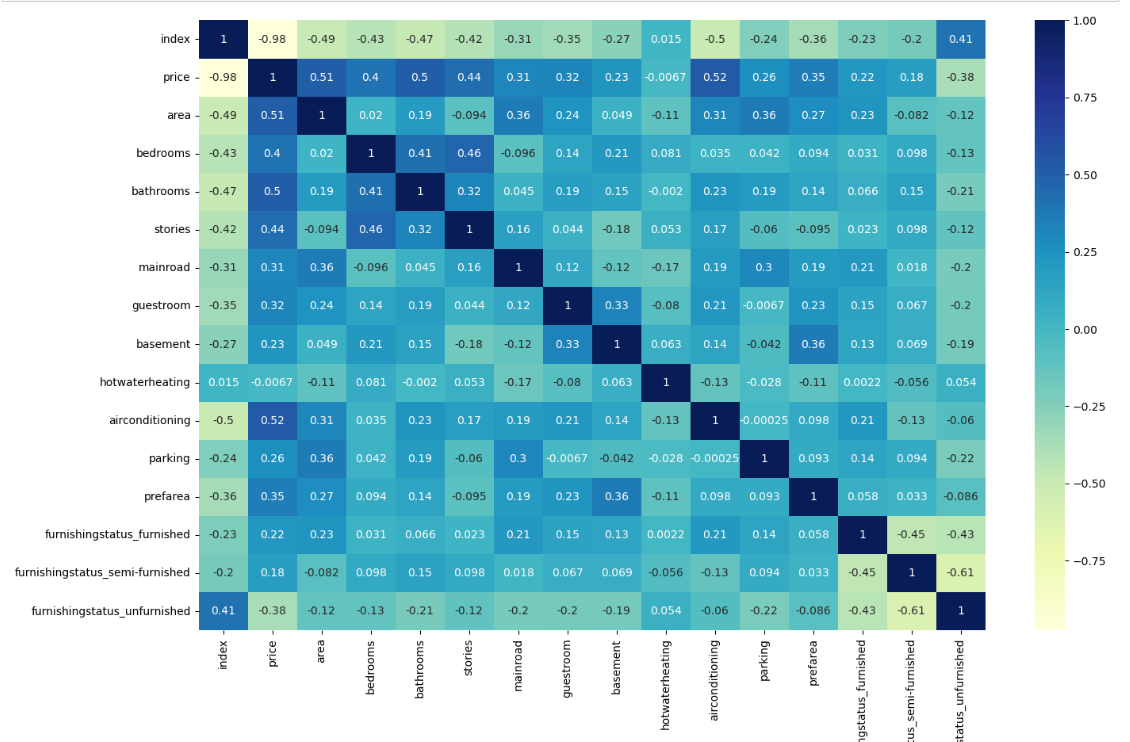
Ridge regression is a regularized linear regression model that employs an additional penalty term, known as the ridge or L2 penalty, to reduce the size of the coefficient estimates. The ridge penalty is equal to the sum of the squared coefficient estimates, and the goal is to minimize the sum of the squared residuals while keeping this constraint in mind. This means that the ridge model shrinks the coefficient estimates towards zero more slowly than the lasso model. This can help the model generalize to new data more effectively. Furthermore, because some of the coefficients may be set to zero during the optimization process, ridge regression can be used to perform feature selection.

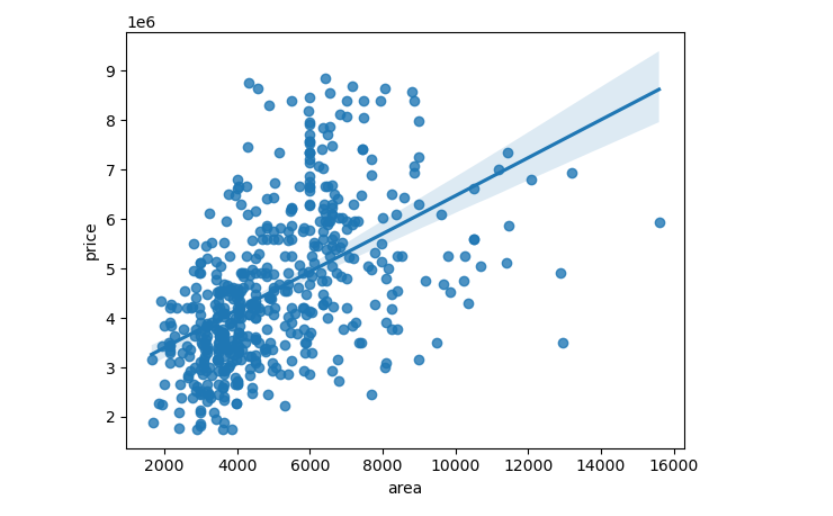
# **6. Assumptions**

Our goal is to create a predictive model that can estimate housing prices based on a given dataset. We will assume that the dataset is clean and reliable, and that the features within it are related to each other in meaningful ways. By leveraging the relationships between different features, our model will be able to accurately predict the prices of new properties with a high degree of accuracy. This will provide valuable insights to potential buyers and sellers, and help them make more informed decisions when it comes to real estate transactions.

# **7. Project Analysis**







## **Model Evaluation**

**Linear Regression**

|  |
| --- |
| Training Scores |
| R2 Score 9.472958e-01 |
| MAE 2.522345e+05 |
| MSE 1.221193e+11 |
| RMSE 3.494557e+05 |

**Ridge Regression**

|  |
| --- |
| Training Scores |
| R2 Score 9.448849e-01 |
| MAE 2.657884e+05 |
| MSE 1.277056e+11 |
| RMSE 3.573592e+05 |

## **Model Prediction**

Sample prediction of test dataset:

**Linear Regression**

|  |  |  |
| --- | --- | --- |
|  | **Actual price** | **Predicted Price** |
| 0 | 2100000 | 2.088947e+06 |
| 1 | 2870000 | 2.611365e+06 |
| 2 | 2380000 | 2.157318e+06 |
| 3 | 2730000 | 2.540647e+06 |
| 4 | 4060000 | 4.109290e+06 |
| 5 | 4410000 | 4.598295e+06 |
| 6 | 4382000 | 4.619372e+06 |
| 7 | 4900000 | 5.105404e+06 |
| 8 | 2940000 | 2.662274e+06 |
| 9 | 2870000 | 2.440030e+06 |

**Ridge Regression**

|  |  |  |
| --- | --- | --- |
|  | **Actual Price** | **Predicted Price** |
| 0 | 2100000 | 2094044.01 |
| 1 | 2870000 | 2611945.43 |
| 2 | 2380000 | 2162949.51 |
| 3 | 2730000 | 2545115.15 |
| 4 | 4060000 | 4115965.98 |
| 5 | 4410000 | 4593047.06 |
| 6 | 4382000 | 4620573.08 |
| 7 | 4900000 | 5100691.86 |
| 8 | 2940000 | 2663802.00 |
| 9 | 2870000 | 2445176.18 |

# 

# **10. Conclusion**

1. Area is the most important feature affecting price.
2. Ridge regression is the best fit model.

# **11. Enhancement Scope**

1. The size of the dataset provided for this project is relatively small, which means that we may need to acquire more data to improve the accuracy of our predictions. Having more data will help us to identify more patterns and relationships in the features, which in turn will allow us to build a more robust and accurate model.
2. There are several important features that are not included in the provided dataset, such as structure\_type, flooring\_type, cladding\_type, ceiling\_type, and so on. These features are highly dependent on the price of the property and can significantly impact the accuracy of our predictions. Therefore, it would be important to include these features in future iterations of the model, in order to improve its overall performance.

# **12. Link to Code**

# **13. References**

1. <https://www.djangoproject.com/>
2. <https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>
3. <https://www.kaggle.com/code/ashydv/housing-price-prediction-linear-regression>
4. <https://dev.mysql.com/doc/>
5. <https://scikit-learn.org/stable/tutorial/index.html>
6. <https://pandas.pydata.org/pandas-docs/stable/getting_started/tutorials.html>
7. <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>