**HOUSING PRICES**

**A MINI PROJECT REPORT**

**Submitted by**

**VASA CHANDANA [RA2011047010080]**

**SHRUTI IYENGAR [RA2011047010105]**

**DEVANSH JAISWAL [RA2011047010133]**

                                                      Under the guidance of

**Dr. Rakesh Kumar M**

(Assistant Professor, Department of Computational Intelligence)

                                     in partial fulfillment for the award of the degree

                                                                  Of

**BACHELOR OF TECHNOLOGY**

                                                                  in

**COMPUTATIONAL INTELLIGENCE**

                                                                  of

**FACULTY OF ENGINEERING AND TECHNOLOGY**



**S.R.M. Nagar, Kattankulathur, Chengalpattu**

**JUNE 2022**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(Under Section 3 of UGC Act, 1956)**

**BONAFIDE CERTIFICATE**

Certified that this project report titled **“HOUSING PRICES”** is the bonafide work of **“VASA CHANDANA [RA2011047010080], SHRUTI IYENGAR [RA2011047010105], DEVANSH JAISWAL [RA2011047010133]”**, who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE                                                                            SIGNATURE**

Dr. RAKESH KUMAR M                                                            Dr. ANNIE UTHRA

**GUIDE                                                                                      HEAD OF THE DEPARTMENT**

Assistant Professor                                                                  Dept. of Computational Intelligence

Dept. of Computational Intelligence

Signature of the Internal Examiner                                          Signature of the External Examiner

**ACKNOWLEDGEMENTS**

I would like to express my deepest gratitude to my guide, **Dr. Rakesh Kumar M**, her valuable guidance, consistent encouragement, personal caring, timely help and providing me with an excellent atmosphere for doing the project. All through the work, in spite of his busy schedule, he has extended cheerful and cordial support to me for completing this project work.

**VASA CHANDANA**

**SHRUTI IYENGAR**

**DEVANSH JAISWAL**

**Chapter 1**

**ABSTRACT**

House price forecasting is an important topic of real estate. The literature attempts to derive useful knowledge from historical data of property markets. Machine learning techniques are applied to analyze historical property transactions in India to discover useful models for house buyers and sellers. Revealed is the high discrepancy between house prices in the most expensive and most affordable areas in the city.

As house price prediction plays an important role in real estate industry many researchers from different ﬁelds are interested in developing a regression model for the house price to obtain an accurate prediction and explore the factors affecting the house price. In this study, we aim to develop an accurate regression model using xgboost algorithms and explain the type of information which has an impact on the house price. For this purpose, we use the House Price dataset available on Kaggle.

This model gives the functionality for buyers, allowing them to search for houses by features or address. When the user will search for the property, original property value and predicted property value will be displayed. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. This prediction system will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction. The property, original property value and predicted property value will be displayed. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. For the price prediction we will be using regression algorithm. The functioning of this project involves a model which accepts customer’s specifications and then uses the application of data mining. This prediction model will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction.

**INTRODUCTION**

Investment is a business activity that most people are interested in this globalization era. There are several objects that are often used for investment, for example, gold, stocks and property. In particular, property investment has increased significantly since 2011, both on demand and property selling. One of the increasing of property demand is because of high population in Indonesia. Indonesian Central Bureau of Statistics states that in East Java 50% of the population of East Java classified as a young population who have age approximately at 30 years old. The result of this census indicates that the younger generation will need a house or buy a house in the future. Based on preliminary research conducted, there are two standards of house price which are valid in buying and selling transaction of a house that is house price based on the developer (market selling price) and price based on Value of Selling Tax Object (NJOP). According to Lim, et al the fundamental problem for a developer is to determine the selling price of a house. In determining the price of home, the developer must calculate carefully and determine the appropriate method because property prices always increase continuously and almost never fall in the long term or short.

The general and standardized real estate characteristics are often listed separately from the asking price and general description because these characteristics are separately listed in a structured way, they can be easily compared across the whole range of potential houses. Because every house also has its own unique characteristics, such as a particular  
view or type of sink, house sellers can provide a summary of all the important features of the house in the description. All given real estate features can be considered by the potential buyers, but it is nearly impossible to provide an automated comparison on all variables due to the large diversity. This is also true in the other direction: house sellers  
have to make an estimation of the value based on its features in comparison to the current market price of similar houses. The diversity of features makes it challenging to estimate an adequate market price. Apart from providing a summary of the important features of the house, the house description is also a means of raising curiosity in the reader,  
or in other words to persuade the person. It is possible that there are certain word sequences in the natural language text that attracts potential buyers more than others. Therefore, there might be a relation between the language used in the description and the price of the property. For example, a description with the word highly can outperform one with the  
word very looking at price fluctuation: the difference between real estate asking- and selling price. This can mean that the word highly is commonly seen in descriptions that show an increase in real estate price while the word very generally leads to a decrease in price. In addition, we can also find words that are distinctive for a certain range in  
selling- or asking price, thus can be used for prediction tasks.

There are several approaches that can be used to determine the price of the house, one of the best approaches is using regression techniques with xgboost library with hyper parameter tuning.

XGBoost is an Extreme Gradient Boosted Trees create several models, taking past trees and factoring in their predictions to create a new tree, with the purpose of minimizing their prediction error. The interpretability of results is good, we can visualize the final trees to determine what variables split the tree and how the thresholds worked. With this XGboost model we use tuned hyper parameters. A hyper parameter is a type of parameter, external to the model, set before the learning process begins. It’s tunable and can directly affect how well a model performs.

This project brings together the latest research on prediction markets to further their utilization by economic forecasters. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project efficiently analyses previous market trends and price ranges, to predict future prices. This topic brings together the latest research on prediction markets to further their utilization by economic forecasters. It provides a description of prediction markets, and also the current markets which are useful in understanding the market which helps in making useful predictions. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project uses data mining algorithm to predict prices by analyzing current house prices, thereby forecasting the future prices according to the users’ requirements.

* 1. **OBJECTIVE**

The objective is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. The functioning involves a model which accepts customers’ specifications and then combines the application of regression techniques xgboost library, a gradient boosting library with hyper parameter tuning. This prediction will help customers to invest in an estate without approaching an agent. It also decreases the risk involved in the transaction. The current property buying or selling is hectic and expensive. As the customer has to roam places and has to pay commission to the Real estate agent. Also, the customer/buyer does not know whether the property is profitable in future or not. Hence, we design a model using regression techniques to overcome the drawbacks of current system.

**1.2 PROBLEM STATEMENT**

Prices of real estate properties are sophisticatedly linked with our economy. Despite this, we do not have accurate measures of house prices based on the vast amount of data available. Proper and justified prices of properties can bring in a lot of transparency and trust back to the real estate industry, which is very important for most consumers.

Hence, naturally, we need a solution to help us to predict the prices of houses and bring the trust to the real estate industry.

* 1. **PROPOSED SOLUTION**

We use XGboost with hyper parameter tuning for creating our model.

**XGBoost** is an optimized distributed gradient boosting library designed to be highly **efficient**, **flexible** and **portable**. It implements machine learning algorithms under the [Gradient Boosting](https://en.wikipedia.org/wiki/Gradient_boosting) framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way.

With this XGboost model we use tuned hyper parameters. A hyper parameter is a type of parameter, external to the model, set before the learning process begins. It’s tunable and can directly affect how well a model performs.

XGBoost is an Extreme Gradient Boosted Trees create several models, taking past trees and factoring in their predictions to create a new tree, with the purpose of minimizing their prediction error.

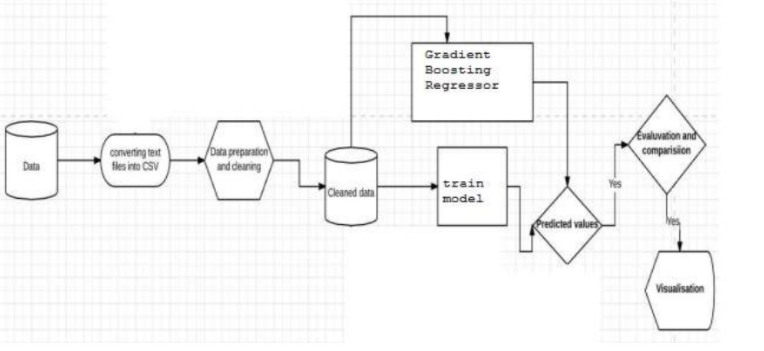
The interpretability of results is good, we can visualize the final trees to determine what variables split the tree and how the thresholds worked.

Speed and accuracy — XG boost is quick and accurate compared to older algorithms.

**CHAPTER - 2**

**PROPOSED METHODOLOGY**

**2.1 ARCHITECTURE DIAGRAM**



**2.2 DESCRIPTION OF PROPOSED MODEL:-**

XGBoost (eXtreme Gradient Boosting) is not only an algorithm. It’s an entire open-source library, designed as an optimized implementation of the Gradient Boosting framework. It focuses on speed, flexibility, and model performances. Its strength doesn’t only come from the algorithm, but also from all the underlying system optimization (parallelization, caching, hardware optimization, etc…).

In most cases, data scientist uses XGBoost with a “Tree Base learner”, which means that your XGBoost model is based on Decision Trees. But even though they are way less popular, you can also use XGboost with other base learners, such as linear model or Dart.

**CHAPTER - 3**

**TOOLS AND SOFTWARES USED**

**3.1 DATASET DESCRIPTION**

MSSubClass: Identifies the type of dwelling involved in the sale.

20 1-STORY 1946 & NEWER ALL STYLES

30 1-STORY 1945 & OLDER

40 1-STORY W/FINISHED ATTIC ALL AGES

45 1-1/2 STORY - UNFINISHED ALL AGES

50 1-1/2 STORY FINISHED ALL AGES

60 2-STORY 1946 & NEWER

70 2-STORY 1945 & OLDER

75 2-1/2 STORY ALL AGES

80 SPLIT OR MULTI-LEVEL

85 SPLIT FOYER

90 DUPLEX - ALL STYLES AND AGES

120 1-STORY PUD (Planned Unit Development) - 1946 & NEWER

150 1-1/2 STORY PUD - ALL AGES

160 2-STORY PUD - 1946 & NEWER

180 PUD - MULTILEVEL - INCL SPLIT LEV/FOYER

190 2 FAMILY CONVERSION - ALL STYLES AND AGES

MSZoning: Identifies the general zoning classification of the sale.

A Agriculture

C Commercial

FV Floating Village Residential

I Industrial

RH Residential High Density

RL Residential Low Density

RP Residential Low Density Park

RM Residential Medium Density

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

Street: Type of road access to property

Grvl Gravel

Pave Paved

Alley: Type of alley access to property

Grvl Gravel

Pave Paved

NA No alley access

LotShape: General shape of property

Reg Regular

IR1 Slightly irregular

IR2 Moderately Irregular

IR3 Irregular

LandContour: Flatness of the property

Lvl Near Flat/Level

Bnk Banked - Quick and significant rise from street grade to building

HLS Hillside - Significant slope from side to side

Low Depression

Utilities: Type of utilities available

AllPub All public Utilities (E, G, W,& S)

NoSewr Electricity, Gas, and Water (Septic Tank)

NoSeWa Electricity and Gas Only

ELO Electricity only

LotConfig: Lot configuration

Inside Inside lot

Corner Corner lot

CulDSac Cul-de-sac

FR2 Frontage on 2 sides of property

FR3 Frontage on 3 sides of property

LandSlope: Slope of property

Gtl Gentle slope

Mod Moderate Slope

Sev Severe Slope

Neighborhood: Physical locations within Ames city limits

Blmngtn Bloomington Heights

Blueste Bluestem

BrDale Briardale

BrkSide Brookside

ClearCr Clear Creek

CollgCr College Creek

Crawfor Crawford

Edwards Edwards

Gilbert Gilbert

IDOTRR Iowa DOT and Rail Road

MeadowV Meadow Village

Mitchel Mitchell

Names North Ames

NoRidge Northridge

NPkVill Northpark Villa

NridgHt Northridge Heights

NWAmes Northwest Ames

OldTown Old Town

SWISU South & West of Iowa State University

Sawyer Sawyer

SawyerW Sawyer West

Somerst Somerset

StoneBr Stone Brook

Timber Timberland

Veenker Veenker

Condition1: Proximity to various conditions

Artery Adjacent to arterial street

Feedr Adjacent to feeder street

Norm Normal

RRNn Within 200' of North-South Railroad

RRAn Adjacent to North-South Railroad

PosN Near positive off-site feature--park, greenbelt, etc.

PosA Adjacent to postive off-site feature

RRNe Within 200' of East-West Railroad

RRAe Adjacent to East-West Railroad

Condition2: Proximity to various conditions (if more than one is present)

Artery Adjacent to arterial street

Feedr Adjacent to feeder street

Norm Normal

RRNn Within 200' of North-South Railroad

RRAn Adjacent to North-South Railroad

PosN Near positive off-site feature--park, greenbelt, etc.

PosA Adjacent to postive off-site feature

RRNe Within 200' of East-West Railroad

RRAe Adjacent to East-West Railroad

BldgType: Type of dwelling

1Fam Single-family Detached

2FmCon Two-family Conversion; originally built as one-family dwelling

Duplx Duplex

TwnhsE Townhouse End Unit

TwnhsI Townhouse Inside Unit

HouseStyle: Style of dwelling

1Story One story

1.5Fin One and one-half story: 2nd level finished

1.5Unf One and one-half story: 2nd level unfinished

2Story Two story

2.5Fin Two and one-half story: 2nd level finished

2.5Unf Two and one-half story: 2nd level unfinished

SFoyer Split Foyer

SLvl Split Level

OverallQual: Rates the overall material and finish of the house

10 Very Excellent

9 Excellent

8 Very Good

7 Good

6 Above Average

5 Average

4 Below Average

3 Fair

2 Poor

1 Very Poor

OverallCond: Rates the overall condition of the house

10 Very Excellent

9 Excellent

8 Very Good

7 Good

6 Above Average

5 Average

4 Below Average

3 Fair

2 Poor

1 Very Poor

YearBuilt: Original construction date

YearRemodAdd: Remodel date (same as construction date if no remodeling or additions)

RoofStyle: Type of roof

Flat Flat

Gable Gable

Gambrel Gabrel (Barn)

Hip Hip

Mansard Mansard

Shed Shed

RoofMatl: Roof material

ClyTile Clay or Tile

CompShg Standard (Composite) Shingle

Membran Membrane

Metal Metal

Roll Roll

Tar&Grv Gravel & Tar

WdShake Wood Shakes

WdShngl Wood Shingles

Exterior1st: Exterior covering on house

AsbShng Asbestos Shingles

AsphShn Asphalt Shingles

BrkComm Brick Common

BrkFace Brick Face

CBlock Cinder Block

CemntBd Cement Board

HdBoard Hard Board

ImStucc Imitation Stucco

MetalSd Metal Siding

Other Other

Plywood Plywood

PreCast PreCast

Stone Stone

Stucco Stucco

VinylSd Vinyl Siding

Wd Sdng Wood Siding

WdShing Wood Shingles

Exterior2nd: Exterior covering on house (if more than one material)

AsbShng Asbestos Shingles

AsphShn Asphalt Shingles

BrkComm Brick Common

BrkFace Brick Face

CBlock Cinder Block

CemntBd Cement Board

HdBoard Hard Board

ImStucc Imitation Stucco

MetalSd Metal Siding

Other Other

Plywood Plywood

PreCast PreCast

Stone Stone

Stucco Stucco

VinylSd Vinyl Siding

Wd Sdng Wood Siding

WdShing Wood Shingles

MasVnrType: Masonry veneer type

BrkCmn Brick Common

BrkFace Brick Face

CBlock Cinder Block

None None

Stone Stone

MasVnrArea: Masonry veneer area in square feet

ExterQual: Evaluates the quality of the material on the exterior

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

ExterCond: Evaluates the present condition of the material on the exterior

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

Foundation: Type of foundation

BrkTil Brick & Tile

CBlock Cinder Block

PConc Poured Contrete

Slab Slab

Stone Stone

Wood Wood

BsmtQual: Evaluates the height of the basement

Ex Excellent (100+ inches)

Gd Good (90-99 inches)

TA Typical (80-89 inches)

Fa Fair (70-79 inches)

Po Poor (<70 inches

NA No Basement

BsmtCond: Evaluates the general condition of the basement

Ex Excellent

Gd Good

TA Typical - slight dampness allowed

Fa Fair - dampness or some cracking or settling

Po Poor - Severe cracking, settling, or wetness

NA No Basement

BsmtExposure: Refers to walkout or garden level walls

Gd Good Exposure

Av Average Exposure (split levels or foyers typically score average or above)

Mn Mimimum Exposure

No No Exposure

NA No Basement

BsmtFinType1: Rating of basement finished area

GLQ Good Living Quarters

ALQ Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinshed

NA No Basement

BsmtFinSF1: Type 1 finished square feet

BsmtFinType2: Rating of basement finished area (if multiple types)

GLQ Good Living Quarters

ALQ Average Living Quarters

BLQ Below Average Living Quarters

Rec Average Rec Room

LwQ Low Quality

Unf Unfinshed

NA No Basement

BsmtFinSF2: Type 2 finished square feet

BsmtUnfSF: Unfinished square feet of basement area

TotalBsmtSF: Total square feet of basement area

Heating: Type of heating

Floor Floor Furnace

GasA Gas forced warm air furnace

GasW Gas hot water or steam heat

Grav Gravity furnace

OthW Hot water or steam heat other than gas

Wall Wall furnace

HeatingQC: Heating quality and condition

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

Po Poor

CentralAir: Central air conditioning

N No

Y Yes

Electrical: Electrical system

SBrkr Standard Circuit Breakers & Romex

FuseA Fuse Box over 60 AMP and all Romex wiring (Average)

FuseF 60 AMP Fuse Box and mostly Romex wiring (Fair)

FuseP 60 AMP Fuse Box and mostly knob & tube wiring (poor)

Mix Mixed

1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet

LowQualFinSF: Low quality finished square feet (all floors)

GrLivArea: Above grade (ground) living area square feet

BsmtFullBath: Basement full bathrooms

BsmtHalfBath: Basement half bathrooms

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

Bedroom: Bedrooms above grade (does NOT include basement bedrooms)

Kitchen: Kitchens above grade

KitchenQual: Kitchen quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Functional: Home functionality (Assume typical unless deductions are warranted)

Typ Typical Functionality

Min1 Minor Deductions 1

Min2 Minor Deductions 2

Mod Moderate Deductions

Maj1 Major Deductions 1

Maj2 Major Deductions 2

Sev Severely Damaged

Sal Salvage only

Fireplaces: Number of fireplaces

FireplaceQu: Fireplace quality

Ex Excellent - Exceptional Masonry Fireplace

Gd Good - Masonry Fireplace in main level

TA Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement

Fa Fair - Prefabricated Fireplace in basement

Po Poor - Ben Franklin Stove

NA No Fireplace

GarageType: Garage location

2Types More than one type of garage

Attchd Attached to home

Basment Basement Garage

BuiltIn Built-In (Garage part of house - typically has room above garage)

CarPort Car Port

Detchd Detached from home

NA No Garage

GarageYrBlt: Year garage was built

GarageFinish: Interior finish of the garage

Fin Finished

RFn Rough Finished

Unf Unfinished

NA No Garage

GarageCars: Size of garage in car capacity

GarageArea: Size of garage in square feet

GarageQual: Garage quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

GarageCond: Garage condition

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Po Poor

NA No Garage

PavedDrive: Paved driveway

Y Paved

P Partial Pavement

N Dirt/Gravel

WoodDeckSF: Wood deck area in square feet

OpenPorchSF: Open porch area in square feet

EnclosedPorch: Enclosed porch area in square feet

3SsnPorch: Three season porch area in square feet

ScreenPorch: Screen porch area in square feet

PoolArea: Pool area in square feet

PoolQC: Pool quality

Ex Excellent

Gd Good

TA Average/Typical

Fa Fair

NA No Pool

Fence: Fence quality

GdPrv Good Privacy

MnPrv Minimum Privacy

GdWo Good Wood

MnWw Minimum Wood/Wire

NA No Fence

MiscFeature: Miscellaneous feature not covered in other categories

Elev Elevator

Gar2 2nd Garage (if not described in garage section)

Othr Other

Shed Shed (over 100 SF)

TenC Tennis Court

NA None

MiscVal: $Value of miscellaneous feature

MoSold: Month Sold (MM)

YrSold: Year Sold (YYYY)

SaleType: Type of sale

WD Warranty Deed - Conventional

CWD Warranty Deed - Cash

VWD Warranty Deed - VA Loan

New Home just constructed and sold

COD Court Officer Deed/Estate

Con Contract 15% Down payment regular terms

ConLw Contract Low Down payment and low interest

ConLI Contract Low Interest

ConLD Contract Low Down

Oth Other

SaleCondition: Condition of sale

Normal Normal Sale

Abnorml Abnormal Sale - trade, foreclosure, short sale

AdjLand Adjoining Land Purchase

Alloca Allocation - two linked properties with separate deeds, typically condo with a garage unit

Family Sale between family members

Partial Home was not completed when last assessed (associated with New Homes)

**3.2 TOOLS DESCRIPTION**

1. **JUPYTER NOTEBOOK**

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at [Project Jupyter](http://jupyter.org/).

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

Notebook documents (or “notebooks”, all lower case) are documents produced by the [Jupyter Notebook App](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-app), which contain both computer code (e.g. python) and rich text elements (paragraph, equations, figures, links, etc…). Notebook documents are both human-readable documents containing the analysis description and the results (figures, tables, etc..) as well as executable documents which can be run to perform data analysis.

A notebook kernel is a “computational engine” that executes the code contained in a [Notebook document](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-document). The ipython kernel, referenced in this guide, executes python code. Kernels for many other languages exist ([official kernels](http://jupyter.readthedocs.org/en/latest/#kernels)).

When you open a [Notebook document](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-document), the associated kernel is automatically launched. When the notebook is executed (either cell-by-cell or with menu Cell -> Run All), the kernel performs the computation and produces the results. Depending on the type of computations, the kernel may consume significant CPU and RAM. Note that the RAM is not released until the kernel is shut-down.

The *Notebook Dashboard* is the component which is shown first when you launch [Jupyter Notebook App](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-app). The *Notebook Dashboard* is mainly used to open [notebook documents](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#notebook-document), and to manage the running [kernels](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html#kernel) (visualize and shutdown).

The *Notebook Dashboard* has other features similar to a file manager, namely navigating folders and renaming/deleting files.

1. **VS CODE**

**Visual Studio Code**, also commonly referred to as **VS Code**, is a [source-code editor](https://en.wikipedia.org/wiki/Source-code_editor) made by [Microsoft](https://en.wikipedia.org/wiki/Microsoft) for [Windows](https://en.wikipedia.org/wiki/Windows), [Linux](https://en.wikipedia.org/wiki/Linux) and [macOS](https://en.wikipedia.org/wiki/MacOS).Features include support for [debugging](https://en.wikipedia.org/wiki/Debugging), [syntax highlighting](https://en.wikipedia.org/wiki/Syntax_highlighting), [intelligent code completion](https://en.wikipedia.org/wiki/Intelligent_code_completion), [snippets](https://en.wikipedia.org/wiki/Snippet_(programming)), [code refactoring](https://en.wikipedia.org/wiki/Code_refactoring), and embedded [Git](https://en.wikipedia.org/wiki/Git). Users can change the [theme](https://en.wikipedia.org/wiki/Theme_(computing)), [keyboard shortcuts](https://en.wikipedia.org/wiki/Keyboard_shortcut), preferences, and install [extensions](https://en.wikipedia.org/wiki/Plug-in_(computing)) that add additional functionality.

In the [Stack Overflow](https://en.wikipedia.org/wiki/Stack_Overflow) 2021 Developer Survey, Visual Studio Code was ranked the most popular developer environment tool, with 70% of 82,000 respondents reporting that they use it.

Visual Studio Code is a source-code editor that can be used with a variety of programming languages, including [Java](https://en.wikipedia.org/wiki/Java_(programming_language)), [JavaScript](https://en.wikipedia.org/wiki/JavaScript), [Go](https://en.wikipedia.org/wiki/Go_(programming_language)), [Node.js](https://en.wikipedia.org/wiki/Node.js), [Python](https://en.wikipedia.org/wiki/Python_(programming_language)), [C++](https://en.wikipedia.org/wiki/C%2B%2B) and [Fortran](https://en.wikipedia.org/wiki/Fortran). It is based on the [Electron](https://en.wikipedia.org/wiki/Electron_(software_framework)) framework, which is used to develop [Node.js](https://en.wikipedia.org/wiki/Node.js) [Web applications](https://en.wikipedia.org/wiki/Web_application) that run on the [Blink layout engine](https://en.wikipedia.org/wiki/Blink_layout_engine). Visual Studio Code employs the same editor component (codenamed "Monaco") used in [Azure DevOps](https://en.wikipedia.org/wiki/Azure_DevOps_Server) (formerly called Visual Studio Online and Visual Studio Team Services).

Out of the box, Visual Studio Code includes basic support for most common programming languages. This basic support includes [syntax highlighting](https://en.wikipedia.org/wiki/Syntax_highlighting), [bracket matching](https://en.wikipedia.org/wiki/Bracket_matching), [code folding](https://en.wikipedia.org/wiki/Code_folding), and configurable snippets. Visual Studio Code also ships with [IntelliSense](https://en.wikipedia.org/wiki/Intelligent_code_completion) for JavaScript, TypeScript, [JSON](https://en.wikipedia.org/wiki/JSON), [CSS](https://en.wikipedia.org/wiki/CSS), and [HTML](https://en.wikipedia.org/wiki/HTML), as well as debugging support for Node.js. Support for additional languages can be provided by freely available extensions on the VS Code Marketplace.

Instead of a project system, it allows users to open one or more directories, which can then be saved in workspaces for future reuse. This allows it to operate as a [language-agnostic](https://en.wikipedia.org/wiki/Language-agnostic) code editor for any language. It supports many programming languages and a set of features that differs per language. Unwanted files and folders can be excluded from the project tree via the settings. Many Visual Studio Code features are not exposed through menus or the user interface but can be accessed via the command palette.

Visual Studio Code can be extended via [extensions](https://en.wikipedia.org/wiki/Plug-in_(computing)), available through a central repository. This includes additions to the editorand language support. A notable feature is the ability to create extensions that add support for new [languages](https://en.wikipedia.org/wiki/Programming_language), [themes](https://en.wikipedia.org/wiki/Theme_(computing)), [debuggers](https://en.wikipedia.org/wiki/Debugger), [time travel debuggers](https://en.wikipedia.org/wiki/Time_travel_debugging), perform [static code analysis](https://en.wikipedia.org/wiki/Static_code_analysis), and add [code linters](https://en.wikipedia.org/wiki/Lint_(software)) using the [Language Server Protocol](https://en.wikipedia.org/wiki/Language_Server_Protocol).

[Source control](https://en.wikipedia.org/wiki/Source_control) is a built-in feature of Visual Studio Code. It has a dedicated tab inside of the menu bar where you can access version control settings and view changes made to the current project. To use the feature you must link Visual Studio Code to any supported version control system ([Git](https://en.wikipedia.org/wiki/Git), [Apache Subversion](https://en.wikipedia.org/wiki/Apache_Subversion), [Perforce](https://en.wikipedia.org/wiki/Perforce), etc.). This allows you to create repositories as well as to make push and [pull requests](https://en.wikipedia.org/wiki/Pull_request) directly from the Visual Studio Code program.

Visual Studio Code includes multiple extensions for [FTP](https://en.wikipedia.org/wiki/FTP), allowing the software to be used as a free alternative for web development. Code can be synced between the editor and the server, without downloading any extra software.

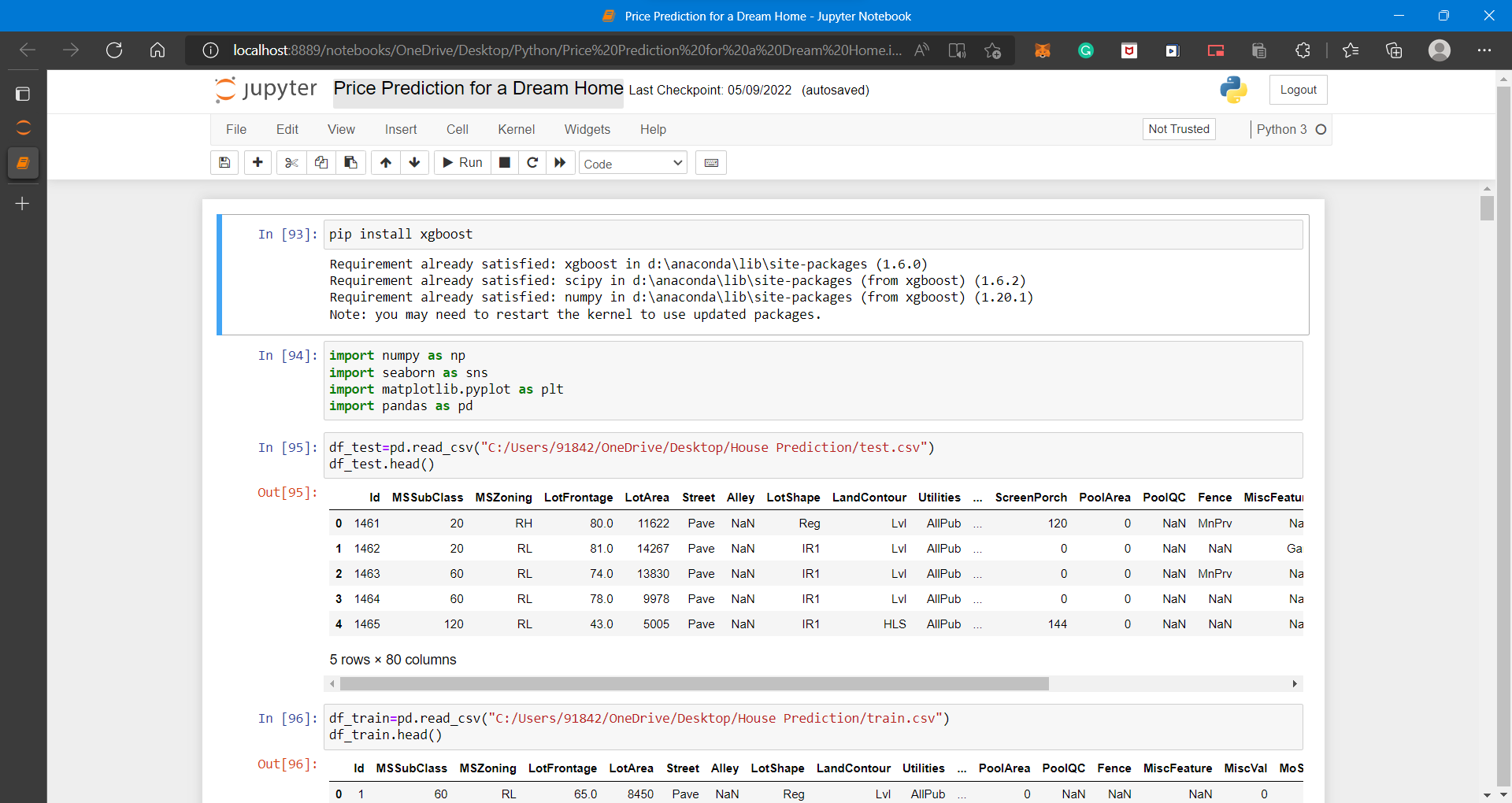
Visual Studio Code allows users to set the [code page](https://en.wikipedia.org/wiki/Code_page) in which the active document is saved, the [newline](https://en.wikipedia.org/wiki/Newline) character, and the programming language of the active document. This allows it to be used on any platform, in any locale, and for any given programming language.

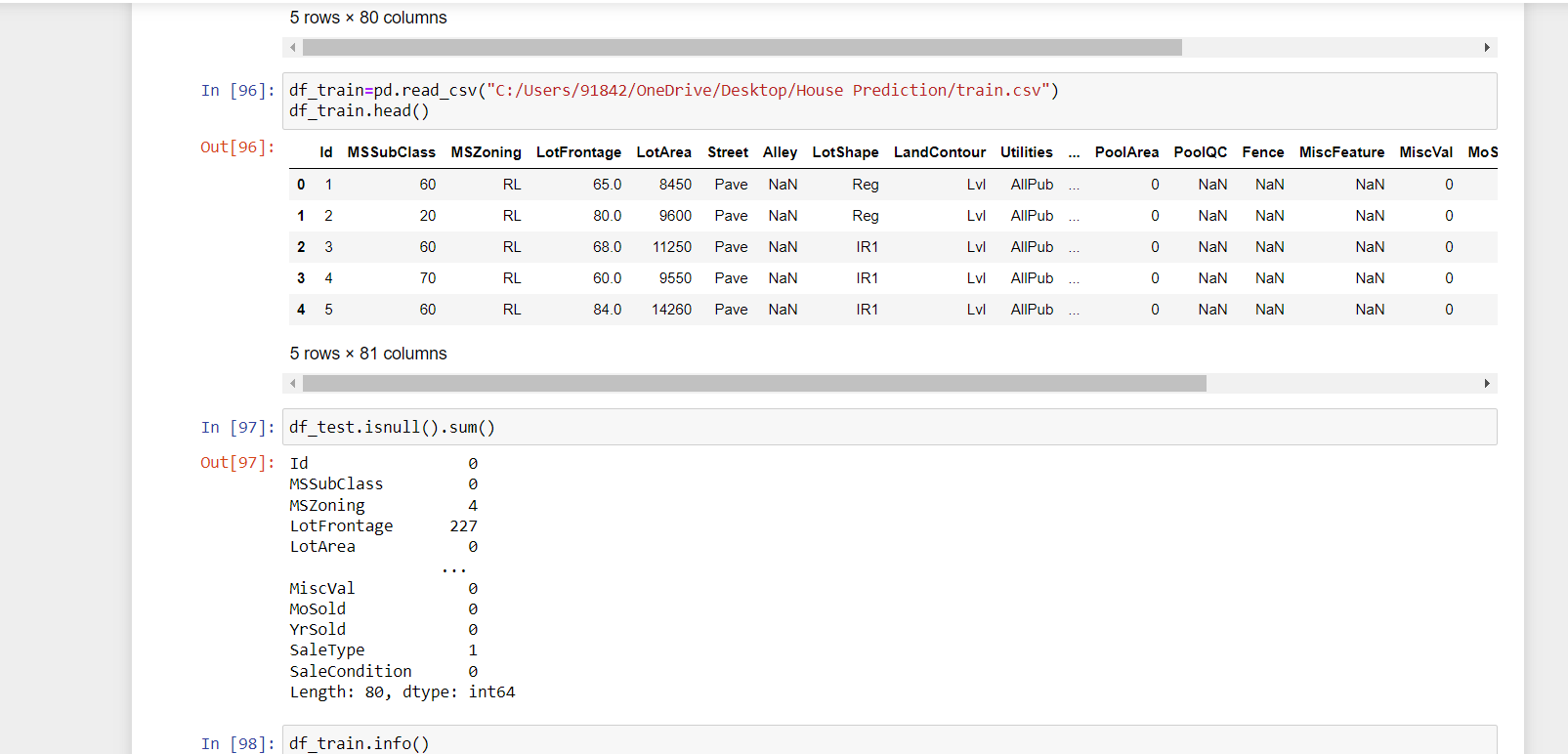
Visual Studio Code [collects usage data and sends it to Microsoft](https://en.wikipedia.org/wiki/Telemetry#Software), although this can be disabled. Due to the open-source nature of the application, the telemetry code is accessible to the public, who can see exactly what is collected.

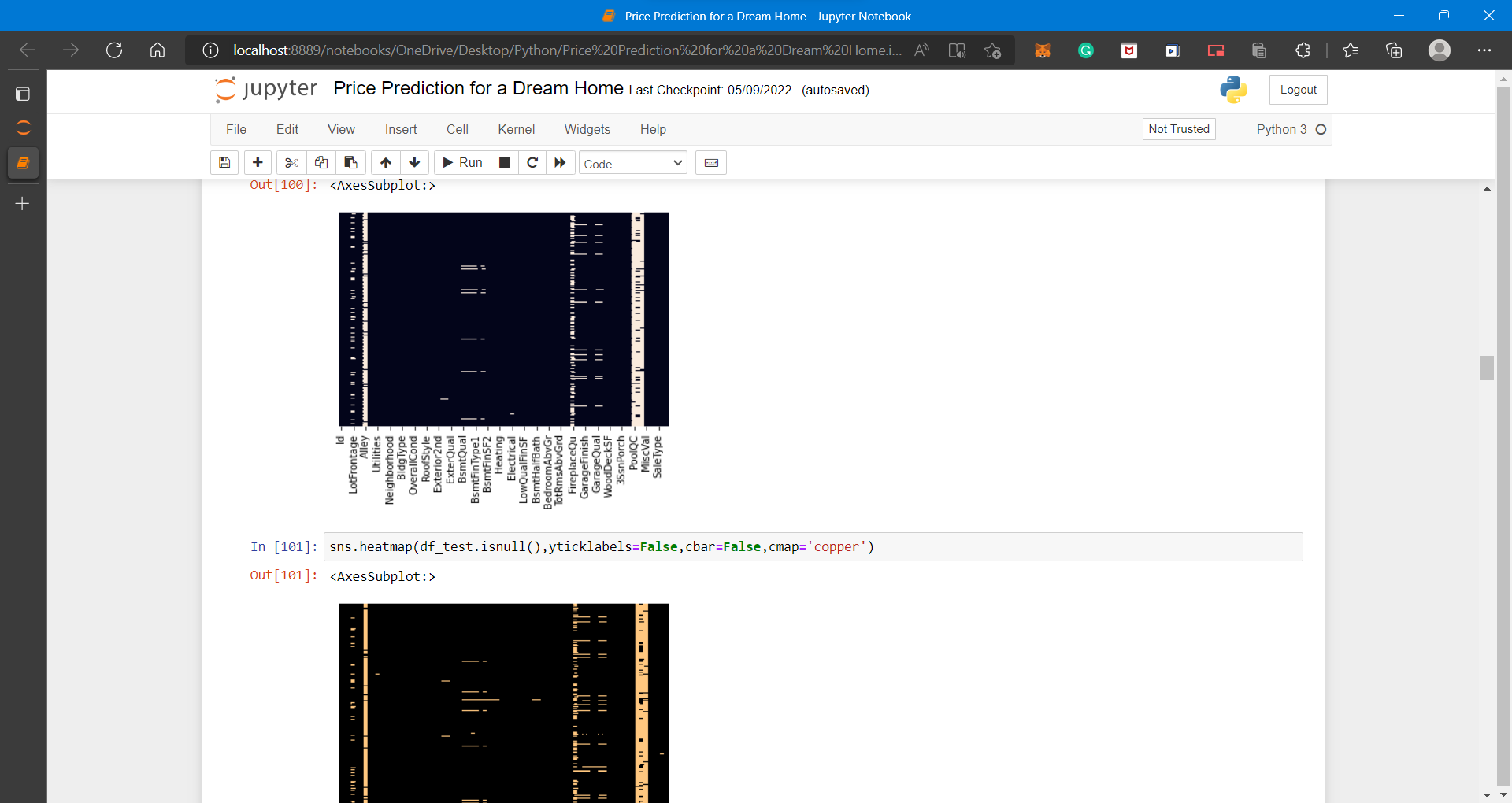
**CHAPTER – 4**

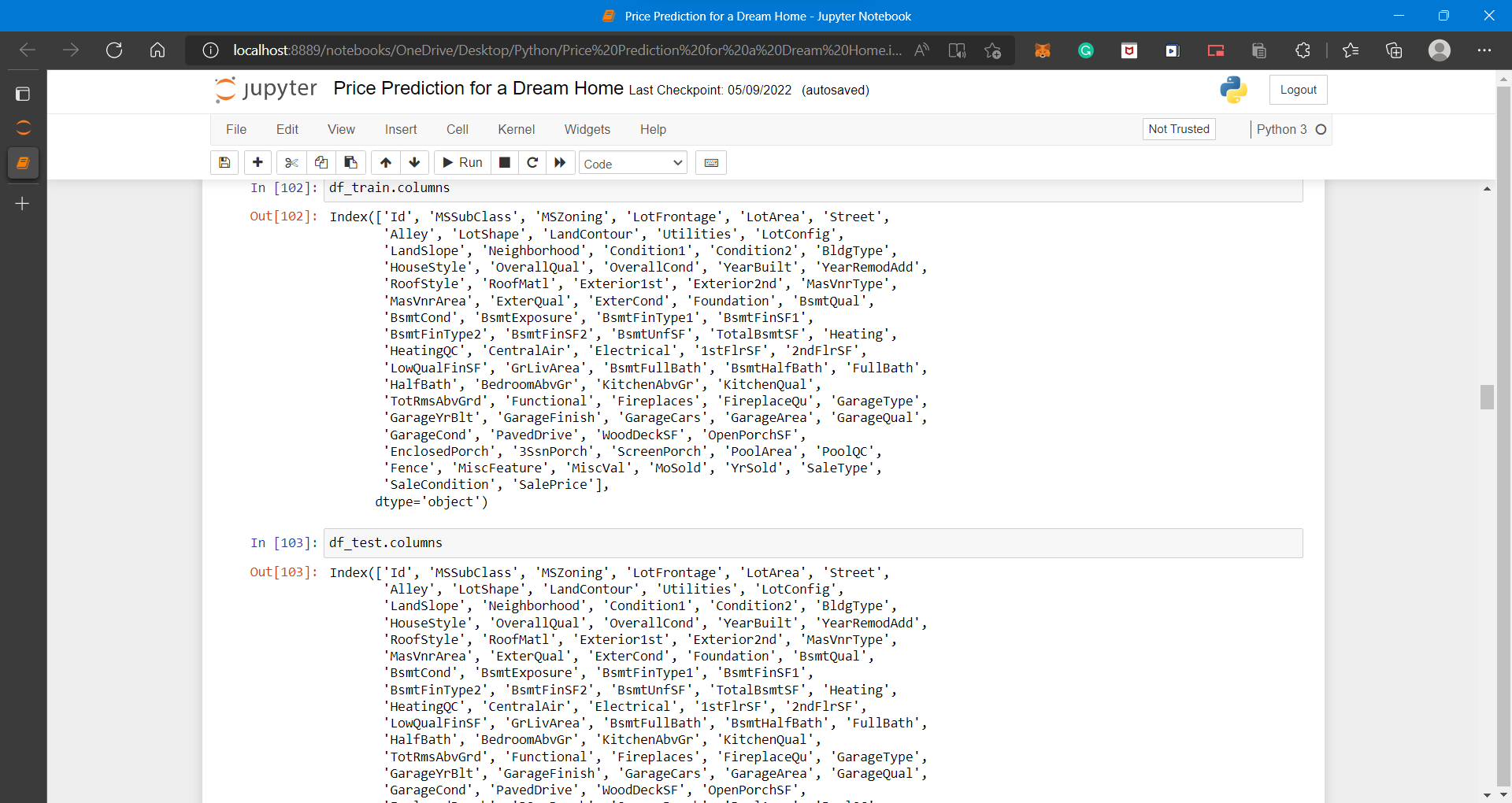
**RESULTS AND DISCUSSION**

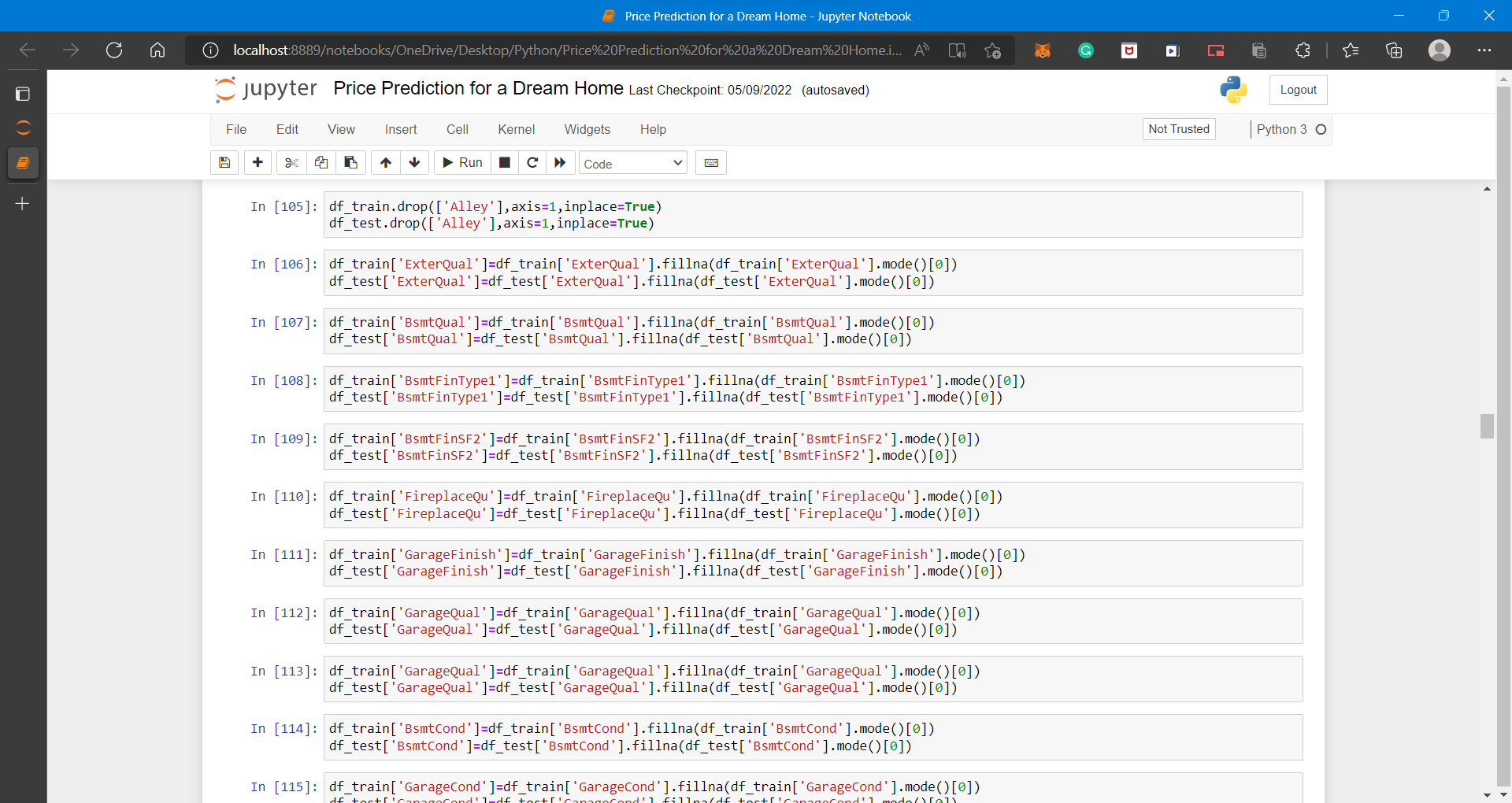
**4.1CODE IMPLEMENTATION:-**

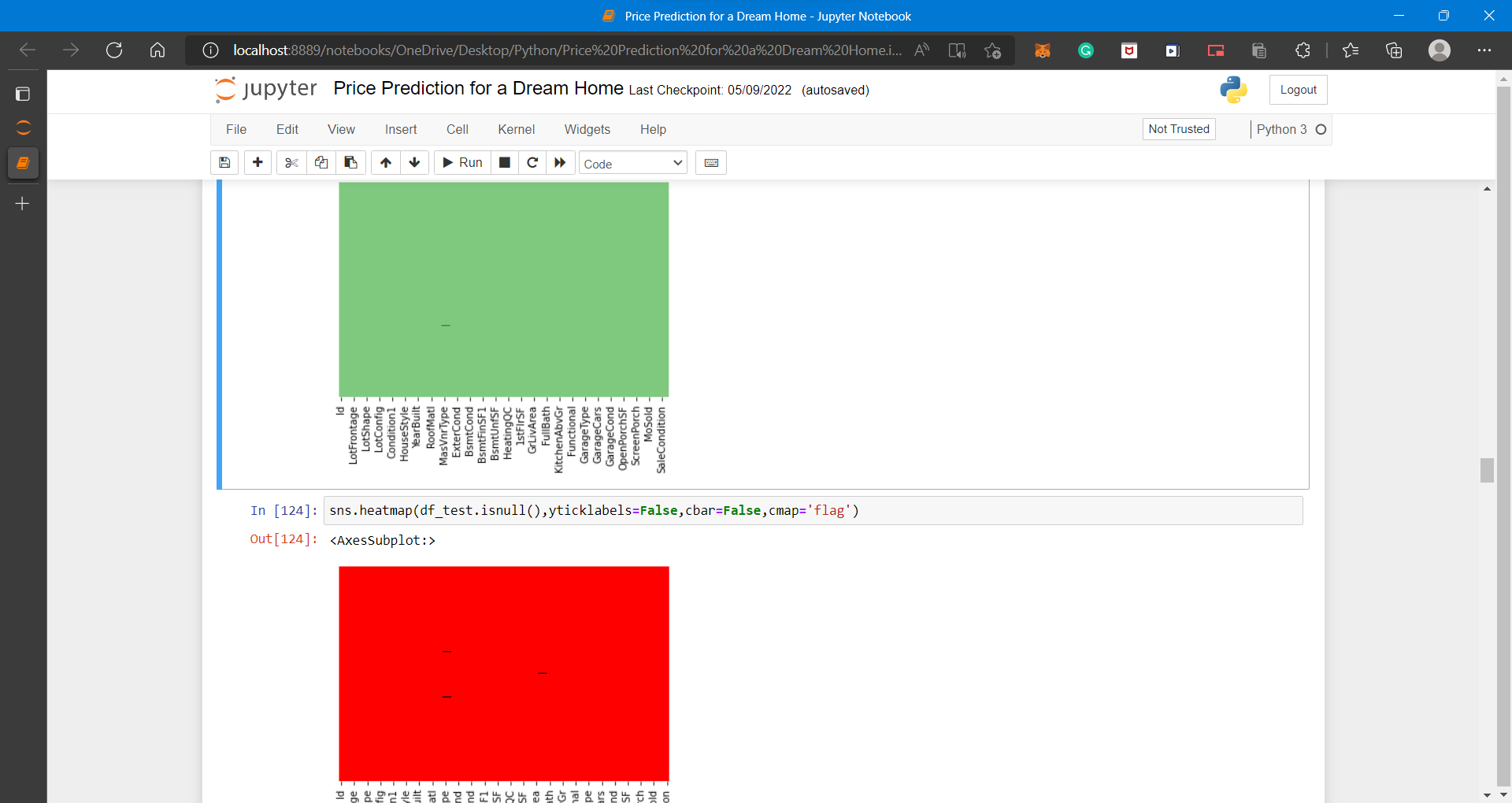
****

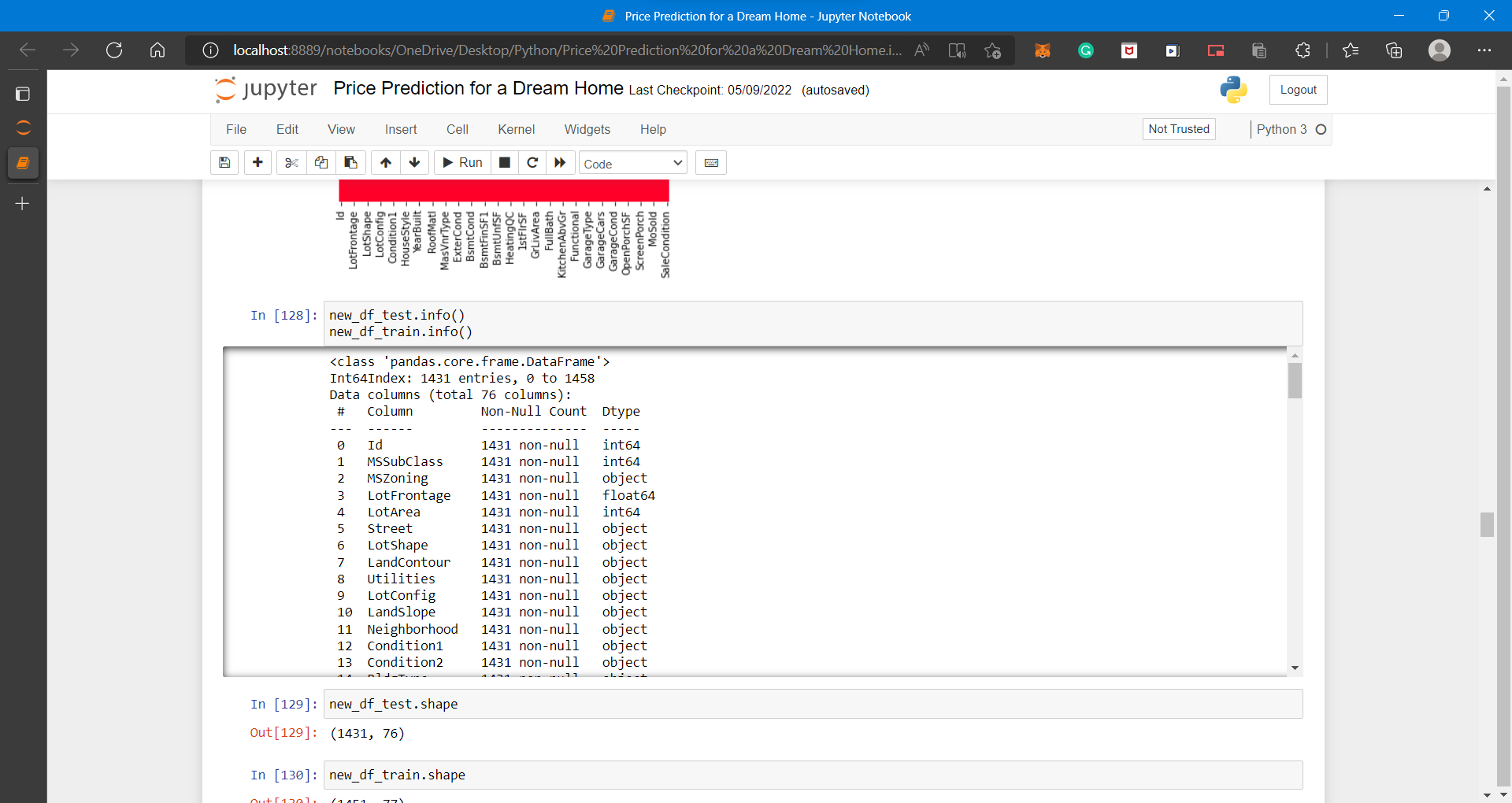
****

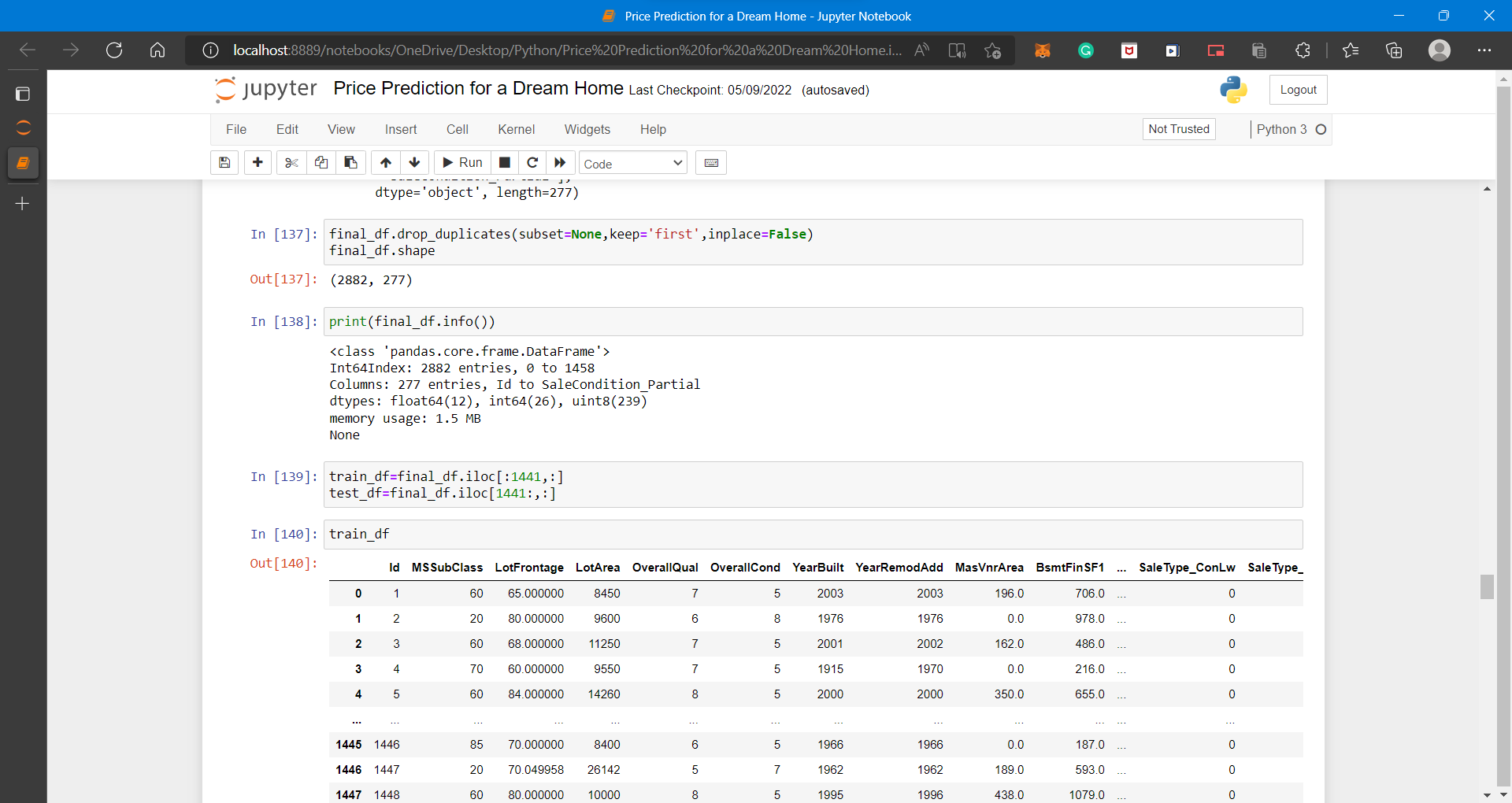
****

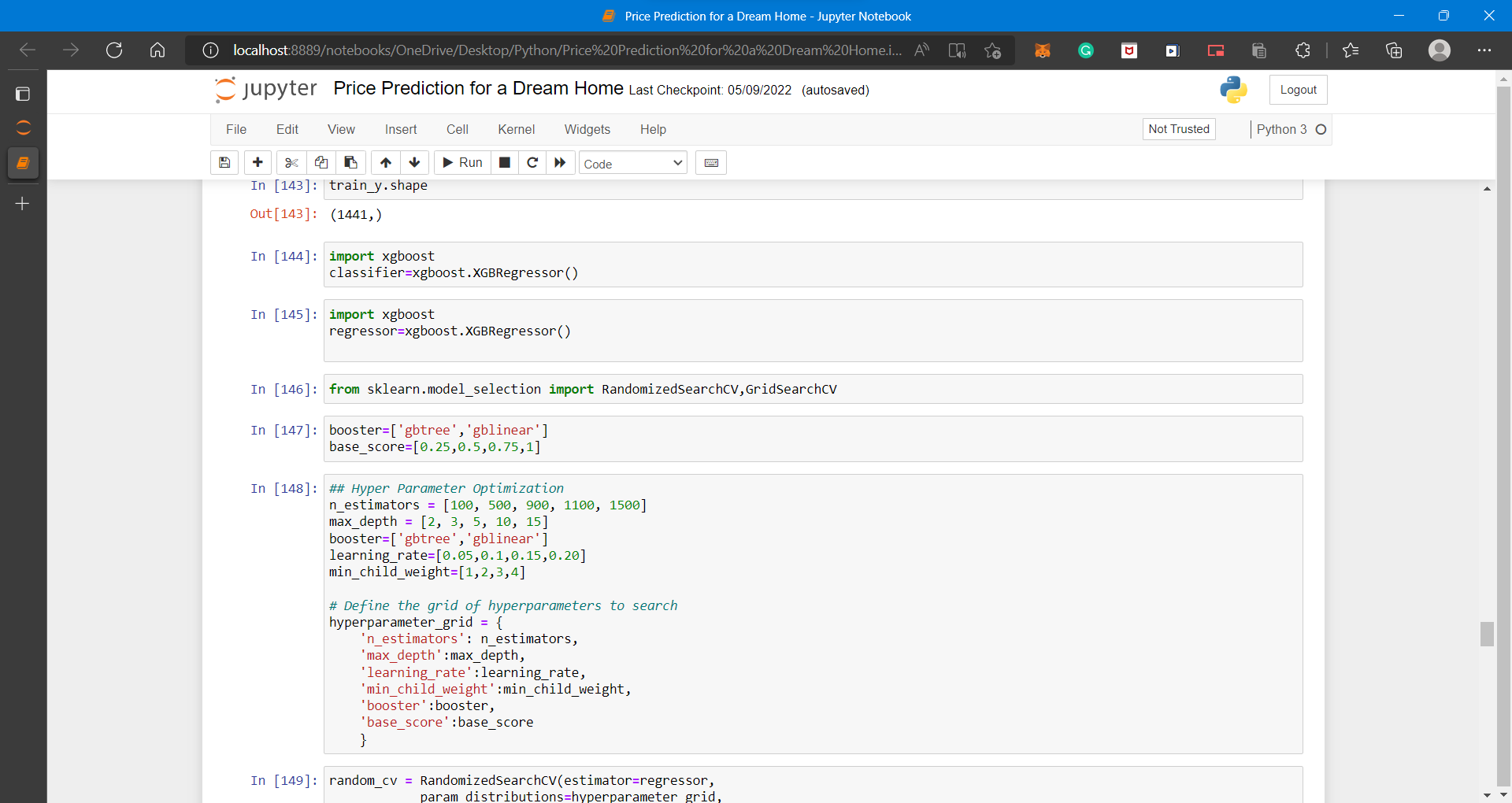
****

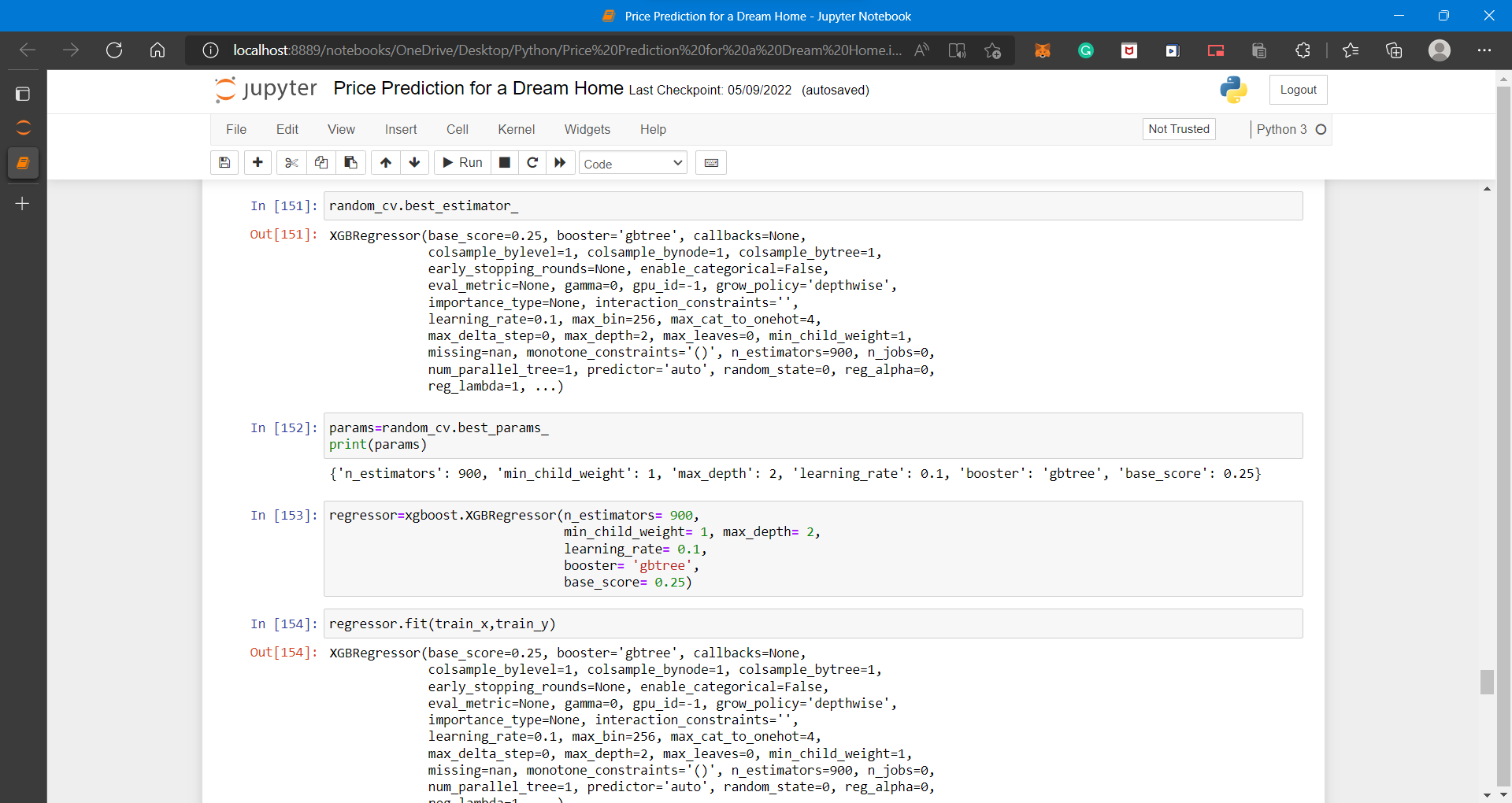
****

****

****

****

****

****

****

**4.2 PERFORMANCE EVALUATION, METRICS:-**

1. **k-fold cross-validation:**  It is a data partitioning strategy so that you can effectively use your dataset to build a more generalized model. The main intention of doing any kind of machine learning is to develop a more generalized model which can perform well on unseen data.
   * + - **Using k-fold cross-validation for evaluating a model’s performance**
       - **Using k-fold cross-validation for hyperparameter tuning**

**2.Cross validation:** Cross-validation is a technique for validating the model efficiency by training it on the subset of input data and testing on previously unseen subset of the input data. *We can also say that it is a technique to check how a statistical model generalizes to an independent dataset*.

**3.** **Rmse:** In machine learning, it is extremely helpful to have a single number to judge a model’s performance, whether it be during training, cross-validation, or monitoring after deployment. Root mean square error is one of the most widely used measures for this. It is a [proper scoring rule](https://sites.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf) that is intuitive to understand and compatible with some of the most common statistical assumptions.

* 1. **RESULTS:-**

Using XGBOOST, our base model provides:

**1)Training Score:** 0.985365509537986

**2)K-fold Value:** 0.88

**3)Mean Cross validation score:** 0.89

**4)Rmse value:** 112156.62661522532

For hyperparameter tuning, we perform many iterations of the entire fivefold cross-validation process, each of which sets different model settings, such as the number of estimators, min samples split, min sample leaf, max features, and max depth. Then, we compare all these models, pick up the best one, train it on the training set, and then evaluate on the test.

**4.4 DISCUSSION:-**

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. But this playground competition's dataset proves that much more influences price negotiations than the number of bedrooms or a white-picket fence.

With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, this competition challenges you to predict the final price of each home.

The model has achieved very good model fitting since most of the predicted values are lying very close or around the red line.

**CHAPTER - 5**

**CONCLUSION**

Improvement in computing technology has made it possible to examine social information that cannot previously be captured, processed and analysed. New analytical techniques of machine learning can be used in property research. This study is an exploratory attempt to use three machine learning algorithms in estimating housing prices, and then compare their results.

We have demonstrated that advanced machine learning algorithms can achieve very accurate prediction of property prices, as evaluated by the performance metrics. Given our dataset used in this paper, our main conclusion is that GBM are able to generate comparably accurate price estimations with lower prediction errors.

Our study has shown that advanced machine learning algorithms (GBM), are promising tools for property researchers to use in housing price predictions. However, we must be cautious that these machine learning tools also have their own limitations. There are often many potential features for researchers to choose and include in the models so that a very careful feature selection is essential.

To conclude, the application of machine learning in property research is still at an early stage. We hope this study has moved a small step ahead in providing some methodological and empirical contributions to property appraisal, and presenting an alternative approach to the valuation of housing prices. Future direction of research may consider incorporating additional property transaction data from a larger geographical location with more features, or analysing other property types beyond housing development.