



HOUSING PROJECT

CONTAINING PROBLEM
STATEMENT

AND

Understanding

Problem Statement : A USA housing company has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The company is looking at prospective properties to buy houses to enter the market.

Build a model using Machine Learning in order to predict the actual value of the prospective properties and decide whether to invest in them or not. For this company wants to know:

- Which variables are important to predict the price of variable?
- How do these variables describe the price of the house?

Understanding : The goal of this statistical analysis is to help us understand the relationship between house features and now these variables are used to predict house price.

Linear regression algorithm is used to predict the relationship(line) among data points. There can be many different (linear or nonlinear) ways to define the relationship. In the linear model, it is based on the intercept and the slope. To find out the most optimal relationship, we need to train the model with the data.

A scatterplot is a good starting point to help in determining the strength of the relationship between two variables. The correlation coefficient is a valuable measure of association between variables. Its value varies between -1 (weak relationship) and 1 (strong relationship).

EDA : (Exploratory Data Analysis)



1. EDA is an approach to analyse the data using Visual technique.
2. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representation.

EDA steps :



1. First, Lets import the data and have a look to see what kind of data we are dealing.
2. Import required libraries.

Dataset preparation [Splitting]

- 1) Data is divided into the Train set and Test set.
 - 2) We use the Train set to make the algorithm learn the data behaviour and then check the accuracy of our model on the Test set.
- X : The columns that are inserted into our model will be used to make prediction.
- Y : Target variable that will be predicted by the feature.

Visualizations

Data visualization is a graphical representation of information and data by using visual element like charts, graph and map.

Data visualization tools provide an accessible way to see and understand trends outliers and pattern in data additionally its provides an excellent way for employees or business owners to present data to non-technical audiences without confusion .

Visualization tools and technologies are essential to analyse massive amounts of information and make data- driven decisions.

Advantages of Visualization :

1. Easily sharing information.
2. Interactively explore opportunities.
3. Visualize patterns and relationships.

Disadvantage of Visualization :

1. Core messages can get lost in translation.
2. Correlation doesn't always mean causation.
3. Inaccurate information.

Steps

And

Assumptions

We used some major steps in this housing project. As per company requirement company want to know the actual value of the prospective properties and decide whether to invest in them or not. For this company wants to know:

- Which variables are important to predict the price of variable?
- How do these variables describe the price of the house?

These steps used in projects :

- 1) Head
- 2) Info
- 3) Tail
- 4) Shape
- 5) Describe
- 6) Column
- 7) Pair plot
- 8) Correlation
- 9) Heat map
- 10) Sk learn Model Selection
- 11) Split

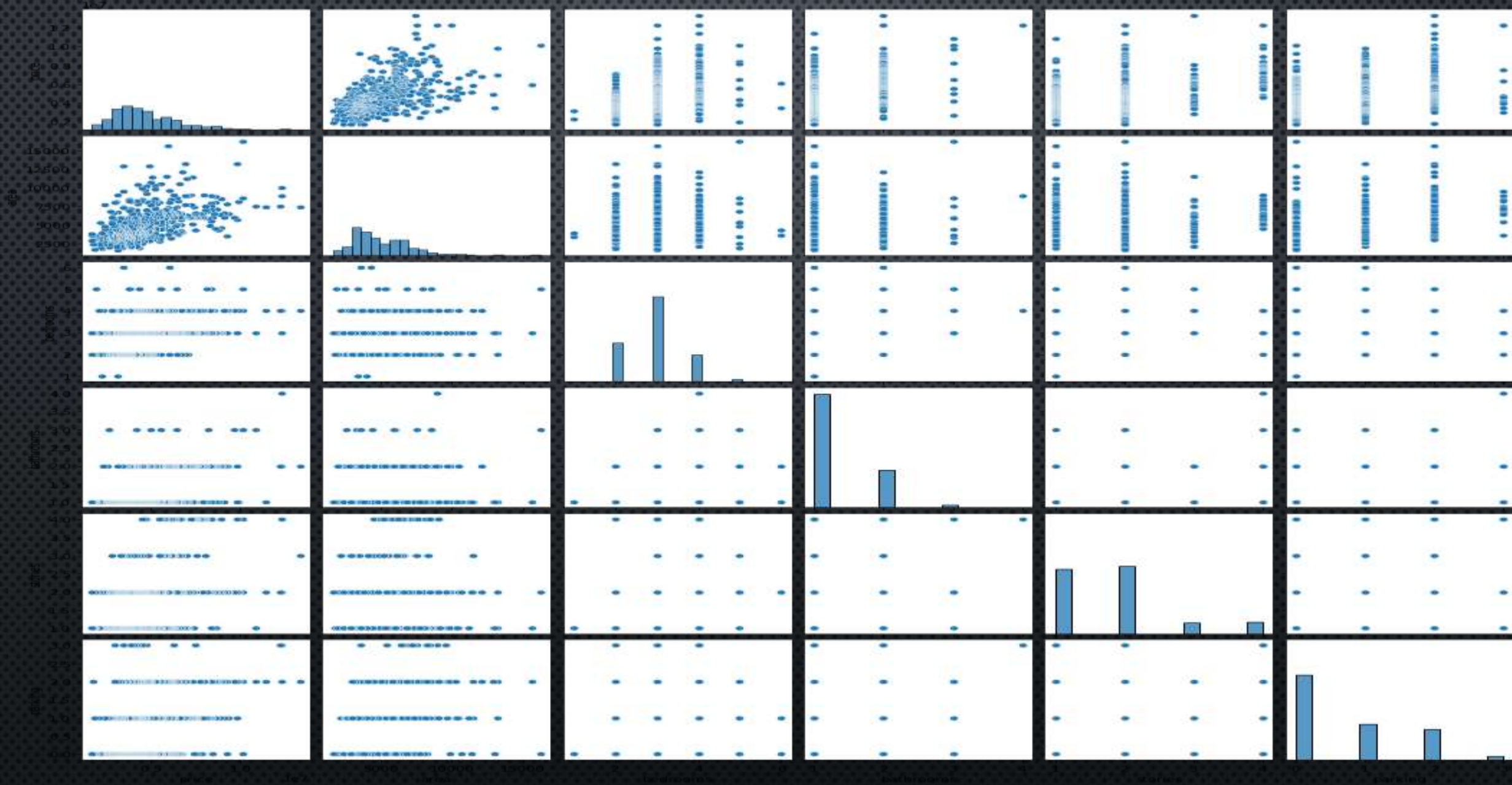
12) Linear Regression

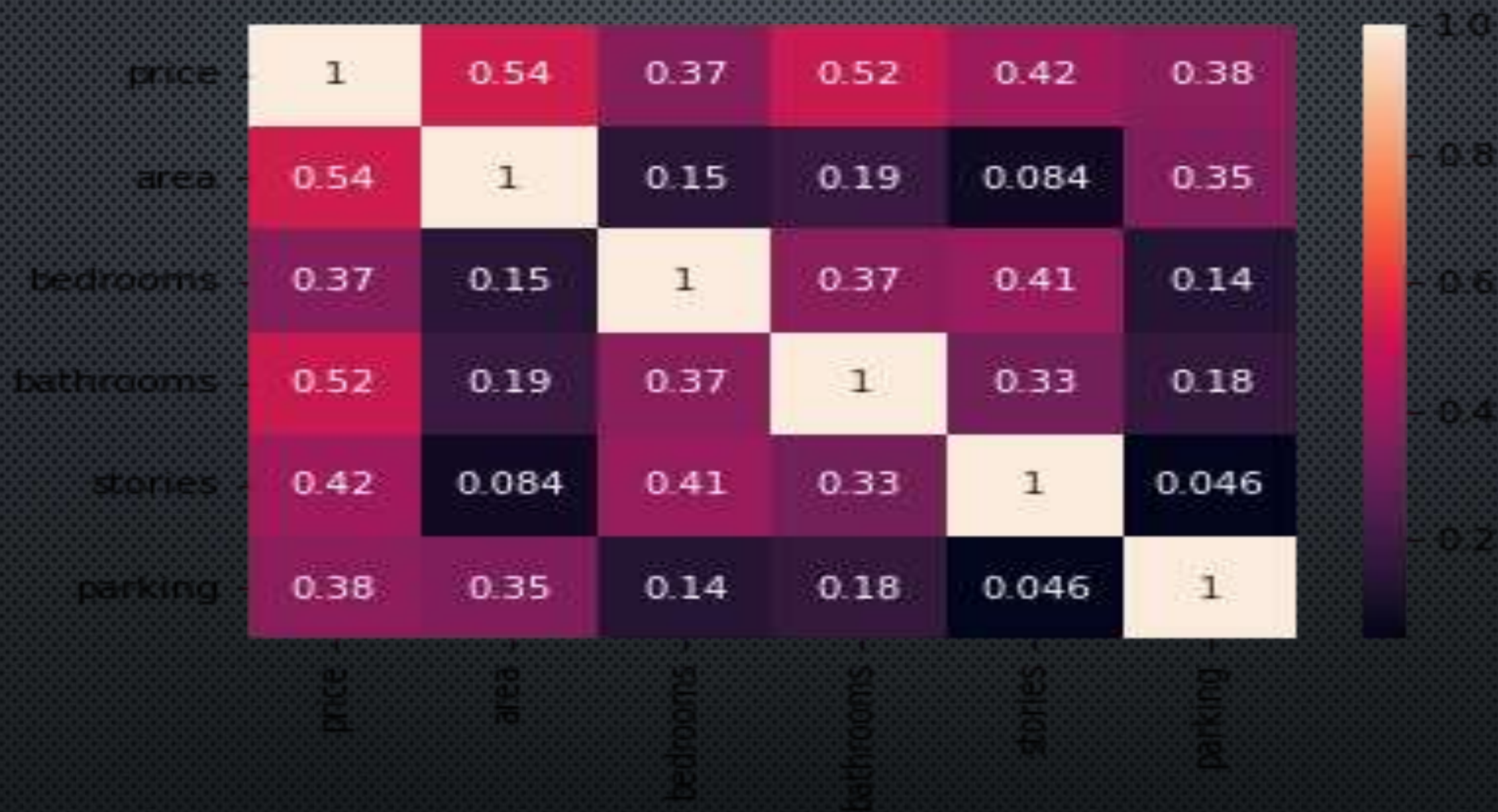
After using these steps in this project I successfully find out housing price.

After using these steps my assumptions that find out correct price of house. I used first step head which shows first five columns for prediction of the house which that a starting step for prediction of house and after that I used info for collecting data this is the second step and after that used tail step which show last five columns because my data is so big and shape step use for find out how much rows and columns in this data and after that use describe step , column step , pair plot step using for a diagram which show that describe step in diagram and correlation step use for searching missing value and heat map used to display the results of eye-tracking tests and sklearn model use for import files for splitting and linear Regression use for estimate value.

Using these steps my assumptions is that company decision is good to invest in Australian market.

Model Dashboard



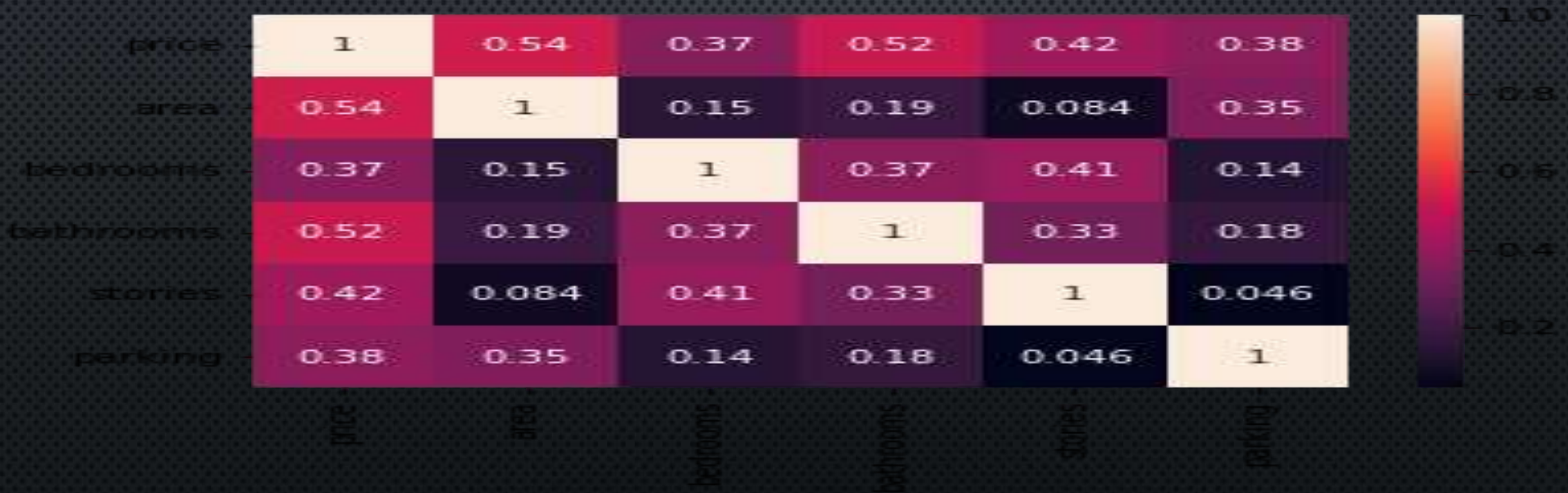


Finalized Model

As per model dashboard :

The Project is completed and the every data show in diagram.

Finalized : I am sharing again a diagram which that show visible clearly.



Conclusion:

Hence that the conclusion of this project that USA based housing company which name is Surprise Housing invest in this project.

Because this investment is good and provide good profit. And after that the company completing his project and sale the house in Australian market in very reasonable price which that the under budget.

Every person wants a good house in a reasonable price and this project is to good for Suuprise Housing Company.