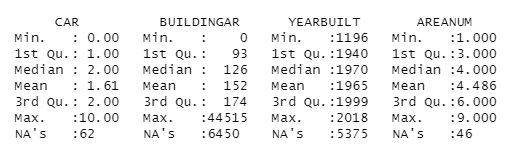
## Data Preprocessing

### Dealing with NULL value

Check NA’ value



Estimate NA value

文本

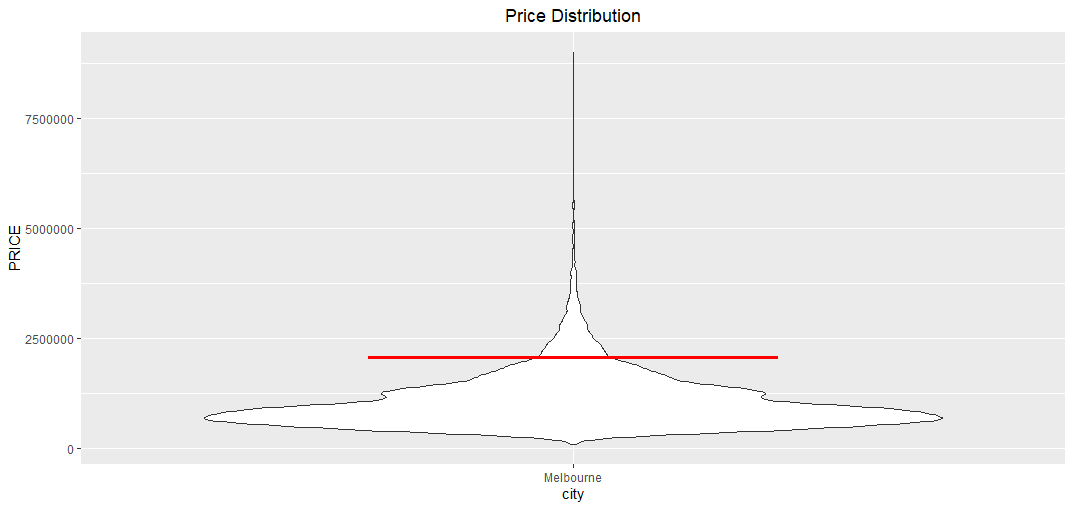
描述已自动生成

### Make level of price

Using the vilint plot could easily see the distribution of the price, we set

split = mean + (mean-min)

as the split point for the price, set greater than the split as high price, otherwise as low price.

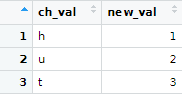


### Remove obvious unrelated features.

Because we have split all the areas into 9 regions as the target for analyzing, so the original features related with region should be removed from the dataset, for this part we do not analyze the time serious, so the time series features should be removed also, finally, the features of SUBURB, ADDRESS, SELLERG, DATE, POSTCODE, COUNCILARE, REGIONNAME should be removed firstly.

### Convert Characters into Numeric

Several algorithms could only deal with numeric values, so all the categorical variables should be converted into numerical features. The features include TYPE, METHOD. The values will be converted as the table below

 表格

描述已自动生成

## Bayesian Network

### literature review regarding the model in context of the background.

A Bayesian network is a graphical representation of probabilistic relationships among a set of variables. are a powerful tool for modeling and reasoning about uncertainty and probabilistic relationships in complex systems. They provide a structured and intuitive way to represent and analyze dependencies between variables, making them a valuable asset in various domains.

### how the model was implemented and its technical result

#### Bayesian Network

图片包含 图表

描述已自动生成图示

描述已自动生成

Parents and children表格

低可信度描述已自动生成

#### Naïve Bayes Classification

Here will use Bayesian networks to analyze the features, then to compare the performance between all features and parents and children features.

Using all features

文本

描述已自动生成

Using parents and children features

文本

描述已自动生成

### conclusion for the model where we interpret the model.

## Classification Model

### Feature selection:

#### literature review regarding the model in context of the background.

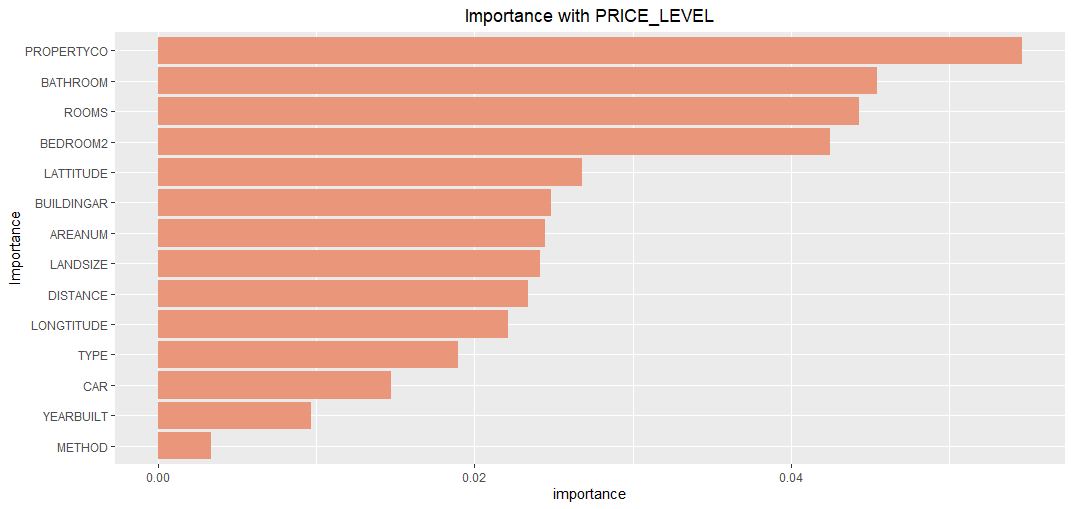
Feature selection is a crucial step in building a classification model for several important reasons:

1. Dimensionality Reduction
2. Improved Model Performance
3. Overfitting Prevention
4. Faster Training and Inference:
5. Interpretability

#### how the model was implemented and it's technical results

1. Importance between input variables and PRICE\_LEVEL

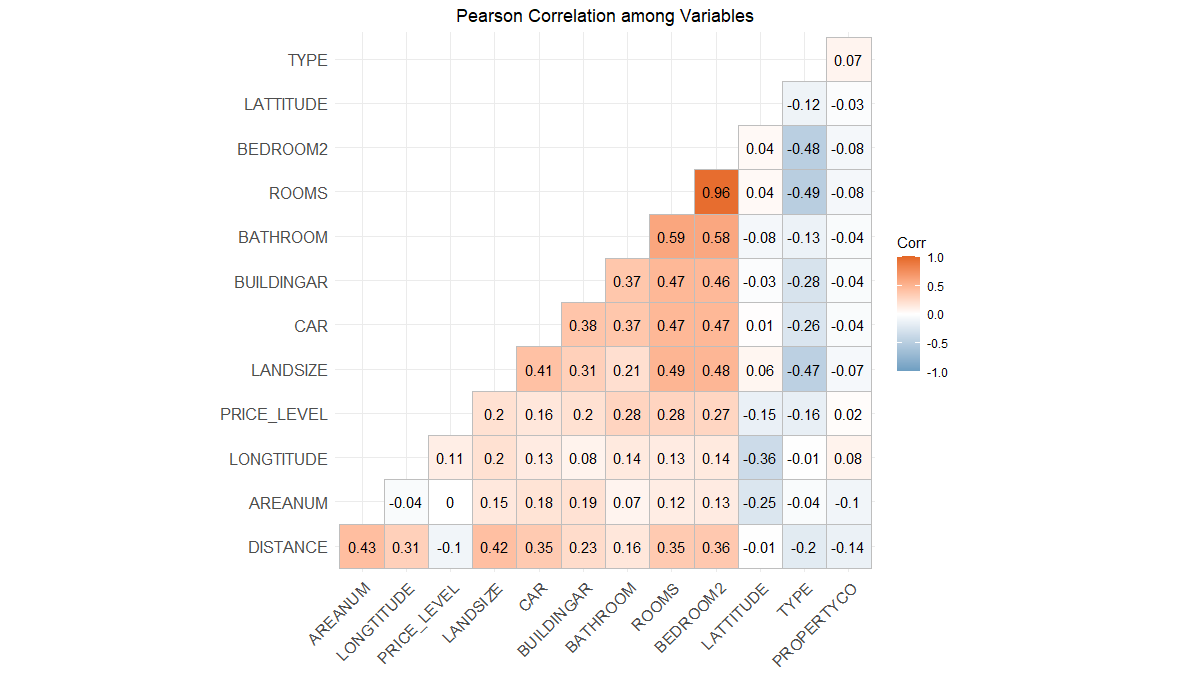
Using the gain information to calculate the importance between input variables with PRICE\_LEVEL, the result like the picture below.



According to the gain importance, the variables PROPERTYCO, METHOD almost have no relationships with PRICE\_LEVEL, so we should deselect the two variables, all other variables will be selected as the input variables.

1. Correlations between input variables.

Because of all the variables has been converted to number type. So here we should select



According to the Pearson relationship among variables, the ROOMS and BEDROOMS2 has highest correlation, so we should deselect one of them, for this case, we should deselect BEDROOMS2. According to the results above, all the input variables are TYPE, METHOD, LATTITUDE, ROOMS, BATHROOM, BUILDINGAR, CAR, LANDSIZE, LONGTITUDE, AREANUM, DISTANCE.

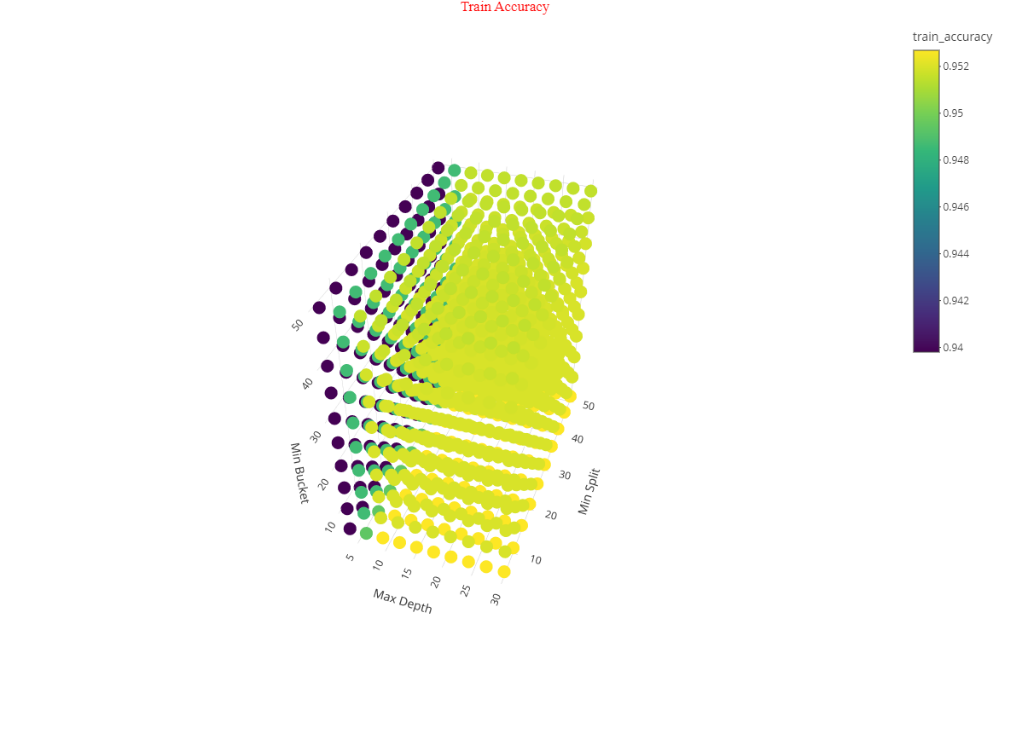
### Decision Tree

#### literature review regarding the model in context of the background.

A decision tree is a popular and widely used machine learning algorithm that is primarily used for classification and regression tasks. It's a tree-like structure that breaks down a complex decision-making process into a series of simpler decisions. Decision trees are simple to understand and interpret, making them a valuable tool for both beginners and experts in machine learning.

#### how the model was implemented and its technical results

using different



图片包含 表格

描述已自动生成

Maxdepth = 9, minsplit = 5, minbucket = 5 as the final model parameters.

#### conclusion for the model where we interpret the model.

### Neural Networks

#### literature review regarding the model in context of the background.

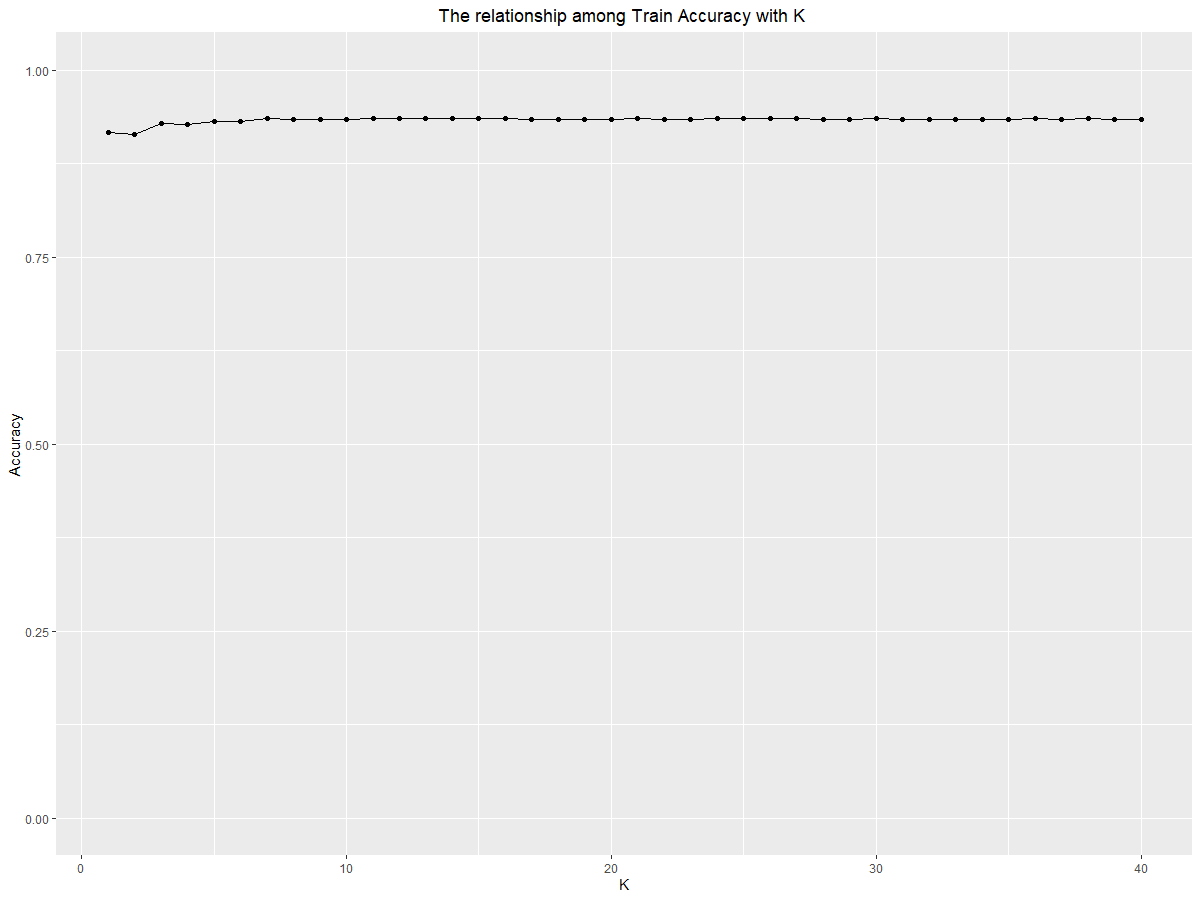
#### how the model was implemented and it's technical results

#### conclusion for the model where we interpret the model.

### K-Nearest Neighbors

#### literature review regarding the model in context of the background.

#### how the model was implemented and it's technical results



Test data

图片包含 图形用户界面

描述已自动生成

Train data

图片包含 图形用户界面

描述已自动生成

#### conclusion for the model where we interpret the model.