THE UNIVERSITY OF WESTERN AUSTRALIA

Predicting Housing Prices in Perth

Presented by
Bharat Guturi

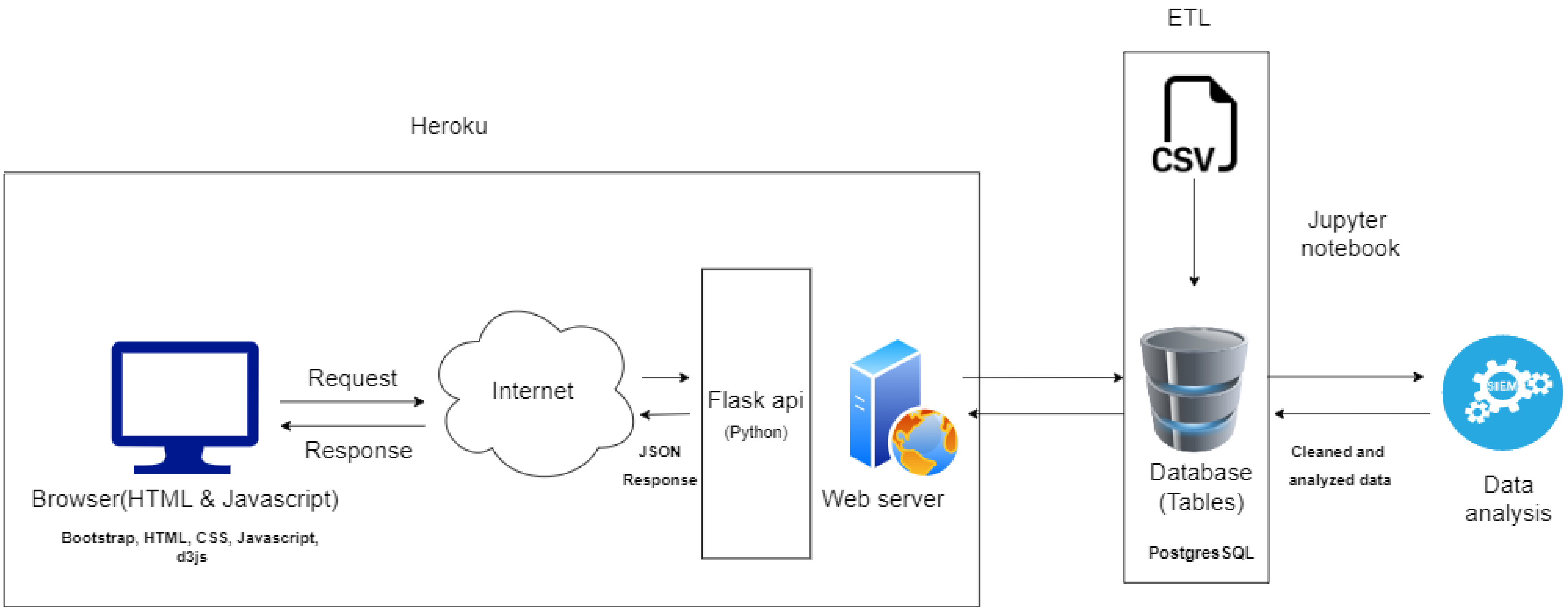
Project Overview and Source of Data

For someone who is contemplating to buy a house, this Project aims to predict the house prices in Perth based on the available historical data available.

The Data is sourced from:

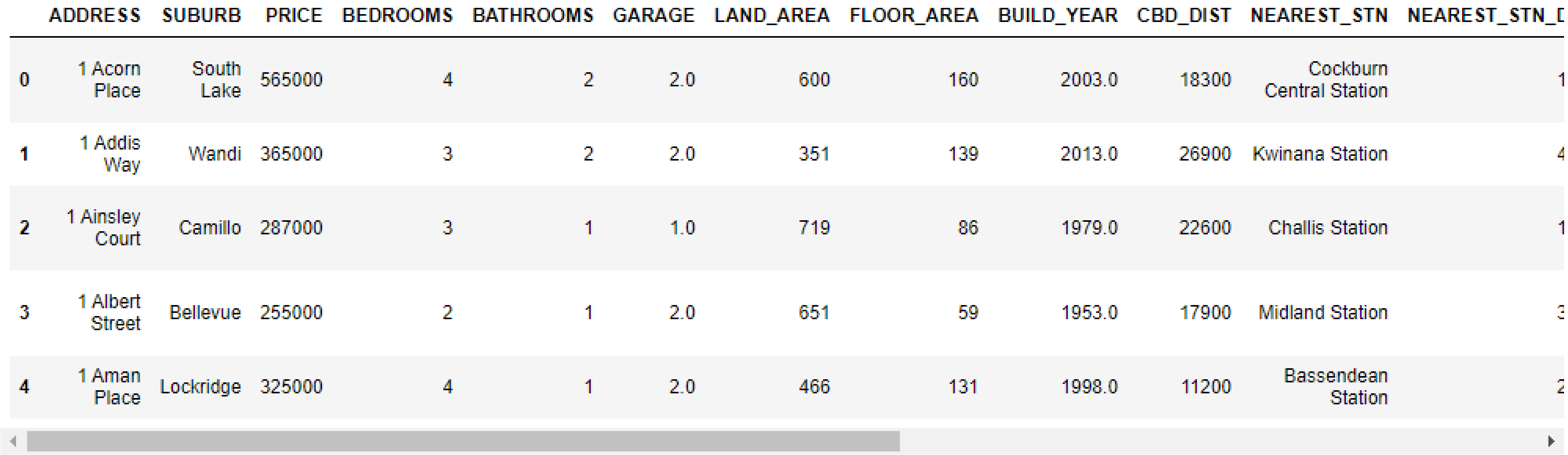
https://www.kaggle.com/datasets/syuzai/perth-house-prices

Architecture diagram



Extract, Transform & load

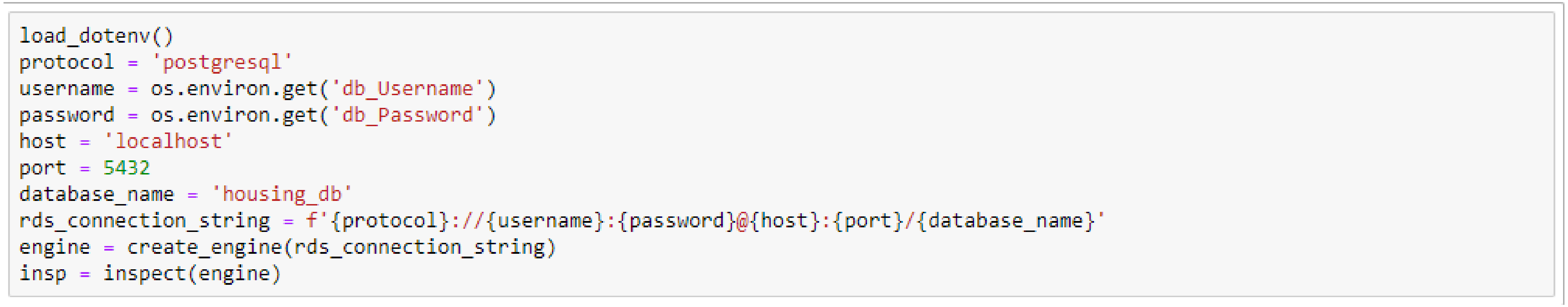
1. Importing Dependencies
2. Storing CSV into Dataframe



	ADDRESS	SUBURB	PRICE	BEDROOMS	BATHROOMS	GARAGE	LAND_AREA	FLOOR_AREA	BUILD_YEAR	CBD_DIST	NEAREST_STN	NEAREST_STN_
0	1 Acorn Place	South Lake	565000	4	2	2.0	600	160	2003.0	18300	Cockburn Central Station	1
1	1 Addis Way	Wandi	365000	3	2	2.0	351	139	2013.0	26900	Kwinana Station	4
2	1 Ainsley Court	Camillo	287000	3	1	1.0	719	86	1979.0	22600	Challis Station	1
3	1 Albert Street	Bellevue	255000	2	1	2.0	651	59	1953.0	17900	Midland Station	3
4	1 Aman Place	Lockridge	325000	4	1	2.0	466	131	1998.0	11200	Bassendean Station	2

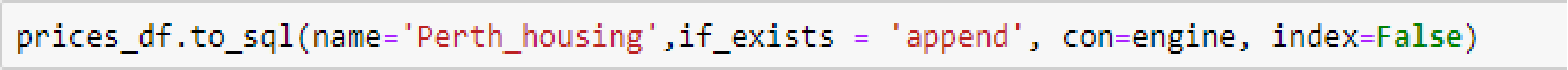
THE UNIVERSITY OF WESTERN AUSTRALIA

3. Connecting to local database



```
load_dotenv()
protocol = 'postgresql'
username = os.environ.get('db_Username')
password = os.environ.get('db_Password')
host = 'localhost'
port = 5432
database_name = 'housing_db'
rds_connection_string = f'{protocol}://{username}:{password}@{host}:{port}/{database_name}'
engine = create_engine(rds_connection_string)
insp = inspect(engine)
```

4. Loading csv data using dataframe to SQL

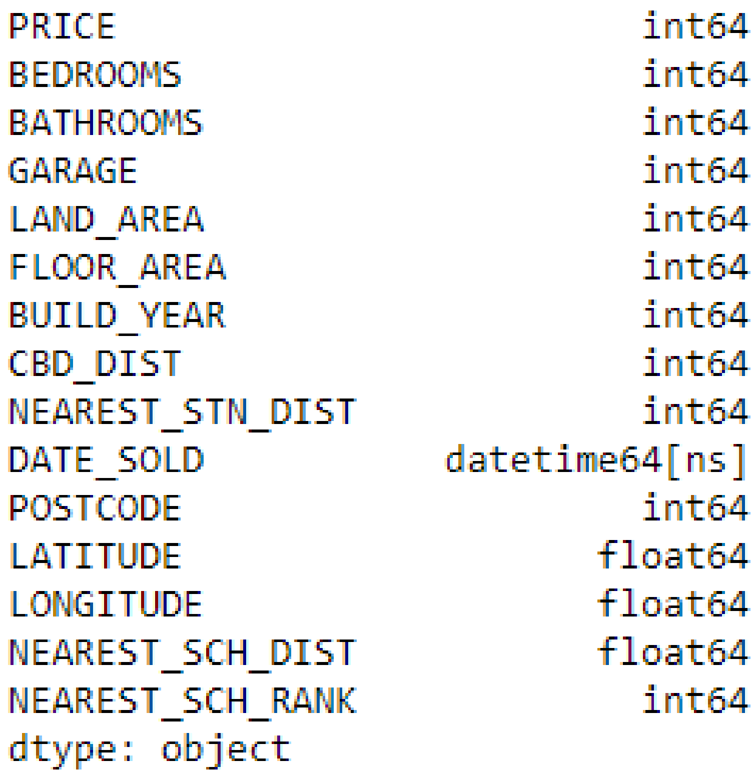


```
prices_df.to_sql(name='Perth_housing',if_exists = 'append', con=engine, index=False)
```

Data Analysis

Data Cleaning:

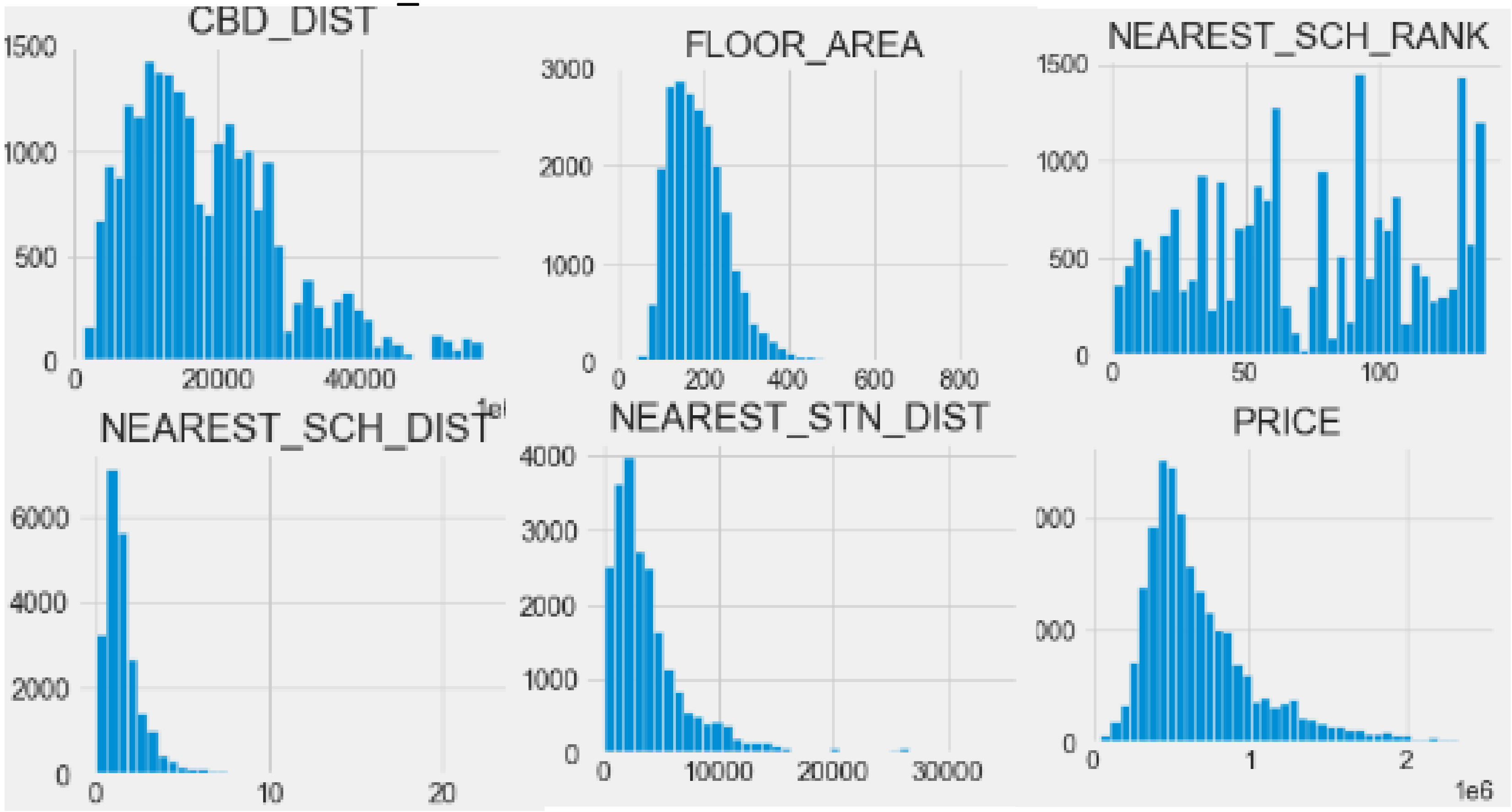
Data Columns and datatypes



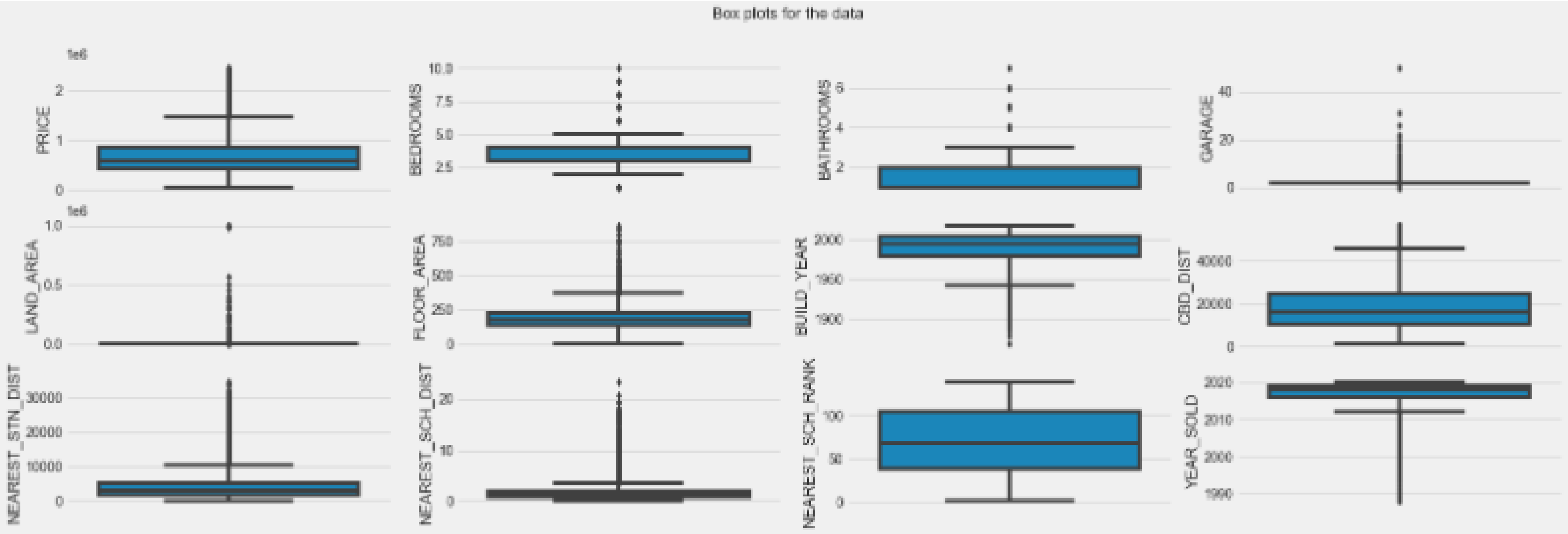
```
PRICE                       int64
BEDROOMS                    int64
BATHROOMS                   int64
GARAGE                      int64
LAND_AREA                   int64
FLOOR_AREA                  int64
BUILD_YEAR                  int64
CBD_DIST                    int64
NEAREST_STN_DIST            int64
DATE_SOLD          datetime64[ns]
POSTCODE                    int64
LATITUDE                  float64
LONGITUDE                 float64
NEAREST_SCH_DIST          float64
NEAREST_SCH_RANK            int64
dtype: object
```

Steps:

1. Removal of rows containing null values in ‘Nearest_SCH_RANK’
2. Changing the ‘DATE_SOLD’ column to ‘MONTH_SOLD’ and ‘YEAR_SOLD’.

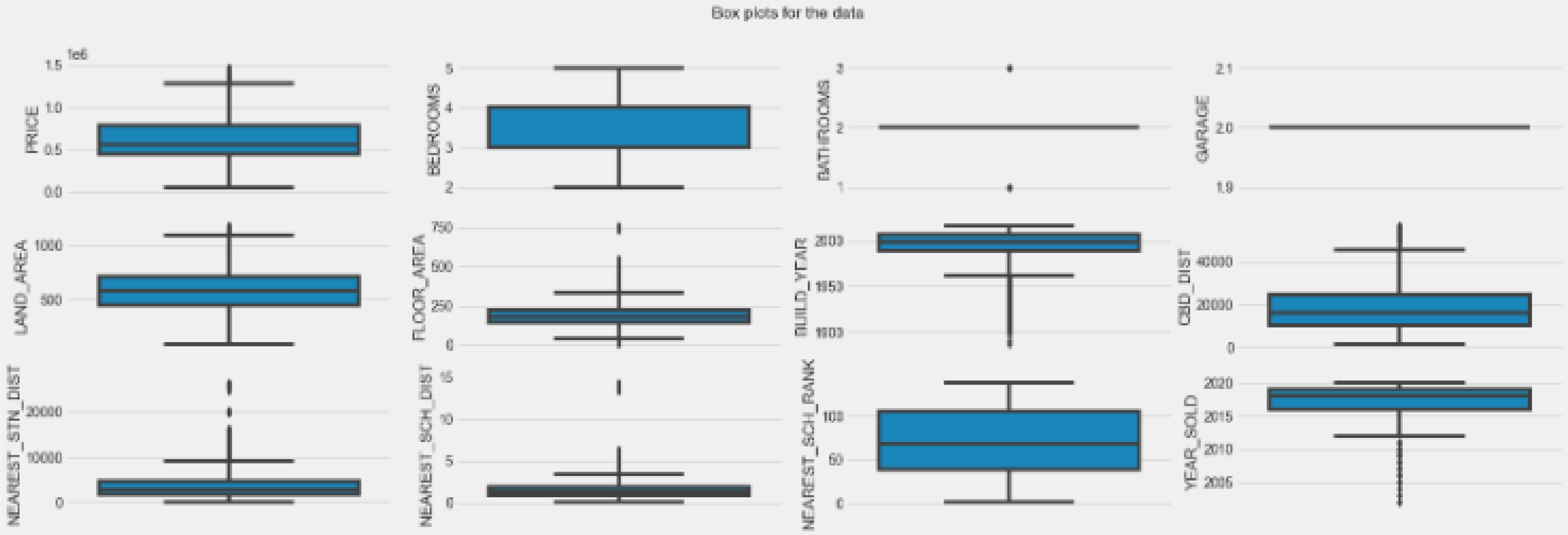


Box Plots

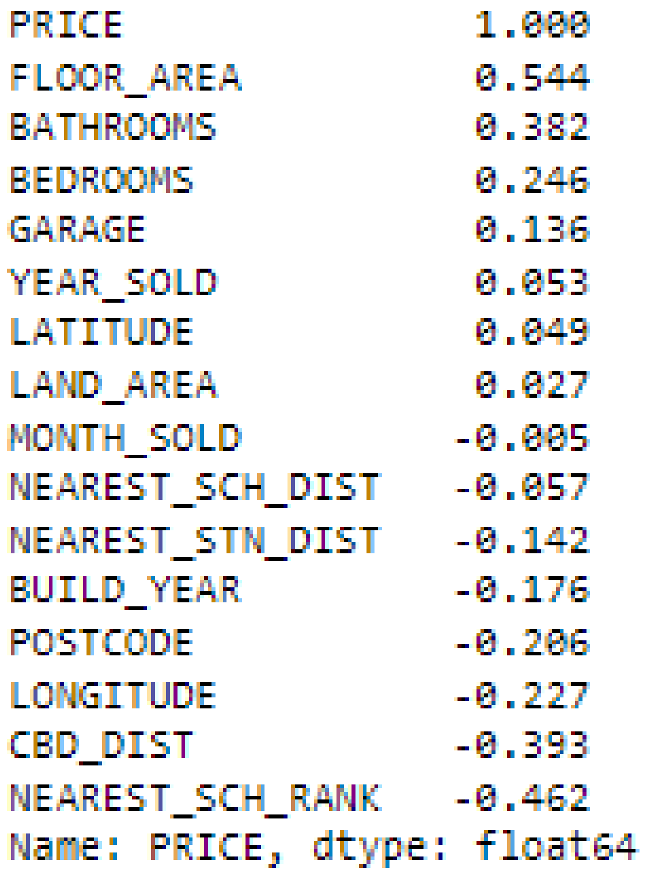


Box plots for the data
1e6
PRICE
0
1
2
BEDROOMS
2.5
5.0
7.5
10.0
BATHROOMS
2
4
6
GARAGE
0
20
40
1e6
LAND_AREA
0.0
0.5
1.0
FLOOR_AREA
0
250
500
750
BUILD_YEAR
1900
1950
2000
CBD_DIST
0
20000
40000
NEAREST_STN_DIST
0
10000
20000
30000
NEAREST_SCH_DIST
0
10
20
NEAREST_SCH_RANK
0
50
100
YEAR_SOLD
1990
2000
2010
2020

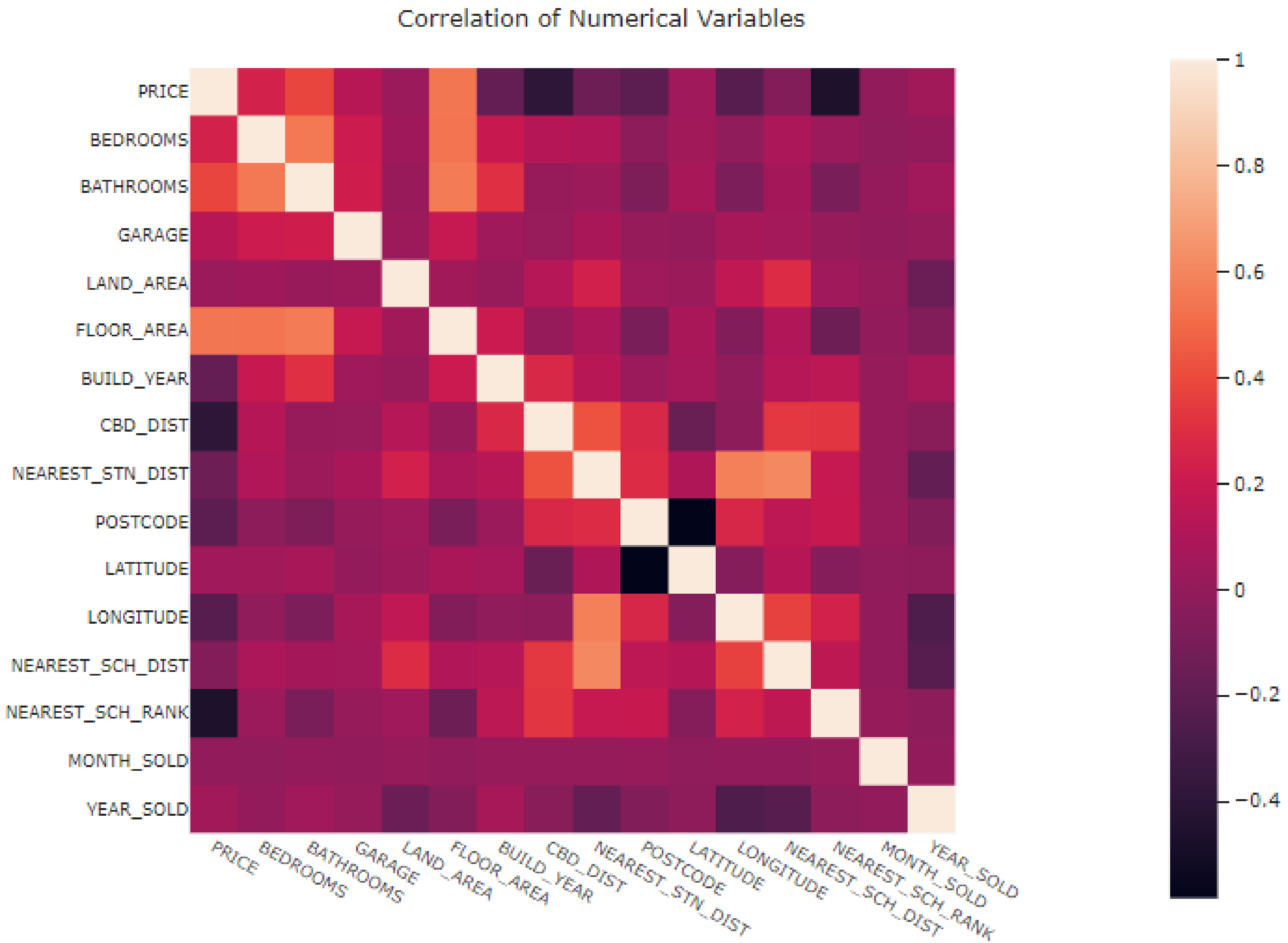
Filtered Data



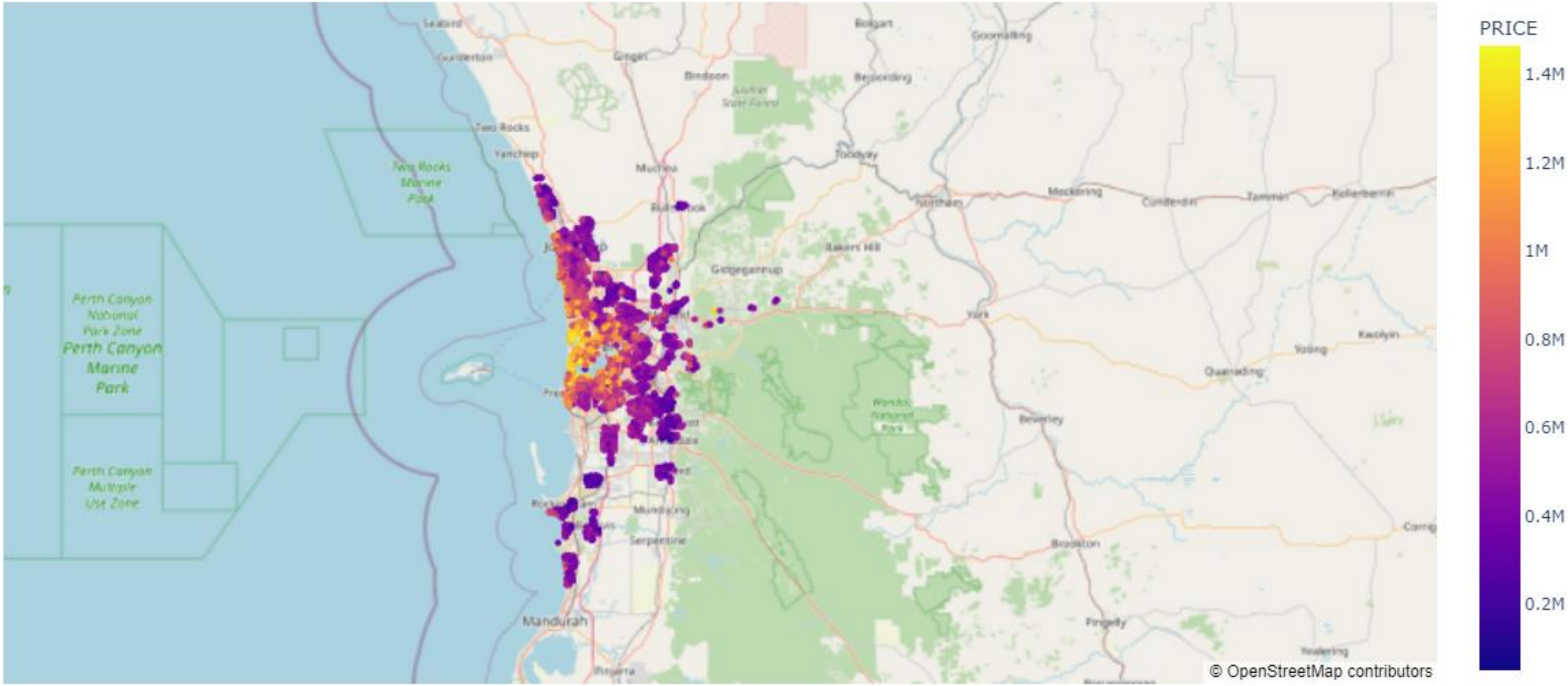
Correlation Matrix



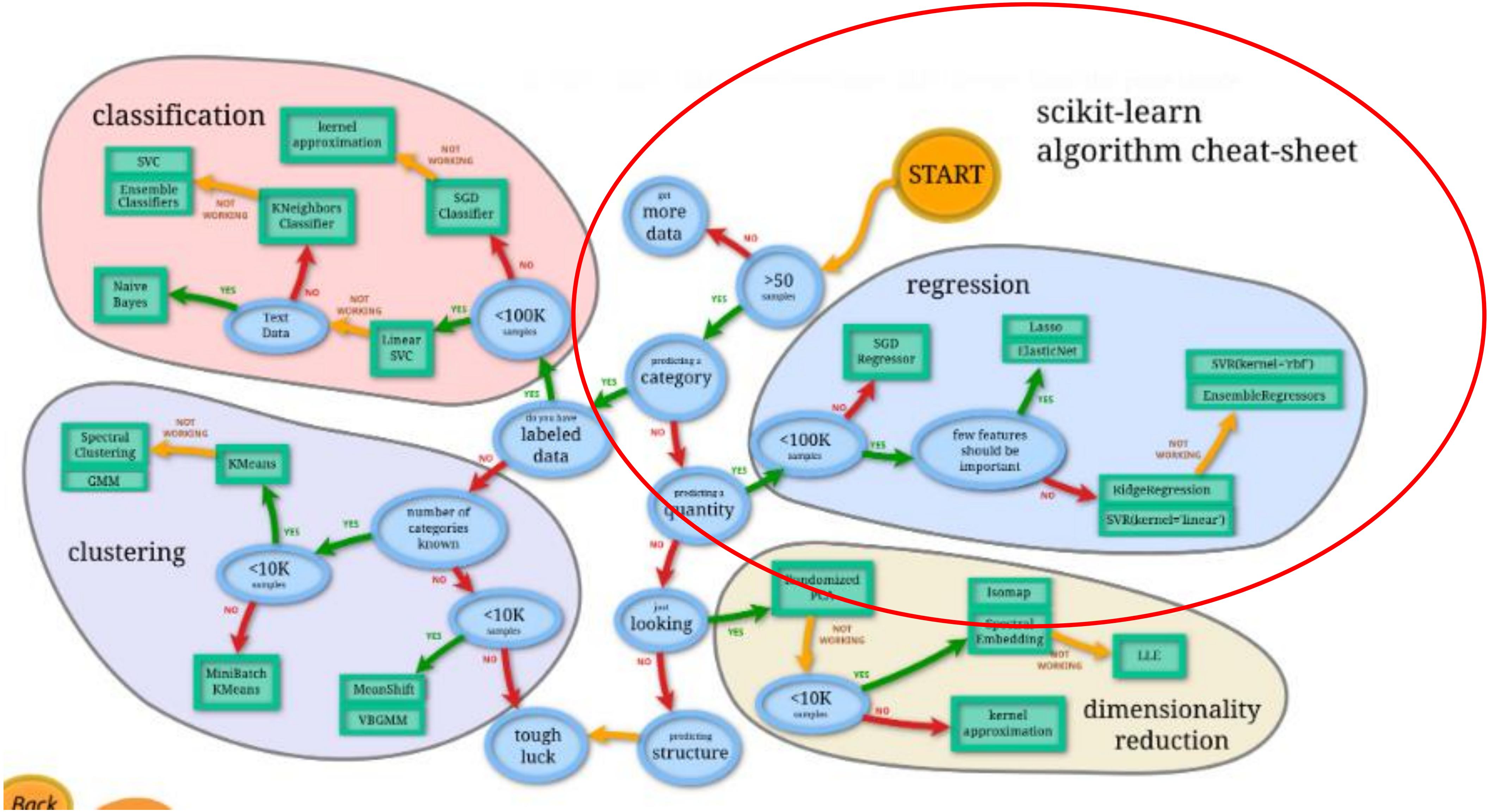
```
PRICE                 1.000
FLOOR_AREA            0.544
BATHROOMS             0.382
BEDROOMS              0.246
GARAGE                0.136
YEAR_SOLD             0.053
LATITUDE              0.049
LAND_AREA             0.027
MONTH_SOLD           -0.005
NEAREST_SCH_DIST     -0.057
NEAREST_STN_DIST     -0.142
BUILD_YEAR           -0.176
POSTCODE             -0.206
LONGITUDE            -0.227
CBD_DIST             -0.393
NEAREST_SCH_RANK     -0.462
Name: PRICE, dtype: float64
```



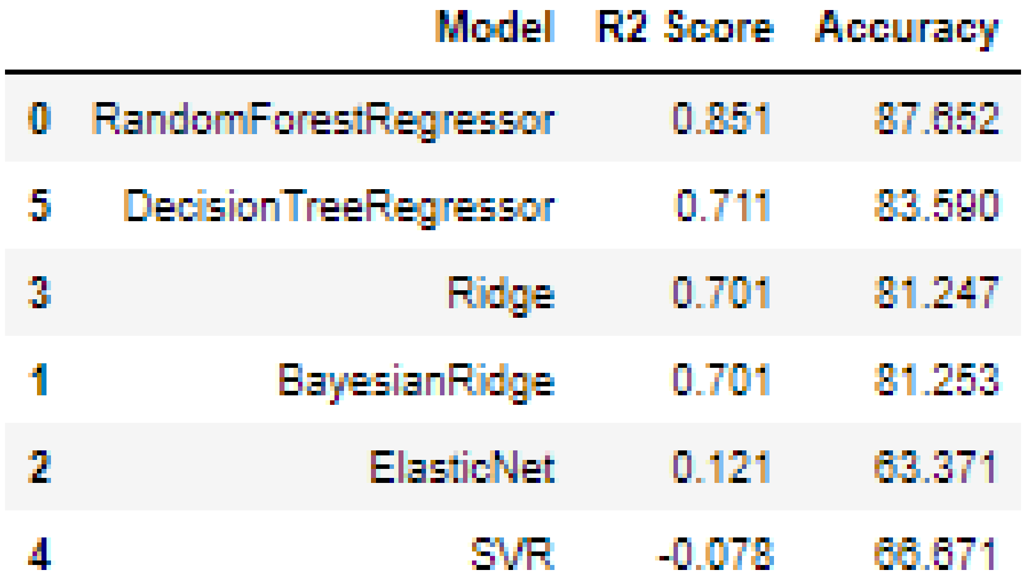
Visualisation of Filtered Data as per the Price



Choosing the right Algorithm

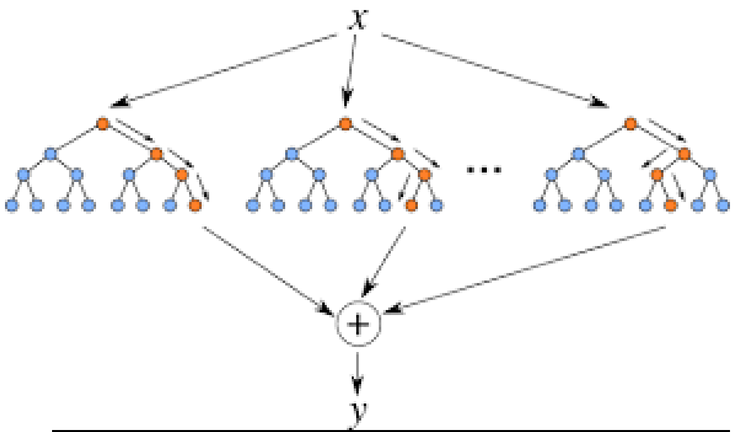


Resulting Accuracy of various Models

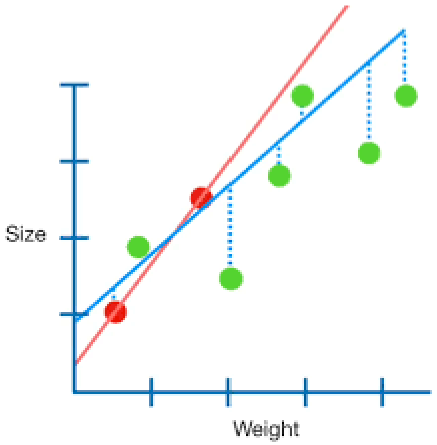


	Model	R2 Score	Accuracy
0	RandomForestRegressor	0.851	87.652
5	DecisionTreeRegressor	0.711	83.590
3	Ridge	0.701	81.247
1	BayesianRidge	0.701	81.253
2	ElasticNet	0.121	63.371
4	SVR	-0.078	66.671

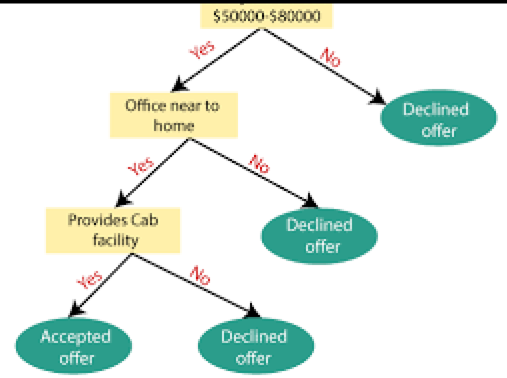
Random Forest Regressor



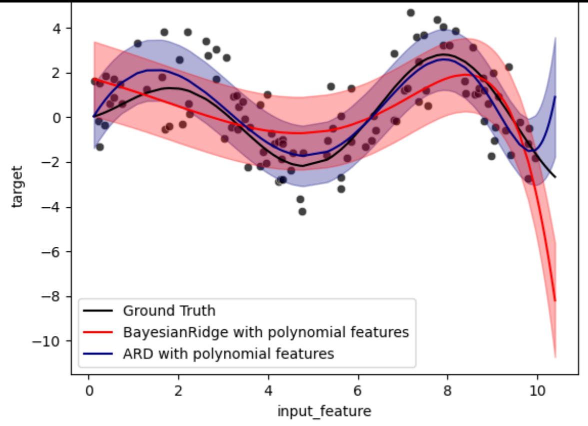
Ridge Regressor



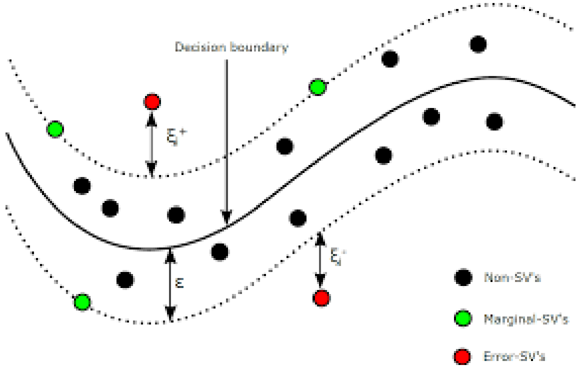
Decision Tree Regressor



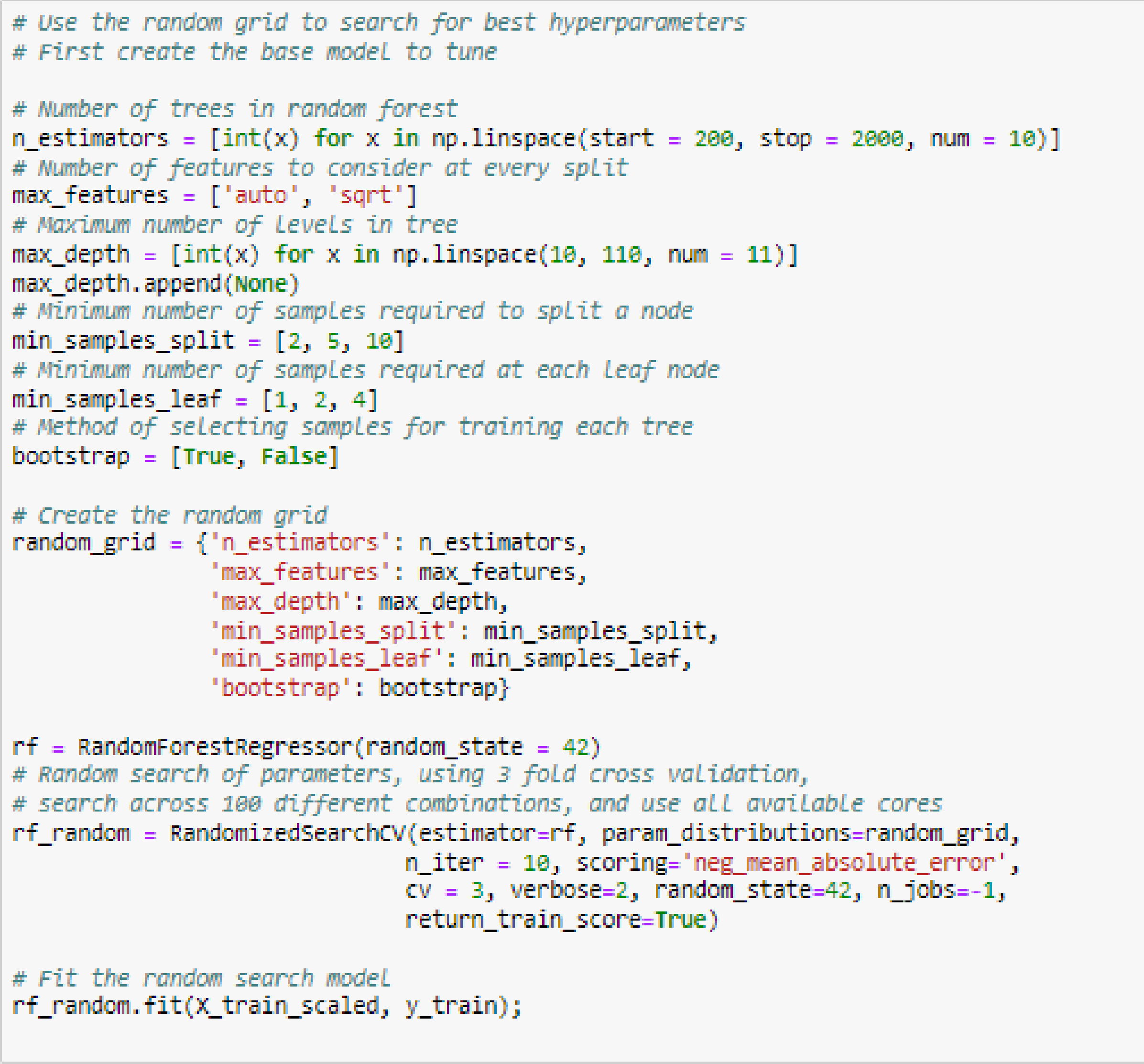
Bayesian Ridge Regressor



SVR

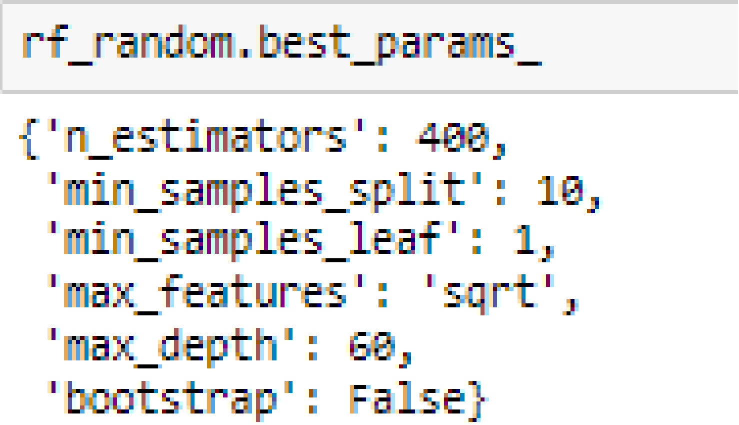


Hyperparameter Tuning



```
# Use the random grid to search for best hyperparameters
# First create the base model to tune

# Number of trees in random forest
n_estimators = [int(x) for x in np.linspace(start = 200, stop = 2000, num = 10)]
# Number of features to consider at every split
max_features = ['auto', 'sqrt']
# Maximum number of levels in tree
max_depth = [int(x) for x in np.linspace(10, 110, num = 11)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [2, 5, 10]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Method of selecting samples for training each tree
bootstrap = [True, False]

# Create the random grid
random_grid = {'n_estimators': n_estimators,
               'max_features': max_features,
               'max_depth': max_depth,
               'min_samples_split': min_samples_split,
               'min_samples_leaf': min_samples_leaf,
               'bootstrap': bootstrap}

rf = RandomForestRegressor(random_state = 42)
# Random search of parameters, using 3 fold cross validation,
# search across 100 different combinations, and use all available cores
rf_random = RandomizedSearchCV(estimator=rf, param_distributions=random_grid,
                              n_iter = 10, scoring='neg_mean_absolute_error',
                              cv = 3, verbose=2, random_state=42, n_jobs=-1,
                              return_train_score=True)

# Fit the random search model
rf_random.fit(X_train_scaled, y_train);
```

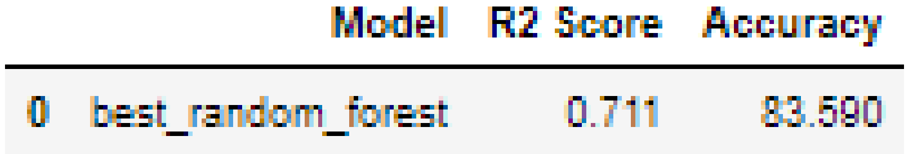
Best Parameters



```
rf_random.best_params_
```

```
{'n_estimators': 400,
 'min_samples_split': 10,
 'min_samples_leaf': 1,
 'max_features': 'sqrt',
 'max_depth': 60,
 'bootstrap': False}
```

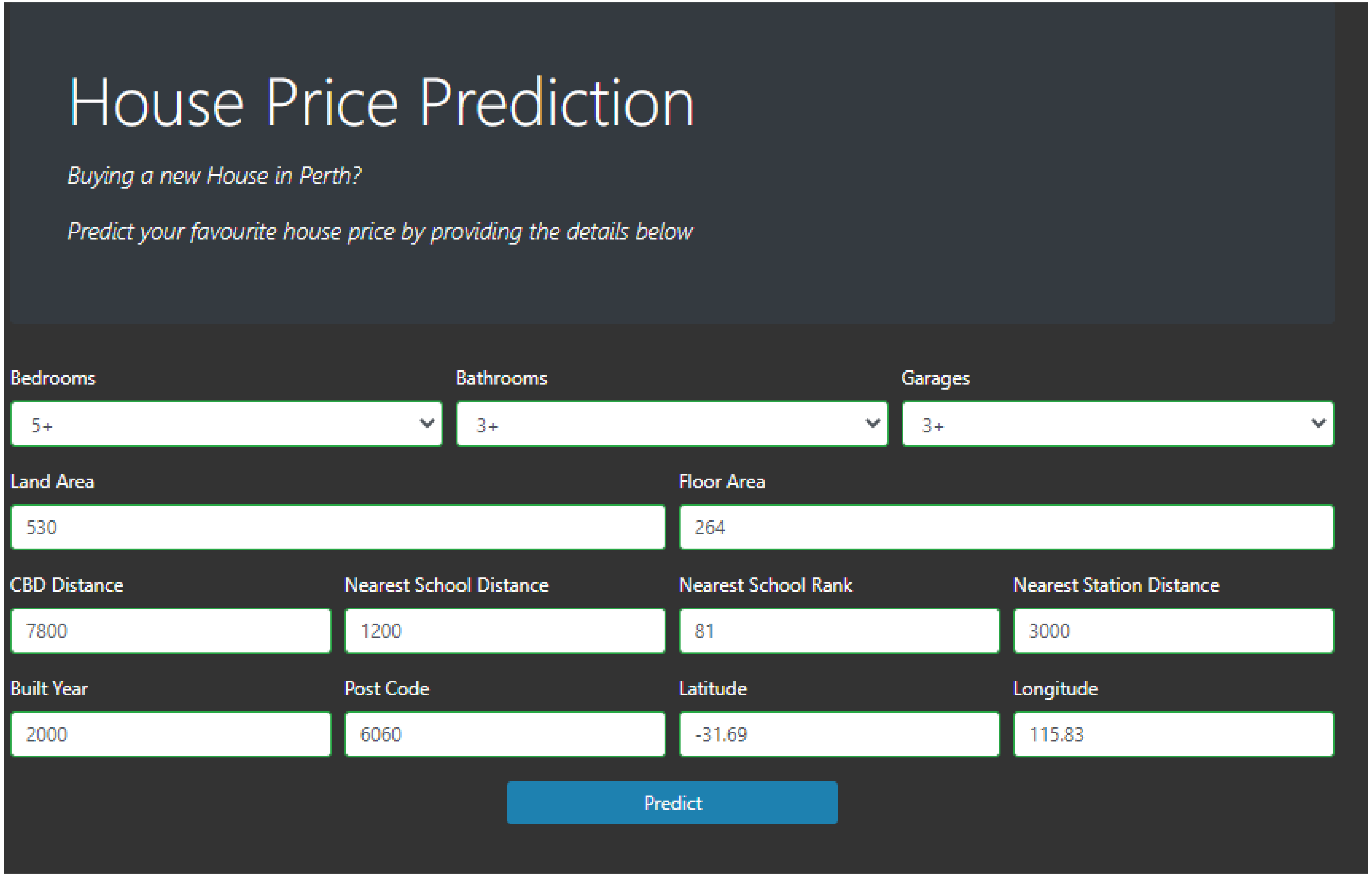
Output



	Model	R2 Score	Accuracy
0	best_random_forest	0.711	83.590

THE UNIVERSITY OF WESTERN AUSTRALIA

Web Application



House Price Prediction

Buying a new House in Perth?

Predict your favourite house price by providing the details below

Bedrooms: 5+

Bathrooms: 3+

Garages: 3+

Land Area: 530

Floor Area: 264

CBD Distance: 7800

Nearest School Distance: 1200

Nearest School Rank: 81

Nearest Station Distance: 3000

Built Year: 2000

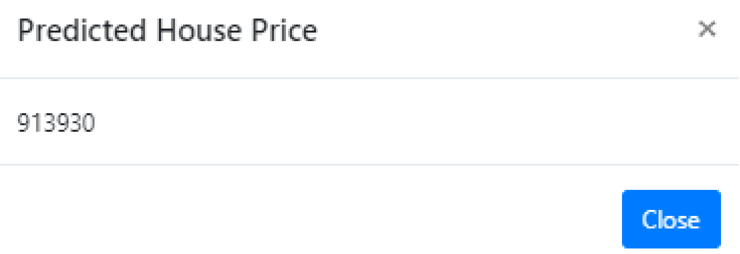
Post Code: 6060

Latitude: -31.69

Longitude: 115.83

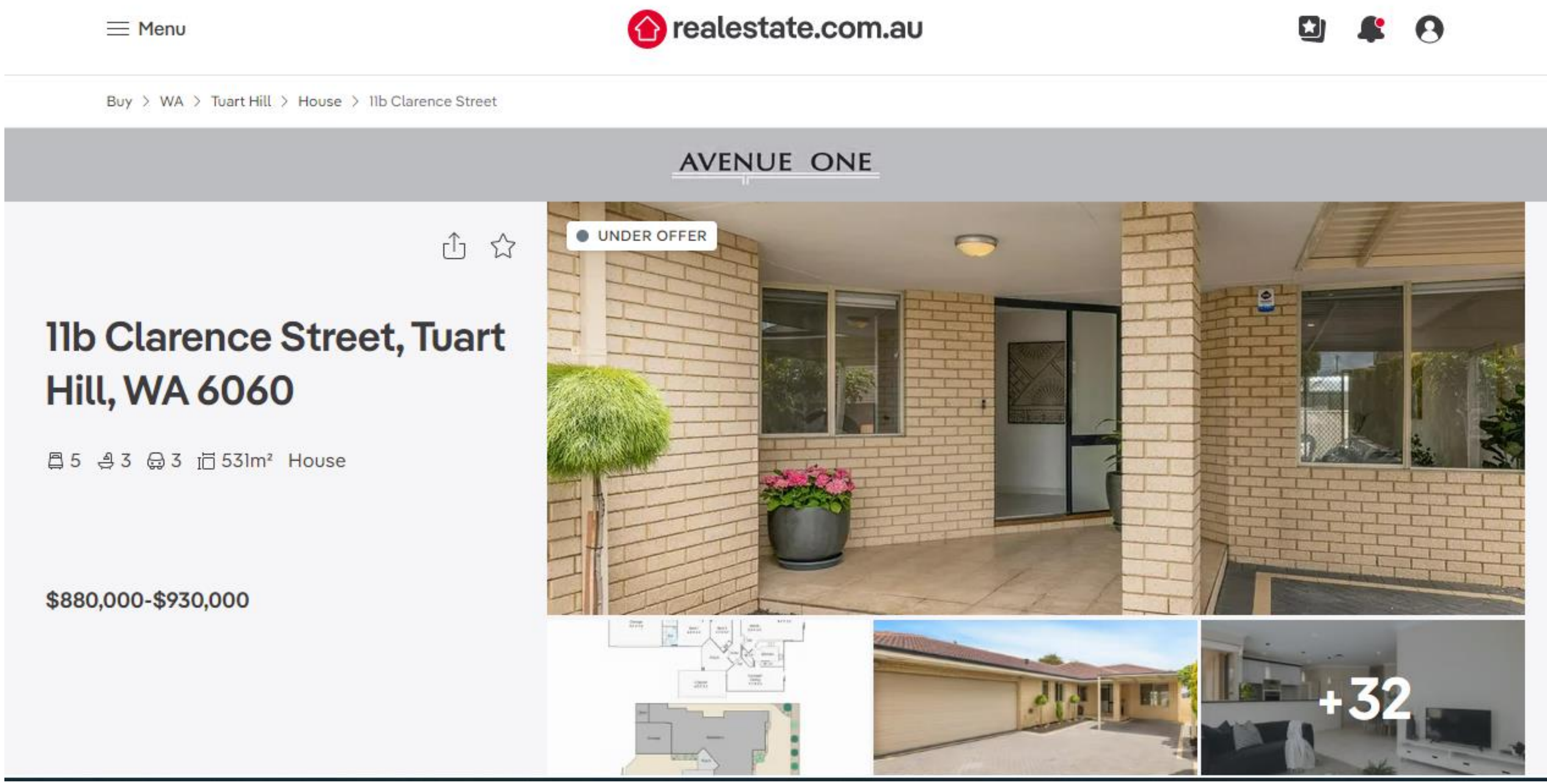
Testing

Case Study: https://www.realestate.com.au/property-house-wa-tuart+hill-140995332



Predicted House Price ×

913930



11b Clarence Street, Tuart Hill, WA 6060

5 3 3 531m² House

$880,000-$930,000

Thank you