

**HOUSING: PRICE PREDICTION**

Submitted by:

VIVEK KUMAR SINGH

**ACKNOWLEDGMENT**

I take great pleasure to thank and acknowledge the help provided by **Flip Robo Technologies**. I extend whole hearted thanks to Mr. Shwetank Mishra who become my Mentor and with whom I worked and learned a lot and for enlightening me with her knowledge and experience to grow with the corporate working. Her guidance at every stage of the Project enabled me to successfully complete this Project which otherwise would not have been possible without his consent encouragement and motivation. Without the support it was not possible for me to complete the report with fullest endeavour.

**INTRODUCTION**

* Business Problem Framing

Houses are one of the necessary need of each and every person around the globe and therefore housing and real estate

market is one of the markets which is one of the major contributors in the world’s economy. It is a very large market

and there are various companies working in the domain. Data science comes as a very important tool to solve problems

in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and

focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling,

recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies.

A US-based housing company named **Surprise Housing** 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.

Review of literature

The academic and empirical literature on modelling and explaininghouse prices covers a wide range of issues. One strand in the empiricalliterature looks at house prices from the perspective of individual

characteristics (size, number of bedrooms, presence or absence of a garage,geographical characteristics such as accessibility, distance to the center or availability of amenities). However, this approach is mainly used to build (quality adjusted) indexes of average house prices in specific markets and areas/cities/countries and not to explain or forecast house prices in relation to

macroeconomic developments. By contrast, this paper aims to explain the observed changes in residential property prices over time, describe their

relationship to the factors of housing demand and supply, and employ these relationships for forecasting purposes.

2. The long-term fundamentals of house prices are mostly chosen on the basis of a demand equation. Many models for house prices assume that the supply of housing is relatively inelastic in the short to medium-run and that it is hence essentially the changes in demand that explain variations in house prices. When housing is treated as a consumption good, its demand is generally a function of a number of variables such as household income,

interest rates, financial wealth, or demographic and labour market factors. Housing can also be seen as an investment good providing a flow of services. This stream of the literature explains the evolution of house prices

as a function of financial variables such as the risk-free rate or alternative investment returns, the housing risk premium and the flow of housing services generally proxied by the rental yield of a property (similar to the dividend yield commonly used in the financial literature). Such models include Aoki, Proudman and Vlieghe (2004) and Piazzesi, Schneider and Tuzel (2007).

Studies examining the performance of credit for house price forecasting point to mixed results. Access to credit and financial conditions are additional

factors influencing house price dynamics. Credit is found to be a very important variable with Favilukis et al. (2012): survey measures of credit

supply explain 53% of the quarterly variation in house price growth in the US (1992-2010). Generally, credit availability is found to have a strong positive

effect on house prices as in Igan and Loungali (2012), Goodhart and Hofmann (2008), Annett (2005) and Tsatsaronis and Zhu (2004). The relationship is not

straightforward though. For example, Annett (2005) found that although credit is statistically insignificant in the short and medium run, it remains one of the

main determinants of house prices in the long-run. Goodhart and Hofmann (2008) and Simigiannis and Hondroyiannis (2009) indicate there is multi-

directional causality between house prices and credit (highlighting the bank credit and the channels). There is some evidence that the impact of private

credit on house prices is stronger in a boom. Gerdesmeier et al. (2011) found that credit and money aggregates lead developments in house prices and that demographic variables, the unemployment rate, disposable income and the debt-to-income ratio affect house prices asymmetrically across the house price cycle, more strongly during booms and less so during busts. Mian and

Sufi (2011) show for the US that the 2002-2005 period is the only period in which one may observe a strongly positive house price growth rate and a

negative income growth rate and explain this with the expansion of subprime mortgage securitization.

Some contributions in the literature also consider supply side factors as determinant of house prices. Supply-side factors useful in forecasting house

prices are primarily real construction costs and construction technology shocks. Building permits are a useful forecasting variable of both construction

volumes and prices as shown in Strauss (2012). Theoretical models like Spiegel (2001) have shown that credit constraints can lead to construction

cycles with developers acquiring land when it is cheap relative to homes in good condition, and developing and selling their land when it becomes expensive. As a result, developer holdings should forecast future expected

returns, and unusually large positive shocks to house prices will lead to abnormally high levels of construction. On the other hand, credit can also

affect the supply side and the housing stock, as banks do not only play a crucial role in the financing of the demand for existing real estate, but also in the lending to developers and the construction sector for the purchase of land and the construction of new buildings.

Changes in structural characteristics can lead to shifts in the demand/supply balance governing house price dynamics. Institutional factors such as access to the housing market, liquidity of the secondary mortgage

market, collateral and bankruptcy legislation (such as legislation regarding

foreclosures) and the fiscal treatment of owning a house as against renting (mortgage interest deductibility,imputing of owner-occupied rents, property

and wealth taxes) are sometimes included in demand equations and could work as demand shifters if they change. Most of these institutional factors are country-specific and vary infrequently over history so most of the time they are introduced in the forecasting equations as dummy variables (as in Tsatsaronis and Zhu, 2004 who discuss the impact of national differences in the mortgage

market on house prices).

3. The short-term momentum of house prices

The empirical literature suggests that house prices are generally driven by momentum, i.e. the observed tendency for rising house prices to rise further (and vice versa), at least in the short-term. Past house prices or past changes in house prices are included as explanatory variables in the overwhelming majority of studies, being also highly significant (across

different time frames or geographies). The results indicate momentum in the short-run (with positive autocorrelation for the first lags in quarterly data) and reversal in the longer run (with negative coefficients for periods above five years). This autocorrelation structure could be seen as a purely empirical necessity because fundamental variables alone are typically not enough to explain house prices, but there are several theoretical models to explain momentum: irrational exuberance and unrealistic expectations of future price appreciation (Shiller 2005, 2009), risk-shifting behavior by banks related to

agency problems and their expectations of continued credit growth (Allen and Gale, 2000), a procyclical behavior of housing sales (Wheaton, 1990), or

down payment constraints in sellers’ reservation prices (Stein, 1995). At thesame time, the autocorrelation structure is typically found to be market

specific and to differ across countries.

The empirical models

Empirical models generally adopt a dynamic approach, linking the evolution of house prices to pre-selected fundamentals and a set of short-run and long-run determinants. A commonly found approach links the level of house prices to a set of short-run and long-run determinants in an Error-Correction Model (ECM) or

Vector Error-Correction Model (VECM). Making use of economic theory for determining long-run equilibria, an ECM explains changes in house prices in terms of changes in explanatory variables and an error-correction term in lagged levels which is interpreted as reflecting long-run disequilibrium responses of house prices. In a VECM all variables are treated as endogenous and multiple cointegration relations can be found. These types of models can be used to determine over/undervaluation of house prices as a deviation to the values implied by the long-run equation, but they can also be used for forecasting purposes by employing the long-run equation as a benchmark and constructing the forecast by assuming reversion towards levels implied by this long-run benchmark. Examples of this approach are

Gattini and Hiebert (2010), Greiber and Setzer (2007) for the euro area, Malpezzi (1999), Painter and Redfearn (2002), McCathy and Peach (2004),

Case et al. (2013) for the US, Apergis and Rezitis (2003) for Greece,

Oikarinen (2012) for Finland and Adams and Füss (2010) for a number of

developed economies. Vector Autoregressive Models (VAR) and Panel VAR (see below) have also been applied in forecasting studies. The focus is

primarily on OECD/industrialized countries as in Tsatsaronis and Zhu (2004),

Assenmacher-Wesche and Gerlach (2008) or Kuethe and Pede (2011).

Bayesian VARs exploit the use of informative priors to shrink the

unrestricted model towards a parsimonious naïve benchmark, therebyreducing parameter uncertainty and improving forecast accuracy; owing to

their enhanced modeling flexibility and better performances when datasets are of limited length they have witnessed a noticeable increase in applications

as in Jarocinski and Smets (2008), Carstensen et al. (2009) and Gupta,

Kabundi and Miller (2009). In many cases, a Bayesian VAR is used to estimate a fully-fledged structural model (most of the time a DSGE specification) as in Gupta et al. (2009).

Panel models are often employed to account either for different regional/urban dynamics or national heterogeneities but they are challenged

by data limitations. This class of models include Poterba, Weil and Shiller (1991) and Schnure (2005) for the US, Englund and Ioannides (1997), Goodhart and Hofmann (2008) for a number of industrialized countries. The

much wider data availability in the US has allowed more granular studies using panel and spatial models. This is an important avenue little employed in

policy analysis where an aggregate national index may hide very large

degrees of regional heterogeneity. Among the studies related to the EU countries, Kajuth, Knetsch and Pinkwart (2013) have shown that German

house price dynamics exhibited fairly pronounced regional heterogeneity in recent years. However, data limitations at the regional level render this

approach quite challenging for practical forecasting purposes. Empirical methods suited for large data sets such as Factor–Augmented VARs

(Eickmeier and Hofman, 2013) and dynamic factor models such as Luciani (2015), Del Negro and Otrok (2007), Ng and Moensch (2011) have also been

used in the literature to model house prices.

* Motivation for the Problem Undertaken

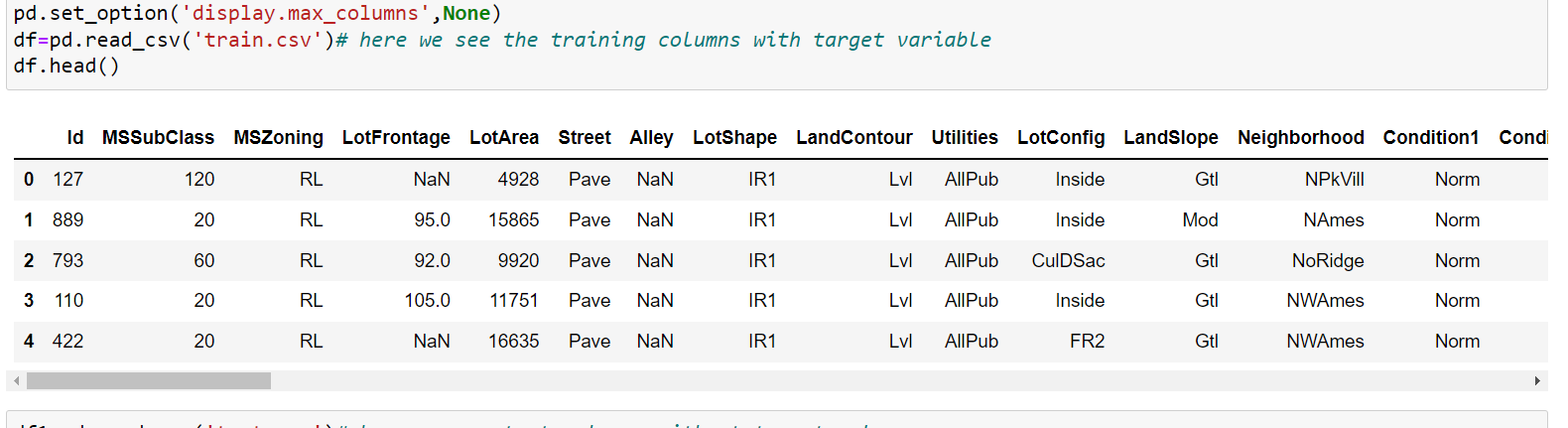
These are the two main problem which I focus during working .

1. Which variables are important to predict the price of

Houses ?

2. How do these variables describe the price of the house?

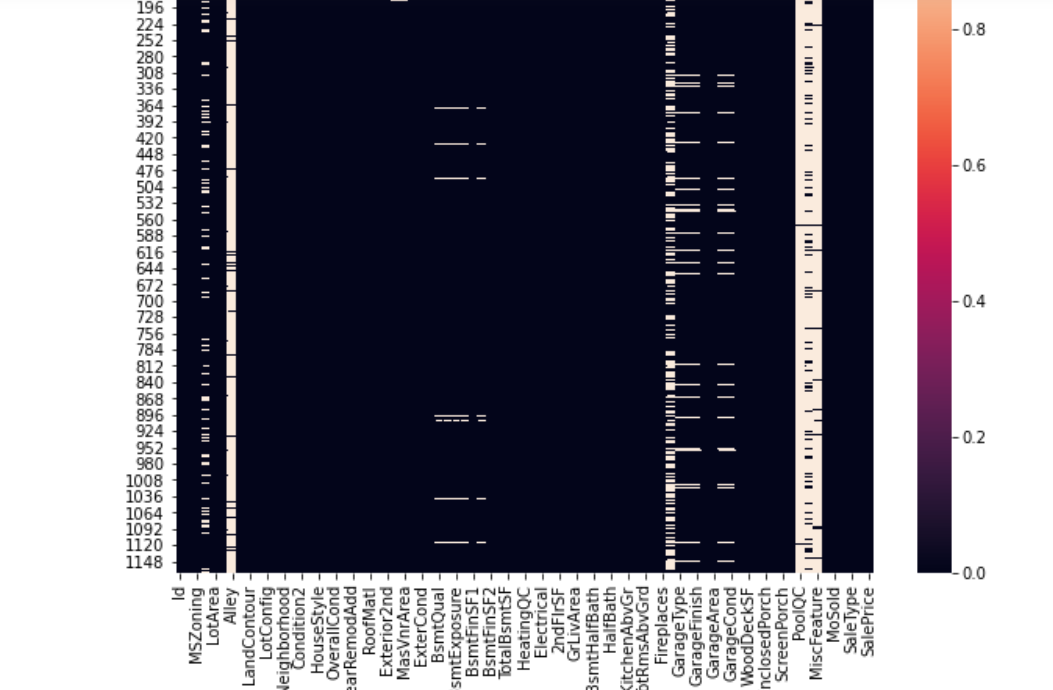
**showing the data in work book.**



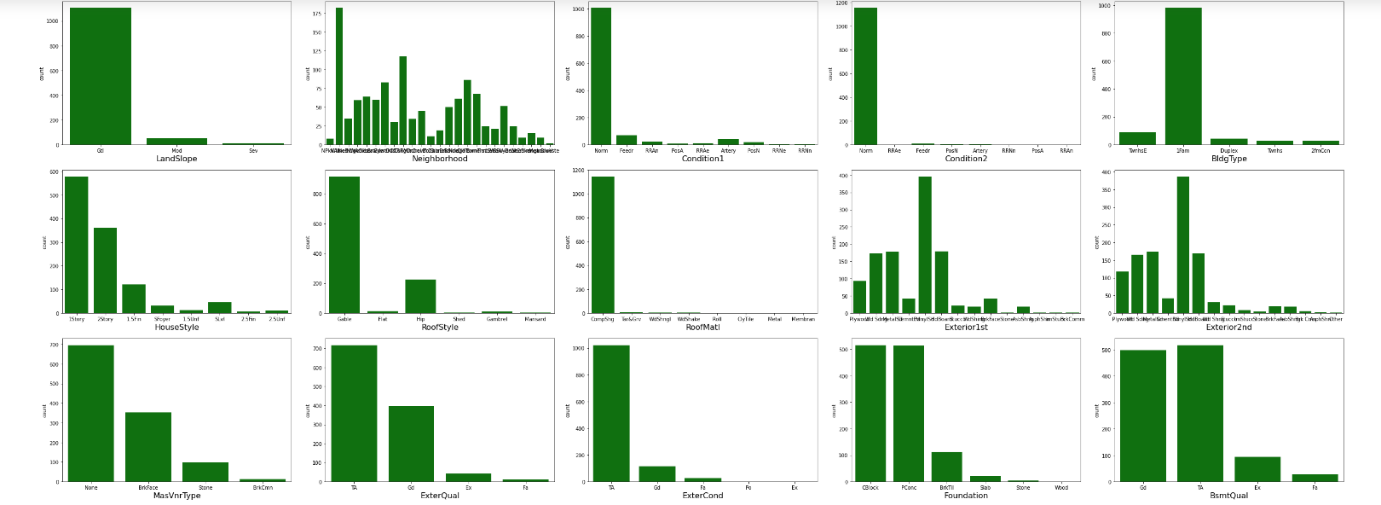
**Checking null values.**

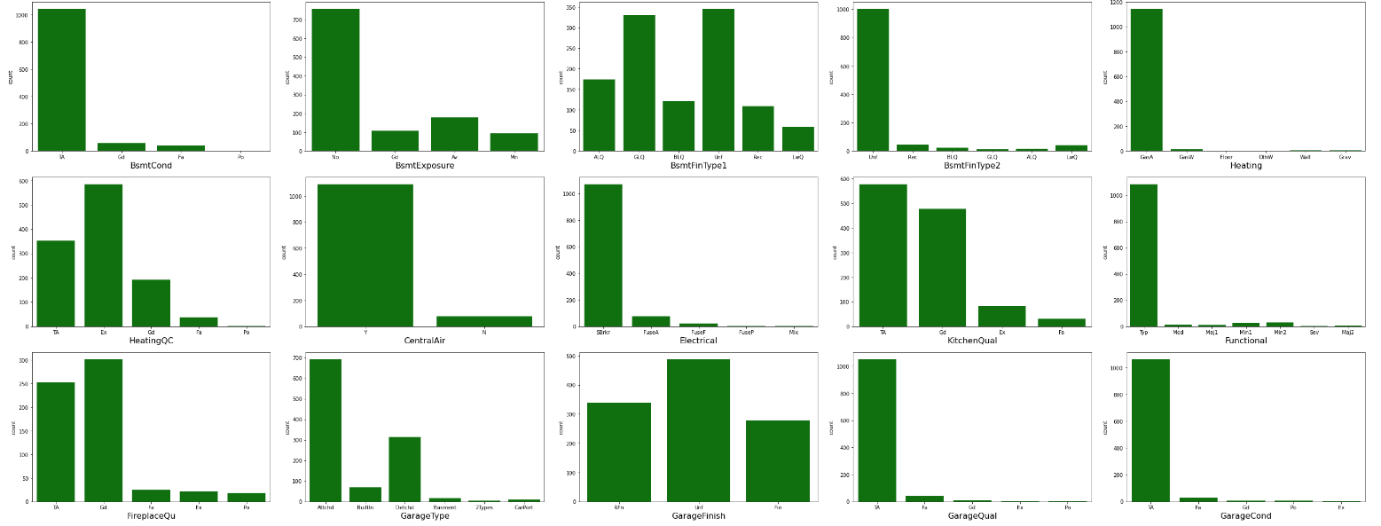
****

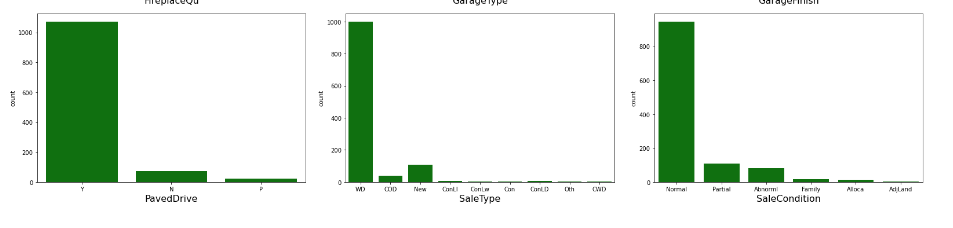
**Mapping the null values**



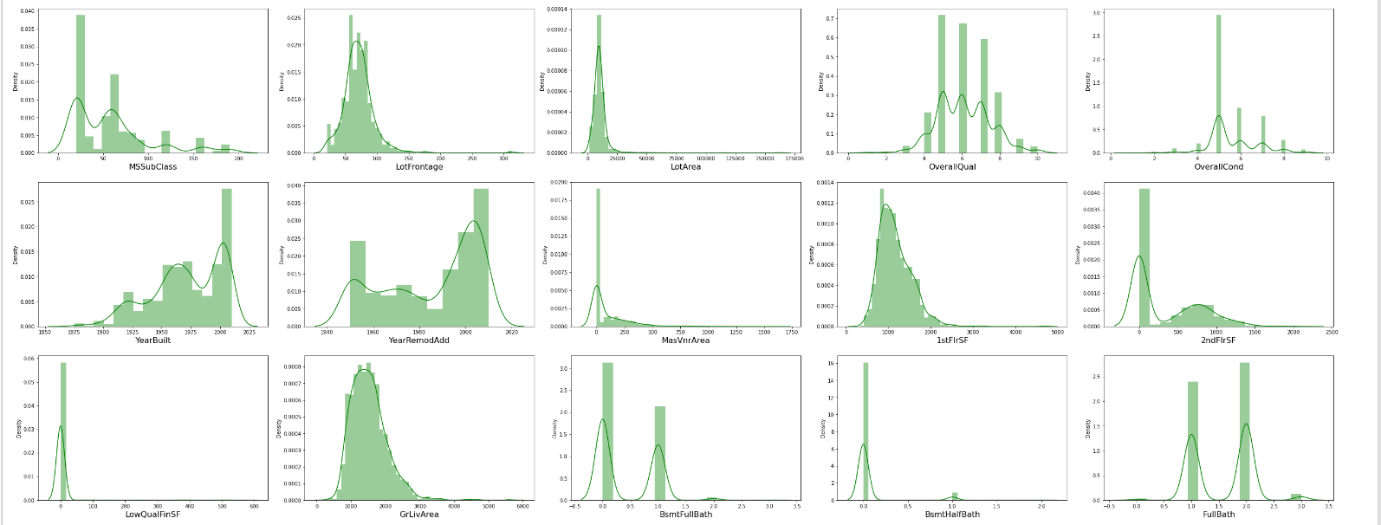
**All the categorical columns**

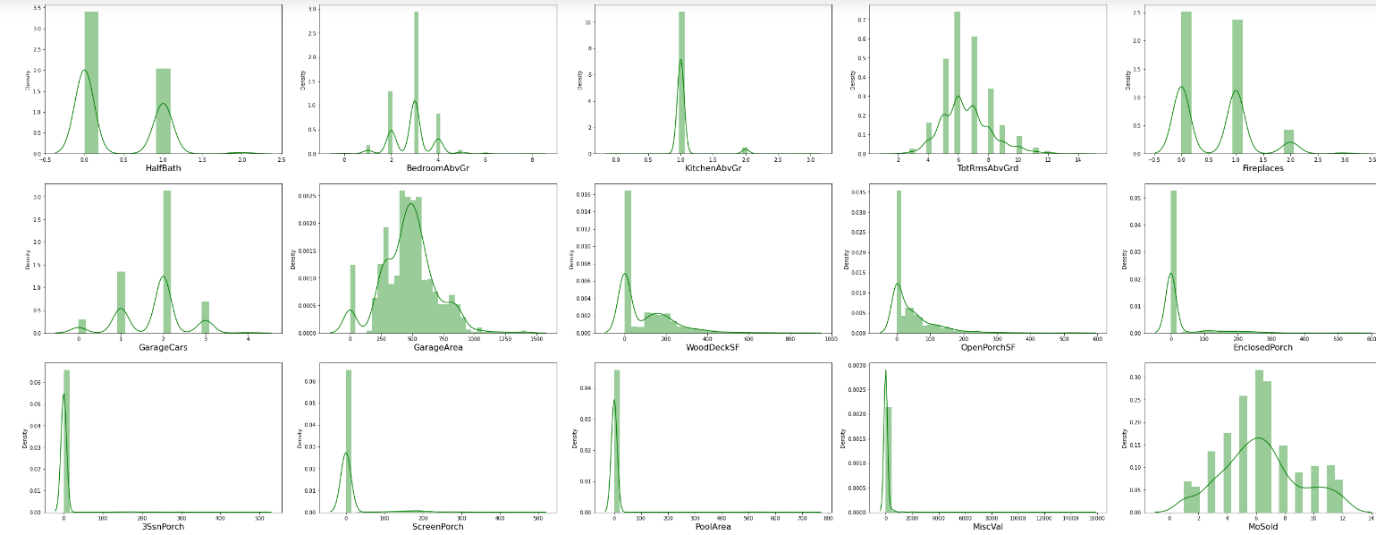


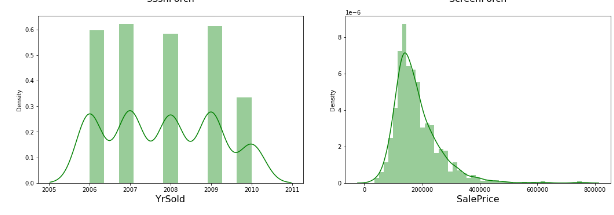




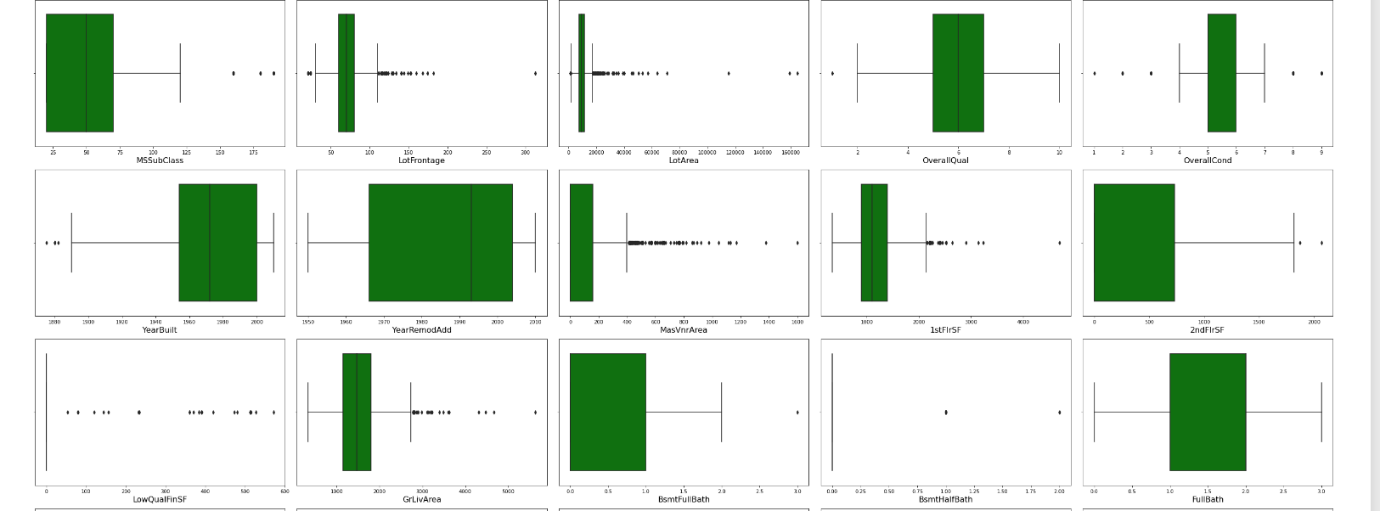
**All the continuous columns showing distribution plot**

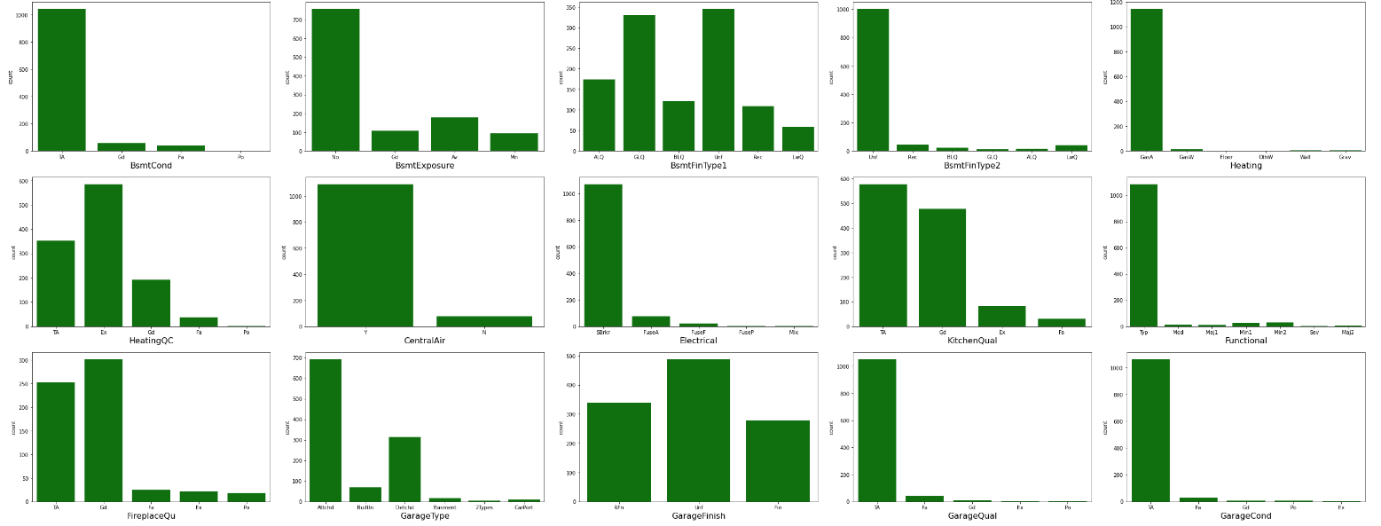


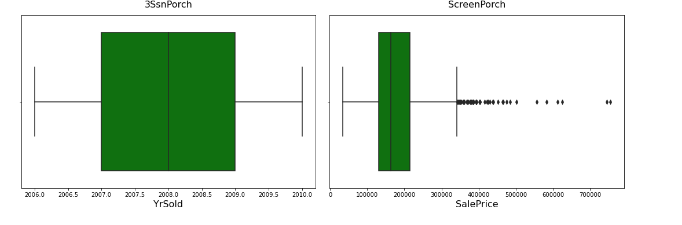




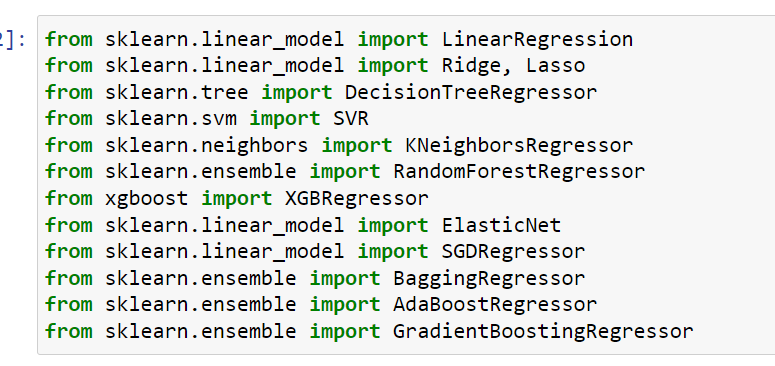
**All the continuous columns showing the outliers**

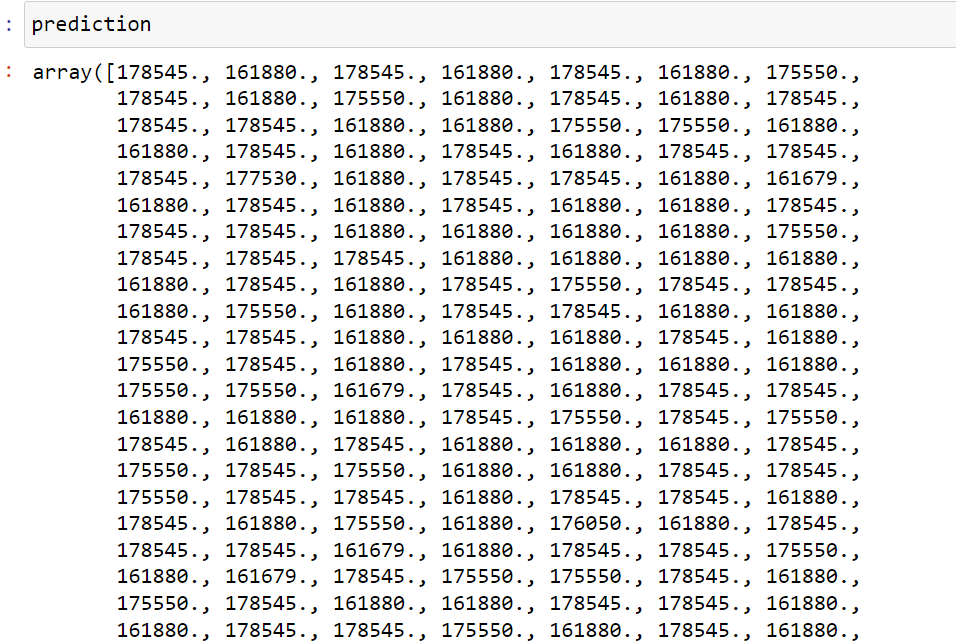






**Model training and building**



**Model prediction** 

Conclusions

In general, the choice of a house price model and its empirical estimation is very much influenced by the quality and availability of data.

While for most developed economies, sufficiently long time series of houseprices and relevant fundamental variables allow error-correction or vector

error-correction models to be estimated, this does not hold for a large number

of Central and Eastern European Countries (CEE). Short time series, insufficient market coverage, lack of quality adjustment or distinction between

old and new dwellings make forecasting much more challenging. The country size and the stage of economic development are identified as factors

determining house price elasticities, with smaller countries and catching-up economies having higher responses to similar sized fundamental changes

than larger and more well developed economies (Mihaljek and Subelyte

2014). Also, Egert and Mihaljek (2007) found that CEE countries tend to close any over- or undervaluation gaps of house prices faster than OECD countries.

References

Annett, A. (2005), House prices and monetary policy in the Euro area,

Chapter 3, IMF, Euro Area Policies: Selected Issues, Available at:

http://www.imf.org/external/pubs/ft/scr/2005/cr05266.pdf

Benjamin, J.D., P. Chinloy,, and G. Donald Jud (2004), Real estate versus financial wealth in consumption, Journal of Real Estate Finance and

Economics, 29, 341-354