Abstract

This study is based to inspect the ability to predict housing prices is important such that investors can make better asset allocation decisions, including the pricing and underwriting of the mortgage. The study intends to explore the possibility of modelling housing price based on Asset-based. Asset pricing theory all stems from one simple concept: price equals expected discounted payoff (rent). The model assumes an association among user cost UC_t, the house selling prices, Pt, and the rents R_t

In this project, first, the data is collected online (Google Doc) via a self-administered questionnaire through the real estate investors in Delhi/NCR and have stored dataset in the database of MS-Access. The dataset has fifteen variables, with twelve analysis variables, which were of both types: qualitative and quantitative. The study was intended to find the relationship of the independent variable (Indicative price) with the dependent variables such as Payoff (Rent), Annual Maintenance Cost and the expected rental increase.

The data was analyzed using CRISP approach. The result was analyzed statistically, to come up to the conclusion that Asset-Based Models can be applied in the case of NCR properties.

The Algorithm has been built in R-Script, and the Model is compared between the two, Simple Multiple Linear Regression and Stepwise Multiple Linear Regression and the results were compared.

Table of the contents:

List of Figures/Plots

List of Tables

Abbreviations used

ANOVA- Analysis of Variance

CRISP-DM- Cross Industry Standard Process for Data Mining.

AIC- Akaike Information Criterion.

Acknowledgement

It gives me immense pleasure to profusely thank my supervisor and guide **Mr. Ashraf Jafri** for the unflagging zeal with which he guided me in carrying out the project work. It has been a proud privilege to work with him. I am deeply indebted for his invaluable guidance, constant encouragement, constructive criticism and inspiring suggestions throughout the course of my project work.

I extend my gratitude to iMT, CDL for giving me this opportunity to research on the topic and learn new things.

I further acknowledge with a deep sense of reverence, my gratitude towards my parents and family members.

Last but not the least to my wife "**Ghuncha Zafar**" who dedicatedly and conscientiously supported me in all my endeavors.

Thanking You,

Mannawar Hussain

Chapter-1

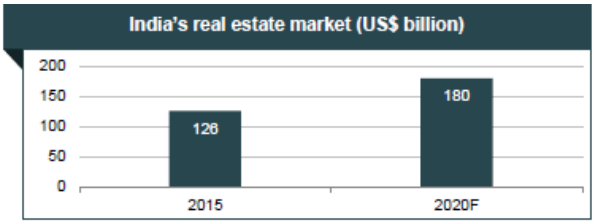
1.0 A Macro-Economic Snapshot of Residential Housing Market of NCR

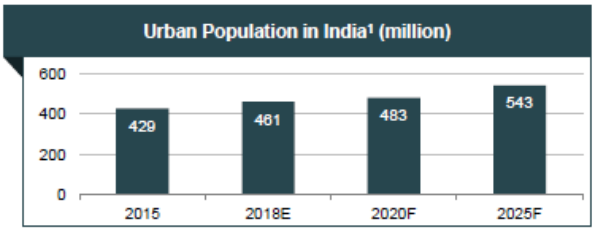
1. Introduction to the Housing Market

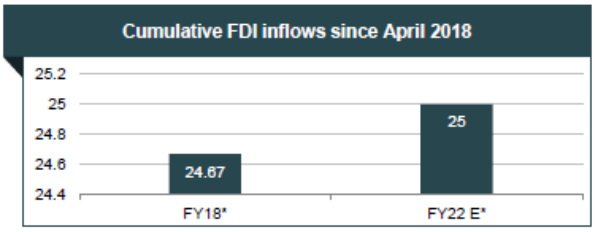
The residential real estate industry in India is complex and dynamic. Varying economic and demographic characteristics across the country result in differences in the housing markets in different cities. The real-estate industry is inextricably linked to socio-economic growth and development. Increasing the availability and affordability of housing can improve the quality of life for citizens, and the overall development of infrastructure improves productivity within the economy. House prices also indirectly impact macroeconomic health through the wealth effect: as house prices rise, homeowners feel wealthier and are likely to increase consumption and boost aggregate demand (Calomiris, Longhofer, & William, 2012).

Housing prices have declined by 7 per cent in the national capital region (NCR) and by 2 per cent in the Mumbai Metropolitan Region (MMR) in the last year, indicating the tepid residential market for the sector reeling under the worst ever slowdown (Business Today, 2019). The year 2018 showed a recovery for the residential sector, with both sales and new supply, gradually reviving across the top 7 Indian cities - Bengaluru, **National Capital Region (NCR),** Mumbai Metropolitan Region (MMR), Chennai, Kolkata, Pune and Hyderabad (NHB, 2018). The year 2018 has brought about a welcome change to the staggering NCR residential market. On one hand, the market witnessed cautious and strategic new launches under the watchful eye of the Real Estate (Regulation and Development) Act, 2016, while on the other green shoots of recovery were seen on the demand side with sales numbers faring well in 2018, indicating a slight revival of confidence in the market.

- Real estate sector in India is expected to reach a market size of US$ 180 billion by 2020. Housing sector is expected to contribute around 11 per cent to India's GDP by 2020.
- Rapid urbanisation bodes well for the sector. The number of Indians living in urban areas is expected to reach 543 million by 2025[1]. More than 70 per cent of India's GDP will be contributed by the urban areas by 2020.
- 4th largest sector in terms of FDI inflows. FDI in the sector[1] stood at US$ 24.67 billion from April 2000 to December 2017. FDI in the sector is estimated to grow to US$ 25 billion by FY22.
- Government of India's Housing for All initiative is expected to bring US$ 1.3 trillion investments in the housing sector by 2025.







*Notes: E – estimated; * from April 2000 to December 2017, [1]Construction development*
Source: Ministry of Tourism, KPMG, World Bank, Census 2011, EY – India's Growth Paradigm 2017, Credai-JLL report, [1]United Nations World Urbanization Prospects 2018

New launches in NCR registered a 35% growth in the number of units launched in 2018, with approximately 75% of the newly launched units falling in Gurugram and Greater Noida. The market saw approximately 15,819 units launched in 2018 compared to 11,726 units in 2017.

Gurugram contributed significantly to the new launches in NCR in H2 2018, taking up 52% of the overall pie, followed by Greater Noida and Ghaziabad. While previously, Gurugram saw new launches in the affordable housing category, the period of H2 2018 had notable developers launching new projects in the above ` 7.5 mn price bracket. New launches in Noida and Greater Noida have also registered a slight revival in 2018 with developers launching new projects majorly in sector 150 in Noida and Sectors 1, 10, Pari Chowk and Zeta 1 in Greater Noida.

Accumulating inventory, debt laden prominent developers and delayed projects in locations such as Greater Noida west, and Noida Greater Noida Expressway has put pressure on the developers to complete projects in hand than launch new ones.

Weighted average prices in the NCR residential market have registered a marginal 2% increase over H2 2017, which effectively translates into a stagnant market. The market has seen a slowdown in prices since 2013 and the scenario has not changed ever since. Low investor interest and an inventory overhang of more than one lakh units have kept the developers cautious of increasing the prices. On the demand side, sales have started to look up in 2018. On a yearly comparison, approximately 40,646 units were sold in NCR in 2018,

registering a growth of 8% over the 2017 sales numbers. The second half of 2018 registered a 10% growth over the same period in 2017 and helped propel the uptick in sales. Our survey suggests that properties that are ready-to-move or are nearing completion are garnering more enquiries from buyers. This can be attributed factors such as policy intervention in the form of Real Estate (Regulation and Development) Act, 2016, which has renewed the developer interest to finish projects at hand rather than launch new ones. Another factor is that the buyer is still not out of the woods and does not want to risk buying an under-construction project.

Greater Noida yet again takes the lion's share of the demand in 2018 with 50% of the overall sales coming from this affordable micro-market, which is distantly followed by Ghaziabad

and Gurugram at 19% and 17%, respectively. Majority of the projects in Greater Noida is in the less than 5 mn price bracket and are driving demand in this market. Regardless of having no major employment pull, the micro-market is not only

attractive to buyers of Noida and Greater Noida but also adjacent tier II cities of Uttar Pradesh such as Mathura, Agra and Aligarh due to its affordability. The second half of 2018 fared well for the Greater Noida market with a more than 100% growth in sales in H2 2018 as compared to the same period in 2017.

Demand in Gurugram has somewhat remained tepid in 2018. Our survey findings suggest that even though there are increased enquires, certain infrastructure delays such as the Dwarka Expressway is holding back interested buyers who are majorly end users. Another factor contributing to this sluggish demand is that the fresh inventory with builders is in direct competition with the inventory lying with investors who are ready to take a hit and exit the property.

NCR Market Snapshot

Parameter	*H2 2018*	*Change YoY*	*2017*	*2018*	*Change YoY*
Launches (housing units)	6,696	-3%	11,726	15,819	35%
Sales (housing units)	22,599	10%	37,653	40,646	8%
Price (weighted average)	₹ 45,640/ sq m (₹ 4,240/ sq ft)	2%	₹44,832/sq m (₹ 4,165/sq ft)	₹ 45,640/sq m (₹ 4,240/sq ft)	2%
Unsold inventory (housing units)	142,007	-15%	166,831	142007	-15%
Quarters to sell	15	-	17	15	-
Age of unsold inventory (in quarters)	22	-	19	22	-

Note- 1 square metre (sq m) = 10.764 square feet (sq ft)

Source: Knight Frank Research

Demand in Noida has also failed to impress in H2 2018. The market registers a YoY 26% decrease in sales compared to H2 2017, market sentiments remain lukewarm.

The quarters to sell unsold inventory (QTS) is the number of quarters required to exhaust the existing unsold inventory in the market. The existing unsold inventory is divided by the average sales velocity of the preceding eight quarters to arrive at the QTS number for that particular quarter. A lower QTS indicates a healthier market. Yet again Ghaziabad and Greater Noida emerge as NCR's better-performing markets, with a QTS of 13 quarters, closely followed by Noida at 14 and Gurugram at 18 quarters. The steady pace of sales and cautious new launches have consequently brought down the unsold inventory by 15% in H2 2018 compared to the same period in 2017. The unsold inventory stands at approximately 142,007 units as of December 2018. Greater Noida and Gurugram account for approximately 65% of the unsold inventory in NCR followed by Ghaziabad and Noida

Greater Noida led sales in 2018. 50% of the total sales in NCR came from this affordable micro-market. The market is not only attractive to buyers of Noida and Greater Noida but also, adjacent tier-II cities of Uttar Pradesh such as Mathura, Agra and Aligarh due to its affordability (Knight Frank, India, 2018)

1.1 Proposal to be investigated

It is widely acknowledged that the value of a house is the mixture of many characteristics. House price prediction thus presents a unique set of challenges in practice.

The homebuyers are mostly concerned with ups and downs of housing prices. During the long process of comparison many buyers might miss the opportunity to buy due to rise in house prices. Also, since houses change hand relatively

infrequently, it becomes necessary to predict the value of house between the sales. These predicted values are used for many purposes including second mortgages and homeowner's insurance.

The asset-pricing model of house prices provides an intuitive method to connect the cost of owning housing units to rents through introducing the concept of user cost, which by definition is comprised of depreciation cost of the housing unit, maintenance fees, mortgage interest payment, and the capital gains from the housing asset.

The analysis of the determinants of housing prices is important because of the housing impact on economic and social factors. Firstly, homeownership has a positive effect on residential mobility, resident's health and other social consequences (DIETZ *et al*., 2003).

Secondly, housing is a good, which is closely connected with other markets and the whole economic status of the country. As houses can be purchased by the mortgages as well as the own funds of the residents; the housing market is especially related to financial sector (JUREVIČIENĖ; OKUNEVIČIŪTĖ; NEVERAUSKIENĖ, 2008).

Finally, the changes in housing prices influence the construction market and other economic variables such as unemployment and inflation (AZBAINIS, 2014).

1.2 Significance of the Problem

The principle underlying an asset-based approach to welfare is that, rather than relying on state-managed social transfers to counter the risks of poverty, individuals accept greater responsibility for their own welfare needs by investing in financial products and property assets which augment in value over

time. These can, at least in theory, later be tapped to supplement consumption and welfare needs when income is reduced, for example, in retirement, or used to acquire other forms of investment such as educational qualifications. (John Doling Richard Ronald, 2010)

1.3 Objectives and the Scope of the Study

The ability to predict housing prices is important such that investors can make better asset allocation decisions, including the pricing and underwriting of mortgages. And therefore, the prime objective of this study is:

- The study aims to investigate the determinants of asset-based housing Pricing Model.
- A Stepwise Multiple Linear Regression Model is employed for predicting the asset-based price.
- The price estimate would further be analysed in terms of possible economic rental gain.

Thus, the objective of this research is to inform an effective investment pathway for social housing.

The data for this study is collected through the primary method. A self-administered structured questionnaire will be prepared, and the response will be collected online through Google Doc from the Property Buyers of NCR location only.

The quantitative variables are presented in summary statistics-mean, standard deviation, maximum and minimum. The categorical variables (nominal) will have frequencies and percentages. All test statistical tests will be two-tailed, tested at α=5% level of significance. The data summary would be presented in statistical

tables, as obtained from R-Script. Suitable graphs will be used to help aid in statistical reporting using ggplot in R.

1.4 Contribution and Impact

The purpose of this research study is to focus on the unusual aspects of the NCR housing market, as well as the characteristics of the market that might be of interest to policy makers. Hence an attempt is made to explain why house prices have not exhibited the same boom and bust cycle in recent years, and why the homeownership rate remains low. Among other things, the study has focused on the lending practices of NCR banks, which require a (20-30) % down payment when purchasing a property. Also, the Asset Based Model can be used in NCR in assessing the value of the collateral, thus banks have much information Housing Finance, Prices, and Tenure in NCR relative to the fair value of a property when granting a loan.

In short: This study investigates the value of housing, which is a specific group of investment assets, in particular the primary valuation drivers of the NCR housing price, by using Asset Based Model.

The main contribution of the Model will be helping in understanding the primary valuation drivers of housing prices will make market participants aware of the size of their risk exposure and can help them to detect early signals of the possibility of investment opportunities. Furthermore, policymakers can use information about the underlying valuation drivers of the house prices to stabilize the market.

1.5 Brief description of the organization (where project is conducted)

Darashaw & Company Pvt. Ltd

Founded in 1926 Darashaw is amongst the oldest continuously managed diversified financial and consultancy houses in India. Built on a tradition of integrity value addition, client commitment and a joy to work with. Darashaw today provides a wide range of financial services to a substantial and diversified client base that includes corporations, financial institutions, retirement trusts, the Government and individuals. Darashaw headquartered in Mumbai and has pan India presence. With a commanding presence in the debt capital markets and asset management, Darashaw today has forayed its expanse into Infrastructure Consulting and the BPO / KPO space.

Chapter-2

2.0 Literature Review

2.0.0 Indian Housing Industry-Pricing Models

Real estate companies in India must improve their condition so that their market share can be improved in this huge competitive environment. The approach which is required is the insights of research into market demand, micro and macro-economic market factors and the study of past sales of sites in a given location. These insights can give an accurate and predictive road map of where the market is headed

According to experts tracking the realty sector, sites such as Magicbricks.com, 99acres.com, Makaan.com, IndiaProperty.com and Common Floor.com are fast becoming the choice of consumers looking for renting a property, as well as for developers. The online real estate market is yet to reach an inflection point as seen in the travel or e-commerce sector. However, with the penetration of the internet, more and more consumers in the urban areas have started using this medium as the first point of search for all their real estate needs. In changing real estate and construction market, the old thumb rules used by industry specialist sometimes stops being valid. There is a fundamental shift in the market rules. Sometimes the investments involved in the industry are very high and these investments are built on speculation. It is very valuable to collect data on the market dynamics in a scientific manner to reduce risk.

Among asset markets, the most heavily researched is the equity market. Two examples include Fama and French (1988) and Poterba and Summers (1988). Using different methodologies both studies find significant evidence of mean

reversion at long horizons. For example, Fama and French conclude that "predictable variation is estimated to be about 40 percent of 3-5-year return variances for portfolios of small firms. Schiffman and Kanuk (2010) mentioned that consumer behavior is 'the behavior that consumers display in searching for purchasing, using, evaluating and disposing of product, service and idea which they expect will satisfy their needs'. Study of decision-making process in buyer behavior has attracted many researchers. It dug upon developing and understanding a logic as to how purchasers exercise their choice between two or more alternative products. Researchers try to understand the stages through which consumers pass when deciding which products or services to purchase and their behavior after that choice. These matters have been extensively investigated over many years by marketers, including Howard and Sheth (1969), Engel et al. (2005), Mowen (1995), Peter and Olson (2001) and Hoyer and MacInnis (1997). Kotler and Amstrong (2005), Mowen (1995) and Engel et al. (2005) suggest that there is a generic decision-making process. Rossi (1980), Law and Warnes (1982) and Livette (2006) have used the same model of decision-making process for evaluating housing options. Many attempts have been made to identify factors that influence the buyer behaviour for residential apartments. The features of the apartment structure itself will be an important determinant of a household choice of residence (Quigley, 1976). Neighborhood quality, local public services and quantity of housing services also affect the choice of residence (Friedman, 1980; Gabriel and Rosenthal, 1989; Hua Kiefer, 2007). Nechyba and Strauss (1997) found that public expenditures, tax, crime rate, commercial activity are factors influencing buying decisions. Connie and Fernando (2001) considered the effects of factors on residential development in PT Delta, Comoro Permai, Dilly as a case study.

They included following 20 factors that might influence buying behavior of a residential dwelling unit: ***Building Quality, Design, Price, Building area, Land area, Affordable down payment, Interest Rate, Payment Period, Clean, Safe, Aesthetic, Electricity, Water, Telephone, Linkage to the Working Place, Accessibility to the Market, Short Distance to the School, Place for Worship, Sport Facilities, Investment and Public Transport.***

Since the housing share moves slowly, a concern with composition risk induces low-frequency movements in stock prices that are not driven by news about cash flow. Moreover, the model predicts that the housing share can be used to forecast excess returns on stocks.

Kanojia, Magar and Jadhav (2016) states in his research paper Hedonic Pricing Method (HPM) is used for estimating the value of a commodity or the demand for a commodity. The method has been widely used in real estate and housing market research in the recent past. In this report, Estimation of residential market price by Hedonic Pricing Method due to environmental service is discussed. Environmental services are a concept from the ecological economics literature. Environmental services refer to qualitative functions of natural non produced assets of land, water and air (including related ecosystem) and their biota, which implies natural environment providing particularly important uses or benefits that can be captured under the concept of "services." It undertakes an examination of the use of this methodology in the recent real estate and housing literature in the preparation of housing predictive model.

The problem is that the real estate sector in India has a massive information irregularity, the companies have all the information and there is no means to confirm if the information they are displaying out is correct.

2.0.1 Statement of Problem

By studying above research work it can be interpreted that there is fair amount of work which has been done to identify the parameters of real estate, but as pointed precisely by the researcher (Prashant Das et al, 2013) the gist of which is the information provided to the end user is not accurate, hence at the same time by looking at current trend of real estate market there is no denial of the fact that many newly launched real estate projects in good locations are losing out on sales because of faulty marketing vision. In this project we will try to scientifically study the effect of various determinant on the final price of the housing, and hence as a result helping developer segment out the niche market for developing customized marketing strategy for each market segment and as a result helping the end user with well-informed decision which will certainly help the real estate industry to grow.

Chapter-3

3.0 Research Methodology

Research in common parlance refers to a search for knowledge. Research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English lays down the meaning of research as "a careful investigation or enquiry especially through search for new facts in any branch of knowledge." Research is, thus, an original contribution to the existing stock of knowledge making for its advancement. Further, it is a systematic search for getting answers to our problems or questions. The answer is from the knowledge.

It is the pursuit of truth with the help of study, observation, comparison and experiment. In short, the search for knowledge through an objective and systematic method of finding a solution to a problem is ***Research.***

3.0.0 RESEARCH APPROACH

The research approach for this study had been Quantitative research that is explaining the phenomena by collecting numerical data that are analyzed based on Statistical Learning System methods (Multiple Linear Regression). The reason why Quantitative research has been used as the study is designed Asset Based Housing Model.

3.0.1 RESEARCH DESIGN

The Research Design used in this study is ***Cross-sectionally used for non-experimental Exploratory Research.*** Generally speaking, these studies are carried on a large sample population at a single point of time, seeks to describe the Mathematical relationships between the constructs (dependent and independent variables).

3.0.2 BASIC THEORY OF MULTIPLE REGRESSIONS

Regression is usually introduced in connection with the theory of correlation, but it is, in reality, a more general, and, in some respects a simpler idea and the regression coefficients are of interest and scientific importance in many classes of data where the correlation coefficient, if used at all, is an artificial concept of no real utility. In many applications, there is more than one factor that influences the response. Multiple regression models thus describe how a single response variable Y depends linearly on several predictor variables.

3.0.3 DATA COLLECTION (RESEARCH INSTRUMENT)

In all research problems, it is often found that data at hand (published sources) are inadequate, and hence, it becomes necessary to collect data that are appropriate (fresh, update and study-related). Data were collected through self-administered structured questionnaire uploaded through Google doc, and the data was captured by collecting information from different Real Estate Agents in NCR namely Gurgaon, Delhi, Noida, and Faridabad. Further, the importance of their input will also be guaranteed confidentiality and anonymity.

3.0.4 STRUCTURED QUESTIONNAIRE

BUYER PROFILE

Name of the Builder \ Company								
Buyer Name (Initial only)								
Profession	Business	1	Private Service	2	Government Service	3	NRI	4

ASSET INFORMATION

Indicative Selling Price (in crore only)										
Location where the Property purchased	Gurgaon		Delhi			Noida			Faridabad	
	1		2			3			4	
The property type concerning the number of bedrooms				The total carpet area in square foot						
Year of Purchase	2019		2018			2017			2016	
	1		2			3			4	
Expected Leasing Period in years	0-5		6-10		11-15			Please write years if above 15		
	1		2		3					
Property purchased (type)	Full Payment					Bank Loan				
	1					2				
Type of Property	Villa		Apartment-Flat			Housing Society			DDA	
	1		2			3			4	
	1	2	3		4	5	6		7	
Expected Payoff	Security Amount				Monthly Rental					
Annual Maintenance Amount				Annual Property Tax						
Percentage of Annual Rental Increment (in % only)				Expected Property Appreciation (in % only)						
Social infrastructure in the vicinity	School		Hospital		Metro		Bus Stop		Market	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

3.0.5 SAMPLING PLAN

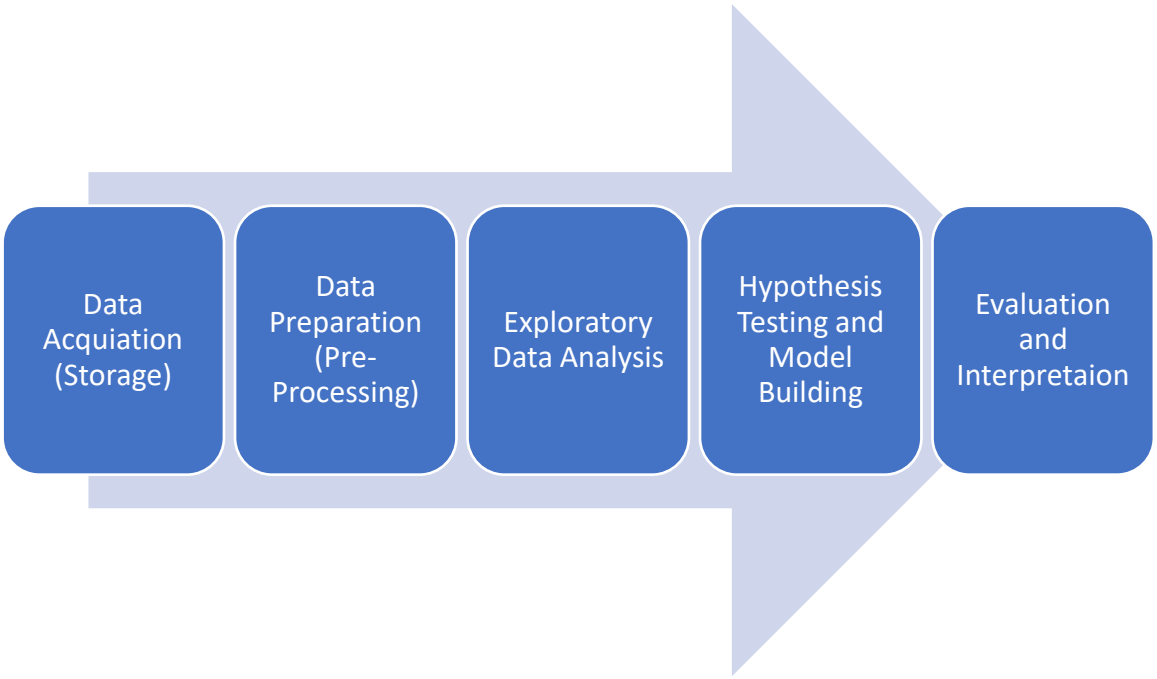
The sampling technique used in this Research is Convenient Sampling Technique. We have selected the observations (records) from the collected database, using the following inclusion criteria:

- Buyers or Investors buying high-valued properties.
- The property used for analysis was mostly ready to move properties.

Therefore, the target population will be a mix of the Potential Buyers or Property Investors of the NCR regions.

3.0.6(a) CRISP-DM MODEL FOR DATA ANALYSIS AND INTERPRETATION

The typical CRISP Model (Cross Industry Standard Process for Data Mining) of a defined framework would be followed for making an informed decision, and therefore the Statistical Learning life cycle would have the following Project Phases in Data Analysis.



3.0.6(b)Data Acquisition: This step refers to Data Storage. In our research study, we would store the data in MS-Excel worksheet, so that the data can be cleaned and validated easily. Also, the dataset is not too large to be stored in any DBMS.

3.0.7 LIMITATION OF THIS RESEARCH STUDY AND ANALYSIS

Statistical Models are based on probabilities and though it uses the techniques of mathematical science. Yet, the results and conclusions drawn are not assumed to be very accurate. Therefore, the study is not a complete solution to any marketing issue as there are many dominant variables between research conclusions and market response. In short: It provides only probable, complex, and abstract solutions whereas traditional management expects simple and concrete solutions.

Chapter-4

4.0 Data Analysis and Inference

4.0.1 Study Dataset

4.0.1.1 Data Source

The study sample data consisted of cleaned 1899 observations (respondents), of both males and females, selected (cleaned) from list of datasets collected from various Real Estate Sellers (Builders, Agencies, and Intermediaries) through Google Doc.

4.0.2 Data Description

The study dataset consists of data for 1899 observations. Each observation contains information on 15 variables. The description of the different variables is here under.

Sr. No	Name of Variable (Attribute)	Definition of the Variable (Attribute)	Section	Descriptions	Type of Variable	Type concerning Analysis
1	Name_B	Name of the Builder/Company	Buyer Profile	This states the builder name or its company which is collected via real estate intermediaries	Character	Not to be analyzed
2	Buyer	Buyer Name (Initial only)		This states the buyer name who purchased property in NCR region	Factor	
3	Proff	Profession		The response of the buyer profile is coded from (1 to 4) 1-Business 2-Private Service 3-Goverment Service 4- NRI	Factor	
4	Price	Indicative Selling Price (in crore only)	Asset Information	This states the price at which any particular properties have been sold in NCR region by the builders	Numeric	Data to be analyzed

Sr. No	Name of Variable (Attribute)	Definition of the Variable (Attribute)	Section	Descriptions	Type of Variable	Type with respect to Analysis
5	Location	Location where the Property purchased	Asset Information	This corresponds to the region in NCR where the actual property is purchased and is coded as below: 1- Gurgaon 2- Delhi 3- Noida 4- Faridabad	Factor	Data to be analyzed
6	Prop_type	Property type with respect to the number of bedrooms		This corresponds to the type of property with regards to bedroom and is coded as below: 1 BHK- 1 2 BHK- 2 3 BHK- 3 4 BHK- 4	Factor	
7	Carpet_area	The total carpet area in square foot		Carpet Area is the area enclosed within the walls, actual area to lay the carpet. This area does not include the thickness of the inner walls	Numeric	
8	Year	Year of Purchase		This relates to year of purchase of property by the end user and are as following: 2019- 1 2018- 2 2017- 3 2016- 4	Factor	
9	Lease_Per	Expected Leasing Period in years		This corresponds to the range of year and is coded as below: (0-5)- 1 (6-10)- 2 (11-15)- 3 (16 & above)- 4	Factor	

Sr. No	Name of Variable (Attribute)	Definition of the Variable (Attribute)	Section	Descriptions	Type of Variable	Type with respect to Analysis
10	Prop_purchase	Property purchased	Asset Information	This refers to the method by which property is purchased and is coded as below: Full Payment- 1 Bank Loan- 2	Factor	Data to be analyzed
11	Type_Prop	Type of Property	Asset Information	This refers to the type of property purchased and is coded as below: villa- 1 Apartment/Flat- 2 Housing Society- 3 DDA-4	Factor	
12	Payoff	Expected Payoff	Asset Information	This corresponds to the expected payoff in the form of monthly rental plus security deposit	Numeric	
13	Main_Amount	Annual Maintenance Amount	Asset Information	This corresponds to the amount of money spent on annual maintenance inclusive of annual property tax	Numeric	
14	Rental_increament	Percentage of Annual Rental Increment (in % only)	Asset Information	This corresponds to the increase in the rental amount of property considering appreciation of property each year	Numeric	
15	Social_infra	Social infrastructure in the vicinity	Asset Information	This corresponds to the availability of social infrastructure in the vicinity of property and as below: School- (Yes-y, No-N) Hospital- (Yes-y, No-N) Metro- (Yes-y, No-N) Bus Stop- (Yes-y, No-	Factor	

				N) Market- (Yes-y, No-N)		

The below is a step-by-step method of Multiple Linear Regression Model applied to the study data.

4.0.3 ALGORITHM FOR DEVELOPMENT OF MULTIPLE LINEAR REGRESSION MODEL				
DATA PREPARATION		**EXPLORATORY DATA ANALYSIS**	**MODEL BUILDING**	**MODEL EVAULATION WITH RESPECT TO COCHRAN ASSET BASED PRICING**
DATA ACQUISITION	**SPLITTING THE MASTER DATASET**			
Data stored in the MS-Excel is transformed in csv format.	The master dataset is divided into two datasets: Study Population and Analysis	The Analysis data variables are explored, by descriptive statistics, the structure of the variables, along with their names	The Linear Regression Model is obtained by using the Stepwise Regression Method. AIC is used to estimate the prediction accuracy of the model, and the criterion being the smallest AIC.	The Regression Models is evaluated to find the dependency of Payoffs to Price.
The master dataset is imported in R environment.		The Analysis Variables are then plotted visually for better representation.	The R-Square and Adjusted R-Square is taken into account. A minimum benchmark of 0.50 is taken for the Model to be accepted.	

The master dataset is cleaning of any missing values, and it was deleted if any.	The Study dataset is checked.		Diagnostic Plots are plotted to understand the residuals, and differences between the actual and the predicted.	
The master dataset was then transformed into a data frame.	The Analysis of dataset is checked.		The correlation among the dependent variables is evaluated by plotting the correlation plots.	
Finally, the master dataset was listed.	Converting the variables into factors in the Study Population.			
The master dataset is checked	Checking again the Study Population dataset.			
	Plotting the Study dataset variables.			

4.0.4 STATSITICAL MODEL OF THE MULTIPLE LINEAR REGRESSION

In this Statistical Study, the Y is the indicative price, or say the dependent (target) variable; that is supposed to linearly dependent to other explanatory variables (location, purchase year,…….., annual maintenance) say $X_{i;}$ and therefore we can write as $Y= X_1\beta_1+X_2\beta_2+….+X_n\beta_n+\varepsilon$(error). The β_1 ,β_2 ,…β_n are the regression coefficient associated with the respective explanatory variables X_1, X_2, …$X_{n.}$

4.0.5 STATSITICAL DEFINITION OF THE MULTIPLE LINEAR REGRESSION

Multiple Linear Regression is an extension (generalization) of simple linear regression. It is one of the most important statistical techniques for business

applications. It's a statistical methodology that helps estimate the strength and direction of the relationship between two or more variables

4.0.6 ASSUMPTION OF THE MULTIPLE LINEAR REGRESSION

- **Linearity:** A linear relationship is assumed between the dependent variable and the independent variables.
- **Normality of Residuals:** Regression residuals (predicted minus observed values) must be normally distributed.
- **Equality of Residual Variance:** When an independent variable is defined by group membership, then the assumption implies that the variance within each group is the same. Hence, in ANOVA terminology, the assumption is also called the homogeneity of variance assumption.
- **Fixed Independent Variables:** In Regression, we assume fixed-effect independent variables model.
- **Absence of Multi-Collinearity:** Absence of multi-collinearity is assumed in the model, meaning that the independent variables are not too highly correlated.

4.0.7 STATSITICAL RATIONALE

This is a predictive data mining problem, and the Multiple Linear Regression has been performed to determine the relative influence of independent (or predictor variables) on the dependent (target) variable, in a mathematical equation form. This model assumes an additive function relating a dependent (target) variable to a set of predictors (independent) variables.

DATA ANALYSIS

The statistical analysis of this study is done by a two-step process:

Exploratory Data Analysis: This includes the process of loading the data in R environment, cleaning, preprocessing, computing the descriptive statistics and plotting for the graphical presentation.

Model Development and Evaluation: This section contains discussions on the model, criteria for selection, and interpretation of the relevant regression parameters.

4.0.8 STUDY POPULATION

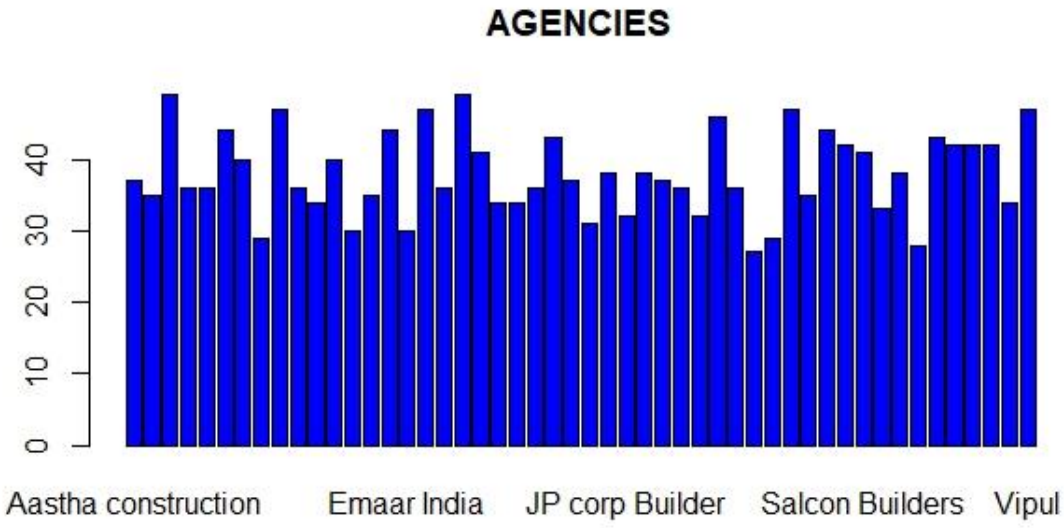
The tables below show the descriptions of the study population, which is arranged to the professions, wherein we find that the most properties are purchased are Government servants, followed by NRIs, Private Service and the last being the Businesspeople. The same is depicted in the pie-chart thereof.

STUDY POPULATION

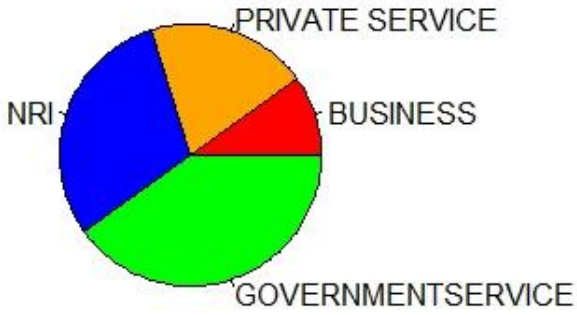
	BUSINESS	PRIVATE SERVICE	GOVERNMENTSERVICE	NRI
Aastha construction	9	8	11	9
Adani Group	7	7	9	12
Adarsh Homes	11	14	12	12
Aman Properties	9	9	8	10
Ambience Group	7	13	9	7
Apex Builder	10	14	9	11
ATFL Builder	8	13	7	12
ATS Green	5	9	6	9
BDI Builder Delhi	7	16	11	13
Bhasin Group	8	8	8	12
CGHS Group	9	8	9	8
Cosmic structures Builders	15	6	12	7
DDA Builder	8	7	9	6
Delhi Infratech	12	7	6	10

DLF group Builder	13	9	8
14			
Emaar India	12	6	5
7			
Garg Properties Builder	20	11	6
10			
GG Realtech Builder	8	12	10
6			
Godrej Properties	10	16	8
15			
Golden Sapphire Builder	12	9	9
11			
Golf Link Builders	13	7	7
7			
Goyal Associates	5	8	9
12			
Gulati Properties Delhi	13	6	13
4			
Gupta Realtors	7	17	6
13			
IFCI Infra Builder	6	11	11
9			
Jain properties	6	9	10
6			
Jayakay Builder	9	8	11
10			
JP corp Builder	8	7	7
10			
Kapil Associates	8	9	8
13			
Kawatra constructions	5	12	10
10			
Khurana Properties	8	11	8
9			
Kiera Homes	8	7	10
7			
Komal Properties	15	5	14
12			
M2k Builder	9	11	11
5			
M3M India	3	10	8
6			
Mahindra Lifespace developer	6	13	5
5			
MG Builders and associates	10	10	12
15			
Navgrow Builder	12	6	7
10			
Oxygreen Galaxy welfare	10	11	12
11			
Revanta Multistate	15	9	7
11			
Salcon Builders	10	11	13
7			
Sobha Limited	17	8	6
2			
Sona Associates	10	8	13
7			
space shapers Builders	6	6	9
7			
Tata Housing development	7	12	18
6			

The Antriksh group	13	12	8
9			
Unity Group	15	8	12
7			
V3 Infratech Builder	13	10	8
11			
Vasant Builder	12	4	11
7			
Vipul	12	10	15
10			

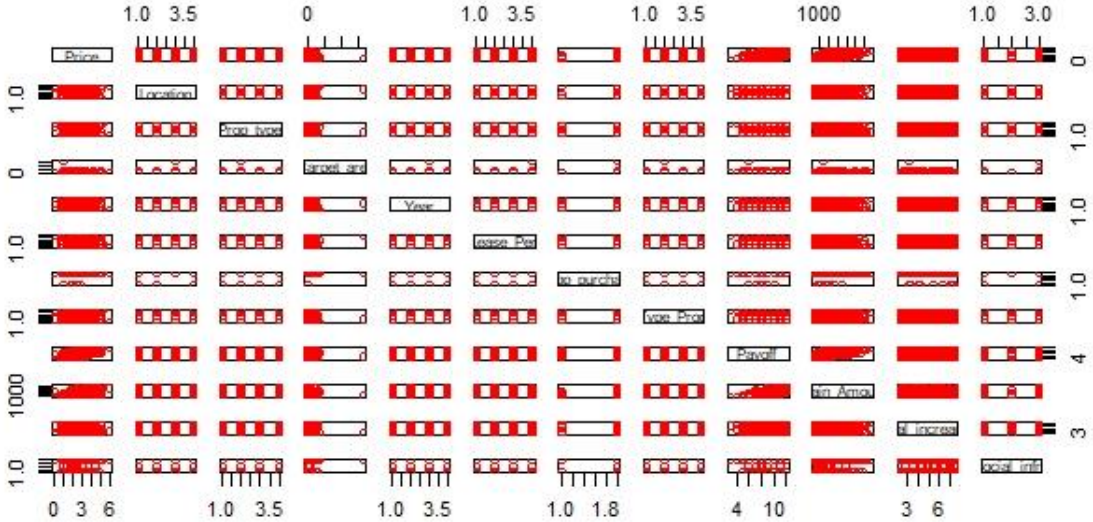


DISTRIBUTION OF PROFESSIONS OF STUDY POPULATION



The visual below shows the statistical distribution of the analysis variables, though not visually clear but can have a bird`s eye view of the variables.

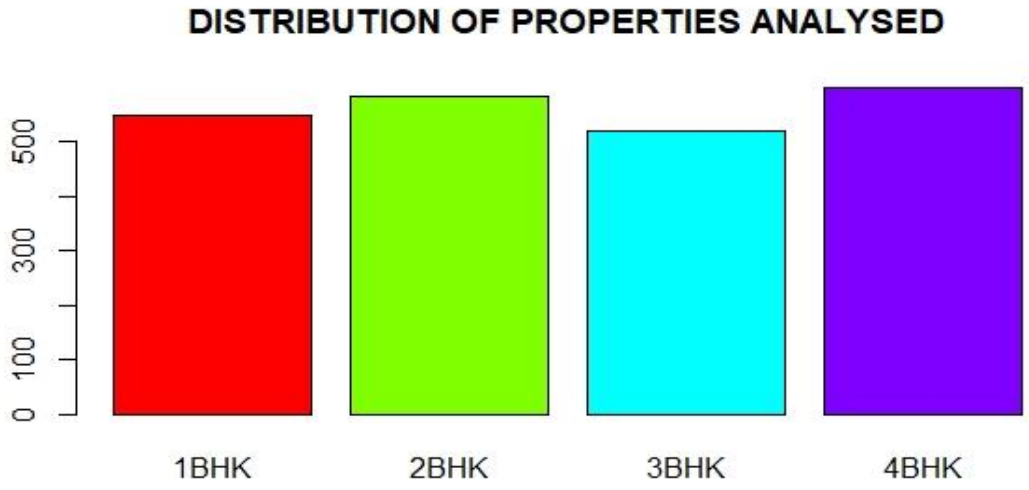
STATISTICAL DISTRIBUTION OF ANALYSIS VARIABLES

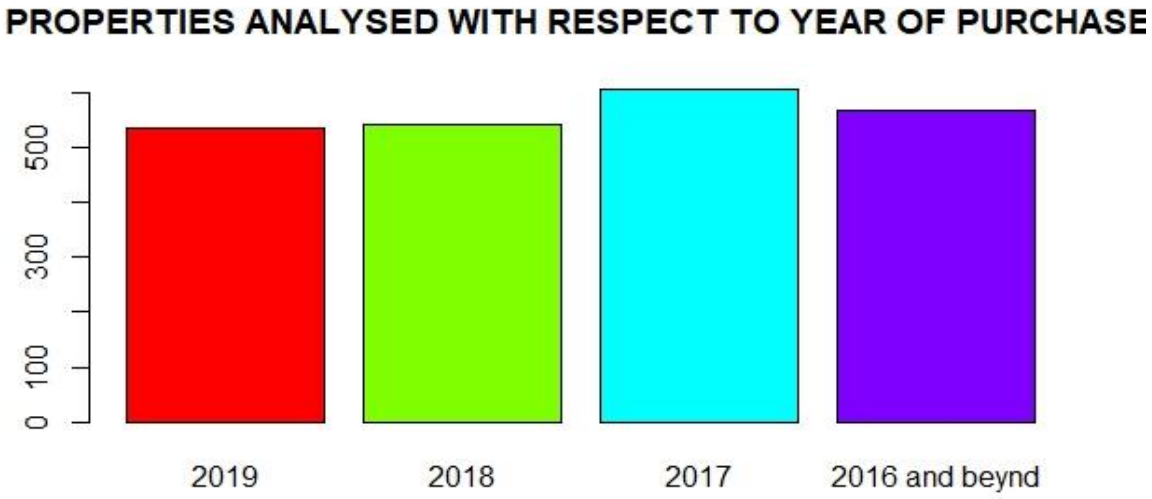


The table below gives the statistical description, wherein we find the 2 BHK is the Model Property, Gurgaon the Model Location and 2017 is the Model Year, the Model numbers of years the people want their property to be leased is (11-15) and the same is depicted the visuals below.

DESCRIPTION OF PROPERTIES ANALYSED LOCATION-WISE

	1BHK	2BHK	3BHK	4BHK
GURGAON	154	174	126	164
DELHI	137	131	126	136
NIODA	142	137	123	146
FARIDABAD	116	141	143	152

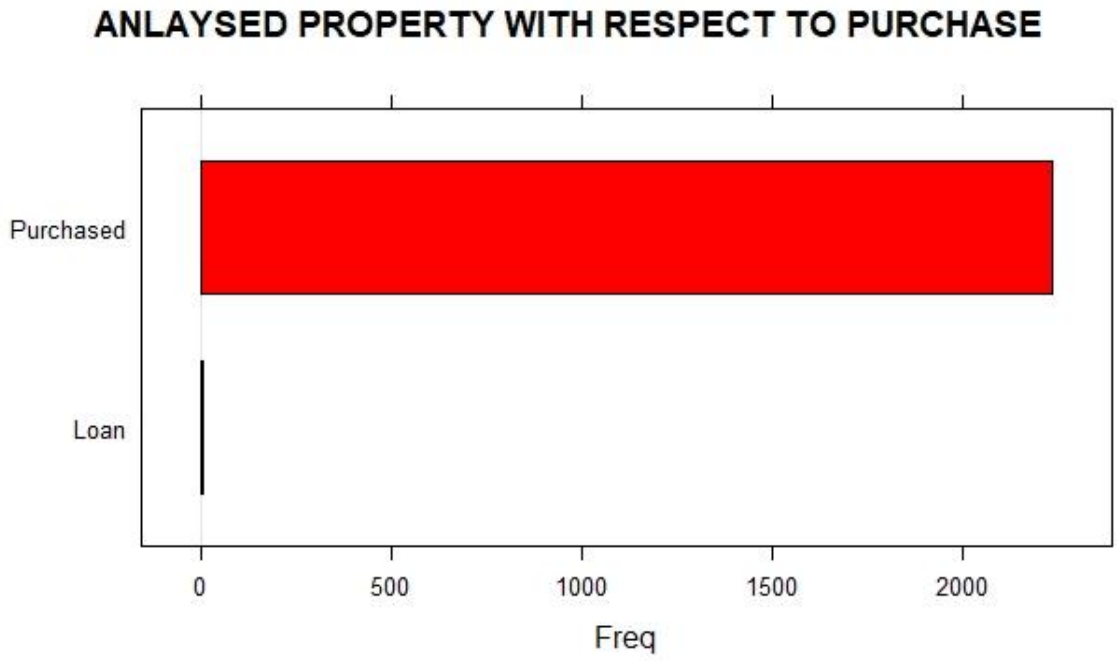

DISTRIBUTION OF PROPERTIES ANALYSED
0
100
300
500
1BHK
2BHK
3BHK
4BHK


PROPERTIES ANALYSED WITH RESPECT TO YEAR OF PURCHASE
0
100
300
500
2019
2018
2017
2016 and beynd

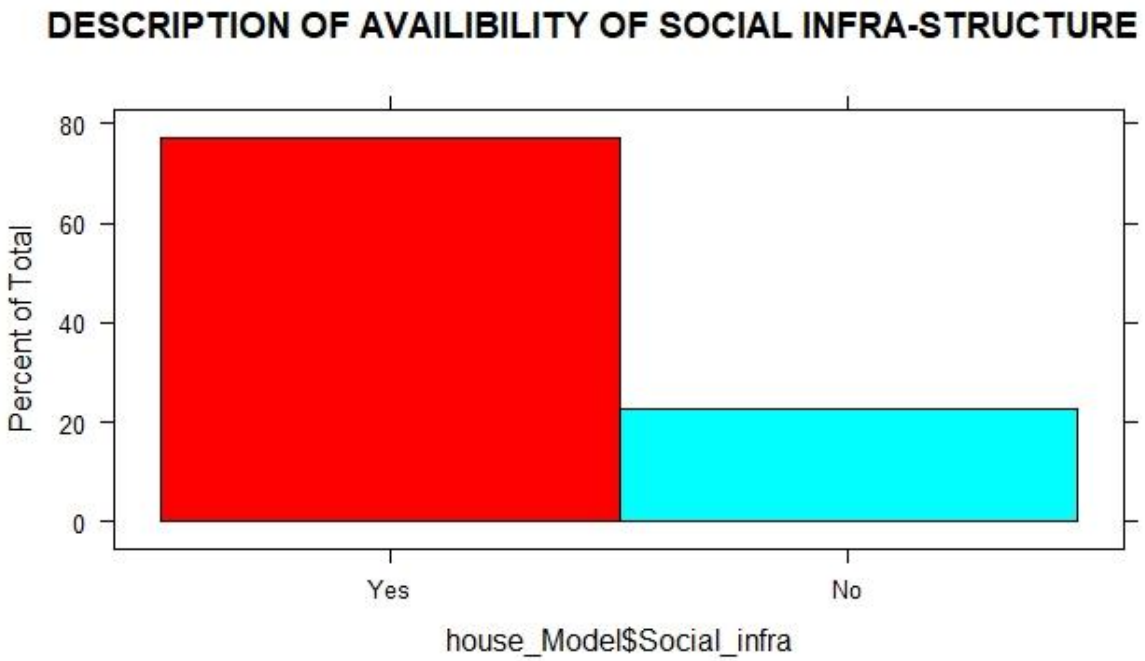
The table below shows that the property analyzed is mostly purchased.

DESCRIPTION OF THE PROPERTIES WITH RESPECT TO PURCHASE.

Loan	Purchased
9	2239



Further, it is observed that the basic social infrastructure such as school, college, market place, hospital, pharmacy, playground, etc.; it is found that majority of them had this in their vicinity.



The table below shows the statistical description of the availability of social infra-structure in the different NCR regions, wherein we find that Delhi and Gurgoan is better when compared with Noida and Faridabad.

DESCRIPTION OF AVAILABILITY OF SOCIAL INFRA-STRUCTURE LOCATION WISE

	Yes	No
GURGAON	466	152
DELHI	400	130
NIODA	434	114
FARIDABAD	437	115

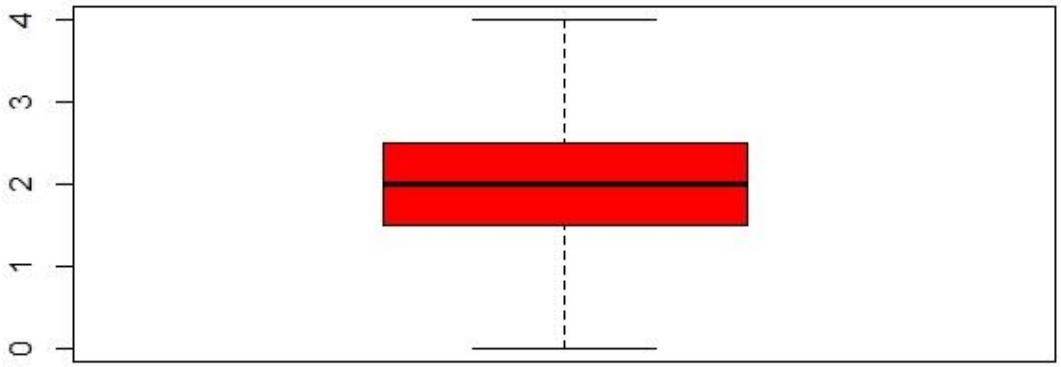
DESCRIPTION OF SOCIAL INFRA-STRUCTURE LOCATION-WISE

The table below is the statistical descriptors of quantitative variables, and the box- plots of it have been visualized below. Also, box-plots are necessary for analysis and modelling. The boxplots below shows that it does not have any outliers, a necessary condition for Regression Models.

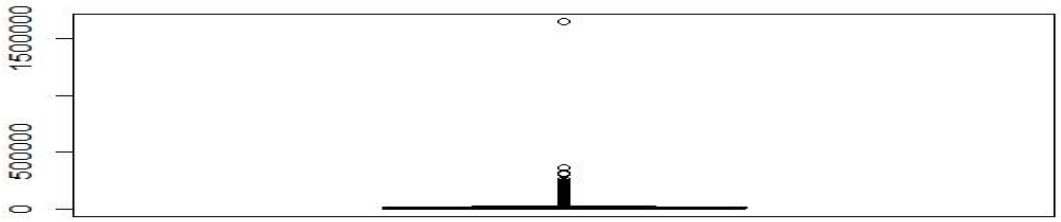
DESCRIPTIVE STATISTICS OFNUMERICAL VARIABLES

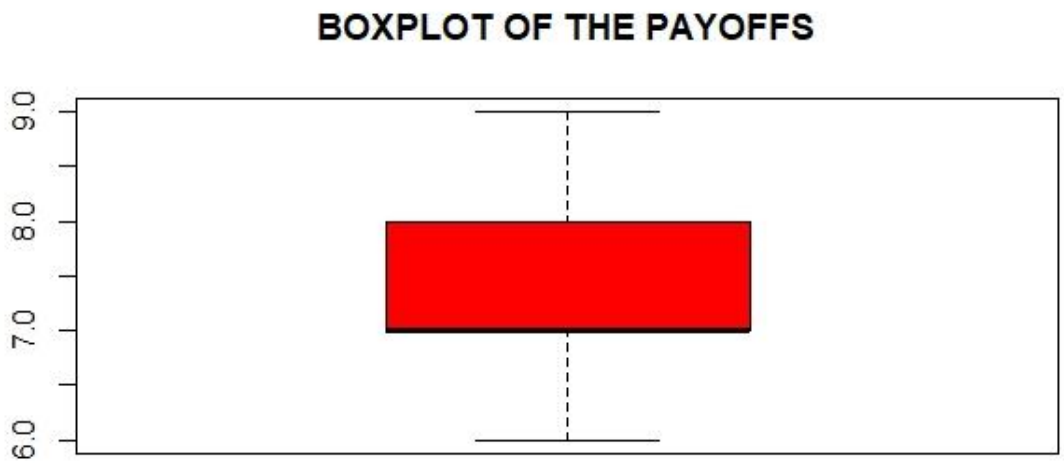
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Price	1	2248	2.08	0.76	2.00	2.03	0.74	0.00	6.0	6.00	0.49	0.81	0.02
Carpet_area	2	2248	15034.06	43958.68	8040.00	8859.05	4091.98	649.00	1651359.0	1650710.00	24.46	858.97	927.14
Payoff	3	2248	7.62	1.18	7.00	7.54	1.48	3.00	12.0	9.00	0.81	1.18	0.02
Main_Amount	4	2248	1769.91	817.23	1555.00	1662.00	644.93	380.00	6720.0	6340.00	1.56	3.61	17.24
Rental_increament	5	2248	4.91	1.50	4.92	4.92	1.88	2.25	7.5	5.25	-0.03	-1.17	0.03

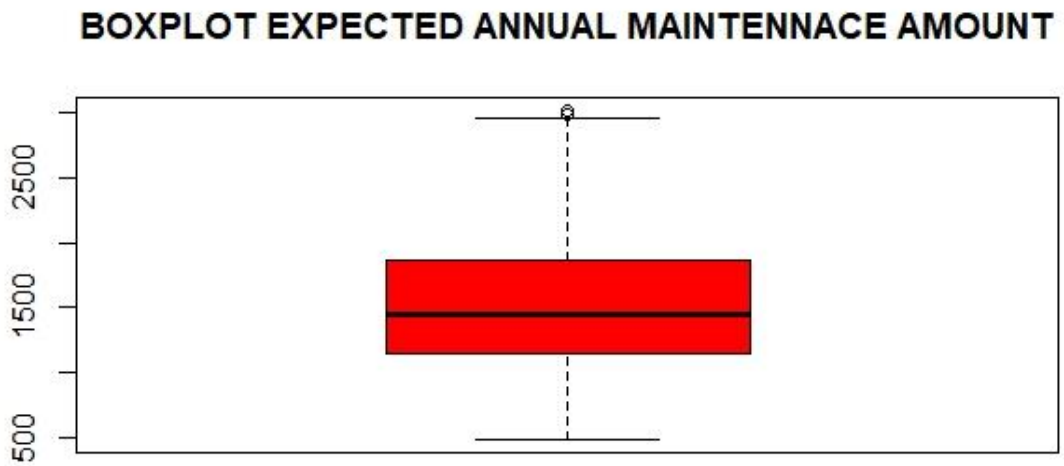
BOXPLOT OF INDICATIVE PRICE

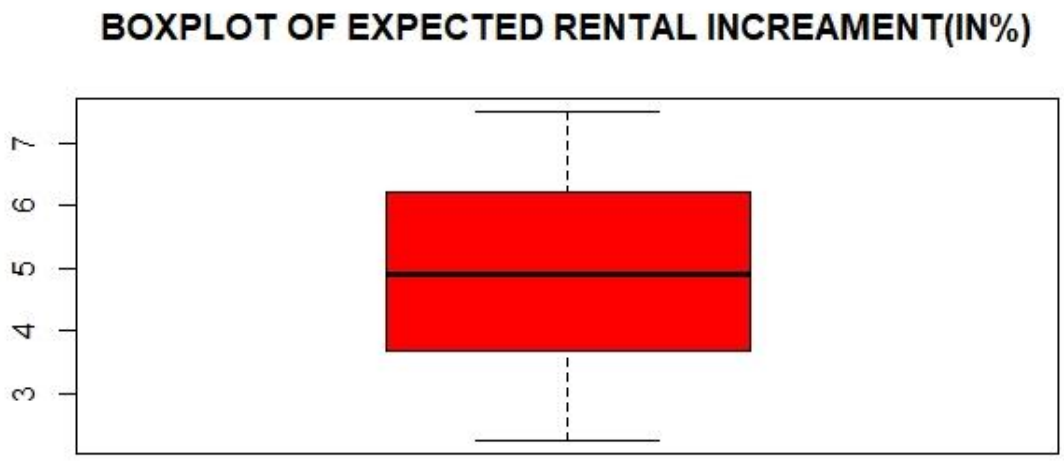


BOX PLOT OF CARPET AREA OF THE PROPERTY




BOXPLOT OF THE PAYOFFS
9.0
8.0
7.0
6.0


BOXPLOT EXPECTED ANNUAL MAINTENNACE AMOUNT
2500
1500
500



4.09 REGRESSION MODEL FOR HOUSE PRICE BASED ON ASSET BASED MODEL

The regression Model here is to investigate the relationship of the Price (Expected / Indicated) with respect to some of the important variables as identified in the Literature review

SIMPLE MULTIPLE LIEAR REGRESSION MODEL

```
Call:
lm(formula = Price ~ ., data = house_Model)

Residuals:
     Min       1Q   Median       3Q      Max
-2.60560 -0.35915 -0.00885  0.31148  1.96307

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -6.863e-02  3.816e-01  -0.180    0.857
Location        4.297e-04  1.026e-02   0.042    0.967
Prop_type      -3.698e-03  1.029e-02  -0.359    0.719
Carpet_area    -1.658e-05  3.380e-06  -4.905 1.02e-06 ***
Year           -8.281e-03  1.051e-02  -0.788    0.431
Lease_Per      -1.298e-02  1.050e-02  -1.236    0.217
Prop_purchase  -9.569e-02  1.799e-01  -0.532    0.595
Type_Prop       1.320e-03  1.031e-02   0.128    0.898
```

```
Payoff            2.298e-01  1.620e-02  14.187  < 2e-16 ***
Main_Amount       4.692e-04  2.667e-05  17.595  < 2e-16 ***
Rental_increment  4.244e-03  7.761e-03   0.547    0.585
Social_infra     -4.435e-02  2.768e-02  -1.602    0.109
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5062 on 1887 degrees of freedom
Multiple R-squared:  0.451,    Adjusted R-squared:  0.4478
F-statistic: 140.9 on 11 and 1887 DF,  p-value: < 2.2e-16
```

ANALYSIS OF VARIANCE

```
Response: Price
                   Df Sum Sq Mean Sq   F value  Pr(>F)
Location            1   0.34   0.343    1.3384 0.24747
Prop_type           1   0.04   0.041    0.1582 0.69089
Carpet_area         1   0.99   0.988    3.8564 0.04970 *
Year                1   1.60   1.597    6.2332 0.01262 *
Lease_Per           1   0.36   0.355    1.3870 0.23906
Prop_purchase       1   0.00   0.002    0.0075 0.93084
Type_Prop           1   0.00   0.001    0.0024 0.96101
Payoff              1 313.55 313.551 1223.5772 < 2e-16 ***
Main_Amount         1  79.70  79.700  311.0150 < 2e-16 ***
Rental_increment    1   0.06   0.059    0.2287 0.63253
Social_infra        1   0.66   0.658    2.5670 0.10928
Residuals        1887 483.56   0.256
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

The Regression Model by Simple Method as we have obtained is as follows:

***Y* (Price-Indicative) = -0.324+0.004(Location)-0.004(Property type)-0.0005(Carpet Area)-0.01(Year of Purchase)-0.010 (Period of Lease) +0.001(Category of Property) +0.22(Payoffs) +0.002(Maintenance Amount)-0.001(Percentage of Rental Increment) +0.03(Availability of Social Infra-structure) +0.005(Property Purchased: Loan/ Cash)**

STATISTICAL PERSPECTIVE OF STEP-WISE MULTIPLE LINEAR REGRESSION

When selecting the model for the multiple linear regression analysis, another important consideration is the model fit. Adding independent variables to a multiple linear regression model will always increase the amount of explained variance in the dependent variable (typically expressed as R^2). Therefore, adding too many independent variables without any theoretical justification may result

in an over-fit model, in poor out-of-sample forecasts. Many formal statistical criteria have been proposed to formalize the trade-off between fit and parsimony. The best-known and most popular one is the Akaike Information Criterion (AIC) .The function step () uses the Akaike Information Criterion to perform Step-wise Regression Model search. However, the search uses backward elimination by default.

MULTIPLE REGRESSION LINEAR MODEL BY STEP-WISE METHOD

```
Start:  AIC=-2573.66
Price ~ Location + Prop_type + Carpet_area + Year + Lease_Per +
    Prop_purchase + Type_Prop + Payoff + Main_Amount + Rental_increment +
    Social_infra

                    Df Sum of Sq    RSS     AIC
- Location           1     0.000 483.56 -2575.7
- Type_Prop          1     0.004 483.56 -2575.7
- Prop_type          1     0.033 483.59 -2575.5
- Prop_purchase      1     0.073 483.63 -2575.4
- Rental_increment   1     0.077 483.63 -2575.4
- Year               1     0.159 483.72 -2575.0
- Lease_Per          1     0.391 483.95 -2574.1
<none>                           483.56 -2573.7
- Social_infra       1     0.658 484.22 -2573.1
- Carpet_area        1     6.165 489.72 -2551.6
- Payoff             1    51.578 535.14 -2383.2
- Main_Amount        1    79.331 562.89 -2287.2

Step:  AIC=-2575.66
Price ~ Prop_type + Carpet_area + Year + Lease_Per + Prop_purchase +
    Type_Prop + Payoff + Main_Amount + Rental_increment + Social_infra

                    Df Sum of Sq    RSS     AIC
- Type_Prop          1     0.004 483.56 -2577.7
- Prop_type          1     0.033 483.59 -2577.5
- Prop_purchase      1     0.073 483.63 -2577.4
- Rental_increment   1     0.077 483.64 -2577.4
- Year               1     0.159 483.72 -2577.0
- Lease_Per          1     0.392 483.95 -2576.1
<none>                           483.56 -2575.7
- Social_infra       1     0.660 484.22 -2575.1
+ Location           1     0.000 483.56 -2573.7
- Carpet_area        1     6.164 489.72 -2553.6
- Payoff             1    51.634 535.19 -2385.0
- Main_Amount        1    79.334 562.89 -2289.2

Step:  AIC=-2577.65
Price ~ Prop_type + Carpet_area + Year + Lease_Per + Prop_purchase +
    Payoff + Main_Amount + Rental_increment + Social_infra

                    Df Sum of Sq    RSS     AIC
```

```
- Prop_type          1     0.033 483.60 -2579.5
- Prop_purchase      1     0.072 483.63 -2579.4
- Rental_increment   1     0.076 483.64 -2579.3
- Year               1     0.161 483.72 -2579.0
- Lease_Per          1     0.393 483.96 -2578.1
<none>                           483.56 -2577.7
- Social_infra       1     0.660 484.22 -2577.1
+ Type_Prop          1     0.004 483.56 -2575.7
+ Location           1     0.000 483.56 -2575.7
- Carpet_area        1     6.179 489.74 -2555.5
- Payoff             1    51.688 535.25 -2386.8
- Main_Amount        1    79.487 563.05 -2290.6

Step:  AIC=-2579.51
Price ~ Carpet_area + Year + Lease_Per + Prop_purchase + Payoff +
    Main_Amount + Rental_increment + Social_infra

                    Df Sum of Sq    RSS     AIC
- Prop_purchase      1     0.073 483.67 -2581.2
- Rental_increment   1     0.079 483.67 -2581.2
- Year               1     0.159 483.76 -2580.9
- Lease_Per          1     0.395 483.99 -2580.0
<none>                           483.60 -2579.5
- Social_infra       1     0.654 484.25 -2578.9
+ Prop_type          1     0.033 483.56 -2577.7
+ Type_Prop          1     0.004 483.59 -2577.5
+ Location           1     0.000 483.60 -2577.5
- Carpet_area        1     6.197 489.79 -2557.3
- Payoff             1    51.775 535.37 -2388.4
- Main_Amount        1    79.454 563.05 -2292.6

Step:  AIC=-2581.23
Price ~ Carpet_area + Year + Lease_Per + Payoff + Main_Amount +
    Rental_increment + Social_infra

                    Df Sum of Sq    RSS     AIC
- Rental_increment   1     0.080 483.75 -2582.9
- Year               1     0.169 483.84 -2582.6
- Lease_Per          1     0.389 484.06 -2581.7
<none>                           483.67 -2581.2
- Social_infra       1     0.645 484.31 -2580.7
+ Prop_purchase      1     0.073 483.60 -2579.5
+ Prop_type          1     0.034 483.63 -2579.4
+ Type_Prop          1     0.003 483.67 -2579.2
+ Location           1     0.001 483.67 -2579.2
- Carpet_area        1     6.163 489.83 -2559.2
- Payoff             1    51.784 535.45 -2390.1
- Main_Amount        1    79.405 563.07 -2294.6

Step:  AIC=-2582.91
Price ~ Carpet_area + Year + Lease_Per + Payoff + Main_Amount +
    Social_infra

                    Df Sum of Sq    RSS     AIC
- Year               1     0.170 483.92 -2584.2
- Lease_Per          1     0.376 484.13 -2583.4
<none>                           483.75 -2582.9
- Social_infra       1     0.627 484.38 -2582.5
+ Rental_increment   1     0.080 483.67 -2581.2
+ Prop_purchase      1     0.074 483.67 -2581.2
+ Prop_type          1     0.037 483.71 -2581.1
+ Type_Prop          1     0.003 483.75 -2580.9
+ Location           1     0.000 483.75 -2580.9
```

```
- Carpet_area       1     6.170 489.92 -2560.8
- Payoff            1    51.975 535.72 -2391.1
- Main_Amount       1    79.326 563.07 -2296.6

Step:  AIC=-2584.25
Price ~ Carpet_area + Lease_Per + Payoff + Main_Amount + Social_infra

                    Df Sum of Sq    RSS     AIC
- Lease_Per          1     0.366 484.28 -2584.8
<none>                         483.92 -2584.2
- Social_infra       1     0.617 484.54 -2583.8
+ Year               1     0.170 483.75 -2582.9
+ Prop_purchase      1     0.084 483.83 -2582.6
+ Rental_increment   1     0.081 483.84 -2582.6
+ Prop_type          1     0.036 483.88 -2582.4
+ Type_Prop          1     0.004 483.91 -2582.3
+ Location           1     0.001 483.92 -2582.2
- Carpet_area        1     6.170 490.09 -2562.2
- Payoff             1    52.073 535.99 -2392.2
- Main_Amount        1    79.524 563.44 -2297.3

Step:  AIC=-2584.81
Price ~ Carpet_area + Payoff + Main_Amount + Social_infra

                    Df Sum of Sq    RSS     AIC
<none>                         484.28 -2584.8
- Social_infra       1     0.625 484.91 -2584.4
+ Lease_Per          1     0.366 483.92 -2584.2
+ Year               1     0.159 484.13 -2583.4
+ Prop_purchase      1     0.077 484.21 -2583.1
+ Rental_increment   1     0.068 484.22 -2583.1
+ Prop_type          1     0.038 484.25 -2583.0
+ Type_Prop          1     0.004 484.28 -2582.8
+ Location           1     0.001 484.28 -2582.8
- Carpet_area        1     6.168 490.45 -2562.8
- Payoff             1    52.450 536.73 -2391.5
- Main_Amount        1    79.242 563.53 -2299.0
> summary(Model2)

Call:
lm(formula = Price ~ Carpet_area + Payoff + Main_Amount + Social_infra,
    data = house_Model)

Residuals:
     Min       1Q   Median       3Q      Max
-2.62259 -0.35794 -0.00789  0.31573  1.94574

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.062e-01  1.057e-01  -2.897  0.00381 **
Carpet_area  -1.656e-05  3.372e-06  -4.911 9.82e-07 ***
Payoff        2.311e-01  1.613e-02  14.322  < 2e-16 ***
Main_Amount   4.678e-04  2.658e-05  17.604  < 2e-16 ***
Social_infra -4.313e-02  2.759e-02  -1.563  0.11815
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5057 on 1894 degrees of freedom
Multiple R-squared:  0.4502,   Adjusted R-squared:  0.449
F-statistic: 387.7 on 4 and 1894 DF,  p-value: < 2.2e-16
```

ANALYSIS OF VARIANCE TABLE

```
Response: Price
               Df Sum Sq Mean Sq   F value  Pr(>F)
Carpet_area     1   0.99   0.991    3.8758 0.04913 *
Payoff          1 315.27 315.269 1232.9928 < 2e-16 ***
Main_Amount     1  79.68  79.683  311.6352 < 2e-16 ***
Social_infra    1   0.62   0.625    2.4438 0.11815
Residuals    1894 484.28   0.256
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The Regression Model by Stepwise Method as we have obtained is as follows:

***Y* (Price-Indicative) = -0.3195- -0.0000005475(Carpet Area)-0.01431(Social Infra-structure) +0.2324(Payoffs)+0.0004(Maintenance Amount)**

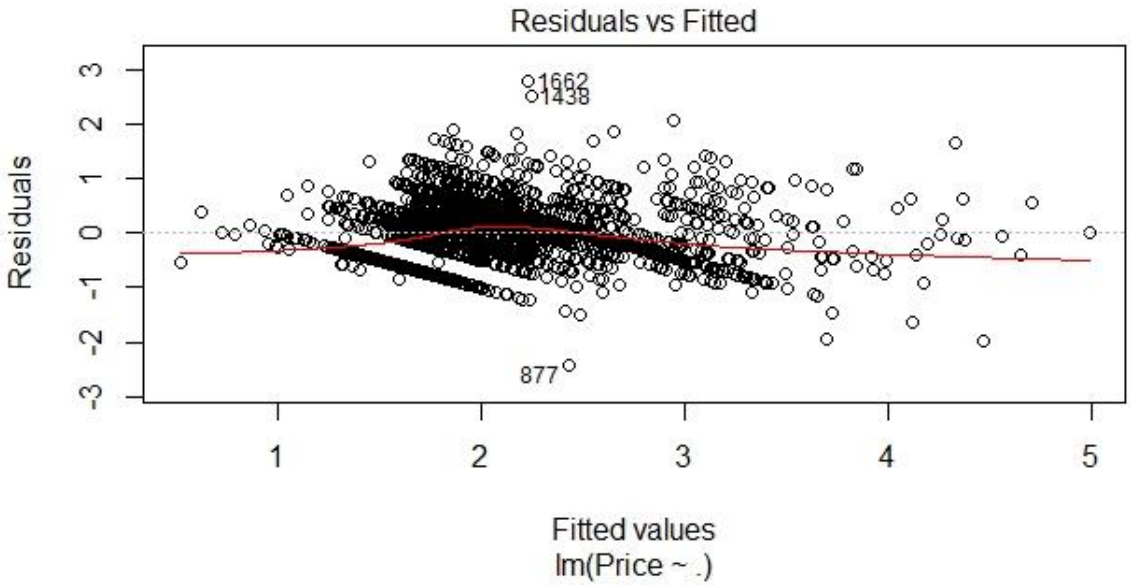
4.10 COMPARISON OF THE REGRESSION MODELS

Comparing the two regression models, namely the simple Multiple Regression and the Stepwise Regression Model, we see that the removal of the other predictors has not significantly affected the explanatory ability of the model.

Statistical Parameters	Model 1(SIMPLE MULTIPLE LINEAR REGRESSION)	Model 2(STEP-WISE MULTIPLE LINEAR REGRESSION)	CRITERION
The residuals (i.e., the errors) of the fit of the linear model to the used data should be as small as possible.	0.5062	0.5057	Smaller the better.
Diagnostics information outputted by R. The adjusted coefficient is more demanding as it takes into account the number of parameters of the regression model.	0.4502	0.451	Higher the better.
R2 coefficients (multiple and adjusted). The adjusted coefficient is more demanding as it takes into account the number of parameters of the regression model	0.449	0.4478	Higher the better.

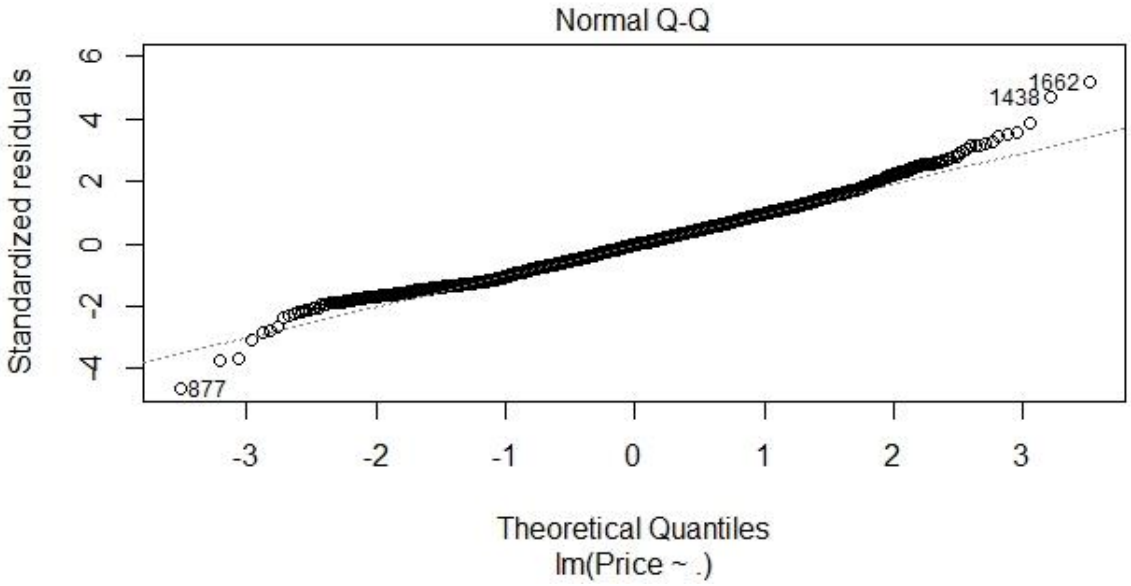
RESIDUAL VS FITTED VALUES PLOT

Ideally, this plot shouldn't show any pattern. As we see no shape (curve, U shape), it suggests linearity in the dataset.



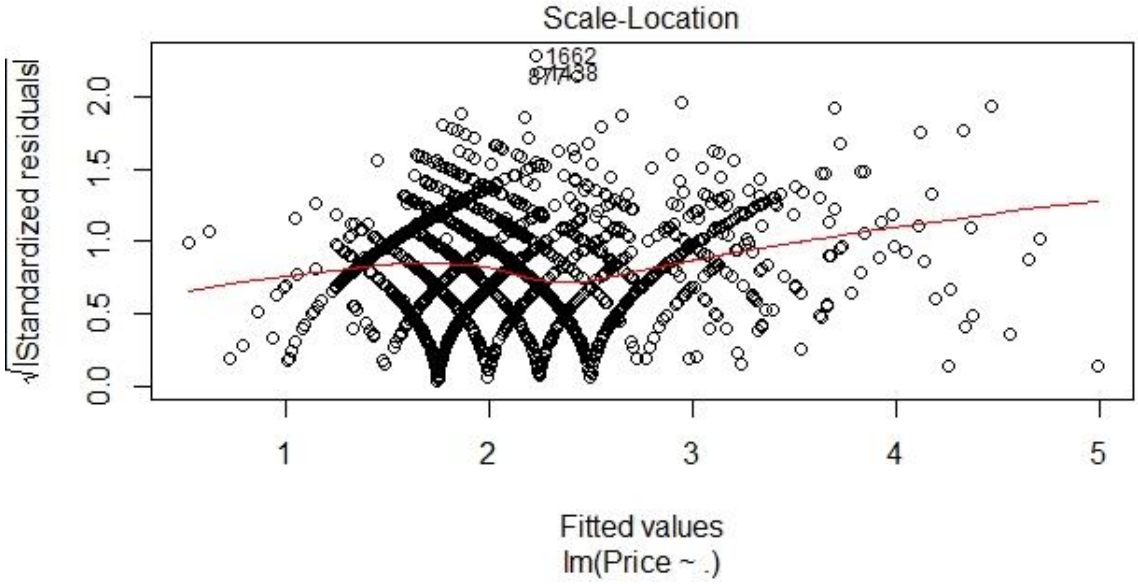
NORMALITY Q-Q PLOT

As the name suggests, this plot is used to determine the normal distribution of errors. It uses standardized values of residuals. Ideally, this plot should show a straight line, which is showing in our case, and therefore we can conclude that the error follows normal distribution.



SCALE-LOCATION

It's also called Spread-Location plot. This plot shows if residuals are spread equally along with the ranges of predictors. This is how we can check the assumption of equal variance (homoscedasticity). Here we see a horizontal line with equally (randomly) spread points.

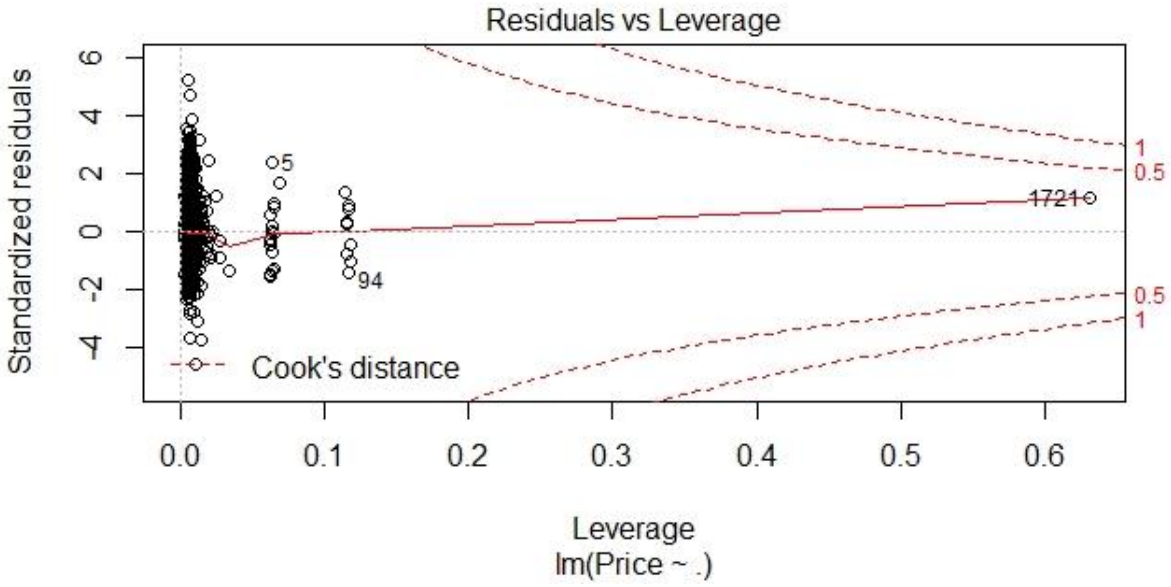


RESIDUALS VS LEVERAGE

This plot helps us to find influential cases (i.e., subjects) if any. Not all outliers are influential in linear regression analysis (whatever outliers mean). Even though data have extreme values, they might not be influential to determine a regression line. That means, the results wouldn't be much different if we either include or exclude them from analysis. They follow the trend in the majority of cases and they don't really matter; they are not influential.

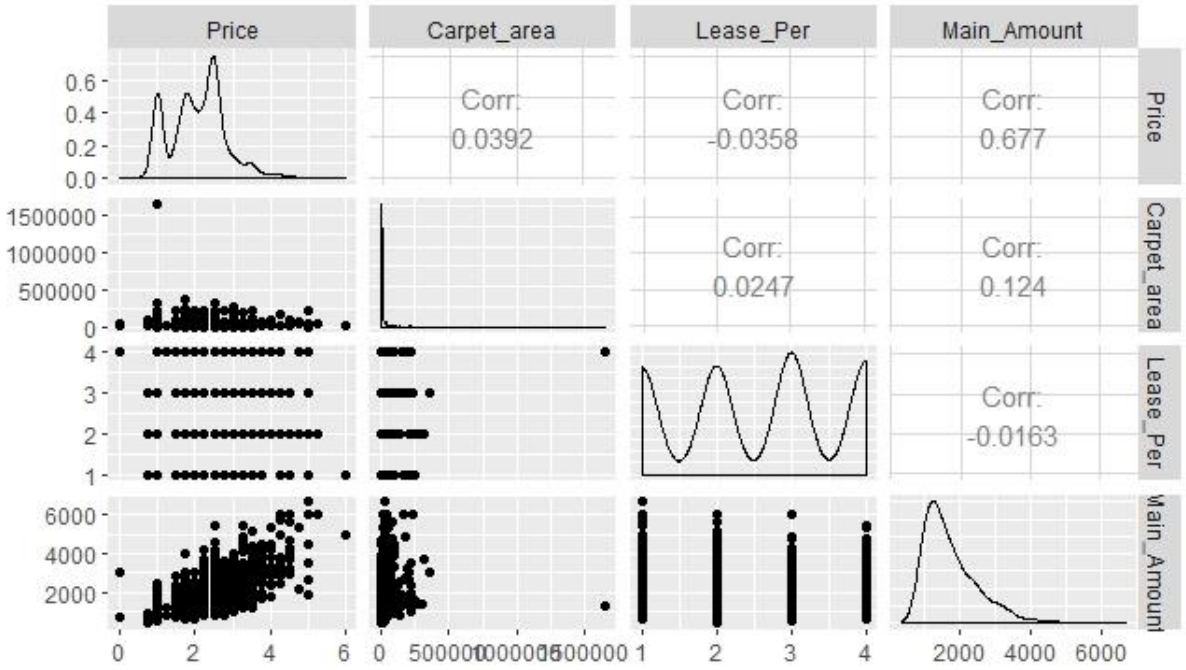
Unlike the other plots, this time patterns are not relevant. We watch out for outlying values at the upper right corner or at the lower right corner. Those spots are the places where cases can be influential against a regression line. Look for cases outside of a dashed line, Cook's distance. When cases are outside of the Cook's distance (meaning they have high Cook's distance scores), the cases are

influential to the regression results. The regression results will be altered if we exclude those cases.

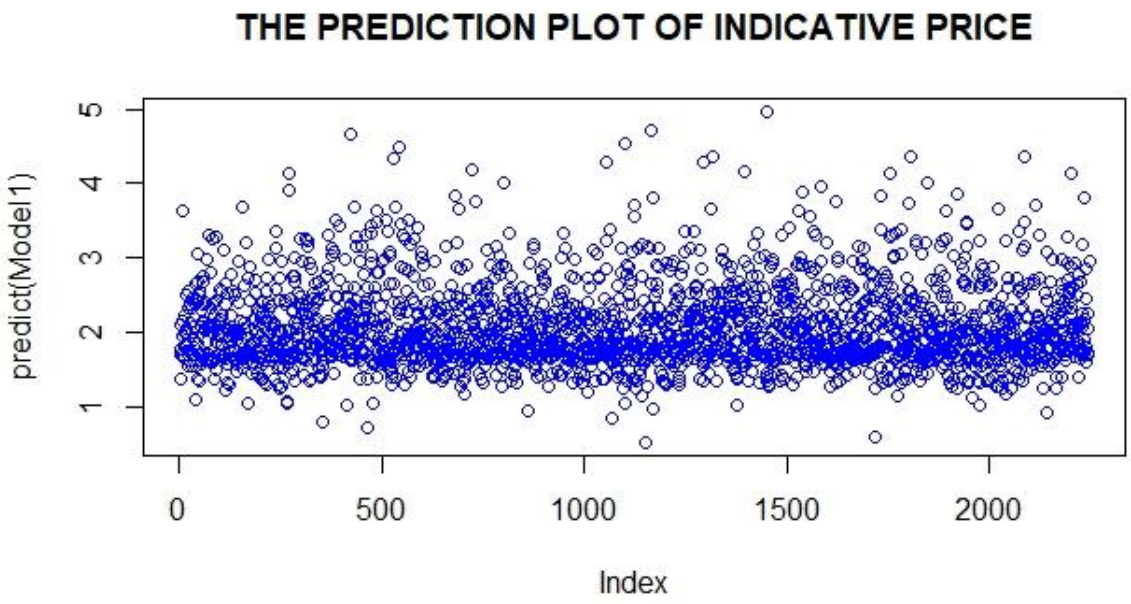


4.11 CORRELATION PLOT TO EVALUATE THE HOUSE PRICE ON ASSET BASED MODEL

To suggest the Asset Based Modeling, it is essential to initial feel for the relationships between the independent variables- and, in particular, Carpet Area (size of the property), Leasing Period, and the Annual Maintenance Cost. It is interesting to observe the correlation plots produced by plotting it clearly emerges that the Price is assumed as Financial Instrument, generating recurring returns seems to be mostly correlated to the Annual Maintenance charges, and has to be negatively correlated with the Leasing Period (lower the charges as the leasing period increases), and yes the Price has to be proportionate to the size (carpet area). However, the regression analysis is to investigate the significance and magnitude of this possible effect.



From the above plots, we understand that the robustness of the model assumptions is met, and therefore, it is safe to assume that the obtained model is relatively sound.



The table below gives the predicted values of the Model in terms of crore.

PREDICTED VALUES OF THE MODEL

1	2	3	4	5	6	7	8
1.721112	2.065244	1.345877	1.686719	2.140916	3.592271	1.908070	1.673600
9	10						
1.671053	1.959944						

4.12 Few Basic terms of Asset Based Modelling

Asset Definition: An asset is any resource owned by the business. Anything tangible or intangible that can be owned or controlled to produce value and that is held by an entity to produce positive economic value is an asset

Asset-Based Modeling: The principle underlying an asset-based approach to welfare is that, rather than relying on state-managed social transfers to counter the risks of poverty, individuals accept greater responsibility for their own welfare needs by investing in financial products and property assets which augment in value over time.

Cochrane Asset Based Model: An investor who thinks about how much to save and consume, and what portfolio of assets to hold. The most basic pricing

equation comes from the first-order conditions to that problem and says that price should be the expected discounted payoff, using the investor's marginal utility to discount the payoff. Asset pricing theory all stems from one simple concept: price equals expected discounted payoff (rent). The model assumes an association among user cost UC_t, the house selling prices, Pt, and the rents R_t $R_t/ P_t=UC_t$

If the rent-to-price ratio exceeds the user-cost rate, renting becomes less attractive than ownership. When the rents equal the costs of owning a state of equilibrium is reached. This rent-to-price relationship fits well in markets where the rental regulation is absent or limited.

Chapter-5

5.0 Results

This study investigates the value of housing, a specific group of investment assets, in particular, the primary valuation drivers of the NCR housing stocks, by employing the Cochrane Econometrics Model.

The study dataset has 2 BHK as the Model Property, Gurgaon as the Model Location, 2017 as the Model Year, the Model numbers of years the people want their property to be leased is (11-15), majority purchased are in cash, and majority have the social infra-structure in their vicinity, which in general comprise of school, hospital, pharmacy, Children Park, etc.

The statistical perspective was to evaluate the relationship of Price (customer willing to pay) and its relationship with payoff (monthly rental income), and the deductions such as annual maintenance cost. To begin with, we have outlined twelve (12) variables. Statistically speaking too many independent variables without any theoretical justification may result in an over-fit model, in poor out-of-sample forecasts. We, therefore, have resorted to Stepwise Multiple Linear Regression, which is a formal statistical criterion which has been proposed to formalize the trade-off between fit and parsimony. The best-known and most popular one is the Akaike Information Criterion (AIC). The function stepwise regression is used using the Akaike Information Criterion to search best performing Regression Model. However, the search uses backward elimination by default.

- ***The Regression Model by Simple Method as we have obtained is as follows:***

***Y* (Price-Indicative) = -0.324+0.004(Location)-0.004(Property type)-0.0005(Carpet Area)-0.01(Year of Purchase)-0.010 (Period of Lease)**

+0.001(Category of Property) +0.22(Payoffs) +0.002(Maintenance Amount)-0.001(Percentage of Rental Increment) +0.03(Availability of Social Infra-structure) +0.005(Property Purchased: Loan/ Cash)

- ***The Regression Model by Stepwise Method as we have obtained is as follows:***

***Y* (Price-Indicative) = -0.3195- -0.0000005475(Carpet Area)-0.01431(Year of Purchase) +0.2324(Payoffs)+0.0004(Maintenance Amount)**

After correlation plots has been produced, Price has clearly emerged as financial instruments. And generating recurring returns seems to be most correlated to the Annual Maintenance charges and is negatively correlated with the Leasing Period (lower the charges as the leasing period increases). Price(dependent-variable) is proportionate to the size carpet area.

Further, this study shows that there has been a significant relationship between housing prices and the expected payoff (rent) in NCR housing stock. It also appears that there is a positive linear relationship between housing prices and year of purchase, the size of the property. Understanding the interaction between housing prices and the rental payoff is of importance since the interdependence is likely to augment boom-bust cycles in the housing economy

Chapter-6

6.0 Conclusion and Future Scope

The asset-based model will help in understanding the primary valuation drivers of housing prices and will make market participants aware of the size of their risk exposure and can help them to detect early signals of the possibility of investment opportunities. Furthermore, policymakers can use information about the underlying valuation drivers of the house prices to stabilize the market. Direct housing investment has some typical disadvantage such as poor liquidity and difficult to evaluate.

one of the several marketing issue individuals in real estate business generally commit is that they didn't remain up to date and they use one-prong strategy (via referrals). With the help of this study, it can help marketing professionals to remain up to date with latest trends in marketing field and they can apply multi-prong strategies like website designing with accurate data information, holding open house etc.

As, from the perspective of homebuyers with the availability of accurate information and data, it becomes quite easy to analyze, compare and take the effective and well-formed decision.

Lastly, we would like to end by quoting Kendall and Stuart, "***A statistical relationship, however strong and however suggestive, can never establish causal connection: our ideas of causation must come from outside statistics, ultimately from some theory or other."***

.

Appendix

Model R Script Documentation

```
#1.libraries used
library(psych)
library(dplyr)
library(corrr)
library(corrplot)
library(RColorBrewer)
library(PerformanceAnalytics)
library(GGally)
#2.Data acquisition
#Reading the data set in csv files
house1<-read.csv(file.choose())
#3.Cleaning the dataset from NA if any
house<-na.omit(house1)
#4. Listing the dataset
house<-as.data.frame(house)
house
#5. cleaning the dataset from NA, if any
house<-na.omit(house)
#6. Data preparation and exploration
str(house)
summary(house)
names(house)
head(house, n=5)
tail(house, n=5)
#7. Splitting the data set into study population and analysis dataset
house_n<-dplyr::select(house,1:3)
house_n<-as.data.frame(house_n)
house_n
house_Model<-dplyr::select(house,-1:-3)
house_Model<-as.data.frame(house_Model)
house_Model
#8. Checking the study dataset
str(house_n)
names(house_n)
psych::headTail(house_n)
#9. Visualization of variables
plot(house_n$Name_B, col="Blue", main="Real estate agencies from where data was collected")
pie(1:4,labels=house_n$Proff,col=c("red","orange","blue","green"),main="DISTRIBUTION OF
PROFESSIONS OF STUDY POPULATION")
#10.converting the variables into the factors
house_n$Name_B<-factor(house_n$Name_B)
house_n$Proff<-factor(house_n$Proff, levels=1:4, labels=c("Business","Private-Service","Government-
Service","NRI"))
house_n$Proff
#11. Summary of study population
table(house_n$Name_B)
table<-ftable(house_n$Name_B, house_n$Proff)
table
#12. Data Exploration of Study Variables Dataset
```

```
DataExplorer::introduce(house_Model)
#13. Converting all the Study variables as in the Data Dictionary
#Checking the Analysis Dataset
str(house_Model)
names(house_Model)
psych::headTail(house_Model)
#14.Descriptive Statistics of Variables of the Study Dataset
plot(house_Model,col="red")
#12. Data exploration of analysis variables datsets
# converting all the variables as in data dictionary
house_Model$Price<-as.numeric(house_Model$Price)
house_Model$Price
house_Model$Location<-factor(house_Model$Location, levels=1:4, labels=
c("GURGAON","DELHI","NIODA","FARIDABAD"))
house_Model$Location
barplot(table(house_Model$Location), col=rainbow(4), main="Property location wise selected for study")
house_Model$Prop_type<-factor(house_Model$Prop_type, level = 1:4, labels
=c("1BHK","2BHK","3BHK","4BHK"))
house_Model$Prop_type
barplot(table(house_Model$Prop_type), col=rainbow(4), main="Distribution of property analyzed")
house_Model$Carpet_area<as.numeric(house_Model$Carpet_area)
house_Model$Carpet_area
house_Model$Year<-factor(house_Model$Year,levels=1:4,labels=c("2019","2018","2017","2016 and before"))
house_Model$Year
barplot(table(house_Model$Year), col=rainbow(4), main="Property Analyzed with respect to year of purchase")
house_Model$Lease_Per<-factor(house_Model$Lease_Per,levels=1:4, labels=c("0-5","6-10","11-15","15 and
above"))
house_Model$Lease_Per
barplot(table(house_Model$Lease_Per), col=rainbow(4), main="Expected year of leasing the property")
house_Model$Prop_purchase<-factor(house_Model$Prop_purchase,levels=1:2,labels=c("Loan","cash
Purchased"))
house_Model$Prop_purchase
table(house_Model$Prop_purchase)
lattice::barchart(house_Model$Prop_purchase,col=rainbow(2), main="ANLAYSED PROPERTY WITH
RESPECT TO PURCHASE")
lattice::histogram(house_Model$Social_infra,col=rainbow(2),main="DESCRIPTION OF AVAILIBILITY OF
SOCIAL INFRA-STRUCTURE")
house_Model$Payoff<-as.numeric(house_Model$Payoff)
house_Model$Payoff
house_Model$Rental_increment<-as.numeric(house_Model$Rental_increment)
house_Model$Rental_increment
house_Model$Main_Amount<-as.numeric(house_Model$Main_Amount)
house_Model$Main_Amount
house_Model$Social_infra<-sample(c(house_Model$Social_infra),2248, replace=TRUE)
house_Model$Social_infra
barplot(table(house_Model$Social_infra), col=rainbow(2), xlab="house_Model$Social_infra", ylab= "Percent
of Total", main="Description of availability of social infrastructure")
#13. checking the analysis dataset
str(house_Model)
names(house_Model)
psych::headTail(house_Model)
#14. Descriptive statistics of the variables of the analysis dataset
plot(house_Model, col="red")
psych::describe(house_Model)
```

```
#15. Descriptive statistics of the continuous variable and plotting the variables
summary(house_Model$Price)
boxplot(house_Model$Price, col="red", xlab= "Price", main="BOX PLOT OF THE IDICATIVE PRICE")
summary(house_Model$Carpet_area)
boxplot(house_Model$Carpet_area, main= "Boxplot of Carpet Area of the property")
summary(house_Model$Main_Amount)
boxplot(house_Model$Main_Amount)
summary(house_Model$Rental_increment)
boxplot(house_Model$Rental_increment)
summary(house_Model$Payoff)
boxplot(house_Model$Payoff, col="Red", main="BOXPLOT OF PAYOFFS")
summary(house_Model$Main_Amount)
boxplot(house_Model$Payoff, col="Red", main="BOXPLOT EXPECTED ANNUAL MAINTENANCE
AMOUNT")
summary(house_Model$Rental_increment)
boxplot(house_Model$Rental_increment, col="Red", main="BOXPLOT EXPECTED ANNUAL
INCREMENT")
#16.Frequency Statistics of the Qualitative Variables and Plotting the Variables
house_Model$Location<-
factor(house_Model$Location,levels=1:4,labels=c("GURGAON","DELHI","NIODA","FARIDABAD"))
house_Model$Location
table(house_Model$Location)
house_Model$Prop_type<-
factor(house_Model$Prop_type,levels=1:4,labels=c("1BHK","2BHK","3BHK","4BHK"))
house_Model$Prop_type
table(house_Model$Prop_type)
lattice::histogram(house_Model$Social_infra,col=rainbow(2),main="DESCRIPTION OF AVAILIBILITY OF
SOCIAL INFRA-STRUCTURE")
#17. Statistical Model development and evaluation multiple linear regression
Model1<-lm(Price~.,data=house_Model)
Model2<-step(Model1,direction = "both",trace=1)
anova(Model2)
summary(Model2)
summary(Model1)
step(Model1,direction = "both",trace=1)
anova(Model1)
predict<-predict(Model1)
#18. Diagnostic analysis of plots
plot(predict(Model1),col="blue")
#19. Transfering the output in CSV FILE
write.csv(predict, file= "Predicted.csv", row.names=FALSE)
#20. Testing the correlation between the contributing varaibles
round(cor(mydata),2)
# It can also be called using the traditional method
# network_plot(correlate(mydata), min_cor=0.5)
mydata<-house_Model[,c('Price','Carpet_area')]
mydata
chart.Correlation(mydata,histrogram="TRUE",pch=10)
pairs.panels(mydata, scale=TRUE)
ggpairs(mydata)
ggcorr(mydata, nbreaks=8, palette='RdGy', label=TRUE, label_size=5, label_color='red')
```

Bibliography

1. Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, The Econometrics of Financial Markets, Princeton, NJ: Princeton University Press

2. Duffie, Darrell, 2001, Dynamic Asset Pricing Theory, 3rd Edition, Princeton, NJ: Princeton University Press

3. Singleton, Kenneth J., 2006, Empirical Dynamic Asset Pricing, Princeton, NJ: Princeton University Press

4. Robin A Dubin, 1998. Predicting housing prices using multiple listings data, Department of economics, case western Reserve university, Cleveland, ohio, 44106

5. Li Yu, Chenlu Jiao, Hongrun Xin, Yan Wang, Kaiyang Wang, 2018, Prediction on Housing Price Based on Deep Learning

6. Friedman, J. (1980) 'A conditional legit model of the role of local public services in residential choice', Urban Studies, Vol. 18, pp.347–258.

7. Ronald, R, Lennartz, C, Kadi, J (2017) What ever happened to asset-based welfare? Shifting approaches to housing wealth and welfare security, Policy & Politics, vol 45, no 2, 173–93, DOI: 10.1332/030557316X14786045239560

8. de Decker, P., & Dewilde, C. (2010). Homeownership and asset-based welfare: the case of Belgium. Journal of Housing and the Built Environment, 25(2), 243-262. DOI: 10.1007/s10901-010-9185-6

9. Lennartz, C., & Ronald, R. (2017). Assetbased Welfare and Social Investment: Competing, Compatible, or Complementary Social Policy Strategies for the New Welfare State? Housing, Theory and Society, 34(2), 201- 220. https://doi.org/10.1080/14036096.2016.1220422

10. https://assets.kpmg/content/dam/kpmg/pdf/2014/09/Decoding-Housing-for-all2022.pdf

11. Diewert, W.E. (2009), "The Paris OECD-IMF Workshop on Real Estate Price Indexes: Conclusions and Future Directions", Price and Productivity Measurement: Volume 1-Housing, pp. 87-116 Trafford Press.

This is to certify that ***MANNAWAR HUSSAIN*** a student of IMT-CDL Ghaziabad has completed the Project work on Asset Based Housing Price Model for NCR, under my guidance and supervision.

I certify that this is original work and has not been copied from any source.

Guide

S.A.JAFRI

NAME: SYED ASHRAF JAFRI

Syed Ashraf Jafri
Email: ashrafjafri66@hotmail.com
Reach: +91-96114-22559

I am a Statistician/ Data Scientist having worked into many different application areas such as Socio-Economic Research, Marketing Research, Biostatistics, Machine Learning, and Government Accounting.

Summary of Qualifications

- Statistics: Extensive experience in Exploratory Data Analysis and Data Inference - Skilled in uncovering insights and identifying patterns. Familiarity with advanced statistical models, tests, and techniques such as GLM (general and generalized linear models), Mixed Models, Multivariate analysis, Sampling Techniques, Survival Analysis, Non-parametric tests such as chi-square, Analysis of variance(ANOVA)- and Analysis of Covariance(ANCOVA) and its application.

- Machine Learning Algorithms (Supervised and Unsupervised): Linear Regression, Logistic Regression, Instance-based Algorithm (Knn), Clustering, Decision Tree, Time-series Analysis, Random Forest, Factor Analysis including Principal Component Analysis and Structure Equation Modeling.

Area of Interest

- ***The Design of Experiments, Sample Survey, Statistical inference, Multivariate Analysis, Regression Analysis, Time-Series Analysis, Machine Learning System, Operations Research***

Statistical Packages / Programming

- Statistical Computing: MS-EXCEL, SPSS-23, SAS-9.4, MINITAB17.0, R3.0.2, Python
- Mathematical / Statistical Modeling: MATLAB
- DBMS: MS-ACCESS and MS-SQL

Education

- M.A. (Statistics) with Specialization in Operations Research from Patna University (Patna)
- PGDCSA from Bihar College of Engineering, (NIT) Patna

· Pursuing PhD in Mathematics & Actuarial Sciences, at BS Abdur Rahman University, Chennai-(2016-2019)

Professional Activities

1. Worked as a Visiting Faculty-Data Analysis at Ed-Excel Centre (GNIIS), Teaching Data Handling and Analysis

2. Presented a Case Study on e-marketing at AMITY BUSINESS SCHOOL.

3. Worked as Consultant Statistician on clinical studies at International Oncology Services Pvt. Limited, Noida.

4. I have provided a Statistical Consultancy for DNB\ PH.D\M.PHIL Students and for publications in Journals.

5. Worked also as a Visiting Faculty at New Shores International College, Bangalore, Teaching Business Research Methods & Numerical and Statistical Methods.

6. Worked as a Consulting Consultant on one Predictive Analytics Project based on Time-Series Analysis for Ex-Village Software Company.

Publications

· A Comparison between Four Immobilization Systems for Pelvic Radiation Therapy using CBCT and Paired Kilo-voltage Portals