SOMAIYA VIDYAVIHAR

K J Somaiya College of Science & Commerce

Somaiya TRUST

K. J. SOMAIYA COLLEGE OF SCIENCE AND COMMERCE (AUTONOMOUS-Affiliated to University of Mumbai) Re-accredited “A’ Grade by NAAC

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that
the following students of TYBSc (Statistics) have successfully completed the project

“ Flat Rent Prices of Mumbai and its Suburb ”

of Discipline Specific Elective Paper (Regression Analysis) as a part of assignment during the academic year 2022-23 under the guidance of Assistant professor Mr. Ashish Mhatre.

Team Members
1) Rupali Keshri
2)Vidhi Murdeshwar

Mr. Ashish Mhatre
(Project Mentor)

Mr. Prashant Shah
(Head of the Department)

Index

Introduction

Regression Analysis is a basic method used in statistical analysis of data. It's a statistical method which allows estimating the relationship between variables. We need to identify dependent variable which will vary based on the value of the independent variable. For example the rent of the house depending on the square feet of the house.

According to the recent reports by CREDAI-MICHI and data analytic firm CRE Matrix, the average housing rent in Mumbai Metropolitan region (MMR) increased by upto 29 percent in the last three and half years.

Also, the average monthly rentals in over 80 micro markets of the MMR rose in a range of 4 percent to 229 percent in August this year as compare to 2018.

We can say that, the real estate industry is currently going through a momentous cycle and the increase in housing rentals give a ray of optimism to both developers and homebuyers since this will encourage more housing sales in the upcoming years.

Methodology

Steps involved in conducting survey:

1. First step was defining our objectives.
2. We surveyed 67 individuals using Google forms.We entered the Data into Excel and cleaned our data, due to missing values, inconsistent data and data beyond our scope of study after cleaning the data we were left with 49 responses.

Scope of the Survey:

In this survey, our target population was anyone residing Mumbai and its suburbs.

Objectives

To determine the factors leading to the flat rent price in mumbai and its suburbs

To determine the rent price of flats in mumbai and its suburbs

Primary data questionnaire

1.Do you live on rent?

2.Distance of your flat from railway station? (in km)

3.Number of rooms in your flat (including hall and kitchen, For Example: In 1BHK there is one hall, one bedroom, one kitchen. i.e. 3 rooms)

4.Area of flat? (in sq.ft)

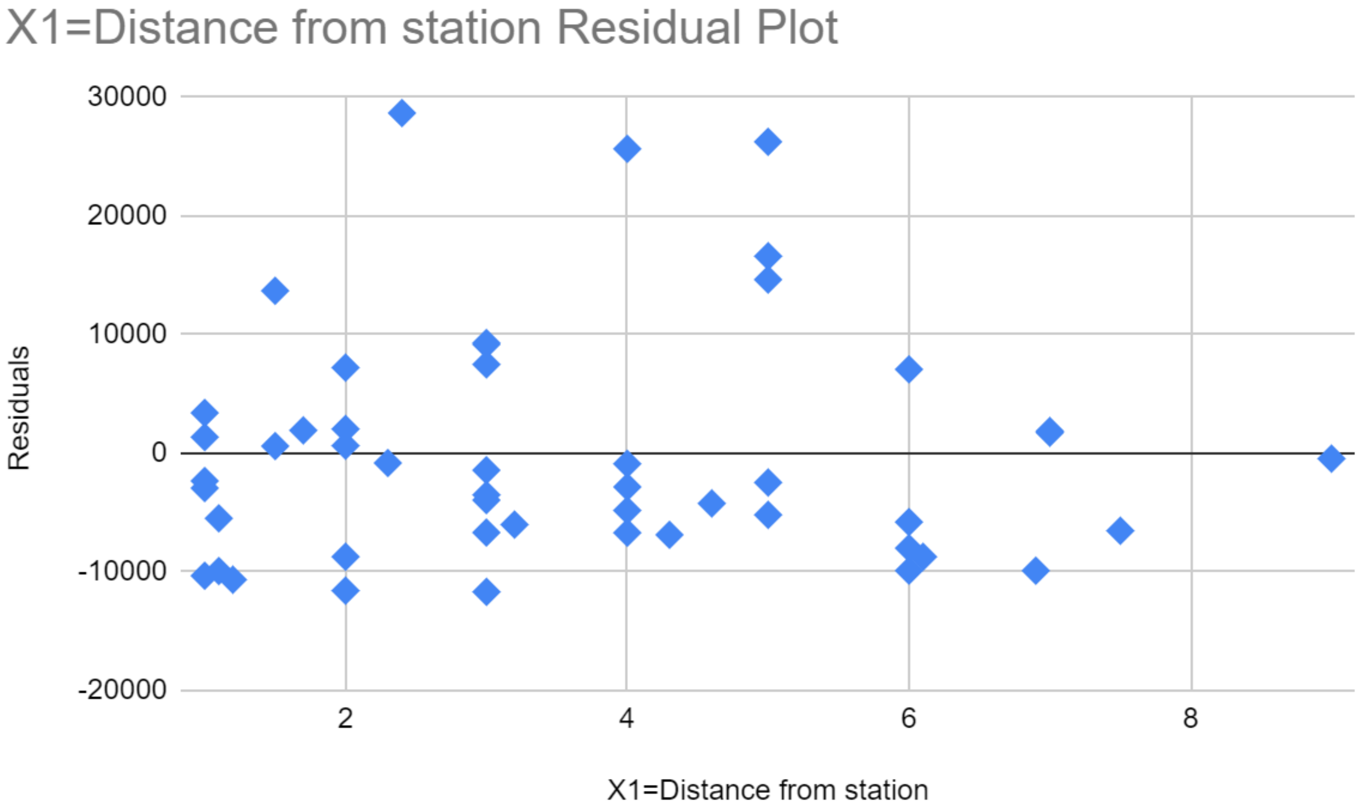
DATA

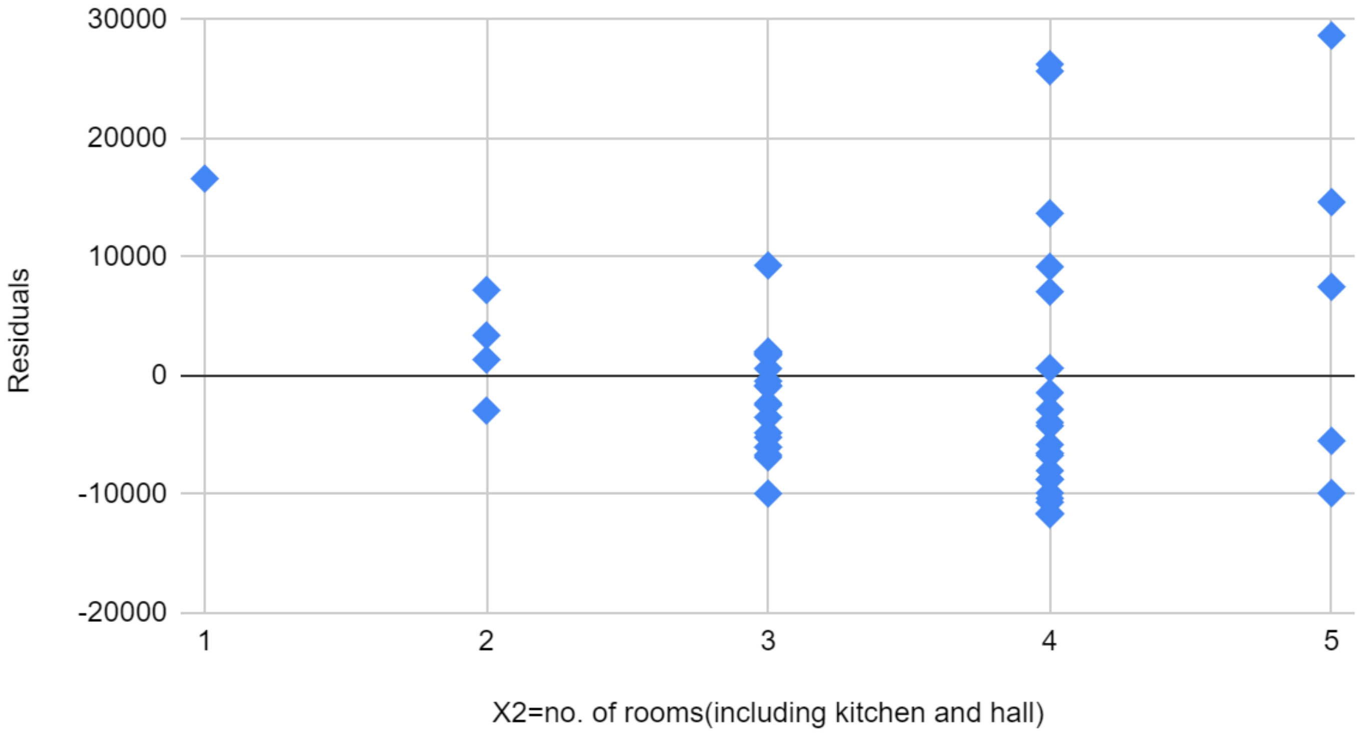
X1=Distance from station	X2=no. of rooms(including kitchen and hall)	X3=area of house(sqft)	Y=Rent of flat(in rupees)
2	4	930	12000
3	3	700	10000
6	4	1050	12000
1.1	3	625	5000
1.7	3	345	17000
1.5	3	250	16000
5	3	550	10000
2	2	400	15000
2	4	1000	9000
3	5	1200	34000
1	2	150	6000
3	4	535	30000
2	3	300	17000
1	2	320	12000
2	4	600	22000
5	5	1100	40000
2.4	5	1500	55000
1.5	4	800	35000
3	3	600	23000
2.3	3	780	13000
1	3	450	13000
1	2	300	10000
5	1	600	15000
7	3	550	13000
1.2	4	725	11000
4	4	950	45000
9	3	720	9000
4	4	1020	12500
1.1	5	1350	22000
3	3	610	7000
4.6	4	800	15000

6	4	750	8500
6.1	4	800	9500
6.9	5	1200	14000
7.5	4	930	10500
4	3	675	12000
3	4	980	16000
4	3	710	8000
3	4	10000	890
7	3	730	12500
4.3	3	580	6000
5	4	910	45000
6	4	950	10000
1	4	960	11000
5	3	690	7000
3	4	810	8600
4	4	950	16500
6	4	990	25000
3.2	3	630	7500

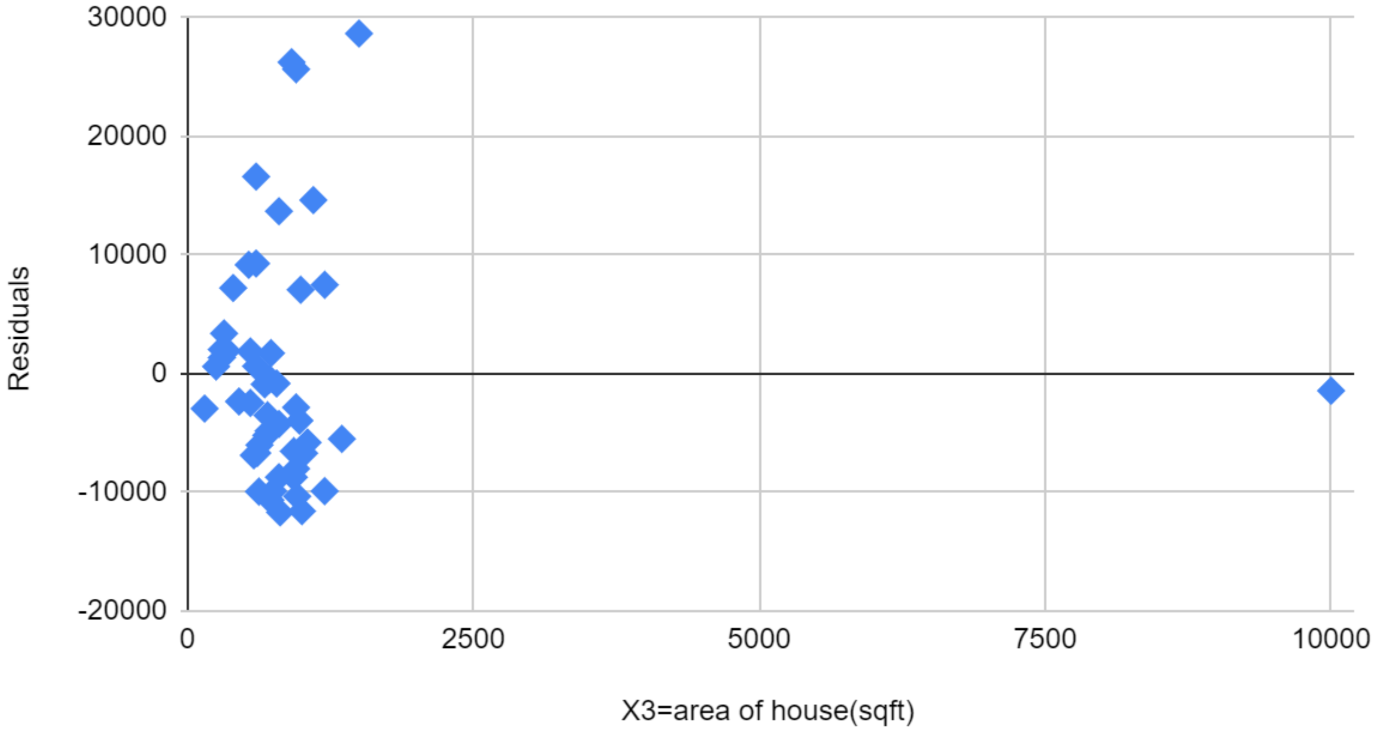
Graphical Representation

Residual plot



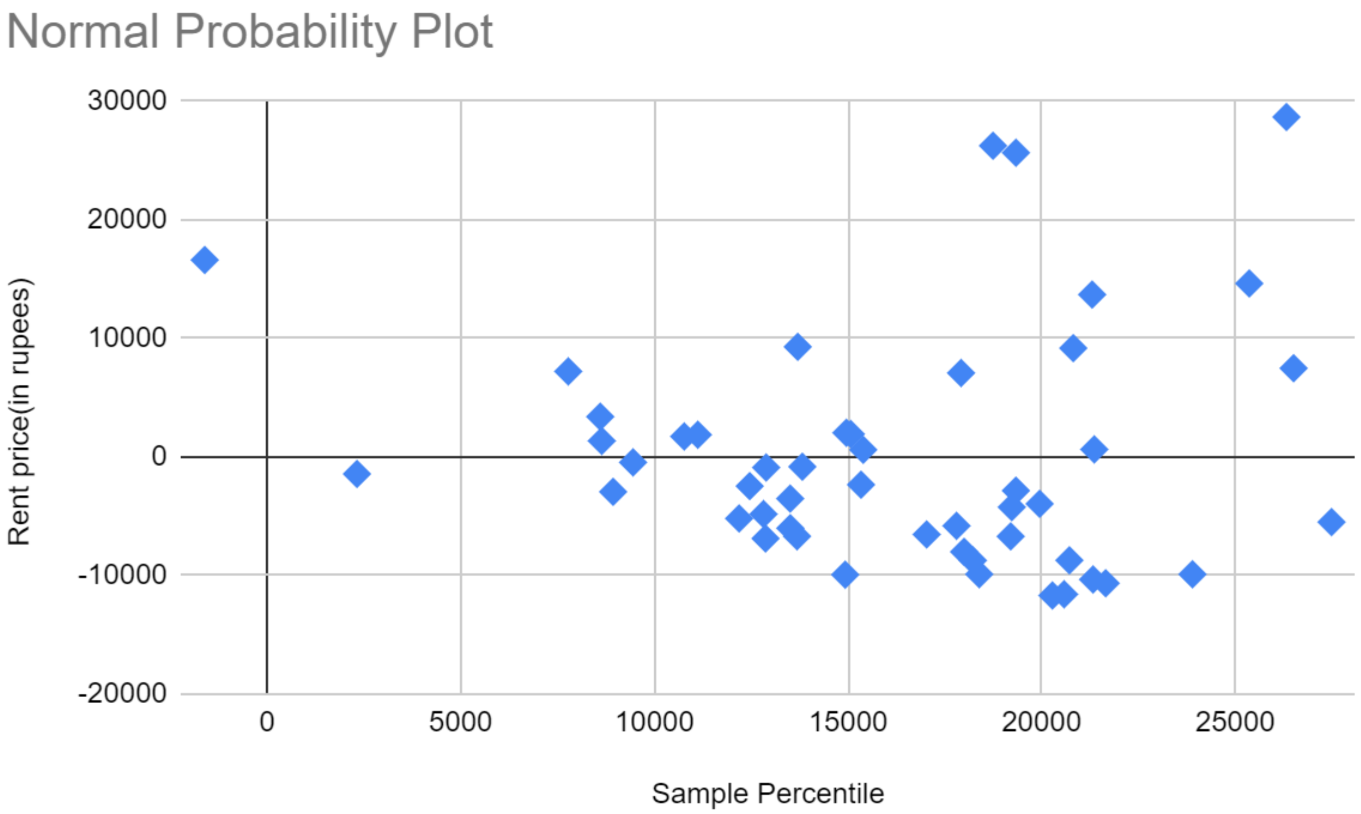


X2=no. of rooms(including kitchen and hall) Residual Plot



X3=area of house(sqft) Residual Plot

Normal Probability Plot



Analysis

From above data we have:

Y= Dependant variable = Rent of flat(rupees)

X1=Independent variable =Distance from station(km)

X2=Independent variable =no. of rooms(including kitchen and hall)

X3=Independent variable=area of house(sqft)

Number of observation=49

The Multiple regression equation for this model is:

Y^ = β0^ + β1^ × X1 + β2^ × X2 + β3^ × X3

Regression Statistics	
Multiple R	0.5182086969
R Square	0.2685402535
Adjusted R Square	0.2197762704
Standard Error	10125.75493
Observations	49

Here , the variability is 26.8540% due to Rent of flat can be explained by regression model.

Overall significance at 5% LOS

H0: β1=β2=β3=0 v/s H1: at least one variable is significant i.e. βi ≠ 0

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1693894280	564631426.8	5.50693845	0.002616869767
Residual	45	4613891083	102530913		
Total	48	6307785363			

Here F cal > F tab

Therefore, we reject H0 at 5% level of significance.

Y= -4078.6129 + (-669.8014514)X1 + 6988.403709X2 + (-1.954411627)X3

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Upper 95%*
Intercept	-4078.612946	6329.075396	-0.6444247685	0.5225715716	-16826.02496	8668.799069	-16826.02496	8668.799069
X1=Distance from station	-669.8014514	718.3575644	-0.9324067632	0.3561022881	-2116.647835	777.0449322	-2116.647835	777.0449322
X2=no. of rooms(including kitchen and hall)	6988.403709	1757.959637	3.975292471	0.0002518875642	3447.691299	10529.11612	3447.691299	10529.11612
X3=area of house(sqft)	-1.954411627	1.117806969	-1.748433925	0.08720854115	-4.205790399	0.2969671453	-4.205790399	0.2969671453

Forward selection

It is a stepwise regression which begins with an empty model and adds in variables one by one.

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	no of rooms	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: rent of flat

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.454[a]	.206	.189	10324.394	1.546

a. Predictors: (Constant), no of rooms

		Anova		

Model	Sum of Squares	df	Mean Square	F	Sig.

1	Regression	1297909572.274	1	1297909572.274	12.176	.001[b]
	Residual	5009875790.991	47	106593101.936		
	Total	6307785363.265	48			

a. Dependent Variable: rent of flat

b. Predictors: (Constant), no of rooms

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics	
						Tolerance	VIF
1	area of house	-.228[b]	-1.732	.090	-.247	.936	1.068
	dist from stn	-.116[b]	-.878	.385	-.128	.978	1.022

Excluded Variables

Model		Collinearity Statistics
		Minimum Tolerance
1	areaofhouse	.936
	distfromstn	.978

a. Dependent Variable: rent of flat

b. Predictors in the Model: (Constant), no of rooms

Assumptions:

Autocorrelation

The durbin Watson statistic is a test for autocorrelation in a regression model’s output.

et	et-1	(et-(et-1))^2	et^2
-8717.796172	-	-	75999970.1
-3509.105686	-8717.796172	27130456.58	12313822.72
-5804.060971	-3509.105686	5266819.76	33687123.76
-9928.309316	-5804.060971	17009424.41	98571325.88
1926.336299	-9928.309316	140532622.7	3710771.538
606.7069045	1926.336299	1741421.74	368093.268
-2462.664527	606.7069045	9421040.987	6064716.575
7223.173083	-2462.664527	93815450.21	52174229.38
-11580.98736	7223.173083	353596449.9	134119268.2
7491.29271	-11580.98736	363751867	56119466.47
-2935.231275	7491.29271	108712402.4	8615582.64
9180.012687	-2935.231275	146779136.3	84272632.93
2039.328212	9180.012687	50989374.77	4158859.554
3397.018701	2039.328212	1843323.466	11539736.06
637.247991	3397.018701	7616334.373	406085.002
14635.45445	637.247991	195949784.1	214196527
28675.73533	14635.45445	197129487.1	822297796.5
13693.22959	28675.73533	224475478.1	187504536.6
9295.453151	13693.22959	19340437.61	86405449.28
-821.6137722	9295.453151	102355043.1	675049.1907
-2337.311496	-821.6137722	2297339.59	5463025.029
1357.930469	-2337.311496	13654813.18	1843975.158
16611.86347	1357.930469	232682472	275954008
1876.938375	16611.86347	217118017.6	3522897.665
-10654.29172	1876.938375	157031727.6	113513932
25660.89496	-10654.29172	1318792784	658481530.3
-451.2087451	25660.89496	681841960.1	203589.3317
-6702.296223	-451.2087451	39076094.65	44920774.66
-5488.168304	-6702.296223	1474106.604	30119991.33
-6685.002733	-5488.168304	1432412.65	44689261.54
-4230.38591	-6685.002733	6025143.747	17896164.95

-9890.384459	-4230.38591	32035583.58	97819704.75
-8725.683733	-9890.384459	1356527.782	76137556.6
-9896.481629	-8725.683733	1370767.715	97940348.64
-6533.888189	-9896.481629	11307034.64	42691694.87
-888.1645256	-6533.888189	31874195.69	788836.2244
-3950.274139	-888.1645256	9376515.287	15604665.78
-4819.760119	-3950.274139	756005.868	23230087.6
-1431.481267	-4819.760119	11480433.58	2049138.617
1728.732468	-1431.481267	9986950.85	2988515.947
-6872.893195	1728.732468	73987964.04	47236660.87
26252.51995	-6872.893195	1097292996	689194803.7
-7999.502134	26252.51995	1173201017	63992034.39
-10328.96527	-7999.502134	5426398.525	106687523.6
-5189.0469	-10328.96527	26418760.9	26926207.73
-11682.52412	-5189.0469	42165246.36	136481369.7
-2839.105037	-11682.52412	78206061.01	8060517.41
7078.674331	-2839.105037	98362347.59	50107630.29
-6011.95421	7078.674331	171364555.6	36143593.42
0.000000005526089808		7614852588	4613891083

d= ∑(et-e(t-1))^2 /∑et^2= 0.000000005526089808

For n= 49, k=3 the critical values from D-W tables table are:

dl= 1.45635 , du= 1.62573

Therefore, positive autocorrelation

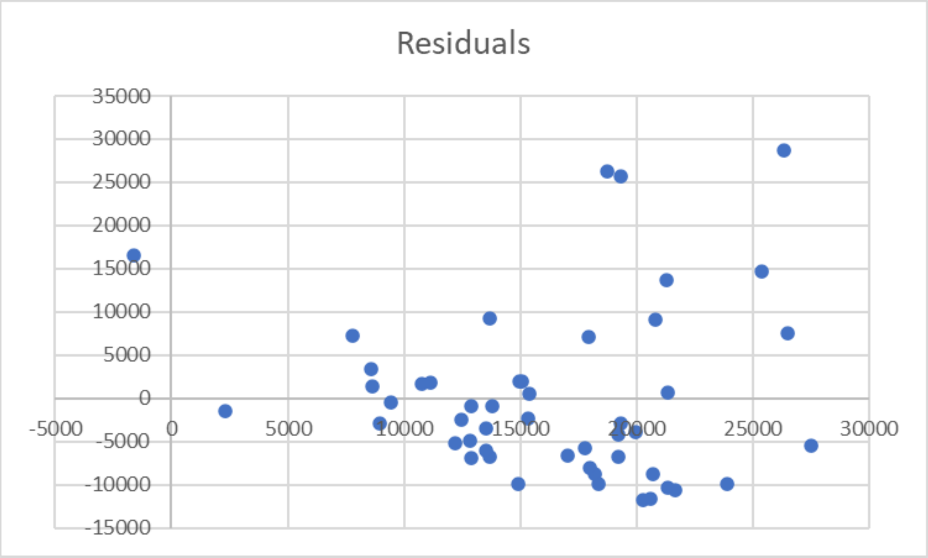
Multicollinearity

	R^2	VIF
x1	0.025765	1.02643
x2	0.088359	1.09692
x3	0.065065	1.06959

If VIF>10 multicollinearity is presents seen from above table,We may conclude that multicollinearity is absent between the above independent variables.

Heteroscedasticity:

In this case, heteroscedasticity is present since the variance of the residuals is unequal over a range of measured values.



Conclusion:

We carried out various regression process with rent of flat as dependent variable (y) amd independent variables as distance from station(x1), number of rooms(x2), are of houses(x3) and so we can say that, positive autocorrelation is present, multicollinearity is absent since VIF<10, and heteroscedasticity (unequal variance) is present.

We can conclude that as the area of the house increases the price of the rent increases, similarly as the distance between the house and the station increases the price of the rent decreases and as the number of rooms in a flat increases so does the price of the rent.

So before renting a house we look after so many objectives and these were just the few necessary ones.