

MAY 31, 2019



REAL ESTATE MODELLING & FORECASTING

EXPLAINING ISTANBUL HOUSING PRICE WITH ECONOMIC
INDICATORS & FORECASTING UNIT PRICES CHANGE

ALPARSLAN YARIKTAS
516171003

Table of Contents

1.	Introduction.....	2
I.	Problem Question.....	2
II.	Literature Review.....	2
2.	Cross Section	4
III.	Setting The Libraries.....	4
IV.	Data Import.....	6
V.	Data Type Observation.....	6
VI.	Understanding Data	7
VII.	See The Missing Values	16
VIII.	Show The Correlation.....	17
IX.	Showing The Density	20
X.	Seeing The Correlation Plot.....	21
XI.	Building A Model (Train-Test)	23
XII.	Residuals - Outliers.....	33
XIII.	Errors	38
3.	Time Series.....	41
XIV.	Set The Libraries	42
XV.	Data Import.....	42
XVI.	Show The Distribution of Housing Price Index	43
XVII.	Clean The Outliers with Tsclean	43
XVIII.	Quarterly and Monthly Moving Averages.....	44
XIX.	Decompose - Forecasting The Data.....	46
XX.	Stationarity - ADF Test.....	46
XXI.	Autocorrelation and Choosing Model Order (ACF).....	47
XXII.	Tune The Model.....	48
XXIII.	Fitting an ARIMA Model	50
XXIV.	Evaluate and Iterate.....	51
XXV.	Forecast The Istanbul Housing Price Changes	53
4.	The Result.....	55
5.	References	56

1. Introduction

I. Problem Question

1. How The Economic Indicators Affect on Housing Unit Price at Istanbul?
2. Which indicators more dominant on the prices?
3. What must be the equation for Istanbul Unit Housing Price?
4. What will be the housing price for 6 months?

II. Literature Review

- The construction industry and other related sub industries are the most crucial aspects of the economy The Turkish housing industry has achieved fast growth in the past fifteen years The macroeconomic importance of the construction industry arises from its multiplier effect It sets 250 sub industries in motion with impacts on both growth and employment.
 - Housing and related interests make up a significant portion of the construction industry in our country. The industry's growth rate is influenced by booms/stagnation observed in housing demand and sales. The industry's sensitivity to the total national growth is also high.
 - As well the price of a home in the housing sector is determined by the balance of housing supply and demand In the short term, the supply of housing is for the most part fixed, meaning that the main variable in determining housing prices is the rise or decline in demand Consequently, in the short term, rising demand for housing causes prices to increase and declining demand causes them to fall.
-

In this study Section - 1 includes “The Regression Analysis” to explain the dependent variable which is **Unit Housing Price in Istanbul** with independent variables below in the table and tried to find **the highest and most effective model** that explaining unit housing price. In Section B, “The ARIMA Analysis” used for forecasting the “Housing Price Changes In Istanbul” for six months. The summary of study as in the below following chart.

Section 1 - Cross Section	Section 2 - Time Series
Regression Analysis	Auto Regressive Integrated Moving Average Analysis
Explaining Istanbul Housing Unit Price	Forecasting Istanbul Housing Price’s Change Next Month

Some variables and description in the below chart.

VARIABLES	DESCRIPTION
DATE	Monthly Date from 2013 to 2019
IST_RPPI	Istanbul Residential Property Price Index
IST_PRC	Istanbul Housing Unit Price TL/m ²
IST_CNST_PRMT	Construction Permits
IST_OCCP_PRMT	Occupational Permits
MG_RT	Mortgage Credit Rates
IST_FGR_SL	Istanbul Housing Sales to Only Foreigners
IST_PRP_SL	Istanbul Total Housing Sales
IST_MRTG_SL	Istanbul Housing Sales by Using Mortgage Loan
CNSTR_TRST	Construction Trust Index
IST_CPI	Istanbul Consumer Price Index
TR_PPI	Turkey Producer Price Index
USD_RT	USD Buying Exchange Rate
RNT_CPI	IST Rent Consumer Price Index
NEMP_RT	Unemployment Rate

Notes : IST_CNST_PRMT - IST_OCCP_PRMT forward some different periods 2,3,4,6,9 etc.

For the reason why using both *Price Index* and *Unit Prices* main purpose of the using both Price Index and Prices on the data set in order to see the correlation between Monthly Property Price Index and Unit Prices and cross check our observations. It can easily expect the correlation coefficient approximately ~ 99% The main value of Housing Prices Index and Unit Prices would be the expecting correlation coefficient which is approximately 0.96 but the monthly percentage changes between them wouldn't be the bigger than 80%. This study also prove this on the next paragraphs.

2. Cross Section

III. Setting The Libraries

Before start the study, the libraries should be import to get rid of the errors and help to analyze all in one for study.

```
library(readr)
library(ggplot2)
library(mlbench)

## Warning: package 'mlbench' was built under R version 3.5.3

library(corrplot)

## Warning: package 'corrplot' was built under R version 3.5.3
## corrplot 0.84 loaded

library(Amelia)

## Warning: package 'Amelia' was built under R version 3.5.3
## Loading required package: Rcpp

## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.5, built: 2018-05-07)
## ## Copyright (C) 2005-2019 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##

library(caret)

## Warning: package 'caret' was built under R version 3.5.3
## Loading required package: lattice
```

```
library(plotly)

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout

library(caTools)
library(reshape2)
library(dplyr)

## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(knitr)

## Warning: package 'knitr' was built under R version 3.5.3

library(ggplot2)
library(forecast)

## Warning: package 'forecast' was built under R version 3.5.3

library(tseries)

## Warning: package 'tseries' was built under R version 3.5.3

library(astsa)

##
## Attaching package: 'astsa'

## The following object is masked from 'package:forecast':
##
##   gas

library(readr)
```

IV. Data Import

This project include different studies that divide main values and percentage change according to the previous months.

IST_RPPI_MODEL

IST_RPPI_MODEL_CG

Main Value of All variables Monthly Percentage Changes Values(%)

```
library(readr)
IST_RPPI_MODEL <- read_csv("IST_RPPI_MODEL.csv",
                           col_types = cols(Date = col_date(format =
"%d.%m.%Y"))))
IST_RPPI_MODEL_CG <- read_csv("IST_RPPI_MODEL_CG.csv",
                              col_types = cols(Date = col_date(format =
"%d.%m.%Y"))))
#View(IST_RPPI_MODEL)
#View(IST_RPPI_MODEL_CG)
```

V. Data Type Observation

All the variables according to data type is “dbl” that is related with numeric values except “Date”. The study is not going to include **dates** in any model at Section-1 on Regression Analysis. Seeing the data types and first 5 column is also important for the data set what imported correctly.

```
head(IST_RPPI_MODEL)

## # A tibble: 6 x 23
##   DATE      IST_PRC IST_RPPI MG_RT IST_FGR_SL IST_PRP_SL IST_MRTG_SL
##   <date>      <dbl>   <dbl> <dbl>   <dbl>      <dbl>      <dbl>
## 1 2013-01-01  2064.    47.8  9.78    138      18235      8423
## 2 2013-02-02  2114.    48.8  9.58    120      18971      8836
## 3 2013-03-03  2157.    49.9  9.31    198      21570     10164
## 4 2013-04-04  2186.    50.9  9.06    209      20791      9726
## 5 2013-05-05  2218.    51.6  8.77    188      22030     10805
## 6 2013-06-06  2263.    52.4  8.3     155      19357      9762
## # ... with 16 more variables: CNSTR_TRST <dbl>, RNT_CPI <dbl>,
## #   IST_CPI <dbl>, TR_PPI <dbl>, USD_RT <dbl>, NEMP_RT <dbl>,
## #   IST_CNST_PT <dbl>, IST_OCCP_PT <dbl>, IST_CNST_PT_2 <dbl>,
## #   IST_OCCP_PT_2 <dbl>, IST_CNST_PT_3 <dbl>, IST_OCCP_PT_3 <dbl>,
```

VI. Understanding Data

In order to understand and examine data generally use five number summary.

Five Number Summary shows *median, mean, min-max* and *Q1 & Q3* values.

This may help us to mining and eliminating and understanding the structure of data and classifying the observations.

For example :

- Housing Unit Prices in Istanbul has minimum value which is 2064 ; Mean Value is 3708 and also the Median Value is 3922. Q3 is 4562 and Max. value 5123. It can be easily understanding this data classified with a good range because there is not any extreme observation.
- Istanbul Unit Price Changes min.value -1.266, max.value 3.756, mean 1.285, median 1.366
- Total Istanbul Housing Sales min.value 11903, max.value 27156, mean value 19512, median value 19305. 3rd Quartile value is 21213 This value shows us there is not any extreme circumstances.
- When it comes to mortgage rates,

Min.	1st Q.	Median	Mean	3rd Q.	Max.
8.30	11.00	12.20	13.11	13.93	28.99

The situation about the outliers will be scrutinized next paragraphs.

1. Unit Housing Price

```
summary(IST_RPPI_MODEL$IST_PRC)
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2064   2775   3922   3708   4562   5123
```

2. Residential Housing Price Index

```
summary(IST_RPPI_MODEL$IST_RPPI)
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      47.81  63.26  86.48  81.55 100.39 103.19
```


3. Property Sales

```
summary(IST_RPPI_MODEL$IST_PRP_SL)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	11903	17398	19305	19512	21213	27156

4. Foreigner Sales

```
summary(IST_RPPI_MODEL$IST_FGR_SL)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	120.0	403.2	540.5	608.1	646.0	2283.0

5. Sales by Mortgage

```
summary(IST_RPPI_MODEL$IST_MRTG_SL)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	987	6230	7150	7109	8672	10805

6. Occupational Permits forward 9 months

```
summary(IST_RPPI_MODEL$IST_OCCP_PT_9)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	552.0	966.2	1121.5	1147.3	1318.5	2024.0

7. Consumer Price Index

```
summary(IST_RPPI_MODEL$IST_CPI)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	220.9	248.1	278.4	285.7	316.6	403.7

8. Mortgage Rates

```
summary(IST_RPPI_MODEL$MG_RT)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	8.30	11.00	12.20	13.11	13.93	28.99

9. Construction Permits

```
summary(IST_RPPI_MODEL$IST_CNST_PT)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	322.0	967.8	1322.5	1371.3	1597.0	4845.0

10. Rent Consumer Price Index

```
summary(IST_RPPI_MODEL$RNT_CPI)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      297.5  327.1   366.3   373.8  415.1   477.5
```

11. Turkey Producer Price Index

```
summary(IST_RPPI_MODEL$TR_PPI)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      206.7  234.6   250.9   271.8  296.1   443.8
```

12. USD Currency Rates

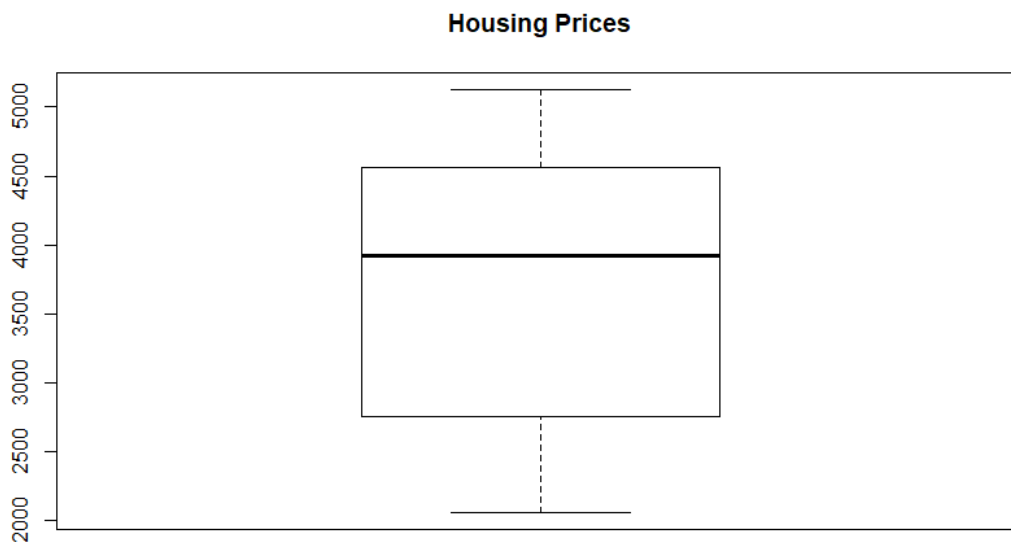
```
summary(IST_RPPI_MODEL$USD_RT)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.770  2.217   2.920   3.055  3.663   6.380
```

```
#summary(IST_RPPI_MODEL)
```

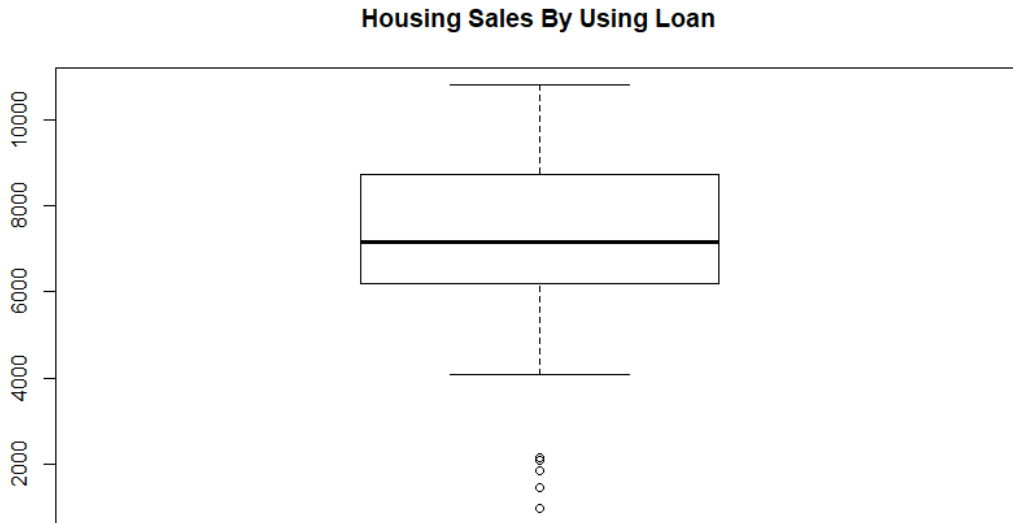
The values with box plot In addition to numeric table, Housing Sales also be showed with box plot.

```
boxplot(IST_RPPI_MODEL$IST_PRC, main = "Housing Prices",
        sub=paste("Outlier rows: ",
boxplot.stats(IST_RPPI_MODEL$IST_PRC)$out))
```



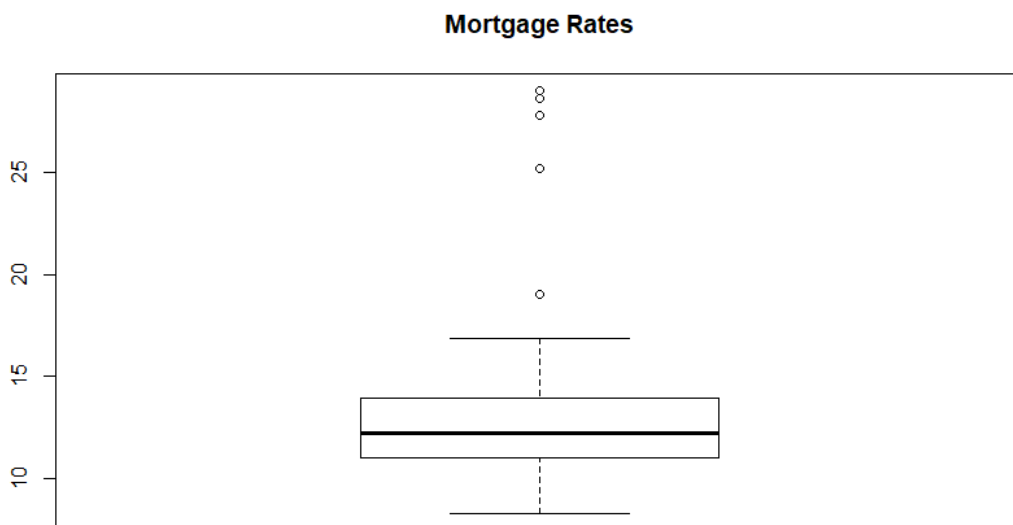
Outlier rows:

```
boxplot(IST_RPPI_MODEL$IST_MRTG_SL, main = "Housing Sales By Using Loan",
        border = par("fg"),
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$IST_MRTG_SL)$out))
```



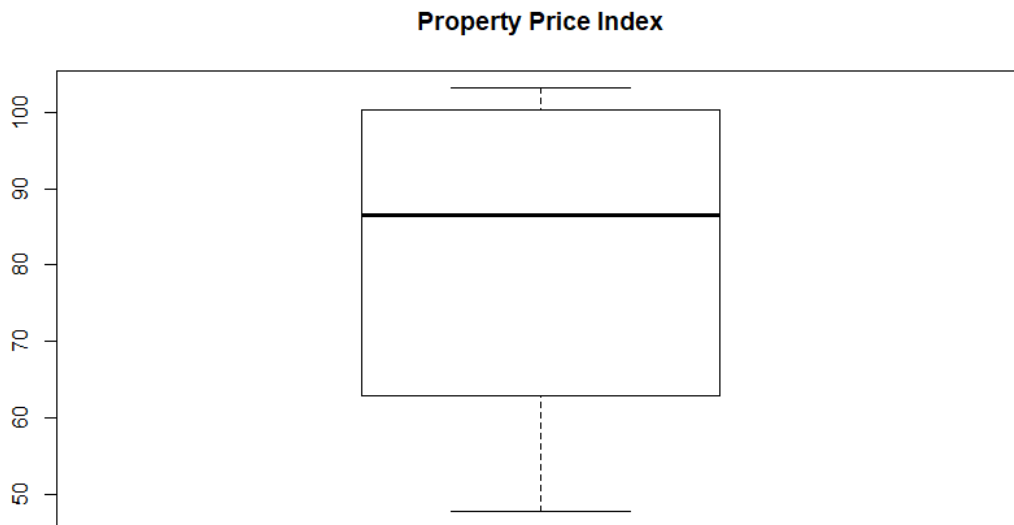
Outlier Rows: 2062

```
boxplot(IST_RPPI_MODEL$MG_RT, main = "Mortgage Rates",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$MG_RT)$out))
```



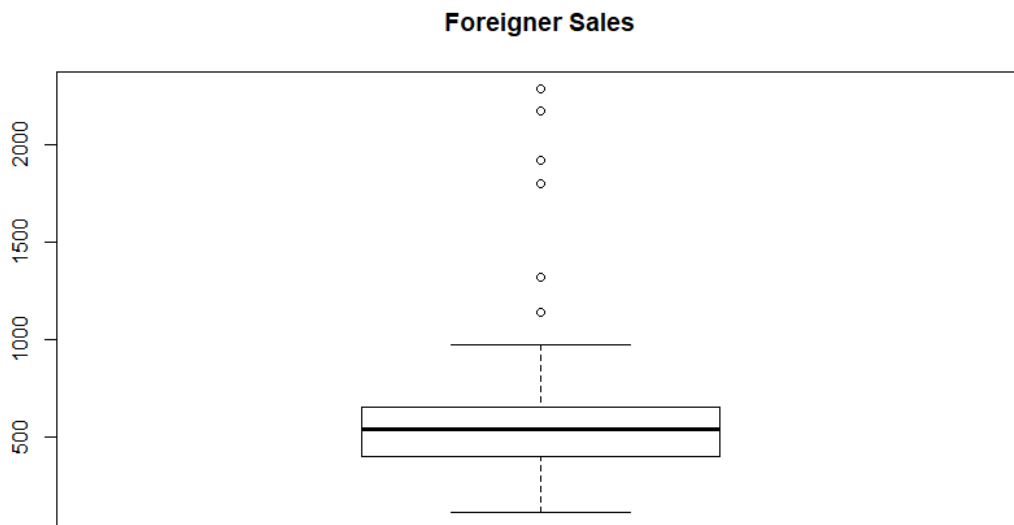
Outlier Rows: 20.02

```
boxplot(IST_RPPI_MODEL$IST_RPPI, main = "Property Price Index",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$IST_RPPI)$out))
```



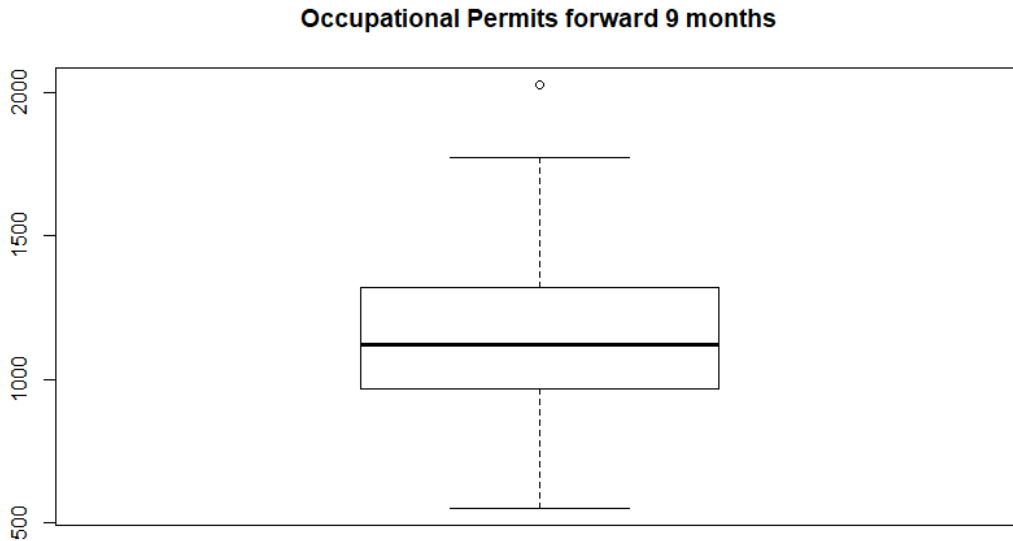
Outlier Rows:

```
boxplot(IST_RPPI_MODEL$IST_FGR_SL, main = "Foreigner Sales",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$IST_FGR_SL)$out))
```



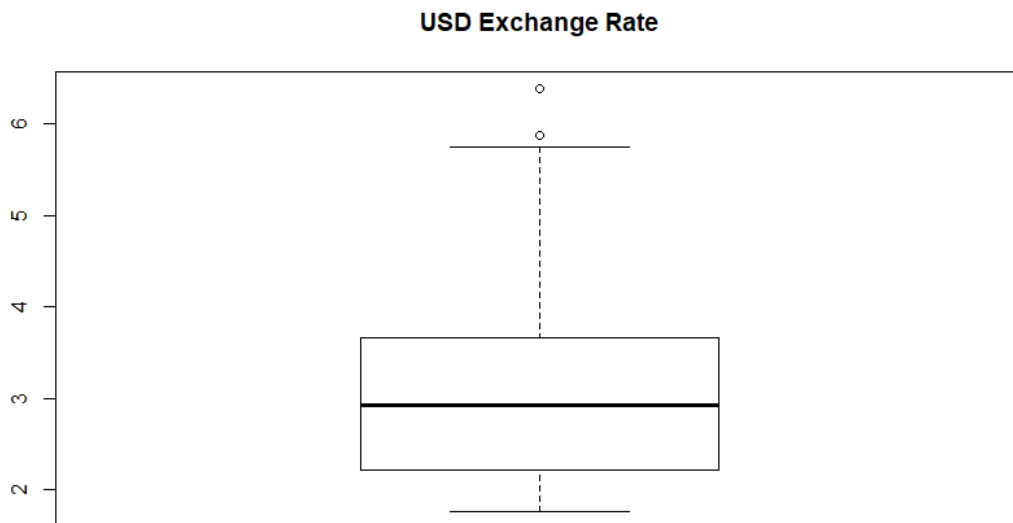
Outlier Rows: **2200**

```
boxplot(IST_RPPI_MODEL$IST_OCCP_PT_9, main = "Occupational Permits forward 9 months",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$IST_OCCP_PT_9)$out))
```



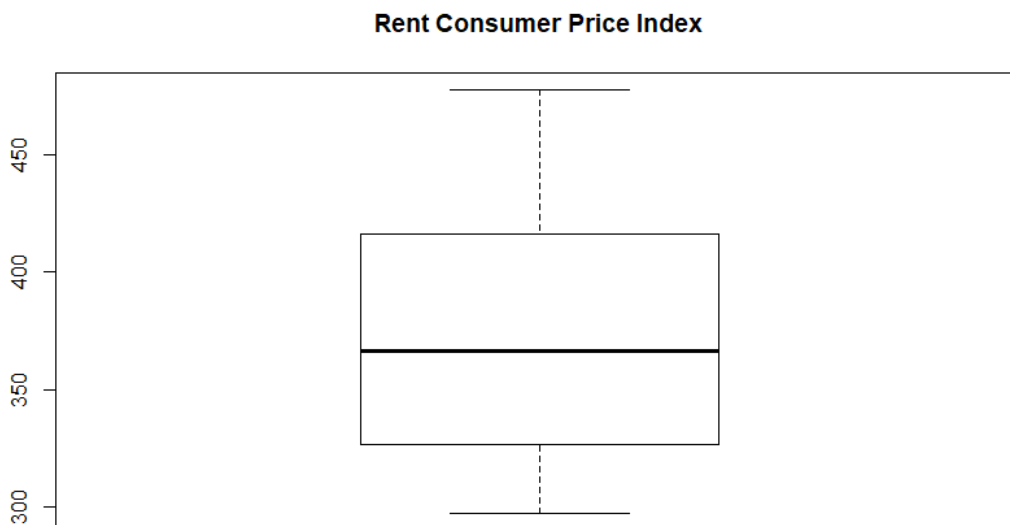
Outlier Rows: 2024

```
boxplot(IST_RPPI_MODEL$USD_RT, main = "USD Exchange Rate",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$USD_RT)$out))
```



Outlier Rows: 6.87

```
boxplot(IST_RPPI_MODEL$RNT_CPI, main = "Rent Consumer Price Index",
        sub=paste("Outlier Rows: ",
boxplot.stats(IST_RPPI_MODEL$RNT_CPI)$out))
```



Outlier Rows:

```
MG_RT_OUT <- boxplot.stats(IST_RPPI_MODEL$MG_RT)$out
MRTG_SL_OUT <- boxplot.stats(IST_RPPI_MODEL$IST_MRTG_SL)$out
data.frame(MG_RT_OUT, MRTG_SL_OUT)

##   MG_RT_OUT MRTG_SL_OUT
## 1    19.02      2082
## 2    25.19      1863
## 3    28.99      2164
## 4    28.63       987
## 5    27.82      1461

IST_FGR_SL_OUT <- boxplot.stats(IST_RPPI_MODEL$IST_FGR_SL)$out
IST_OCCP_PT_9_OUT <- boxplot.stats(IST_RPPI_MODEL$IST_OCCP_PT_9)$out
USD_RT_OUT <- boxplot.stats(IST_RPPI_MODEL$USD_RT)$out

data.frame(IST_FGR_SL_OUT, IST_OCCP_PT_9_OUT, USD_RT_OUT)

##   IST_FGR_SL_OUT IST_OCCP_PT_9_OUT USD_RT_OUT
## 1           1321           2024         6.38
## 2           1141           2024         5.87
## 3           1799           2024         6.38
## 4           2283           2024         5.87
## 5           1922           2024         6.38
## 6           2170           2024         5.87
```

Mortgage Rates and Mortgage Sales have 5 outliers, Foreigner Sales 6, Occupational permits 1 and 2 outliers in Usd Exchange Rates.

The outliers distracting model confidence, outliers should remove from dataset. In the study outliers used because of the size.

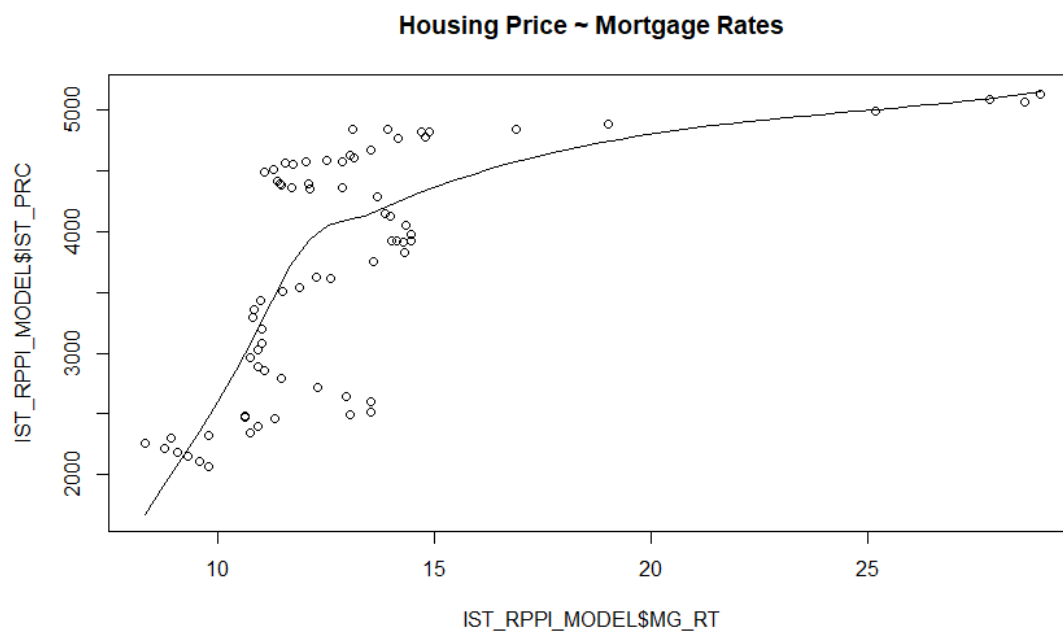
An alternative solution may offer remove the outliers and **fill with median value** of column.

When It comes to *dependent - independent variables*, It can be easily showing distribution in scatter plot.

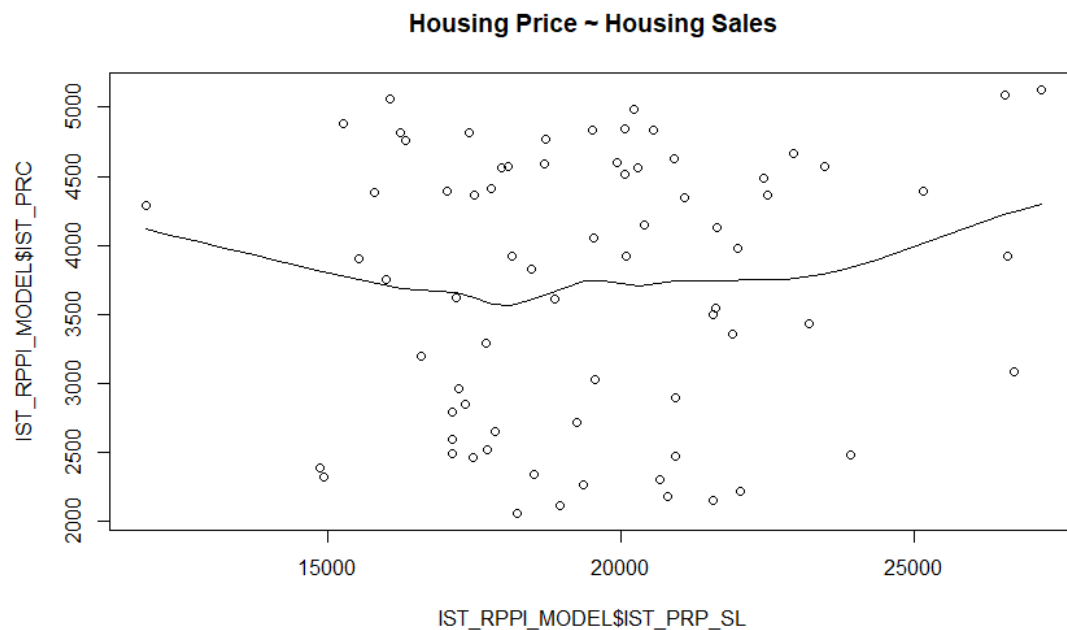
For example :

Some distrubitons about Housing Price and other variables showing with the scatter plot.

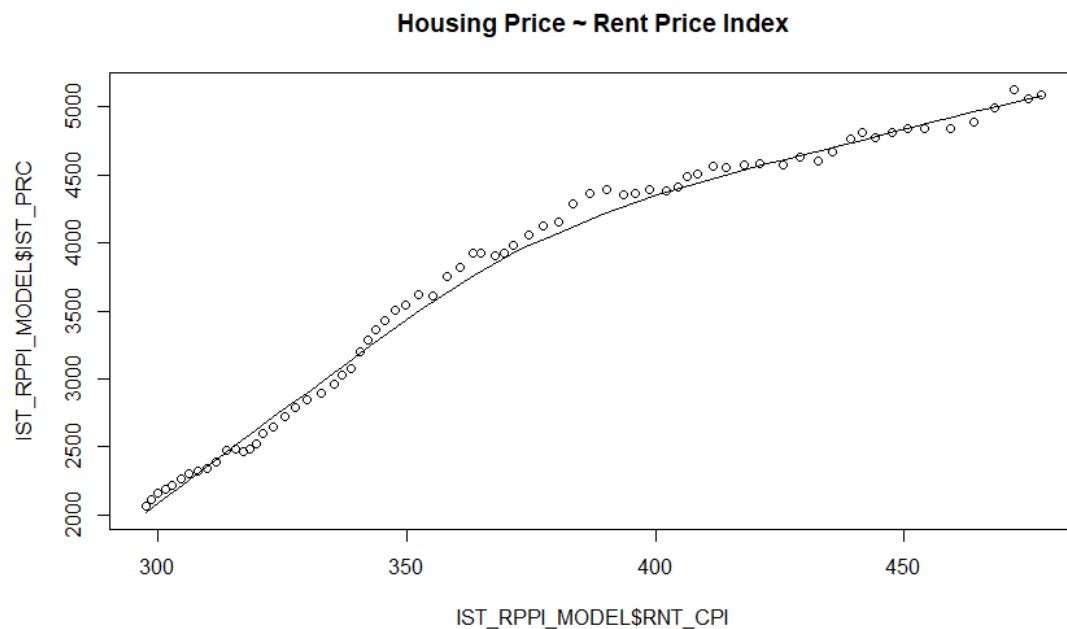
```
scatter.smooth(x = IST_RPPI_MODEL$MG_RT,  
              y = IST_RPPI_MODEL$IST_PRC,  
              main = "Housing Price ~ Mortgage Rates")
```



```
scatter.smooth(x = IST_RPPI_MODEL$IST_PRP_SL,  
              y = IST_RPPI_MODEL$IST_PRC,  
              main = "Housing Price ~ Housing Sales")
```



```
scatter.smooth(x = IST_RPPI_MODEL$RNT_CPI,
               y = IST_RPPI_MODEL$IST_PRC,
               main = "Housing Price ~ Rent Price Index")
```



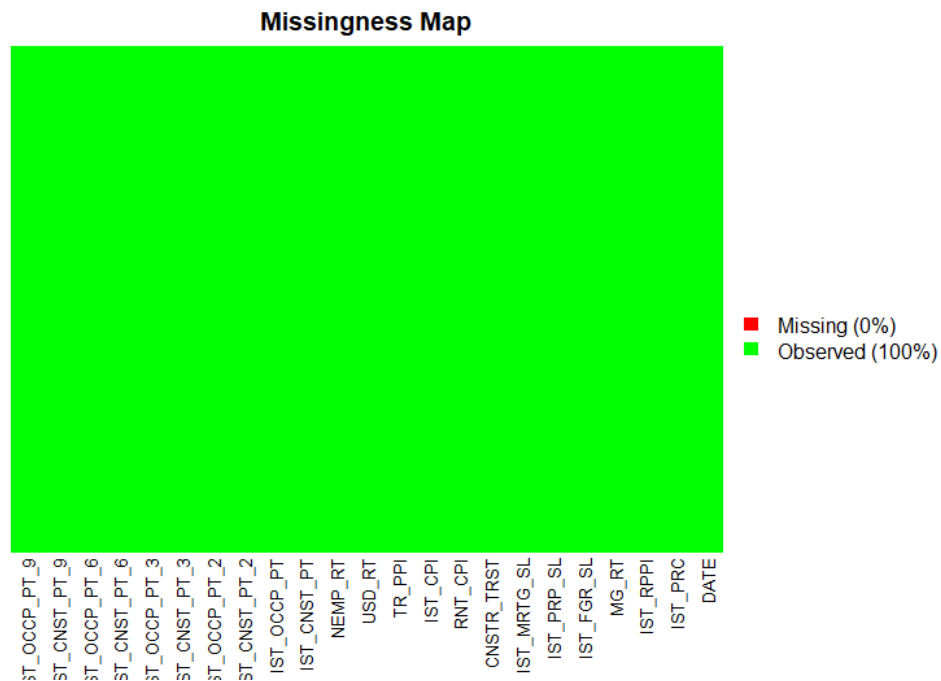
- This plot also sort out to correlation. The relation between price & mortgage rates also positive correlation but taking into consideration that supply and demand equilibrium, expecting when mortgage rates down prices would be higher because of the demand would increasing.

- When it comes to Turkey considering that there is an imbalance situation. People always wanting to get a housing neither mortgage rates increase nor mortgage rates decrease.
- The study emphasis this situation on next paragraphs at the Correlation.

In order to see the missing observations, easily get this code by using *Amelia Package*. In this study, any missing value the model including.

VII. See The Missing Values

```
# in order to see the sum of missing values
# colSums(is.na(IST_RPPI_MODEL))
# for show the plot of the missing values
missmap(IST_RPPI_MODEL, col=c('red', 'green'),
        y.at=1, y.labels = '', legend = TRUE)
```



Note: The study doesn't consist any missing observations. On the some literature text that filling the missing values with median value of the variables.

VIII. Show The Correlation

Correlation table including related coefficients below in the chart.

```
# coefficients tables
res_mn <- cor(select(IST_RPPI_MODEL, -DATE))
res_mn <- data.frame(round(res_mn, 2))
head(res_mn, n=6)
```

##	IST_PRC	IST_RPPI	MG_RT	IST_FGR_SL	IST_PRP_SL	IST_MRTG_SL	
## IST_PRC	1.00	1.00	0.59	0.64	0.10	-0.55	
## IST_RPPI	1.00	1.00	0.53	0.58	0.09	-0.50	
## MG_RT	0.59	0.53	1.00	0.91	0.13	-0.85	
## IST_FGR_SL	0.64	0.58	0.91	1.00	0.35	-0.69	
## IST_PRP_SL	0.10	0.09	0.13	0.35	1.00	0.33	
## IST_MRTG_SL	-0.55	-0.50	-0.85	-0.69	0.33	1.00	
##	CNSTR_TRST	RNT_CPI	IST_CPI	TR_PPI	USD_RT	NEMP_RT	IST_CNST_PT
## IST_PRC	-0.37	0.96	0.93	0.83	0.88	0.64	-0.16
## IST_RPPI	-0.31	0.94	0.90	0.78	0.84	0.63	-0.11
## MG_RT	-0.75	0.68	0.77	0.84	0.80	0.48	-0.32
## IST_FGR_SL	-0.74	0.72	0.80	0.85	0.82	0.50	-0.20
## IST_PRP_SL	-0.23	0.11	0.13	0.13	0.13	0.11	0.21
## IST_MRTG_SL	0.62	-0.64	-0.70	-0.75	-0.72	-0.42	0.35
##	IST_OCCP_PT	IST_CNST_PT_2	IST_OCCP_PT_2	IST_CNST_PT_3			
## IST_PRC	0.23	-0.12	0.22	-0.09			
## IST_RPPI	0.24	-0.07	0.24	-0.04			
## MG_RT	0.22	-0.28	0.09	-0.23			
## IST_FGR_SL	0.26	-0.24	0.00	-0.12			
## IST_PRP_SL	0.44	-0.14	-0.27	0.20			
## IST_MRTG_SL	0.00	0.21	-0.21	0.22			
##	IST_OCCP_PT_3	IST_CNST_PT_6	IST_OCCP_PT_6	IST_CNST_PT_9			
## IST_PRC	0.25	0.02	0.40	0.06			
## IST_RPPI	0.27	0.04	0.41	0.08			
## MG_RT	0.08	-0.20	0.19	-0.30			
## IST_FGR_SL	0.02	-0.17	0.29	-0.26			
## IST_PRP_SL	-0.10	0.04	0.12	0.11			
## IST_MRTG_SL	-0.14	0.18	-0.13	0.35			
##	IST_OCCP_PT_9						
## IST_PRC	0.44						
## IST_RPPI	0.45						
## MG_RT	0.28						
## IST_FGR_SL	0.33						
## IST_PRP_SL	0.19						
## IST_MRTG_SL	-0.26						

This table show us the correlation coefficients between Istanbul Unit Housing Prices and the other variables.

The best correlation equivalent in order to explain the housing unit prices :

1. RNT_CPI : 0.96
2. IST_CPI : 0.93
3. USD_RT : 0.88
4. TR_PPI : 0.83
5. NEMP_RT : 0.64
6. IST_FGR_SL : 0.64
7. MG_RT : 0.59
8. IST_MRTG_SL : -0.55
9. IST_OCCP_PT_9 : 0.44
10. CNSTR_TRST : -0.37
11. IST_CNST_PT : -0.16

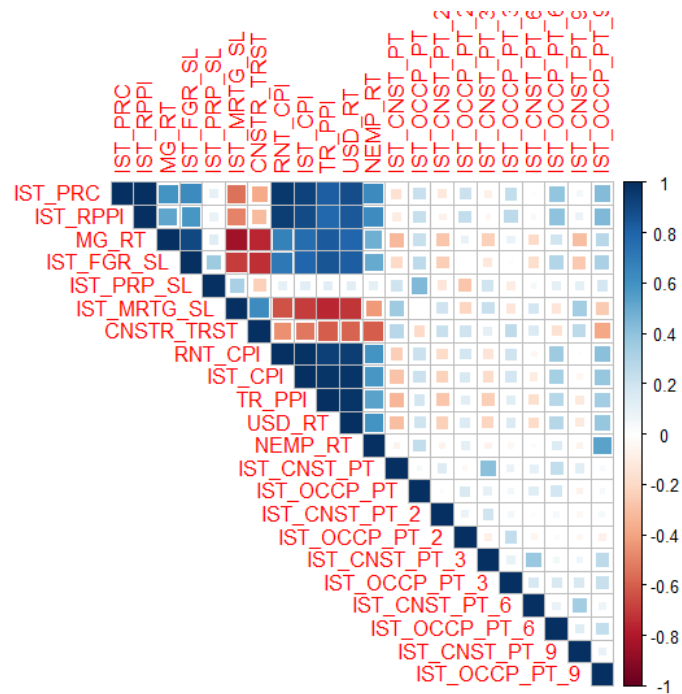
The Index which are Rent Consumer Price Index, Istanbul Consumer Price Index, USD Rate and Turkey Producer Price Index affecting The Housing Unit Prices very strongly correlated. So that try to find the best model in this study, include strong correlated variables in the model on next paragraphs.

```
res_cg<- cor(select(IST_RPPI_MODEL_CG, -DATE))
res_cg <- data.frame(round(res_cg, 2))
head(res_cg, n=1)

##          IST_PRC IST_RPPI MG_RT IST_FGR_SL IST_PRP_SL IST_MRTG_SL
## IST_PRC          1      0.62 -0.05      0.07      0.02      0.02
##          CNSTR_TRST RNT_CPI IST_CPI TR_PPI USD_RT NEMP_RT IST_CNST_PT
## IST_PRC          0.06  -0.17   0.08  -0.07  -0.11  -0.15  -0.22
##          IST_OCCP_PT IST_CNST_PT_2 IST_OCCP_PT_2 IST_CNST_PT_3
## IST_PRC          -0.15      -0.26      -0.01      -0.13
##          IST_OCCP_PT_3 IST_CNST_PT_6 IST_OCCP_PT_6 IST_CNST_PT_9
## IST_PRC           0.04      -0.18      -0.01      -0.19
##          IST_OCCP_PT_9
## IST_PRC          -0.1
```

This study also proved that the marker of Price and Price Index is positive correlated but first case Unit Price and Price Index correlated equal, The correlation coefficient between Unit Price Changes and Property Price Index Changes nearly 0.62, different from each others.

```
## plot the correlations
corrplot(cor(select(IST_RPPI_MODEL, -DATE)),
         method = "square", type = "upper", sig.level = .01)
```



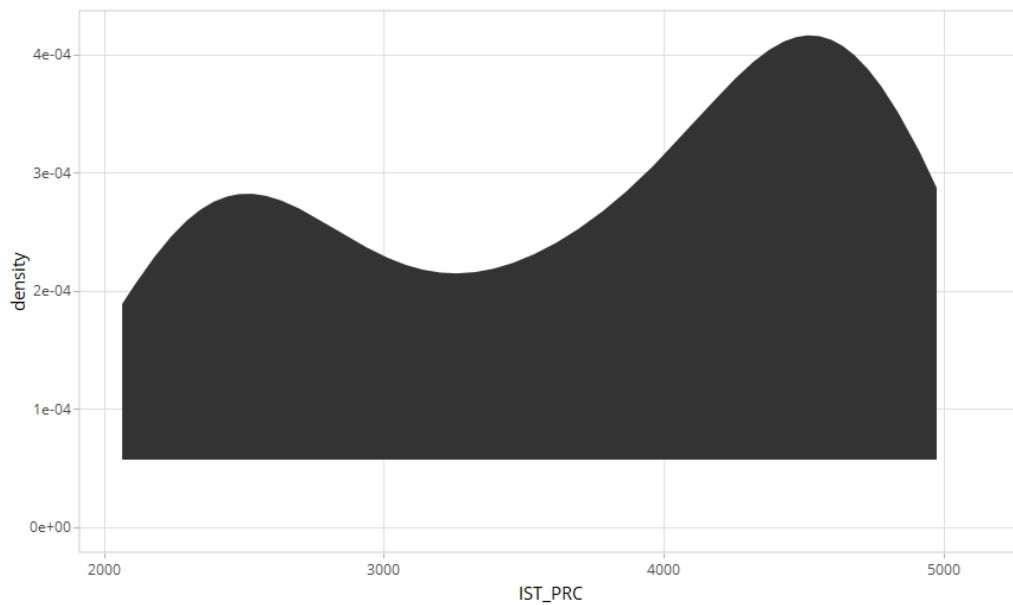
```
#corrplot(cor(select(IST_RPPI_MODEL_CG, -DATE)),
#         method = "square", addrect = 2, sig.level = .01, type = "upper")
```

1. The main value's correlation coefficient was ~ 1.00 but in this case the monthly changes different than each others when considered the actual values, the monthly change's coefficients different from main values. First, we look up the Unit Price and Property Price Index's coefficient is approximately $\sim 100\%$ but we look the monthly change of the both variables coefficient is $\sim 62\%$, So when we consider why the monthly changes different than main values.
This two main dependent variables is the same. It can be explained that Price Index is publishing after The Unit Prices maybe 1-2 month, Price Index can be influenced by the other things. According to housing market the index publishing after 1 and 2 months published unit prices and Index be affecting from other main Index like Rent, CPI, PPI etc.
2. Secondly, we should look the Construction-Occupational Permits the best correlation with Housing Prices is
 - The direct and not postponed value of Construction Permits $\sim -16\%$
 - Held over value for 9 months of Occupational Permits $\sim 44\%$

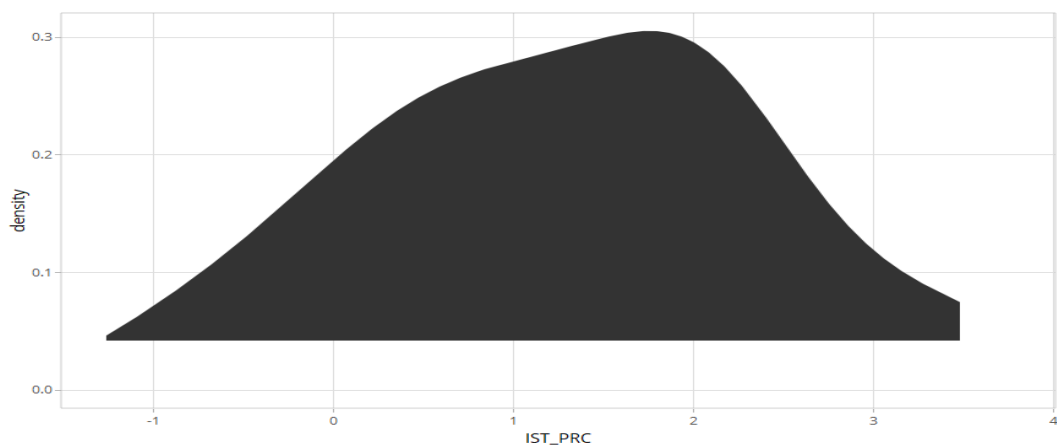
*So, It would be think that when a construction start and all permits be get , how long will this pre-construction and construction process extend or when this new supply get in the market? The study also show that **Housing Market Prices**, much more influenced by Occupational Permits which published data forward for 9 months.*

IX. Showing The Density

```
ggplotly(IST_RPPI_MODEL %>%  
  ggplot(aes(IST_PRC)) +  
  stat_density() +  
  theme_light())
```



```
ggplotly(IST_RPPI_MODEL_CG%>%  
  ggplot(aes(IST_PRC)) +  
  stat_density() +  
  theme_light())
```



The Property Price has two major range -bimodal distribution- one of the between 55-60, the other is 95-100 because of the time process .
Prices always getting increase time to time and also the monthly price changes have one modal distribution. It can be easily reach mod, median, mean values show in “Five Number Summary section”.

Istanbul Housing Price monthly change has one modal,normal distribution. Showing The Five Number Summary of Price and Price Change belowing chart.

```
summary(IST_RPPI_MODEL$IST_PRC)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2064   2775   3922   3708   4562   5123

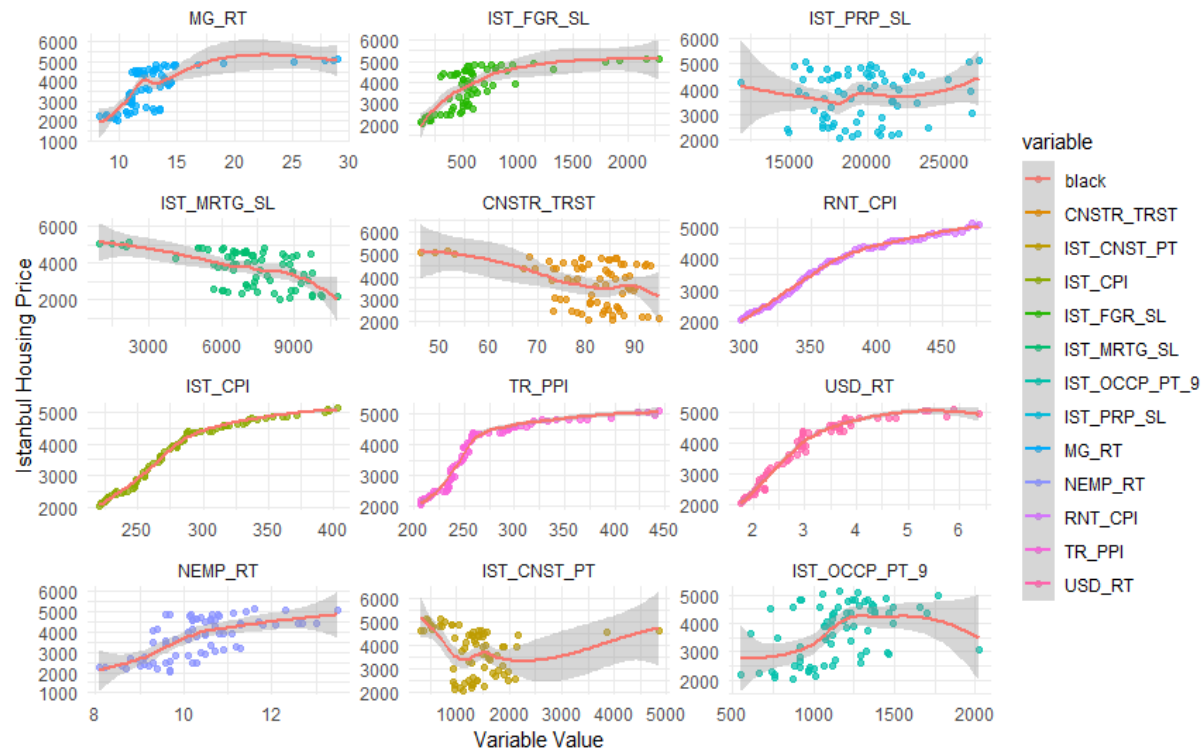
summary(IST_RPPI_MODEL_CG$IST_PRC)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     -1.266   0.463   1.336   1.285   2.084   3.758
```

X. Seeing The Correlation Plot

In order to plot all variables related with the Housing Prices :

```
IST_RPPI_MODEL %>%
  select(c(IST_PRC, MG_RT, IST_FGR_SL, IST_PRP_SL,
           IST_MRTG_SL, CNSTR_TRST, RNT_CPI,
           IST_CPI, TR_PPI, USD_RT,
           NEMP_RT, IST_CNST_PT, IST_OCCP_PT_9)) %>%
  melt(id.vars = "IST_PRC") %>%
  ggplot(aes(x = value, y = IST_PRC, colour = variable)) +
  geom_point(alpha = 0.7) +
  stat_smooth(aes(colour = "black")) +
  facet_wrap(~ variable, scales = "free", ncol = 3) +
  labs(x = "Variable Value",
       y = "Istanbul Housing Price") +
  theme_minimal()
```



The correlation between all variables *Positive & Negative Correlation* would easily showed above a plot with Istanbul Housing Prices.

Positive Correlation Negative Correlation

Mortgage Rates	Construction Trust Index
Foreigner Sales	Mortgage Sales
Rent CPI	
USD Rate	
TR PPI	
IST CPI	
NEMP Rate	

UNEMP rate is also expecting negative correlation because the purchasing power is decreasing so anyone would not want to buy a house this means demand will decrease and price should be expecting down. In this case this not go through what we expecting.

XI. Building A Model (Train-Test)

Setting a Seed Random Number

Seed number should be selected to determine the same results, accuracy, errors every run command. In this study, 2019 is specified.

```
#set.seed(2019)
#split <- sample.split(IST_RPPI_MODEL, SplitRatio = 0.70)
#train_mn <- subset(IST_RPPI_MODEL, split == FALSE)
#test_mn <- subset(IST_RPPI_MODEL, split == FALSE)
#set.seed(2019)
#split <- sample.split(IST_RPPI_MODEL_CG, SplitRatio = 0.70)
#train_cg <- subset(IST_RPPI_MODEL_CG, split == TRUE)
#test_cg <- subset(IST_RPPI_MODEL_CG, split == FALSE)
```

Split is meaning the portion of the training data. For a study train set determined 0.70 of all data set this should be increase in small data sets as well test set specified 0.30

Notes : Split and seed number didn't use for the model why the model is weak for observations ~ 72

Building Different Models and See Residuals

```
model1_mn <- lm(IST_PRC ~ MG_RT + IST_FGR_SL +
                IST_MRTG_SL + CNSTR_TRST +
                RNT_CPI + USD_RT +
                IST_CNST_PT + IST_OCCP_PT_9 ,
                data = IST_RPPI_MODEL)
summary(model1_mn)

##
## Call:
## lm(formula = IST_PRC ~ MG_RT + IST_FGR_SL + IST_MRTG_SL + CNSTR_TRST +
##     RNT_CPI + USD_RT + IST_CNST_PT + IST_OCCP_PT_9, data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -474.41 -183.41  -54.82  173.76  450.79
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.953e+03  7.345e+02  -6.743  5.6e-09 ***
## MG_RT        2.651e+01  2.520e+01   1.052   0.297
## IST_FGR_SL   -1.726e-01  2.013e-01  -0.858   0.394
## IST_MRTG_SL   3.880e-02  2.880e-02   1.347   0.183
```



```
## CNSTR_TRST      3.926e+00  4.885e+00   0.804   0.425
## RNT_CPI         2.169e+01  1.901e+00  11.414 < 2e-16 ***
## USD_RT         -1.857e+02  1.159e+02  -1.603   0.114
## IST_CNST_PT     7.382e-02  4.622e-02   1.597   0.115
## IST_OCCP_PT_9   1.591e-01  1.215e-01   1.310   0.195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 232.7 on 63 degrees of freedom
## Multiple R-squared:  0.9483, Adjusted R-squared:  0.9417
## F-statistic: 144.4 on 8 and 63 DF,  p-value: < 2.2e-16
```

Model's P-values not good what expecting, Model-1 says us only Rent CPI is affecting Price, The data including both one digit numbers and four&five digit numbers. An alternative solution to get rid of big numbers to start logarithm process for big numbers. Built a new one more model which is **Model-2 with logarithm process**.

```
model2_mn <- lm(log10(IST_PRC) ~ MG_RT + log10(IST_FGR_SL) +
                log10(IST_MRTG_SL) + CNSTR_TRST + RNT_CPI + USD_RT +
                log10(IST_CNST_PT) + log10(IST_OCCP_PT_9) ,
                data = IST_RPPI_MODEL)
summary(model2_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ MG_RT + log10(IST_FGR_SL) +
log10(IST_MRTG_SL) +
##      CNSTR_TRST + RNT_CPI + USD_RT + log10(IST_CNST_PT) +
log10(IST_OCCP_PT_9),
##      data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.077355 -0.026468 -0.003827  0.024054  0.083776
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.8108140   0.2877130   6.294 3.34e-08 ***
## MG_RT           0.0006012   0.0035592   0.169  0.8664
## log10(IST_FGR_SL) 0.0738485   0.0331655   2.227  0.0296 *
## log10(IST_MRTG_SL) 0.0595833   0.0620361   0.960  0.3405
## CNSTR_TRST      0.0005521   0.0007025   0.786  0.4349
## RNT_CPI         0.0026682   0.0002929   9.110 4.17e-13 ***
## USD_RT         -0.0322541   0.0169629  -1.901  0.0618 .
## log10(IST_CNST_PT) 0.0560448   0.0250875   2.234  0.0290 *
## log10(IST_OCCP_PT_9) 0.0624600   0.0476259   1.311  0.1945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.03452 on 63 degrees of freedom
## Multiple R-squared: 0.9298, Adjusted R-squared: 0.9208
## F-statistic: 104.2 on 8 and 63 DF, p-value: < 2.2e-16
```

Model-2 increased the p-value of variables. When looking model-2, there are four different variables in confidence p-values as well USD Rate close to line.

In Model-2 used logarithm on three or more digit values.

```
model3_mn <- lm(log10(IST_PRC) ~ log10(MG_RT) + log10(IST_FGR_SL) +
                log10(IST_MRTG_SL) + log10(CNSTR_TRST) +
                log10(RNT_CPI) + log10(USD_RT) +
                log10(IST_CNST_PT) ,
                data = IST_RPPI_MODEL)
summary(model3_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ log10(MG_RT) + log10(IST_FGR_SL) +
##     log10(IST_MRTG_SL) + log10(CNSTR_TRST) + log10(RNT_CPI) +
##     log10(USD_RT) + log10(IST_CNST_PT), data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.059546 -0.021530 -0.001008  0.017957  0.082533
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.45896    0.69953  -3.515 0.000813 ***
## log10(MG_RT)    0.17492    0.11004   1.590 0.116859
## log10(IST_FGR_SL) 0.02709    0.02995   0.905 0.369053
## log10(IST_MRTG_SL) 0.13435    0.05131   2.618 0.011015 *
## log10(CNSTR_TRST) 0.09797    0.09546   1.026 0.308622
## log10(RNT_CPI)   1.89303    0.29121   6.501 1.39e-08 ***
## log10(USD_RT)    0.03683    0.14308   0.257 0.797697
## log10(IST_CNST_PT) 0.05401    0.02128   2.539 0.013571 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02972 on 64 degrees of freedom
## Multiple R-squared: 0.9471, Adjusted R-squared: 0.9413
## F-statistic: 163.8 on 7 and 64 DF, p-value: < 2.2e-16
```

In Model-3, all variables reduced with logarithm so that decrease digits but in this model there isn't enough confidence than Model-2. So keep continuous with Model-2 and remove the highest p-value which is mortgage rates plus mortgage sales. Add the property sales rather than mortgage sales.

```

model4_mn <- lm(log10(IST_PRC) ~ log10(IST_FGR_SL) +
                log10(IST_PRP_SL) + CNSTR_TRST + RNT_CPI + USD_RT +
                log10(IST_CNST_PT) + log10(IST_OCCP_PT_9) ,
                data = IST_RPPI_MODEL)
summary(model4_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ log10(IST_FGR_SL) + log10(IST_PRP_SL) +
##     CNSTR_TRST + RNT_CPI + USD_RT + log10(IST_CNST_PT) +
log10(IST_OCCP_PT_9),
##     data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.079379 -0.023790 -0.006316  0.029584  0.051114
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.4238074   0.3063565    7.912 4.61e-11 ***
## log10(IST_FGR_SL)  0.0867433   0.0294176    2.949  0.00445 **
## log10(IST_PRP_SL) -0.1132032   0.0668606   -1.693  0.09530 .
## CNSTR_TRST        0.0007370   0.0006209    1.187  0.23957
## RNT_CPI           0.0027436   0.0002770    9.904 1.54e-14 ***
## USD_RT           -0.0413958   0.0152992   -2.706  0.00872 **
## log10(IST_CNST_PT)  0.0719461   0.0244949    2.937  0.00460 **
## log10(IST_OCCP_PT_9) 0.0652729   0.0461248    1.415  0.16188
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03394 on 64 degrees of freedom
## Multiple R-squared:  0.931, Adjusted R-squared:  0.9235
## F-statistic: 123.4 on 7 and 64 DF, p-value: < 2.2e-16

```

When considered the Construction Trust Index's p-value so higher than others. It can be remove from model, try to find more accuracy model in Model-5.

```

model5_mn <- lm(log10(IST_PRC) ~ log10(IST_FGR_SL) +
                log10(IST_PRP_SL) + RNT_CPI + USD_RT +
                log10(IST_CNST_PT) + log10(IST_OCCP_PT_9) ,
                data = IST_RPPI_MODEL)
summary(model5_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ log10(IST_FGR_SL) + log10(IST_PRP_SL) +
##     RNT_CPI + USD_RT + log10(IST_CNST_PT) + log10(IST_OCCP_PT_9),
##     data = IST_RPPI_MODEL)
##
## Residuals:

```

```
##           Min           1Q           Median           3Q           Max
## -0.083918 -0.024270 -0.006217  0.031252  0.051470
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.5382655   0.2917025    8.702 1.66e-12 ***
## log10(IST_FGR_SL)  0.0804707   0.0290301    2.772 0.007260 **
## log10(IST_PRP_SL) -0.1238142   0.0664687   -1.863 0.067018 .
## RNT_CPI           0.0028985   0.0002451   11.824 < 2e-16 ***
## USD_RT           -0.0500706   0.0134835   -3.713 0.000427 ***
## log10(IST_CNST_PT)  0.0801521   0.0235731    3.400 0.001156 **
## log10(IST_OCCP_PT_9) 0.0490817   0.0442006    1.110 0.270905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03404 on 65 degrees of freedom
## Multiple R-squared:  0.9295, Adjusted R-squared:  0.923
## F-statistic: 142.9 on 6 and 65 DF,  p-value: < 2.2e-16
```

Model-5 good correlated to explain the prices but It is not enough because of the big p-values. In model-6, get rid of the Occupational Permits.

```
model6_mn <- lm(log10(IST_PRC) ~ log10(IST_FGR_SL) +
                log10(IST_PRP_SL) + RNT_CPI + USD_RT +
                log10(IST_CNST_PT) ,
                data = IST_RPPI_MODEL)
summary(model6_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ log10(IST_FGR_SL) + log10(IST_PRP_SL) +
##     RNT_CPI + USD_RT + log10(IST_CNST_PT), data = IST_RPPI_MODEL)
##
## Residuals:
##           Min           1Q           Median           3Q           Max
## -0.07949 -0.02566 -0.00592  0.03070  0.05494
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.6541108   0.2728896    9.726 2.27e-14 ***
## log10(IST_FGR_SL)  0.0888126   0.0280907    3.162 0.002370 **
## log10(IST_PRP_SL) -0.1250391   0.0665768   -1.878 0.064784 .
## RNT_CPI           0.0029514   0.0002409   12.253 < 2e-16 ***
## USD_RT           -0.0522155   0.0133680   -3.906 0.000223 ***
## log10(IST_CNST_PT)  0.0811789   0.0235965    3.440 0.001012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0341 on 66 degrees of freedom
```

```
## Multiple R-squared:  0.9282, Adjusted R-squared:  0.9227
## F-statistic: 170.6 on 5 and 66 DF,  p-value: < 2.2e-16
```

When Supply-Demand Equilibrium considered: Property Sales expecting increase, Price would be get higher. Because of the confidence level of Property Sales the variable's marker say us the negative situation. In Model-7, removed the Property Sales.

```
model7_mn <- lm(log10(IST_PRC) ~ log10(IST_FGR_SL) +
                RNT_CPI + USD_RT +
                log10(IST_CNST_PT) ,
                data = IST_RPPI_MODEL)
summary(model7_mn)

##
## Call:
## lm(formula = log10(IST_PRC) ~ log10(IST_FGR_SL) + RNT_CPI + USD_RT +
##     log10(IST_CNST_PT), data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.083967 -0.027947 -0.005618  0.030005  0.069826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.1745802   0.0981212   22.162 < 2e-16 ***
## log10(IST_FGR_SL)  0.0722039   0.0271606    2.658 0.009810 **
## RNT_CPI          0.0029860   0.0002447   12.204 < 2e-16 ***
## USD_RT          -0.0521542   0.0136177   -3.830 0.000285 ***
## log10(IST_CNST_PT) 0.0732571   0.0236503    3.098 0.002850 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03474 on 67 degrees of freedom
## Multiple R-squared:  0.9243, Adjusted R-squared:  0.9198
## F-statistic: 204.7 on 4 and 67 DF,  p-value: < 2.2e-16
```

Model-7 is final outcome in which more comprehensive and meaningfully than other models. The model prove that :

- If Total Foreigner Sale increase than Price would be higher
- If Rent Consumer Price Index would be higher than Price get increase
- When it come to USD Rate, considered the Housing Market at this moment. USD currency extremely high and Housing Prices getting decrease. So this model actually prove that.
- Construction Permits affecting housing prices positive.
- Another different model for explaining **Mortgage Sales** :

```

model8_mn <- lm(log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
               log10(IST_PRC) + IST_RPPI +
               CNSTR_TRST + RNT_CPI + IST_CPI +
               TR_PPI + USD_RT + NEMP_RT +
               log10(IST_CNST_PT) +
               log10(IST_OCCP_PT_9) ,
               data = IST_RPPI_MODEL)
summary(model8_mn)

##
## Call:
## lm(formula = log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
##     log10(IST_PRC) + IST_RPPI + CNSTR_TRST + RNT_CPI + IST_CPI +
##     TR_PPI + USD_RT + NEMP_RT + log10(IST_CNST_PT) + log10(IST_OCCP_PT_9),
##     data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.169169 -0.023805 -0.000374  0.046476  0.182193
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.039615   3.320589   2.120  0.03822 *
## MG_RT         -0.046521   0.006010  -7.740  1.5e-10 ***
## log10(IST_FGR_SL)  0.107414   0.080741   1.330  0.18852
## log10(IST_PRC)   -1.037223   1.043871  -0.994  0.32446
## IST_RPPI        0.004044   0.009811   0.412  0.68167
## CNSTR_TRST     -0.001159   0.001750  -0.662  0.51025
## RNT_CPI        -0.006574   0.003119  -2.108  0.03928 *
## IST_CPI        0.015465   0.005552   2.786  0.00717 **
## TR_PPI        -0.006651   0.003196  -2.081  0.04180 *
## USD_RT         0.019703   0.065725   0.300  0.76540
## NEMP_RT       -0.017519   0.013246  -1.323  0.19107
## log10(IST_CNST_PT)  0.062833   0.056462   1.113  0.27029
## log10(IST_OCCP_PT_9) 0.106697   0.103936   1.027  0.30881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06588 on 59 degrees of freedom
## Multiple R-squared:  0.8997, Adjusted R-squared:  0.8793
## F-statistic: 44.1 on 12 and 59 DF, p-value: < 2.2e-16

```

The model can't explain what the study want. Remove the high p-values which are IST_RPPI and CNSTR_TRST from model.

```

model9_mn <- lm(log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
               log10(IST_PRC) +
               RNT_CPI + IST_CPI +
               TR_PPI + USD_RT + NEMP_RT +

```

```

log10(IST_CNST_PT) +
log10(IST_OCCP_PT_9) ,
data = IST_RPPI_MODEL)
summary(model9_mn)

##
## Call:
## lm(formula = log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
##     log10(IST_PRC) + RNT_CPI + IST_CPI + TR_PPI + USD_RT + NEMP_RT +
##     log10(IST_CNST_PT) + log10(IST_OCCP_PT_9), data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.178090 -0.027195 -0.001252  0.040338  0.189463
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.804990   1.307289   4.440 3.84e-05 ***
## MG_RT          -0.044687   0.005255  -8.504 6.00e-12 ***
## log10(IST_FGR_SL)  0.116321   0.075503   1.541  0.12858
## log10(IST_PRC)    -0.710977   0.494219  -1.439  0.15538
## RNT_CPI         -0.005563   0.002643  -2.105  0.03945 *
## IST_CPI          0.015315   0.005132   2.984  0.00409 **
## TR_PPI          -0.007335   0.002448  -2.996  0.00395 **
## USD_RT           0.040256   0.051781   0.777  0.43991
## NEMP_RT         -0.012860   0.010775  -1.193  0.23731
## log10(IST_CNST_PT)  0.064872   0.049501   1.311  0.19493
## log10(IST_OCCP_PT_9) 0.105658   0.102583   1.030  0.30709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06506 on 61 degrees of freedom
## Multiple R-squared:  0.8988, Adjusted R-squared:  0.8823
## F-statistic: 54.21 on 10 and 61 DF,  p-value: < 2.2e-16

```

Remove from model USD_RT, NEMP_RT, CNST & OCCP Permits, RNT_CPI

```

model10_mn <- lm(log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
log10(IST_PRC) +
IST_CPI +
TR_PPI ,
data = IST_RPPI_MODEL)
summary(model10_mn)

##
## Call:
## lm(formula = log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
##     log10(IST_PRC) + IST_CPI + TR_PPI, data = IST_RPPI_MODEL)
##

```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.187080 -0.039010  0.002511  0.045031  0.221918
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.554099    1.022878   5.430  8.7e-07 ***
## MG_RT          -0.040984    0.004293  -9.546  4.7e-14 ***
## log10(IST_FGR_SL) 0.203193    0.062403   3.256  0.00178 **
## log10(IST_PRC)   -0.591437    0.399044  -1.482  0.14306
## IST_CPI         0.005162    0.002764   1.868  0.06627 .
## TR_PPI         -0.004112    0.001765  -2.330  0.02288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06659 on 66 degrees of freedom
## Multiple R-squared:  0.8854, Adjusted R-squared:  0.8767
## F-statistic: 102 on 5 and 66 DF, p-value: < 2.2e-16
```

Final step IST_PRC and IST_CPI are not in confidence p-value, remove from model.

```
model11_mn <- lm(log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
                 TR_PPI,
                 data = IST_RPPI_MODEL)
summary(model11_mn)

##
## Call:
## lm(formula = log10(IST_MRTG_SL) ~ MG_RT + log10(IST_FGR_SL) +
##     TR_PPI, data = IST_RPPI_MODEL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.221901 -0.038148  0.006199  0.037142  0.214488
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.0644591    0.0999960  40.646 < 2e-16 ***
## MG_RT          -0.0447582    0.0037657 -11.886 < 2e-16 ***
## log10(IST_FGR_SL) 0.1971191    0.0520759   3.785 0.000327 ***
## TR_PPI         -0.0006944    0.0002850  -2.437 0.017437 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0677 on 68 degrees of freedom
## Multiple R-squared:  0.8779, Adjusted R-squared:  0.8725
## F-statistic: 163 on 3 and 68 DF, p-value: < 2.2e-16
```


Final outcome show us Property Sales with Mortgage is highly negative correlated with Mortgage Rates what It is easily expected. Turkey Producer Price Index also negative correlated with Mortgage Sales but Foreigner Sales strongly correlated with Foreigner Sales. When Foreigner Sales increase, Mortgage Sales also increasing.

For this study created and built different models. First of all in Model-1 include all variables which have logical correlation coefficients.

Housing Price Model

- Created model with the most logical variables in Model-1
- Eliminated one by one according to the P-Value.
- Housing Price Model

$$R^2 = 0.91$$

$$\begin{aligned} & \log_{10}(\text{HousingPrice}) \\ = & 0.07x\log_{10}(\text{FRGNRSL}) + 0.003x\text{RentCPI} - 0.05x\text{USDRate} + 0.07x\log_{10}(\text{CNSTPRMT}) \\ & + 2.175 \end{aligned}$$

Mortgage Sale Model

1. Model 11 is explaining the **Housing Sales By Using Loan**, all big numbers like thousand reduced with log10. To many irrelevant p-values observed in model 8. Related Index numbers removed from model.
2. P-values which is bigger than 0.05 eliminated one by one and the final model.
3. Final model

$$R^2 = 0.83$$

$$\log_{10}(\text{Sales}) = 0.197x\log_{10}(\text{FRGNRSL}) - 0.045x\text{MGRT} - 0.0007x\text{TRPPI} + 4.064$$

```
##### 9.3. Residential Property Price Index Monthly Changes - Creating
Different Models and See Residuals -
#model1_cg <- lm(IST_PRC ~ MG_RT + IST_FGR_SL + IST_PRP_SL + IST_MRTG_SL +
CNSTR_TRST + RNT_CPI + IST_CPI + TR_PPI + USD_RT + NEMP_RT + IST_CNST_PT +
IST_OCCP_PT_9 , data = train_cg)
#summary(model1_cg)

#model2_cg <- lm(IST_PRC ~ MG_RT + IST_FGR_SL + IST_PRP_SL + IST_MRTG_SL +
CNSTR_TRST + RNT_CPI + IST_CPI + TR_PPI + USD_RT + NEMP_RT + IST_CNST_PT +
IST_OCCP_PT_9 , data = train_cg)
#summary(model2_cg)

#model3_cg <- lm(IST_PRC ~ IST_FGR_SL + IST_PRP_SL + IST_MRTG_SL + RNT_CPI +
IST_CPI + TR_PPI + IST_CNST_PT + IST_OCCP_PT_9 , data = train_cg)
#summary(model3_cg)

#model4_cg <- lm(IST_PRC ~ RNT_CPI + IST_FGR_SL , data = train_cg)
#summary(model4_cg)

#This model include the change rates of Housing Unit Prices and other
variables. So try to explain Unit Prices with different undependent
variables.
```

XII. Residuals - Outliers

For all model errors below in the table according to AIC & BIC.

```
library(stats)
AIC(model1_mn,
    model2_mn,
    model3_mn,
    model4_mn,
    model5_mn,
    model6_mn,
    model7_mn,
    model8_mn,
    model9_mn,
    model10_mn,
    model11_mn)
```

```
##          df      AIC
## model1_mn 10  999.4975
## model2_mn 10 -270.0226
## model3_mn  9 -292.4694
## model4_mn  9 -273.3428
## model5_mn  8 -273.7747
## model6_mn  7 -274.4216
## model7_mn  6 -272.6729
## model8_mn 14 -173.6773
## model9_mn 12 -177.0729
## model10_mn 7 -178.0671
## model11_mn 5 -177.5276
```

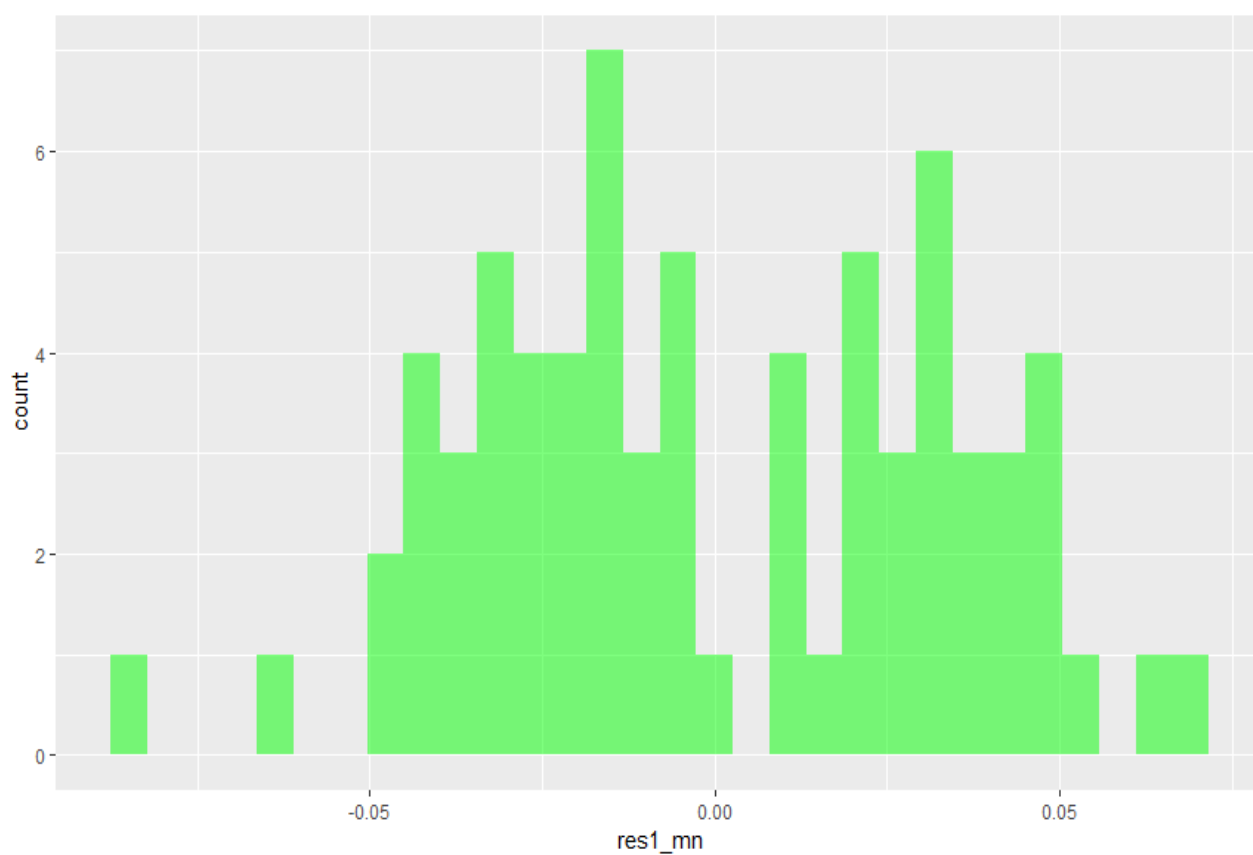
```
BIC(model1_mn,
     model2_mn,
     model3_mn,
     model4_mn,
     model5_mn,
     model6_mn,
     model7_mn,
     model8_mn,
     model9_mn,
     model10_mn,
     model11_mn)
```

```
##          df      BIC
## model1_mn 10 1022.2642
## model2_mn 10 -247.2560
## model3_mn  9 -271.9794
## model4_mn  9 -252.8529
## model5_mn  8 -255.5613
## model6_mn  7 -258.4850
## model7_mn  6 -259.0129
## model8_mn 14 -141.8039
## model9_mn 12 -149.7529
## model10_mn 7 -162.1304
## model11_mn 5 -166.1443
```

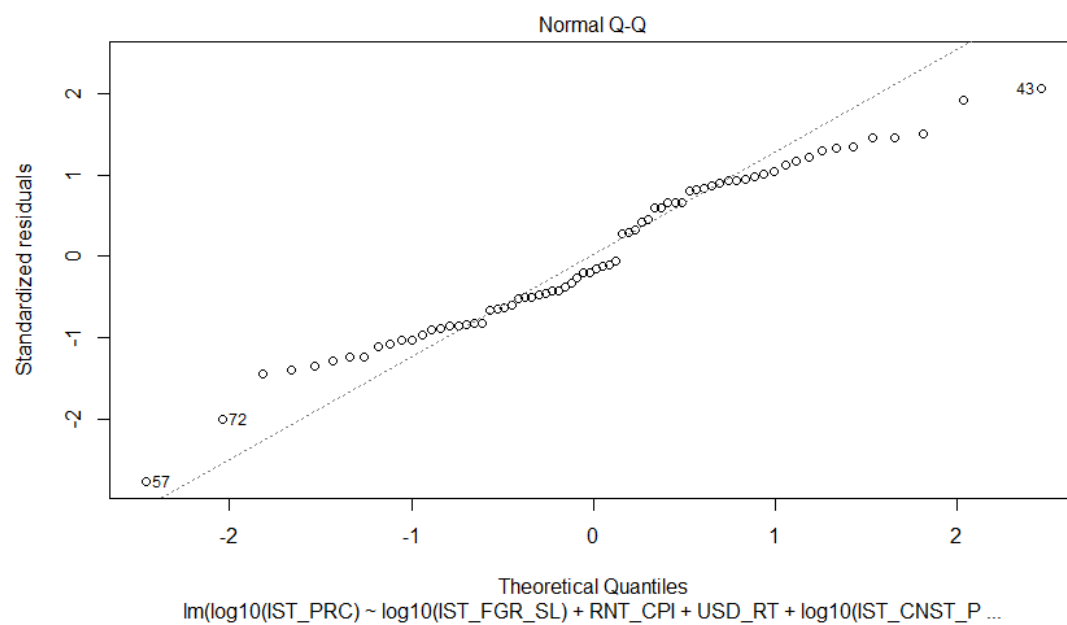
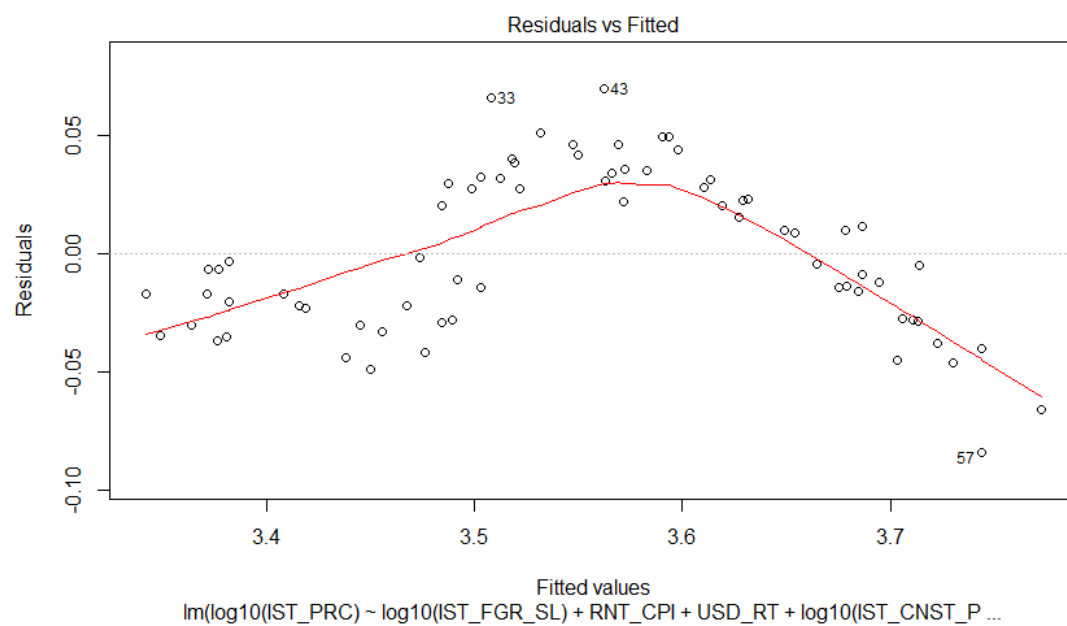
Related with Model-7 Residuals, Outliers showed on the plot.

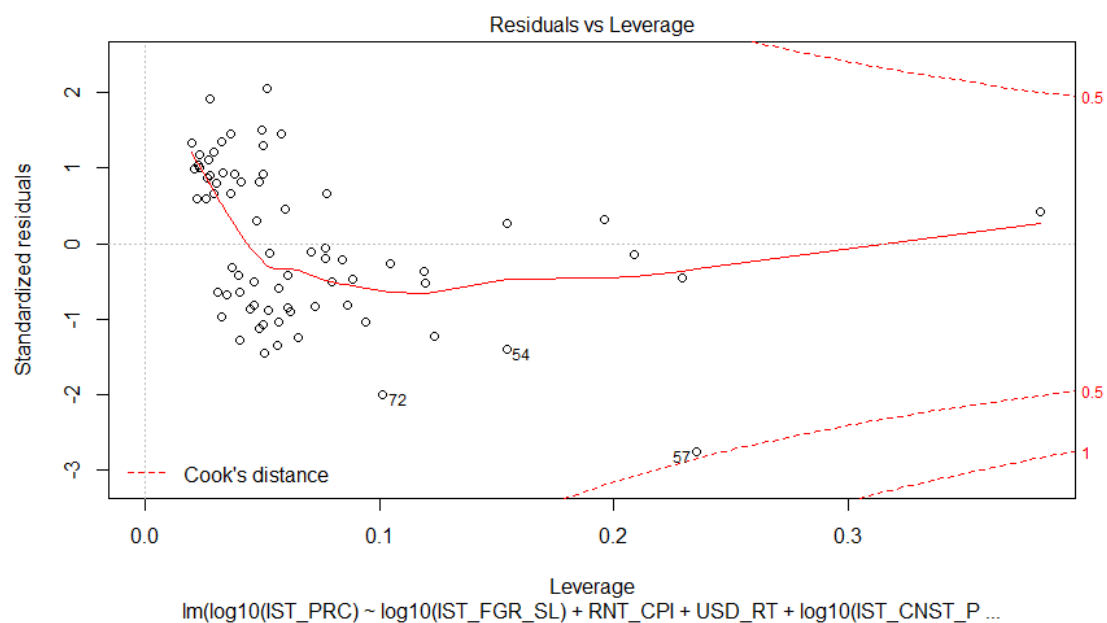
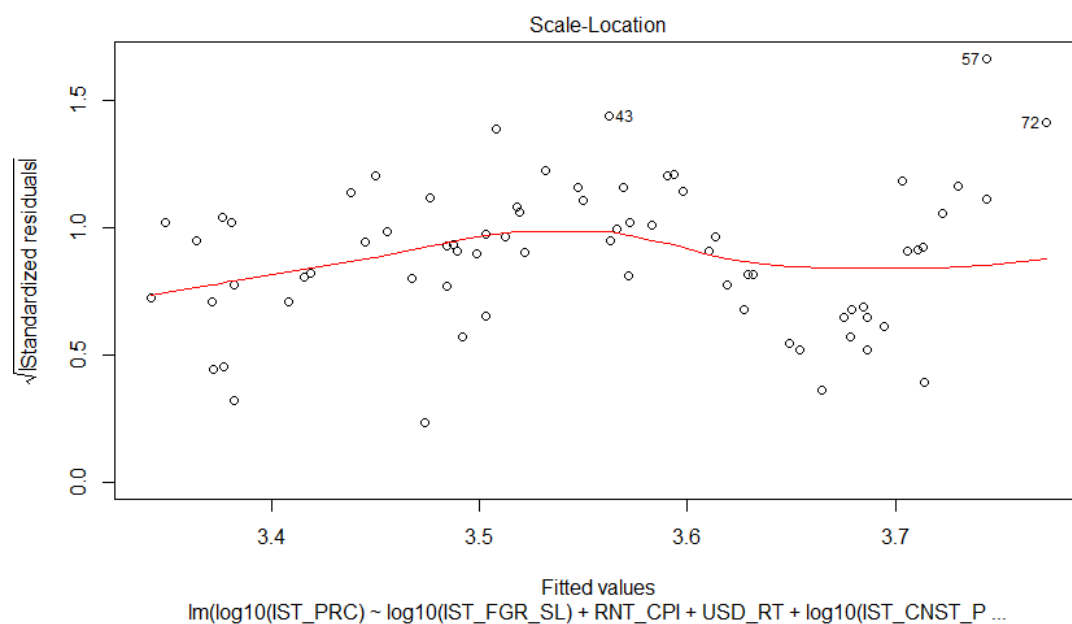
```
## model2
res1_mn <- residuals(model7_mn)
res1_mn <- as.data.frame(res1_mn)
ggplot(res1_mn, aes(res1_mn)) +
  geom_histogram(fill = 'green', alpha = 0.5)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
plot(model7_mn)
```





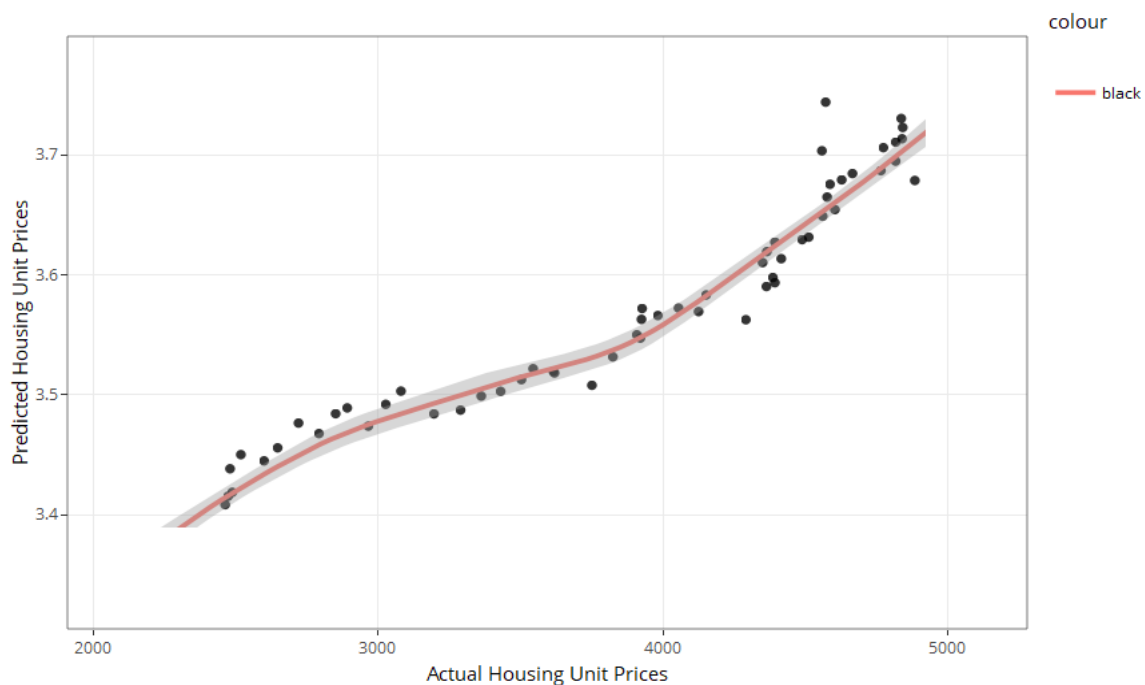
```
##model2
## res2 <- residuals(model2)
##
## res2 <- as.data.frame(res2)
##
## ggplot(res2, aes(res2)) + geom_histogram(fill = 'green', alpha = 0.5)
##
## plot(model2)
```

```

IST_RPPI_MODEL$predicted.IST_PRC <- predict(model7_mn, IST_RPPI_MODEL)
pl1_mn <- IST_RPPI_MODEL %>%
  ggplot(aes(IST_PRC, predicted.IST_PRC)) +
  geom_point(alpha = 0.80) +
  stat_smooth(aes(colour = 'black')) +
  xlab('Actual Housing Unit Prices') +
  ylab('Predicted Housing Unit Prices') +
  theme_bw()
ggplotly(pl1_mn)

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

```



XIII. Errors

Root Mean Square Error of Model-7:

```

error <- log10(IST_RPPI_MODEL$IST_PRC-IST_RPPI_MODEL$predicted.IST_PRC)
rmse <- sqrt(mean(error)^2)
data.frame(error)

##      error
## 1  3.313962
## 2  3.324350
## 3  3.333263
## 4  3.339033

```

```
## 5 3.345260
## 6 3.354024
## 7 3.361043
## 8 3.364759
## 9 3.369697
## 10 3.378417
## 11 3.392814
## 12 3.394049
## 13 3.391059
## 14 3.395388
## 15 3.400648
## 16 3.414459
## 17 3.422334
## 18 3.434210
## 19 3.445506
## 20 3.454532
## 21 3.460680
## 22 3.471628
## 23 3.480521
## 24 3.488120
## 25 3.504160
## 26 3.516639
## 27 3.526142
## 28 3.534838
## 29 3.543987
## 30 3.549053
## 31 3.558151
## 32 3.557635
## 33 3.573646
## 34 3.582072
## 35 3.592887
## 36 3.593561
## 37 3.591520
## 38 3.593327
## 39 3.599627
## 40 3.607498
## 41 3.614967
## 42 3.617730
## 43 3.632097
## 44 3.639345
## 45 3.642284
## 46 3.638012
## 47 3.639483
## 48 3.642314
## 49 3.641572
## 50 3.644485
## 51 3.651632
## 52 3.653857
## 53 3.658556
## 54 3.658292
```



```
## 55 3.660048
## 56 3.660989
## 57 3.659547
## 58 3.664869
## 59 3.662634
## 60 3.668403
## 61 3.677587
## 62 3.682265
## 63 3.678380
## 64 3.682320
## 65 3.684525
## 66 3.684302
## 67 3.684049
## 68 3.688326
## 69 3.697681
## 70 3.709240
## 71 3.703702
## 72 3.706080

data.frame(rmse)

##      rmse
## 1 3.552557
```

Tried for each observations to estimate the real and predicted value in Model-7 In this case, using logarithm process for the errors because of the model include Housing Price with Log10. So this errors say us differences between Real Housing Price and The model which built to explain housing price.

3. Time Series

Autoregressive Integrated Moving Average

In statistics and econometric, and in particular in time series analysis, an auto regressive integrated moving average (ARIMA) model is a generalization of an auto regressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where **data show evidence of non-stationarity**, where an **initial differencing step** (corresponding to the “integrated” part of the model) can be applied one or more times to eliminate the non-stationary.

The AR part of ARIMA indicates that the evolving variable of interest is **regressed on its own lagged** (i.e., prior) values. The MA part indicates that the **regression error** is actually a linear combination of *error terms* whose values occurred contemporaneously and at various times in the past. The I (for “integrated”) indicates that the data values have been replaced with the difference between their values and the previous values (and this difference process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted $ARIMA(p,d,q)$ where parameters p , d , and q are non-negative integers, p is the order (number of time lags) of the auto regressive model, d is the degree of difference (the number of times the data have had past values subtracted), and q is the order of the moving-average model. Seasonal ARIMA models are usually denoted $ARIMA(p,d,q)(P,D,Q)_m$, where m refers to the number of periods in each season, and the uppercase P, D, Q refer to the auto regressive, difference, and moving average terms for the seasonal part of the ARIMA model.

When two out of the three terms are zeros, the model may be referred to based on the non-zero parameter, dropping “AR”, “I” or “MA” from the acronym describing the model. For example, $ARIMA(1,0,0)$ is $AR(1)$, $ARIMA(0,1,0)$ is $I(1)$, and $ARIMA(0,0,1)$ is $MA(1)$.

ARIMA models can be estimated following the Box-Jenkins approach.

This analysis help us to understand the housing price changes will be in 6 months

XIV. Set The Libraries

Setting the libraries which will using on analyze.

```
library(ggplot2)
library(forecast)
library(tseries)
library(astsa)
library(readr)
```

XV. Data Import

The data include monthly changes of unit housing prices at Istanbul

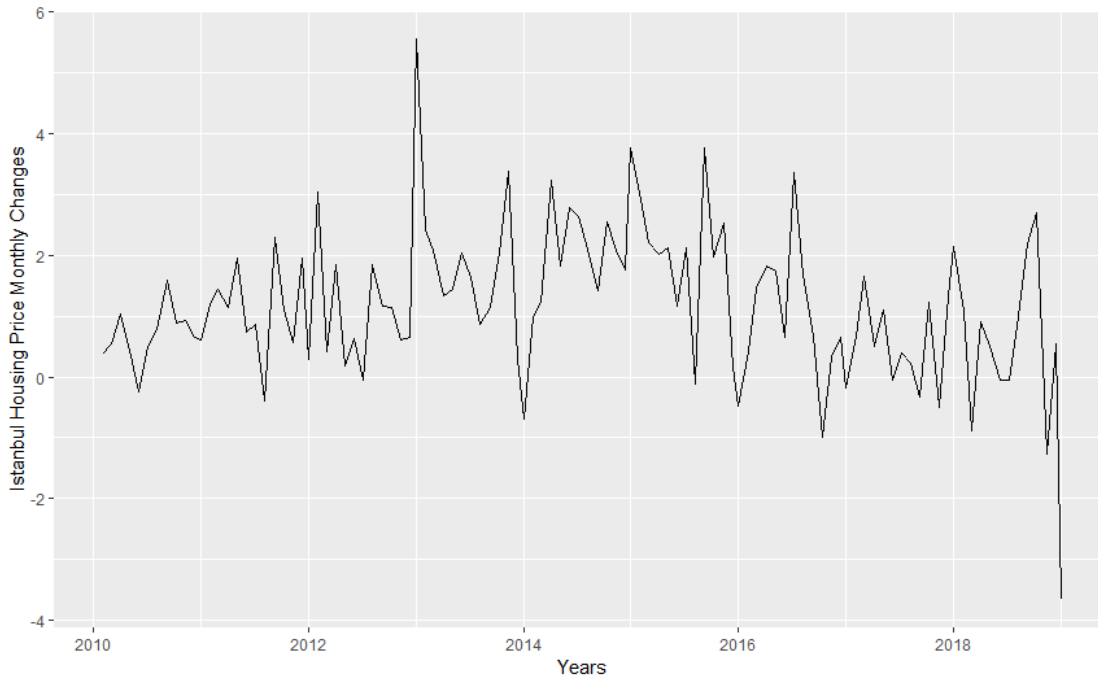
```
library(readr)
IST_UNTPRC_FRCST <- read_csv("IST_UNTPRC_FRCST.csv",
                             col_types = cols(Date = col_date(format =
"%d.%m.%Y"))))
head(IST_UNTPRC_FRCST)

## # A tibble: 6 x 2
##   DATE          IST_PRC_CG
##   <date>          <dbl>
## 1 2010-02-02      0.37
## 2 2010-03-03      0.54
## 3 2010-04-04      1.04
## 4 2010-05-05      0.44
## 5 2010-06-06     -0.25
## 6 2010-07-07      0.5
```

XVI. Show The Distribution of Housing Price Index

Distribution of Monthly Housing Price Changes:

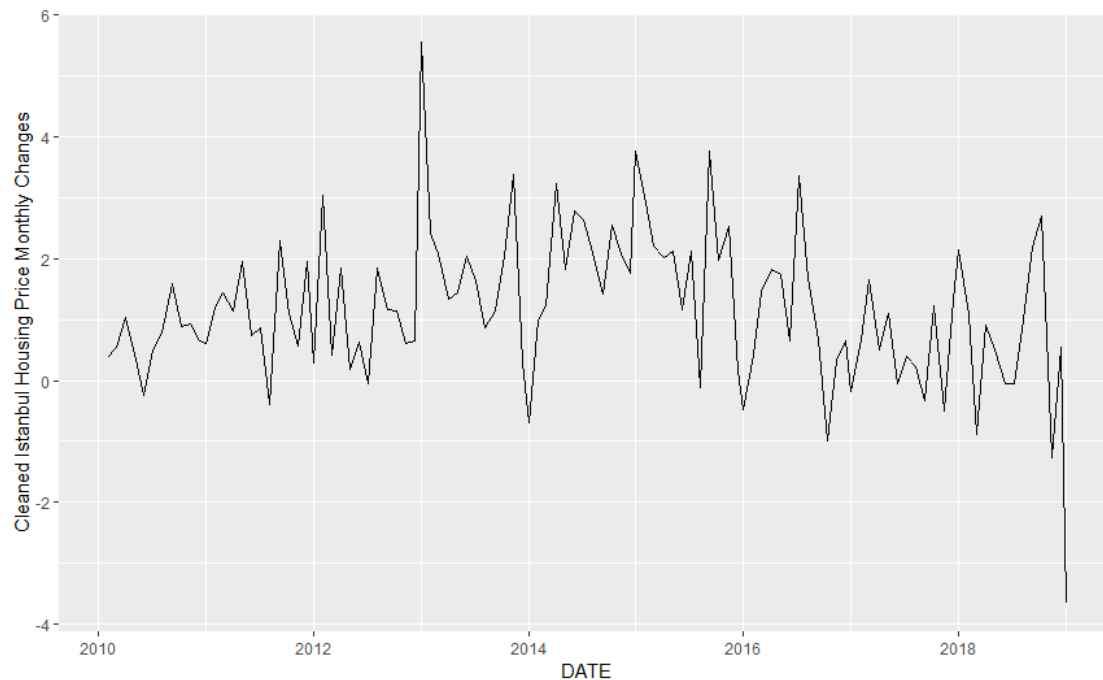
```
ggplot(IST_UNTPRC_FRCST, aes(Date, IST_PRC_CG)) +  
  geom_line() + scale_x_date('Years') +  
  ylab("Istanbul Housing Price Monthly Changes") +  
  xlab("")
```



XVII. Clean The Outliers with Tsclean

Cleaning the trend of monthly changes to get linear, handle with extreme outliers.

```
count_ts = ts(IST_UNTPRC_FRCST[, c('IST_PRC_CG')])  
IST_UNTPRC_FRCST$clean_cnt = tsclean(count_ts)  
ggplot() +  
  geom_line(data = IST_UNTPRC_FRCST, aes(x = DATE, y = clean_cnt)) +  
  ylab('Cleaned Istanbul Housing Price Monthly Changes')  
  
## Don't know how to automatically pick scale for object of type ts.  
Defaulting to continuous.
```



XVIII. Quarterly and Monthly Moving Averages

Divide into data some different periods to forecast more accuracy.

[illegible]

```

Average")) +
  geom_line(data = IST_UNTPRC_FRCST, aes(x = DATE,
                                          y = cnt_ma6,
                                          colour = "6 Monthly Moving
Average")) +
  geom_line(data = IST_UNTPRC_FRCST, aes(x = DATE,
                                          y = cnt_ma12,
                                          colour = "12 Monthly Moving
Average")) +
  ylab('Istanbul RPPI Monthly Changes')

## Don't know how to automatically pick scale for object of type ts.
## Defaulting to continuous.

## Warning: Removed 2 rows containing missing values (geom_path).

## Warning: Removed 2 rows containing missing values (geom_path).

## Warning: Removed 6 rows containing missing values (geom_path).

## Warning: Removed 12 rows containing missing values (geom_path).

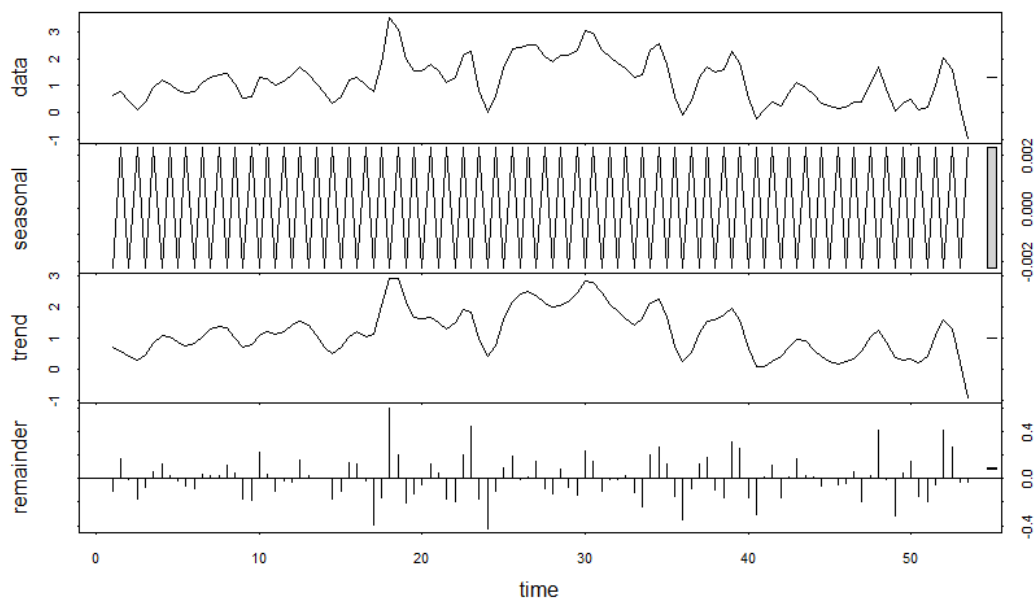
```



XIX. Decompose - Forecasting The Data

Decomposing data to get rid of seasonality. Chooses the frequency two. This means Auto Regressive Term and Moving Averages according to two months.

```
count_ma = ts(na.omit(IST_UNTPRC_FRCST$cnt_ma2),
              frequency=2)
decomp = stl(count_ma,
             s.window = "periodic")
deseasonal_cnt <- seasadj(decomp)
plot(decomp)
```



- Data : The trend isn't in balance and very variable term.
- Seasonality: Not in seasonality
- Trend: Changes are going down for the last 5-6 periods.
- Remainder: Show us the lag of data. In the 16th period.

XX. Stationarity - ADF Test

```
adf.test(count_ma,
         alternative = "stationary")

##
## Augmented Dickey-Fuller Test
```

```
##
## data: count_ma
## Dickey-Fuller = -2.5629, Lag order = 4, p-value = 0.3432
## alternative hypothesis: stationary
```

Now, testing model with ADF - Augmented Dickey Fuller Test, p-value bigger than .05

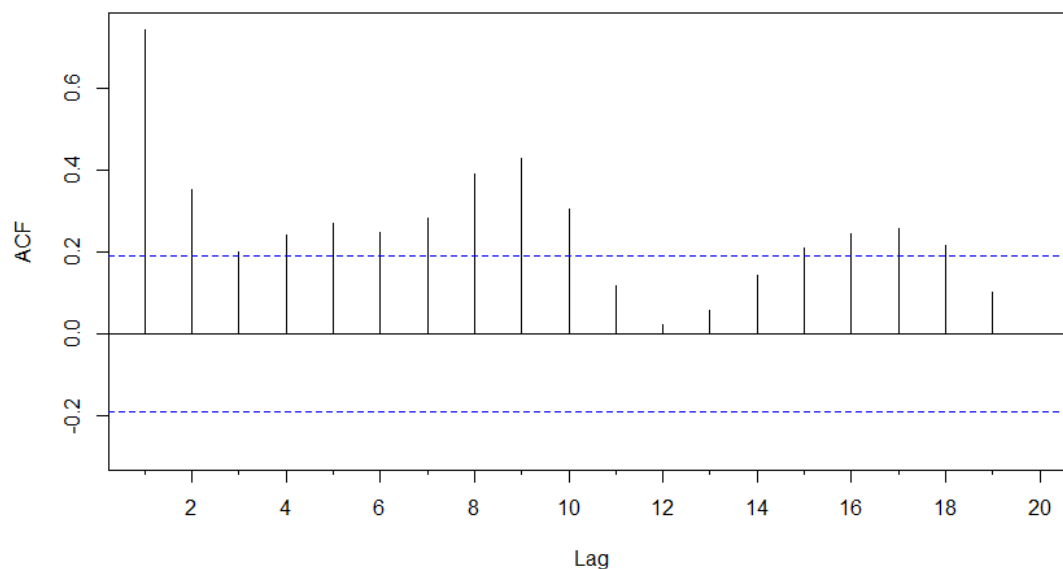
XXI. Autocorrelation and Choosing Model Order (ACF)

Auto correlation, also known as **serial correlation**, is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the **time lag** between them. The analysis of auto correlation is a mathematical tool for finding repeating patterns, such as the presence of a periodic signal obscured by noise, or identifying the missing fundamental frequency in a signal implied by its harmonic frequencies. It is often used in signal processing for analyzing functions or series of values, such as time domain signals.

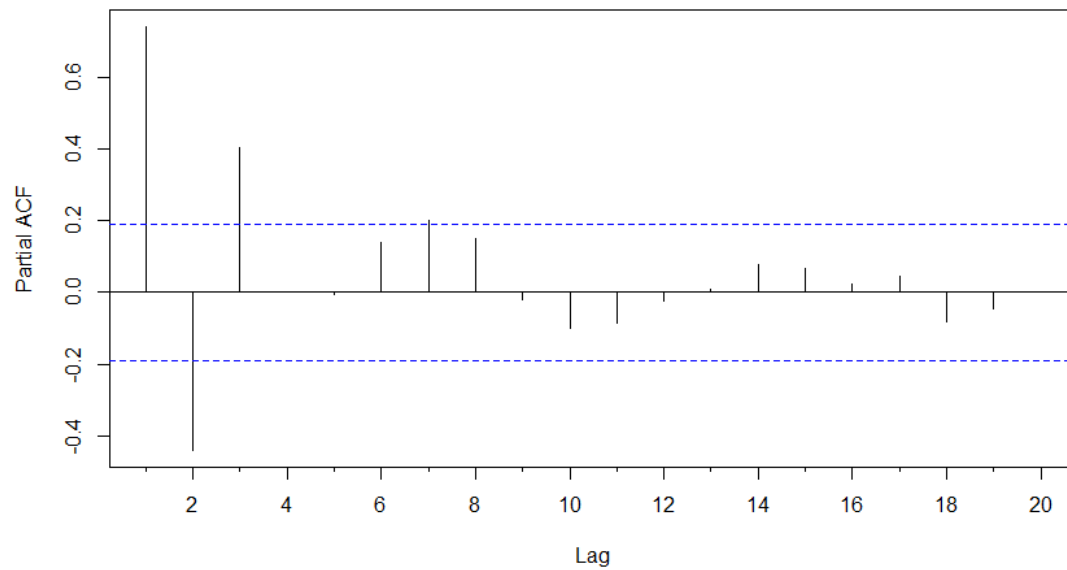
Different fields of study define autocorrelation differently, and not all of these definitions are equivalent. In some fields, the term is used interchangeably with autocovariance.

Unit root processes, trend stationary processes, auto regressive processes, and moving average processes are specific forms of processes with auto correlation.

```
##autocorrelation
Acf(count_ma, main = "", plot = TRUE)
```




```
Pacf(count_ma, main = "", plot = TRUE)
```

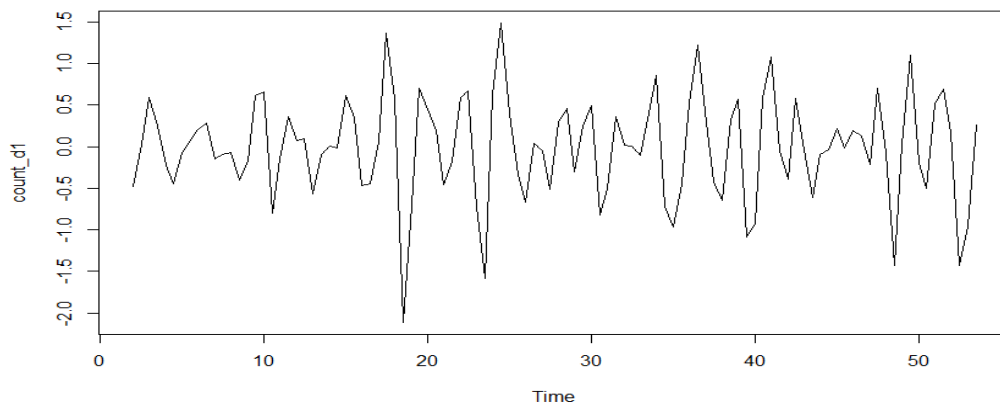


```
#Acf(count_ma, main = "", plot = FALSE)  
#Pacf(count_ma, main = "", plot = FALSE)
```

There are no stationary in the model, Lags are keeping decrease untill 18th periods.

XXII. Tune The Model

```
count_d1 = diff(deseasonal_cnt, differences = 2)  
plot(count_d1)
```



```

adf.test(count_d1, alternative = "stationary" )

## Warning in adf.test(count_d1, alternative = "stationary"): p-value smaller
## than printed p-value

##
## Augmented Dickey-Fuller Test
##
## data: count_d1
## Dickey-Fuller = -6.2587, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary

```

If the p value was bigger than 0.05, Differences should be use again. The model's p-value ~ .01

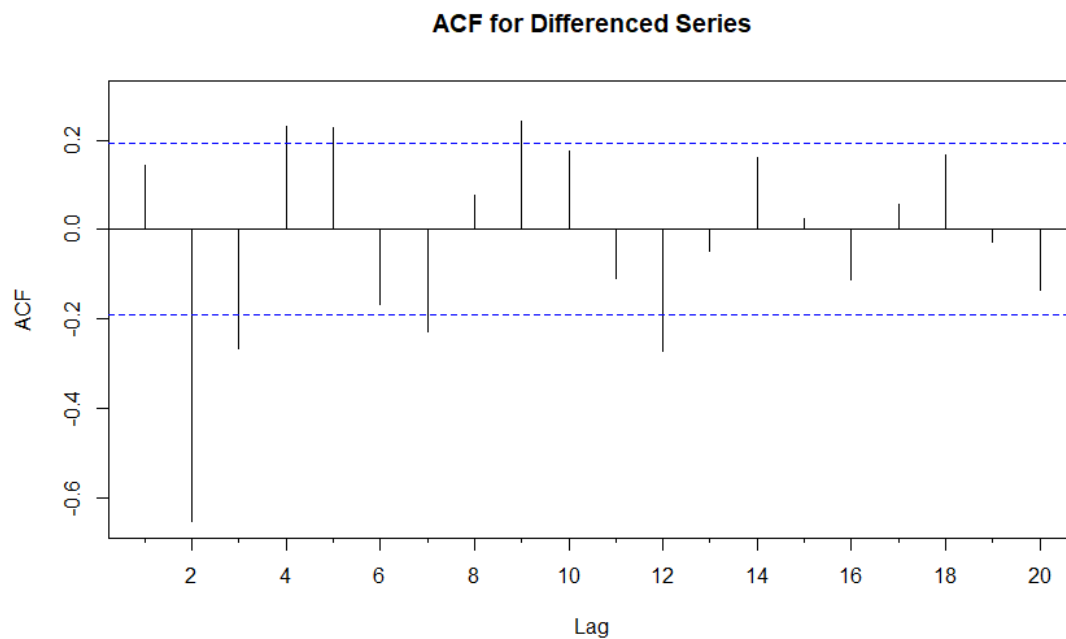
This process using errors to forecast previous periods.

```

## count_d2 = diff(count_d1, differences = 1)
## plot(count_d2)
## adf.test(count_d2, alternative = "stationary" )

Acf(count_d1,
    main = "ACF for Differenced Series")

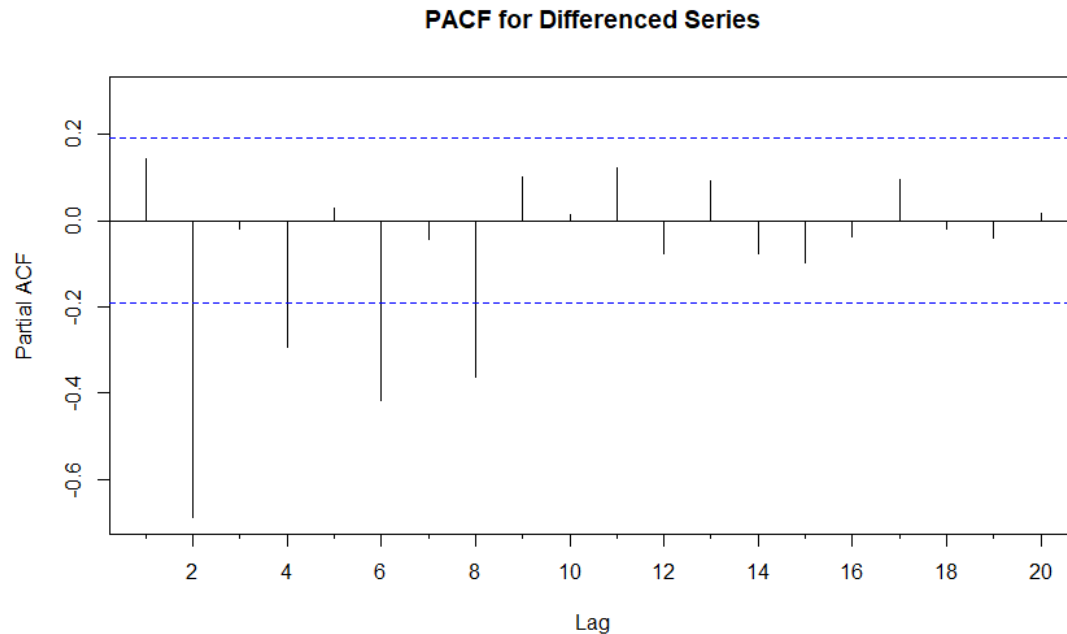
```



```

Pacf(count_d1,
    main = "PACF for Differenced Series")

```



```
#Acf(count_d1,
#   main = "ACF for Differenced Series", plot = FALSE)
#Pacf(count_d1,
#   main = "PACF for Differenced Series", plot = FALSE)
```

Lags can be observed in 3rd months because of the model array order is 3. To justify the model lag max should be 3. At 9th month's lag isn't in trust line. Fitting and Autoarima should be try.

XXIII. Fitting an ARIMA Model

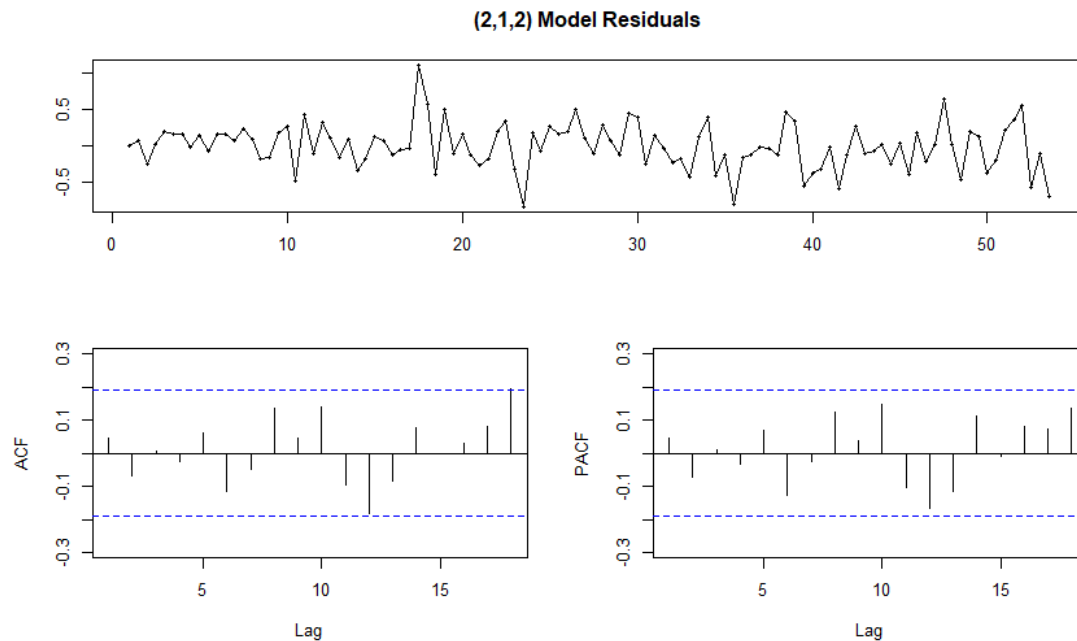
```
auto.arima(deseasonal_cnt, seasonal = FALSE)

## Series: deseasonal_cnt
## ARIMA(2,1,2)
##
## Coefficients:
##      ar1      ar2      ma1      ma2
##      0.7675 -0.5081  0.0884 -0.7289
## s.e. 0.0983  0.0912  0.0804  0.0771
##
## sigma^2 estimated as 0.1038: log likelihood=-29.71
## AIC=69.42  AICc=70.03  BIC=82.69
```

Auto arima say to model use Auto Regressive Term and Moving Average Two, Difference period is one.

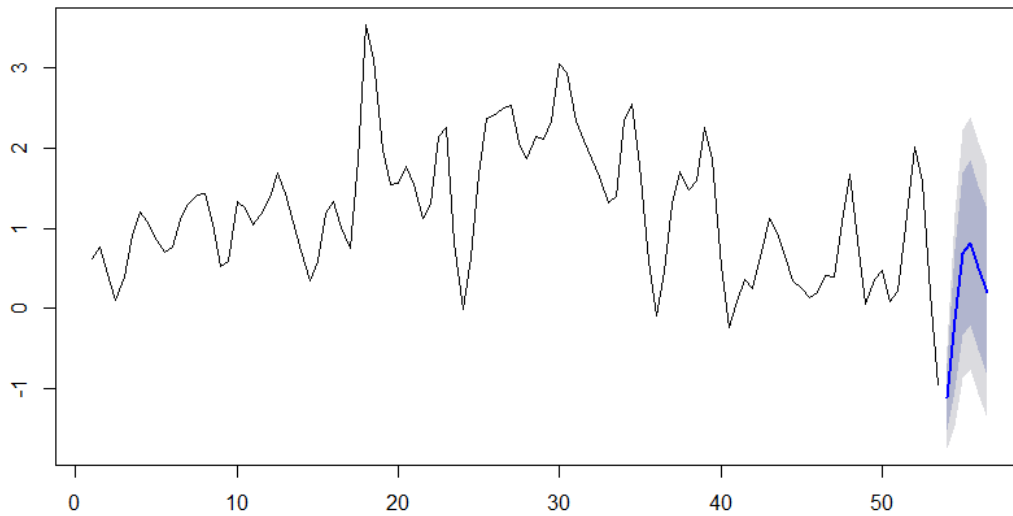
XXIV. Evaluate and Iterate

```
fit1 <- auto.arima(deseasonal_cnt,  
                  seasonal = FALSE)  
tsdisplay(residuals(fit1), lag.max = 18,  
          main = "(2,1,2) Model Residuals")
```

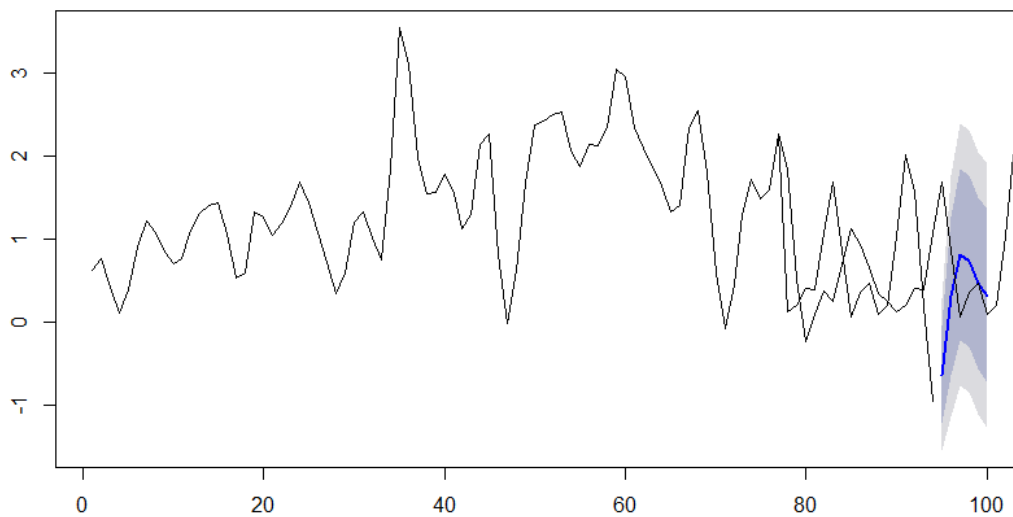


```
#fit2 = arima(deseasonal_cnt, order = c(1,1,1))  
#fit2  
#tsdisplay(residuals(fit2), lag.max = 12, main = "Seasonal Model Residuals")  
  
fcast <- forecast(fit1, h = 6)  
plot(fcast)
```

Forecasts from ARIMA(2,1,2)



```
hold <- window(ts(deseasonal_cnt), start = 60)
fit_no_holdout = arima(ts(deseasonal_cnt[-c(78:89)]),
                      order = c(2,1,2))
fcast_no_holdout <- forecast(fit_no_holdout, h = 6)
plot(fcast_no_holdout, main = "")
lines(ts(deseasonal_cnt))
```



```
fit_w_seasonality = auto.arima(deseasonal_cnt, seasonal=TRUE)
fit_w_seasonality
```

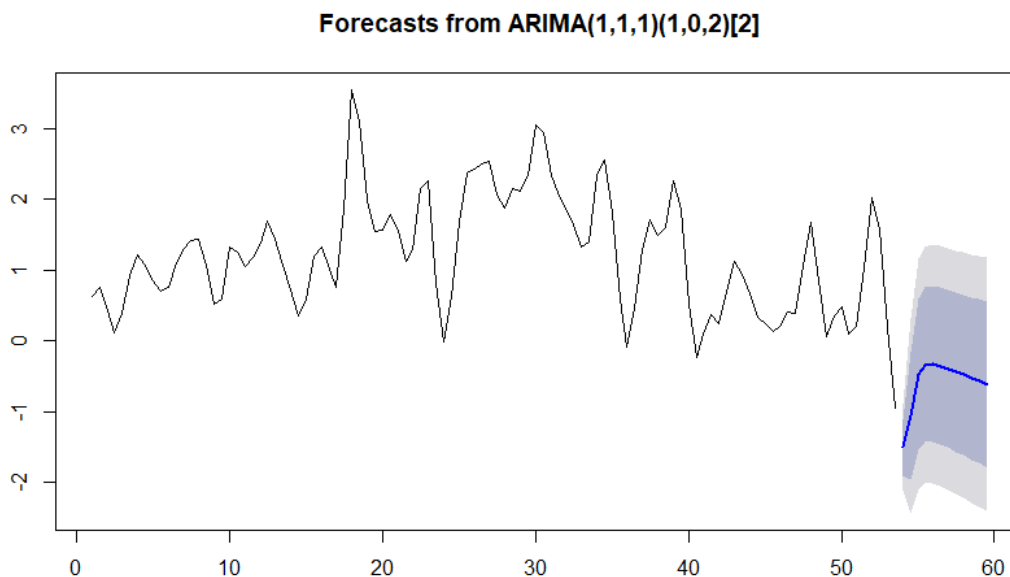
```
## Series: deseasonal_cnt
## ARIMA(1,1,1)(1,0,2)[2]
##
## Coefficients:
##          ar1      ma1      sar1      sma1      sma2
##      0.2992  0.7402  0.9749 -1.8786  0.9037
## s.e.  0.1703  0.1726  0.0448  0.0838  0.0823
##
## sigma^2 estimated as 0.09516:  log likelihood=-26.99
## AIC=65.98   AICc=66.84   BIC=81.9
```

XXV. Forecast The Istanbul Housing Price Changes

Actual Value of The Price Index

DATE	IST RPPI	CHANGE	UNIT PRICES	UNIT PRICE CHANGE
1.01.2019	100.15	-1.84	4900.15	-3.66
2.02.2019	99.63	-0.52	4702.03	-4.04
3.03.2019	99.34	-0.29	4657.59	-0.95

```
seas_fcast <- forecast(fit_w_seasonality, h=12)
plot(seas_fcast)
```



```
summary(seas_fcast)
```

```

##
## Forecast method: ARIMA(1,1,1)(1,0,2)[2]
##
## Model Information:
## Series: deseasonal_cnt
## ARIMA(1,1,1)(1,0,2)[2]
##
## Coefficients:
##          ar1      ma1      sar1      sma1      sma2
##          0.2992  0.7402  0.9749  -1.8786  0.9037
## s.e.      0.1703  0.1726  0.0448   0.0838  0.0823
##
## sigma^2 estimated as 0.09516:  log likelihood=-26.99
## AIC=65.98   AICc=66.84   BIC=81.9
##
## Error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.02015907 0.2996248 0.2265602 -2.891127 35.16316 0.3256143
##              ACF1
## Training set 0.009417765
##
## Forecasts:
##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 54.00      -1.5099912 -1.905514 -1.1144684 -2.114891 -0.9050913
## 54.50      -1.0575998 -1.955976 -0.1592238 -2.431547  0.3163477
## 55.00      -0.4742754 -1.539578  0.5910271 -2.103515  1.1549642
## 55.50      -0.3413893 -1.432975  0.7501962 -2.010825  1.3280466
## 56.00      -0.3257026 -1.427128  0.7757226 -2.010187  1.3587819
## 56.50      -0.3616087 -1.470237  0.7470197 -2.057109  1.3338921
## 57.00      -0.3958208 -1.511893  0.7202516 -2.102706  1.3110646
## 57.50      -0.4456388 -1.569952  0.6786740 -2.165127  1.2738493
## 58.00      -0.4834252 -1.617247  0.6503963 -2.217455  1.2506051
## 58.50      -0.5333204 -1.677822  0.6111815 -2.283685  1.2170442
## 59.00      -0.5705563 -1.727195  0.5860823 -2.339482  1.1983697
## 59.50      -0.6193195 -1.789356  0.5507173 -2.408736  1.1700974

```

4. The Result

The result of the study “Housing Prices in Istanbul” are more impose by **Consumer and Producer Index related with Rent, USD Currency Rate, Construction Permits and Foreigner Sales**. Some Index and Rates directly affecting Prices&Sales. The final model below with the equation:

$$\begin{aligned} & \log_{10}(\text{HousingPrice}) \\ = & 0.07x\log_{10}(\text{FRGNRSL}) + 0.003x\text{RentCPI} - 0.05x\text{USDRate} + 0.07x\log_{10}(\text{CNSTPRMT}) \\ & + 2.175 \end{aligned}$$

$$\log_{10}(\text{Sales}) = 0.197x\log_{10}(\text{FRGNRSL}) - 0.045x\text{MGRT} - 0.0007x\text{TRPPI} + 4.064$$

Regression analysis proved that the other independent variables may affecting less than consumer and procuder index.

Time Series Analysis also proved that Housing Market in Istanbul will be negative growth for 6 months as well will be decreasing The Housing Prices for 6 months.

5. References

1. Linear Regression, Selva Prabhakaran
 2. Linear Regression with Boston Housing Data set, Sukesh Kumar Pabba
 3. Introduction to Forecasting with ARIMA in R, Ruslana Dalinina
 4. https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average
 5. [https://en.wikipedia.org/wiki/Autocorrelation\](https://en.wikipedia.org/wiki/Autocorrelation)
 6. Overview Of The Real Estate And Housing Industry , May 2018
 7. Kaya, Asli , “Determining of The Factors Affects Housing Price in Turkey with Hedonic Price Model”, CBRT Expertise Qualification Thesis Unpublished,Ankara , 2012
-