Serhat Ekli
2020402051

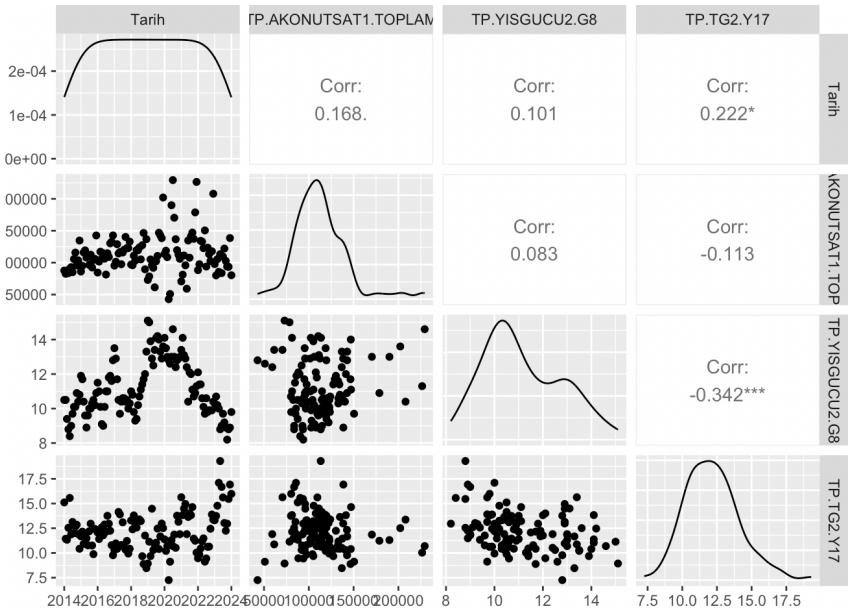
# HW1-REPORT

## Introduction

In this homework, I tried to demonstrate my time series data manipulation and regression skills on the data I chose from https://evds2.tcmb.gov.tr/ and https://trends.google.com/trends/. From EVDS, I chose 3 types of series, which are unemployment rate, people's expectations of buying a car in the next 12 months, and total house sales statistics in Turkey. I chose 2 possible variables for each one of them that may have a effect on those. In addition to those variables, I use the Google trends data by choosing a relevant keyword and try to predict the series.

## Data Manipulation and Visualization

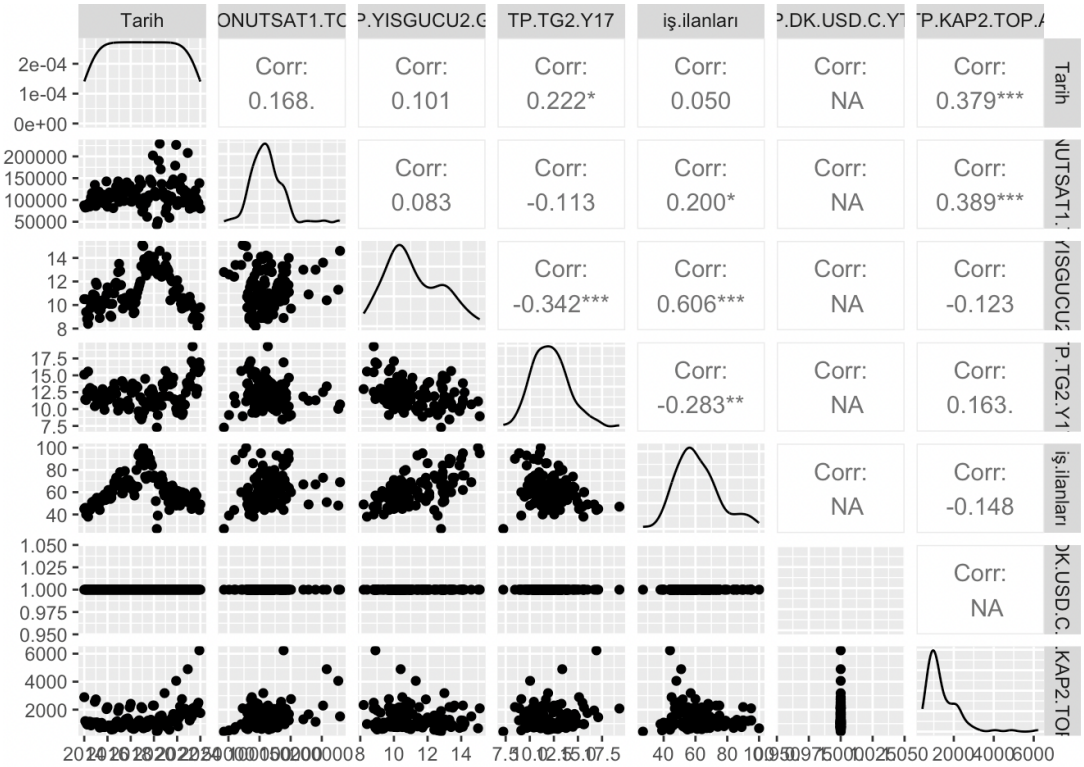
First of all, I checked whether the time series I selected are correlated or not. For this, I made a correlation matrix for the variable named "data", which includes the time series that I selected. After printing it "ggpairs(data)", I saw that their absolute correlation coefficients are less than 0.5.



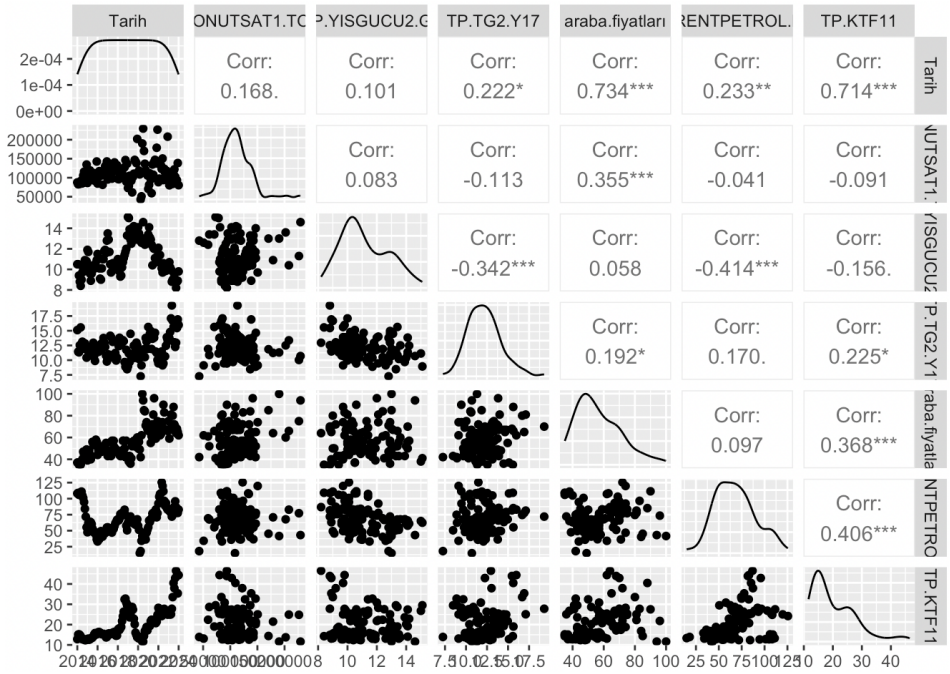
Then, I loaded the “issizlik_orani”, “araba_alma_beklentisi”, and “konut_satis” data from EVDS, and csv files from Google Trends; and made some modification in order to utilize them. Below, you can find the variables that I used for each time series.

- issizlik_orani: “TP.DK.USD.C.YTL”, which is US dollars cross rate; “TP.KAP2.TOP.A”, which is the total number of companies closed; and “iş ilanları” keyword
- araba_alma_beklentisi: “TP.BRENTPETROL.EUBP”, which is Europe Brent Spot Price FOB; “TP.KTF11”, which is vehicle loan interest rate; and “araba fiyatları” keyword
- konut_satis: “TP.DK.EUR.C.YTL”, which is Euro cross rate; “TP.KTF12”, which is mortgage interest rate; and “konut kredisi” keyword

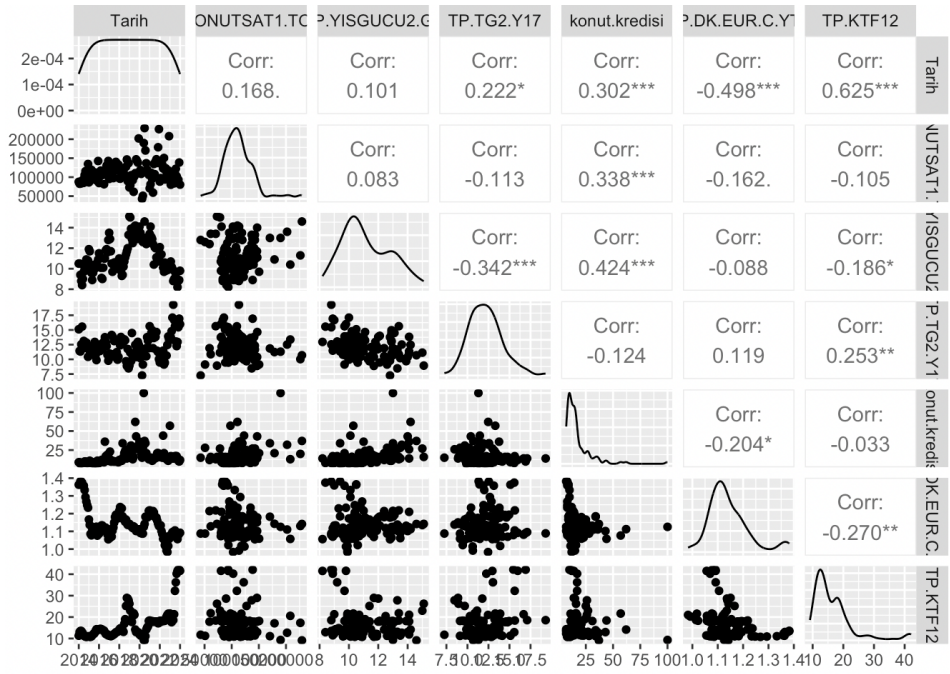
By looking at the consolidated correlation matrices, I found out that unemployment rate is correlated with the “iş ilanları” Google search. Also, the total numbers of companies closed is negatively correlated with unemployment rate, which is not surprising; but, it is not that significant, just -0.123.



Even though they are not significant, people's expectations of buying a car in the next 12 months are positively correlated with the brent price, vehicle loan interest rate, and "araba fiyatları" Google search.




While house sales statistics is positively correlated with "konut kredisi" Google search, it is negatively correlated with Euro cross rate, and mortgage interest rate.

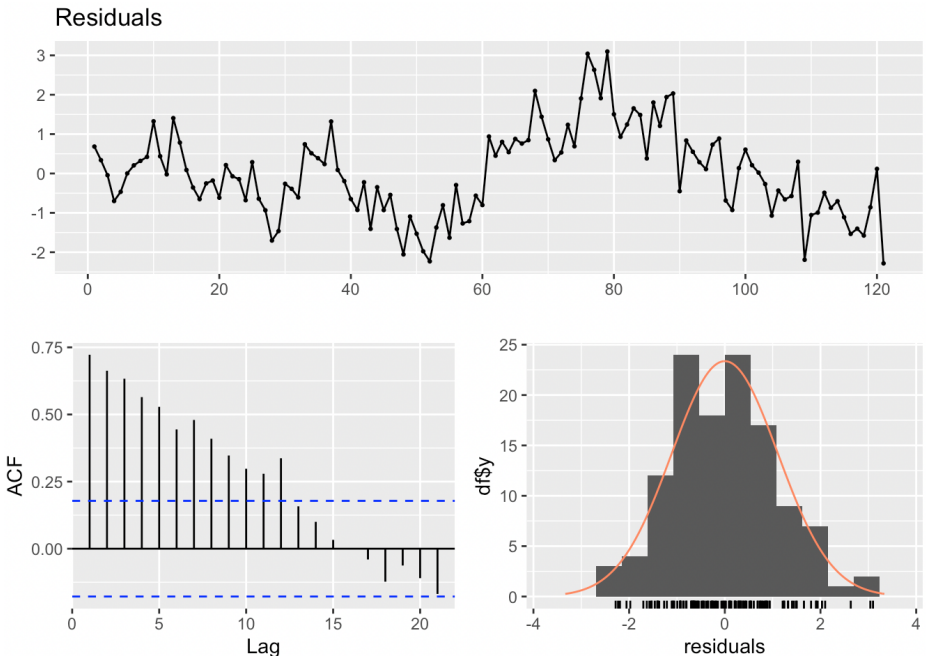


Serhat Ekli
2020402051

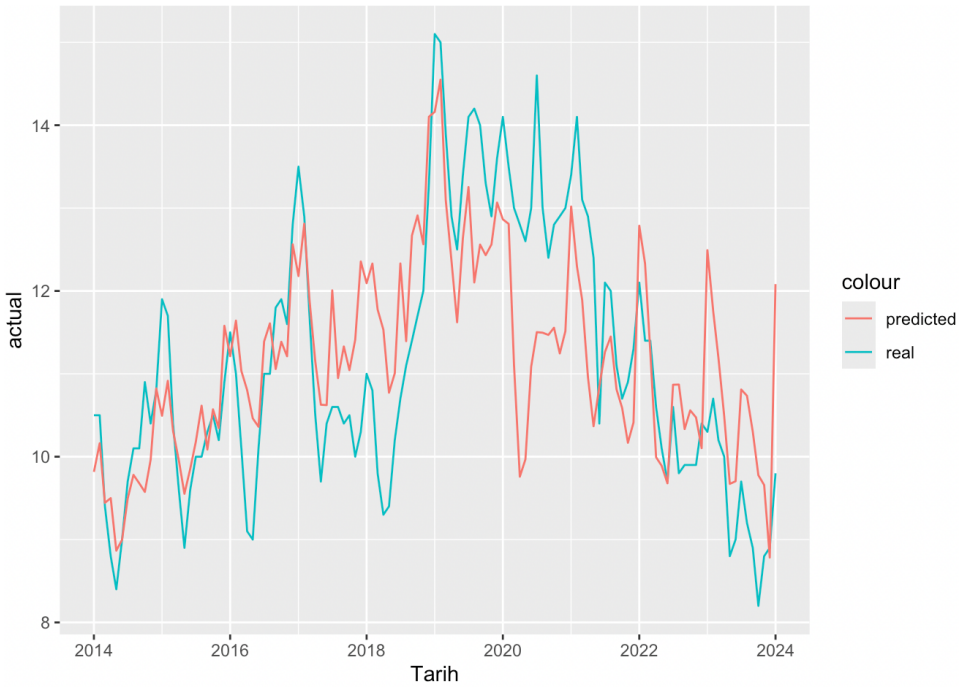
## Time Series Regression Analyses

In this part, I handled the data preprocessing and visualization tasks for unemployment rates, expectation of buying a car, and house sales, by using the other time series that I had. I converted the dataset into a data table, by adding trend, year, and month columns. Also, I fitted LM model for each of them, checked the residuals and ACF plots to assess the adequacy of the regression models, implemented the Ljung-Box test, and visualized the actual and predicted trends over time.

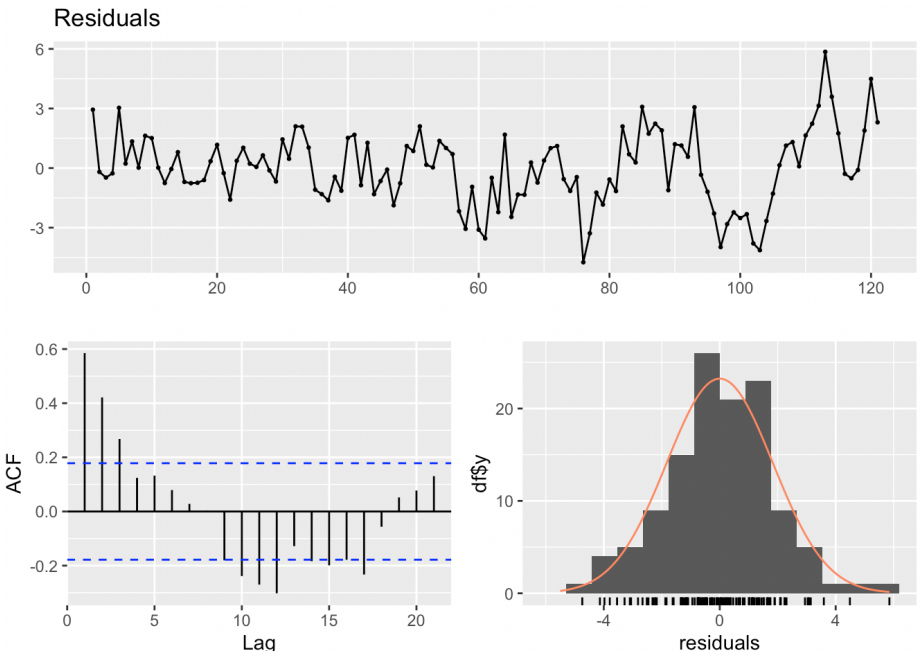
For the unemployment rate data, I had an R-squared value of 0.5223, which means that 52% of the variation in the dependent variable can be explained by the independent variables. However, when I looked at the residuals, even though they may be regarded as normally distributed, I saw that there is a strong autocorrelation, which means that making predictions and inferences may be misleading.



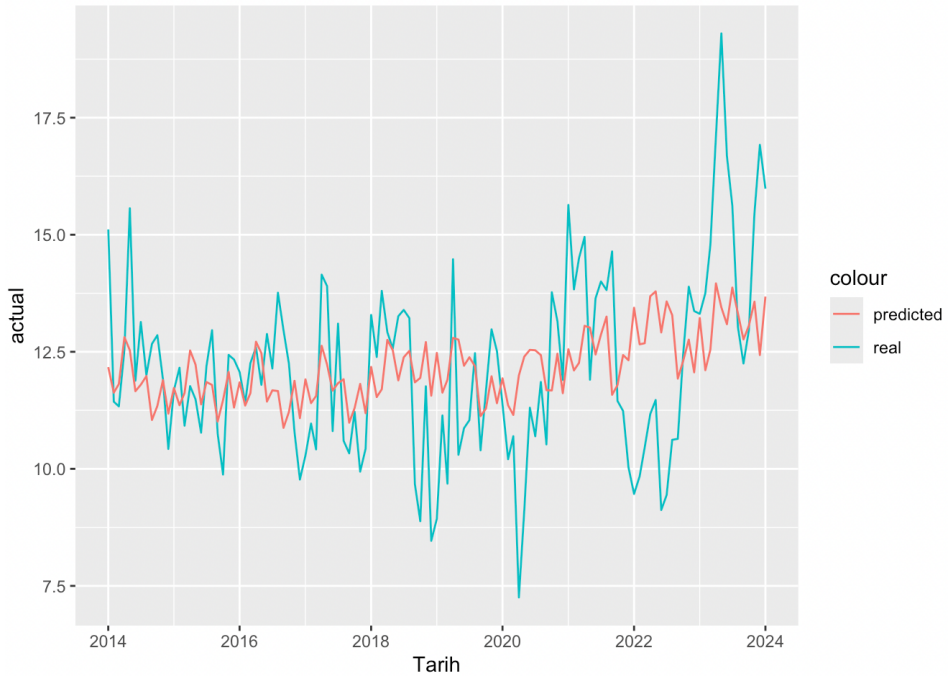
Serhat Ekli
2020402051



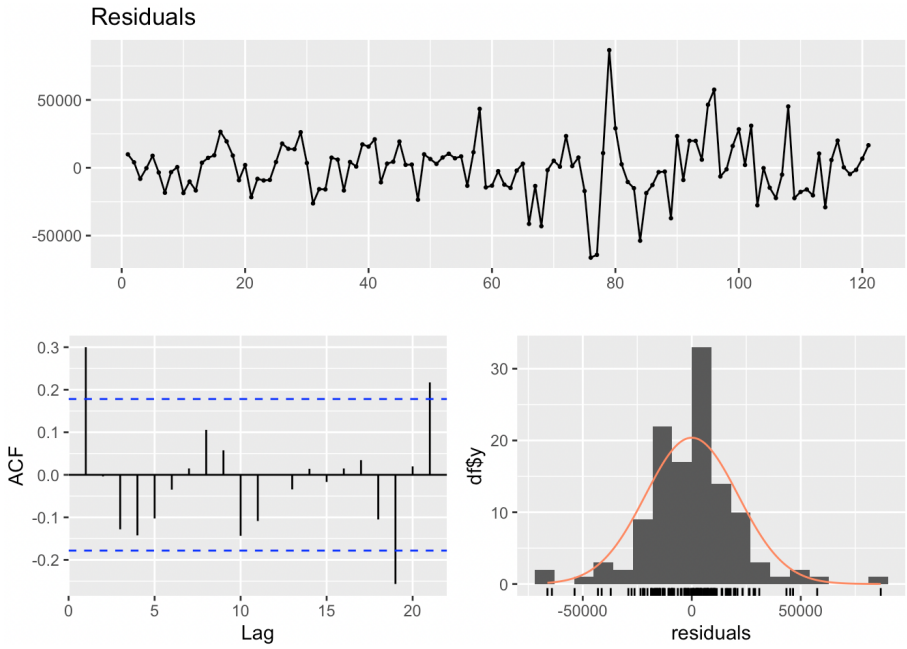

For the people's expectations of buying a car in the next 12 months, residual graphs seemed okay but I had a very low R-squared value. Also, when I looked at predicted and real data, it can be understood that predicted and real values do not move together, and it is an unreliable model.



Serhat Ekli
2020402051




For the house sales data, I again had an R-squared value of 0.51, and there was not a strong autocorrelation. When I looked at the actual and predicted graph, I saw that the predictions moved together with the actual data when there was an up. Also, for the real values, there was a strong down and up in 2020, which may be explained by the lockdown for Covid-19 pandemics.



Serhat Ekli
2020402051


## Conclusion

This homework provided me a valuable insight into time series data manipulation and regression topics. Being exposed to R programming language during the homework, I learned how to prepare and visualize the data. Also, by making some comments on time series and forecasting, I believe that I become more familiar with the lecture topics.