

# DSM 5012 Home Work –II

Fırat Melih Yılmaz

201990092

## 1: Gerekli Paketlerin Yüklenmesi

```
library(tidyverse) # Easily Install and Load the 'Tidyverse', CRAN v1.3.0
library(zoo) # S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations), CRAN v2.3
library(fpp2) # Data for "Forecasting: Principles and Practice" (2nd Edition), CRAN v2.3
library(forecast) # Forecasting Functions for Time Series and Linear Models, CRAN v8.12
library(gridExtra) # Miscellaneous Functions for "Grid" Graphics, CRAN v2.3
library(stargazer) # Well-Formatted Regression and Summary Statistics Tables, CRAN v5.2.2
library(xtable) # Export Tables to LaTeX or HTML, CRAN v1.8-4
```

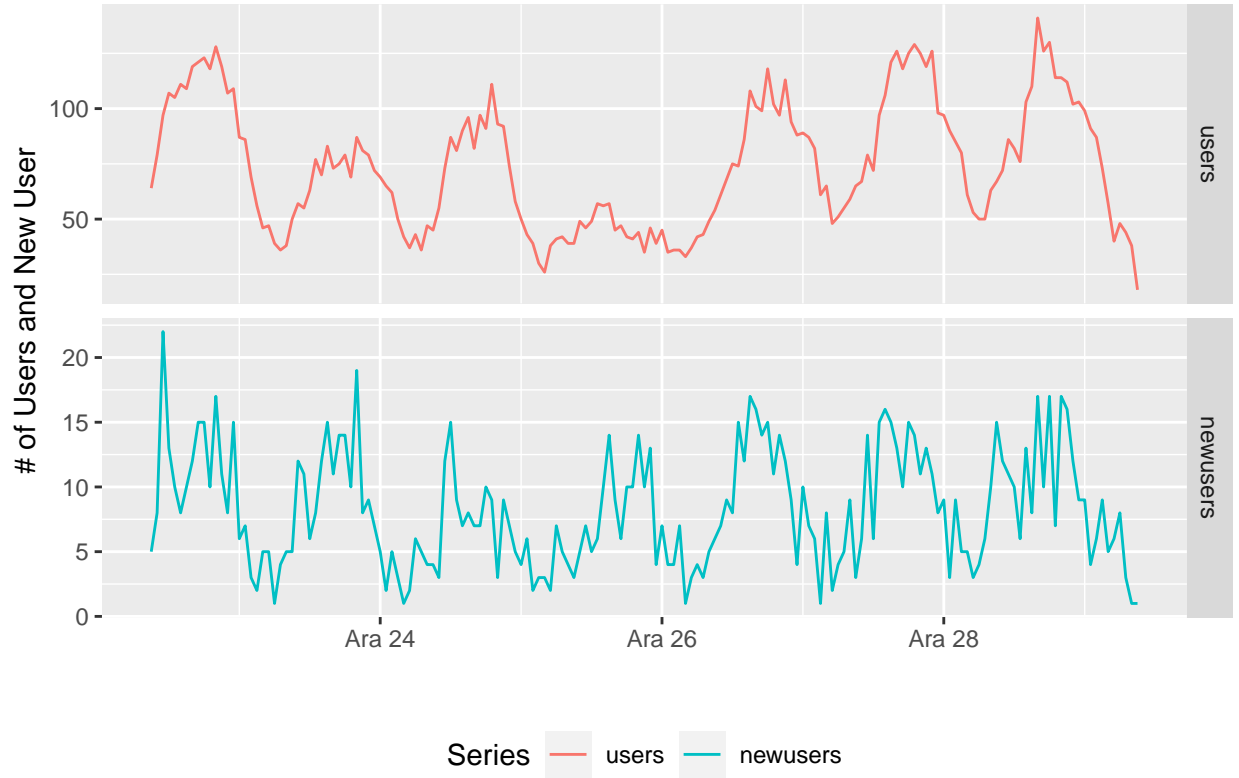
## 2: Verilerin Yüklenmesi ve Düzenlenmesi

```
users_data <- read.csv('data/timeseries.csv')
users_data <- users_data[,-1]
users_date <- as.POSIXct(as.character(users_data$time, format="%Y-%m-%d %H:%M"))
users_data <- zoo(users_data[,c('users', 'newusers')], order.by = users_date)
```

## 3: Veriler üzerinde Keşifçi Veri Analizi uygulanması yapılması.

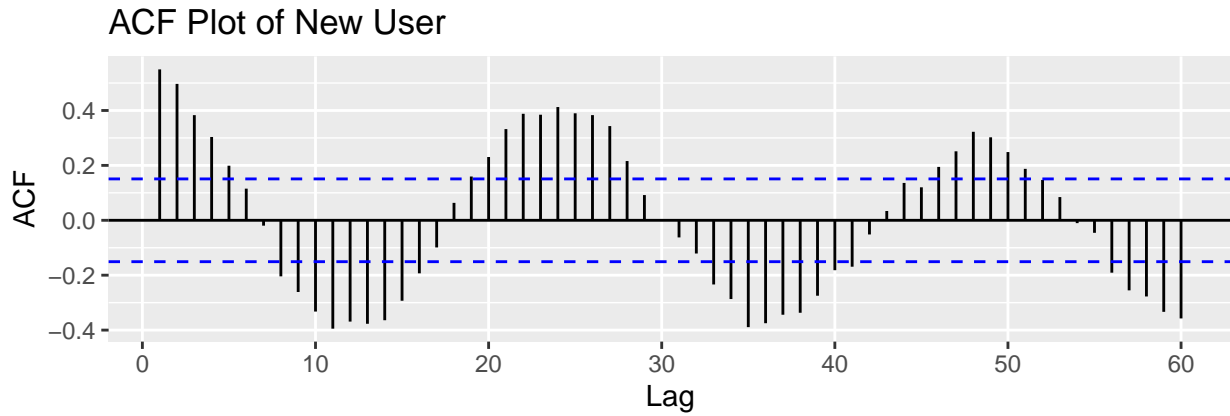
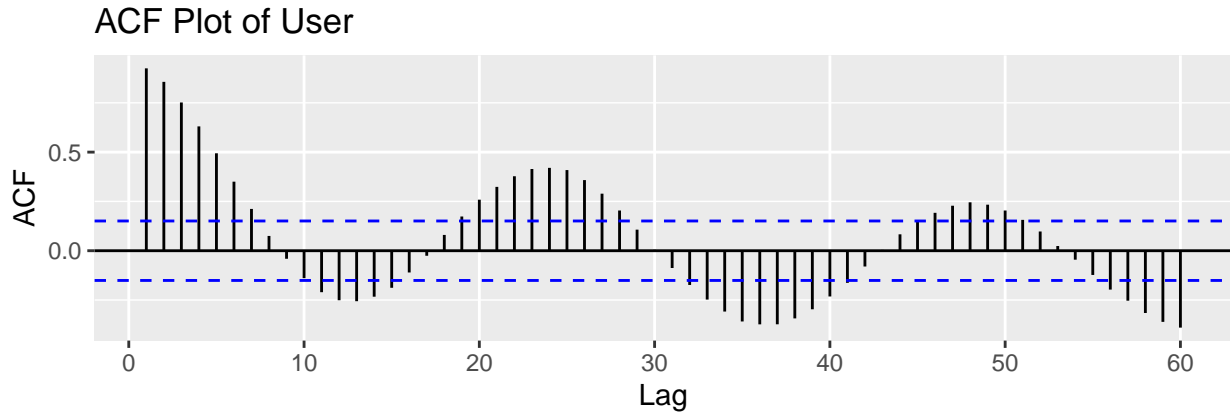
```
autoplot(users_data, facets = Series ~ .) + ggtitle("User and New User Count by Hour") +
  ylab("# of Users and New User") + xlab("") + facet_free() + theme(legend.position = 'bottom')
```

## User and New User Count by Hour



Her iki zaman serisini incelediğimizde her ikisinde de trend yok denecek kadar azken verilede mevsimsellik etkisinin olduğu görülmektedir. Ayrıca serilerde varyans sabit olması nedeniyle serilerin toplamsal olduğu anlaşılmaktadır. *newuser* serisinde ise mevsimselliğe ek olarak rassal etki ciddi derecede mevcuttur. Söz konusu bu bileşenlere daha yakından bakmak için serilere ait *ACF* ve *PACF* grafiklerine göz atalım.

```
user_acf <- ggAcf(users_data$users, lag.max = 60) + ggtitle("ACF Plot of User")
newuser_acf <- ggAcf(users_data$newusers, lag.max = 60) + ggtitle("ACF Plot of New User")
grid.arrange(user_acf, newuser_acf)
```



Her iki seride de otokorelasyon yavaşca azaldığı için bir durağan olmayan durum söz konusudur ancak, bu durum beşinci gecikmeden itibaren ortadan kalkması nedeniyle durağan olmayan durum çok güçlü değildir. Buna ek olarak belirli bir gecikmeden sonra tekrar otokorelasyonun ortaya çıkması seride mevsimsel etkinin olduğuna işaretler.

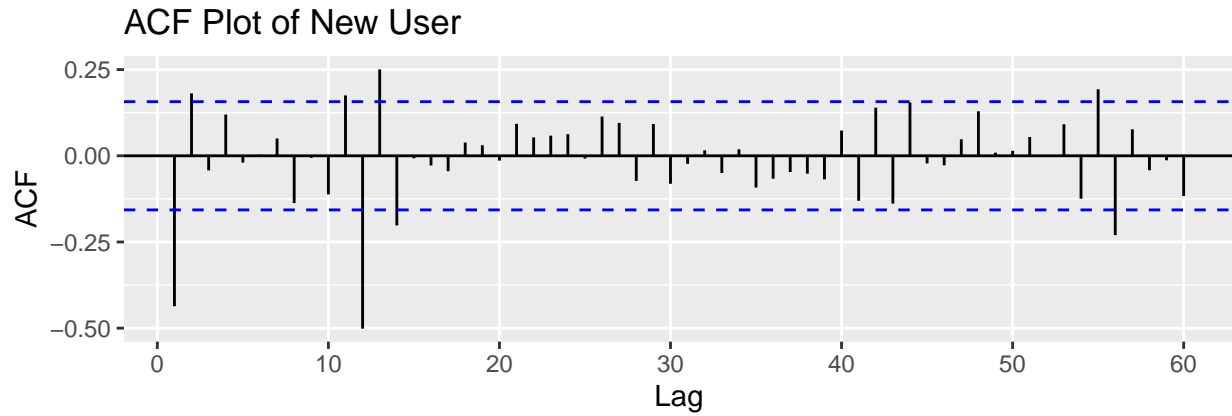
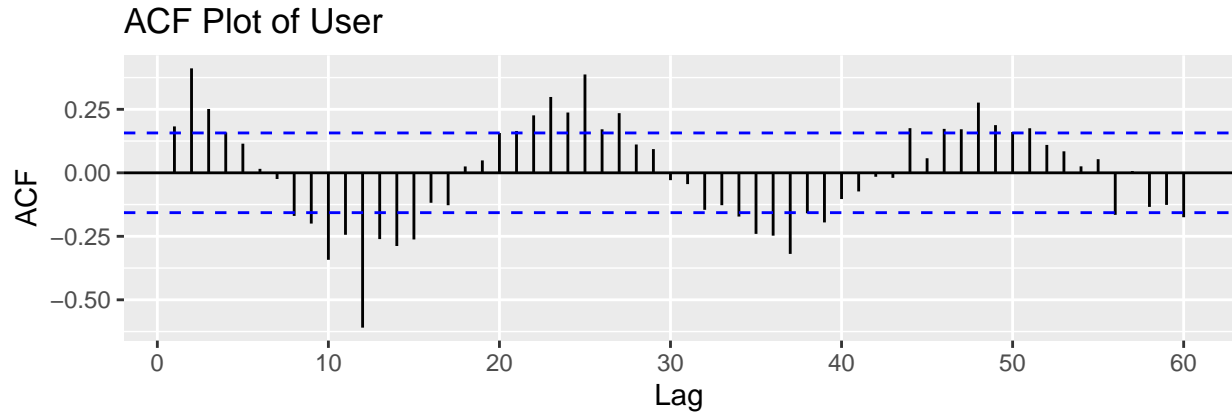
Serileri durağan hale getirmek için birinci farkı alalım ve durağan hale gelip gelmediğini inceleyelim.

```
first_diff <- diff(users_data, 1)
diff_user_acf <- ggAcf(first_diff$users, lag.max = 60) + ggtitle("ACF Plot of User")
diff_newuser_acf <- ggAcf(first_diff$newusers, lag.max = 60) + ggtitle("ACF Plot of New User")
grid.arrange(diff_user_acf, diff_newuser_acf)
```



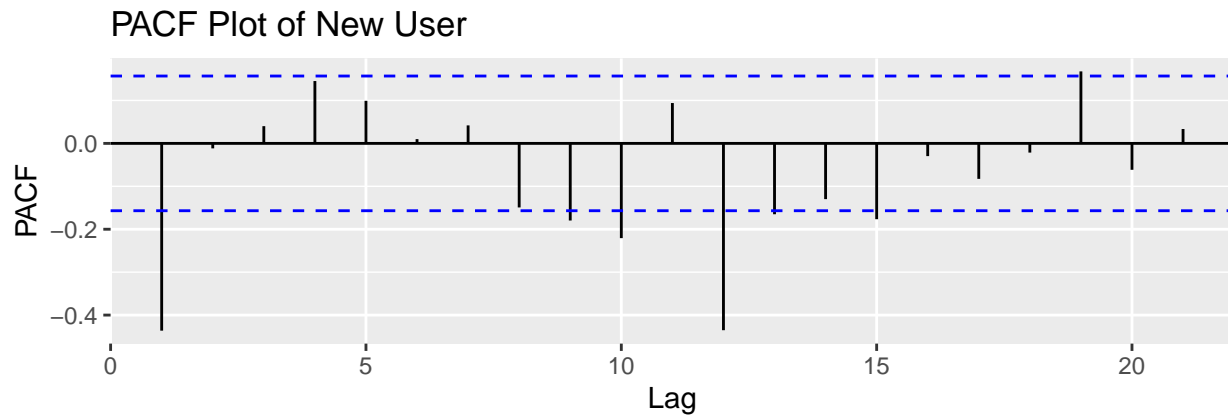
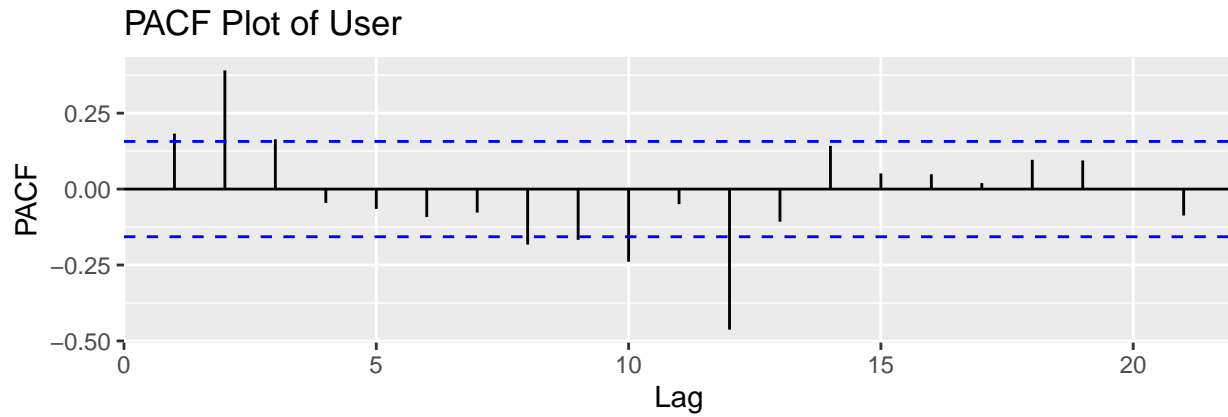
Her iki seri birinci fark alımından sonra durağan hale gelmiştir. İkinci seride her ne kadar birinci serideki nazarı mevsimsellik ortadan kalkmış gibi görünsede aslında ikinci seride de mevsimsellik vardır ve her iki seride de mevsimselliğin periyodu onkidir. Mevsimselliği ortadan kaldırmak için birinci farka ek olarak mevsimsel fark alalım.

```
second_diff <- diff(first_diff,12)
diff_user_acf2 <- ggAcf(second_diff$users, lag.max = 60) + ggtitle("ACF Plot of User")
diff_newuser_acf2 <- ggAcf(second_diff$newusers, lag.max = 60) + ggtitle("ACF Plot of New User")
grid.arrange(diff_user_acf2, diff_newuser_acf2)
```



Mevsimsel farkın ardından mevsimselliğinde ortadan kalktığı görülmektedir ve ilk seri için MA(2) veya MA(3) süreci olduğu görülmektedir. İkinci seride ise MA(1) sürecinin olduğu görülmektedir.

```
diff_user_pacf <- ggPacf(second_diff$users) + ggtitle("PACF Plot of User")
diff_newuser_pacf <- ggPacf(second_diff$newusers) + ggtitle("PACF Plot of New User")
grid.arrange(diff_user_pacf, diff_newuser_pacf)
```



*PACF* grafiğini incelediğimizde ise verilerimizde ilk veri için AR(2), ilk veri için AR(1) süreci olduğu görülmektedir. Elde ettiğimiz bu bilgiler doğrultusunda ilk fark ve mevsimsel farkı dikkate alarak aday modellerimizi kuralım.

#### 4: Modellerin Kurulması

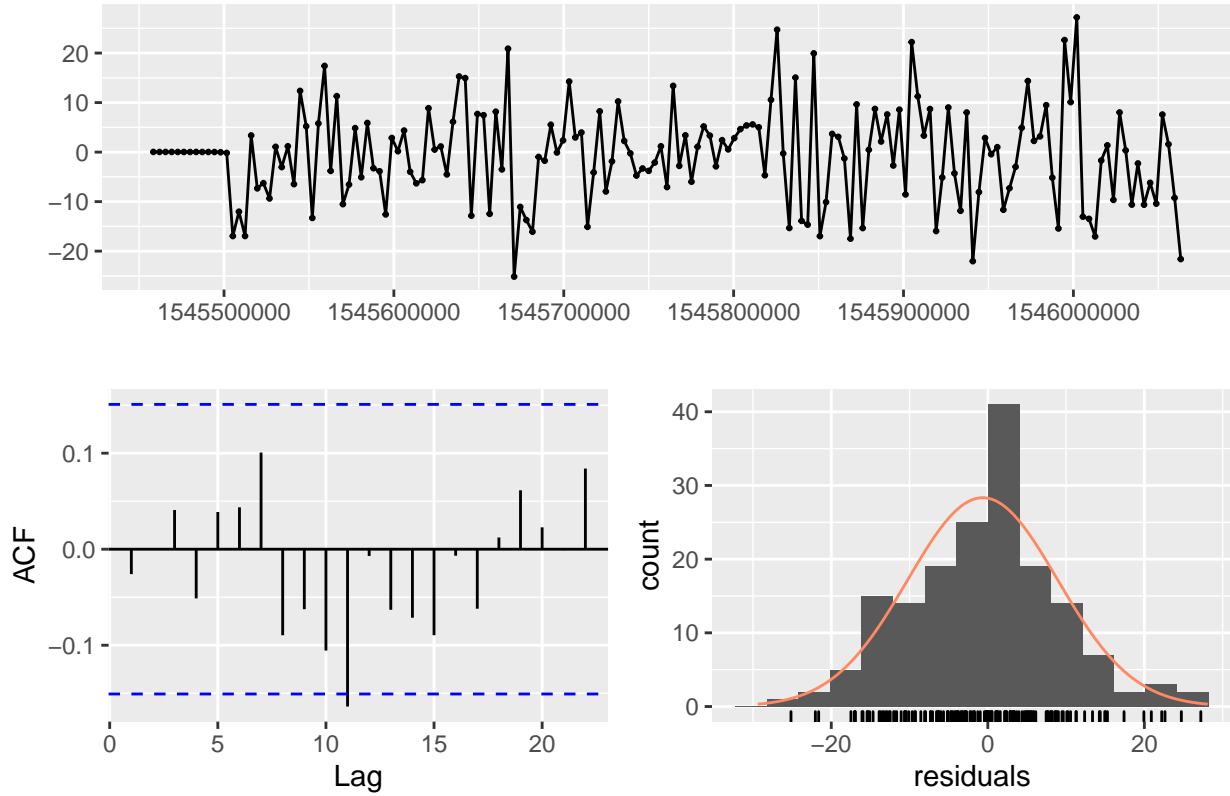
```
users_model <- arima(users_data$users, order = c(2,1,1),
                     seasonal = list(order = c(2,1,1), period =12))

checkresiduals(users_model, test = FALSE)
```

Table 1: SARIMA models for New Users

<i>Dependent variable:</i>	
ARIMA(2,1,1)(2,1,1)	
AR(1)	0.334** (0.163)
AR(2)	0.333*** (0.079)
MA(1)	-0.449*** (0.165)
SAR(1)	-0.269** (0.128)
SAR(2)	0.034 (0.128)
SMA(1)	-0.808*** (0.116)
Observations	156
Log Likelihood	-589.100
$\sigma^2$	99.358
Akaike Inf. Crit.	1,192.200
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Residuals from ARIMA(2,1,1)(2,1,1)[12]



İlk seri için önerdiğimiz ARIMA(2,1,1)(2,1,1) modelinin parametreleri anlamlı olmasının yanı sıra Ljung-Box testine göre hatalar arasında herhangi bir otokorelasyon yoktur.

```
newusers_model1 <- arima(users_data$newusers, order = c(1,1,2),
                          seasonal = list(order = c(1,0,1), period =12))
```

İkinci seri için önerdiğimiz ARIMA(1,1,2)(1,0,1) modeli ise yine bu seri için en uygun modeldir ve Ljung-Box testine göre kalıntılar arasında bir korelasyon yoktur. Bu modelleri kullanarak 12 saat sonrasını tahmin edecek olursak:

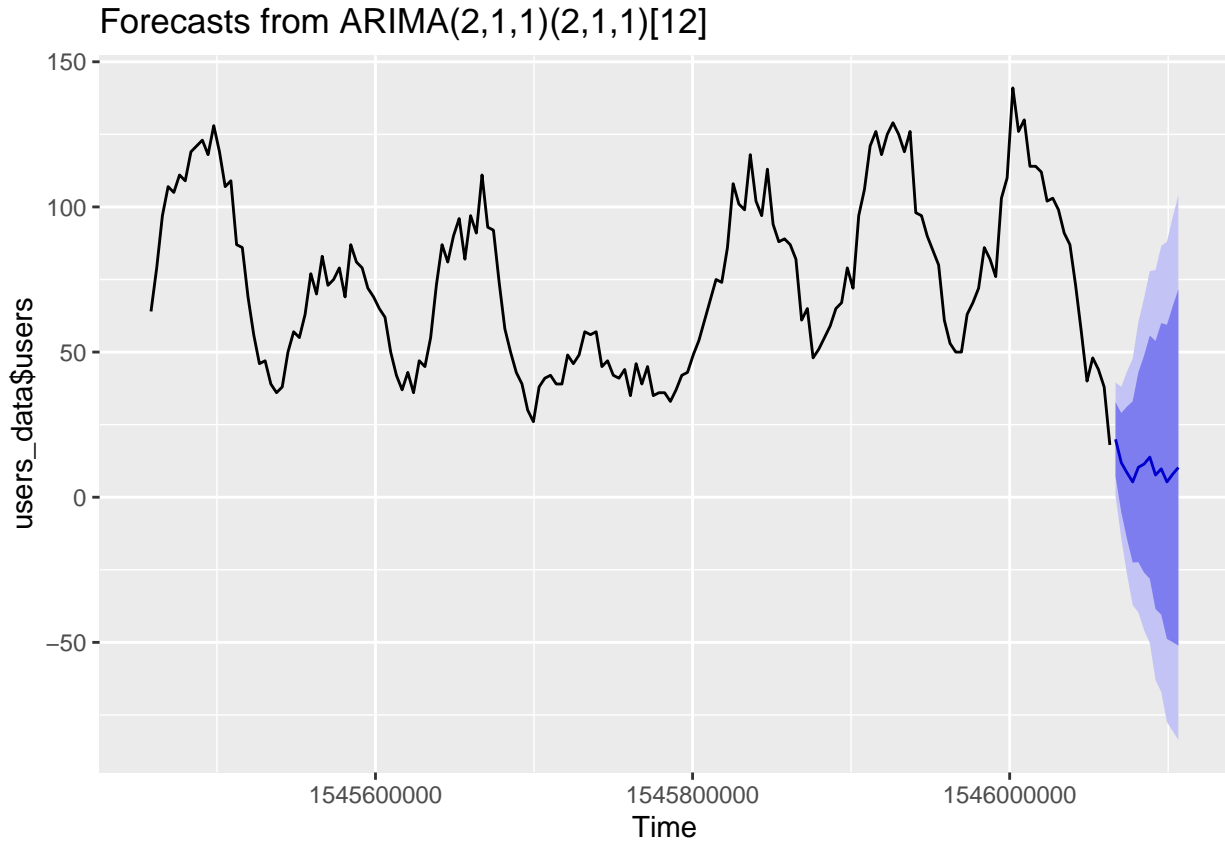
## S5: Öngöründe Bulunulması.

```
user_forecast <- forecast(users_model, h = 12)
autoplot(user_forecast)
```



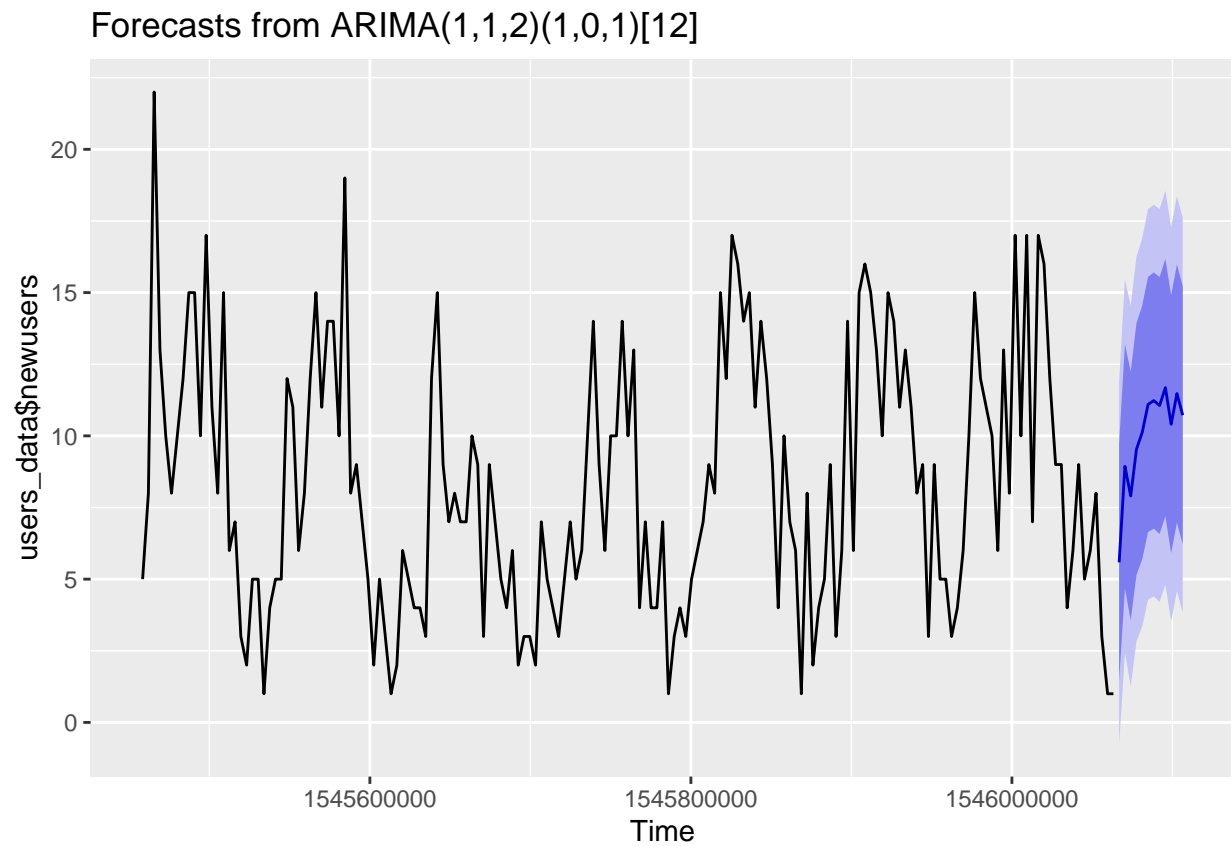
Table 2: SARIMA models for New Users

<i>Dependent variable:</i>	
ARIMA(1,1,2)(1,0,1)	
AR(1)	0.818*** (0.102)
MA(1)	-1.585*** (0.150)
MA(2)	0.585*** (0.148)
SAR(1)	-0.999*** (0.003)
SMA(1)	0.977*** (0.049)
Observations	168
Log Likelihood	-443.444
$\sigma^2$	10.007
Akaike Inf. Crit.	898.888
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	



	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
1546066800	19.95	7.17	32.73	0.40	39.49
1546070400	11.92	-5.15	28.99	-14.18	38.02
1546074000	8.45	-14.33	31.22	-26.38	43.28
1546077600	5.27	-22.47	33.02	-37.16	47.71
1546081200	10.32	-22.40	43.05	-39.73	60.38
1546084800	11.42	-25.98	48.83	-45.79	68.64
1546088400	13.81	-28.09	55.71	-50.27	77.89
1546092000	7.64	-38.53	53.80	-62.97	78.24
1546095600	9.76	-40.47	60.00	-67.06	86.59
1546099200	5.28	-48.83	59.39	-77.47	88.04
1546102800	7.92	-49.90	65.73	-80.50	96.33
1546106400	10.24	-51.11	71.59	-83.59	104.06

```
newuser_forecast <- forecast(newusers_model1, h = 12)
autoplot(newuser_forecast)
```



	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
1546066800	5.59	1.44	9.73	-0.75	11.92
1546070400	8.94	4.68	13.20	2.42	15.45
1546074000	7.90	3.56	12.24	1.27	14.54
1546077600	9.54	5.15	13.93	2.82	16.25
1546081200	10.13	5.70	14.55	3.36	16.90
1546084800	11.10	6.64	15.55	4.29	17.91
1546088400	11.23	6.76	15.70	4.40	18.07
1546092000	11.06	6.58	15.54	4.21	17.91
1546095600	11.69	7.19	16.18	4.82	18.55
1546099200	10.40	5.91	14.90	3.53	17.28
1546102800	11.48	6.97	15.98	4.59	18.36
1546106400	10.72	6.22	15.22	3.84	17.61