

# Referandum Denemeler

Deniz Kenan KILIÇ

09.03.2017

## Contents

.....	1
Referandum Denemeler .....	1
No table of figures entries found.....	2
1 GİRİŞ .....	3
2 VERİLERİN DÜZENLENMESİ .....	3
3 SONUÇ.....	10

### **No table of figures entries found.**

Figure 1: Verilerin Okunması ve Düzenlenmesi .....	3
Figure 2: "Evet" ve "Hayır" Verileri .....	5
Figure 3: "Evet" ve "Hayır" Verileri Bar Çizimi .....	6
Figure 4: İllere göre 26. dönem seçim ve tweeter verileri arası farklar .....	7
Figure 5: İllere göre 26./25. dönem seçim ve tweeter verileri arası farklar .....	8
Figure 6: İllere göre 26. ve 25. dönem ile tweeter verileri arası farklar .....	9
Figure 7: İllere göre 26./25. dönem "evet" ve "hayır" puanlamaları yüzdeleri .....	9
Figure 8: İllere göre Nüfus Dağılımı.....	10

## 1 GİRİŞ

Bu raporda referanduma ait denemeler için 25. Ve 26. dönem seçim verileri ile “21 Ocak 2017” ve “8 Mart 2017” tarihleri arasında çekilen saatlik, lokasyon bazlı tweeter verileri kullanılmıştır.

## 2 VERİLERİN DÜZENLENMESİ

Veriler .csv formatında okunup, seçim sonuçları için “NA” değerleri “0” değeri ile değiştirilmiştir. Örnek kodlar Figure 1’de yer almaktadır. Seçim verilerinden “evet” ve “hayır” puanı belirlemek için, oy sayıları şu şekilde ele alınmıştır;

$$Evet = Oy_{Akp} + \frac{Oy_{Mhp}}{2},$$
$$Hayır = Oy_{Chp} + \frac{Oy_{Mhp}}{2} + Oy_{Hdp} + Oy_{Saadet}.$$

```
1 Secim_25 <- read.csv(file="2016_Secim_25.Donem.csv", header=TRUE, sep=";")
2 Secim_26 <- read.csv(file="2016_Secim_26.Donem.csv", header=TRUE, sep=";")
3 Secim_25_eksik <- read.csv(file="2016_Secim_25.Donem_eksik.csv", header=TRUE, sep=";")
4 Secim_26_eksik <- read.csv(file="2016_Secim_26.Donem_eksik.csv", header=TRUE, sep=";")
5 Tweets_loc_perc <- read.csv(file="il_tweet_perc.csv", header=TRUE, sep=";")
6 # Tw_vs_sec26 <- read.csv(file="Tw-7gun vs Secim 26.csv", header=TRUE, sep=";")
7 # Tw_hours <- read.csv(file="referandum.csv", header=TRUE, sep=";")
8 # Tweets_locations_perc <- read.csv(file="tweets_hours.csv", header=TRUE, sep=";")
9
10 # Analysis for Secim_25
11 Secim_25[is.na(Secim_25)] <- 0
12 summary(Secim_25)
13 names(Secim_25)
14 head(Secim_25$akp)
15
16 Secim_25["evet_sayi"] <- NA
17 Secim_25["hayir_sayi"] <- NA
18 Secim_25["evet_yuzde"] <- NA
19 Secim_25["hayir_yuzde"] <- NA
20 Secim_25$evet_sayi <- Secim_25$akp+(Secim_25$mhp/2)
21 Secim_25$hayir_sayi <- Secim_25$chp+(Secim_25$mhp/2)+Secim_25$hdp+Secim_25$saadet
22 Secim_25$evet_yuzde <- 100*(Secim_25$evet_sayi/(Secim_25$evet_sayi+Secim_25$hayir_sayi))
23 Secim_25$hayir_yuzde <- 100*(Secim_25$hayir_sayi/(Secim_25$evet_sayi+Secim_25$hayir_sayi))
24
25 # Analysis for Secim_26
```

Figure 1: Verilerin Okunması ve Düzenlenmesi

Seçim sonuçları için belirlenen bu puanlama biçimleri ufak oynamalar ile düzeltilebilir. Örneğin Mhp’ye ait ağırlıklar değiştirilebilir. Bbp’ye ait oylar eklenmemiştir çünkü puanlamaya etkisi çok azdır. Aşağıda 25. Dönem seçimlere ait verilerin özet hali verilmiştir. Aynı şekilde 26. Dönem seçimi için de bu özetlere

bakılmıştır. Özet değerlere göre diğer partilerin puanlamaya etkisinin düşük olacağı görüldüğü için hesaplamada kullanılmamışlardır.

```
> summary(Secim_25)

      il      akp      mhp      chp      saa
det
  Ā±ĀŸdĀ±r : 1   Min.   : 3907   Min.   : 1485   Min.   : 851   Min.
: 207   Min.   : 448   Min.   :0
  Ā±sparta : 1   1st Qu.: 49019  1st Qu.: 20267  1st Qu.: 16464  1st Qu.
: 2125  1st Qu.: 2850  1st Qu.:0
  Āřanakkale: 1   Median : 105787  Median : 45360  Median : 37125  Median
: 4430  Median : 14292  Median :0
  ĀřankĀ±rĀ±: 1   Mean   : 210904  Mean   : 84948  Mean   : 131320  Mean
: 10728  Mean   : 65688  Mean   :0
  Āřorum     : 1   3rd Qu.: 213321  3rd Qu.:103254  3rd Qu.: 115609  3rd Qu.
: 9486  3rd Qu.: 64287  3rd Qu.:0
  aĀŸrĀ±     : 1   Max.   :3282111  Max.   :883716  Max.   :2357017  Max.
:172619  Max.   :1001314  Max.   :0
  (Other)    :75

      btp      dp      dsp      hkp      kp
millet
  Min.   : 42   Min.   : 16.0   Min.   : 38.0   Min.   : 60.0   Min.   :
4.0   Min.   : 7.0   Min.   : 20
  1st Qu.: 266  1st Qu.: 173.0  1st Qu.: 263.0  1st Qu.: 221.0  1st Qu.:
30.0  1st Qu.: 47.0  1st Qu.: 228
  Median : 578  Median : 396.0  Median : 533.0  Median : 400.0  Median :
52.0  Median : 106.0  Median : 446
  Mean   : 1090  Mean   : 853.3  Mean   : 965.3  Mean   : 666.4  Mean   :
156.9  Mean   : 197.1  Mean   : 1806
  3rd Qu.: 1133  3rd Qu.:1328.0  3rd Qu.: 1221.0  3rd Qu.: 782.0  3rd Qu.:
129.0  3rd Qu.: 244.0  3rd Qu.: 1266
  Max.   :14355  Max.   :4339.0  Max.   :11606.0  Max.   :6640.0  Max.   :
3359.0  Max.   :1652.0  Max.   :33648

      dyp      hak_par      ldp      alan      latitu
de
  Min.   : 0.0   Min.   : 0.0   Min.   : 0.0   Min.   : 74   Min.   :3
6.40   Min.   :26.41  Min.   : 11.0
  1st Qu.: 0.0   1st Qu.: 152.0  1st Qu.: 0.0   1st Qu.: 4575  1st Qu.:3
7.88   1st Qu.:30.56  1st Qu.: 41.0
  Median : 136.0  Median : 415.0  Median : 126.0  Median : 7228  Median :3
9.63   Median :34.96  Median : 59.0
  Mean   : 311.8  Mean   : 650.7  Mean   : 300.6  Mean   : 8771  Mean   :3
9.31   Mean   :35.13  Mean   : 122.1
  3rd Qu.: 416.0  3rd Qu.: 797.0  3rd Qu.: 320.0  3rd Qu.:12007  3rd Qu.:4
0.65   3rd Qu.:39.50  3rd Qu.: 109.0
  Max.   :3293.0  Max.   :6610.0  Max.   :4230.0  Max.   :41001  Max.   :4
2.02   Max.   :44.00  Max.   :2758.0

      nĀ.fus      plaka      telefon      vali
  Min.   : 7855   Min.   : 1   212:216: 1   Āřmer doĀŸanay : 1
  1st Qu.: 236978  1st Qu.:21  222 : 1   Āřmer faruk koĀřak : 1
  Median : 421766  Median :41  224 : 1   Āřzdemir Āřakacak : 1
  Mean   : 932519  Mean :41  226 : 1   abdulkadir yazĀ±cĀ±: 1
  3rd Qu.: 993442  3rd Qu.:61  228 : 1   abdullah erin : 1
  Max.   :14657434  Max.   :81  232 : 1   ahmet ĀřĀ±nar : 1
  (Other):75   (Other):75
```

Seimler ile ilgili veriler dzenlendikten sonra, Tweeter verilerinde bulunmayan iki il seim verilerinden de ıkarılmıřtır.

Geriye kalan 79 ile gre “25. Dnem Evet”, “26. Dnem Evet”, “Tweeter Evet”, “25. Dnem Hayır”, “26. Dnem Hayır” ve “Tweeter Hayır” verileri tek bir veri seti haline Figure 2’deki gibi getirilmiřtir.

	iller	25.donem_evet	26.donem_evet	tweet_evet	25.donem_hayir	26.donem_hayir	tweet_hayir
1	adana	41.63469	46.32169	58.41219	58.36531	53.67831	41.587811
2	adana	61.29036	72.06611	86.91983	38.70964	27.93389	13.080169
3	afyonkarahisar	66.50246	73.33708	64.87310	33.49754	26.66292	35.126904
4	aksaray	74.82106	82.16889	86.90969	25.17894	17.83111	13.090307
5	amasya	57.51580	62.45109	76.64399	42.48420	37.54891	23.356009
6	ankara	52.27114	56.71145	57.90238	47.72886	43.28855	42.097616
7	antalya	46.68959	50.78798	62.96811	53.31041	49.21202	37.031887
8	ardahan	31.80046	39.95131	91.04859	68.19954	60.04869	8.951407
9	artvin	48.59308	50.03497	73.85214	51.40692	49.96503	26.147860
10	aydin	38.97236	42.54380	53.67003	61.02764	57.45620	46.329966
11	aydin	17.45869	28.27191	90.68768	82.54131	71.72809	9.312321
12	balikesir	50.07424	54.09913	77.43436	49.92576	45.90087	22.565636
13	bartın	50.72329	58.94683	56.22318	49.27671	41.05317	43.776824
14	batman	18.88272	27.81648	74.53052	81.11728	72.18352	25.469484
15	bayburt	76.95556	84.83202	87.50000	23.04444	15.16798	12.500000
16	bilecik	51.04084	54.51620	75.93361	48.95916	45.48380	24.066390
17	bingl	52.37930	67.37593	82.51834	47.62070	32.62407	17.481663
18	bitlis	36.40757	47.55615	82.48588	63.59243	52.44385	17.514124
19	bolu	61.02292	69.68099	51.36364	38.97708	30.31901	48.636364
20	burdur	56.93065	63.50514	58.41584	43.06935	36.49486	41.584158

owing 1 to 20 of 79 entries

onsole C:/Users/deniz.kilic/Desktop/Big Data/R Denemeler/Referandum/

```
secim_25_eksik["evet_sayi"] <- NA
secim_25_eksik["hayir_sayi"] <- NA
secim_25_eksik["evet_yuzde"] <- NA
secim_25_eksik["hayir_yuzde"] <- NA
```

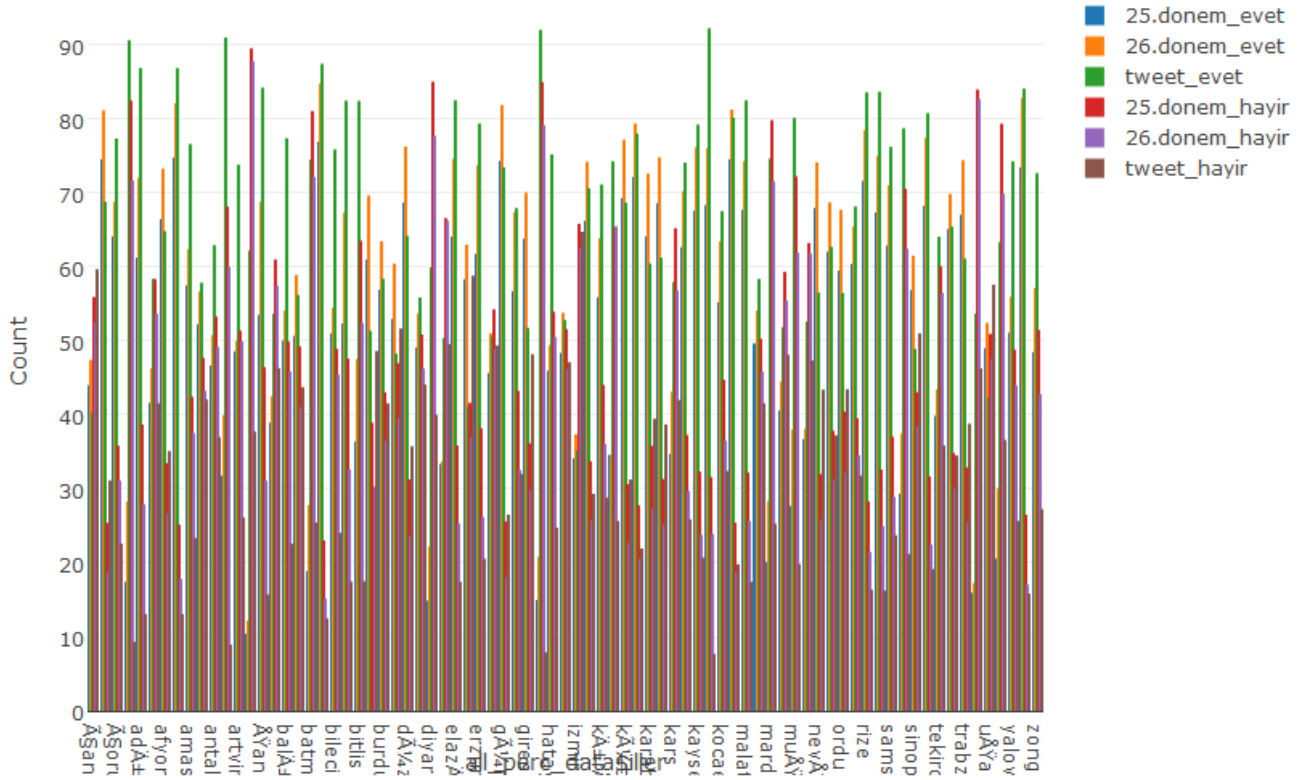
Figure 2: "Evet" ve "Hayır" Verileri

Bu veri setine ait bar izimi Figure 3’de gzlmektedir. Her il iin “25. Dnem Evet”, “26. Dnem Evet”, “Tweeter Evet”, “25. Dnem Hayır”, “26. Dnem Hayır” ve “Tweeter Hayır” verileri izdirilmiřtir. Grafie bakıldıėında 26. Dnem seimlerinden elde edilen puanlamanın tweeterden ekilen verilerin aėrılıėına daha yakın olduėu gzlmektedir. Ayrıca beklendiėi gibi il nfusu arttıka tweeter aėrılıėı ve seimlere gre puanlama deėeri yakınlařmaktadır. Tweeter deėerlerinin 26. dnem seimleri ve 25. dnem seimleri ile koralasyonlarına bakıldıėında, 26. dnem deėerleri ile koralasyonunun daha fazla olduėu gzlmektedir.

```
# Plot of all "evet" and "hayir" data for 79 locations
p1 <- plot_ly(all_perc_data, x = ~all_perc_data$iller, y = ~all_perc_data$`25.donem_evett`, type = 'bar', name = '25.donem_evett')
add_trace(y = ~all_perc_data$`26.donem_evett`, name = '26.donem_evett') %>%
add_trace(y = ~all_perc_data$tweet_evett, name = 'tweet_evett') %>%
add_trace(y = ~all_perc_data$`25.donem_hayir`, name = '25.donem_hayir') %>%
add_trace(y = ~all_perc_data$`26.donem_hayir`, name = '26.donem_hayir') %>%
add_trace(y = ~all_perc_data$tweet_hayir, name = 'tweet_hayir') %>%
layout(yaxis = list(title = 'count'), barmode = 'group')
p1

cor(all_perc_data$tweet_evett, all_perc_data$`25.donem_evett`)
cor(all_perc_data$tweet_evett, all_perc_data$`26.donem_evett`)
cor(all_perc_data$tweet_hayir, all_perc_data$`25.donem_hayir`)
cor(all_perc_data$tweet_hayir, all_perc_data$`26.donem_hayir`)
```

```
> cor(all_perc_data$tweet_evett, all_perc_data$`25.donem_evett`)
[1] 0.1607889
> cor(all_perc_data$tweet_evett, all_perc_data$`26.donem_evett`)
[1] 0.2335128
> cor(all_perc_data$tweet_hayir, all_perc_data$`25.donem_hayir`)
[1] 0.1607889
> cor(all_perc_data$tweet_hayir, all_perc_data$`26.donem_hayir`)
[1] 0.2335128
```



26. dönem ve 25. dönem değerleri ile tweeter değerleri arası farka baktığımızda 26. Döneme ait farkların toplamının daha az olduğunu görmekteyiz.

```
> sum(difference_evett)
[1] 1243.445
> sum(difference_evett_2)
[1] 1491.03
```

İllere göre 26. dönem ve tweeter verileri farklarının bar çizimi Figure 4'te görülmektedir. Grafiğin sol tarafında yüksek farka sahip iller genellikle Hakkari, Ardahan, Şırnak, Batman, Mardin, Siirt, Diyarbakır, Tunceli, Bitlis, Van gibi doğu illeridir.

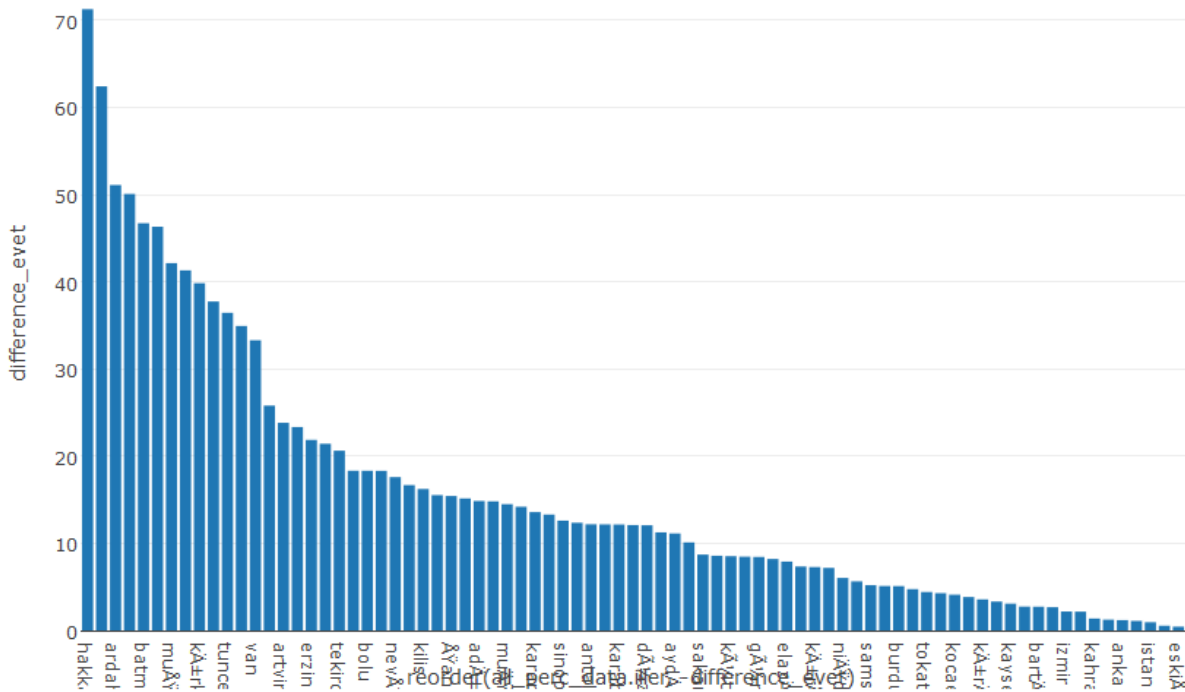


Figure 4: İllere göre 26. dönem seçim ve tweeter verileri arası farklar

Figure 5'e bakıldığında ise illere göre hem 26. dönem için hem de 25. dönem için farklar görülmektedir. Büyük bir çoğunlukta 25. dönem farkları 26. dönem farklarından yüksektir. Figure 6'da belirtilen "difference\_evett" kolonu illere göre 26. dönem puanları ile tweeterden gelen yüzde puanlar arası farkı, "difference\_evett\_2" ise illere göre 25. dönem puanları ile tweeterden gelen yüzde puanlar arası farkı göstermektedir.

Figure 7'de 25. ve 26. Döneme ait puanladığımız "evett" ve "hayır" değerlerine ait yüzdeler her il için çizilmiştir. Bu grafiğe baktığımızda, İstanbul ve Ankara gibi büyük illerin değerlerinin gerçek seçim sonuçlarına çok yakın çıktığı görülmektedir. Büyük şehirler seçilerek örneklem oluşturmak daha sağlıklı görülmektedir. İller küçüklükçe "evett" ve "hayır" arası puanlama yükselmektedir.

```

111 library(plotly)
112 evet_compare_secim <- data.frame(Secim_25$il, Secim_25$evet_yuzde, Secim_26$evet_yuzde, Secim_25$hayir_yuzde, Secim_26$hayir_yuzde)
113 p2 <- plot_ly(evet_compare_secim, x = ~Secim_25$il, y = ~Secim_25$evet_yuzde, type = 'bar', name = 'Secim_25$evet_yuzde')
114 add_trace(y = ~Secim_26$evet_yuzde, name = 'Secim_26$evet_yuzde') %>%
115 add_trace(y = ~Secim_25$hayir_yuzde, name = 'Secim_25$hayir_yuzde') %>%
116 add_trace(y = ~Secim_26$hayir_yuzde, name = 'Secim_26$hayir_yuzde') %>%
117 layout(yaxis = list(title = 'Count'), barmode = 'group')
118 p2
119
120
121 p3 <- plot_ly(Secim_25, labels = ~Secim_25$il, values = ~Secim_25$A.fus, type = 'pie') %>%
122 layout(title = 'Secim Percentages',
123 xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),
124 yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))
125 p3

```

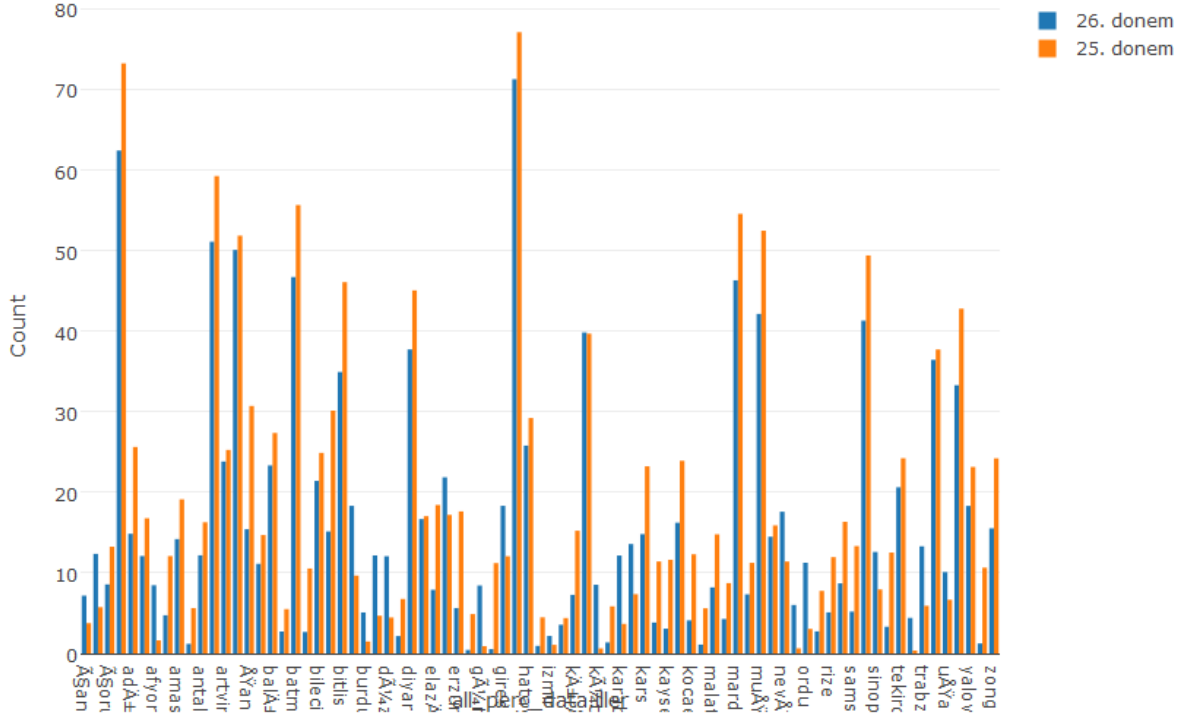


Figure 5: İllere göre 26./25. dönem seçim ve tweeter verileri arası farklar



	all_perc_data.iller	difference_evet	difference_evet_2
1	adana	12.0905006	16.7775000
2	adana	14.8537222	25.6294757
3	afyonkarahisar	8.4639815	1.6293658
4	aksaray	4.7408036	12.0886379
5	amasya	14.1928985	19.1281887
6	ankara	1.1909327	5.6312427
7	antalya	12.1801372	16.2785260
8	ardahan	51.0972853	59.2481305
9	artvin	23.8171699	25.2590605
10	aydin	11.1262371	14.6976695
11	aydin	62.4157696	73.2289905
12	balikesir	23.3352381	27.3601231
13	bartin	2.7236548	5.4998863
14	batman	46.7140337	55.6477962
15	bayburt	2.6679779	10.5444428
16	bilecik	21.4174123	24.8927732
17	bingol	15.1424022	30.1390336
18	bitlis	34.9297246	46.0783054
19	bolu	18.3173531	9.6592811
20	burdur	5.0893017	1.4851915

Figure 6: İllere göre 26. ve 25. dönem ile tweeter verileri arası farklar

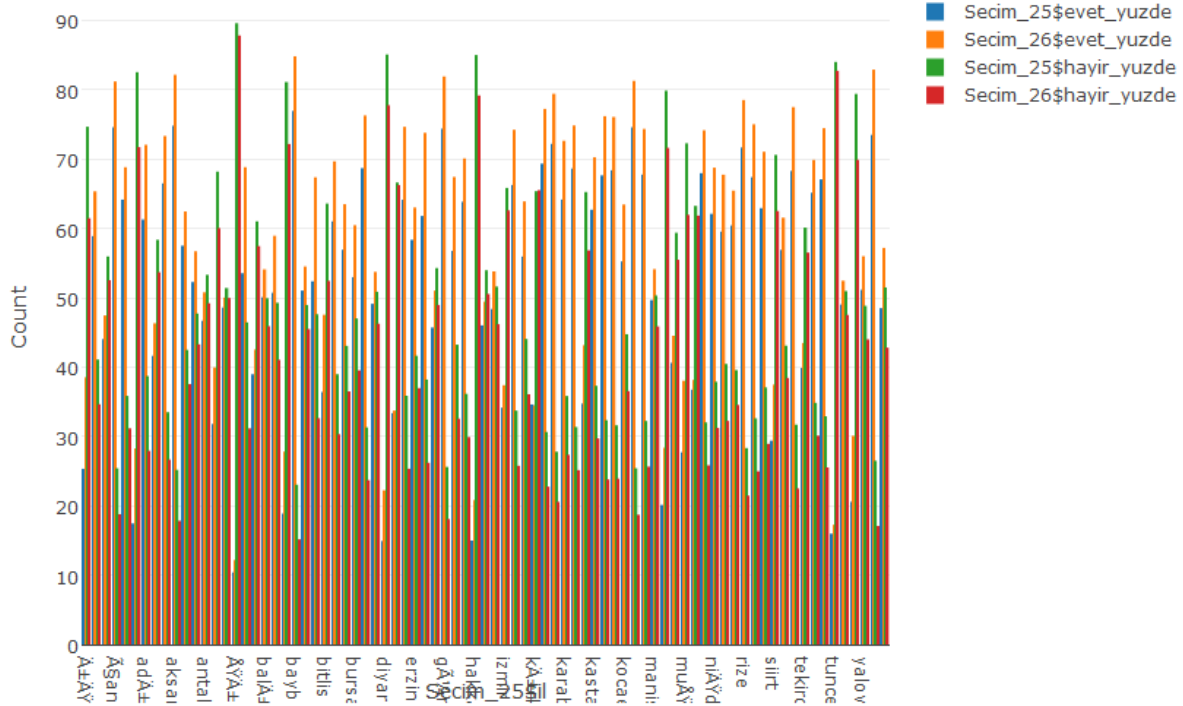


Figure 7: İllere göre 26./25. dönem "evet" ve "hayır" puanlamaları yüzdeleri

### 3 SONUÇ

İller büyükçe örneklem doğruluğu artmaktadır, bu sebeple il büyüklüklerine göre bir tek yüzde tahmini hesaplaması yapılabilir. Figure 8’de illere göre nüfus oranları görülmektedir. Bu grafikte en çok fark değeri veren doğu illerinin büyük bir kısmı 2%’lik dilimin altında kalmaktadırlar.

Tüm illere göre seçim puanlaması yapıp, tweeter verilerinden elde edilen puanlama ile karşılaştırıldıktan sonra tahmini bir puan belirlemek için yüksek fark yaratan iller analizlerden çıkarılabilir. Çok yüksek nüfusa sahip illerden gelen tweet verileri (hacim olarak), seçimlerden elde edilen puanlama değerleri ile yakın çıkmaktadır. Ancak çok sayıda seçim sonucu olmadığı için, zaman serisi modeli uygulamak zordur. Daha basit bir tahmin yöntemi kullanılabilir. Bunun ile ilgili bazı modeller ilerleyen süreçte denenecektir.

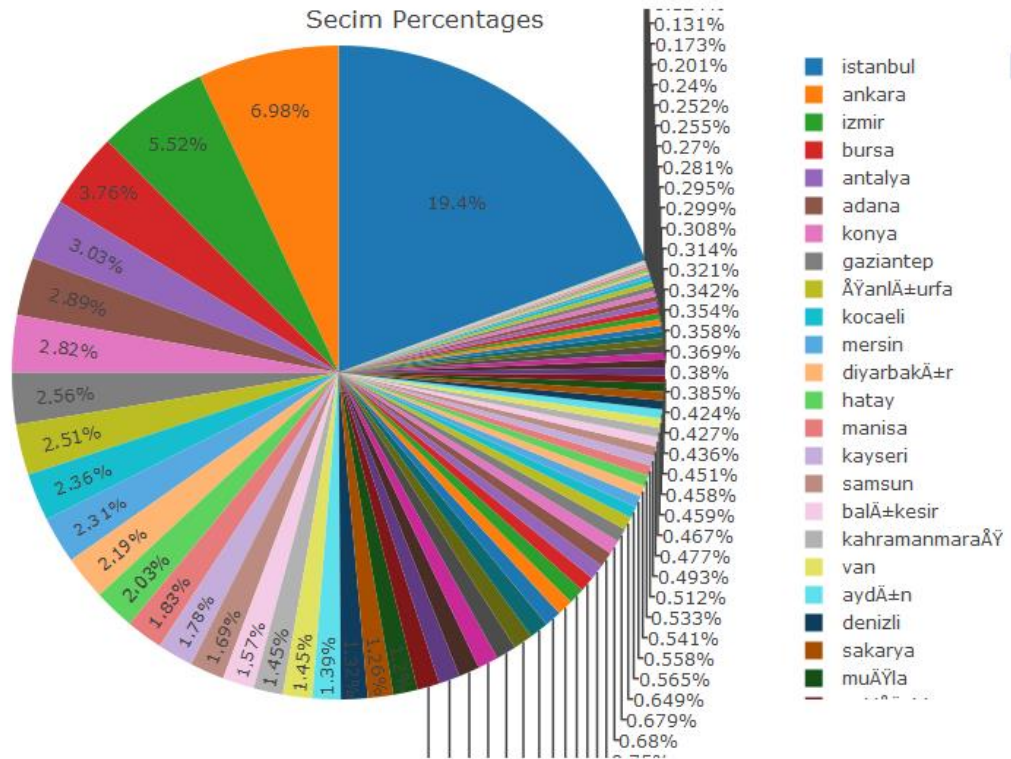


Figure 8: İllere göre Nüfus Dağılımı