

**1950-2017 YILLARI ARASINDAKİ FORMULA 1 VERİLERİNİN
DÜZENLENMESİ VE GÖRSELLEŞTİRİLMESİ**

OĞUZHAN GÜNDÜZ

gnoguz@outlook.com

1. Verinin Tanıtılması

Proje ödevi için kullanılan veri; FIA (Fédération Internationale de l'Automobile) yani Uluslararası Otomobil Federasyonu tarafından 1950 yılından bu güne her sene düzenlenen motor sporları kategorisine ait "Formula 1" yarışlarının, 1950 ile 2017 yılları arasındaki bilgilerdir. Kullanılan veri "<https://data.world/>" internet adresinden, "F1AllData.csv" isimli dosya olarak indirilmiştir. Veri, 23 sütun, 23.642 satırdan oluşmaktadır. Aşağıda her bir sütunda hangi bilgilerin olduğu maddeler halinde sıralanmıştır.

- 1. season: 1950-2017 arasında Formula 1 sezonlarıdır. Yıl olarak verilmiştir.
- 2. round: 0 sezonda kaçınıcı yarışın yapıldığının bilgisi.
- 3. raceName: Yapılan yarışın ismi.
- 4. date: Yapılan yarışın tarihi.
- 5. circuitId: Yarışılan pistin tek kelimelik ismi.
- 6. circuitName: Yarışılan pistin tam ismi.
- 7. number: Pilotlara ait numaralar.
- 8. position: Pilotların yarış hangi sıradan başladığı bilgisi.
- 9. positionText: Yarışı tamamlayamayan pilotların harf ile belirtildiği sıralama bilgisi.
- 10. points: Pilotların yarış sonunda aldıkları puan bilgisi.
- 11. driverId: Yarışan pilotların tek kelimelik ismi.
- 12. givenName: Pilotların isimleri.
- 13. familyName: Pilotların soy isimleri.
- 14. full name: Pilotların isim ve soy isimleri.
- 15. dateOfBirth: Pilotların doğum tarihleri.
- 16. nationality: Pilotların ait oldukları milletler.
- 17. constructorId: Pilotların kullandıkları araçların hangi takıma ait oldukları. Motor sağlayıcıları olarak da isimlendirilebilir.
- 18. Constructor name: Takımların ismi.
- 19. Constructor nationality: Takımların hangi millete ait olduklarının bilgisi.
- 20. results.grid: Pilotların yarış hangi sırada tamamladıklarının bilgisi.
- 21. Results.laps: Pilotların yarışın bitimine kadar ne kadar tur attıkları bilgisi.
- 22. results.status: pilotların yarış bitirip bitiremediklerinin, lider pilottan fark yedikleri tur sayıları, yarış bitiremeyen pilotların hangi sebep yüzünden bitiremediği bilgileri.
- 23. Circ Count: Veri setinde bulunan her yarışın en başında verilen pist isimleri.

Kod:

```
##
```

```
F1data <- read.csv("F1AllData.csv")
```

```
summary(F1data)
```

```
##
```

Çıktı:

```
season      round      raceName      date      circuitId
Min. :1950   Min. : 1.000   Length:23641   Length:23641   Length:23641
1st Qu.:1975   1st Qu.: 4.000   Class :character   Class :character   Class :character
Median :1989   Median : 8.000   Mode :character   Mode :character   Mode :character
Mean :1988   Mean : 8.137
3rd Qu.:2003   3rd Qu.:12.000
Max. :2017   Max. :21.000

circuitName      number      position      positionText      points
Length:23641     Min. : 0.00   Min. : 1.00   Length:23641     Min. : 0.000
Class :character  1st Qu.: 7.00   1st Qu.: 7.00   Class :character  1st Qu.: 0.000
Mode :character   Median :15.00   Median :13.00   Mode :character   Median : 0.000
Mean :16.96       Mean :13.08
3rd Qu.:23.00     3rd Qu.:19.00
Max. :208.00      Max. :39.00
NA's :6

driverId      givenName      familyName      fullName      dateOfBirth
Length:23641   Length:23641   Length:23641   Length:23641   Length:23641
Class :character   Class :character   Class :character   Class :character   Class :character
Mode :character   Mode :character   Mode :character   Mode :character   Mode :character

nationality      constructorId      Constructor.name      Constructor.nationality      results.grid
Length:23641     Length:23641     Length:23641     Length:23641     Min. : 0.00
Class :character   Class :character   Class :character   Class :character   1st Qu.: 5.00
Mode :character   Mode :character   Mode :character   Mode :character   Median :11.00
Mean :11.26
3rd Qu.:17.00
Max. :34.00

results.laps      results.status      Circ.Count
Min. : 0.0   Length:23641   Length:23641
1st Qu.:20.0   Class :character   Class :character
Median :52.0   Mode :character   Mode :character
Mean :45.3
3rd Qu.:66.0
Max. :200.0
```

Yukarıda “summary()” fonksiyonu kullanılarak üzerinde çalışılan verinin hakkında bilgiler verilmiştir. 23 değişken bulunan veride, 16 değişken karakter sınıfına aittir.

2. Her Bir Sezondaki Pilot Sayıları

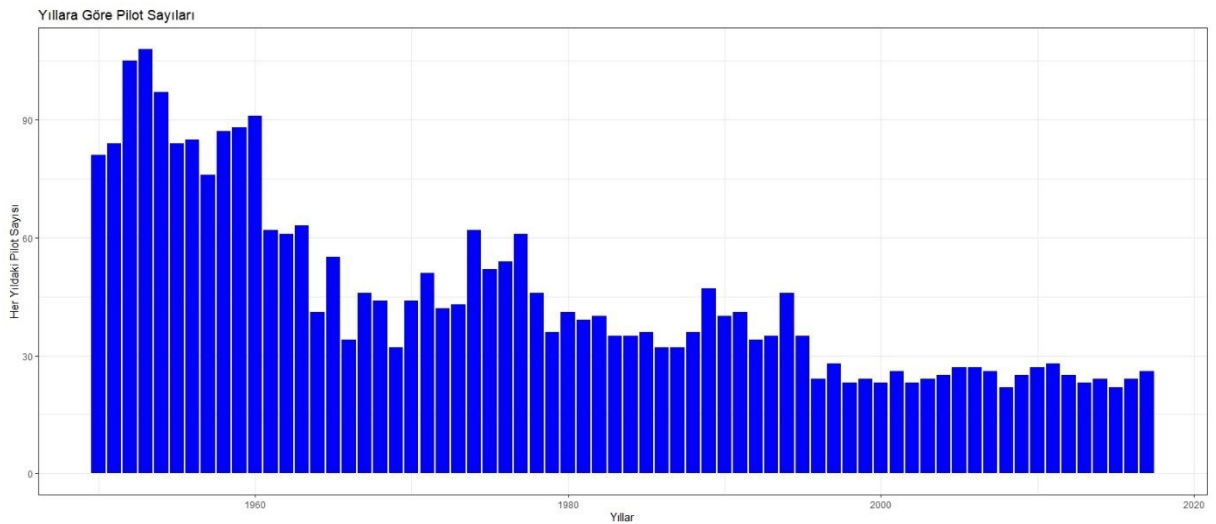
Bu bölümde her bir sezondaki yarışan pilot sayıları yıllara göre verilmiştir. Şekil 1’deki grafiğe bakıldığında bir sezondaki yarışan pilot sayısı giderek düşmektedir. Grafikte görüldüğü üzere pilot sayıları sporun geliştiği son yıllarda bile dalgalı bir hal almıştır. 2003 yılından sonra her sezon 20 pilotun yarışması gerekirken bazı yıllarda bu rakamın üzerine çıkmıştır. Bunu sebebi ise yarışan her takımın yedek pilot bulundurmasıdır. Böylece pilotlardan birisinin yerine geçen yedek pilotun ismi de o sezon yarışan pilotlar listesine eklenmektedir [1]. Bu nedenle de grafik sabit değil dalgalı durumdadır.

Kod kısmında ilk olarak her bir sezondaki pilotlar tablo olarak elde edilmiştir. Daha sonrasında görselleştirme işlemi ile Şekil 1'deki grafik elde edilmiştir. Şekil 2'de ise tablo halinde her sezondaki pilotlar gösterilmiştir.

Kod:

```
##  
one_season_pilot_number <- F1data %>% group_by(season, full.name) %>%  
  summarise(number_of_pilot = n())%>%  
  group_by(season)%>%  
  summarise(number_of_pilot = n())  
one_season_pilot_number <- as.data.frame(one_season_pilot_number)  
datatable(one_season_pilot_number)  
  
ggplot(data = one_season_pilot_number)+  
  geom_col(aes(x=season,  
    y = number_of_pilot, ),fill = "blue", size=5)+  
  labs(x = "Yıllar",  
    y = "Her Yıldaki Pilot Sayısı",  
    title = "Yıllara Göre Pilot Sayıları")+  
  theme_bw()  
##
```

Çıktı:



Şekil 1

	season	number_of_pilot
1	1950	81
2	1951	84
3	1952	105
4	1953	108
5	1954	97
6	1955	84
7	1956	85
8	1957	76
9	1958	87
10	1959	88

Şekil 2

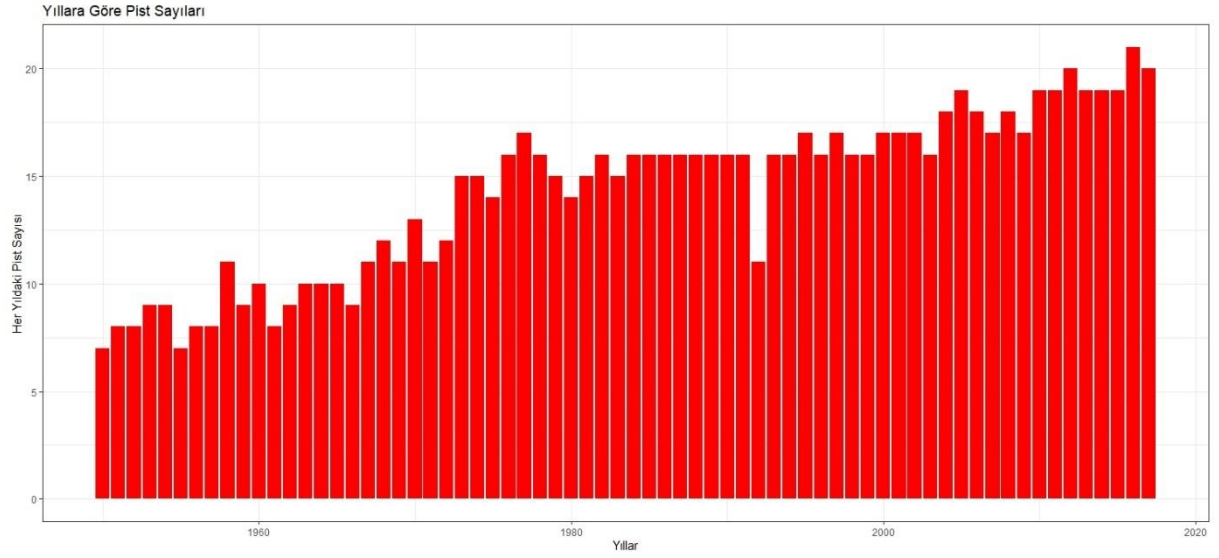
3. Her Bir Sezondaki Pist Sayısı

Bu bölümde her bir sezondaki pist sayısının yıllara göre değişimi incelenmiştir. Şekil 3'deki grafikte pist sayılarının değişimi gösterilmektedir. Gelişen teknoloji ile her yıl yarışılan pist sayısındaki artış grafikte gözlemlenmektedir. Şekil 4'te ise pist sayılarının tablo olarak gösterimi bulunmaktadır.

Kod:

```
##
ne_season_circuit_number <- F1data %>% group_by(season, circuitName) %>%
  summarise(number_of_circuit = n())%>%
  group_by(season)%>%
  summarise(number_of_circuit = n())
one_season_circuit_number
ggplot(data = one_season_circuit_number)+
  geom_col(aes(x=season,
               y = number_of_circuit ),fill = "red", size=5)+
  labs(x = "Yıllar",
       y = "Her Yıldaki Pist Sayısı",
       title = "Yıllara Göre Pist Sayıları")+
  theme_bw()
datatable(one_season_circuit_number, options = list(scrollX=T))
```

##



Şekil 3

	season	number_of_circuit
1	1950	7
2	1951	8
3	1952	8
4	1953	9
5	1954	9
6	1955	7
7	1956	8
8	1957	8
9	1958	11
10	1959	9

Şekil 4

4. 2000-2017 Yılları Arasında Yarışan Takımların Ülke Dağılımı

Şekil 5'te hangi takımın hangi millete ait olduğunu belirten nokta grafiği gösterilmektedir. Buna göre yarışmada en fazla takımın ait olduğu millet İngilizlerdir.

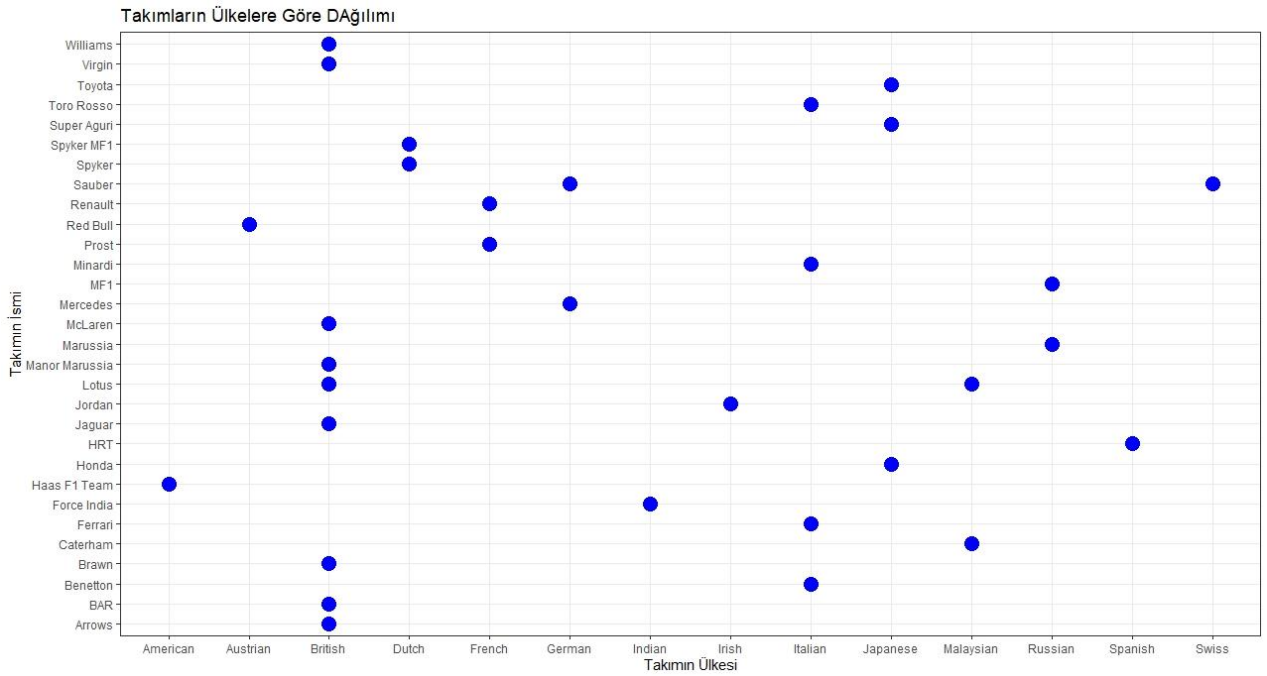
Kod:

##

```
F1data %>%filter(season>2000) %>% select(Constructor.name, Constructor.nationality)
%>%
```

```
ggplot()+
  geom_point(aes(x= Constructor.nationality,
    y = Constructor.name), color = "blue", size = 5)+
  labs(x = "Takımın Ülkesi",
    y = "Takımın İsmi",
    title = "Takımların Ülkelere Göre DAğılımı")+
  theme_bw()
##
```

Çıktı:



Şekil 5

5. Her Bir Sezonun Şampiyon Pilotunun Bulunması

Bu bölümde her bir sezonda hangi pilotun sezon birincisi olarak tamamladığı bulunmuştur. Şekil 6’da tablo olarak gösterilen sonuçlarda “total” olarak isimlendirilen değişken; şampiyon olan pilotun o sezon topladığı toplam puanı göstermektedir. Şekil 7’de ise şampiyon pilotlar görsel olarak nokta grafik ile gösterilmiştir.

Kod:

```
##
sampionlar<-F1data %>% select(season,full.name,results.grid, points) %>%
```

```

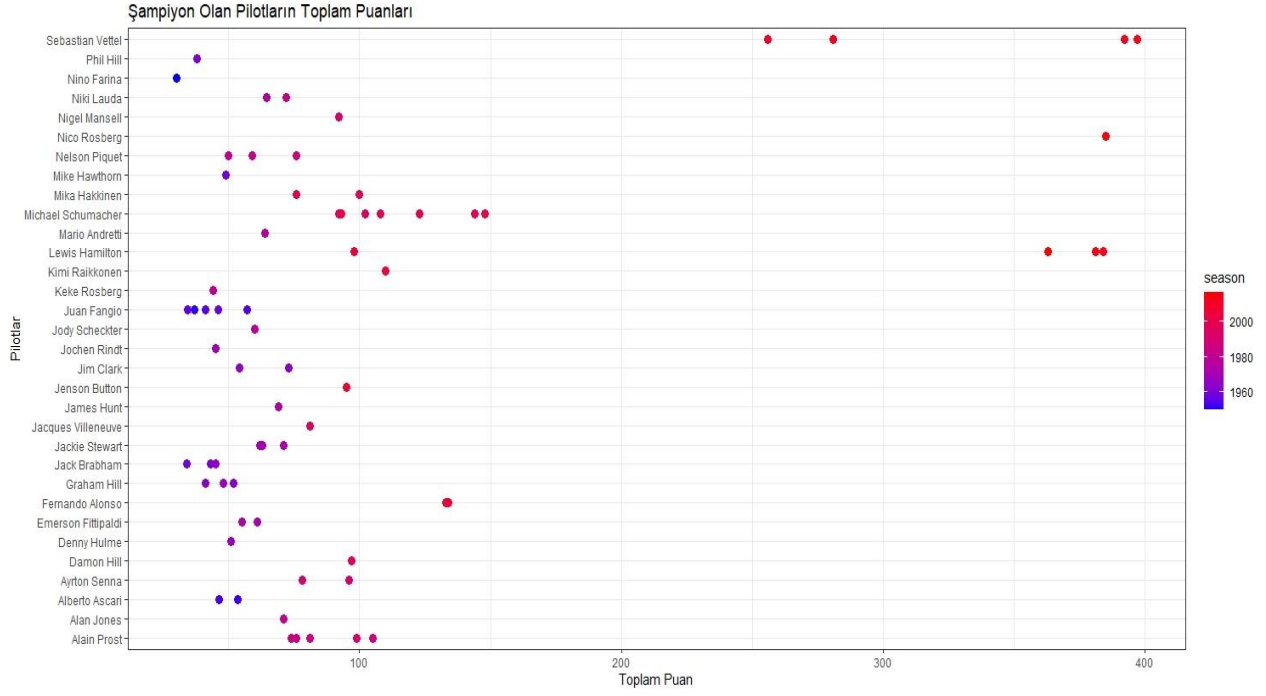
group_by(full.name, season) %>% summarise(total = sum(points)) %>%
group_by(season) %>%
top_n(1) %>%
arrange(season)
datatable(sampiyonlar)
ggplot(data = sampiyonlar)+
geom_point(aes(x = total,
               y = full.name,
               color = season), size=3)+
scale_colour_gradient(low = "blue", high = "red", na.value = NA)+
labs(x = "Toplam Puan",
     y = "Pilotlar",
     title = "Şampiyon Olan Pilotların Toplam Puanları")+
theme_bw()
##

```

Çıktı:

	full.name	season	total
61	Sebastian Vettel	2010	256
62	Sebastian Vettel	2011	392
63	Sebastian Vettel	2012	281
64	Sebastian Vettel	2013	397
65	Lewis Hamilton	2014	384
66	Lewis Hamilton	2015	381
67	Nico Rosberg	2016	385
68	Lewis Hamilton	2017	363

Şekil 6



Şekil 7

6. Pilotların Yarışlardaki Yaşları

Bu bölümde her bir pilotun yarıştığı yarıştaki yaşı hesaplanmıştır. İlk olarak verideki “dateOfBirth” değişkeninde bulunan bozuk tarihler düzeltilmiştir. Bazı tarihlerin “-” işareti ile ayrılması ve yıl-ay-gün yerleşimlerinin farklı olması nedeniyle düzenleme gerektirmektedir. Verinin düzenlenmesinin ardından “/” işareti ile ayrılan tarih verisi birbirinden ayrılmış ve yarışın tarihinde pilotun doğum tarihinin çıkarılması ile veriye “pilot_age” değişkeni eklenmiştir. Şeki 8’deki tabloda verinin bir kısmı gösterilmektedir.

Kod:

```
##
# "-" ile ayrılan tarihlerin düzeltilmesi
for (i in 1:length(F1data$dateOfBirth)-1){
  f1split <- strsplit(F1data$dateOfBirth[i], split = "")
  f1split <- unlist(f1split)
  if("-" %in% f1split){
    f1split <- strsplit(F1data$dateOfBirth[i], split = "-")
    f1split <- unlist(f1split)
    f1split_1 <- f1split[1]
    f1split_2 <- f1split[2]
    f1split_3 <- f1split[3]
```

```

f1split[1]<- f1split_2
f1split[2]<- f1split_3
f1split[3]<- f1split_1
f1splitx <- paste(f1split[1],f1split[2],f1split[3])
f1splitx <- gsub(pattern = " ", "/", f1splitx)
F1data$dateOfBirth[i] <-f1splitx
}
}
F1data$dateOfBirth
# Gün,ay ve yılın ayrılması
F1data$race_day <- str_split(F1data$date,"/",simplify = T)[,1]
F1data$race_month <- str_split(F1data$date,"/",simplify = T)[,2]
F1data$race_year <- str_split(F1data$date,"/",simplify = T)[,3]
F1data$pilot_day <- str_split(F1data$dateOfBirth, "/",simplify =T)[,1]
F1data$pilot_month <- str_split(F1data$dateOfBirth, "/",simplify =T)[,2]
F1data$pilot_year <- str_split(F1data$dateOfBirth, "/",simplify =T)[,3]
F1data %<>% relocate(race_day, .after=date)
F1data %<>% relocate(race_month, .after=race_day)
F1data %<>% relocate(race_year, .after=race_month)
F1data %<>% relocate(pilot_day, .after=dateOfBirth)
F1data %<>% relocate(pilot_month, .after=pilot_day)
F1data %<>% relocate(pilot_year, .after=pilot_month)
datatable(F1data, options = list(scrollX=T))
F1data$race_year <- as.integer(F1data$race_year)
F1data$pilot_year<- as.integer( F1data$pilot_year)
for(i in 1:length(F1data$dateOfBirth)-1){
  F1data$pilot_age[i] = (F1data$race_year[i] - F1data$pilot_year[i])
}
F1data %<>% relocate(pilot_age, .after=pilot_year)
datatable(F1data, options = list(scrollX=T))
##
Çıkt:

```

familyName	full.name	dateOfBirth	pilot_day	pilot_month	pilot_year	pilot_age
FARINA	Nino Farina	10/30/1906	10	30	1906	44
FAGIOLI	Luigi Fagioli	06/09/1898	06	09	1898	52
PARNELL	Reg Parnell	7/2/1911	7	2	1911	39
CABANTOUS	Yves Cabantous	10/8/1904	10	8	1904	46
ROSIER	Louis Rosier	11/5/1905	11	5	1905	45
GERARD	Bob Gerard	1/19/1914	1	19	1914	36
HARRISON	Cuth Harrison	7/6/1906	7	6	1906	44
ETANCELIN	Philippe Etancelin	12/28/1896	12	28	1896	54
HAMPSHIRE	David Hampshire	12/29/1917	12	29	1917	33
FRY	Joe Fry	10/26/1915	10	26	1915	35

Şekil 8

7. Her Bir Pilotun Kazandığı Yarış Sayısı

Bu bölümde her bir pilotun kazandığı yarış sayısı hesaplanmıştır. İlk olarak “Lead_pilot()” fonksiyonu yazılarak, girilen pilot isminin ne kadar yarışı kazandığı bulunmaktadır. Bu yazılan fonksiyon her bir pilot için uygulanmıştır. Şekil 9’da tablo halinde pilotların kazandıkları yarış sayısı verilmiştir. Şekil 10’da ise grafik halinde pilotların başarısı görselleştirilmiştir. Grafiğin daha net anlaşılması için kazanılan yarış sayı 5’ten büyük olacak şekilde filtrelenmiştir.

Kod:

```
##
#Fonksiyon
Lead_pilot <- function(x){
  x<- toupper(x)
```

```

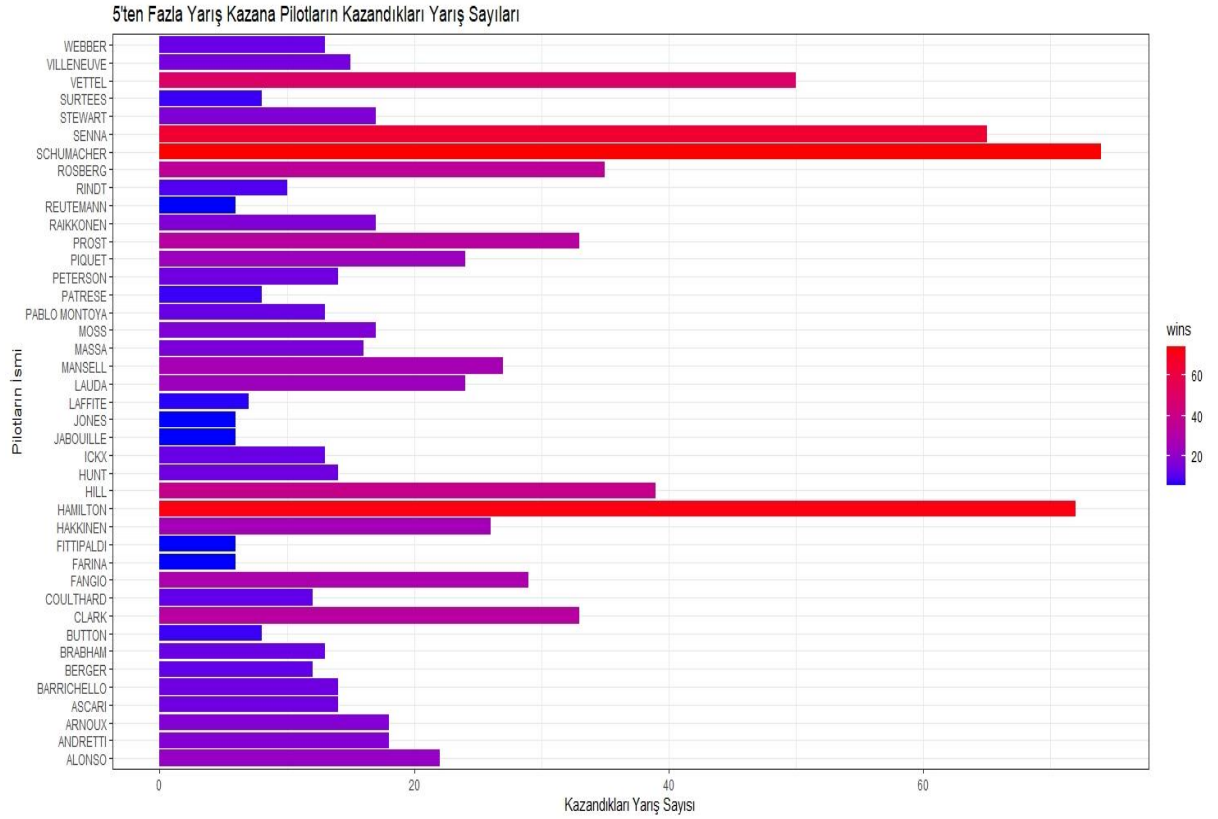
counter <- 0
F1data$familyName <- toupper(F1data$familyName)
F1data$familyName
if(!is.character(x)){
  stop('Bu Fonksiyon Sadece Pilot İsimleri İle Çalışmaktadır\n',
    ' classes:\n',
    'x: ', class(x))
}
if( x %in% F1data$familyName){
}else{
  stop('Aradığınız Kişi Pilotlar Arasında Yok')
}
for (i in 1:length(F1data$familyName)-1) {
  if (x %in% F1data$familyName[i]){
    if (F1data$results.grid[i] == 1){
      counter<-counter+1
    }
  }
}
wins<- counter
wins
}
#test
Lead_pilot("SCHUMACHER")
#Pilot İsimlerinin ve Kazandıkları Yarış Sayılarının Birleştirilmesi
pilots<- F1data %>% select(familyName, results.grid) %>% group_by(familyName)
%>%
  summarise(race_number_for_each_pilot = n())
pilot_name <- pilots$familyName
wins <- sapply(pilot_name, Lead_pilot)
wins <- as.data.frame(wins, colnames(wins))
wins<- wins$wins
wins <- data.frame(pilot_name,wins)
datatable(wins)
#Görselleştirme

```

```
wins %>% filter(wins > 5) %>%
  ggplot()+
  geom_col(aes(x=wins,
               y = pilot_name,
               fill = wins) ,size=2 )+
  scale_fill_gradient(low = "blue", high = "red", na.value = NA)+
  labs(x = "Pilotların İsmi",
       y = "Kazandıkları Yarış Sayısı",
       title = "5'ten Fazla Yarış Kazanan Pilotların Kazandıkları Yarış Sayıları")+
  theme_bw()
##
Çıktı:
> #test
> Lead_pilot("SCHUMACHER")
[1] 74
```

	pilot_name	wins ▼
654	SCHUMACHER	74
328	HAMILTON	72
661	SENNA	65
742	VETTEL	50
351	HILL	39
621	ROSBERG	35
154	CLARK	33
586	PROST	33
243	FANGIO	29
464	MANSELL	27

Şekil 9



Şekil 10

8. Pilotların Yarış Başına Aldıkları Puanlar

Bu bölümde, bir önce bölümde yazılan “Lead_pilot()” fonksiyonuna benzer şekilde, “Races_pilot()” fonksiyonu yazılarak her pilotun yarıştığı yarış sayısı elde edilmiştir. Ardından her pilotun topladığı toplam puan yarış sayısına bölünerek, pilotların yarış başına aldıkları puanlar elde edilmiştir. Şekil 11’deki tabloda pilotların yarış başına aldıkları puanlar gösterilmiştir. Şekil 12’de ise nokta grafiğinde pilotların yarış başına aldıkları puanlar gösterilmektedir.

Kod:

```
##
```

```
#katıldıkları yarışların hesaplanması için fonksiyon
```

```
Races_pilot <- function(x){
```

```
  x<- toupper(x)
```

```
  c <- 0
```

```
  F1data$familyName <- toupper(F1data$familyName)
```

```
  F1data$familyName
```

```
  for (i in 1:length(F1data$familyName)-1) {
```

```
    if (x %in% F1data$familyName[i]){
```

```
      c<-c+1
```

```

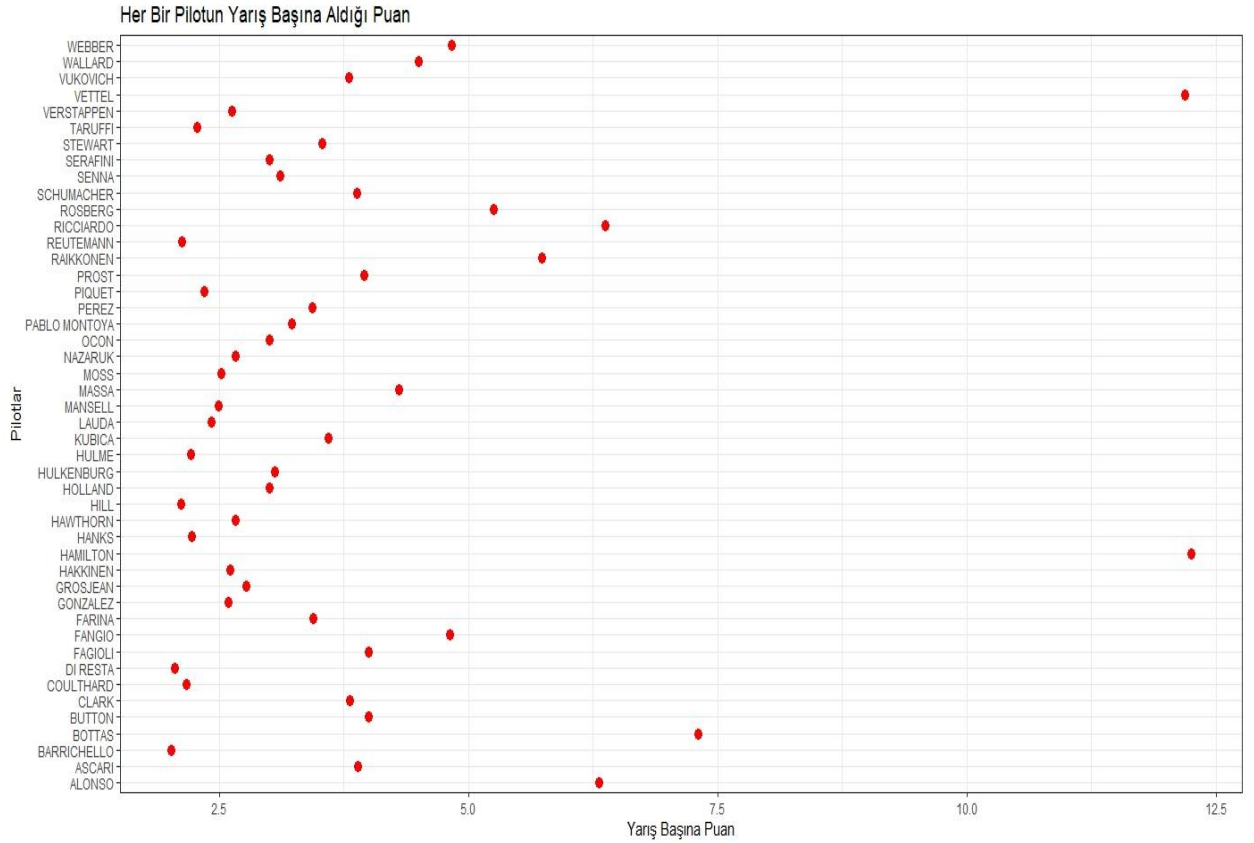
    }
  }
  races <- c
  races
}
#test
Races_pilot("SCHUMACHER")
pilots<- F1data %>% select(familyName, position) %>% group_by(familyName) %>%
  summarise(race_number_for_each_pilot = n())
pilot_name <- pilots$familyName
races <- sapply(pilot_name, Races_pilot)
races <- as.data.frame(races, colnames(races))
races<- races$races
races <- data.frame(pilot_name,races)
races
total_points<- F1data %>% select(familyName,position, points) %>%
  group_by(familyName) %>% summarise(total_points = sum(points))
total_points
yaris_basi_puan<- right_join(total_points,races, by = c("familyName"="pilot_name"))
datatable(yaris_basi_puan)
yaris_basi_puan<-yaris_basi_puan %>% mutate(yaris_basi_puan = total_points/races) %>%
  arrange(desc(yaris_basi_puan))
datatable(yaris_basi_puan)
yaris_basi_puan %>% filter(yaris_basi_puan>2) %>%
  ggplot()+
  geom_point(aes(x= yaris_basi_puan,
    y = familyName), color = "red", size = 3)+
  labs(x = "Yarış Başına Puan",
    y = "Pilotlar",
    title = "Her Bir Pilotun Yarış Başına Aldığı Puan")+
  theme_bw()
##
Çıktı:

> #test
> Races_pilot("SCHUMACHER")
[1] 483

```

	familyName	total_points	races	yaris_basi_puan
1	HAMILTON	2610	213	12.2535211267606
2	VETTEL	2425	199	12.1859296482412
3	BOTTAS	716	98	7.30612244897959
4	RICCIARDO	816	128	6.375
5	ALONSO	1849	293	6.31058020477816
6	RAIKKONEN	1565	273	5.73260073260073
7	ROSBERG	1754	334	5.25149700598802
8	WEBBER	1047.5	217	4.82718894009217
9	FANGIO	279.14	58	4.81275862068965
10	WALLARD	9	2	4.5

Şekil 11



Şekil 12

Kaynaklar

- [1] https://tr.wikipedia.org/wiki/Formula_1_sezonlar%C4%B1_listesi
- [2] <https://data.world/sportsvizsunday/sports-viz-sundays-2018/workspace/file?filename=F1+All+Data.xlsx>