

Şampiyonluk Mücadelesi Olan Superbowl'u Kazanmayı Etkileyen Faktörler Nelerdir?

Suna Temelcioğlu*

Özet

Amerika Birleşik Devletlerinde popüler spor organizasyonu olan Superbowl etkinliğinin kendine has kuralları bulunmaktadır. Turnuvada 32 takım, 2 konferansa ayrılmakta ve bu konferanslar da kendi içerisinde 4 takım oluşturulmaktadır. Konferans içerisinde yapılan eşleşmeler sonucunda playoff turuna katılmaya hak kazanan takımlar arasında son karşılaşmalar yapılarak kupanın sahibi bulunmaktadır. Ancak takımların playoff turuna kalıp kalmayacağı belirlenirken alışlagelmiş kurallar dışında etmenler bulunmaktadır. Alışlagelmiş lig kurallarından farklı olarak galibiyete göre sıralama yapılmamaktadır. Sıralamada tarafların galibiyet mağlubiyet sayıları yanında averajları, kazanma marjları, defans ve hücum sıralamaları, basit sıralamaları, kazanma marjları gibi istatistikler de hesaba katılmaktadır. Bu çalışmada takımların, Superbowl kupasını kazanabilmek için hangi kriterlere önem vermesi gerektiği incelenmiştir. Bu inceleme yapılırken Github üzerinden bulunan organizasyona 2000 ve 2019 yılları arasında katılan takımların istatistiklerinden oluşan veri seti kullanılmıştır. Yer alan değişkenlerin hepsinin nihai hedef olan kupayı kazanmaya etki ağırlığı farklıdır. Hangi değişkenin takımı nihai hedefe götürme konusunda önem taşıdığı, hangi değişkenlerin ise nihai hedefin sağlanmasında istatistiki olarak anlam göstermediğinin incelendiği bu çalışmada her takımın sezon boyunca katıldığı 16 karşılaşmanın en az yarısını kazanmasının kupayı kazanmada etkisinin olup olmadığı t testi ile, kazanma sayısı ile üretilen skor arasında bir ilişki olup olmadığı doğrusal regresyon analizi ile, takımın savunma sıralamasında iyi olup olmamasının playoff turuna kalmasında etkisi olup olmadığı Anova testi ile araştırılmıştır. Araştırma neticesinde bulunan sonuçlar neticesinde takımların Superbowl kupasını kazanmak için wins ve points for değişkenine önem vermesi gerektiği, defensive ranking değişkeninin playoff turuna katılma ile ters yönlü ilişkisi olduğu tespit edilmiştir.

*19080245, [Github Repo](#)

1 Final Hakkında Önemli Bilgiler

2 Giriş

BBu çalışmada Github üzerinden bulunan veri seti ile 2000 ve 2019 yılları arasında Amerikan Futbol Liginde yarışan takımların verileri değerlendirildi. Çalışma kapsamında her yıl ligde yer alan 32 takımın galibiyet ve mağlubiyet sayıları, maçlarda aldıkları toplam puanlar ile kaybettikleri toplam puanlar ve bu puan farkı ile belirlenen averajları, takımların ligdeki basit sıralamaları, hücum sıralamaları ve savunma sıralamaları, rakiplerin gücü, galibiyet marjları, playoff'a kalıp kalmama ve Superbowlu kazanıp kazanmama durumları her yıl için ayrı ayrı değerlendirildi.

2.1 Çalışmanın Amacı

Bu araştırma ile Amerika Birleşik Devletlerinin en popüler spor organizasyonu olan Superbowl etkinliğinde kupayı kazanmayı etkileyen faktörleri araştırmak hedeflenmiştir. Superbowl Öncesinde ülkenin farklı şehir ve eyaletlerinden gelen 32 takım, 2 konferansa ayrılır (AFC ve NFC). Konferans içerisindeki 16 takım da, 4er kişiden oluşan 4 takım oluşturur. Her konfe- ranstan 6 takım Playoffs'a kalmaya hak kazanır. Playoffs'a kalmaya hak kazanan takımlar, her grubun 1.si ve en iyi 2 adet 2.si olmaktadır. Playoffta eleme usulüyle birer maç oyna- nır ve kalan son 2 takım arasında yapılan maç ile Amerikan Futbol Liginin şampiyonluk mücadelesi olan Superbowl'un kazananı belirlenir. Superbowl'u kazanmak için ilk gereklilik olan playoff'a kalmak için sadece galip olmak yeterli gelmemektedir. Galip gelinen takımın ne kadar kuvvetli olduğu, diğer takımların atmış oldukları gol sayısı gibi doğrudan takımı etkilemeyen faktörler de gruptan çıkmak için önemli kriterler arasında yer almaktadır. Kullanılan veri setinde yer alan wins değişkeni, takımların sezonda kazandığı karşılaşma sayısını, loss değişkeni takımların sezon boyunca kaybettiği karşılaşma sayısını, points for değişkeni takımların sezon boyunca ürettikleri skoru, points against değişkeni rakiplerin ürettikleri skoru, points differential değişkeni ise skorların birbirinden çıkarılması ile bulunan averajı ifade etmektedir. Margin of victory takımların her maçta kazanma marjını ifade eder. Kazanma marjı, takımın sezon boyunca kazandığı/kaybettiği karşılaşmaları, karşılaşma başına ortalama kaç sayı önde/geride tamamladığının hesaplanmasıdır. Strenght of schedule takımların sezon boyunca karşılaştıkları rakiplerinin kazanma oranlarını ifade eder. Bu değişken Superbowl turnuvasında her takım birbiri ile karşılaşmadığından daha güçlü takımlarla karşılaşan takımlar aleyhine bir dezavantaj oluşmasını engellemektedir. Simple rating basit sıralamayı, Offensive ranking değişkeni hücum sıralamasını, Defensive ranking değişkeni, savunma sıralamasını, Playoffs değişkeni takımların o sezonda playoff turuna kalıp kalmadığını, Sb winner değişkeni, her sezon Superbowl kupasını kazanan takımı ifade etmektedir. Gruptan çıkmak için bakılan kriterlerin birbiri ile ilişkisinin araştırıldığı bu çalışma ile, takımların Superbowl'u kazanmasını etkileyen kriterleri neler olduğu araştırılacaktır.

2.2 Literatür

Amerika Birleşik Devletleri'ndeki en popüler düzenli televizyon yayını olan Super Bowl, Amerika Birleşik Devletleri'nde her yıl düzenlenen, National Football League'in sezon şampiyonunu belirleyecek olan şampiyonluk maçıdır. kalan son 2 takım arasında yapılan bu maç ile önceki yılın yaz aylarında başlamış olan sezon sonlanmakta ve sezonun şampiyonu belirlenmektedir. @Redelmeier Super Bowl, ilk elli yılında Amerika Birleşik Devletleri'nde sıradan bir lig şampiyonası maçından ibaretken, günümüzde ülke sınırları içerisinde hem bir tatil günü hem de bir mega etkinlik olarak kutlanmaktadır. Dyreson M. (2017) Kimlik teorisine göre bireyler kendilerini çeşitli sosyal gruplara üyeliklerine göre tanımlar. Brown (2000) Mevcut bir çalışmada, gelişmiş kolektif benlik saygısı, bir spor etkinliğine ev sahipliği yapıldığında bireyin bir topluluk olarak artan moralini ifade eder. @Brewer Bu nedenle halkın Super Bowl'a olan ilgisi incelenmesi gereken bir durumdur. Amerika'da Super Bowl'u ortalama olarak 100 milyon seyirci izlemektedir. Super Bowl Amerikan televizyon yayınlarında tarihin en fazla reytingini almaktadır. @Ceviker SuperBowl, Amerika Birleşik Devletleri vatandaşları için sadece bir final maçından ibaret değildir. Final maçının yayınlandığı tarihte herkes televizyon başındadır. Hatta yapılan bir araştırma ile Super Bowl televizyon yayınından sonra araba kazaları neticesinde yaşanan ölümlerin yüzde 41'lik göreceli artış oranında arttığı tespit edilmiştir. @Redelmeier

3 Veri

Çalışmada kullanılan veri setinin kaynağı Github'dır, ham veri üzerinde herhangi bir işlem yapılmamıştır. Veri seti ile ilgili özet istatistikler aşağıda verilmiştir.

```
library(readxl)
library(tidyverse)
library(dplyr)
library(readxl)
superbowl_data <- read_excel("~/superbowl_data.xlsx")
View(superbowl_data)
```

3.1 Özet İstatistikler

```
summary(superbowl_data)
```

##	team	team_name	year	wins
##	Length:638	Length:638	Min. :2000	Min. : 0.000
##	Class :character	Class :character	1st Qu.:2005	1st Qu.: 6.000
##	Mode :character	Mode :character	Median :2010	Median : 8.000
##			Mean :2010	Mean : 7.984

```

##                               3rd Qu.:2015   3rd Qu.:10.000
##                               Max.      :2019   Max.      :16.000
##      loss      points_for  points_against  points_differential
##  Min.      : 0.000   Min.      :161.0   Min.      :165.0   Min.      : -261.00
##  1st Qu.: 6.000   1st Qu.:299.0   1st Qu.:310.0   1st Qu.: -75.00
##  Median : 8.000   Median :348.0   Median :347.0   Median :   1.50
##  Mean   : 7.984   Mean   :350.3   Mean   :350.3   Mean   :   0.00
##  3rd Qu.:10.000   3rd Qu.:396.0   3rd Qu.:391.5   3rd Qu.:  72.75
##  Max.   :16.000   Max.   :606.0   Max.   :517.0   Max.   : 315.00
##  margin_of_victory  strength_of_schedule  simple_rating
##  Min.      :-16.300000   Min.      :-4.600000   Min.      :-17.400
##  1st Qu.: -4.700000   1st Qu.: -1.100000   1st Qu.: -4.475
##  Median :  0.100000   Median : 0.000000   Median :  0.000
##  Mean   : -0.001881   Mean   : 0.001097   Mean   :  0.000
##  3rd Qu.:  4.575000   3rd Qu.: 1.200000   3rd Qu.:  4.500
##  Max.   : 19.700000   Max.   : 4.300000   Max.   : 20.100
##  offensive_ranking  defensive_ranking      playoffs      sb_winner
##  Min.      :-11.700000   Min.      :-9.800000   Min.      :0.0000   Min.      :1.000
##  1st Qu.: -3.175000   1st Qu.: -2.400000   1st Qu.:0.0000   1st Qu.:1.000
##  Median :  0.000000   Median : 0.100000   Median :1.0000   Median :1.000
##  Mean   : -0.000157   Mean   : -0.001097   Mean   :0.6238   Mean   :1.031
##  3rd Qu.:  2.700000   3rd Qu.: 2.500000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   : 15.900000   Max.   : 9.800000   Max.   :1.0000   Max.   :2.000

```

4 Yöntem ve Veri Analizi

4.1 ANOVA testi

ANOVA testi, playoffs ve defensive ranking arasında yapılmıştır. `data <- data.frame(superbowl_datadefensive_ranking)`
`model <- lm(superbowl_dataplayoffs ~ factor(superbowl_datadefensive_ranking), data = data)`
`anova_result <- anova(model)`
`print(anova_result)`

Defensive ranking değişkeni ile playoffs değişkeni arasında anlamlı bir ilişki olup olmadığının test edilmesi için ANOVA testi yapılmıştır. Yapılan ANOVA testi sonuçları incelendiğinde playoffs değişkeninin defensive_ranking değişkenine göre istatistiksel olarak anlamlı bir ilişkisi olduğunu görülmektedir. ANOVA analizinde, faktör değişken (defensive_ranking) ile yanıt değişkeni (playoffs) arasındaki varyansın farklılığının anlamlılığını test ederiz. ANOVA tablosunda, `factor(superbowl_data$defensive_ranking)` satırı faktör değişkenin etkisini temsil ederken, Residuals satırı hata varyansını temsil eder. Df sütunu ise, serbestlik derecelerini gösterir. Sum Sq sütunu, toplam kareler toplamını temsil eder. Mean Sq sütunu, ortalama kareler toplamını gösterir. F value sütunu, faktör etkisinin istatistiksel anlamlılığını ifade eden F istatistiğini verir. `Pr(>F)` sütunu, p değerini gösterir. Analiz neticesinde elde edilen sonuçlar, faktör değişken olan defensive_ranking'in playoffs değişkenine etkisi olduğunu ve

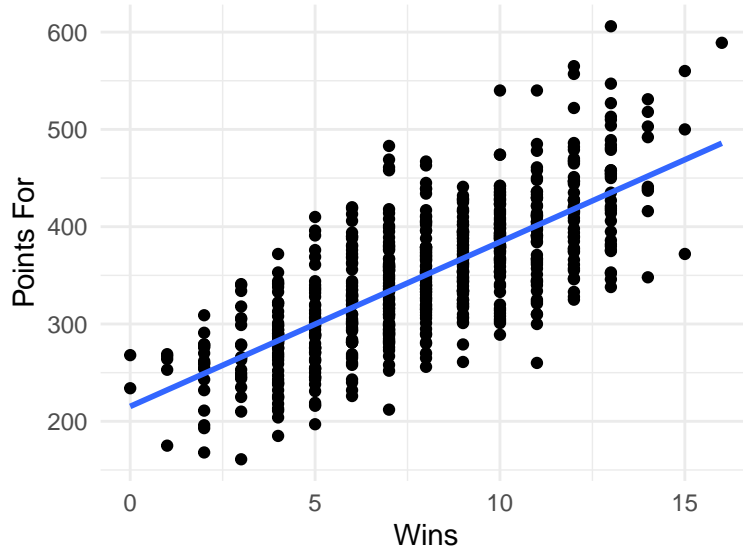
bu etkinin istatistiksel olarak anlamlı olduğunu göstermektedir. $Pr(>F)$ değeri 0.001'e yakın olduğu için, bu ilişkinin tesadüfi olmadığı ve anlamlı olduğu sonucuna varabiliriz.

##Basit Doğrusal Regresyon

```
data <- data.frame(superbowl_data$wins, superbowl_data$points_for)
model <- lm(superbowl_data$points_for ~ superbowl_data$wins, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = superbowl_data$points_for ~ superbowl_data$wins,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -141.260  -34.645   -0.049   28.813  170.933
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    215.3192     5.3691   40.10  <2e-16 ***
## superbowl_data$wins  16.9037     0.6274   26.94  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.83 on 636 degrees of freedom
## Multiple R-squared:  0.533, Adjusted R-squared:  0.5323
## F-statistic: 726 on 1 and 636 DF, p-value: < 2.2e-16
```

```
library(ggplot2)
ggplot(data, aes(x = superbowl_data$wins, y = superbowl_data$points_for)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "Wins", y = "Points For") +
  theme_minimal()
```



Bu araştırma sırasında wins ve points for değişkeni arasında basit doğrusal regresyon analizi yapılmıştır. Yapılan regresyon analizi sonucunda wins değişkeninin points for değişkeni üzerinde anlamlı bir etkiye sahip olduğu gözlemlenmiştir. Katsayılar bölümünde wins değişkeninin eğim (slope) katsayısı 16.9037 olarak hesaplanmıştır bu katsayı her bir ekstra kazanılan maç için ortalama olarak puanların 16.9037 artacağını ifade etmektedir. Ayrıca, kesme noktası (intercept) katsayısı olan Intercept için tahmin edilen değer 215.3192'dir ki bu katsayı da kazanılan maç sayısı sıfır olduğunda beklenen ortalama puanın 215.3192 olduğunu ifade etmektedir. Analize ayrıca R-kare (R-squared) değeri 0.533 olarak hesaplanmıştır, yani modelin açıkladığı varyansın %53.3'ünü wins değişkeni açıklamaktadır ki bu modelin veri setindeki değişikliğin büyük bir kısmını açıkladığını göstermektedir. Son olarak, analiz neticesinde F istatistiği 726 olarak hesaplanmış ve p-değeri $< 2.2e-16$ (çok düşük) olarak bulunmuştur. Bu da, wins değişkeninin puanları açıklamada istatistiksel olarak anlamlı olduğunu göstermektedir. Bu sonuçlar, wins ve points for değişkeni arasında pozitif bir ilişki olduğunu ve kazanılan her maçın takımın puanlarını artırdığını göstermektedir

##t-testi

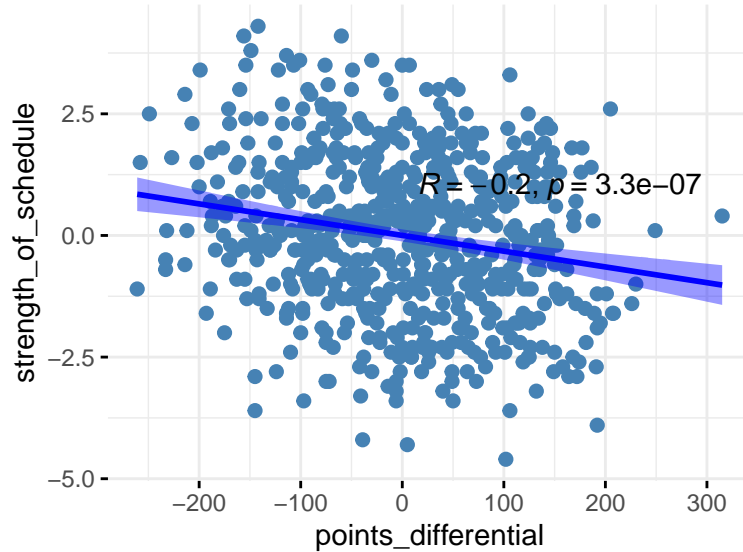
```
t.test(superbowl_data$wins~ superbowl_data$sb_winner)
```

```
##
##  Welch Two Sample t-test
##
## data:  superbowl_data$wins by superbowl_data$sb_winner
## t = -11.696, df = 25.007, p-value = 1.238e-11
## alternative hypothesis: true difference in means between group 1 and group 2 is not e
## 95 percent confidence interval:
##  -4.693501 -3.288052
## sample estimates:
## mean in group 1 mean in group 2
##      7.859223      11.850000
```

Her takım her sezonda 16 karşılaşma yapmaktadır. Takımların sezon boyunca çıktıkları karşılaşmaların yarısını kazanmasının Superbowl kupasını kazanması ile ilişkisi olup olmadığının araştırılması için t testi uygulanmıştır. Yapılan t testinde, H_0 :Kazanan takımların ortalama kazanma(win) sayısı = 8'dir.Yapılan testin sonuçlarına bakıldığında, t değeri: -11.696, serbestlik derecesi (df): 25.007, p değeri: 1.238e-11 (çok düşük bir p değeri), hipotez testi sonucunda p değeri, 0.05 anlamlılık düzeyinden çok daha düşük bulunmuştur. Bu da kazanan takımların ortalama win sayısı ile 8 arasında anlamlı bir fark olduğunu gösterir. Örneklemeye verilerine dayanarak elde ettiğimiz tahminlere göre, grup 1'in (sb_winner = 0, yani kaybeden takımlar) ortalama win sayısı 7.859223, grup 2'nin (sb_winner = 1, yani kazanan takımlar) ortalama win sayısı ise 11.85'tir. Sonuç olarak, bu t-testi sonucuna dayanarak, kazanan takımların ortalama win sayısının 8'den anlamlı bir şekilde farklı olduğunu söyleyebiliriz.

##Scatter Plot

```
library("ggpubr")
ggscatter(superbowl_data,
  x = "points_differential",
  y = "strength_of_schedule",
  color = "steelblue",
  add = "reg.line",
  add.params = list(color = "blue"),
  conf.int = TRUE,
  cor.coef = TRUE,
  cor.coef.coord = c(15, 1),
  cor.method = "pearson",
  ggtheme = theme_minimal())
```



##Normallik için Shapiro Wilk Testi

```
shapiro.test(superbowl_data$points_differential)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  superbowl_data$points_differential  
## W = 0.99391, p-value = 0.01136
```

p değeri ($p = 0.01136$) 0.05 anlamlılık düzeyinden küçüktür, bu da verilerin dağılımının normal dağılımdan önemli ölçüde farklı olduğunu gösterir. Başka bir deyişle, normal dağılım olduğu varsayılmaz.

```
shapiro.test(superbowl_data$strength_of_schedule)
```

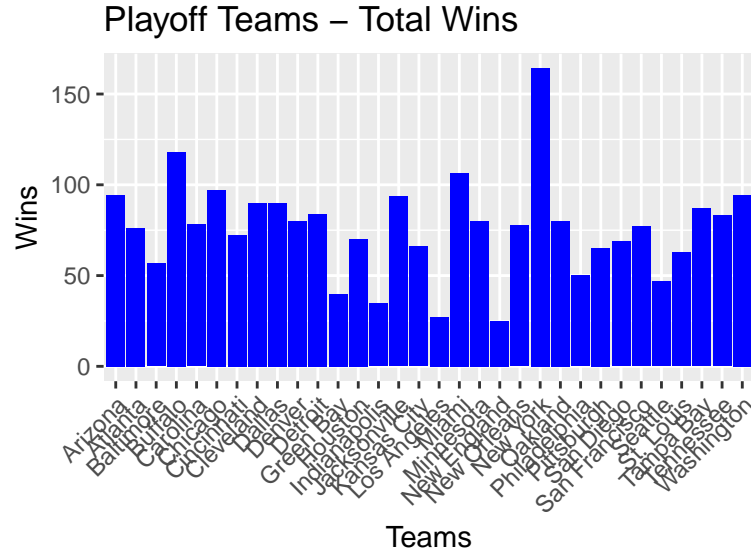
```
##  
## Shapiro-Wilk normality test  
##  
## data:  superbowl_data$strength_of_schedule  
## W = 0.99604, p-value = 0.1088
```

p değeri ($p = 0.1088$) 0.05 anlamlılık düzeyinden büyüktür, bu da verilerin dağılımının normal dağılımdan önemli ölçüde farklı olmadığını gösterir. Başka bir deyişle, normal dağılım olduğu varsayılır.

Veri setlerimizden birisi normal dağılımdan gelirken diğerinin normal dağılımdan gelmemesi halinde Spearman korelasyon testi yapmamız gerekir. Spearman korelasyonu, sıralı veriler arasındaki ilişkiyi değerlendirmek için kullanılır ve değişkenlerin dağılımının normal olmasını gerektirmez.

##Sütun grafiği çizimi

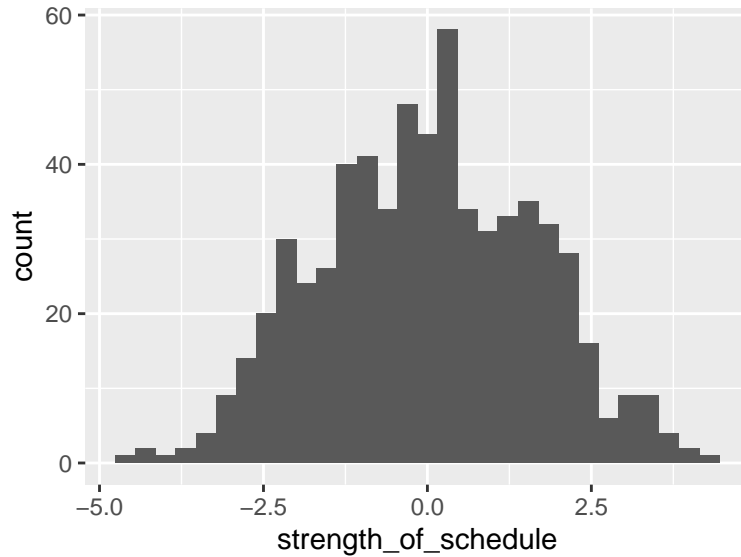
```
library(ggplot2)  
playoff_teams <- superbowl_data[superbowl_data$playoffs == 1, ]  
playoff_wins <- playoff_teams$wins  
data <- data.frame(Team = playoff_teams$team, Wins = playoff_wins)  
ggplot(data, aes(x = Team, y = Wins)) +  
  geom_bar(stat = "identity", fill = "blue") +  
  xlab("Teams") +  
  ylab("Wins") +  
  ggtitle("Playoff Teams - Total Wins") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Bu sütun grafiğinde x ekseninde playoffa kalan takımlar ve toplam galibiyet sayıları görünmektedir.

##Histogram Strength of schedule değişkeninin histogramı şeklindeki gibidir. y ekseninde frekans, x ekseninde strength of schedule değişkeni yer almaktadır.

```
library(ggplot2)
ggplot(superbowl_data) +
  geom_histogram(aes(x = strength_of_schedule))
```



5 Sonuç

Yukarıda yapılan istatistiki analizler neticesinde, takımların Superbowl kupasını kazanabilmesi için wins değişkeninin önem arz ettiğini, takımların galibiyete önem göstermeleri gerektiği sonucuna götürmektedir. Bir diğer analiz neticesinde wins değişkeni ile points for değişkeni arasında pozitif yönlü bir ilişki olduğu, points for değişkenindeki artışın wins değişkenini olumlu yönde etkilediği sonucuna ulaşılmıştır. Bu sonuç, Superbowl kupasını almak isteyen takımların skor üreterek karşılaşmalarda kazanması ve topladığı galibiyetler neticesinde Superbowl galibi olmasının istatistiki olarak anlamlı olduğunu göstermektedir. Yapılan bir diğer analiz neticesinde ancak defensive ranking değişkeni ile Superbowl'u kazanmanın gerekli koşulu olan playoff turuna kalma arasında ters yönlü ilişki olduğu sonucuna ulaşılmıştır. Bu analiz de, takımların Superbowl kupasını kazanabilmek için defensive ranking değişkenine vermeleri gereken önemin, diğer değişkenlere nazaran arasında alt sıralarda olduğu sonucuna götürmektedir.

Gruplardan çıkmak için bakılan kriterlerin birbiri ile ilişkisinin araştırıldığı bu çalışma ile, takımların Superbowl'u kazanmasını etkileyen kriterleri neler olduğu sorusuna cevap vermeye çalışılmıştır. Bulunan sonuçlar baz alınan değişkenlerin hepsinin aynı etki düzeyine sahip olmadığı, hatta bazı değişkenlerin ters yönlü etkiye neden olduğunu göstermiştir. Veri setinde yer alan değişkenlere takımların gruplardan çıkarak playoff turuna kalması ve neticesinde Superbowl kupasını kazanmasında belli bir kural sıralaması ile bakılmaktadır. Birbirini etkileyen değişkenlere farklı aşamalarda kural sırası gözetilerek bakıldığından aslında ilişkili olan, birbirini doğrudan etkileyen değişkenler sonucu aynı oranda etkileyememiştir. Bu nedenle kurallara göre daha önemli görünen değişken verilere göre daha önemsiz bulunabilmiştir. İleride bu çalışma, Amerikan Futbol Liginin grup sıralaması kuralları gözetilerek, kurallar ile belirlenen değişkenlerin öncelik sıralaması gözetilerek daha faydalı hale getirilebilir.

6 Kaynakça

- Brown, R. (2000). Social identity theory: Past achievements, current problems and future challenges. *European Journal of Social Psychology*.
- Dyreson M., H. P. &. (2017). Super Bowl Sunday: A National Holiday and a Global Curiosity. *The International Journal of the History of Sport*.