

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.

GITHUB REPO BAĞLANTINIZI BU DOSYANIN 37. SATIRINA YAZINIZ!

Proje gönderimi, Github repo linki ile birlikte ekampus sistemine bir zip dosyası yüklenerek yapılacaktır. Sisteme zip dosyası yüklemeyen ve Github repo linki vermezseniz ara sınav ve final sınavlarına girmemiş sayılırsınız.

*19080245, [Github Repo](#) # Final Hakkında Önemli Bilgiler

Proje klasörünüzü sıkıştırdıktan sonra (OgrenciNumarasi.zip dosyası) 9 Haziran 2023 23:59'a kadar ekampus.ankara.edu.tr adresine yüklemeniz gerekmektedir.

Daha fazla bilgi için proje klasöründeki README.md dosyasını okuyunuz.

1 Giriş

Bu 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 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 kazanımı 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 gruplardan 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. Gruplardan çıkmak için bakılan kriterlerin birbiri ile ilişkisinin araştırıldığı bu çalışma ile, takımların Superbowl'u kazanmasını etkileyen kriterleri neler olduğu araştırılacaktır.

3 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 maçdır. 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

4 Veri

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

4.1 Verileri import etme

```
library(tidyverse) library(readxl) superbowl_data <- read_excel("C:/Users/begum/OneDrive/Masaüstü/s  
View(superbowl_data) superbowl_data
```

4.2 Özet istatistikler

```
summary(superbowl_data) library(skimr) skimr::skim_without_charts(superbowl_data)
```

4.3 ANOVA testi

4.4 ANOVA testi, playoffs ve defensive ranking arasında yapılmıştır.

```
data <- data.frame(superbowl_data$defensive_ranking, superbowl_data$playoffs) model  
<- lm(superbowl_data$playoffs ~ factor(superbowl_data$defensive_ranking), data = data)
```

```
anova_result <- anova(model) print(anova_result)
```

4.5 Basit Doğrusal Regresyon

```
data <- data.frame(superbowl_datawins, superbowl_datapoints_for) model <- lm(superbowl_datapoints_for ~  
data = data) summary(model)
```

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

4.6 Normallik için Shapiro Wilk Testi

```
shapiro.test(superbowl_datapoints_differential)shapiro.test(superbowl_datapoints_strength_of_schedule)
```

5 t-testi

```
t.test(superbowl_datawins ~ superbowl_datapoints_winner)
```

6 Yöntem ve Veri Analizi

Yapılan ANOVA testi sonuçları incelendiğinde playoffs değişkeninin defensive_ranking değişkenine göre istatistiksel olarak anlamlı bir ilişkisi olduğunu göstermektedir. ANOVA analizinde, faktör değişken (defensive_ranking) ile yanıt değişkeni (playoffs) arasındaki varyansın farklılığının anlamlılığı test edilmiştir. ANOVA tablosunda, factor(superbowl_data\$defensive_ranking) satırı faktör değişkenin etkisini temsil ederken, Residuals satırı hata varyansını temsil eder. Df sütunu, özgürlük 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.

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 analizinde "Kazanılan maçlar (superbowl_datawins)veattubasketsayısı(points_for) için tahmin edilen değer 16.9037'dir.

Ayrıca, kesme noktası (intercept) katsayısı olan Intercept için tahmin edilen değer 215.3192'dir. Bu, kazanılan maç sayısı sıfır olduğunda beklenen ortalama puanın 215.3192 olduğunu gösterir.

R-kare (R-squared) değeri 0.533 olarak hesaplanmıştır, yani modelin açıkladığı varyansın %53.3'ünü kazanan maçlar değişkeni açıklamaktadır. Bu, modelin veri setindeki değişikliğinin büyük bir kısmını açıkladığını göstermektedir.

Son olarak, F istatistiği 726 olarak hesaplanmış ve p-değeri $< 2.2e-16$ (çok düşük) olarak bulunmuştur. Bu da, kazanan maçlar değişkeninin puanları açıklamada istatistiksel olarak anlamlı olduğunu göstermektedir.

Bu sonuçlar, kazanan maçlar ve attığı puanlar arasında pozitif bir ilişki olduğunu ve kazanan her maçın takımın puanlarını artırdığını göstermektedir.

$$Y_t = \beta_0 + \beta_N N_t + \beta_P P_t + \beta_I I_t + \varepsilon_t$$

Yapılan t testinde, H_0 :Kazanan takımların ortalama kazanma(win) sayısı = 8'dir. Bu t-test sonucunda elde edilen bilgiler şunlardır: 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 düşüktür. Bu da kazanan takımların ortalama win sayısı ile 8 arasında anlamlı bir fark olduğunu gösterir. Hipotez testi sonucunda p değeri, 0.05 anlamlılık düzeyinden çok daha düşüktür. Bu da kazanan takımların ortalama win sayısı ile 8 arasında anlamlı bir fark olduğunu gösterir. Örneklem 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.

Yapılan Shapiro-Wilk normallik testinde `superbowl_datapoints_differentialdeikeniiincelendiindepdeeri(p=0.01136)0.05anlamlılıkdzeyindenkktr, budaverilerindailımınınnormaldalımdannemlildefarklıolduun` değişkeni incelendiğinde 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.

7 Regresyon Analizi

```
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())
```

8 Normallik için Shapiro Wilk Testi

`shapiro.test(superbowl_data$points_differential)` 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) 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 diğeri gelmiyor. Bu durumda Spearman korelasyon testi yapmalıyız. Spearman korelasyonu, sıralı veriler arasındaki ilişkiyi değerlendirmek için kullanılır ve değişkenlerin dağılımının normal olmasını gerektirmez.

8.1 Sütun grafiği çizimi

```
ggplot(data, aes(x = superbowl_data$Team, y = superbowl_data$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))
```

8.2 Histogram

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

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

8.3 Playoffa kalan takımların kazanma sayıları

```
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 analiz yapılırken playoffa kalan takımların toplam galibiyet sayıları sütun grafiğinde gösterilmiştir.

9 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. Bu çalışmada Pearson Korelasyon Analizi yapmak hedeflenmiştir ancak değişkenlere Normallik için Shapiro Wilk Testi uygulandıktan sonra değişkenlerin ikisinin de normal dağılımdan gelmediği görülmüştür. 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.

Kaynakça bölümü Rmarkdown tarafından otomatik olarak oluşturulmaktadır. Taslak dosyada Kaynakça kısmında herhangi bir değişikliğe gerek yoktur.

Taslakta bu cümleden sonra yer alan hiçbir şey silinmemelidir.

10 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*.