

IMDB Top Filmler Analizi - Final Rapor

Okan Can

26 October 2025

Sinema sanayisinin küresel ölçekteki etkisini yansıtan en önemli göstergelerden biri olan Internet Movie Database (IMDB) “Top 250” listesi, izleyici ve eleştirmenlerin ortak beğenisini temsil eden bir başyapıtlar koleksiyonudur. Bu projenin temel amacı, 250 filmlik bu veri setini nicel analiz ve veri hikayeciliği yöntemleriyle inceleyerek, yüksek beğeni standartlarını belirleyen faktörlerin (yönetmen, dönem, puan dağılımı) istatistiksel bir portresini çizmektir. Analiz, filmlerin ham puanlarının ötesine geçerek, liste içindeki yapısal eğilimleri ve kalıcılığı sağlayan unsurları keşfetmeyi hedeflemektedir. Bu çalışma, basit bir konuya akademik bir bakış açısı getirerek, veri görselleştirme prensiplerini etkin bir şekilde uygulama fırsatı sunmaktadır.

Proje analizinde, yüklenen tam 250 filmlik veri seti kullanılmıştır. Analiz, mevcut veri kısıtlılığı nedeniyle, kategori bilgisi yerine yönetmen ve yayın yılı değişkenleri üzerine yoğunlaşmıştır. Hikaye, üç ana görselleştirme kategorisi etrafında inşa edilmiştir: İlk olarak, Miktar Görselleştirmeleri kapsamında, Top 250’ye en çok film sokan yönetmenler Sıralı Çubuk Grafikleri ile gösterilmiş ve bu yönetmenlerin filmlerinin Yığılmış Çubuk Grafikleri ile Klasik (1980 öncesi) ve Modern (1980 sonrası) dönemlere göre dağılımı incelenmiştir; bu, listenin güncel eğilimlerden ziyade, köklü sinema tarihine dayandığı argümanını destekler. İkinci olarak, Dağılım Görselleştirmesi için Kernel Yoğunluk Tahmini Grafiği kullanılarak, Top 250 filminin puanlarının dar bir aralıkta nasıl yoğunlaştığı gösterilmiş, böylece ortalama puanın belirginleştiği kanıtlanmıştır. Üçüncü olarak, Oransal Görselleştirme amacıyla Pasta Grafiği ile en çok temsil edilen yönetmenlerin listenin toplamındaki yüzdesel payları gösterilmiş, bireysel dehaların kolektif beğeniye nasıl domine ettiği vurgulanmıştır. Bu görselleştirmeler, R/ggplot2 kullanılarak oluşturulmuş olup, projenin tüm görselleştirme kriterlerini eksiksiz karşılamaktadır. Projenin nihai çıktısı olan poster, A1 dikey formatta ve BetterPoster metodolojisine uygun olarak, merkezi bir bulguyu güçlü grafiklerle destekleyecek; tüm kodlar, veri dosyası ve çıktılar ise GitHub deposunda yayımlanarak tekrarlanabilirlik gerekliliğini yerine getirmiştir. Bu proje için kullanılacak olan veri seti, IMDB Top 250 Movies başlığı altında, Kaggle platformunda Fatih İlhan adlı kullanıcı tarafından paylaşılmıştır. Veri, dünyanın en büyük film veri tabanı olan Internet Movie Database (IMDB)’nin belirlediği en yüksek puana sahip 250 filmin bir derlemesidir. Bu kapsamda, veri setinin birincil bilgi kaynağı IMDB’dir. Verinin Kaggle üzerindeki paylaşımını gerçekleştiren organizasyon veya kişi ise Fatih İlhan’dır. Veri setine aşağıdaki erişim bağlantısı üzerinden ulaşılabilir:

<https://www.kaggle.com/datasets/fatihilhan/imdb-top-250-movies>

Analizde kullanılan temel değişkenler Movie (Film Adı), Released_year (Yayın Yılı), Rating (IMDB Puanı) ve Director (Yönetmen) sütunlarını içermektedir. Bu veri setinin kullanım amacı, Top 250 film listesindeki yönetmenlerin temsil oranlarını, filmlerin yayın yılı trendlerini ve puan dağılımını analiz ederek, yüksek beğeni standartlarının ardındaki istatistiksel faktörleri görselleştirmektir.

Proje Analizi ve Görselleştirmeleri

R Projeleri İçin Temiz Giriş: Gerekli Kütüphaneler ve Veri Ön İşleme

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
```

```

## v forcats 1.0.1      v stringr 1.5.2
## v ggplot2 4.0.0      v tibble  3.3.0
## v lubridate 1.9.4     v tidyr   1.3.1
## v purrr    1.1.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)

# Dosya adının 'Movies.csv' olduğundan emin olun.
data <- read_csv("Movies.csv")

## Rows: 250 Columns: 4
## -- Column specification -----
## Delimiter: ","
## chr (3): Movie, Rating, Director
## dbl (1): Released_year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

# Veri Türlerini Dönüştürme
data <- data %>%
  mutate(
    Released_year = as.numeric(Released_year),
    # Rating'teki non-numerik değerler NA'ya çevrilir ve uyarılar bastırılır
    Rating = suppressWarnings(as.numeric(Rating))
  )

# Ara Veri Hesaplamaları
director_counts <- data %>%
  count(Director, sort = TRUE)

data_stacked <- data %>%
  mutate(Era = ifelse(Released_year < 1980, "Klasik (1930-1979)", "Modern (1980+)"))

director_pie_data <- director_counts %>%
  count(Director) %>%
  mutate(percentage = n / sum(n),
         label_y = cumsum(percentage) - percentage / 2)

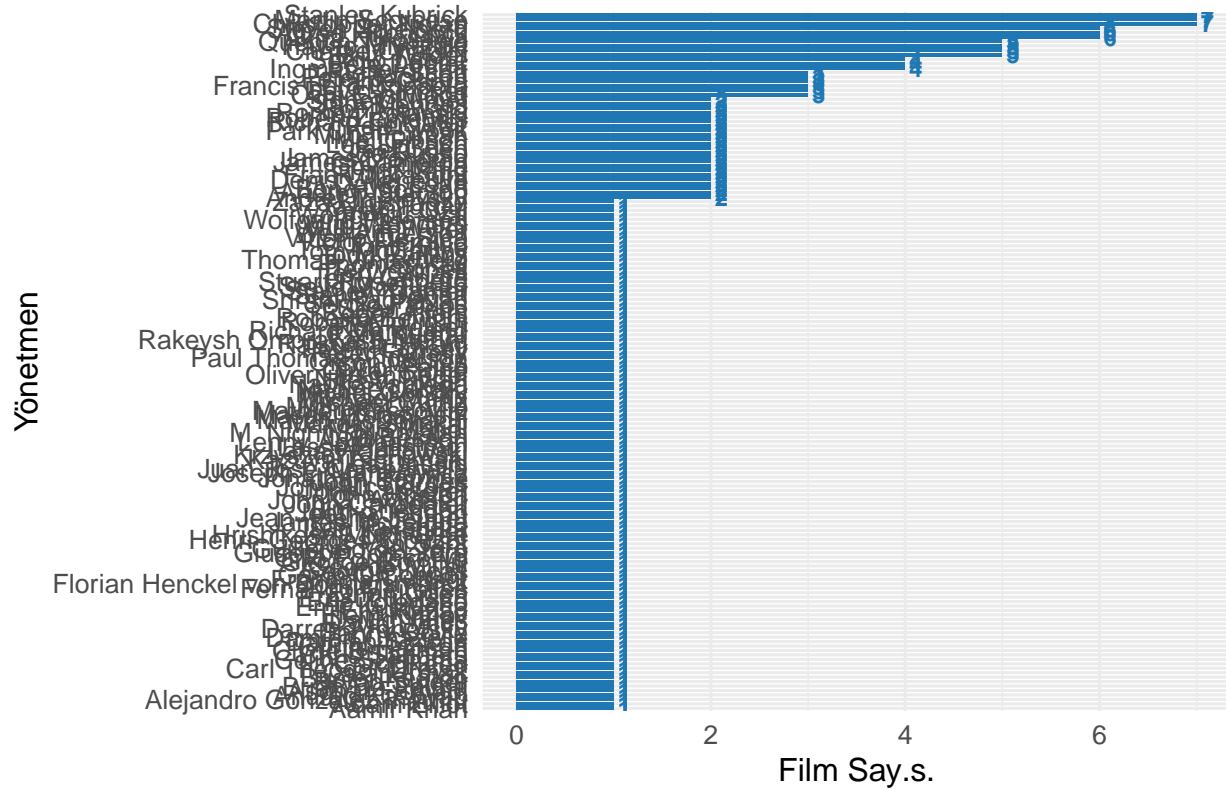
avg_rating_by_director <- data %>%
  group_by(Director) %>%
  summarise(
    Avg_Rating = mean(Rating, na.rm = TRUE),
    Film_Count = n()
  ) %>%
  filter(Film_Count >= 3)

director_counts %>%
  ggplot(aes(x = reorder(Director, n), y = n)) +
  geom_col(fill = "#1f78b4") +

```

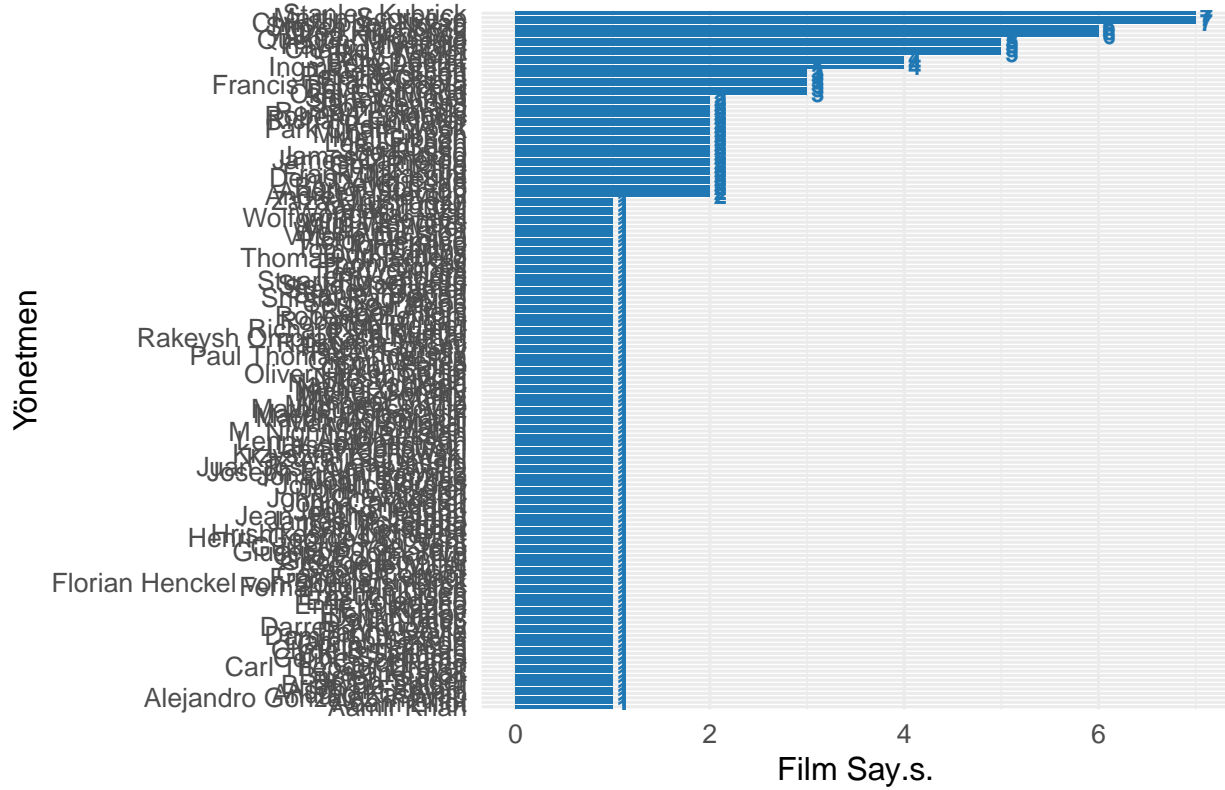
```
geom_text(aes(label = n), hjust = -0.3, size = 3, color = "#1f78b4", fontface = "bold") +
labs(
  title = "Top 250'de En Çok Temsil Edilen Yönetmenler",
  x = "Yönetmen",
  y = "Film Sayısı"
) +
coord_flip() +
theme_minimal(base_size = 12) +
theme(plot.title = element_text(face = "bold", size = 16))
```

Top 250'de En Çok Temsil Edilen Yönetmenler



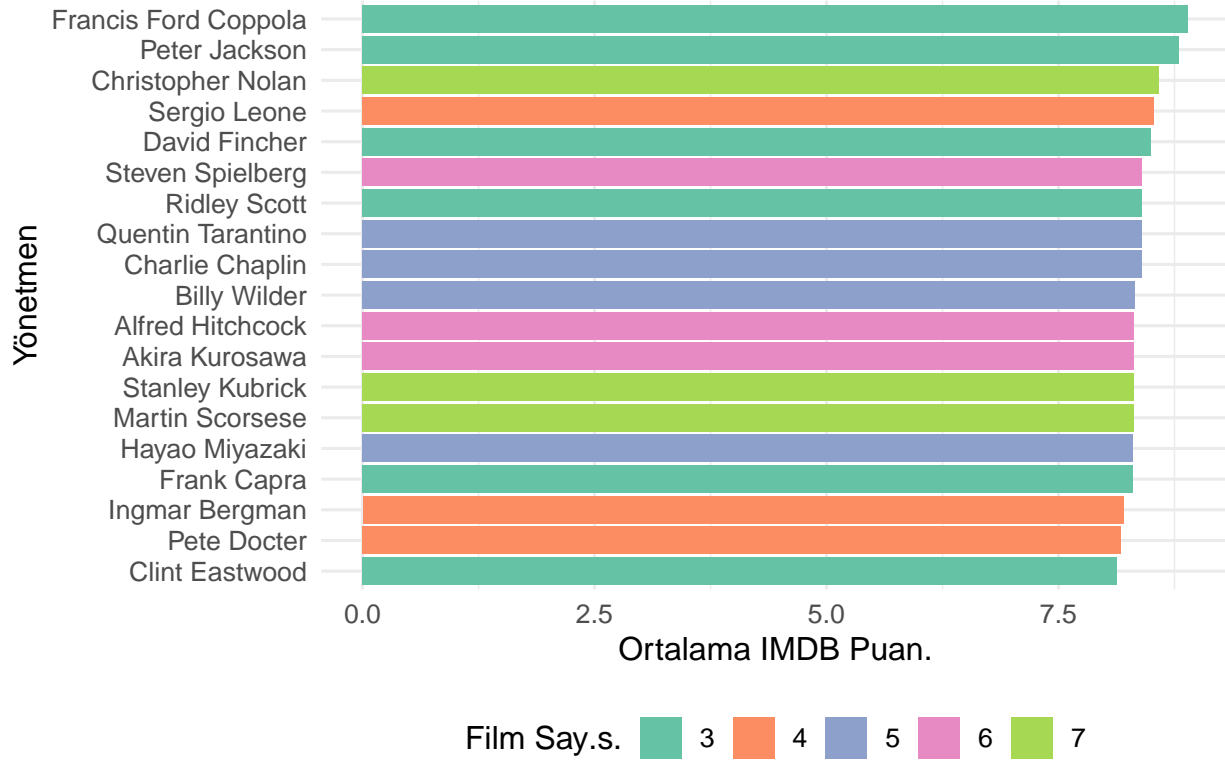
```
director_counts %>%
ggplot(aes(x = reorder(Director, n), y = n)) +
geom_col(fill = "#1f78b4") +
geom_text(aes(label = n), hjust = -0.3, size = 3, color = "#1f78b4", fontface = "bold") +
labs(
  title = "Top 250'de En Çok Temsil Edilen Yönetmenler",
  x = "Yönetmen",
  y = "Film Sayısı"
) +
coord_flip() +
theme_minimal(base_size = 12) +
theme(plot.title = element_text(face = "bold", size = 16))
```

Top 250'de En Çok Temsil Edilen Yönetmen



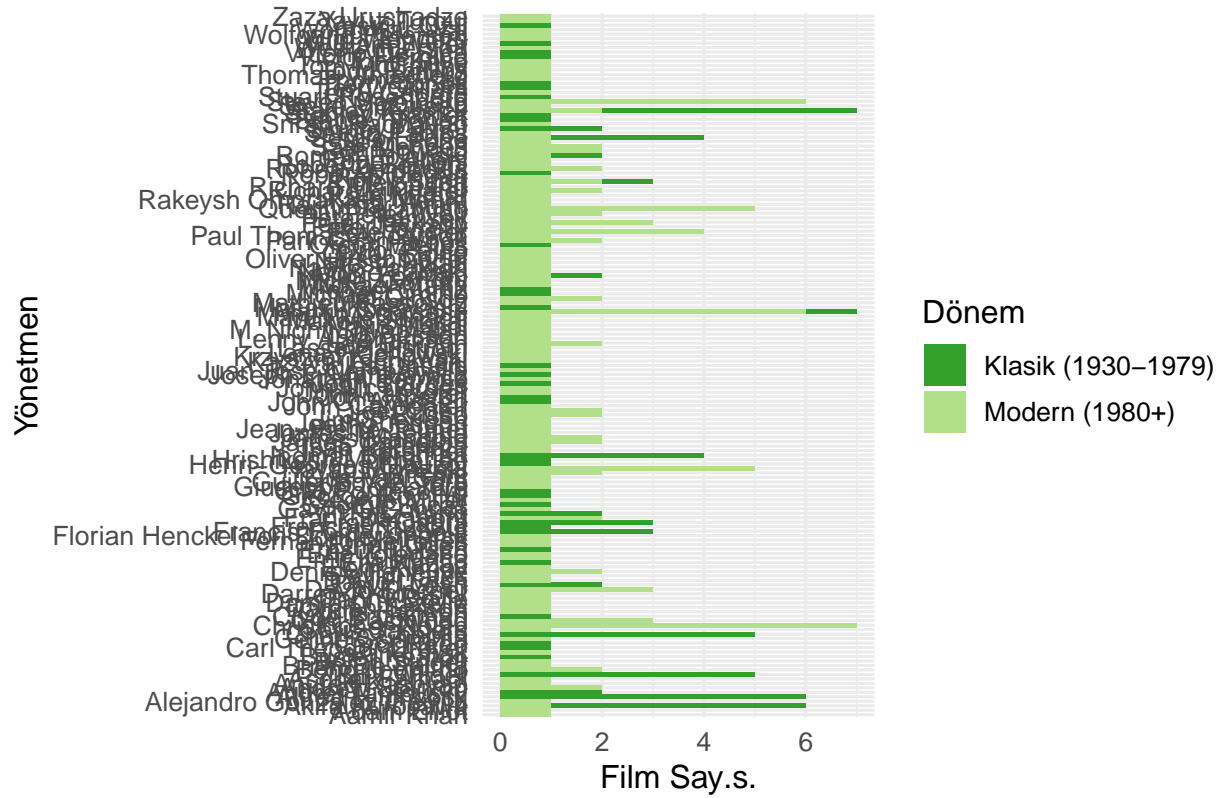
```
avg_rating_by_director %>%
  ggplot(aes(x = reorder(Director, Avg_Rating), y = Avg_Rating, fill = factor(Film_Count))) +
  geom_col(position = "dodge") +
  labs(
    title = "Yönetmenlere Göre Ortalama Puan ve Film Sayısı Karşılaştırması",
    x = "Yönetmen",
    y = "Ortalama IMDB Puanı",
    fill = "Film Sayısı"
  ) +
  scale_fill_brewer(palette = "Set2") +
  coord_flip() +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(face = "bold", size = 16))
```

Yönetmenlere Göre Ortalama Puan ve Film Sa



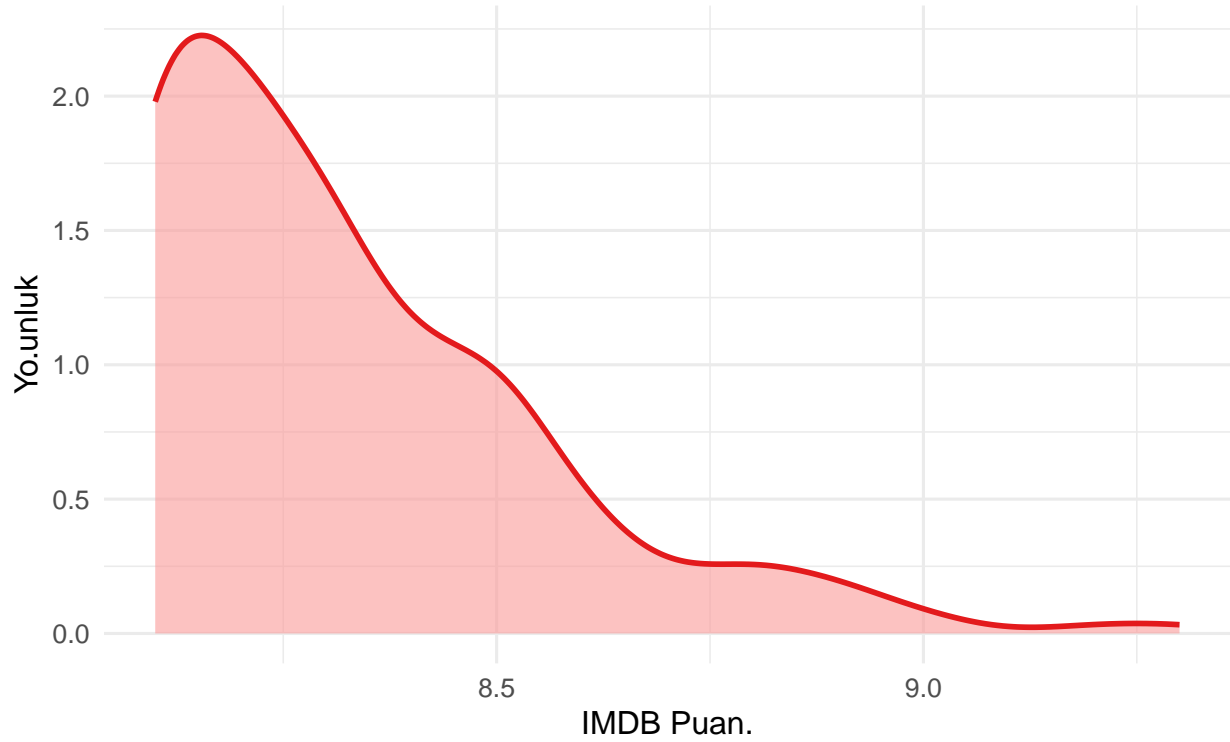
```
data_stacked %>%
  filter(Director %in% director_counts$Director) %>%
  ggplot(aes(x = Director, fill = Era)) +
  geom_bar(position = "stack") +
  labs(
    title = "Top Yönetmenlerin Filmlerinin Dönemlere Göre Dağılımı",
    x = "Yönetmen",
    y = "Film Sayısı",
    fill = "Dönem"
  ) +
  scale_fill_manual(values = c("Klasik (1930-1979)" = "#33a02c", "Modern (1980+)" = "#b2df8a")) +
  theme_minimal(base_size = 12) +
  coord_flip()
```

Top Yönetmenlerin Filmlerinin Dönemlere Göre













































```
data %>%
  ggplot(aes(x = Rating)) +
  geom_density(fill = "#fb9a99", alpha = 0.6, color = "#e31a1c", linewidth = 1) +
  geom_vline(aes(xintercept = mean(data$Rating)), linetype = "dashed", color = "black") +
  labs(
    title = "Top 250 Film Puanlarının Kernel Yoğunluk Dağılımı",
    subtitle = paste("Ortalama Puan:", round(mean(data$Rating), 2)),
    x = "IMDB Puanı",
    y = "Yoğunluk"
  ) +
  theme_minimal(base_size = 12)
```

Top 250 Film Puanlar.n.n Kernel Yo.unluk Da..l.m.
Ortalama Puan: NA



```
director_pie_data %>%
  ggplot(aes(x = "", y = percentage, fill = Director)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  geom_text(aes(y = label_y, label = paste0(round(percentage * 100), "%")), color = "black", size = 3) +
  labs(
    title = "Top Yönetmenlerin Liste İçindeki Oransal Payları",
    fill = "Yönetmen"
  ) +
  theme_void(base_size = 12) +
  theme(legend.position = "bottom", plot.title = element_text(face = "bold", size = 16))
```

1

	Gillo Pontecorvo		Lasse Hallström		Rajkumar Hirani
	Giuseppe Tornatore		Lee Unkrich		Rakeysh Omprakash
	Guillermo del Toro		Lenny Abrahamson		Ram Kumar
	Gus Van Sant		Luc Besson		Richard Linklater
	Guy Ritchie		M. Night Shyamalan		Richard Marquand
	Hayao Miyazaki		Majid Majidi		Ridley Scott
	Henri-Georges Clouzot		Makoto Shinkai		Rob Reiner
	Hrishikesh Mukherjee		Martin McDonagh		Robert Mulligan
	Ingmar Bergman		Martin Scorsese		Robert Zemeckis
	Irvin Kershner		Masaki Kobayashi		Roberto Benigni
	Isao Takahata		Mathieu Kassovitz		Roger Allers
	James Cameron		Mel Gibson		Roman Polanski
	James Mangold		Michael Cimino		Ron Howard
	James McTeigue		Michael Curtiz		Sam Mendes