

IMDB Top 250 Film Raporu

Özet

Bu projede, Python dili ve Selenium kütüphanesi kullanılarak IMDb'nin Top 250 filmlerine ait veriler (film isimleri, sıralama, IMDb puanı, tür, yorum sayısı, puanlanma sayısı, yönetmen, senarist, bütçe, hasılat, çıkış yılı, süre, ülke) web scraping yöntemiyle çekilmiştir. Çekilen veriler Excel'e aktarılıp, hatalı ve eksik veriler üzerinde gerekli ön işlemler uygulanarak temizlenmiş ve düzenlenmiştir. Sonrasında, elde edilen veriler Power BI kullanılarak görselleştirilmiştir. Projenin temel amacı, doğru ve temiz veriler elde ederek filmlerle ilgili anlamlı analizler yapmaktır.

Bu rapor, projenin teknik adımlarını ve veri hazırlama sürecini belgelerken, elde edilen verilerden çıkarılan analitik bulgular **"IMDB Top 250: Grafiklerle Analiz"** başlıklı çalışmada detaylandırılmıştır.

Giriş

- **Motivasyon ve Problem Tanımı:**

Bu projede amaç, Python kullanarak web scraping yöntemiyle veri çekmek, elde edilen verileri temizleyip düzenlemek ve görselleştirme araçlarıyla analiz edilebilir hâle getirmektir. Veri kaynağı olarak, hem popülerliği hem de gerçek veriye dayanması nedeniyle IMDb'nin Top 250 listesi seçilmiştir. Veri çekimi sonrasında karşılaşılan eksik ve tutarsız bilgiler, gerekli düzenlemelerle giderilerek analiz için uygun hale getirilmiştir.

- **Hedefler:**

- Doğru veri toplama tekniklerini öğrenmek ve uygulamak
- Web scraping, veri temizleme ve veri görselleştirme süreçlerini deneyimlemek
- Power BI aracılığıyla verileri görsel olarak analiz edebilmek
- Film verilerinin analizine dayanarak sektörle ilgili bazı temel gözlemler yapmak

Veri Seti ve Ön İşleme

- **Veri Kaynağı ve Temel Özellikler:**

Proje, IMDb'nin Top 250 filmler listesinden elde edilen veriler üzerine kuruludur. Toplanan değişkenler; film adı, sıralama, IMDb puanı, film türleri, kullanıcı yorum sayısı, puanlanma sayısı, yönetmen adı, senarist adı, bütçe, hasılat, çıkış yılı, film süresi ve ülke bilgisini içermektedir.

- **Ön İşleme Adımları:**

- **Veri Temizleme:**

- Hatalı ve eksik veriler belirlendi ve düzeltildi.

- IMDb'deki bütçe ve hasılat verileri, sıklıkla yanıltıcı ve hatalı olabilmektedir. Bu veriler, doğruluk açısından sorunlar taşıdığından, analiz ve görselleştirme aşamalarında kullanılmamış ve veri setinden çıkarılmıştır.
- **Veri Dönüşümü:** Gerekli format dönüşümleri yapılarak veri analiz için uygun hale getirildi.
- **Aykırı Değer Analizi:** Veri setinde mevcut olabilecek aykırı değerler incelenip, gerekirse düzeltmeler yapıldı.
- **Kullanılan Araç ve Kütüphaneler:**
 - **Python** (Web Scraping)
 - **Excel** (Veri işleme ve düzenleme)
 - **Power BI** (Veri görselleştirme)

Veri İşleme ve Analiz Yöntemleri

- **Kullanılan Analiz/Metodoloji Teknikleri:**

Projede öncelikle web scraping yöntemi kullanılarak IMDb verileri çekildi. Sonrasında, verilerin temizlenmesi, dönüştürülmesi ve düzenlenmesi işlemleri uygulandı. Elde edilen temiz veri seti, Power BI ile görselleştirildi.
- **Parametre ve Optimizasyon Yöntemleri:**
 - Selenium ile element seçimlerinin ve veri çekme parametrelerinin doğru belirlenmesi
 - Dinamik içeriklerin yönetimi için ek kontroller ve bekleme süreleri ayarlandı
- **Veri Kalitesi ve Analiz Yeterliliği:**

Veri bütünlüğü ve görselleştirme sonuçları üzerinden değerlendirme yapıldı. Projenin başarısı, verinin doğruluğu ve analiz için yeterliliği ile ölçüldü.

Sonuçlar ve Tartışma

- **Ana Bulgular:**
 - IMDb'nin Top 250 film verileri başarıyla çekildi ve Excel'e aktarıldı.
 - Hatalı ve eksik veriler, gerekli ön işlemlerle giderildi.
 - Power BI ile oluşturulan grafikler, verilerin trendlerini ve önemli noktalarını başarılı şekilde ortaya koydu.
- **Yorum ve Tartışma:**
 - **Karşılaşılan Zorluklar:** Dinamik web içerikleri nedeniyle veri çekme sırasında bazı aksaklıklar yaşandı; eksik verilerin tespiti ve düzeltilmesi zaman alıcı oldu.

- **Geliştirme Önerileri:**

- Web scraping sürecinin daha da otomatikleştirilmesi ve element seçimi süreçlerinde kısaltılmış yöntemlerin uygulanması
- Ek veri kaynaklarının entegre edilerek daha kapsamlı analiz yapılması

Sonuç ve Gelecek Çalışmalar

- **Projenin Genel Çıktısı:**

IMDB Top 250 filmlerinden elde edilen veriler başarılı bir şekilde çekildi, ön işlem den geçirildi ve Power BI aracılığıyla görselleştirildi. Proje, veri toplama ve görselleştirme aşamalarında önemli kazanımlar sağladı.

- **Ek Analiz Fikirleri:**

- Çekilen verilerin düzenli güncellenmesi için otomatik veri akışı oluşturulması
- Makine öğrenmesi modelleri ile film performans tahminleri yapılması

- **Profesyonel Alandaki Potansiyel Uygulamalar:**

Sinema sektöründeki temel trend analizlerine dair başlangıç düzeyinde veriler sunarak, gelecekteki daha kapsamlı çalışmalara zemin oluşturabilir.