Veri setindeki her sütun için yapılabilecek feature engineering önerileri:

1. \*\*Unnamed: 0\*\*:

- Bu sütun indeks gibi gözüküyor. Eğer veri setinin doğal bir indeksi varsa, bu sütun kaldırılabilir.

2. \*\*show\_id\*\*:

- Bu sütun benzersiz bir tanımlayıcıdır. Analizlerde veya modellemede genellikle kullanılmaz. Ancak, benzerlik analizi veya veri birleştirme işlemleri için saklanabilir.

3. \*\*type\*\*:

- Kategorik bir değişkendir. One-Hot Encoding veya Label Encoding uygulanabilir. Örneğin, 'Movie' ve 'TV Show' gibi sınıfları binary olarak kodlayabilirsiniz.

4. \*\*title\*\*:

- Bu metinsel veriyi doğal dil işleme (NLP) teknikleri ile işleyebilirsiniz. Örneğin, TF-IDF, Bag of Words, veya BERT gibi embedding yöntemleriyle anlam çıkartılabilir.

- Title uzunluğunu yeni bir özellik olarak ekleyebilirsiniz.

5. \*\*director\*\*:

- Eksik değerler mevcut, bu nedenle eksik değerler doldurulabilir veya "Unknown" olarak kodlanabilir.

- Yönetmen popülaritesini hesaba katmak için yönetmen başına toplam yapım sayısını bir özellik olarak ekleyebilirsiniz.

6. \*\*cast\*\*:

- Aynı şekilde eksik değerler "Unknown" olarak kodlanabilir.

- Oyuncuların popülaritesine göre yeni özellikler oluşturulabilir.

- Cast sayısını bir özellik olarak ekleyebilirsiniz.

7. \*\*country\*\*:

- Eksik değerler doldurulabilir.

- Bu sütunu One-Hot Encode yaparak ülkeleri ayrı özellikler olarak işleyebilirsiniz.

8. \*\*date\_added\*\*:

- Tarih formatına dönüştürülebilir. Yıl, ay ve gün bilgilerini ayrı sütunlar halinde ekleyebilirsiniz.

- Filmin platforma eklenme zamanına göre sezon bilgisi çıkartılabilir (örneğin, "Kış" veya "Yaz").

9. \*\*release\_year\*\*:

- Bu sayısal bir veridir ve doğrudan kullanılabilir.

- Zamanla ilgili analizler yapmak için, yılın on yıllık aralıklar halinde gruplandırılması veya çıkarılma yılından bugüne kadar geçen süreyi hesaplamak faydalı olabilir.

10. \*\*rating\*\*:

- Kategorik bir değişken olduğundan One-Hot Encoding veya Label Encoding yapılabilir.

- Belirli bir grupta olanlar için binary sınıflandırma yapılabilir (örneğin, 'G', 'PG', 'PG-13' gibi çocuk dostu olanlar ve diğerleri).

11. \*\*duration\*\*:

- Bu sütun hem dakika hem de sezon sayısı içeriyor gibi görünüyor. Sezon ve dakika değerlerini ayırarak her biri için ayrı sütunlar oluşturabilirsiniz.

- Süreyi sayısal bir forma çevirmek faydalı olacaktır (örneğin, "97 min" -> 97, "2 Seasons" -> 2).

12. \*\*listed\_in\*\*:

- Bu metin tabanlı veriyi birden fazla kategori içerdiği için, kategorileri ayırıp One-Hot Encoding uygulayabilirsiniz.

- Tür sayısını hesaplayarak yeni bir özellik eklenebilir.

13. \*\*description\*\*:

- Doğal dil işleme yöntemleri uygulanabilir. TF-IDF veya embedding teknikleriyle anlam çıkarılabilir.

- Metin uzunluğu gibi basit bir özellik de çıkarılabilir.