

دیتاست انتخاب شده مربوط به نتفلیکس میباشد. نتفلیکس یکی از محبوبترین پلتفرمهای پخش رسانه و ویدئوست که بیش از ۸۰۰۰ فیلم یا برنامه تلویزیونی در پلتفرم خود دارد. این مجموعه داده جدولی شامل فهرستی از تمام فیلمها و نمایشهای تلویزیونی موجود در نتفلیکس، همراه با جزئیاتی مانند بازیگران، کارگردانان، رتبهبندی، سال انتشار، مدت زمان و... است.

در ادامه ستونهای موجود در دیتاست توضیح داده شده است و در انتها موارد خواسته شده در هر فاز آورده شده است.

عنوان	شرح
show_id	شناسه منحصر به فرد برای هر فیلم / نمایش تلویزیونی
type	شناسه - یک فیلم یا نمایش تلویزیونی
title	عنوان فیلم / نمایش تلویزیونی
director	کارگردان فیلم
cast	بازیگران درگیر در فیلم / نمایش
country	کشوری که فیلم / نمایش در آن تولید شده است
date_added	تاریخ اضافه شدن آن به نتفلیکس
release_year	سال انتشار واقعی
rating	رتبه بندی تلویزیونی فیلم / نمایش
duration	مدت زمان کل - بر حسب دقیقه یا تعداد فصل
listed_in	دسته بندی فیلم / نمایش
description	توضیحات

فاز اول:

در فاز اول لازم است برای این دیتاست، مقایسه ی آماری ای بین تعداد فیلم هر کارگردان با متوسط (میانگین) تعداد فیلم کارگردانان انجام شود.

فاز دوم:

در فاز دوم لازم است برای این دیتاست، الگویی میان "director" و "cast" با ژانر آن ("listed_in") استخراج نمایید.

فاز سوم:

در فاز سوم لازم است برای این دیتاست، طبقه بندی با فرض label بودن "listed_in" به صورتی انجام شود که بتوان برای داده های تست از روی description ژانر فیلم (label) را پیش بینی کرد.