Assignment 1: Examine Movie Dataset

Deepti V. Khandagale

G # G01353007

AIT 580 DL4

1.   Identify the data type for each of the attribute/column, and explain your reason of selecting data type for each attribute/column.

Table 1

Data types of the attributes

Attributes	Data type	Reasons
director_name	Nominal	Director name column contains the name of the director of the movies. There is no order or size of data, it is a nominal data type
duration	Ratio	Duration column contains the length of the movie in minutes. As the data is comparable and measurable, it is ratio data type
actor_2_name	Nominal	Actor 2 name column contains the name of the second actor in the movie. There is no order or size of data, it is a nominal data type
genres	Nominal	Genres column contains the category of the movie. There is no order or size, it is a nominal data type
actor_1_name	Nominal	Actor 1 name column contains the name of the first actor in the movie. There is no order or size, it is a nominal data type
movie_title	Nominal	Movie tittle column contains the name of the movie. There is no order or size, it is a nominal data type
num_voted_users	Ratio	Number of voted users column contains the count of votes by the user which is measurable and continuous TAs the data is measurable and continuous, it is ratio data type
actor_3_name	Nominal	Actor 3 name column contains the name of the third actor in the movie. There is

Attributes	Data type	Reasons
		no order or size of data, it is a nominal data type
movie_imdb_link	Nominal	Movie IMDB link column contains the address where the assigned movie will open and is discrete for each movie, hence it is a nominal data type
num_user_for_reviews	Ratio	Number of users for reviews column contains the count of users who has reviewed the movie. As the data is measurable and continuous, it is a ratio data type
language	Nominal	Language column contains the language type used in the movie. There is no order or size of data, it is a nominal data type
country	Nominal	Country column contains the country where the movie is produced. There is no order or size of data, it is a nominal data type
title_year	Interval	Title year column contains the release date of the movie. As the data is continuous, it is a interval data type
imdb_score	Ordinal	IMDB score column contains the rating of the movie measured on a scale. As the data is ordered and discrete, it is ordinal data type

2. Give one sample value for each data type from the dataset

There are four data type found in the Movie dataset. The data types with the sample values are as follows:

Nominal data type

The Nominal data type in the movie dataset are the director name, actor 1 name, actor 2 name, actor 3 name, genres, movie title, movie IMDB link, language and country.

The sample value for each of the data type is as follows:

Director name: Gore Verbinski

Actor 2 name: Ruth Wilson

Genres: Action|Adventure|Western

Actor 1 name: Johnny Depp

Movie title: The Lone Ranger¬†

Actor 3 name: Tom Wilkinson

Movie IMDB link: http://www.imdb.com/title/tt1210819/?ref_=fn_tt_tt_1

Language: English

Country: USA

Ordinal data type

The Ordinal data type in the movie dataset is the IMDB score.

The same sample value for the ordinal data type is as follows:

Imdb score: 6.5

Interval data type

The Interval data type in the movie dataset is the title year.

The sample value for the interval data type is as follows:

Title year: 2013

Ratio data type

The Ratio data type in the movie dataset are the duration of the movie, number of voted

users and the number of users for reviews.

The sample value for the ratio data type is as follows:

Duration: 150

Number of voted users: 181792

Number of users for reviews: 711