9/30/2020
Lab 1

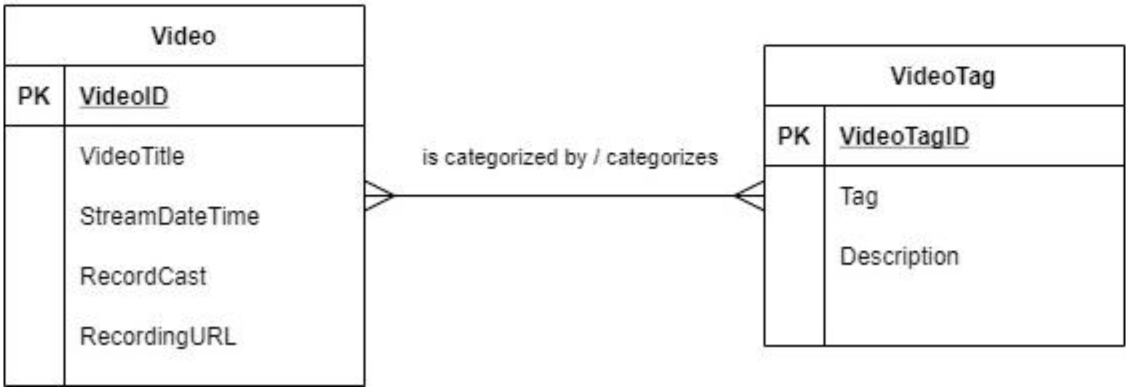
4 challenges to fitting this data into the relational data model

1. There are not usernames or user tags on two of the rows. As far as the dataset is concerned, there are two videos out there that belong to blank. Not having a unique identifier is a critical issue in this dataset because there is no way to tell who owns the video.

2. In the followers column, there is a mix of some cells that contain numbers and some text. From the looks of it, "7" is a string and is being treated the same way that "Prince Humperdink" is. I doubt that this is actually the case so this is confusing and needs to be clarified.

3. The values in the video duration column are presented in several different formats. This could be problematic. For example, what if I was an end user of this data and wanted to total up all of the video durations. There is no way to do this unless each of the cells have a consistent format.

4. Stream dates are not in the same format. It looks like some are in DD/MM/YY while others are in MM/DD/YY. Not to mention, one of the cells has multiple stream dates separated with a comma delimiter and two are blank. This needs to be cleaned up if analysis is ever to be performed successfully.


Video
PK
VideoID
VideoTitle
StreamDateTime
RecordCast
RecordingURL
is categorized by / categorizes
VideoTag
PK
VideoTagID
Tag
Description