|  |  |  |
| --- | --- | --- |
| **Name** | : | Sathish Kumar Rajendiran |
| **Current Date** | : | 01/14/2020 |
| **Lab Number** | : | 01 |

**Part 1: Lab 01 – VidCast Database**

* **Data Challenges Involved in building a Relational Data Model**
  1. Missing data values on several fields across rows
     + *User Name is missing in two rows*
     + *User Tags is missing in two rows*
     + *Followers is missing in two rows*
     + *Stream Date is missing in two rows*
     + *Video Duration is missing in a row*
     + *File Location is missing in two rows*
  2. No Constraint in place, thus allowing Null/empty values
  3. Difficult to identity Uniqueness/Primary Key candidates ex. VideoID/VideoTagID from the data sheet
  4. Inconsistent Format/Data Type
     + Ex. “*Followers*” Column has Numbers as well Name of the followers.
     + *“Stream Date”* has mixed format. Unsure dd-mm-yyyy or mm-dd-yyyy or mmm -dd-yyyy. This has to be consistent and be specific to one global time standard.
     + “*Duration*” – has to follow either duration as in minutes, or hh:mm: ss or any specific format
  5. “Charlie and the Biebs” user have three videos contained in single row; Not sure what is corresponding to what or what duration each has. May be the row has to be split into different rows. Cardinality is not defined properly. It is possible that 1 to many or many to many is possible in this scenario.
  6. General Recommendation is to have proper column names. i.e. without space or use “\_” instead.
  7. Typo or wrong spelling /data entry errors in “*User Tags*” column

**Part 2: Lab 01 – Basic ERD using Draw.IO**

A screenshot of a cell phone

Description automatically generated