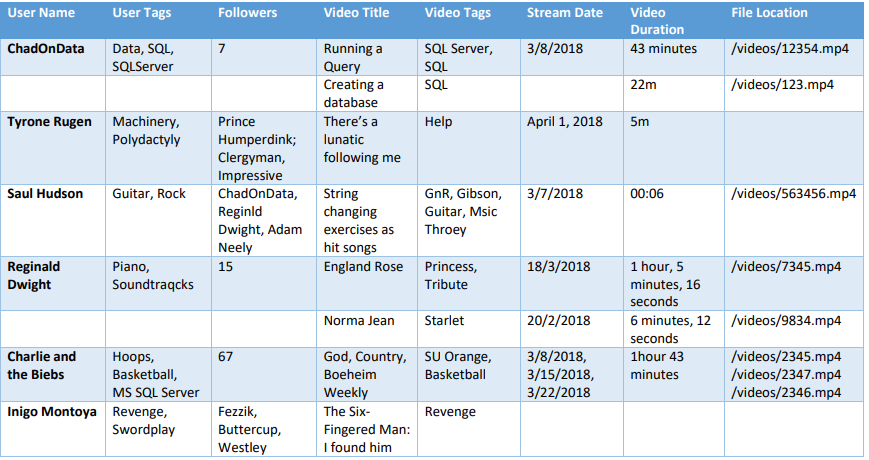
Nick Videtti

IST-659 Week 1 Lab Assignment

**VidCast Data Sheet:**



**Part 1:**

There are few issues with these data that may present a challenge while fitting them into a relational data model. One of the most important attributes of a set of data is for each column to have the same format and has only one value per row. It is also usually important to have no missing values in a column, unless a value truly does not exist for that specific row. We will assume that there are values that exist in every row for each column. Based on this, we will note the following issues with the VidCast data currently.

User Name - 2 missing values

User Tags - 2 missing values, multiple values per row may also present an issue in data analysis using this column

Followers - 2 missing values, differing formats between rows (some in list format, some in numerical, we would likely want to make them all numerical)

Video Title - no obvious issues

Video Tags - multiple values per row may also present an issue in data analysis using this column

Stream Date - 2 missing values, differing formats between rows, multiple values per row may also present an issue in data analysis using this column

Video Duration - 1 missing value, differing formats between rows

File Location - 2 missing values, multiple values per row may also present an issue in data analysis using this column

**Part 2:**

