Nick Videtti
IST-659 Week 1 Lab Assignment

VidCast Data Sheet:

User Name	User Tags	Followers	Video Title	Video Tags	Stream Date	Video Duration	File Location
ChadOnData	Data, SQL, SQLServer	7	Running a Query	SQL Server, SQL	3/8/2018	43 minutes	/videos/12354.mp4
			Creating a database	SQL		22m	/videos/123.mp4
Tyrone Rugen	Machinery, Polydactyly	Prince Humperdink; Clergyman, Impressive	There's a lunatic following me	Help	April 1, 2018	5m	
Saul Hudson	Guitar, Rock	ChadOnData, Reginld Dwight, Adam Neely	String changing exercises as hit songs	GnR, Gibson, Guitar, Msic Throey	3/7/2018	00:06	/videos/563456.mp4
Reginald Dwight	Piano, Soundtraqcks	15	England Rose	Princess, Tribute	18/3/2018	1 hour, 5 minutes, 16 seconds	/videos/7345.mp4
			Norma Jean	Starlet	20/2/2018	6 minutes, 12 seconds	/videos/9834.mp4
Charlie and the Biebs	Hoops, Basketball, MS SQL Server	67	God, Country, Boeheim Weekly	SU Orange, Basketball	3/8/2018, 3/15/2018, 3/22/2018	1hour 43 minutes	/videos/2345.mp4 /videos/2347.mp4 /videos/2346.mp4
Inigo Montoya	Revenge, Swordplay	Fezzik, Buttercup, Westley	The Six-Fingered Man: I found him	Revenge			

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: