Joseph Scherreik

INFO 365

WA10

# **Requirements Analysis**

Application Description/Plan

This data will be used to make decisions about student loans. The best way to analyze all of this data would be to follow the basic characteristics of quality data. The data needs to be unique, accurate, consistent, timely, concurrent, conform, and referentially integrant. Since these are our bounds, we need to go into detail about making these tables much more consistent, concurrent, and have better referential integrity.

Another good thing to check for is repetitive data. Some of the data in the large tables seem to be overlapping a bit. There might a be a good way to check for things that might be similar to each other and removed duplicates of similar information. This will make the data more unique, accurate, consistent, and timely.

Business Rules

* Colleges can have student dept from all different sources
* Different groups/regions/etc have different levels of debt

# **Potential Issues with Tables**

Objective

Fix tables relating to student.

Fix tables relating to aid/repayment.

Issues with student

Once this is all figured out there seems to be a lot of issues with how the student data is represented. It does not follow any of our characteristics of quality data and needs to be reworked. Most of the data seems unreadable or the relations are very complicated to understand to the average user. Lots of the columns are hardly populated with any information.

Issues with aid/repayment

The information in these tables is somehow related to one another. This needs to be properly determined.

Possible Solutions

The best possible solution would be to go through the original data to get a feel of how it exactly is laid out. Lots of these columns make no sense to the average user. It’s extremely difficult to even understand what they represent. They need different names and most likely a whole new schema to properly represent the data. The tables that need the most attention are “aid” and “repayment”. The data required is spread in these tables surrounded by data that may or may not be important at all. Clean up anything that is only used by a few entries and slow work until the data starts to make more logical sense. If more tables are needed create more tables and relations in order for the data to be easier to read.

There seems to be some tables inside the database like, “menonly”, ”womenonly”, ”hsi”, ”hbcu”, etc, that only contain a value of yes or no. This can easily be changed to a Boolean in the schema for the tables that contain a column that references the table. Having separate tables for columns that have many values is acceptable for example “relaffil” and “region”. In these there are multiple possible values represented by an ID that can be referenced back to these tables. When you have a boolean value, you should just add one to the table directly. This will help reference integrity and conformity.