**Notes on grading CSCI 5333.4 HW #5**

**Problems:**

* Country name declared not long enough to hold 52 characters.
* Data type size longer than necessary: e.g FLOAT(50) for measured values, VARCHAR(20) for gender.
* Float values were converted into integer values.
* Use Excel to MS SQL Server upload facility and thus not storing results in MySQL. Example: “I have used Microsoft SQL Server Management Studio IDE.”
* I didn’t find any candidate keys in the given all five data sets.
* Sex is a multivalued –attribute.
* Created a surrogate key. Although it is not harmful, there is no need to do so.
* ENGINE = InnoDB is not needed at the end of the table as InnoDB is the default.
* Consider 0 as the value 0 and not null.
* There is an intentional replacement of 0 with 999999 which is incorrect data wise.
* Listing only the primary key, not all candidate keys.
* Mention using a surrogate key in the previous question but do not use one here. Mismatch in code and documentation leads to confusion.
* creating a new dataset having Referencecode, Referencearea as attributes: Good.

**Some examples to be discussed in the class:**

“Some consists observation value=0, in database we will consider it as ‘0’ but not NULL.”

“There are no candidate keys since it’s difficult to keep, due to lack of unique data in it.”

“In this dataset the observation values for the country HolySee are zero. I thought it as correct information and didn’t remove these records because there may be chance of zero enrollments during that period.”

“In the year 2003 the observation value for the countries Argentina and Brazil are in decimal format I truncated it as number because the units of measurement is number”

“Primary Education:

Primary\_Education (PrimaryId, Reference\_area, Time\_period, Sex, Observation\_value).

Sex(PrimaryId, Male, Female, Allgender).

Primary Key: PrimaryId”

CREATE TABLE `c533304fa11\_\_\_\_\_`.`second` (

`Reference` VARCHAR( 30 ) NOT NULL ,

`Time\_period` INT( 30 ) NOT NULL ,

`Sex` VARCHAR( 30 ) NOT NULL ,

`Observation\_value` FLOAT( 30 ) NOT NULL

) ENGINE = INNODB;

“1. Primary Education

CREATE TABLE `c533304fa11\_\_\_\_\_\_\_\_`.`Primary Education` (

`Reference\_Area` TEXT NOT NULL ,

`Time\_Period` INT( 20 ) NOT NULL ,

`Sex` TEXT NOT NULL ,

`Observation\_value` INT( 20 ) NOT NULL

) ENGINE = INNODB;”

Any record with null values should be eliminated from the table structure.

`Observation Value` varchar(11) DEFAULT NULL

CREATE TABLE [dbo].[PEE](

[Reference\_Area] [varchar](150) NOT NULL,

[Time\_Period] [varchar](150) NOT NULL,

[Sex] [varchar](150) NOT NULL,

[Age\_group] [varchar](150) NOT NULL,

[Units\_of\_measurement] [varchar](150) NOT NULL,

[Observation\_Value] [varchar](150) NOT NULL

) ON [PRIMARY]

GO

**Values are inserted as:**

INSERT INTO [GID\_Raw].[dbo].[PEE]

([Reference\_Area]

,[Time\_Period]

,[Sex]

,[Age\_group]

,[Units\_of\_measurement]

,[Observation\_Value])

VALUES

(<Reference\_Area, varchar(150),>

,<Time\_Period, varchar(150),>

,<Sex, varchar(150),>

,<Age\_group, varchar(150),>

,<Units\_of\_measurement, varchar(150),>

,<Observation\_Value, varchar(150),>)

GO

CREATE TABLE [dbo].[GID](

[GID] [bigint] IDENTITY(1,1) NOT FOR REPLICATION NOT NULL,

[CountryID] [bigint] NOT NULL,

[TimePeriod] [varchar](150) NOT NULL,

[Sex] [varchar](150) NOT NULL,

[Units] [varchar](150) NULL,

[TypeID] [bigint] NOT NULL,

[Value] [varchar](150) NULL,

CONSTRAINT [PK\_GID\_CountryID] PRIMARY KEY CLUSTERED

(

[GID] ASC

)WITH (PAD\_INDEX = OFF, STATISTICS\_NORECOMPUTE = OFF, IGNORE\_DUP\_KEY = OFF, ALLOW\_ROW\_LOCKS = ON, ALLOW\_PAGE\_LOCKS = ON) ON [PRIMARY]

) ON [PRIMARY]

CREATE TABLE Primaryeducation (

Referencecode VARCHAR 3,

Timeperiod INTEGER 4,

Sex VARCHAR 11,

Observationvalue INTEGER 16,

FOREIGN KEY (Referencecode) REFERENCES Country(Referencecode)

PRIMARY KEY (Referencecode,Timeperiod,Sex));

“I removed the records containing the Percent value in Tertiary Education Enrolment dataset, because we can retrieve these values from the other records.”