**Homework Assignment #2 - Math 290.2**

**Group Members**:

* James Mancuso
* Christian G.
* Michael Velez

1. In order to populate the data for my partners and I, we first needed to find out what our id numbers were and than we were able to fill out our information regarding question #1.

/\*Create the homework table with a primary key constraint on the id field called pk\_homework\*/

create table homework (

"id" bigint,

"homework\_name" varchar,

"posted\_date" timestamp,

"due\_date" timestamp,

"homework\_duration\_minutes" bigint,

constraint "pk\_homework" primary key ("id")

);

/\*Creating the homework\_submission table with

\* - a primary key constraint on the id field called pk\_homework\_submission

\* - a foreign key constraint on the homework\_id field called fk\_homework\_id referencing the id field in the homework table

\* \*/

create table homework\_submission(

"id" bigint,

student\_id bigint,

homework\_id bigint,

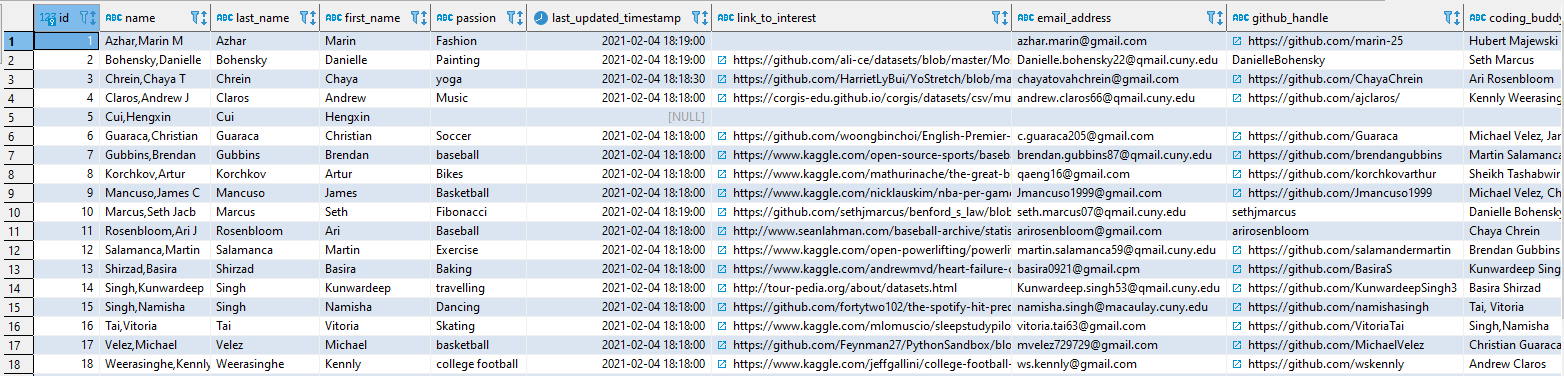
primary key ("id"),

constraint "fk\_homework" foreign key (homework\_id) references homework(id)

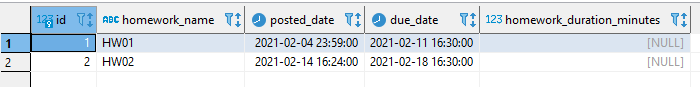
constraint “”fk\_id” foreign key (student\_id) references class\_roster(id)

);

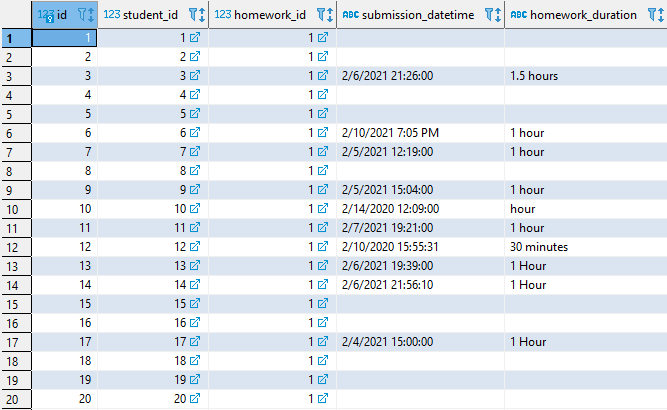
**Class Roster Table:**



**Homework:**

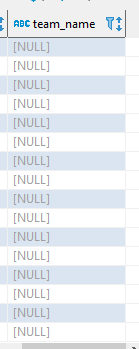


**Homework Submission:**



1. **Current size:** 8.05 MB 8,445,415 bytes 16,395

1. The solution we decided on was to create a column to represent the teams. We created another column because we had to consider a few factors: size, general complexity, and query complexity. Having a column specifically for the teams reduces the amount of data that needs to be stored. It also makes it easier to query the data and to understand the data.



1. .

create table YELLOW\_TAXI\_TRIP\_DATA

(vendorID smallint,

tpep\_pickup\_datetime timestamp,

tpep\_dropoff\_datetime timestamp,

passenger\_count smallint,

trip\_distance integer,

ratecodeID integer,

store\_and\_fwd\_flag char(1),

PULocationID integer,

DOLocationID integer,

payment\_type smallint,

fare\_amount decimal(10,2),

extra decimal(10,2),

mta\_tax decimal(10,2),

tip\_amount decimal(10,2),

tolls\_amount decimal(10,2),

improvement\_surcharge decimal(10,2),

total\_amount decimal(10,2)

);

copy yellow\_taxi\_trip\_data(

vendorid,

tpep\_pickup\_datetime,

tpep\_dropoff\_datetime,

passenger\_count,

trip\_distance,

ratecodeid,

store\_and\_fwd\_flag,

pulocationid,

dolocationid,

payment\_type,

fare\_amount,

extra,

mta\_tax,

tip\_amount,

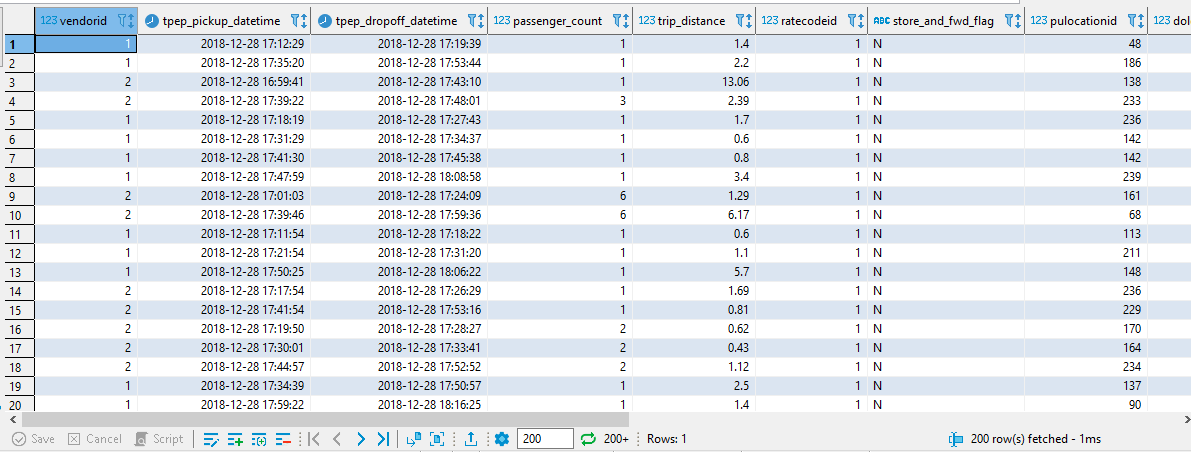
tolls\_amount,

improvement\_surcharge,

total\_amount)

from program 'cmd /c "type D:\Downloads\2018\_Yellow\_Taxi\_Trip\_Data.csv"' delimiter ',' csv header;

**Screenshot of data:**



1. The CSV size is 9.71 GB while the base file was 12.71 GB. The base file is bigger and the reason we believe this is due to the datatypes of the columns. Since some of the columns’ data types require a certain amount of bytes which accounts for more space. The data stored in the csv may be stored as one data type, making it memory efficient.
2. 1 ZB = 256,800,000

100,000 TB hd = 25,680,000

10,000 TB hd = 2,568,000

1000 TB hd = 256,800

100 TB hard drive = 25,680

10 TB hard drive = $2,568

1 TB = 256.8

**Visualization of Conversion:**

