/\* ----------------------------------------------------------------------- \*/

/\*DATE: December 20, 2015\*/

/\*PROGRAM: sample\_load\_code.sql\*/

/\*AUTHOR: Tate Garibyan\*/

/\*DESCRIPTION: Basic sample code to upload raw data in the form of csv into local database. Step 1 creates the

structure (shell) for the table in which raw data will be uploaded. Step 2 imports the data into the tables.

Step 3 creates the master dataset table that will be used for analysis. Step 4 exports the master data set

to an Excel file.\*/

/\* ----------------------------------------------------------------------- \*/

/\*The following code is designed for your local PostgreSQL server (version 9.4 or 9.5 Beta as of this writing), in the FinalProject database we created together in class. \*/

/\*STEP 1:\*/

/\*DROP TABLE IF EXISTS is a cautionary step to clear out any tables with these names, Dataset\_1 and Dataset\_2. You can use whatever

name you want that best identifies each dataset, however, stay consistent with your naming conventions\*/

DROP TABLE IF EXISTS Dataset\_1;

DROP TABLE IF EXISTS Dataset\_2;

/\*CREATE TABLE code creates a shell for the data that will be uploaded. The column names should match what is in the raw data file (csv), or at the very least be in the order you intend to import.

Data types must be compatible with the raw data. Refer to class resources and internet resources for full range of data types and proper

syntax.

NOTE: We are ultimately joining two or more datasets, so this needs to be done for each dataset that is uploaded.\*/

CREATE TABLE Dataset\_1

(tripduration INT --syntax is column\_name DATATYPE

,starttime TIMESTAMP

,stoptime TIMESTAMP

,start\_station\_id varchar

,start\_station\_name varchar

,start\_station\_latitude FLOAT

,start\_station\_longitude FLOAT

,end\_station\_id varchar

,end\_station\_name varchar

,end\_station\_latitude FLOAT

,end\_station\_longitude FLOAT

,bikeid varchar

,usertype varchar

,birth\_year varchar

,gender varchar);

CREATE TABLE Dataset\_2

(tripduration INT

,starttime TIMESTAMP

,stoptime TIMESTAMP

,start\_station\_id varchar

,start\_station\_name varchar

,start\_station\_latitude FLOAT

,start\_station\_longitude FLOAT

,end\_station\_id varchar

,end\_station\_name varchar

,end\_station\_latitude FLOAT

,end\_station\_longitude FLOAT

,bikeid varchar

,usertype varchar

,birth\_year varchar

,gender varchar);

/\*STEP 2: Imports the csv and fills in the table shell with the data. You need to provide the entire

file path to where your csv is saved so pgAdmin knows where to pull the data from to fill your your table. If you run

into error messages at this point, read the messages carefully. This may be a result of improper read/write access to the file,

or incompatible data types\*/

/\*WINDOWS USERS ONLY: Import data from CSV file \*/

copy Dataset\_1 from 'C:\Users\[...].csv' with delimiter ',' csv header;

copy Dataset\_2 from 'C:\Users\[...].csv' with delimiter ',' csv header;

/\*MAC USERS ONLY: Import data from CSV file\*/

copy Dataset\_1 from '/Users/Tate/[...].csv' with delimiter ',' csv header;

copy Dataset\_2 from '/Users/Tate/[...].csv' with delimiter ',' csv header;

/\*Run some basic select all statements to make sure your data loaded correctly\*/

SELECT \* FROM Dataset\_1;

SELECT \* FROM Dataset\_2;

/\*STEP 3: Create the master data set. The join is embedded in the CREATE TABLE code. Embedding it in CREATE TABLE is optional. You can also just export the results of any joins. Creating the permanent table just helps you preserve the master data set and makes it easier to query.\*/

DROP TABLE IF EXISTS Dataset\_Master;

CREATE TABLE Dataset\_Master as

(

--join query goes here

);

/\*Run a select all on your master dataset to make sure your table looks good and everything turned out as expected.\*/

SELECT \* FROM Dataset\_Master;

/\*STEP 4: Export the master data set to a csv. Use code below or use Export in drop-down\*/

/\*WINDOWS USERS ONLY: Export table to excel file\*/

COPY Dataset\_Master TO 'C:\Users\[...].csv' delimiter ',' csv header;

/\*MAC USERS ONLY: Export table to excel file\*/

COPY Dataset\_Master TO '/Users/Tate/[...].csv' delimiter ',' csv header;

/\* \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* \*/

/\*Sample code if you want to alter table columns\*/

/\*

ALTER TABLE CITIBIKE\_0414

ADD TABLE\_MONTH VARCHAR;

ALTER TABLE CITIBIKE\_0414

ADD TABLE\_YEAR VARCHAR;

\*/

/\*Sample code if you want to update/add values in a column\*/

/\*

UPDATE CITIBIKE\_0414 SET TABLE\_MONTH = 'APRIL';

UPDATE CITIBIKE\_0414 SET TABLE\_YEAR = '2014' ;

\*/