**Homework Assignment #3 - Math 290.2**

**Group Members**:

* James Mancuso
* Christian G.
* Michael Velez

1. Individual write-ups uploaded to google sheets.

create table qcmath290.public."2018\_Yellow\_Taxi\_Trip\_Data\_gz"(

"vendorID" bigint,

"tpep\_pickup\_datetime" varchar,

"tpep\_dropoff\_datetime" varchar,

"passenger\_count" bigint,

"trip\_distance" float8,

"PULocationID" bigint,

"DOLocationID" bigint,

"ratecodeID" bigint,

"store\_and\_fwd\_flag" varchar,

"payment\_type" bigint,

"fare\_amount" float8,

"extra" float8,

"mta\_tax" float8,

"improvement\_surcharge" float8,

"tip\_amount" float8,

"tolls\_amount" float8,

"total\_amount" float8);

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,

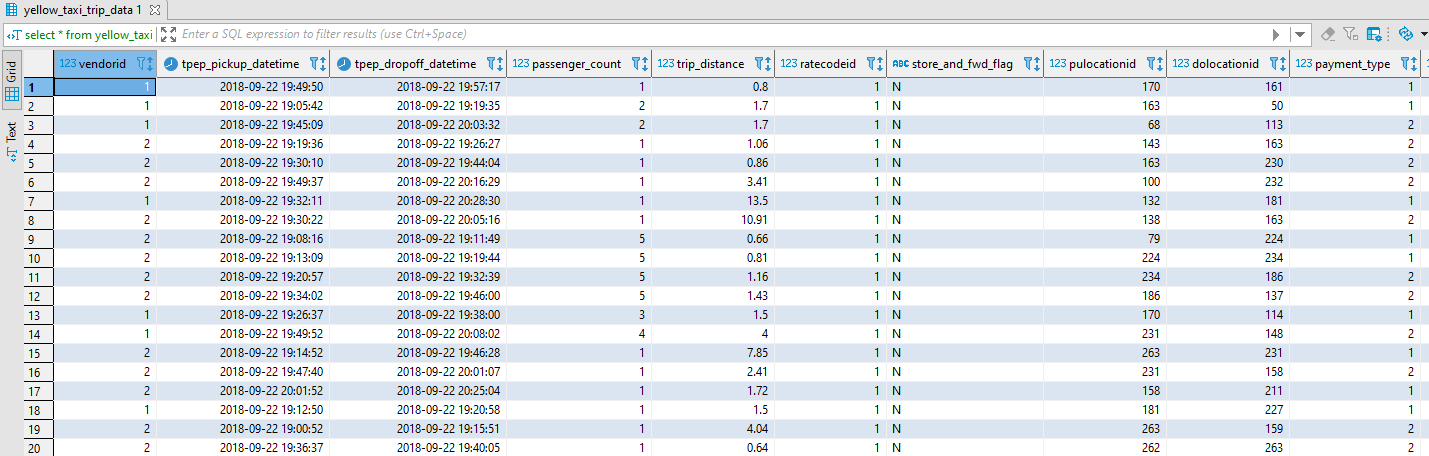
tip\_amount,

tolls\_amount,

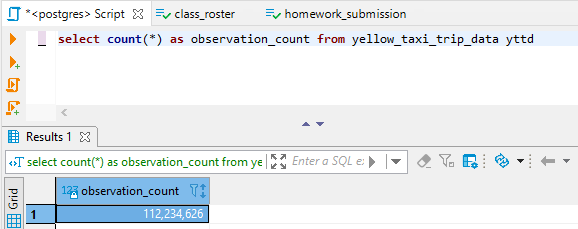
improvement\_surcharge,

total\_amount)

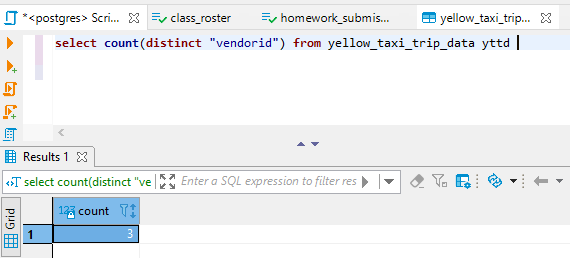
from program 'cmd /c "type D:\Downloads\2018\_Yellow\_Taxi\_Trip\_Data.csv"' delimiter ',' csv header;



select count(\*) as observation\_count from yellow\_taxi\_trip\_data yttd



select count(distinct "vendorid") from yellow\_taxi\_trip\_data yttd



* This would **not** be a good candidate for a primary key because there are 3 unique values within the attribute of a dataset that contains over 112 million tuples.