Peter Antonaros
Lamae Maharaj
Math 290 Homework 2

1) Exercise (Complete)

- Populated tables with required information
- The main detail of importance was that I referenced multiple tables without really thinking about it because that's what made sense. This is why we need to explicitly define foreign keys and their connections, so the computer can replicate this human combined steps.

2) Exercise (Complete)

- Downloaded the newly populated google sheets as csv files.
- Created the required relations
- homework table script was fine
- homework_submission table script needed the additional foreign key reference to student_id from the class_roster table

```
create table qcmath290.public.class_roster (
 "id" bigint
,"name" varchar
,"last_name" varchar
,"first_name" varchar
,"passion" varchar
,"link_to_interest" varchar
,"email_address" varchar
,"github_handle" varchar
,"goodreads_link" varchar
,"operating_system" varchar
,"coding_buddy_name" varchar
,"group_id" integer
,constraint "id" primary key ("id")
);
```

```
create table homework (
"id" bigint,
"homework_name" varchar,
"posted_date" timestamp,
"due_date" timestamp,
"homework_duration_minutes" bigint,
constraint "pk_homework" primary key ("id")
);
```

Peter Antonaros
Lamae Maharaj
Math 290 Homework 2

```
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_student_id" foreign key (student_id) references class_roster(id)
);
```

	id	name	last_name	first_name	passion	link_to_interest	en
1	1	Antonaros,Peter	Antonaros	Peter	Cancer Diagnostics	https://www.kaggle.com/uciml/breast-cancer-wis	peter.
2	2	Eltabakh,Amir	Eltabakh	Amir	F1	F1 Drivers Data	amira
3	3	Farber,Shoshana	Farber	Shoshana	Shopping	Mall Customer Data	shosh
4	4	Garg,Sachin	Garg	Sachin			
5	5	Hossain,Faria	Hossain	Faria			fariah
6	6	Khan,Samin	Khan	Samin			samin
7	7	Khatun,Miss P	Khatun	Miss			prityk
8	8	Lutz,Sasha	Lutz	Sasha			
9	9	Maharaj,Lamae A	Maharaj	Lamae	Crypto / Data Science	datasciencecrypto	lamae
10	10	Steele,Benjamin Matthew	Steele	Benjamin		discord.js	bentz
11	11	Weerasinghe,Kennly	Weerasinghe	Kennly	Plant Genetics and Physiology		ws.ke

	id	homework_name	posted_date	due_date	homework_duration_minutes
1	1	hw_01	2022-02-04 00:00:00.000	2022-02-17 00:00:00.000	[NULL]
2	2	hw_02	2022-02-17 00:00:00.000	2022-02-24 00:00:00.000	[NULL]

	id	student_id	homework_id	submission_datetime	homework_duration_minutes
1	1	2	1	02/12/22	120
2	2	11	1	02/16/22	120
3	3	2	2		[NULL]
4	4	11	2		[NULL]
5	5	4	1	02/17/22	180
6	6	10	2	2/22/22	[NULL]
7	7	5	1	02/17/22	140
8	8	1	1	02/17/22	45
9	9	1	2	02/22/22	100

3) Exercise (Complete)

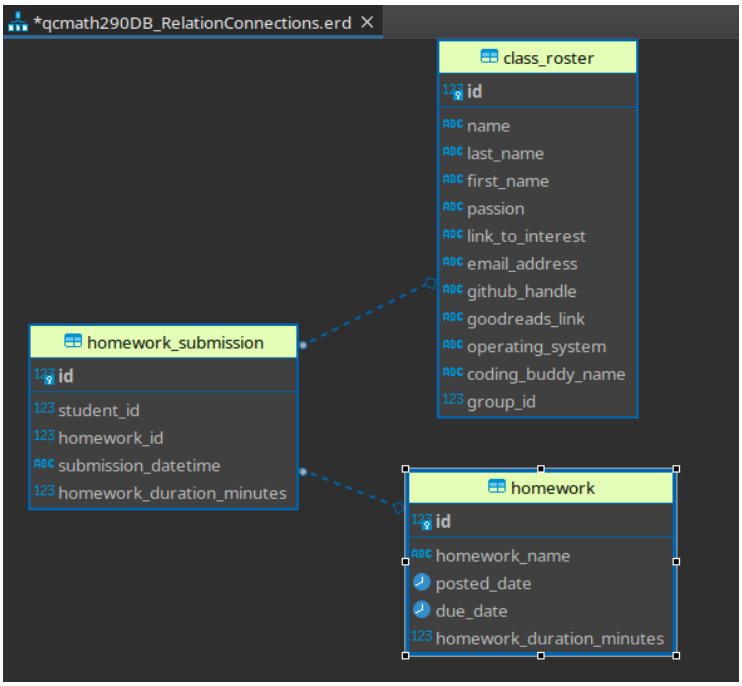
- At this point our qcmath290 database is 8721 kilobytes, prior to importing the taxi dataset

List of databases

Name	Owner	Encoding	Collate	Ctype	Access privileges	Size	Tablespace	Description
postgres	postgres	UTF8	en_US.UTF-8	en_US.UTF-8		8569 kB	pg_default	default administrative connection database
qcmath290	postgres	UTF8	en_US.UTF-8	en_US.UTF-8		8721 kB	pg_default	

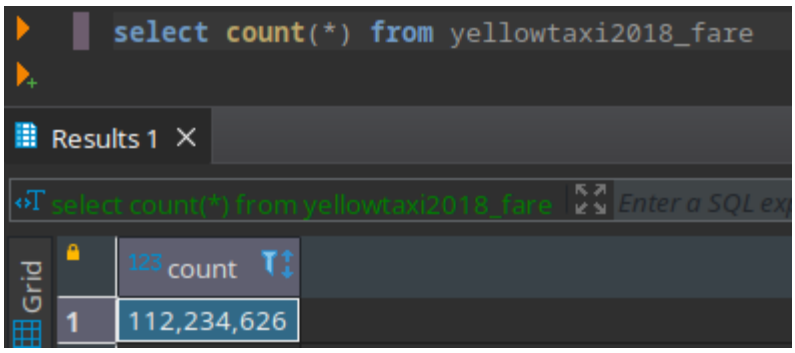
Peter Antonaros
Lamae Maharaj
Math 290 Homework 2

And here is what the current status of the database looks like in the form of an ER diagram



4) Exercise (Complete)

- Created the schema for 2018 taxi information data set while, the data set itself downloaded
- Copied the CSV to the table and ran a count statement to check number of rows



5) Exercise (Complete)

- Now the size of out database is ~18gb
- The row space overhead for SQL relations are larger than traditional CSV files. This follows suit with the classic Computer Science sacrifice, higher compute performance relies on greater storage/memory usage, and lower compute performance allows for less storage/memory usage (Generally a consistent rule to keep in mind) We are getting higher performance with the SQL relation (18gb) rather than the 9.7gb CSV file. There is also the factor of what data types are used for the columns which can have an impact on overall size.

Peter Antonaros
Lamae Maharaj
Math 290 Homework 2

Name	Owner	Encoding	Collate	Ctype	Access privileges	Size	Tablespace	Description
postgres	postgres	UTF8	en_US.UTF-8	en_US.UTF-8		8569 kB	pg_default	default administrative connection database
qcmath290	postgres	UTF8	en_US.UTF-8	en_US.UTF-8		18 GB	pg_default	

6) Exercise (Complete)

- Lets make 59 ZB of data a bit easier to visualize/understand

A zettabyte is equal to 10^{21} bytes. That means if every byte was equal to 1 meter, then we could travel the distance of the entire milky way, plus some!

For some visualization, look at the following image!