DATA 514 Week 2 Worksheet

These exercises represent the in-lecture and section questions for week 2. Reasonable answers will receive credit - please answer as fully and correctly as possible. For SQL question use correct SQL, for query questions provide complete answers when possible.

During Lecture

Answer this set of questions when prompted during lecture. You may discuss with your classmates.

1. What does the pseudo code ?

```
sum = 0
foreach row in Payroll:
  if row.Job == 'TA':
    sum = sum + row.Salary
output(sum)
```

2. What is returned by each of the following count queries:

```
SELECT COUNT(*) FROM Payroll;
SELECT COUNT(job) FROM Payroll;
SELECT COUNT(salary) FROM Payroll;
```

3. What is returned by the follow query for Francis? and for Leslie?

```
  SELECT Name, AVG(salary)
    FROM Payroll
GROUP BY Name;
```

4. What is the output of the following query?

```
  SELECT Job
    FROM Payroll
GROUP BY Job
  HAVING MAX(Salary) > 60000;
```

5. What is the output of each of the following queries?

```
  SELECT Job, MAX(Salary)
    FROM Payroll
GROUP BY Job
  HAVING MIN(Salary) > 55k;

  SELECT Job, MAX(Salary)
    FROM Payroll
   WHERE MIN(Salary) > 55k
GROUP BY Job;

  SELECT Job, MAX(Salary)
    FROM Payroll
   WHERE Salary > 55k
GROUP BY Job;
```

6. Write a query to return everyone who owns a Civid AND a Ferrari

---

```
SELECT .....
```

7. At which stage (FJWGHOS) did Frances' tuple disappear? Quinn's typle disapppear?

---

8. What is your custom acronym for FJWGHOS ??

Section and Practice Questions

Prework

1) From the course website, download two data files. You can do this in a variety of ways, including right-clicking on the links at the assignment page or using wget at the command line interface (CLI) of your choice. (For example `wget https://courses.cs.washington.edu/courses/csed514/25sp/assignments/employeedata1pay.csv`)
2) Use your CLI and the change directory command, navigate to the file directory where you stored the database. For example `cd "~/Desktop/Data Science/data"`. Notice the double quotes surrounding the entire path term. Use these when you have spaces. Alternatively you can run `cd ~/Desktop/'Data Science'/data` to get the same output. Note the single quotes around the term 'Data Science'. On windows the commands would be `cd "C:\Users\[insert user]\Desktop\Data Science\data"` or `cd C:\Users\[insert user]\Desktop\'Data Science'\data`. 3) In the CLI, in the folder containing the csv file with extension .csv, run the command `sqlite3 employees.db` to open an empty database to load the csv file into. 4) Run the following commands: `CREATE TABLE salaries (id int PRIMARY KEY, first_name varchar(30), last_name varchar(30), salary int);`
`CREATE TABLE titles (id int PRIMARY KEY, first_name varchar(30), last_name varchar(30), title varchar(30));`
`.mode csv`

```
.import [file name] salaries
.import [file name] titles
```

Question 1

What are the two most common titles?

Question 2

What is the average pay for software engineers?

Question 3

For each title that has an average salary of less than $7000, how many employees have that title?

Question 4

Given tables created with these commands:

```
CREATE TABLE A (a int);
CREATE TABLE B (b int);
INSERT INTO A VALUES (1), (2), (3), (4);
INSERT INTO B VALUES (3), (4), (5), (6);
```

Part 1

What's the output for the following:

```
SELECT *
  FROM A INNER JOIN B
       ON A.a=B.b;
```

Part 2

What's the output for the following:

```
SELECT *
  FROM A RIGHT OUTER JOIN B
       ON A.a=B.b;
```

Part 3

What's the output for the following:

```
SELECT *
  FROM A LEFT OUTER JOIN B
       ON A.a=B.b;
```

Part 4

What's the output for the following:

```
SELECT *
  FROM A FULL OUTER JOIN B
       ON A.a=B.b;
```

Part 5

What type of join does the following code produce?

```
SELECT * FROM A INNER JOIN B;
```

Question 5

Consider the following over simplified Employee table:

```
CREATE TABLE Employees (id int, bossId int);
```

Suppose all employees have an id which is not null. How would we find all distinct pairs of employees with the same boss? Provide your sql query below.

Question 6

For the following tables:

```
CREATE TABLE Movies (id int PRIMARY KEY, name varchar(30), budget int,
                     gross int, rating int, year int);
CREATE TABLE Actors (id int PRIMARY KEY, name varchar(30), age int);
CREATE TABLE ActsIn (mid int REFERENCES Movies(id),
                     aid int REFERENCES Actors(id));
```

Part 1

What is the number of movies, and the average rating of all movies that the actor ”Patrick Stewart” has appeared in?

Part 2

What is the minimum age of an actor who has appeared in a movie where the gross of the movie has been over $1,000,000,000?

Part 3

What is the name and budget of each movie released in 2017 whose oldest actor is less than 30?