Assignment #2: SQL, Correlation & Causation
Sample Solution

Part 1: SQL on Kickstarter data (Same as HW1 questions)

Problem 1. Which was the project with the smallest goal? Return the name of the project and the goal amount.

```
Select *
From Kickstarter
Where goal = (Select min(goal) From Kickstarter)
```

Problem 2. What is the total sum of goal amounts for projects where the pledged amount was greater or equal than the goal amount? Return the total sum.

```
Select sum(goal)
From Kickstarter
Where pledged >= goal
```

Problem 3. Which was the most ambitious project (highest goal) where the pledged amount was greater or equal than the goal amount? Return the project's name and goal amount.

```
Select *
From Kickstarter
Where pledged >= goal
and goal = (Select max(goal) From Kickstarter Where pledged >= goal)
```

Problem 4. Which country has the highest percentage of successful projects (status = successful)? Return the country's name and the percentage rate of success.

```
Select K1.country, 1.0 * numsuccess / total as successrate
From
(Select country,  count(status)  as total
From Kickstarter
group by country) as K1,
(Select country,  count(status)  as numsuccess
From Kickstarter where status = 'successful'
group by country) as K2
Where K1.country = K2.country
Order by successrate desc
```

Problem 5. Which are the top 5 most popular categories (From highest to lowest) in terms of number of backers? Return the 5 categories along with the number of backers for each category.

```
Select category, sum(backers_count) as sum_backers
From Kickstarter
Group by category
Order by sum_backers desc
limit 5
```

Problem 6. What is the average pledge amount by continent? Return the average pledge amount for Europe, North America and Oceania.

```
Select Continent, avg(pledged)
From
Kickstarter, CountryContinents
Where Kickstarter.country = CountryContinents.country
Group by Continent
```

Part 2: SQL on Football data

Problem 7. Find all games where Seattle was the home or away team. For each game, return the two teams and scores.

```
Select home, homescore, away, awayscore
From Football
Where home = 'Seattle' or away = 'Seattle'
```

Problem 8. Find all games where the home team scored more than 40 points and the away team scored fewer than 10 points. For each game, return the home and away team, their scores, and the prediction for the game; return the result sorted by the home-team score in descending order.

```
Select home, homescore, away, awayscore, prediction
From Football
Where homescore > 40 and awayscore < 10
Order By homescore desc
```

Problem 9. Find all teams who lost by more than 20 points as the home team during the first four weeks of some year (i.e., Week < 5). Return the teams in alphabetical order and return each team only once.

```
Select Distinct Home
From Football
Where AwayScore - HomeScore > 20 and week < 5
Order By Home
```

Problem 10. Find the average prediction for games where the home team won.

Select avg(prediction) From Football Where HomeScore > AwayScore

Problem 11. In each of the three years (1998, 1999, 2000), how many games were there where the home team was predicted to win, i.e., prediction > 0, and the actual point spread (Homescore - Awayscore) was more than three times the prediction? Your result should have three tuples, each containing the year and number of games.

```
Select year, count(*)
From Football
Where prediction > 0 and HomeScore - AwayScore > 3*prediction
Group By year
```

Problem 12. Find all pairs of teams where the two teams played each other in 1998 and 1999 in the same configuration (the same team was home and the same team was away), and in 1998 the home team won while in 1999 the away team won. Return the pairs of teams.

```
Select F1.home, F1.away
From Football F1, Football F2
Where F1.year = 1998 and F2.year = 1999 and F1.home = F2.home and F1.away = F2.away
        and F1.homescore > F1.awayscore and F2.homescore < F2.awayscore
```

Part 3: SQL on Schoolkids data

Problem 13. Find all students whose goal is Sports but they ranked Sports as fourth in impact on popularity. Return the entire tuple for each such student.

*Select * From Schoolkids Where Goal = 'Sports' and Sports = 4*

Problem 14. Count how many 4th graders think Looks are more important than Grades. Return just one number.

Select count(*) From Schoolkids Where grade = 4 and Looks < Grades

Problem 15. For each of the three goals (Grades, Sports, Popular), what is the average age of students who have that goal? Your result should have three tuples, one each with the goal and the average age for that goal.

Select goal, avg(age) From Schoolkids Group By goal

Problem 16. Find all 4th graders in Urban schools where there is no 5th grader in their same school who has the same goal and same ranking of popularity impact factors. Return each student's grade, school, and goal.

```
Select grade, school, goal
From Schoolkids S1
Where grade = 4 and type = 'Urban' and Not Exists
```

```
(Select * From Schoolkids S2
 Where S2.grade = 5 and S2.school = S1.school and S2.goal = S1.goal
 and S2.grades = S1.grades and S2.sports = S1.sports and S2.looks = S1.looks
 and S2.money = S1.money)
```

Problem 17. Find the oldest students. For each one return their gender, grade, age, and school.

```
select grade, age, school
From Schoolkids
Where age >= (Select max(age) From Schoolkids)
```