SEEM2460 Introduction to Data Science
Assignment 2

Assignment 2 is due by **5:00 pm, February 29th** (Thu) 2024.
You need to submit your work along with the VeriGuide receipt on the blackboard website.

Question 1: SQL Practice (20 marks)

Table 1 is a set of daily weather records, where attributes "Air Pressure" and "Temperature" are the weather status in the morning, and the prediction column "Rain" records the raining conditions in the afternoon ("1" for "rain" and "0" for "no rain"). Table 2 shows additional weather data of the raining records in Table 1, where 1/0 in "Rain Yesterday" attribute refers the previous day is/isn't a raining day. You are asked to write SQL to get required results. You can use "Table1" and "Table2" as relational table names.

ID	Air Pressure (9 am)	Temperature (9 am)	Rain (3 pm)
1	High	Cold	1
2	Low	Warm	1
3	High	Warm	0
4	High	Warm	0
5	Low	Warm	1
6	High	Warm	0
7	High	Warm	0
8	High	Warm	0
9	High	Warm	0
10	Low	Cold	1
11	Low	Warm	1
12	High	Warm	0
13	High	Warm	0
14	Low	Cold	1
15	Low	Warm	1
16	High	Warm	0
17	Low	Warm	1
18	Low	Cold	1
19	High	Warm	0
20	High	Cold	1

Table 1: Daily weather records.

ID	Wind Direction (9 am)	Wind Speed (9 am)	Rain Yesterday	Rain (3 pm)
1	73.40	10.65	1	1
2	179.50	7.02	0	1
5	166.40	3.85	0	1
10	211.80	2.01	0	1
11	100.20	2.13	0	1

14	187.30	11.81	1	1
15	184.50	6.87	0	1
17	179.00	5.26	0	1
18	150.50	2.33	1	1
20	151.00	3.31	0	1

Table 2: Additional weather data of the raining records in Table 1

(a) Find the record ID where the Air Pressure is “High” and Temperature is “Cold”. Please write both **SQL** and corresponding **algebra** expression. (10 marks)

(b) Write a SQL with **natural join of** Table1 and Table2, then find tuples where the previous day is not a raining day. (5 marks)

(c) Natural join Table1 and Table2, then find the Air Pressures and Wind Speeds of the records of which Wind Direction is less than 185, Wind Speed is larger than 3 and the Temperature is “Cold”. (5 marks)

Question 2 (Supervised learning) (10 points)

The outside air temperature (Celsius) is related to the daily revenue (dollar) of an ice cream store. The following link is a dataset containing information about the outside air temperature and the revenue for an ice cream shop on 500 different days.

https://drive.google.com/file/d/1ZxudIGw5l67dTMe_pFmo4tvkbFIICkp2/view?usp=sharing

Please use linear regression to predict the revenue of the ice cream store using the temperature attribute. Then plot the regression line your get and show its coefficients (5 points). Considering the temperature is 26.6 Celsius, please predict the revenue of an ice cream store using your regression function (5 points).

You can use R programming language (refer to pages 60-61 of Week 3A Course Slides), or any other programming languages feasible for this question (e.g., Python). Note that code plagiarism is not allowed, you should work on your own code.