Games DS homework test

Question 1

A gaming company stores its customers, transactions, and campaigns data in 3 tables with the following format.

customers table			transactions table				campaigns table		
id	**Name**	**RegDate**	**id**	**customerid**	**value**	**timestamp**	**id**	**customerid**	**Campdate**
1	Xin	2018-01-01	1	2	0.71	2018-03-04 12:45	1	5	2019-01-01
...	...	...	...	...	...	...	...	...	...

Write SQL queries which retrieve following data

1. All customers with no transactions;
2. For each customer output **mean** and **maximum** number of days between two consecutive transactions. Return "-1" for customers with no or only one transaction; e.g. if a player had transactions on '2019-01-01', '2019-01-02' and '2019-01-05' then number of days between transactions is 1 and 3, average being (1+3)/2 = 2 and maximum being 3.

Question 2

A gaming company is hiring a data scientist. Applicants are asked to participate in two online tests: (a) SQL and (b) Coding. For each of the tests, applicants are given a score between 0 and 100. Applicants with a total score ("SQL"+"coding") of more than 100 are invited for an interview.

For the question, assume that applicants' scores in these two online tests are independently uniformly distributed random variables between 0 and 100. Explain and quantify your answers to the following questions:

1. What is the correlation between SQL and Coding scores of **all applicants**?
2. What is the correlation between SQL and Coding scores of **applicants who passed the test**?

Question 3

Order the following situations based on the expected power of the statistical test intended to identify the difference between two groups. Explain your reasoning.

1. Datasets come from normal distributions with equal unknown variances
2. Datasets come from normal distributions with unequal known variances
3. Datasets come from two unknown distributions
4. Datasets come from normal distributions with unequal known variances and known unequal means
5. Datasets come from gamma distributions with unknown parameters

Question 4

Write a function in **R or Python** which takes a vector of length L consisting only of 0s and 1s as input and returns the largest number of consecutive 1s from the vector as an integer, e.g. the function takes [1,1,0,1,1,1,0,0,1] and returns 3.