Question 1:

Given that D is the data set of all NCSU employees. For the sake of this question, let us assume that there are 5 employees in total and their salaries are reported as follows:

A	10K
B	20K
C	30K
D	40K
E	50K

Given an objective function *f* as the mean of the salaries in D. Therefore, whenever a query is made to the database D, the result is added with Laplacian noise as *f(D)* + ξ.

a. In order to calculate the sensitivity of the mean function, we will need to identify what the largest change in the dataset D is when one of the tuples have changed. When we look at this dataset D, all the salaries belong in the range [10k,50k]. The average salary of this dataset can be computed as $(\sum_{i=1}^{5} Sal_i)/5$ = 30K.

   Now, by observation in the database, we can say that the maximum change in the average when a tuple is removed occurs when either the largest or the smallest value of the salary is removed. Therefore, when we remove 10K, we have the average salary as as $(\sum_{i=2}^{5} Sal_i)/4$ = 25K.
   Similarly, when we remove 50K, we have the average salary as $(\sum_{i=1}^{4} Sal_i)/4$ = 25K.

   From the above calculations, we can see that the highest variation in the average occurs as a difference of 5K in salary.

   Therefore, the sensitivity of the function f is 5K.

b. Now, in order to achieve ε-differential privacy using Laplace noise, we know that the λ parameter is calculated as follows:

$$\lambda \geq |h - h'|/\varepsilon$$

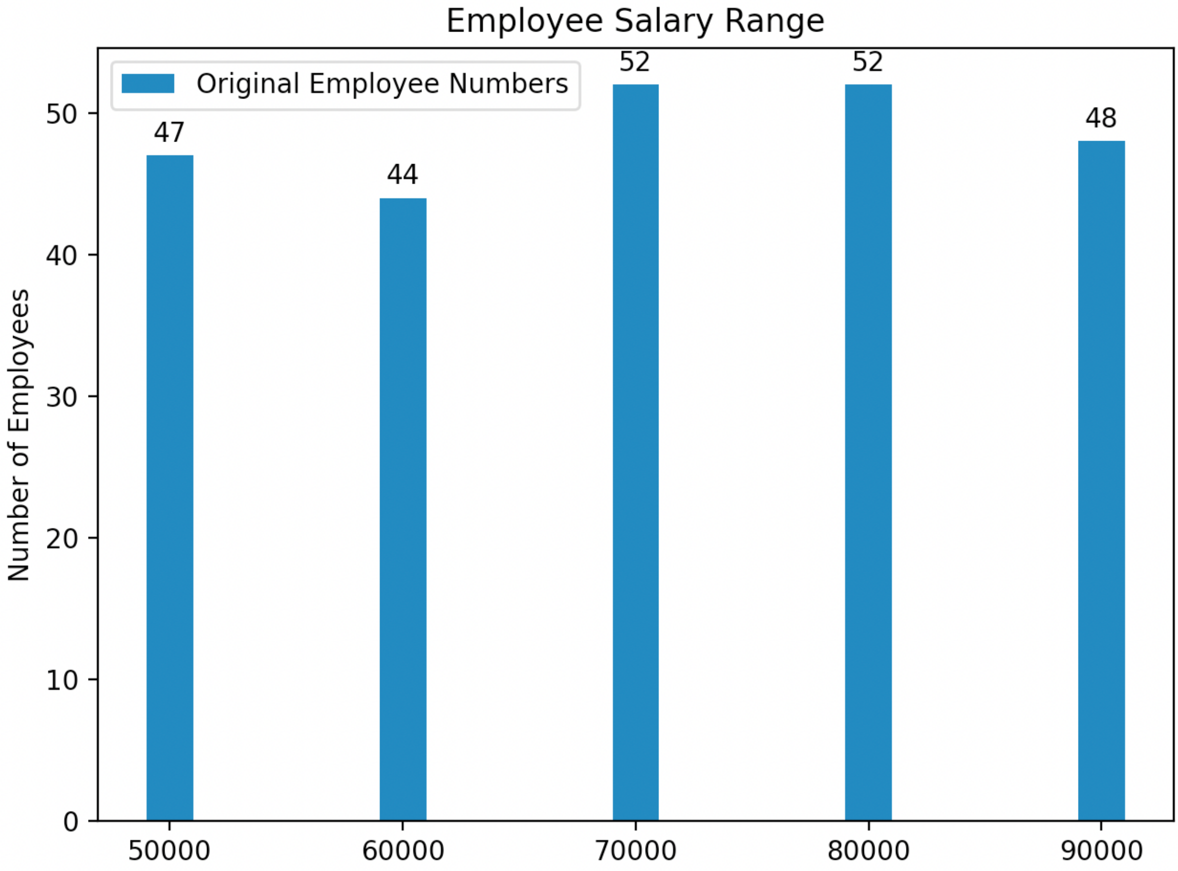
   Assuming that ε is 0.1, we can say that $\lambda \geq 5k/0.1$. Therefore, we can pick:

$$\lambda = 50K$$

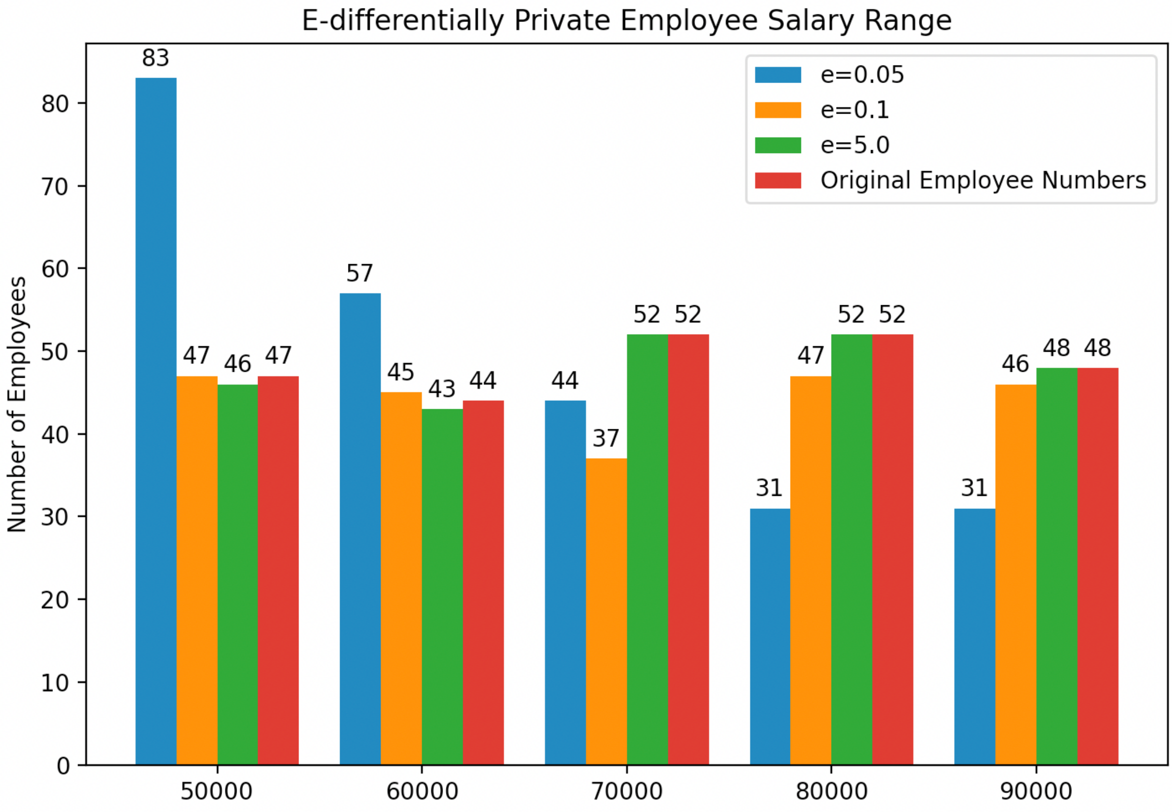
Question 2:

a. The histogram for the salary of the employees was calculated and plotted using the following ranges:
   50k-60k
   60k-70k
   70k-80k
   80k-90k
   90k-100k

   The corresponding histogram is shown below:



b. The e-differentially private histogram is plotted below that shows how the numbers vary based on the values of ε = 0.05, 0.1, 5.0. The reference of the original dataset is also plotted. Since we know that the laplace noise can be negative as well, care is taken to ensure that we round to 0 in case of a negative value.



c. As the values of ε are increased, we can see that the values of the number of employees are approaching closer to that of the unperturbed histogram. We can conclude that the utility of the histogram increases as you increase ε from 0.05 to 5.

Note: The code is provided in the submission files along with a readme and requriements.txt.

Question 3:

The given data is as follows:

0	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1

Converting this to a Square matrix, we have the following:

0	1	1	0
0	1	0	1
1	0	0	1
1	0	1	1

The data bit that we need to retrieve is shaded in green in the above matrix. The i and j values for this desired element is: i=2, j=3.

The data on S1 and S2 would be:
S1:

0	1	1	0
0	1	0	1
1	0	0	1
1	0	1	1

S2:

0	1	1	0
0	1	0	1
1	0	0	1
1	0	1	1

It is known that for the 2-server $O(n^{1/2})$ PIR scheme, we need to generate Q1 which is a random sequence of 4 bits.
For this assignment, we will assume Q1 = 1010.

Now, we perform a dot product with each row of S1 with Q1 and xor the result across. We have the following:

$1.0 \oplus 0.1 \oplus 1.1 \oplus 0.0 = 1$
$1.0 \oplus 1.0 \oplus 1.0 \oplus 1.0 = 0$
$1.1 \oplus 0.0 \oplus 1.0 \oplus 1.0 = 1$
$1.1 \oplus 0.0 \oplus 1.1 \oplus 1.0 = 0$

Now, Q2 for S2 can be derived from Q1, by flipping the jth bit of Q1. Therefore, we have Q2 = 1011.

Now, we perform a dot product with each row of S2 with Q1 and xor the result across. We have the following:

$1.0 \oplus 0.1 \oplus 1.1 \oplus 1.0 = 1$
$1.0 \oplus 0.1 \oplus 1.0 \oplus 1.1 = 1$
$1.1 \oplus 0.0 \oplus 1.0 \oplus 1.1 = 0$
$1.1 \oplus 0.0 \oplus 1.1 \oplus 1.1 = 1$

Now that we have both the columns that have been computed for S1 and S2, we need to take the ith element from each of these columns and do an XOR operation. The ith elements of the two columns are highlighted above in green.

Therefore, $1 \oplus 0 = 1$. We now have the value that we needed to retrieve from the database, and it was retrieved using the 2-Server $O(n^{1/2})$ PIR scheme.