**COMP 430 ASSIGNMENT 2 REPORT, Çisem Özden 69707**

**PART 1: Privacy Proofs**

a) $$\frac{P[A(D)=O]}{P[A(D')=O]} = \underbrace{\frac{P[A_1(D)=O_1]}{P[A_1(D')=O_1]}}_{\leq e^{\varepsilon_1}} \cdot \underbrace{\frac{P[A_2(D)=O_2]}{P[A_2(D')=O_2]}}_{\leq e^{\varepsilon_2}} \cdot \ldots \cdot \underbrace{\frac{P[A_n(D)=O_n]}{P[A_n(D')=O_n]}}_{\leq e^{\varepsilon_n}}$$

Multiplying them, we get $\leq e^{(\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_n)} = e^{\sum_{i=1}^{n}\varepsilon_i}$

It means that the sequential composition of algorithms $A_1, A_2 \ldots, A_n$ satisfies $\sum_{i=1}^{n}\varepsilon_i$ -DP.

b) Let D be a dataset whose number of records greater than $e^{\varepsilon}$.
Let D' be the neighboring dataset of D formed by removing 1 record from D. Assuming number of records of D' is smaller than $e^{\varepsilon}$, we have the following:

$$\frac{P[A(D)=\text{"large"}]}{P[A(D')=\text{"small"}]} = \frac{1}{0} > e^{\varepsilon}$$

Thus, it is not $\varepsilon$-DP.

c) Using the definition of $\varepsilon$-DP,

$$\frac{P[A(D)=O]}{P[A(D')=O]} = \frac{P[q(D)+Lap(\varepsilon)=O]}{P[q(D')+Lap(\varepsilon)=O]} = \frac{P[Lap(\varepsilon)=O-q(D)]}{P[Lap(\varepsilon)=O-q(D')]} = \frac{\frac{1}{2\varepsilon}\cdot e^{\frac{|O-q(D)|}{\varepsilon}}}{\frac{1}{2\varepsilon}\cdot e^{\frac{|O-q(D')|}{\varepsilon}}}$$

$$= e^{\frac{|O-q(D)|-|O-q(D')|}{\varepsilon}} \leq e^{\frac{|O-q(D)-O+q(D')|}{\varepsilon}} = e^{\frac{|q(D')-q(D)|}{\varepsilon}} \leq e^{\frac{S(q)}{\varepsilon}}$$

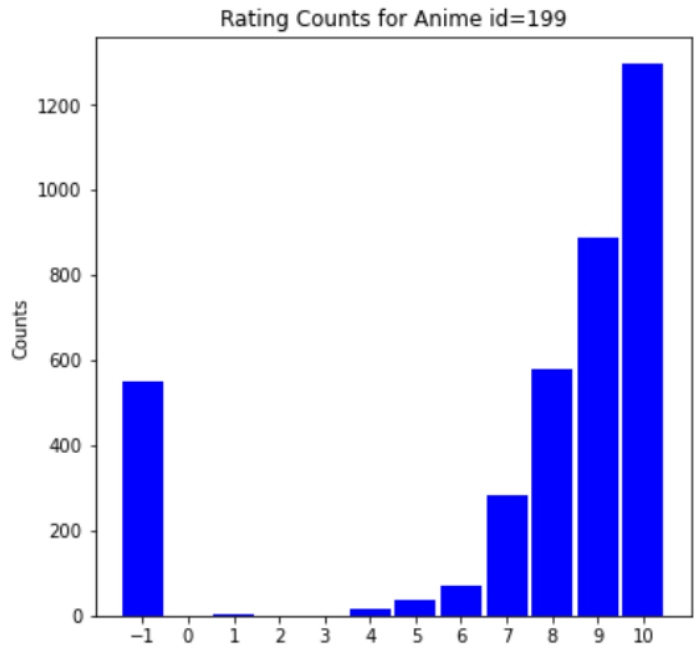
↓ Using Triangular Inequality

For this algorithm to satisfy $\varepsilon$-DP, $e^{\frac{S(q)}{\varepsilon}} \leq e^{\varepsilon}$

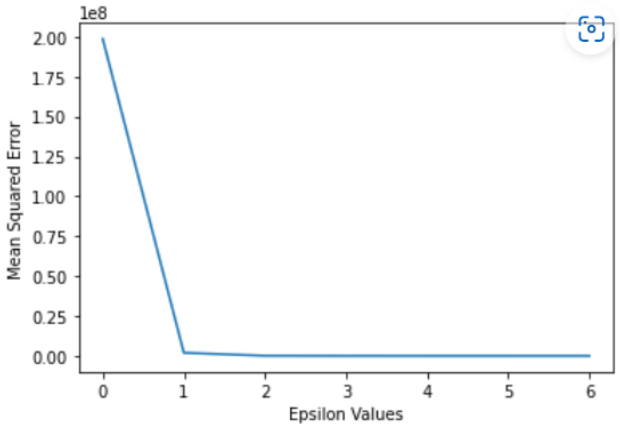
↳ $\frac{S(q)}{\varepsilon} \leq \varepsilon \rightarrow S(q) \leq \varepsilon^2$ should be satisfied for it to satisfy $\varepsilon$-DP.

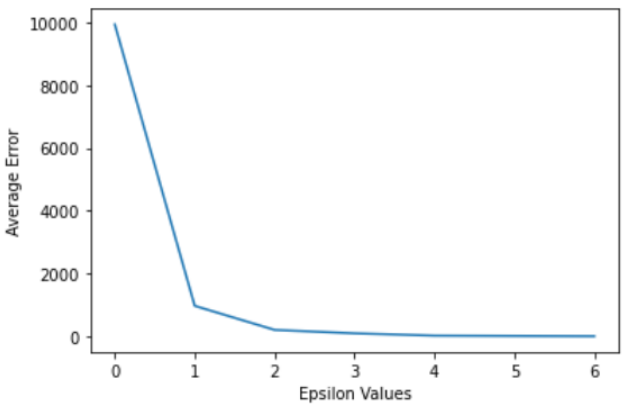
**PART 2: DP Implementation**

**Task 1.a:**




**Task 2.d:**

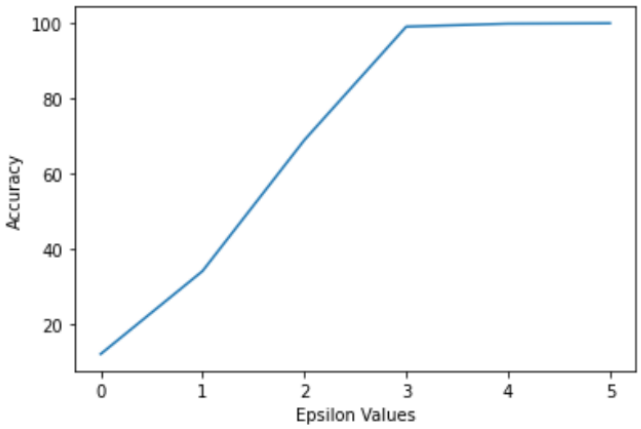



```
**** LAPLACE EXPERIMENT RESULTS ****
**** AVERAGE ERROR ****
eps =  0.0001  error =  9944.212768212294
eps =  0.001  error =  969.7800926126154
eps =  0.005  error =  202.3853675881972
eps =  0.01  error =  92.70191021175326
eps =  0.05  error =  20.22900587029423
eps =  0.1  error =  10.551274529245667
eps =  1.0  error =  1.0204497185176375
**** MEAN SQUARED ERROR ****
eps =  0.0001  error =  198645339.73315877
eps =  0.001  error =  1918894.9521268327
eps =  0.005  error =  85125.2431929259
eps =  0.01  error =  17881.086316402667
eps =  0.05  error =  782.2500891812119
eps =  0.1  error =  215.44034038128993
eps =  1.0  error =  2.0520175478486107
```

According to the result of Laplace experiment, we see that epsilon value is inversely proportional to average error and mean squared error. As the epsilon value increases, we observe decrease in the average and mean squared error.

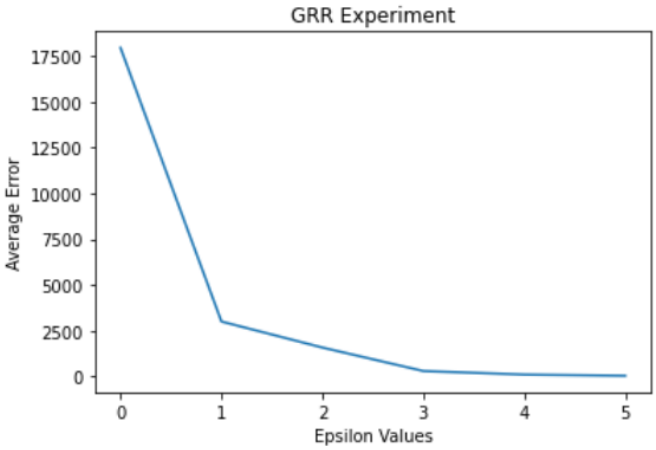
**Task 2.f:**

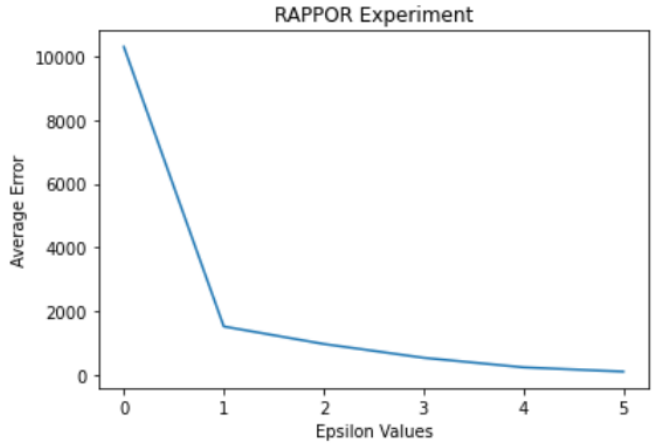


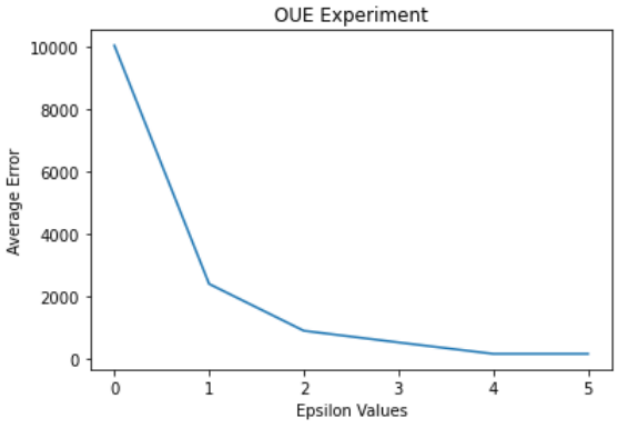

```
**** EXPONENTIAL EXPERIMENT RESULTS ****
eps =  0.001  accuracy =  11.9
eps =  0.005  accuracy =  34.0
eps =  0.01  accuracy =  68.9
eps =  0.03  accuracy =  99.1
eps =  0.05  accuracy =  99.9
eps =  0.1  accuracy =  100.0
```

From the result of Exponential Mechanism experiment, it is seen that epsilon value and the accuracy are directly proportional. As the epsilon value increases, we observe an increase in the accuracy as well.

**PART 3: LDP Implementation**






```
GRR EXPERIMENT
e=0.1, Error: 17956.14
e=0.5, Error: 3000.68
e=1.0, Error: 1578.69
e=2.0, Error: 295.96
e=4.0, Error: 103.08
e=6.0, Error: 39.30
**************************************************
RAPPOR EXPERIMENT
e=0.1, Error: 10304.47
e=0.5, Error: 1513.53
e=1.0, Error: 963.83
e=2.0, Error: 528.98
e=4.0, Error: 228.69
e=6.0, Error: 94.83
**************************************************
OUE EXPERIMENT
e=0.1, Error: 10060.74
e=0.5, Error: 2405.91
e=1.0, Error: 901.67
e=2.0, Error: 527.14
e=4.0, Error: 161.37
e=6.0, Error: 116.26
```

According to results of the experiment we've conducted, it is clearly seen that average error value decreases as the value of epsilon increases. This conclusion is common for all the protocols we've implemented. When the value of epsilon increases average value decreases for GRR, RAPPOR, and QUE protocols.

From this result, it wouldn't be right to conclude that this algorithm is better or that algorithm is worse. For some values of epsilon, one protocol is better than others, for some other values of epsilon, another protocol is better. Moreover, I also encountered that one protocol is better for all epsilon values compared to other protocols. However, when I run the code multiple times, I saw that this situation does not hold always as well. Since the results are based on randomness, it is not surprising that not conclude on the best protocol, at least for couple of experiments.