There are many nice dragons of three main colors (red, green, blue) living in a Dragon City. Each dragon has a collection of treasures which can be measured in a special currency unit called dragon coins.

Once the dragons decided to find out which of the three tribes (red, green, or blue) is the most wealthy. For this, they counted the total wealth of all dragons of each color. First of all, they constructed a dataset dragons of records with attributes name, color, wealth, where name is the dragon’s name, color is the dragon’s color (red, green, or blue), and wealth is the total wealth of the dragon, measured in dragon coins. A pictorial example of the dataset (with dragon names omitted) is given in Figure 1.

The dragons are interested in the following queries:

Q1: How many dragon coins does the most wealthy tribe possess?

```sql
SELECT MAX(tribe_wealth) FROM ( 
    SELECT color AS tribe_color,
       SUM(wealth) AS tribe_wealth 
    FROM dragons 
    GROUP BY color 
) ;
```

Q2: Which tribe is the most wealthy?

```sql
SELECT color FROM ( 
    SELECT color AS tribe_color,
       SUM(wealth) AS tribe_wealth 
    FROM dragons 
    GROUP BY color 
) ORDERED BY tribe_wealth DESC LIMIT 1;
```

It is known in advance that the wealth of a single dragon always fits into the following ranges:

- **red**: from 80 to 110 coins.
- **green**: from 50 to 90 coins.
- **blue**: from 100 to 120 coins.
1 No Differential Privacy (1p)

The answers to queries Q1 and Q2 were first released without applying differential privacy mechanisms. The answer to Q1 was 15950, and the answer to Q2 was green. The organizers thought that publishing just the aggregated answers is good enough to preserve privacy. However, these answers were noticed by an evil Dragonfighter who in addition could acquire the following data:

- The Drake has exactly the same amount of coins as his 9 friends who are of the same color as Drake is.
- All the other dragons that are of the same color as Drake (except his friends) together hold 15000 coins.

What could be the color of Drake? The wealth of Drake? It is fine if the attacker still cannot guess something exactly. Justify your answer.

NB! The actual dataset is not the one of depicted in Figure 1.

2 Choosing a suitable $\epsilon$ (3p)

The dragons decided to enhance the outputs of Q1 and Q2 with $\epsilon$-differential privacy. For this, they are planning to apply separately an $\epsilon_1$-DP mechanism to Q1, and an $\epsilon_2$-DP mechanism to Q2. They want to decide how small $\epsilon_1$ and $\epsilon_2$ should be.
1. Let $y$ be the observed noisy output (after applying an $\epsilon$-DP mechanism). The dragons want that the probability of getting the same output $y$ in the case where the dataset does contain Drake’s data would be at most 1.9 times the probability of getting $y$ in the case where the dataset does not contain Drake’s data, assuming that all the other records are the same in both cases. Which privacy level $\epsilon$ should the applied DP mechanism satisfy?

2. Let the output $y$ consist of two separate outputs $y = (y_1, y_2)$, where $y_1$ is an $\epsilon_1$-DP answer to $Q_1$, and $y_2$ is an $\epsilon_2$-DP answer to $Q_2$. Which values of $\epsilon_1$ and $\epsilon_2$ should one take so that the output $y$ would be $\epsilon$-DP, assuming that the value of $\epsilon$ is obtained in the previous point? If there are multiple possibilities, you can suggest any of them.

3. The dragons want to take into account that the evil Dragonfighter may know in advance certain correlations inside groups of dragons, such as having the same color. How should the $\epsilon_1$ and $\epsilon_2$ obtained in the previous point be changed so that the output $y$ would provide $\epsilon$-DP for groups of size up to 10?

3 Differential Privacy for Q1 (3p)

The dragons decided to enhance the output of Q1 with $\epsilon_1$-differential privacy.

1. Let Laplace mechanism be applied with privacy parameter $\epsilon_1 = 0.5$. Assuming that the dataset looks like in Figure 1, compute the probability density function of the distribution of noisy output at the points 750, 450, 600, 0, $-100$.

2. Let $\epsilon_2 = 0.5$. What is the probability that the noise will stay in the range $[-100, 100]$?

3. The dragons released by mistake many instances of a noisy output of Q1, which can be found in dpSamples_Q1.csv.
   - What could be (approximately) the true output of Q1?
   - What is the actual level of differential privacy after releasing all these noisy outputs, assuming that each single output is $\epsilon$-DP for $\epsilon = 0.5$?
     
     Hint: It is difficult to analyse how much exactly the particular set of outputs leaks. However, you can find an upper bound by applying composition theorem.

4 Differential Privacy for Q2 (3p)

The dragons decided to enhance the output of Q2 with $\epsilon_2$-differential privacy. The output of the query Q2 was protected using exponential mechanism. The dragons decided to measure the “goodness” of the color output by Q2 by the total wealth of the dragons sharing that color. That is, the score function $Q(\text{dragons, color})$ is defined as the output of the query

$$\text{SELECT SUM(wealth) FROM dragons WHERE color}=\text{color};$$
where $\text{color} \in \{\text{red, green, blue}\}$ is the set of possible outputs, and $\text{dragons}$ is the dataset.

1. Let exponential mechanism based on score function $Q$ be applied with privacy parameter $\epsilon_2 = 0.5$. Assuming that the dataset looks like in Figure 1, find the probabilities of exponential mechanism returning the color $\text{red, green, blue}$.

2. Let $\epsilon_2 = 0.5$. What is the probability that the difference in wealth of the true winner (true output of $Q_2$) and the reported winner (output of $Q_2$ after applying exponential mechanism) will stay in the range $[-1000, 1000]$?

3. The dragons released by mistake many instances of a noisy output of $Q_2$, which can be found in $\text{dpSamples}_Q2.csv$.
   - What could be the true output of $Q_2$?
   - Can the attacker infer that there are no red dragons in the database?