Part 1 - a

No. DP is not about micro data,
it is about macrodata.

Since we are told all questions are answered with DP, I assume this is a valid DP.
Valid DP means that all database entries added noise with randomisation algoritms. Since noise added with randomization, even if we know speacial queries that expose Ali's identifiers due to the randomization nature of Algorithm we cannot deduce the rule of noise.

Part 1-b

$$DP \Rightarrow \frac{Pr[A(D)=0]}{Pr[A(D')=0]} \leq e^{\varepsilon}$$

Since it is said they are independent algorithm their probability doesnt affect the other's. So we can calculate their propebility by multiplying them.

$$\underbrace{\frac{Pr[A_1(D)=0]}{Pr[A_1(D')=0]}}_{\leq e^{\varepsilon_1}} \cdot \underbrace{\frac{Pr[A_2(D)=0]}{Pr[A_2(D')=0]}}_{\leq e^{\varepsilon_2}} \cdots \underbrace{\frac{Pr[A_n(D)=0]}{Pr[A_n(D')=0]}}_{\leq e^{\varepsilon_n}}$$

$$\leq e^{\varepsilon_1+\varepsilon_2 \cdots \varepsilon_n}$$

$$\leq e^{\sum_{i=1}^{n} \varepsilon_i}$$

Problem 2

$$\underbrace{\frac{\Pr[A(v_1)=y]}{\Pr[A(v_2)=y]} \leq e^{a \cdot d(v_1, v_2)}}_{a\text{-MLDP imposition}}$$

$$\Pr[\psi(v_1)=y] = \frac{e^{-\frac{a\,d(v_1,y)}{2}}}{\sum_{z \in u} e^{-\frac{a\,d(v_1,z)}{2}}}$$

$$\Pr[\psi(v_2)=y] = \frac{e^{-\frac{a\,d(v_2,y)}{2}}}{\sum_{z \in u} e^{-\frac{a\,d(v_2,z)}{2}}}$$

$$\frac{\Pr(\psi(v_1)=y)}{\Pr(\psi(v_2)=y)} = \frac{e^{-\frac{a\,d(v_1,y)}{2}}}{\sum_{z \in u} e^{-\frac{a\,d(v_1,z)}{2}}} \cdot \frac{\sum_{z \in u} e^{-\frac{a\,d(v_2,z)}{2}}}{e^{-\frac{a\,d(v_2,y)}{2}}}$$

$$e^{-\frac{a}{2}\left(d(\vartheta_2,Y)-d(\vartheta_1,Y)\right)} \cdot \frac{\sum_{z\in\mathcal{U}} e^{-a\frac{d}{2}(\vartheta_2,z)}}{\sum_{z\in\mathcal{U}} e^{-a\frac{d}{2}(\vartheta_1,z)}} -$$

Triangular equality

$$d(\theta_2,z) \leq d(\vartheta_2,\vartheta_1) + d(\vartheta_1,z)$$

To make nominator max choose the equal case.

$$= e^{-\frac{a}{2}\left(d(\vartheta_2,Y)-d(\vartheta_1,Y)\right)} \cdot \frac{\sum_{z\in\mathcal{U}} e^{-\frac{a}{2}\left(d(\vartheta_2,\vartheta_1)+d(\vartheta_1,z)\right)}}{\sum_{z\in\mathcal{U}} e^{-\frac{a}{2}d(\vartheta_1,z)}}$$

$$= e^{-\frac{a}{2}\left(d(\vartheta_2,Y)-d(\vartheta_1,y)\right)} \cdot e^{-\frac{a}{2}d(\vartheta_2,\vartheta_1)} \cdot \cancel{\frac{\sum_{z\in\mathcal{U}} e^{-\frac{a}{2}d(\vartheta_1,z)}}{\sum_{z\in\mathcal{U}} e^{-\frac{a}{2}d(\vartheta_1,z)}}}$$

Use another triangular equity

$$d(\vartheta_2,Y) \leq d(\vartheta_2,\vartheta_1) + d(\vartheta_1,Y)$$

$$d(\vartheta_2,Y) - d(\vartheta_1,Y) \leq d(\vartheta_2,\vartheta_1)$$

Choose equaility again

$$= e^{-\frac{a}{2}(d(\theta_2,\theta_1))} \cdot e^{-\frac{a}{2}(d(\theta_2,\theta_1))}$$

$$= e^{-a(d(\theta_2,\theta_1))}$$

Use symmetry

$$= e^{-a(d(\theta_1,\theta_2))} \leq e^{a(d(\theta_1,\theta_2))}$$

Since we know

$a>0$ and $d(\theta_1,\theta_2) \geq 0$

It satisfy the imposition of a-MLDP.

Comp430- Project 2 Report

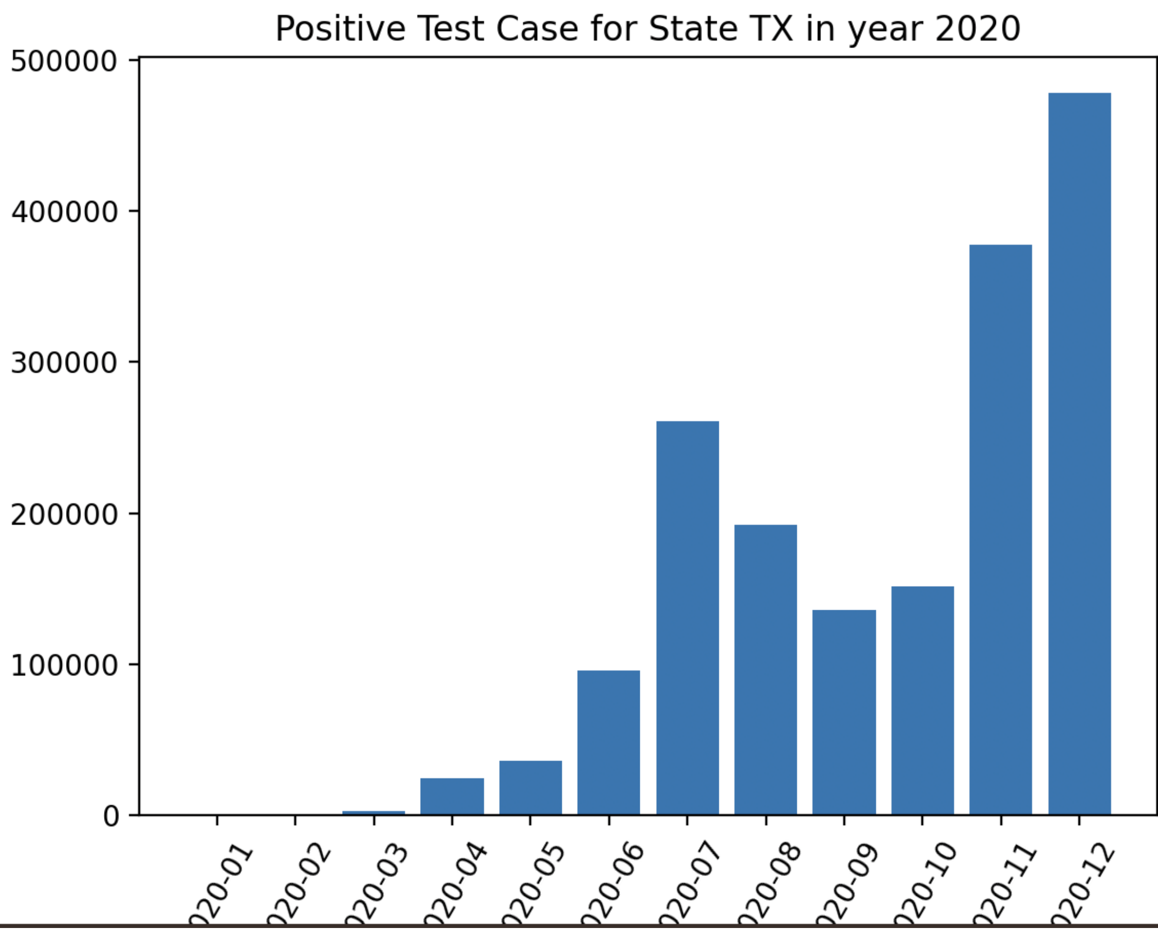
Question 3

Batuhan Arat

68665

Task 1

a)



b)N gives us a sensitivity value. Because as it said in the pdf neighbouring datasets are obtained by addition or removal of one individual. Since one can have maximum N times get covid. The removal or addition of one individual can have impact of max N.

d)

Epsilon is our privacy coeffienct, and it infers the how private the data is. While epsilon is getting bigger we are having much more non-privacy. So epsilon’s getting bigger decrease the privacy. When it is getting smaller we are having much more noise, so we are having much more privacy. So while epsilon is getting bigger, we have less noise, therefore our original data is much more lookalike with our noisy data so we have less error

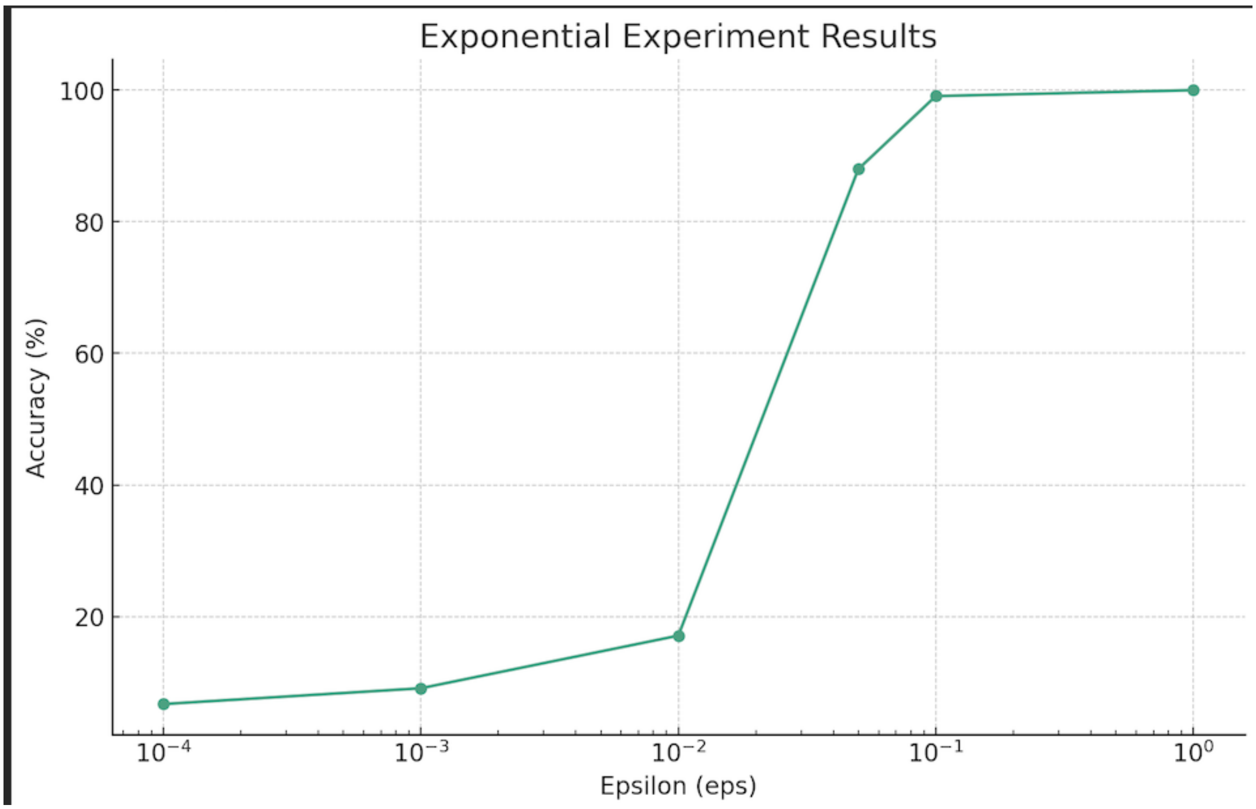
Epsilon	Error
0.0001	23102.402408
0.0010	2499.359105
0.0050	392.228835
0.0100	218.668255
0.0500	42.262269
0.1000	19.541668
1.0000	2.059238

e)

N directly associated with our sensitivity. While N is getting bigger our privacy is getting bigger too. Increase in sensitivity means much more noise to satisfy. Therefore while our N is getting bigger, our privacy is also getting enriched and error is increasing because of the increased noise.

N	Error
1	1.921794
2	4.149880
4	8.952032
8	16.810416

Task 2



Epsilon is a coefficient of a how private our data is. While it is getting bigger we are having less privacy. In that graph we can see while epsilon is getting bigger, our data is much more accurately close to our original data so we have more accuracy