5a.

See table below for part a:

Part	Model	PPV	NPV	FPR	FNR
b	Overall	0.81	0.79	0.14	0.28
b	Group 0	0.83	0.82	0.12	0.23
b	Group 1	0.75	0.71	0.17	0.4
c	Overall	0.81	0.79	0.14	0.28
c	Group 0	0.83	0.82	0.12	0.23
c	Group 1	0.75	0.71	0.17	0.4
	G=0				
d	Overall	0.85	0.88	0.12	0.15
	G=1				
d	Overall	1	1	0	0

5b.

The classifier appears to incorrectly give out loans to group 1 (higher FNR/FPR than group 0). I would rather be a member of group 0 when applying for a loan because I would be more confident that my loan application would be considered fairly and accurately. Group 0's metrics are most similar to the overall metrics, probably because the training sample consisted predominantly of those in group 0.

5c.

This model was given the group data to train on, whereas the part b model was not. The decision boundary and fairness metrics do not significantly differ from those of part b despite this difference. This surprised me a little because I would have expected including a new metric to change the classifier's predictions or perhaps make it more biased. The fact that little changed shows that group is already accounted for by our model in part b via its relationship to the other variables used to fit the data.

5d.

These models are different in that each one is only trained on data from one group. If I were a member of group 0, I would rather be classified by the part d classifier because it has lower false negative and false positive rates. If I were a member of group 1, I would also rather be classified by the part d classifier because it has much lower false negative and false positive rates.

5e.

One challenge is that it will be much harder to ensure fairness along 8 attributes than 1 – these additional attributes will constrain the model much more significantly than just one attribute would. Another challenge is that it might be hard to get data for each combination of subgroup – with 8 different subgroups (even assuming they are just binary which is untrue) you have to ensure that you have enough sample data from 256 different groups of people. In addition, some of these groups likely have low representation in the population already, so collecting a "fair" dataset to train on will be quite challenging.