The dataset I have chosen is the COMPAS Recidivism Risk Score Data and Analysis from ProPublica. This dataset can be found on their GitHub: https://github.com/propublica/compas-analysis/. I will specifically be using the compass-scores-two-years.csv, which contains 7,215 rows and 53 columns. It contains a number of categorical variables that will be useful for training the model such as sex, age category, race, decile score, violent decile score, charge degree, etc. The key feature the data contains is a binary variable, which represents if the person went back to jail within two years of their charge. I will use the various categorical features described as inputs for the logistic regression model and predict if the person will return to jail in two years.

The first article we read for this class was on biased machine learning algorithms having a say in someone's jail sentence. That article spiked my interest greatly because biased models are a major issue today. The COMPAS algorithm which spits out the decile score has been questioned for its racial biases. There are many examples and articles about the algorithms predictions and how they are biased. I want to investigate if I can accurately predict whether an inmate will return within two years of their release, without the biased score. I will also then spend time investigating what biases lie within my own model. I also would like to create a model that uses the COMPAS decile score and see if my model reflects the same biases or if they are different. The last thing I would like to investigate would be generating a model which does not have race as an input feature, as this seems to be the subject of controversy for the recidivism algorithms. It does not seem feasible to create a model with no biases for this problem, but it would be interesting to see how much I can mitigate them and how they change with the various inputs. I predict that my model will hold similar biases after investigating the data, but we will find out.

Fass, Tracy L., et al. "The LSI-R and the COMPAS: Validation data on two risk-needs tools." *Criminal Justice and Behavior* 35.9 (2008): 1095-1108.

Brennan, Tim, William Dieterich, and Beate Ehret. "Evaluating the predictive validity of the COMPAS risk and needs assessment system." *Criminal Justice and Behavior* 36.1 (2009): 21-40.

Dressel, Julia, and Hany Farid. "The accuracy, fairness, and limits of predicting recidivism." *Science advances* 4.1 (2018): eaao5580.

Van Berkel, Niels, et al. "Crowdsourcing perceptions of fair predictors for machine learning: a recidivism case study." *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019): 1-21.

Tollenaar, Nikolaj, and P. G. M. Van der Heijden. "Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176.2 (2013): 565-584.

Zeng, Jiaming, Berk Ustun, and Cynthia Rudin. "Interpretable classification models for recidivism prediction." *arXiv preprint arXiv:1503.07810* (2015).

Wang, Caroline, et al. "In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction." *arXiv preprint arXiv:2005.04176* (2020).