Final Project

FAQ

General questions:

How do I submit my team member names?
See the directions in the Final Project instructions PDF.

Can I use any dataset?
*Datasets that are **NOT ALLOWED** to be used in the final project:*

Dataset Name	**URL**
Mental Health in Tech Survey: Survey on Mental Health	https://osmihelp.org/research/
in the Tech Workplace in 2019	https://health.data.ny.gov/dataset/Hospital-Inpatient-Discharges-SPARCS-De-Identified/22g3-z7e7
Information on deaths that occur in custody or during the process of arrest in California	https://openjustice.doj.ca.gov/data
toxicity-per-attribute (unintended-ml-bias-analysis)	https://github.com/conversationai/unintended-ml-bias-analysis/tree/master/data
Google News dataset	https://code.google.com/archive/p/word2vec/
German Credit Data Set	https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)
Taiwan Credit Data Set	https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients
Portuguese Bank Marketing Data Set	https://archive.ics.uci.edu/ml/datasets/Bank+Marketing
Compas (ProPublica recidivism)	https://github.com/propublica/compas-analysis
Adult census income	https://archive.ics.uci.edu/ml/datasets/adult
UCI Census Income Dataset	http://archive.ics.uci.edu/ml/datasets/Census+Income
Compas dataset	https://www.kaggle.com/datasets/danofer/compass
IBM HR Analytics Employee Attrition & Performance	https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

How many observations and protected classes does my dataset need to have?
At least 500 observations, 2 protected class variables and 2 dependent variables. If you only have one dependent variable, the second one can be derived (see the next question).

What is a derived dependent variable?
A derived dependent variable is created by calculating or categorizing other variables present in the dataset. For example, if your dataset contained the columns Primary_Income and Secondary_Income, the variable Total_Income (Primary_Income + Secondary_Income) would be a derived dependent variable.

How should we handle Nulls in the dataset? Can we do data manipulation? Or should we drop rows with NaNs?
It's entirely up to you and your team how you want to handle this. We've had teams in the past take both approaches. Please note the approach used in your report.

Are we allowed to aggregate attributes/combine groups into our own buckets for the final project?
You can combine groups as you need to. Please mention them in your report.

Step 3

"For each protected class variable, select two fairness metrics and compute the fairness metrics associated with your privileged/unprivileged groups as a function of each of your two dependent variables..." Can we use the same dependent variables for each fairness metric/protected class combination?
You must use the same dependent variables.

Are we allowed to use both Statistical Parity Difference and Disparate Impact as separate fairness metrics? Or are these too similar and we should use something else like Equal Opportunity, Demographic Party, etc.?
Yes, you may use both Statistical Party Difference and Disparate Impact!

Is it possible that fairness metrics calculated on the original and transformed testing datasets are the same?
Yes, this is possible (i.e. the change is 0.0). If this happens, note it in your report.

Step 4

Are we supposed to use 50% for the random split into training/test or does the test size just have to contain at least a certain amount of samples (example at least 100 samples)?
It's up to you to decide. The common split ratio is 70-30% or 80-20% train/test split. You can still use 50-50% too.

Step 4 states:
"For the next set of questions, you are allowed to code up your own algorithm, modify open-source

code that wasn't developed for this course, or modify code found from the AI Fairness 360 Open Source Toolkit to work with your dataset…" Does "you are allowed to code up your own algorithm" mean we can leverage any package we like (e.g. scikit-learn)?
Yes, feel free to use or modify algorithms according to your needs. Please mention it in your report.

For the classifier that we train, should the independent variables include the protected class variables only, or could they include all the independent variables in the original dataset?

You should include relevant features that your classifier needs to give a good result.

What is the table with changes supposed to look like in Step 4?
Below is a valid example (feel free to combine into one table):

Sex (Independent) to Approved (Dependent) Fairness Metrics

	Disparate Impact	Change compared to previous
Original Dataset	1.00200	NA
After Transforming Dataset	1.00200	No change
After Training Classifier on Original Dataset	1.00125	No change / very minimal positive change
After Training Classifier on Transformed Dataset	1.00173	No change / very minimal negative change

	Statistical Parity Difference	Change compared to previous
Original Dataset	0.00199	NA
After Transforming Dataset	0.00199	No change
After Training Classifier on Original Dataset	0.00125	No change / very minimal positive change
After Training Classifier on Transformed Dataset	0.00172	No change / very minimal negative change

Step 5

Are we required to generate all graphs and figures strictly using Python?
Students are encouraged to use Python, but other software such as Excel can be used to generate the graphs.