Introduction

* Overview Bias in Machine Learning
  + Intervention Types
  + Pre-Processing, In-Processing, Post-Processing

Main Topics

* GANs Overview
* Adversarial Network De-biasing Strategy
* Employee Attrition Overview

Problem

* Predict Attrition with data set containing age
* Design an Adversarial Network Structure to do this
* Reduce bias between age groups, while minimizing impact on accuracy

Results

* It worked
* Did not trade too much accuracy for fairness
* Brought groups more into alignment

Conclusion

* Explore Fairness in relation to business problem/context
* Mindful of each stage, bias can be – reintroduced or amplified
* Don’t rest on laurels – these techniques do work, but still requires human perspective

Visuals

* 7-group bar chart/histogram
* Histogram of Attrition % (??)
* Summary tables (accuracy and demographic parity)
* Simple summary of pre-process, in-process, post-process