## Introduction

Risk assessment algorithms and biased results are not pleasant combination. In this paper we revisit the famous ‘Machine Bias’ article published by ProPublica in May 2016 about how the COMPAS systems used for pre-trial risk assessment is bias towards a race. Our work questions and corrects rudimentary assumption ProPublica made which turns out to be an error. We also make an effort to build a fair assessment score for the given data only based on criminal activities, eliminating the race or any racial effect in the outcome.

## Literature review

United States has the world’s highest incarceration rate of 655 per 100,000 of population as per the world prison brief data. To control this rate and inflow of the prisoners, states of New York, Wisconsin, California, Florida and few other jurisdictions uses **COMPAS** (developed and owned by Northpointe), an acronym for *Correctional Offender Management Profiling for Alternative Sanctions* which provides certain risk score for an offender to access its likelihood of being a recidivist. These scores are used by US courts to grant the pre-trial release too.

ProPublica, an American non-profit newsroom based out of New York, presented a study in May 2016 and claimed that COMPAS generated ‘Risk Scores’ are biased against African-American race group. For this study they looked at more than 10,000 criminal defendants in Broward County, Florida, and compared their predicted recidivism rates with the rate that actually occurred over a two-year period. To do so, they also collected data about public incarceration from the ‘Florida Department of Corrections’. By joining these two data sets with a defendant’s first name, last name and Date of birth ProPublica assess about 11000 records.

Their analysis was at fault in terms of sampling error and terminologies as they swapped the cause-effect which inflated the result. The article ‘*False Positives, False Negatives, and False Analyses’ : A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.”,* stateshow ProPublica has failed to test the bias as per standard and used faulty statistical references and benchmarks to produce the result.

Moreover recently, ProPublica’s COMPAS data was revisited by Matias Barenstein, published on July 08, 2019. He highlights a classic data sampling mistakes made by ProPublica. In which ProPublica included extra 40% (1000) recidivists in the analysis which inflated the results of racial bias by 24%. In order to check for recidivism, ProPublica should have used a cut off on the COMPAS screening date (at 1st April 2014). This would ensure that these people are observed for next two years, as the database contains data through March 2016.

## Data

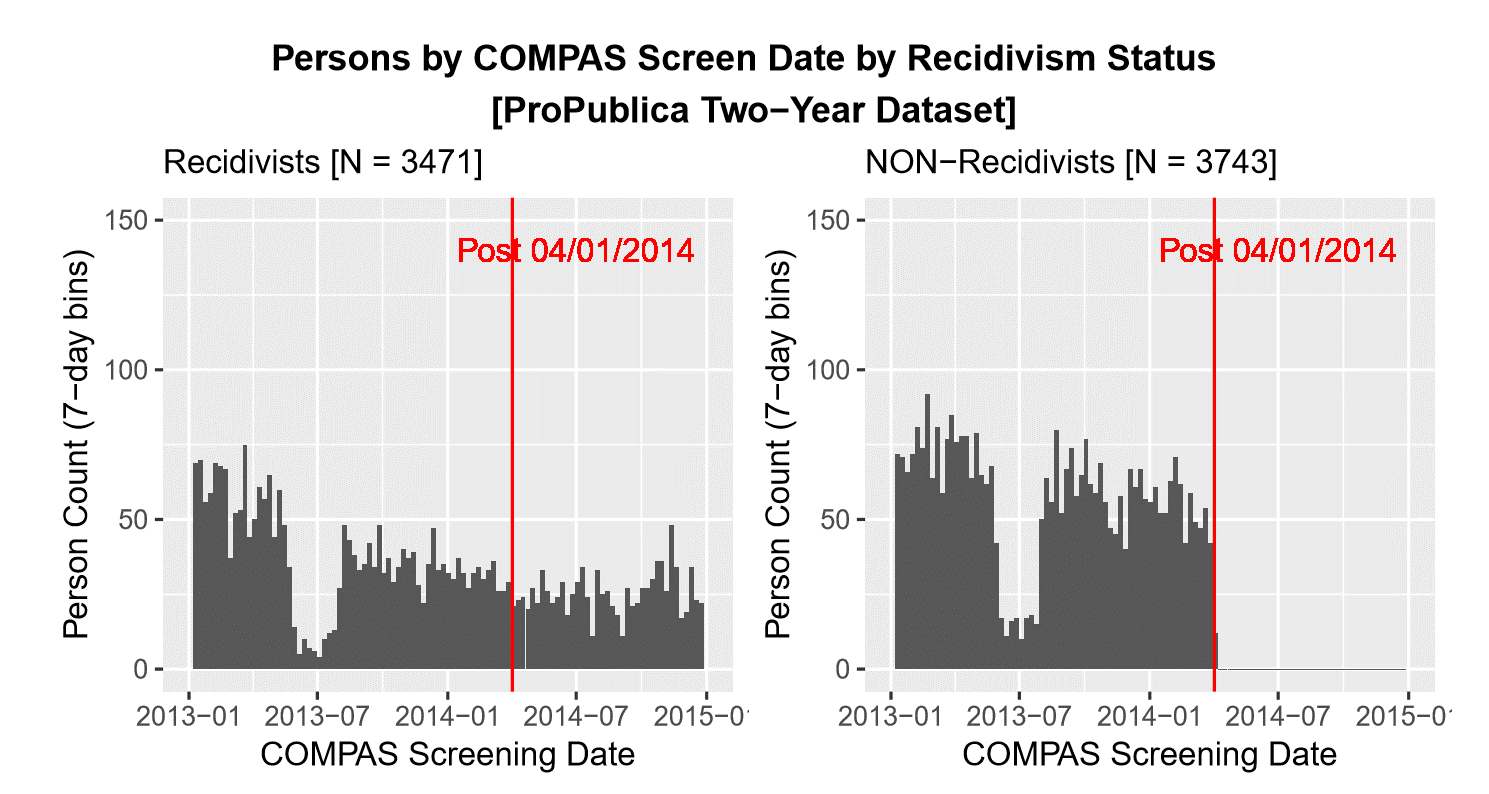
## Objective

We studied and recreated ProPublica’s analysis and based on that study, we defined 3 modes of analysis:

* To study the marginal effects of available attributes in data on the decile score (risk score) generated using COMPAS which is used by Courts to tag an offender with ‘low’, ‘medium’, or ‘high’ risk of recidivism
* To build ‘fair score’ from the available data that only concerns to the criminal history and current crime description of an offender and no other social factors, and compare its variation with original ‘decile score’
* To compare the predictive power and error rate of ‘decile score’ and ‘fair decile score’ for recidivism

## Method

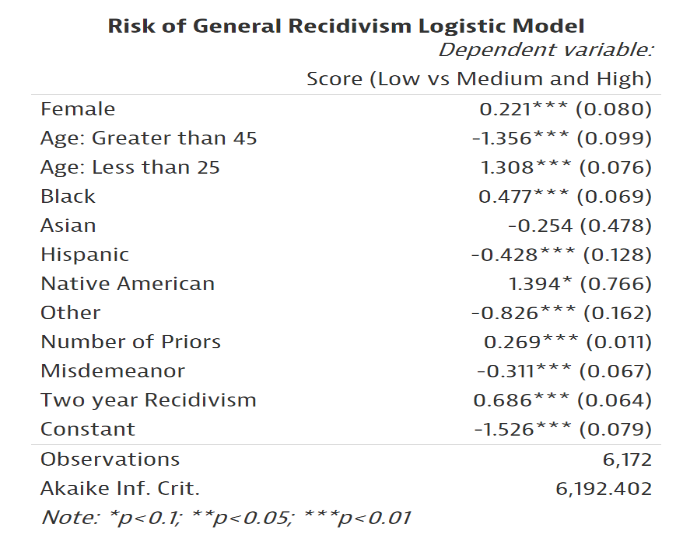
ProPublica provided a copy of database they worked with on their GitHub repository. They had acquired this dataset from Broward County, FL. It includes the pretrial defendant’s data from January 2013 to December 2014. They also got the data from Broward county’s Sherriff’s office and federal jail to observe these defendants for recidivism till April 2016. We sourced the database copy and extracted various tables which had information collected from Broward county COMPAS system.

Firstly, we replicated the data pre-processing and model built by ProPublica. We notice there are 2 errors in analysis done by ProPublica. A data dictionary is provided in annexure.

**1. Sampling Bias**

ProPublica selected just the recidivist population (over 1000 more records) but removed the corresponding non- recidivist population for given timeframe. This made the sample used for analysis bias by having more recidivist population and hence the results were bound to inflate.

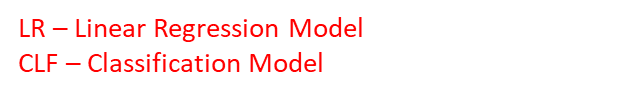
Figure 1.Segregation of defendants as recidivist and non-recidivist across compass screening date



**2. Cause-Effect mix up**

ProPublica collected data from Broward county’s Sherriff’s office so they could know if the defendant has committed and arrested again. This information was stored under ‘Two-year Recidivism’ variable in the data. This information was used from data in order to predict the of risk score for defendants. This data was not available while generating risk scores for any defendant in first place, so logically cannot be used in analysis. The act of recidivism (Effect) is used to predict the cause of recidivism (Cause).

Figure 2 A snippet from ProPublica's analysis indicating usage of Two\_year\_recidivism for analysis of Decile score



## Data pre-processing and Feature Engineering:

We used RStudio to extract data from database copy. Primarily we worked on *Compas\_system* and *charge\_degree* table. We joined these two tables by a unique key made by first name, last name, and date of birth of the criminal. The charge degree was following below distribution and were ordered as M3 (misdemeanor level 3) being the least and felony level 1 being the most heneous crime

Combined data and retained required features

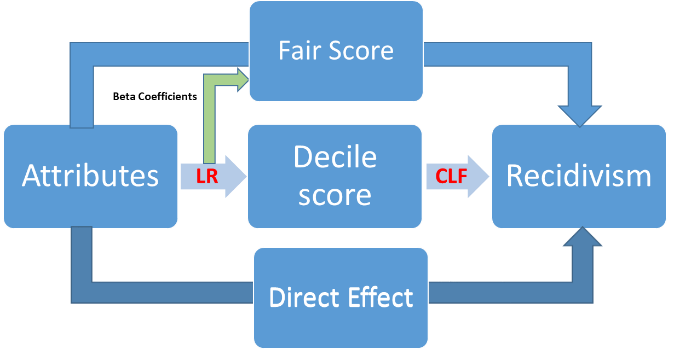
Computed Length of Stay in jail from jain\_in and jail\_out variables

Extracted drug\_involvement flag from charge description

Refactored race, charge degree, to a different

Removed records with unidentified charge degrees

## Our Analysis:

As stated in the objective, our analysis was aimed to build a fair risk score which is not racially bias and has which is also a good predictor of act of recidivism.

## Building Fair Risk Score

Since the claim is that the decile score provided by COMPAS system is biased towards African-American population, we intended to build a fair score which would not account for any racial factor. So, we studied the individual and combined effects of attributes on decile score.

Figure 3 Methodology used for building fair score and predicting act of recidivism

1. Decile score as a function of crime and demographic factors

*Decile score =*

1. Decile score as a function of crime and demographic factors with sex and race interaction

*Decile score =*

1. Decile score as function of crime related factors

*Decile score = )*

1. Decile score as a function of just Race and sex interaction term

*Decile score =*

Meanwhile we defined fair score as ‘A risk score which only accounts for the crime related attributes and activities of a

## Building fair\_score using crime related factors

As per the literature review and research done, our hypothesis for an ideal ‘Fair Risk Score’ for a criminal should only concern his/her past and current criminal records and risk score should be independent of socio-economic and demographic factors.

In the above section where we analyzed various combination of the decile score, we particularly chose one which just related to the offender’s criminal data available with us. i.e. *priors count, juv\_fel\_count, juv\_misd\_count, charge\_degree.*  We computed the Beta coefficients for this

Data Acquisition 🡪 Problem and analysis process

1. Excluded “Two-year recidivism” from the model. It is a feature used by ProPublica which is a future variable. It indicates if the arrested person committed Crime in 2 years after COMPAS screening. It will not be available at the time of screening. So, we are not including the same in our risk prediction model.
2. Excluded “Crime factor” from the model and used another variable which defined the severity of the crime more precisely. Crime factor is a feature used by which has two distinct values F(Felony) and Misdemeanour(M). Instead of that, we have used charge degree feature which includes different levels of felony and misdemeanours, like F1, F2, etc.
3. Created a binary variable to find the defendant’s involvement in substance abuse. Used charge description to find if the defendant was charged for any drug possession or use. We have used this feature in prediction of “risk of recidivism” since studies show that people involved in substance abuse tend to re-offend to either fulfil their drug requirements or under the influence of drugs.

**Results**

**Conclusion**

I Introduction:

Importance, goals, overview, literature review.

M Methods

Scope of your methods (test1, 2, )

R Results

Results for each method

If you were able to achieve the goals?

D Discussion

Contribution to existing research

Significance of each result

C Conclusion

Summary of objectives

Major findings

Implications of finding