Review of our project

Our project aims to help the prison to find out which prisoners are ready to go back to their communities with relatively low risk. The current COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system, which is designed to determine the risk of further crimes of prisoners and give them corresponding terms of imprisonment, seems to have a huge bias towards people of color and minorities. We want to check this hypothesis and determine if it is true based on our research.

Dataset details and preliminary analysis

For the compas-scores-two-years (53 columns) and compas-scores-two-years-violent (54 columns) datasets (the former includes COMPAS general crime scores for criminals and the latter includes violent scores), there are many columns we don't need, e.g., name, first, last, date of birth etc. We only need to choose those columns that may affect the final scores.

Preprocessing

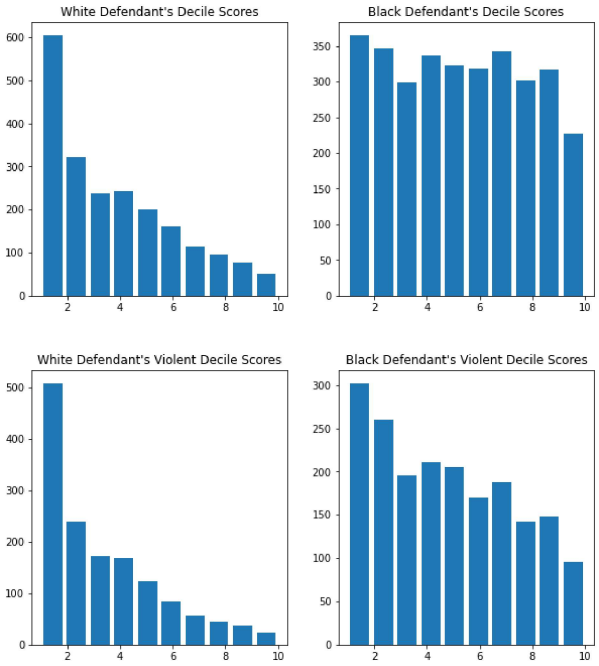
For the first few analysis, based on common sense, we pick **age**, **crime charge degree**, **race**, **age category**, **score as text**, **sex**, **priors count of crimes**, **time stay in jail**, **decile score**, and **recidivism** as our variables.

1. For "compas-scores-two-years" with specific columns described above, there are **921** rows with null in **7214** rows. After picking out those rows with NaN, we found most NaN from score_text col. Also, we remove rows in which charge date of a defendants COMPAS scored crime not within 30 days starting from the arrested day of each criminal; rows with re_recid = -1, which means no COMPAS case at all; rows of ordinary traffic offenses -- those with a c_charge_degree of 'O' -- will not result in Jail time. After doing so, we have datasets without NaN with shape **(6172, 13)**.
2. For "compas-scores-two-years-violent", we use same method to remove null rows and end up with a datasets of shape **(4015, 13)**.

```
African-American    0.477114
Caucasian           0.362935
Hispanic            0.088308
Other               0.063433
Asian               0.006468
Native American     0.001741
Name: race, dtype: float64
```

We can see most of prisoners are *Caucasian* and *African-American*, thus we will mainly focusing on difference between scores of these two groups.

Histograms based on our datasets



As we can see from above, both datasets show strong evidence that white defendents tend to have lower score compare with their counterpart. To test racial disparities, we applied a logistic regression model to quantify.

Logistic regression analysis

We apply a logistics regression on the data with High/Low score_text as Y and other variables as Xs. We choose this algorithm because the output variable is binary and there is no obvious relationship between other variables.

Since this is only an test of disparity, we pick only 5 columns for our model, i.e., **race**, **sex**, **age_cat**, **is_recid**, and **c_charge_degree**.

Below is the results for both datasets.

Logit Regression Results

Dep. Variable:	y_bias	**No. Observations:**	3821
Model:	Logit	**Df Residuals:**	3814
Method:	MLE	**Df Model:**	6
Date:	Mon, 01 Nov 2021	**Pseudo R-squ.:**	0.2212
Time:	16:02:05	**Log-Likelihood:**	-1763.2
converged:	True	**LL-Null:**	-2263.9
Covariance Type:	nonrobust	**LLR p-value:**	4.548e-213

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3241	0.125	-10.616	0.000	-1.569	-1.080
race[T.Caucasian]	-1.0834	0.092	-11.730	0.000	-1.264	-0.902
sex[T.Male]	0.1348	0.114	1.178	0.239	-0.089	0.359
age_cat[T.Greater than 45]	-1.0813	0.133	-8.140	0.000	-1.342	-0.821
age_cat[T.Less than 25]	0.5824	0.098	5.943	0.000	0.390	0.774
is_recid[T.1]	1.6158	0.088	18.323	0.000	1.443	1.789
c_charge_degree[T.M]	-0.7448	0.093	-8.013	0.000	-0.927	-0.563

Logit Regression Results

Dep. Variable:	y_bias	**No. Observations:**	3821
Model:	Logit	**Df Residuals:**	3814
Method:	MLE	**Df Model:**	6
Date:	Mon, 01 Nov 2021	**Pseudo R-squ.:**	0.2212
Time:	16:02:05	**Log-Likelihood:**	-1763.2
converged:	True	**LL-Null:**	-2263.9
Covariance Type:	nonrobust	**LLR p-value:**	4.548e-213

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.3241	0.125	-10.616	0.000	-1.569	-1.080
race[T.Caucasian]	-1.0834	0.092	-11.730	0.000	-1.264	-0.902
sex[T.Male]	0.1348	0.114	1.178	0.239	-0.089	0.359
age_cat[T.Greater than 45]	-1.0813	0.133	-8.140	0.000	-1.342	-0.821
age_cat[T.Less than 25]	0.5824	0.098	5.943	0.000	0.390	0.774
is_recid[T.1]	1.6158	0.088	18.323	0.000	1.443	1.789
c_charge_degree[T.M]	-0.7448	0.093	-8.013	0.000	-0.927	-0.563

We use the coefficients of intercept and Caucasian to compute the possibility of Caucasian to be marked 'High' are 60% and 63% less likely as African-American defendents being marked in these two datasets.

FN and FP rate

Intuitively, there do exist racial disparity among these two groups. But we need to check the FN and FP rate for the test.

For this test, we remove rows that recidivsm information less than 2 years. The remaining datasets has shape **(7214, 53)**.

All Defendants		
	Low	High
Survived	2681	1282
Recidivated	1216	2035
FP rate: 32.35		
FN rate: 37.40		
PPV: 0.61		
NPV: 0.69		
LR+: 1.94		
LR-: 0.55		

Black Defendants		
	Low	High
Survived	990	805
Recidivated	532	1369
FP rate: 44.85		
FN rate: 27.99		
PPV: 0.63		
NPV: 0.65		
LR+: 1.61		
LR-: 0.51		

White Defendants		
	Low	High
Survived	1139	349
Recidivated	461	505
FP rate: 23.45		
FN rate: 47.72		
PPV: 0.59		
NPV: 0.71		
LR+: 2.23		
LR-: 0.62		

As we can see from above, the FN and FP rate indicate that the COMPAS algorithm tends to have bias issue wrt race.

What we will do in future

We plan to redo the logistic regression by using feature selection method we have learnt such as lasso and ridge. Also, we will apply bagging method on tree model to see if it will generate results without racial bias. Further, we planned to fit some other models we are interested in.