Data is from compas-scores-two-years.csv
7214 individuals, 53 columns containing both numerical and categorical information. Columns are explained below along with some basic statistics about their contents.

The COMPAS score is either low (1-3), medium (4-6), or high (7-10). According to the ProPublica article, both medium and high are predictive of recidivism. When performing our analysis we could predict arrest, violent recidivism, misdemeanor, felony, high risk of violence, and high risk of arrest.

Throughout the data set, the following abbreviations are used.
c stands for COMPAS and refers to the initial offense when the COMPAS score was calculated.
r stands for recidivism and refers to the next offense after the COMPAS score was calculated.
v stands for violent in order to differentiate violent crimes from non-violent ones (vr is therefore violent recidivism) because there are separate predictions for violent crimes.

Categorical columns:
'name': Name
 'first': First name
 'last': Last name
 'compas_screening_date',
 'sex': Male or Female
 'dob': Date of birth
 'age_cat': Greater than 45, 25-45, Less than 25
 'race': Other, Black, White, Hispanic, Native American, Asian
 'c_jail_in': Date that they checked into jail the first time (when the COMPAS score was calculated)
 'c_jail_out': Date checking out of jail
 'c_case_number': 13 digit alphanumeric string referring to the case where the COMPAS score was calculated
 'c_offense_date': Date of COMPAS scoring offense
 'c_arrest_date': Date of COMPAS scoring arrest
 'c_charge_degree': Felony or Misdemeanor for COMPAS offense
 'c_charge_desc': Description of the COMPAS offense charges, e.g. Grand Theft in the 3rd Degree
 'r_case_number': 13 digit alphanumeric string referring to the recidivist case
 'r_charge_degree': Felony or Misdemeanor for recidivist offense
 'r_offense_date': Date of recidivist offense
 'r_charge_desc': Description of the recidivist offense
 'r_jail_in': Date for checking into jail for recidivist offense
 'r_jail_out': Date for checking out of jail for recidivist offense
 'vr_case_number': Violent recidivist case

'vr_charge_degree',
'vr_offense_date',
'vr_charge_desc',
'type_of_assessment': Risk of recidivism
'score_text': Low, medium, or high
'screening_date',
'v_type_of_assessment': Risk of violence
'v_score_text',
'v_screening_date',
'in_custody',
'out_custody':

	name	first	last	compas_scre	sex	dob	age_cat	race	c_jail_in	c_jail_out
count	7214	7214	7214	7214	7214	7214	7214	7214	6907	6907
unique	7158	2800	3950	690	2	5452	3	6	6907	6880
top	anthony smit	michael	williams	3/20/13	Male	8/31/89	25 - 45	African-Ame	6/22/13 7:50	9/28/13 2:10
freq	3	149	83	32	5819	5	4109	3696	1	3

	c_case_num	c_offense_d	c_arrest_dat	c_charge_de	c_charge_de	r_case_num	r_charge_de	r_offense_da	r_charge_de	r_jail_in
count	7192	6055	1137	7214	7185	3471	3471	3471	3413	2316
unique	7192	927	580	2	437	3471	10	1075	340	972
top	13000025MI	1/14/13	2/6/13	F	Battery	13023854MI	(M1)	12/8/14	Driving Licen	5/27/14
freq	1	26	9	4666	1156	1	1201	12	258	9

	r_jail_out	vr_case_num	vr_charge_d	vr_offense_c	vr_charge_d	type_of_asse	score_text	screening_da	v_type_of_a	v_score_text	v_screening	in_custody	out_custody
count	2316	819	819	819	819	7214	7214	7214	7214	7214	7214	6978	6978
unique	938	819	9	570	83	1	3	690	1	3	690	1156	1169
top	2/18/14	14010274CF	(M1)		8/15/15	Battery	Risk of Recid	Low	3/20/13	Risk of Violer	Low	3/20/13	1/22/14	1/1/20
freq	9	1	344	6	329	7214	3897	32	7214	4761	32	20	61

Numerical columns:
'id': Id of the person
'age': Age
'juv_fel_count': Number of juvenile felonies
'decile_score': COMPAS score
'juv_misd_count': Number of juvenile misdemeanors
'juv_other_count': Number of other juvenile crimes
'priors_count': Number of prior arrests
'days_b_screening_arrest',
'c_days_from_compas',
'is_recid': Whether or not the person was arrested again within 2 years
'r_days_from_arrest': How long until the recidivist arrest
'violent_recid': Not used
'is_violent_recid': Whether or not the recidivism was violent
'decile_score.1': Same as decile_score
'v_decile_score': COMPAS score for judging violent offenses
'priors_count.1',
'start',
'end',
'event',
'two_year_recid': Slightly different but similar to is_recid??? Not sure exactly what

	id	age	juv_fel_coun	decile_score	juv_misd_co	juv_other_cc	priors_count	days_b_scre	c_days_from	is_recid
count	7214	7214	7214	7214	7214	7214	7214	6907	7192	7214
mean	5501.25575	34.8179928	0.06723039	4.50956474	0.09093429	0.10937067	3.47241475	3.30476328	57.7313682	0.48114777
std	3175.70687	11.8889218	0.47397152	2.85639568	0.48523851	0.5015858	4.88253835	75.8095048	329.740215	0.4996791
min	1	18	0	1	0	0	0	-414	0	0
25%	2735.25	25	0	2	0	0	0	-1	1	0
50%	5509.5	31	0	4	0	0	2	-1	1	0
75%	8246.5	42	0	7	0	0	5	0	2	1
max	11001	96	20	10	13	17	38	1057	9485	1

	r_days_from	violent_recid	is_violent_re	decile_score	v_decile_sco	priors_count	start	end	event	two_year_recid
count	2316	0	7214	7214	7214	7214	7214	7214	7214	7214
mean	20.2694301		0.11352925	4.50956474	3.69184918	3.47241475	11.4650679	553.436651	0.38286665	0.45065151
std	74.8716681		0.31726064	2.85639568	2.51014801	4.88253835	46.9545631	399.020583	0.48611988	0.49759326
min	-1		0	1	1	0	0	0	0	0
25%	0		0	2	1	0	0	148.25	0	0
50%	0		0	4	3	2	0	530.5	0	0
75%	1		0	7	5	5	1	914	1	1
max	993		1	10	10	38	937	1186	1	1

ProPublica used these general factors in their logistic regression model.

Risk of General Recidivism Logistic Model

Dependent variable:

Score (Low vs Medium and High)

Female	0.221*** (0.080)
Age: Greater than 45	-1.356*** (0.099)
Age: Less than 25	1.308*** (0.076)
Black	0.477*** (0.069)
Asian	-0.254 (0.478)
Hispanic	-0.428*** (0.128)
Native American	1.394* (0.766)
Other	-0.826*** (0.162)
Number of Priors	0.269*** (0.011)
Misdemeanor	-0.311*** (0.067)
Two year Recidivism	0.686*** (0.064)
Constant	-1.526*** (0.079)
Observations	6,172
Akaike Inf. Crit.	6,192.402

Note: *p<0.1; **p<0.05; ***p<0.01

Proposed binarized categories:
Is male
Age: <=18, 18-22, <=20, <=22, <=25, 24-30, 24-40, >=30, >=40
Black
Asian
Native_american

Other
juvenile felonies - 0, 1-3, >=3, >=5
total juvenile crimes - same thing
total past crimes - 0 , 1-3, >=3, >=5, >=10
total violent crimes...
age first jail time - <=18, <=24, 18-24, >=25, >=40