Homework 4 Report

Jae Young Kim

Applied Data Science, University of Southern California,

Los Angeles, California 90089, USA

(Dated: March 25, 2021)

Abstract

Data of Ethnic group concentration, education level, income, valence of the area is given. To make an unbiased model predicting valence with other predictor features, models were built. Through this research, four models were built, Ground Truth Model, Ethnic Group Aware Model, Ethnic Group Blind Model and Fair Model. The original Ground Truth model was not biased. However, bias is found on Ethnic Group Aware Model and Ethnic Group Blind Model. After building fair model with given functions, bias is not found on the Fair Model.

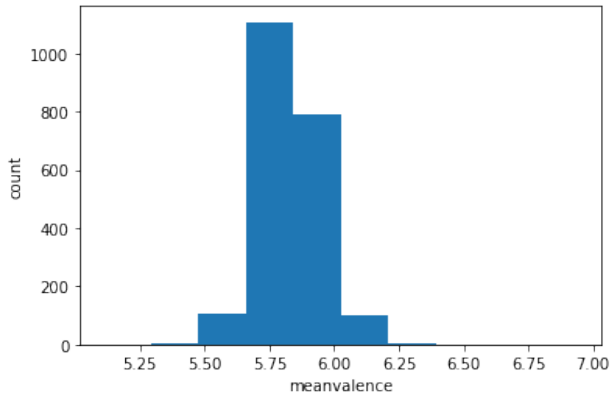


FIG. 1. Histogram of Meanvalence

I. INTRODUCTION

Through this research, four models were built on the chirper-happiness data set. The data set contains ethnic group concentration (totalGroup1 and totalGroup2), education level (percent bachelorPlus), income (households meanIncome) and target feature, happiness (meanvalence). The four models built were Ground Truth Model, Ethnic Group Aware Model, Ethnic Group Blind Model and Fair Model. Bias was not found on the first model. However, after training linear regression model, bias were found on Ethnic Group Aware Model and Ethnic Group Blind Model. However, after applying helper functions from a paper on debiasing linear models(available at : Learning Fair and Interpretable Representations via Linear Orthogonalization), it was confirmed that the bias almost disappeared from the model.

II. DATA EXPLORATION AND PREROCESSING

To explore the data, histogram of target feature, "meanvalence" was plotted in Figure 1 and the mean value of 'meanvalence' is 5.83.

Before training model, standard scaler was applied to the data and column "Id2" was dropped since "Id2" is just the id of the data.

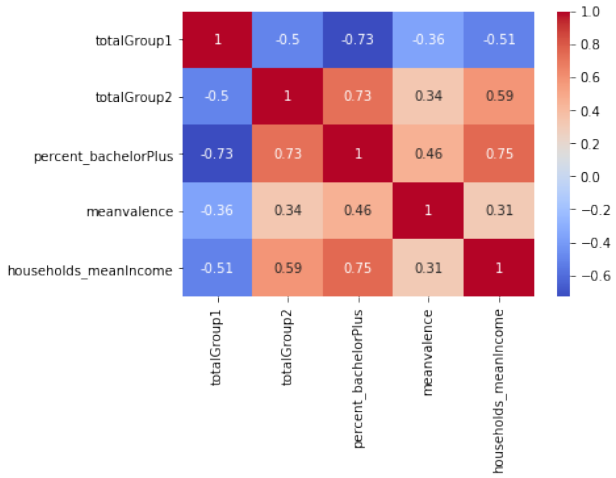


FIG. 2. Heat map of Correlation Matrix

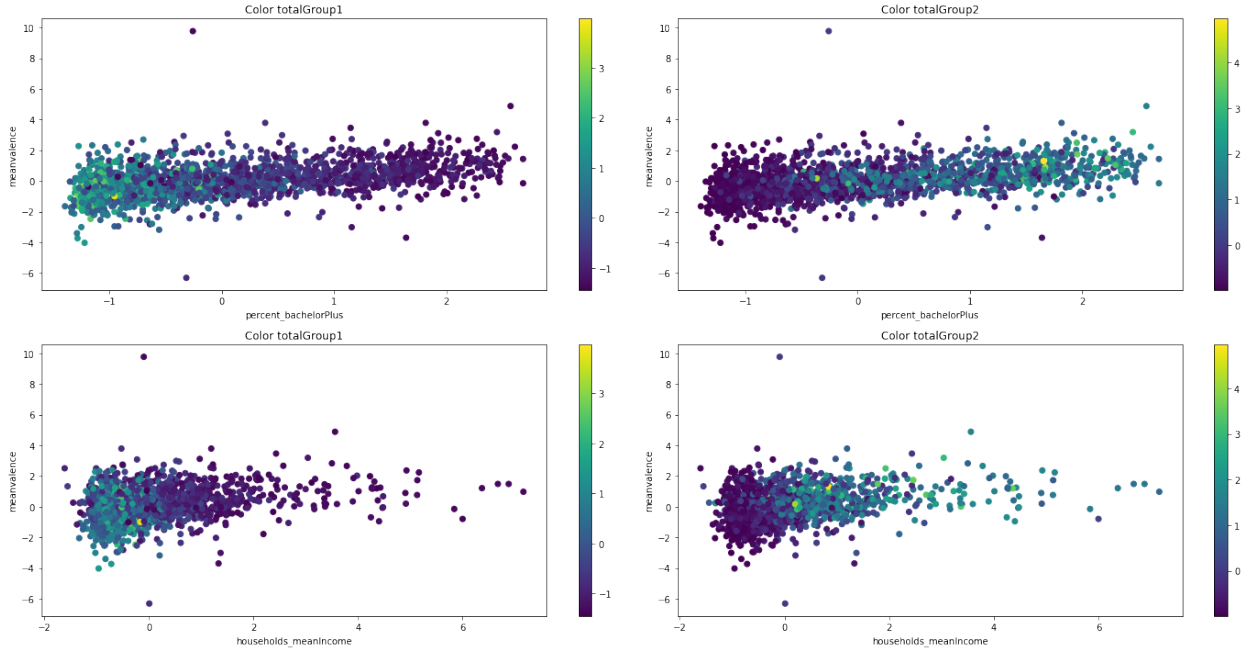
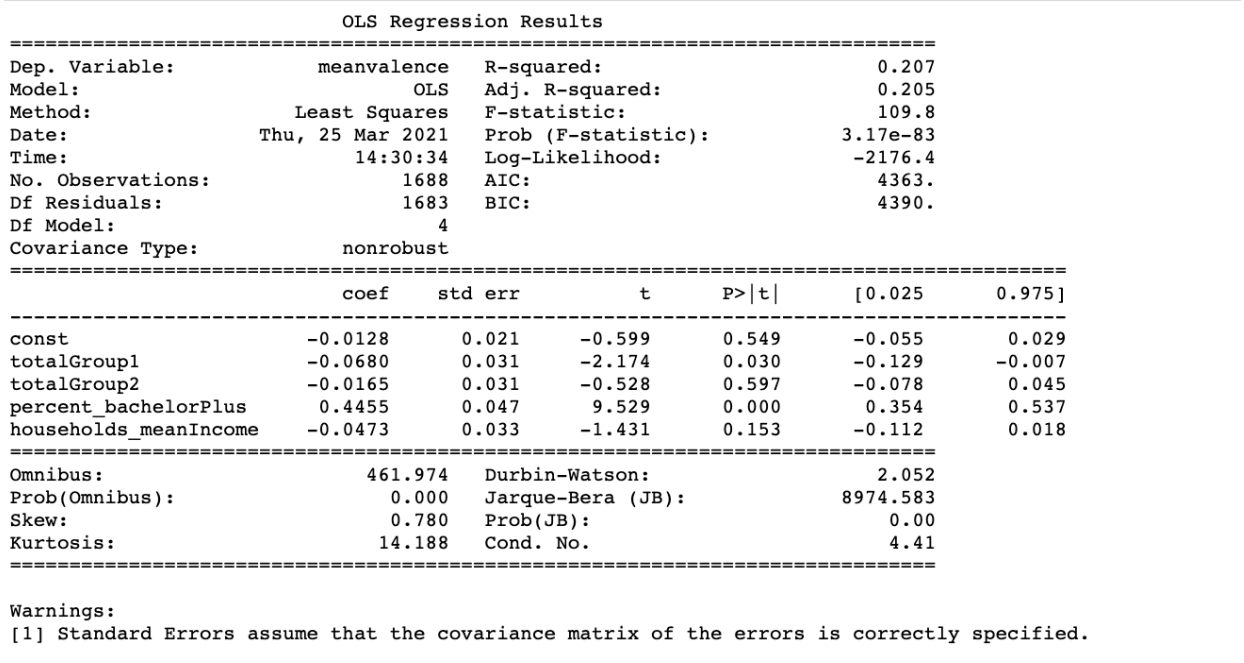


FIG. 3. Ground Truth model

III. GROUND TRUTH MODEL

Correlations and scatter plots of ground truth model are generated. From the heat map of correlation matrix in Figure2, the correlation of totalGroup1 is negative to other features while other features have positive correlation to each other. This trend is visualized in Figure3. The color distribution of left side plots were opposite from the right side plots. Left plots were colored base on totalGroup1 and right plots were colored base on totalGroup2.



```
                            OLS Regression Results
==============================================================================
Dep. Variable:            meanvalence   R-squared:                       0.207
Model:                            OLS   Adj. R-squared:                  0.205
Method:                 Least Squares   F-statistic:                     109.8
Date:                Thu, 25 Mar 2021   Prob (F-statistic):           3.17e-83
Time:                        14:30:34   Log-Likelihood:                -2176.4
No. Observations:                1688   AIC:                             4363.
Df Residuals:                    1683   BIC:                             4390.
Df Model:                           4
Covariance Type:            nonrobust
============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                       -0.0128      0.021     -0.599      0.549      -0.055       0.029
totalGroup1                 -0.0680      0.031     -2.174      0.030      -0.129      -0.007
totalGroup2                 -0.0165      0.031     -0.528      0.597      -0.078       0.045
percent_bachelorPlus         0.4455      0.047      9.529      0.000       0.354       0.537
households_meanIncome       -0.0473      0.033     -1.431      0.153      -0.112       0.018
==============================================================================
Omnibus:                      461.974   Durbin-Watson:                   2.052
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8974.583
Skew:                           0.780   Prob(JB):                         0.00
Kurtosis:                      14.188   Cond. No.                         4.41
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

FIG. 4. Statistics of Ethnic Group Aware Model

Specifically, the scatter plots in Fig3 shows that totalGroup1 and totalGroup2 have opposite trend. As percent_bachelorPlus increases, number of Group1 people dropped while the number of Group2 people increased. Similarly, when households mean income increases, Group1 people decreases while Group2 people increases. This trend means that generally, the education level and income level of Ethnic group1 are lower than those of Ethnic group2. However, in the Ground Truth Model, bias is rarely found in Figure3. The plots showed horizontal symmetry.

To sum up, the Ground Truth Model showed relatively small bias in terms of "meanvalence". It can be checked from correlation heat map in figure2. Correlation of (totalGroup1, meanvalence) and (totalGroup2, meanvalence) were -0.36 and 0.34 for each and they are smaller than correlation of other features.

IV. ETHNIC GROUP AWARE MODEL

Ethnic Group Aware Model is built with all of the predictor variables including totalGroup1 and totalGroup2. Linear regression model is generated. As standard scaled meanvalence 5.8 is -0.25, the scatter plots were divided into two parts base on meanvalence = -0.25.

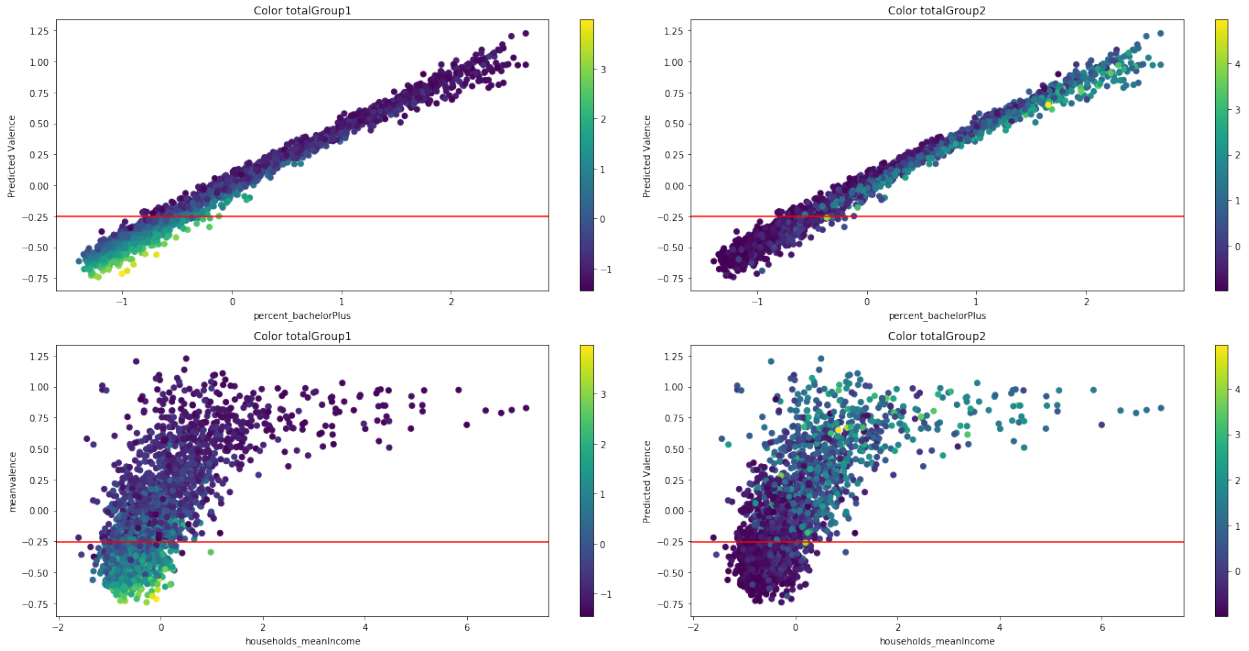


FIG. 5. Ethnic Group Aware Model

The statistics of linear regression model in Figure4 showed R-squared score 0.207 and pvalue showed that totalGroup1 has linear correlation with "meanvalence" but totalGroup2 does not have linear correlation with "meanvalence". P-value of "Households mean Income" was also higher than 0.05 and it means there isn't linear relation between "meanvalence" and "Households mean Income". However, the model showed that there is positive relation between "percent bachelorPlus" and "meanvalence".

From Figure 5, the Ethnic Group Aware Model shows obvious horizontal asymmetry compared to the Ground Truth Model. Based on the red line which means predicted Valence = -0.2, the color distribution , percent bachelor Plus, households mean income trends were all severely different. It means that there is bias on the model based on the two ethnic groups because the ethnic group distribution was different base on the mean valence.

From Figure 6, as the plots are divided, the difference of the trend based on the ethnic group can be seen more obviously. The shape and color concentration of the upper plots(predicted scaled meanvalence ¡ -0.2) and the shape of the lower plots(predicted scaled meanvalence ¿= -0.2) looks different.

Figure 7 shows the heat map of correlation between the features. "ypred" is predicted scaled meanvalence. The correlation of (totalGroup1, ypred) and (totalGroup2, ypred) are each -0.8, 0.69 and they are much higher than those of Figure 2(Ground Truth Model).

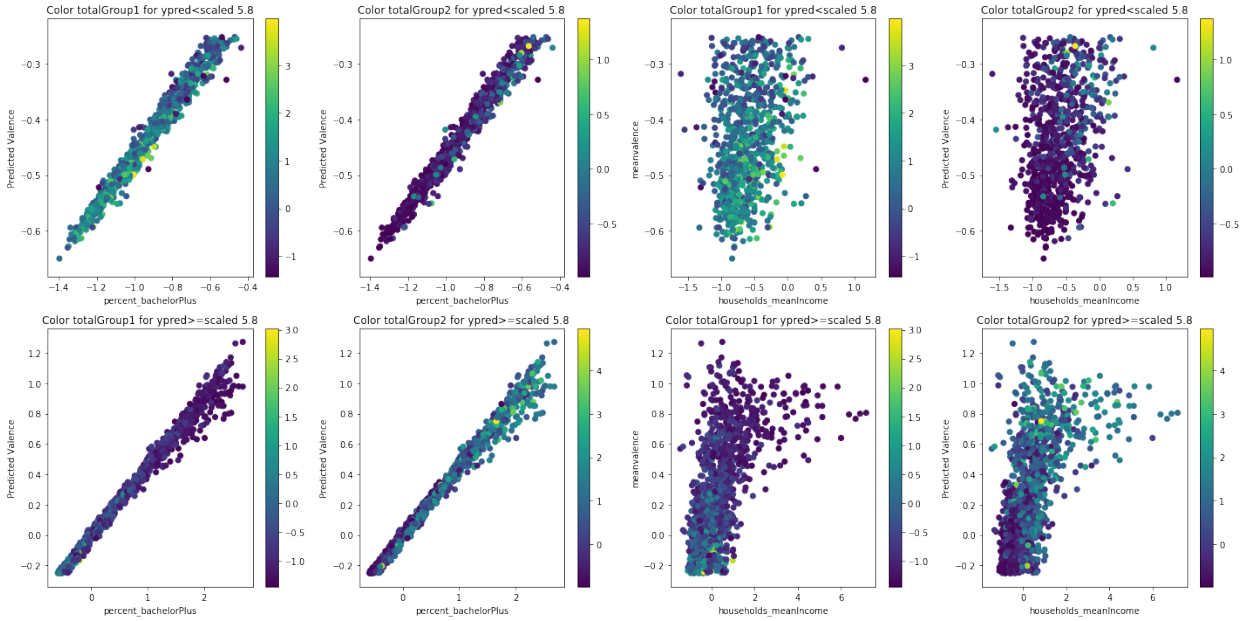


FIG. 6. Ethnic Group Aware Model split base on predicted scaled meanvalence -0.2

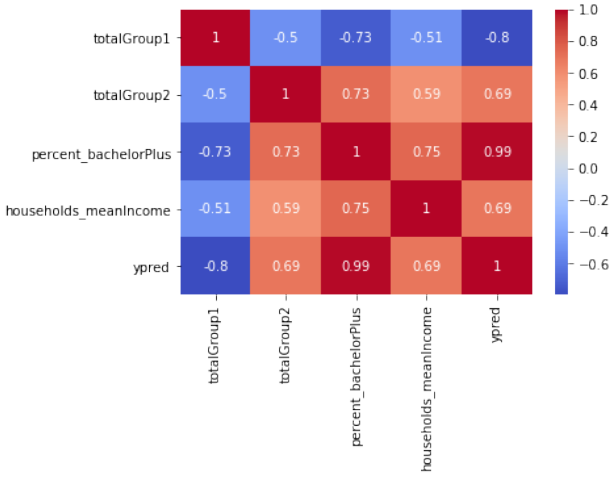
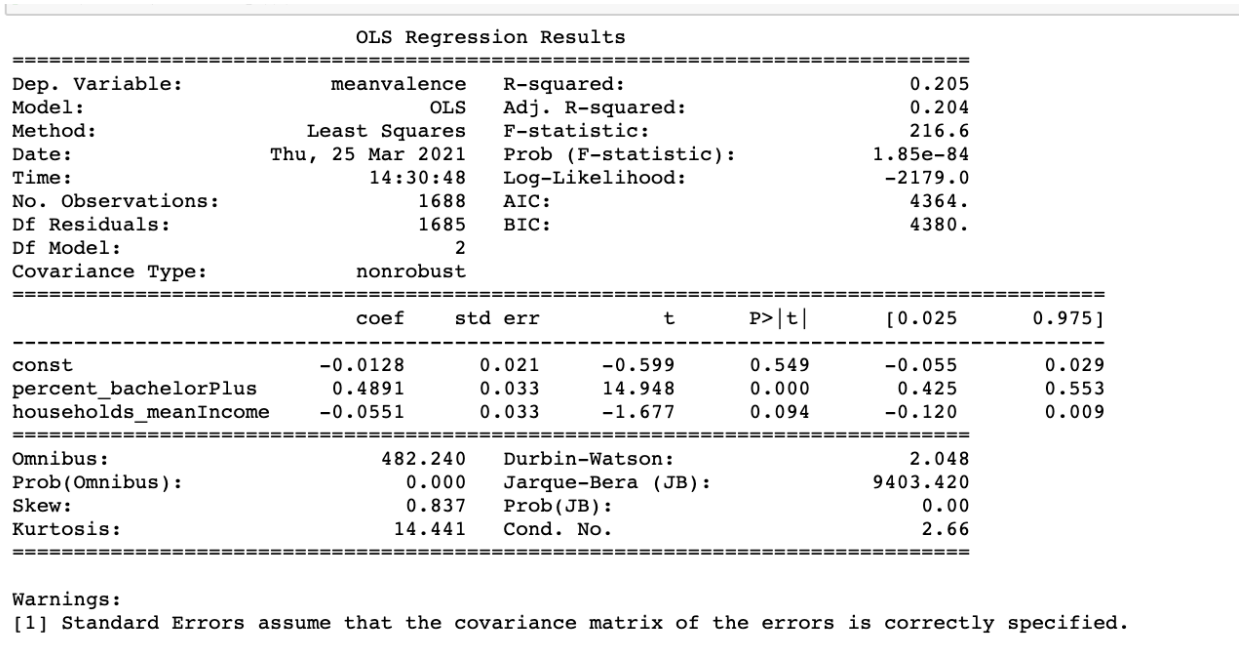


FIG. 7. Heat map of correlation matrix of Ethnic Group Aware Model

In Figure 2, (totalGroup1, meanvalence) and (totalGroup2, meanvalence) were -0.36 and 0.34 for each. This shows the linear relation between the target variable and Ethnic group features became stronger and the bias became stronger than Ground Truth Model.

To sum up, bias exists to both groups when the predicted valence threshold is set to 5.8. If predicted scaled mean valence is lower than -0.2, Group1 was concentrated and if predicted scaled mean valence is larger than -0.2, Group2 was concentrated.



OLS Regression Results

Dep. Variable:	meanvalence	R-squared:	0.205
Model:	OLS	Adj. R-squared:	0.204
Method:	Least Squares	F-statistic:	216.6
Date:	Thu, 25 Mar 2021	Prob (F-statistic):	1.85e-84
Time:	14:30:48	Log-Likelihood:	-2179.0
No. Observations:	1688	AIC:	4364.
Df Residuals:	1685	BIC:	4380.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0128	0.021	-0.599	0.549	-0.055	0.029
percent_bachelorPlus	0.4891	0.033	14.948	0.000	0.425	0.553
households_meanIncome	-0.0551	0.033	-1.677	0.094	-0.120	0.009

Omnibus:	482.240	Durbin-Watson:	2.048
Prob(Omnibus):	0.000	Jarque-Bera (JB):	9403.420
Skew:	0.837	Prob(JB):	0.00
Kurtosis:	14.441	Cond. No.	2.66

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

FIG. 8. Statistics of Ethnic Group Blind Model

V. ETHNIC GROUP BLIND MODEL

Ethnic Group Blind Model is built with all of the predictor variables without totalGroup1 and totalGroup2. Linear regression model is generated.

In Figure 8, the statistics of Ethnic Group Blind Model were shown and they were similar to the statistics of Ethnic Group Aware Model. The R-squared score was 0.205 and p-values showed that "percent bachelorPlus" has linear relationship with the predicted mean valence but, "households mean Income" showed there is no linear relationship with the predicted mean valence.

The scatter plots in Figure 9 also shows similar result to Ethnic Group Aware Model. Group1 concentrated area shows low predicted mean valence while Group2 concentrated area shows high predicted mean valence.

Compared to Ground Truth Model, Ethnic Group Blind Model shows stronger linear relationship with the predictor variables.

Figure 10 shows the heat map of correlation between the features. "ypred" is predicted scaled meanvalence. The correlation of (totalGroup1, ypred) and (totalGroup2, ypred) are each -0.73, 0.72 and they are much higher than those of Ground Truth Modle(Figure 2). In Figure 2, (totalGroup1, meanvalence) and (totalGroup2, meanvalence) were -0.36 and

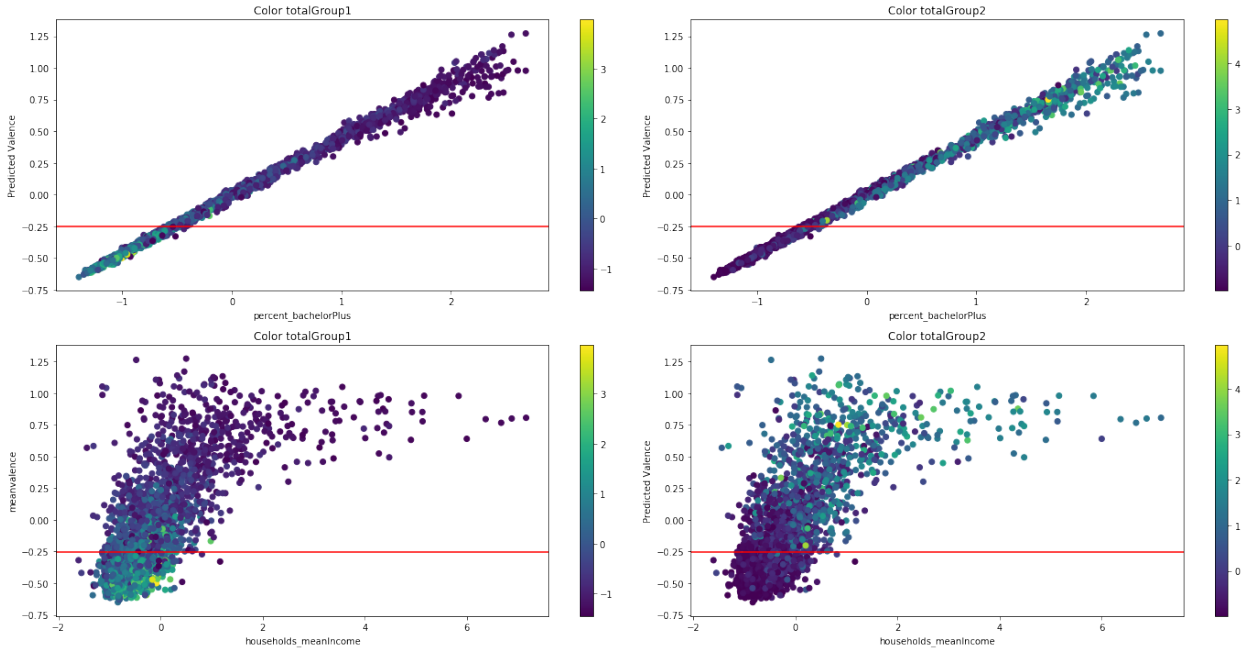


FIG. 9. Ethnic Group Blind Model

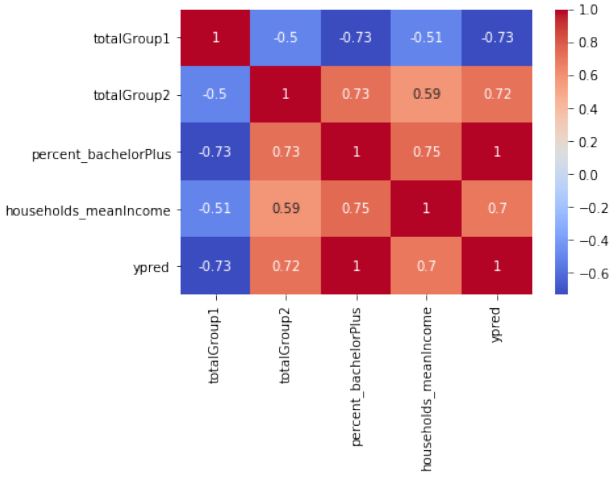


FIG. 10. Heat map of correlation matrix of Ethnic Group Blind Model

0.34 for each. This shows the linear relation between the target variable and Ethnic group features became stronger and the bias became stronger than Ground Truth Model.

To sum up, bias is still found on the Group Blind Model. Hiding protected variables(totalGroup1, totalGroup2) could not prevent the model from having the bias identified in Ethnic Group Aware Model.

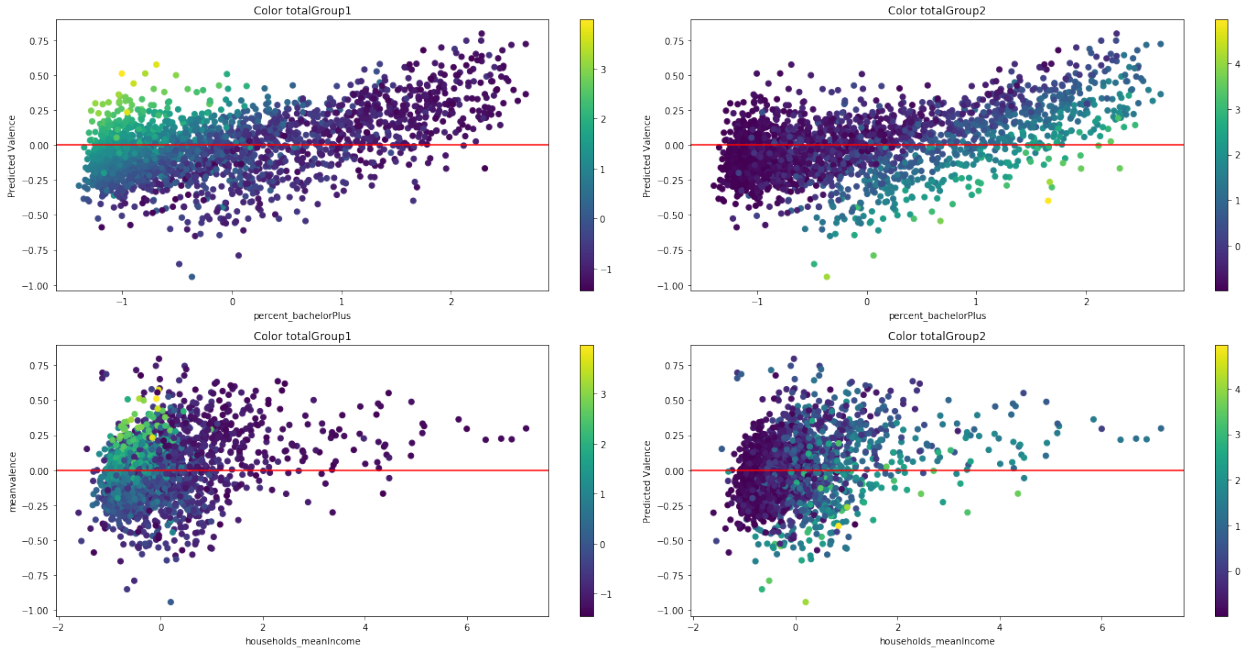


FIG. 11. Fair Model with lambda 0

VI. FAIR MODEL

To remove bias from Group Aware Model, helper function is applied to the target feature of Group Aware Model. Two versions were used. One version with lambda 0 and the other version with lambda 1. When lambda is 0, the function removes bias completely and when lambda is 1, the result is same as before.

In Figure 11, Fair model with lambda 0 is plotted. It shows that the bias is pretty much removed compared to the Group Aware Model. The color distribution shows horizontal symmetry compared to the plot of Group Aware Model in Figure 5.

On the other hand, in Figure 12, Fair model with lambda 1 shows same plot as Group Aware Model in Figure 5.

The heatmap in Figure 13 shows that the correlation of fair meanvalence with Ethnic group features are low. Compared to the previous correlation heat maps, the heat map in Figure 13 shows absolutely lower correlation of (totalGroup1, meanvalence fair) and (totalGroup2, meanvalence fair). In other words, ethnic group features has almost no linear correlation with meanvalence in the Fair model with lambda 0.

To sum up, the ethnic bias shown in the previous models were removed in the fair model with lambda 0.

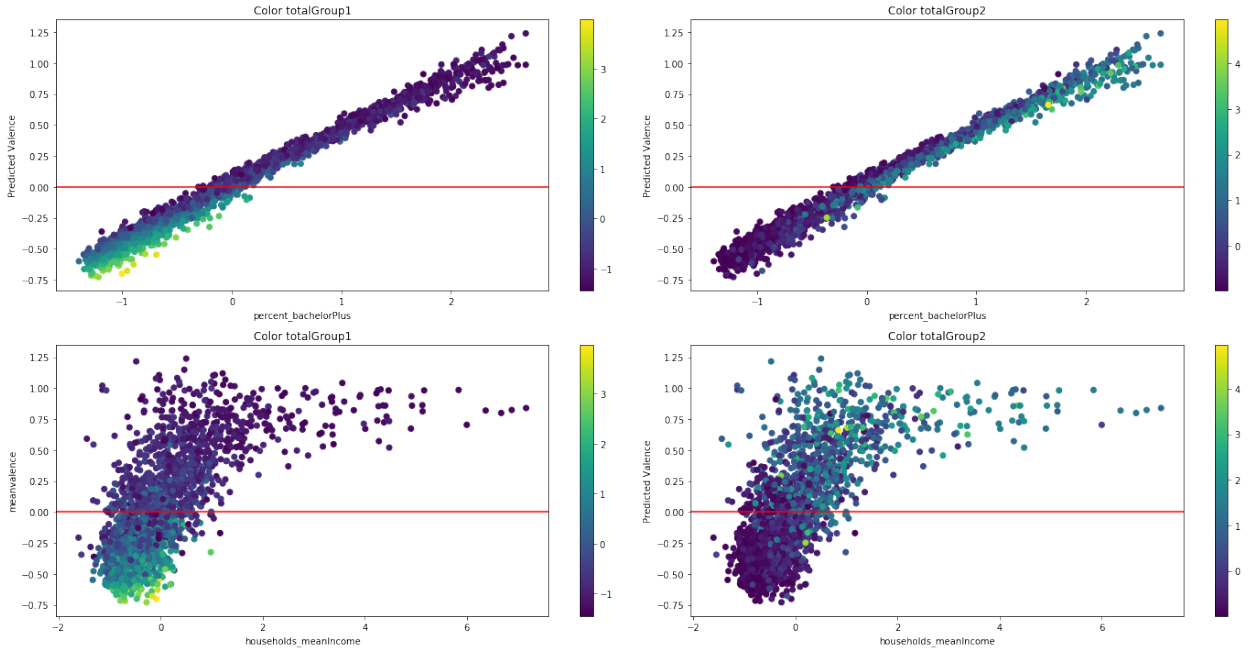


FIG. 12. Fair Model with lambda 1

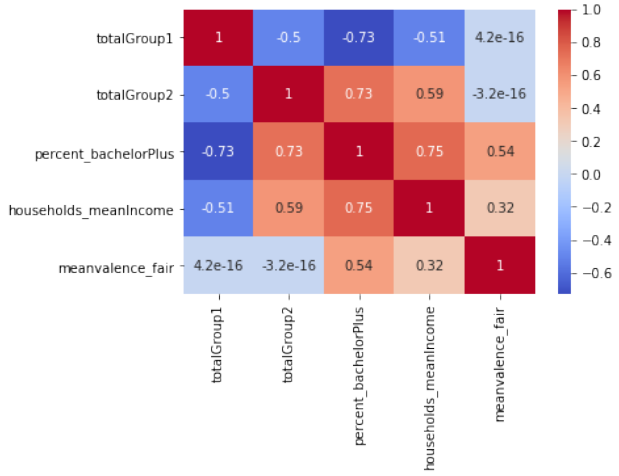


FIG. 13. Heatmap of the correlation of Fair Model with lambda 0

VII. CONCLUSION

To find the bias in the model and generate fair model on the chirper-happiness data, four models were generated. Even though bias was small in the Ground Truth model, high bias were found in Ethnic Group Aware model and Ethnic Group Blind model. The mean valence was different when the concentration of the ethnic group varies. Finally, fair model was built and through the model, bias is removed.

DATA AVAILABILITY

https://github.com/USC-DSCI-552-Spring2021/dsci552-spring2021-32416d-ps4-jeayoung114/data

CODE AVAILABILITY

Code is available at
https://github.com/USC-DSCI-552-Spring2021/dsci552-spring2021-32416d-ps3-jeayoung114/code/hw4