Predicting Happiness scores fairly for Chirper Inc.

Saurabh Jain

Department of Computer Science, University of Southern California,

Los Angeles, California 90089, USA

(Dated: March 25, 2021)

Abstract

The primary task of this assignment is to design a linear model that can fairly predict happiness and identify areas with average happiness less than 5.8. Happiness, also known as "valence" has been measured for a select few counties in the nation. The models will help Chirper Inc. in predicting valence and the impact of socio-demographic features on it. Correlation between the predicted outcome and independent predictors can be used as a proxy for measuring fairness. Our goal is to ensure that valence predictions are not biased towards one of the ethnic group variables. We define protected variables as ones along which biases can be observed in the outcome. In our case, these are – totalGroup1 and (or) totalGroup2. The analysis in this assignment consists of 4 different models namely: Ground Truth Model, Ethnic Group Aware Model, Ethnic Group Blind Model, and Fair Model.

It can be concluded that the model was not as biased for the ground truth model. Making Ethnic Group Aware Model and Ethnic Group Blind Model and comparing scatter plots and correlation heat maps, it can be said that these two models were biased equally towards both the ethnic groups (*totalGroup1* and *totalGroup2*). The fair model was successfully able to debias the Ethnic Group aware model and the fair model's scatter plots and correlation heat maps were visualized to be quite similar to the ground truth model, which suggests that the fair model is fair. Ethnic Group Blind model does not change the bias identified in the previous model by hiding the protected variables. We can conclude this since the correlation between meanvalence and percent_bachelorPlus and the correlation between meanvalence and household_meanIncome does not change significantly compared to the Ethnic Group Aware Model. The scatter plots also remain quite similar to the Ethnic Group Aware Model.

The threshold gives a great insight on which model is bias and which model is not towards the ethnic groups.

I. INTRODUCTION

This assignment is related to fairness and bias. The primary task of this assignment is to design a linear model that can fairly predict happiness and identify areas with average happiness less than 5.8. Happiness, also known as "valence" has been measured for a select few counties in the nation. The models will help Chirper Inc. in predicting valence and the impact of socio-demographic features on it. In the study, mean square error is not the primary focus. Instead, the focus is on building a fair model. Because the primary focus is the measure of fairness, three models are being built – race aware, race blind and fair.

Scatter plots of predicted outcome with respect to independent variables conditioned on ethnic group were generated for each model and compared to the ground truth model. This led us to compare and compute which model is more fair and how the fair model compares with the ground truth model.

II. DATA EXPLORATION

The data consisted of the following columns:

- **Id2** which gives the Region Id
- **totalGroup1** is the number of people in ethnic group 1
- **totalGroup1** is the number of people in ethnic group 2
- **percent_bachelorPlus** is the percentage of population with at least a bachelor's degree
- **households_meanIncome** gives us the mean income per household in the region and
- **meanvalence** gives us the mean happiness in the region

totalGroup1 and **totalGroup1** are considered as ethnic group variables or protected variables. **meanvalence** is considered as the predicted outcome variable and **percent_bachelorPlus** and **households_meanIncome** are considered as independent variables.

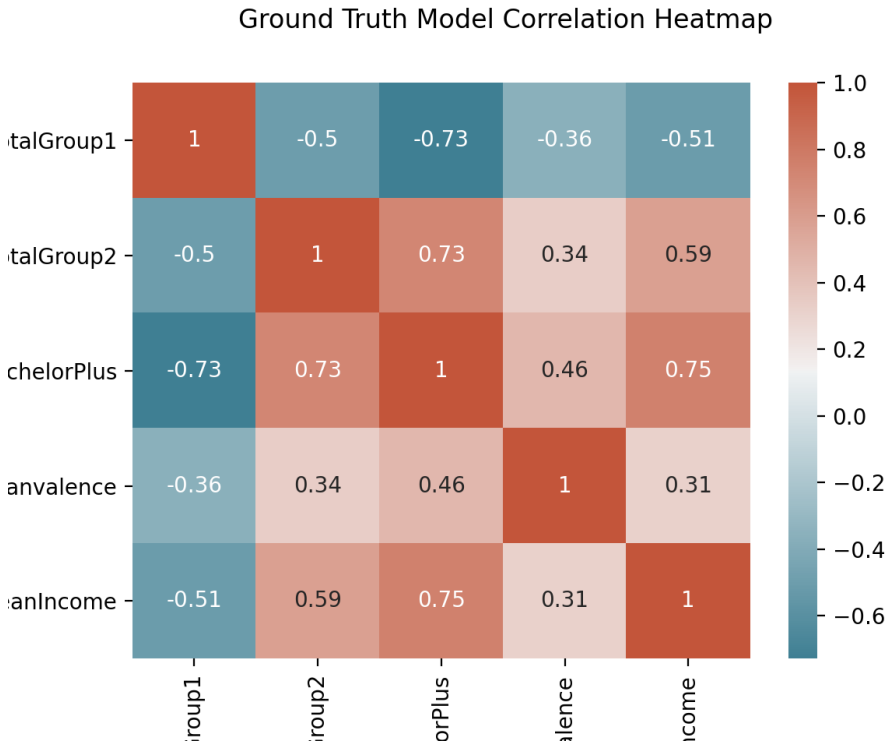

FIG. 1. Ground Truth Model Correlation heat map

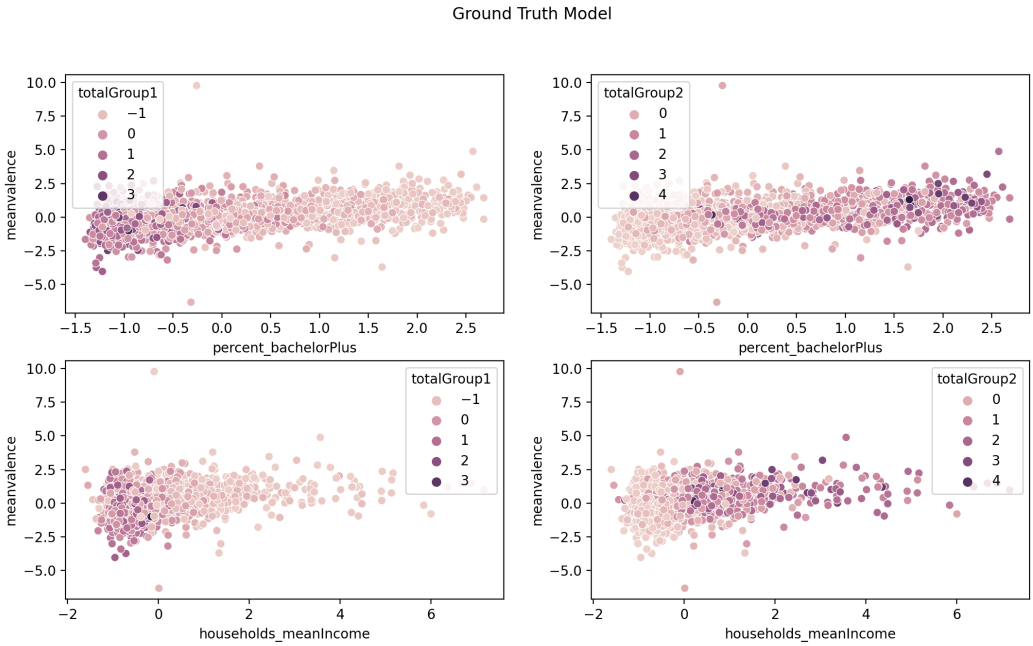

FIG. 2. Ground Truth Model Correlation heat map

III. DATA PREPROCESSING

Data was read using Pandas DataFrames and scaled using sklearn's StandardScaler. Correlation was computed for all the above mentioned variables except 'Id2'. Correlation heat map and scatter plots were generated for the same. This is regarded as the ground truth model. Four scatter plots were generated for meanvalence; two with respect to percent_bachelorPlus and two with households_meanIncome, both conditioned on totalGroup1 and totalGroup2 each. The heat map is displayed in Fig. 1. The Ground Truth Model can be visualized in scatter plots shown in Fig. 2.

Just from the scatter plots, we can visualize that there is not much of a bias towards one of the ethnic groups. The meanvalence with respect to percent_bachelorPlus for totalGroup1 is a flat one, very similar to that of totalGroup2. Similarly, for households_meanIncome, the graphs are similar to each other and the trend is a flat one which suggests that there is not much of a bias.

IV. MODEL SELECTION

The following models were implemented after the ground truth model was constructed:

1. **Ethnic Group Aware Model (Prediction Model I)** – A linear regression model was built using all predictors, on a percentage of the data provided. Valence was predicted on the entire data set. Fairness of predicted valence was also analyzed, by comparing the correlations and scatter plots of Group Aware Model with Ground Truth Model. The predicted valence threshold is set to 5.8. The correlation heat map is displayed in Fig. 3. The Ethnic Group Aware Model can be visualized in scatter plots shown in Fig. 4.

2. **Ethnic Group Blind Model (Prediction Model II)** – A linear regression model was built without using totalGroup1 and totalGroup2 as predictors, on a percentage of the data provided. Valence was predicted on the entire data set and fairness of predicted valence was also analyzed by comparing the correlations and scatter plots of Group Blind Model with Ground Truth Model. The correlation heat map is displayed in Fig. 5. The Ethnic Group Blind Model can be visualized in scatter plots shown in Fig. 6.

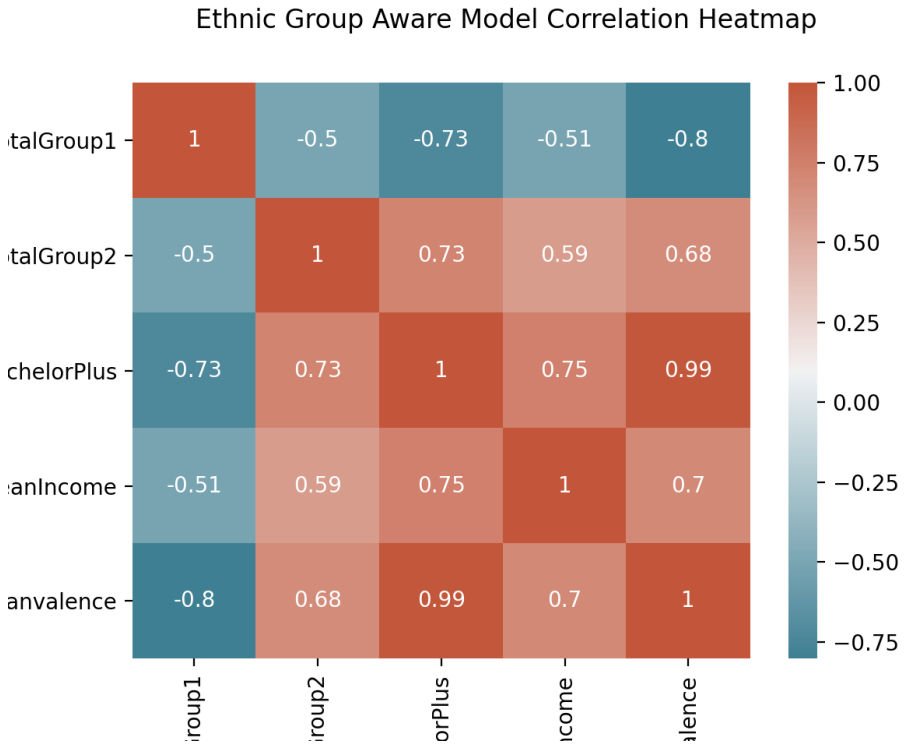

FIG. 3. Ethnic Group Aware Model Correlation heat map

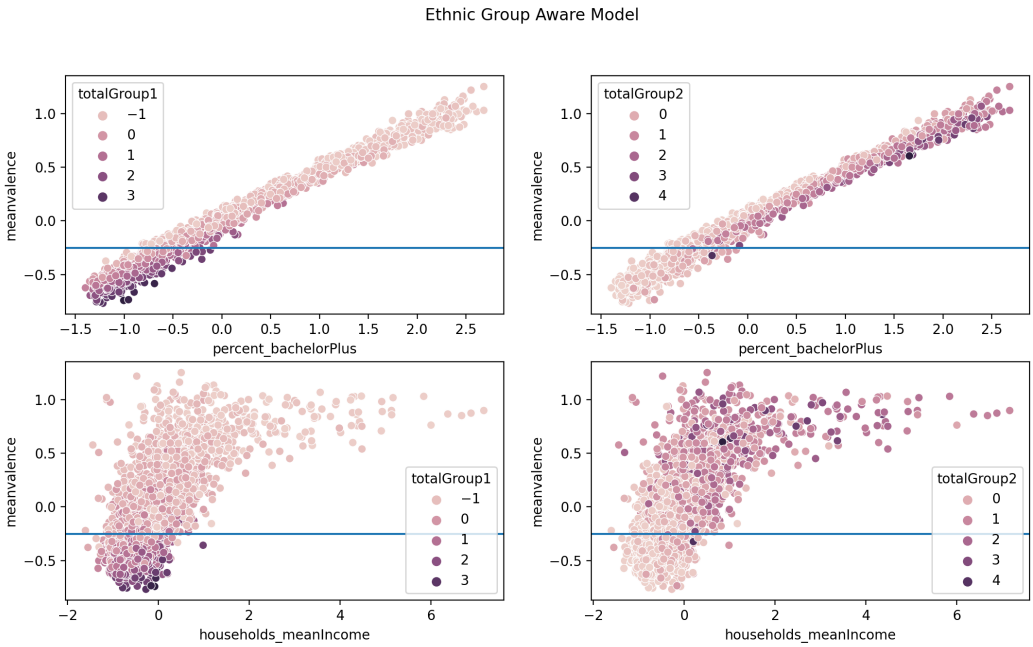

FIG. 4. Ethnic Group Aware Model Correlation heat map

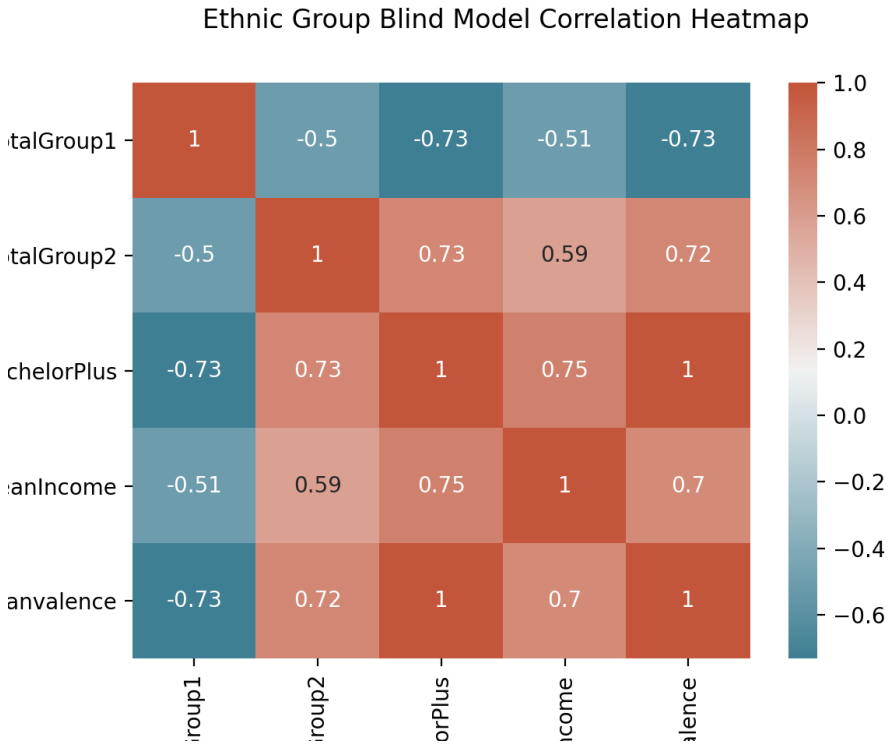

FIG. 5. Ethnic Group Blind Model Correlation heat map

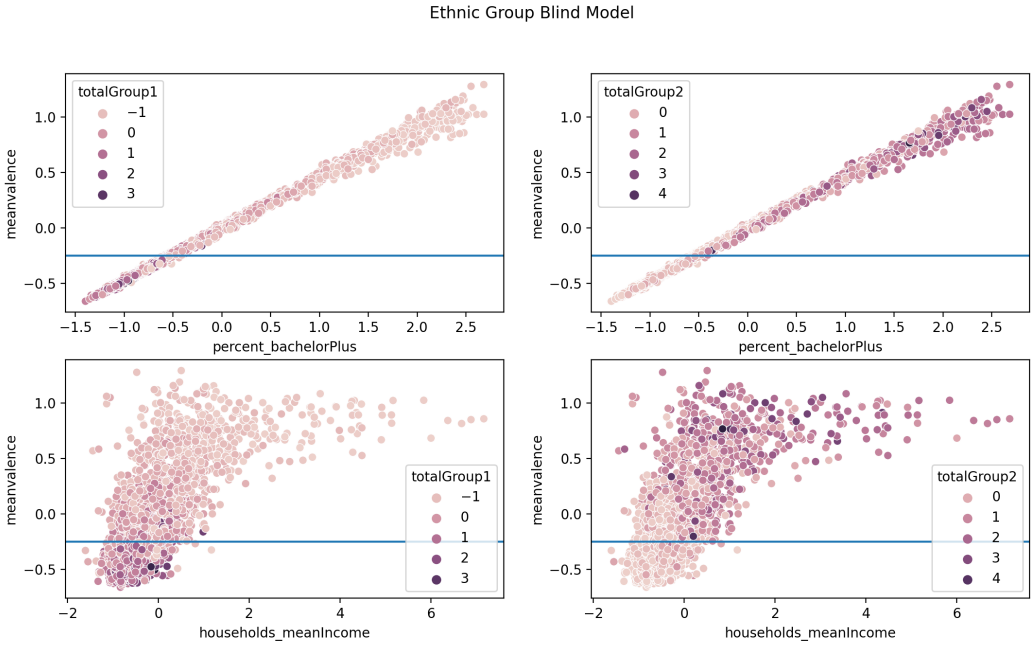

FIG. 6. Ethnic Group Blind Model Correlation heat map

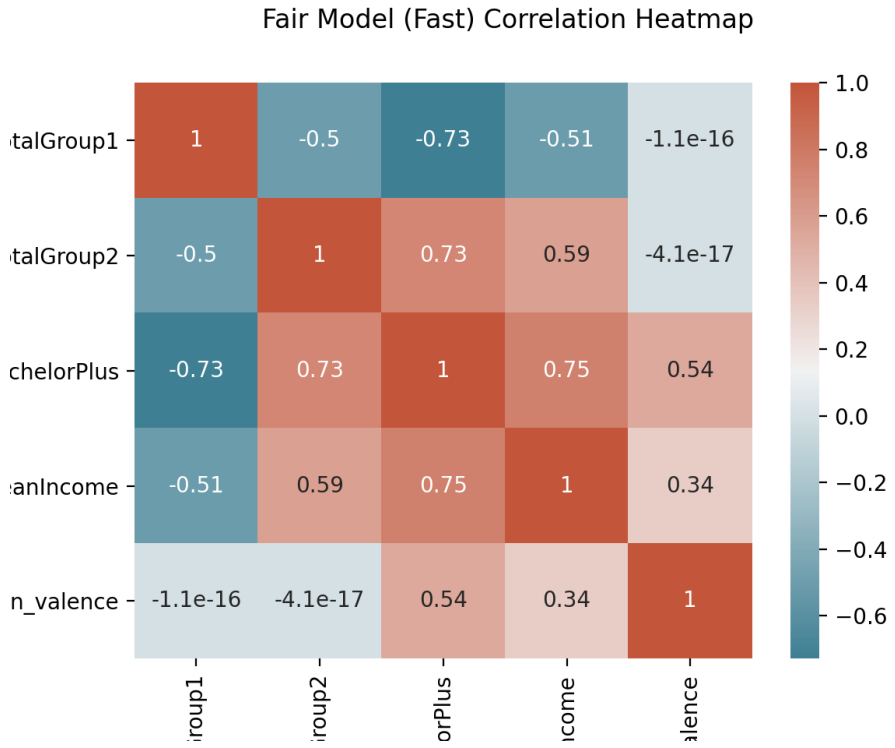

FIG. 7. Fair Model (Fast) Correlation heat map

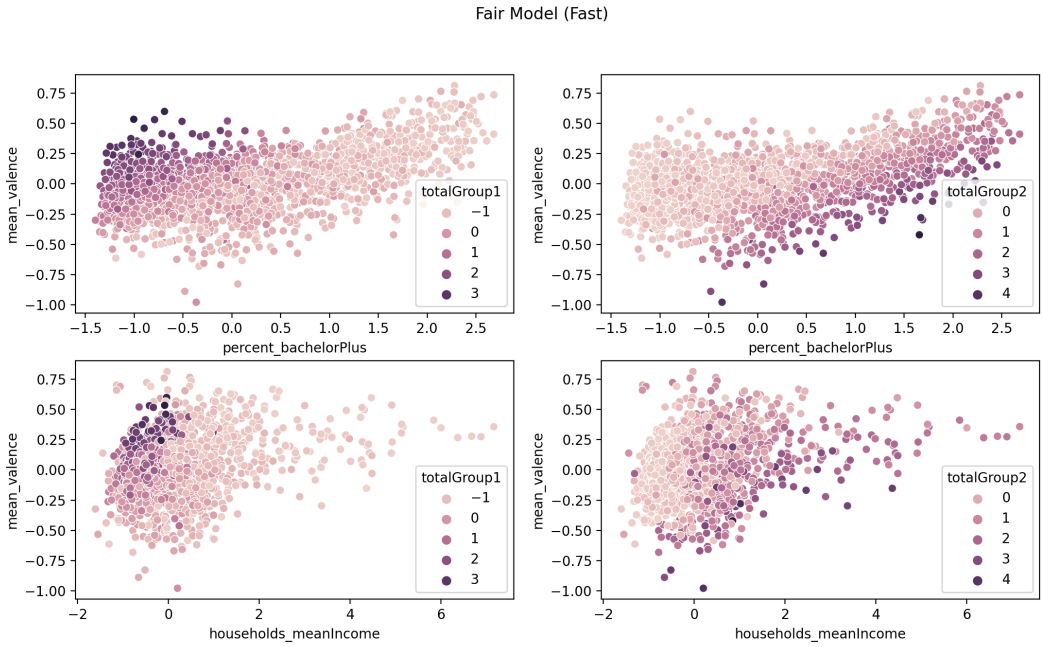

FIG. 8. Fair Model (Fast) Correlation heat map

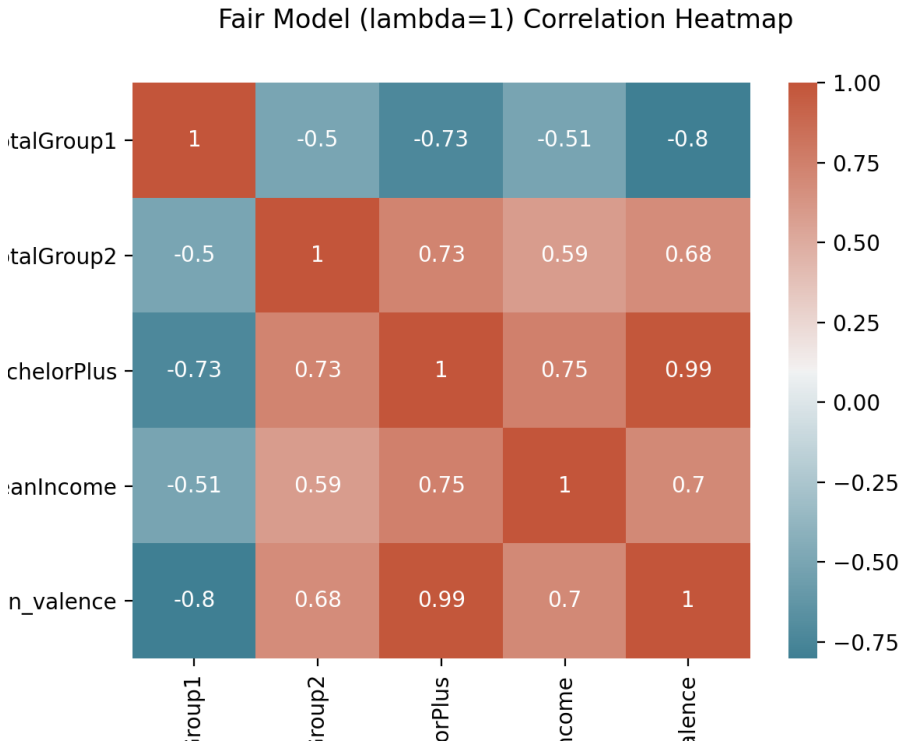

FIG. 9. Fair Model (Lambda = 1) Correlation heat map

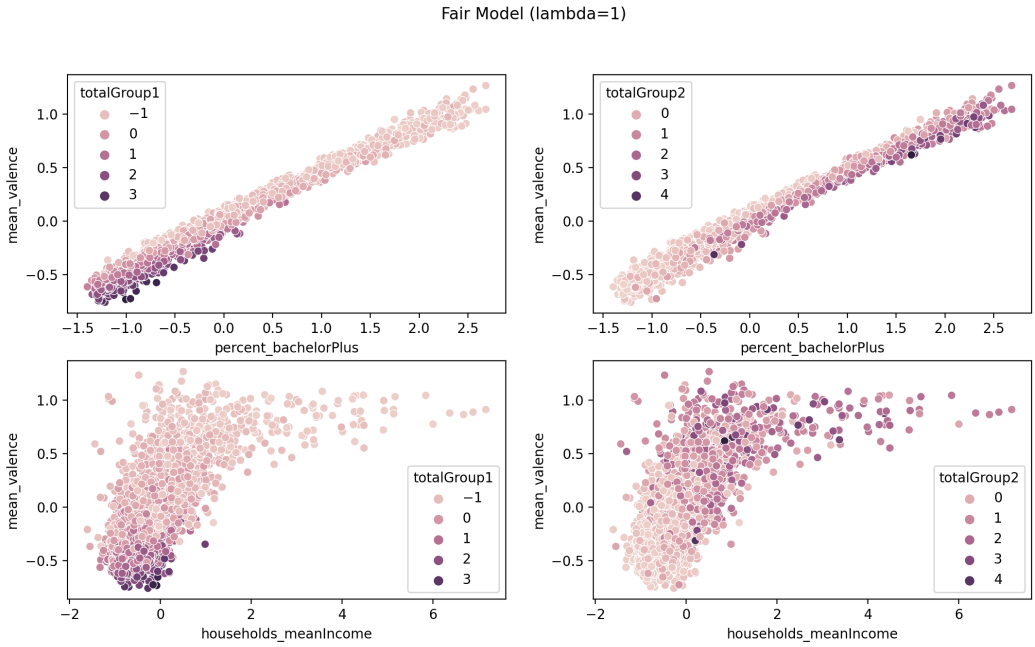

FIG. 10. Fair Model (Lambda = 1) Correlation heat map

3. **Fair Model** - This model was used to debias the outcome variable, which is meanvalence. Ethnic Group Aware Model (Prediction Model I) was used to debias the outcome variable. A couple of algorithms were used, which were provided by the instructors of the course: *gen_latent_fast* and *gen_latent_nonparam_regula*. The second algorithm had an extra input variable, *lambda*, which was set as 0 and 1. 0 meant totally fair and 1 meant same as outcome. The Fair model correlation heat map is displayed in Fig. 7. The Fair model (Fast) can be visualized in scatter plots shown in Fig. 8.

The Fair model (lambda = 1) correlation heat map is displayed in Fig. 9. The Fair model (lambda = 1) can be visualized in scatter plots shown in Fig. 10.

V. MODEL EVALUATION AND INTERPRETATION

1. **Ethnic Group Aware Model (Prediction Model I)** – As shown in Fig. 3 and in Fig. 4, the correlation between meanvalence and percent_bachelorPlus and the correlation between meanvalence and household_meanIncome increases compared to the Ground Truth Model which shows there is an increase in bias. As seen in the scatter plots, the scatter plot below the threshold is different to the one above the threshold, which suggest that there is bias between the two ethnic groups. The darker points of the scatter plot are also on either side of the threshold for both ethnic groups, which suggest that there is equal bias towards either ethnic groups.

   Since the data set was scaled, the new threshold was calculated to be -0.252 using the equation Eq. 1:

$$newthreshold = \frac{5.8 - mean(meanvalence)}{standarddeviation(meanvalence)} \tag{1}$$

2. **Ethnic Group Blind Model (Prediction Model II)** – As shown in Fig. 5 and in Fig. 6, the correlation between meanvalence and percent_bachelorPlus and the correlation between meanvalence and household_meanIncome increases compared to the Ground Truth Model which shows there is an increase in bias. As seen in the scatter plots, the scatter plot below the threshold is different to the one above the threshold, which suggest that there is bias between the two ethnic groups. The darker

points of the scatter plot are also on either side of the threshold for both ethnic groups, which suggest that there is bias. Hiding the protected variables do not change the bias identified in the previous model. We can conclude this since the correlation between meanvalence and percent_bachelorPlus and the correlation between meanvalence and household_meanIncome does not change significantly compared to the Ethnic Group Aware Model. The scatter plots also remain quite similar to the Ethnic Group Aware Model.

3. **Fair Model (Fast)** - This model was used to debias the outcome variable, which is meanvalence. Ethnic Group Aware Model (Prediction Model I) was used to debias the outcome variable. A couple of algorithms were used, which were provided by the instructors of the course: *gen_latent_fast* and *gen_latent_nonparam_regula*. The second algorithm had an extra input variable, *lambda*, which was set as 0 and 1. 0 meant totally fair and 1 meant same as outcome. The Fair model correlation heat map is displayed in Fig. 7. The Fair model (Fast) can be visualized in scatter plots shown in Fig. 8.

   The scatter plots for this model are quite similar to the ground truth model which confirms that there was not much of a bias in ground truth model. The Ethnic Group Aware model mean valence values were used to debias and therefore the correlation heat map values are also similar to the ground truth model.See Fig. 1 and Fig. 7. Therefore, we were successfully able to debias the model.

4. **Fair Model (Lambda = 0)** - This model behaves exactly similar to the Fair Model (Fast). Lambda value set as 0 means that the new model is now completely fair, and the graphs are similar to the Fair Model (Fast) ones.

5. **Fair Model (Lambda = 1)** - Lambda value set to 1 means that the model is quite unfair. The scatter plots and the correlation heat map suggest the same since this model's plots and heat map are identical to the Ethnic Group Aware Model's scatter plots and heat map, respectively. See Fig. 9 and Fig. 10.

VI. CONCLUSIONS

The primary task of this assignment was to design a linear model that can fairly predict happiness and identify areas with average happiness less than 5.8. Happiness, also known as "valence" has been measured for a select few counties in the nation. The models will help Chirper Inc. in predicting valence and the impact of socio-demographic features on it. It can be concluded that the model was not as biased for the ground truth model. Making Ethnic Group Aware Model and Ethnic Group Blind Model and comparing scatter plots and correlation heat maps, it can be concluded that these two models were biased equally towards both the ethnic groups (*totalGroup1* and *totalGroup2*). The fair model was successfully able to debias the Ethnic Group aware model and the fair model's scatter plots and correlation heat maps were visualized to be quite similar to the ground truth model, which suggests that the fair model is fair.

Ethnic Group Blind model does not change the bias identified in the previous model by hiding the protected variables. We can conclude this since the correlation between meanvalence and percent_bachelorPlus and the correlation between meanvalence and household_meanIncome does not change significantly compared to the Ethnic Group Aware Model. The scatter plots also remain quite similar to the Ethnic Group Aware Model.

The threshold gives a great insight on which model is bias and which model is not towards the ethnic groups.

DATA AVAILABILITY

Data is available at blackboard namely:

- chirper-happiness.csv

CODE AVAILABILITY

Code is available at `https://github.com/USC-DSCI-552-Spring2021/dsci552-spring2021-32416d-ps4-sjain681`

- ground-truth.py
- ethnic-group-aware.py

- ethnic group-blind.py
- fair-model.py
- fair-model-0.33, 0.67.py
- fair-model-0.5.py

ACKNOWLEDGMENTS

I would like to thank professor Kristina Lerman, professor Keith Burghardt, the TAs and the graders of the course DSCI 552 for all their efforts towards this assignment.

Appendix A: References

1. "Sklearn.linear_model.LinearRegression." Scikit, scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html.
2. "Statsmodels.regression.linear_model.OLS." Statsmodels, www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.OLS.html.