Write-up Question 0:

https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html
https://aegis4048.github.io/mutiple_linear_regression_and_visualization_in_python
https://www.datacourses.com/evaluation-of-regression-models-in-scikit-learn-846/
https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/
https://ogletree.com/insights/unintentional-discrimination-what-every-employer-needs-to-know-about-disparate-impact-claims/
https://www.hirevue.com/blog/hiring/what-is-adverse-impact-and-why-measuring-it-matters

Write-up Question 1 (5 points): Describe the type of model you built, note your prior experience (or lack thereof) with machine learning, and also provide the the root mean squared error (RMSE) of your model. We'll have an informal competition (with neither prizes nor points awarded) for the most accurate model.

I have only built one machine learning model in this capacity in Mobile Computing where we built a binary predictive model, so my experience is lacking. I spent most of my time reading about and setting up the data to be used in the model training (as the previous set of instructions noted that this was an important step). I ultimately used a multiple linear regression model. In regard to the root mean squared error of the model, I got 353.85108034146725. When considering the price range of the Air BnBs (0 to 1000 USD), a RMSE of approximately 354 seems to be relatively good!

Write-up Question 2 (5 points): Please describe which features (columns of the data) you chose to include in, or exclude from, your model. Why did you choose to include/exclude these particular features? Consider both the overall performance of the model and the ethical issues we've discussed in this class in making your decision.

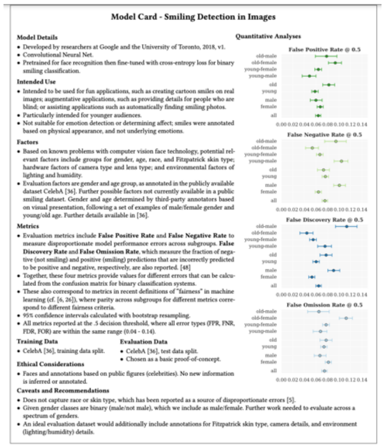
Regarding the privacy concerns with the data, I began by considering and subsequently removing any data that would have been considered PII or data that could link the Air BnB to a particular account or host. To achieve this, I simply dropped the columns that contained the id of the listing, the host id, the host name, and the name of the listing. I found that these were also not necessary when considering the dummy variables I wanted to use later on. I also worked to clean up the data by first adding in 0s for the spots where the review per month was null to ensure that the values were properly reflective of the meaning of the data. Additionally, I removed the neighborhood group column which I later, inadvertently, mimic with my community column. I dropped the neighborhood group column because there were no values present for and of the rows, making the column unnecessary to the desired outcome of the code. I also dropped any rows that contained NaN values after this because of the implication that there was a lacking of meaningful, usable data. I then decided to attempt to “bin” the latitude and longitudinal values in a sense. I did so by rounding either coordinate attribute to one decimal in the date frame. Moving on in the code, I also noticed, after plotting my bar chart graphs, that there were over 1000 entries where the “availability_365” column did not have any availability (the Air BnB is available 0 nights of the 365 days in the year) which means that they were not currently being rented out. I took this as an indicator that these values may skew the data as well, so I removed all occurrences of availability_365 being 0. After looking at the various value_counts for each of the columns that appeared to have some consistency in the data being presented, I finally landed on the three attributes of the data that I would ultimately use to train the multivariable model. These three variables are the coordinates (including both latitude and longitude values), neighborhood (or, later described, my grouping of them), and room types. I chose these three variables because I thought they would be a) the most beneficial *once binned or dummy variable-ified* and b) would be the most useful based on my external knowledge and beliefs about the housing and rental market/pricing in general. I handle outliers and data sparsity in the latitude and longitude values (please reference code and commenting for specifics!) before moving on to work on the neighborhood values. For the neighborhoods, I "binned" them using a new column I call community, which classifies each neighborhood associated with a given row with one of seven Chicago neighborhoods. This creates a better layout for the dummy variables (one that is not insanely overwhelming to understand)! I left room type untouched as it was already pretty straight forward and free from any glaring outliers. I then get dummy values for the given variables mentioned (latitude, longitude, room type, and community group) to then use in my model training and testing data.

Write-up Question 3 (5 points): What data pre-processing steps, if any, did you follow, and why? What data type (numerical data vs. dummy-coded categorical data) did you use for the different columns you included in your model? Why?

Many of my preprocessing steps were mentioned in Question 3! Briefly, removing columns that do not add value either due to PII issues or containing all null values, binning data in various manners to make the creation of the dummy variables more understandable, and rounding of decimals helped to clean the data in an appropriate way. Additionally, I used dummy-coded categorical training data for the various columns used in the model because I was attempting to use a multiple variable linear regression model which was the easiest to understand in terms of dummy variables. Additionally, dummy variables were required for this type of model training in particular because many of the categories (such as community) were string values which did not work and threw errors when attempting to train the models, thus dummy variables were required.

Write-up Question 4 (10 points): Included in Assignment 6 Folder

Write-up Question 5 (5 points): Now, describe in prose your process for making your model card, focusing on what you chose to include in (and what you chose to exclude from) your model card.



Model Card - Smiling Detection in Images

Model Details

- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use

- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors

- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Metrics

- Evaluation metrics include **False Positive Rate** and **False Negative Rate** to measure disproportionate model performance errors across subgroups. **False Discovery Rate** and **False Omission Rate**, which measure the fraction of negative (not smiling) and positive (smiling) predictions that are incorrectly predicted to be positive and negative, respectively, are also reported. [48]
- Together, these four metrics provide values for different errors that can be calculated from the confusion matrix for binary classification systems.
- These also correspond to metrics in recent definitions of "fairness" in machine learning (cf. [6, 26]), where parity across subgroups for different metrics correspond to different fairness criteria.
- 95% confidence intervals calculated with bootstrap resampling.
- All metrics reported at the .5 decision threshold, where all error types (FPR, FNR, FDR, FOR) are within the same range (0.04 - 0.14).

Training Data

- CelebA [36], training data split.

Evaluation Data

- CelebA [36], test data split.
- Chosen as a basic proof-of-concept.

Ethical Considerations

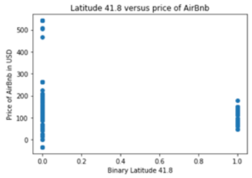
- Faces and annotations based on public figures (celebrities). No new information is inferred or annotated.

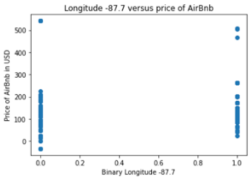
Caveats and Recommendations

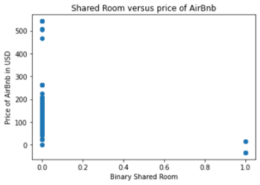
- Does not capture race or skin type, which has been reported as a source of disproportionate errors [5].
- Given gender classes are binary (male/not male), which we include as male/female. Further work needed to evaluate across a spectrum of genders.
- An ideal evaluation dataset would additionally include annotations for Fitzpatrick skin type, camera details, and environment (lighting/humidity) details.

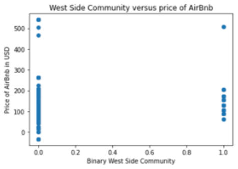
For the Model Card, I wanted to focus specifically on the information that was *excluded* from the training and testing data, and why I made the decisions I did in regards to the exclusion of said data. Regarding the layout of the Model Card, I used the example Model Card from the example paper and included information relevant to the list presented from the same research paper. The information described in my Model Card (which is included in the Assignment6 folder submission) pertains to the nature of the model, the breakdown and cleaning of the data, how said data was used to train the linear regression model, and what the predictions of the model are in terms of quantitative analysis.

I used information outputs from the code written for quantitative analysis, but I excluded the use of graphs as the linear regression model scatter plots for each variable/column were not yielding meaningful visuals for this Model Card. Consider the scatter plot graphs below. For brevity, I chose one of each latitude, longitude, room type, and community variable type, despite there are multiple of each. In each case, we can see how, in chosing binary/dummy variables, the plots of the graphs are not meaningful and, thus, would not mean much as a visual to be included in the Model Card, whereas, in many other cases, such visuals would yield more visually meaningful results.



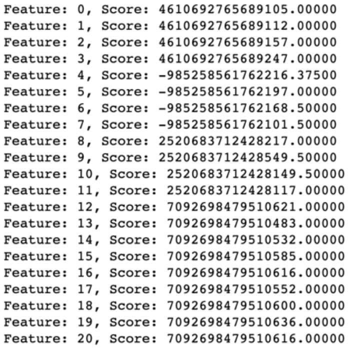






Write-up Question 6 (5 points): Briefly discuss what input features were important in your model's predictions. What does this exercise reveal about Airbnb pricing in Chicago?

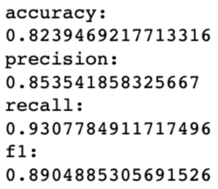
Based on the way I trained my model and the inputs used, I learned that the prcing in Chicago is heavily influenced by the location of the rental space. Based on not only the feature scores from the *communities* but also the *latitude values* (associated with features 0 – 3), the location in proximity to the city center seems to have the largest impact on the price. This is not surprises considering both supply and demand as well as the general knowledge of expensive real estate being closer to the city center (which is considered a by product of supply and demand).


```
Feature: 0, Score: 4610692765689105.00000
Feature: 1, Score: 4610692765689112.00000
Feature: 2, Score: 4610692765689157.00000
Feature: 3, Score: 4610692765689247.00000
Feature: 4, Score: -985258561762216.37500
Feature: 5, Score: -985258561762197.00000
Feature: 6, Score: -985258561762168.50000
Feature: 7, Score: -985258561762101.50000
Feature: 8, Score: 2520683712428217.00000
Feature: 9, Score: 2520683712428549.50000
Feature: 10, Score: 2520683712428149.50000
Feature: 11, Score: 2520683712428117.00000
Feature: 12, Score: 7092698479510621.00000
Feature: 13, Score: 7092698479510483.00000
Feature: 14, Score: 7092698479510532.00000
Feature: 15, Score: 7092698479510585.00000
Feature: 16, Score: 7092698479510616.00000
Feature: 17, Score: 7092698479510552.00000
Feature: 18, Score: 7092698479510600.00000
Feature: 19, Score: 7092698479510636.00000
Feature: 20, Score: 7092698479510616.00000
```

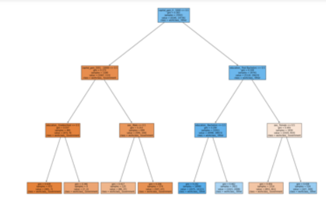
Since the higher the score means the more impact the given feature has on the effect on the model, based on the feature importance score outputs, the most impactful/important features seem to be features 12 – 20 which correlate to *all* of the community groups which were a derivative/grouping of the neighborhoods.

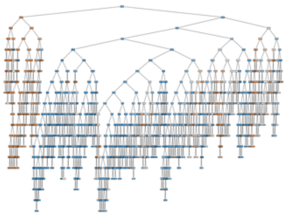
Write-up Question 7 (5 points): Report your decision tree model's accuracy, precision, recall, and F1 score based on your test set. You should see a pretty high accuracy. Does this mean that this is a good model?

Considering *all* available statistics, I would consider this model to be very strong when considering the predictions. Looking at just accuracy, however, is not sufficient as a greater accuracy alone does not necessarily indicate a better machine learning algorithm for the prediction of the binary variable(s) in question. To know if the model is a good predictor for the data, we must consider all of the given values.


```
accuracy:
0.8239469217713316
precision:
0.853541858325667
recall:
0.9307784911717496
f1:
0.8904885305691526
```

Write-up Question 8 (5 points): Visualize your decision tree. Include the decision tree in your writeup. Examine the facets that contribute to these predictions. What issues do you notice?

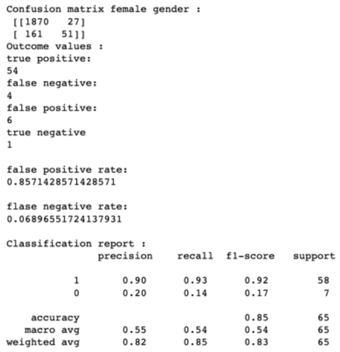




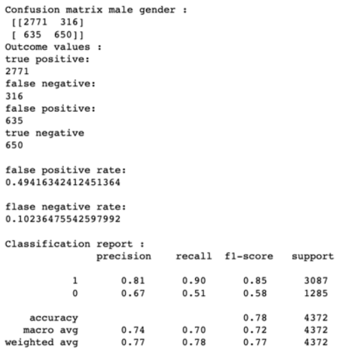
Initially, I did not set a depth for the tree, and the image on the left was what I got! This is clearly not super helpful, so I then set the dept to 3 which yielded much more readable and useful results. Something I found that the Gini value/index was quite high (as the Gini index ranges from 0 to 0.5 where the smaller the Gini value, the more pure and well-split the data is). This means that (especially when disregarding the education_ Gradeschool node and following children, the Gini nodes are almost always over half of the possible Gini index values which indicates a relative impure and unbalanced decision tree splitting.

Write-up Question 9 (10 points): For your proposed model, plot the false positive rate (FPR) and false negative rate (FNR) by gender and (separately) also by race. In your write-up, include your graphs and describe in prose any patterns you see. Does your model seem fair through this lens?

Gender Confusion Matrices:



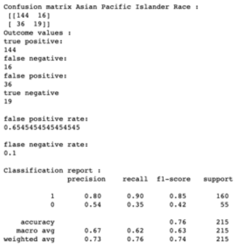
```
Confusion matrix female gender :
 [[1870   27]
 [ 161   51]]
Outcome values :
true positive:
54
false negative:
4
false positive:
6
true negative
1

false positive rate:
0.8571428571428571

flase negative rate:
0.06896551724137931

Classification report :
              precision    recall  f1-score   support

           1       0.90      0.93      0.92        58
           0       0.20      0.14      0.17         7

    accuracy                           0.85        65
   macro avg       0.55      0.54      0.54        65
weighted avg       0.82      0.85      0.83        65
```



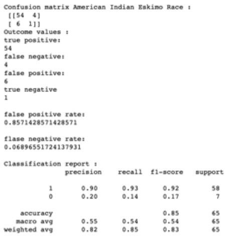
```
Confusion matrix male gender :
 [[2771  316]
 [ 635  650]]
Outcome values :
true positive:
2771
false negative:
316
false positive:
635
true negative
650

false positive rate:
0.49416342412451364

flase negative rate:
0.10236475542597992

Classification report :
              precision    recall  f1-score   support

           1       0.81      0.90      0.85      3087
           0       0.67      0.51      0.58      1285

    accuracy                           0.78      4372
   macro avg       0.74      0.70      0.72      4372
weighted avg       0.77      0.78      0.77      4372
```

Considering the false positive and false negative rates for either gender, I would say the classifier is better at classifying negative values than it is classifying positive values. Additionally, it misidentifies women positivity of income at 85% where that same statistic is 49% in men. The false negative rates are also approximately 10% or less which indicates accuracy in the negative classification of this algorithm. These values are also relatively similar (6% vs. 10% from men to women respectively). When considering the gendered breakdown of these confusion matrices and rates, it seems that the classifier is better at predicting the outcome of the males rather than the females so, for the purpose of this question, it seems that the classifier has a pattern of misidentifying women and properly identifying men when considering bot the false negative and false positive rates in tandem.

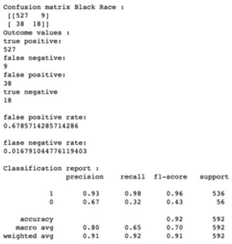
<u>Race Confusion Matrices:</u>



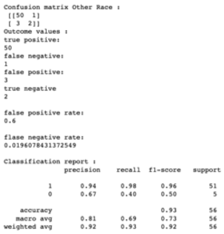
```
Confusion matrix Asian Pacific Islander Race :
 [[144  16]
 [ 36  19]]
Outcome values :
true positive:
144
false negative:
16
false positive:
36
true negative
19

false positive rate:
0.6545454545454545

flase negative rate:
0.1

Classification report :
              precision    recall  f1-score   support

           1       0.80      0.90      0.85       160
           0       0.54      0.35      0.42        55

    accuracy                           0.76       215
   macro avg       0.67      0.62      0.63       215
weighted avg       0.73      0.76      0.74       215
```



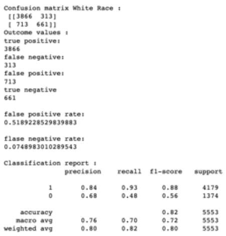
```
Confusion matrix American Indian Eskimo Race :
 [[54  4]
 [ 6  1]]
Outcome values :
true positive:
54
false negative:
4
false positive:
6
true negative
1

false positive rate:
0.8571428571428571

flase negative rate:
0.06896551724137931

Classification report :
              precision    recall  f1-score   support

           1       0.90      0.93      0.92        58
           0       0.20      0.14      0.17         7

    accuracy                           0.85        65
   macro avg       0.55      0.54      0.54        65
weighted avg       0.82      0.85      0.83        65
```



```
Confusion matrix Black Race :
 [[527   9]
 [ 38  18]]
Outcome values :
true positive:
527
false negative:
9
false positive:
38
true negative
18

false positive rate:
0.6785714285714286

flase negative rate:
0.016791044776119403

Classification report :
              precision    recall  f1-score   support

           1       0.93      0.98      0.96       536
           0       0.67      0.32      0.43        56

    accuracy                           0.92       592
   macro avg       0.80      0.65      0.70       592
weighted avg       0.91      0.92      0.91       592
```



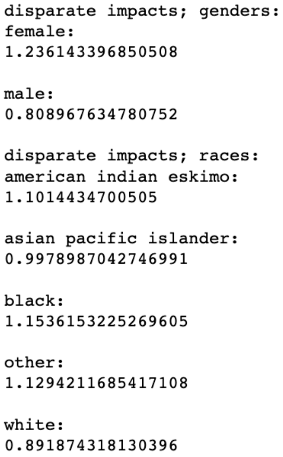
```
Confusion matrix Other Race :
 [[50  1]
 [ 3  2]]
Outcome values :
true positive:
50
false negative:
1
false positive:
3
true negative
2

false positive rate:
0.6

flase negative rate:
0.0196078431372549

Classification report :
              precision    recall  f1-score   support

           1       0.94      0.98      0.96        51
           0       0.67      0.40      0.50         5

    accuracy                           0.93        56
   macro avg       0.81      0.69      0.73        56
weighted avg       0.92      0.93      0.92        56
```



```
Confusion matrix White Race :
 [[3866  313]
 [ 713  661]]
Outcome values :
true positive:
3866
false negative:
313
false positive:
713
true negative
661

false positive rate:
0.5189228529839883

flase negative rate:
0.0748983010289543

Classification report :
              precision    recall  f1-score   support

           1       0.84      0.93      0.88      4179
           0       0.68      0.48      0.56      1374

    accuracy                           0.82      5553
   macro avg       0.76      0.70      0.72      5553
weighted avg       0.80      0.82      0.80      5553
```

When considering the confusion matrices above for the various races, we see that the classifier has the lowest false positive rate with White gender and the highest false positive rate with American Indian Eskimos. Additionally, regarding the false negative rate, Black gender has the smallest rate whereas Asian Pacific Islander race has the highest value. For the purpose of the question, it seems that relative to each other, the false negative rate is standardized, varying no more than 9% which is comparatively low, whereas the false positive rate has a larger difference of 34%. This does not necessarily show a pattern as the high and low values for gender do not fall in the same race categories, but these values do indicate a certain level of unpredictability in the classification.

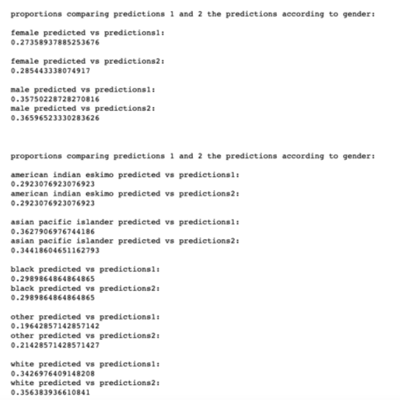
Write-up Question 10 (7.5 points): One proposed fairness definition is **<u>disparate impact in outcomes</u>** <u>(Links to an external site.)</u>. Disparate impact is defined as the ratio Pr[prediction = 1 | member of a particular group]/Pr[prediction = 1 | not member of a particular group]. Note that you can also formaulate this ratio based on privileged and unprivileged groups; for simplicity, calculate this ratio for each group where the denominator represents all other groups. **Consider *prediction = 1* to represent a predicted income above $50,000.** If this ratio is less than some threshold tau, it is considered to have disparate impact. Evaluate your proposed model using the disparate impact metric looking first at gender, then at race. Explain what you conclude from these measurements. Should we be concerned by these results? Why or why not?



```
disparate impacts; genders:
female:
1.236143396850508

male:
0.808967634780752

disparate impacts; races:
american indian eskimo:
1.1014434700505

asian pacific islander:
0.9978987042746991

black:
1.1536153225269605

other:
1.1294211685417108

white:
0.891874318130396
```

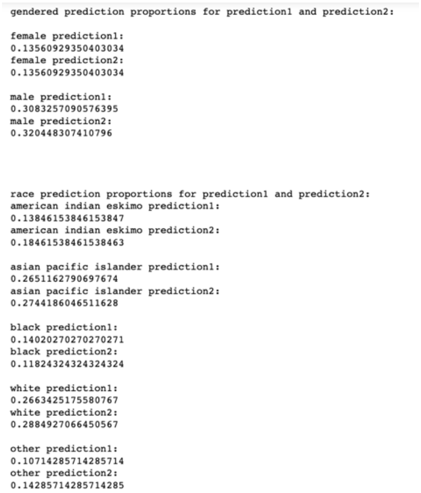
Considering the 4/5 or 80% rule (and corresponding tau value) regarding disparate impact, the numbers from the model indicate that there is no disparate impact to be concerned about, since (when converted to percentages). The two groups for either gender or race that are closest to a disparate impact are male (gender) and white (race) which I found to be particularly interesting as these are groups that are not typically associated with "unprivileged groups" which is the purpose of the use of a disparate impact analysis. Based purely on the values seen on the right, alone, there is no cause for concern regarding the ethical nature of the prediction model.

Write-up Question 11 (7.5 points): The column predictions1 and predictions2 in the dataframe contains predictions from two black-box models we've constructed. Examine the predictions in light of gender and race. Do they seem fair by the metrics you have considered so far?



```
proportions comparing predictions 1 and 2 the predictions according to gender:

female predicted vs predictions1:
0.27358937885253676

female predicted vs predictions2:
0.285443338074917

male predicted vs predictions1:
0.35750228728270816
male predicted vs predictions2:
0.36596523330283626


proportions comparing predictions 1 and 2 the predictions according to gender:

american indian eskimo predicted vs predictions1:
0.2923076923076923
american indian eskimo predicted vs predictions2:
0.2923076923076923

asian pacific islander predicted vs predictions1:
0.3627906976744186
asian pacific islander predicted vs predictions2:
0.34418604651162793

black predicted vs predictions1:
0.2989864864864865
black predicted vs predictions2:
0.2989864864864865

other predicted vs predictions1:
0.19642857142857142
other predicted vs predictions2:
0.21428571428571427

white predicted vs predictions1:
0.342697640914820 8
white predicted vs predictions2:
0.356383936610841
```

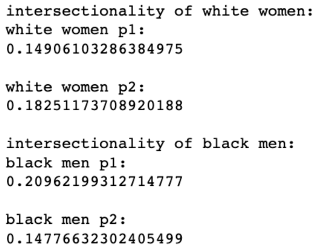
Using prediction1 and prediction2 columns in comparison with the predictions made with the model, it seems that the model proportions could be argued as being relatively similar from group to group as seen in the proportions present on the left. In my opinion, however, there is a discrepancy when considering the proportions. The largest difference seen here is with prediction1 versus the predictions made by the model for "Other" race and "White" predicted model values and predicitons2 with other being the smallest proportion of incomes over $50,000 and white being the largest proportion of incomes over $50,000 when compared with each respective prediction1 or prediction2 array. Additionally, when we consider the genders, the discrepancy between the largest and smallest values are also, in my opinion, quite large (too large, but not as large as with the race discrepancy). The "Female" predicted versus prediction1 array and "Male" prediction2 and the predictions from the model have the highest difference with "Female" having a smaller similarity proportion between the two previously stated prediction arrays and "Male" having the largest similarity proportion.



```
gendered prediction proportions for prediction1 and prediction2:

female prediction1:
0.13560929350403034
female prediction2:
0.13560929350403034

male prediction1:
0.3083257090576395
male prediction2:
0.320448307410796



race prediction proportions for prediction1 and prediction2:
american indian eskimo prediction1:
0.13846153846153847
american indian eskimo prediction2:
0.18461538461538463

asian pacific islander prediction1:
0.2651162790697674
asian pacific islander prediction2:
0.2744186046511628

black prediction1:
0.14020270270270271
black prediction2:
0.11824324324324324

white prediction1:
0.2663425175580767
white prediction2:
0.2884927066450567

other prediction1:
0.10714285714285714
other prediction2:
0.14285714285714285
```

When considering the values seen in the figure on the right, I computed the proportion of predictions for both prediction1 and prediction2 where the value was indicted to be over $50000 income. We can see that the discrepancy in gender is largest when comparing the proportion of "Over $50,000 values" in prediction1 array for "Female" and "Male" prediction2 proportions with the former being the smallest and the latter being largest with a difference of approximately 18.5%. For race, this discrepancy is largest when comparing "Other" prediction1 array proportion and "White" prediction2 proportions. This differences rests at approximately 19% with the "White" prediction2 proportion of values estimated to be over $50,000 incomes to be higher than that of the prediction1 array for "Other". I would consider both of these values to indicate a level of inequality and problematic tendencies within the prediction1 and prediction2 arrays.

Write-up Question 12 (5 points): "Intersectionality" is a term coined by Kimberlé Crenshaw used to refer to the phenomena in anti-discrimination law where plaintiffs who brought cases alleging discrimination on two bases (e.g., sex and race) would lose cases because the sex-based discrimination claims and the race-based discrimination claims would be evaluated separately, rather than considering the intersection between different demographic categories. For example, Black women alleging discrimination in a seniority system were denied relief because the seniority system did not disadvantage Black employees (compared to all non-Black employees) when ignoring gender. It also did not disadvantage women (compared to all non-women) when ignoring race. However, the seniority system **did** discriminate against the intersectional category of Black women (compared to all data subjects who were not Black women). **Analyze the performance of your proposed model through this intersectionality lens.** Since the number of comparisons to make grows exponentially with the number of intersectional groupings, please **choose two intersectional groups you consider to be of particular importance** and evaluate the models you've developed with respect to those two groups. What do you conclude?

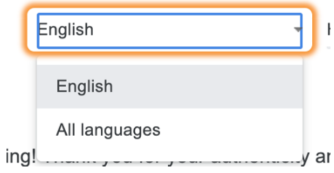


```
intersectionality of white women:
white women p1:
0.14906103286384975

white women p2:
0.18251173708920188

intersectionality of black men:
black men p1:
0.20962199312714777

black men p2:
0.14776632302405499
```

I chose white women and black men as the groups to evaluate intersectionality. On the right is the outcomes when finding the proportions of evaluations from prediction1 and prediction2 arrays when considering when women are white and when men are black for either category. In this instance when only white women are considered, they have a higher proportion of income predicitons over $50,000 when compared to all women but fewer income predictions of this same category when compared to the proportions associated with "White" predictions in general. When considering black men intersectionality, there is, overall, a higher proportion when comparing the values for prediction1 and prediction2 arrays to the "Black" race category. Additionally, the values indicate fewer predictions of incomes over $50,000 when comparing black men to men. This indicates that women that are *not* white are evaluated with incomes over $50,000 less frequently than women that *are* white and predictions in the "White" race category have a higher proportion of incomes predicted over $50,000 when women are not considered (considering both men and women). Additionally, for black men, these values indicate that the proportion of all men is higher than the proportion of men who are black and the proportion of "Black" race predictions. These proportion values indicate a certain discriminatory quality in the prediction model and dataset (or within society as a whole).

Task 3 (Exploration): Screen Readers (25 points)

Write-up Question 13 (5 points): In a few sentences, state which screen reader you've downloaded (and for which browser), and then briefly describe the key features of the screen reader in a few sentences based on your experience using it.



I chose to use the Google Chrome extension screen reader linked in the prompt for Task 3. I initially found that the screen reader gives a bit of detail not just about the text itself but about the HTML of the text (saying descriptors such as stating when text is a header, when I am viewing a new tab, and when I am engaging with the screen as a whole versus a specific element on the page). However, when there is a dropdown menu option, I notice that the options are not listed out verbally. For example, the image on the right shows two options, English and

All languages, but the reader states that there are two options which does not seem particularly helpful for people who have visual disabilities that prevent them from reading their screen. Additionally, any of the verbal descriptions include information that does not seem useful particularly when considering subtext and symbols.

There is also a functionality in the top right corner that allows the user to "skip to main content," but I could see this posing an issue as, if a user is visually impaired, they would not be able to locate this button in a timely manner. It is also small, such that it would be hard to click without the potential for clicking other things on the page mistakenly.

When I had the screen reader reading my Jupyter Notebook, I was truly so confused. It was reading so many attributes on the page that did not seem meaningful or helpful to someone who is visually impaired or who otherwise struggles to use a computer. I could also be completely wrong in this regard, as I have been privileged enough to be able to use my browser without the need for an aid.

Upon looking at the overview for this screen reader, the visual examples it gives seem to be from a very antiquated version of Google, which makes me think that this extension may not have been developed much in recent years. The overview of the extension also goes on to state that the screen reader is "built using only web technologies such as html and JavaScript" which is apparent as the screen reader typically verbalizes page attributes that are not seen but (as I infer) seemingly speaks urls and other page attributes that I assume are written in the code but not seen visually on the page.

Write-up Question 14 (10 points): What aspects of web browsing do you feel are captured **well** by this screen reader?

The textual elements and ordering of the page seem to be done well. Additionally, the *idea* of the "skip to main content" is a good inclusion as the page in its entirety will be read out. The screen reader will also read out the tab and window that the page being read is on, which allows for increased accessibility as the user would know more about the page they are working on and what to expect from it.

Write-up Question 15 (10 points): What aspects of web browsing do you feel are captured **poorly**, or missed entirely, by this screen reader? These are, of course, aspects of the experience of browsing the web that are typically inaccessible to people who are blind or low vision.

I think this application has a lot to improve on, but I also recognize that is easier said than done. Especially considering the strides that many companies are making regarding accessibility and user experience, however, I feel like this extension has a long way to go. Based on the images that are meant to "sell" the extension to the user and the actual implementation of the screen reader, it seems like the extension has been ignored and is antiquated. Additionally, I think the screen reader does a poor job of annunciation and describing where on the page it is reading, which seems like it could pose an issue for someone who is attempting to use a given webpage but is hard of seeing. Additionally, the descriptions that the screen reader gives are sometimes very generic or too specific and detailed at the wrong times. Such examples are reading URLs or underlaying information regarding the HTML corresponding to certain elements on the page and failing to read out options for certain drop-down menus (respectively).

Overall, the idea and beginning implementation for the screen reader is a spectacular idea and is, on a very basic level, good for those who are hard of seeing to make internet browsing experiences more accessible, but I think the screen reader from Google in this case has a long way to go before it is usable in a truly meaningful way to users.