**19CSE304 – Fundamentals of Data Science**

**ASSIGNMENT 3**

Gajula Sujan Sagar

**[CB.****EN.U4CSE20219]**

The Dataset has 3 Attributes Ethnicity, Eligible, Panels:

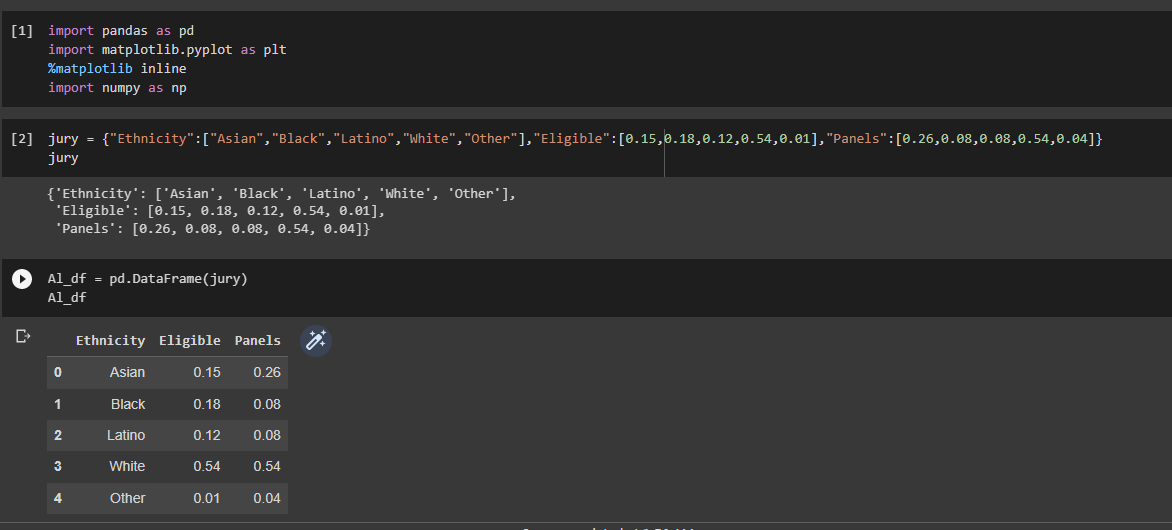
Ethnicity describes about the ethnic categories of people in Alameda.

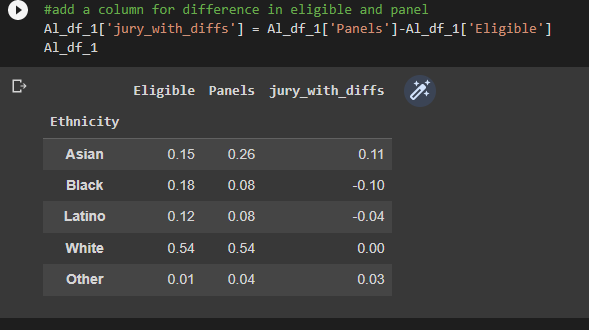
Eligible describes about the fraction of people eligible grouped by ethnicity

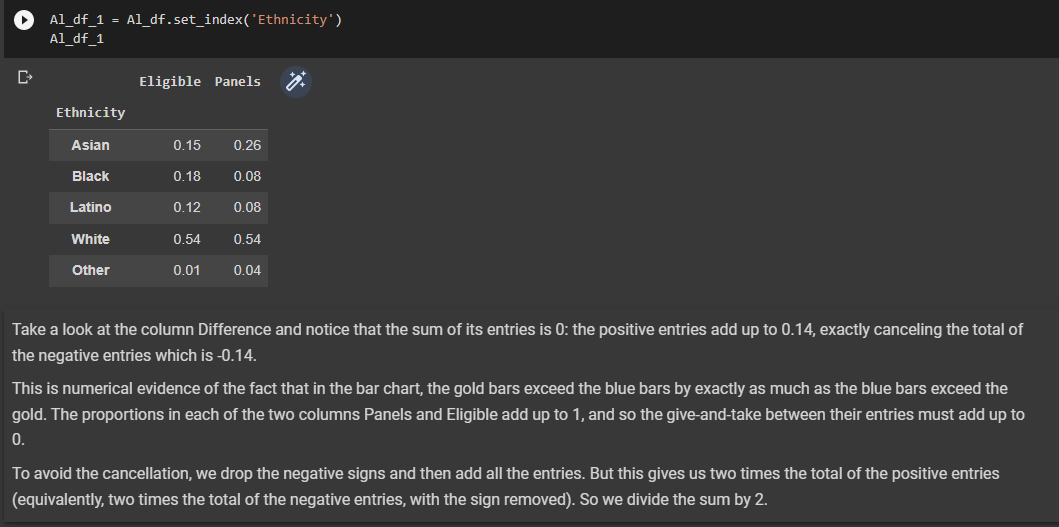
Panels describes about the proportion of people currently chosen for the panel.

##### Null Hypothesis: -panels were selected at random from the population of eligible jurors.

##### Alternate Hypothesis: -panels were not selected at random

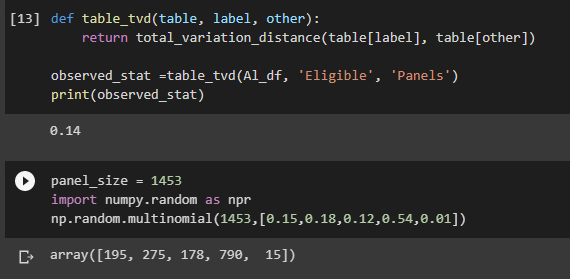






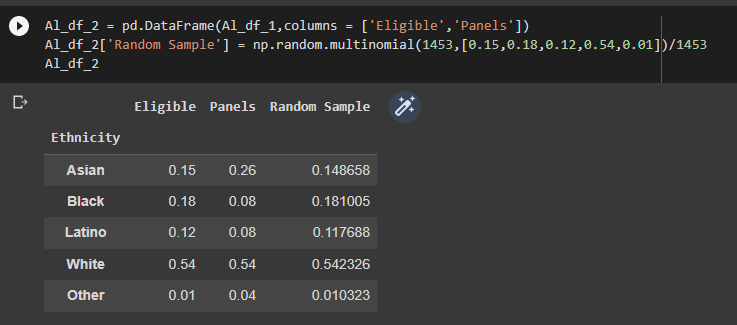


So, it would only logically make sense if the number of members chosen in random that are in excess are same as the one in deficit. 14%, -14% in our case.

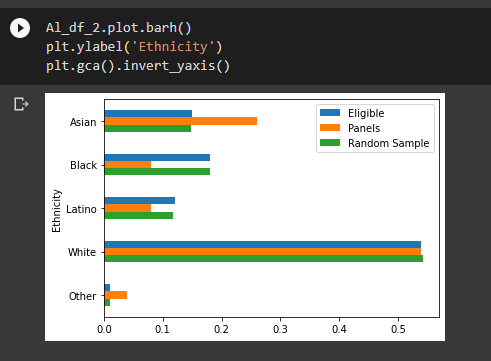


This quantity 0.14 is the total variation distance (TVD) between the distribution of ethnicities in the eligible juror population and the distribution in the panels.

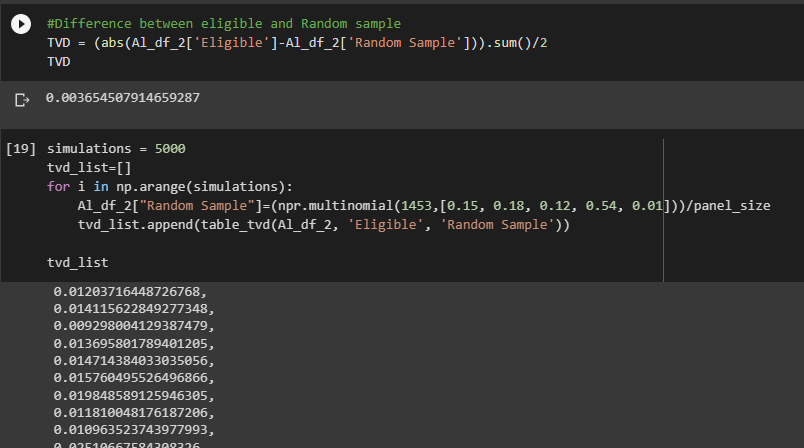
We could have obtained the same result by just adding the positive differences. But our method of including all the absolute differences eliminates the need to keep track of which differences are positive and which are not.

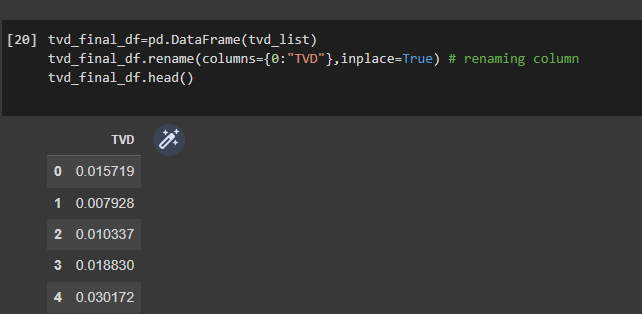


the distribution of the random sample is close to the distribution of the eligible population and is different from the distribution of the panels.



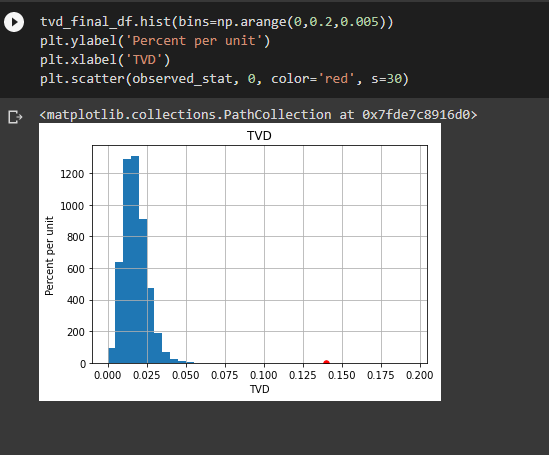
The green bar are closer in size to the blue bars than the orange bars are. The randomsample resembles the eligible population, but the panels don't.





Now, we find out the Difference between Random value and the Actual Eligible value, 0.01209222 in our case.

Repeat this task 5000 more times, and find this difference each time and store it in a dataframe.



Plot a histogram to visualize such a huge data easily and also use scatter plot to plot a point of the observed difference, and hence to Reject our Null Hypothesis. This is because we have sufficient proof to prove that the Alternative hypothesis turns out to be True, meaning, there was a clear bias.

**INFERENCE: The difference which we find is not related with the actual eligible value. It's far away from the value. So, here the alternative hypothesis has enough validations to prove its correctness and null hypothesis is rejected. Hence, the jury panel is not picked at random and this ensures that it favours to a particular community or group and underrepresents the other.**