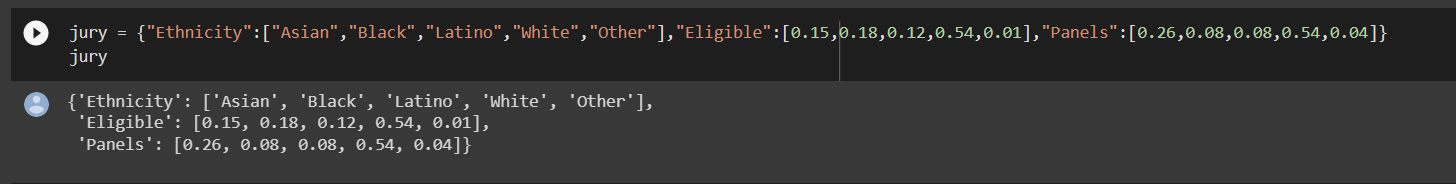
CB.EN.U4CSE20203

ASSIGNMENT -3

Problem:Jury selection in Alameda county

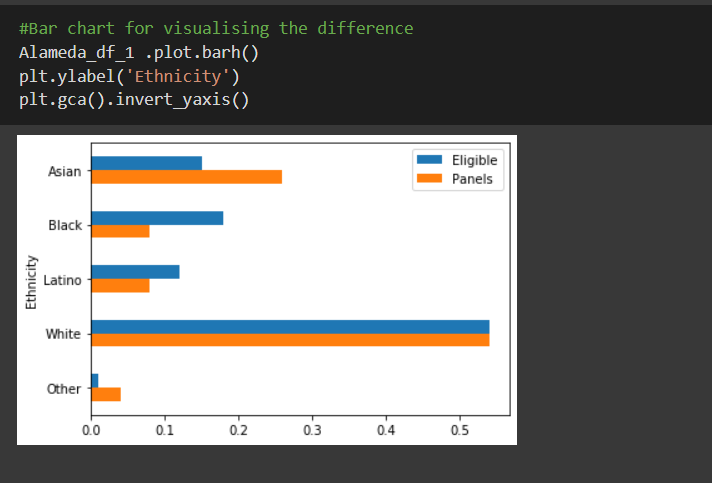


There are 3 attributes:

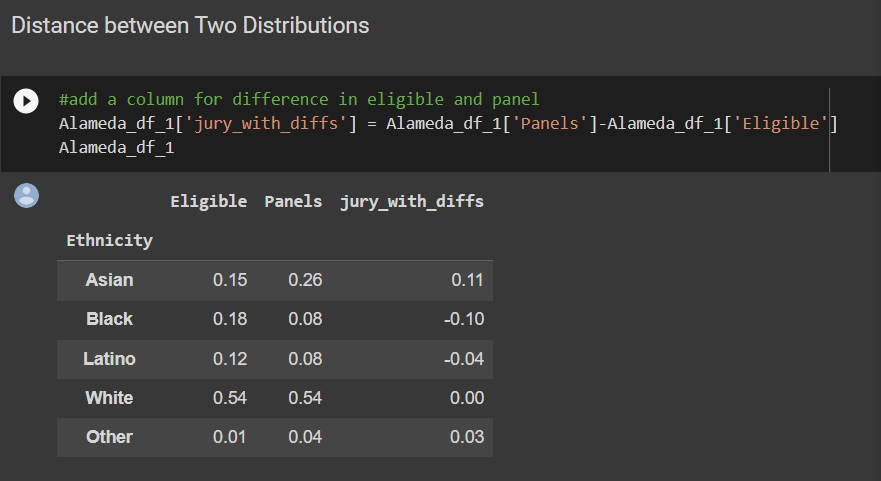
* Eligible: provides information of proportion of eligible members
* Panel: gives information about people currently chosen for the panel
* Ethnicity :gives information about different ethnic groups in alameda

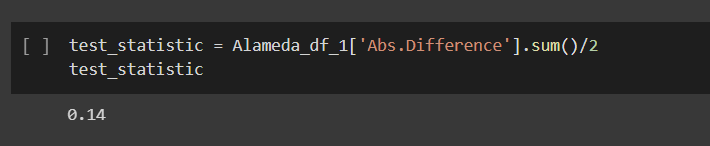
##### Null Hypothesis:-panels were selected at random from the population of eligible jurors.

Alternate Hypothesis:-panels were not selected at random



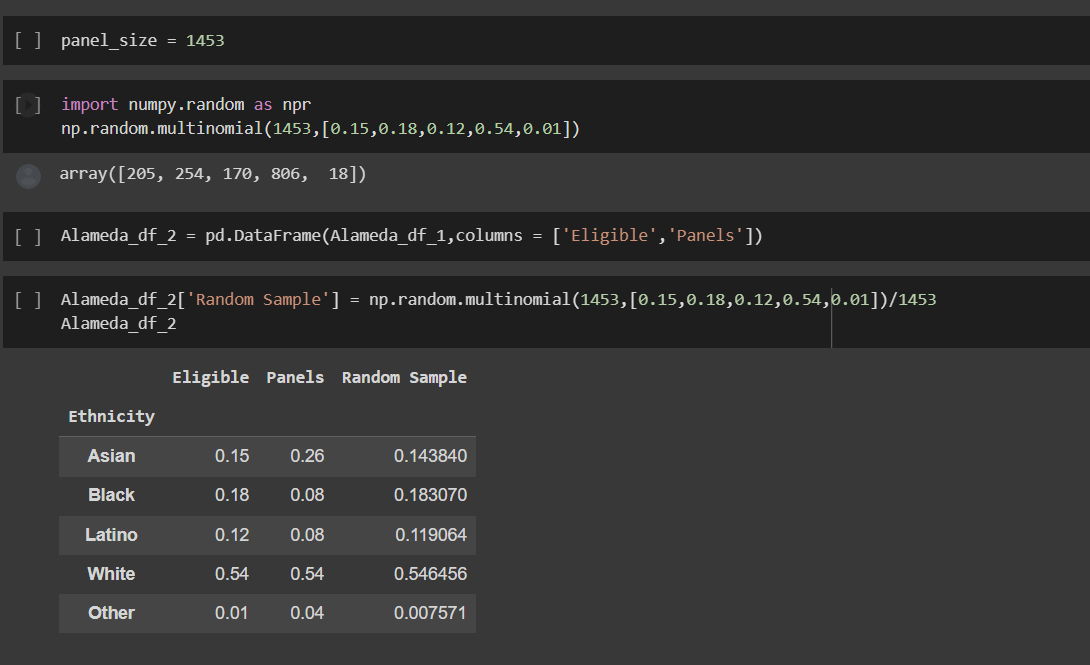
From the above bar graph we can analyze that the eligible and panel members are not equal for majority of the ethnicity. So we can the difference between them.



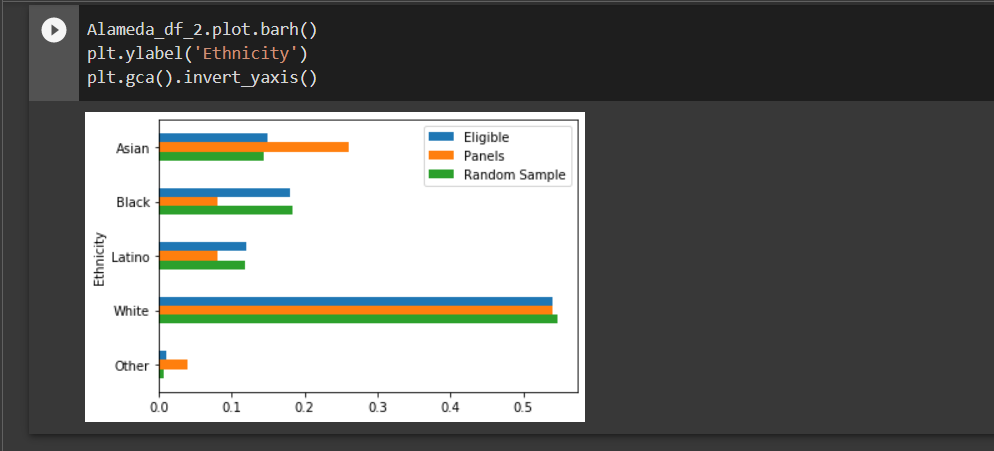


column Difference and notice that the sum of its entries is 0: the positive entries add up to 0.14, exactly canceling the total of the negative entries which is -0.14.

From the value we can of the fact that in the bar chart, the gold bars exceed the blue bars by exactly as much as the blue bars exceed the gold. The proportions in each of the two columns Panels and Eligible add up to 1, and so the give-and-take between their entries must add up to 0.

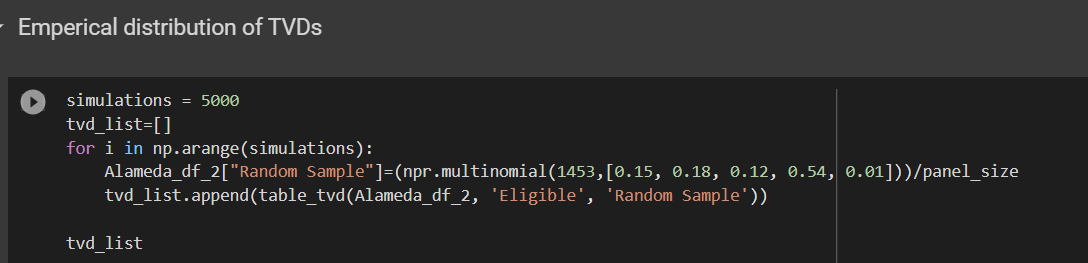


the distribution of the random sample is close to the distribution of the eligible population and is different from the distribution of the panels.

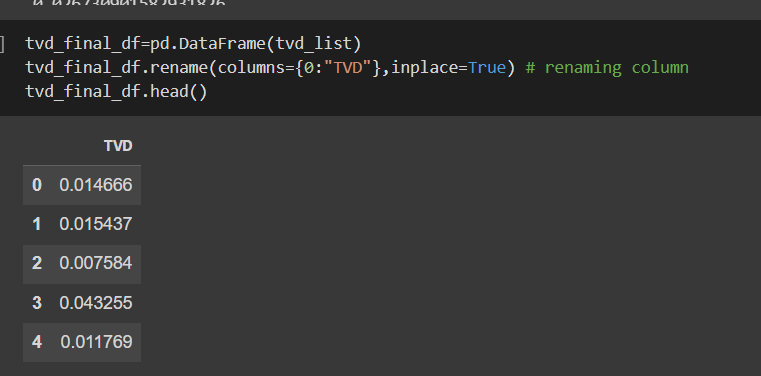


The overall distance between a random sample size of 1453 and the population of potential jurors varies for each row in df\_4.

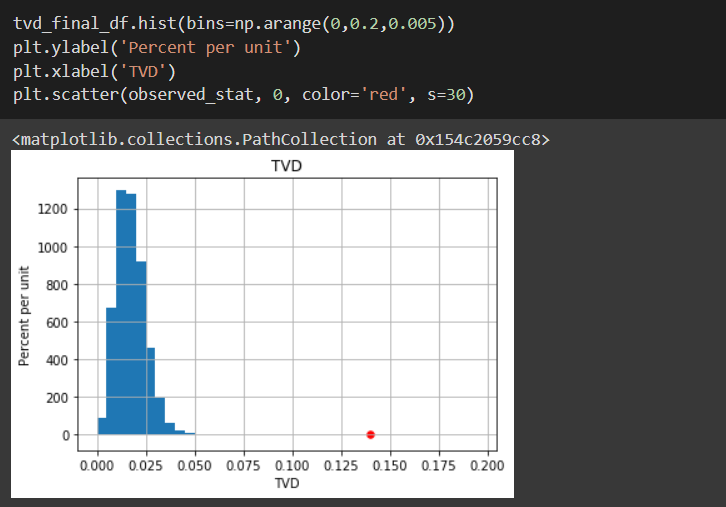
The green bar are closer in size to the blue bars than the orange bars are. The random sample resembles the eligible population, but the panels don't.



The overall distance between a random sample size of 1453 and the population of potential jurors varies for each row in df\_4.



each row of the column above contains the total variation distance between a random sample of size 1453 (the size of the panels) and the population of eligible jurors.



The overall distance between a random sample size of 1453 and the population of potential jurors varies for each row in df\_4.

The histogram of the simulated distances (pulling 1453 jurors from the pool of eligible applicants at random) rarely deviates from the eligible jurors by more than 0.05, in contrast to the observed data, which show a variance of 0.14, which is far off the distribution.

As a result, the statistics' projected values, which were based on the supposition of random selection, are in conflict with the facts in the panels. The panels thereby misrepresented the distribution made available to prospective jurors.

The null hypothesis is rejected