Assignment 3

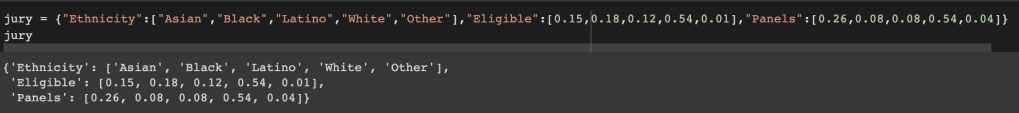
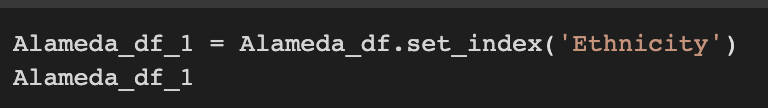
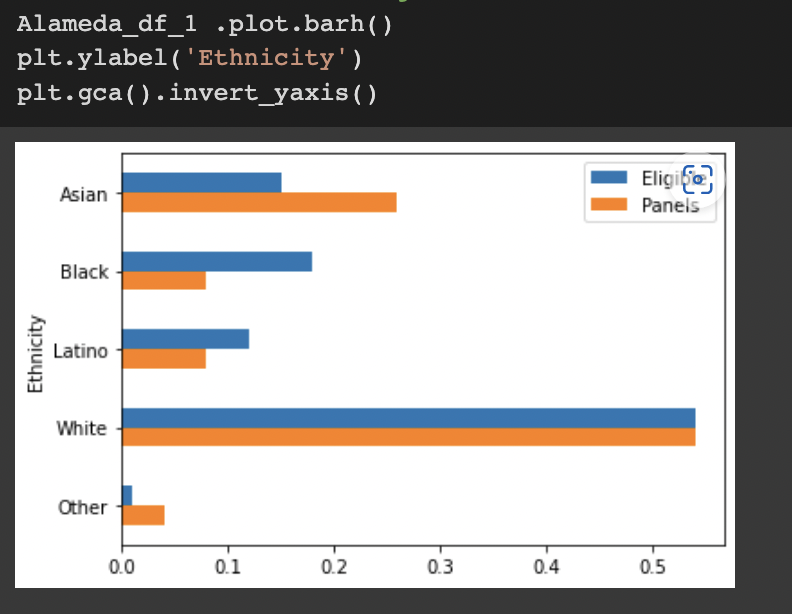
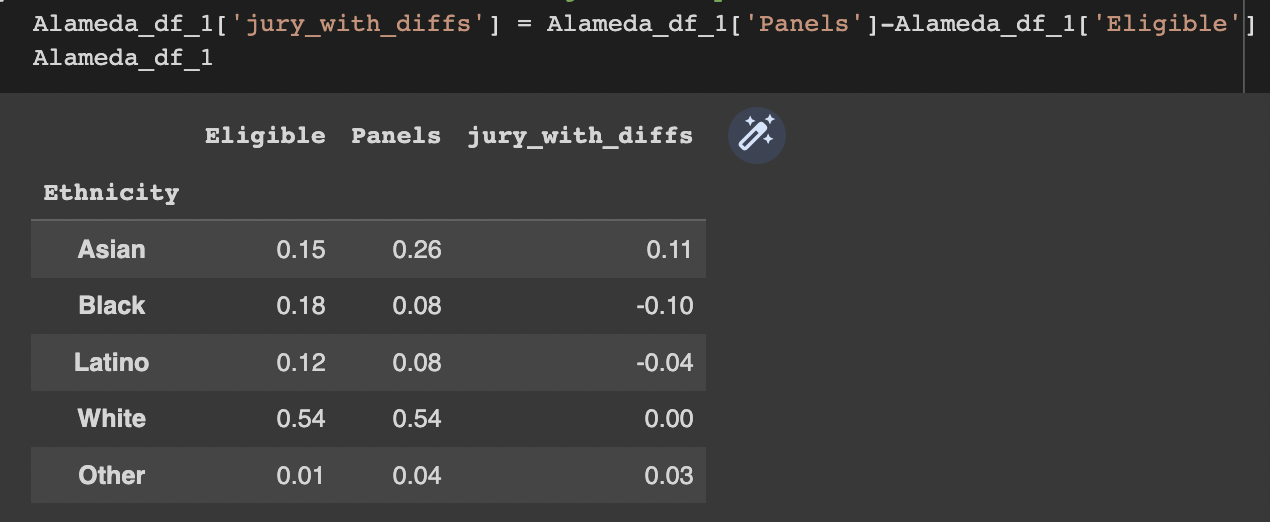
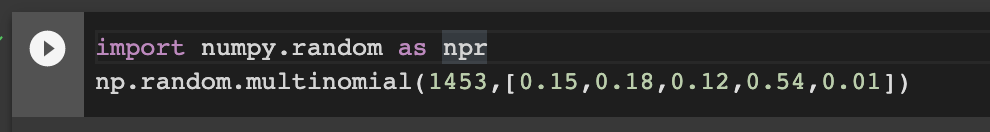
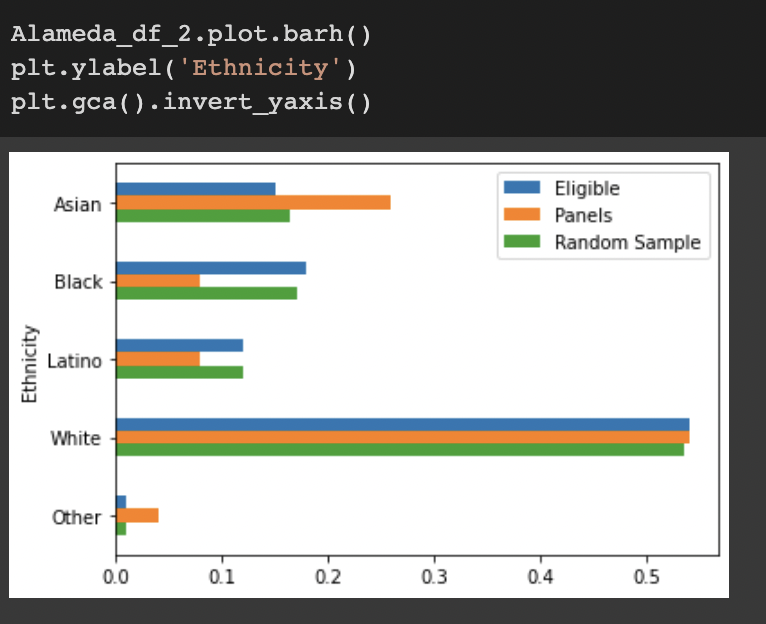
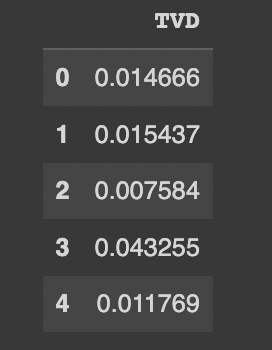
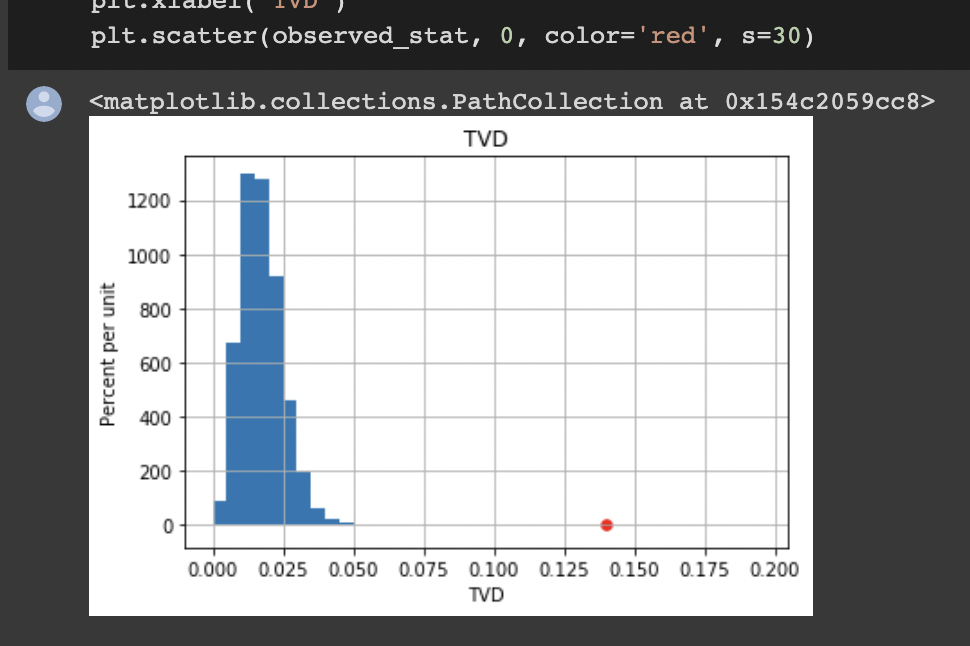
Case study

Kowshek

CB.EN.U4CSE20233

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

Alternate hypothesis: Panels weren’t selected at random

1. Creating the dataframe to work with   
     
     
     
   
2. Plotting a bar graph to see the difference in eligible jury members and the ones actually chosen clearly   
     
   
3. Creating a column that gives us the difference between the eligible people and the ones chosen   
     
     
     
   We convert the jury\_with\_diffs values to absolute difference and divide by 2 to get the total variation distance for all ethnicities   
     
   TVD = 0.14
4. Using random sampling with 1453 random people who are eligible jurors   
     
     
     
     
     
   On plotting the bar plot, we see the green line is closer to the orange line. From this we can see whites are the highest, followed by blacks.  
     
   The TVD we get with this is 0.009525120440467962   
     
   And we find the TVD for each ethnicity on comparing with random sample  
     
   
5. The empirical histogram of the simulated distances shows that drawing 1453 jurors at random from the pool of eligible candidates results in a distribution that rarely deviates from the eligible jurors' race distribution by more than about 0.05.  
     
     
     
   This shows that picking the jurors based on their eligibility shouldn’t result in a value different the distribution of the random 1453 people but it did. Hence, we can say the panels weren’t selected at random and may be biased.

Hence we reject the null hypothesis