**Examining Racial Discrimination in the US Job Market**

**Background**

Racial discrimination continues to be pervasive in cultures throughout the world. Researchers examined the level of racial discrimination in the United States labor market by randomly assigning identical résumés to black-sounding or white-sounding names and observing the impact on requests for interviews from employers.

### **Data**

In the dataset provided, each row represents a resume. The 'race' column has two values, 'b' and 'w', indicating black-sounding and white-sounding. The column 'call' has two values, 1 and 0, indicating whether the resume received a call from employers or not.

Note that the 'b' and 'w' values in race are assigned randomly to the resumes when presented to the employer

**Exercise**

Perform a statistical analysis to establish whether race has a significant impact on the rate of callbacks for resumes.

1. What test is appropriate for this problem? Does CLT apply?
2. What are the null and alternate hypotheses?
3. Compute margin of error, confidence interval, and p-value. Try using both the bootstrapping and the frequentist statistical approaches.
4. Write a story describing the statistical significance in the context or the original problem.
5. Does your analysis mean that race/name is the most important factor in callback success? Why or why not? If not, how would you amend your analysis?

Please use the link below for code review and for the details analysis

[racial\_discrimination.ipynb](https://github.com/rivasjmr/Springboard/blob/master/racial_discrimination.ipynb)

**Process**

The first step was doing anexploratory data analysis (**EDA**) to summarize the main characteristics.

Used the bootstrap hypothesis test and frequentist statistical approach

Applied a two-sample z-test

null and alternate hypotheses:

H0: W = B (call back rate)

H1: W ≠ B (call back rate)

**Conclusion**

CLT applies because we assume that the samples are representative of the population. The dataset size, 4870, is sufficient sample size and by randomly assigning identical résumés to black-sounding or white-sounding names assumed to be independent

What are the null and alternate hypotheses?

Answer: null and alternate hypotheses:

H0: W = B (call back rate)

H1: W ≠ B (call back rate)

The low value of the p-value indicates to reject the H0 hypothesis.

The callback rate differs based on race. But other factors can also be factors

in the callback rate that were not measured i.e. geography, sex, age, ...

To amend the analysis would look at other factors as mentioned above.