**Inferential Statistics**

In order to explore the data further and to determine any significant interactions within the data, I chose to explore four relationships. The first relationship I explored was whether there was a difference in the prevailing wages of applications that were certified or denied. For this analysis, I chose to use a t-test with a null hypothesis that there is not difference between the mean prevailing wages for applications that were certified or denied. The resulting alternative hypothesis is that there is a difference in the mean wages of the two groups. The mean wage of certified applications was $72,553.44 and the mean wage of denied applications was $2,245,810.64. I next performed a t-test with unequal variance as the sample sizes for the two groups were unequal (certified = 2,818,282; denied = 94,267). The resulting p-value of the t-test was extremely small, 7.67e-166. Such an incredibly small p-value leads me to reject the null hypothesis that the mean wage for certified and denied applications are equal.

The second analysis that I performed tested whether the wages for full-time and part-time positions were equal. As with the above analysis, I performed a t-test with unequal variance. There were 2,501,021 applications for full-time positions with a mean wage of $155,125.41 and 411,528 applications for part times positions with a mean wage of $68,549.35. Once again, the resulting p-value of 4.66e-102 was incredibly small. This again led me to reject the null hypothesis that there was no difference in mean wages for full-time and part-time positions.

Next, I wanted to test whether there was a difference between the populations of counties that for certified and denied applications. This could elucidate whether applications are more or less likely to be approved in either larger or smaller municipal areas. Again, I used a t-test with unequal variance. The average population for certified applications was 1,720,650 (n=2,721,739) and the average for denied applications was 1,993,234 (n=85107). Again, the p-value was incredibly small, essentially zero, leading me to reject the null hypothesis that there was no difference in county population size for applications that were certified and denied.

Finally, I looked at whether there was a correlation between prevailing wage and county population. I generated a least squares line as well as a Pearson R in order to determine if a correlation was evident. The correlation coefficient for this relationship is -0.00039. This indicates that there is essentially no correlation between the two variables.

All of the code for the above analyses can be found here: <https://github.com/Liptoni/Springboard/blob/master/H1B_Capstone/jupyter_notebooks/capstone1_Inf_Stats.ipynb>