Executive Summary

When opening a new business, one of the first major decisions made is the location. In New York City, home of over 200,000 businesses, this includes the choice of a borough for an operation (City of New York). An aspiring business owner in NYC might ask: is there a relationship between the boroughs and the number of licensed businesses in that borough? This analysis assesses if there is a statistically significant relationship between a business's address borough and its business license status. The null hypothesis for this analysis is that there is no significant relationship between borough and license status. The alternative hypothesis is that there is a significant relationship between borough and license status.

Data Analysis Process

The data extraction began by downloading the Legally Operating Businesses dataset from NYC OpenData in a CSV file. It was opened in a Jupyter Lab notebook environment using Python 3. The data was inspected, and many features were determined to be irrelevant to this study. Redundant and irrelevant features were dropped, leaving type, status, and borough to be further analyzed. Unique and null values were investigated, and rows with null values were dropped. In the step of dropping null values, the feature type was delimited to only one value, and was excluded from further analysis. Redundant values in the feature borough were renamed to preserve consistency in the data. After data cleaning was complete, the data was visualized graphically and numerically to determine distribution, and then a Chi-Square test of independence was performed.

Findings

The test resulted in a Chi-Square statistic of 2351.3, a p-value of 0.0, and a degree of freedom of five. For this analysis, the standard p-value threshold of 0.05 is sufficient (University of

Southhampton, 2022). As the p-value from the above test is less than the threshold, the null hypothesis is rejected. The alternative hypothesis is accepted. This analysis has determined that the borough in which a business operates does have a statistically significant relationship with the status of the business's license.

Limitations

One limitation of the data set were 93,262 rows with null values for the independent feature. The cleaning of these values delimited the study. Redundant features were a challenge in this data, which was overcome by selecting the more descriptive feature containing the same information to preserve value. A limitation of the Chi-Square test is its lack of ability to determine causality (University of Utah Department of Sociology, 2022). In other words, the test can tell us there is or is not a relationship, with no insight as to why.

Proposed Actions

I recommend considering Manhattan as a business location. In addition to showing a statistically significant relationship between borough and license status, this study also shows that Manhattan has a high number of active licenses compared to the Bronx, Staten Island, and outside NYC. Yet, it has room to grow in active license compared to Brooklyn and Queens. This makes it a good candidate location for a new active business license (Heaslip, 2020). While the relationship between city borough and license status is statistically significant, further study is needed to determine a cause for that relationship. One recommendation for further study would be to include more features in further analysis to determine potential relational causes. For instance, how many of the licenses labeled inactive in each borough are also expired? A second recommendation for further study is to use the Chi-Square test findings as potential feature selection for a classifier model to gain further business insights.

Expected Benefits

This analysis expects to provide business owners with additional insight into the location of their business. The Chi-Square test provides statistical evidence of a relationship between borough and the number of active business licenses, indicating that further investigation would be beneficial.

References

City of New York. (n.d.). *Small Business First.* Retrieved January 11, 2022, from

https://www1.nyc.gov/assets/smallbizfirst/downloads/pdf/small-business-first-report.pdf

Heaslip, E. (2020, January 10). *Location, Location, Location: How to Decide Where to Start Your Business*. Retrieved from uschamber.com:

https://www.uschamber.com/co/start/strategy/where-to-start-a-business

University of Southhampton. (2022). *Chi Square*. Retrieved from University of Southhampton:

https://www.southampton.ac.uk/passs/full_time_education/bivariate_analysis/chi_square.

page

University of Utah Department of Sociology. (2022). The Chi-Square Test for Independence.

Salt Lake City, Utah, USA.