Beyond L&M and CX: Are There Other Influences on Restaurant Prices and Ratings?

L&M: labor and materials CX: customer experience

Project Team Members: Baljit Alang, Rick Cavalla, Marc Donatiello, Haarris Haque, Mark Visco

**Initial Question**

Do median age, average household income and percent of foreign born in a zip code have a relationship with the average price levels and average customer ratings of restaurants in that zip code?

**Hypothesis Statement**

If a zip code’s average household income is low, median age is low or percent of foreign-born is high, then the average customer rating and average price level of restaurants in that zip code will be lower than a zip code with a high average household income, high median age and low percent of foreign-born.

**Null Hypothesis**

A zip code’s average household income, median age and percent of foreign-born are unrelated to average customer rating and average price level of restaurants in that zip code.

**Questions**

1. Do higher income zip codes have higher-priced restaurants?
2. Do higher income zip codes have higher-rated restaurants?
3. Do zip codes with a younger median age have lower-priced restaurants?
4. Do zip codes with a younger median age have lower-rated restaurants?
5. Do zip codes with a higher percentage of foreign-born people have lower-priced restaurants?
6. Do zip codes with a higher percentage of foreign-born people have lower-rated restaurants?

**Data Scope - Essex County, NJ Analysis**

Demographic and restaurant data for 31 zip codes in the 21 municipalities in Essex County, NJ:

1. Belleville 12. Millburn
2. Bloomfield 13. Montclair
3. Caldwell 14. Newark
4. Cedar Grove 15. Nutley
5. East Orange 16. Orange
6. Essex Fells 17. Roseland
7. Fairfield 18. Short Hills
8. Glen Ridge 19. South Orange
9. Irvington 20. Verona
10. Livingston 21. West Orange
11. Maplewood

The data for the zip codes in Essex County was visualized in six scatterplots with each plot point representing one of the zip codes.

**Data Scope - NJ State Analysis T-Test**

Demographic and restaurant data for the Top 25 and Bottom 25 zip codes for household income, median age and foreign-born percentage in the state of New Jersey. For each of the three demographic measures, a t-test was the method used to compare the Top 25 to the Bottom 25 to see if there was a statistically significant difference between the two groups in terms of average price and average customer rating. For example, a list of the "average price" for each of the 25 youngest zip codes was the first parameter of the t-test, and a list of the "average price" for each of the 25 oldest zip codes was the second parameter.

**Data Sources**

*Google Places (Maps) API* - for the following data for restaurants in each zip code:

Price ratings

Customer ratings

*US Census API* - *2017 American Community Survey* for the following demographic data for each zip code in Essex County:

Median Age

Foreign Born

Household Income

**Data Limitations**

*Google Places API*

Query results are limited to the 20 most proximate results to the zip code reference point.

*US Census API* - *2017 American Community Survey*

This analysis references zip codes as opposed to municipalities because the US Census Data is organized by zip code. This resulted in the sample including multiple zip codes for some municipalities.. Also, the geographic range of zip codes sometimes extends across two or more municipalities.

Note: The *Yelp! Fusion API* was not utilized because queries “only return business info for businesses that have Yelp user-generated content added to them. Specifically, they must have at least one review." Data is not provided for restaurants that do not have at least one customer review.

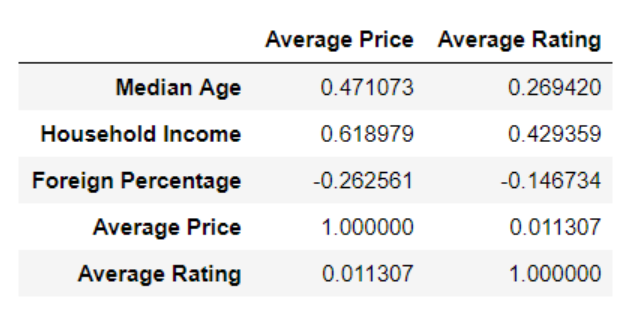
**Summary of Findings - Essex County, NJ**

In general, the R-values from the Pandas corr() function and the scatterplots depicting zip codes in our Essex County dataset indicate a relationship between restaurant prices and income, median age and foreign-born percentage. The strongest correlation is between household income and average price. The R-values and plots show little-to-no relationship between customer ratings and household income, median age and foreign-born percent.

In summary:

* Lower income zip codes generally have lower-priced restaurants.
* Younger zip codes generally have lower-priced restaurants.
* Zip codes with a relatively higher-percentage of foreign-born residents generally have lower-priced restaurants.

Here are the R-values from the Pandas corr() function:



Note 1: See further below for a more detailed summary, with scatterplots, of the Essex County data.

Note 2: See the Appendix at the end of this document for the full set of raw data used in the Essex

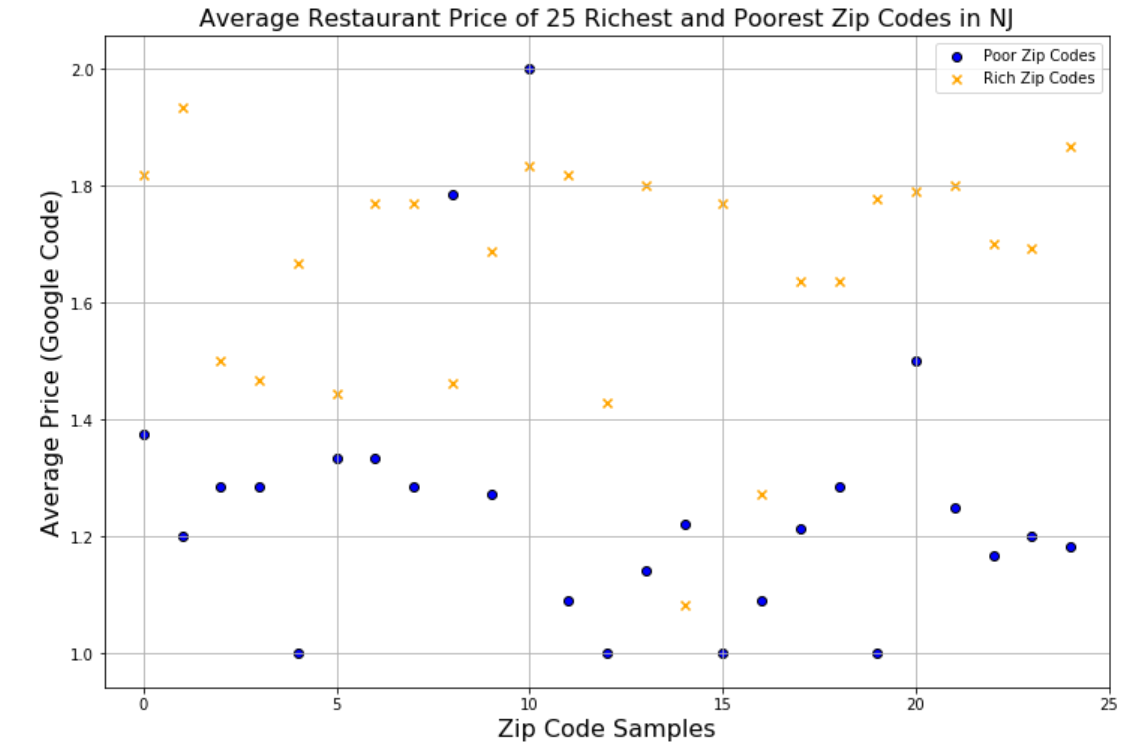
County analysis.

**Summary of Findings - State of New Jersey (Top 25/Bottom 25)**

The six t-tests that were conducted revealed only the difference between household income and average price was statistically significant with higher income zip codes having higher-priced restaurants and lower income zip codes having lower-priced restaurants.



The scatterplot for household income and average price shows a significant difference between the data for the lower income zip codes and the higher income zip codes.



**Detailed Summary of Findings - Essex County**

1. ***Income v. Restaurant Price Level***

*Do higher income zip codes have higher-priced restaurants?*

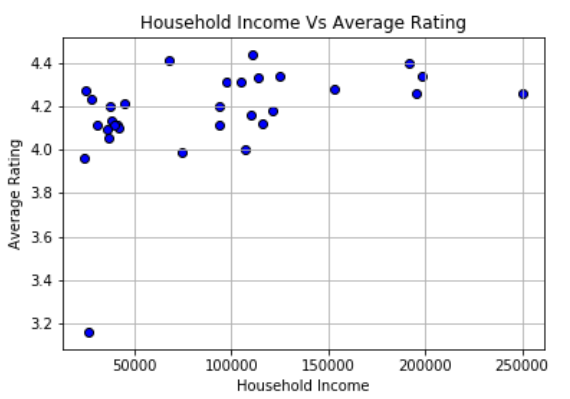


Our data showed zip codes with average household income of $50,000 or less have lower-priced restaurants when compared with zip codes with average household income of more than $50,000. An increase in the average price of restaurants is seen as a zip code’s average household income increases. Interestingly, zip codes in the dataset with households earning $125,000 or more only have restaurants with a price rating of 1.5 or higher.

Average household income in our dataset ranged from $0-$250,000. The average price ratings were based off Google Places’ dollar sign rankings of one dollar sign ($) to four dollar signs ($$$$). Our dataset, however, did not have any restaurants rated three dollar signs ($$$) or higher. The outlier for Household Income is Short Hills at $250,000.

***2. Income v. Restaurant Ratings***

*Do higher income zip codes have higher-rated restaurants?*

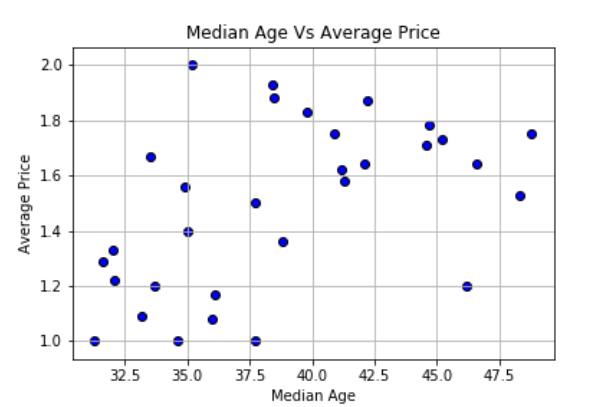


The data shows the average customer rating at restaurants in all but one of the zip codes was roughly 4.0 or higher regardless of average household income level.

Average household income in our data set ranged from $0-$250,000. The average rating is scaled from 3 to 4.6 with the maximum rating possible being a 5. The outlier for Household Income is Short Hills at $250,000. The outlier for Average Rating is a Newark zip code at 3.16.

3. ***Age v. Average Price***

*Do zip codes with a younger median age have lower-priced restaurants?*



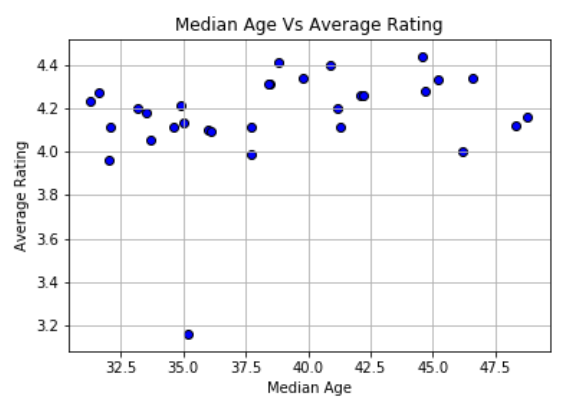
Our analysis reveals the median age of a zip code is related to average price of the restaurants in that zip code. Price appears to increase with age with the exception of a slight decrease beyond 45 years of age.

All but two Essex County zip codes with a median age of 30-37.5 have restaurants with an average price rating of 1.0-1.4. All but three of the zip codes with a median age between 37.5-45 have restaurants with an average price rating of roughly 1.6-2.0 or higher. Zip codes with a median age of 45-50 show a tapering off in average price rating to 1.2-1.8.

Ages were scaled from 30-50 years. The average restaurant prices are scaled from one dollar sign ($) to four dollar signs ($$$$). Our data did not contain any zip codes with an average price rating of less than three dollar signs ($$$).

4. ***Age v. Average Rating***

*Do zip codes with a younger median age have lower-rated restaurants?*

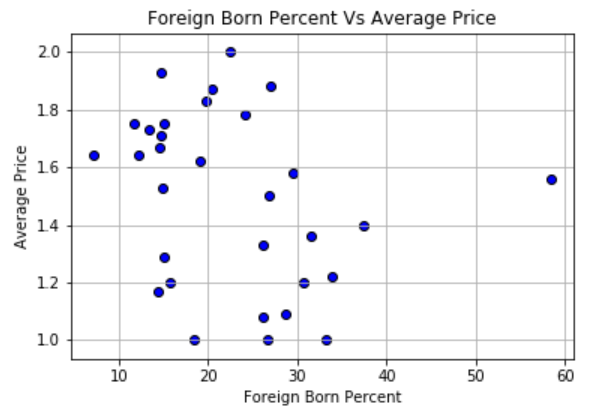


Overall, our research revealed there is not a relationship between the median age in a zip code and average ratings of restaurants in that zip code. The data does show zip codes with a median age of about 38 or higher have slightly higher ratings than zip codes with a median age below 38.

Median age is scaled from 30-50 years. Average customer rating is measured on a scale of 1-5, with 5 being highest. The outlier for Average Rating is a Newark zip code at 3.16.

5. ***Foreign Born v. Average Price***

*Do zip codes with a higher-percentage of foreign-born residents have lower-priced restaurants?*

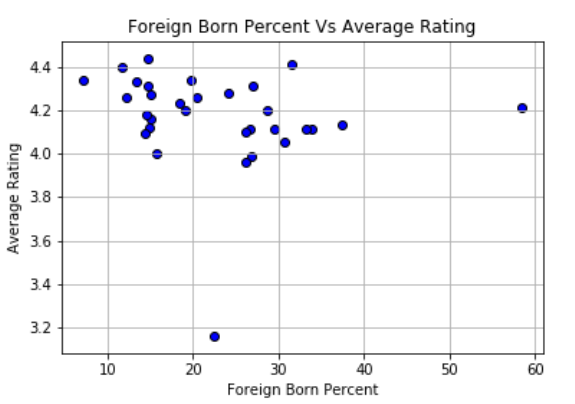


Our analysis illustrated zip codes with fewer foreign-born residents have higher-priced restaurants and zip codes with a higher percentage of foreign-born residents have lower-priced restaurants.

Notes: Zip codes are scaled from having 0%-60% foreign-born residents. Average restaurant prices are scaled from one dollar sign ($) to four dollar signs ($$$$). Our data did not contain any zip codes with an average price rating of less than three dollar signs. The outlier for Foreign Born Percent is a zip code in Newark at 58.36%.

6. ***Foreign-born Percentage v. Average Rating***

*Do zip codes with a higher percentage of foreign-born residents have lower-rated restaurants?*



The data does not show a relationship between the foreign-born percentage of a zip code and the quality of restaurants in that zip code. Regardless of the foreign-born percentage, all but one of the zip codes have average restaurant ratings between roughly 4.0 and 4.5.

Notes: Restaurants are rated on a scale from 1-5, with 5 being the highest possible rating. The outlier for Foreign Born Percent is a zip code in Newark at 58.36%.

**Conclusion**

Our research leads us to the conclusion, based on the r-values, p-values and scatterplots, that there is a relationship between average restaurant prices and household income, median age and percentage of foreign-born. The strongest correlation is household income to average price. The results of our analysis indicate there is little to no correlation between restaurant ratings and the three demographic measures.

**Appendix - Raw Data**

Following is the raw data used in our analysis. This is the data behind the scatterplots in this document.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Zip Code** | **City** | **Median Age** | **Foreign Born** | **Household Income** | **Average Restaurant Price** | **Average Restaurant Rating** |
| 07003 | Bloomfield | 37.7 | 12653 | 74961 | 1.5 | 3.99 |
| 07004 | Fairfield | 46.2 | 1188 | 107417 | 1.2 | 4 |
| 07006 | Caldwell | 45.2 | 3389 | 114096 | 1.733333333 | 4.33 |
| 07009 | Cedar Grove | 48.3 | 1862 | 115926 | 1.533333333 | 4.122222222 |
| 07017 | E. Orange | 37.7 | 9478 | 41322 | 1 | 4.105263158 |
| 07018 | E. Orange | 36 | 7455 | 42335 | 1.083333333 | 4.1 |
| 07021 | Essex Fells | 46.6 | 148 | 198750 | 1.636363636 | 4.345 |
| 07028 | Glen Ridge | 42.1 | 931 | 195714 | 1.636363636 | 4.257894737 |
| 07039 | Livingston | 44.7 | 7139 | 153381 | 1.777777778 | 4.283333333 |
| 07040 | Maplewood | 39.8 | 4785 | 125036 | 1.833333333 | 4.34 |
| 07041 | Millburn | 38.5 | 1949 | 104786 | 1.875 | 4.305263158 |
| 07042 | Montclair | 38.4 | 3822 | 97920 | 1.933333333 | 4.305263158 |
| 07043 | Montclair | 40.9 | 1467 | 191449 | 1.75 | 4.4 |
| 07044 | Verona | 44.6 | 2039 | 110745 | 1.714285714 | 4.436842105 |
| 07050 | Orange | 35 | 11213 | 38506 | 1.4 | 4.133333333 |
| 07052 | W. Orange | 41.3 | 13939 | 93954 | 1.583333333 | 4.105555556 |
| 07068 | Roseland | 48.8 | 887 | 110208 | 1.75 | 4.16 |
| **Zip Code** | **City** | **Median Age** | **Foreign Born** | **Household Income** | **Average Restaurant Price** | **Average Restaurant Rating** |
| 07078 | Short Hills | 42.2 | 2632 | 250001 | 1.866666667 | 4.26 |
| 07079 | S. Orange | 33.5 | 2364 | 121637 | 1.666666667 | 4.185 |
| 07102 | Newark | 32 | 3229 | 24353 | 1.333333333 | 3.961111111 |
| 07103 | Newark | 31.3 | 5845 | 27859 | 1 | 4.233333333 |
| 07104 | Newark | 33.7 | 13468 | 36547 | 1.2 | 4.047368421 |
| 07105 | Newark | 34.9 | 29777 | 45227 | 1.555555556 | 4.210526316 |
| 07106 | Newark | 33.2 | 9385 | 37421 | 1.090909091 | 4.195 |
| 07107 | Newark | 32.1 | 11642 | 30779 | 1.222222222 | 4.105555556 |
| 07108 | Newark | 31.6 | 3208 | 25192 | 1.285714286 | 4.266666667 |
| 07109 | Belleville | 38.8 | 11036 | 67782 | 1.363636364 | 4.410526316 |
| 07110 | Nutley | 41.2 | 5350 | 94128 | 1.625 | 4.2 |
| 07111 | Irvington | 34.6 | 17829 | 39734 | 1 | 4.105263158 |
| 07112 | Newark | 36.1 | 3676 | 35765 | 1.166666667 | 4.088888889 |
| 07114 | Newark | 35.2 | 2919 | 26326 | 2 | 3.161111111 |