[CUISINE COMPARISON OF COLLEGE CITIES] April 15, 2020

Data Science Capstone Project by IBM/Coursera
Prepared by Mary Ann Peterman

1. Introduction

1.1. Summary

In this project I will examine cuisine and restaurant styles in American college towns. Cities with major universities also have other colleges. This data analysis seeks to determine if college town restaurants share popularity in dining types. Answers are essential for a successful new restaurant or food truck enterprise in these communities. Restaurants, in all towns across the melting pot of America, represent a wide variety of cuisines. Dining and fun eats are a draw to college towns in addition to other activities and venues. What can we learn from them?

1.2. Business Problem

Choice of cuisine is an initial step in restaurant design and planning. Marketers and developers for restaurant industry would benefit from this analysis. Both locations are expected to serve a dominance of typical American fare. What other cuisines do similar restaurant visitors share? What kind of restaurant would be popular in a college town? Are there cuisine trends? What types of restaurants are highly rated? From this inquiry, a new restaurant model could be chosen or assessment to move ahead.

2. Data

2.1. City Selection

Initial research was made with city-data.com available data on two particular cities;
South Bend, Indiana and **Tallahassee, Florida** are comparable midsized college towns. South Bend, Indiana was chosen for having the **University of Notre Dame** located just to the north in unincorporated Notre Dame. Tallahassee is home to **Florida State University** which is just to the west. Both colleges are adjacent and close to active downtowns. Preliminary examinations show that metro populations of both are approximately 320,000 and demographics are similar. Both metro areas include more than 3 other traditional colleges. Both cities have very similar outdoor and cultural venues and cost of living. Also, both have international airports and proximity to interstate traffic.

2.2. Data Sources

Generational age data for each city was taken from the U.S. Census Bureau here. Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-Year Estimates, Table: S0101, Age and Sex. *(Note: The city specific census tables used are no longer available due to reorganization of the website on March 30, 2020.)* Census data age information was segmented into 5 year incremented groups, not what needed. Children under 18 were excluded from this study. Metropolitan populations were not studied, only city limits.
Github was used for Capstone project data repository. Geographical coordinates for locations of universities and cities and maps were created with from geopy.geocoders, Nominatim (OpenStreetMaps).

Foursquare and Zomato APIs returned approximately more than 160 restaurants within 8,000 meters of downtowns. Foursquare returned; restaurant name, category, lat, longs. Foursquare and Zomato APIs returned 178 restaurants within 8,000 meters of each downtown. Category is a restaurant type differentiating from Zomato's cuisine information only. Zomato returned; venue, latitude, longitude, ratings and cuisines. Ratings are a normalized distribution 5 point scale. Restaurants that are not rated were excluded from the subject data. Cuisines list out each type of cuisine served at the restaurant venue.

2.3. Data Cleaning

Population data was not grouped into generation groups on any site explored. Data downloaded was grouped into the desired generational groups for each city in Excel and saved on Github as .csv for import. South Bend, Indiana did not include this unincorporated area. Notre Dame's, an unincorporated city, population data had to be downloaded and added to South Bend for population counts.

Each city's data had to be cleaned separately and described as followed. Foursquare API supplied restaurant name, category, latitude and longitude. However, to obtain ratings and cuisines premium calls were required. Zomato API, was much easier to access. After doing initial API calls and comparing all information from these sources it was noted that the information from Foursquare was more accurate as well as the names. Further verification of duplicates was checked against restaurant searches. Data was combined by mapping to each venue and location.

3. Methodology

3.1. Get rating and trending information on Foursquare and Zomato APIs

Foursquare's ratings for venues are only available for premium calls. (However, it is has less errors in venue names than Zomato.) My goal is to obtain a thorough list of restaurant types and ratings. A more advantageous solution is using Zomato as a data source which gives rating with no added cost. In order to match venues between the two acquired lists and for mapping, latitude and longitude is requested. Total restaurants used in study analysis were South Bend, 157 and Tallahassee, 171.

In researching API's rating information to harvest, it was learned that 5 point

Foursquare categories are = restaurant types. Restaurant types are listed for each venue. Objective was planning information for

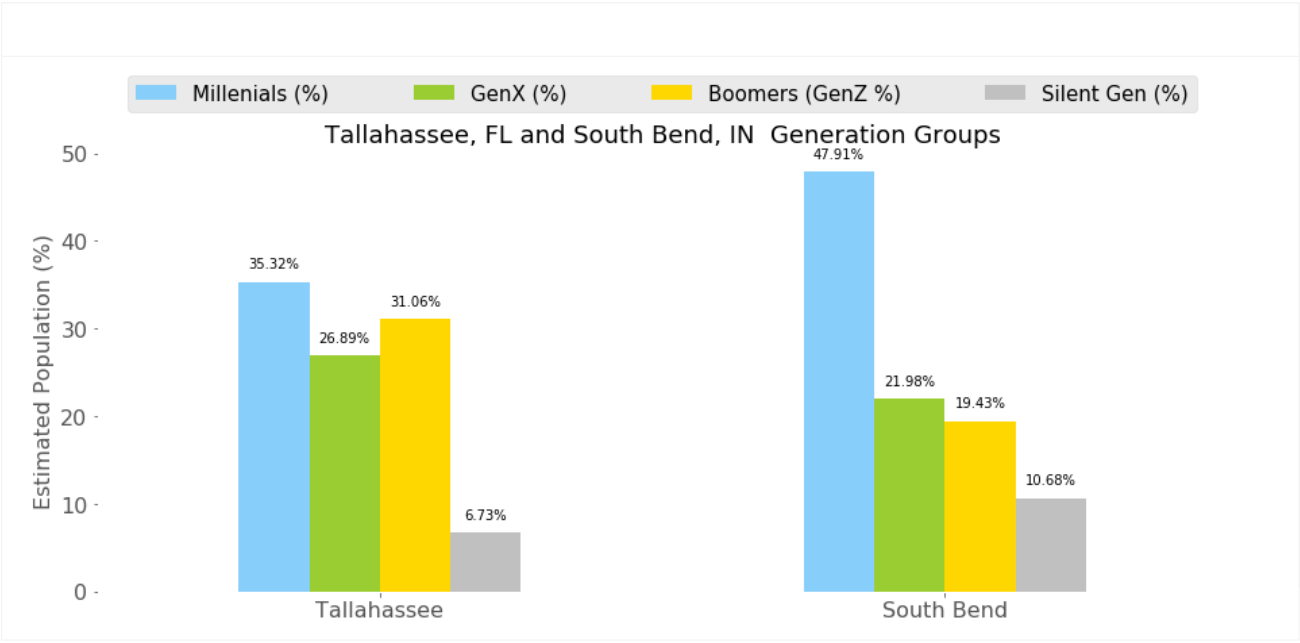
Exploratory analysis looked at the quantity of the cuisines by ratings.

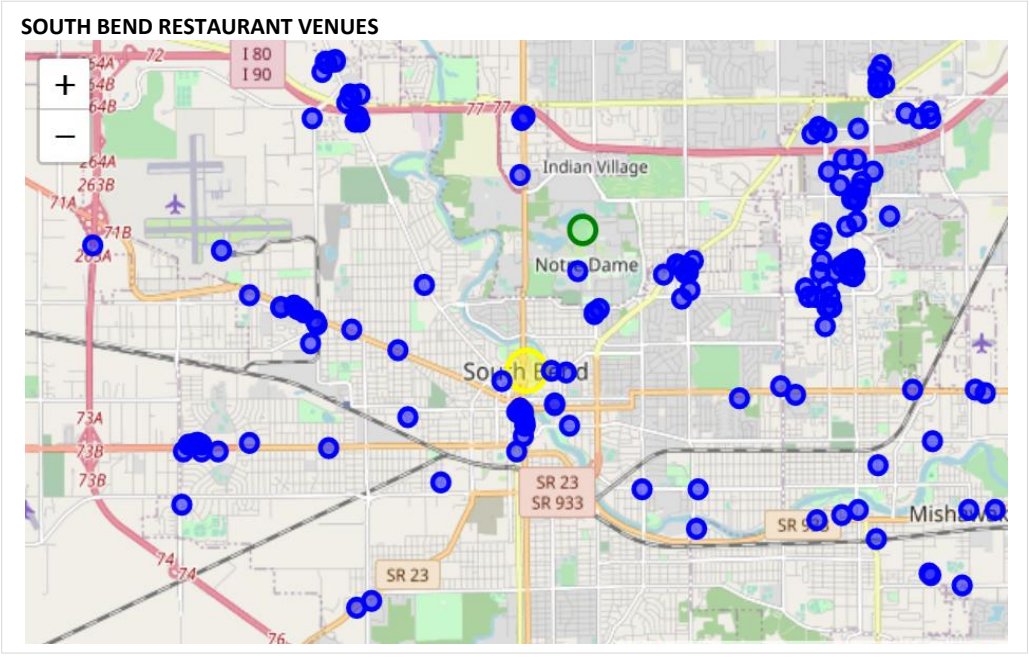
3.2. Analysis

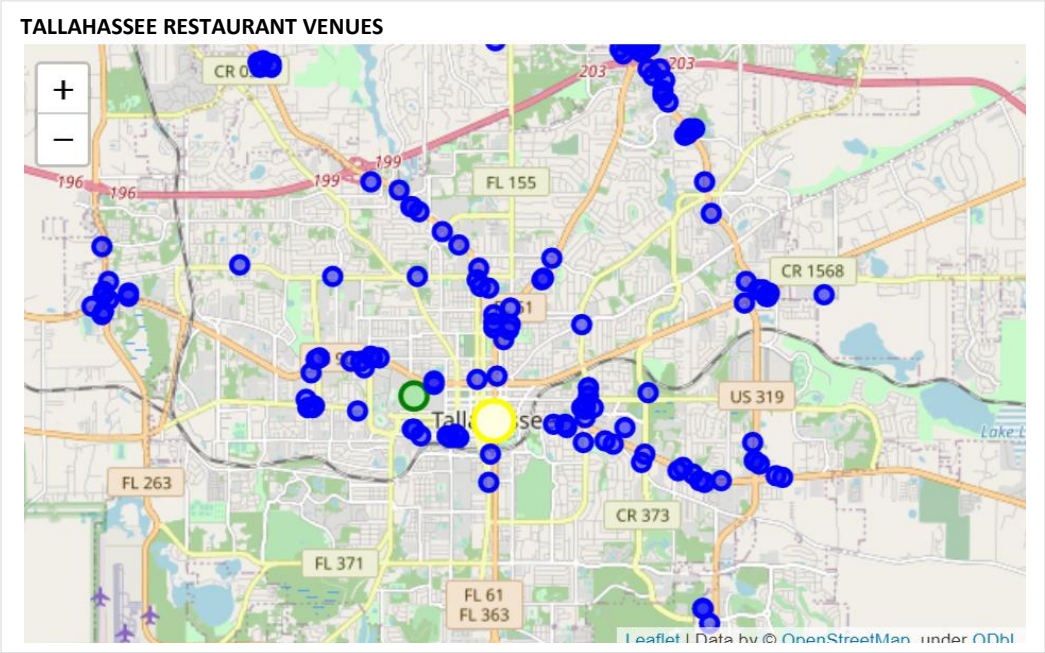
A. Customer Demographics

First objective is to verify both cities are comparatively sized and populated. General groups are very important to a food service marketer. The 3 primary generation groups (Millennials, GenXers and Boomers) make up 73% of dollars spent. Summarized demographic information was already readily available. However, generational age groups were not. Census data had to be grouped accordingly and visualized with bar chart for each city. Percentages are calculated for comparison between two different total populations.

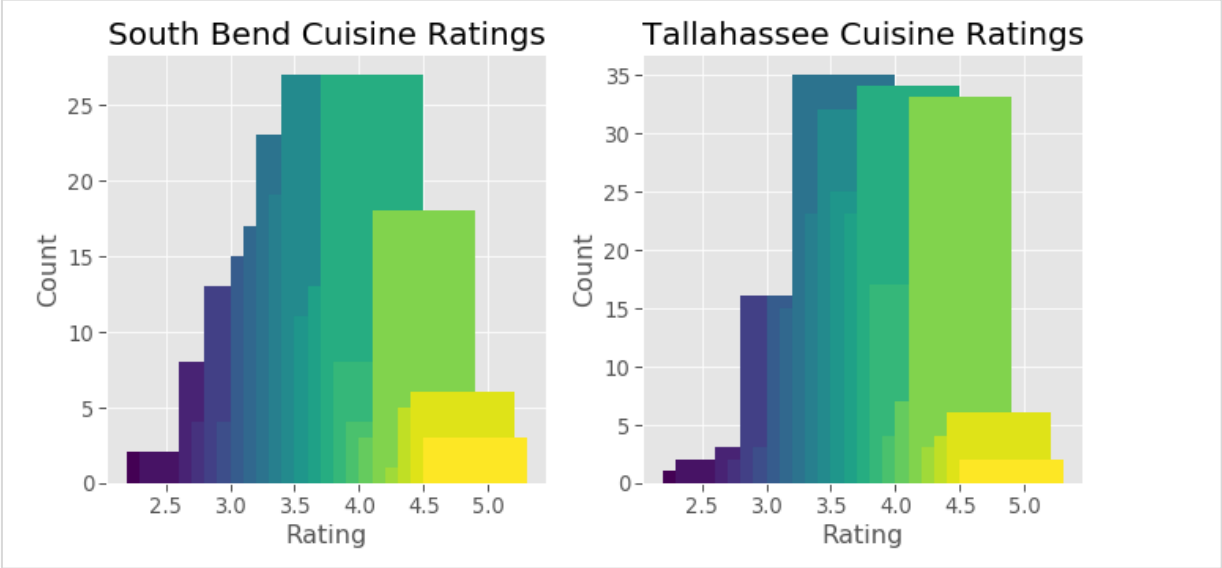
City	Millenials (%)	GenX (%)	Boomers (GenZ %)	Silent Gen (%)
Tallahassee	35.32	26.89	31.06	6.73
Southbend	47.91	21.98	19.43	10.68







3.4. Cuisine Rating Spread



4. Results

4.1. Population

The hypothesis of these comparisons was that there would be a clear dominate millennial population in college towns who frequent restaurants. Population segmentation analysis revealed that there are proportionately more millennials in South Bend than Tallahassee. Analysis shows millennials make the largest population group in both cities. However, Tallahassee's population is more evenly distributed between the focus generation groups.

B. Restaurants

Restaurant ratings do not correlate to the frequency of restaurants in either city. Trending information is only gathered on a weekly basis so conscientious reviews serve as a best test available at this time. Specific clarity on methodology of how APIs collect trending is not publically available. Check-in and voting data is likewise not clear and different methodology between restaurant venue APIs.
Tallahassee has 12 more restaurants than South Bend within 3.28 miles/8500 meters.

C. Cuisine

TALLAHASSEE

	cuisine	counts
27	American	37
38	Fast Food	21
16	Sandwich	20
4	Burger	18
46	Sandwich	15
25	Vegetarian	14
41	Italian	13
32	Breakfast	11
15	Pizza	10
45	Pizza	10

SOUTH BEND

	cuisine	counts
37	Fast Food	33
50	Thai	14
19	Sandwich	14
41	Italian	12
29	American	12
35	Chinese	11
46	Pizza	10
47	Sandwich	9
18	Pizza	8
45	Mexican	8

5. Discussions

5.1. Observations

Several findings break common logic assumptions. College towns were selected to find a larger millennial population. Understanding the restaurant needs of the 3 primary dollar spending generations will help marketing be focused.

The quality of the Zomato API data for South Bend had a lot to be desired. The data is unusable. For example the cuisine of burger joints, fried chicken joints and American restaurants were identified as Thai. We ended up with a disproportional amount of Thai restaurants. We cannot compare the college town cuisine

Restaurants that are not favored have a high presence as much as lower rated restaurants. A successful business would take note from this study what foods and restaurant types are better favored

5.2. Recommendations

Taking this a step further I need a better method to determine guest likes to refine cuisine choices. A deeper study could be made to find more sources. Identifying, if there are any, unique customer psychographics would be enlightening. Also, knowledge of marketing to the generation groups needs to be examined and restaurant models identified for business.

6. Conclusion

Cuisine comparisons were mixed with no clear trend identified.

The API cuisine data harvested from Zomato proved to be inaccurate so no comparison could be drawn.

Both city's data showed that the quantity of restaurants has no correlation on highly rated ones.