IDENTIFYING NEW EXPANSION OPPORTUNITIES FOR NY BASED PATSY'S PIZZERIA

Coursera Capstone Project

Farukh Zaripov
March 11, 2019

AGENDA

- Helping Parcy's Pizzeria to expand to new locations

- Data Acquisition

- Analytic Approach

- ML Findings

- Recommendations

HELPING PARCY'S PIZZERIA TO EXPAND TO NEW LOCATIONS

- Since its origination in 1933, Patsy's was able to build a loyal client base and became a regular hang out place for many celebrities including Frank Sinatra, Dean Martin and Tony Bennett.

- Despite its popularity, Patsy's still operates at a single location at

- 2287 First Avenue New York, NY 10035.

- The ML model will allow comparing various neighborhoods (on zip code level) based on common sets of features and recommend the best match to current successful location.

DATA ACQUISITION

Data sources that are used for feature engineering

Lat/long for all US zip codes (https://gist.github.com/erichurst/7882666)

    The data was used primarily to generate a map with the recommended new locations.

Database of active US businesses based on zip codes. (https://www2.census.gov/programs-surveys/cbp/datasets/2016/zbp16detail.zip?#)

    The file provides a breakdown of US businesses by zip code based on number of employees

IRS tax return for every US zip codes (https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2016-zip-code-data-soi).

    The file provides number of returns, which approximates the number of households and groups total tax returns into the following bins based on the income.
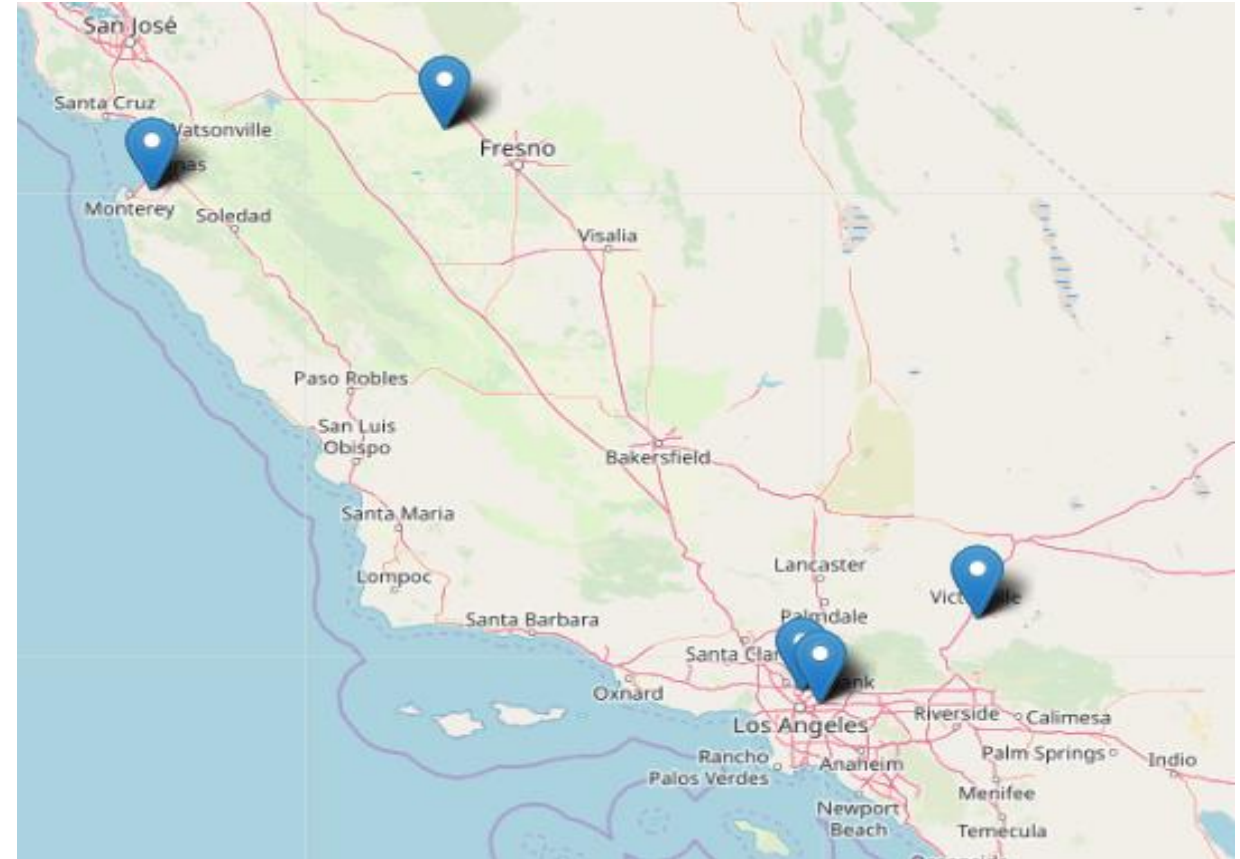
The data from all sources was combined, merged and aggregated into one file with 19 columns

ANALYTIC APPROACH

- Hierarchical Clustering was used to identify similarities between different geo locations and place them into appropriate clusters.

- Advantages it provides for this particular case:
  - No need to worry about identifying clusters and let the model to do the job
  - It allows visualizing the distances between different zip codes and helps to immediately identify the closest zip codes to the original location based on their feature similarities.

ML FINDINGS

- The same ML model was used three times against decreasing subset of data. After each run, I would focus on the cluster that contain the actual NY location and run it again for the subset of the data.

- The final cluster contained only 6 members including original location and five new potential locations.

ML FINDINGS (CONTINUED)

The model produced result that greatly correlates to the existing location patterns.

ZIPCODE	10035	91205	91803	92395	93637	93955
1 USD_25	7040	7350	5910	7660	6120	5470
2 USD_25_50	3890	4240	3770	4110	4520	4470
3 USD_50_75	1950	1980	1900	2040	2110	2210
4 USD_75_100	910	1090	1180	1100	1090	1150
5 USD_100_200	1020	1220	1560	1160	1210	1270
6 USD_OVER_200	320	240	310	220	270	200
7 BUS_TOTAL	3918	4488	4682	3608	3790	3546
8 EMPL_1_4	2256	2832	2954	1790	1738	1828
9 EMPL_5_9	648	696	708	780	876	770
10 EMPL_10_19	414	468	504	516	540	426
11 EMPL_20_49	348	348	342	354	408	342
12 EMPL_50_99	150	102	108	84	138	132
13 EMPL_100_249	78	30	48	54	60	36
14 BUS_EMPL_250_499	12	12	18	18	18	6
15 BUS_EMPL_500_999	6	0	0	12	6	6
16 BUS_EMPL_OVER_1000	6	0	0	0	6	0
Correlation Coefficient		0.997493	0.977515	0.994565	0.987003	0.977373

RECOMMENDATIONS

Additional analysis using FourSquare produced the table below with top 10 venues per each new location.

targeted zip codes	1st Most	2nd Most	3rd Most	4th Most	5th Most	6th Most	7th Most	8th Most	9th Most	10th Most
91205	Fast Food Restaurant	Bakery	Thai Restaurant	Convenience Store	Pizza Place	Mediterranean Restaurant	Sandwich Place	Park	Asian Restaurant	Donut Shop
91803	Asian Restaurant	Mexican Restaurant	Pizza Place	Fried Chicken Joint	Italian Restaurant	Szechuan Restaurant	Spa	Fast Food Restaurant	Vietnamese Restaurant	Hot Dog Joint
92395	Convenience Store	Café	Construction & Landscaping	Karaoke Bar	Pizza Place	Electronics Store	Donut Shop	Farm	Fast Food Restaurant	Fried Chicken Joint
93637	Mexican Restaurant	Gym / Fitness Center	Farm	Gift Shop	Wings Joint	Department Store	Donut Shop	Electronics Store	Fast Food Restaurant	Fried Chicken Joint
93955	Business Service	American Restaurant	Trail	Korean Restaurant	Basketball Court	Golf Course	Greek Restaurant	Grocery Store	Gun Range	Cuban Restaurant

✓ As per the table above, every one of the new locations seem to be very accepting of the diverse world cuisine.
✓ Strong feature correlation and FourSquare based findings provide strong reasons to believe that Patcy's pizzeria has a great chance of succeeding at any of the five new locations.