Identifying Fast Food Franchise Restaurants in Houston, Texas to Pursue for Purchase

Applied Data Science Capstone Project by Joe McReynolds

The goal of the project is to identify poorly rated fast-food franchise restaurants as candidates to purchase and turn around to profitability. A work flow utilizing Foursquare data will be designed and executed. The results will be reported with emphasis on visual displays.

TABLE OF CONTENTS

INTRODUCTION / BUSINESS PROBLEM

The business problem is identifying underperforming "fast food" franchise restaurants in Houston, Texas for restaurant investors to consider for purchase and turn around to profitability. The plan is to leverage Foursquare location data to design a work flow that helps commercial real estate investors locate fast-food franchise restaurants to purchase for investment. Two main areas will be the focus.

1. Underperforming or poorly rated restaurants will be identified because poorly rated restaurants are more likely to cost less to purchase than high performing properties. They may also be more available.
2. Along with identifying that a low performing group, restaurants and the neighborhoods they are in will be grouped for similar qualities to evaluate how that specific franchise typically performs in the type of neighborhood in which it is located in Houston. This will identify the best areas to look for investment in as well.

The Foursquare data service is well suited for this task because it not only locates the properties but also contains venue statistics and ratings for each restaurant.

The value of the project is to give investors another line of evidence to add to their decision process and should improve their chance of buying the underperforming business that can become profitable.

DATA

Two main data sources will be used in the project. Foursquare data will be combined with local Houston neighborhood data found in web research to complete the project.

1. Houston Neighborhood Spatial Data

Houston neighborhood spatial data will be used to segment Houston into areas that through cluster analysis will be grouped into neighborhoods with similar venues found in the Foursquare data.

Information on Houston neighborhoods is located at the following web page:
https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods

The salient information will be scraped from this web page, and latitude longitude will be generated using geocoder packages imported from the geopy library.

88 "super" neighborhoods exist in the Houston area.
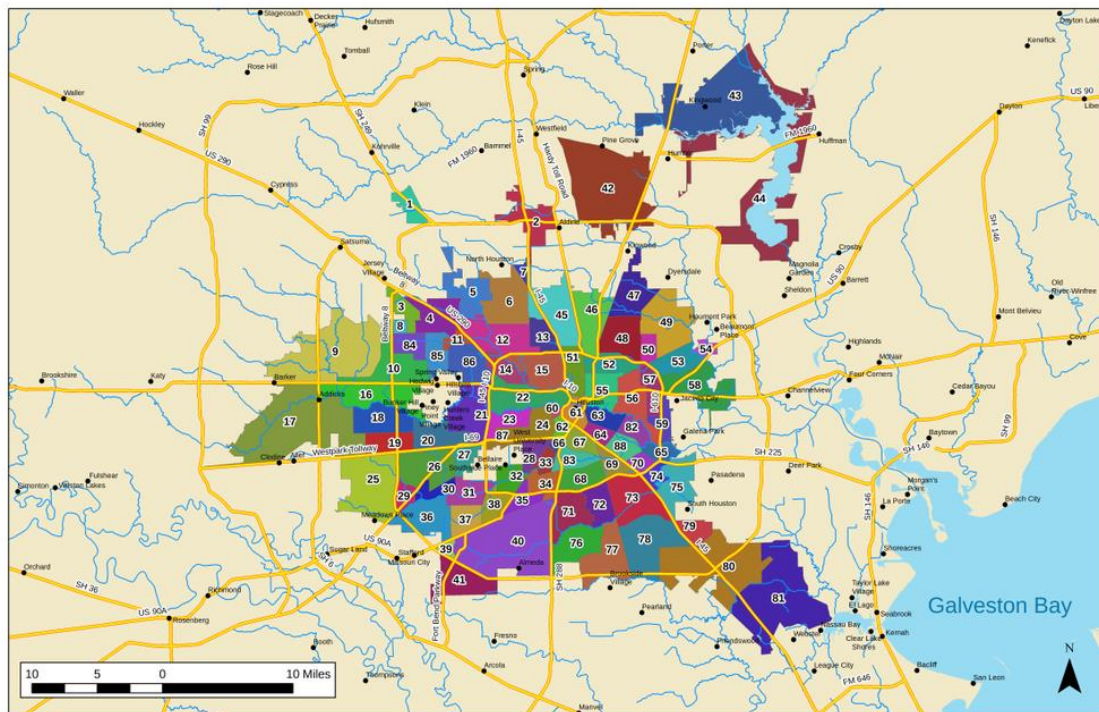
2. Foursquare Venue data

Foursquare data will used in multiple parts of the workflow. Specifically two data types will be utilized from Foursquare: general Venue category data and Venue specific rating data.

- **General Venue category data** will be gathered for each neighborhood in preparation for cluster analysis to group similar neighborhoods.

- **Venue specific rating data** will be used after cluster analysis to analyze the performance of individual fast food restaurants.

The Foursquare data service is well suited for this task because it not only locates the properties but also contains venue statistics and ratings for each restaurant.

METHODOLOGY

Developed a workflow to identify underperforming "fast food" franchise restaurants in Houston, Texas for restaurant investors to purchase. It is a four part process.

PART 1. **Neighborhood Polygon Preprocessing**

Houston Neighborhood areas were downloaded as shapefiles from City of Houston. First they were transformed to json format to create an interactive map of Houston's Neighborhood using folium maps. Second, centroid points for each neighborhood polygon were calculate to use in the Foursquare explore API because it functions best searching around a latitude longitude point.

PART 2. **General Venue Category Neighborhood Comparison**

The neighborhoods were first compared based on the all types of venue categories and those categories prevalence in each neighborhood. Foursquare data was downloaded, preprocessed and run through a cluster analysis to group neighborhoods with similar venue categories. A K-Cluster algorithm was used in the analysis. Comparison of similar neighborhoods within cluster groups should allow for more accurate comparisons. This is the style analysis that we have done in class for Toronto and New York.

PART 3. **Fast Food Restaurant Franchise Neighborhood Comparison**

As well as comparing general category composition, neighborhoods were compared based on only fast food franchise composition and prevalence. Foursquare data was further filtered down to a set of common fast food franchises and run through another cluster analysis to again group similar neighborhoods for comparison. A K-Cluster algorithm was used in the analysis. Comparison of similar neighborhoods within cluster groups using only fast food venues should also improve comparisons.

PART 4. **Fast Food Restaurant Ratings Analysis for select Neighborhoods**

Finally, for rating comparison, a small subset of similar neighborhoods was selected based on the cluster analyses from Part 2 and 3. Foursquare ratings data are part of their premium service and our free service limits the amount of requests that can be made on this data. Ideally, I would have done all neighborhoods.

With neighborhood's venues filtered to only fast food restaurants, the venue id for each individual restaurant was submitted to the Foursquare API for detailed information that includes ratings and how many "likes" customers had given a venue. Results were sorted by neighborhood and printed to evaluate restaurants in comparable neighborhoods that are lower rated compared to similar restaurants. Results were also sorted by fast food franchise and printed to evaluate for restaurants among the same franchise venues for anomalously low ratings.

Note: Getting fast food restaurants as a single category is not a trivial task, and I built our own fast food category by filtering for specific fast food franchise names. In our WhataBurger example, the raw download data had the venue categorized both as a "burger joint" and as "fast food restaurant". This a common occurrence in the Foursquare database and has proven challenging to always import the "fast food" category into the pandas databases for analysis. It is also worth noting that some peculiarities exist in querying venue data within Foursquare. Fast Food Query results frequently pull more than "fast food" venues.

The investor can compile an initial list of restaurants to begin studying in further to pursue for purchase.

ANALYSIS

PART 1. Neighborhood Polygon Preprocessing

The first step to analyze Houston Neighborhoods was to find and load Polygons that defined neighborhoods in Houston. The interactive map provides pop up labels to identify individual neighborhoods. The yellow dots represent the calculated centroid points that will be used to gather venue data from Foursquare in the Part 2 and 3. See map below and note the neighborhoods are randomly colored to show their areas and the Sharpstown label has been selected to show what the pop labels look like
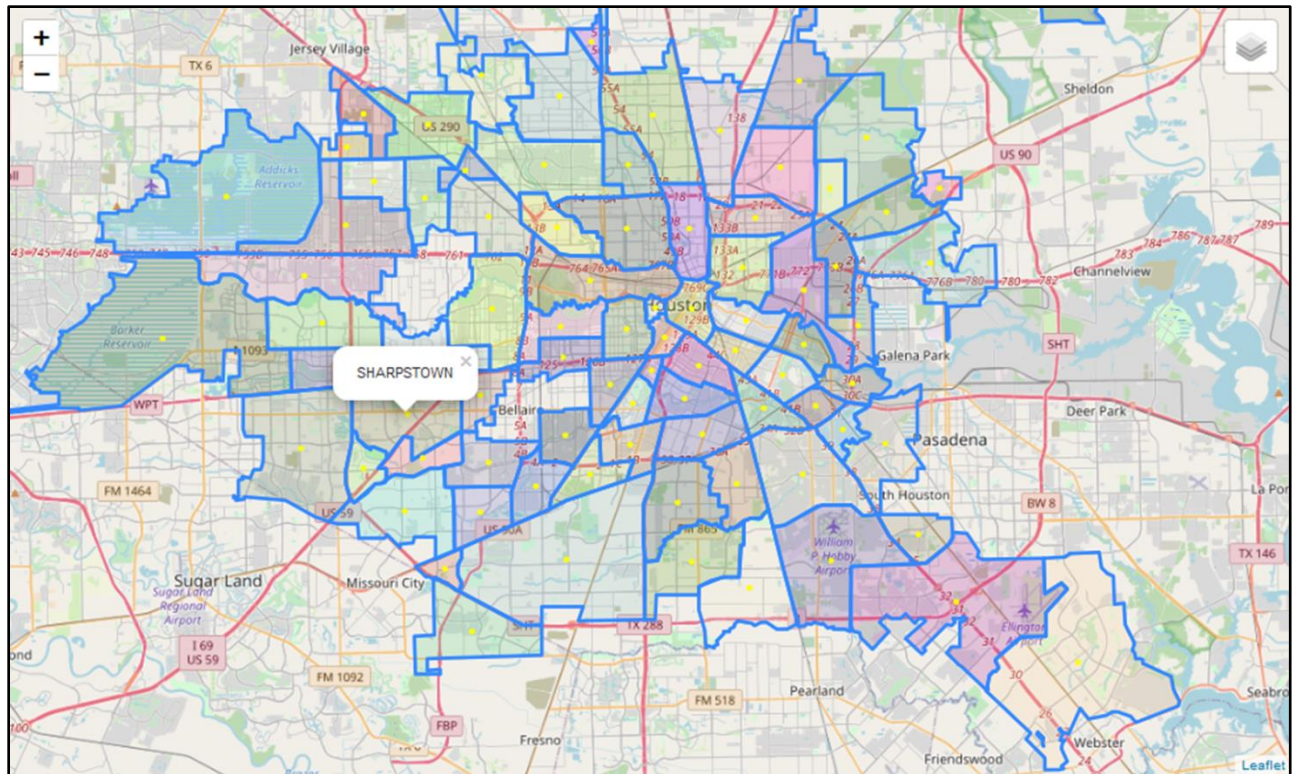
City of Houston Neighborhood Map

PART 2. General Venue Category Neighborhood Comparison

For K-Cluster analysis, 5009 individual venues were collected over the 88 neighborhoods of Houston. 332 unique categories were observed in the dataset. The cluster analysis identified 5 clusters or similarity groups.
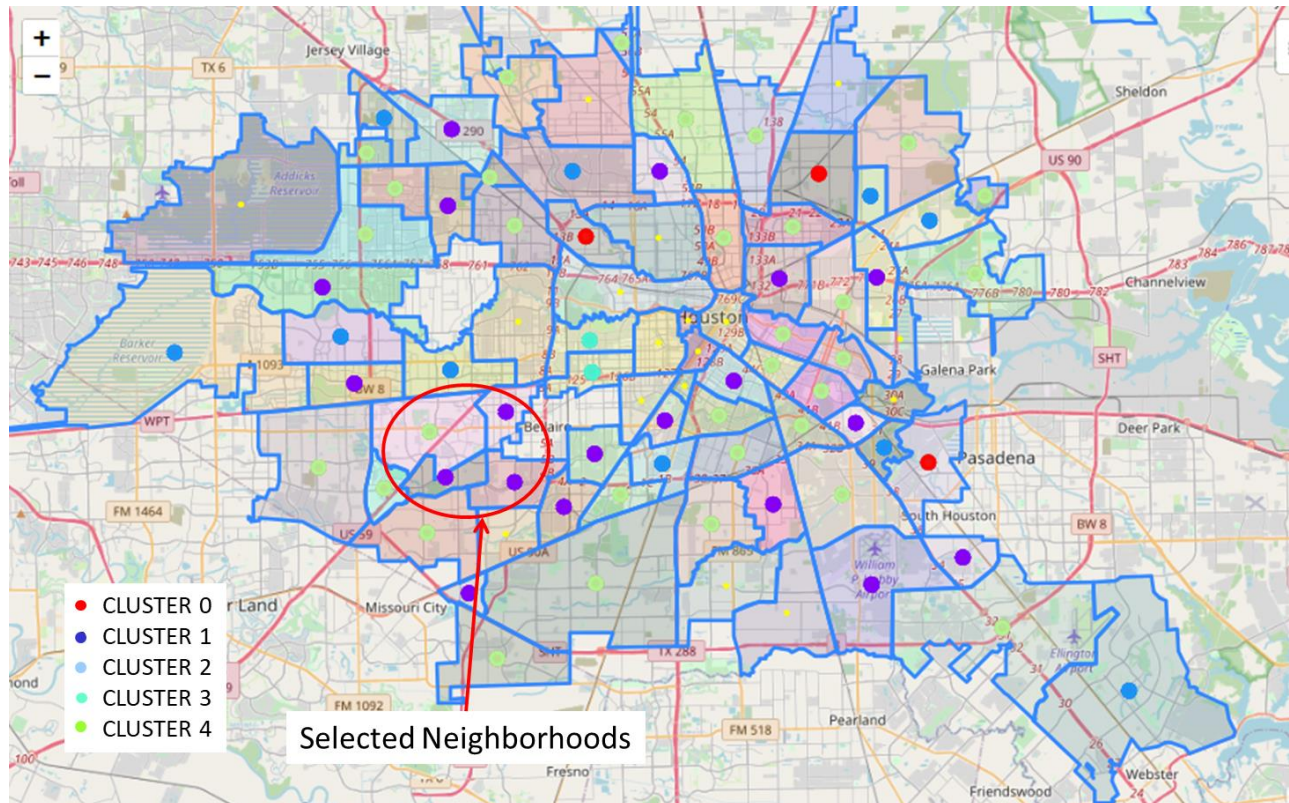
Results of Neighborhoods Clustering for All Venue Categories

Cluster 0 is dominated by Mexican restaurants and other typical fast food restaurants.

Cluster 1 is dominated by shopping and service oriented businesses with some restaurants

**Cluster 2** is dominated by discount shopping stores mixed with many styles of restaurants.

Cluster 3 is dominated by Mexican restaurants and a mix of other restaurants and stores.

Cluster 4 is composed of Parks and other recreations venues.

Cluster 5 is composed of a mix of restaurants, shopping, and personal service businesses,

PART 3. Fast Food Specific Neighborhood Comparison

For K-Cluster analysis, 350 individual venues were collected over the 88 neighborhoods of Houston. 71 unique fast food franchises were filtered to make the dataset. The cluster analysis identified 4 clusters or similarity groups, but the group differences a slight. Most Neighborhoods had a good mix of fast food restaurants with some shuffle for the most common venues.
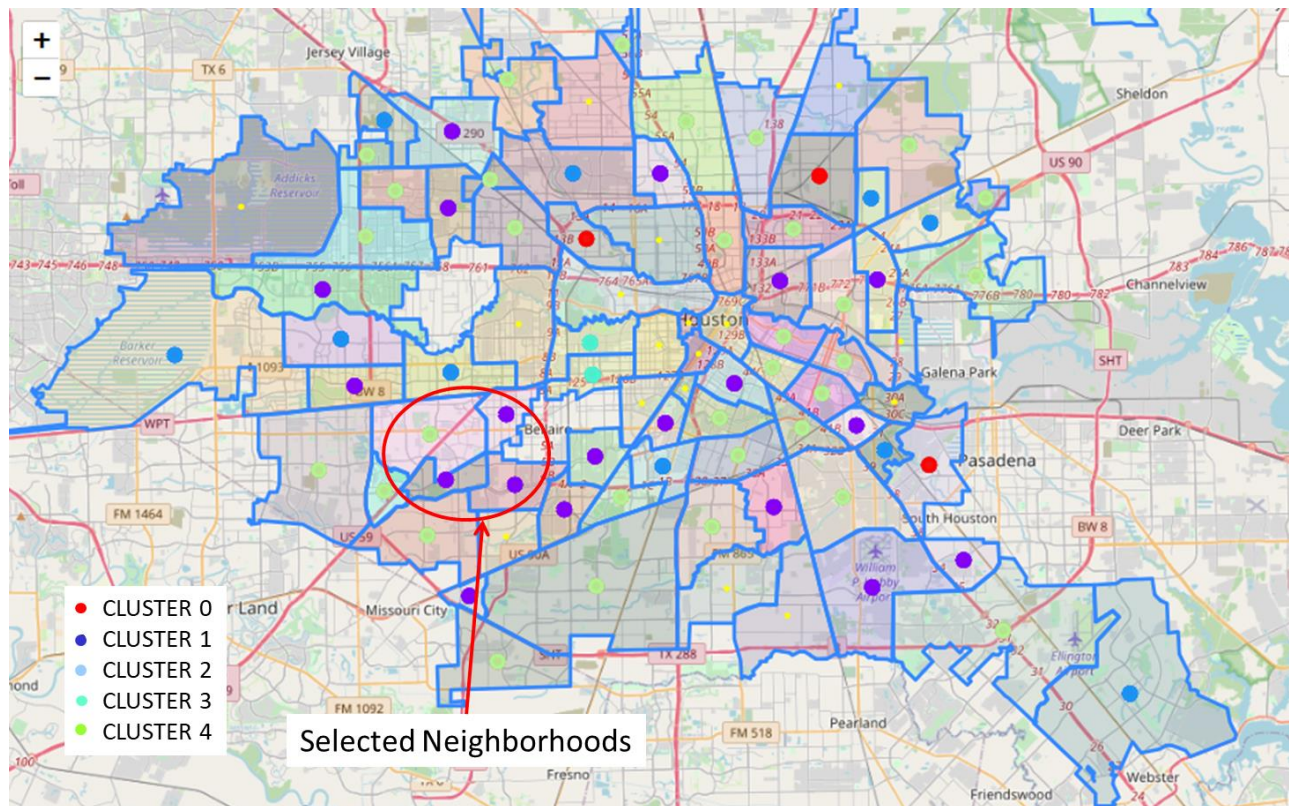
Results of Neighborhoods Clustering for All Venue Categories

Cluster 0 has WhataBurger, McDonalds, and Burger King as the most common venues.

Cluster 1 has Subway, What A Burger, and McDonalds as the most common venues.

Cluster 2 has Subway, WhataBurger, and Wendy's as the most common venues.

Cluster 3 has What A Burger, and Wendy's and Taco Bell as the most common venues.

Cluster 4 has Subway, WhataBurger, and Sonic as the most common venues.

PART 4. Fast Food Restaurant Ratings for select Neighborhoods

We limit the number of Neighborhoods to study in Part 4 because venue detail data with ratings is part of Foursquare premium and strictly limits how much data can get pulled at a time and per day..

The selected Neighborhoods are: **Braeburn, Meyerland, Gulfton, and Sharpstown**

These neighborhoods were selected because they are in the same or similar cluster or grouping for both the fast food venue analysis and the overall venue category analysis. They are also all on the west side of Houston. All this should prove to make a stronger comparison.

Report of Selected Fast Food Restaurants Grouped by Neighborhood and sorted by Rating

First, the selected fast food venues grouped by neighborhood and printed in report to help find unusually low rated restaurant relative to other restaurants in Neighborhood (See Below). The McDonalds in Sharpstown is the only restaurant that appears rated significantly lower than the rest. It is also apparent Whataburger is the favorite in the area with high ratings and lots of "likes"s.

```
SELECTED FAST FOOD RESAURANTS GROUPED BY NEIGHBORHOOD SORTED BY RATING
        Neighborhood              Venue   Rating Likes
    6       BRAEBURN        Whataburger      7.4     2
    7       BRAEBURN            SUBWAY       6.6     0
    8       BRAEBURN            SUBWAY       6.6     0
    9       BRAEBURN   Churchs Chicken       6.6     3
    10      BRAEBURN            SUBWAY       6.3     3


        Neighborhood              Venue   Rating Likes
    11       GULFTON        Whataburger      8.1    32
    12       GULFTON            SUBWAY       6.8     2
    13       GULFTON   Jack in the Box       7.1     8


         Neighborhood               Venue   Rating Likes
    14   MEYERLAND AREA        Whataburger     7.8    47
    15   MEYERLAND AREA   Jack in the Box      7.2     7
    16   MEYERLAND AREA            SUBWAY      6.3     5
    17   MEYERLAND AREA          McDonalds     6.7     8
    18   MEYERLAND AREA            SUBWAY      6.6     0


      Neighborhood               Venue   Rating Likes
    0    SHARPSTOWN        Whataburger     7.4     2
    1    SHARPSTOWN   Churchs Chicken      7.3     9
    2    SHARPSTOWN            SUBWAY      6.6     9
    3    SHARPSTOWN   Churchs Chicken      6.6     3
    4    SHARPSTOWN        Burger King     6.2     6
    5    SHARPSTOWN          McDonalds     5.6     8
```

Report of Selected Fast Food Restaurants Grouped by Fast Food Franchise and sorted by Rating

Second, Use to find unusually low rated restaurant relative to similar franchises across similar neighborhoods

Second, selected fast food venues are grouped by fast food franchise and printed in report to help find unusually low rated restaurants relative to same franchises in Neighborhood (See Below). Again, the McDonalds in Sharpstown stands out as rated significantly lower than the other McDonalds. There are two Churchs chicken in Sharpstown and one is lower than the other. This Churchs Chicken is another poorly rated. Whataburger still looks like the favorite in the area with high ratings and lots of "likes'.

```
SELECTED FAST FOOD RESAURANTS GROUPED BY FRANCHISE SORTED BY RATING
     Neighborhood          Venue  Rating Likes
4    SHARPSTOWN  Burger King      6.2      6


   Neighborhood               Venue  Rating Likes
1    SHARPSTOWN  Churchs Chicken       7.3      9
3    SHARPSTOWN  Churchs Chicken       6.6      3
9      BRAEBURN  Churchs Chicken       6.6      3


        Neighborhood             Venue  Rating Likes
15   MEYERLAND AREA  Jack in the Box       7.2      7
13           GULFTON  Jack in the Box      7.1      8


        Neighborhood     Venue  Rating Likes
17   MEYERLAND AREA  McDonalds      6.7      8
5        SHARPSTOWN  McDonalds      5.6      8


        Neighborhood  Venue  Rating Likes
12           GULFTON  SUBWAY     6.8      2
2         SHARPSTOWN  SUBWAY     6.6      9
7          BRAEBURN  SUBWAY      6.6      0
8          BRAEBURN  SUBWAY      6.6      0
18   MEYERLAND AREA  SUBWAY      6.6      0
10          BRAEBURN  SUBWAY      6.3      3
16   MEYERLAND AREA  SUBWAY      6.3      5


        Neighborhood        Venue  Rating Likes
11           GULFTON  Whataburger      8.1     32
14   MEYERLAND AREA  Whataburger      7.8     47
0        SHARPSTOWN  Whataburger      7.4      2
6          BRAEBURN  Whataburger      7.4      2
```

CONCLUSIONS

My key conclusions are as follows:

1) The McDonalds in Sharpstown was the lowest rated both for all fast food restaurants in Sharpstown and similar Neighborhoods analyzed. It is also the lowest rated of McDonalds. This venue should be on your list of underperformers.
2) The lower rated Churchs Chicken in Sharpstown is also on the list of underperformers.
3) Most restaurants had very similar ratings and no other venues stood out as low rated in our select area
4) Whataburger stands out as best overall rated restaurant. Consider a New Whataburger in a Neighborhood lacking Whataburger.
5) The "Likes" information was only useful in a qualitative sense and best identified the most popular individual restaurants such as the Whataburgers in Gulfton and Meyerland.

Again, I did not analyze venue ratings for the entire City or a larger area because of the issue with foursquare limiting the ratings data as a premium service. With a paying account more might be done.

Identifying lower rated restaurants would allow the investor to build a list of restaurants to begin looking into further detail to pursue for purchase. With Whataburgers being noticeably higher rated, it might be better to start a new Whataburger