

COURSERA DATA SCIENCE CAPSTONE  
BATTLE OF THE NEIGHBORHOODS

*WHERE IS THE BEST PLACE TO PARK A FOOD  
TRUCK IN FRANKLIN COUNTY, OHIO?*

JORDAN PIERRE

# CONTENTS

<b>BACKGROUND</b>	<b>3</b>
<b>PROBLEM DESCRIPTION</b>	<b>3</b>
<b>DATA</b>	<b>3</b>
<b>METHODOLOGY</b>	<b>4</b>
<b>RESULTS</b>	<b>9</b>
<b>DISCUSSION</b>	<b>9</b>
<b>CONCLUSION</b>	<b>9</b>

# COURSERA APPLIED DATA SCIENCE CAPSTONE

Jordan Pierre

## I. BACKGROUND

One of the major reasons I am drawn to data science and analytics is to cultivate the ability to use data to obtain a competitive advantage in business. I am interested in the entrepreneurship space and am intrigued by the unique advantages of a food truck business model, especially their portability. Besides having great food at a fair price, the success of a food truck depends heavily on where it is parked because people tend not to drive to food trucks as they would a regular restaurant. As a result, the ideal place to park a food truck would be a place where hungry people already are, where there is heavy foot traffic and, ideally, relatively few restaurants (competition) nearby. Some ideal locations for a food truck might be near a bar, a sports venue, a music venue, an office complex, or college campuses.

I am also a student at the Ohio State University and live in Columbus, OH, so I thought it would be fun and interesting to use data analytics to determine the best place for my theoretical food truck in my local area.

## II. PROBLEM DESCRIPTION

Based on my interests and experience, I wanted to use data science to answer the following question:

*“What area is the best place to park a food truck in Franklin County, Ohio?”*

To answer this question, we’ll assume that we have a permit to park anywhere in Franklin County.

## III. DATA

The data used to answer this question will consist of:

1. City, County, Zip Code, and Coordinate information of the entire US a from CSV from OpenDataSoft.com.
2. Categorical and Locational information for venues in Franklin County, Ohio, using the FourSquare API.

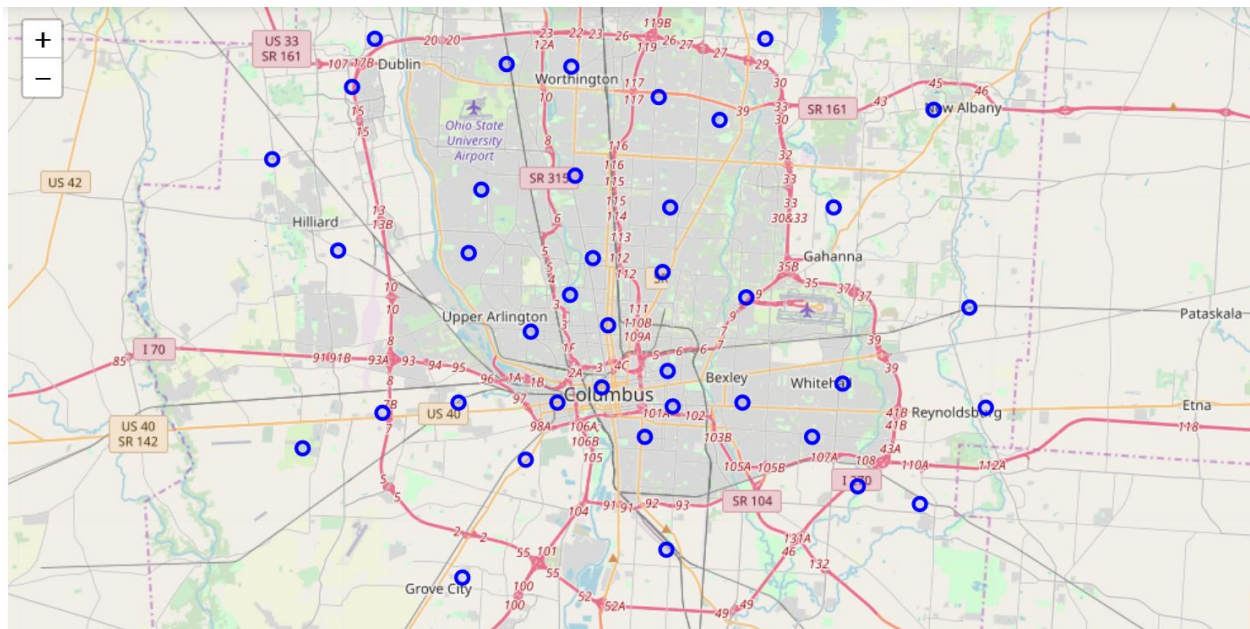
3. Population Size, Family Income, and other demographic information of Ohioans from the IRS website.

#### IV. METHODOLOGY

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

In order to investigate the best area to park a food truck in central Ohio, I had to prepare my data sets such that all irrelevant data (those relating to other areas) were removed. I made a list of all zip codes in Ohio, then selected only those in Franklin County, along with the name of the associated city. Then I loaded coordinate data (latitude and longitude) for the entire US, and then joined the coordinate data on zip code, leaving me with a dataframe whose columns were zip code, city, latitude, and longitude, and each row corresponded to some zip code in Franklin County.

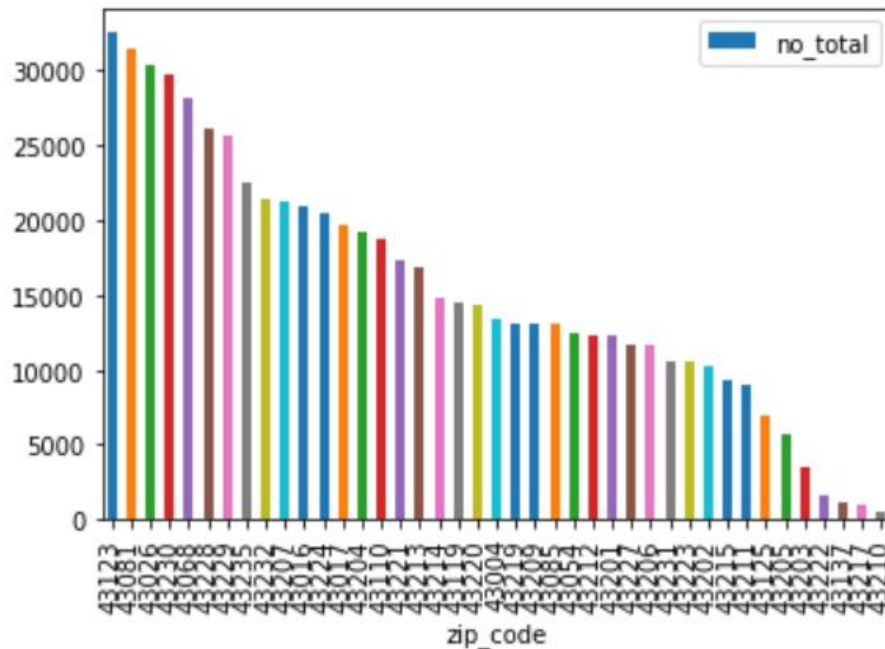
I then visualized the coordinates of the zip codes in Franklin County on a map using Folium.



I then collected data about the types of venues in each zip code using the FourSquare API. I created a dataframe of each zip code and their 10 most popular venue categories. The category and popularity of venues in an area is important to the food truck problem because we want to park our truck in an area where people already go to and stay long enough to be hungry and ideally, we want as little competition as possible.

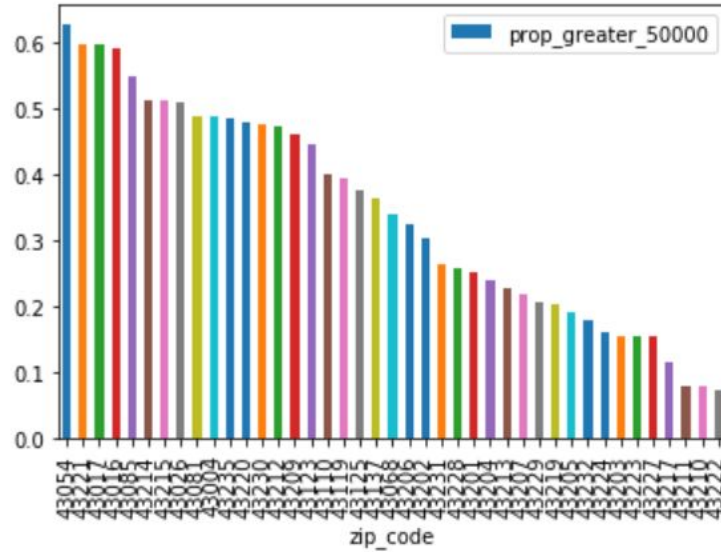
	zip_code	#1 Most Common Venue Category	#2 Most Common Venue Category	#3 Most Common Venue Category	#4 Most Common Venue Category	#5 Most Common Venue Category	#6 Most Common Venue Category	#7 Most Common Venue Category	#8 Most Common Venue Category	#9 Most Common Venue Category	#10 Most Common Venue Category
0	43002	Gas Station	Supermarket	Pharmacy	Cosmetics Shop	Video Store	Deli / Bodega	Fabric Shop	Donut Shop	Dive Bar	Discount Store
1	43004	Light Rail Station	Garden Center	Women's Store	Flower Shop	Fast Food Restaurant	Farmers Market	Fabric Shop	Donut Shop	Dive Bar	Discount Store
2	43016	Gym	Swim School	Sandwich Place	Gastropub	Women's Store	Deli / Bodega	Fabric Shop	Donut Shop	Dive Bar	Discount Store
3	43017	Baseball Field	Gym	Football Stadium	Office	Cafeteria	Basketball Court	Women's Store	Farmers Market	Fabric Shop	Donut Shop
4	43026	Pizza Place	Women's Store	Department Store	Fast Food Restaurant	Farmers Market	Fabric Shop	Donut Shop	Dive Bar	Discount Store	Diner
5	43054	Park	History Museum	Shop & Service	Comic Shop	Department Store	Farmers Market	Fabric Shop	Donut Shop	Dive Bar	Discount Store

After getting the information about the types of venues in each area, I collected data that could reveal what kind of people live in those areas. To do that, I took reported IRS data for all residents of Ohio and joined it on the relevant zip codes. From there, I categorized the number of people in each zip code that reported an annual earnings of less than \$25,000, \$25,000-\$50,000, \$50,000-\$75,000, \$75,000-\$100,000, \$100,000-\$200,000, or greater than \$200,000, as well as the total number of reported earnings in that zip code. Then I added a column ranking each zip code by its total reportings size relative to the other Franklin County zip codes and graphed a bar chart.



After plotting the total sizes, I realize that the total number of people who reported earnings to the IRS was not a good indicator of who my potential customers are, because people with low incomes are less likely to spend money at my food truck. I believe people who make more than \$50,000 per year are more likely to eat out than people who make less than \$50,000 per year, so I made a new column counting the number of people in the area that are “well-off,” meaning they make more than \$50,000 annually. I then ranked the zip codes by the number of well-off residents as well as ranking them by the proportion of people who are well-off relative to the total number of reported earnings. This is relevant because we want to park in a location where there are a lot of people nearby and a lot of people who can afford to eat at my food truck.

	zip_code	well_off_rank	prop_greater_50000	no_greater_50000	no_total
4	43054	1.0	0.627926	7780	12390
30	43221	2.0	0.598616	10380	17340
2	43017	3.0	0.597468	11800	19750
1	43016	4.0	0.592381	12440	21000
7	43085	5.0	0.548734	7150	13030
25	43214	6.0	0.514189	7610	14800



After exploring the financial demographic information, I combined the financial information about the people in each zip code with the type of venues located in each zip code.

The venue category information had to be categorized and converted into numbers to be used in a mathematical model, so I manually allocated the venue categories shown in the top 5 wealthiest zip codes into the categories: bad, okay, good, and great. I categorized bad venues as venues that do not attract many people or they are in direct competition with my food truck. Okay venues are those that attract people, but those people might not stay long enough to want to eat. Good venues are those that attract people and those people stay long enough that they will probably want food. Great venues are those that attract a lot of people that will definitely be hungry because they are either drinking or stay so long that they will want to eat. I categorized all the nominal values and assigned a value of 0 for bad venues, 1 for okay venues, 2 for good venues, and 3 for great venues. I also normalized the total number of people who earn more than \$50,000 in the area by dividing the zip code's number of people who earn more than \$50,000 by the value of the zip code with the maximum number of people who earn more than \$50,000.

I then calculated an overall score for each zip code based on the equation:

$$food\ truck\ score_i = \frac{wealthy_i}{max\{wealthy\}} * \sum_{j=1}^{10} venue\ score_{ij}$$

$food\ truck\ score_i = food\ truck\ score\ for\ zip\ code,\ i$

$wealthy_i = number\ of\ people\ who\ make\ more\ than\ \$50,000\ in\ zip\ code,\ i$

$max\{wealthy\} = maximum\ number\ of\ people\ who\ make\ more\ than\ \$50,000\ in\ all\ zip\ codes$

$venue\ score_{i,j} = score\ of\ the\ jth\ most\ popular\ venue\ in\ zip\ code,\ i$



## V. RESULTS

	zip_code	Score
2	43017	16.845659
4	43054	12.876206
28	43221	11.668810
1	43016	11.000000
7	43085	10.574759

The top five best zip codes for our food truck, based on the equation above was Dublin (43017) with a score of 16.85, New Albany (43054) at 12.87, Upper Arlington (43221) at 11.67, Dublin (43016) at 11.00, and Worthington (43085) at 10.57.

Based on the income levels and types of venues, Dublin, OH (43017) would be the best place to park our food truck in Franklin county. It has a large and wealthy population, and the most popular venues are office, baseball field, and football stadium--all of which attract a lot of hungry people who stay for a long time.

## VI. DISCUSSION

Dublin (43017) was shown to be the best place to park our food truck as it has many office buildings and sports venues, which attract many people who stay for a long time and are or become hungry. It's score was actually 30.85% higher than the second place zip code, New Albany (43054). The range between the score of the second highest scoring zip code and 5th highest scoring zip code was 2.30 points, about half of the difference between the first highest scoring zip code and second highest zip code, which was 3.97. This shows that Dublin (43017) was the clear winner.

## VII. CONCLUSION

According to the scoring system used, which accounts for the normalized number of individuals in each zip code who make more than \$50,000 per year and the top 10 most popular venues in each zip code according to the FourSquare website, Dublin (43017) was found to be the best choice to park a food truck in Franklin County, Ohio. This is because it's wealthy population, number of office complexes, and prevalence of sports venues, all of which are great conditions for foot traffic for our food truck.