

Indianapolis Foodtrucks

James Wang

2023-10-12

Introduction and Dataset

Let's begin by loading in our dataset and the libraries we'll need to analyze it.

```
library(tidyverse)
library(dplyr)
indi_foodtrucks = read.csv("Indianapolis_Food_Trucks_Final_Data.csv")
```

This dataset was obtained by scraping the results of a Google Map search corresponding to the following url using PhantomBuster, a cloud based web scraping tool.

After manually cleaning out food trucks that didn't actually exist or went out of business, the dataset was left with 85 food trucks. Food trucks that were missing a value for the `rating` field were given the average rating of the rest of the food trucks in the data set and had their `reviewCount` field set to 0.

The original format of the hours for each day of the week was a string. This report only needs the hours for Saturday and Sunday so the `saturday` and `sunday` fields were each parsed and separated into two fields representing opening and closing times in 24 hour time format. Note here that some food trucks are open past midnight of the same day so cases where the closing time is before the opening time represent closing times that take place the next day. All of this pre-cleaning work was done in Excel.

In this report, the variables of interest are:

- `rating`
- `reviewCount`
- `category`
- `address`
- `saturday_open` and `saturday_close`
- `sunday_open` and `sunday_close`

Overall, our goal in this report is to identify the best food trucks for each type of cuisine, analyze the locations of these “best food trucks,” and select a few that will allow a tourist to experience a diverse sample of Indianapolis’s food truck cuisine.

To answer that larger question, we will first find answers to these sub-questions using the dataset:

1. What are the must visit food trucks according to user reviews?
2. What cuisines are popular with Indianapolis food trucks?
3. What are the most common hours of operation for Indianapolis food trucks?

Thankfully we have a wonderful data set to help us provide informed answers to those questions. Here are the first few rows of the dataset as a preview:

```
head(data.frame(indi_foodtrucks$title, indi_foodtrucks$rating, indi_foodtrucks$category))
```

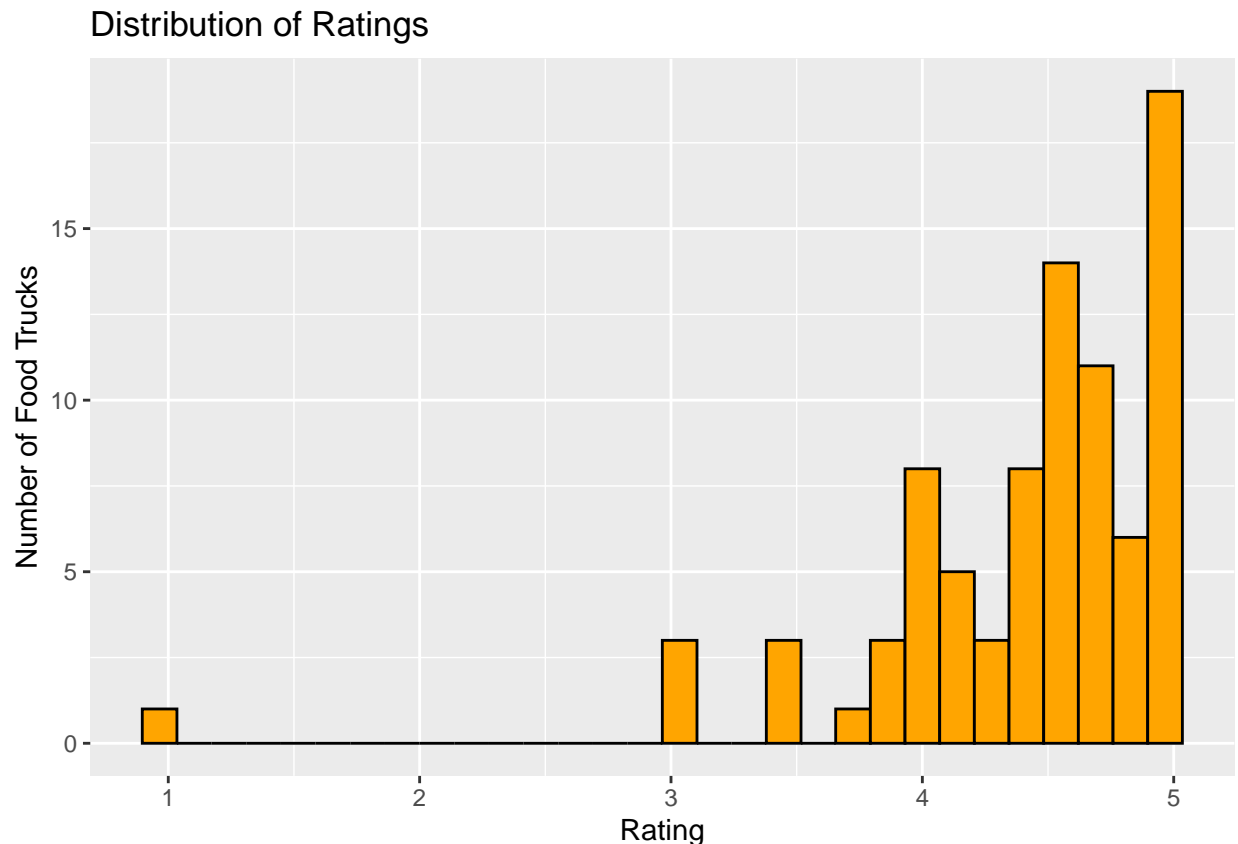
```
##          indi_foodtrucks.title indi_foodtrucks.rating
## 1   Paco's Taqueria - Food Truck                3.4
## 2      J's Lobster & Fish Market                3.5
## 3      El Viejon Food Truck                    1.0
## 4      Food Truck Don Luis                    3.0
## 5 Antojitos Guzman food truck #1                3.5
## 6      Huge Impact Restaurant                  3.9
##   indi_foodtrucks.category
## 1      Taco restaurant
## 2      Seafood restaurant
## 3      Mexican restaurant
## 4      Mexican restaurant
## 5      Restaurant
## 6      Restaurant
```

What are the Must Visit Food Trucks According to User Reviews?

Food trucks that serve delicious food tend to get high ratings from users. So why don't we use rating as a way to identify the most popular food trucks in Indianapolis?

Let's begin with some basic EDA on the `rating` variable:

```
indi_foodtrucks %>%  
  ggplot(aes(x = rating)) +  
  geom_histogram(fill = "orange", color = "black") +  
  labs(title = "Distribution of Ratings",  
        x = "Rating", y = "Number of Food Trucks")
```

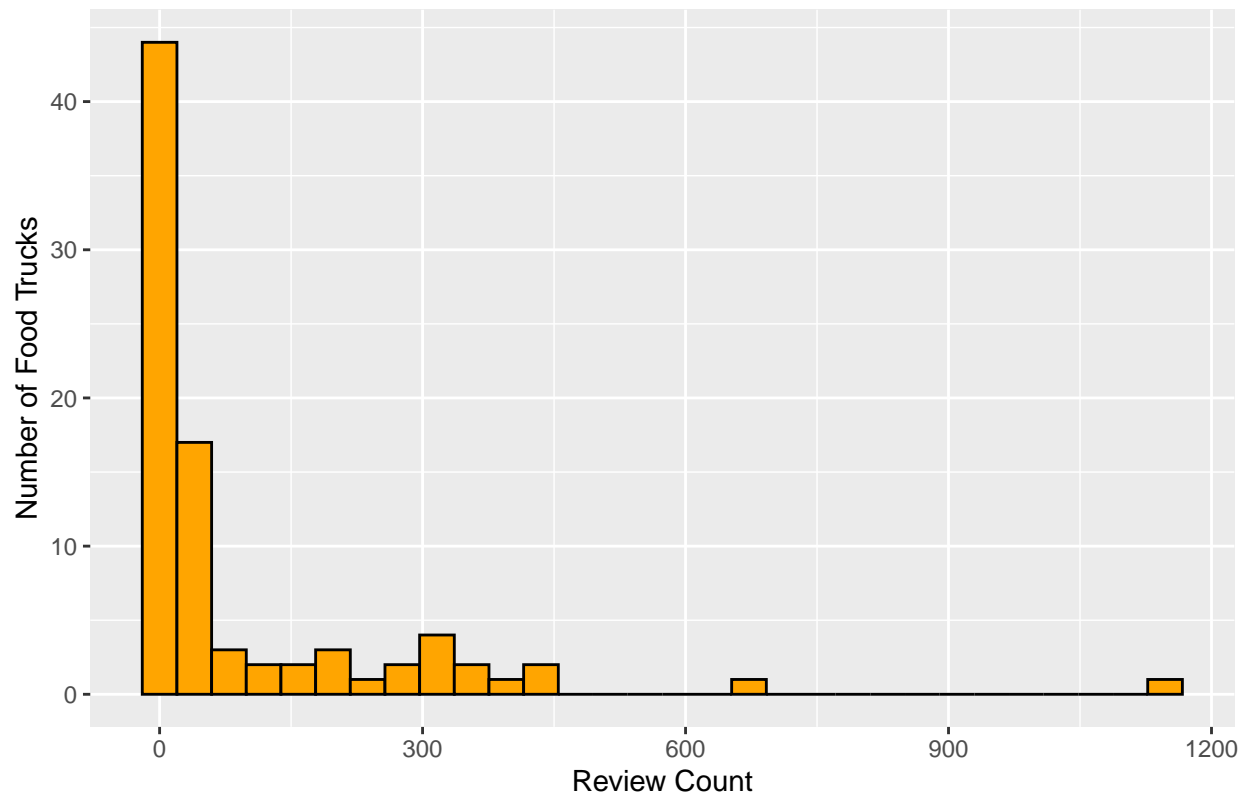


From this simple histogram, it seems that `rating` has three peaks at around 4, 4.5, and 5. In general, the distribution of `rating` has a heavy left skew which gives us reason to think that Indianapolis food trucks are highly rated and probably all sell delicious food. Good for us! We also notice a single outlier food truck with a rating of 1. Probably avoid that one.

However, not every customer leaves a review so we should also look at the distribution of `reviewCount`

```
indi_foodtrucks %>%  
  ggplot(aes(x = reviewCount)) +  
  geom_histogram(fill = "orange", color = "black") +  
  labs(title = "Distribution of Review Count",  
        x = "Review Count", y = "Number of Food Trucks")
```

Distribution of Review Count



Unfortunately, it seems that the majority of the food trucks in our data set have their ratings based off of a very small number of reviews, shown by the fact that the distribution of `reviewCount` has just a single peak at around 0 reviews. From this result, we might want to filter for just food trucks that have a `reviewCount` above a certain reliable threshold. While food trucks with just a few reviews might be hidden treasures, we don't want to take that risk for our newbie friendly Foodie Weekend. Additionally, we notice two outliers with an incredibly high number of reviews. If their ratings are good, maybe we've found our candidate food trucks!

Based on the results of our EDA, let's create a filtered data frame that only contains food trucks with a `rating` that's higher than 4 and at least a 10 for `reviewCount`. We'll also sort this new data frame by `rating` and `reviewCount` so that the first few rows are the highest rated food trucks with the highest number of reviews.

```
highly_rated_indi_foodtrucks = indi_foodtrucks %>%  
  filter(rating > 4 & reviewCount > 10)  
sorted_highly_rated_indi_foodtrucks = highly_rated_indi_foodtrucks %>%  
  arrange(desc(rating), desc(reviewCount))
```

Now let's take a look at our candidate food trucks:

```
head(data.frame(sorted_highly_rated_indi_foodtrucks$title, sorted_highly_rated_indi_foodtrucks$rating,  
  sorted_highly_rated_indi_foodtrucks$reviewCount), n = 10)
```

```
## sorted_highly_rated_indi_foodtrucks.title  
## 1 ClusterTruck - Broad Ripple  
## 2 ClusterTruck - Carmel
```

```
## 3          Garcia's Hot Dogs
## 4      ClusterTruck - Castleton
## 5          ClusterTruck
## 6      Chris' Ice Cream and Mexican Food
## 7      Tacos Tierra caliente Restaurant
## 8          ChefTLC
## 9          Gulas food truck
## 10         Emergency Taco Truck
##      sorted_highly_rated_indi_foodtrucks.rating
## 1          4.9
## 2          4.9
## 3          4.9
## 4          4.9
## 5          4.8
## 6          4.8
## 7          4.8
## 8          4.8
## 9          4.8
## 10         4.7
##      sorted_highly_rated_indi_foodtrucks.reviewCount
## 1          350
## 2          307
## 3          278
## 4          144
## 5          1147
## 6          680
## 7          118
## 8          44
## 9          13
## 10         104
```

Uh oh, why are so many of the top food trucks the same? This is where manual inspection is critical. A quick Google search reveals that ClusterTruck (Confused Google Maps because Truck is in its name) and Chris' Ice Cream and Mexican Food are not food trucks! Google Maps tricked us. Let's drop those from our dataset and sorted dataset and get our candidate food trucks again.

```
not_foodtrucks = c("ClusterTruck - Broad Ripple", "ClusterTruck - Carmel", "ClusterTruck - Castleton",
                  "ClusterTruck", "Chris' Ice Cream and Mexican Food")
indi_foodtrucks = subset(indi_foodtrucks, !(title %in% not_foodtrucks))
sorted_highly_rated_indi_foodtrucks = subset(sorted_highly_rated_indi_foodtrucks, !(title %in% not_foodtrucks))
head(data.frame(sorted_highly_rated_indi_foodtrucks$title, sorted_highly_rated_indi_foodtrucks$rating,
                sorted_highly_rated_indi_foodtrucks$reviewCount), n = 10)
```

```
##      sorted_highly_rated_indi_foodtrucks.title
## 1          Garcia's Hot Dogs
## 2      Tacos Tierra caliente Restaurant
## 3          ChefTLC
## 4          Gulas food truck
## 5      Emergency Taco Truck
## 6      Birrieria Iturbidense
## 7      King gyros food truck
## 8      El Pastor Taco Truck
## 9      Gaucho's Fire Express
```

```

## 10 Talkin Turkey Real Cajun Food Truck & Catering
##      sorted_highly_rated_indi_foodtrucks.rating
## 1              4.9
## 2              4.8
## 3              4.8
## 4              4.8
## 5              4.7
## 6              4.7
## 7              4.7
## 8              4.7
## 9              4.7
## 10             4.7
##      sorted_highly_rated_indi_foodtrucks.reviewCount
## 1              278
## 2              118
## 3              44
## 4              13
## 5              104
## 6              86
## 7              46
## 8              46
## 9              43
## 10             42

```

Much better! A quick verification Google Search confirms that our candidates are all food trucks.

What Cuisines are Popular with Indianapolis Food Trucks?

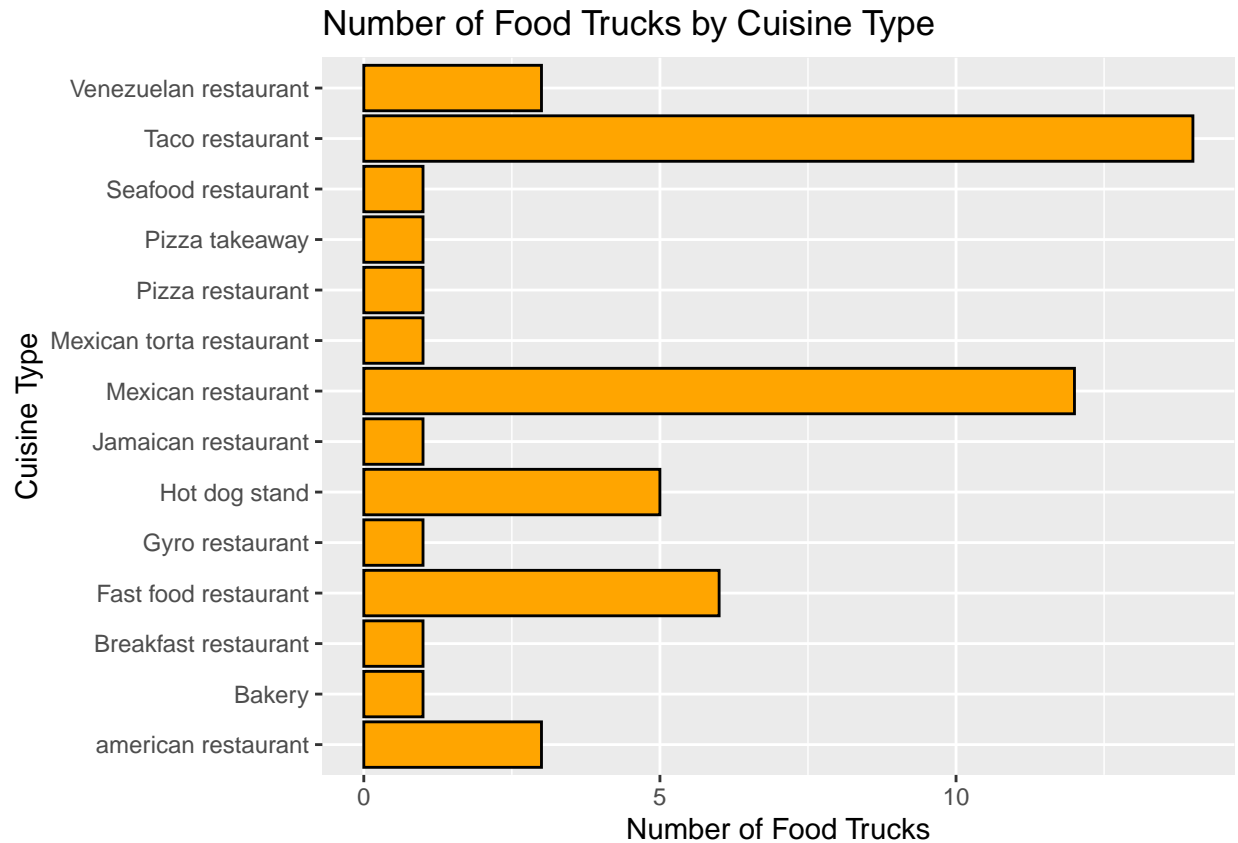
Tourism is all about new experiences so on our foodie trip, we want to experience a diverse selection of Indianapolis cuisine. First, let's get all of the unique cuisine types offered by Indianapolis food trucks using the `category` variable.

```
unique(indi_foodtrucks$category)
```

```
## [1] "Taco restaurant"      "Seafood restaurant"
## [3] "Mexican restaurant"   "Restaurant"
## [5] "Fast food restaurant" "Gyro restaurant"
## [7] "Hot dog stand"        "Venezuelan restaurant"
## [9] "Food producer"        "american restaurant"
## [11] "Mexican torta restaurant" "Takeout Restaurant"
## [13] "Jamaican restaurant"  "Breakfast restaurant"
## [15] "Mobile caterer"       "Pizza restaurant"
## [17] "Caterer"              "Bakery"
## [19] "Delivery Restaurant"  "Pizza takeaway"
```

That's quite a lot of them and many of them feel quite similar to each other. First, we should get rid of ridiculous categories like restaurant that tell us nothing about cuisine. Then let's visualize their frequencies with a bar chart.

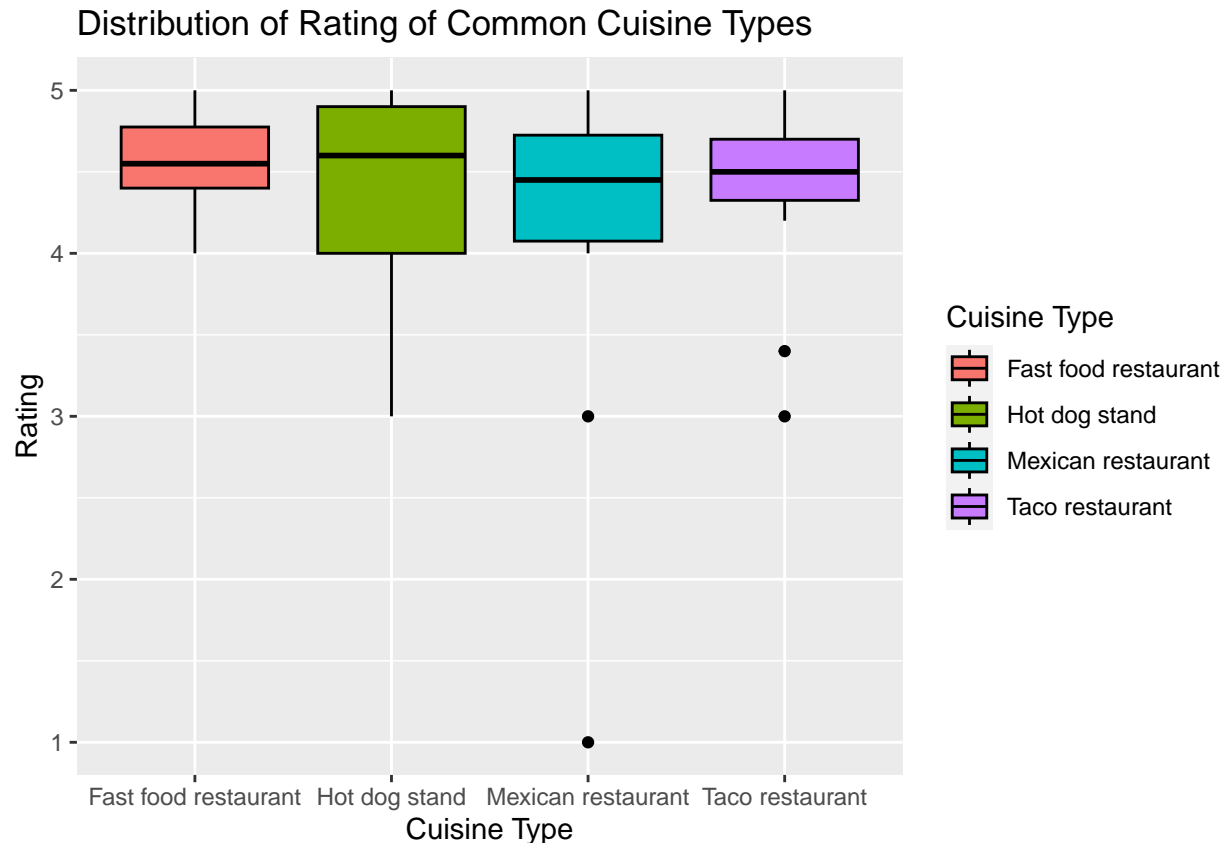
```
banned_categories = c("Restaurant", "Delivery Restaurant", "Caterer", "Food producer", "Takeout Restaurant")
indi_foodtrucks_cuisine = subset(indi_foodtrucks, !(category %in% banned_categories))
indi_foodtrucks_cuisine %>%
  ggplot(aes(x = category)) +
  geom_bar(fill = "orange", color = "black") +
  labs(title = "Number of Food Trucks by Cuisine Type",
       x = "Cuisine Type", y = "Number of Food Trucks") +
  coord_flip()
```



It appears that Taco, Mexican, Hot Dog, and Fast Food are the most common types of food trucks in Indianapolis. Given this information, we figure out which one of those categories has the highest rating. First some EDA between `category` and `rating`.

```
common_cuisines = c("Taco restaurant", "Mexican restaurant", "Hot dog stand", "Fast food restaurant")
common_cuisines_indi_food_trucks = subset(indi_foodtrucks, category %in% common_cuisines)

common_cuisines_indi_food_trucks %>%
  ggplot(aes(x = category, y = rating)) +
  geom_boxplot(aes(fill = category), color = "black") +
  scale_fill_discrete(name = "Cuisine Type") +
  labs(title = "Distribution of Rating of Common Cuisine Types",
       x = "Cuisine Type", y = "Rating")
```

Looking at this plot, it appears that the median rating of Hot dog stands is slightly higher than that of the rest of the cuisine types. Going off of our earlier EDA, it makes sense that the ratings are all quite high for each of the four common cuisine types. We do note the 4 low rating outliers for Mexican and Taco restaurants.

But this is all graphical and the difference in median could simply be a matter of random variation. An ANOVA test will tell us for sure!

```
oneway.test(rating ~ category, data = common_cuisines_indi_food_trucks)
```

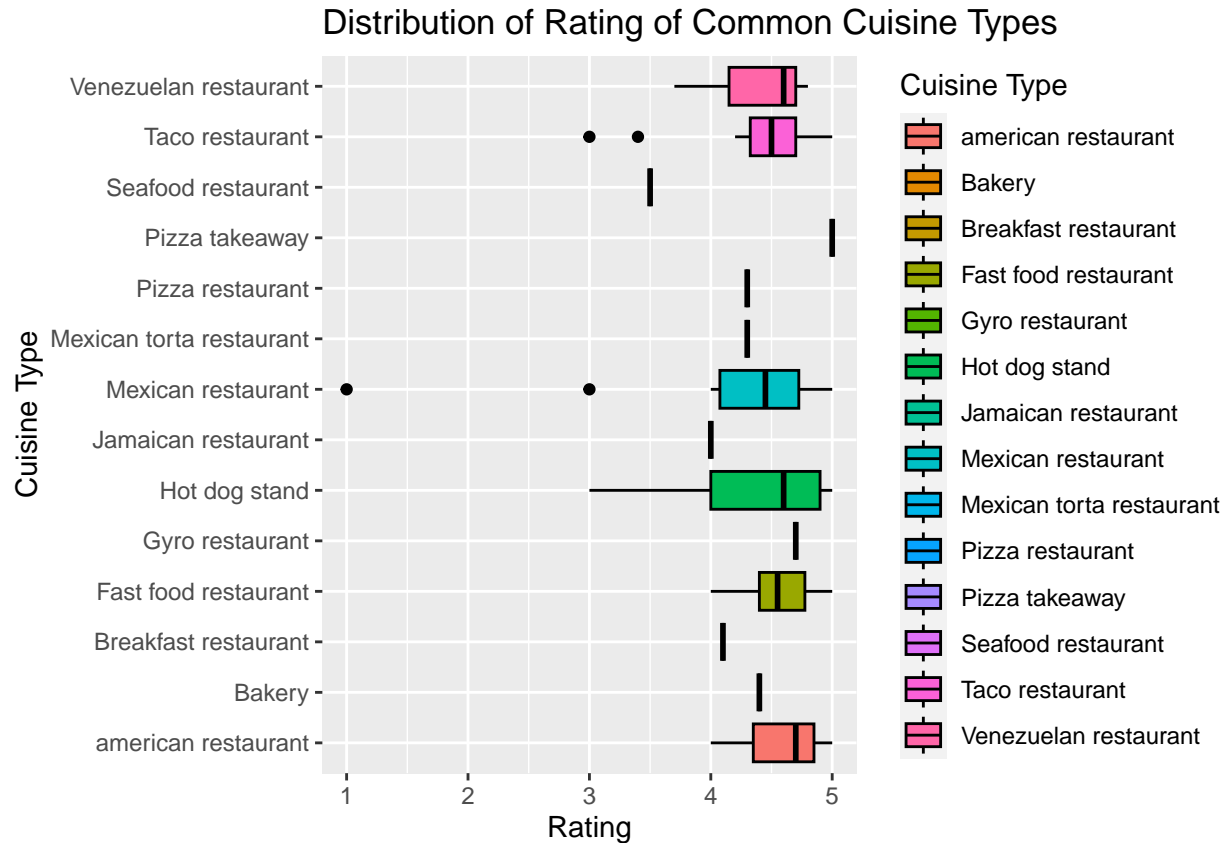
```
##
## One-way analysis of means (not assuming equal variances)
##
## data: rating and category
## F = 0.59511, num df = 3.000, denom df = 13.027, p-value = 0.6292
```

Since the p-value of the one-way ANOVA test is greater than the standard alpha threshold of 0.05, we do not reject the null hypothesis that the ratings are the same for all four categories of cuisine types. Therefore, it doesn't seem that Indianapolis has a specialty cuisine between its four most common.

What about for all cuisines? Is there a highly rated cuisine there? Unfortunately we can't formally answer this question because of the limitations of our small dataset (The ANOVA assumptions are violated!). A graphic will have to suffice.

```
indi_foodtrucks_cuisine %>%
  ggplot(aes(x = category, y = rating)) +
  geom_boxplot(aes(fill = category), color = "black") +
```

```
scale_fill_discrete(name = "Cuisine Type") +
labs(title = "Distribution of Rating of Common Cuisine Types",
      x = "Cuisine Type", y = "Rating") +
coord_flip()
```



Inferring based on the previous ANOVA test and this graphic, it does not seem like there is empirically any difference in the rating for most cuisines. Therefore, we can try all kinds of cuisines without worrying that Indianapolis isn't known for them.

What are the Most Common Hours of Operation for Indianapolis Food Trucks?

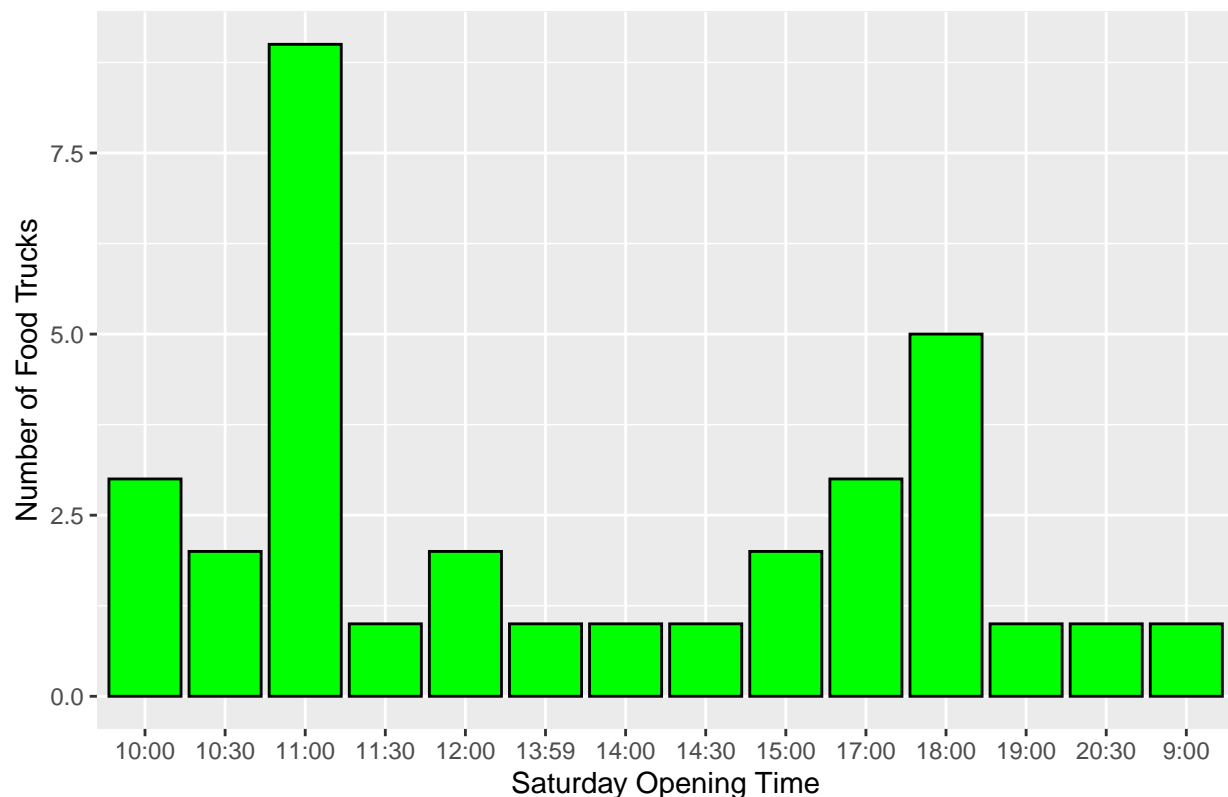
Recall that the end goal of this report is to help us plan a weekend foodie trip to Indianapolis. The most important piece of information we need to build an itinerary is when food trucks actually open and close in Indianapolis.

We begin by looking at the distribution of `saturday_open`, `saturday_close`, `sunday_open`, and `sunday_close`.

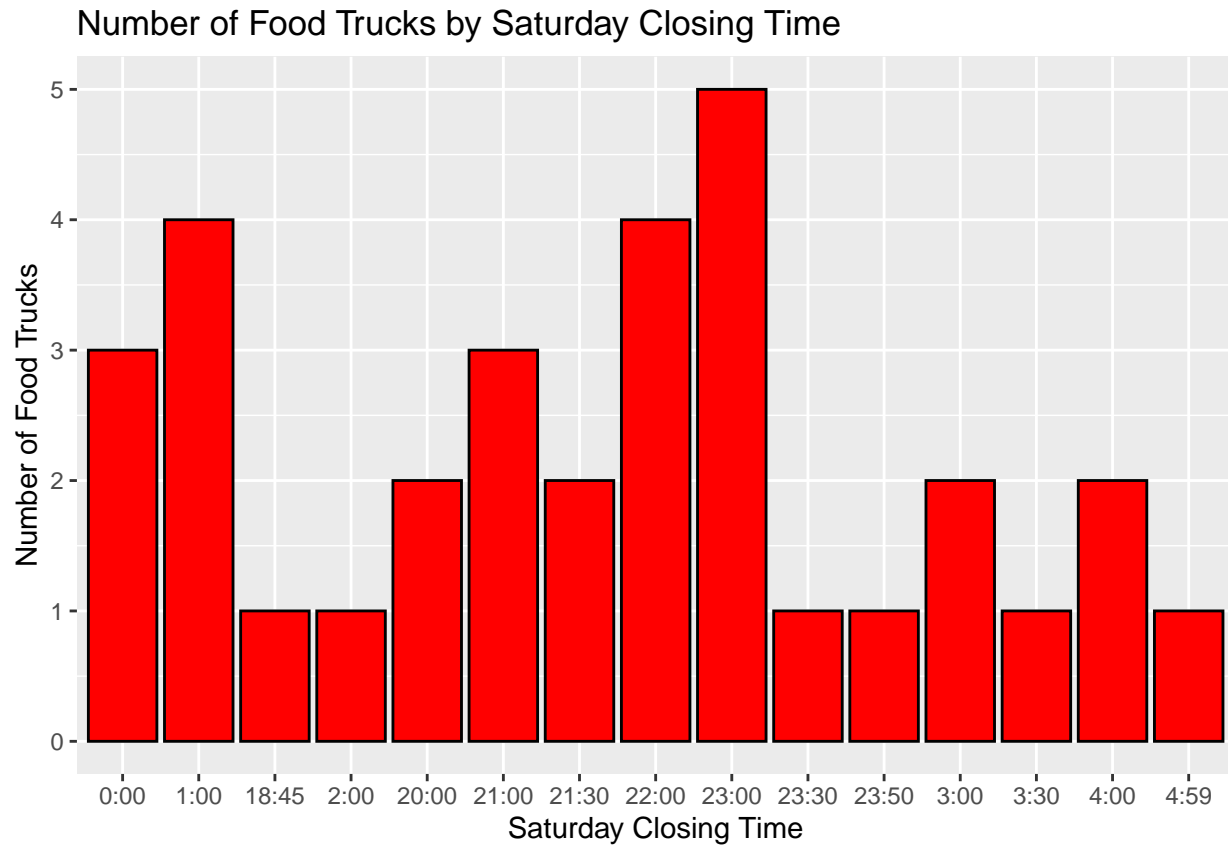
```
banned_hours = c("Closed", "", "Open 24 hours")
indi_foodtrucks_hours = subset(indi_foodtrucks, (!(saturday %in% banned_hours) & !(sunday %in% banned_h
indi_foodtrucks_hours_sat_open = indi_foodtrucks_hours %>%
  arrange(saturday_open)
indi_foodtrucks_hours_sat_close = indi_foodtrucks_hours %>%
  arrange(saturday_close)
indi_foodtrucks_hours_sun_open = indi_foodtrucks_hours %>%
  arrange(sunday_open)
indi_foodtrucks_hours_sun_close = indi_foodtrucks_hours %>%
  arrange(sunday_close)

indi_foodtrucks_hours_sat_open %>%
  ggplot(aes(x = saturday_open)) +
  geom_bar(fill = "green", color = "black") +
  labs(title = "Number of Food Trucks by Saturday Opening Time",
       x = "Saturday Opening Time", y = "Number of Food Trucks")
```

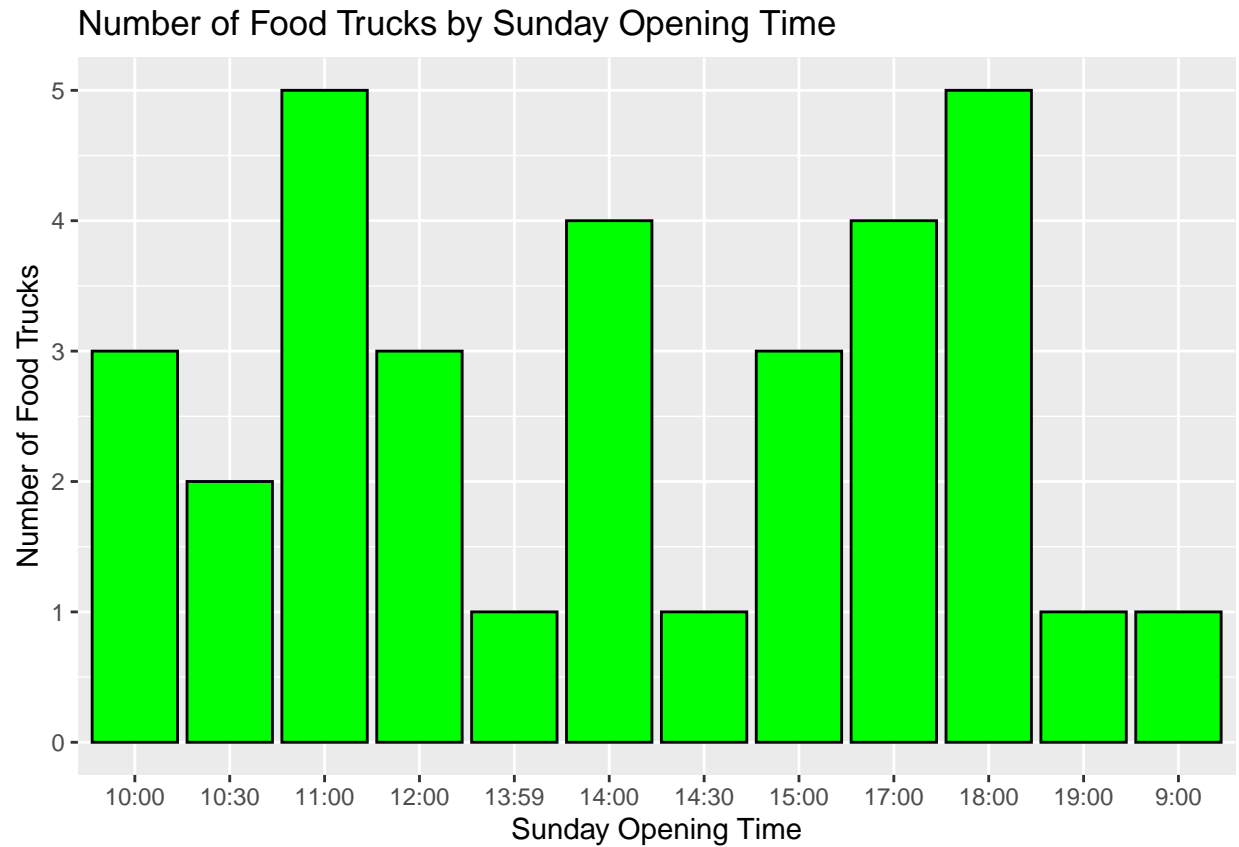
Number of Food Trucks by Saturday Opening Time



```
indi_foodtrucks_hours_sat_close %>%
  ggplot(aes(x = saturday_close)) +
  geom_bar(fill = "red", color = "black") +
  labs(title = "Number of Food Trucks by Saturday Closing Time",
       x = "Saturday Closing Time", y = "Number of Food Trucks")
```

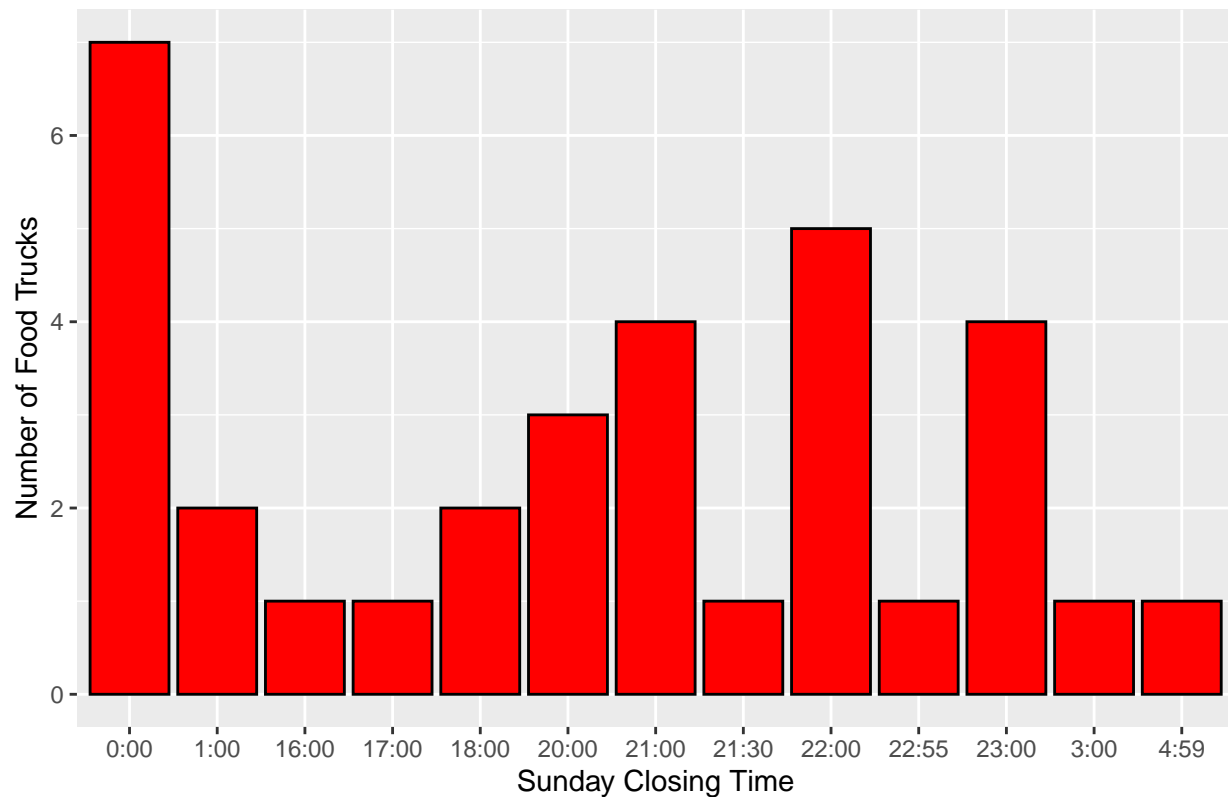


```
indi_foodtrucks_hours_sun_open %>%
  ggplot(aes(x = sunday_open)) +
  geom_bar(fill = "green", color = "black") +
  labs(title = "Number of Food Trucks by Sunday Opening Time",
       x = "Sunday Opening Time", y = "Number of Food Trucks")
```



```
indi_foodtrucks_hours_sun_close %>%  
  ggplot(aes(x = sunday_close)) +  
  geom_bar(fill = "red", color = "black") +  
  labs(title = "Number of Food Trucks by Sunday Closing Time",  
        x = "Sunday Closing Time", y = "Number of Food Trucks")
```

Number of Food Trucks by Sunday Closing Time



Keep in mind that since we are using 24 hour time and due to the limitations of the dataset, 0 for closing refers to midnight and early hours like 1 and 3 refer to the first hours of the next day. 1 for `sunday_close` would represent 1 am on the following Monday. Regardless, our EDA tells us that the most popular starting times for Indianapolis food trucks on Saturday are 11 am and from 3 to 6 pm. The most popular closing times for Saturday are exactly or after midnight, or from 10 to 11. The opening times for Sunday are a bit more distributed but we can note down 10 am to 12 pm, 2 pm, and from 3 pm to 6 pm as popular start times. The closing times for Sunday are mostly at midnight with other popular closing times being from 9 pm to 11 pm.

Therefore, as a general guideline, we can plan our trip in the time frame from 11 am each day to midnight of the same day. For safety, we will likely try to send our tourists back to their hotel by 8 pm, which also avoids brushing up on closing time (food trucks eager to close might prepare food with less care and use less fresh ingredients).

Creating a Plan

Having conducted all of this analysis, let's finalize some plans for the Foodie Weekend. As a reminder, here are the top 10 food trucks based on our part 1 analysis, this time with hours.

```
head(data.frame(sorted_highly_rated_indi_foodtrucks$title, sorted_highly_rated_indi_foodtrucks$rating,
                 sorted_highly_rated_indi_foodtrucks$reviewCount, sorted_highly_rated_indi_foodtrucks$category,
                 sorted_highly_rated_indi_foodtrucks$saturday, sorted_highly_rated_indi_foodtrucks$sunday,
                 sorted_highly_rated_indi_foodtrucks$address), n = 10)
```

```
##      sorted_highly_rated_indi_foodtrucks.title
## 1                      Garcia's Hot Dogs
## 2          Tacos Tierra caliente Restaurant
## 3                      ChefTLC
## 4                      Gulas food truck
## 5          Emergency Taco Truck
## 6          Birrieria Iturbidense
## 7          King gyros food truck
## 8          El Pastor Taco Truck
## 9          Gaucho's Fire Express
## 10 Talkin Turkey Real Cajun Food Truck & Catering
##      sorted_highly_rated_indi_foodtrucks.rating
## 1                      4.9
## 2                      4.8
## 3                      4.8
## 4                      4.8
## 5                      4.7
## 6                      4.7
## 7                      4.7
## 8                      4.7
## 9                      4.7
## 10                     4.7
##      sorted_highly_rated_indi_foodtrucks.reviewCount
## 1                      278
## 2                      118
## 3                      44
## 4                      13
## 5                      104
## 6                      86
## 7                      46
## 8                      46
## 9                      43
## 10                     42
##      sorted_highly_rated_indi_foodtrucks.category
## 1                      Hot dog stand
## 2                      Mexican restaurant
## 3                      Food producer
## 4          Venezuelan restaurant
## 5                      Mexican restaurant
## 6                      Restaurant
## 7                      Gyro restaurant
## 8                      Taco restaurant
## 9                      Fast food restaurant
```

```

## 10 Mobile caterer
## sorted_highly_rated_indi_foodtrucks.saturday
## 1 10:30 am - 6:45 pm
## 2 11 am - 12:59 am
## 3 Closed
## 4
## 5 3 pm - 12 am
## 6 6 pm - 1 am
## 7 10:30 am - 1 am
## 8 11 am - 11 pm
## 9 10 am - 10 pm
## 10
## sorted_highly_rated_indi_foodtrucks.sunday
## 1 10:30 am - 6 pm
## 2 Closed
## 3 11 am - 1 pm
## 4
## 5 Closed
## 6 6 - 11 pm
## 7 10:30 am - 12 am
## 8 2 - 10 pm
## 9 10 am - 9 pm
## 10
## sorted_highly_rated_indi_foodtrucks.address
## 1 5102 E 16th St, Indianapolis, IN 46218, United States
## 2 5035 W Washington St, Indianapolis, IN 46241, United States
## 3
## 4 4002 E Washington St, Indianapolis, IN 46201, United States
## 5 7001 Pendleton Pike, Indianapolis, IN 46226, United States
## 6 Lot Food Truck, 6302 W Washington St, Indianapolis, IN 46241, United States
## 7 4585 S Harding St, Indianapolis, IN 46217, United States
## 8 1053997, Indianapolis, IN 46201, United States
## 9 906 Carrollton Ave Suite 320, Indianapolis, IN 46202, United States
## 10 Commercial Kitchen, 4550 Central Ave, Indianapolis, IN 46205, United States

```

Using this information, we have created a two day plan that takes us to a total of 6 food trucks with distinct cuisine offerings. It is located in the file: *Indianapolis_Trip_Itinerary*

We have purposefully only listed planned times for the food truck visits because we believe in the importance of letting the tourists explore the city on their own. Indianapolis is filled with attractions and only the tourists themselves know what they want to see. Given the Xterns' excellent technical abilities, they should be able to search for fun attractions in Indianapolis en route to each of the pre-set destinations and have fun in the plenty of downtime our planned trip gives them. Should the Xterns feel particularly ambitious, the `highly_rated_indi_foodtrucks` data set in this report gives them plenty of supplementary options. For the same reason, this trip is by car, giving the Xterns maximum freedom.

Regardless our pre-planned route will take them past museums, parks, Universities, and even the airport, all while giving them a first-class tour of Indianapolis's diverse neighborhoods. Xterns will be sure to gain a new appreciation for Indianapolis's wonderful food, all while having the time of their lives!