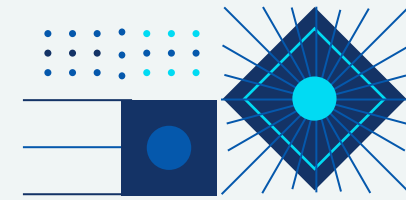




# Improving the Health & Safety of Dining Establishments in Butler County



Magda Odrowaz, Riley Nguyen, Matt Smuda & Drew Nanda



# Meet the Team:



**Drew Nanda**  
Beavercreek, OH  
B.S. Marketing '22  
M.S. Sports Analytics '23



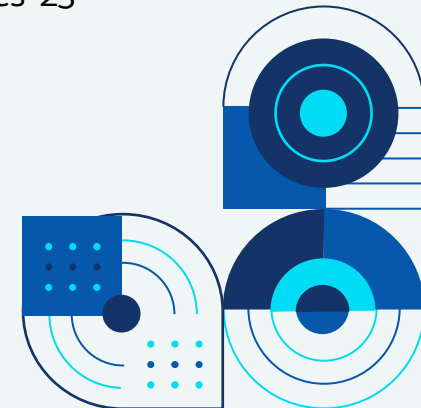
**Magda Odrowaz**  
Buffalo Grove, IL  
B.S. Kinesiology & Health '22  
M.S. Business Analytics '23

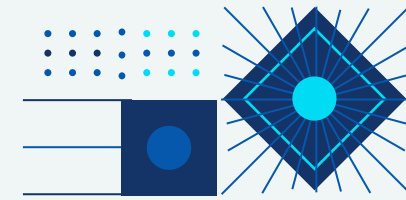


**Riley Nguyen**  
Ho Chi Minh, Vietnam  
B.S. Business Analytics '23  
M.S. Business Analytics '23



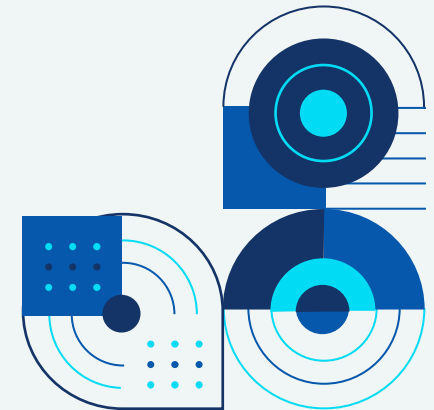
**Matt Smuda**  
Palos Park, IL  
B.S. Finance '22  
M.S. Sports Analytics '23

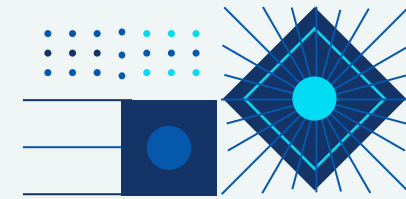




# Agenda

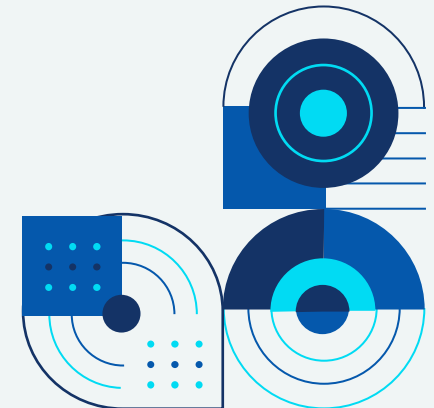
1. Background
2. Project Scope
3. Data Preparation
4. Exploratory Analysis
5. Modeling
6. Conclusion
7. Limitations

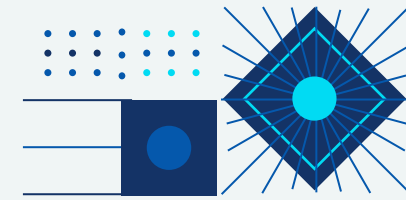




# Background

- Foodborne illnesses due to unsafe food handling are a significant public health problem
  - Known foodborne disease agents are estimated to cause approximately 9.4 million illnesses each year in the United States (CDC)
- Butler County Health Department inspects the community's establishments at least twice a year and/or when complaints are filed

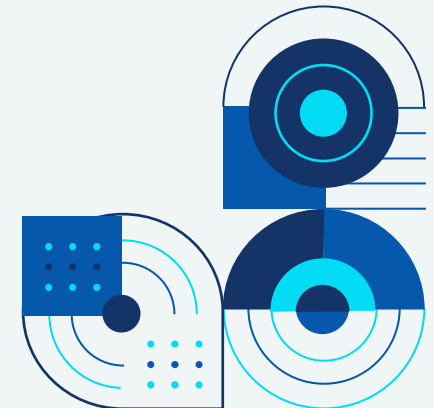


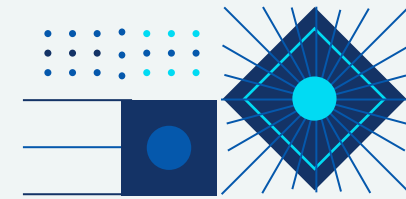


# Background

- Gastrointestinal disease...
  - Missed days of work/school
  - Life threatening to immunocompromised population
  - Increase risk of hospitalization

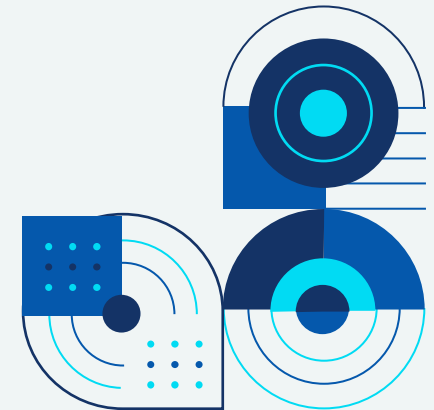
**Risk** mitigated with frequent inspections and inspections due to complaints.



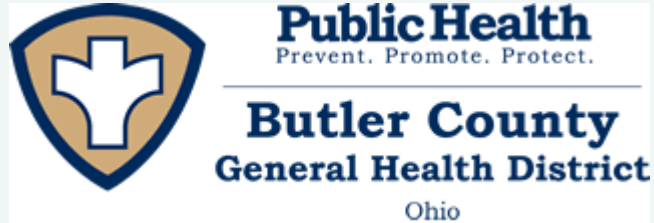
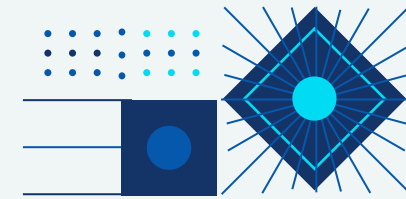


# Project Goals

- **Our goal** is to build a model that predicts when inspections and/or intervention by the Butler County Health Department are necessary
- **Provide** a tool to the Butler County Health Department to efficiently audit and monitor eateries in Butler County
- **Analyze** sentiment in Yelp reviews to connect public opinions with the risk at eateries
- **Advise** the public about where they are eating



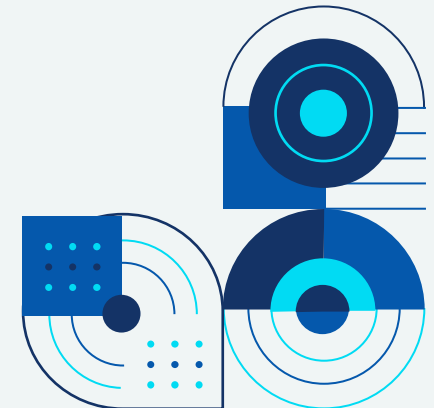
# Data Sources



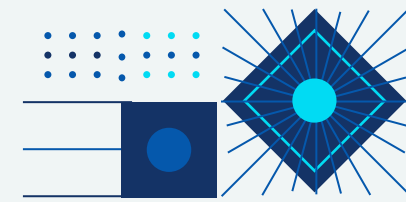
- Health Inspection Records
- Facility information
  - Phone number
  - Address
- Critical & Non-Critical Inspection counts
- Inspection Type



- String type Yelp reviews
- Restaurant Name
- Phone Number
- Address
- Date of Review
- Rating







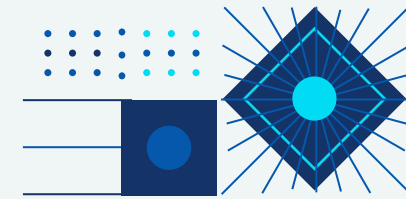
# Data Preparation: Integrating Yelp and Inspections

185 restaurants

|    | A   | B                 | C            | D        | E                | F          | G              | H              | I              | J                 | K                                | L                    | M               | N                  |  |
|----|-----|-------------------|--------------|----------|------------------|------------|----------------|----------------|----------------|-------------------|----------------------------------|----------------------|-----------------|--------------------|--|
| 1  |     | Facility Name_x   | Non-Critical | Critical | Location         | Unnamed: 0 | Facility Type  | Risk Rating    | Phone Number   | Restaurant Nam    | Address                          | Review               | Number of Revie | Non_Critical_Risks |  |
| 2  | 0   | Bee's Buffet      | 0            | 0        | 725 Nilles Rd,Fa | 329        | Commercial FSC | Risk Level IV  | (513) 858-6497 | Bee's Buffet      | 725 Nilles Rd Fa                 | I'll pretext this wi | 23              | Good               |  |
| 3  | 4   | Aangan Sweets     | 10           | 1        | 9020 Sutton Pl,H | 1025       | Commercial FSC | Risk Level III | (513) 807-8703 | Aangan Sweets     | 9020 Sutton Pl H                 | I stopped in for s   | 16              | Bad                |  |
| 4  | 5   | Acapulco          | 9            | 3        | 5953 Boymel Dr,  | 46         | Commercial FSC | Risk Level IV  | (513) 874-5777 | Acapulco          | 5953 Boymel Dr                   | We have been g       | 76              | Bad                |  |
| 5  | 7   | Al Zaytuna Grill  | 13           | 7        | 3425 Princeton F | 331        | Commercial FSC | Risk Level IV  | (513) 737-2000 | Al Zaytuna Grill  | 3425 Princeton F                 | This place is am     | 99              | Bad                |  |
| 6  | 8   | Aladdin's Eatery  | 5            | 3        | 9344 Union Cen   | 1027       | Commercial FSC | Risk Level IV  | (513) 874-1302 | Aladdin's Eatery  | 3664 Edwards Rd Ste A Cincinna   |                      | 100             | Good               |  |
| 7  | 18  | Another Broken    | 5            | 2        | 7701 Voice Of A  | 1036       | Commercial FSC | Risk Level IV  | -513           | Another Broken    | 7701 Voice Of America Centre Dr  |                      | 100             | Good               |  |
| 8  | 20  | Applebee's Grill  | 1            | 2        | 5331 Pleasant A  | 53         | Commercial FSC | Risk Level IV  | (513) 939-1790 | Applebee's Grill  | 5331 Pleasant A                  | Very good they e     | 46              | Good               |  |
| 9  | 22  | Arby's            | 5            | 3        | 558 Nilles Rd,Fa | 54         | Commercial FSC | Risk Level III | (513) 829-5253 | Arby's            | 2 Lynn Ave Oxfo                  | If there were 6 st   | 8               | Good               |  |
| 10 | 31  | Asiana Korean F   | 11           | 6        | 6875 Fountains   | 1041       | Commercial FSC | Risk Level IV  | (513) 755-9888 | Asiana Korean F   | 6875 Fountains Blvd Ste G West   |                      | 94              | Bad                |  |
| 11 | 35  | BC's Bottle Lodg  | 0            | 0        | 7121 Liberty Cer | 519        | Commercial RFE | Risk Level II  | (513) 847-6866 | BC's Bottle Lodg  | 7121 Liberty Cer                 | I have been goin     | 32              | Good               |  |
| 12 | 37  | Bob Evans         | 11           | 5        | 5076 College Co  | 804        | Commercial FSC | Risk Level IV  | (513) 523-8866 | Bob Evans         | 3524 Indian Mea                  | I have seen Bob      | 30              | Bad                |  |
| 13 | 47  | Wings and Rings   | 14           | 2        | 5486 Dixie Hwy,I | 68         | Commercial FSC | Risk Level IV  | -513           | Wings and Rings   | 4782 Plum Sq W                   | Excellent!! order    | 33              | Bad                |  |
| 14 | 50  | The Back Porch    | 6            | 2        | 9626 Princeton C | 1047       | Commercial FSC | Risk Level IV  | (513) 874-2432 | The Back Porch    | 9626 Princeton Glendale Rd Wes   |                      | 100             | Bad                |  |
| 15 | 51  | Bargos Grill & Te | 11           | 3        | 7105 Dixie Hwy,I | 59         | Commercial FSC | Risk Level III | (513) 860-5500 | Bargos Grill & Te | 7105 Dixie Hwy                   | This review will k   | 37              | Bad                |  |
| 16 | 52  | Barn N Bunk - O   | 0            | 0        | 3677 Wayne Ma    | 979        | Commercial RFE | Risk Level I   | (513) 988-9211 | Barn N Bunk - O   | 3677 Wayne Ma                    | Love the fall fest   | 13              | Good               |  |
| 17 | 57  | Bella Sweets      | 0            | 0        | 4877 Princeton F | 337        | Commercial FSC | Risk Level III | -513           | Bella Sweets      | 4877 Princeton F                 | This morning we      | 7               | Good               |  |
| 18 | 58  | Berd's Grill & Ba | 0            | 0        | 500 Wessel Dr,V  | 61         | Commercial FSC | Risk Level IV  | (513) 227-7115 | Berd's Grill & Ba | 500 Wessel Dr V                  | Great local sport    | 50              | Good               |  |
| 19 | 61  | Big Buls          | 0            | 3        | 2461 Ross Millvi | 338        | Commercial FSC | Risk Level IV  | (513) 844-2857 | Big Buls          | 2461 Ross Millvi                 | First time here a    | 22              | Good               |  |
| 20 | 62  | Big Dog's Pizza   | 32           | 6        | 7967 Cincinnati  | 1053       | Commercial FSC | Risk Level III | (513) 779-8300 | Big Dog's Pizza   | 7967 Cincinnati                  | Great food Just l    | 76              | Bad                |  |
| 21 | 71  | Buffalo Bob's Fa  | 16           | 3        | 6130 Dixie Hwy,I | 67         | Commercial FSC | Risk Level IV  | (513) 829-8118 | Buffalo Bob's Fa  | 6130 Dixie Hwy                   | I love the food h    | 1               | Bad                |  |
| 22 | 72  | Buffalo Wild Win  | 17           | 5        | 3417 Princeton F | 343        | Commercial FSC | Risk Level IV  | (513) 737-2999 | Buffalo Wild Win  | 10 East Walnut S                 | I see some low r     | 30              | Bad                |  |
| 23 | 99  | Burger King       | 5            | 0        | 7782 Dudley Dr,  | 1067       | Commercial FSC | Risk Level IV  | (513) 475-3700 | Burger King       | 5093 College Co                  | I loved Burger Ki    | 5               | Good               |  |
| 24 | 146 | C & C Lounge      | 3            | 2        | 2518 Cincinnati  | 965        | Commercial FSC | Risk Level II  | (513) 738-9931 | C & C Lounge      | 2518 Cincinnati                  | Smoking hasn't l     | 3               | Good               |  |
| 25 | 147 | Cassano's Pizza   | 2            | 1        | 4761 Dixie Hwy,I | 77         | Commercial FSC | Risk Level III | (513) 829-8880 | Cassano's Pizza   | 4761 Dixie Hwy                   | This is an odd pl    | 47              | Good               |  |
| 26 | 149 | Cavu Coffee       | 4            | 6        | 7755 Cox Ln,We   | 1076       | Commercial FSC | Risk Level III | (513) 755-2288 | Cavu Coffee       | 7755 Cox Ln West Chester, OH 4   |                      | 100             | Good               |  |
| 27 | 152 | Chipotle Mexica   | 0            | 1        | 7717 Voice Of A  | 1085       | Commercial FSC | Risk Level IV  | (513) 779-3615 | Chipotle Mexica   | 1 W High St Ste                  | I always enjoy th    | 75              | Good               |  |
| 28 | 176 | Captain D's       | 33           | 7        | 490 Nilles Rd,Fa | 74         | Commercial FSC | Risk Level IV  | -513           | Captain D's       | 490 Nilles Rd Fa                 | Great service for    | 9               | Bad                |  |
| 29 | 178 | Caruso's Ristora  | 4            | 2        | 6765 Dixie Hwy,I | 76         | Commercial FSC | Risk Level IV  | (513) 882-3210 | Caruso's Ristora  | 6765 Dixie Hwy Fairfield, OH 450 |                      | 100             | Good               |  |
| 30 | 184 | Chick-fil-A       | 5            | 0        | 7733 Voice Of A  | 1083       | Commercial FSC | Risk Level IV  | (513) 779-1767 | Chick-fil-A       | 3403 Princeton F                 | The service is th    | 28              | Good               |  |

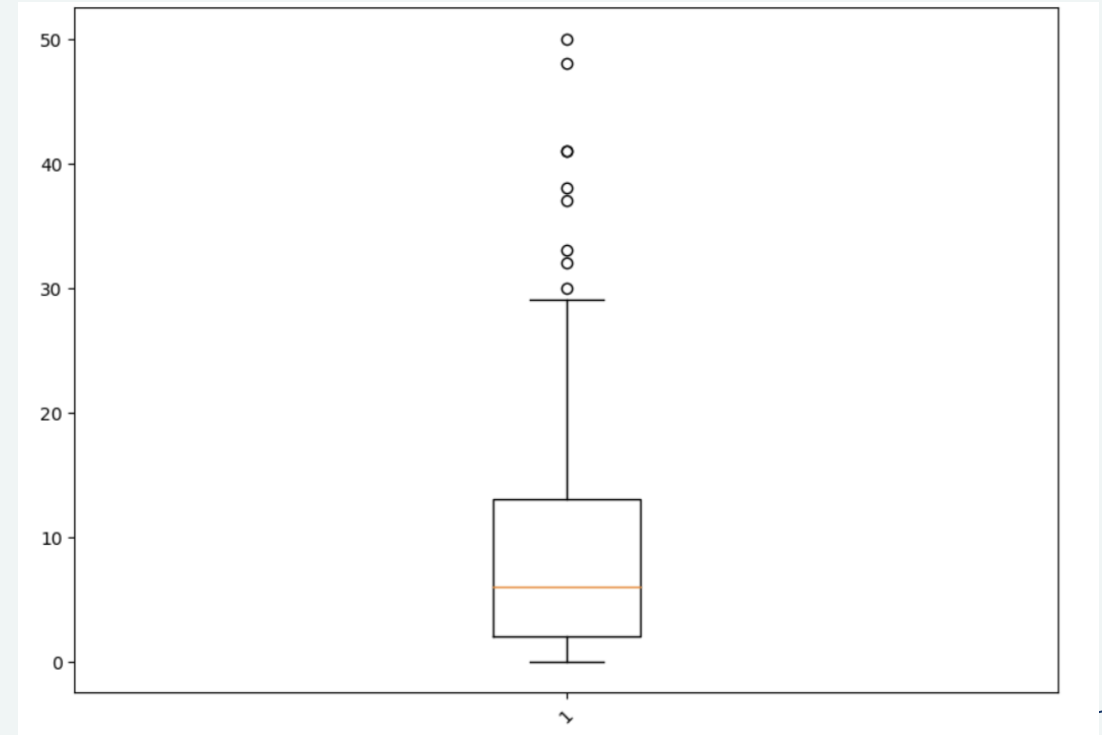






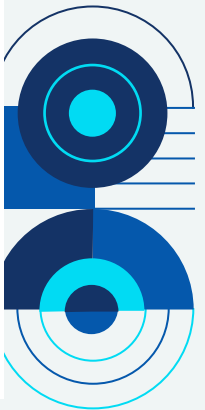
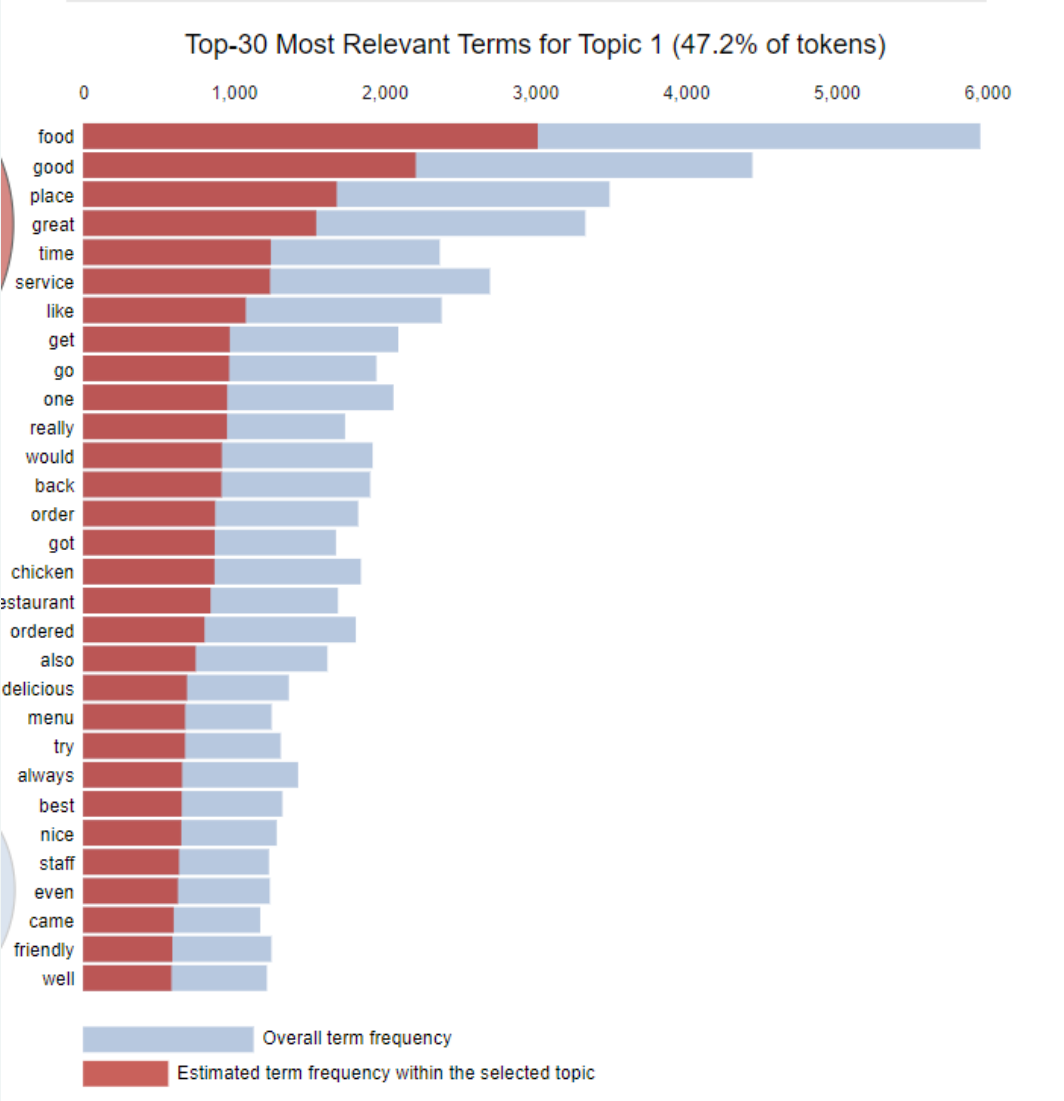
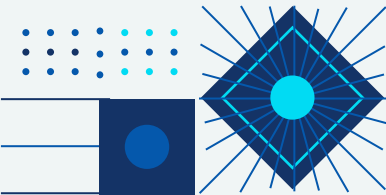
# Data Preparation: Variable Information

- 185 restaurants
- Critical & Non critical violation counts
- Address - City
- Restaurant Name
- Yelp Rating from customers
- Risk Rating- established by Butler County
- Reviews
- # of reviews
- Non critical violation count **classifier**
  - Benchmark = 6 = median of non critical violation frequency
  - **Good** = non critical violations < 6
  - **Bad** = non critical violations > 6



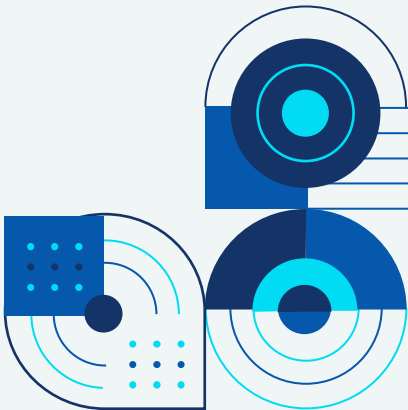
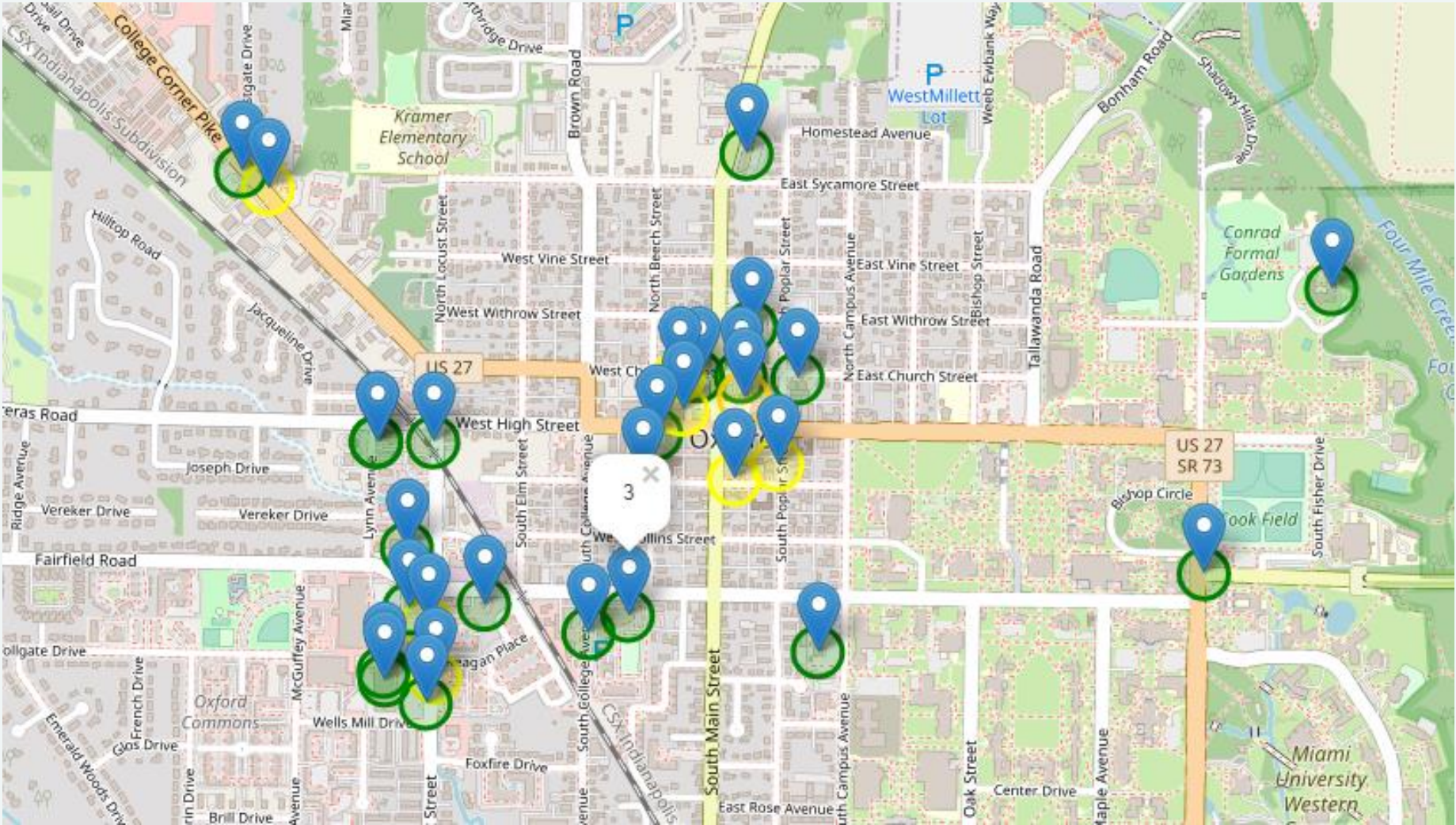
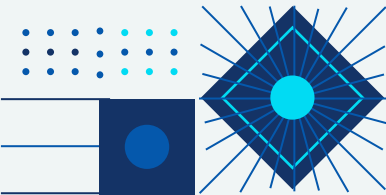


# Exploratory Analysis: Topic Modeling

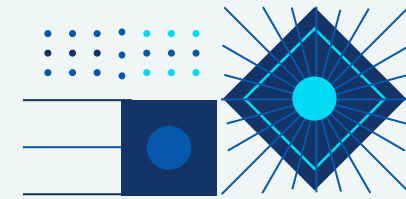




# Exploratory Analysis: Map



# Model Building

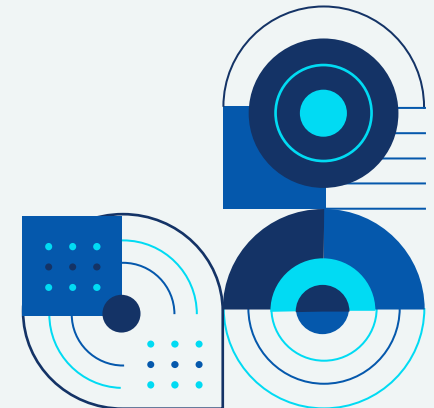


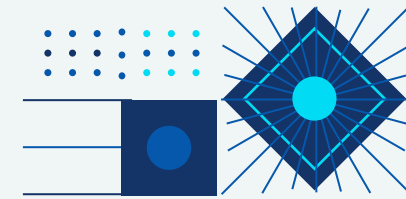
## Predictor Variables

- Restaurant Name
- Yelp Name
- Risk level
- Number of Reviews
- Non Critical Risks
- City
- Sentiment of Reviews
  - ◆ Positive
  - ◆ Negative
  - ◆ Neutral

## Response Variables

- Non critical risks
  - ◆ Good if  $< 6$  violations
  - ◆ Bad if  $> 6$  violations





# Evaluation of Model

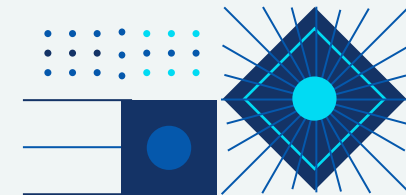
Model type used: Random Forest Classification

Overall Accuracy: 0.5541

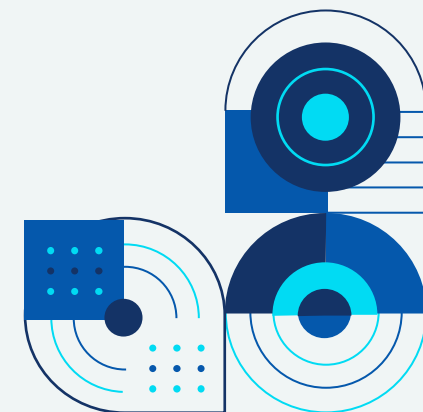
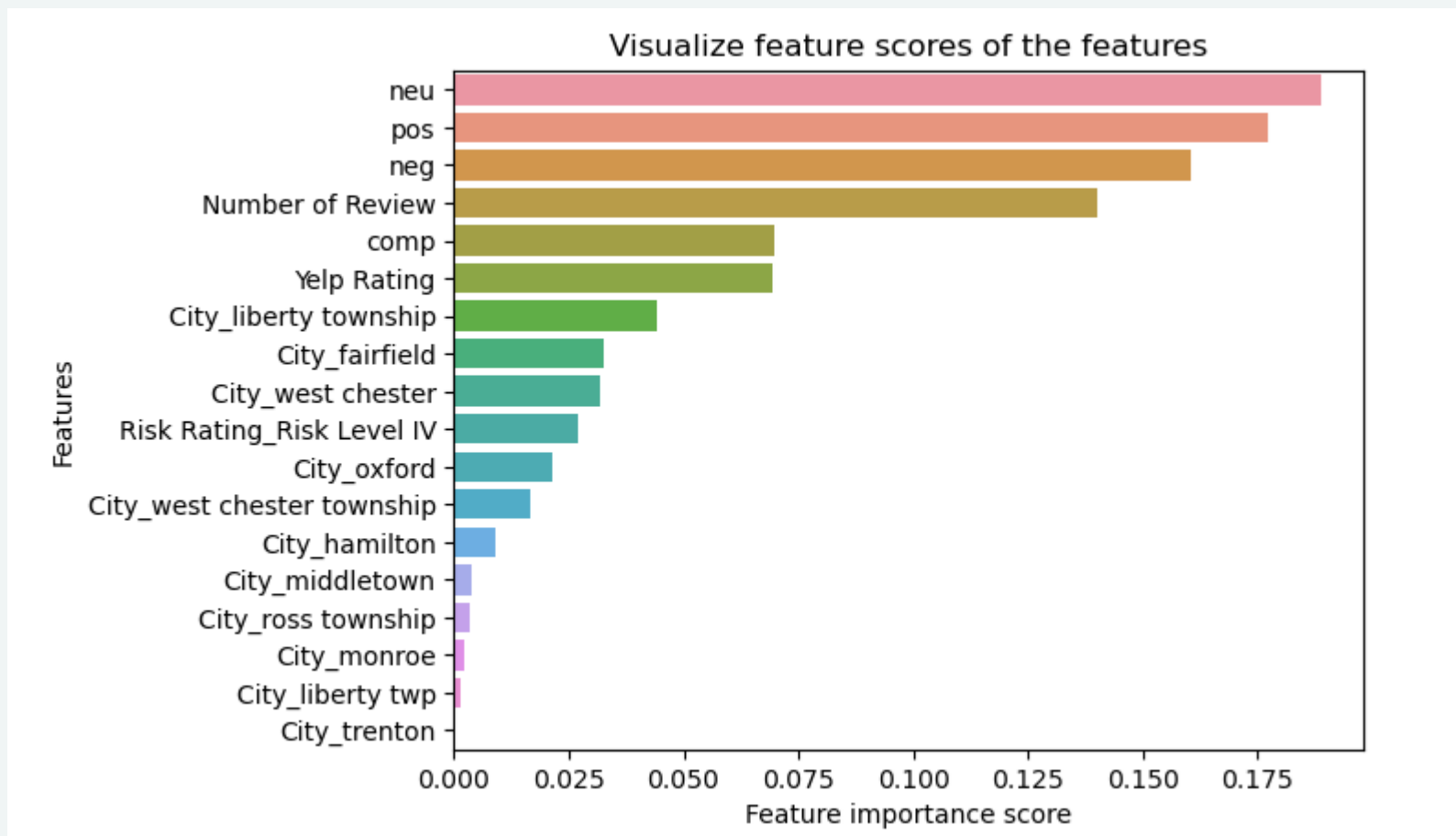
|      | Precision | Recall |
|------|-----------|--------|
| Bad  | 0.6       | 0.51   |
| Good | 0.53      | 0.60   |

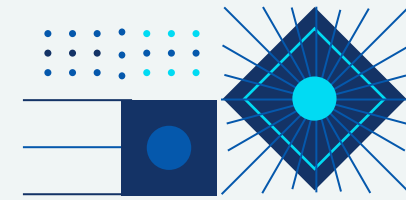






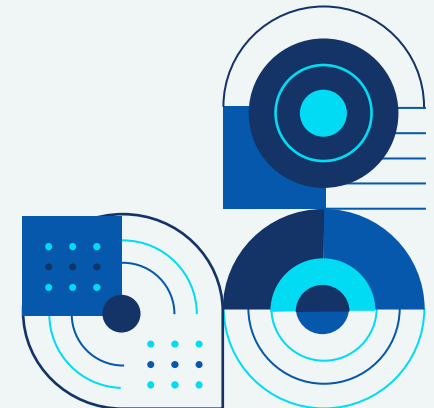
# Evaluation of Model - Feature Importance

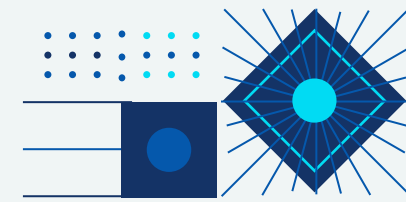




# Conclusions

- Yelp Rating and Reviews sentiment did contribute to the prediction of Good/Bad restaurants
- Customers are not as interested in sanitary attributes as they are in food taste and service in Yelp reviews
- Frequency of historical critical and non critical violation counts are most indicative of establishment risk

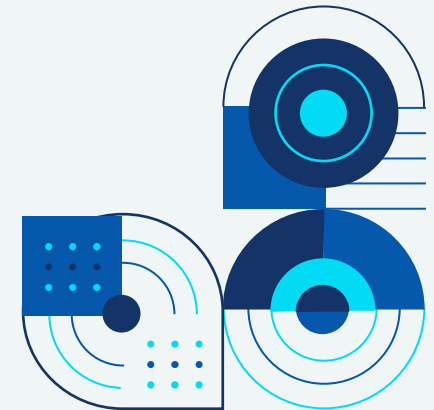


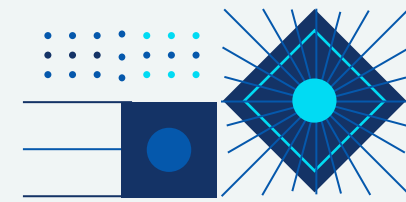


# Limitations

- Limited number of observations in the dataset
  - Focus limited to just Butler County
- Positive sentiment yelp reviews far outweigh the negative sentiment reviews
  - Yelp reviews that violated their terms of service can be removed
  - These are likely more negative reviews, containing inappropriate language
- Negative Yelp reviews scoring lower negative sentiment scores due to their use of words like “good”, even when in a negative context

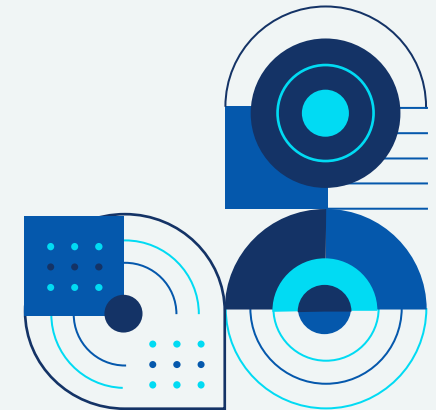
Google

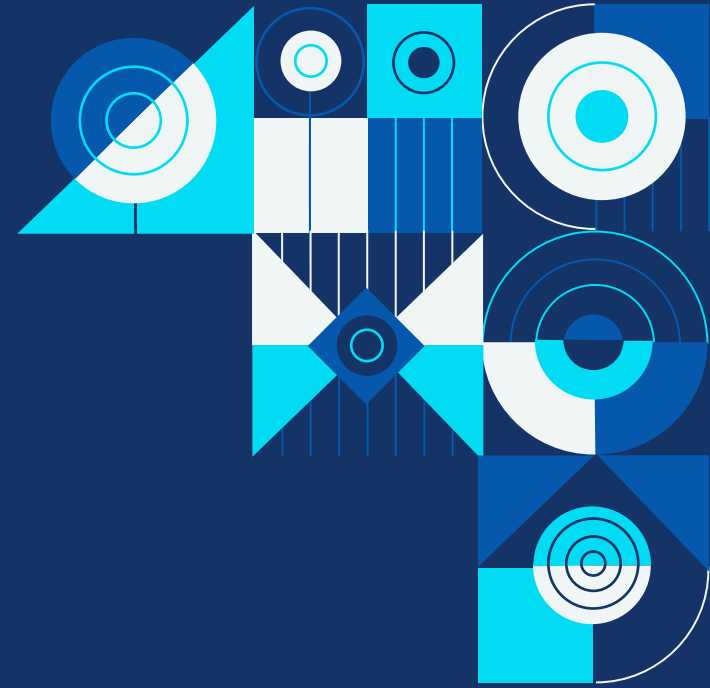
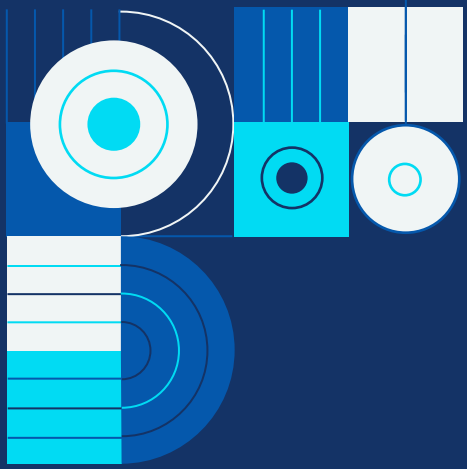




# Looking Forward

- Focus on a more highly populated area
  - Opportunity for more observations within the data
- Find other reviews sources (Google reviews, etc.) that may provide more for the analysis





Thank you!