# The Heart of Campustown

## Analyzing Green Street's Most Popular Restaurants

Pranav Chandra
pranavc5@illinois.edu

Kavya Moharana
kavyam3@illinois.edu

Sriya Mikkilineni
sriyam2@illinois.edu

Aryan Vaswani
aryangv2@illinois.edu

## ABSTRACT

Green Street's wide variety of restaurants and cafes is an integral part of student life at the University of Illinois at Urbana-Champaign. Known as the "Heart of Campustown", these establishments provide an important cultural and social hub for students, faculty, and residents. We intend to conduct sentiment analysis of the Yelp reviews for multiple major restaurants on this historical street. We will analyze online reviews over time periods to determine how students and food enthusiasts perceive these restaurants. Through this evaluation, we can see how the reception of these various eateries has changed over the past decade.

## INTRODUCTION

We split this project into three major areas: the retrieval of Yelp reviews from online databases, the sentiment analysis per time period for each restaurant's reviews, and statistical analysis to develop graphs on how students feel overall about the restaurants. Our research on the following major topics guided us throughout the project.

To gather an understanding of the different methods behind sentiment analysis, we reviewed multiple pieces of literature and narrowed down which techniques to focus on. Taboada defines the term sentiment as "positive or negative evaluation expressed through language" [2]. Sentiment analysis has a variety of applications, including determining whether an online review has a positive connotation or a negative one. A research study by Liu compared the efficiency of several sentiment analysis tools and processes on Yelp reviews [3]. They found that using a binary bag-of-words representation with the addition of bi-grams had positive effects on the performance of their analysis models. A binary bag-of-words model records the number of occurrences of each bag that is created for each term without accounting for order or grammar [4]. It allows us to add "meaning" to each term that is used for extra parameters. Using bi-grams involves taking two consecutive terms as a pair for each term of the vocabulary to construct the bag of words.

Furthermore, the imposition of minimum frequency constraints and text normalization has positive effects. Minimum frequency constraints allow us to filter out terms that may not hold significant semantic meaning in our analysis and improve the robustness of the model. Text normalization involves modifying the texts to account for casing, tokenization, stop words, and other parameters. We intend to implement these four techniques in our project to improve the efficiency of our model.

## DESCRIPTION

The intended goal of our project is to utilize Yelp reviews to gauge the popularity of restaurants on Green Street using a variety of methods. The first of these methods will be to illustrate the overall sentiment of restaurants relative to each other over given time segments using a graph. Additionally, we aim to produce a ranking of the current favorite and most-disliked restaurants. Lastly, we intend to articulate, with reasonable precision, whether each of the group members' favorite restaurants is similarly viewed with respect to sentiment by the data we scraped as a fun addition to the project.

## ACTIVITES

1. Decide which Campustown restaurants to include in our sentiment analysis (2 hours)

2. Create a web scraping algorithm to glean data from Yelp reviews for relevant restaurants (13 hours)

3. Remove irrelevant reviews from data to perform data cleaning (5 hours)

4. Label data required for training (10 hours)

5. Decide which sentiment analysis techniques to use (5 hours)

6. Utilize sentiment analysis techniques to ascertain positive, neutral, and negative sentiments for relevant reviews (15 hours)

7. Use generated sentiment analysis to produce the above-described graph and ranking (3 hours)

8. Produce a final report (2 hours)

## TASK

We will create a script that web scrapes and/or utilizes APIs to access data from Google Reviews and Yelp to download reviews of a list of restaurants (that we will predefine) over multiple years. We will analyze each downloaded review and take the average of each restaurant's sentiments (0 for purely negative through 1 for purely positive) for each year. We will utilize these yearly averages of each restaurant to construct a line plot that explains the trend in sentiments.

## DATA

We utilized two different datasets during this project: one for the testing of five different models and one for the evaluation of our actual project. To understand why we needed two models, we need to look at Yelp's GraphQL API documentation [5]. For this project, we have identified 12 restaurants on Green Street for which we were interested in understanding peoples' sentiments towards them over time. Ideally, we would pull a high number of reviews from each of the restaurants and get an understanding of sentiments over the years. The data is available and as Figure 1 shows, there is a lot of data at our disposal.

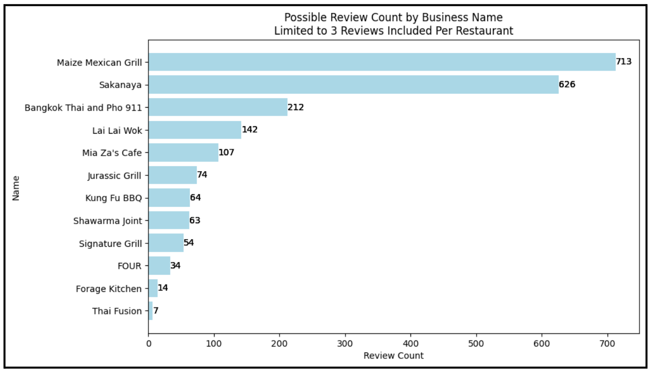
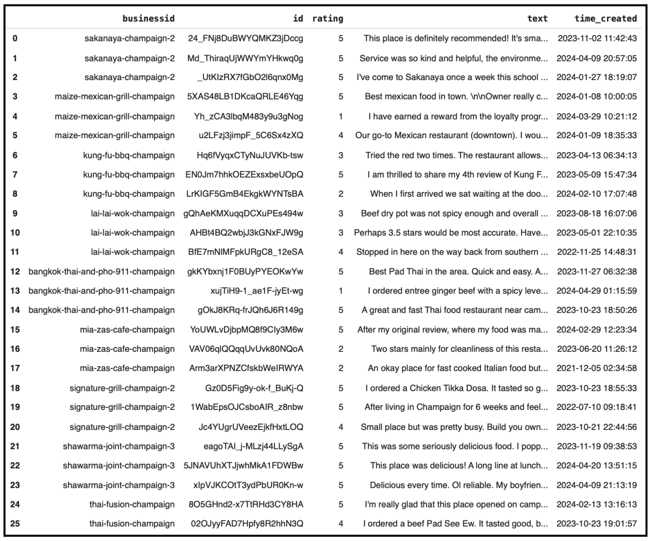


Figure 1

However, because Yelp API is still in its beta stage, we are limited to only the first three reviews on each restaurant's Yelp page. This severely limits our ability in trying to understand Champaign's sentiments towards these restaurants over the years. However, our code still shows sentiments over a couple of months and when the API limits increase, we would only have to re-run the code as is to get all the data.

The other option we considered was to use a web scraper to parse through each restaurant's Yelp page and pull the necessary data for each review. However, we faced trouble as Yelp's website isn't very standard. It also becomes increasingly difficult to parse because there is no standard reviews page that shows all the data we need in a clear, easy, parsable format.

Therefore, as shown in Figure 2, we are left with these 12 restaurants and we have been able to pull three reviews for each restaurant using GraphQL queries for each restaurant, leaving us with a dataset of 36 documents that we will later use after selecting our model.



| | businessid | id | rating | text | time_created |
|---|---|---|---|---|---|
| 0 | sakanaya-champaign-2 | 24_FNj8DuBWYQMKZ3jDccg | 5 | This place is definitely recommended! It's sma... | 2023-11-02 11:42:43 |
| 1 | sakanaya-champaign-2 | Md_ThiraqUjWWYmYHkwq0g | 5 | Service was so kind and helpful, the environme... | 2024-04-09 20:57:05 |
| 2 | sakanaya-champaign-2 | _UtKIzRX7fGbO2l6qnx0Mg | 5 | I've come to Sakanaya once a week this school ... | 2024-01-27 18:19:07 |
| 3 | maize-mexican-grill-champaign | 5XAS48LB1DKcaQRLE46Yqg | 5 | Best mexican food in town. \n\nOwner really c... | 2024-01-08 10:00:05 |
| 4 | maize-mexican-grill-champaign | Yh_zCA3lbqM483y9u3gNog | 1 | I have earned a reward from the loyalty progr... | 2024-03-29 10:21:12 |
| 5 | maize-mexican-grill-champaign | u2LFzj3jimpF_5C6Sx4zXQ | 4 | Our go-to Mexican restaurant (downtown). I wou... | 2024-01-09 18:35:33 |
| 6 | kung-fu-bbq-champaign | Hq6fVyqxCTyNuJUVKb-tsw | 3 | Tried the red two times. The restaurant allows... | 2023-04-13 06:34:13 |
| 7 | kung-fu-bbq-champaign | EN0Jm7hhkOEZExsxbeUOpQ | 5 | I am thrilled to share my 4th review of Kung F... | 2023-05-09 15:47:34 |
| 8 | kung-fu-bbq-champaign | LrKIGF5GmB4EkgkWYNTsBA | 2 | When I first arrived we sat waiting at the doo... | 2024-02-10 17:07:48 |
| 9 | lai-lai-wok-champaign | gQhAeKMXuqqDCXuPEs494w | 3 | Beef dry pot was not spicy enough and overall ... | 2023-08-18 16:07:06 |
| 10 | lai-lai-wok-champaign | AHBt4BQ2wbjJ3kGNxFJW9g | 3 | Perhaps 3.5 stars would be most accurate. Have... | 2023-05-01 22:10:35 |
| 11 | lai-lai-wok-champaign | BfE7mNlMFpkURgC8_12eSA | 4 | Stopped in here on the way back from southern ... | 2022-11-25 14:48:31 |
| 12 | bangkok-thai-and-pho-911-champaign | gkKYbxnj1F0BUyPYEOKwYw | 5 | Best Pad Thai in the area. Quick and easy. A... | 2023-11-27 06:32:38 |
| 13 | bangkok-thai-and-pho-911-champaign | xujTiH9-1_ae1F-jyEt-wg | 1 | I ordered entree ginger beef with a spicy leve... | 2024-04-29 01:15:59 |
| 14 | bangkok-thai-and-pho-911-champaign | gOkJ8KRq-frJQh6J6R149g | 5 | A great and fast Thai food restaurant near cam... | 2023-10-23 18:50:26 |
| 15 | mia-zas-cafe-champaign | YoUWLvDjbpMQ8f9CIy3M6w | 5 | After my original review, where my food was ma... | 2024-02-29 12:23:34 |
| 16 | mia-zas-cafe-champaign | VAV06qlQQqqUvUvk80NQoA | 2 | Two stars mainly for cleanliness of this resta... | 2023-06-20 11:26:12 |
| 17 | mia-zas-cafe-champaign | Arm3arXPNZCfskbWeIRWYA | 2 | An okay place for fast cooked Italian food but... | 2021-12-05 02:34:58 |
| 18 | signature-grill-champaign-2 | Gz0D5Fig9y-ok-f_BuKj-Q | 5 | I ordered a Chicken Tikka Dosa. It tasted so g... | 2023-10-23 18:55:33 |
| 19 | signature-grill-champaign-2 | 1WabEpsOJCsboAIR_z8nbw | 5 | After living in Champaign for 6 weeks and feel... | 2022-07-10 09:18:41 |
| 20 | signature-grill-champaign-2 | Jc4YUgrUVeezEjkfHxtLOQ | 4 | Small place but was pretty busy. Build you own... | 2023-10-21 22:44:56 |
| 21 | shawarma-joint-champaign-3 | eagoTAI_j-MLzj44LLySgA | 5 | This was some seriously delicious food. I popp... | 2023-11-19 09:38:53 |
| 22 | shawarma-joint-champaign-3 | 5JNAVUhXTJjwhMkA1FDWBw | 5 | This place was delicious! A long line at lunch... | 2024-04-20 13:51:15 |
| 23 | shawarma-joint-champaign-3 | xIpVJKCOtT3ydPbUR0Kn-w | 5 | Delicious every time. Ol reliable. My boyfrien... | 2024-04-09 21:13:19 |
| 24 | thai-fusion-champaign | 8O5GHnd2-x7TtRHd3CY8HA | 5 | I'm really glad that this place opened on camp... | 2024-02-13 13:16:13 |
| 25 | thai-fusion-champaign | 02OJyyFAD7Hpfy8R2hhN3Q | 4 | I ordered a beef Pad See Ew. It tasted good, b... | 2023-10-23 19:01:57 |

Figure 2

For the model selection, we needed a bigger dataset that would allow us to both fine-train and test the model. We could then use our chosen metrics on these models and evaluate which specific model to use on our own Yelp dataset. On Kaggle [1], we found a dataset of yelp-reviews already labeled using positive and negative. The test set included 38000 documents, of which we used 10000 documents. As explained later, we pulled pre-trained models from the internet and then tested their accuracies using the test set. We pulled the data using a simple system command 'kaggle datasets download -d hhalalwi/yelp-light --unzip'. This command downloads three files. For our purposes, we only used the "raw_train.csv" file.

In total, pulling all the Yelp data we needed using GraphQL requests took about 2 minutes and the Kaggle command took 4 seconds.

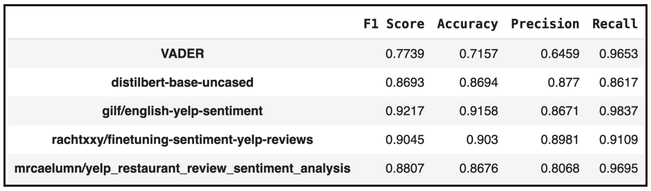
## MODEL

For this project, we tested five different models among which one was utilized on our own dataset to produce the analysis we set out in the first place. We used NLTK's VADER (Valence Aware Dictionary and sEntiment Reasoner) module and four HuggingFace models:

- distilbert-base-uncased
- gilf/english-yelp-sentiment
- rachtxxy/finetuning-sentiment-yelp-reviews

- mrcaelumn/yelp_restaurant_review_sentiment_analysis

It is important to note that each of these models have already been trained using yelp review data. But, we still need to test each of the models to select the one that would perform best for our analysis. We used four different metrics (F1 score, accuracy, precision, and recall) to determine the best-performing model. For the actual metric calculations, we used scikit-learn's premade functions (the y_true input was taken from the CSV file and the y_pred input was just the output from each of the models). As shown in Figure 3, the best model according to every metric we used was the "gilf/english-yelp-sentiment" model. Choosing this specific model was very easy as the only metric that it didn't perform the best in was recall.



| | F1 Score | Accuracy | Precision | Recall |
|---|---|---|---|---|
| VADER | 0.7739 | 0.7157 | 0.6459 | 0.9653 |
| distilbert-base-uncased | 0.8693 | 0.8694 | 0.877 | 0.8617 |
| gilf/english-yelp-sentiment | 0.9217 | 0.9158 | 0.8671 | 0.9837 |
| rachtxxy/finetuning-sentiment-yelp-reviews | 0.9045 | 0.903 | 0.8981 | 0.9109 |
| mrcaelumn/yelp_restaurant_review_sentiment_analysis | 0.8807 | 0.8676 | 0.8068 | 0.9695 |

Figure 3

## ANALYSIS

After choosing the model, we proceeded with our own analysis of the Yelp data. First, we created a simple plot that used the "star-rating" of every review as opposed to anything text related. This was simply a way to provide a baseline for further analysis. Figure 4 displays the ratings of every restaurant over time.

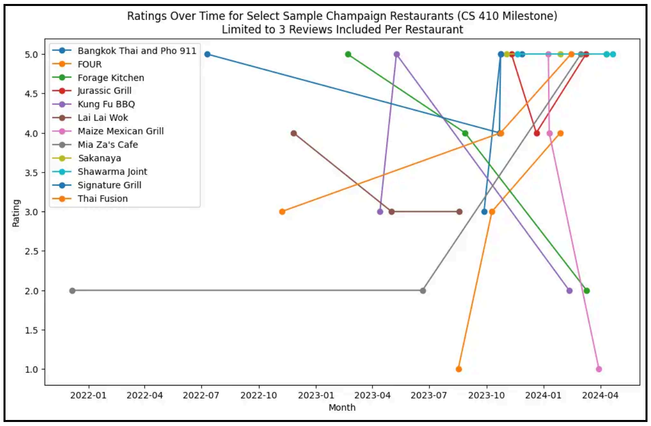


Figure 4

Next, we used the sentiment values from the model and plotted them over a time period of 2 years for each restaurant, as shown in Figure 5. There are some notable trends we can infer from this graph. First, we can see a dip in sentiments towards Mia Za's during September 2023, which was right around the time when they came under fire for hygiene issues. We can also see that Signature Grill has been on the rise which could be attributed to their release of new specials this year. Finally, we can see Forage Kitchen has had an overall decline in sentiment which could be attributed to the general demand for a vegan-friendly restaurant on Green Street possibly declining over time.

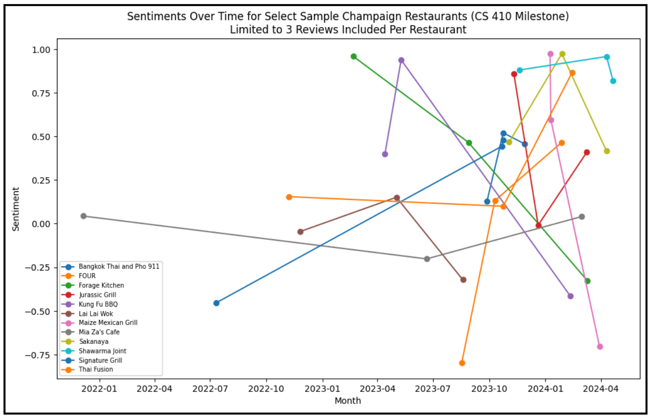


Figure 5

## REFERENCES

[1] Alalawi, H. 2021. Yelp_light. Kaggle. https://www.kaggle.com/datasets/hhalalwi/yelp-light.

[2] Maite Taboada. 2015. https://doi.org/10.1146/annurev-linguistics-011415-040518

[3] Siqi Liu. 2020. Sentiment Analysis of Yelp Reviews: A comparison of techniques and models. https://arxiv.org/abs/2004.13851

[4] Wisam A. Qader, Bilal I. Ahmed, and Musa M. Ameen. 2019. https://ieeexplore.ieee.org/abstract/document/8950616/

[5] Yelp Inc. (2023) *Yelp API GraphQL Basic Usage*, *Yelp Developer Portal*. Available at: https://docs.developer.yelp.com/docs/graphql-basic-usage (Accessed: 29 April 2024).