Felix Zheng
Kieran McHugh
CPT_S 475

Team 6: Detecting Dine-out Patterns

Abstract

Detecting Dine-Out Patterns is about understanding consumer dine-out habits by performing feature extraction on time-stamped geospatial data to identify “dine-out sessions.” The project integrates two datasets: (1) individual GPS location traces and (2) a Washington State restaurant/outlet database. By classifying each individual’s geospatial points against 50-meter spatial buffers around restaurants and then applying temporal constraints, we constructed a reliable method for detecting true restaurant visits. As a secondary goal, we also wanted to categorize visits into breakfast, lunch, or dinner windows and compare visit frequencies across different times, as well as differentiate between a standard dine-in visit and a fast food or takeout visit. We gained various insights into these categories, one of which is that dinner is the most common meal period for restaurant visits, as shown in the “Meal Period Visits by Day of Week” graph, and is especially common in Tuesday–Thursday dining frequencies, as well as what proportion of overall restaurant visits in the dataset are for takeout.

Introduction

Detecting dine-out behavior from raw GPS data is challenging because individual geospatial data must be transformed into meaningful events such as restaurant visits. This problem is important for understanding consumer behavior, urban food access, and population-level activity patterns. Our approach combines individual time-stamped geospatial data with a geospatial restaurant dataset, using spatial buffers and temporal dwell-time constraints to infer dine-in episodes. We further classify visits into meal periods to study temporal dining patterns at scale. Compared to existing related classification studies, our method emphasizes a “correct” dine-in classifications to improve visit reliability in Washington State. From our dine-in classifications, we found dine-in visits to be clustered in high population areas, and that fast food visits comprised about 1/8 of the visits overall in the dataset. From our temporal-based categories of dine-ins, one strongly supported insight was that dinner was the most common type to dine-in.

Problem Definition

As stated above, our primary goal is to use the information provided in both datasets to determine which individuals visited a particular restaurant. However, further insights can be extrapolated from this data beyond just whether a restaurant was visited or not. Using the additional details provided about restaurants as well as the individuals' location data we can begin to discover behavioral patterns in our dataset, like when certain individuals prefer to eat their meals. Understanding the trends and habits involved with which people visit a restaurant and when could be very beneficial for a restaurant owner as it would help them to identify potential commonalities their visitors share, which would in turn both help an owner to better cater to their main clientele while also discovering areas where they could potentially expand their reach should they desire to do so.

Models/Algorithms/Measures

Both of the datasets included in this project provided location data, but it was formatted in such a way that it would be difficult to use this raw data alone to determine an individual's proximity to a restaurant. As such, the first step was to perform a transformation on the location data, converting them to sf (Spatial Feature) points using the sf library in R. A unified representation of location data would provide the foundation for our detection algorithm going forward.

Now having the location data in a more manageable form, our devised method for determining a visit begins by determining the geographic locations of each restaurant and then programmatically drawing a circular buffer zone with a radius of 50 meters around each one. We can then use the location data of the individuals in our dataset to check if they happened to be within this radius at any point.

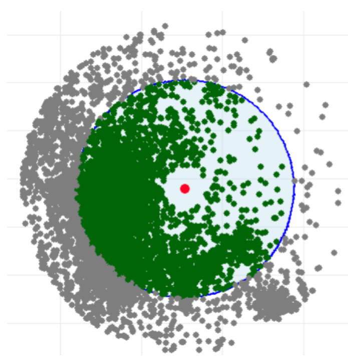

Figure 1: Checking for points within a restaurant's buffer zone

The above method was somewhat simple, and after further exploratory data analysis it was found that there were some inaccuracies in the individuals' dataset, primarily due to some points being missing for extensive periods of time. As such, it was necessary to further refine what should constitute a visit. This was addressed by applying a minimum number of points in the buffer zone to be flagged as a visit. Implementing this requirement resulted in a significant reduction in reported potential visits, but lent much more credibility to the visits that were found.

However, simply checking if an individual entered the buffer zone is also not a particularly accurate representation of a visit as it is highly prone to over-report potential visits. For this reason it was necessary to develop further constraints to help narrow the definition of a visit. To achieve this, the timestamps associated with the individuals' location data were used, and a minimum time requirement was applied to any points that fall within the zone. Individuals would need to spend a certain amount of time within the buffer zone to qualify as having visited the restaurant. Additionally, a maximum threshold was also applied to deal with potential edge cases such as someone living near a restaurant or possibly working in one. Our finalized classification method is done using feature engineered dwell-time blocks (≥12 minutes, ≤120 minutes, ≥3 points inside a restaurant buffer) and timestamp decomposition, we identified dine-in episodes and removed noise such as drive-by points or prolonged stays from living/working inside a buffer zone. Then in the case of overlapping buffer zones, we only created a visit for the closest restaurant, which makes the assumption that the data points all have the same closest restaurant. This provided a basic foundation for determining whether a visit occurred.

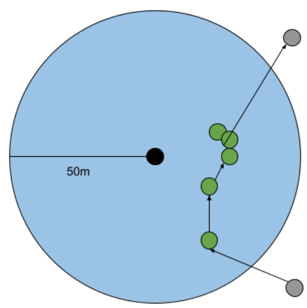

Figure 2: What a "visit" looks like after refinement

The combination of these constraints produced a much smaller but also much more accurate list of potential restaurant visits by the individuals in the dataset. Having a list of visits that we believed could stand up to a fair degree of scrutiny, we felt it appropriate to begin moving on to the secondary objectives we had set out for ourselves at the start of the project.

The first of these objectives was determining the windows for when a visit occurred (or put in simpler terms, what meal period). As mentioned earlier, our dataset provides an associated timestamp with each individual's location points. Furthermore, by this point we were already checking these timestamps when determining whether a visit occurred at all, so determining what meal period the visit occurred during simply consisted of checking the start time of the visit and comparing that against our predetermined time ranges for each period. We then appended a Meal Period attribute to each visit, which would be labeled "breakfast," "lunch," "dinner," or "late," depending on the time at which the visit occurred. In doing this, we were able to provide a more comprehensive picture of what each particular visit looked like and could then apply these trends to extrapolate common behaviors from the dataset.

However, when evaluating our data we began noticing that certain restaurants appeared to be underperforming in terms of visits, especially locations that would intuitively be popular locations (like Starbucks or McDonald's). It occurred to us that while our current method for determining a visit was capable of determining whether someone simply spent time nearby a restaurant without entering, this had the unintended consequence of under-reporting takeout and fast food visits, where an individual would likely spend a much shorter amount of time within the restaurant's buffer zone.

Fortunately for us, our restaurant dataset had a "fastfood" boolean value associated with each location, making it easy to determine which restaurants in particular would likely suffer from our current implementation's oversights. We then created and ran a modified version of our previous algorithm on these locations in particular with a more relaxed minimum time requirement of only a few minutes (though still with the 3 minimum point constraint from before) and saw results that were more in line with the rest of the restaurants. We then classified the findings of this algorithm as "fast food visits" and included those in the comprehensive list of visits as well. By maintaining the instances of fast food visits separately we could then also perform analysis on those particular visit types as well, such as at what time in the day they were most likely to occur and the average duration of a fast food restaurant visit as opposed to that of a dine-in visit.

Both of these secondary findings allowed us to draw more conclusions about the individuals in our dataset and thus gave us greater insights into the common trends and behaviors of people overall.

Implementation/Analysis

As mentioned previously, we are analyzing two datasets simultaneously for this project: one with restaurant data and another with the location data of about 20 individuals along with associated timestamps for each location point. Our goal for the project is relatively straightforward: given these two datasets, did the individuals visit a restaurant at any point during the recording process? Given the duration over which location points were recorded and the number of individuals that participated in the collection process, we hypothesized that we could identify numerous potential visits and at the very least a few dozen visits with a great degree of confidence.

The very first step was to normalize the data between the two, especially in regards to the locations within both datasets, as this component is essential for detecting a visit. This was accomplished by converting the locations of both the restaurants and the individuals to SF (Simple Feature) points. This provided a unified way to represent both spatially, which provided the foundation for our further implementation into visit detection.

Unfortunately there is not an entirely foolproof method available to us to evaluate if all of our flagged potential visits are indeed correct; the method with the lowest possible chance of error would be to simply have the individuals themselves verify whether or not they visited a particular restaurant, but this is not only unreliable given the time that has passed since this data was collected, but the data itself is also anonymized, so this would be impossible.

However, there are still some avenues we can take to validate whether we are getting the right results. The most straightforward of them is to simply compare the results of our most refined version to those of our previous iterations. As listed in the algorithms section above, our first iteration was not very “picky” when it came to classifying a visit in the dataset. As such, when running a later version of the visit detection algorithm with heightened requirements for a visit, we would intuitively expect the overall total flagged visits to be lower than the previous algorithm’s output. This was a quick (if somewhat inelegant) way to verify whether our data was generating the expected output.

We also compared the visit counts of individual restaurants to one another as a way to ensure our findings were in line with what we would expect to see. While a certain degree of variance was of course expected, we generally expected visits to be fairly normalized across the restaurants, with popular locations making up the majority of visits and smaller venues maintaining a relatively consistent visit count with one another. However, as discussed in the following Results and Discussion section, this was not the case with our original implementation, requiring us to take additional considerations into account with our approach.

Results and Discussion

In terms of strengths, our approach is interpretable because it is based on human inferences and relies on clearly defined spatial and temporal rules (buffer distance, dwell time, and minimum point counts) that make detected visits easy to explain, justify, and adjust. By combining GPS traces with a curated restaurant outlet dataset and filtering out noise such as drive-bys and brief stops, our classification of a visit is robust and accounts for likely noise factors. We can leverage our domain knowledge of human habits when going out to correctly identify dine-out sessions and to avoid incorrect classifications. The pipeline scales well to large datasets using efficient spatial joins and grouped aggregations, enabling population-level analysis. Additionally, defining visits in this structured way supports additional feature extractions based on visits, including meal-period patterns, fast-food versus non–fast-food comparisons, and spatial clustering of dining visits.

In terms of weaknesses, our approach is dependent on fixed thresholds (e.g., buffer size and dwell-time limits), which may not generalize equally across different environments or sampling rates. In dense urban areas with closely spaced restaurants, attributing visits to a single establishment remains ambiguous despite overlap resolution because the geospatial data is not perfect and users can be moving around while inside as well. The approach is also sensitive to GPS noise and data sparsity, which can fragment true visits or exclude them entirely. Finally, without labeled data, the method cannot directly validate visit accuracy and misses and it limits us to just unsupervised learning.

Additionally, there were also some oversights with the original implementation that we had not at first considered. As mentioned in the algorithms section, we did begin to notice some discrepancies in our findings, particularly that certain locations seemed to be under-reporting their visits despite being a popular location that we would instead expect to see many more visits in. We found that this error was the product of an oversight in the original implementation, particularly that fast food and takeout restaurants would be underrepresented in total visits due to the minimum time constraints we had applied when determining a visit. While this time constraint provided a good way to separate what would likely otherwise be individuals simply commuting by from actual visits, it had the unintended consequence of filtering out visits to fast food and take out establishments. This was resolved by applying a shorter minimum dwell time to locations that fell into the fast food category, and after doing so we were able to achieve results that were more in line with our original hypothesis and more reflective of the actual behaviors of the individuals in the dataset.

Overall, by the conclusion of our work on the project we were able to produce a list of potential visits that we were confident in, supporting our basic hypothesis that there were visits to be found in the dataset. Furthermore, we were able to gather insights into our secondary objectives as well, such as finding which meal periods were most popular and whether this fluctuates with the day of the week, as well as determining which proportion of the visits in our dataset took place at fast food locations. Some diagrams of our findings are included below:

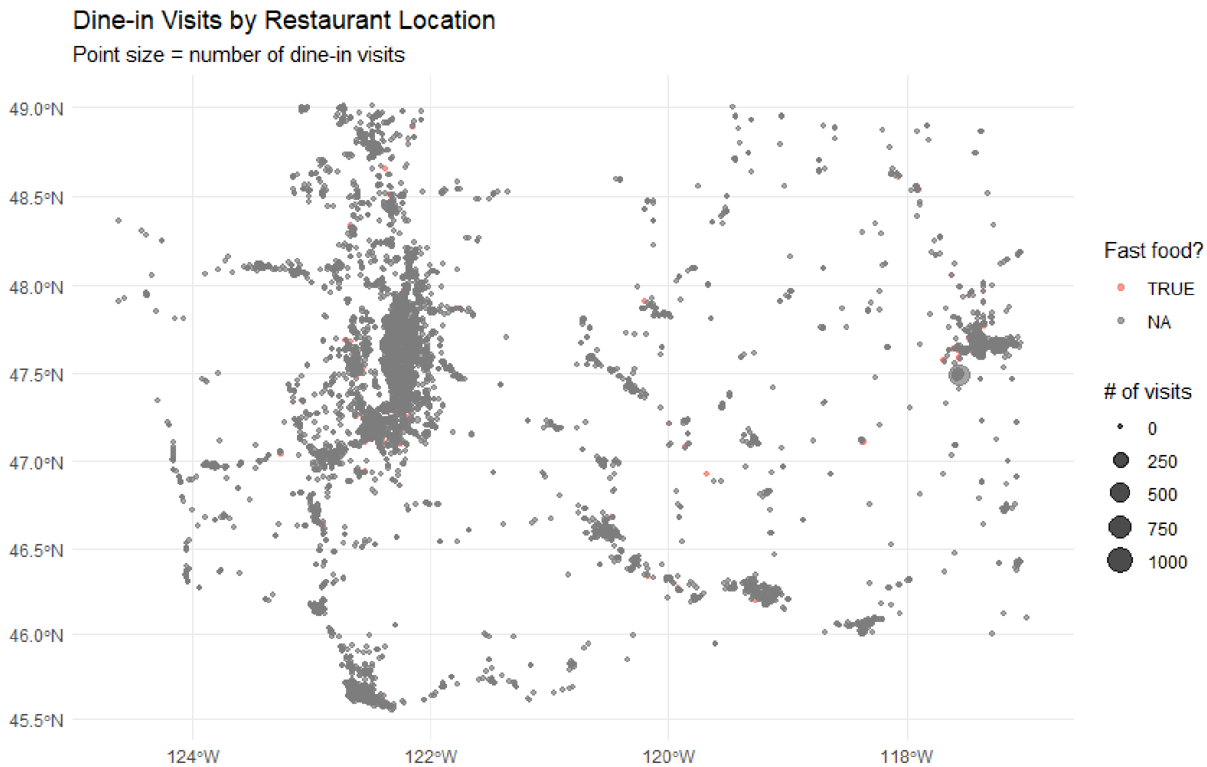

Diagram 1: Dine-in visits by Restaurant Location

A spatial representation of our primary objective, with dine-in visits displayed across Washington State. Each dot depicts a restaurant, and the dot's size corresponds to the number of total visits that were determined to have occurred at that restaurant. As one would intuitively expect, the dot sizes increase proportionally with the overall population at their respective locations as there are more people visiting restaurants at those locations.

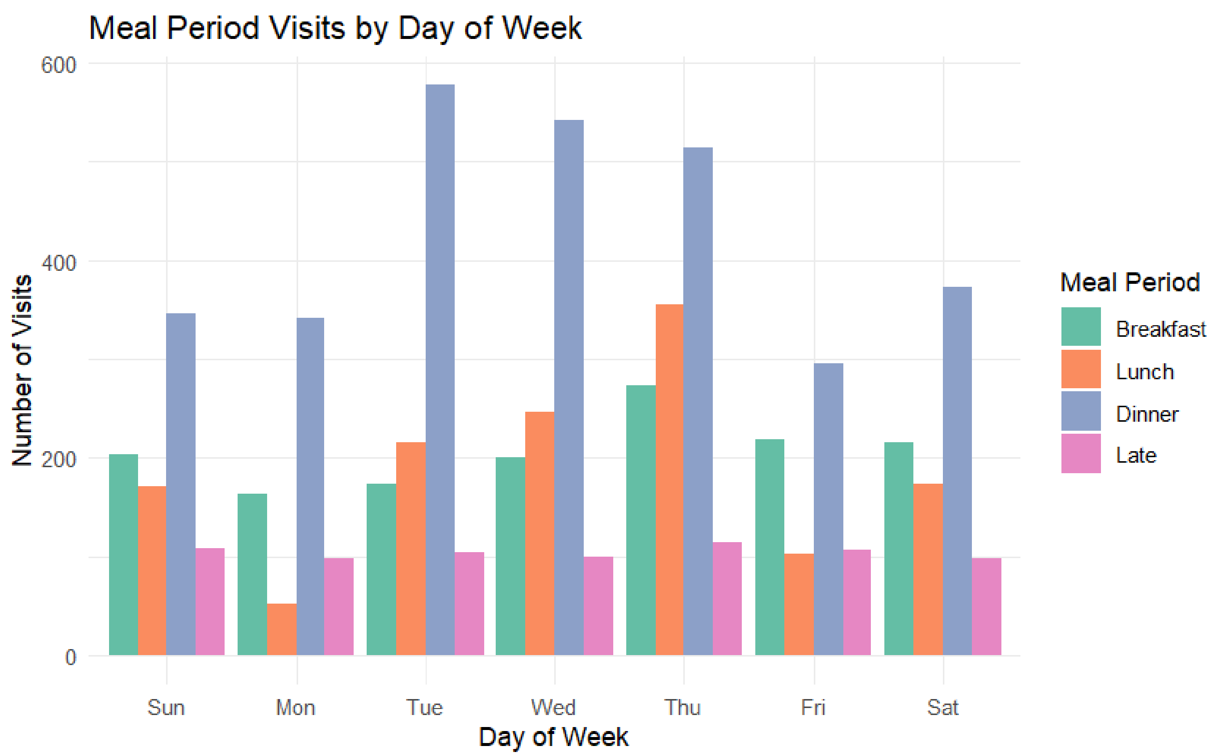

Diagram 2: Meal Period Visits by Day of Week

As determining which meal periods were most popular throughout the days of the week was one of our secondary objectives, we created visualization to depict these findings. As can be seen in the bar graph above, dinner is the most popular meal to eat out, and the most common days of the week to do so are from Tuesday to Thursday. Lunch is the second most popular meal, with the majority of visits also occurring within the range of Tuesday to Thursday. This can likely be attributed to people working these days of the week and not having as much time to prepare and eat something at home.

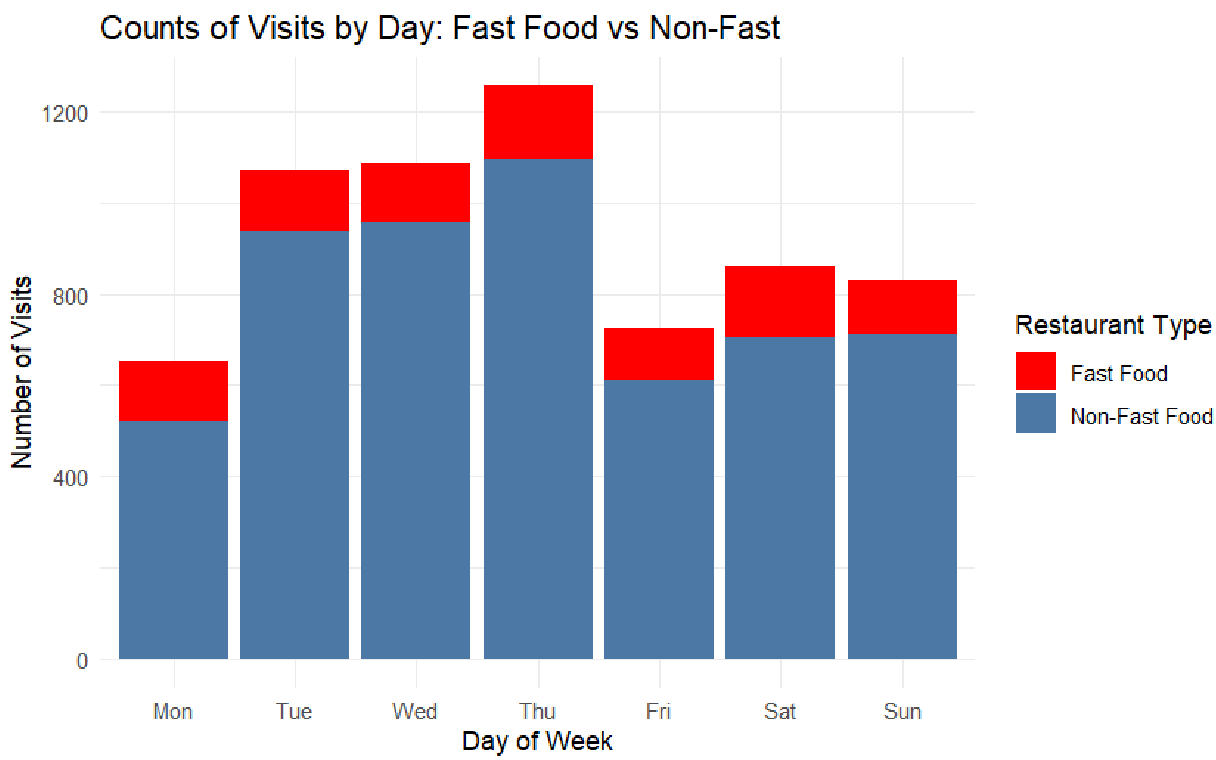

Diagram 3: Counts of Visits by Day of Week: Fast food vs Non-Fast Food

Another one of our secondary objectives was to be able to classify a fast food/takeout visit and then compare those visits to normal overall dine-in visits. Here the comparison of the two is visualized, with fast food visits appearing in red and non-fast food visits appearing in blue. A notable feature of this dataset is that whether a fast food visit occurs is highly dependent on the habits and behaviors of an individual. Certain people are more predisposed to visiting these fast food restaurants and thus make up a large portion of these visits on their own, while others may have few (if any) fast food visits associated with them.

Related Work

Huang, Yue, and Li (2010) address a closely related problem but in a broader sense. From the abstract, they created an "algorithm to figure out how the intersections of trajectories and spatial-temporal attractiveness prisms indicate the potential possibilities for activities" (Huang et al. 2010). In other words, identifying human activities from spatial and temporal "attractiveness" of points of interest (POIs). Their approach models how likely a person is to perform a certain activity at a POI based on temporal patterns (e.g., typical times for work, leisure) and spatial context, and then infers activity types from raw movement data.

In contrast, our work focuses specifically in Washington State on dine-out behavior at restaurants, rather than general activity recognition, and we pair individual GPS traces with a restaurant outlet dataset to directly detect restaurant visits. Instead of learning activity types from POI attractiveness, we use geometric buffers and dwell-time constraints (distance, time in buffer, number of points) to define a “visit,” and then classify each visit into meal periods (breakfast, lunch, dinner) and fast-food vs. non–fast-food categories.

So while Huang et al. provides a method for inferring general activities from timestamped geospatial data, our approach narrows the scope to Washington state, restaurant dine-ins, adds domain-specific constraints to improve visit precision, and emphasizes population-level dining patterns (meal-period frequencies, fast-food share) rather than a broad category of activities.

Conclusion

In this project, we developed an interpretable and robust method for detecting dine-out behavior by integrating timestamped GPS traces with a geospatial restaurant dataset and applying constraints determined by our domain knowledge of dine-in habits as well as likely noise that would be misclassified as dine-ins. Our refined approach successfully identified dine-in patterns across Washington State, revealing that dinner is the most common meal period and that visits cluster around major population centers. We also addressed an initial under-reporting of fast-food and takeout visits by relaxing dwell-time requirements for these establishments, resulting in more accurate and representative visit counts. While our findings support our hypotheses and demonstrate the effectiveness of rule-based inference for unlabeled mobility data, limitations remain due to fixed thresholds, GPS noise, and ambiguity in dense restaurant regions. Future work could incorporate further feature extractions based on the visit feature, improve validation through labeled datasets, and expand the framework to study broader categories of human activity and POI interactions.

Bibliography

Huang, Lian & Li, Qingquan & Yue, Yang. (2010). Activity identification from GPS trajectories using spatial temporal POIs' attractiveness. 27-30. 10.1145/1867699.1867704.

Amram, Ofer. “Web Mapping and Data Visualization.” 10 Sept. 2025, Pullman, WA, WSU.

Repository Link

https://github.com/zhengfel26/CPTS_475_dine_out_patterns