
STARCH MADNESS: AN OBJECTIVE ANALYSIS OF THE SUBJECTIVE QUALITY OF FRENCH FRIES IN DOWNTOWN SANTA MONICA

Adam “Fry” Brownell
Retina.ai
adam@retina.ai

May 1, 2019

ABSTRACT

It all started with a team lunch. We always ordered fries for the table. After several lunches, we began to ask the question: What are the best French fries near our Santa Monica office? How do you even determine “best”? As a Data Science company, we knew this couldn’t be determined simply by eating fries at various restaurants—we had to structure our analysis around scientific rigor and specified constraints, so that we could be confident in our findings. . . . If the results weren’t reproducible through peer review, what was the point? And thus, Starch Madness was born; **a contest wherein a bunch of data nerds objectively decided which restaurant served the subjectively best French fries in Santa Monica.**

1 Introduction

In 1802, Thomas Jefferson served “potatoes served in the French manner” at a White House dinner. From there, the French fry was born. Potatoes cut into slices and deep fried became a staple in American cuisine.

Despite being most associated with burgers and fast food, French fries are served in restaurants of all shapes and sizes. Fries are just as likely to be seen on the menu of an English pub as a high-end Asian Fusion restaurant; albeit at vastly different price points. With their ubiquity, the question must be asked: “Who serves the best French fry?”

The simplicity of the food recipe makes this question especially interesting, as there are only a few variables a chef can play with to craft their ideal French fry. In addition, some institutions produce mass quantities of fries per day—does this mass production reduce the quality of the fry? Does the fact that they must focus a large portion of their business on producing enjoyable fries create a better quality than a restaurant known for higher end cuisine?

We would like to take this moment to address the research elephant in the room—if this research is even important in the first place. We should propose that in

fact, yes. The pursuit of knowledge is admirable regardless of scale, and there has yet to be any studies conducted on the French fry quality of the downtown Santa Monica area. We may not have millions in French fry funding (yet) at our disposal, but we have the tenacity and drive (and palettes) to see this project through. It is our pleasure to add a tiny morsel of new knowledge into the human corpus.

2 Method

2.1 Location

Firstly, we had to choose the fries that would be entered into the contest. The simplest sampling method was to select fries that could be purchased at a location that was less than 10min walk away from our office. Luckily, our office was located in the heart of downtown Santa Monica and there was no shortage of fry options; in fact there were far too many. To narrow down the locations, we simply typed “best French fries” into Yelp and chose the top 12 results. Several of the locations were eventually replaced due to fry style (see section 2.2) but this methodology yielded a strong crop of contenders.













Restaurant	False Identity	Media Reference	Restaurant	False Identity	Media Reference
Burger Lounge		It's Always Sunny in Philadelphia	Umami Burger		Pulp Fiction
Barney's Beanery		Star Wars	The Misfit		Arrested Development
HiHo Burger		SpongeBob SquarePants	Little Ruby		Breaking Bad
McDonalds		Parks & Recreation	The Great American Fries		Harry Potter
The Independence		House of Cards	Johnny Rockets		American Psycho
Plan Check		Bob's Burgers	Steak & Shake		Toy Story

Table 1: Identities given to each contestant to remove location bias

To avoid bias from judges based on where the French fries were from, each location was randomly assigned a false identity based on popular T.V shows. Table 1 is the key mapping the real restaurants to their false identity. Henceforth all restaurants will be referred to by their false identity names.

2.2 Style & Procurement

There are many types of French fries, and unfortunately comparing different styles of French fries is like comparing apples and oranges, and thus we had to standardize our French fry types. We settled on the most common style—salted potato fries. This excluded popular variants such as waffle fries and sweet potato fries, but still allowed creativity and diversity amongst the participants.

Judging one batch of fries at a time was not a feasible option, as that would have required quadruple the amount of time the current experiment was allotted as well as made tournament play (2.3) impossible. This increase in time would have inevitably resulted in judging fatigue. We instead attempted to gather as many fries as required per round at one time without degrading the integrity of each fry batch. Despite our earnest efforts to procure the fry batches as close to simultaneously as possible, our retrieval was not without flaws and some may argue the distances a fry batch travelled correlates inversely with the batch's score. This conjecture is exhaustively reviewed in our discussion section 4.3.

Individuals in charge of picking up the French fries were instructed to order the fries by saying, “Can I please have an order of fries?” Whichever style the

American Sitcom Restaurant Division	Animated Restaurant Division	Reckless Greed Restaurant Division	Classic Movie Restaurant Division
  	  	  	  

Table 2: Group Play Divisions. “Classic” movie distinction is subjective (but correct)

restaurant produced from this order was considered their typical fry style.

2.3 Experimental Design

Similar to the FIFA World Cup tournament design, Starch Madness was structured with two sections: group play and tournament play. Adequate performance in the group play stage qualified you for tournament play.

Group Play Stage. Each of the 12 participating French fries were randomly assigned into 4 groups of 3 contestants, called divisions, to be rated during the group play stage. The scoring used was the FAT System, defined in section 2.4. Judges were instructed to use the FAT system instead of comparing fries from within a division to reach a score. This is an important distinction as producing an objective score for each fry is a stronger indicator of fry strength than a comparative one. In addition, if a particular division has three excellent (or terrible) batches, they should be admitted into tournament play on their strength alone, rather than their division’s overall strength. Table 2 outlines the divisions.

Tournament Play Stage. While not the most scientifically rigorous way to compare French fries, it certainly is entertaining. During this stage the top 8 highest rated French fry batches were seeded and placed in a single elimination bracket. Each head-to-head match-up was a blind vote of which fry the judge preferred. Votes

were cast through pen and paper before the vote was folded and dropped in a secure receptacle (cup) after being properly secured. While table talk was acceptable, no judge was allowed to discuss or hint at their voting decisions prior or after voting. There were a total of 7 certified FAT judges participating in the tournament stage, ensuring that there is never a tie in the voting. Through the tournament play, a winner is crowned.

2.4 The FAT System

One of the first obstacles set forth when judging French fries is how to standardize scoring. And to do this, one must answer a deep philosophical question of, “what makes a good French fry?” Our team of fry-experts narrowed the quality of a French fry down into 3 categories: flavor, appearance, and texture; brilliantly named the “FAT System.” Each of these represents a core attribute of a good French fry. By rating batches based on these metrics, we can decide with confidence which fry is best. Each of these metrics is scored on a 5-point scale, with 1 being the lowest and 5 being the highest. Averaging these three numbers produces a fry’s overall quality and helps the judges make important voting decisions during tournament play.

2.5 Judges

The French fries were judged by the capable Retina.ai team, whose love of fries is only surpassed by their love of data collection and analysis.

Restaurant	Total Score	Flavor	Appearance	Texture
Little Ruby	4.44	4.66	4.50	4.16
Burger Lounge	3.92	3.92	3.92	3.92
The Great American Fries	3.67	3.86	3.29	3.86
Barney's Beanery	3.43	3.42	3.57	3.29
The Independence	3.27	2.83	3.83	3.16
HiHo Burger	3.00	3.66	3.00	2.33
Plan Check	2.72	3.16	2.83	2.16
McDonalds	2.43	2.43	2.43	2.43
Johnny Rockets	2.33	2.16	2.83	2.00
Umami Burger	2.00	1.86	2.42	1.71
Steak & Shake	2.00	1.33	2.66	2.00
The Misfit	1.21	1.21	1.21	1.21

Table 3: Group Play Scoring Results. Those highlighted in green qualified for Playoffs

3 Results

3.1 Group Play Results

After the group play stage, and each contestant had been graded on the FAT System, we produce Table 3 cataloging the results of the Fry judgments.

Those highlighted in green received acceptably high enough FAT scores to qualify for the tournament round. Those not highlighted fell below the McDonald's line, named for the fact that not only was McDonald's the lowest scored fry that qualified, but also with an average score of 2.43/5 McDonald's can be considered an average, mediocre fry. It should also be noted that Little Ruby had an absolutely incredible score, scoring nearly a half point higher than any other contender and a near perfect score of 4.44/5. An additional interesting result was that Little Ruby and Burger Lounge scored higher than Great American Fries, a food truck that exclusively serves French fries.

All contestants above the McDonald's line were entered the seeded tournament play resulting in the tournament structure seen in Figure 1. False identities are used because this is the same bracket judges saw during the contest.

Little Ruby (*Los Pollos Hermanos*) clearly won the Group Play Stage, and entered the tournament a heavy favorite.

3.2 Tournament Stage Results

After the first round, *Los Pollos Hermanos* and *Mos Eisley Cantina* won as projected, with *Los Pollos Hermanos* sweeping *JJ's Diner* 7-0. However, in the bottom two matchups the both fry-offs were upsets. *Krusty Krab* won decisively with a 6-1, while *Bob's Burgers* squeaked out a victory 4-3 in a hotly debated heat.

Krusty Krab knocking out *The Leaky Cauldron*, and doing so effectively, is very surprising because as stated in the previous section, *The Leaky Cauldron* (Great American Fries) only sells fries—and yet they couldn't win a single game in the playoffs.

The Final Four went as expected, with #1 *Los Pollos Hermanos* beating #4 *Mos Eisley Cantina* and #6 *Krusty Krab* dominating #7 *Bob's Burger* 7-0. The Championship game was an instant ESPN classic. Seed #6 *Krusty Krab*, with only 1 vote against it the entire tournament, took on Seed #1 *Los Pollos Hermanos*, whose group play score was significantly higher than anyone else—and 1.44 points above their championship opponent. But in the end, after intense debate and judging, the underdog took the championship 4 – 3.

Hi Ho Burger (*Krusty Krab*) was crowned victor of Starch Madness!



Figure 1: Starch Madness Results

4 Discussion

Now that we have objective scores for each fry, we are given the opportunity to see the impact of certain attributes on French fry quality, derived from FAT score averages.

4.1 Does Price Imply Quality?

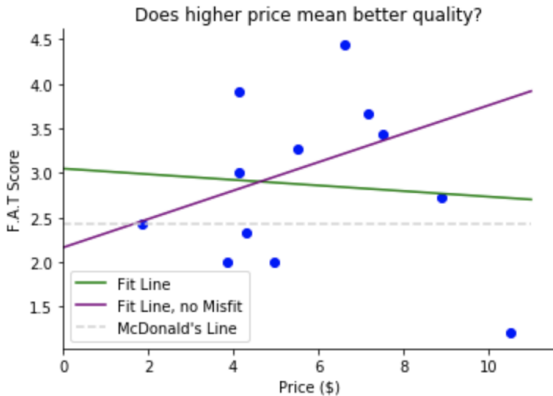


Figure 2: The Relationship between Price and Fry Quality

When looking at French fry price vs quality, there is one thing that sticks out quite clearly: the most expensive French fry in the contest, The Misfit, was also voted the worst. In fact, the Misfit is such an outlier that inclusion of it as a data point actually changes the relationship between fry price and quality from positive to negative. The correlation between price and quality without the Misfit is moderately positive (0.40 Pearson Correlation Coefficient) and the correlation with the Misfit is none to negative (-0.08 Pearson Correlation Coefficient). To sum this up nicely:

- Price does imply quality, but only up to \$10
- The Misfit (and perhaps any/all fries that are above \$10) are a rip off – You are better off going to McDonalds

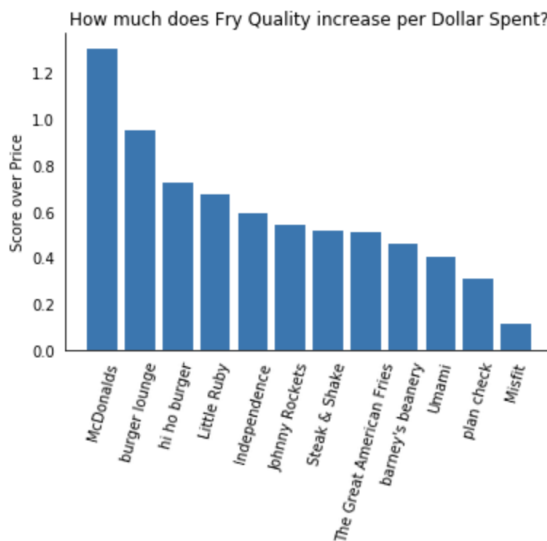


Figure 3: A Visualization of Price-for-Quality Ratio

Speaking of McDonalds—the cheapest option is actually the best option for the price you get. While absolute quality does increase as price increases, it does not increase to the point that you increase the quality per dollar spent. In fact, a dollar spent at McDonald’s produces 38% more fry quality than the next restaurant. If you’re looking for the best “bang for your buck,” McDonald’s is the best option.

4.2 Did External Factors influence our Voting?

Comedic False Identity. The false identities were added to the restaurants so that people wouldn’t be biased based on their prior experiences with that restaurant when voting. However, in our attempt to mask their identities, we may have introduced new bias through our associations with the cinematic universes that the fictional restaurants are from. Specifically, what if a restaurant was given a comedic identity, and because of our positive association with that movie/show, we graded it higher? Or conversely, what if a restaurant was given a very serious movie/show identity, and because of this, the judges scrutinized the fry more closely? Would people treat associations to Silence of the Lambs more negatively than Dumb and Dumber?

The answer is no. Splitting fries by comedy, we get a *t*-test *p*-value of 0.52. The comedic value of the false identity did not influence our voting.

Weather. We conducted these experiments during February and March—the heart of winter in Los Angeles. It reached a chilling 58 degrees during some of our rounds—did this influence our votes? Did the cold weather make the warm fries taste even better, and therefore result in higher FAT Scores? Or did the overcast weather make us somber, and therefore more critical of the fries we ate?

The answer is no. Splitting fries by if the weather as above 60 degrees or not, we get a *t*-test *p*-value of 0.74. The weather outside did not influence our voting

Day of the Week. Many of the Starch Madness events occurred on different days of the week. Perhaps the judges are harsher at the start of the week, especially on Mondays, and progressively get nicer as the week progresses. Or maybe by the end of the week the judges are tired and their judging abilities are impaired. Maybe judging on Fridays, lovingly named FryDays, influenced judges’ opinions. Does day of the week change judging in any way?

The answer is no, a *p*-value of 0.55 rejects this hypothesis.

4.3 Does Distance indicate Quality?

A major discussion amongst judges during scoring was whether the distance from our office implied if the fries would be better or worse. The reasoning behind this is simply because fries are best when they are fresh and hot, and if we had to walk them back to the office for several minutes, they could lose more heat than a batch procured closer to our office. There are strong arguments for both sides. In support of this theory, Umami burger is one of the farthest and lowest rated fries, Little Ruby, Plan Check, and HiHo all made the playoffs and are very close to our office. However, Misfit, the closest restaurant, is the lowest score, and Great American Fries is far but still very high quality. Clearly, the only way to solve this is with statistics.

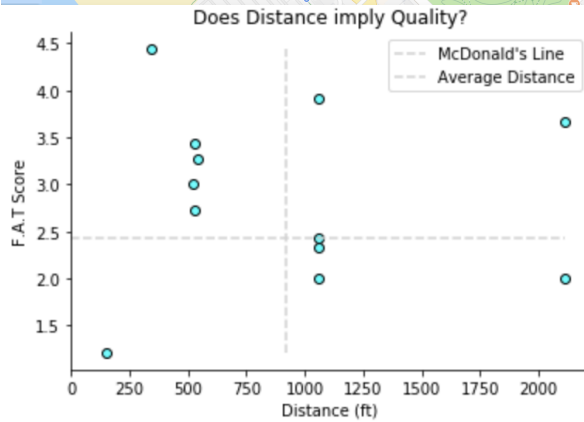
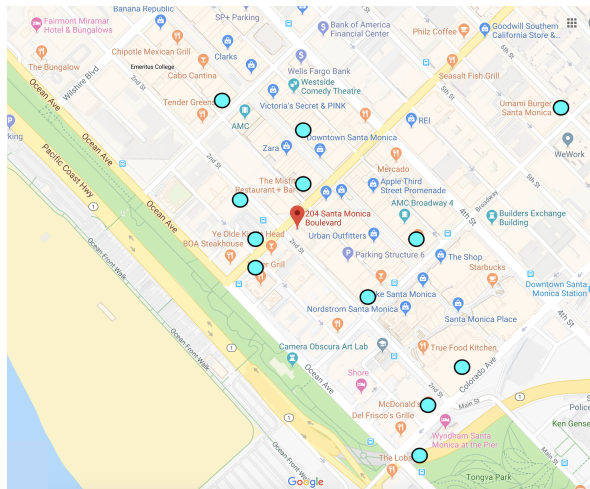


Figure 4: Location proximity and its Relationship to Quality. Quadrants named moving clockwise from top-left.

We charted where each restaurant was and found the Euclidean distance of each from our office, marked on the map. Firstly, even by removing The Misfit from the data, only 10% of variance can be explained through a linear or exponential relationship ($0.1 R^2$ value), so the hypothesis that distance degrades quality is rejected.

What the distance-quality chart does produce is an interesting bucketing strategy for these restaurants. By bisecting the data by both the McDonald's line and the average distance, we can create 4 quadrants/segments. Quadrant I can be labelled "Easy Bets," where each restaurant is close and of high quality. Quadrant II can be labelled "Worth the Journey" since they are farther away, but still high quality. Quadrant III can be labelled "Not Worth the Journey" for having below-McDonalds' quality at above average distance. Finally, Quadrant IV can be labelled "Deceptively Bad" because while they are close to the office, they are not even the short trip. The Quadrant IV name also stems from the fact that we frequent the Misfit often since it is so close, and we have grown to believe that their fries are quite good. Our studies show this is incorrect, and our frequent visits have simply biased us.

4.4 Can Yelp Ratings indicate F.A.T. Score?

Yelp is a service that prides itself on crowd-sourced ratings for restaurants. Many often check Yelp first before deciding if they should eat a particular location. Can we use the same strategy when selecting a location to dine on the best French fries? One would hope that as the overall Yelp score increases, so does the quality of the French fry.

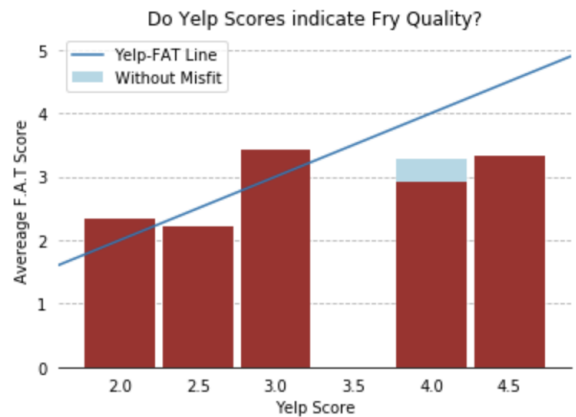


Figure 5: The Relationship between Yelp Score and Fry Quality

Unfortunately, this is not the case. As seen in Figure 5, highly rated Yelp eaters fail to produce on average fries that are as highly rated as their restaurant. Even when we cut out the Misfit outlier, good Yelp restaurants fall below the Yelp-FAT Line, which simply charts a 1:1 relationship between Yelp score and FAT score—a restaurant above the Yelp-FAT line (such as 3.0 Yelp score) produces fries of a higher caliber that the Yelp score would suggest. Our explanation for the

4.0+ restaurants failing to produce 4.0+ fries is that the Yelp score includes not only other foods, but also the ambiance of a restaurant. High Yelp score restaurants tend to have a better dining experience than lower scoring locations, despite their fries being of an equal or lesser quality.

4.5 Addressing Potential Confounding Variables

Playoff Rankings. Everyone loves an underdog story, and so it is possible that despite having blinded each restaurant during group and tournament play, knowing the rankings could have compelled voters to choose the underdog when a contest was close. Fry quality in tournament play was higher than in group play, and so this is entirely possible. How to control for this in the future is to avoid telling rankings, but also changing the false identities several times throughout the contest, thereby removing any bias judges have procured during previous tastings.

Judge Acumen. To address the concern that the judges themselves were bias and not indicative of scores of those not involved in the research, this is also possible. However, each judge has passed the French Fry Judge Certification Program (Figure 6) and therefore their expertise should be considered adequately accurate for French fry quality.



Figure 6: Logo of Certification Program all judges received

Sauce Accompaniment. Fries do not often stand on their own. They are regularly paired with another American delicacy: ketchup. It is quite likely that there are varying styles of ketchup served throughout the

French Fry landscape, and each of these permutations may have influenced our results. Moreover, perhaps certain restaurants took a deeper look into the French fry and ketchup relationship and specially crafted a sauce to fit their unique fry flavor palette. This immediately relevant to this study because there was an lively internal debate on whether fry quality should be judged with or without condiments. Some argued for fry purity, citing that condiments distracted from the overall quality of the fry itself. Others insisted that a strong fry could demonstrate quality with and without condiments— with ketchup being a core element of the fry experience. The research team eventually settled on the latter argument, and allowed each french fry batch to be paired with its given sauce. In addition, we instructed judges to consider the fry above the ketchup in terms of quality, as they were to judge the quality of the french fry with condiment rather than the success of the pairing itself. In future studies, and as french fry science progresses further, a standard fry-ketchup research method should be established by an international community of fry researchers; but until this day we would like to advocate for our own research methodologies as optimal.

5 Conclusion

Ranking French Fries is a relevant subject to all readers, and we hope to continue our research beyond the Santa Monica area. While our resources confined us to Santa Monica and single elimination, next steps could be an expansion of both region and trials. We are all interested if these results hold up across all restaurant locations, or if there are locations that claim to have the best French Fry that wish to enter the challenge. If any reader knows of research grants given in the area of French Fries, please reach out to our research team.

For those unconvinced that the research topic of Fry Quality of Santa Monica French Fries was worthy of publication, we would strongly argue that the pursuit of knowledge should always be encouraged, no matter how small the impact or specialized the application. Maybe in a few weeks you can read our upcoming paper on Hot Wings quality and be further convinced.

For those worried that we have written an entire research paper on french fries, do not worry. The entire experience was delicious.

We hope that all those who read it feel empowered to conduct their own French fry quality assessment as a form of peer review, and publish any interesting results found. The research area of French Fry Quality is young and ripe for more publications.