



15.072 FINAL PROJECT REPORT

Topic: How should Boston restaurants allocate content creation resources (e.g. time, effort, number of posts) across Instagram, Facebook and Tiktok?

Submitted by: Jason Jia, Stephanie Sha, Maria Besedovskaya, Pavena
Vongkhammi

1. Scope and Idea

Key Question: How should Boston restaurants allocate content creation resources (e.g. time, effort, number of posts) across Instagram, Facebook and Tiktok?

2. Data

Engagement metrics dataset: We manually collected data for 96 restaurants in Boston. 100% of the restaurants have an Instagram account, 96% have a Facebook account, and 49% have a TikTok account. We obtain social media metrics (number of likes, views, comments, followers) for Instagram, Facebook and Tiktok over 6 months (April 2023 - September 2023) and group the restaurants into 3-4 equal classes (quantiles) based on their metrics for each period.

The histogram of overall engagement metrics is shown below. Note that the distribution of Instagram and TikTok are relatively similar, while Facebook has much lower engagement on average.

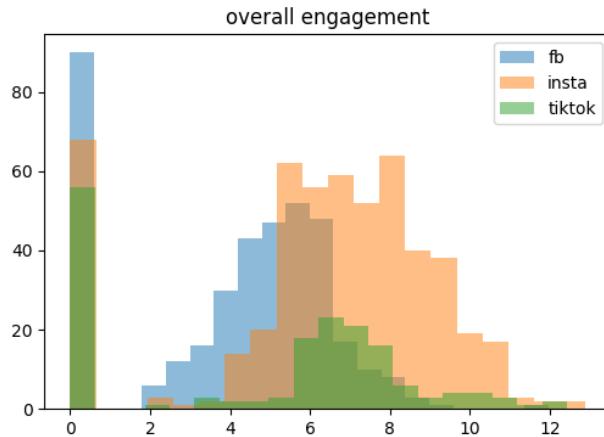


Figure 1: Overall Engagement

3. Methods and Models

Ensembles (CatBoost, XGBoost): to predict the next month's class for each restaurant, we developed a model to forecast each social media metric across Instagram, Facebook, and TikTok.

Multi-Armed Bandits (Epsilon-greedy): to get a recommendation of the share of content creation resources restaurants should put into each social media platform (see Appendix 1 for details). The model seeks to optimize overall engagement, defined as $\log(\text{number of views} + 5 * \text{number of likes} + 10 * \text{number of comments} + 1)$. We compare the best epsilon-greedy bandit out of a range of epsilons against a baseline where restaurants spend 60% of their time on Instagram, 30% on Facebook and 10% on Tiktok. We used this baseline because it is a good approximation of the relative amount of content appearing in each social media platform.

For the baseline we used Time Series based models: Random Walk (the metric for the next month is approximately equal to the metric in the last month), ARIMA (the future metric can be predicted as the linear combination of its past values) and VAR (the future metric is dependant on the past values of the same and other metrics). Because of the short time series for each variable and overall consistency of the metrics over time, the Random Walk model works the best. To obtain the AUC we used the q-cut to group the metric values and define whether the predictions fall in the same group as the real values. The ensemble models outperform the baseline in 17%.

4. Model Results

Key results

Restaurants should strongly consider posting more on TikTok even if it comes at the expense of posting less on Instagram and Facebook, because it can improve overall engagement by 30% on average. The full distribution of improvements from baseline to epsilon-greedy methods is shown in the figure below:

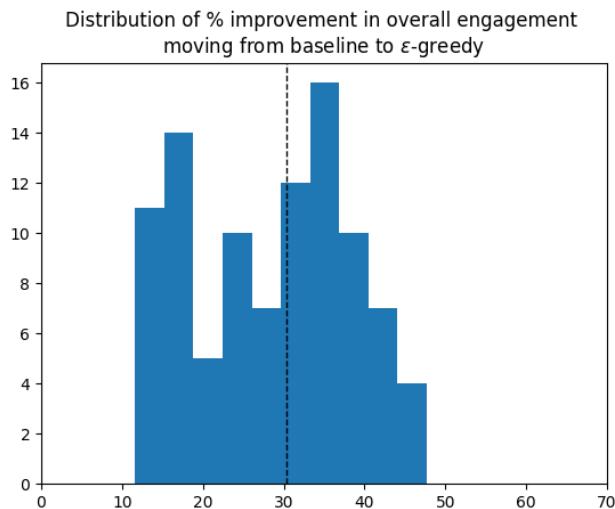


Figure 2: Distribution of improvement percentage in overall engagement

The epsilon-greedy bandit also recommends spending 152% more resources on Tiktok, 16% less resources on Instagram and 18% less resources on Facebook relative to baseline, on average. This gives an average optimal distribution of resources of 50.2% on Instagram, 24.6% on Facebook and 25.2% on Tiktok.

The full distribution is shown in the figure below:

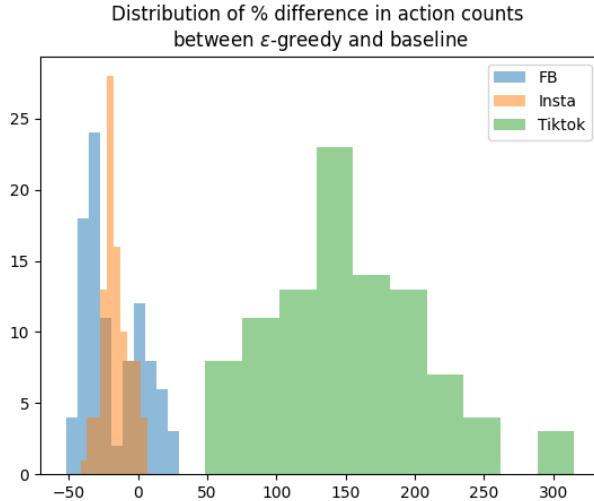


Figure 3: Distribution of difference percentage

Finally, we note that the optimal strategy is not to make content on the social media channel with the highest estimated reward all the time (ϵ = 0), but to do so with a high probability of $1 - \epsilon$ and make content on a random social media channel with a small probability of ϵ ($\epsilon > 0$). For example, the optimal ϵ for the restaurant below is 0.15, not 0. This is because rewards are unknown and change over time. Coupled with uncertain estimates, it is optimal to keep exploring and choose a different social media platform sometimes. This accurately reflects reality in which social media engagement metrics are highly volatile and unpredictable.

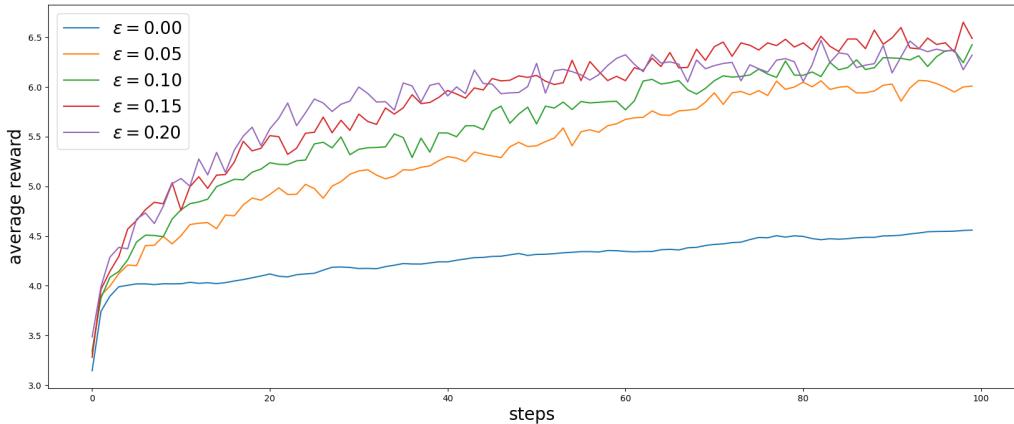


Figure 4: Time path of rewards for epsilon-greedy bandits with varying epsilons

Additional results

The CatBoost model performs well on our dataset: our results show 70-80% AUC for Instagram, 72-88% AUC for Facebook, and 60-72% AUC for TikTok.

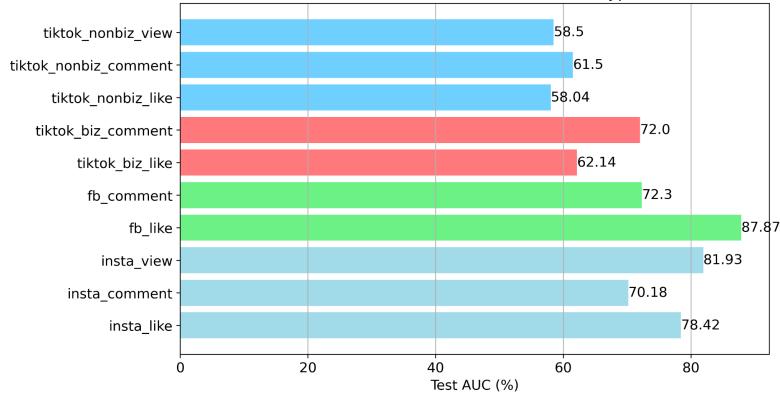


Figure 5: Model Performance

5. Web Interface

User-friendly web interface: We developed a Streamlit-based front-end that enables restaurants to assess their social media follower count in comparison to similar establishments in the Boston region. The platform also allows restaurants to upload their social media engagement statistics which are then fed into our models to predict the restaurant's social media engagement for the upcoming month, and generate a recommendation of how the restaurant should spend its time across the 3 different platforms.

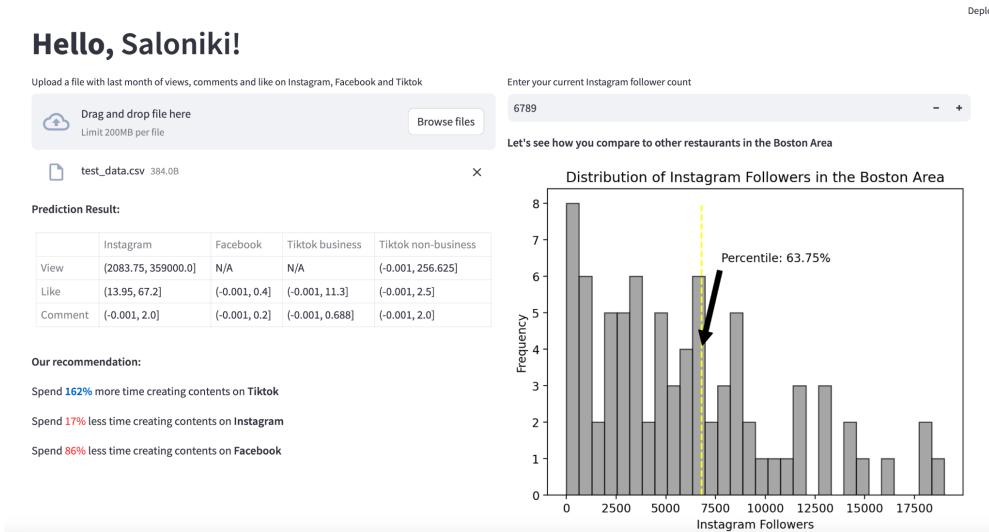


Figure 6: Web Interface

6. Impact

Potential future collaboration with Life Alive Organic Cafe

We pitched our findings to Kaitlyn Mailly, the social media coordinator of Life Alive Organic Cafe, a Boston-based restaurant specializing in healthy vegetarian options, and she expressed strong interest in our project as her team is also considering ramping up activity on Tiktok relative to Instagram and Facebook, which currently serve as their main social media channels (see Appendix 2). We are currently discussing how we can further collaborate with the marketing team on this problem.

Appendix

Appendix 1: Multi-Armed Bandits Model

We first get overall engagement quantiles for each of the social media platforms: 4 for Instagram, 4 for Facebook and 3 for Tiktok. We also get their standard deviations. We then read in the final predicted probabilities for the number of likes for the month t+1 (as a proxy for overall business engagement), and use them as initial probabilities to the model.

We use a 3-armed bandit problem where the arms represent Instagram, Facebook and TikTok respectively. Taking an action here can be interpreted as creating a post on the selected social media platform, or more broadly a unit of content creation effort.

We get an initial action value by first drawing a quantile using the initial probabilities of being in each quantile, then draw an overall engagement number randomly from the list of overall engagement numbers in that quantile. Since numbers with higher frequency appear multiple times in the list, we maintain the relative probabilities in the quantile.

Then, when action A_t is selected at time step t, a reward R_t is drawn uniformly from bounds $[Q_*(A_t) - SD(\text{quantile of } Q_*(A_t))/2, Q_*(A_t) + SD(\text{quantile of } Q_*(A_t))/2]$. If this is negative, we floor it at 0. This also becomes the true action value at the next time step.

The action values are estimated with a constant step size $\alpha = 0.1$ rather than using a sample average, because the action values are non-stationary and we want to put more emphasis on learning from more recent rewards: $Q_{t+1} = Q_t + \alpha(R_t - Q_t)$.

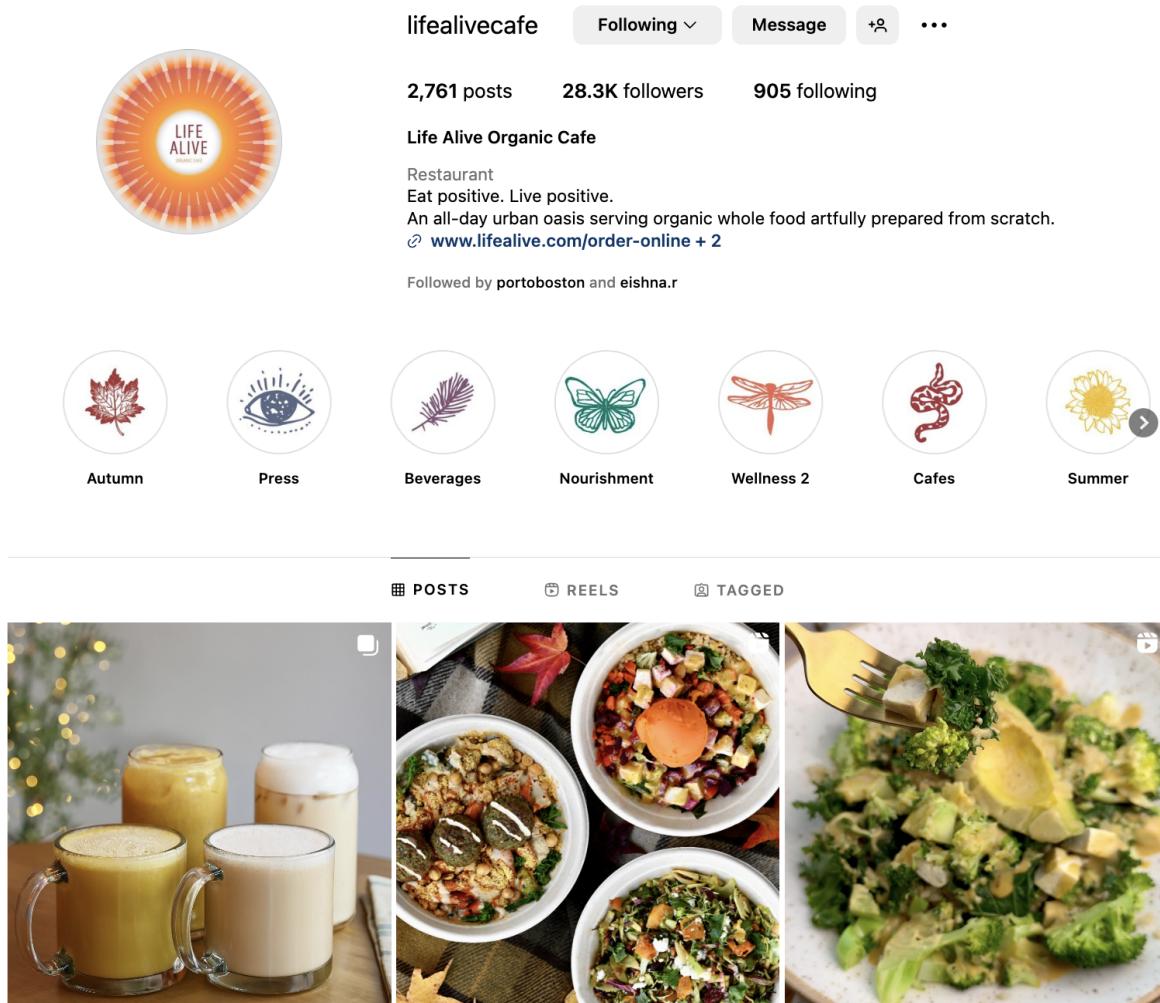
The action is selected using the epsilon-greedy method: select the action with the highest estimated action value (1-epsilon) of the time, and a random action epsilon of the time.

We use 100 time steps and keep track of the average reward of 500 runs of this model for each restaurant over time. We try out epsilons in the range [0, 0.05, 0.1, 0.15, 0.2].

Appendix 2: Social media profile pages for Life Alive Organic Cafe

We note that Life Alive has much more followers on Instagram and Facebook than Tiktok.

Instagram:



Facebook:

The screenshot shows the Facebook profile of "Life Alive Organic Cafe". The cover photo features a circular logo with the words "LIFE ALIVE ORGANIC CAFE" and a background of colorful, stacked items. The page name "Life Alive Organic Cafe" is displayed with 13K likes and 13K followers. Navigation tabs include Posts, About, Mentions, Reviews, Followers, Photos, and More. A sidebar on the left contains an "Intro" section with a brief description and links to the page, phone number, email, website, price range, and rating. Below this is a "Photos" section showing a grid of images related to the cafe's offerings. A main post on the right shows two glasses of warm, frothy beverages (likely lattes) with a Christmas tree in the background. The caption reads: "Welcome December! It's officially cozy season. Wrap yourself in the cozy embrace of our seasonal lattes. Experience the warming notes of our new Chai Latte, made with aromatic spices. Or indulge in the oh-so-smooth Butternut Spice Latte—it's a celebration of positive eating in every sip. Enjoy hot or iced." The post has 4 likes.

TikTok:

The screenshot shows the TikTok profile of "lifealivecafe". The profile picture is the same circular logo as the Facebook page. The bio reads: "We're here to renew your energy & connection to life through vibrant nourishment". Statistics show 55 Following, 390 Followers, and 755 Likes. The "Videos" tab is selected, showing a grid of five clips. The first video, titled "How the beloved Swami ...", shows a bowl of vegetables and has 3031 likes. The second, "Lover Bowl #lifealive ...", shows a bowl of grains and has 2682 likes. The third, "Coconut coffee ☕️ ...", shows a close-up of coffee being prepared and has 3821 likes. The fourth, "New Spooky Pumpkin Sp...", shows a smoothie with a spoon and has 796 likes. The fifth, "Starting October 3rd ...", shows a smoothie being poured and has 718 likes. A "Get app" button is visible at the bottom right.