Analyzing Data to Identify Factors that Affect the Collection of Free Food Items

Anonymous Author(s)

ABSTRACT

Food waste is a huge problem internationally, with economic, environmental, and humanitarian impact. Olio is a London-based app that is trying to tackle the food waste problem by allowing users to give food to other users on the platform for free. An analysis of Olio listings over a 17 month period shows that about 40% of the unique items listed were not collected. In this project, I aim to understand what factors affect an item's collectability by comparing various features of listings that were collected versus those that were not.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

hci, data analysis, food waste, sharing economy

ACM Reference Format:

Anonymous Author(s). 2019. Analyzing Data to Identify Factors that Affect the Collection of Free Food Items. In *COMPASS '19: ACM SIGCAS Conference on Computing and Sustainable Societies, July 03–05, 2018, Accra, Ghana.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Nearly one third of the food produced in the world is lost or wasted [4][7]. Recent research suggests that if food waste was a country, it would be the third highest contributor to greenhouse gases after China and the United States [3]. Additionally, a large percentage of people around the globe face food shortages, driving some people to look to food redistribution as an intervention [1][6].

Olio is a London-based app that aims to tackle this problem by allowing users to redistribute food to others. People with extra food can post it online then others can reach out to collect it. All food is offered for free, with the main goal of eliminating food waste. Although Olio has had success in the United Kingdom, and most items are arranged for pickup within 10 minutes of being posted, nearly 40% of unique items listed on the platform are not collected. It is likely that many of these items end up wasted. If we are to tackle food waste via redistribution networks, then we must understand why some items are not collected. This information can help Olio better tailor its platform so that more food is collected, and will also help drive additional interventions for food waste.

COMPASS '19, July 03–05, 2019, Accra, Ghana

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Table 1: Listings in London and Jersey

Location	Total	Collected	Not Collected
London	64,974	44,498 (68.49%)	20,476 (31.51%)
Jersey	28,842	17,831 (61.82%)	11,011 (38.18%)

2 METHODOLOGY

For this analysis, I used a dataset provided by Olio that included anonymized information for all of their listings posted between April 1, 2017 and September 9, 2018. I focused on Olio's two largest markets — London and Jersey, which are two different areas in the United Kingdom. The total number of listings and breakdown of those that were collected and not is shown in Table 1.

I chose to focus on separate markets for two reasons. First, most users would probably not travel a long distance to collect free food. Additionally, it is likely that different locations have different offerings and preferences for food, and it is important to understand those in order to provide suggestions to Olio and its users and to maximize the number of items that are collected.

For the first part of my analysis, I examined various aspects of the time at which a listing was posted. These included the day of the week, month of the year, and hour of the day. I compared the amount of items that were collected versus not collected on each of those factors. I also calculated the mean and standard deviation of the percentage of items that was collected to determine if any particular time seemed to be an issue for collection.

For the second part of my analysis, I looked further into the specifics of the items to try and understand why so many were not collected. I first examined the different food types based on work by Makov, et al. that divided the listings into 13 different categories [5]. I then performed text analysis on the title, description, and collection notes of each item, to determine whether the inclusion of certain words or phrases had an effect on an item's collection. These text fields are all written by the user who lists the item. The fields provide information about the ingredients, source of the item, and potential restrictions for collecting the item.

To perform the linguistic analysis I used the Natural Language Toolkit to clean the text, separate it into individual tokens (word-like units surrounded by white space), and get word frequencies [2]. I collected frequencies for the titles, descriptions, and collection notes across both locations and collection conditions. I also removed special characters, tokens that were only numeric, and tokens that were less than three characters long.

I calculated the percentage of collected or uncollected listings that contained the token to determine which tokens may be related to an item's collection state. I then repeated this for bigrams (sets of two tokens) and trigrams (sets of three tokens).

Table 2: Frequent Terms with Discrepancy in Usage across Collection Conditions

Location	Field	Term	Collected Rank	Not Collected Rank	Collected Usage	Not Collected Usage
London	description	'open'	50	**8**	4.31%	**7.65%**
London	description	'opened'	54	**11**	3.95%	**7.07%**
Jersey	title	'mill'	NR	**18**	0%	**2.34%**

3 RESULTS

3.1 Timespan Analysis

The day of the week a listing was posted did not seem to be a contributing factor for collection rates. In London, month also did not seem to have an effect. However, in Jersey both November and December had collection rates well below the first standard deviation, with just over 50% collected in each of those months (mean of 62.9% collection rate, standard deviation of 7.83). Data from additional years will be needed to see if this is in fact a statistically significant trend.

Hour of the day provided interesting results, especially as it compared across the two locations. In Jersey, items posted between 1 and 5am had less than a 50% collection rate, although there were only 66 items posted in that timeframe. In London, every hour had at least a 60% collection rate.

3.2 Categories

Examining the different food categories and their collection rates across both locations also provided interesting results. In London, every food category except frozen food and baby food had at least a 60% collection rate. Sandwiches and baked goods had two of the highest collection rates in London at 73.25% and 68.51% respectively. Conversely in Jersey, nearly half of the baked goods are not collected and sandwiches, baby food, dairy, and protein also all had less than a 60% collection rate. Figure 1 shows the collection rates for the ten categories with the most listings across the two different locations.

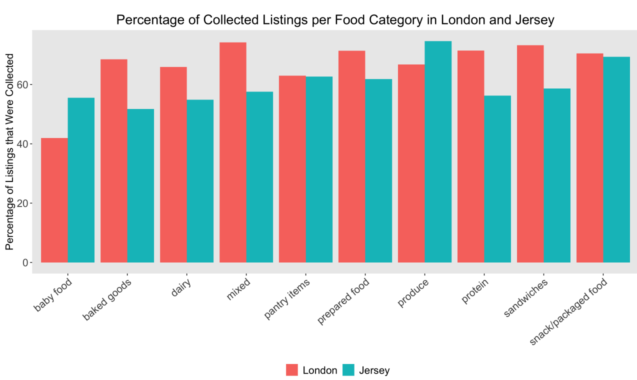

Figure 1: This graph shows the collection rates for the top ten categories in both London and Jersey.

3.3 Text Analysis

To determine which findings were meaningful, I examined the 30 most frequent terms for each of the three text fields, grouped by the collection status and location. I compared the rank (with the most frequent word being 1) and the percentage of collected or uncollected listings that contained that term. I then identified the set of most frequent terms whose rank differed by 30 or more spots and usage differed by more than 2% between collected and uncollected items. These findings are shown in Table 2.

4 DISCUSSION

The results suggest that the factors affecting a listing's collectibility are specific to the listing's location.

In London, the type of food seems to be one of the main factors in an item's collection. This is seen from the variation in collection rates across the different food categories. In addition, the words that seemed to have an impact on an item's collection were not related to the type of food, but rather it's state. Items including the words 'open' and 'opened' were not collected nearly 2 out of 3 times. This indicates that users are willing to collect some food items that have already been opened, but not others. Further analysis will show the types of opened items that users are willing to collect and how those differ from the opened items that are not collected.

In Jersey, both time and food type seem to affect an item's collectibility. Unlike London, baked goods are not highly desirable for collection. In fact, 6 of the 13 different categories have a collection rate of less than 60%, versus just one category in London. In addition, the single term that seemed to have an impact on an item's collection in Jersey was found in the title and referred to the type of food, rather than the state of the item. This suggests that users in Jersey either have different food preferences or use Olio to collect different things than London users do.

Although there were not many terms found to impact an item's collection, the textual analysis provided other useful results. In London the most common bigrams in collection notes across all listings were 'bring bag' and 'please bring'. For descriptions the most common bigrams included 'bring bag' and 'bring container'. Because these terms showed up so frequently in both collection conditions, it suggests that bringing one's own bag or container is not a deterrent for food collection in London.

For next steps, I plan to examine data in different London neighborhoods to see if there are distinctions at a more local level. I also want to see if there are trends based on multiple factors, such as words used and time of day. Finally, I want to build a model to predict if an item will be collected or not. This model can help Olio improve its collection rate and further reduce food waste.

REFERENCES

[1] Madina A. Abdulkadyrova, Andzor H. Dikinov, Hassan E. Tajmashanov, Lomali A. Shidaev, and Eliza A. Shidaeva. 2016. Global Food Security Problems in the Modern World Economy. *International Journal of Environmental and Science Education* 11, 12 (2016), 5320—5330. https://eric.ed.gov/?id=EJ1115556

[2] Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python.* O'Reilly Media Inc.

[3] Jenny Gustavsson, Christel Cederberg, Ulf Sonesson, Robert van Otterdijk, and Alexandre Meybeck. 2011. Global food losses and food waste — Extent, causes and prevention.

[4] M. Kummu, H. de Moel, M. Porkka, S. Siebert, O. Varis, and P.J. Ward. 2012. Lost food, wasted resources: Global food supply chain losses and their impacts on freshwater, cropland, and fertiliser use. *Science of the Total Environment* 438, 1 (January 2012), 477–489. https://doi.org/10.1016/j.scitotenv.2012.08.092

[5] Tamar Makov, Alon Shepon, Jonathan Krones, and Marian Chertow. 2019. Who Wants My Half Eaten Artisanal Sandwich? Uncovering the Structure and Dynamics of the Sharing Economy. Work in Progress.

[6] Margaret C. Nelson, Scott E. Ingram, Andrew J. Dugmore, Richard Streeter, Matthew A. Peeples, Thomas H. McGovern, Michelle Hegmon, Jette Arneborg, Keith W. Kintigh, Seth Brewington, Katherine A. Spielmann, Ian A. Simpson, Colleen Strawhacker, Laura E. L. Comeau, Andrea Torvinen, Christian K. Madsen, George Hambrecht, and Konrad Smiarowski. 2016. Climate challenges, vulnerabilities, and food security. *PNAS* 113, 2 (Jan. 2016), 298 — 303. https://doi.org/10.1073/pnas.1506494113

[7] Julian Parfitt, Mark Barthel, and Sarah Macnaughton. 2010. Food waste within food supply chains: quantification and potential for change to 2050. *Philosophical Transactions of the Royal Society B* 365, 1554 (September 2010). https://doi.org/10.1098/rstb.2010.0126