# Data Analyst Test Report

In order to investigate how users interact with our food and recipe content, this report will demonstrate the analysis in three main parts: content analysis, campaign analysis, and time/frequency analysis.

## 1. Content Analysis

---

### 1.1 Views distribution

First we can take a look at the distribution of views in different topics/formats.

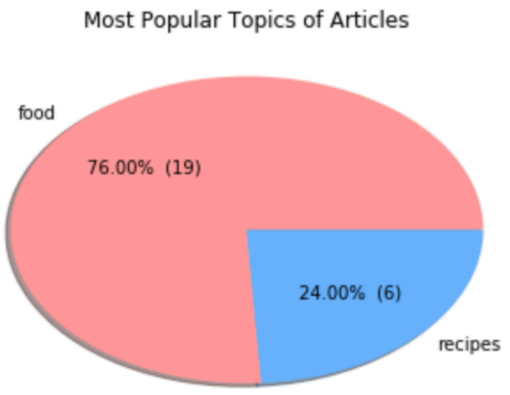
| topic | food | recipes | wellness | travel | sweepstakes | beauty | total |
|---|---|---|---|---|---|---|---|
| **views** | 67222 | 55768 | 101 | 30 | 10 | 1 | 123132 |

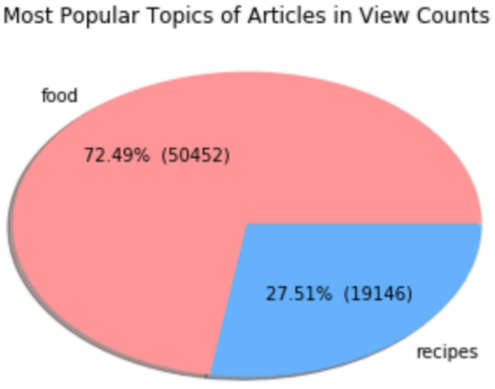
Food gains a little more views than recipes within this week, but overall we can say that they are equally popular. As for format distribution, the same pattern applies to list and recipe, plus an article format that takes a small part of the whole population.

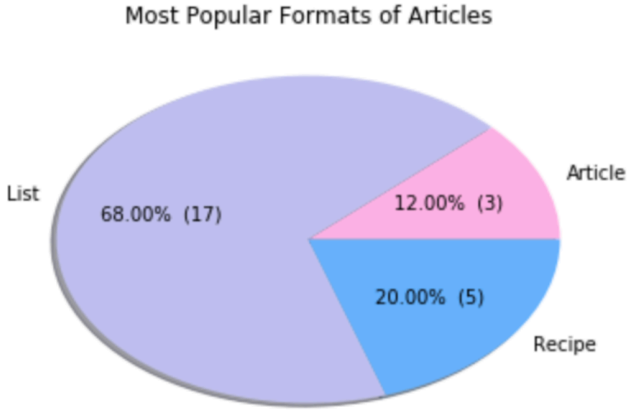
| format | List | Recipe | Article | Video | Content Hub | Guides | total |
|---|---|---|---|---|---|---|---|
| **views** | 61783 | 53269 | 7683 | 215 | 170 | 12 | 123132 |

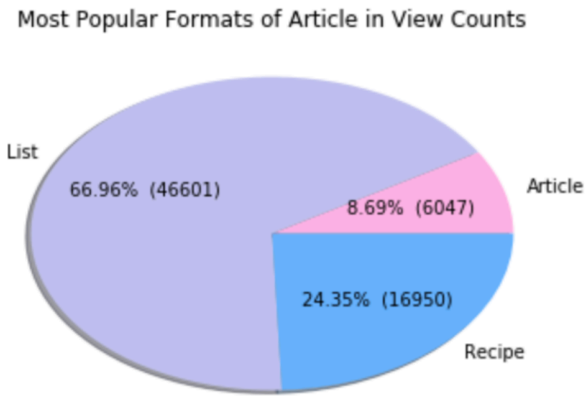
---

### 1.2 Content Popularity



I counted the Top 25 articles with the highest views. There is a distinct pattern that most of these articles are in the food topic and a list format. I used four pie charts to demonstrate the patterns showing above, which is more intuitional to compare.

Based on this pattern, **I recommend that we can push more articles from food topic and list format to customers in order to stimulate growth in page views**. Meanwhile, content creators are encouraged to write more articles following the same rule.

From the article titles of recipe section, I also noticed an interesting thing that dessert recipes appear frequently. Since PureWow's main audience is mostly females, it's definitely a good idea to publish more content in dessert recipes to grab their attentions.

Most Popular Articles in Recipes

| page_title | article_publication_date | click times |
|---|---|---|
| 3-Ingredient Strawberry Magic Shell Ice Cream Recipe - PureWow | 6/26/19 | 7768 |
| Cherry Tomato Galette Recipe - PureWow | 6/25/19 | 5458 |
| 8 Easy Single-Serve Recipes for Feasting Solo - PureWow | 6/17/19 | 2196 |
| Cacio e Pepe Cauliflower Recipe - PureWow | 1/19/18 | 1610 |
| Spicy Corn Carbonara Recipe - PureWow | 8/22/17 | 1179 |
| 30-Minute Angel Food Cupcakes Recipe - PureWow | 6/24/19 | 935 |
| No-Bake Berry Tiramisu Recipe - PureWow | 5/11/17 | 772 |
| 15-Minute Gazpacho Recipe - PureWow | 7/27/17 | 746 |
| Baking Sheet Quesadilla Recipe - PureWow | 5/23/16 | 677 |
| Baked Caprese Chicken Skillet Recipe - PureWow | 4/12/17 | 657 |
| Cold Sesame Cucumber Noodles Recipe - PureWow | 8/9/17 | 655 |
| Grilled Angel Food Cake with Blueberry Sauce Recipe - PureWow | 8/3/17 | 652 |
| Keto Guacamole Burger Bites Recipe - PureWow | 8/13/18 | 552 |
| Mini Spanish Tortilla with Zucchini Recipe - PureWow | 7/20/17 | 540 |
| No-Churn Mint Chocolate Chip Ice Cream Recipe - PureWow | 6/20/19 | 534 |

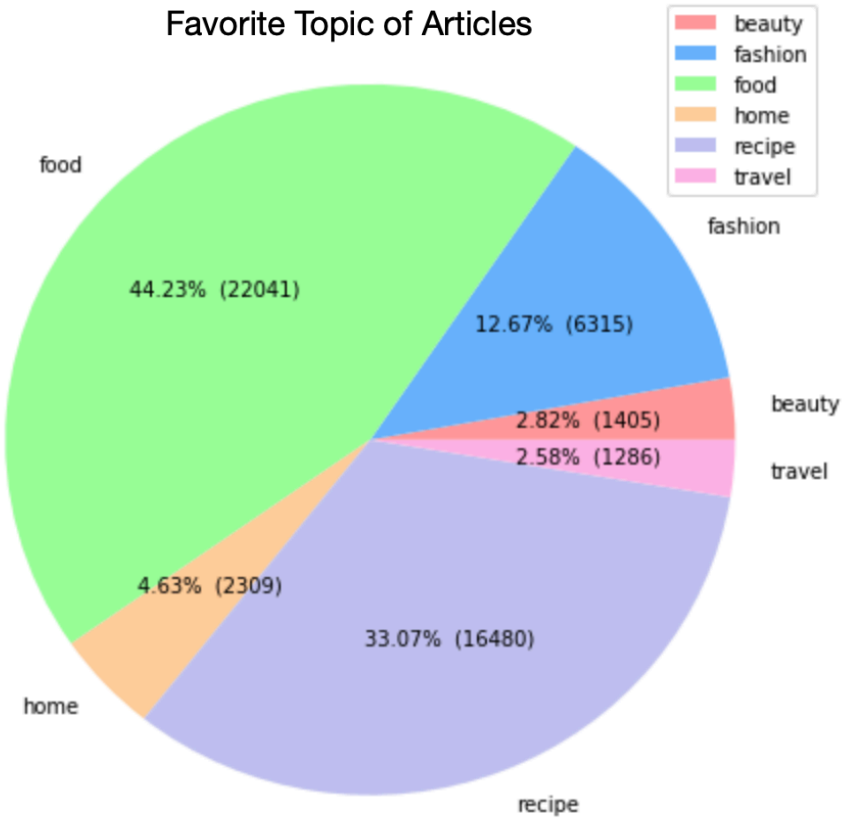
Another insight I found from the table above is recipe posts have no certain timeliness issues, since most of the popular recipe articles viewed in this week are old posts.

## 1.3 Interest-based customer segmentation

After grouping customers' viewing history in food and recipes, I combined it with their prior views information from the subscription dataset. According to the new subscription dataset, I labeled each customer with their favorite topics based on the highest views. Since some

customers have the same highest views in different topics, I gave them labels up to three for the convenience of further analysis.

In my opinions, it's important to label our customers since we can precisely target each person's content preference and make accurate recommendations based on group information. **For future work, I think it's viable to push articles which have received recognition within the group to customers with the same label, then check the conversion rate to evaluate the strategy's performance.**




# 2. Campaign Analysis

---

## 2.1 Campaign channels

In the article views dataset, email campaign takes the largest proportion and there is a part of data missing campaign information. For the convenient of further analysis, I used "unknown" to fill out all the NaN blocks. The distribution of all campaign channels is shown below:

| Campaign | email | unknown | social | syndication | cpc | sms | HUFFPOBRUNCH | sweeps | display | curatorcrowd | yummly | referral | foodandwine | 3Demail |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Views** | 79169 | 41774 | 899 | 803 | 419 | 38 | 11 | 9 | 3 | 3 | 1 | 1 | 1 | 1 |

It's obvious "unknown" channels won the second place with a huge lead, which is worthy studying in detail as well. So I listed all the referrer pages that came from "unknown":

Page referrers in unknown campaign

| | |
|---|---|
| https://www.purewow.com | 34437 |
| https://www.google.com | 763 |
| https://www.msn.com | 549 |
| https://www.yahoo.com | 474 |
| https://www.pinterest.com | 291 |
| https://admin.purewow.com | 61 |
| http://www.msn.com | 43 |
| http://www.pinterest.com | 31 |
| https://r.search.yahoo.com | 30 |
| https://www.bing.com | 27 |

Most of the "unknown" views came from the homepage that is probably a direct website access from our loyal customers. For the other views, search engines such as google, yahoo, and bing also contributed certain amount of clicks, however, they are much fewer than homepage visit. **If the company tends to run a marketing campaign in search engines, we can use the ranking list to decide the best channel to drive more traffic.**

Since I'm curious about the performance of social network campaign, I also listed the referrer pages that came from category "social":

Page referrers in social campaign

| | |
|---|---|
| https://www.facebook.com | 172 |
| https://www.pinterest.com | 141 |
| https://m.facebook.com | 124 |
| https://www.purewow.com | 120 |
| https://l.facebook.com | 17 |
| http://m.facebook.com | 13 |
| android-app://m.facebook.com | 7 |
| https://l.instagram.com | 6 |
| http://l.facebook.com | 6 |
| http://instagram.com | 6 |

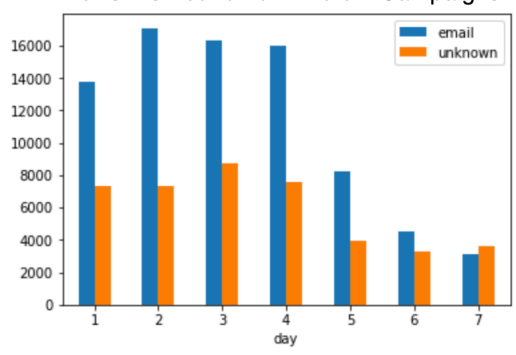
It seems that Facebook and Pinterest are the main contributors of social network visit. I noticed that there are share buttons to Facebook and Pinterest at the bottom of each article page. **We can provide more entries to share to other sites for more social media exposure.**

---

## 2.2 Campaigns in weekly pattern

I plotted the views distribution of campaign email and "unknown" in a week, and I discovered visits from "unknown" campaign stay more stable in weekends while email visits experienced a huge drop. **In order to drive more traffic in weekends, we could develop more campaign strategies in search engines and social media, which could be more efficient than email since people don't open their email box as often as usual during weekends.**

Views Distribution of Different Campaigns



## 2.3 Article content in campaigns

The most popular topics in different campaign channels have a huge variance as the tables shown below. Food posts receive plenty of views comparing to recipes in email campaign, while recipes are trending in “unknown” channels. **Based on this insight, we can send more articles on food topic in newsletter when focusing on email campaign. Meanwhile, we can increase the share of recipe articles when conducting other campaigns.**

`Email views in different topics`

| topic | food | recipes | wellness |
|---|---|---|---|
| **views** | 55063 | 24019 | 87 |

Unknown views in different topics

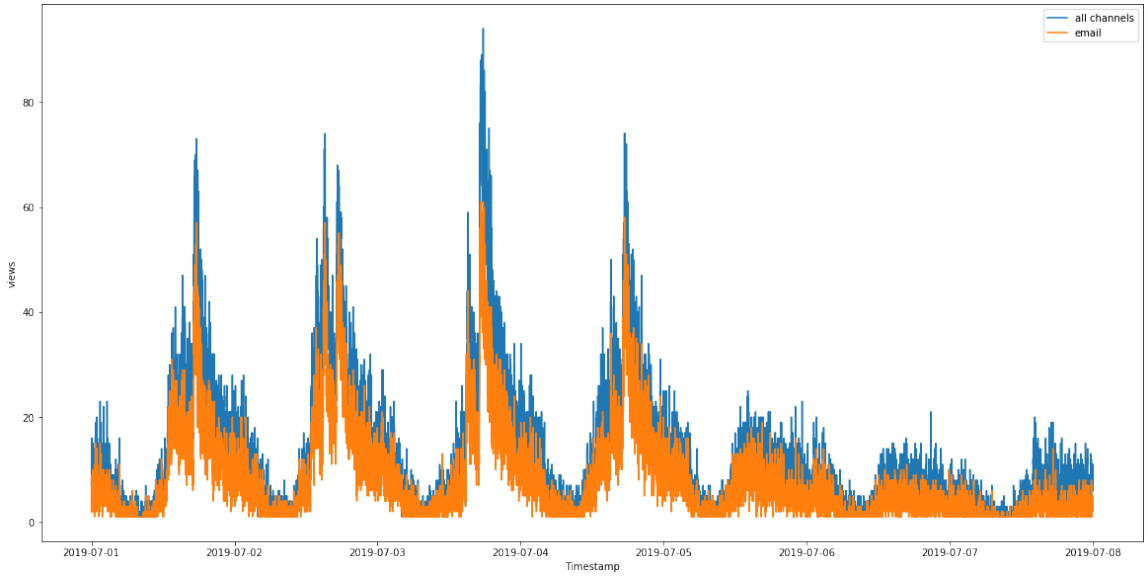
| topic | recipes | food | travel | wellness | sweepstakes | beauty |
|---|---|---|---|---|---|---|
| **views** | 31390 | 10339 | 30 | 13 | 1 | 1 |

# 3. Visit time and frequency analysis

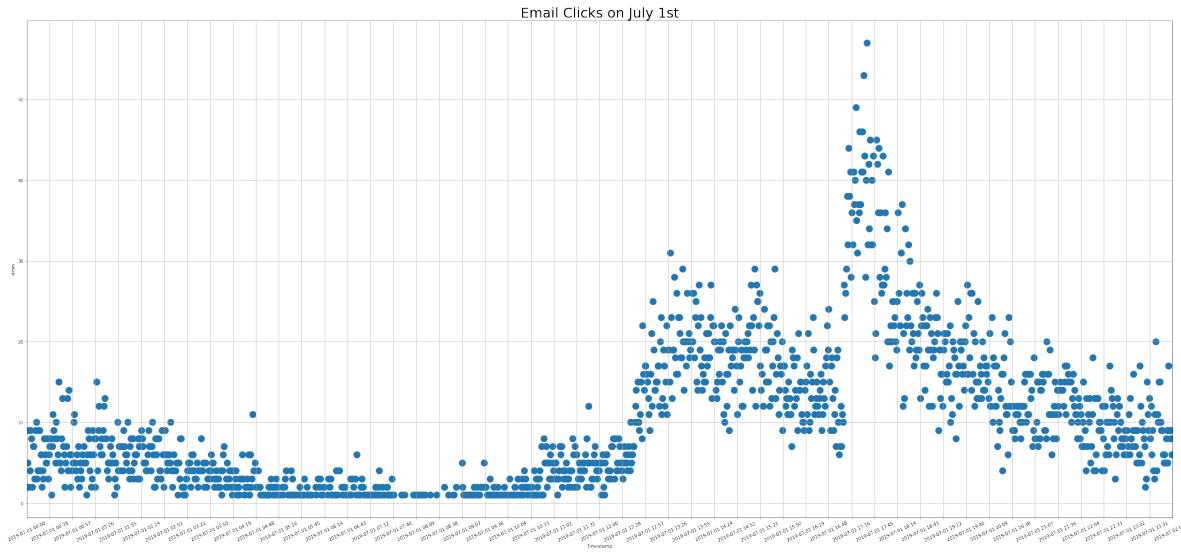
## 3.1 Time-based strategy in email

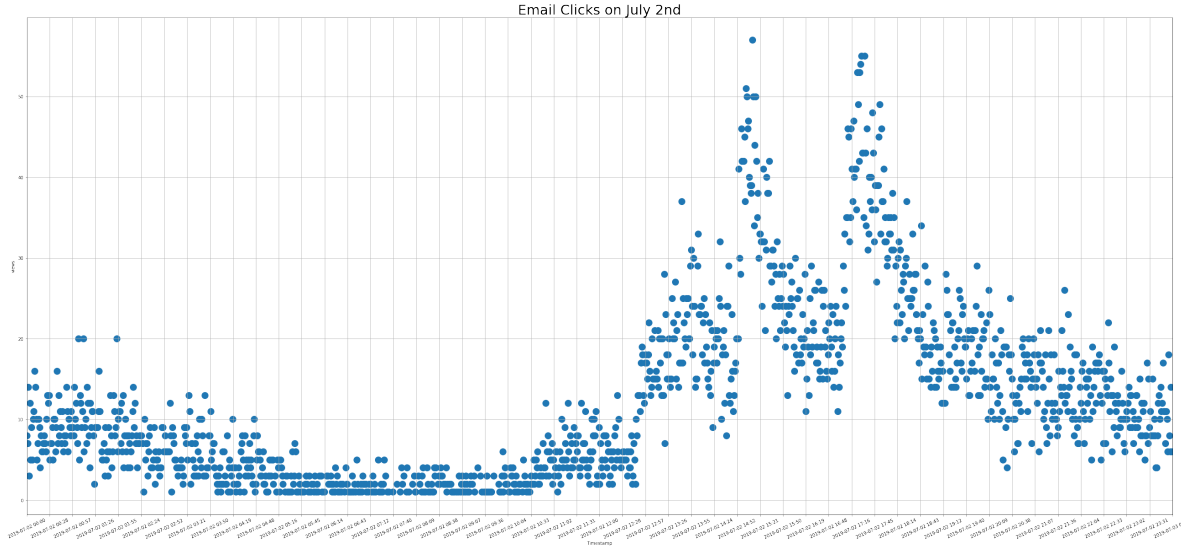
Since visits from emails take the vast majority and its time factor is easy to control, this time analysis is mainly based on email strategy. I visualized the views dataset in a time series plot and compare the email views with the whole population, which are following the same pattern.

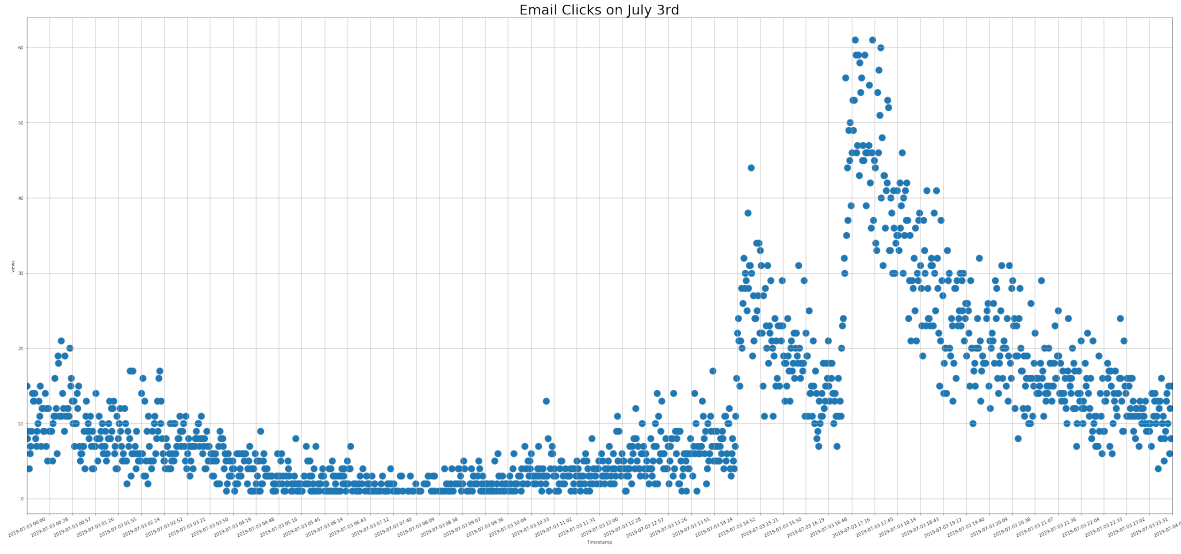
Views Count from July 1, 2019 to July 7, 2019

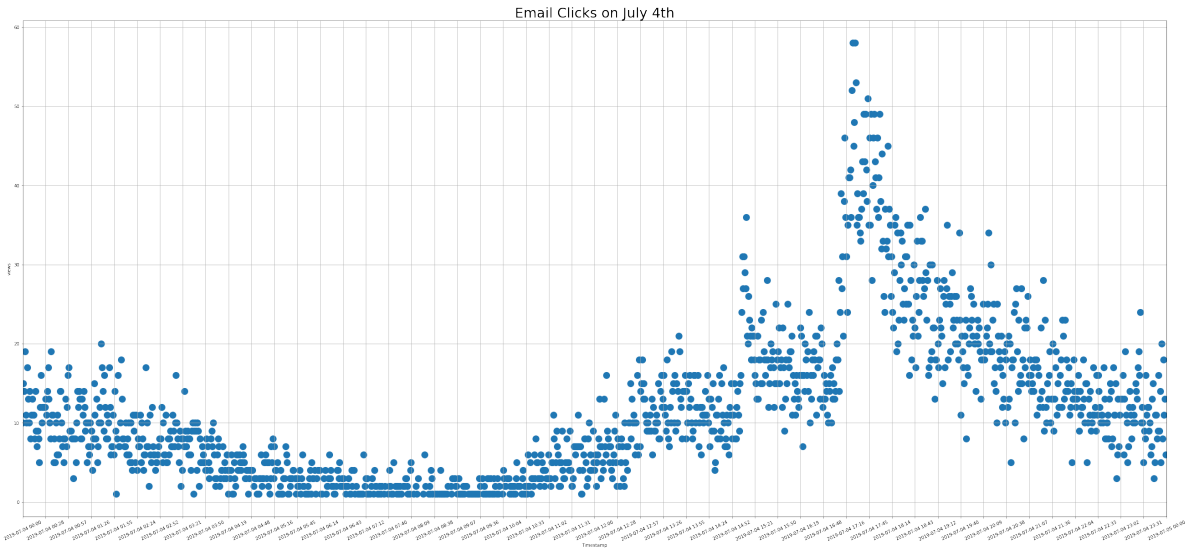



Based on the page views of each timestamp in this week, everyday has the similar pattern that they have at least a peak around 80 views from Monday, July 1, 2019, to Thursday, July 4, 2019. Because July 4th is a national holiday, the last three days, Friday, July 5, 2019, to Sunday, July 7, 2019, could be a long weekend to many people, which explains why page views distributed so differently than the previous days. In order to find out the best time to send out email newsletters precisely, I decided to break it down and observe the time pattern day by day.

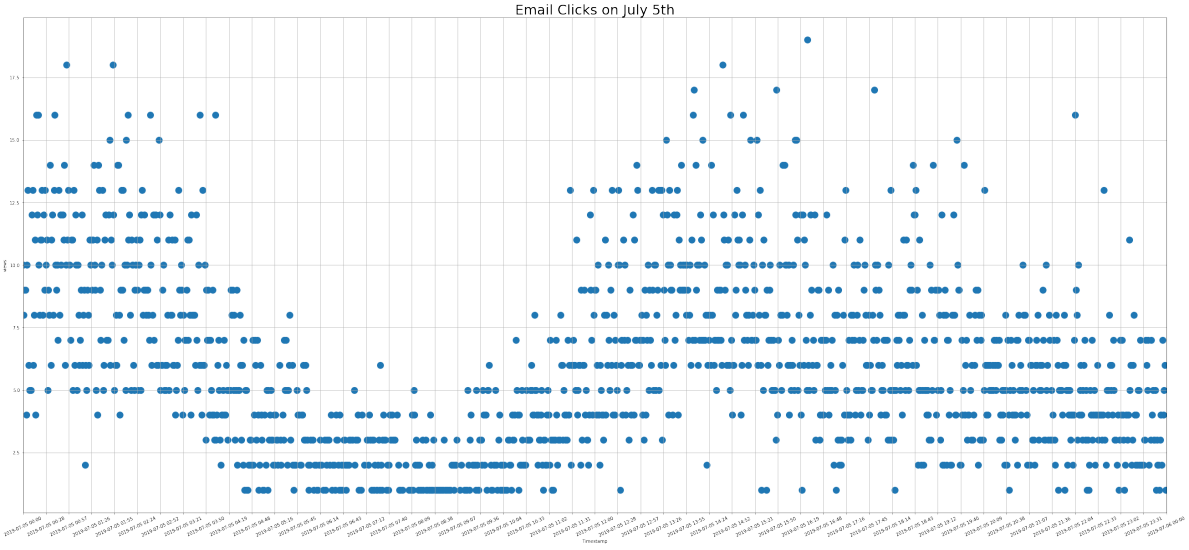

Email Clicks on July 1st
Timestamp

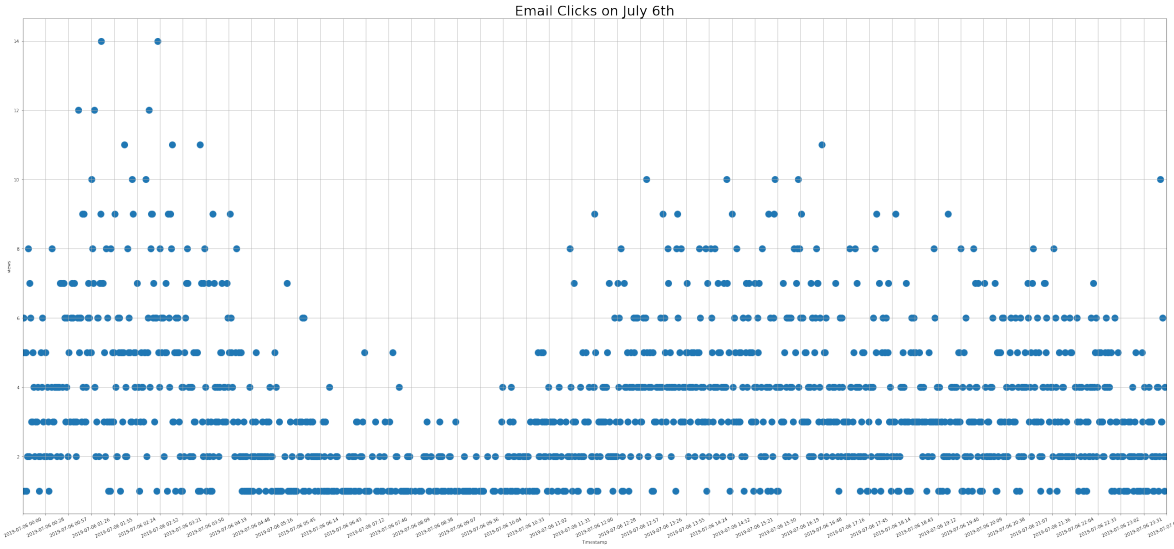

Email Clicks on July 2nd
Timestamp

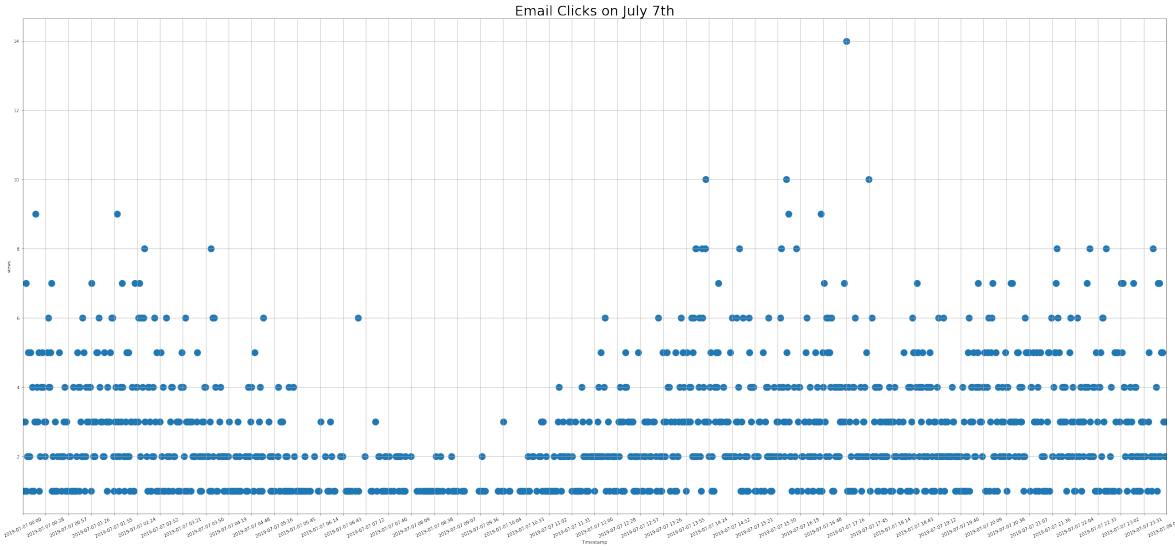

Email Clicks on July 3rd
Timestamp




Based on the scatter plots, we can see that between Monday to Thursday, the page views remain steady in the morning and gradually slide up and down between 1pm to 4pm. Then an abrupt increase starts at 5pm, jumping to 60 approximately, and slowly bounce back around 10 in the end of the day. **According to the finding, a simple assumption can be made that we can start sending out our email newsletters before lunch time in weekdays to make sure it appears not far away from the top when our customers open their email inbox.** If we could obtain more data such as email sending frequency, and email sending time of each page views data, we could draw a more solid conclusion.





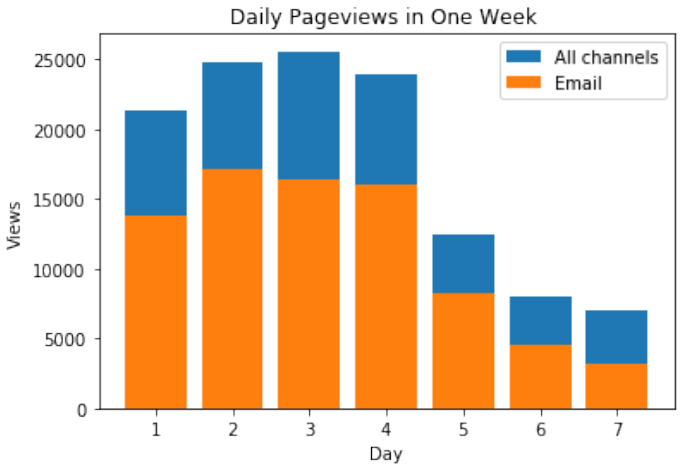



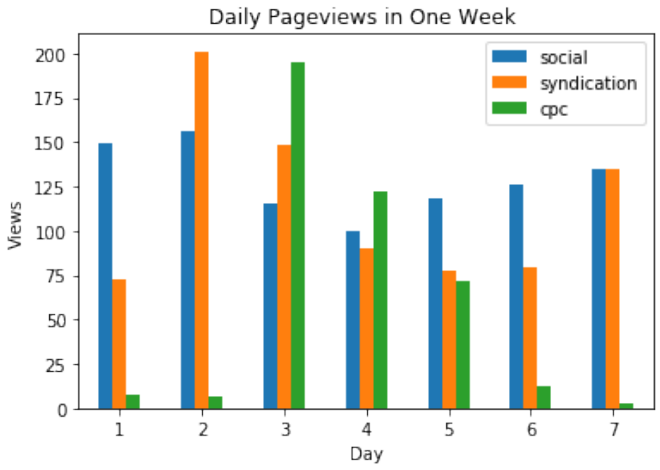

The last three days don't have a distinct pattern since all data points are clearly discrete. Again, this could be caused by the holiday, which can be easily verified with at least one entire month data. Besides, the page view fluctuates within a small range in the same day. Therefore, we cannot capitalize this finding when making our email strategy.

## 3.2 Daily views distribution of different channels

The first bar chart indicates that views from email and all channels are following the same pattern that the first four days remain stable and the last three days experienced a major drop. The second chart includes views from social, syndication, and cpc, which gained 899, 803, and 419 views respectively in this week. Although these three channels represent a small part of views, their patterns provide some reference for further analysis. Views from social channels stay steady through the entire week while cpc even beat them all on Wednesday and Thursday.

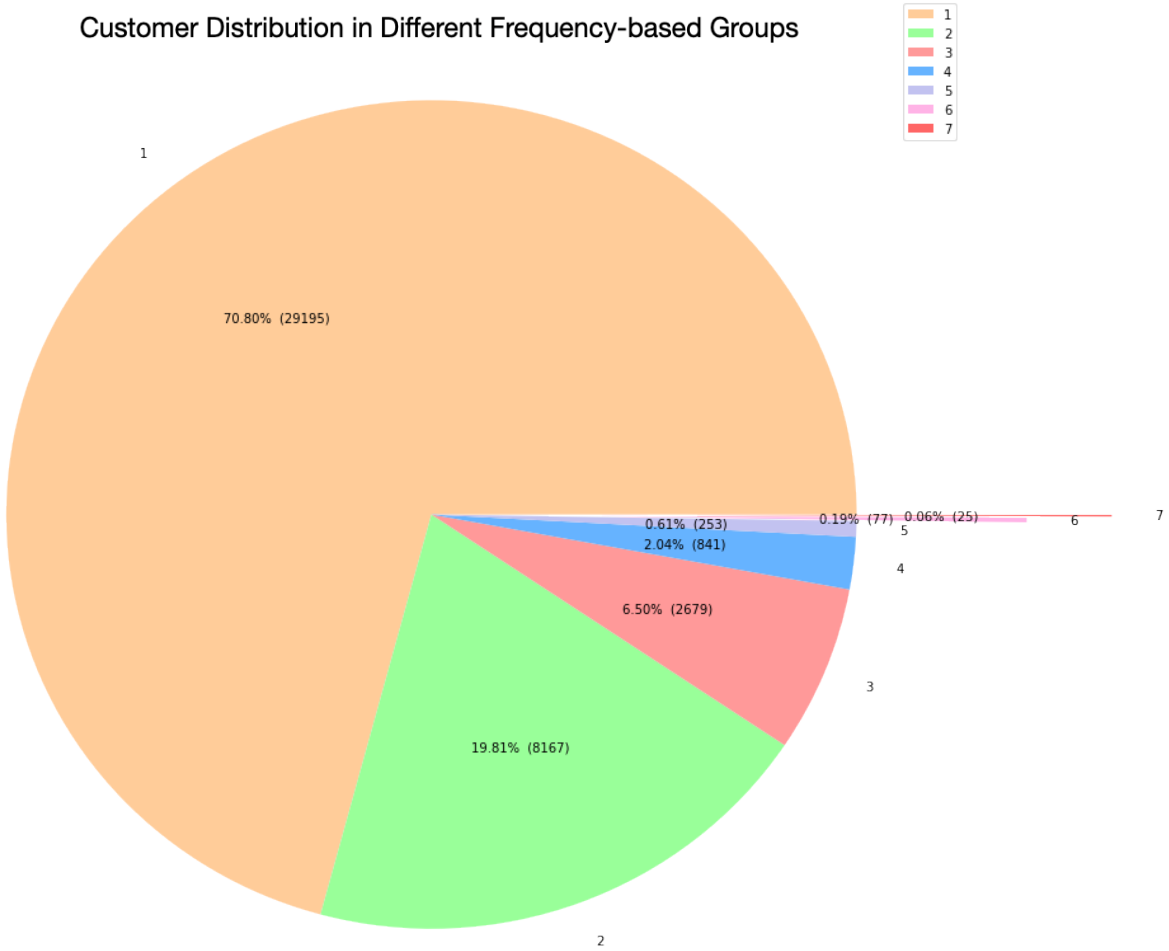
If more views data from these channels can be obtained, we can verify these insights to develop periodic marketing strategies and target campaign channels accurately.



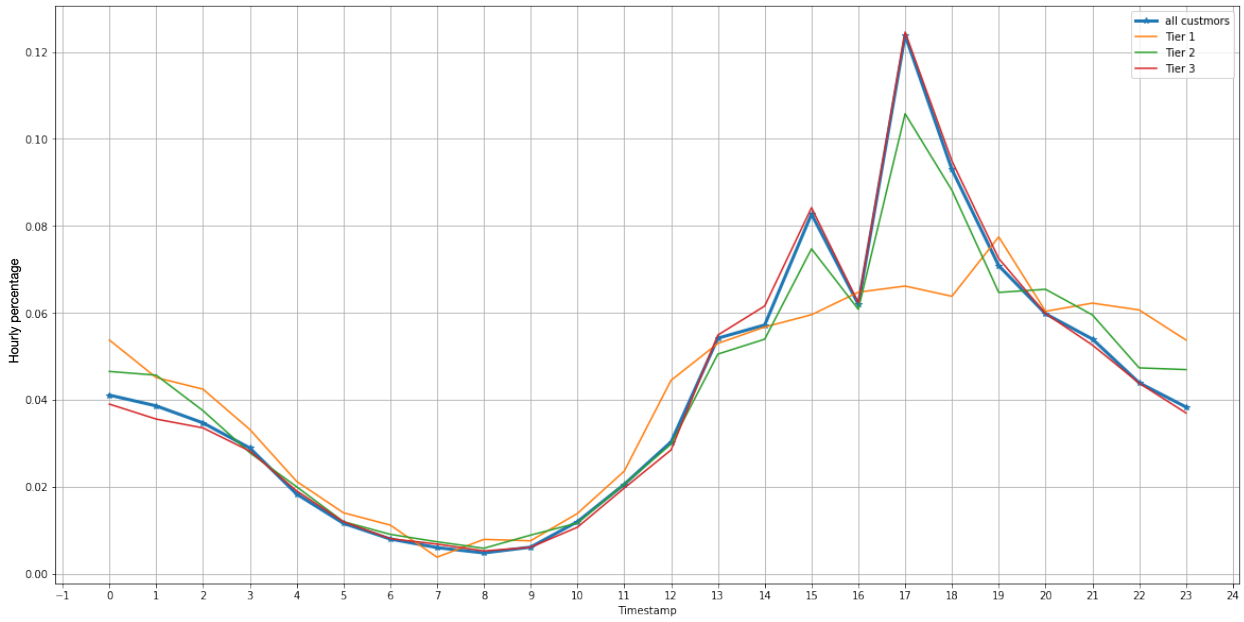

## 3.3 Frequency-based customer segmentation

Another method to categorize our customers is to group them based on their weekly visit frequency. From visiting every day (labeled 7) to once a week (labeled 1), customers are distributed as follows:



Because the distribution of different customers is highly imbalanced and some groups only have a few people, I decided to regroup them into three tiers. If a customer visits our webpage more than four days in a week, they must be strongly interested in our content and loyal to our brand; I put them into tier 1. For customers who have visited three or four days, they belong to tier 2. The rest of customers who only visited twice or once a week, they are divided into tier 3. This segmentation is meant to investigate different patterns hidden behind the loyal group, so that we can apply the insight into the strategy to help drive engagement and loyalty. Then I visualized the page views data from different groups into a hourly time series plot.

Hourly Views Percentage of Different Customer Tiers




In contrast to other tiers, tier 1 customers have steady increase and decrease in page views between 1pm to 10pm, instead of a dramatic fluctuation. In addition, the line of tier 2 customers also locate below tier 1. It shows that time factor has less impact on loyal customers than other customers during this period. This customer segmentation analysis can be extended further based on different content preference, frequent visit channels, and seasonality patterns if we can obtain more information.

# 4. Conclusion

According to the analysis above, we can summarize four data informed recommendations to our marketing team as follows:

1. In order to stimulate growth in page views, we can push more articles in food topic and list format to customers. Meanwhile, content creators are encouraged to write more articles following the same rule.
2. We can send more articles on food topic in newsletter when focusing on email campaign. Meanwhile, we can increase the share of recipe articles when conducting other campaigns.

3. In order to drive more website traffic in weekends, we could develop more campaign strategies in search engines and social media, which could be more efficient than email campaign since people don't open their email box as often as usual during weekends.
4. We can start sending out our email newsletters before lunch time in weekdays to make sure it appears near the top when our customers open their email box.

If we have a chance to know more about our customers, I would like to collect their demographic data and time on page data.

1. With demographic data, we can conduct research on customers' preference related to their occupation, education, and other aspects. This information can help us improve the email content personalization. This task can be achieved by building a prediction model with the help of machine learning algorithms.
2. With time on page data for each user, we can find out which articles that users are fully engaged in and interested in. These articles will increase customers' exposure to good content and help improve campaign performance.