Ann Arbor Summer Festival Visitor Survey Final Report

By Vicky Wang (viwa@umich.edu)

Project Overview

Ann Arbor Summer Festival at Top of the Park is a free, yearly event held in Ann Arbor, Michigan over the course of two weeks in mid June. Ann Arbor Summer Festival also puts on ticketed events in venues such as Hill Auditorium for artists to perform. This year, a visitor survey was administered to visitors to A2SF events, either through in-person or email/online methods. The following questions were asked in the survey.

1. What is your zip code?
2. What is your age?
3. With what gender do you most identify?
4. What is your education level?
5. Which of the following describes your race and/or ethnicity (Note: we recognize the following list is not all-encompassing for how people may identify. Please select all that apply, and in addition, you may further self-describe race and/or ethnicity using the 'Other' option.)?
6. What was your gross household income during the last year?
7. How many times did you attend Top of the Park this season?
8. What were the biggest factors that influenced your decision whether or not to attend? Select all that apply.
9. What were your favorite offerings this season? (Choose up to three)
10. On average, how much money did you spend at Top of the Park this season (drinks, food, merchandise, etc.)?
11. A2SF is an independent, community-supported 501(c)3 nonprofit organization, with individual donors contributing $340,000 or 20% of our yearly budget. How likely are you to donate to our organization?
12. On a scale of 1-10, how likely are you to attend one of our events in the future?
13. On a scale of 1-10, how likely are you to recommend our events to a friend or family member?
14. On a scale of 1-10, how satisfied are you with the Top of the Park music programming?
15. On a scale of 1-10, how satisfied are you with family offerings (KidZone, Kids Rock) at Top of the Park?
16. On a scale of 1-10, how do you feel about the overall trajectory that A2SF is heading in?
17. Is there anything else you would like to share about Top of the Park or A2SF as an organization?
18. What would make it easier for you to participate in A2SF events?
19. How did you hear about A2SF at Top of the Park?
20. If you would like to be entered to win a $50 gift card, please provide your email address.
21. If you provided your email, do you want to be added to our mailing list?

Project Timeline *January - April 2024*

CTAC student team consulted with A2SF to design a survey and disemmination plan for the Summer 2024 season.

June 14 - 30 2024

CTAC students supported in-person data collection measures during the summer festival. Data was not collected on June 19 and 22, and the festival was dark on June 17 and 24.

August - September 2024

CTAC student team analyzed the in-person and postseason data, combining the two sets into a single 2024 A2SF dataset following robust cleaning.

Survey Data Cleaning The in-person survey had 183 respondents and the postseason survey had 349 responses. After bot removal of 162 responses from the postseason survey, the final postseason survey sample size was n = 187.

A problem that arose during the analysis for post-season data was the presence of bots. Despite inclusion of reCAPTCHA tests in the CrowdSignal survey platform, approx. 150 responses on the post-season survey indicated bot responses. Even after cleaning, there is not a 100% guarantee the remaining responses are human, as we cannot depend solely on short response langauge to assess whether a respondent is a bot or not. To address these bots during cleaning, we use a combination of methods, per this website. The methods we used were:

1. Removing respondents from the same IP Address
2. Removing unusually short response duration entries
3. Reviewing open-ended questions

However, this does not fully remove responses, so other less proven filters were also applied. For example, if country was not the United States, then the respondents should not have included a zip code.

In the future, there are several methods to prevent this from occurring. Some have already been implemented, but there is merit to adding additional safeguards. Existing methods (KEEP THESE!): * reCAPTCHA inclusion * collect IP Address * collect Start/End time/dates * include mandatory open-ended questions

To include in the future:

- "Honeypot Questions" include a single test question that consists of white text on a white background. Humans should skip this, but bots will fill it out.
- Image question. Bots cannot typically read images, so taking a screenshot of a text question and uploading that image makes it more challenging to respond.
- Enable more CrowdSignal restrictions:

1. One response per computer
2. IP restriction
3. Email restriction

Understanding Respondent Demographics The results demonstrate that while ethnicity aligns with the expected Washtenaw County results, race and ethnicity clearly demonstrates that respondent demographics were not in line. Using the following strategies outlined may help standardize the results and determine whether this is an accurate reflection of the visitor demographic of A2SF events, or if this was due to surveyor error. Data Collection

Post-season and in-season analysis yields some different visitor demographic results. For example, in-person respondents skewed significantly younger, while post-season respondents were much older overall. This was in part due to the surveyors were closer in age to the 18-24 category, so they may have felt more comfortable talking to similar-aged individuals. Additionally, the race and ethnicity demographics did not closely align with Washtenaw County expectations, either indicating sampling was non-representative or the population A2SF serves does not align with County demographics. Possible ways to address this: * recruit volunteers of many ages/backgrounds to walk around the festival and talk to patrons. * train volunteers to speak with everyone. * stratify sampling. for example, have surveyors walk around, and speak with every 5th person they go past. * Set a target population in advance to the survey collection period. For Washtenaw County, the collected demographics are as follows:

- Age:

1. 0-19: 26.0%

2. 20-34: 26.1%
3. 35-49: 19.4%
4. 50-64: 18.2%
5. 65+: 20.5%

- Median Income: $79,665
- Families - $112,532
- Married-couple families - $131,226
- Nonfamily households - $44,446
- Educational Attainment:

1. High school - 14.3%
2. Some college, no degree - 16.6%
3. Associate - 6.7%
4. Bachelor's - 27.0%
5. Graduate/professional degree - 30.9%

- Race/Ethnicity:

1. American Indian and Alaska Native alone: 0.30%
2. Asian alone: 8.54%
3. Black or African American alone: 10.87%
4. Hispanic or Latino (of any race): 5.26%
5. Native Hawaiian and Other Pacific Islander alone: 0.05%
6. White alone: 65.42%
7. Some Other Race alone: 1.88%
8. Two or More Races: 7.44%

Important Takeaways Race/Ethnicity: There is no statistical difference in the race demographic breakdown between in-person and online-collected survey responses, but there is a difference between both individual survey source datasets and the current Washtenaw County percentages. While the survey results appear to accurately reflect the racial/ethnic demographics of the audience A2SF currently serves, they do not fully align with the broader demographics of Washtenaw County (American Community Survey 2020). If A2SF aims to engage the entire county, it may be beneficial to examine these differences more closely and consider targeted outreach strategies to better reach and include those demographic groups that are currently underrepresented.

Visitor Hometowns: A2SF aims are to service Washtenaw County, and this is well-supported by the significant majority of visitors coming from Washtenaw County zip codes.

Age: Older populations more frequently responded online, while younger populations were over-sampled in-person, likely because canvassers shared that characteristic.

References

- 2021 American Community Survey
- Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity

Summary of Cleaning Process

Post-Season survey data was collected via email, social media, and other online methods. As a result, despite reCAPTCHA inclusion, there were high numbers of bot responses within the data set. Initial survey analysis required cleaning and removing these responses from the dataset.

A separate survey on CrowdSignal collected participant data for the post-season survey. This was cross-joined with the post-season survey on primary key "Respondent.ID" to determine the IP Addresses for all respondents. Then, IP address analysis could be used to determine bot presence.

First, we grouped by IP.Address in order to count multiple responses for the same IP address. We shouldn't be expecting more than one response from a single address, so we removed any duplicates from the dataset.

Following IP verification, we further examined specific responses that may have evaded this filter. Another strong indicator of bots was unusually short response duration. Because Time.Taken is a parameter within the post-season-respondents (participants) survey, we used these numbers to determine unusually short responses. Through this, we found that there is no evidence for bots present based on extremely short response times only. However, after examining post_season_ip to a new file, "post_season_cleaned_ip_addresses.csv," there were clearly still bots within the data based on responses and short answer indicators.

To filter further, we looked at non-US auto-collected countries and manual zip code entries that were not 5 characters. Since auto-collected countries outside of the United States are unlikely given the scope and target audience of the survey, we removed those responses. Additional manual review of row responses for non-US locations indicated bot behavior such as illogical free response answers.

Non-US countries should not have filled out the zip code question. The "Anything to share" column was used as an additional verification of bot-indicating behavior.

Zip codes should be 5 characters in length. When we see typos like "4103" instead of "48103", it is possible this is a typo, but we should not assume this to be a zip code in 48103.

Because some United States-based responses may also be bots, we have to be creative to identify additional ones. To do this, we used a function to identify states and zip codes and determine whether the zip codes, which respondents manually inputted, match with the states/countries, which was automatically collected from respondents' device locations.

Following identification of "Match" or "Not Match", we filtered the full dataset so it only contains "Match" results. Then, we created a fully cleaned file that can be used in further post-season analysis and data visualization methods.

Final cleaned results may still some suspicious responses because the short answer does not align well with the question asked, but these are few in number.*

Notably suspect: **304885124, 304722909, 304882174, 304751759**. However, these managed to pass ALL of the following:

- Non-duplicate IP Address sources
- Normal response duration
- United States
- Zip code length is 5 characters
- State (auto-collected) and Zip code (manual-collected) match

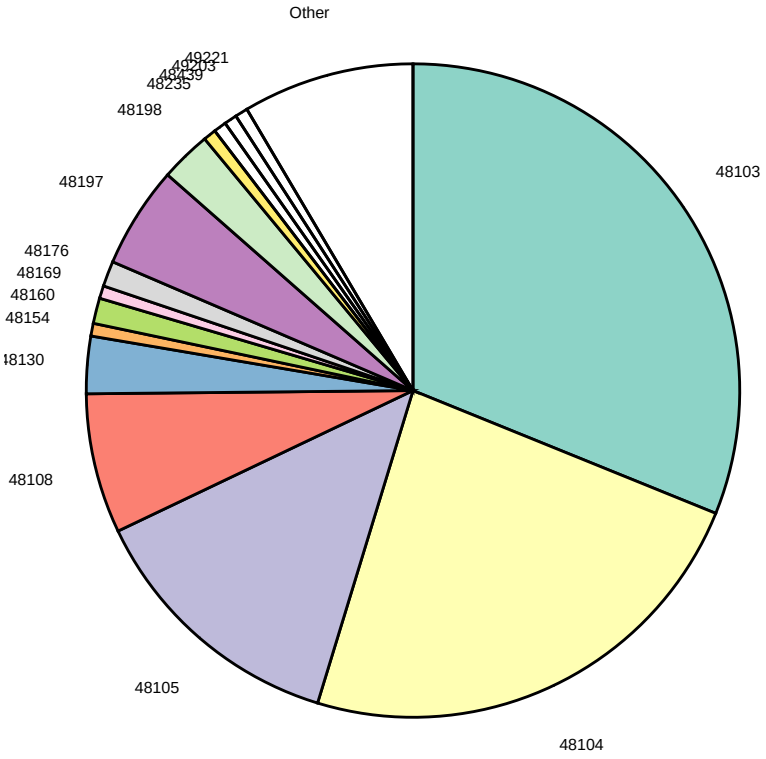
As a result, we will leave these responses in the dataset, but should exercise caution when performing analysis.

The final output of this cleaned file was created and named "post_season-processed.csv."

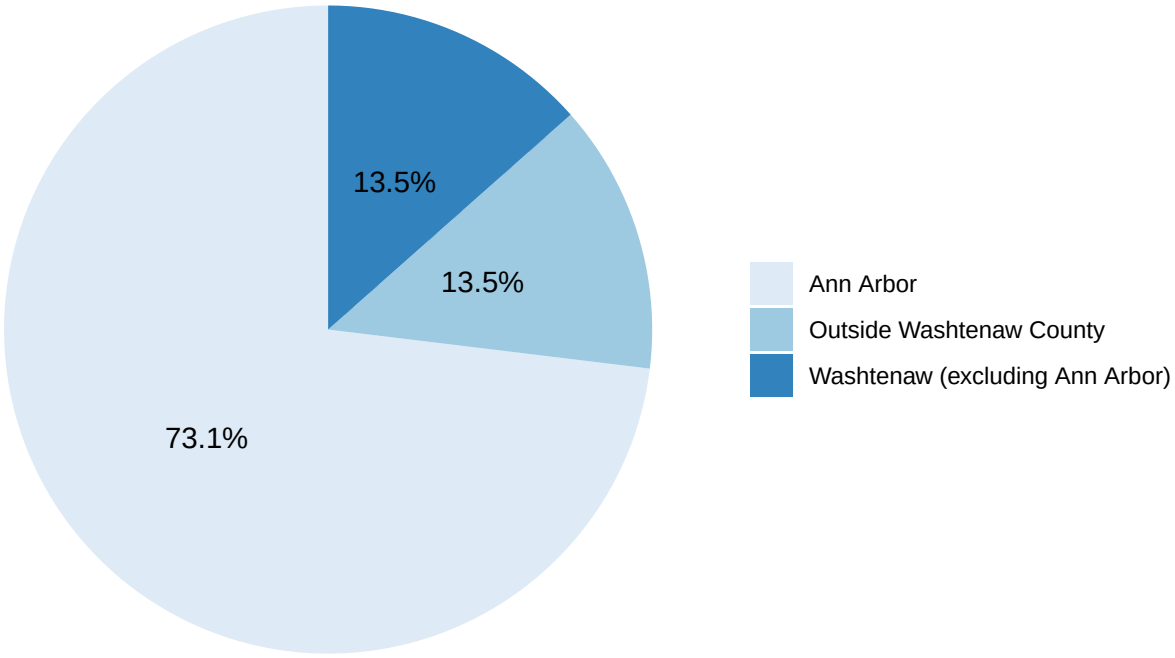
Data Visualization

Where are A2SF Visitors From? *Summary of all represented zip codes of A2SF visitors.*

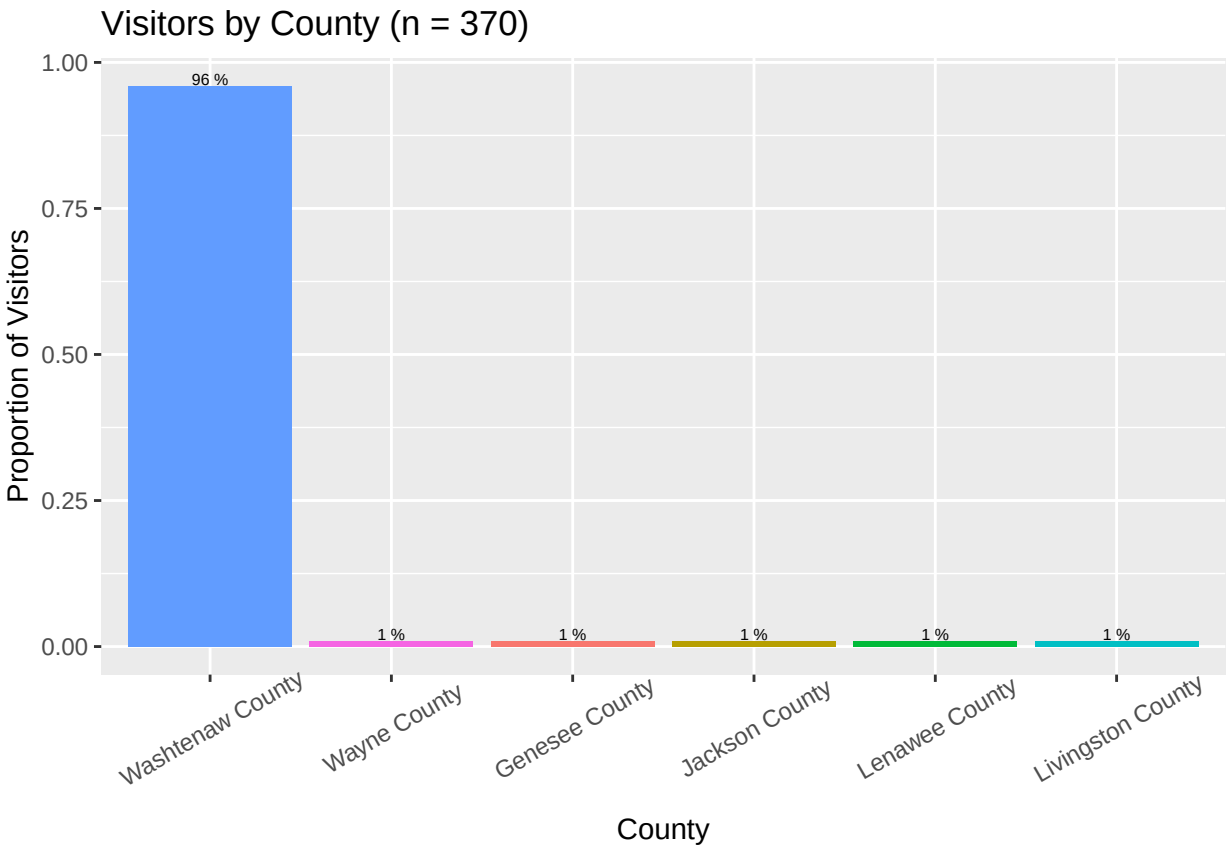
Visitors by Zip Code



Simplified Summary

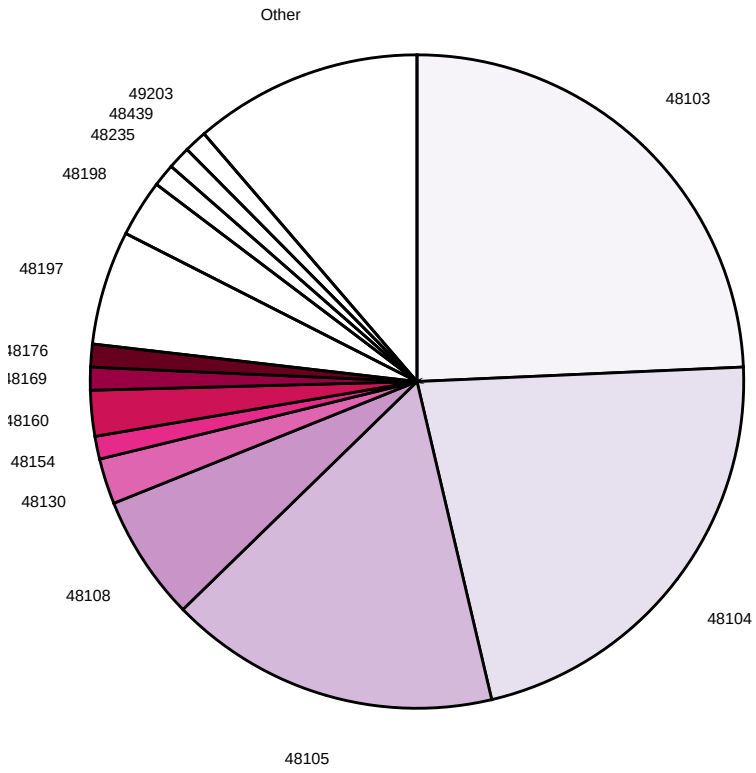


All visitors by county. Numbers above bars represent the proportion of visitors from represented counties.



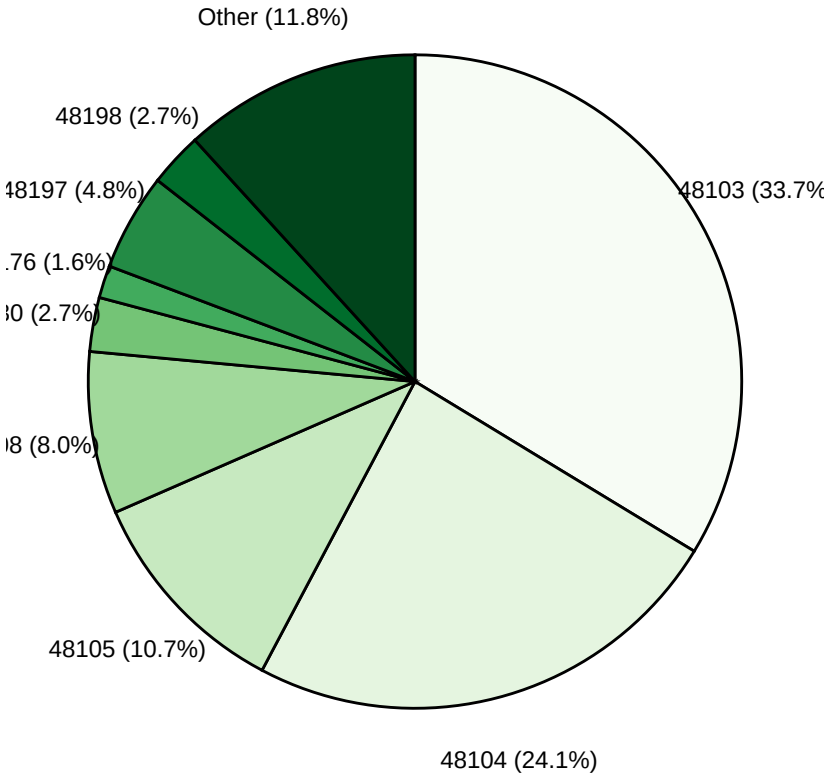
In-season Visitors

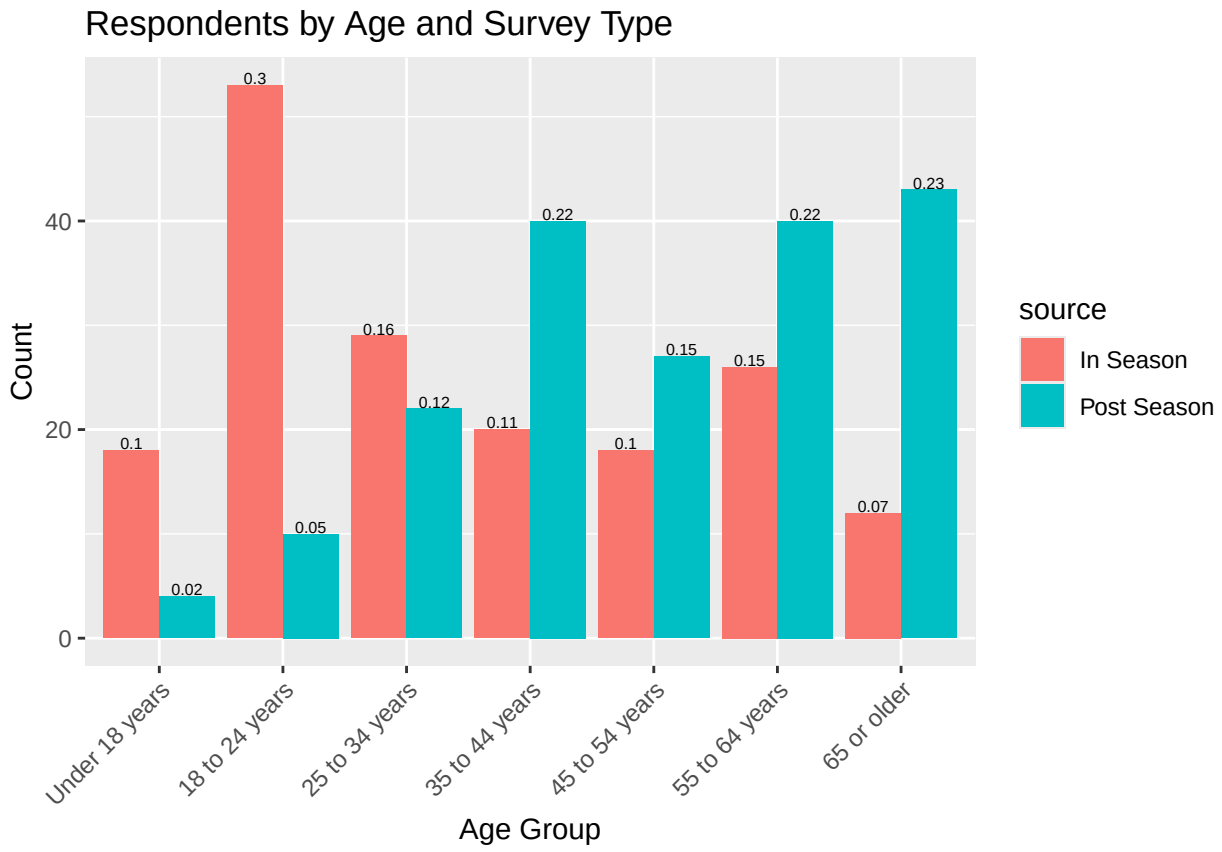
Rationale: non-5 digit zip codes were omitted, as it is difficult to determine the typing error committed to lead to an abnormal zip code. Additionally, collapsed all zip codes collected with count == 1 into an "Other" category in order to better visualize most common visitor homes.



Post-season Visitors

Visitors by Zip Code



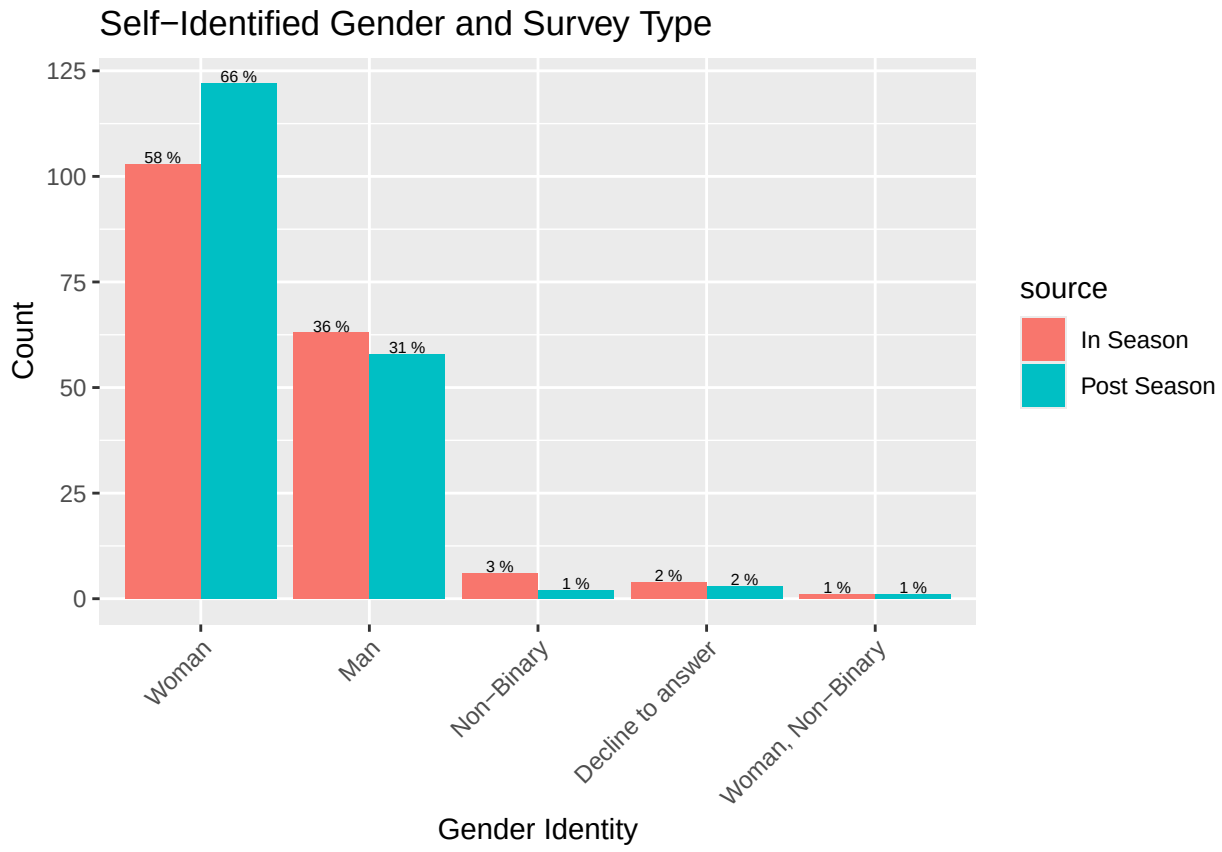


Age Analysis

Older age groups answer in higher proportions during the post-season, while younger groups were over-sampled in the in-season group. Pearson's Chi-squared test indicates that there is a difference in respondents by age group between the two survey sources.

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 67.905, df = 7, p-value = 3.913e-12
```

Gender Analysis *Visitors by their self-identified gender.*



Other self-described gender identities.

Survey	Other Identity	Count
In Season	Gender fluid	1

Race/Ethnicity Analysis *All unique race/ethnicity identity selections.*

Race/Ethnicity
American Indian or Alaska Native
American Indian or Alaska Native, Hispanic, Latino or Spanish origin, White
American Indian or Alaska Native, White
Asian
Asian, Hispanic, Latino or Spanish origin
Asian, White
Asian, White, Mixed heritage
Black or African American
Black or African American, Asian
Black or African American, Hispanic, Latino or Spanish origin
Black or African American, White
Black or African American, White, Mixed heritage
Decline to answer
Hispanic, Latino or Spanish origin
Hispanic, Latino or Spanish origin, White
Middle Eastern and North African (MENA)

Race/Ethnicity
Middle Eastern and North African (MENA), Mixed heritage
Mixed heritage
White
White, Middle Eastern and North African (MENA)
White, Mixed heritage

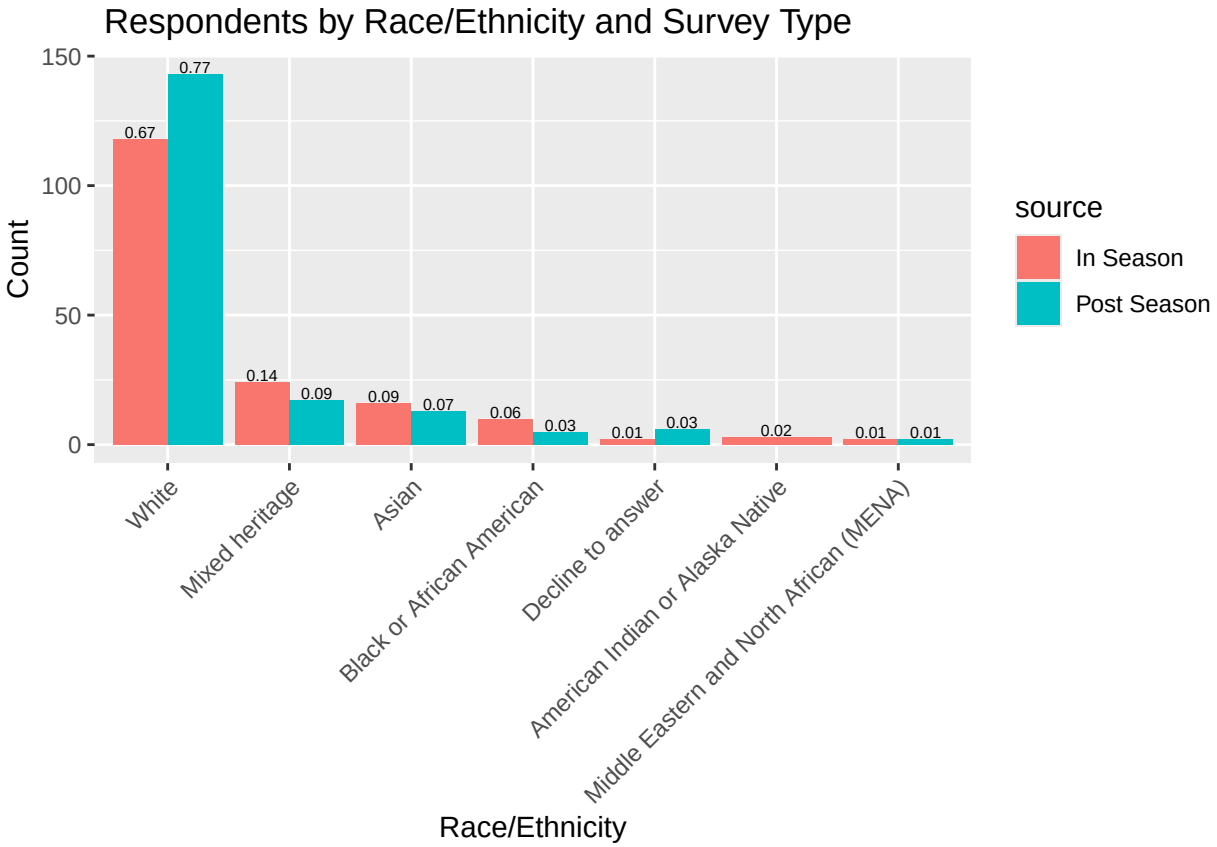
Ethnicity question- Hispanic,Latino, Spanish or Not?

Ethnicity	Count	Percent (%)
Hispanic, Latino or Spanish origin	23	6.32
NOT Hispanic, Latino or Spanish origin	341	93.68

Ethnicity does not align with Washtenaw County ($p < 0.10$).

```
##
##  1-sample proportions test with continuity correction
##
## data:  observed_count out of total_count, null probability expected_proportion
## X-squared = 0.047284, df = 1, p-value = 0.8279
## alternative hypothesis: true p is not equal to 0.0526
## 95 percent confidence interval:
##  0.03122254 0.10553287
## sample estimates:
##          p
## 0.05882353
```

Race with mixed heritages collapsed into a single general category.



Race	Frequency	Percent (%)
White	261	72.30
Mixed heritage	41	11.36
Asian	29	8.03
Black or African American	15	4.16
Decline to answer	8	2.22
Middle Eastern and North African (MENA)	4	1.11
American Indian or Alaska Native	3	0.83

Race/ethnicity does not vary by source.

This could indicate that the sampling method was unbiased and the population of patrons is relatively accurately represented.

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 23.058, df = 20, p-value = 0.286
```

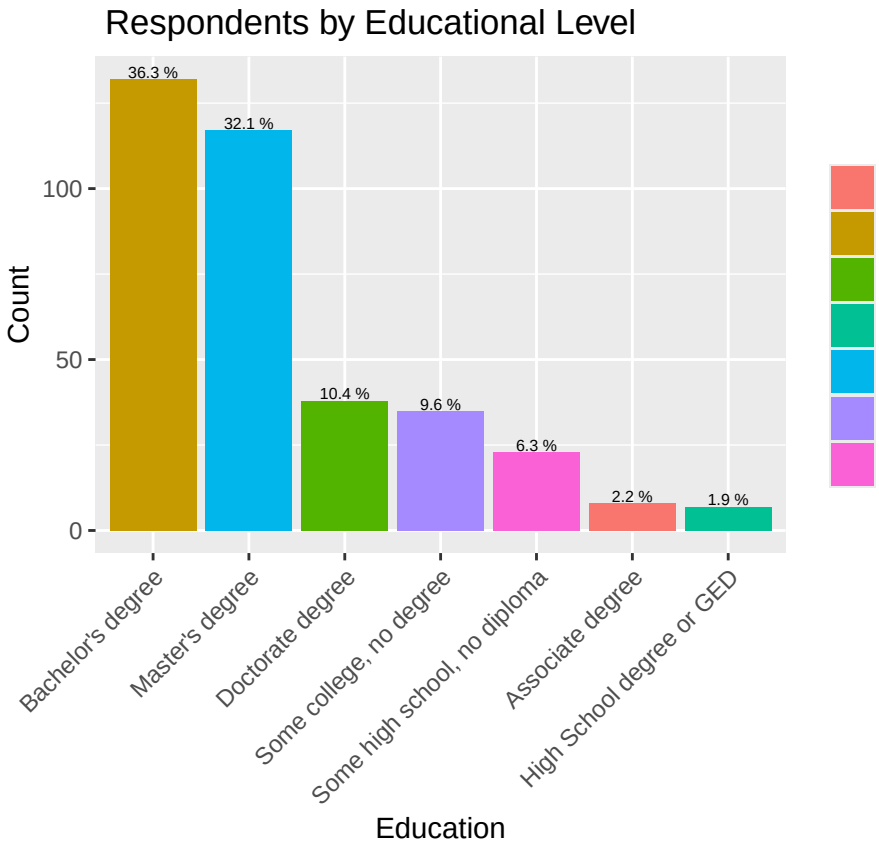
Race/Ethnicity representation does not align with Washtenaw County demographics ($p < 0.0001$).

```
##
##  Chi-squared test for given probabilities
```

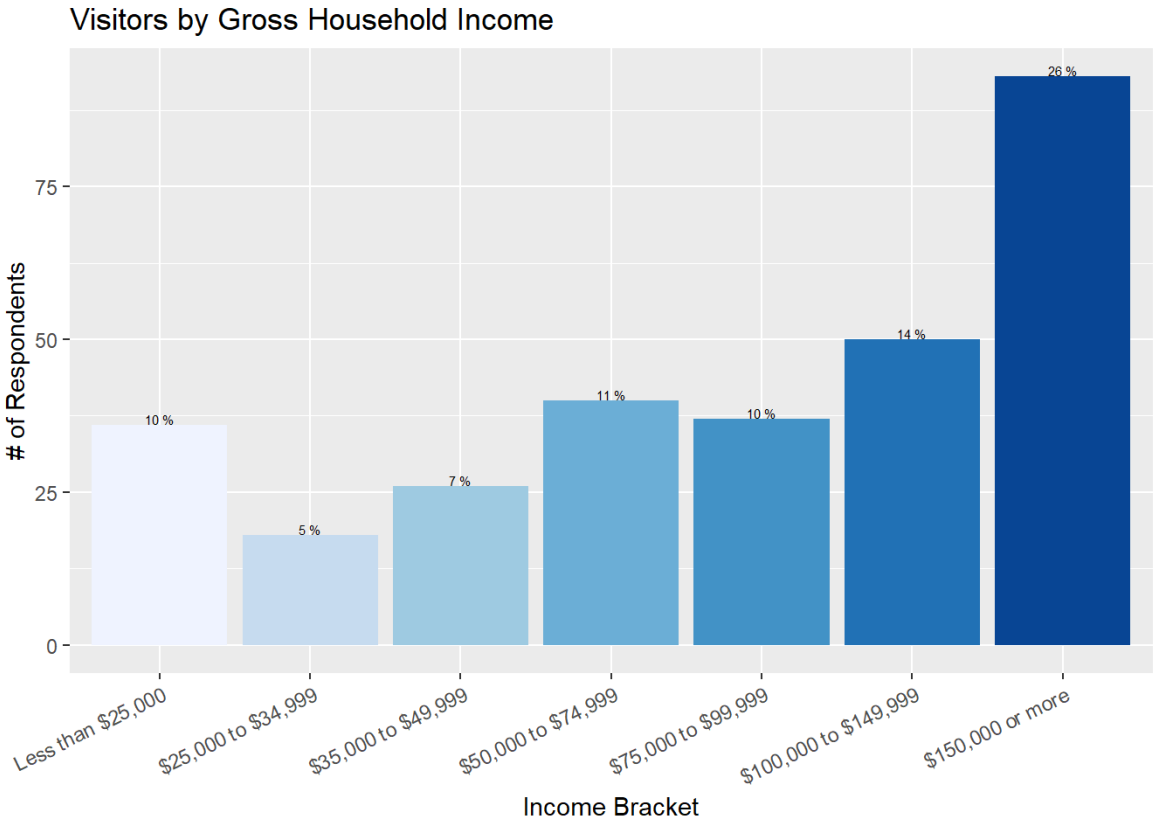
```
##
## data:  observed
## X-squared = 88.114, df = 8, p-value = 1.122e-15
```

Other race/ethnicity identities, as self-described by respondents.

Survey	Other Identities	Count
In Season	Carribbean	1
In Season	Jewish	2
Post Season	Filipino	1
Post Season	Italian-American	1
Post Season	Race is a construct.	1



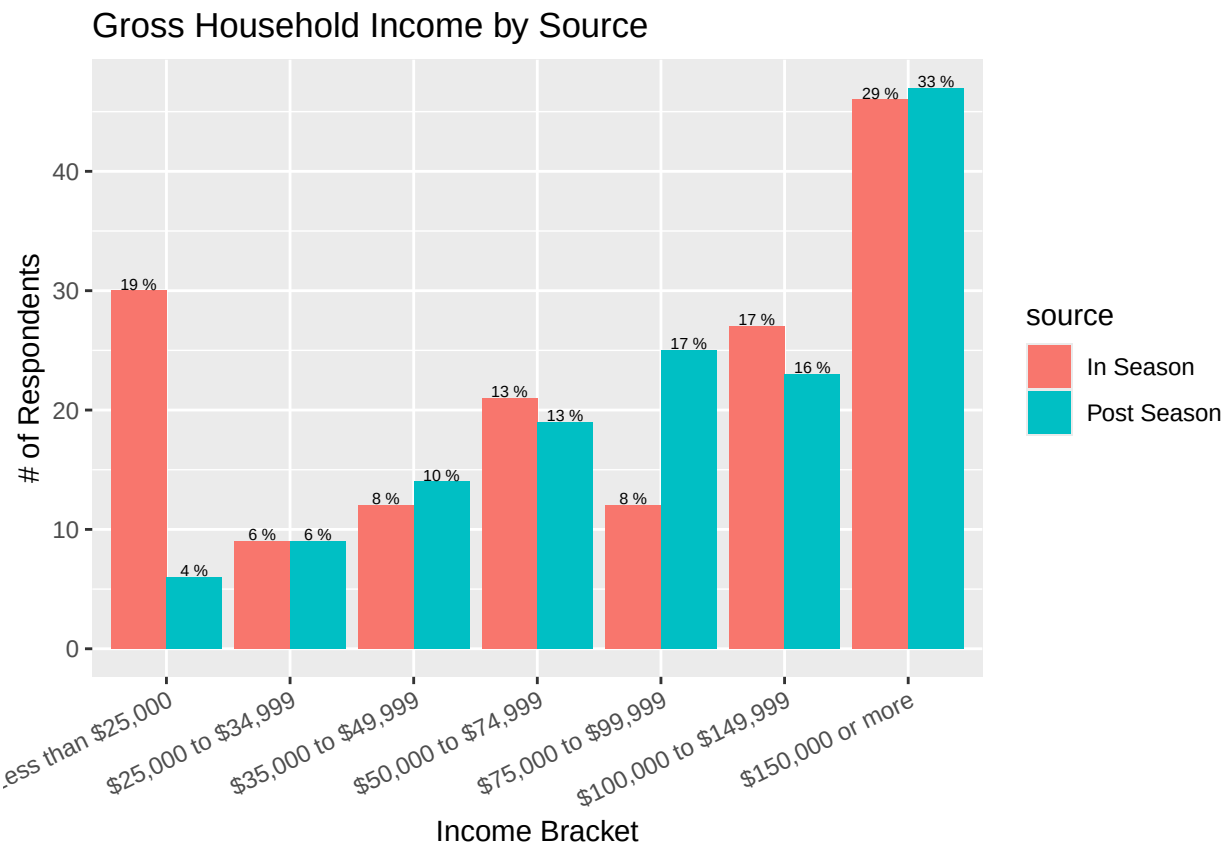
Educational Level



Gross Household Income

Grouped by source

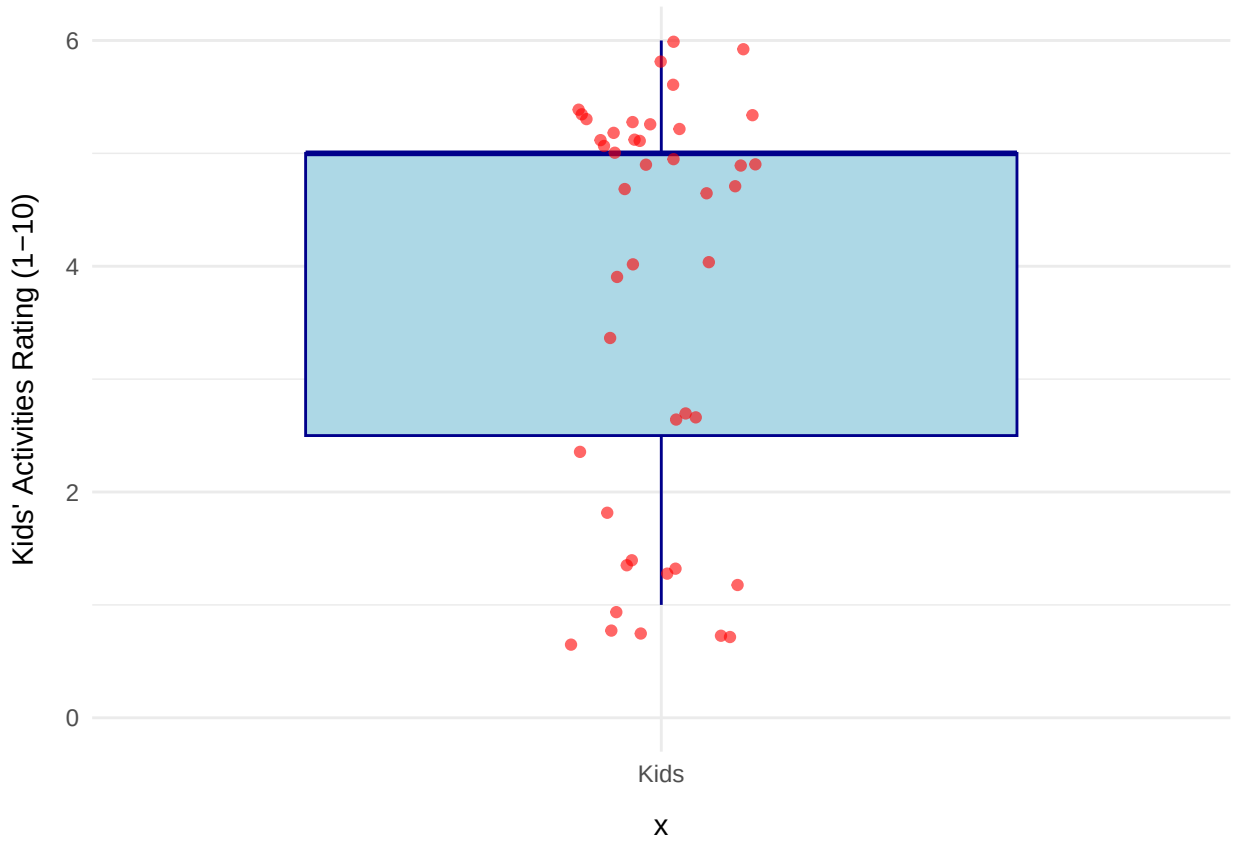
More individuals making under 25,000 were most likely present within the in-person source because younger populations were targeted by outreach.



Family Demographics There were 200 individuals who responded to the family question about satisfaction with kids' activities at A2SF.

Also, 11 respondents specifically mentioned Kids' Activities as one of their 3 favorite offerings.

Average satisfaction was **7.58** with a standard deviation of **2.50**.



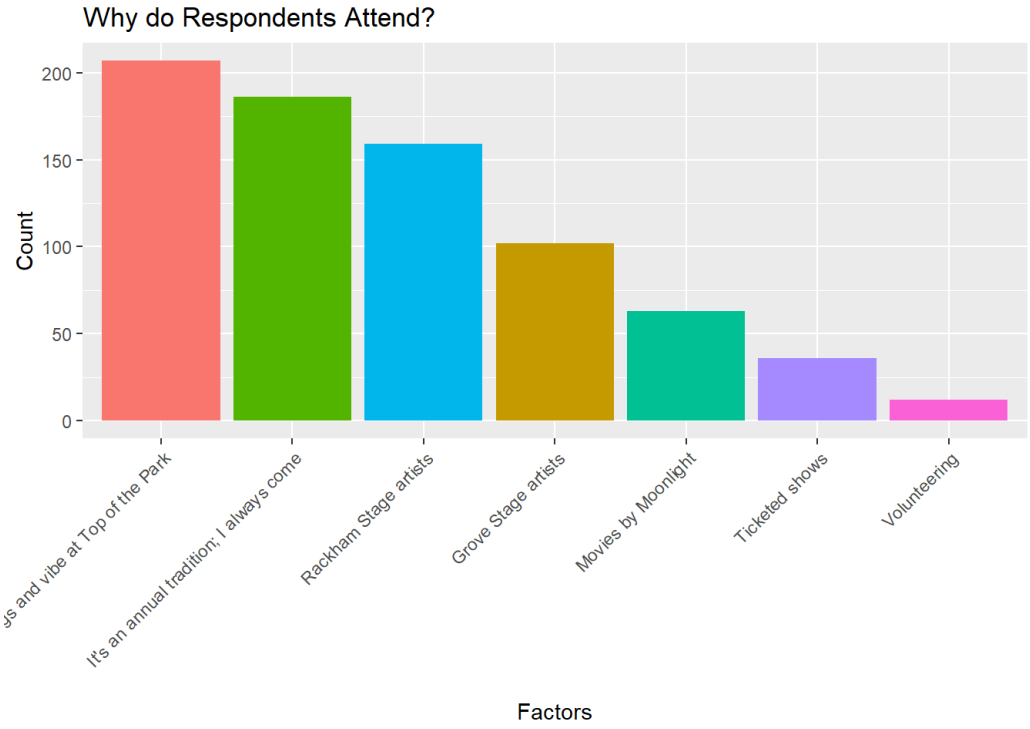
Of these respondents, the age groups indicate both adults and the children that used kids' activities participated in the survey alike. It would be interesting to dive further into these individuals' responses to see what they have to say in the short answer sections.

Age Group	Count
35 to 44 years	43
18 to 24 years	39
25 to 34 years	26
45 to 54 years	26
55 to 64 years	25
65 or older	25
Under 18 years	14

The following table shows individuals who mentioned kids' activities as one of their top activities and their short answer responses.

Age	Top 3 Favorite Offerings	Kids' Activities Satisfaction	Anything else to share?	Making it easier to participate	How did you hear?
Under 18 years	Music at Top of the Park, Movies by Moonlight, Kids' Activities (under age 11) in KidZone	7	Nope :)	Location	From Naya Shalom Loftus
35 to 44 years	Special Outdoor Attractions (circus, street performances), The Grove Beverage Garden, Kids' Activities (under age 11) in KidZone	10	N/a	N/a	Been attending for years
65 or older	Music at Top of the Park, Special Outdoor Attractions (circus, street performances), Kids' Activities (under age 11) in KidZone	5	No	Nothing	Friend
45 to 54 years	Music at Top of the Park, Special Outdoor Attractions (circus, street performances), Kids' Activities (under age 11) in KidZone	10	Stop using diesel generators. Gross. It's 2024!	Don't have mandatory fields for voluntary surveys	Live here
18 to 24 years	Retreat Classes on Power Center lawn (yoga, tai chi, etc.), Kids' Activities (under age 11) in KidZone	5	N/A	Better Mexican food.	The university of Michigan
35 to 44 years	Music at Top of the Park, Retreat Classes on Power Center lawn (yoga, tai chi, etc.), Kids' Activities (under age 11) in KidZone	8	Wish there were more food options, vendors, and seating. I wish it provided opportunities for local/independent artists (but not Art fair prices-too expensive for most).	Getting the schedule in advance to carve out time to go	Unsure
25 to 34 years	Music at Top of the Park, Special Outdoor Attractions (circus, street performances), Kids' Activities (under age 11) in KidZone	7	Love it!	I don't really even know about others beyond TOP	Branding
55 to 64 years	Music at Top of the Park, Food Selections, Kids' Activities (under age 11) in KidZone	10	Love David Zinn	Will attend	Onlune

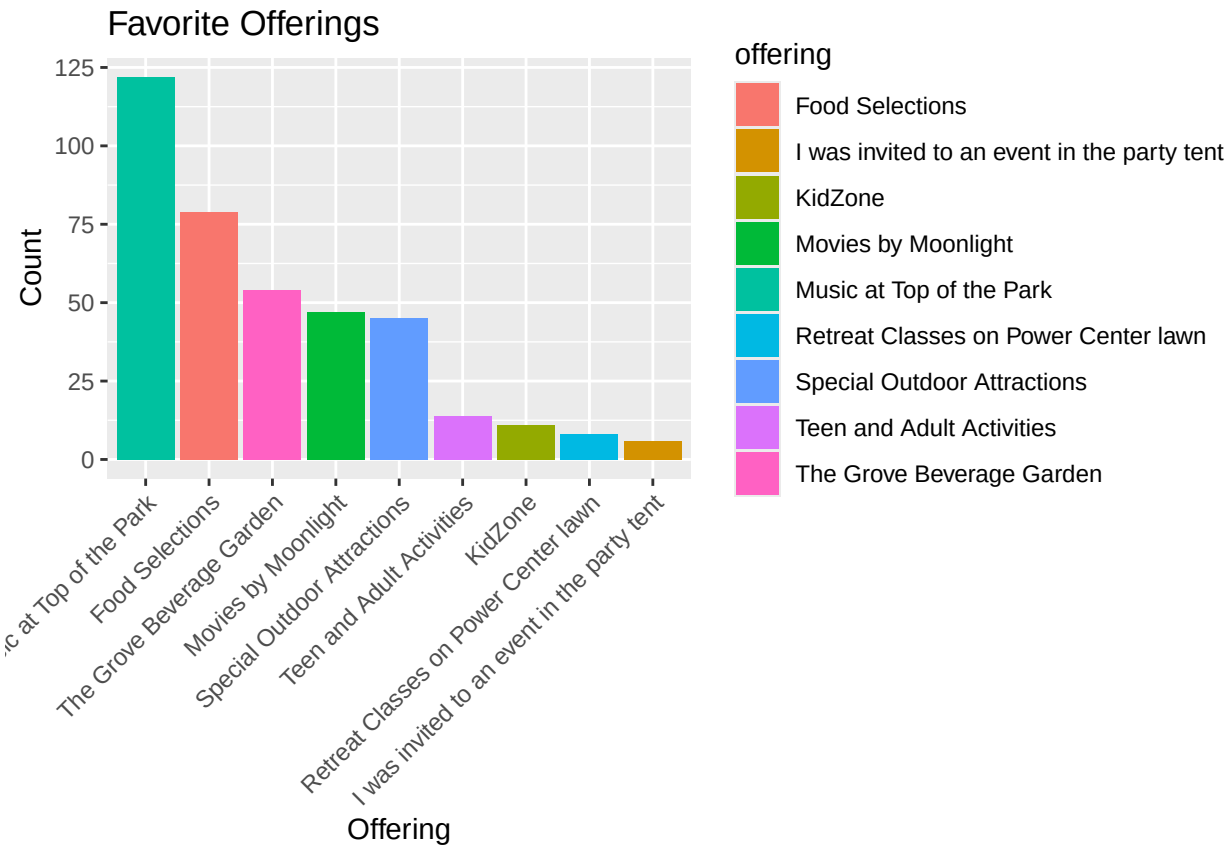
Age	Top 3 Favorite Offerings	Kids' Activities Satisfaction	Anything else to share?	Making it easier to participate	How did you hear?
45 to 54 years	Special Outdoor Attractions (circus, street performances), The Grove Beverage Garden, Kids' Activities (under age 11) in KidZone	10	We love these weeks! It's part of what makes A2 magical!! THANK YOU!!	Earlier events	Family
18 to 24 years	Kids' Activities (under age 11) in KidZone, Teen and Adult Activities (age 11+) in the Annex	10	N/a	Free parking and more games for the kids	Saw it
45 to 54 years	Special Outdoor Attractions (circus, street performances), The Grove Beverage Garden, Kids' Activities (under age 11) in KidZone	9	Expand fenced in area for adult beverages so adults can have a drink while on the lawn or in front of the stage	N/a	Returning attendees

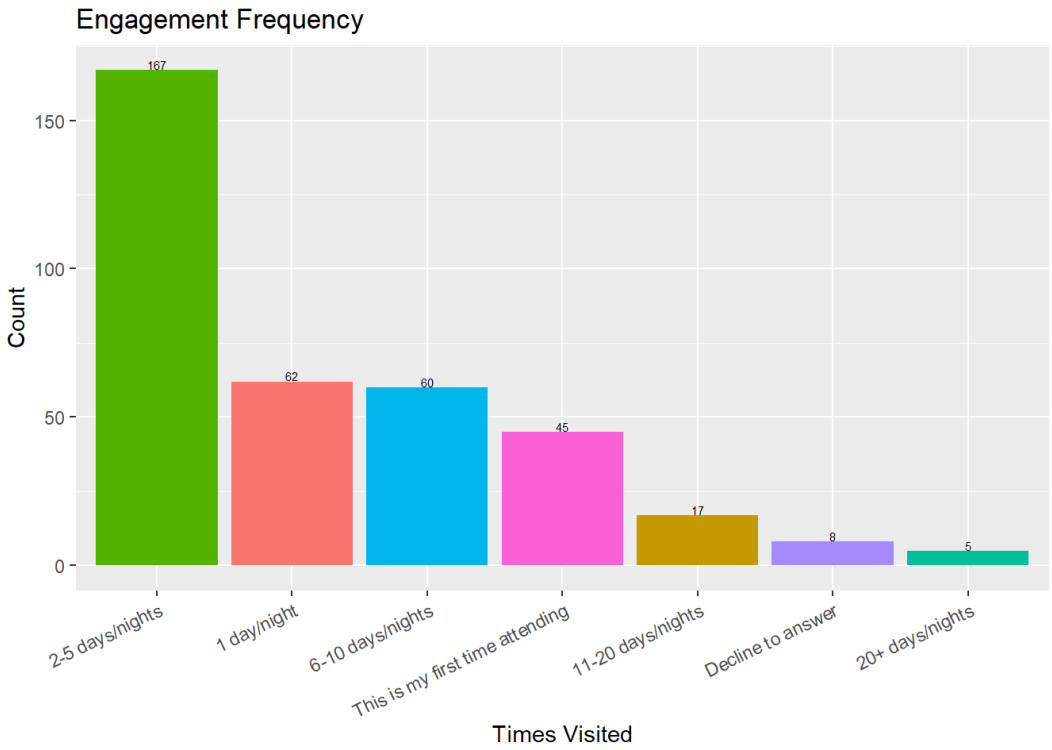


Programming Analysis - Favorite Offerings

All favorite offering counts by respondents.

Offering	Frequency	Percent (%)
Music at Top of the Park	122	30.58
Food Selections	79	19.80
The Grove Beverage Garden	54	13.53
Movies by Moonlight	47	11.78
Special Outdoor Attractions	45	11.28
Teen and Adult Activities	14	3.51
KidZone	11	2.76
Retreat Classes on Power Center lawn	8	2.01
I was invited to an event in the party tent	6	1.50
CHELAS	1	0.25
Can Chelas be allowed to sell orchatas?	1	0.25
Great overall vibe	1	0.25
Gregory Alan Isakov!!!! and Mon Rovia!	1	0.25
Juggling	1	0.25
Reflex	1	0.25
Spoken word night	1	0.25
This survey	1	0.25
Ticketed offerings have fallen off significantly in volume and alignment with my intetrests.	1	0.25
Tuesday's Tastings	1	0.25
loved the silent films	1	0.25
meet up with friends, people I have not seen for a long while; listen to god entertainment, dance outdoors	1	0.25
trivia!	1	0.25


Favorite Offerings
Count
125
100
75
50
25
0
Offering
ic at Top of the Park
Food Selections
The Grove Beverage Garden
Movies by Moonlight
Special Outdoor Attractions
Teen and Adult Activities
KidZone
Retreat Classes on Power Center lawn
I was invited to an event in the party tent
offering
Food Selections
I was invited to an event in the party tent
KidZone
Movies by Moonlight
Music at Top of the Park
Retreat Classes on Power Center lawn
Special Outdoor Attractions
Teen and Adult Activities
The Grove Beverage Garden



How many times did respondents attend Top of the Park this season?

Other Satisfaction Ratings Welch Two Sample T test is met, meaning there is no difference in responses between in-person and online survey respondents.

Patrons' likelihood to donate (1-10) is **6.43** with a standard deviation of **3.28.**

Patrons' likelihood to recommend to family or friends (1-10) is **9.07** with a standard deviation of **1.76.**

Patrons' likelihood to attend again in the future (1-10) is **8.96** with a standard deviation of **1.81.**

Patrons' satisfaction with music (1-10) is **8.19** with a standard deviation of **1.87.**

Patrons' satisfaction with A2SF's overall trajectory (1-10) is **8.06** with a standard deviation of **2.05.**

Future Recommendations

Surveying visitor race and ethnicity

Moving forward, separate race and ethnicity into 2 distinct questions:

1. "Do you/do you not identify as of Hispanic, Latino, or Spanish origin?"
2. "With what racial identities do you identify? Select all that apply."

The aim of this question is to assess ethnicity as is currently accepted per U.S. Census. It certainly is not the most comprehensive definition, but it is the current accepted method. This was established in 1997, so it is fairly out of date and may change in the coming years.

Per OMB, there are five categories for data on race:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White

Instead of allowing a multiracial category, the OMB adopted a standard of allowing respondents to select one or more races when they self-identify.

Meanwhile, there are two categories for data on ethnicity:

- Hispanic or Latino or Spanish Origin
- Not Hispanic or Latino or Spanish Origin

Breaking these questions down this way allows for ease of analysis in future visitor surveys.