DATA-DRIVEN APPROACH TO VILLAGER POPULARITY WITHIN ANIMAL CROSSING NEW HORIZONS

W200 Spring 2021 | Section 12

Project 2

Team Members: Karl Eirich, Tina Hsieh, Kasha Muzila

Overview

In the year 2020, Nintendo released a new game called "Animal Crossing: New Horizons." During the first few months of its release, this game became very popular, primarily due to the pandemic and its unique way of creating a community of people based on characteristics within the game. These characteristics range from villager popularity to catching rare items. However, we will only focus on villager popularity and correlations that could cause the villager's popularity for this analysis.

A villager is a non-player character (NPC) within the game that interacts with the player, items, or events. There are about 400 villagers that the player can have, and they range from different animal species and personalities. Figure 1 gives an example of what a villager could look like.

Figure 1 - Player interacting with a villager (NPC) within Animal Crossing: New Horizons

Many public forums give rankings per month on which villager is the most popular. However, these public forums mostly go off of survey-based data in order to do these rankings. They do not look into other causes on why a particular villager would be more popular over another. Our analysis will explore in-game data and scraping Twitter to determine if other correlations cause a villager's popularity.

Key Terms

- Popularity -- The amount that the villager is talked about, whether favorable for that villager or unfavorable.
- Villager -- A non-player character (NPC) within the game that interacts with the player, items, or events.
- In-game data -- Data that was directly pulled from Animal Crossing: New Horizons

Focusing Questions

Our focusing questions divide our report into 2 sections: in-game information that could influence villager popularity and a deeper look into villager popularity from Twitter and in-game events.

- What in-game information should we know about villagers and that could influence their popularity?
- Are there any correlations between game events and villager popularity by scraping tweets about villagers from Twitter?

Dataset

Our primary dataset used within our analysis is a frequently updated public spreadsheet that houses in-game data. It includes 35 sub-datasets, all in CSV format. The following sub-datasets are what we focused on for in-game information that we should know about villagers, and that could influence their popularity:

- Villagers.csv
    - Summary data on the background information of villagers such as name, species, gender, personality, hobby, and birthday
- Recipes.csv
    - Summary data on all recipes within the game, such as where and when to obtain recipes, materials needed to make the recipe, recipe value, and category it falls under.
- Housewares.csv
    - Summary data of all items that can only be used within the player's house, specifically, items that can only be placed on the floor. This data includes information such as the name of the item, color, value, and where and when to obtain it.
- Wall-mounted.csv
    - Summary data of all items that can only be used within the player's house, specifically, items that can only be placed on the walls. This data includes information such as the name of the item, color, value, and where and when to obtain it.
- Miscellaneous.csv
    - Summary data of all miscellaneous items that do not fall within a specific category. These items can be placed anywhere within the game. This data includes information such as the name of the item, color, value, and where and when to obtain it.
- Tools.csv
    - Summary data of all items that can be used as a tool within the game, such as an ax to cut down trees, magic wands to change the player's outfits, and a fishing rod to catch fish. This data includes name, color variation, amount of uses, value, and where and when to obtain the item.
- Dress-up.csv
    - Summary data of all body clothing that villagers and the player can wear within the game. This data includes name, color variation, value, typical gender of clothing type, clothing style, and where and when to obtain clothing items.
- Headwear.csv
    - Summary data of all headwear that villagers and the player can wear within the game. This data includes name, color variation, value, typical gender of clothing type, clothing style, and where and when to obtain clothing items.
- Wallpaper.csv
    - Summary data of all wallpapers that can be used to decorate the player's house. This data includes name, value, color, type, and where and when to obtain the wallpaper.
- Seasons and Events.csv
    - Summary data of all in-game seasonal events. Seasonal events within the game unlock special items that the player can collect to use for recipes, to decorate, or as a collector's item. This data includes the season's name and type of season when it occurs during the calendar year.

Our supplemental dataset used within our analysis is a Twint based query. Twins is a Twitter scraping tool that allows for scraping tweets from Twitter without using Twitter's API. With its search operators, we pulled a total of 596K popular tweets since March 2020 with hashtags #AnimalCrossing or #ACNH.

Data Validation and Preparation

Since there are 397 villagers and each villager has its personality, we can validate our data by comparing how players can obtain recipes from different villager personalities. We verified that the villager and recipe data matched the villagers' personality values after joining datasets and did not find any other issues with our source data.

Villager Personality	Villager Count
Normal	61
Lazy	61
Cranky	55
Snooty	55
Jock	55
Peppy	51
Smug	35
Big Sister	24

Recipe Source	Recipe Count
All	67
Peppy	35
Cranky	35
Lazy	35
Smug	34
Normal	34
Big Sister	34
Jock	34
Snooty	33

Table 1 - Example Validation Tables. Villager personality vs. villager count (left). Recipe source from villager personality vs recipe count (right).

From a preparation perspective, our main task was to analyze and/or merge CSV's that could explain why a villager could be popular. Most of the data were merged with Villagers.csv to analyze any additional in-game information correlated with a villager, such as recipes or items that they can give to the player throughout the game.

We also did a sanity check on the Season and Events data, and there were a total of 89 in-game events. We found some missing dates and duplicated event names with consecutive dates. We assumed that if the event name was the same, the event was on during that period. So, we performed some data cleaning, split the event date column, which contains start and end dates, into two columns with datetime format for further analysis. See Table 2 of the count of in-game events as well as an example of the cleaned up data on in-game events.

```
1 dfev.Type.value_counts()
```

```
Nook Shopping event     39
Special event           18
Zodiac season           12
Crafting season          8
Shopping season          8
Basegame event           4
Name: Type, dtype: int64
```

	Name	Type	Year	Dates	Start	End
0	young spring bamboo	Crafting season	2020	February 25 – May 21	February 25	May 21
1	cherry-blossom petals	Crafting season	2020	April 1 – April 10	April 1	April 10
2	summer shells	Crafting season	2020	June 1 – August 31	June 1	August 31
3	acorns and pine cones	Crafting season	2020	September 1 – December 10	September 1	December 10
4	mushrooms	Crafting season	2020	November 1 – November 30	November 1	November 30

Table 2 – Count of all type of in-game events (left). Output of cleaned up data on in-game events (right).

Exploratory Questions

Section 1: What in-game information should we know about villagers and that could influence their popularity?

In-Game Information About Villagers

We wanted to explore in-game data that would give us more background information about villagers we might need to know before scraping Twitter for any more correlations in their popularity with in-game events. The first thing that we looked into was how many villagers and villager species were there in the game. We found that there are a total of 397 villagers and 35 species (see Figure 2).

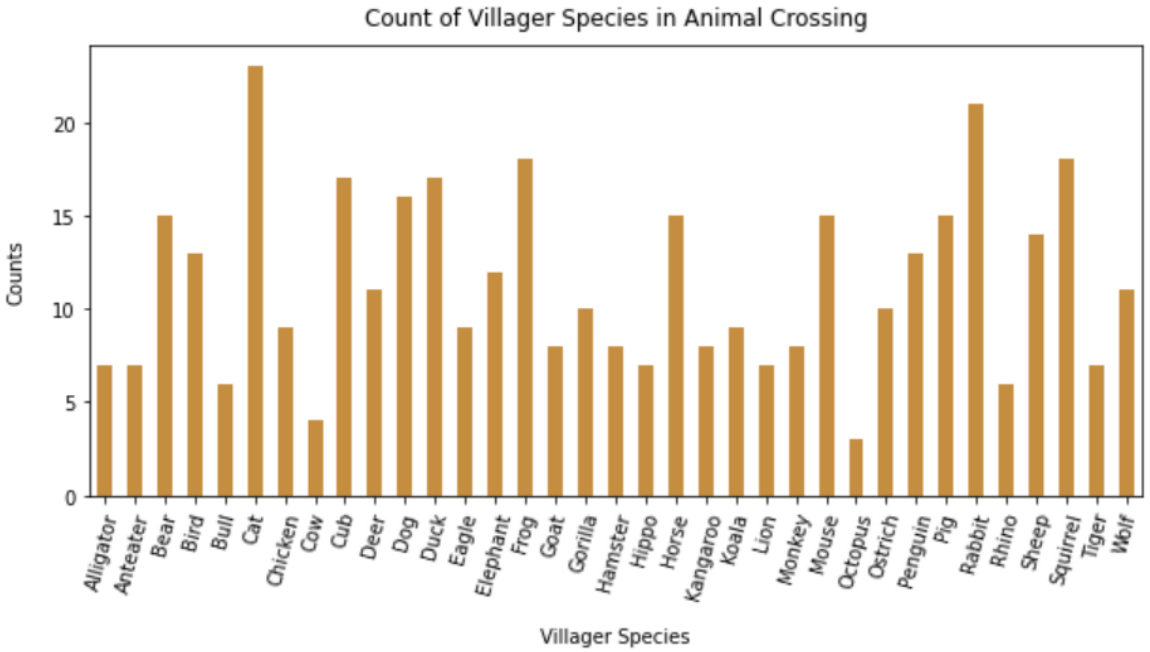

Figure 2 - Number of villager species within Animal Crossing: New Horizons.

We can see from Figure 2 that there are several species with a higher count than others. This could correlate with villager popularity. However, we do not have direct access to the game's server in order for us to verify this. For now, we will assume that these higher counts were due to the game developer's decision to add more of one species over others.

Historically, Nintendo has made an emphasis on Animal Crossing's villager personalities and their birthdays in order for the player to become more immersed in previous games. This makes players feel like the villagers they take care of are unique to the player, such as celebrating the villager's unique birthday and experiencing their clever prompts when interacting with them. With Animal Crossing: New Horizons, we wanted to know the villager's genders, birthdays, and personalities to understand how this might relate to villager popularity.

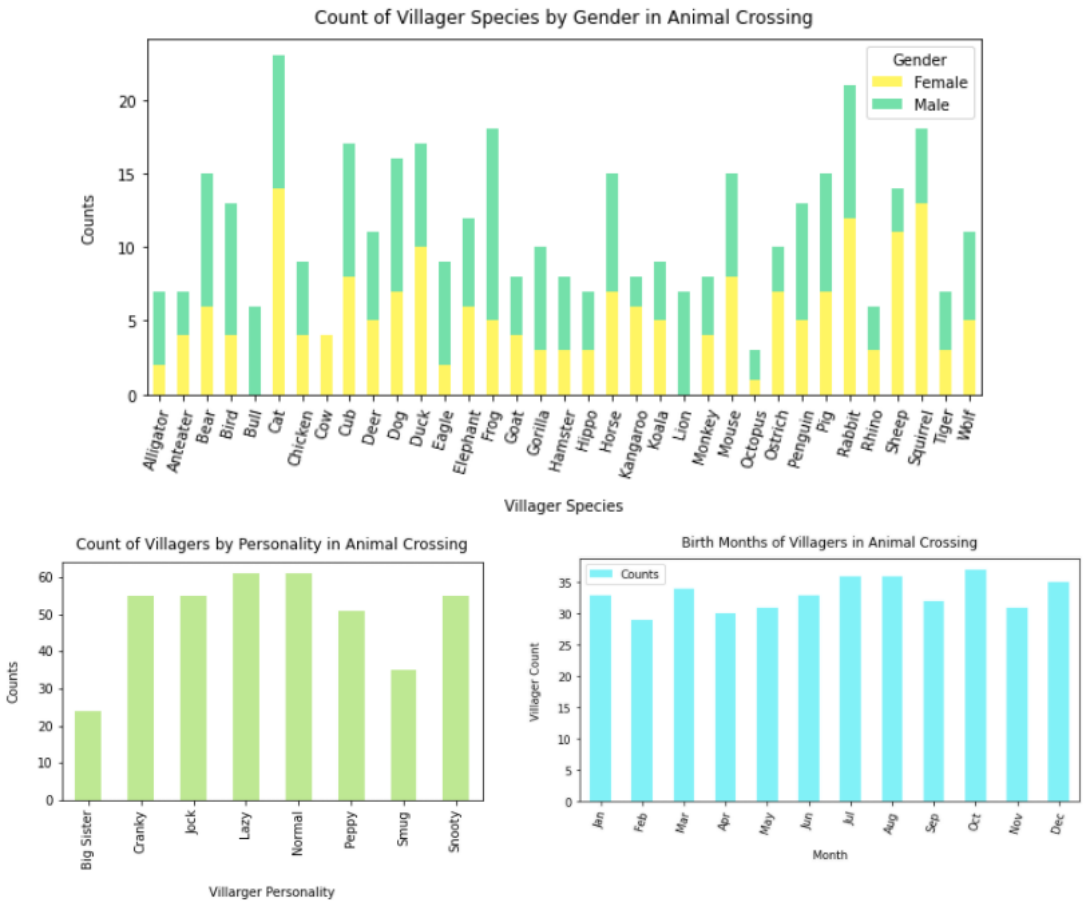

Figure 3 - Villager species and how gender count (top). Count of the different types of villager personalities (bottom left). Number of villagers' birthdays that are in a specific month of the year (bottom right).

Figure 3 shows 191 female species and 206 male species, that there are less smug and big sister personalities and that birthdays seem to be well distributed. This gives us enough background information that we would need to know about the villagers themselves before seeing any correlations in their popularity.

Villagers vs. In-Game Items

The currency used in the game is called "bells," which the player will accrue a certain amount of debt to pay back. Therefore, the player needs to come up with multiple streams of income. One of the game features that the player can use to help pay off their debt is items they can receive from villagers valued in "bells." We wanted to analyze which villager personality would give the player the most value in items to see if that makes a particular villager more popular.

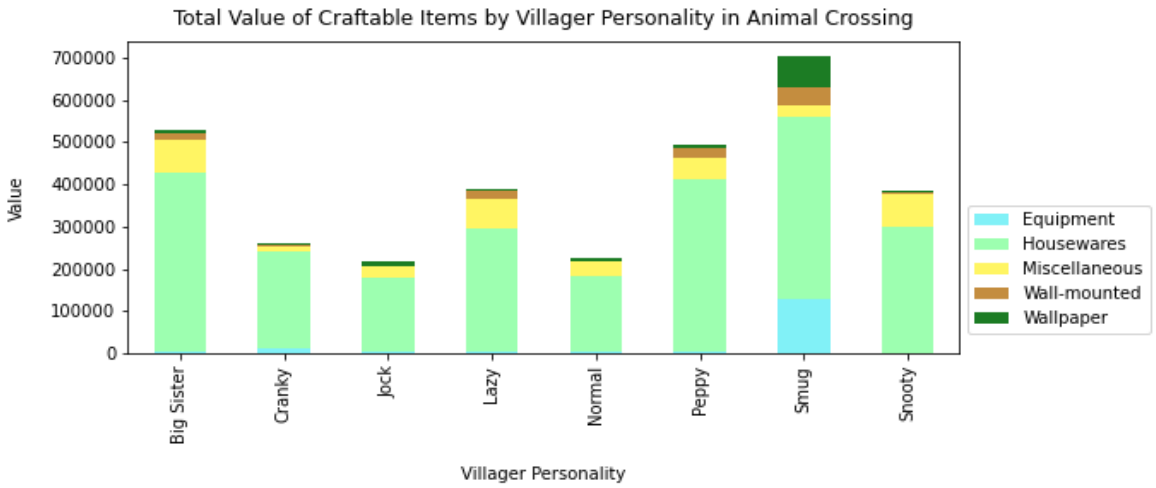

Figure 4 - Total value of craftable items that each villager personality can give to the player.

We can see from Figure 4 that the smug villager personality can give the most valued items to the player to help them pay off their debt in the game. This could be a correlation as to why smug personalities could be more popular than others. However, later we will discuss which villager and their personalities are popular based on the number of tweets to see if it can provide more evidence to this correlation.

Section 2: Are there any correlations between game events and villager popularity by scraping tweets about villagers from Twitter?

Amount of Tweets Scraped from Release to Current Date

Before we dove into specific Twitter areas, we wanted to understand how much data we could scrape by looking at hashtags #ACNH and #AnimalCrossing. This will give us a good idea if there is enough data for us to look at in order for us to do proper analysis.

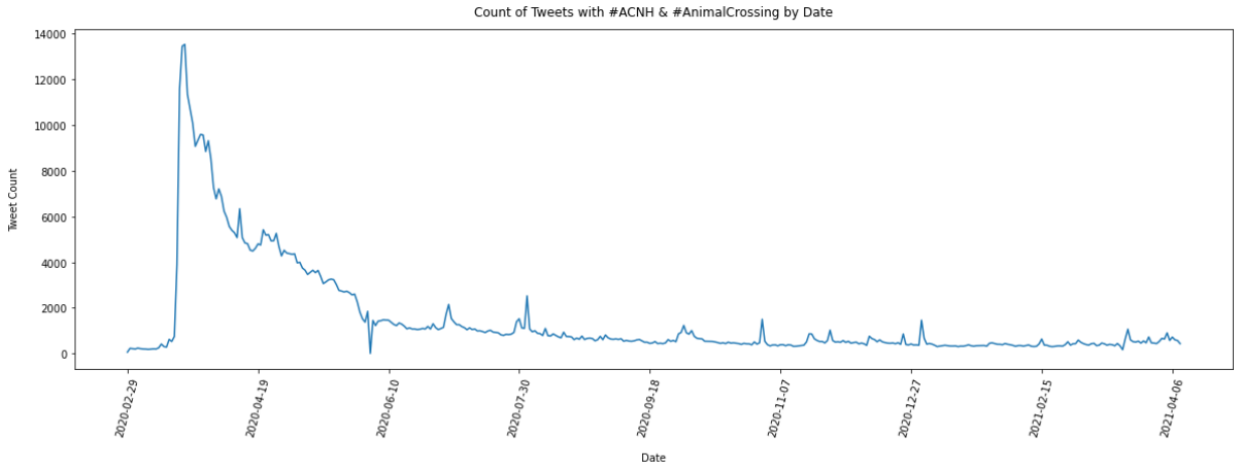

Figure 5 - Number of tweets found on Twitter with hashtags #ACNH and #AnimalCrossing. The timeline is from the game's release up to the current date.

We can conclude from Figure 5 that we have about 596K tweets, giving us sufficient data to use in order for us to find correlations in villager popularity versus in-game events.

Villager Popularity versus Twitter Data

Players often tweet about their favorite villagers by mentioning the villagers' names or posting images. We decided to focus on tweets with villagers' names to understand their popularity. Going through some initial

exploratory analysis and trial rounds helped us refine the rules, eliminate noises and get relevant results to answer our questions.

	Villager	TweetCt
298	Raymond	1960
228	Marshal	1308
329	Sherb	742
186	Judy	611
116	Dom	595
...	...	...
350	T-Bone	6
395	Rilla	6
160	Greta	5
54	Boyd	4
220	Maelle	2

397 rows × 2 columns

Table 3 - Number of tweets scraped per villager. The higher the number of tweets, the more popular that villager.

Based on Table 2, we can see that the top 5 villagers are the following: Raymond (smug personality), Marshal (smug personality), Sherb (lazy personality), Judy (snooty personality), and Dom (jock personality). Since Raymond and Marshal are both smug personality types, the total value of craftable items that smug personalities can give to the player from Figure 4 can provide more evidence as to why the smug personality types are more popular than others we see in Table 2.

We then wanted to analyze the top 5 villager tweets to see if they had any abnormalities that could have caused them to be more popular during certain times after the game's release. The time-series graphs offered a clear visualization of the trends based on tweets' count with these top 5 villagers mentioned.

Figure 6 - Number of tweets scraped for the top 5 villagers. The number of tweets were scraped from the game's initial release to the current date.

After analyzing Figure 6, we can see that all of the top 5 villagers had many tweets about them during the first few months of the game's initial release, and then the tweets started to decrease. However, we can see some spikes for Raymond and Marshal around October last year and some spikes for Sherb and Judy around late January and early March this year. However, Dom seemed to not have any significant spikes since the game's release. We will conclude that Dom's popularity is subjective and varies by time in the player's perspective.

To understand what could have caused these spikes, we analyzed all of the in-game season events by creating a Gantt chart using the Plotly library. Plotly is a popular interactive visualization tool, and its *create_gantt* function enabled us to track the timeline and duration of different events. There are multiple event types in Animal Crossing, and the main ones are nook shopping season, crafting season, and special events. With color codes, we can quickly identify the event types on the timeframe.

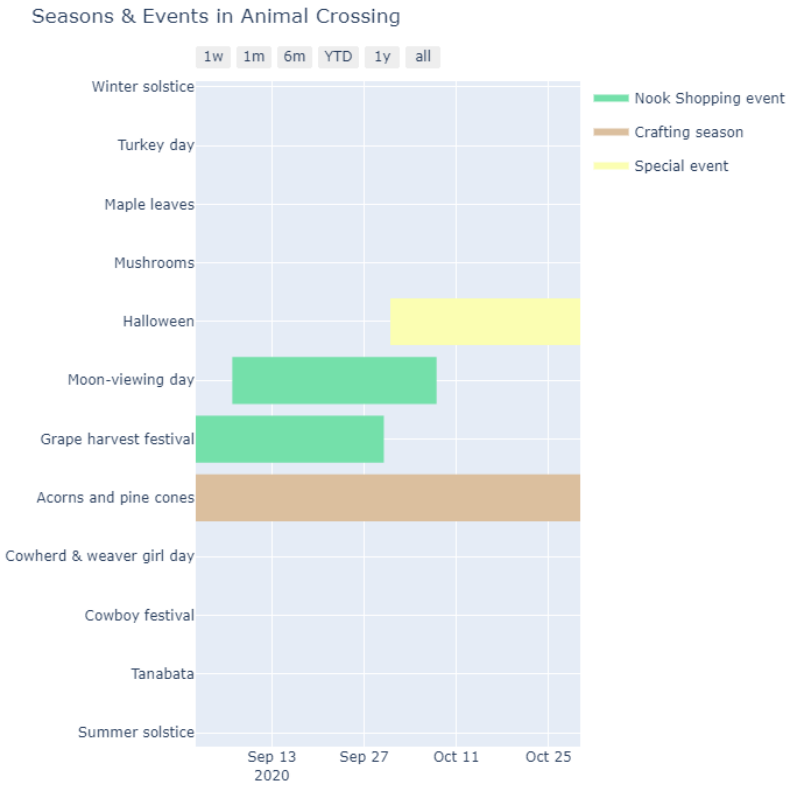

Figure 7 - Gantt chart of all of the seasons and events within Animal Crossing: New Horizons. The timeline is between September and October of 2020.

In Figure 7, we can see that the only events happening in the game around the time that Raymond and Marshal had tweet spikes were Halloween (special event), Moon-Viewing Day (crafting season), Grape Harvest Festival (crafting season), and Acorns and Pine Cones (crafting season). Because this was around the holiday season, this could be an indicator of why they became more popular at that time. However, this is still inconclusive.

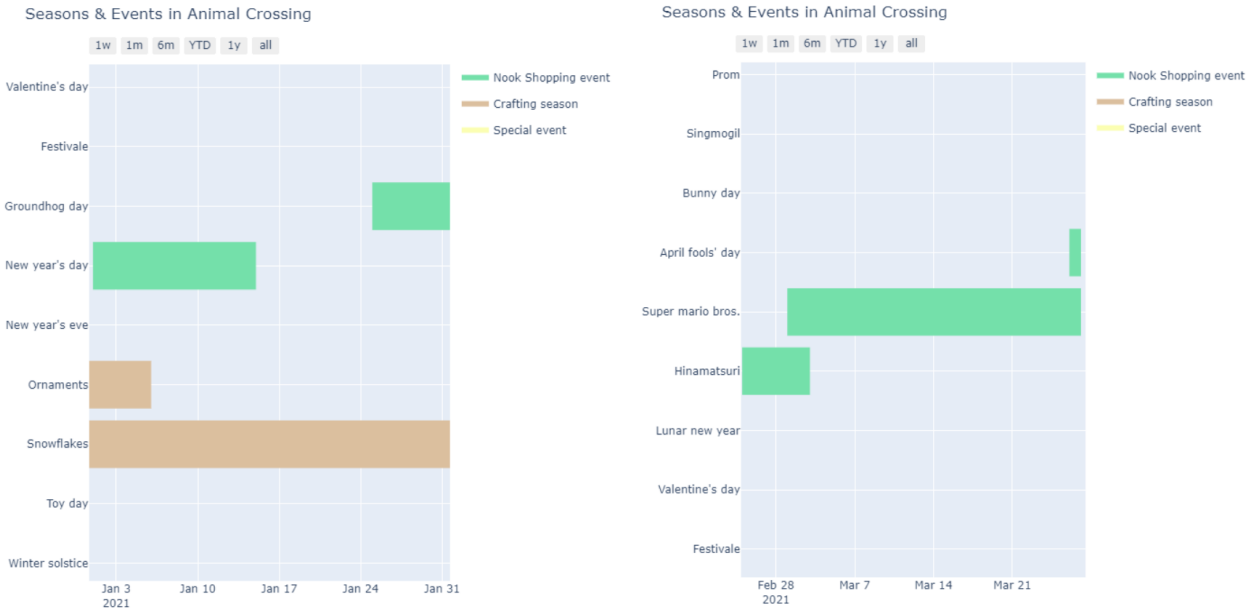

Figure 8 - Gantt chart of all of the seasons and events within Animal Crossing: New Horizons. The timeline is the month of January (left) and early March (right) of 2021.

If we compare the tweet spikes we see from Figure 6 on Sherb and Judy, then compare it with the Gantt charts we generated, we can see that the in-game events were Snowflakes (crafting season), Ornaments (crafting

season), New Year's Day (Nook Shopping event), Groundhog Day (Nook Shopping event), Hinamatsuri (nook shopping event), and Super Mario Bros. (nook shopping event). There are several events here that could have caused Sherb or Judy to be popular around this time. However, we cannot see any leading cause for their popularity based on comparing the Gantt charts.

When drilling down in the data to analyze the dates of Raymond, Marshal, Sherb, and Judy's spike, it reveals that each of their respective spikes occurred on their birthday. Looking at Figure 9, we can see Raymond has a 6x increase in tweets on his birthday, October 1st. There is a similar increase for Marshal and Sherb, who both received an 8x increase in tweet traffic on their birthdays, and Judy saw a ~4.5% increase in tweet mentions on her birthday. However, the reason as to why Marshal and Sherb's tweets were more than Raymond and Judy's is inconclusive. We will assume that their magnitude of increase is subjective to the player.

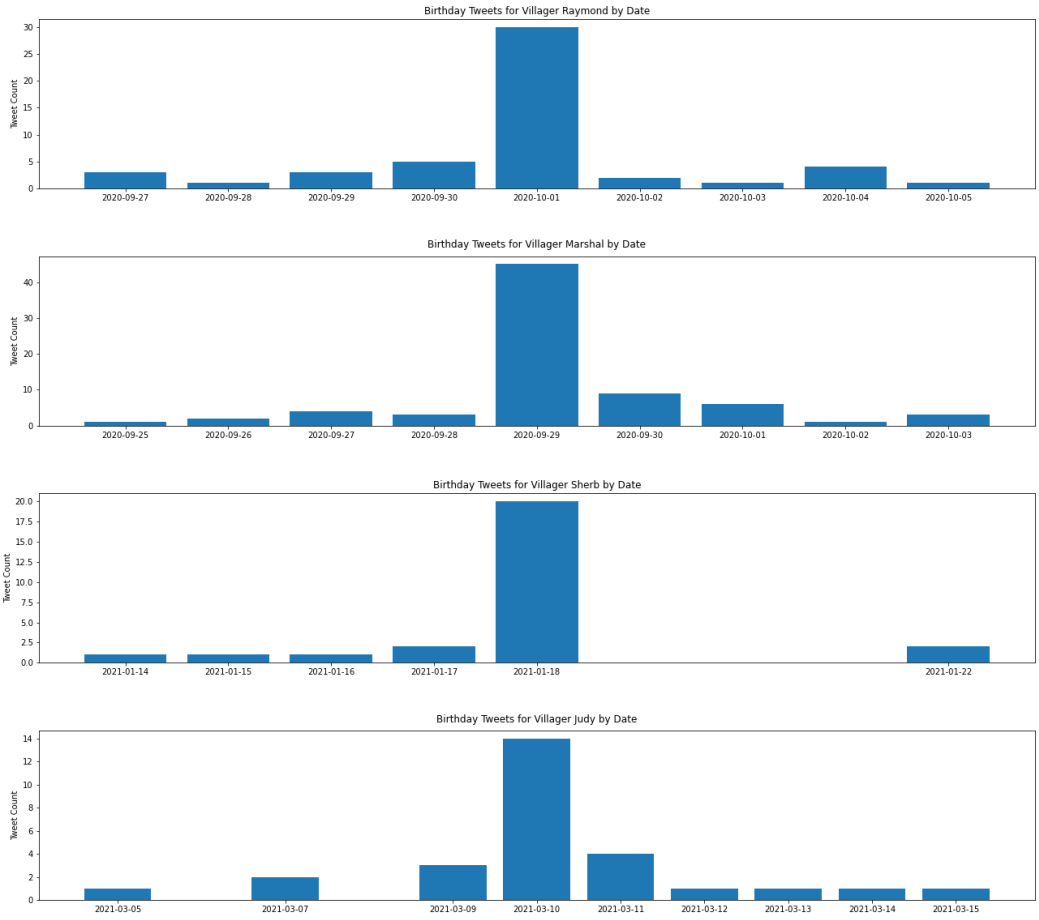

Figure 9 - Number of tweets for the top 4 villagers' birthdays.

Conclusion

Animal Crossing: New Horizons became a cultural touchstone during the pandemic's early months and has been a gaming staple since its release. Exploring the tweets mentioning Animal Crossing's most popular villagers consistently showed a spike on the villager's birthday. Though their birthday is not a perfect indicator of popularity, the fifth most popular villager did not increase mentions on their birthday. All other holidays did not have a demonstrable effect on the villager's Twitter mentions. Many lists and polls regarding villager popularity seem to be subjective based on the community or author. Still, through Twitter analysis, we were able to identify the popular villagers and a consistent indicator of their popularity.