Animal Crossing New Horizons: A Comprehensive Analysis

Background, Intent, Motivation and Questions

Prepared By:
Jacqueline Skunda
**Tamara Qawasmeh**
Malini Varadarajan

Background: Animal Crossing New Horizons (ACNH) is Nintendo's social simulation game launched in March 2020, around the same time the COVID-19 pandemic began shutting down the world. Animal Crossing has been around for over 20 years, but this past year the game exploded in popularity with 31 million players because it was a great lockdown distraction. ACNH allows players to set up an alternate life on a pleasant island that is visited by interesting animal characters termed "villagers". Villagers are the animal friends that move in to the island as the title suggests. They have unique personalities and are the main attraction in the game. Players can fish, find fossils, catch bugs and other exotic sea creatures then sell them to Nook's Cranny for bells (ACNH currency). Players grow various plants, flowers and fruit trees on their island, making it their own, while sharing their island magic with other players that visit them.

Intent: Given that the game was so wildly popular this past year, we wanted to investigate some of the ACNH related data to paint a data driven picture of the game itself. We assembled ACNH datasets from multiple sources to understand the game at a high level. Analysis commenced by looking at the villagers in the game to examine what features drive the highly favored villagers. By examining Google search totals along with the ACP player polls we hope to see if the Google search totals are representative of player poll rankings and look at any disparities. We also wanted to examine the highest selling insects and fish in the game and look at their endangered species status to see if there was correlation between the game developers incentivizing the catching then selling of endangered species via higher bell rewards. We are interested in understanding how game developers have designed villagers personalities and the overall game during our data analysis.

Motivation: At a time when many of us were presented with Covid-19 datasets daily, we wanted to focus on the lighter side of the pandemic. A game like ACNH united several players and families on a virtual island where players found some much needed mental health healing on the zen islands. We ourselves are avid ACNH game players ourselves and took to it quite a bit during COVID.

Questions: [1] What can we learn about the game's creators and gamer preferences from the ACNH focused google searches? [2] Why is a Villager popular? Is it based on species/personality/gender? [3] Do certain personality traits enable popularity? [4] Are there any unintended ethical consequences as it pertains to the villagers and critters (fish and insects)?

Github

New Horizons

Welcome to Animal Crossing

1
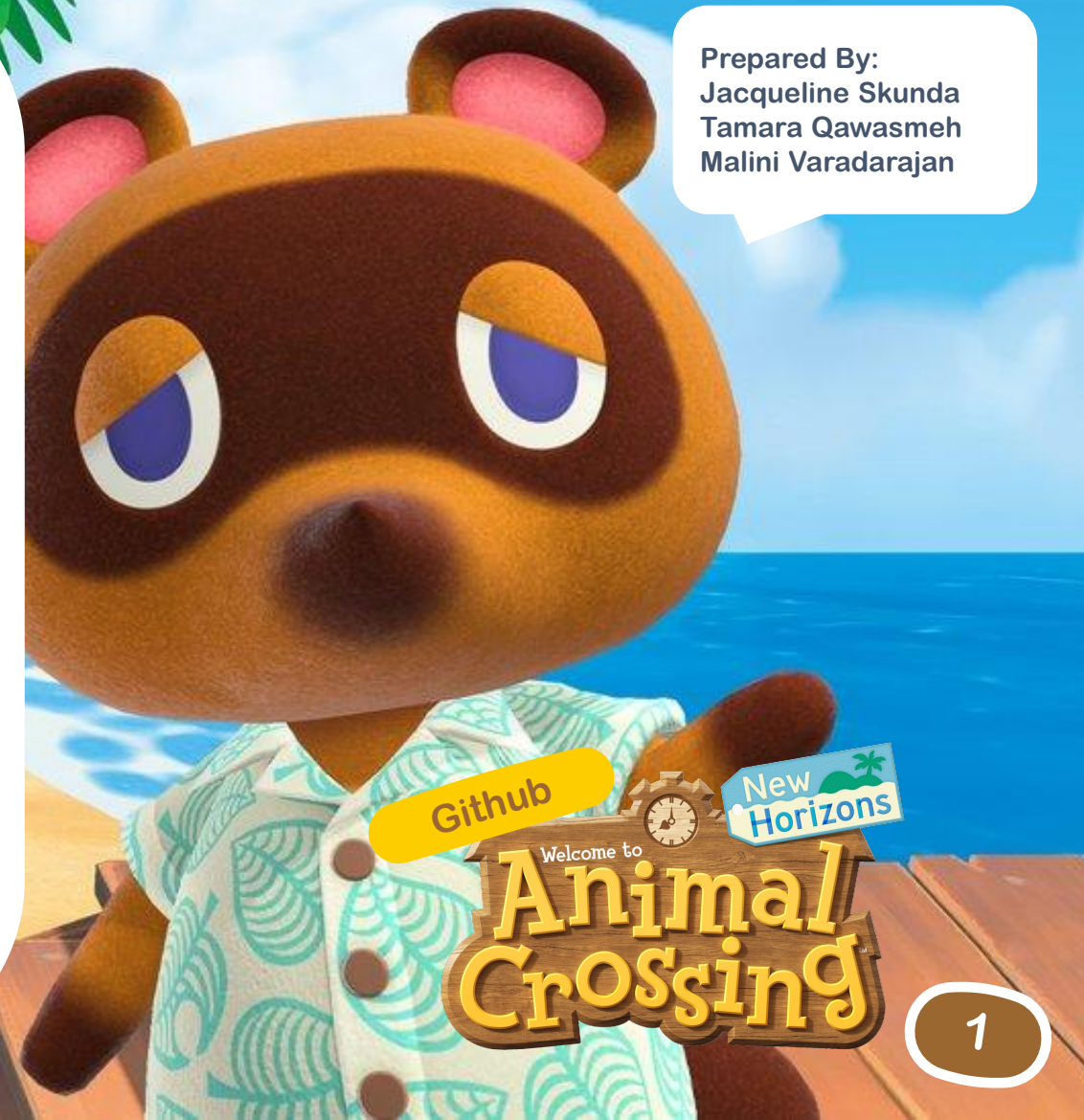
Animal Crossing New Horizons Comprehensive Analysis

Data Sources

Primary Dataset

Animal Crossing New Horizons Catalog

Description: curated data set about the various Animal Crossing life simulation game items, features, collectibles, houseware, fashion items, and villagers.

Size: 3.7 MB in total size, made up of 31 small csv files, we will only be considering the fish, insects and villagers csv files for this analysis which are 24.5 KB, 25.4 KB and 78.7 KB

Variables: name (villager and common name of critter), sell, personality, gender, species, where found

Format: CSV

Access: Download

Location: Animal Crossing New Horizons Catalog.

Secondary Data Sources

Google Trends API

Description: represents historical searches and number conducted on a global level based on parameter payload supplied of "villager name + Animal Crossing".

Size: 26.1 KB; 789 rows returned

Variables: query & value (returns dictionary of dataframes)

Format: CSV

Access & Location: Call via pytrends library python: Pytrends Docs

Time Frame: (Game release date) 2020-03-20 - 2021-11-24

ACP Player Poll Results

Description: Multiple monthly Excel files containing the votes per villagers from a public poll.

Size: [2021] 1.66 MB, [2020] 3.94 MB

Variables: villager(name), tally(votes)

Format: CSV

Access: Download

Location: Villager Popularity Tier List

Time Frame: June 2020 - October 2021

Nookipedia

Description: Animal Crossing version of Wikipedia. Info on villagers, fish, insects and various island features.

Size: ~37.8 KB; 160 rows returned

Variables: ACNH critter to scientific name

Format: HTML returned list per item

Access: Beautiful Soup (BS4) to scrape Nookipedia by critter name to fetch ACNH listed scientific name in <i> tag.

Location: www.nookipedia.com

Time Frame: N/A

Wikipedia

Description: Online encyclopedia written/maintained by community of volunteers through a model of open collaboration. (Free to use)

Size: 37.8 KB; 132 rows returned

Variables: Scientific fish & insect names, IUCN Conservation status

Format: Returned dataframe per item

Access: pd.read_html, base url + scientific name

Location: www.wikipedia.org/wiki/

Time Frame: N/A

Animal Crossing New Horizons Comprehensive Analysis

Google Trends API

Goal: Find the number of times each villager was searched and pertaining "search" to see global popularity for each villager historically. Google Trends was utilized via the Pytrends library.

1. Cleaned villager names (str.replace) and created a list that adds "Animal Crossing" to the villager name to build the payload search per list item.
2. For loop used to extract the payload and appended results to a dictionary, "trends", per villager. The pytrend related queries are stored as key. To pull the query and value from the related query, we parsed the 'trends' dictionary for the 'rising' key.
3. We created a new dictionary to update the key as the villager search name and the values derived from the rising (query and value).
4. Pytrends returns a dictionary of dataframes when a search item has content. To handle the keys which had a item type of None (returned) a if/else statement nested inside a for loop was used
5. A final list was created then appended to a empty dataframe; pd.concat was used to create final dataframe and the final_villager_list_v2.csv.

ACP Player Poll Results

Goal: Take all polls spanning June 2020 through October 2021, and combine by villager name and total votes conducted monthly (17 Excel Files).

1. A issue arose around the monthly polls being in 2 different formats for each year. To work around this issue we used 2 different for loops to read in each Excel file to a dataframe then set the columns and appended to an empty df for the year. We removed (via filter) any Villagers that were Nan.
2. For each year df, we utilized a groupby by on Villager with a sum on Tally (votes) then reset the index to create a new dataframe. We combined the 2020 and 2021 results via pd.concat.
3. We cleaned spelling iterations of villagers using str.replace() to fix the names then recreated the groupby, reset the index and made the final file, ACP_FinalDF.csv.
4. The polls has 397 villagers while the game data has 391. We set the names from both to a list and used the set() method to subtract the lists from each other. It turned out that the ACP Poll data contained 6 villagers from a Sanrio update which is not considered for this analysis.

Merged Trends and Polls

Goal: Merge Google trends data and poll results with villagers.csv on 'Name' to create the final dataframe,

1. Lambda expression used to create column, 'Keep', where its purpose is to see if the villagers names were in the Google search query, if 'True', 'Keep' was updated to True otherwise set to False.
2. Usage of np.where() to update 'Keep' to False when 'query' referred to leaves, trees, fruits or diy recipes for villagers Maple, Peaches and Cherry. We fixed June update vs. June the villager (set to False) since all Google searches, 'query', referenced a update not the villager.
3. 'Name' & 'query' were cleaned. When 'query' was null used fillna and updated to 'unknown', then set "Keep" to False.
4. Once cleaned, a filter was applied to keep rows where 'Keep' column value was True. A groupby was done to get the total Google searches by villager name in a new dataframe.
5. Results above are merged on 'Name' with the villagers.csv. Then the poll results are merged onto the same df. Overall popularity field is created (sum of Google searches and ACP Poll votes), final file, Final_Villagers_ACNH.csv, is created.

Data Analysis Process Goals & Manipulation

Nookipedia (Critters)

Goal: To take fish and insects and concat them into one dataframe and use common name to scrape Nookipedia for scientific name.

1. We used str.replace() to fix the common names for web scraping/searching. The critters names were then put to a list and a for loop was used to attach the Nookipedia base url plus the critter name from the list.
2. Beautiful soup was used to scrape Nookipedia for the scientific name based on the common name via .find of text equal to 'Name:' and the html <i> tag.
3. The results of multiple lists are zipped together and then put into a critters dataframe. From there a .map() was applied to the 'title' column to pull out the scientific name at index 0 based on if it is a single list or list of lists.
4. The common name was kept and appended to a list so the critters dataframe maintained the common name during the scraping to match up to the concatenated dataframe. The common name was merged into one final df, acnh_critters_sci_name.csv.
Note: if no scientific name is on Nookipedia based on common name of fish or insect the resultant name is Unknown

Final Critters Data

Goal: To take the final critters dataframe and clean the scientific names while extracting the IUCN Red List Category (conservation status) from Wikipedia.

1. We utilized multiple regex along with str.replace to properly clean the scientific name from the leftover excess html or brackets.
2. Export the data frame into a clean dataset with all critters (fish and insects) with common name and scientific name. This creates the dataframe, critters_chk_final.csv.

Scrape Wikipedia

Goal: Scraped Wikipedia for conservation status (IUCN Red List Category) and develop visualizations from final dataframe.

1. Using critters_chk_final.csv, we took the scientific name column and created a list where a lambda expression was utilized with str.replace to take out spaces and replace with underscores so name could be used in a for loop.
2. For loop with Wikipedia base URL plus the scientific name used inside pd.read_html to extract the conservation status from the Wikipedia table, all pushed to 1 dataframe.
3. Used np.where & .values to fetch indices location equal to conservation status from returned data frame.
4. If no conservation status listed, the species (fish or insect) assumed to be of least concern. Regex used to clean status.
5. Critters analysis initialized by feature engineering the Red List Category by adding category code (int) via hierarchical status from least to greatest concern and pearson correlation found using .corr() method.
6. 'Where/How' field summarized to see where Red List critters were located via .apply & .get methods then visualized with median sell price.

**Villager Data Analysis:** The key goal of our analysis was to glean a better understanding of which villagers were popular in the ACNH game. Who was a favorite villager, did this mean the villager was Googled heavily? Alternatively, for a 'most searched' villager, why was this villager searched? Can we examine their personality traits and other features to enable a data driven portrait of why players prefer certain villagers? If we only utilized ACP Poll gaming data to decide who a favorite Villager is, then we would be painting a biased picture. The ACP website polling data for villager popularity is generated by avid game players who vote monthly based on their gaming experiences. We wanted to examine Google searches for villagers as an unbiased way of attributing popularity to villagers based on a general global sampling of all players, not just avid gamers. In other words, the Google searches are agnostic of ACP Poll gaming website driven biases. For our purposes, both datasets have meaningful information about villager popularity, but when we join the two, we are able to truly see who a popular villager is, driven by two independently generated datasets measuring two different ways of attributing a villagers' popularity. Of course, the caveat being, higher Google searches may coincide with how one may try to get a 'rare' villager on their island, making the villager popular potentially because of their rarity or on the opposite spectrum a search conducted based on a highly 'disliked' villager.

Villagers by Species and Personality: The **Popular Villagers: By Species & Personality bar chart** on the right, shows cat villagers are the most popular species based on overall popularity (Google Searches + ACP Poll Results) with cubs (bear) coming in second and deers coming in third. On the y-axis, each bar represents the nominal categorical variable of villager species and color as a nominal variable bar point to the personality traits amongst villagers for the given species. The y-axis is a quantitative variable for total count of google searches and ACP polling votes for a given villager (aka overall popularity). The size of the color stacked bar chart signifies a quantitative overall popularity of a personality trait by species. With this visual we can easily examine the most dominant personality traits amongst the villager species in ACNH. Based on the overall popularity ranking of the stacked bar chart by species and personality we can see that there seems to be a preference for smug or snooty personalities driven largely by cats. Overall, **3,752,037** total google searches and votes were for snooty cats and **3,923,865** for smug cats. Normal personality traits are found to be of higher value for cubs coming in at **2,869,752**. Analysis shows ~**14%** of overall popularity for snooty personalities for highly favored villagers and ~**12%** for smug villagers regardless of species. This is shocking since snooty (female) villagers tend be vain, pompous, fashionable or cranky which can lead to conflicts with other villager types on your island [3]. The most popular personality type is lazy at ~**17%**, second being Normal at ~**16%** and Big sister coming in lowest at just ~**7%** (regardless of species). Lazy's high ranking could be due to their conflicting personalities on islands which would lead to more searches as to why there are issues amongst their island villagers peaceful homeostasis.

Popular Villagers : By Species & Personality

Personality: Big Sister, Cranky, Jock, Lazy, Normal, Peppy, Smug, Snooty
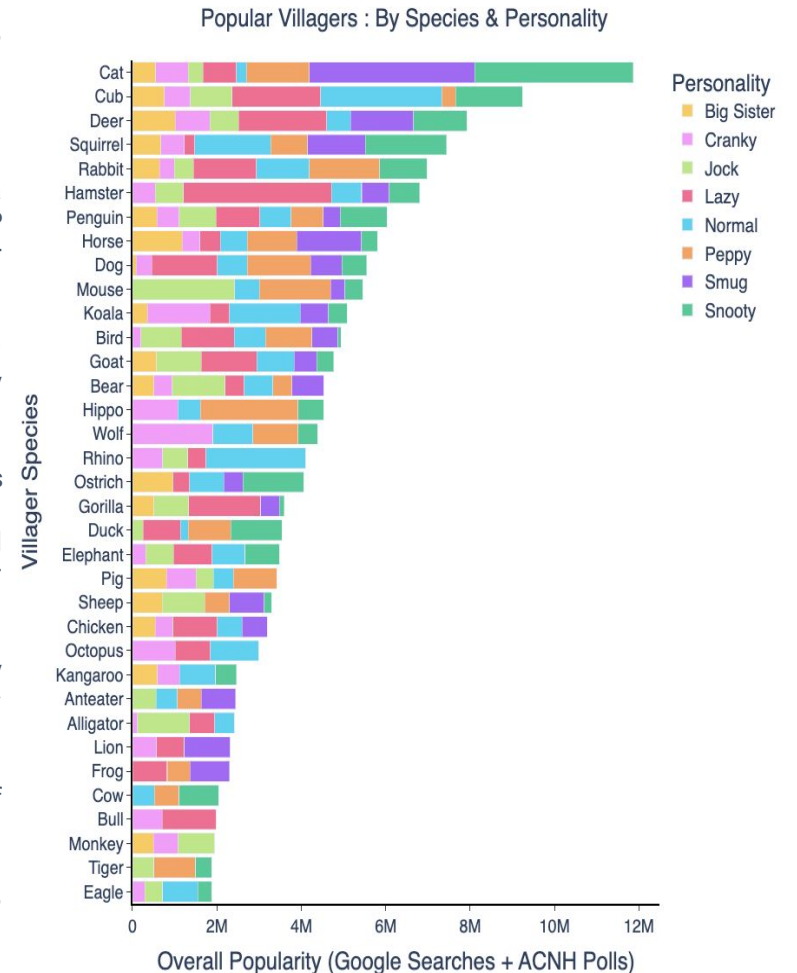
Villagers by Popularity, Google Searches versus ACP Poll Results: In this bubble chart on the right we are looking to find which Villager was both highly google searched as well as secured the highest votes in the ACP polls. The y-axis shows the total number of google searches for a given Villager and it is in very high numbers, in the millions. While the x-axis shows the number of times a Villager was voted as being favorite in the ACP poll results. Each dot is a Villager colored by their personality type. The position of the dot along with size of the dot on the y-axis and x-axis reflects their total searches in google and votes in ACP polls respectively.

Raymond, a cat villager is the most popular Villager at approximately 4 million google searches and over 25,000 votes in the ACNH polls. Which is consistent with online articles about Raymond being the 'King' of ACNH Villager characters[7]. He wears a sports business themed outfit with a smug personality making him very popular. Smug villagers are very sweet and caring and is no wonder, one of the most popular cats in our analysis. it also seems like players are really trying very hard to bring Raymond on their island, which may explain some of the "high" google searches for Raymond.

Marshal, a squirrel villager is almost just as popular as Raymond in the game, also sporting a Smug personality but not as searched on google as Raymond. Marshal is one of the cutest Villagers of the ACNH game and is not a surprise that he is ACNH poll popular with almost over 26000 votes.

Clay, a lazy Hamster Villager is not voted very high in the ACNH polls, only around 4000 votes, but Clay is searched in google 3.5 Million times. As a lazy villager, Clay may be very welcome on many islands due to his laid back and lazy lifestyle and getting along with most villagers except the jock (active + healthy) villagers but can get into conflicts with others.

Ankha is a snooty Cat Villager and based on our analyses, is both highly searched in google with around 2.5 Million google searches and bagged over 10,000 ACNH poll votes. We think that the rude and snooty personality of Ankha combined with her charming Egyptian inspired personality


Top 10 Villagers by Google Searches vs ACNH Polls
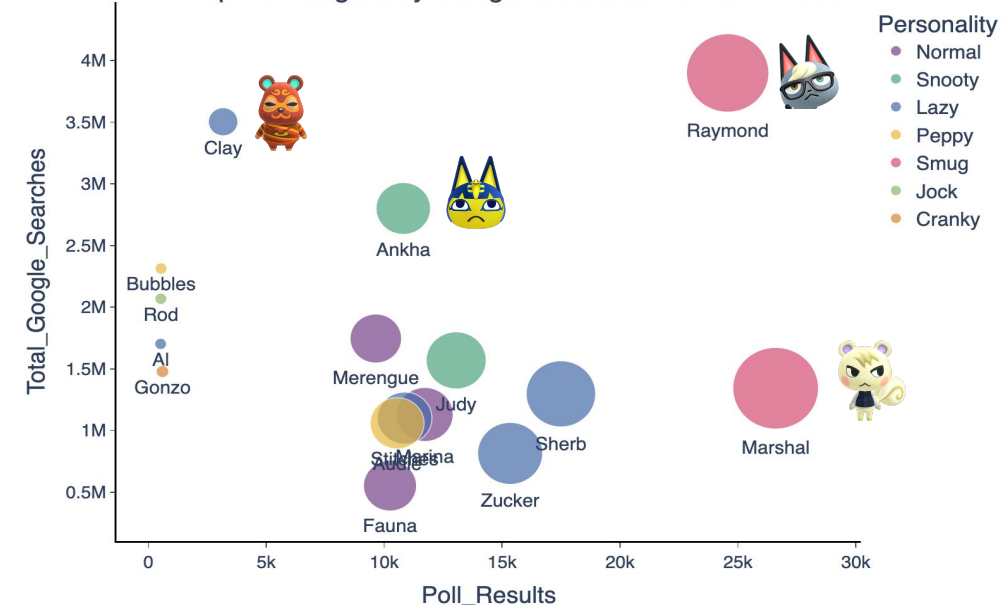
Disparities: We can see with the visualization a large disparity between the avid gamer ACP Polls and total Google searches with the villager, Clay. As mentioned Clay, is highly Googled but low in favorability amongst players. He is a lazy personality type, this conflicts with other villagers as they will not understand or become annoyed or angered by a lazy villager's inability to do simple tasks such as clean. Villager personality conflicts lead to disputes, and eventually unhappiness on the island amongst villagers resulting in lower island ranking or causing other villagers wanting to leave. This leads to higher Google searched villagers but lower ranking in terms of ACP Polls. There are various reddit posts about *such* disliked villagers who cause conflicts. Highly favored versus highly Googled can mean two different things. Overall our analysis findings conclude, popular Google searched villagers are generally highly favored in ACP Poll rankings.

Distribution of Critters in ACNH
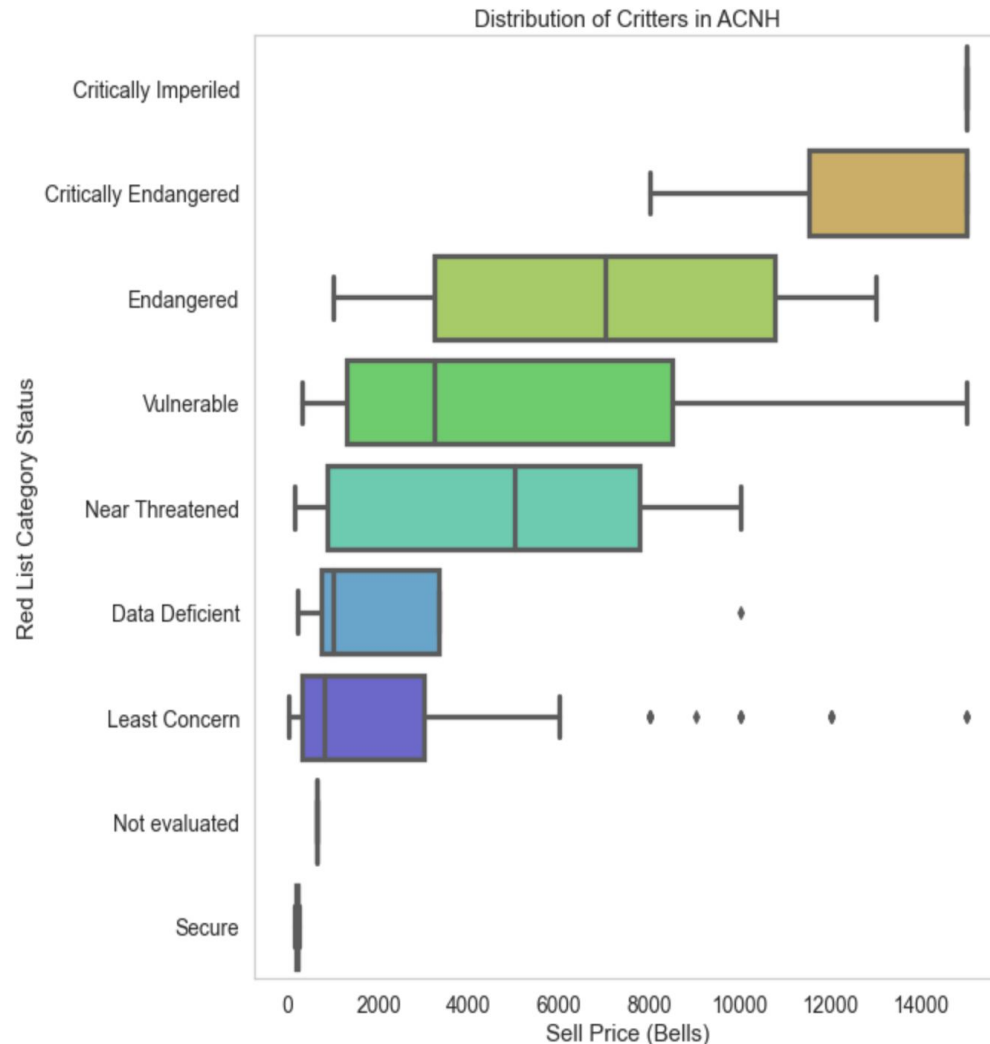
Critter Data Analysis: Our overarching goal for the critter analysis in the game was to see if the game was inadvertently encouraging the selling of endangered species? In order to achieve this goal we used the critter common name (insect or fish) and scraped Nookipedia to fetch the game listed scientific name. With the scientific name we scraped Wikipedia the by scientific name per critter and fetched the conservation status listed by the IUCN Red List Category on Wikipedia table page to see if there is a connection for the species status (i.e. critically endangered, endangered, least concern, etc). If there are endangered species in the game are they perpetuating the selling of the species further by giving them higher cost (bells) when sold? If so, what does this say in the real world for those of us playing the game? Are we turning a blind eye since it's a life simulation game of cartoon-like characters and what are the ethical concerns with this?

Critters by Red List Category and Sell Price: We used a standard box plot, created with Seaborn as a data visualization tool to see If they are perhaps pushing the catching and selling of endangered species within the game. The box plot shows the trend, range of sell price (in Bells), on the x-axis as a quantitative variable and the conservation status from the IUCN Red List Category as nominal variable on the y-axis. Color is applied as a nominal variable to each IUCN Red List Category. The box plot also display the IQR along with outlier sell prices per category. It finely demonstrates that our initial hypothesis of the game design is valid and gives the impression of rarity based on sell price for endangered critters. This caused a need for further deeper dive into correlation via a pearson correlation coefficient as a source of confirmation and where the critters of higher conservation status are being caught for later sell.

Correlation between Red List Category and Sell Price (Bells): In order to find the correlation between the Red List Category and the Sell price we made use of an assignment of category grade based on Red List IUCN Category hierarchical status with 1 through 9, 1 being Secure and 9 being Critically Imperiled. We then applied pearson corr method to the Sell price (in bells) and the Red List Category Grade which showed a positive pearson **correlation coefficient is 0.464954**. This coincides with our assumptions that the game developers unintentionally assigned higher value sell (bell) prices to endangered species and inadvertently are perpetuating the catching and selling of high conservation status species.

Knowing that there is a positive correlation between Sell price (Bells) and Red List Category we decided to look into where these critters are predominantly being caught Did the game developers have an even spread across the island and waters or were specific habitats found to have critters of higher conservation status then others. If so why?

In order to achieve the 'Where Found', habitats, we created a dictionary to simplify the 'Where/How' field on the file and utilized a dictionary with a lambda expression and .apply on Where/How based on value to key in the dictionary to update with the new dictionary value field to create a new field 'Where Found'. From here we looked at the median price by location and IUCN Red List Category status.

Sell Price by IUCN Status and Where Found: Treemap charts are great to visualize hierarchical data by using nested rectangles. We used Plotly Express to produce a nested treemap examining the relationships between median sell price grouped by IUCN status and where found (habitat) within the game. Based on this graph, we can see that critters (fish and insects) of least concern are more common and found relatively equally in different areas of the game such as in rivers or flying. The median sell price is also found to be near the minimum range. Critters with a threatened or higher status do prove to be sold for more bells based on how critical the status; such that critically endangered critters had the highest median sell price and are found in habitats of water like rivers, pounds or in the ocean sea. Additionally, these critters were found less frequently in connection to their higher bell value median sell price.

Why: We do not believe that the correlation between the conservation status and sell price were intentional by the game devs, although unfortunate, it could be for a global common ground of players and being familiar with the critters being caught in the game. If we examine the majority of threatened + Red List Category critters we find that most of them are larger shadow critters found in water; either ocean, sea or in rivers. We presume this coincides with most large sea creatures such as whales, tuna, marlins, great whites and hammerhead sharks.

Critters habitats by Red List Category and Mean Sell Value
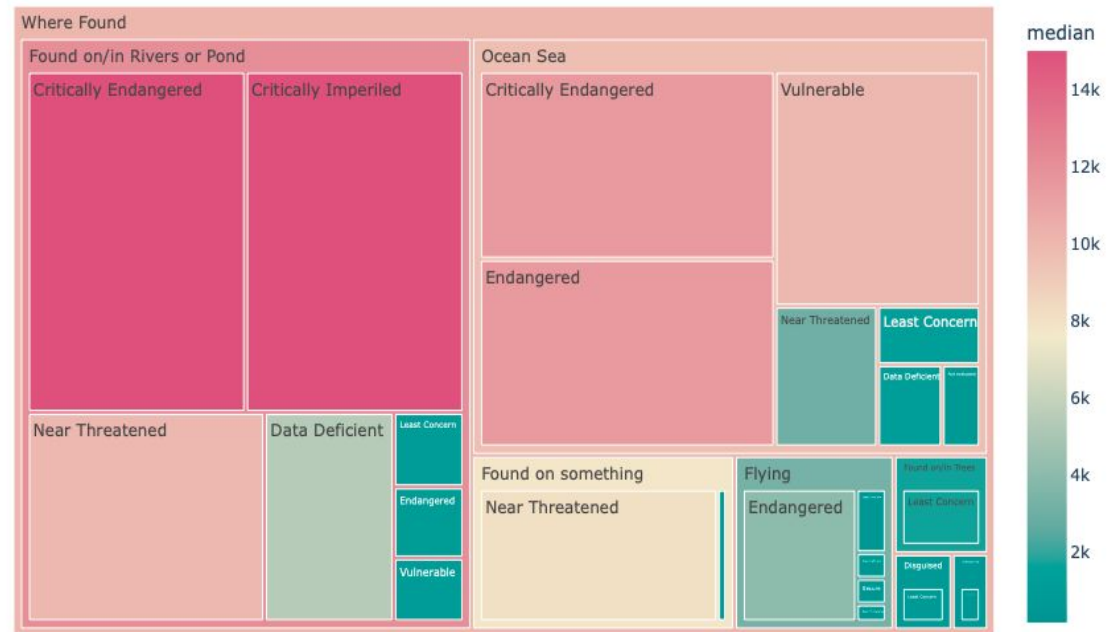
In actuality, many marine species are decreasing at an alarming rate and are considered endangered. All manner of marine species are affected, including whales, dolphins, porpoises, seabirds, sea turtles, sharks, other fish, and marine invertebrates. The consequences of loss of species in the marine environment are significant, both ecologically and economically [5]. It is sad to say but all of the oceans inhabitants have suffered from their homes being destroyed via overfishing, pollution and global warming.   Choosing commonality among players familiarity with larger sea creatures they are catching does inadvertently highlight the ecological perils of our worlds marine life today. Although the attention to detail within the game space has highlighted these ethical concerns to truly say they did this on purpose is not possible.

Conclusion, Further Use and Expansion

In conclusion, we believe the game developers did a decent job at creating a well-balanced game. The gender of villagers were equally balanced between male and female as well as their personality traits. As for popularity of villagers and their personality traits or species, this is more of a phenomenon surrounding preferred rare characters versus adverse characters that cause problems between villagers on the island. The sell price of critters (fish and insects) did correlate to be based on real world conservation status such that endangered critters sold for more bells than critters of least concern status as there was a overall positive correlation between sell price and IUCN Red List Category. This forgetfulness of a real world conservation issue does bring up the concern of ethical consideration and aligns with our initial expectations of the ACNH critter game data. Although a sore spot in findings to see the game is inadvertently perpetuating the catching and selling of endangered species we believe the correlation might be due to the game developers needing common species everyone globally would be familiar with.We all are familiar with a great white shark and the game does need 'larger" creatures to catch when fishing. Is it ethically wrong, yes, but is it a feature for commonality to game players? This is truthfully yet to be determined.

Further analysis could have been done by applying a sentiment score to the catchphrase and personality trait of each villager using the NLTK library in python. It would be interesting to examine if catchphrases or personality description sentiment values correlate to a villagers overall popularity value. As well as looking at Nookazon, a black market for buying rare villagers to compare highest selling villagers to our Google and ACP Poll results.

Consideration for further expansion may include an overarching game analysis to understand the easiest ways to earn bells in the game for quicker island expansion and how to achieve higher island ratings. This could be done in various ways by looking at the best crops to plant (i.e. Turnips) or activities during the day, such as fishing on a specific part of your island at a select time, then proceeding to catch bugs or dig for fossils at X time of day. Or a combination of crops to plant and when to sell said crops along with a daily itinerary to create the highest amount of bells per day. If the data was available looking into the counterfeit artwork being sold on the islands would be an interesting topic to dive into.

Animal Crossing New Horizons Comprehensive Analysis

Statement of Work

Jacqueline Skunda

Data collection for villagers via use of API usage for Google Trends/pytrends, cleaning, manipulation. ACP poll results, combining datasets to get total votes per villager. Creation of Overall Popularity field for villagers. Use of BS4 to scrape the Nookipedia. Wikipedia via pd.read_html to fetch the IUCN conservation status red list category. Further, data manipulation to clean and extract out needed information and merge back with the ACNH primary datasets. Correlation critters to red list category.

Malini Varadarajan

Research into Animal crossing data and potential handles for interesting analyses. Summarization of finding and explanations from data analysis. Slide document design, layout, presentation and data visualizations. PM lead. ACP Poll to Google trends deeper analysis and visualization creation. Considerations for sampling biases in the ACP Poll Villager vote results versus global Google Trends results. Critters analysis summarization of findings.

Tamara Qawasmeh

Researched to find the ACNH Primary data source. Deep dive/research into ethical considerations surrounding the questions and results from collection, linking and cleaning of datasets and data visualizations. Citations. Formulation and reviewing of final code, project turn in prep along with overall document flow. Report design and editing.

Sources and Citations

[1] En.wikipedia.org. 2022. Animal Crossing: New Horizons - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Animal_Crossing:_New_Horizons>

[2] Pypi.org. 2022. Pytrends. PyPI. [online] Available at: <https://pypi.org/project/pytrends/#credits>

[3] Fandom.com. 2022. Snooty. [online] Available at: <https://www.animalcrossing.fandom.com/wiki/Snooty>

[4] Plotly.com. 2022. Treemap Charts in Python. [online] Available at: <https://plotly.com/python/treemaps>

[5] Animal Welfare Institute, 2022, Endangered Marine Species. [online] Available at: <www.awionline.org/content/endangered-marine-species>

[6] Animal Crossing: New Horizons - The most popular villagers. <https://www.imore.com/animal-crossing-new-horizons-most-popular-villagers>

Github Link —>