Animal Crossing Villagers: An Exploratory Data Analysis

Business Problem

In the video game Animal Crossing, players have known to become emotionally attached to villagers. Also, some villagers are more well liked than others. Is there a reason? In this exploratory data analysis, we will look a villager's attributes to determine if there are any patterns associated with an animal's popularity to its characteristics.

Background/History

Animal Crossing New Horizons is a social simulation video game created by Nintendo for the Nintendo Switch Console. In the game, the player is a human who is moving to a deserted resort island populated with anthropomorphic animals, called villagers.

Data Explanation

I will be getting my data set from the Animal Crossing New Horizons Catalog available on Kaggle. This data set contains the villager's name, species, gender, personality, hobby, birthday, catchphrase, favorite song, style 1 and style 2, color 1, color 2, wallpaper, flooring, furniture list, a filename and unique entry id.

Resource Link: https://www.kaggle.com/jessicali9530/animal-crossing-new-horizons-nookplaza-dataset?select=villagers.csv

I will be getting my animal crossing villager ranking from an open survey. This article ranks each of the animal crossing villagers into TIER 1 – TIER 6, with TIER 1 being the most desirable.

Reference: https://www.animalcrossingportal.com/games/new-horizons/guides/villager-popularity-list.php#/

The github is available for this project at:

https://github.com/RachelONelson/680/tree/main/Project%201

Methods

Cleansing and prepping the data

Checking data types

All data types are objects with the exception of the villager's "Ranking" which is an integer.

Reviewing and handling of missing values

No missing values found in the villager data frame, no action required. There were additional villagers listed in the rank list, which were dropped after the two lists were merged.

Looking for duplicate data

No duplicate data was found in the data frame. No further action required.

Creating dummy variables

Dummy variables were created for all object data types. Also used an ordinal encoder
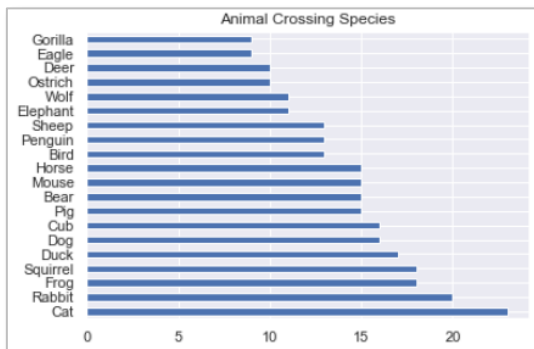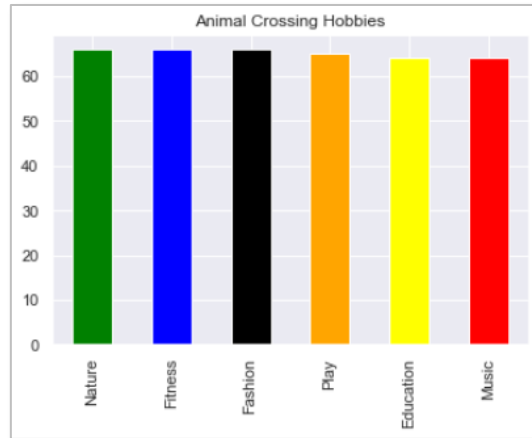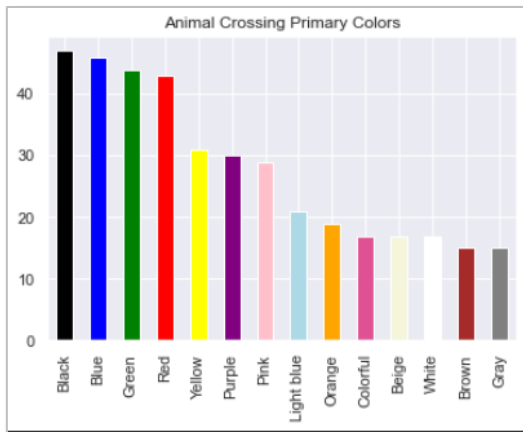
Analysis

Correlation Matrix

Correlation analysis revealed correlation between Gender and Personality. Further investigation showed that Personality Types are divided by Gender.
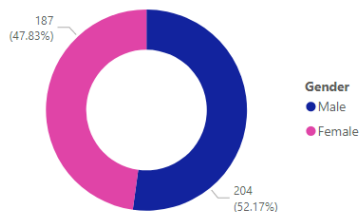
Charts Reviewed

Reviewed bar charts for animal species, personality types, hobbies, primary styles, and primary colors.
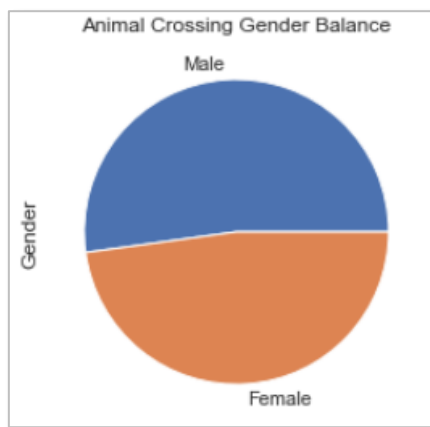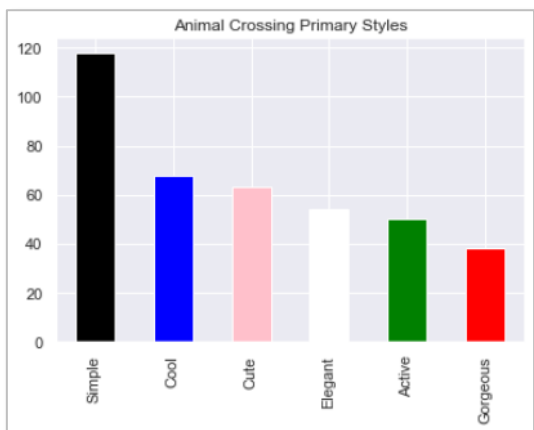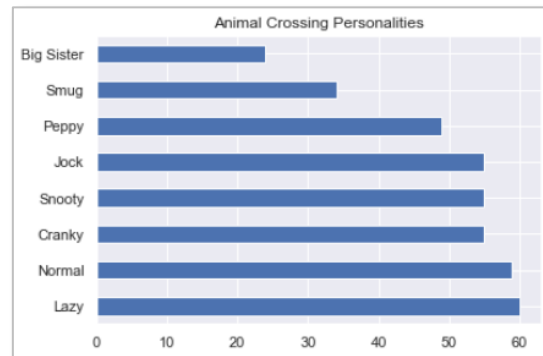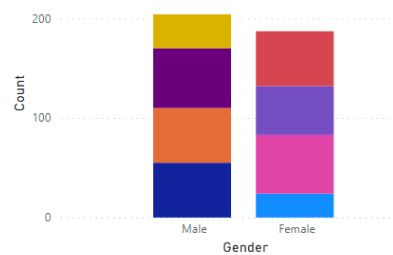
Reviewed charts by gender:
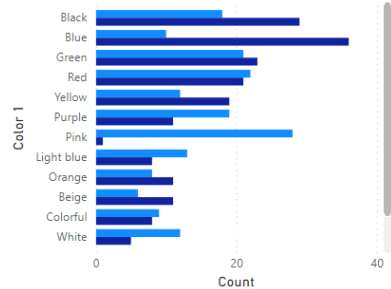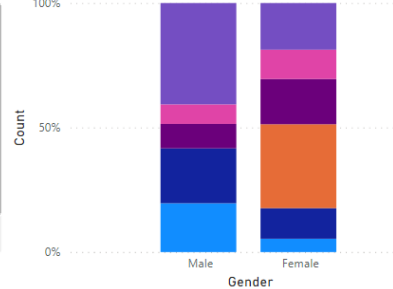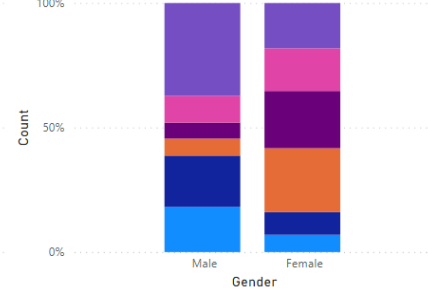
Modeling

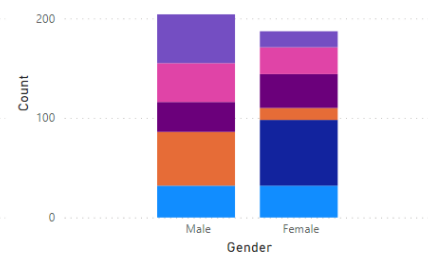The goal is to see if we can predict a villager's popularity using their attributes.

Tree Classification Models

For this analysis, I decided to use a decision tree and random forest models to see what the error rate would be.

When reviewing the models, the error rate was near 50% for both models, so they were not effective in predicting the animal's tier based off of the other known factors.

Rattle 2022-Jan-09 13:28:09 emera

Conclusion/Findings

Gender plays a huge role in a villager's personality, hobby and style:

- Personality Type is tied directly the gender of each animal crossing villager.
  - Male personality types include Cranky, Jock, Lazy and Smug
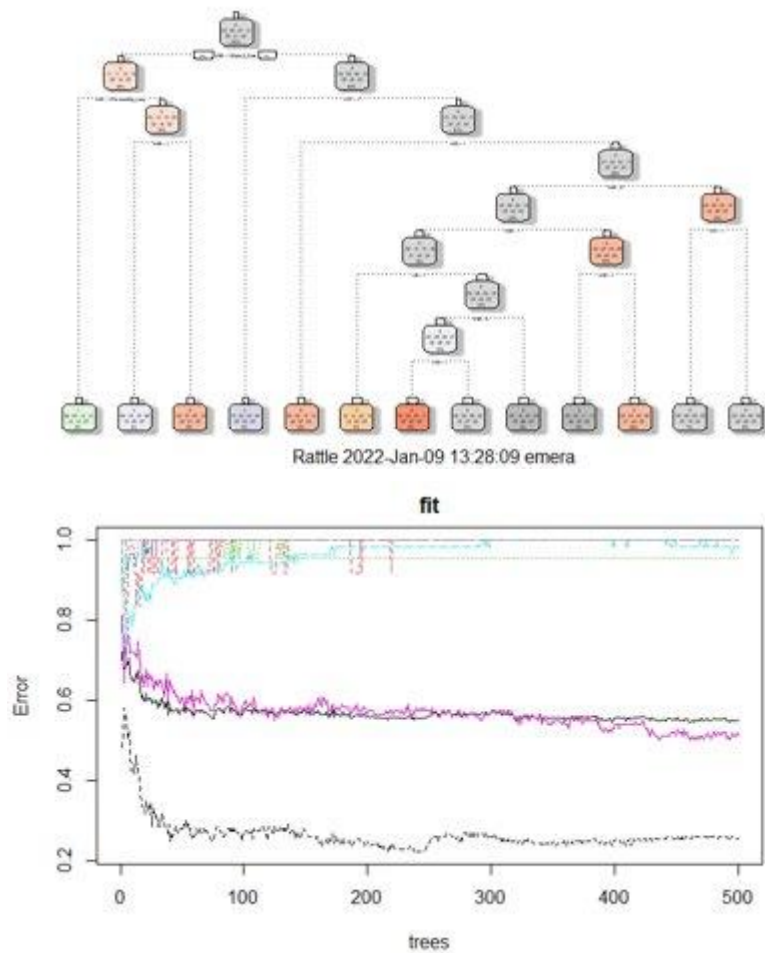 - Female personality types include Snooty, Peppy, Normal and Big Sister
- Only Female villagers have the "Fashion" hobby
- Only Female villagers have the "Cute" style as their primary style
- The Color "Pink" was only assigned to one male villager

Using villager attributes to determine the villager's desirability resulted in an R-Squared value was 0.562. However, we were able to determine that the most significant factors in desirability are the hobby and gender of the villager.

Assumptions

I had to assume that the villager ranking is a good representation of the population of animal crossing and how they rate their villagers. I also am assuming that any new villagers that have been added to the game since the original release follow the same guidelines/trends as the original population of the 391 original villagers in the analysis.

Limitations

I was limited to the villagers that were captured at the time of original data set creation. This included 391 unique villagers. This data is also limited to the specific game: Animal Crossing New Leaf.

Challenges

The biggest challenge in dealing with this data, is all factors are categorical/text. There were 0 numerical values in the data set. While some can be interpreted as binary (male/female), other factors needed to be either encoded as ordinal or translated to dummy variables. Dealing with this information while still finding meaningful patterns was a challenge.

Future Uses/Additional Applications

This information can be used for future characterization development of animal crossing villagers. It can also be used to determine what types of new traits each animal crossing villagers might adopt in future adaptations of the game. This data can also be used for game balancing and might even have some psychology applications.

Recommendations

To ensure popularity of animals is properly demonstrated in the collected values, I recommend doing a random survey of animal crossing players to rank animal crossing villagers.

Implementation Plan

No implementation is planned.

Ethical Assessment

The ethical question, which was not anticipated at the beginning of the analytics, was how gender is directly tied to a villager's personality trait and hobbies. This could be interpreted as creating/perpetuating gender stereotypes in the video game.

References

Li, J. (2021, June 8). *Animal Crossing New Horizons Catalog*. Kaggle. Retrieved December 19, 2021, from https://www.kaggle.com/jessicali9530/animal-crossing-new-horizons-nookplaza-dataset?select=villagers.csv

Villager popularity tier list - animal crossing: New horizons. Animal Crossing Portal. (n.d.). Retrieved December 19, 2021, from https://www.animalcrossingportal.com/games/new-horizons/guides/villager-popularity-list.php#/