Animal Crossing New Horizons Villagers

Topic

This project will include exploratory data analysis of Animal Crossing New Horizons Villagers. I will also be looking to see if I can predict an animal's personality based on other known factors.

Business Problem

I will be looking into many questions regarding the animal crossing villagers including:

- How Many Villagers are there?
- What types and how many of each species of animal
- Are the genders balanced?
- What personality types are there?
- Can you predict an animals personality based on other known factors?

Datasets

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

Methods -

I will be using a variety of analysis methods in this exploratory data analysis including

- Cleansing and prepping the data
  - Checking data types
  - Reviewing and handling of missing values
  - Looking for duplicate data
  - Creating dummy variables
- Analysis
  - Looking for outliers using boxplots
  - Histograms to observe the distribution of data
  - Bar charts to see data balance among categories
  - Correlation Matrix + Scatterplots to look for correlation between variables
- Modeling
  - Regression and Tree models

Ethical Considerations

Since I am dealing with video game data, there should be not too much in the area of ethical considerations. However, this data was translated into real-life data, there could be ethical considerations around using a person's species/race, age, and gender to predict anything.

Challenges/Issues

Most of my data is string-based, so I will need to make some dummy variables. I also need to break down the furniture sets into a different data set since there are going to be a 1 to many relationships and right now, they are all in one column.

References -

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