Final Project

Yixuan Chen

1. Motivation

Animal Crossing is one of the most popular switch game in 2020. The game's relaxed pace, gorgeous visuals, soothing soundtrack, and peaceful vibes provided a welcome contrast to the onslaught of harrowing news in the real world, and players took full advantage of that escapism. For my final project, I would like to analyze the most profitable way to play this game. Basically, the easiest way to make money is to catch and sell animals. In this project, two main kinds of species will be analyzed, fish and bugs.

The purpose of this project is that given a specific time, the result will show one or several locations to go to by finding the largest expect profit of selling species which could be found at each location.

It should be noted that all animals will show up in their routine time and time in this game is corresponding to one in the real world and poeple in different hemisphere will have different experiences.

2. Data Sources Description

The first two datasets I found for web scraping were:

https://animalcrossing.fandom.com/wiki/Fish_(New_Horizons)

https://animalcrossing.fandom.com/wiki/Bugs_(New_Horizons).

These two datasets include all fish and bugs the game has. We could get the basic information from them such as name, price, location, and image.

For the data achieved from external public API, the host will be

https://acnhapi.com/v1a/.

We could get the rarity of all species which we may use as a probability of show-up rate.

3. Data Handling

The packages needed are listed before.

In [1]:
```
import pandas as pd
import seaborn as sns
import re
import numpy as np
from statistics import mean
import math
import full_dataframe as frame
```

Read in datasets

In [2]:
```
data_north  = frame.default_north()
data_south  = frame.default_south()
```

Clean datasets

This process removes the additional information of string columns.

In [3]:
```
def data_clean(dataset):
    for i in range(dataset.shape[0]):
        dataset['Price'][i] = dataset['Price'][i].replace(',','')
        location = re.sub(r'\([^()]*\)', '', dataset['Location'][i]).rstrip()
        dataset['Location'][i] = location
        rarity = re.sub(r'\([^()]*\)', '', dataset['Rarity'][i]).rstrip()
        dataset['Rarity'][i] = rarity
        if dataset['Category'][i] == 'Fish':
            shadowsize = re.sub(r'\([^()]*\)', '', dataset['Shadow size'][i]).rstrip()
            dataset['Shadow size'][i] = shadowsize
    return dataset
```

In [4]:
```
data_north_clean = data_clean(data_north)
data_south_clean = data_clean(data_south)
```

Create a new variable 'Catch Probability'

Due to different shadow sizes and rarity, there still exists some probability that players can't catch the animial even they are at the right location at right time. Therefore, a new variable is created to measure the probability of catching the species. For common species, there is 90% chance they will show up. For uncommon ones, 50%; rare ones, 30% and for ultra-rare ones, the probability is 10%. Also, fish will have different shadow sizes in river and players may miss them if they have too small shadow sizes. So from 6 to 1, the prabability of player find the species decrease from 95% to 70% by 5% and for 'narrow' size, the probability will be 60%. Players may need some skills to catch the species, so there is 90% chance of success for each attempt of catching fish and 85% for catching bugs. Therefore, the catch probability will be the product of three probabilies.

And the expected profit for each species will be catch probability times price.

```
In [5]: def prob(dataset):
            dataset['Price'] = pd.to_numeric(dataset['Price'])
            dataset['showup_prob'] = dataset['Rarity'].map({'Common':0.9,'Uncommon':0.5,'Rare':0.3,'Ultra-rare':0.1,'None':0})
            dataset['find_prob'] = dataset['Shadow size'].map({'6':0.95,'5':0.9,'4':0.85,'3':0.8,'2':0.75,'1':0.7,'Narrow':0.6}
            dataset['find_prob'] = dataset['find_prob'].fillna(1)
            dataset['success_prob'] = dataset['Category'].map({'Fish':0.9,'Bug':0.85})
            dataset['catch_prob'] = dataset['showup_prob'] * dataset['find_prob'] * dataset['success_prob']
            dataset['Price'] = dataset['Price'].replace(',','')
            dataset['profit'] = dataset['catch_prob'] * dataset['Price']
            dataset = dataset[dataset.catch_prob != 0]
            return dataset
```

```
In [6]: data_north_ready = prob(data_north_clean)
        data_south_ready = prob(data_south_clean)
```

Find optimal place according to time and month

According to the expected profit of each species and its show-up month and time, we could find the optimal place for players to go to at specific time. The main idea for this process is to find all species which will show-up and categorized them by location. Then we could find the mean profit could be achieved for different places and will go to one with highest profit. This process will find three optimal places for each hour.

```
In [7]: def find_location(month,time,north = True):
            if north == True:
                dataset = data_north_ready
                monthname = 'North_month'
            else:
                dataset = data_south_ready
                monthname = 'South_month'
            locationlist = []
            rank_1 = 'First'
            rank_2 = 'Second'
            rank_3 = 'Third'

            locationname_1 = float('nan')
            profit_1 = 0
            species_1 = 0
            rarity_1 = float('nan')
            price_1 = float('nan')
            category_1 = float('nan')


            locationname_2 = float('nan')
            profit_2 = 0
            species_2 = 0
            rarity_2 = float('nan')
            price_2 = float('nan')
            category_2 = float('nan')

            locationname_3 = float('nan')
            profit_3 = 0
            species_3 = 0
            rarity_3 = float('nan')
            price_3 = float('nan')
            category_3 = float('nan')

            for location in list(dataset.groupby('Location')):
                locationname = location[0]
                sub_dataset = location[1]
                profitlist = []
                maxspecies = ''
                maxrarity = ''
                maxprice = 0
                maxcategory = ''

                for i in sub_dataset.index.tolist():
                    if month in sub_dataset[monthname][i]:
                        if time in sub_dataset['Time'][i]:
                            profitlist.append(sub_dataset['profit'][i])
                            if sub_dataset['Price'][i] > maxprice:
                                maxprice = sub_dataset['Price'][i]
                                maxspecies = sub_dataset['Name'][i]
                                maxrarity = sub_dataset['Rarity'][i]
                                maxcategory = sub_dataset['Category'][i]
                if not profitlist:
                    meanprofit = 0
                else:
                    meanprofit = mean(profitlist)

                if meanprofit > profit_1:
                    locationname_1 = locationname
                    profit_1 = meanprofit
                    species_1 = maxspecies
                    rarity_1 = maxrarity
                    price_1 = maxprice
                    category_1 = maxcategory


                if meanprofit > profit_2 and meanprofit < profit_1:
                    locationname_2 = locationname
                    profit_2 = meanprofit
                    species_2 = maxspecies
                    rarity_2 = maxrarity
                    price_2 = maxprice
                    category_2 = maxcategory
                    rank_2 = 'Second'

                if meanprofit > profit_3 and meanprofit < profit_2:
                    locationname_3 = locationname
                    profit_3 = meanprofit
                    species_3 = maxspecies
                    rarity_3 = maxrarity
                    price_3 = maxprice
                    category_3 = maxcategory
                    rank_3 = 'Third'

                locationlist = [locationname_1,rank_1,profit_1,species_1,rarity_1,price_1,category_1,
                                locationname_2,rank_2,profit_2,species_2,rarity_2,price_2,category_2,
                                locationname_3,rank_3,profit_3,species_3,rarity_3,price_3,category_3]
            return locationlist
```

Create final dataframe for analysis

From the above function, we find most 3 profitable places with expected profit, most valuable species and its price and category. So we are going to include all the information and create a dataset for final analysis.

In [8]:
```
month = np.repeat(np.repeat(range(1,13),3),24)
time = list(np.repeat(range(0,24),3))*12
timedict = {'Month':month,'Time':time,'Optimal Location':''*3*12*24,
            'Rank':''*3*12*24,'Optimal Profit':''*3*24*12,'Most Valued Species':''*3*24*12,
            'Rarity':''*3*24*12,'Price':''*3*24*12,'Category':''*3*24*12}
optimallocation = pd.DataFrame(timedict)
```

In [9]:
```
def location_summary(dataset,north = True):
    for i in range(12*24):
        optimallist = find_location(dataset['Month'][i],dataset['Time'][i],north)
        for j in range(3):
            for k in range(7):
                dataset.iloc[i*3+j,k+2] = optimallist[j*7 + k]
    return dataset
```

In [10]:
```
north = location_summary(optimallocation,north = True)
north['Optimal Profit'] = pd.to_numeric(north['Optimal Profit'])
north.head()
```

Out[10]:

	Month	Time	Optimal Location	Rank	Optimal Profit	Most Valued Species	Rarity	Price	Category
0	1	0	On the Ground	First	4207.50000	Tarantula	Common	8000	Bug
1	1	0	Pier	Second	2180.25000	Blue marlin	Rare	10000	Fish
2	1	0	Sea	Third	956.38125	Barreleye	Ultra-rare	15000	Fish
3	1	1	On the Ground	First	4207.50000	Tarantula	Common	8000	Bug
4	1	1	Pier	Second	2180.25000	Blue marlin	Rare	10000	Fish

In [11]:
```
south = location_summary(optimallocation,north = False)
south['Optimal Profit'] = pd.to_numeric(north['Optimal Profit'])
south.head()
```

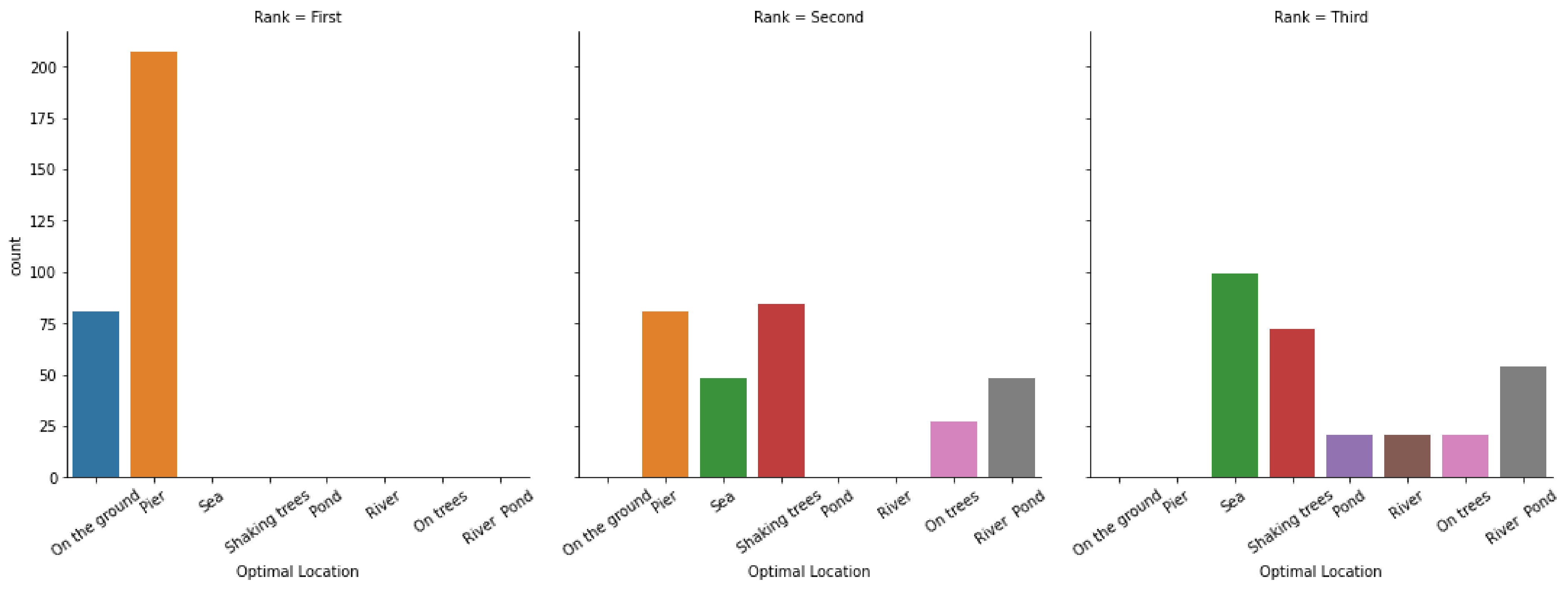
Out[11]:

	Month	Time	Optimal Location	Rank	Optimal Profit	Most Valued Species	Rarity	Price	Category
0	1	0	On the ground	First	2329.000000	Scorpion	Common	8000	Bug
1	1	0	Pier	Second	1948.500000	Blue marlin	Rare	10000	Fish
2	1	0	Sea	Third	1333.480263	Great white shark	Rare	15000	Fish
3	1	1	On the ground	First	2329.000000	Scorpion	Common	8000	Bug
4	1	1	Pier	Second	1948.500000	Blue marlin	Rare	10000	Fish

4. Data visualization and Analysis

In [12]:
```
g1 = sns.catplot(x = 'Optimal Location',col = 'Rank',data = north, kind='count')
g1.set_xticklabels(rotation=35)
```

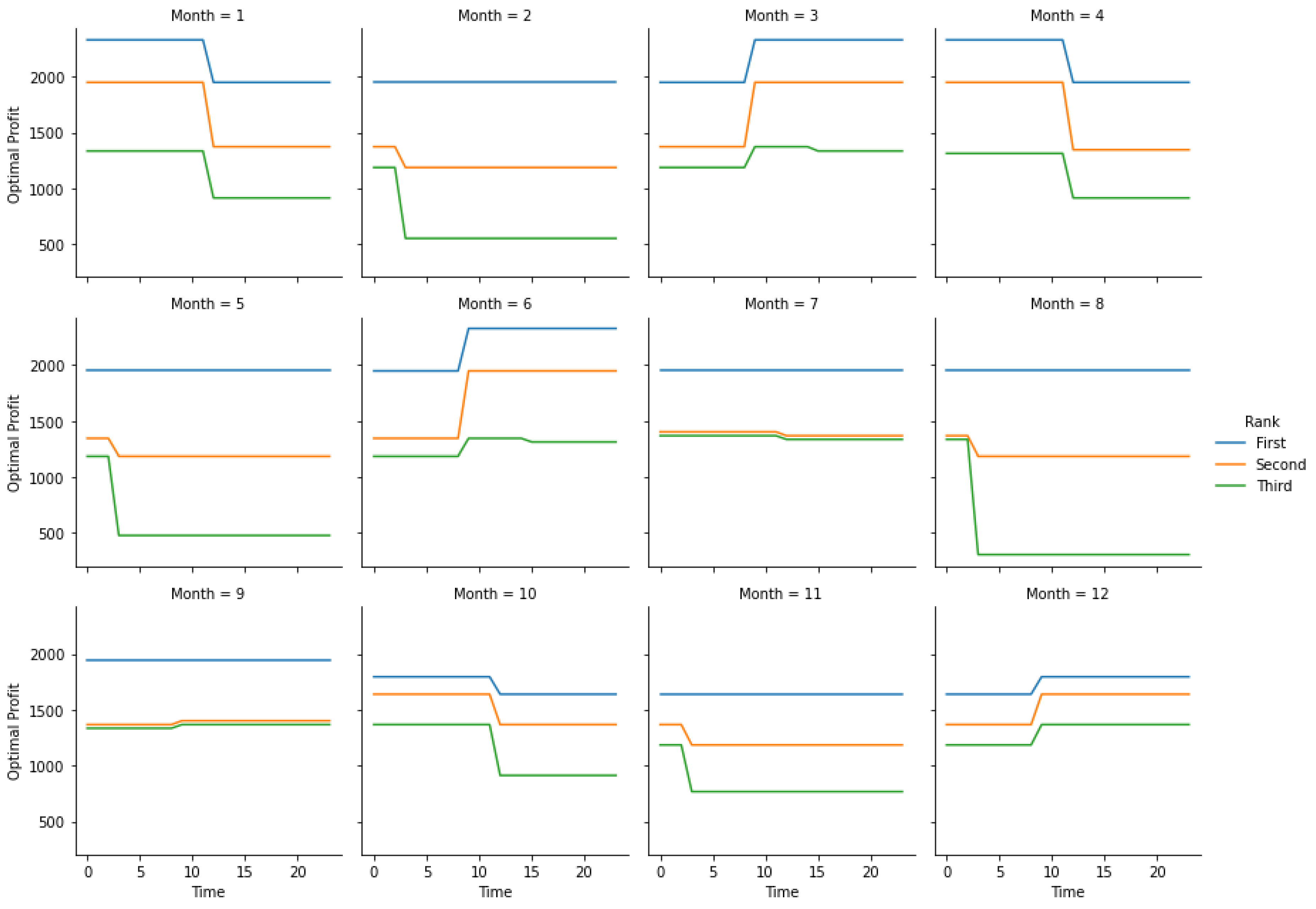
Out[12]:
```
<seaborn.axisgrid.FacetGrid at 0x7fa26af7f2b0>
```



This plot shows that the most profitable places for the whole year seem to be "on the ground" and "pier". If player want to make money quicker, they should always be around these two places, expecially around 'pier' which is also second most profiable places for more than 140 times. Also, if players get bored with same locations, there are other choices to explore and make money at the same time such as "sea" or "shaking trees".

In [13]:
```
g2 = sns.FacetGrid(north,col = 'Month',hue = 'Rank',col_wrap = 4)
g2.map(sns.lineplot,'Time','Optimal Profit')
g2.add_legend(title='Rank')
```

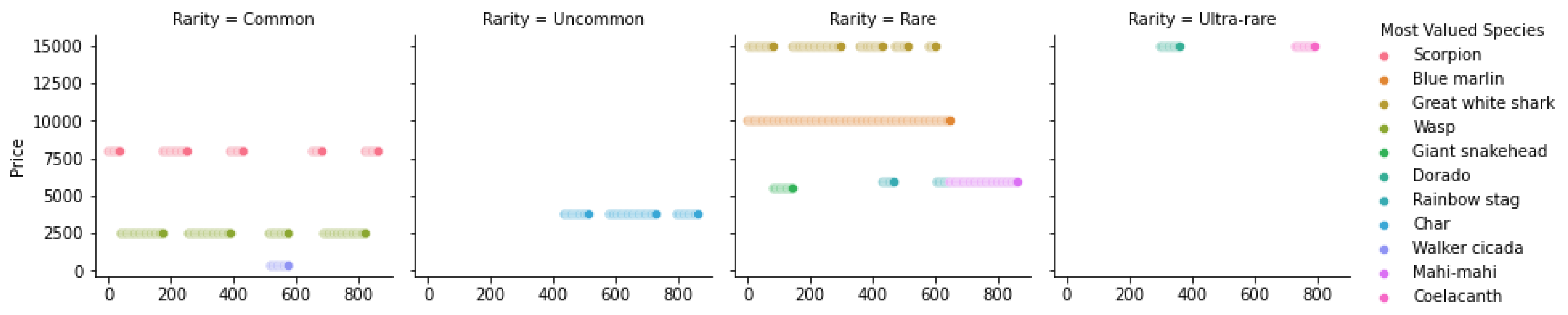
Out[13]:
```
<seaborn.axisgrid.FacetGrid at 0x7fa26aefd2b0>
```



This plot shows that the relationship between optimal profit and time. We could see that the best time for making money is in early-half days of January and late-half days of March with a expected profit of more than 4000. Also, players could make good earnings in April, June, July, September, October and December. In Feburary, May, August and November, players could try different things or explore the game as the expected profit is lower than other months.

In [14]:
```
north_g3 = north
north_g3['rownumber'] = north.index
g3 = sns.FacetGrid(north,col = 'Rarity',hue = 'Most Valued Species',
                   col_order = ['Common','Uncommon','Rare','Ultra-rare'])
g3.map(sns.scatterplot,'rownumber','Price')
for i in range(4):
    g3.axes[0,i].set_xlabel('')
g3.add_legend(title='Most Valued Species')
```

Out[14]:
```
<seaborn.axisgrid.FacetGrid at 0x7fa26b944190>
```



According to this plot, we could find that generally the rarer a species is, the more expensive it will be. Rare species tend to the most profitable kind for most seasons. Ultra-rare animals only show up in November as it seems unlikely for ultra-rare ones to have a lower price. The highest price for fish and bugs are 15000.