The Nobel Prize has been among the most prestigious international awards since 1901. Each year, awards are bestowed in chemistry, literature, physics, physiology or medicine, economics, and peace. In addition to the honor, prestige, and substantial prize money, the recipient also gets a gold medal with an image of Alfred Nobel (1833 - 1896), who established the prize.

The Nobel Foundation has made a dataset available of all prize winners from the outset of the awards from 1901 to 2023. The dataset used in this project is from the Nobel Prize API and is available in the `nobel.csv` file in the `data` folder.

In this project, you'll get a chance to explore and answer several questions related to this prizewinning data. And we encourage you then to explore further questions that you're interested in!

```python
# Loading in required libraries
import pandas as pd
import seaborn as sns
import numpy as np

# Read in the Nobel Prize data
nobel = pd.read_csv('data/nobel.csv')

#Question 1: What is the most commonly awarded gender and birth country?
top_gender = nobel['sex'].value_counts().index[0] #counted values of each value of
sex aand selected the information in the first position as that shows the gender
with the highest value count
top_country = nobel['birth_country'].value_counts().index[0] #did same thing as step
1 but for birth country
print("\n The gender with the most Nobel laureates is :", top_gender)
print(" The most common birth country of Nobel laureates is :", top_country)

#Question 2: Which decade had the highest ratio of US-born Nobel Prize winners to
total winners in all categories?
nobel['usa_born_winner'] = nobel['birth_country'] == 'United States of America'
#created new column that gave us whether birth country of winner was USA or not
nobel['decade'] = (np.floor(nobel['year'] / 10) * 10).astype(int) #used np.floor to
create a decade column so that we could group the winners by which decade they won
in. Using np floor allows us to get the decade without rounding up to the next
decade. MAde sure to convert the decade to type integer
prop_usa_winners = nobel.groupby('decade', as_index=False)['usa_born_winner'].mean()
#grouped this new data by decade without letting it be the index of the data, and
obtained average number of usa born winners from each of those decades
max_decade_usa = prop_usa_winners[prop_usa_winners['usa_born_winner'] ==
prop_usa_winners['usa_born_winner'].max()]['decade'].values[0] #Identified the
decade with the highest proportion of US-born winners by selecting max value
print(prop_usa_winners)

#Question 3: Which decade and Nobel Prize category combination had the highest
proportion of female laureates?
nobel['female_winner'] = nobel['sex'] == 'Female' #created coluimn that allowed us
to see whether the winner was female or not.
prop_female_winners = nobel.groupby(['decade', 'category'], as_index=False)
['female_winner'].mean() #Calculated the proportion of female laureates and made
sure to group by decade and category
max_female_decade_category =
prop_female_winners[prop_female_winners['female_winner'] ==
prop_female_winners['female_winner'].max()][['decade', 'category']] # Found the
decade and category with the highest proportion of female laureates by setting the
data so that it is equal to the decade and categpry that had the .max number of
female winners.
```

```python
max_female_dict = {max_female_decade_category['decade'].values[0]:
max_female_decade_category['category'].values[0]} # Created a dictionary with the
decade and category pair
print(max_female_dict)

#Question 4: Who was the first woman to receive a Nobel Prize, and in what category?
nobel_women = nobel[nobel['female_winner']] #created data set that showed only
female winners using new dataset from previous question.
min_row = nobel_women[nobel_women['year'] == nobel_women['year'].min()] #we want the
first woman to win so we set the year equal to the minimum year shown in this data
set, since our data only includes women winners anyways
first_woman_name = min_row['full_name'].values[0] #We select the full name of the
winner since our data now reflects all of the information of the first female winner
and set it as a new value
first_woman_category = min_row['category'].values[0] #create a value that shows only
the category of that first female winner.
print(f"\n The first woman to win a Nobel Prize was {first_woman_name}, in the
category of {first_woman_category}.") #using the two new values we created above, we
create the new sentence printed

#Question 5: Which individuals or organizations have won more than one Nobel Prize
throughout the years?
counts = nobel['full_name'].value_counts() #counted number of times each name
appears in the dataset
repeats = counts[counts >= 2].index #set counts so that only names that appear 2 or
more times are listed
repeat_list = list(repeats) #created a list using the new previous stem that shows
names that appear twice as a list

print("\n The repeat winners are :", repeat_list)
```

```
 The gender with the most Nobel laureates is : Male
 The most common birth country of Nobel laureates is : United States of America
     decade  usa_born_winner
0       1900         0.017544
1       1910         0.075000
2       1920         0.074074
3       1930         0.250000
4       1940         0.302326
5       1950         0.291667
6       1960         0.265823
7       1970         0.317308
8       1980         0.319588
9       1990         0.403846
10      2000         0.422764
```

```
11    2010       0.314050
12    2020       0.360000
{2020: 'Literature'}
```

 The first woman to win a Nobel Prize was Marie Curie, née Sklodowska, in the
category of Physics.

 The repeat winners are : ['Comité international de la Croix Rouge (International
Committee of the Red Cross)', 'Linus Carl Pauling', 'John Bardeen', 'Frederick
Sanger', 'Marie Curie, née Sklodowska', 'Office of the United Nations High
Commissioner for Refugees (UNHCR)']