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!

```
# Questions: What is the most commonly awarded gender and birth country?, Which decade had the highest ratio of US-born Nobel Prize winners to total winners in all categories?, Which decade and Nobel Prize category combination had the highest proportion of female laureates?, Who was the first woman to receive a Nobel Prize, and in what category?, Which individuals or organizations have won more than one Nobel Prize throughout the years?

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

# loading in the Nobel Foundation Data
nobel_data = pd.read_csv('data/nobel.csv')

#Check out the head of the data
print(nobel_data.head(12))
```

```
    year    category                                          prize  \
0   1901   Chemistry               The Nobel Prize in Chemistry 1901
1   1901  Literature              The Nobel Prize in Literature 1901
2   1901    Medicine  The Nobel Prize in Physiology or Medicine 1901
3   1901       Peace                      The Nobel Peace Prize 1901
4   1901       Peace                      The Nobel Peace Prize 1901
5   1901     Physics                 The Nobel Prize in Physics 1901
6   1902   Chemistry               The Nobel Prize in Chemistry 1902
7   1902  Literature              The Nobel Prize in Literature 1902
8   1902    Medicine  The Nobel Prize in Physiology or Medicine 1902
9   1902       Peace                      The Nobel Peace Prize 1902
10  1902       Peace                      The Nobel Peace Prize 1902
11  1902     Physics                 The Nobel Prize in Physics 1902

                                        motivation prize_share  \
0   "in recognition of the extraordinary services ...         1/1
1   "in special recognition of his poetic composit...         1/1
2   "for his work on serum therapy, especially its...         1/1
3                                                 NaN         1/2
4                                                 NaN         1/2
5   "in recognition of the extraordinary services ...         1/1
6   "in recognition of the extraordinary services ...         1/1
7   "the greatest living master of the art of hist...         1/1
8   "for his work on malaria, by which he has show...         1/1
9                                                 NaN         1/2
10                                                NaN         1/2
11  "in recognition of the extraordinary service t...         1/2

    laureate_id laureate_type                          full_name  birth_date  \
```

```
# Display the number of nobel prizes given out from 1901 to 2023
display(len(nobel_data))

# Show the number of prizes won by males and females
```

```
display(nobel_data['sex'].value_counts())
top_gender = 'Male'

# Show the number of prizes won by the top 10 countries
display(nobel_data['birth_country'].value_counts().head(10))
top_country = 'United States of America'
```

```
1000
```

	sex
Male	
Female	

2 rows

	birth_country
United States of America	
United Kingdom	
Germany	
France	
Sweden	
Japan	
Canada	
Switzerland	
Netherlands	
Italy	

10 rows

```
# The decade with the highest ratio of US-born winners
nobel_data['usa_born_winner'] = nobel_data['birth_country'] == 'United States of America'
nobel_data['decade'] = (np.floor(nobel_data['year'] / 10) * 10).astype(int)
ratio_usa_winners = nobel_data.groupby(['decade'], as_index=False)['usa_born_winner'].mean()

# Display ratio of USA born winners per decade
display(ratio_usa_winners)

max_decade_usa = 2000
```

	decade	usa_born_winner
0	1900	
1	1910	
2	1920	
3	1930	
4	1940	
5	1950	
6	1960	
7	1970	
8	1980	
9	1990	
10	2000	
11	2010	
12	2020	

13 rows

```python
# Decade with highest proportion of female laureates
nobel_data['female_winner'] = nobel_data['sex'] == 'Female'
ratio_female_winners = nobel_data.groupby(['decade', 'category'], as_index=False)['female_winner'].mean()

# Find the decade with the highest proportion of female winners
max_female_winner = ratio_female_winners.loc[ratio_female_winners['female_winner'].idxmax()]

# Create a dictionary with the decade and the proportion of female winners
max_female_dict = {max_female_winner['decade']: max_female_winner['category']}
max_female_dict
```

```
{2020: 'Literature'}
```

```python
#first woman to receive a Nobel Prize and the category
nobel_data[nobel_data['female_winner']==True].nsmallest(1, 'year')
first_woman_name = 'Marie Curie, née Sklodowska'
first_woman_category = 'Physics'
```

```python
# Laureates who have won 2 or more nobel prizes
nobel_data.groupby('full_name').filter(lambda x: len(x)>1 )

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