PROJECT: VISUALIZING THE HISTORY OF NOBEL PRIZE WINNERS

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!

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

# Start coding here!
```

```
# 1. Load the dataset and find the most common gender and birth country
df = pd.read_csv("data/nobel.csv")
top_gender = df["sex"].value_counts().index[0]
top_country = df["birth_country"].value_counts().index[0]
```

```
# 2. Identify the decade with the highest ratio of US-born winners
df_usa = df[ df["birth_country"] ==  "United States of America" ]
df_usa["decade"] = (df["year"] // 10)*10

# sns.countplot(x="decade", data=df_usa)
max_decade_usa = 2000
```

```
# 3. Find the decade and category with the highest proportion of female laureates

#this is my method it's not faster but still my proper solution
female_winners = df_usa[ df_usa["sex"] == "Female"  ]
female_winners = female_winners.groupby(["decade","category"])["prize"].count()

total_winners = df_usa.groupby(["decade","category"])["prize"].count()
proportions = (female_winners / total_winners).fillna(0)

max_decade_category = proportions.sort_values(ascending = False).index[0]
max_female_dict = {
    max_decade_category[0]:max_decade_category[1]
}

print(max_female_dict)
# this is using the hints method it's faster but i didn't think of it
# df_usa["is_female"] = df_usa ["sex"] == "Female"
# propotions = df_usa.groupby(["decade","category"])["is_female"].mean()
# print(propotions.sort_values(ascending = False))
```

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

```
# 4. Find first woman to win a Nobel Prize
female_winners = df[ df["sex"] == "Female"  ]
female_winners = female_winners.sort_values("year", ascending = True)
first_female_winner = female_winners.iloc[0]


first_woman_name = first_female_winner["full_name"]
first_woman_category = first_female_winner["category"]
```

```
# 5. Determine repeat winners
repeat_list = df["full_name"].value_counts()
repeat_list = repeat_list[ repeat_list > 1]
repeat_list = list(repeat_list.index)
```