Project 5: Hurricane Study

Hurricanes often count among the worst natural disasters, both in terms of monetary costs and, more importantly, human life. Data Science can help us better understand these storms. For example, take a quick look at this FiveThirtyEight analysis by Maggie Koerth-Baker: [Why We're Stuck With An Inadequate Hurricane Rating System] (you should all read FiveThirtyEight, btw!).

For this project, you'll be analyzing data in the `hurricanes.csv` file. We generated this data file by writing a Python program to extract stats from this page: https://en.wikipedia.org/wiki/List_of_United_States_hurricanes. By the end of this semester, we'll teach you to extract data from websites like Wikipedia for yourself.

This project will focus on **conditional statements** and **loops**. To start, download `project.py`, `test.py` and `hurricanes.csv`. You'll do your work in Jupyter Notebooks this week, producing a `main.ipynb` file. You'll test as usual by running `python test.py` to test a `main.ipynb` file (or `python test.py other.ipynb` to test a notebook with a different name).

We won't explain how to use the `project` module here (the code in the `project.py` file). The lab this week is designed to teach you how it works.

This project consists of writing code to answer 20 questions. If you're answering a particular question in a cell in your notebook, you need to put a comment in the cell so we know what you're answering. For example, if you're answering question 13, the first line of your cell should contain `#q13`.

Questions and Functions

For the first four, you don't have to define any functions of your own. Instead you should just make use of the functions provided in the file `project.py` by calling the corresponding function that you need to solve a particular problem.

Q1: How many records are in the dataset?

Q2: What is the name of the hurricane at index 0?

Q3: How many deaths were caused by the hurricane at index 110?

Q4: How much damage (in dollars) was done by the hurricane at index 1?

Be careful! In the data, the number was formatted with a suffix, but you'll need to do some processing to convert it to this: `1430000000`.

While not required, you may wish to write a general function that handles "K", "M", and "B" suffixes (it will be handy later).

Q5: Is there a hurricane named Flossy?

To get full credit on this one, you are required to use a `break` to finish your loop early if Flossy is found.

Hint: here's a loop that prints every hurricane name. Consider adapting the code?

```
for i in range(project.count()):
    print(project.get_name(i))
```

Q6: How many hurricanes were named Floyd?

Write your code such that it counts all the variants (e.g., "Floyd", "FLOYD", "floyd", etc.).

Q7: How many total deaths are represented in the dataset?

Q8: What were the total damages across all hurricanes in the dataset, in dollars?

Remember that "K" stands for thousand, "M" stands for million, and "B" stands for billion! These may appear in the dataset, but the answer you compute (`792890014998`) should not use them.

Q9: What is the fastest MPH ever acheived by a hurricane?

Q10: What is the name of that fastest hurricane?

Q11: In what year did that fastest hurricane occur?

Q12: What is the slowest MPH in the dataset?

Function Suggestion:

We suggest you complete a function something like the following to answer the next several questions (this is not a requirement if you prefer to solve the problem another way):

```
# return name of deadliest hurricane over the given date range
def worst_in_range(year1, year2):
    worst_idx = 0
    for i in range(project.count()):
        pass # TODO: finish this code!
    return project.get_name(worst_idx)
```

Q13: what was the deadliest hurricane in the entire dataset?

You may assume all years are between 1900 and 2100 (this assumption applies to all questions).

Q14: what was the deadliest hurricane in or before 2016?

Q15: what was the deadliest hurricane between 2005 and 2016 (inclusive)?

Function Suggestion:

We suggest you complete a function something like the following to answer the next several questions (this is not a requirement if you prefer to solve the problem another way):

```
def decade_deaths(decade):
    pass
```

Hint: what is `year - year%10`? Try evaluating after putting different values in a `year` variable.

Q16: how many people died in the decade starting in 2010?

Q17: how many people died in the decade starting in 2000?

Q18: how many people died in the decade starting in 1990?

Q19: how many people died in the decade starting in 1980?

Q20: what was deadliest decade in the dataset?

Report the start year for that decade. Only consider round numbers For example, the ten years starting in 1990 should be considered as a possible worst decade, but the ten years starting in 1991 should not be.

Good luck with your hurricanes project! :)