

Homework 0

Due Thursday, January 30th at 11:59pm

The goal of this assignment is to provide you with some practice coding in Python and manipulating data.

You are encouraged to discuss the assignment in general with your classmates, and may optionally collaborate with one other student. If you choose to do so, you must indicate with whom you worked. Multiple teams (or non-partnered students) submitting the same code will be considered plagiarism.

Code must be written in a reasonably current version of Python (>3.0), and be executable from a Unix command line. You are free to use Python's standard modules for data structures and utilities, as well as the pandas, scipy, and numpy modules.

Data Wrangling with Python

For reasons I can't quite explain, I keep a detailed record of every time I put gasoline in my car. For this assignment, you will be writing Python code to read, process, and analyze this data using common Python tools and idioms. Whenever possible, you are encouraged to take advantage of Python's extensive library of built-in modules ("Don't reinvent the wheel!"). Remember, the [Python docs](#) are your friend.

Data Format

The file named `mustard_data.csv` contains almost 16 years of refueling records in [comma-separated value](#) (csv) format. The file has five columns (date, mileage, location, gallons, price per gallon) and 427 rows (including the header row). Note that some of the rows are missing one or more fields. When dealing with these cases, you should skip any rows with missing data.

Code

The included file `mustard_analytics.py` contains skeleton code that you'll be modifying. For each of the exercises, you'll need to flesh out the body of one of the function stubs (see the comments in the code for more detail). You are welcome to define additional functions if you'd like, but do not change the names or signatures of the existing functions.

Your program must be executable from a Unix(-like) command line. For example:

```
purple:~/$ python mustard_analytics.py /path/to/mustard_data.csv
```

Grading

We will run your program and examine the output for correctness. Note that we may run it on a data file with a different number of rows (but the same column format), so your code should be able to handle data files of arbitrary lengths. Your grade will be determined by how many of the exercises achieve the correct output, with partial credit being awarded.

What to Submit

You should submit the following to Gradescope:

- A modified `mustard_analytics.py`
- A `readme.txt`, containing
 - Your name
 - A fun fact about yourself
 - Notes or warnings about what you got working, what is partially working, and what is broken