# **INF 405 & 505, Fall 2021**

# **Programming Assignment One (PA1)**

**Assigned:** Monday, August 23, 2021

**Due:** Monday, September 13, 2021 (before midnight)

**Description: Reading in and answering simple queries about COVID-19 Pandemic Mask wearing Data from the Summer of 2020.**

This program will use as its primary resource a file of mask use behaviors based upon data compiled by the NY Times in the Summer of 2020. The program will read that in (as well as an auxiliary file, described below). The program will then present a simple interface to the user. This will allow the user to ask for statistics for the entire country, or for a specified US State-level entity (*i.e.* the 50 States, plus the District of Columbia and Puerto Rico). The program will repeat this until the user chooses to quit.

The input files

* Mask data by county: <https://github.com/nytimes/covid-19-data/tree/master/mask-use>
* State FIPS codes: State-geocodes-v2018.csv (Available on Blackboard)
* State names and abbreviations: State\_abbreviations.csv (Available on Blackboard)

The mask data file is explained on the Github page.

The FIPS code is the key to identifying counties and states. The FIPS code for which state a county is in is the first two characters of that county’s FIPS identifier (*e.g.* 17009 is a county in State “17” which is Illinois).

Specific Program Behavior

1. The program will read in the three files listed above. This will give it the data it needs to answer the different types of user queries.
2. The program will then operate in a loop. In the loop, the user can specify three different types of queries.
   1. Whole US data.
   2. Data for a specified US State
   3. Exit the program

The queries are detailed as follows:

* Whole US data:
  + By aggregating the data for all the counties, give the nationwide average of each of the mask wearing category percentages.
  + For every State, calculate and give
    - The statistics for each of the mask wearing category percentages in that state
    - The County with the highest best mask wearing percentage in that State. Best mask wearing percentage is the sum of the FREQUENTLY and ALWAYS percentages.
    - The County with the highest worst mask wearing percentage in that State. Worst mask wearing percentage is the sum of the NEVER and RARELY percentages.
  + Using the same method just described, output
    - the State with the highest best mask wearing percentage (aggregated)
    - the State with the highest worst mask wearing percentage (aggregated).
* Specified State Data
  + The user specifies a state by its two-letter code (e.g. “NY” for “New York”) and the program gives the same State data as is supplied for it for the nationwide query.
* Exit
  + If the user requests the program exit, the program terminates.
  + Otherwise it goes back into the loop prompting the user for its query.

## Hints:

* The individual underlying problems here are not difficult when considered one-by-one. The main complexity is that there is a lot going on. You will have to be organized, especially in your initial design of the program to keep things clear.
* You should test your program against some hand-calculated data for some of the States to help increase the likelihood that it is working. Pick States with a small number of Counties to make your life easier.
* Please check the Discussion Board for this program for questions and answers, clarifications, *etc*.

## Notes

* This should be a complete program in Python. You can organize it into a single file, or multiple files.
* The program must use comments to describe its overall operation and to identify the specific operation of significant lines in the code.
* This is an individual assignment, so please do your own work. Except as explicitly noted, no collaborating with others, looking up or soliciting answers from the internet or other sources.
* If you have difficulties, please use the Discussion Board set up for the assignment on Blackboard, ask questions, via Blackboard, or see the instructor.
* Please follow the submission details given below to get credit for this program.

## Submission

You must submit your source language file(s) containing your program via Blackboard. *Please* don’t email them to the instructor.

## Grading

The program is graded based upon *specifications*. Specifications are specific aspects of the assignment that are important. Each specification has a certain number of points associated with it. Each specification is all-or-none (that is, you get either zero or full credit for each specification). The total of your specification scores is your score for the program (out of 100)

### The Specifications

Programming

* Uses Valid basic Python (no PANDAS or other libraries) (4 points)
* The Code is Commented
  + Inline (2)
  + Overall (2)

The Assignment

* Reads in the mask data file (4)
* Reads in the State FIPS code file (4)
* Reads in the State Abbreviations file (4)
* Correctly processes the mask data file (4)
* Correctly processes the State FIPS code file (4)
* Correctly processes the State Abbreviations file (4)
* Contains the main loop (4)
* Implements the Whole US Option
  + Breaks data down by State (4)
  + Provides mask category for the State (4)
  + Mask category for the State is correct (4)
  + Provides Best County data for the State (4)
  + Best County data for the State is correct (4)
  + Provides Worst County data for the State (4)
  + Worst County data for the State is correct (4)
* Implements the Specified State Option
  + Correctly identifies the specified State (4)
  + Correctly handles an invalid State name (4)
  + Provides mask category for the State (4)
  + Mask category for the State is correct (4)
  + Provides Best County data for the State (4)
  + Best County data for the State is correct (4)
  + Provides Worst County data for the State (4)
  + Worst County data for the State is correct (4)
* Implements the Exit Option (4)