

## Computer Project #05

### Assignment Overview

This assignment focuses on the design, implementation and testing of Python programs to process data files.

It is worth 45 points (4.5% of course grade) and must be completed no later than 11:59 PM on Monday, February 22<sup>nd</sup>.

### Assignment Deliverable

The deliverable for this assignment is the following file:

`proj05.py` – the source code for your Python program

Be sure to use the specified file name and to submit it for grading via the **handin system** before the project deadline.

### Assignment Background

What do people earn in different jobs? What is the average annual pay for all full-time employees in the US? These are interesting questions. We grabbed a file of data named `national_M2014_dl.txt` from the data.gov website so you can answer those questions.

### Assignment Specifications

1. The program will prompt for a data file name. If the file is not found, keep prompting until the file is found. (Hint: See Code Listing 5.6 in the text, i.e. use a `while True` loop with a `try-except` block in the loop and a `break` to exit the loop. Note that since the text was written there is a new, more specific exception `FileNotFoundError` to use instead of the broader `IOError` used in that code listing).
2. The input file `national_M2014_dl.txt` is formatted as follows. The file is formatted by columns so string slicing is the way to extract data. Except for the header line (first line), all other lines are formatted as follows:
  - Occupation code: columns 0-10
  - Occupation title: columns 10-120
  - All other fields each occupy 13 columns

- Entry 7 (starting in column 172) is the Average Salary—the value used for this programming project. (Hint: these values are read as strings so they must be converted to numbers.)
  - Entry 3 (starting in column 120) is the Occupational Group. For this project we are only interested in rows that are coded “detailed” in this column—data in all other rows are ignored.
  - Entry 2 (starting in column 10) is the Occupation description. For this project that is the entry that will be referenced when searching for keywords (see item 3 next).
3. The program will prompt the user for a keyword. Your program will print a table listing all the occupations that contain the keyword, print the highest and lowest paying occupations that have that keyword in its title, and print the average salary and number of occupations having that keyword. Salaries and occupations should be formatted nicely in columns with commas in the numerical values. If there is only one matching occupation neither a max nor a min should be output; also, the word “occupation” should be singular in that case. If at the keyword prompt you simply hit enter, the keyword will be the empty string and all occupations will be selected. See the sample below. If the keyword doesn’t exist in the file, print an error message and do not do any calculations.

(Mathematical note: the average we are expecting in this project is a simple average of the values from the rows selected by the keyword. This is a simplification to make your programming easier and it is mathematically incorrect because we should be taking into account how many people have that average salary in calculating the overall average—that is, taking the average of averages is usually meaningless. You should convince yourself why this is so – feel free to ask your instructor or TA for clarification.)

4. The inputs are case insensitive, so if user enters “computer”, “COMPUTER” or “CompUter”, your program should find all the occupations having “computer” in the title. Keep in mind that title of occupations in the file are not all in lower case.

### **Assignment Notes**

1. You cannot use Python’s lists for this project.
2. Items 1-6 of the Coding Standards will be enforced for this project.
3. Lab06 should be a big help so check it out before doing this project
4. Project should work with any input file that has the same format. Your project will be tested using multiple input files that have different number of lines.
5. Sometimes there is no data in the file for average salary: there will be a single asterisk (\*) character in the field—ignore those values in your calculations. Trying to convert

that string to a number will generate an error so you need to check for that case before trying to convert. You can use a conditional (*if*) or use a *try-except* block—whichever you prefer.

6. Keyword input must be case insensitive so if a user enters “computer”, “COMPUTER” or “CompUter”, your program should find all the occupations having “computer” in the title. Keep in mind that title of occupations in the file are not all in lower case.

## Suggested Procedure

- *Solve the problem using pencil and paper first.* You cannot write a program until you have figured out how to solve the problem. This first step may be done collaboratively with another student. However, once the discussion turns to Python specifics and the subsequent writing of Python statements, you must work on your own.
- Divide-and-conquer! I start programs that read data from a file by simply opening the file and printing all the contents. Since there is a header line I skip that next (hint: try `readline`). Next I print out all the data from the columns that I am interested in (Entries 2, 3, and 7 in this project). With that framework you can now select by keyword and print out those columns. Then count and find the average. Then min and max.
- Here is an algorithm to find the minimum value of a collection of numbers that you are reading through. Start with a huge value and whenever you find a smaller value make that new value the minimum value. After working your way through all the numbers you end up with the minimum value. Finding the maximum is similar.

```
minimum_value = 10e10 # something huge
if new_value < minimum_value:
    minimum_value = new_value
```

If you want to also gather some information about the minimum value, e.g. the occupation title in this case, simply add a variable to hold that information and update the information whenever you have found a new minimum value.

- Use the **handin system** to turn in the first version of your solution.
- Cycle through the steps to incrementally develop your program:
  - Edit your program to add new capabilities.
  - Test the program and fix any errors.
  - Regularly use the **handin system** to submit the current version of your solution. That way, if something happens and you cannot submit a final version there will be a partial version within handin so you can get partial credit.

- Be sure to log out when you leave the room, if you're working in a public lab.

### Sample Output

```
Enter a file name: xxxxx
File not found. Try again.
```

```
Enter a file name: national_M2014_dl.txt
```

```
Enter an occupational key word: waiter
```

```
Salary    Occupation
$ 21,640   Waiters and Waitresses
```

```
Across 1 occupation the average salary was $21,640
```

```
In [66]:
```

```
runfile('/Users/enbody/Documents/cse231/SS16/Projects/Project05/proj05_rje.py', wdir='/Users/enbody/Documents/cse231/SS16/Projects/Project05')
```

```
Enter a file name: national_M2014_dl.txt
```

```
Enter an occupational key word: real estate
```

```
Salary    Occupation
$ 65,880   Property, Real Estate, and Community Association Managers
$ 58,010   Appraisers and Assessors of Real Estate
$ 80,420   Real Estate Brokers
$ 55,530   Real Estate Sales Agents
```

```
Max: $ 80,420   Real Estate Brokers
```

```
Min: $ 55,530   Real Estate Sales Agents
```

```
Across 4 occupations the average salary was $64,960
```

```
In [67]:
```

Enter a file name: national\_M2014\_dl.txt

Enter an occupational key word: computer

Salary	Occupation
\$136,280	Computer and Information Systems Managers
\$113,190	Computer and Information Research Scientists
\$ 87,320	Computer Systems Analysts
\$ 82,690	Computer Programmers
\$ 79,770	Network and Computer Systems Administrators
\$100,710	Computer Network Architects
\$ 51,500	Computer User Support Specialists
\$ 66,140	Computer Network Support Specialists
\$ 85,520	Computer Occupations, All Other
\$110,650	Computer Hardware Engineers
\$ 99,660	Electronics Engineers, Except Computer
\$ 80,730	Computer Science Teachers, Postsecondary
\$ 41,240	Computer Operators
\$ 30,200	Office Machine Operators, Except Computer
\$ 38,450	Computer, Automated Teller, and Office Machine Repairers
\$ 37,920	Computer-Controlled Machine Tool Operators, Metal and Plastic
\$ 50,200	Computer Numerically Controlled Machine Tool Programmers, Metal and Plastic

Max: \$136,280 Computer and Information Systems Managers

Min: \$ 30,200 Office Machine Operators, Except Computer

Across 17 occupations the average salary was \$76,010

... 5001 ...

This is the tail end of the 800+ lines when nothing was entered at the keyword prompt (except a carriage return, of course).

```
$ 55,250 Subway and Streetcar Operators
$ 59,950 Rail Transportation Workers, All Other
$ 40,970 Sailors and Marine Oilers
$ 79,180 Captains, Mates, and Pilots of Water Vessels
$ 41,130 Motorboat Operators
$ 74,600 Ship Engineers
$ 46,210 Bridge and Lock Tenders
$ 21,610 Parking Lot Attendants
$ 22,660 Automotive and Watercraft Service Attendants
$ 46,540 Traffic Technicians
$ 70,820 Transportation Inspectors
$ 27,060 Transportation Attendants, Except Flight Attendants
$ 35,940 Transportation Workers, All Other
$ 34,020 Conveyor Operators and Tenders
$ 53,550 Crane and Tower Operators
$ 45,640 Dredge Operators
$ 44,160 Excavating and Loading Machine and Dragline Operators
$ 47,500 Loading Machine Operators, Underground Mining
$ 48,820 Hoist and Winch Operators
$ 33,320 Industrial Truck and Tractor Operators
$ 23,340 Cleaners of Vehicles and Equipment
$ 27,180 Laborers and Freight, Stock, and Material Movers, Hand
$ 30,630 Machine Feeders and Offbearers
$ 23,040 Packers and Packagers, Hand
$ 55,430 Gas Compressor and Gas Pumping Station Operators
$ 46,700 Pump Operators, Except Wellhead Pumpers
$ 48,590 Wellhead Pumpers
$ 36,030 Refuse and Recyclable Material Collectors
$ 54,830 Mine Shuttle Car Operators
$ 44,540 Tank Car, Truck, and Ship Loaders
$ 37,840 Material Moving Workers, All Other
```

Max: \$246,320 Anesthesiologists

Min: \$ 19,030 Cooks, Fast Food

Across 816 occupations the average salary was \$55,383