Project: How diverse is this group? – An Analytic Tool

Problem Description and Solution

For any group, given their demographic information such as gender, race/ethnicity, age, and similar, this program can calculate what percentage is M/F, White/Black/Hispanic /Asian/etc, 20-30/30-40/40-50/etc years old. This analysis can be used in hiring decisions or bias investigations in companies, as well as looking at data from other groups, like a census. Code can be easily edited to add, change, or remove the categories of data to suit any dataset, making this a versatile base program.

The end user would need to acquire a database, compiled by asking the group of people in question to answer some simple questions about their demographics. This will include: Gender [options: Male, Female, Other], Race/ethnicity [White, Black or African American, Hispanic, American Native, Asian, Other], Age [Input as integer], Hourly wage [Input as integer].

Technical Approach

Input: This database will have the input values for each question on one line per person, answers separated by a comma. The path of the .txt file containing the database will be entered by the end user.

>Each line will be read one by one, split and stripped, and contents will be a tuple.

>This tuple will be the value, whereas each entry will have a key assigned – starting at 1, moving through integers.

>There will be a function to count the number of times an item appears in the results. This will be converted into a percentage.

>For some categories, such as age or salary, a similar code will look at ranges of data.

>There will be exception-handling and error messages.

Output: The program will show a list of statistics of the demographics, and will ask the end user if they would like to save this information. If answered yes, the program will create a new document with the calculated results, formatted for easy transfer to a spreadsheet program for data visualization. The end user will be able to name this new file.

Tests

There will be different test files, including one blank, and two with content which would not be usable by the program (e.g. other text and an image), as well as two with different data.