**This program will allow the user to view gender pay ratios in the US and sort information by state and by year.**

**Pseudocode w/ added Abstraction:**

**START**

**IMPORT SQL**

Establish **CONNECTION** to salary database

Create **CURSOR** to communicate with data

Ask user to **INPUT** the state they want to know about and assign to **VARIABLE**

**DEFINE** variable to be equal to user input

Use **CAPITALIZE** method to clean user input

**DEFINE** the select statement query as **VARIABLE**

Execute the **QUERY**

**IF** Query **NOT** in results

**PRINT** (“No data available”)

**GET** the set of results

**CLEAN** results by stripping/removing unwanted rows/columns

**ITERATE** through the results for column/row values

Create **FOR** loop to iterate through all results

**CLOSE** connection to database

Create **VARIABLE** called ‘gap’from disparity in salaries from results

**DEFINE** variable function with two parameters

Set **PARAMETERS** to be based off male and female salaries

Set parameter **TYPE** to be int

**RETURN** average value based off difference between parameters

**PRINT** variable for user to see wage ‘gap’

**RETURN** query

**END**

**NEW PSEUDO (Note: the current scope of our project and the amount of coding we can do with the database we have currently is very limited. We need to spend more time coding, but as it is, we need a new database with more information first. One solution we have come up with is to use census data from** [**https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-people.html**](https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-people.html)**. Using the databases of table p-6, we can sort by gender, race, region, and year, define a gender gap variable, and return a value specific to the user’s inputs. We would no longer be able to sort by state, but the additional information would more than make up for that. We can also possibly generate graphs showing the wage gap over time for a specific group in a specific region. We currently have not had time to code this but we have begun putting together the pseudocode should this change be approved and we all collectively decide to continue down this route.)**

**DEFINE a variable function with three parameters: race, region, and year**

* **DEFINE** variable function with two parameters
  + Set **PARAMETERS** to be based off male and female salaries
  + Set parameter **TYPE** to be int
  + **RETURN** average value based off difference between parameters
* **READ database file corresponding to race variable**
* **RETURN values for region by year**
* **RETURN wage\_gap**

Ask user to **INPUT** the race, region, and year they want to know about:

* **PRINT** all regions, so users know region options
* **DEFINE** variable to be equal to user input
* Use **CAPITALIZE** method to clean user input and set it to be the same as the values/titles from the databases

**USE** user input for race, region, and year as parameters for defined function

**PRINT** wage\_gap

**GRAPH** relevant group’s wage gap over time using matplotlib based off of user inputs

Program needs to be able to:

-pull information from the database

-have some form of user input in order to get a proper query

-sort through/query information based off user input (race, region, year)

-output information and graphs

research:

<https://pynative.com/python-mysql-select-query-to-fetch-data/>

This website was used to understand SQL queries, something which as of yet we have not incorporated but which we may be using in our final project.

<https://code.tutsplus.com/tutorials/database-handling-in-python--cms-25645>

<https://statusofwomendata.org/state-by-state-rankings-on-womens-and-mens-employment-and-earnings-2013/>