**READ ME:**

Before running this program, you have to install matplotlib package in the computer, instructions are given below:

**To install matplotlib on a Windows system, open a Command Prompt window and enter this command:**pip install matplotlib

**To install matplotlib on a Mac or Linux system, open a Terminal window and enter this command:** sudo pip3 install matplotlib

**Following are the modules needed to be import to run the program:**

Import **matplotlib.pyplot as pt** where pyplot is the module and pt is the alias name, import **csv module** in order to use DictReader and DictWriter class to read and write the csv file and plot Gender\_Comparison, AnimalStrain\_Comparison and import another module **Area\_File** in order to write area file and use that file to plot CageVsStatus and ParameterVsStatus comparison.

**Below is the file name used for Main\_project\_code.py:**

Final\_project.csv

**Below is the file name will create after importing Area\_File module in the Main\_project\_code.py and used to plot CageVsStatus also ParameterVsStatus:**

Area.csv

**Below are the file names that will create after Main\_project\_code.py run:**

CancerReady.csv

CancerNotReady.csv

gender\_comparison.png

animal\_strain\_comparison.png

CageVsStatus.png

ParameterVsStatus.png

**Unit of measurement used:**

Cubic Millimeter – Length and Width values are taken from the Final\_project file and volume has been ignored.

**Introduction:**

In this project we are using Final\_project.csv file provided by the research company where they do cancer testing on the tumor those are greater then 500 cubic millimeters in area found in either side of mouse body.

So,the data file provided by the research company has the following data, on the top row were the header[Cage,AnimalStrain,Parameter,Width,Length,Sex] and below that row were the values related to the headers respectively. From this data file we have decided to calculate area in cubic millimeter taking length and width provided in the Final\_project.csv file and created to file one CancerReady and another one is CancerNotReady.

**Step by Step Explanations:**

**In the 1st step:** We are opening the Final\_project.csv file to read and write two different file that is CancerReady.csv and CancerNotReady.csv files where in CancerReady.csv file we have added two columns one column is for area which will have area value greater then equal to 500 and other column will have status value that is Ready. Now in the other File that is CancerNot Ready.csv file we have again added two columns, one column is for area which will have area value lesser then 500 cubic millimeter and other column will have status value that is NotReady.

**In the 2nd step:** With the help of matplotlib library we are showing gender comparison in order to show which gender (Female or Male) helped more for cancer testing purpose. In this part we are opening CancerReady.csv file and reading the file using DictReader and after that using for loop to iterate through the row Sex and see if the row has F then increment the counter allotted to the female variable else increment the counter allotted to the male counter. Once we got the value we have called the matplotlib to plot pie chart and show in percentage which gender (Female or Male) helped more for cancer testing purpose.

**In the 3rd step:** With the help of matplotlib library we are showing animal strain comparison in order to show which strain (NOG Vs AthymicNude) helped more for cancer testing purpose. In this part we are opening CancerReady.csv file and reading the file using DictReader and after that using for loop to iterate through the row Strain and see if the row has NOG then increment the counter allotted to the NOG variable else increment the counter allotted to the Athymic Nude counter. Once we got the value we have called the matplotlib to plot pie chart and show in percentage which strain (NOG Vs AthymicNude) helped more for cancer testing purpose.

**In the 4th Step:** In order to show the comparison between Cage and Status that is which cage is ready for the cancer testing and which cage is not ready for the cancer testing, we have written a module Area\_File.py in this we are reading the Final\_project.csv file in order to write Area.csv file. In Area.csv file will have headers [Cage,AnimalStrain,Parameter,Width,Length, Sex, Area, Status] including all the original values along with all the area values(below 500 and greater then equal 500) and Status(Ready and NotReady).

In this we have used for loop to iterate through the Final\_project file and csv module classes DictReader(Read data in dictionary form – Key:Value pair ) and Dictwriter(Write data where the fieldname are the key or header and below them are the value).We have calculated the area using Width and Length from the Final project and write it in the Area.csv file along with status.

In the main\_project\_code.py we have imported the Area\_File.py module which will write the area.csv file. In the CageVsStatus function we are reading the data from area.csv(Cage values and Status values) in order to show which Cage is ready for testing and which cage is not ready for testing. With the help of matplotlib library we have plotted Bar Chart to show the result graphically. In this function that is CageVsStatus we have used list to append values those are index[0](Cage) and from the index[8](Status) in the empty list. After getting the list we have plotted cage values in the x-axis and status in the y-axis. In this we have used 70 data set.

**In the 5th Step:** In this function we are showing comparison between ParameterVs Status. In this we reading the Area.csv file using DictReader which has been written by importing the Area\_File. We are using for loop to iterate through the row Parameter and Status and see if the row has the situation one (Parameter – Right or left equal to Ready) then increment the counter allotted to this situation one variable else increment the counter allotted to the situation two(Parameter – Right or left equal to NotReady)counter. Once we got the value we have called the matplotlib to plot pie chart and show in percentage the comparisons between Parameters(Left & Right) and Status(Ready)Vs Status Not Ready.

At the end of the we are calling main for execution.