

Postgraduate Certificate in Cloud Native Computing

Postgraduate Certificate in Software Design with Artificial Intelligence

**Postgraduate Diploma in in Software Design with Cloud Native Computing**

Postgraduate Diploma in Software Design with Artificial Intelligence

**MSc in Cloud Native Computing**

**MSc in Software Design with Artificial Intelligence**

Applied Scripting Languages

**Assignment Part 3: Analysis and Visualisation**

**Student ID: A00279933**

***Death and Confirmed Cases with Correlation, variance and Covariance with Visualisation of both Cases Completed.***

Contents

[Contents 1](#_Toc55933026)

[Additional Analysis, Visualisation, Results and Conclusions 2](#_Toc55933027)

[Program Design 6](#_Toc55933028)

[Unit Testing 9](#_Toc55933029)

[Reflective Learning Log 13](#_Toc55933030)

[References 19](#_Toc55933031)

**ADITIONAL ANALYSIS, VISUALISATION, RESULTS AND CONCLUSIONS**

**ADDITIONAL ANALYSIS**

In the Novel Corona virus dataset **[1],[2]** in this part of the assignment, Additional Analysis such as Correlation, Variance and Covariance **[3]** has been carried out and the virtualization with both Death and Confirmed cases of the covid-19 dataset. First let us discuss about the correlation. Correlation is statistics which is about the relationship between (which can be casual or not a casual) two either random variable or bivariate data in statistics. Program has been divided into three parts, First- Main file, Second -Function File, Third- Test File.

**Correlation:**

Therefore, I have created a **“calc\_corr”** function for the correlation, which is to calculate the Both Death and Confirmed cases of the covid-19 data, First is to calculate the x-means and y-means of the variables, from that using list comprehension calculate the x-deviations and y-deviations, using the zip function to combine the two list (x,y deviations)later doing the both squared deviation of x and y values, and finally using the sum of x-y deviations and the sq. root of the y deviations correlation is calculated Below is the result of the Correlation between the death and the Confirmed case dataset variables.



**Variance:**

I have created a **“calc\_Varience”** function for thevariance, which is to calculate the how much the data for a variable varies from its mean, first is to calculate the mean, next is to find theith observation of the variable and then calculate the Ith variable-mean times squared, next is to calculate the length of the observation ( Variable records) -1 .And now the Variance is calculated as the sum of Ith variable-mean times squared to the length of the variable minus 1. Below is the result of the variance of Death case and Confirmed cases of the covid-19 dataset.



**Covariance:**

I have created a” **Calc\_covar”** function for thecovariance, which is to calculate the relation between two variables which are scale dependent, i.e., which depends upon how much does the variable change with respect to another variable changes. Below snippet represents the covariance**.,** First is to calculate themean of the x and y variable values, next is to find the first observation i.e.: ith of the x and y value and next is to find the number of observations, i.e..: length of the variable minus-1, finally it is calculated as sum of ith of the x- mean of x times ith of the y- mean of y to the number of observations minus 1. Below is the Result of the Covariance of the death case and the Confirmed case of the covid-19 dataset.



**VISUALISATION**

in Novel coronavirus dataset, Virtualization has bee carried out, plots such as, box plot, violin plot, scatter plot, Pie Chart

**SCATTER PLOT:**

This scatter plot shows the how is the difference between the Death and Confirmed Cases of the Novel corona 2019.database

Chart, line chart

Description automatically generated

**VIOLIN PLOT**

Violin Plot: This shows the Death of the Novel corona Virus dataset of the first 20 instances

Chart

Description automatically generated

**PIE CHART**

Pie Chart for the Death cases has been done, since the cases of the death are extremely high, I have taken first 1000 instances for the image virtualization

**Chart, pie chart

Description automatically generated**

**BOX PLOT.**

Box plot has been done between the Confirmed cases and the death cases of the first 1000 instances

**Chart, box and whisker chart

Description automatically generated**

**OVERALL RESULTS:**

Below are the overall results of the Covid-19 Confirmed and the Death Cases, I would like to note the readers that the Least recorded death has been given -178 because of the mismatch adjusted data updating on the dataset similarly Least recorded in confirmed cases as well

**Out File Results**

**Text

Description automatically generated**

**Program Results:**

**Text

Description automatically generated**

**RESULTS AND CONCLUSIONS**

Results have shown how far the Covid-19 has been affected the people and its very sad to look about the death caused because of Covid-19 and from studying about the both Death and Confirmed cases of the Covid-19, we can tell that from January 2020 till the November 15, 2020, death cases percentage is 3.4% when compared to the total infect of nearly 65 Million people confirmed, from this study which tells about the decline of the spread has started.

**PROGRAM DESIGN**

In the Novel corona virus dataset, in the Part-3 of assignment we are performing the analysis of the death and confirmed variables of the dataset and we analyse the calculations by creating separate functions such as Mean, Median, Mode, Range, Standard Deviation, Correlation, Variance, Covariance and the General function: Minimum, Maximum, length of the record, percentage of the Death and Confirmed Cases of Covid-19 dataset and we performed the Virtualization of both variables.

* BOX PLOT
* PIE CHART
* VIOLIN PLOT
* SCATTER PLOT

OUTPUT

COVID-19 DATABASE

**DEATH AND CONFIRMED**

MAX, MIN, LENGTH, PERCENTAGE

**Functions used in the program**

**Function “get\_file”:**

This function which reads the input datafile “COVID\_19” dataset and it has the empty list of variables and which are appended with comma separated to read each line in the datafile and which is returned Death, Confirmed and Province

**Function “calc\_mean”:**

This function which is created for the calculation of the **mean** with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the file is a zero-division error and type error. And it returns the mean. The purpose of this function is to calculate the mean (Average) of the total death cases of worldwide COVID-19 data and confirmed cases, which is calculated by the sum of the number by total length of number

**Function “calc\_median”**

This function which is created for the calculation of the **median** with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And it returns the **median**. The purpose of this function is to find the Median of the death case and confirmed cases, where the records are ordered ascending from first to last record available on the dataset and it takes the middle value divided by 2 only if the totoal record is even, when it is odd it takes the exactly the middle position, this can be determined as an approximitae of an avaerage

**Function “calc\_mode”**

This function which is created for the calculation of the **mode** with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And the function returns the **mode.** The purpose of the mode function is to calculate the Mode, the most frequent death case and confirmed cases record in the Covid 2019 dataset.

**Function” calc\_Range”**

This function which is created for the calculation of the **Range with** the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And the function is returned by the **range.** The purpose of the range function is to calculate the Range is the diffeence between the maxamium number of death case minus the minimum number of death case similarly to find diffeence between the maxamium number of Confirmed case minus the minimum number of confirmed case

**Function** **“calc\_standardDeviation”**

This function which is created for the calculation of the **standard Deviation** with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And returns the **s**tandard Deviation. The purpose of the standard deviation is to calculate thethe spred out of the data and it tells about how far each of our observed Death case value and confirmed case value is far from the meanof the death case value and confirmed case value of Covid-19 dataset.

**Function** **“calc\_general”**

This function which is created for the calculation of the **general** with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. The purpose of the **general** function is to calculate the maxamium number of death case minus the minimum number of death case similarly to find diffeence between the maxamium number of Confirmed case minus the minimum number of confirmed case and function is retuned by maxamium, minimum, length.

**Function** **“calc\_varience”**

This function which is created for the **variance** calculation of the with the parameter is **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And returns the **variance**. The purpose of the variance is to calculate the how much the data for a variable varies from its mean.

**Function** **“calc\_covar”**

This function which is created for the calculation of the **Covar** with the parameter **values** in list and mean function is made with the try bloc. And except to print error if the calculation is a zero-division error and type error. And it returns the **Covar,** the purpose of the functions covar is towhich is to calculate the relation between two variables which are scale dependent, i.e., which depends upon how much does the variable change with respect to another variable changes.

**Function** **“calc\_corr”**

This function which is created for the calculation of the **mode** with the parameter **values** in list and mean function is made with the try bloc and except to print error if the calculation is a zero-division error and type error. And it returns the **corr.** The purpose of the **Corr** function is to calculate the relation between two variables which are scale dependent, i.e., which depends upon how much does the variable change with respect to another variable changes.Here we use Both Death and Confirmed cases of the covid-19 data

**Function** **pie\_chart, barchart, scatterplot, violinplot**

The function for the pie chart, scatter plot, box plot, violin plot has been created and called in the main file and which are used to teel about the Death and the confirmed cases overall -

**UNIT TESTING**

In the unit testing part of the assignment we have, created the test case functions such as

calc\_mean, calc\_median, calc\_mode, calc\_Range, calc\_standardDeviation, calc\_general, calc\_varience, calc\_covar, calc\_corr and performed each individual testing.

Each case is tested here.

**Test\_Calc\_Mean**:

Here the testing of the Mean function is carried out by giving a new **Test\_Calc\_Mean** function and then assert the main function from **Calc\_mean** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the calc\_mean with the numbers and check where it will return the true or false.

Below is the snippet of the results.

Text

Description automatically generated

**Test\_calc\_median:**

Here the testing of the Mean function is carried out by giving a new **Test\_Calc\_Median** function and then assert the main function from **calc\_median** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_median** with the numbers and check where it will return the true or false.

Below is the snippet of the results.

**Text

Description automatically generated**

**Test\_Calc\_Mode,**

Here the testing of the Mean function is carried out by giving a new **Test\_Calc\_Mode** function and then assert the main function from **Calc\_Mode** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the calc\_mode with the numbers and check where it will return the true or false.

Below is the snippet of the results.

Text

Description automatically generated

**Test\_Calc\_Range:**

Here the testing of the Mean function is carried out by giving a new **Test\_Calc\_Range** function and then assert the main function from **calc\_range** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_range** with the numbers and check where it will return the true or false.

Below is the snippet of the results.

****

**Test\_calc\_standardDeviation:**

Here the testing of the Mean function is carried out by giving a new **Test\_ standardDeviation** function and then assert the main function from **calc\_standardDeviation** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_standardDeviation** with the numbers and check where it will return the true or false.

In the Standard deviation test, we used 3 types of test cases and each with numbers and the letters, which showed the correct working of the function,

Below is the snippet of the results

**A picture containing graphical user interface

Description automatically generated**

**Test\_calc\_corr:**

Here the testing of the Mean function is carried out by giving a new **Test\_calc\_corr** function and then assert the main function from **calc\_corr** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_corr** with the numbers and check where it will return the true or false.

In the Correlation test function, we have created 4 test cases, each with number instances and last with the letter, the results have shown the exact function test output, below is the result snippet for the reference.

**Text

Description automatically generated**

**Test\_calc\_general:**

Here the testing of the Mean function is carried out by giving a new **Test\_calc\_general** function and then assert the main function from **calc\_general** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_general** with the numbers and check where it will return the true or false.

Here the Test\_calc\_general has the three internal functions in it such as the Minimum, Maximum, length, so here to test we have given all three values in order.

Below is the snippet of the results

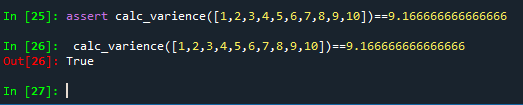
**Text

Description automatically generated**

**Test\_calc\_varience:**

Here the testing of the Mean function is carried out by giving a new **Test\_calc\_varience** function and then assert the main function from **calc\_varience** and given numbers to test, then we need to run the program, and which will not print any: means it is correct and we can further test by taking only the **calc\_varience** with the numbers and check where it will return the true or false.

Below is the snippet of the results

****

**Test\_calc\_covar:**

Here the testing of the Mean function is carried out by giving a new **Test\_calc\_covar** function and then assert the main function from **calc\_covar** print any: means it is correct and we can further test by taking only the **calc\_covar** with the numbers and check where it will return the true or false.

Here in the Covariance, we are giving two parameter values

Below is the snippet of the results

**Text

Description automatically generated**

**REFLECTIVE LEARNING LOG**

**Entry Number**: 1

**Date:** 11th December 2020

**Objective:** File Loading and adding additional analysis

**DESCRIPTION OF WORK DONE**

After finishing the part 2 assignment, the plan was to analyse further with adding additional variables which is Confirmed Cases of Covid-19 dataset and perform virtualization and additional analysis. After starting the program and added the functions for the Mean, Median, Mode, Range, General: Maximum, Minimum, Total Length of the record and Percentage. Analysis of Standard Deviation, Correlation, Variance and Covariance is the additional work.

Program is divided into 3 parts, first is the main part (for interactive Menu board, calling all functions from the function file part), second is the functions and third is the test part.

**REFLECTIONS**

After running the program faced with packed and unpack value errors,

Below code snipped are error which caused during the when the list of variables is declared in the definition i.e.:

Text

Description automatically generated

Which return Value Error and too many values to unpack (expected 2) this type of similar error we have encountered during the assignment part-1 and which was solved by verifying the variables which are declared correctly in the return of the function or not and this error was nullified

Program started to run but it returned such as Division by zero Error (Zero Division Error), Mean cannot be calculated and looking at the try and except block which returned it

Text

Description automatically generated

While looking the code, there was no problem and when I looked the dataset to know about the problem, thereby found that the dataset was empty and that is the reason which returned Zero Division Error and entered the exact dataset Csv File. After this program was performing well based upon the basic analysis which was divided into main and function files

For Additional Analysis, I calculated correlation, Variance which is a measurement of how much the data for a variable varies from mean of it. Below snippet represents the variance.

A picture containing text, watch

Description automatically generatedText

Description automatically generated

And Covariance which is the measurement of the relation between two variables which are scale dependent, i.e., which depends upon how much does the variable change with respect to another variable changes. Below snippet represents the covariance**.[3]**

A picture containing text

Description automatically generatedText

Description automatically generated

After this I encountered again the problem while running the covariance and the variance.by converting these formulas into written code. For the variance I faced the same problem and later on only when I looked my code, I found that I have written as var= sum(x-mean) \*\*2 for x in values)/ len(values)-1, which caused the similar error which is shown below (Error Image) and I have corrected as (len(values)-1) which is nothing but N-1

Corrected Code: For Variance:

Text

Description automatically generated

Error Image: For Covariance

Text

Description automatically generated

After understanding the error from the N-1, and in covariance we calculate the mean1.And x\_cov which is the first observation – mean, which is squared, similarly for y-cov we are doing and mean 2, for N which is the length (number of observations), I have declared sum =-0 for so that it counts and using the for loop for I in range of the length of the variable (since both x and y will be having same length) checking the dataset and keep this mind then performed. then summ = summ (first value[i] of the x variable times the first value [i] of the y variable and finally covar = sum/(N-1)

Below is the code snippet for the covar after error:

Text

Description automatically generated

From this I understood to look back once I program and to verify code and to solve immediately during coding and checking do that it does not prologue the error and reduces the stress of the error and you can get the performance improvement and motivation. And I understood the new concept of the variance and covariance which was not implemented in the class work and its very important to know about during the statistics part. Along with the correlation which is done as an additional work

**PLANNING**

My next plan is to implement the virtualization, interactive menu and printing the out file automatically in the loop without any special menu and save the file in the application directory.

**Entry Number:** 2

**Date:** 11th December 2020.

**Objective:** Virtualization, Front-End and

**DESCRIPTION OF WORK DONE**

After finishing the Analysis, we need to do the virtualization using the Math plot library function and we will be using the plots like violin plot, Box plot, Scatter plot, Pie chart

Next task is to make the interactive menu by using the while loops and, to print the out file automatically

**REFLECTIONS**

since I have divided the program into 2 different parts I was bit confused how do we integrate the outfield and then I decided to declare the out file before the while loop so that all loops print automatically without any separate menu.

Below is the snippet howe I used the out file

Text

Description automatically generated

This snippet for calling the function **Pie \_Chart, Box\_Plot, Scatter\_Plot, Violin\_Plot,**

Text

Description automatically generated

Printing the Charts for the out file directly without any separate menu on the choice:3 Of the program

**Text

Description automatically generated**

Which worked well and the output filer was saved, and it run well Below is the output result,

Graphical user interface, text, application, email

Description automatically generated

And python documentation creating was the bit challenging step and here is the snippet of the pydoc generated

A picture containing text

Description automatically generated

**PLANNING**

In the next part of the assignment, Will be implemented using the Powerful library like pandas, NumPy, and virtualization and I have planned to add the speech recognition function for the interactive menu.

**REFERENCES**

**[1] Kaggle Dataset:**

[**https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset**](https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset)

**[2] World Health Organization:**

[**https://www.who.int/health-topics/coronavirus#tab=tab\_1**](https://www.who.int/health-topics/coronavirus#tab=tab_1)

**[3] Variance and Covariance Additional Analysis:**

**https://www.pythonfordatascience.org/variance-covariance-correlation/**