**Data Science and AI Module – Continuous Assessment 2023(100%)**

**BSc (Hons) Applied Computing Yr4 & BSc (Hons) Computing Yr4**

**Part B – Data Science (50%)**

Select a data set related to your chosen AI application, it may be simple.

# Document the Process description (description of dataset, question you are trying to answer, aims, steps in cleaning, feature selection, justification of algorithms/packages used) (15%)

## Description of dataset

The data set that I was going to use is as follows:

COVID-19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_State.csv

This dataset is of covid19 reported cases which included confirmed admitted by age of 18 to 80+ patients and categorized as confirmed, confirmed coverage, suspected and others etc.

**Question you are trying to answer:**

1. Create a bar-plot represents all the previous day admitted adult covid19 confirmed cases from 18-80+?
2. Create a bar-plot represents all the previous day admitted adult covid19 suspected cases from 18-80?
3. Comparison between previous day admitted adult covid19 confirmed cases from 60 to 80+?
4. Comparison between previous day admitted adult covid19 suspected cases from 60 to 80+?
5. Comparison previous day admission adult covid19 between suspected and confirmed from the age of 60 to 80?

**Aims:**

My first aim is to find out the records of the previous day admitted adult covid19 confirmed cases from 18-80+ and create a bar-plot which describe this information.

Second aim is to find out the records of all the previous day admitted adult covid19 suspected cases from the age of 18y/o to 80 year old.

Third aim is make a comparison between previous day admitted adult covid19 confirmed cases from the age of 60 to 80+ and same of suspected cases form 60 y/o to 80 y/o?

Fourth aim is to make a comparison between suspected and confirmed admission adult covid19 on previous day from the age of 60 to 80.

**Feature selection:**

To perform the feature selection, we will chose the selective columns to findout the previous day admitted adult covid19 confirmed cases from 18-80+ and previous day admitted adult covid19 suspected cases from 18-80

To perform this task either you use drop() function from pandas library or use dataFrame() from pandas which allow to choose selective columns to perform data cleaning.

We use dataFrame() methods to choose the columns because the dataset have many columns.

**Justification of algorithms/packages used:**

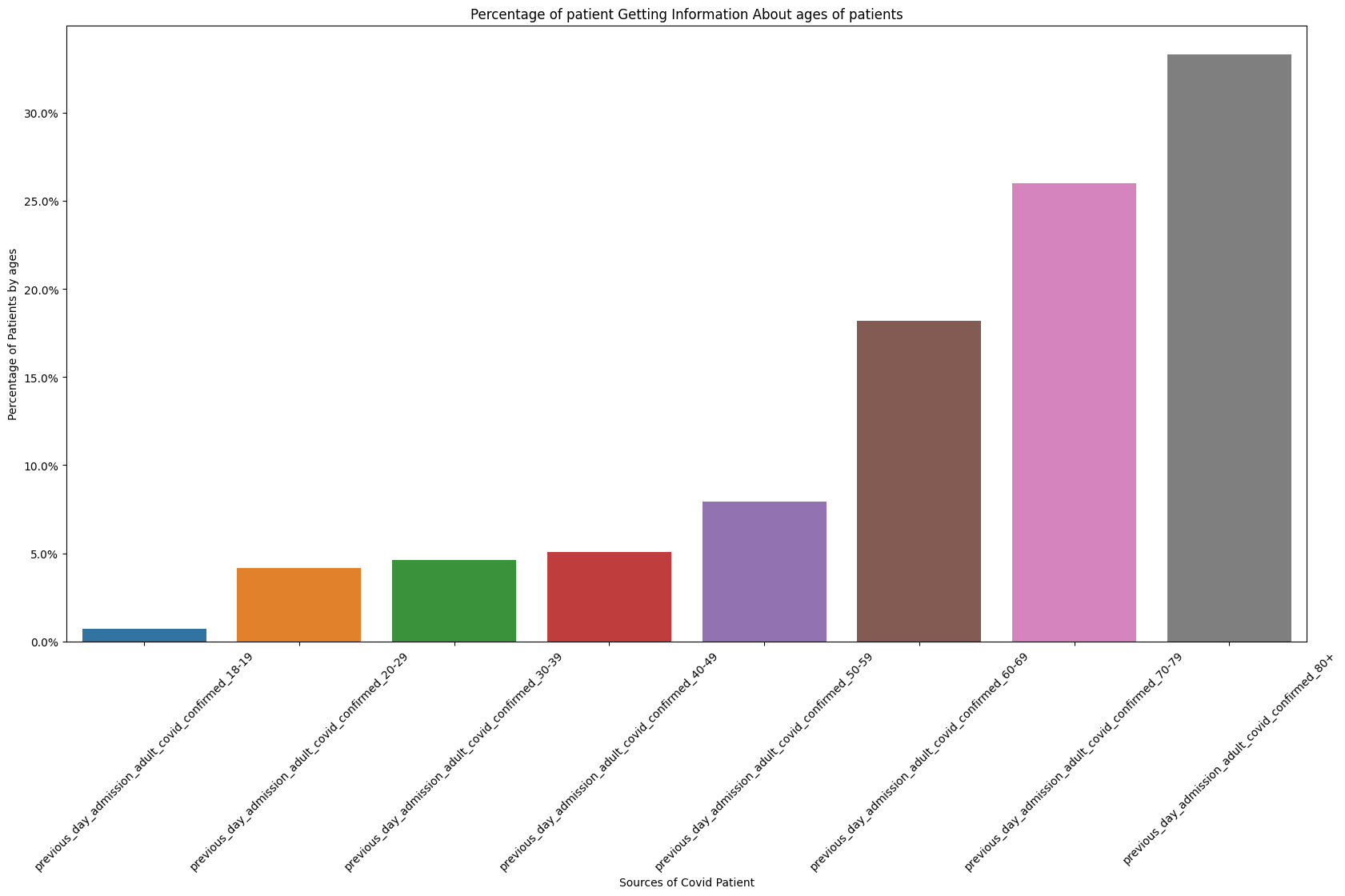
The package I will use are numpy, pandas, and for making bar-plot I will use sea-born.

# Code (python notebook file) (20%)

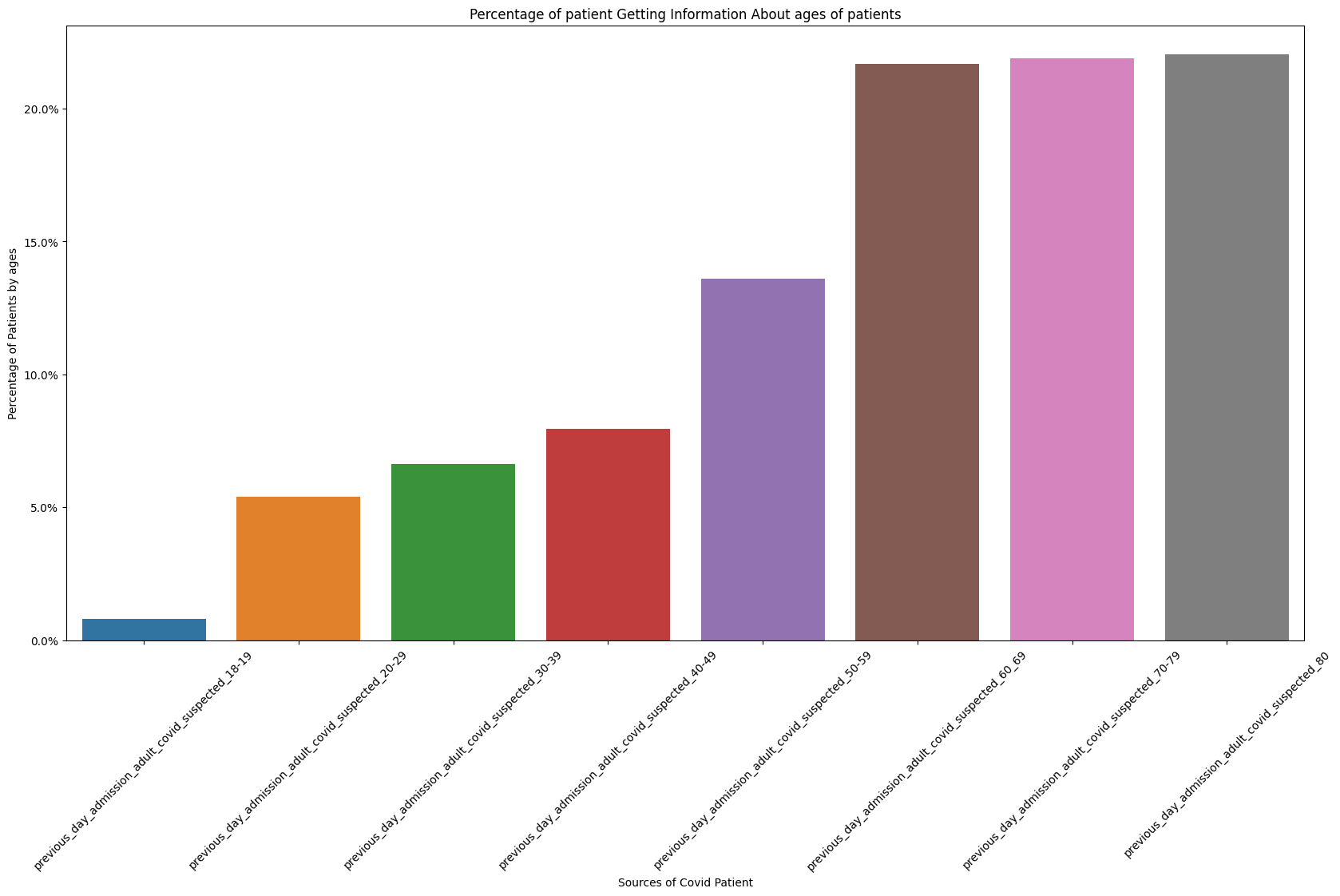
Code files has been uploaded on the link

# Document Results and Evaluation – Graphs and description/interpretation (15%)

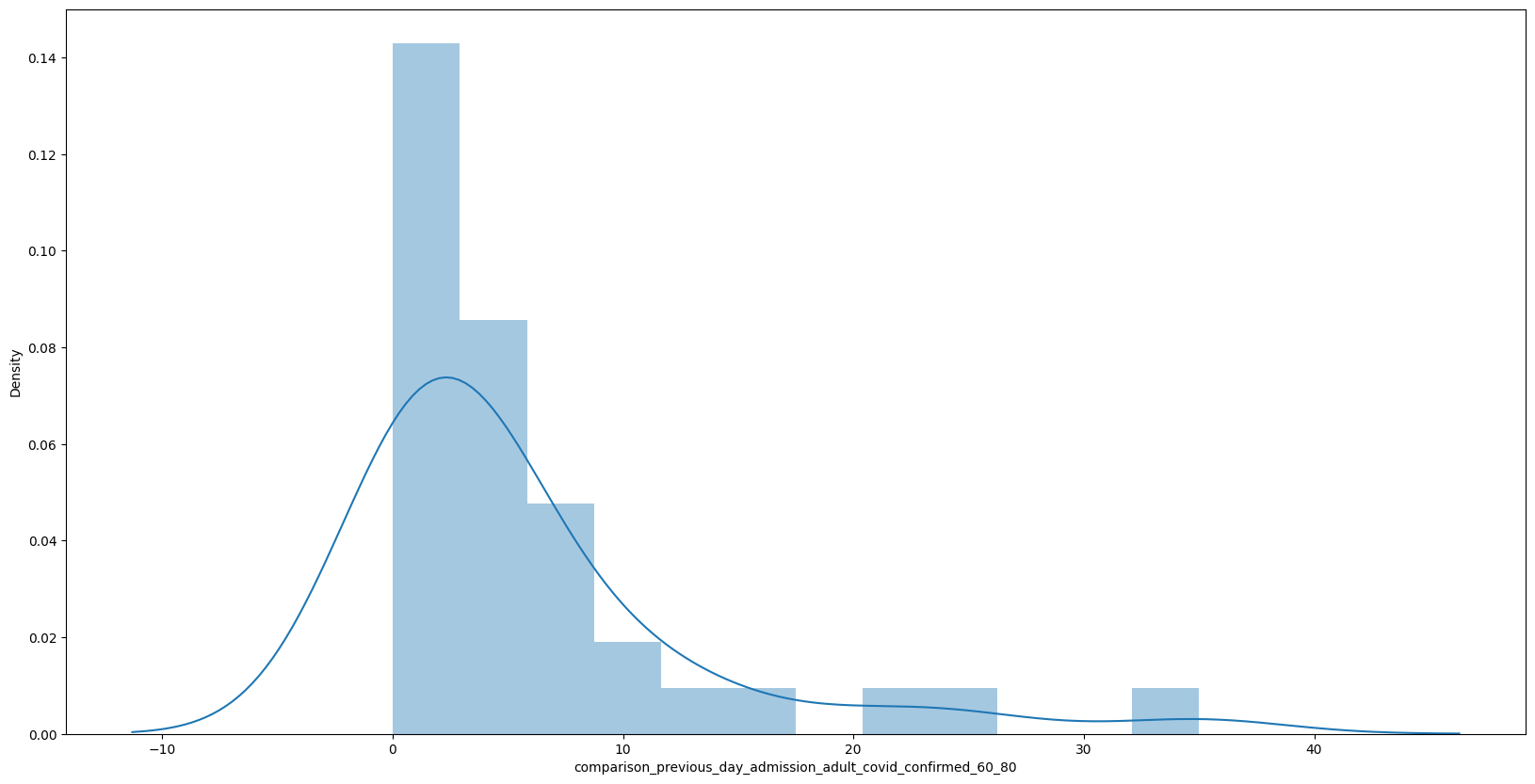
After analysing the aim of first questions and data cleaning the records of the previous day admitted adult covid19 confirmed cases from 18-80+ and making a bar-plot which describe the admitted. The bar-plot describe that the ratio of previous day admitted adult covid19 confirmed cases from the age of 80+ are the 30% higher and the ratio of the age of 70 to 79 are from the 25% higher and the ratio of adult covid-19 patient admitted from the age of 60 to 69 are of 20%. And the ratio of confirmed covid19 cases from the age of 18 to 39 is very low, about 5% only. These results evaluate that patients from the age of 60 to 80+ got more covid19 cases than the ages below and the 18 to 39 years of patients have very low ratio only 5%.



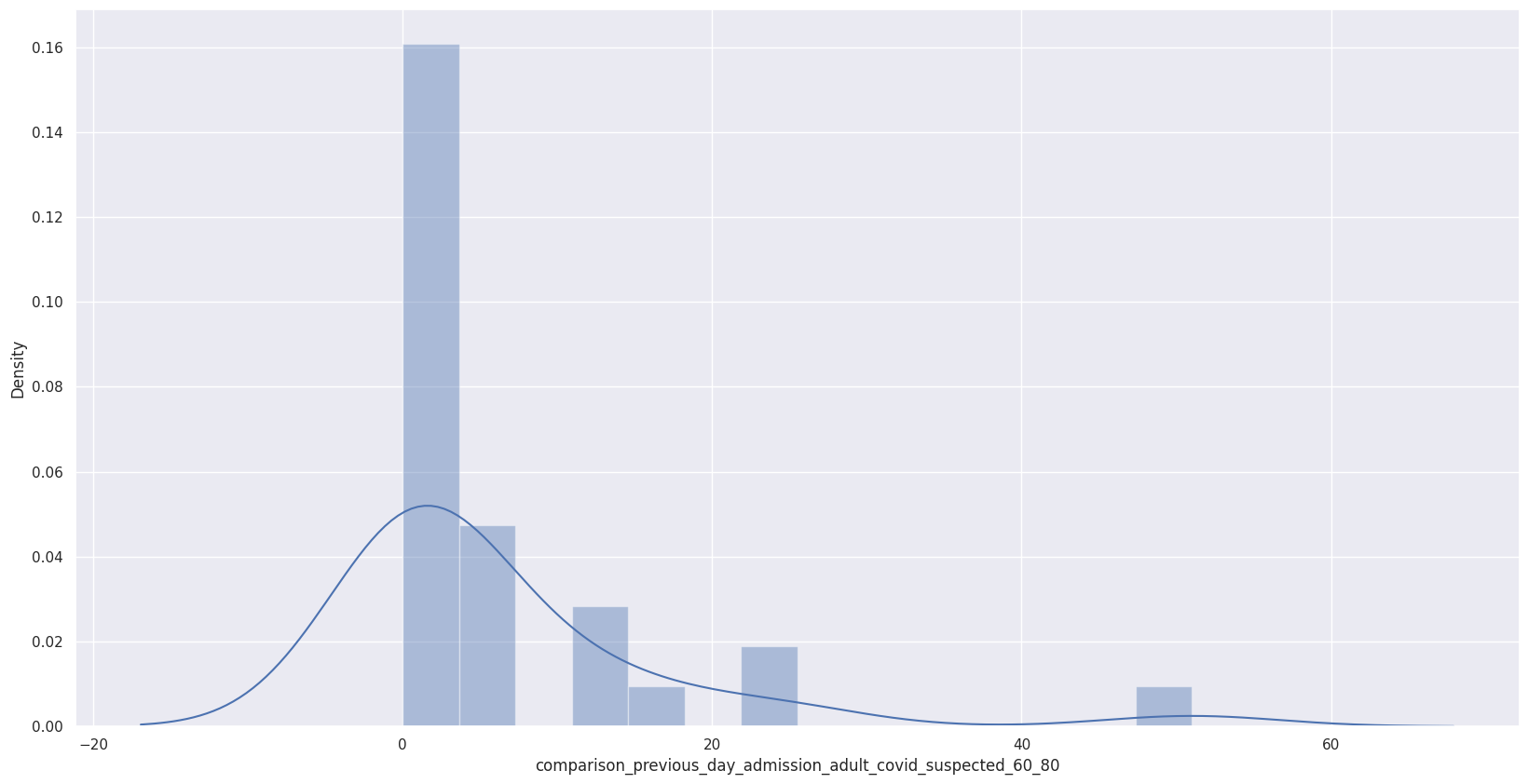
Analysing the aim of second questions and data cleaning the records of the previous day admitted adult covid19 suspected cases from 18-80+ and making a bar-plot which describe the admitted. The bar-plot describe that the ratio of previous day admitted adult covid19 suspected cases from the age of 50 to 80+ are the 20% higher and the ratio of the age of 40 to 49 are from the 15% lower. And the ratio of confirmed covid19 cases from the age of 18 to 39 are very low than 10% only. These result evaluate that patients from the age of 60 to 80+ got more covid19 cases than the ages below and the 18 to 39 years of patients have very low ratio only 5%.



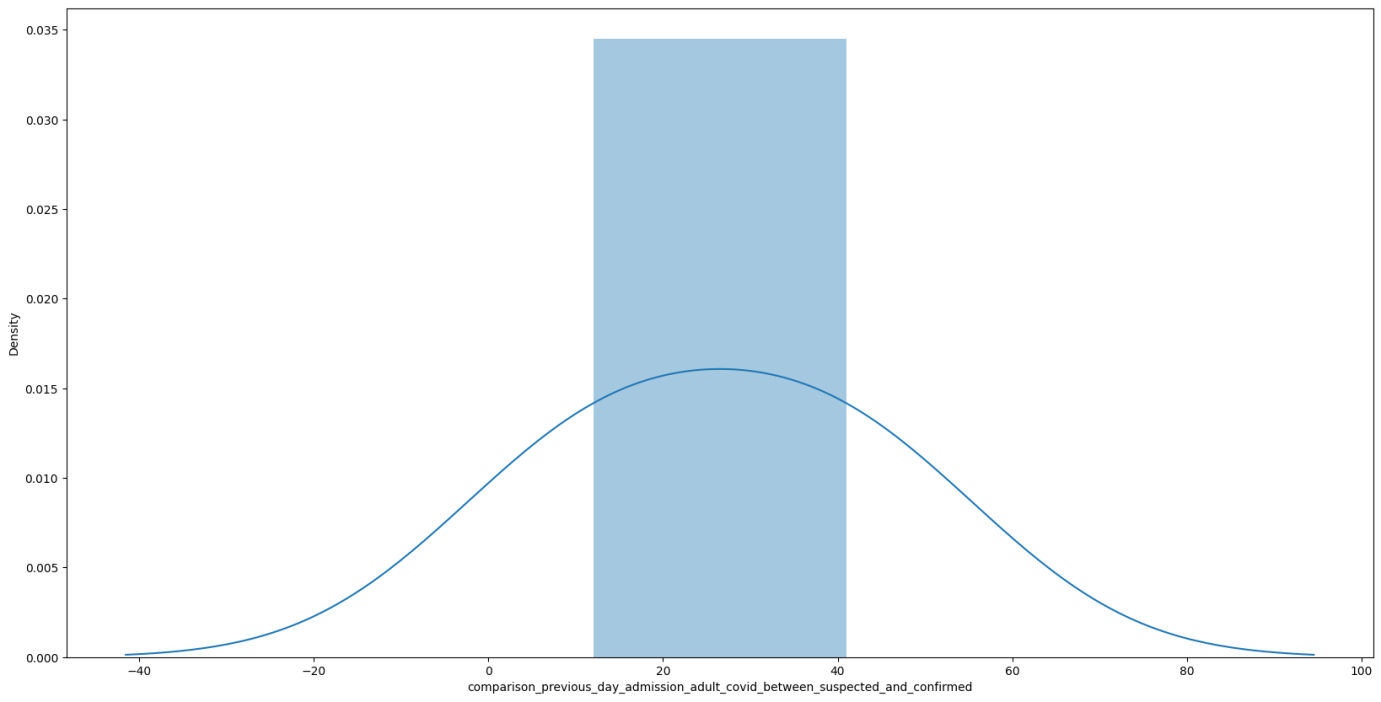
Analysing the aim of third questions and data cleaning the records of the previous day admitted adult covid19 confirmed cases from 60-80+ and making a distplot which describe the admitted. These result evaluate the comparison patients from the age of 60 to 80+ got more covid19 cases than the ages below and the 18 to 59 years of patients.



Analysing the aim of fourth questions and data cleaning the records of the previous day admitted adult covid19 suspected cases from 60-80 and making a distplot which describe the admitted. These result evaluate the comparison of suspected patients from the age of 60 to 80+ got more covid19 cases than the ages below and the 18 to 59 years of patients.



Analysing the aim of fifth questions and data cleaning the records of the comparison between suspected cases from 60-80 and confirmed cases from the age of 60-80+ admitted. and making a distplot which describe the admitted. These result evaluate that suspected patients from 60-80 are more comparison between suspected patients from the age of 60 to 80+ got more covid19 cases than the ages below and the 18 to 59 years of patients.



Submission of text files is via Turn-It-In on blackboard and submission of code files is also via blackboard. **Part A must be submitted on Wednesday 5th April 2023** and **Part B to be submitted by Wednesday 26th April 2023**. This must be the students own work and cases of plagiarism will receive a mark of zero. Failure to submit by the specified dates will incur penalties as per the QAH (15% for up to one week late, 30% for up to two weeks and a mark of zero if received after that time). Your lecture will explain the requirements in more detail during class.