# COVID-19 US County JHU Data & Demographics

# Introduction:

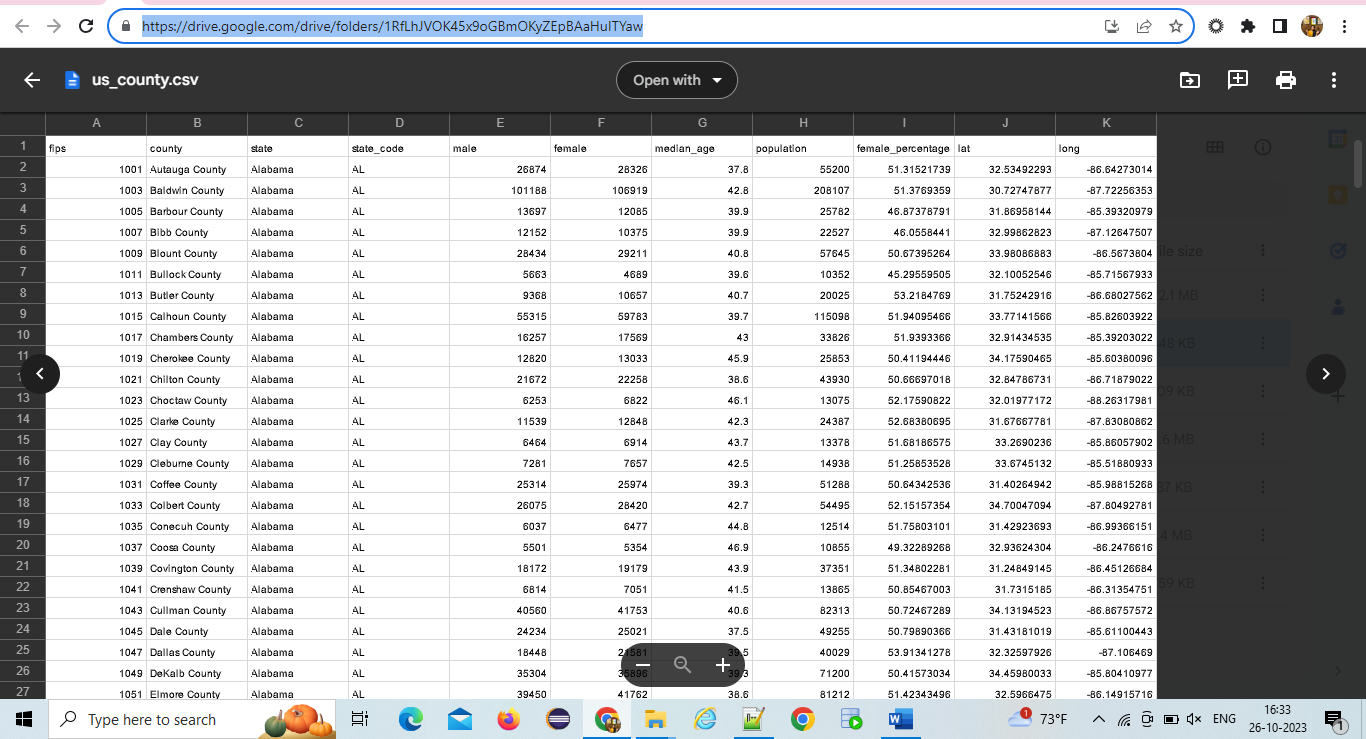
The United States of America has recently, had the most reported COVID-19 cases and this dataset that I have taken gives a piece of detailed information about the country, state, male, female, age group, and demographics information such as latitude and longitude. To perform this research, I used this dataset.

**DATASET LINK:**

<https://drive.google.com/drive/folders/1RfLhJVOK45x9oGBmOKyZEpBAaHuITYaw>

**US\_COUNTY.CSV**

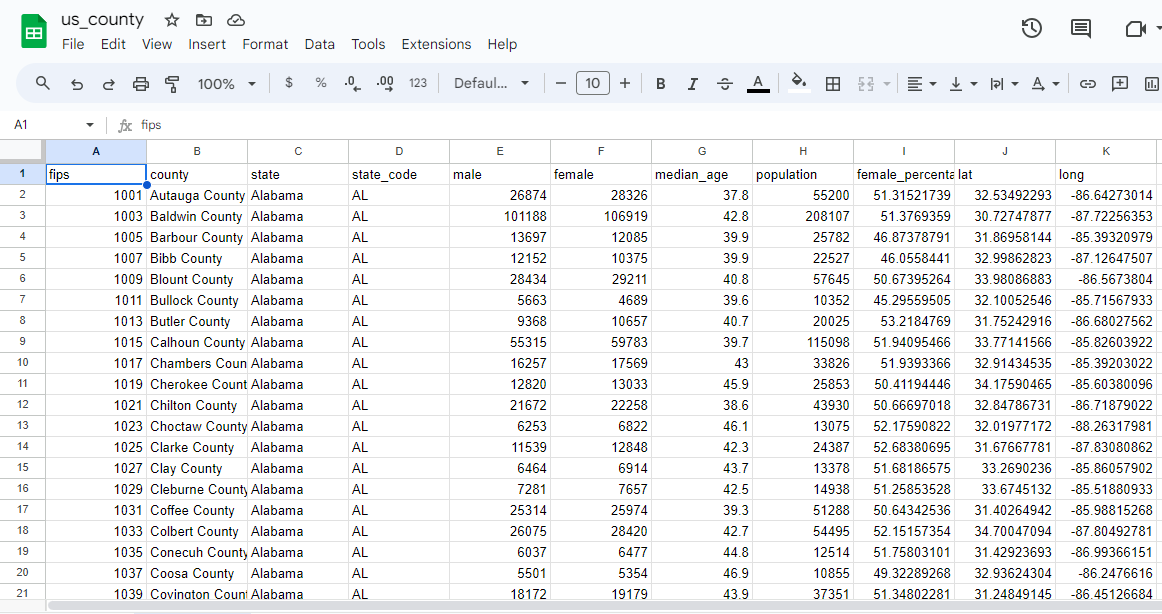
**The main objective of this analysis is** to find out the patterns within the dataset to get a further understanding of the data. I also wanted to leverage it to choose a machine algorithm for predicting the survival rate of patients during the period of COVID-19.



The dataset consists of demographic information population information (Such as male and female rates) and age information.

**Data attributes: Fips, County, State, State code, male, female, median age, population, female\_percentage, lat, long.**

So totally my dataset has 3220 rows \* 11 columns with no null values. The columns have a title/heading, which makes them readable.



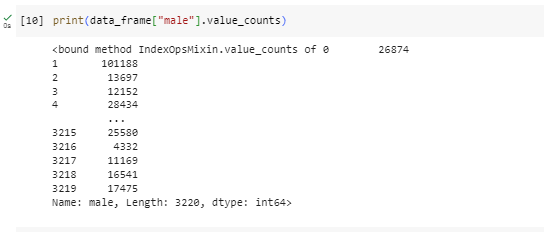
Observations of dataset:

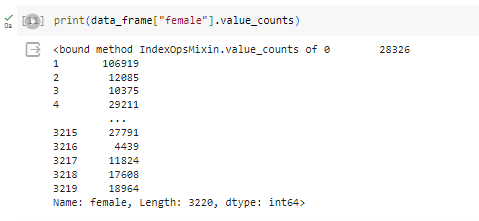
* It has all the states in the United States of America.
* The data includes patients whose ages range from 30 to 60.
* The data also contains fips code, latitude, and longitude details for easy understanding of the location details.

# Dataset and Code Description:

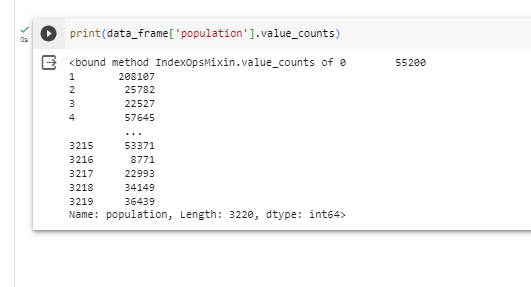
This data contains the total population, male and female.

**Explanation:** This code helps us to know the total count of males from different states.

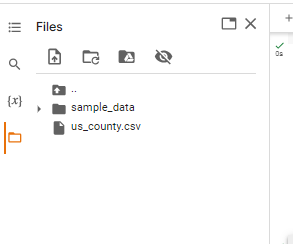
**Explanation**: This code helps us to know the total count of females from different states.



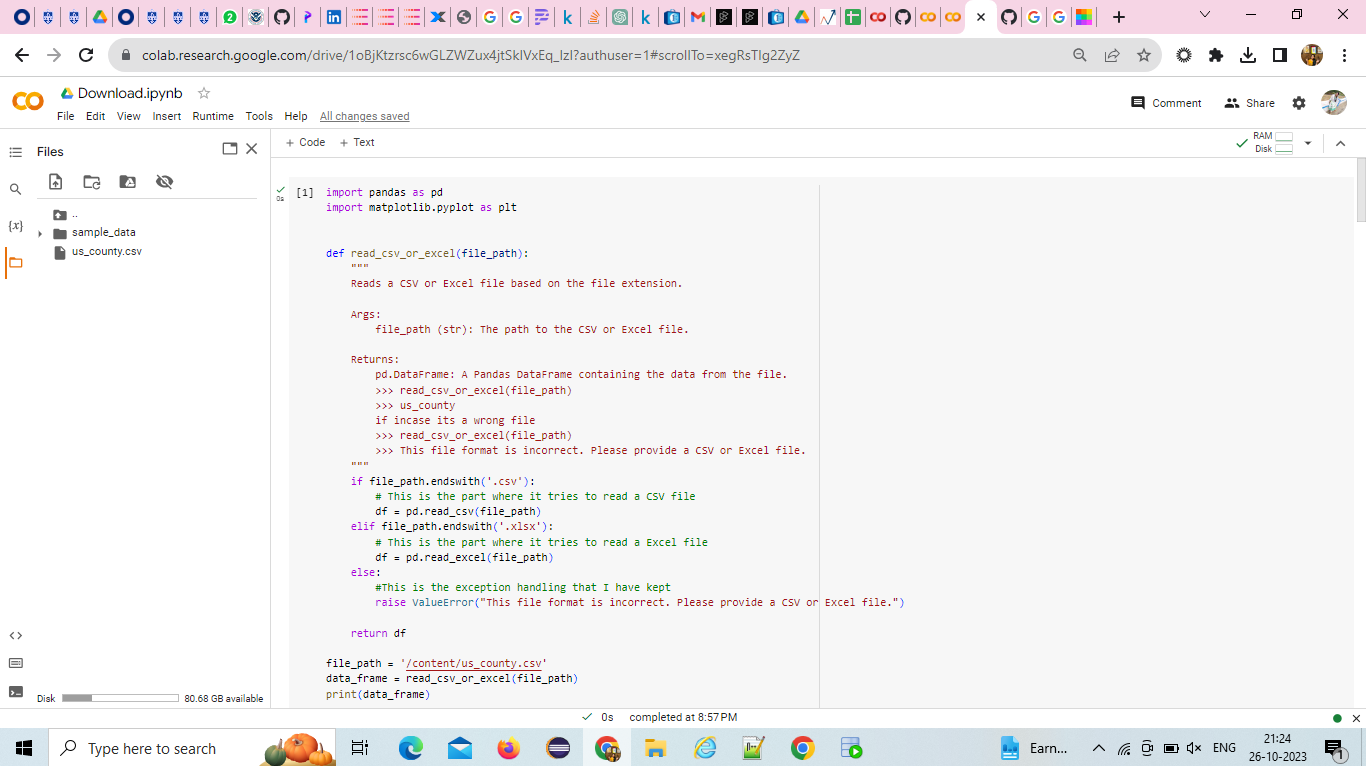
**Explanation**: This code helps us to know the total count of population from different state

**Important note:**

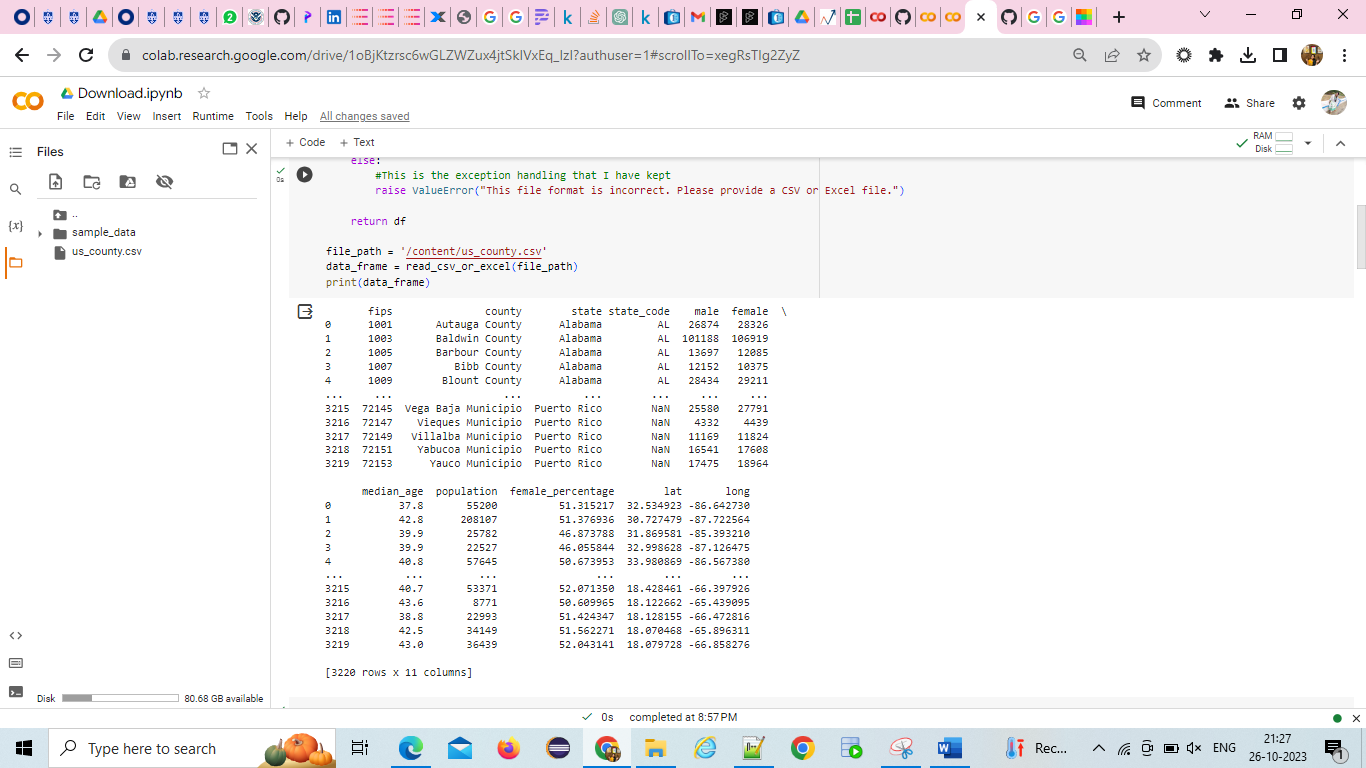
Before performing this code, we need to down the dataset and upload it in the Google Colab environment.



**Code:** This code helps me to read a CSV or Excel file in order to due EDA

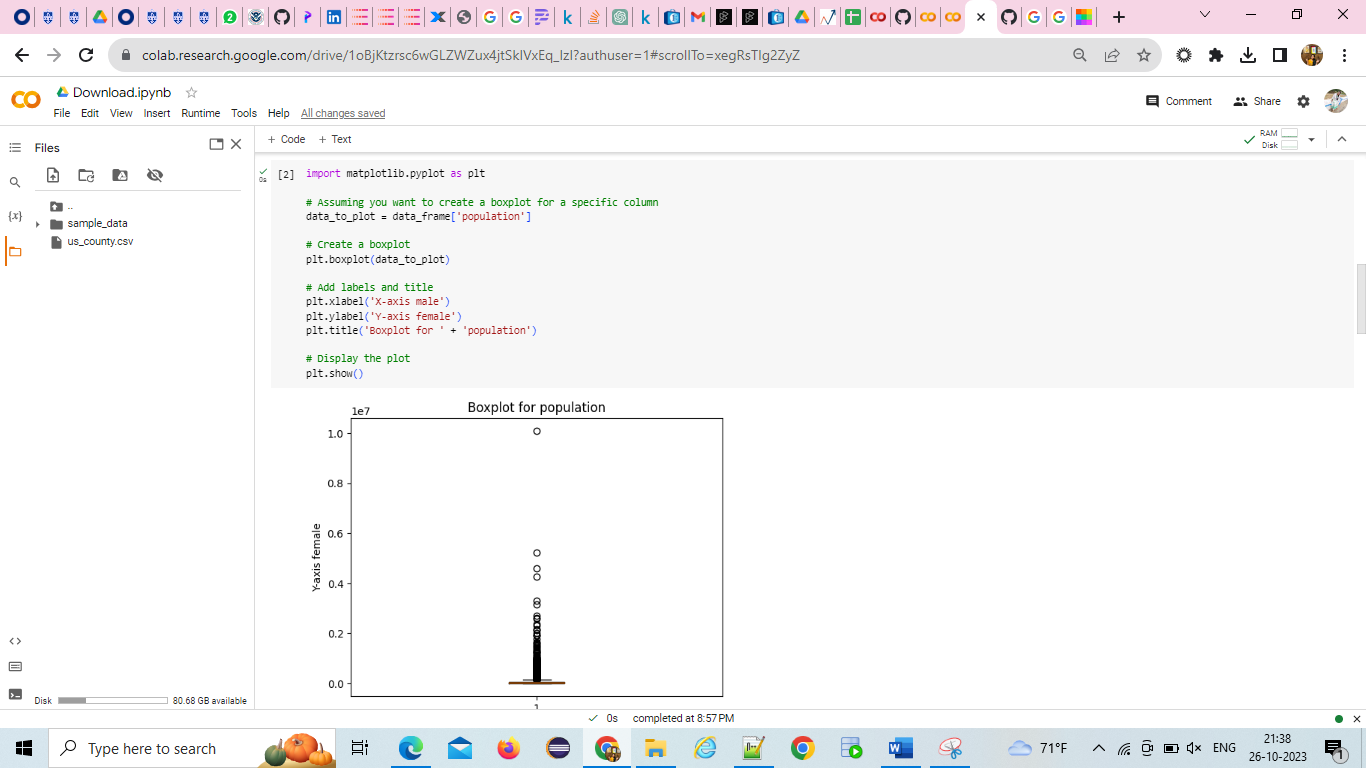


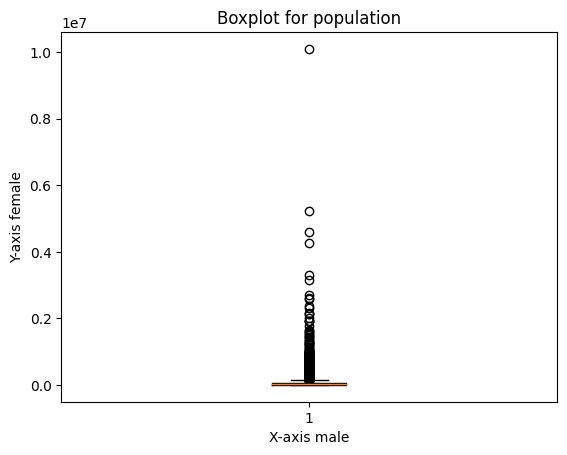
**Output:**



# Boxplot Graph:

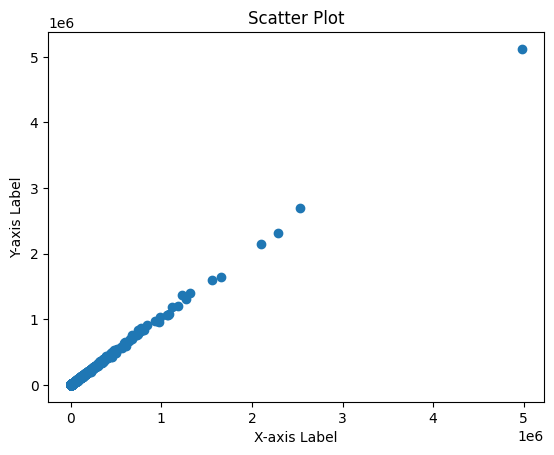
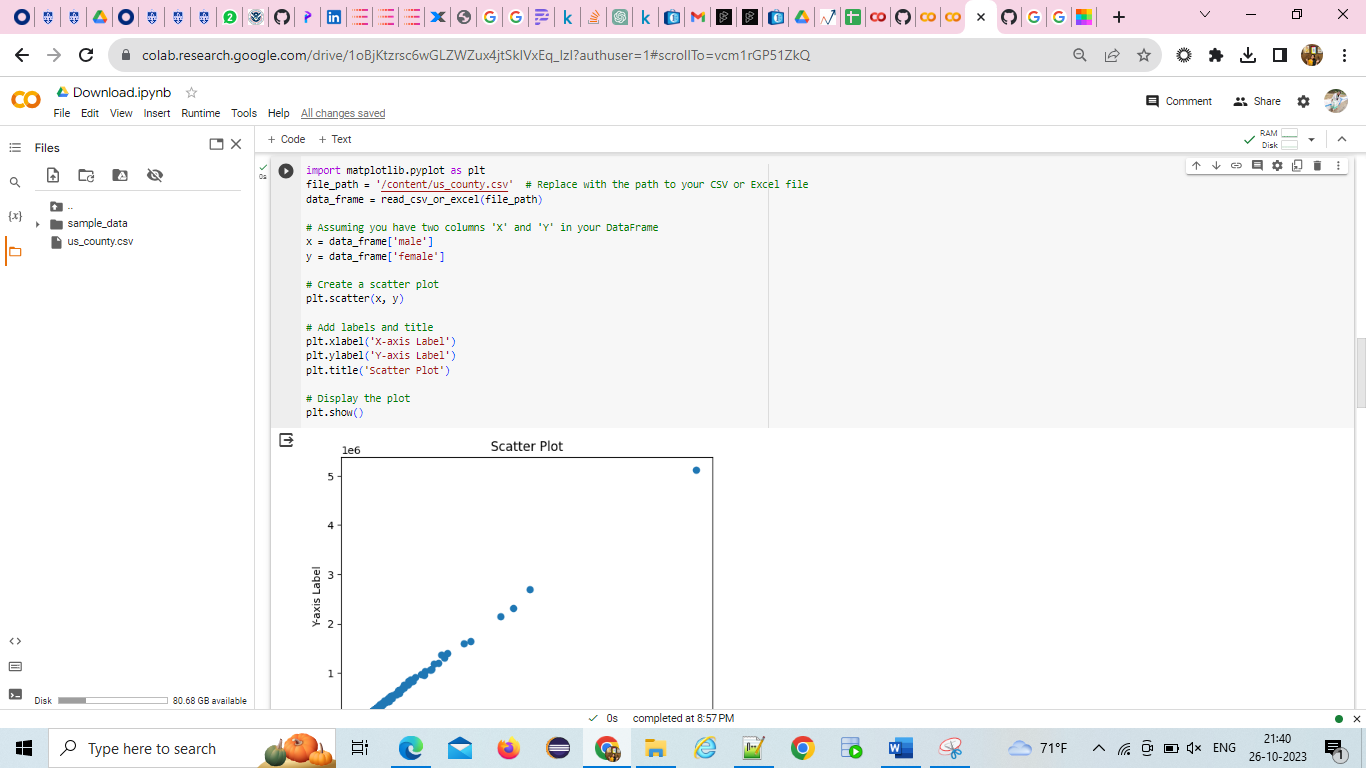
This graph shows a clear understanding of the male and female ratio





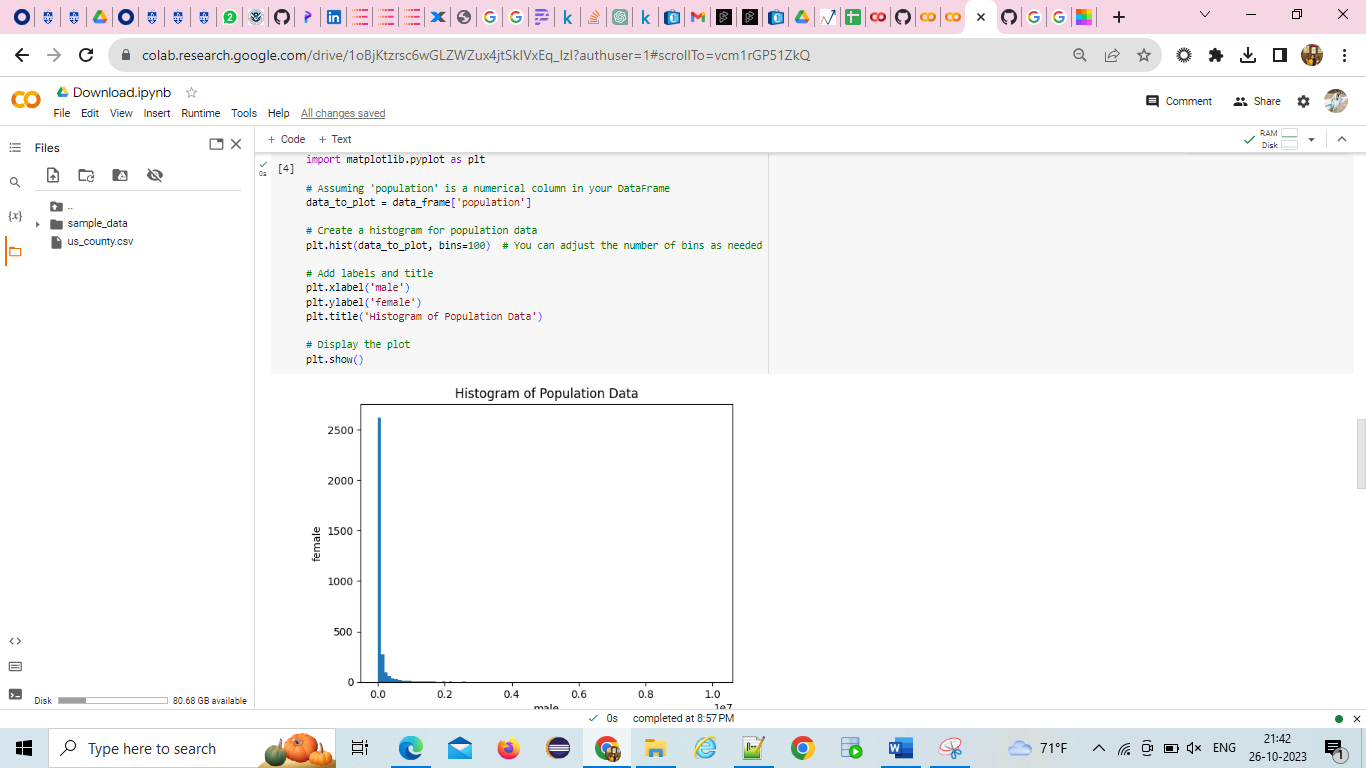
Scatterplot**:**

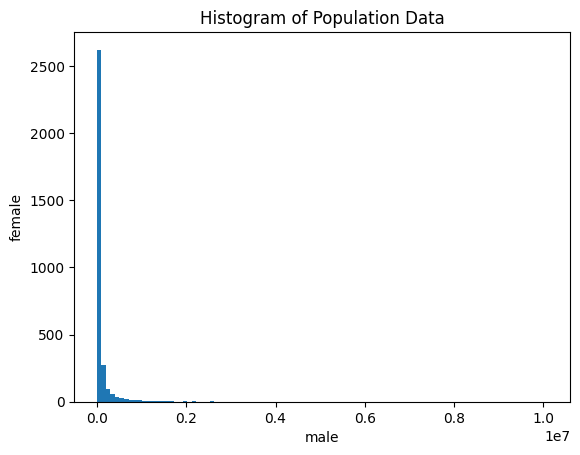
This graph shows a clear understanding of the male and female ratio.



# Histogram:

This graph shows a clear understanding of the male and female ratio





**Important Links:**

**Dataset Link:**

<https://drive.google.com/drive/folders/1RfLhJVOK45x9oGBmOKyZEpBAaHuITYaw>

<https://docs.google.com/spreadsheets/d/1OVgcN0T2npE5nRc9RTND8tUP9znStHVZJwMrOthtqDo/edit#gid=1650272371>

**GitHub Link:**

<https://github.com/santhiya-hds5210/ORES-5160-EDA>

**Drive Link:**

<https://drive.google.com/drive/folders/1W8AiXxbgTYK-HOXSPKjee9qGdj_Ari1O>

**Appendix**:

* <https://www.google.com/search?q=what+is+eda+in+data+science&oq=what+is+EDA+inn&gs_lcrp=EgZjaHJvbWUqCQgBEAAYDRiABDIGCAAQRRg5MgkIARAAGA0YgAQyCQgCEAAYDRiABDIJCAMQABgNGIAEMgkIBBAAGA0YgAQyCQgFEAAYDRiABDIJCAYQABgNGIAEMgkIBxAAGA0YgAQyCQgIEAAYDRiABDIJCAkQABgNGIAE0gEJMTE4MjhqMGo3qAIAsAIA&sourceid=chrome&ie=UTF-8>
* <https://www.kaggle.com/datasets/headsortails/covid19-us-county-jhu-data-demographics?select=us_county.csv>
* <https://stackoverflow.com/questions/18039057/pandas-parser-cparsererror-error-tokenizing-data>
* <https://chat.openai.com/c/8da6a9dc-bee7-4983-9bf9-7530b2178d31>
* <https://www.kaggle.com/code/masoudfaramarzi/basics-of-accesing-data-from-urls-using-pandas>
* <https://www.forefront.ai/app/chat/new>
* <https://www.numbeo.com/quality-of-life/rankings_by_country.jsp>
* <https://www.analyticsvidhya.com/blog/2022/03/exploratory-data-analysis-with-an-example/>
* <https://docs.google.com/spreadsheets/d/1OVgcN0T2npE5nRc9RTND8tUP9znStHVZJwMrOthtqDo/edit#gid=1650272371>
* <https://canvas.slu.edu/courses/45377/assignments/343230>
* <https://colab.research.google.com/drive/1Yr_FH_rjTCW7741e1rArixu4ZWL02FGC#scrollTo=ZfIbVsMyiqOI>
* <https://github.com/santhiya-hds5210/ORES-5160-EDA>
* <https://www.google.com/search?q=scatter+plot&oq=scatter&gs_lcrp=EgZjaHJvbWUqDQgBEAAYgwEYsQMYgAQyDwgAEEUYORiDARixAxiABDINCAEQABiDARixAxiABDIKCAIQABixAxiABDINCAMQABiDARixAxiABDINCAQQABiDARixAxiABDIKCAUQABixAxiABDINCAYQABiDARixAxiABDIHCAcQABiABDIKCAgQABixAxiABDINCAkQABiDARixAxiABNIBCDMzOTdqMGo3qAIAsAIA&sourceid=chrome&ie=UTF-8>
* <https://www.google.com/search?q=boxplot&oq=boxpl&gs_lcrp=EgZjaHJvbWUqDAgBEAAYQxixAxiKBTIGCAAQRRg5MgwIARAAGEMYsQMYigUyDwgCEAAYQxiDARixAxiKBTIKCAMQABixAxiABDIJCAQQABhDGIoFMgcIBRAAGIAEMgkIBhAAGEMYigUyCQgHEAAYQxiKBTIJCAgQABhDGIoFMgcICRAAGIAE0gEIMzEwNmowajeoAgCwAgA&sourceid=chrome&ie=UTF-8>