# Project Introduction

A country’s commonplace of living is usually measured by financial gain (Gross Domestic Product (GDP)) per person, with higher per-person financial gain indicating higher living standards. various measures utilized by development practitioners embrace the Human Development Index and accessibility and quality of education and health. The Commission on the measure of Economic Performance and Social Progress (CMEPSP) (2009) defines well–being as a third-dimensional construct involved with not solely financial gain however additionally non–economic aspects of life like political voices and governance, social networks, and relationships, the atmosphere and security. The multi– spatial property suggests that there's no single indicator to summarise it, so requiring the institution of a system that captures all the relevant dimensions. The question of whether or not or not value could be a smart indicator of national economic well–being has been debated ofttimes in economic analysis (Nordhaus and James Tobin, 1973; Zolotas, 1981; Boarini et al, 2006). the subject has been self-addressed during a huge analysis literature (Nordhaus and Kokkelenberg, 1999; CMEPSP, 2009; Chiripanhura, 2010), leading to proposals to develop and emphasize alternative measures of national economic well–being. value is criticized as a poor indicator of a nation’s economic well–being (though it absolutely was ne'er designed for that (Vanoli, 2005) as a result of it doesn't live some activities within the assembly boundary well, and it excludes some determinants of welfare that's outside the assembly boundary (Allin, 2007). Mean analysis of National Accounts indicators, like consumption, income, and wealth, to point economic well–being suffers from the downside that it masks spatial arrangement variations, so making a control of rising material living standards of all members of society, which can not be the case. to beat this limitation, CMEPSP (2009) (Recommendation 4) planned the accrued use of median analysis to supplement mean analysis. this is often as a result of the median isn't sensitive to extreme values (very massive or tiny values) at either finish of the financial gain distribution. This paper examines the various footage of fabric living standards conferred by mean and median financial gain measures. It uses real value per person because the benchmark and compares this with home mean and median incomes. It builds on the applied mathematics reasons why the median could also be higher than and/or got to supplement the mean by conflict that once measure the amount of national economic well-being, focus ought to get on the ‘typical’ instead of the ‘average’ home. it's going to even be essential to possess each average analyses so as to induce a fuller image of living standards.

# Analysis OBjectives

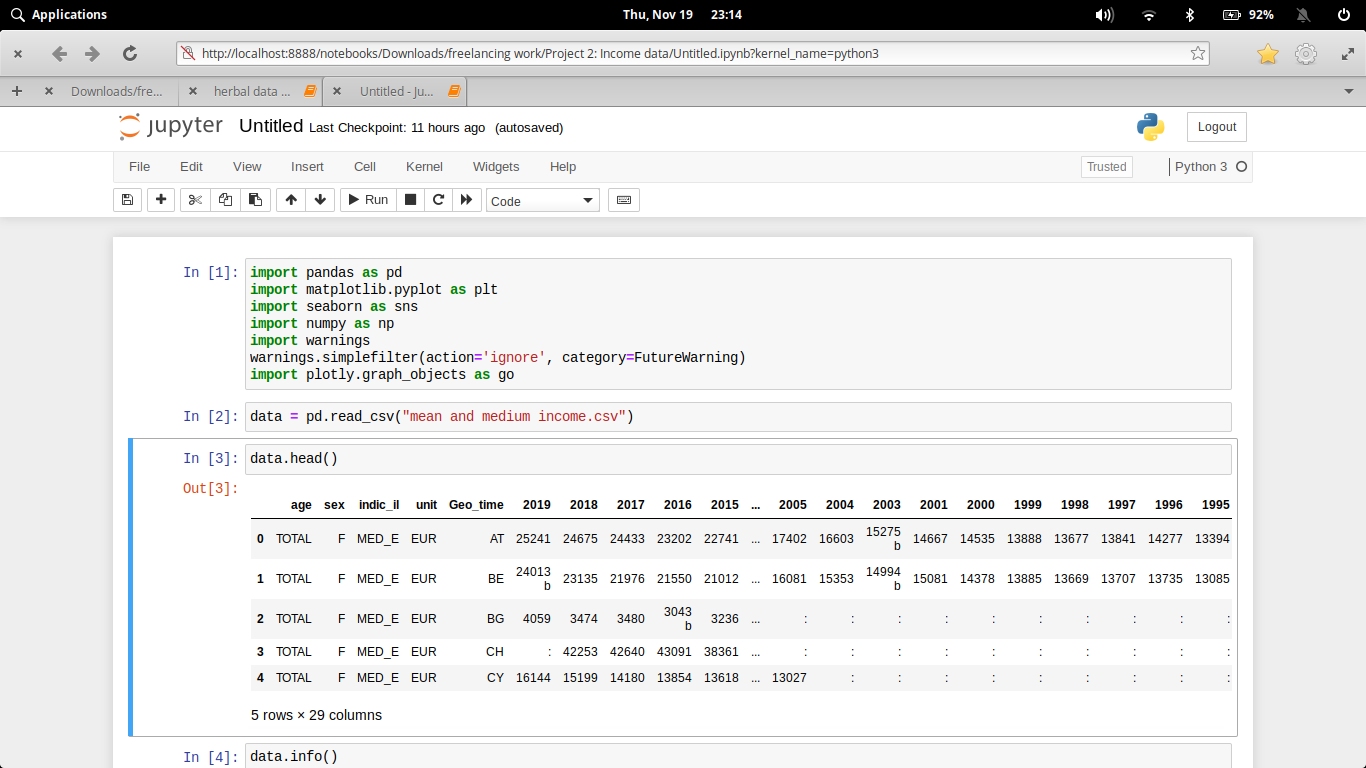
Here are the list of question whose analysis which be given below..

1. What is the average mean salary of dfferent sex(gender) and in total in different consecutive years?
2. What is the average mean salary in different geograpfic locations?
3. What is the median salary of different sex(gender) and in total in different consecutive years?
4. What is the average median salary in different geographic locations?
5. Mean vs Median graph and Why mean is choosen over median to calculate the average number?
6. What is the percentage rise in salary for different sex(gender) over a period of time?
7. What is the maximun percentage rise in salary over these different years?

# Data Acquisition and Cleaning

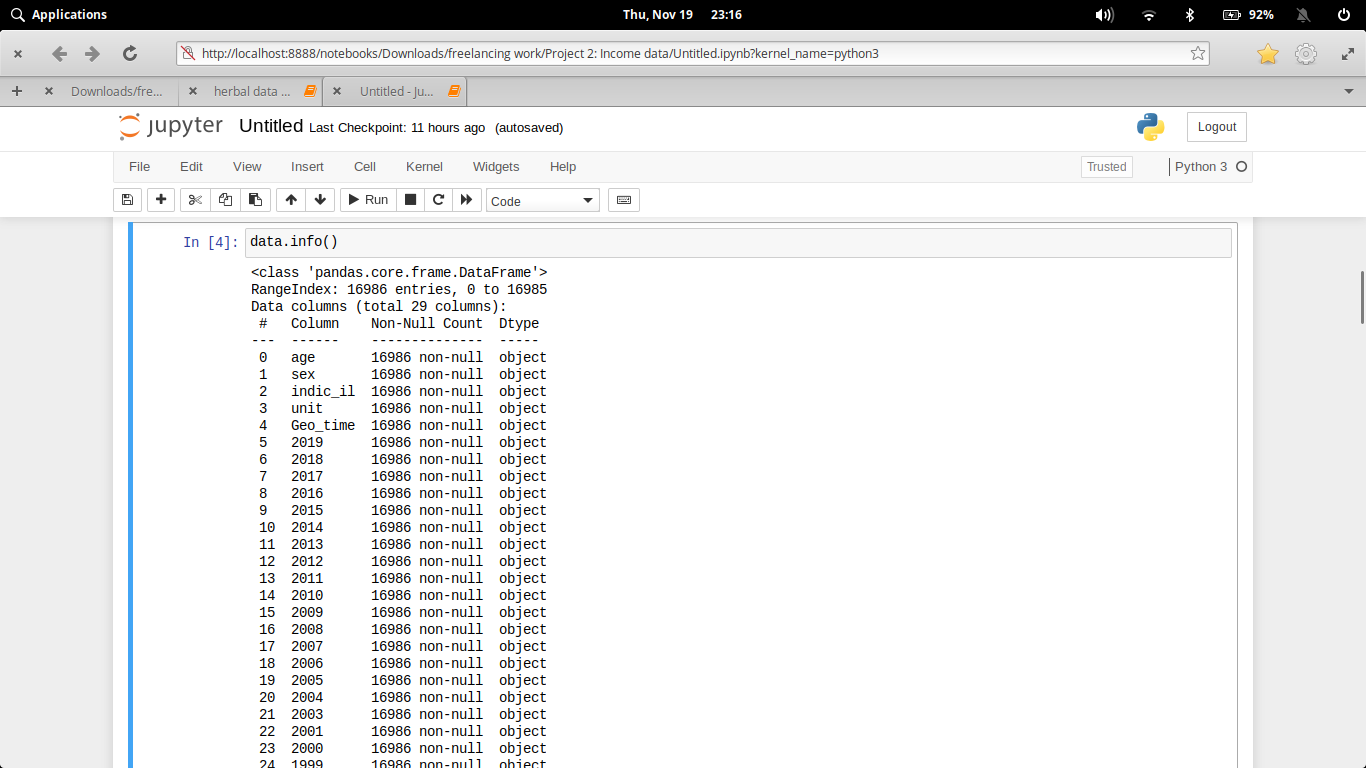
**Code to read the data from Excel / CSV / HTML.**

To read the dataset in csv format, we will load it into Pandas data frame but first let’s import the pandas library and set an alias by typing **“import pandas as pd”**. After importing the library with the alias **“pd”**, let us load the .csv file using the following line of code:



Here we have import our csv files and read through pandas library.

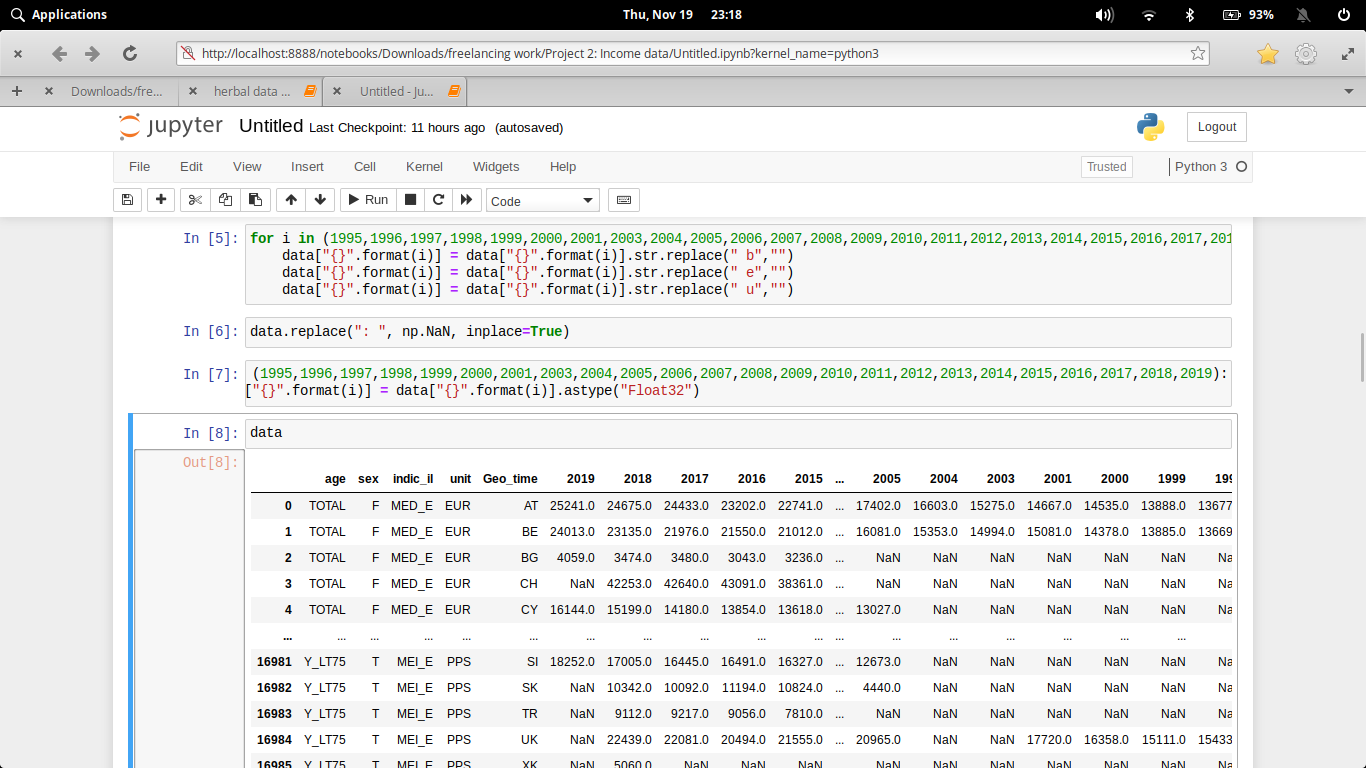
Here the csv file can be read through (Pandas library) and store in **data Dataframe**. The Dataframe can be shown through **.head()**. The number of rows we want to show, that number we have to pass in head parentheses as an argument.

Now if we want to describe our dataframe for our better understanding to know the stats. and other parameter that our dataset should follow

**Clean the unnecessary data, by removing, replace the missing data and renaming the columns.**

Dataset generally contains some null value, which is generally caused by misplacing some values. So its necessary to clean this mess from our dataset for better visualization

In the first diagram (data.head()) we see that some of the columns of this dataset contains “: “ value.



In the process of data cleaning we first remove some variables in different value columns. Then subsequently we fill the “: “ values column with inumpy.NaN so that the mean/median values will not much get effected. Finally after all process we convert every column from object to float, for better calculation.

**why data clean needed (for your data)**

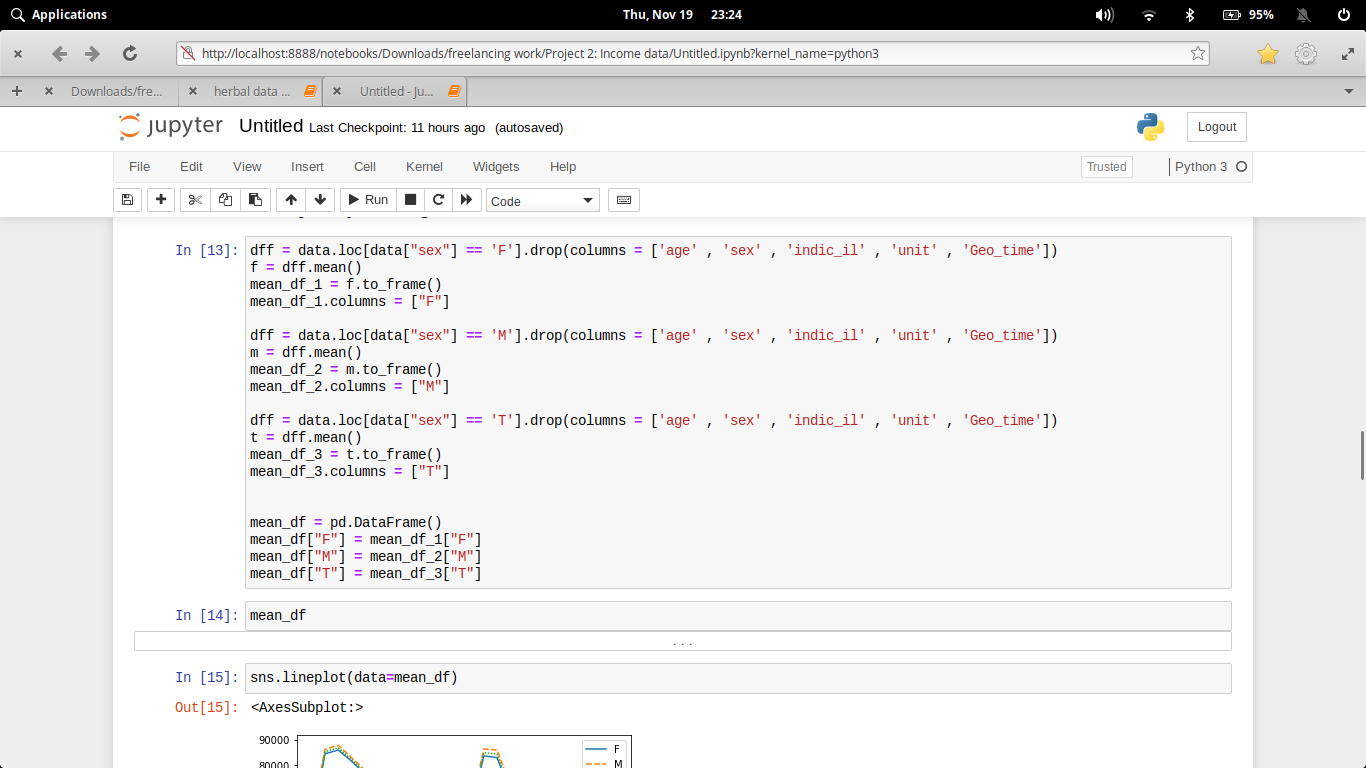
Data cleansing or scrubbing or appending is the procedure of correcting or removing inaccurate and corrupt data. This process is crucial and emphasized because wrong data can drive a business to wrong decisions, conclusions, and poor analysis, especially if the huge quantities of big data are into the picture.

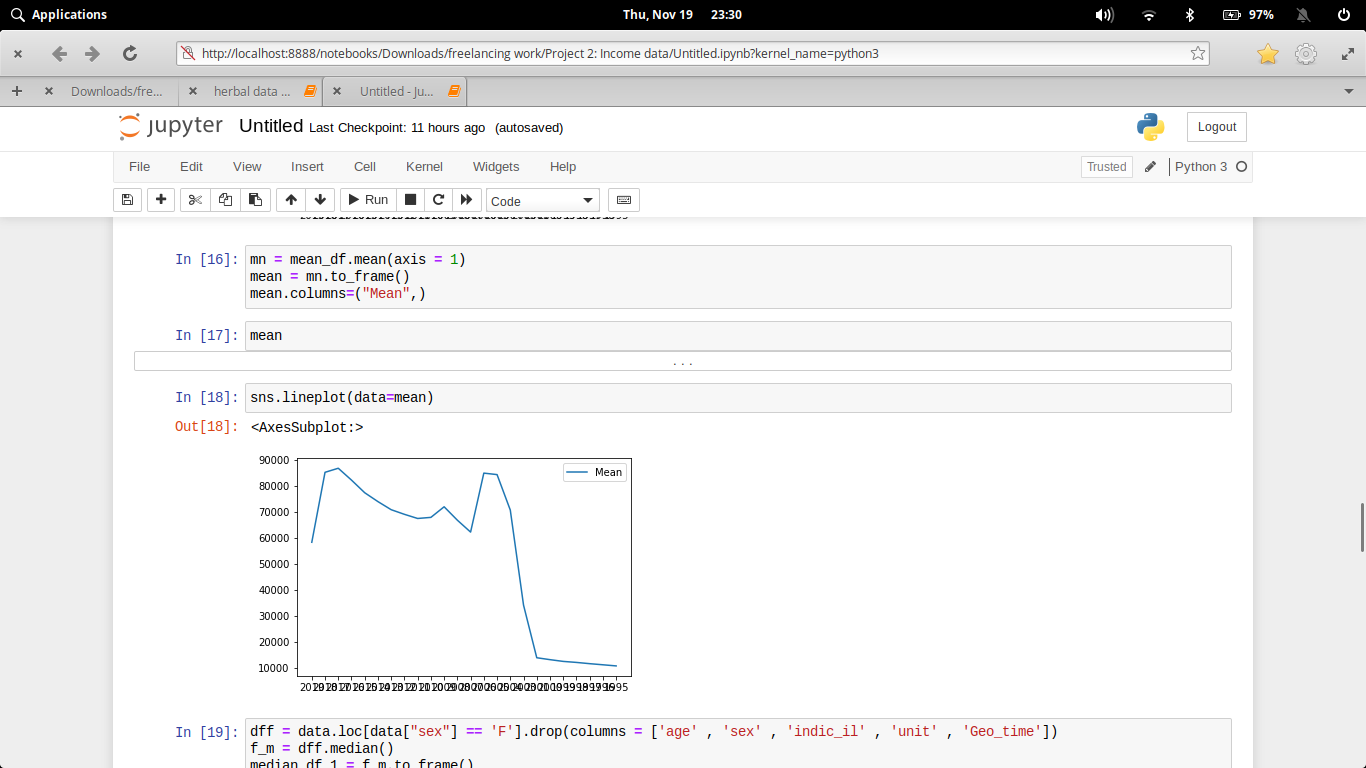
# Data and Exploratory Analysis

**Code and its output with Explanation**

1. **What is the average mean salary of dfferent sex(gender) and in total in different consecutive years?**

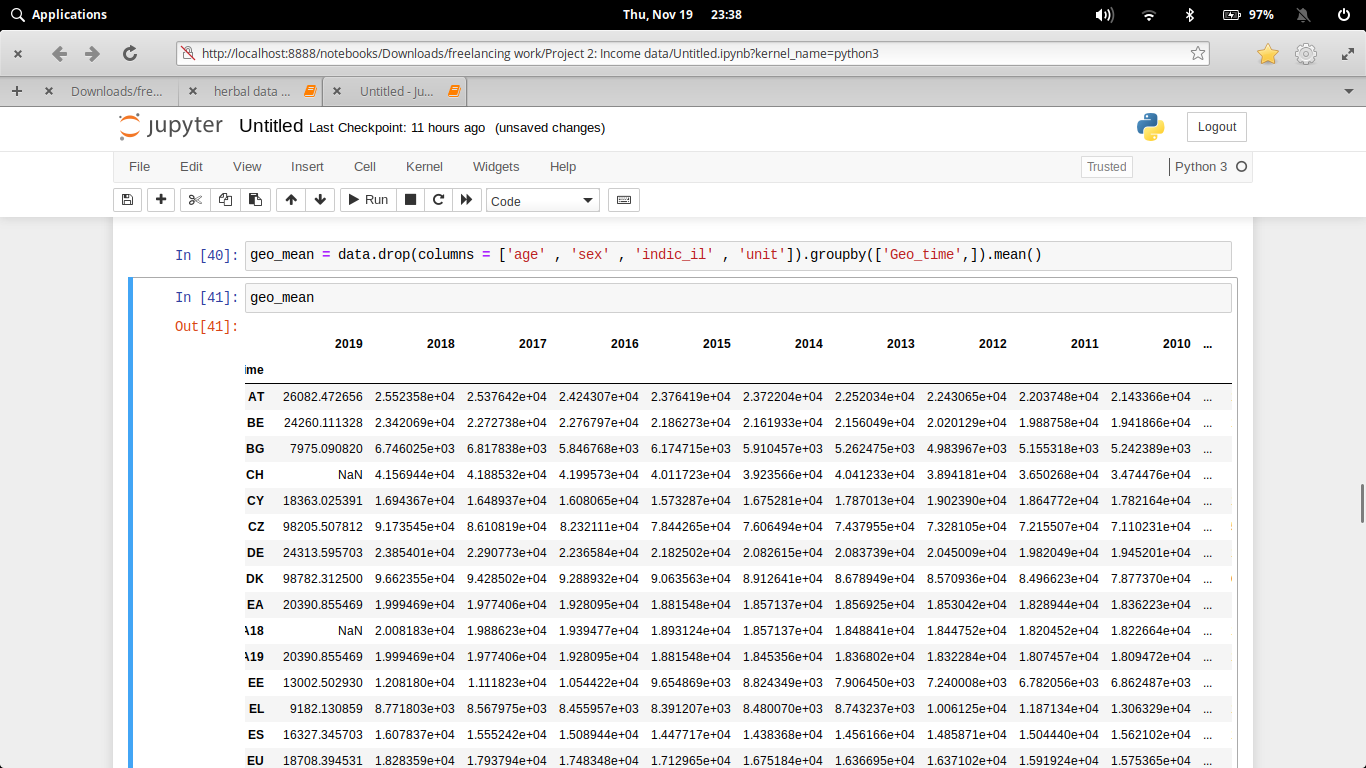
Here, mean\_df describes the mean salary of different sex(gender) and mean describes the overall mean value.

****

****

1. **What is the average mean salary in different geograpfic locations?**

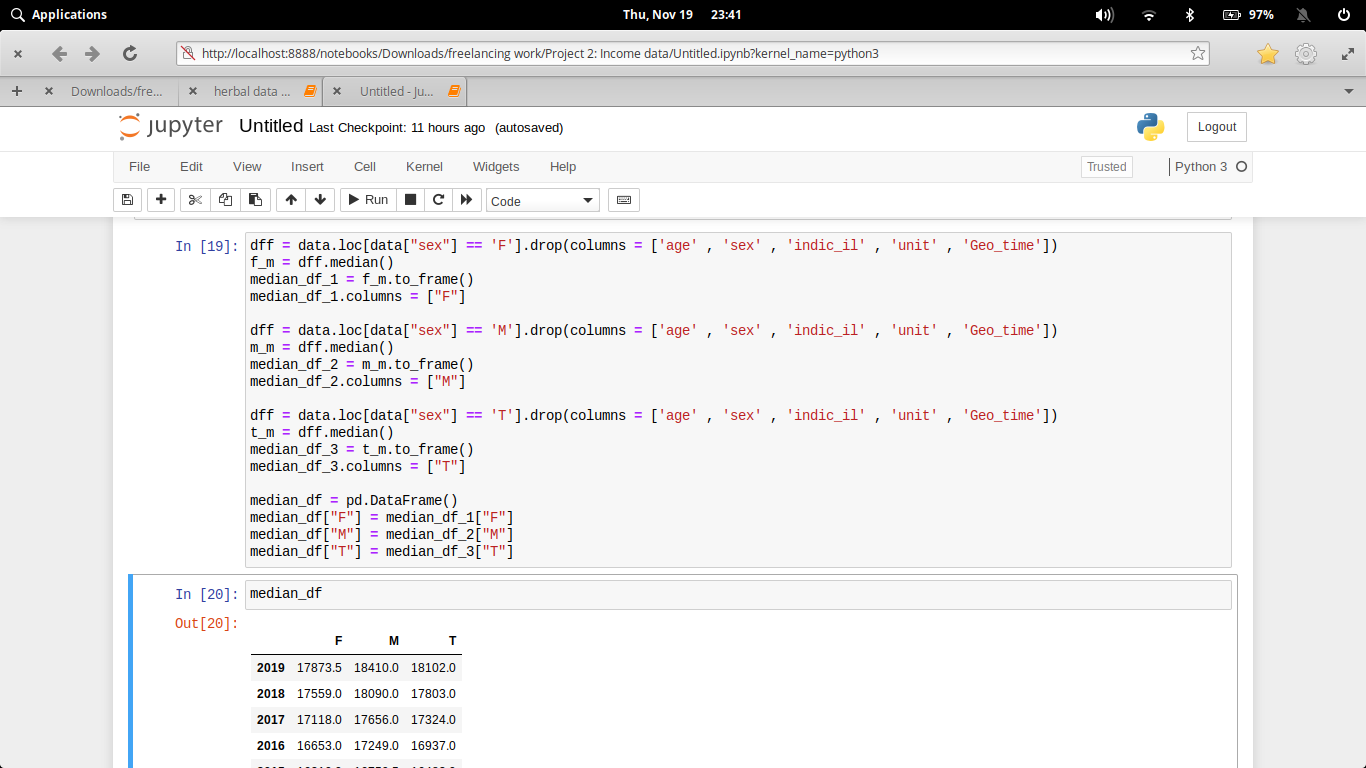
From this, we want to find out that which geo\_location has highest mean salary, irrespective of the no.of people living there.



1. **What is the median salary of different sex(gender) and in total in different consecutive years?**

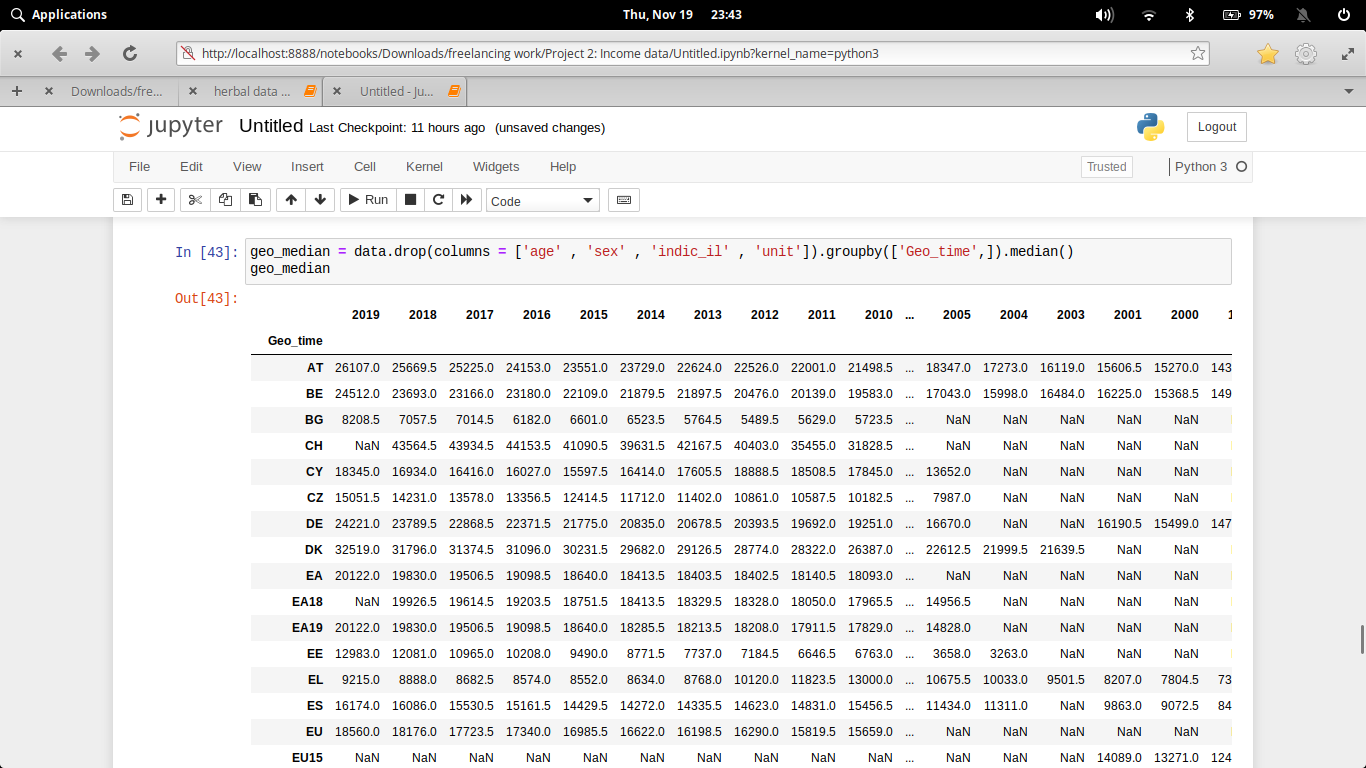
Our motto is to find the median salary of the whole country.

Here, mean\_df describes the mean salary of different sex(gender) and mean describes the overall mean value.



1. **What is the average median salary in different geographic locations?**

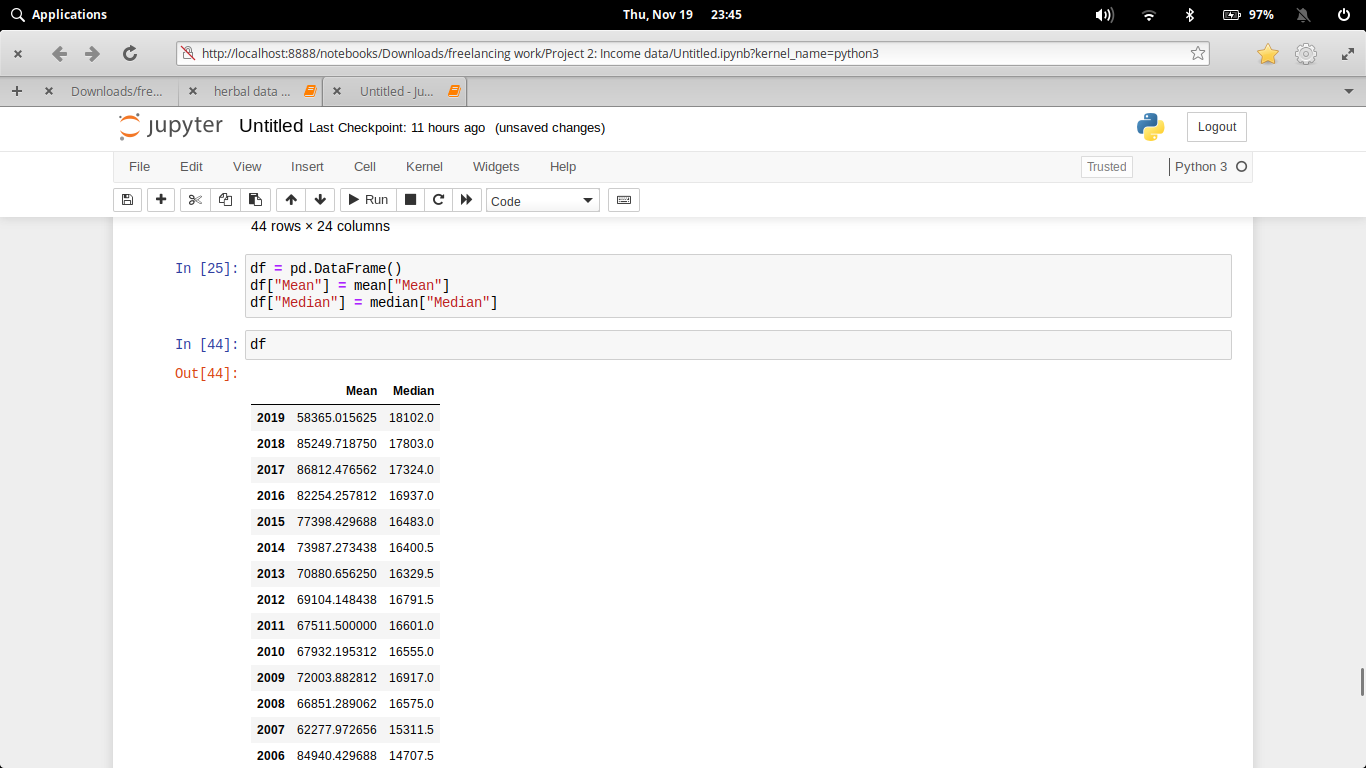
From this, we want to find out that which geo\_location has highest median salary, irrespective of the no.of people living there.



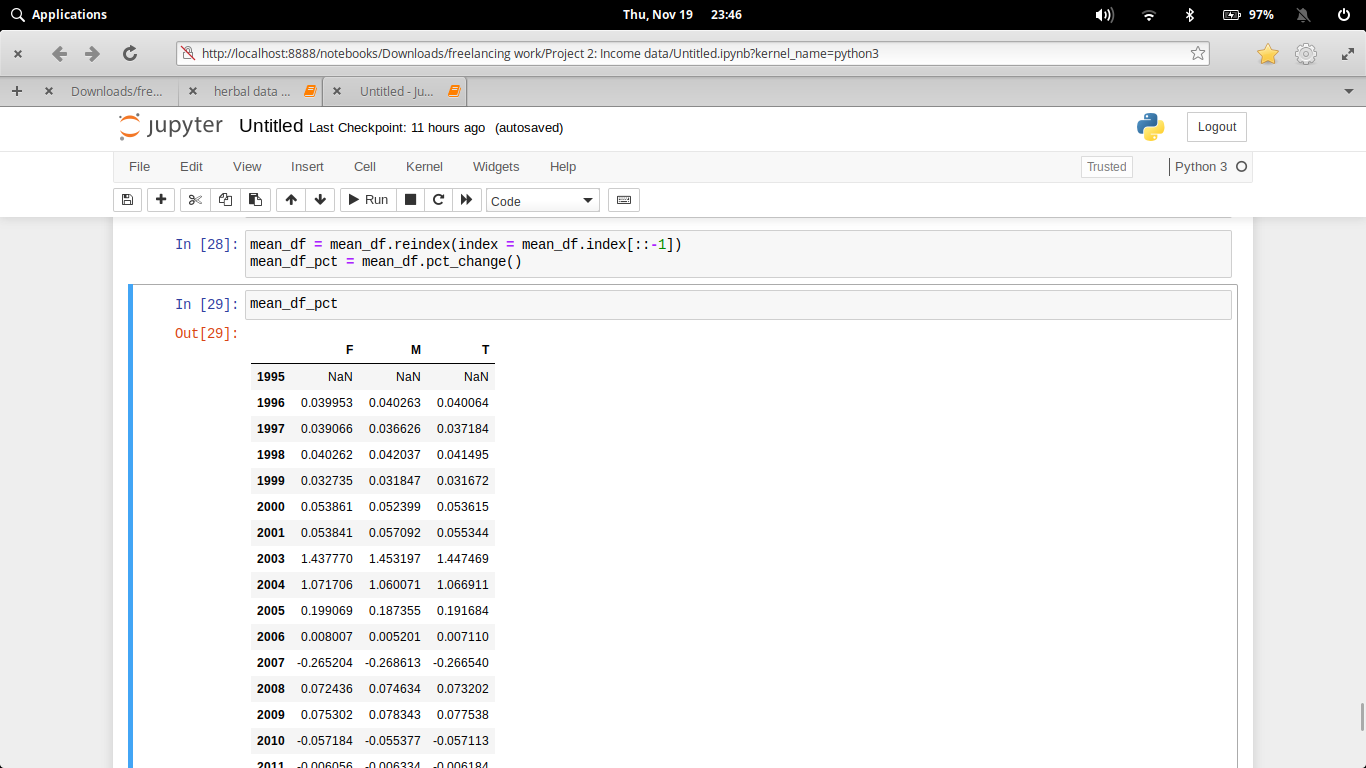
1. **Mean vs Median graph and Why mean is choosen over median to calculate the average number?**

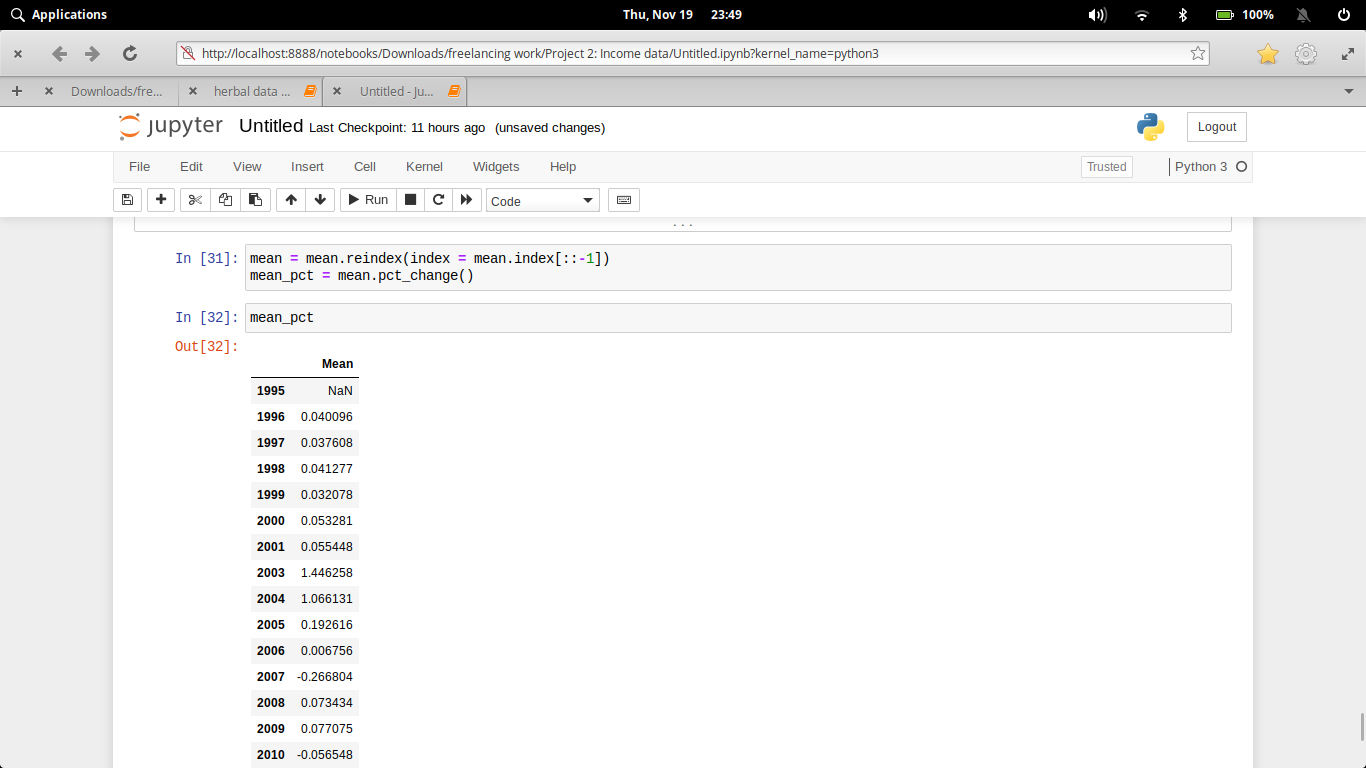
Mean value depends on every data from the dataset ulike median, which is not.

Therefore, for a large amount of data, mean is the optimum solution.



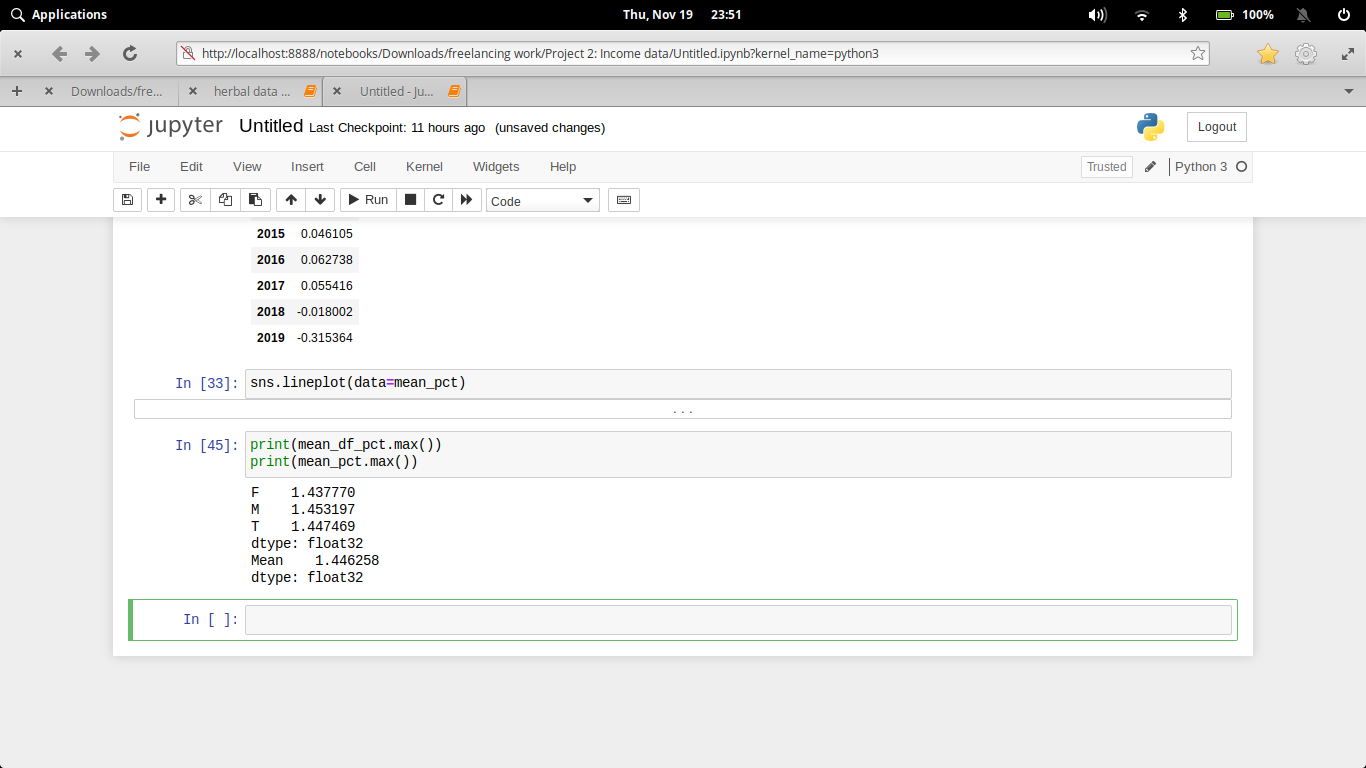
1. **What is the percentage rise in salary for different sex(gender) over a period of time?**

A decent percentage rise in salary is sustainiable, more than a limit rise in average salary is a sign of infation.

****

1. **What is the maximun percentage rise in salary over these different years?**

If the percentage rise according to the gender and mean percentage are somewhat similar, means that the country has given equal opportunity to all the genders.

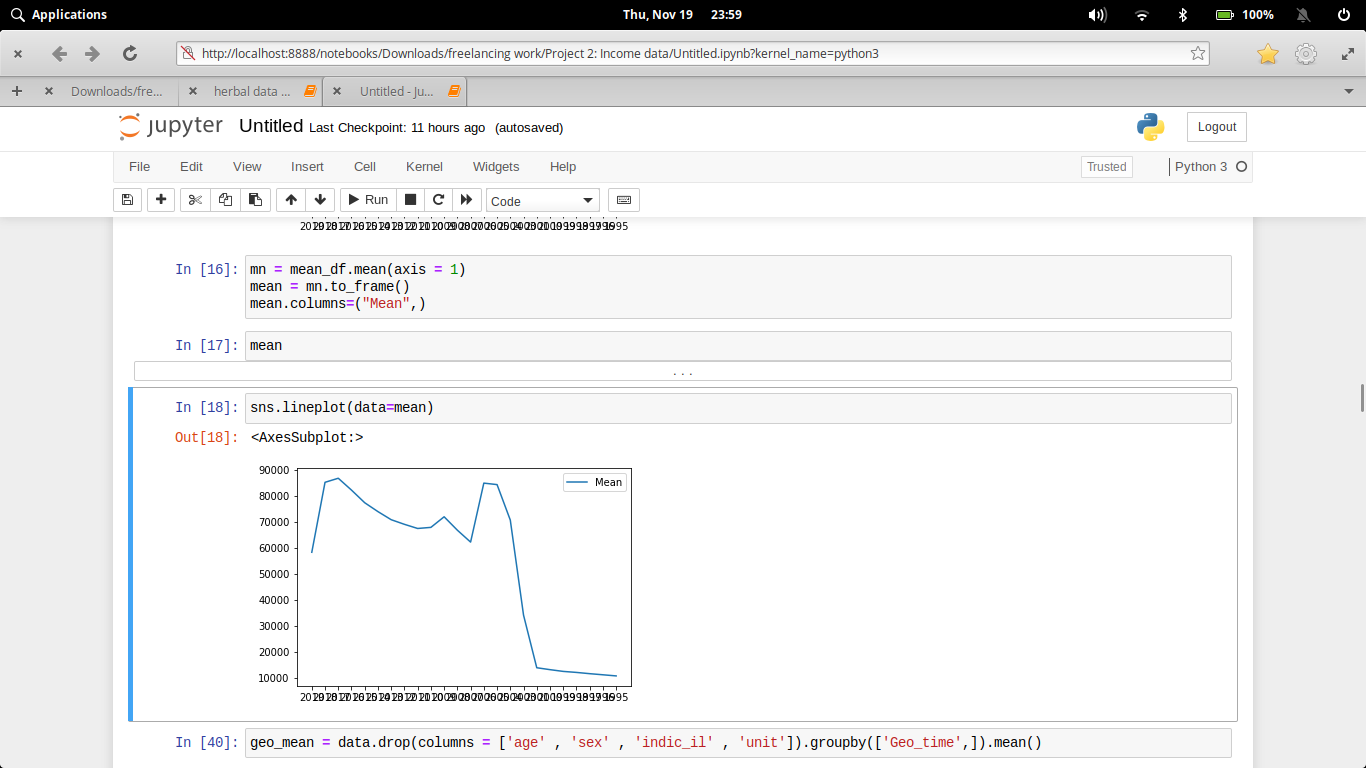
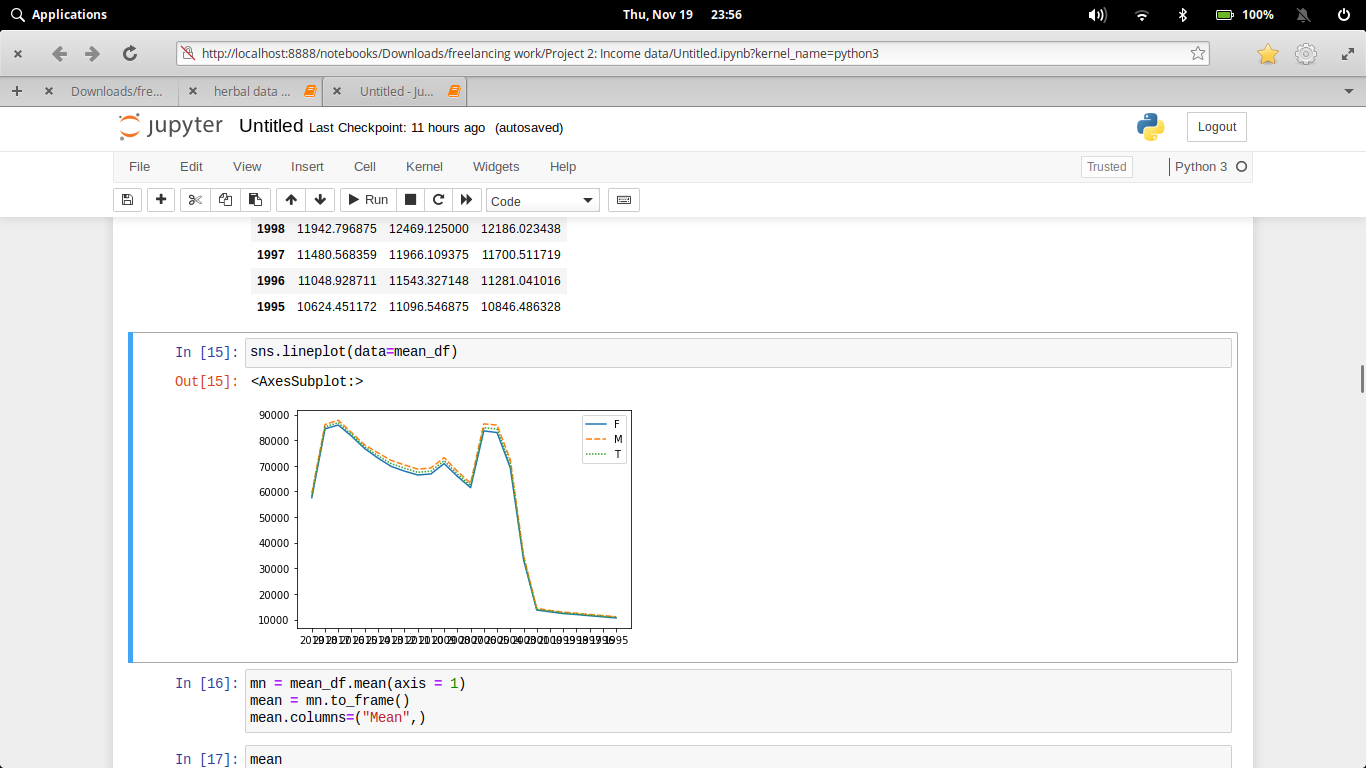


# Data Analysis – Visualization

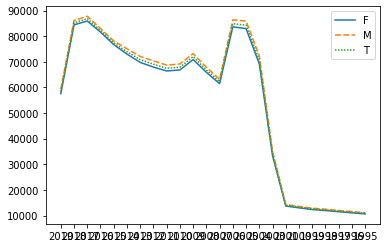
**Code and its output with visualization**

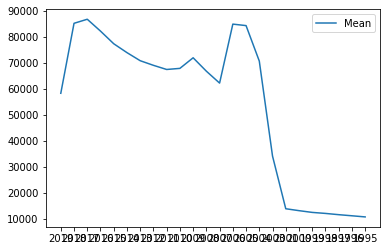
1. **What is the average mean salary of dfferent sex(gender) and in total in different consecutive years?**

**Code:-**



**Output:-**

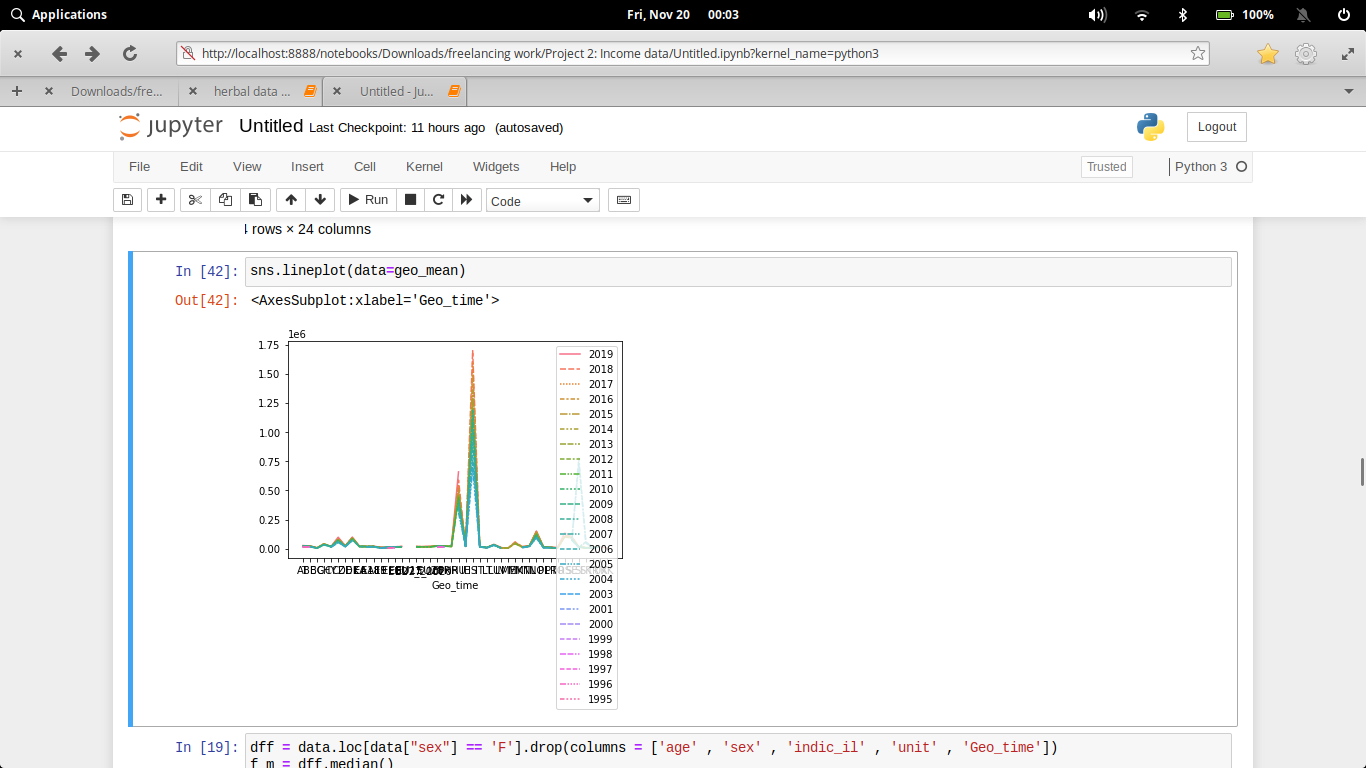
****

****

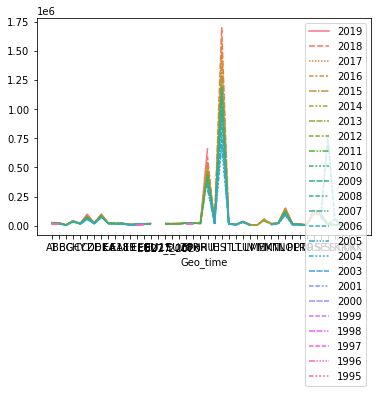
This visualization shows that either we don’t have sufficient data before 2002 or, the average income was very less before 2002, and increases exponentially thereafter.

1. **What is the average mean salary in different geograpfic locations?**

**Code:-**

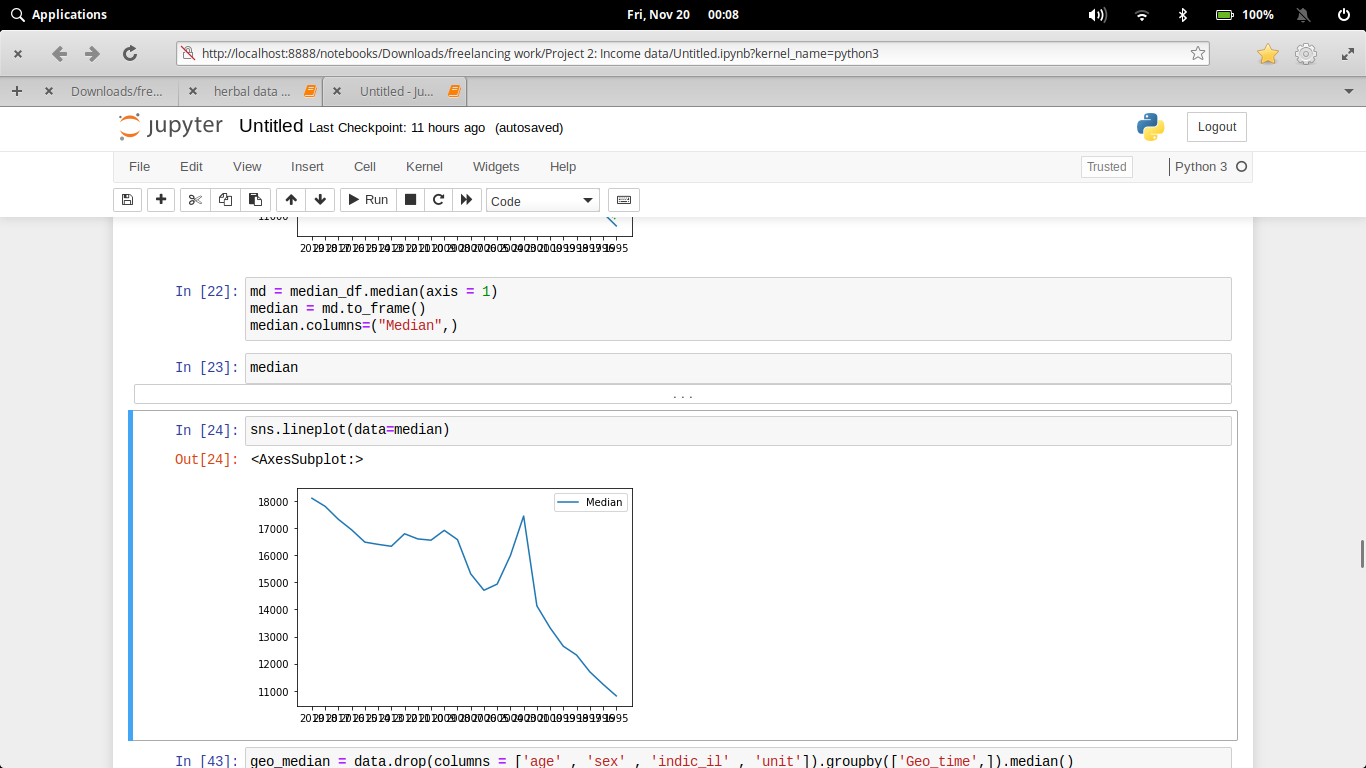
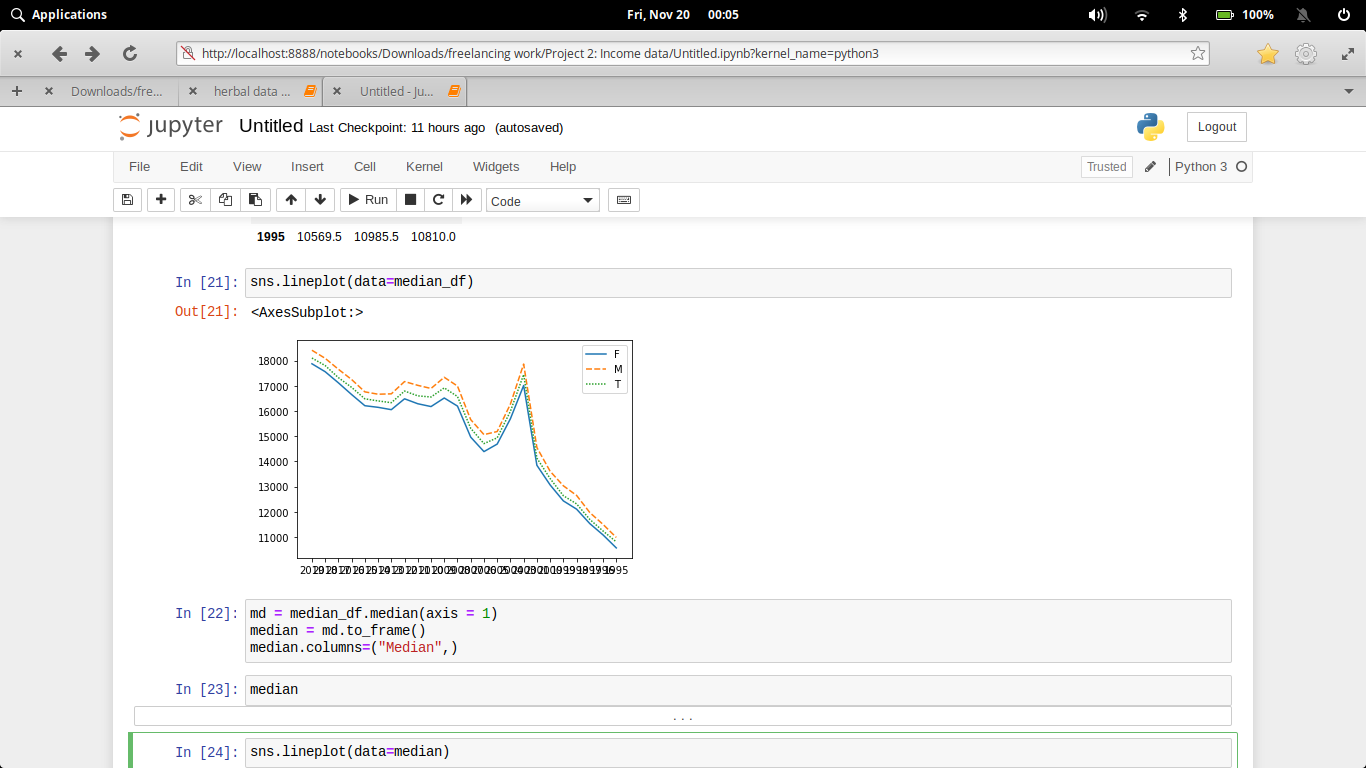
****

**Output:-**

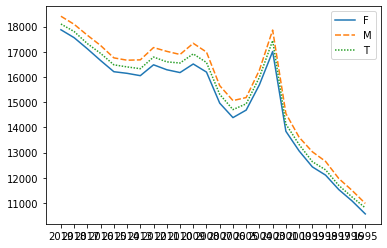
****

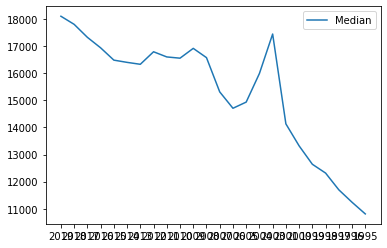
1. **What is the median salary of different sex(gender) and in total in different consecutive years?**

**Code:-**



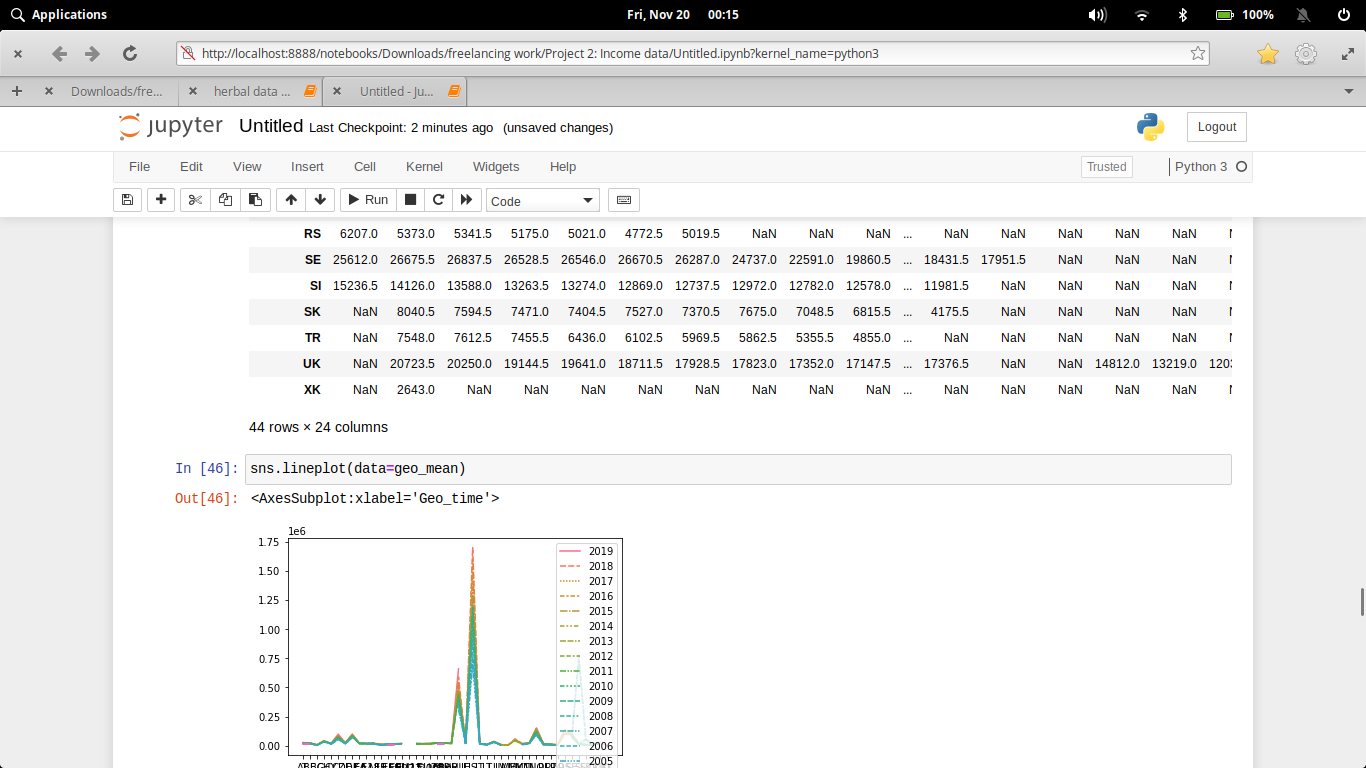
**Output:-**



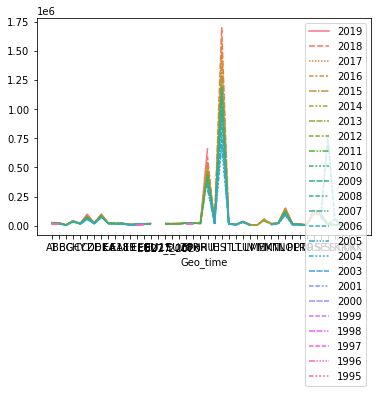


1. **What is the average median salary in different geographic locations?**

**Code:-**

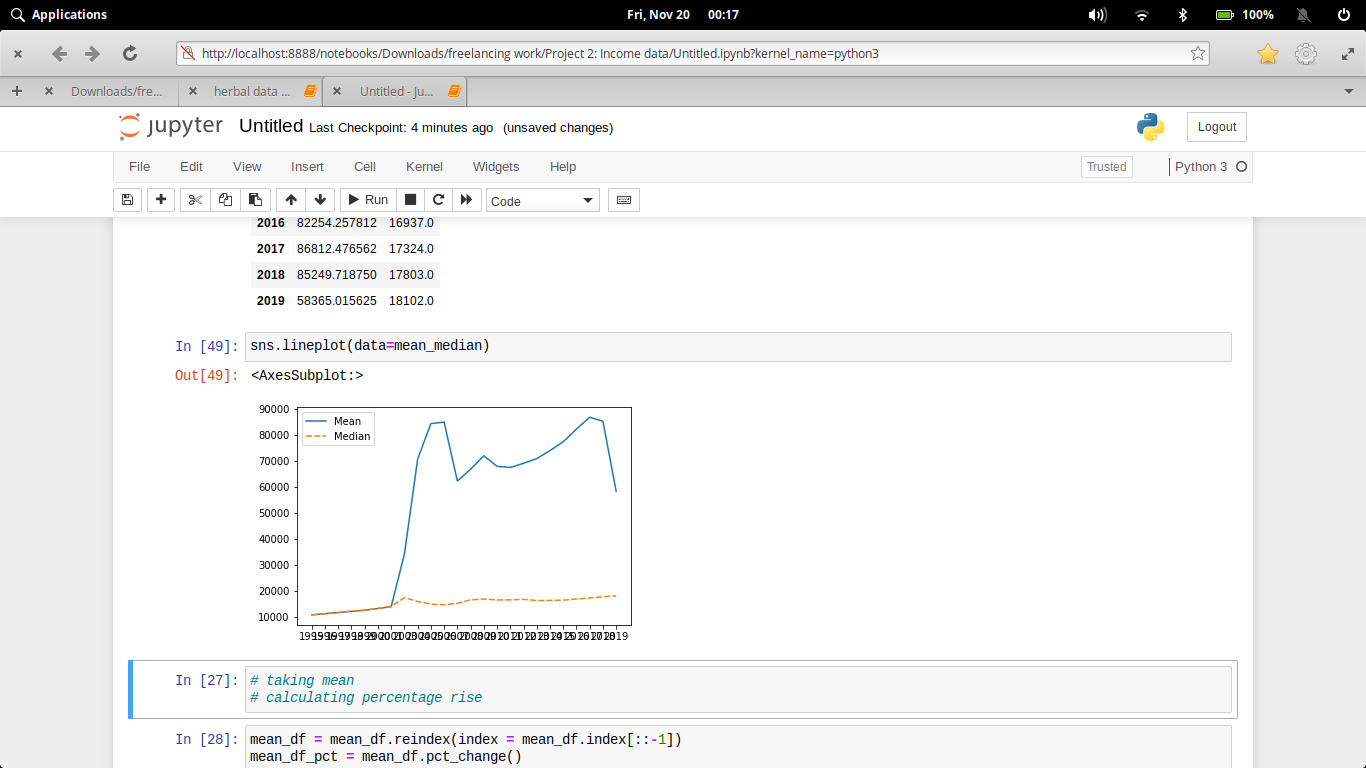
****

**Output:-**

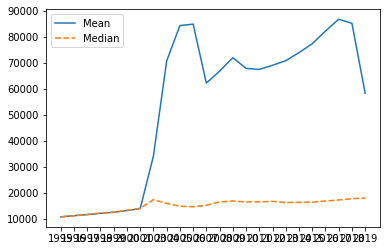
****

1. **Mean vs Median graph and Why mean is choosen over median to calculate the average number?**

**Code:-**

****

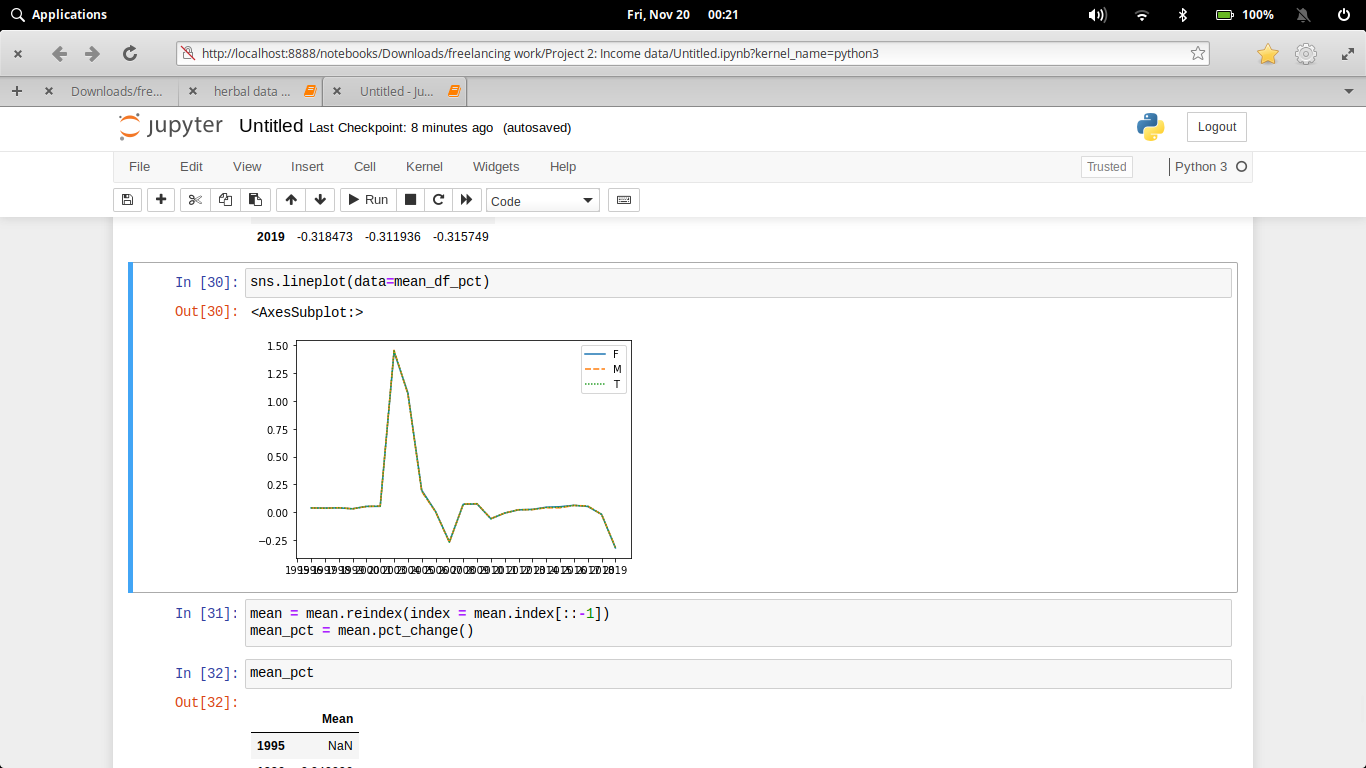
**Output:-**

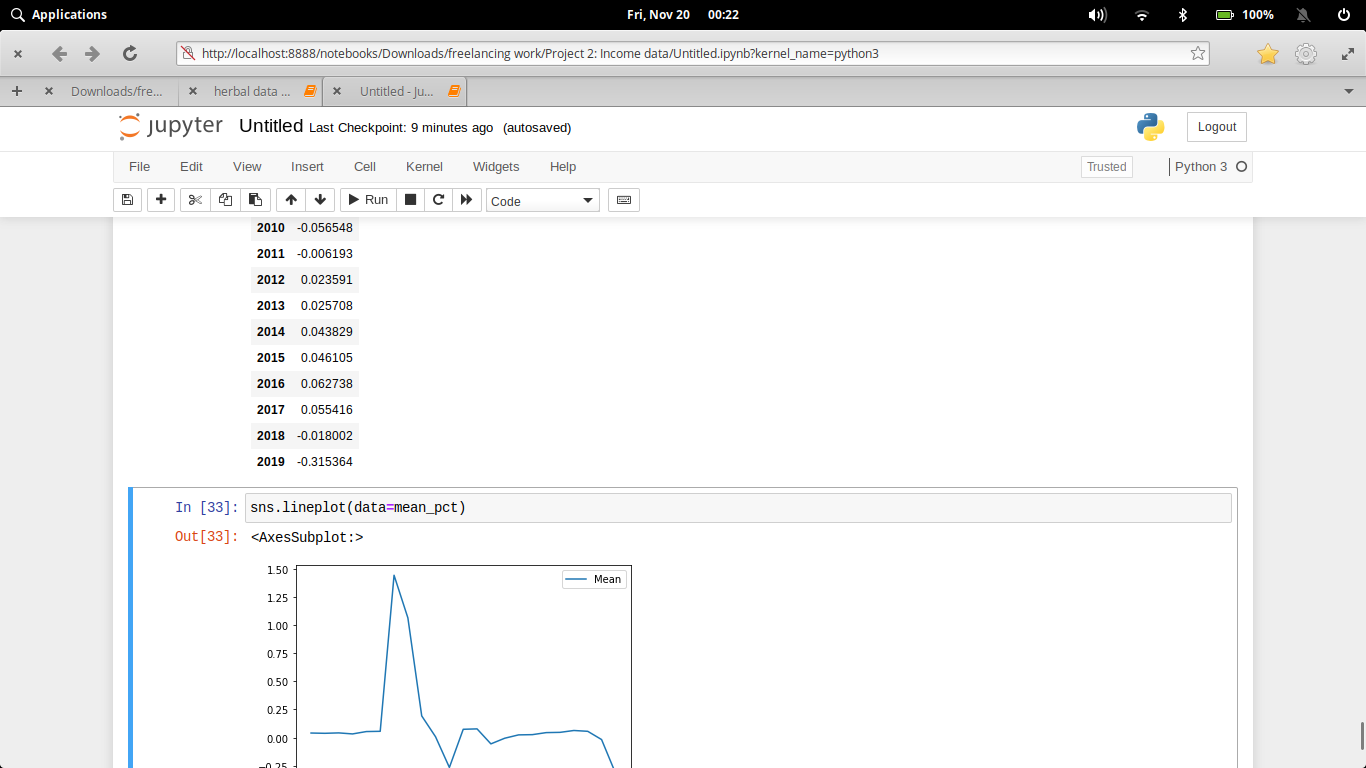


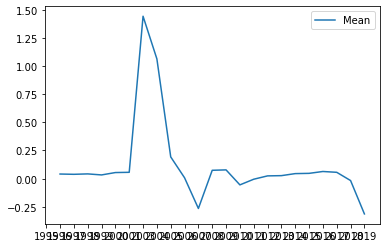
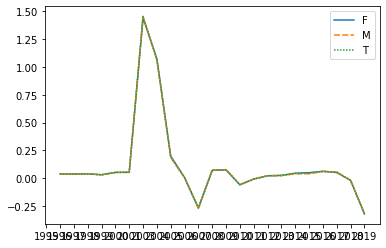
From this graph, we know that before 2002 we don’t have sufficient data, but after that, the median was quiet low flat as compared to mean. Because, mean depends on the whole.

1. **What is the percentage rise in salary for different sex(gender) over a period of time?**

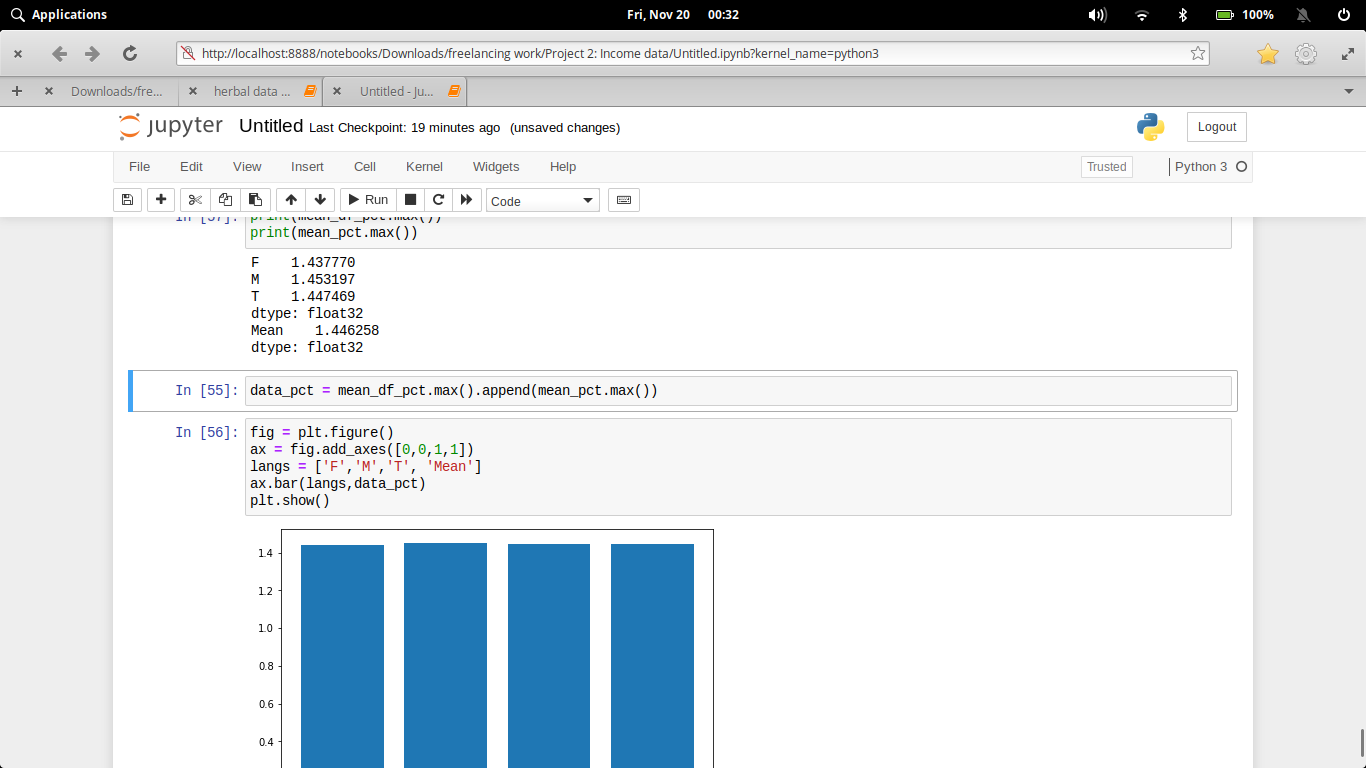
**Code:-**

****

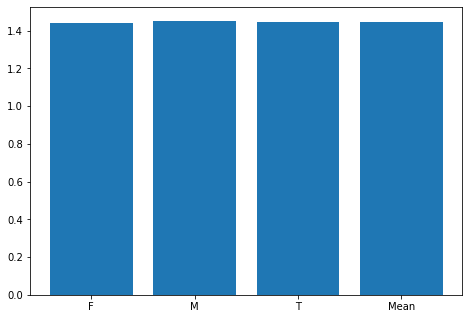


**Output:-**

1. **What is the maximun percentage rise in salary over these different years?**

**Code:-**

**Output:-**

****

Maximum percentage rise in salary is 1.4 each, and in year 2003.

# Executive Summary

**CONCLUSION**

From the above observations, we concluded:

1. Due to non availability of data, maximum increase in income is around 2003 which is 1.4 times followed by 0.19 times in 2005.

2. The female male and trans have equal rise in income suggesting that there is no gender discrimination.

3. From the groupby based on location, Switzerland has the maximum income per capita, over all the years followed by Denmark, Austria, Germany and Belgium.

# References

* https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php
* https://data.europa.eu/euodp/en/data
* https://www.purplemath.com/modules/meanmode.htm