## WORLD HAPPINESS RESEARCH REPORT

*A*

***Major Project Report***

*Submitted*

*In partial fulfillment*

*For the award of the Degree of*

## BACHELOR OF TECHNOLOGY

***In Department of Computer science and Engineering***

### Submitted By: Guided By: Submitted To :

### Palash Siyal Mr. Aaditya Mr. Gaurav

### Srushti Choudhri Maheshwari Kumawat



### Department of Computer Science and Engineering TECHNO INDIA NJR INSTITUTE OF TECHNOLOGY

RAJASTHAN TECHNICAL UNIVERSITY

**2022**

### TECHNO INDIA NJR INSTITUTE OF TECHNOLOGY



## CERTIFICATE

This is to certify that this project report **“World Happiness Research Report”** is the confide work of **“ Srushti Choudhri & Palash Siyal ”** who have carried out the project work under my supervision. I approve this project for submission of the Bachelor of Technology in the **Department of Computer Science and Engineering, Techno India NJR Institute of Technology**, affiliated to Rajasthan Technical University, Kota.

##### Mr. Aaditya Maheshwari Mr. Gaurav Kumawat

**Assistant Professor & Head Head of Department**

Department of Computer Science Department of Computer Science

## ABSTRACT

The World Happiness Research Report is a Data Science project portraying a detailed report for understanding how the happiness of citizens is affected by simple factors such as physical/mental health, work-life, and family support.

                The project will be based on the available global data on national happiness and reviewed based on the emerging knowledge & science of happiness, showing that the quality of people’s lives can be reasonably and authentically assessed by a variety of personalized well-being measures, collectively resulting in the happiness report.

# More and more nations are starting to assess the well-being of citizens—for example, by measuring life satisfaction and meaning in life. These measures can tell us who in society is suffering or getting better and what factors are responsible for increasing/decreasing happiness.

## ACKNOWLEDGEMENT

It gives me immense pleasure to express my deepest sense of gratitude and sincere thanks to my highly respected and esteemed guide **Mr. , Dept. of CSE TINJRIT** for their valuable guidance, encouragement and help for completing this work. Their useful suggestions for this whole work and co-operative behavior are sincerely acknowledged.

I would like to express my sincere thanks to **Mr. , Dept. of CSE TINJRIT** for giving me this opportunity to undertake this project.

I also wish to express my indebtedness to my parents as well as my family member whose blessings and support always helped me to face the challenges ahead.

At the end I would like to express my sincere thanks to all my friends and others who helped me directly or indirectly during this project work.

**Place: Udaipur**

**Date:**

## 

## TABLE OF CONTENTS

**Table of Contents**

[Abstract](#_TOC_250027) 3

[Acknowledgement 4](#_TOC_250026)

[Table of Contents 5](#_TOC_250025)

[List of Figures 7](#_TOC_250023)

[List of Symbols 8](#_TOC_250022)

CHAPTER – 1 Introduction 9-12

* 1. [Introduction 10](#_TOC_250020)
  2. [Scope 10](#_TOC_250019)
  3. [Document Overview 11](#_TOC_250019)
  4. [Parameters 11](#_TOC_250019)
  5. [Limitations 12](#_TOC_250019)

CHAPTER – 2 Project Requirement Specification 13-36

* 1. Hardware Specification 14
  2. Software Specification 19

CHAPTER – 3 Data Dictionary 37-42

3.1.[Dataset Elements 38](#_TOC_250006)

3.2.[Dataset Elements Description 38](#_TOC_250005)

CHAPTER - 4 Screen Shots 42-71

* 1. [Importing Dataset 43](#_TOC_250004)
  2. [Country Division 44](#_TOC_250003)
  3. [Ladder Score 46](#_TOC_250003)
  4. [GDP 50](#_TOC_250003)
  5. [Social Support 54](#_TOC_250003)
  6. [Healthy Life Expectancy 58](#_TOC_250003)
  7. [Freedom 62](#_TOC_250003)
  8. [Perception of Corruption 66](#_TOC_250003)

CHAPTER – 5 Limitations & Conclusion 70-73

* 1. [Limitations 71](#_TOC_250002)
  2. [Conclusion 71](#_TOC_250000)

CHAPTER – 6 Bibliography 74

### List of Figures

* Fig 2.1 - [Anaconda GUI 17](#_TOC_250000)
* Fig 2.2 - [Jupyter Notebook GUI 18](#_TOC_250000)
* Fig 2.3 - [Matplotlib UI Menu 23](#_TOC_250000)
* Fig 2.4 - [Matplotlib & NumPy 24](#_TOC_250000)
* Fig 2.5 - [Matplotlib & Pandas 24](#_TOC_250000)
* Fig 2.6 - [Line Plot 25](#_TOC_250000)
* Fig 2.7 - [Histogram 25](#_TOC_250000)
* Fig 2.8 - [3D Plot 25](#_TOC_250000)
* Fig 2.9 - [Line Plot 25](#_TOC_250000)
* Fig 2.10 - [Contour Plot 25](#_TOC_250000)
* Fig 2.11 - [Image Plot 25](#_TOC_250000)

**List of Symbols**

* AMD ………………………………………………… Advanced Micro Devices
* CPU …………………………………………………. Central Processing Unit
* RAM …………………………………………………. Random Access Memory
* HDD …………………………………………………. Hard Disk Drive
* SSD ……………………………………………………Solid State Drive
* OS ……………………………………………………. Operating System
* WWW ………………………………………………… World Wide Web
* HTML ………………………………………………… Hyper Text Markup Language
* CSS……………………………………………………. Cascading Style Sheets
* NumPy ………………………………………………… Numerical Python
* Pandas …………………………………………………. Python Data Analysis Library
* MATLAB ……………………………………………… Matrix Library
* Sklearn ………………………………………………… Scikit learn
* GUI ……………………………………………………. Graphical User Interface
* SQL ……………………………………………………. Structured Query Language

# ***CHAPTER : 1 INTRODUCTION***

**Introduction**

###### *Introduction*

The World Happiness Research Report is a Data Science project portraying a detailed report for understanding how the happiness of citizens is affected by simple factors such as physical/mental health, work-life, and family support.

                The project will be based on the available global data on national happiness and reviewed based on the emerging knowledge & science of happiness, showing that the quality of people’s lives can be reasonably and authentically assessed by a variety of personalized well-being measures, collectively resulting in the happiness report.

# More and more nations are starting to assess the well-being of citizens—for example, by measuring life satisfaction and meaning in life. These measures can tell us who in society is suffering or getting better and what factors are responsible for increasing/decreasing happiness.

###### *Scope*

Scope of this project is worldwide some of the points proving that:-

* The Happiness Index is a tool for the use of researchers, community organizers and policy makers seeking to understand and enhance individual happiness, community well-being, social justice, economic equality, and environmental sustainability.
* The Index reflects a new worldwide demand for more attention to happiness as a criterion for government policy
* The Index is also a tight slap on the face of policy makers, who think that pursuit of a healthy GDP number alone is sufficient to mark the happiness of a nation.

###### *Document Overview*

This document is meant for describing all the features and procedures that were followed while developing the system.

This document specially mentions the details of the project how it was developed, the primary requirement, as well as various features and functionalities of the project and the procedures followed in achieving these objectives.

World Happiness Report is responsible for recording the factors affecting human happiness all around the world which is why this report can be used in figuring out the reasons behind a country having more unhappy people, government can plan different policies for increasing the overall happiness of the country.

###### *Parameters*

There are many factors with strong impact on happiness including quality of work, social trust, freedom of choice and political participation by making goals that work towards strong communities with quality of employment, mental and physical health, support of family and quality education.

###### *Limitations*

* There is huge difference between evaluations and experiences of well-being. For example, Columbia came 37th rank in 2018 world happiness report but 1st by daily emotional experience.
* World Happiness Report are more appropriate for measuring national-level rather than individual-level happiness.

***CHAPTER : 2***

# ***PROJECT REQUIREMENT SPECIFICATION***

### HARDWARE REQUIRMENT SPECIFICATION

###### *Generation of Processor*

The processor is the main part of your computer. You must consider buying newer-generation processors. AMD and Intel processors are the best. AMD 5th Gen and Intel 11th Gen are the latest in the market. Intel 8th Gen or 10th Gen can also be considered. Processors advance with new generations. They give you the increased processing power, new hardware compatibility, thermal management, power efficiency, and much more.



#### Number of Cores and Threads

Cores(hardware) are the number of independent CPUs in a single chip. Threads are the instructions that are processed by a single CPU core. Almost every task you do in machine learning or data science needs parallel computations. So, in that case, the number of cores and threads of your CPU affect the performance significantly. Random forest algorithm in machine learning performs parallel computation. So, if you have a higher number of cores in your processor those can be put to use and random forest then can-do faster computations.

#### 3. Cache Memory

Cache memory is the buffer between CPU and RAM. It is used to store those instructions which are frequently used so that they are available immediately to the CPU when needed again. It is usually measured in Megabytes (MB). Higher the cache memory, the faster your computer. An 8 MB of cache memory is recommended.

#### 4. Base Clock speed: Frequency

The base frequency is defined as the minimum speed of the processor.

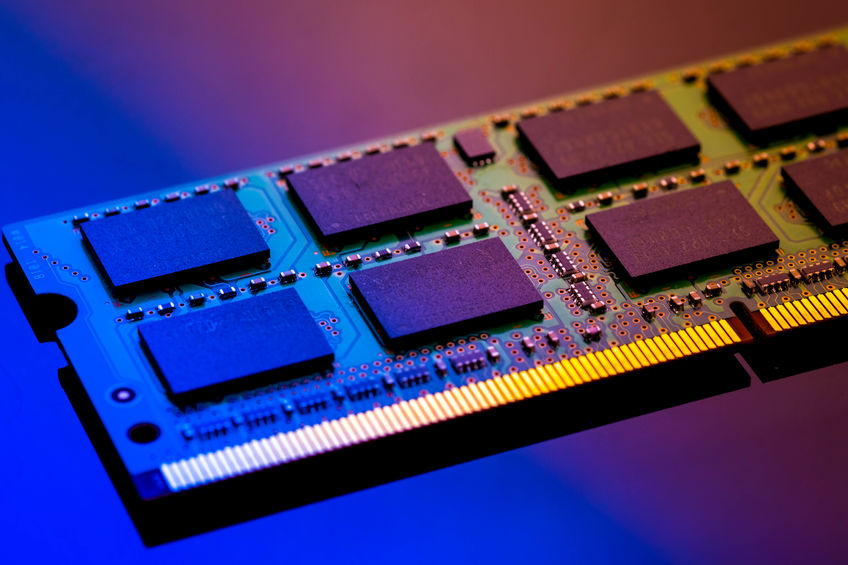
The higher the base frequency, the faster is the processor.

It is measured in Gigahertz (GHz).

#### 5. RAM size:

I think this is the first thing people look at while buying a laptop or computer and it is important also. But it should be very clear that increasing the RAM size is not going to increase the speed of your computer much.

Higher RAM allows you to multi-tasking. So, while selecting RAM you should go for 8GB or greater. 4GB is a strict no because more than 60 to 70% of it is used by Operating System and the remaining part is not enough for Data science tasks. If you can afford then go for 12 GB or 16GB RAM that is best.



*6. Storage type:*

Laptops come with HDDs i.e., Hard Disk Drives. They are really very slow. A laptop even with i7 and HDD is very slow. HDDs take much time to open and load a program because they have mechanical parts which delay the processing of information and reduces durability.  So, the best alternative for this is SSD (Solid State Drive). SSDs are much powerful than HDDs. They do not have any moving parts and provide good performance. Hence, try to go for SSD, the difference you will see is amazing.

7. Storage size:

Buying a laptop with SSD is already costly. Generally, HDDs come with 1TB but an SSD of 1TB may not be affordable so go for 512GB. But do not go below this.

*9. Brand of GPU:*

One might be aware that NVIDIA and AMD are the two major brands of Graphic Cards. A separate GPU is preferred because then you have a separate graphic card in which the average GPU has more than 100 cores whereas standard CPUs provide you only 4 or 8 cores.

TensorFlow deep learning library uses CUDA which compiles only on NVIDIA graphics cards. So, if you are going for deep learning tasks, recommended is to go for an NVIDIA GPU of 1650 or higher.



*10. Blue light filtering feature:*

Machine learning and data science students spend hours of time in front of their laptops. And as you know that almost every electronic display emits harmful blue lights. So, it damages your eyes. And your eyes are really important and you should keep good care of them. Hence go for laptops or monitors certified low-blue light and flicker-free screens. This is very important and shouldn’t be ignored.

*11. Display size:*

The display size must be 15.6 inches or 17.3 inches. Do not go below this. If you can afford go for bigger screens or monitors and connect them to your laptop. Display resolution should be a full HD (1080p) or an HD (720p) display.

*12. Operating System (OS):*

Windows 10 is recommended for data science students because it is user-friendly and configuring anything won’t take much effort. But you can go for another operating system as well like Linux. Install it separately or you can also run it in a virtual box.

*13. Reliability of laptop:*

Reliability is how often your laptop fails when operating. Some very common and annoying issues are blue screen errors, unexpected shutdowns, and other hardware-related issues. So, to avoid all these issues always go for a branded laptop.

### SOFTWARE REQUIRMENT SPECIFICATION

1. *Supported Web Browsers*

* ***Google Chrome:***

Google Chrome browser is an open-source program for accessing the World Wide Web (WWW) and running Web-based applications.

The Google Chrome Web browser is based on the open-source Chromium project. Google released Chrome in 2008 and issues several updates a year. It is available for Windows, Mac OSX, Linux, Android and iOS operating systems. The Google Chrome browser takes a sandboxing-based approach to Web security. Each open website runs as its own process, which helps prevent malicious code on one page from affecting others (or the computer operating system at large). The browser also supports Web standards such as HTML5 and cascading style sheets (CSS).



* ***Mozilla Firefox:***

Firefox is a Web browser that is smaller, faster, and in some ways more secure than the Mozilla browser from which much of its code was originally derived. Compared to Internet Explorer, the most popular Web browser, Firefox gives users a cleaner interface and faster download speeds.

Although not as extensive as Chrome's add-on library, Firefox, as open-source software with a huge number of active contributors, also features an incredible number of useful extensions. Firefox also has a sync feature to see your open and recent tabs, browsing history, and bookmarks across all your devices.

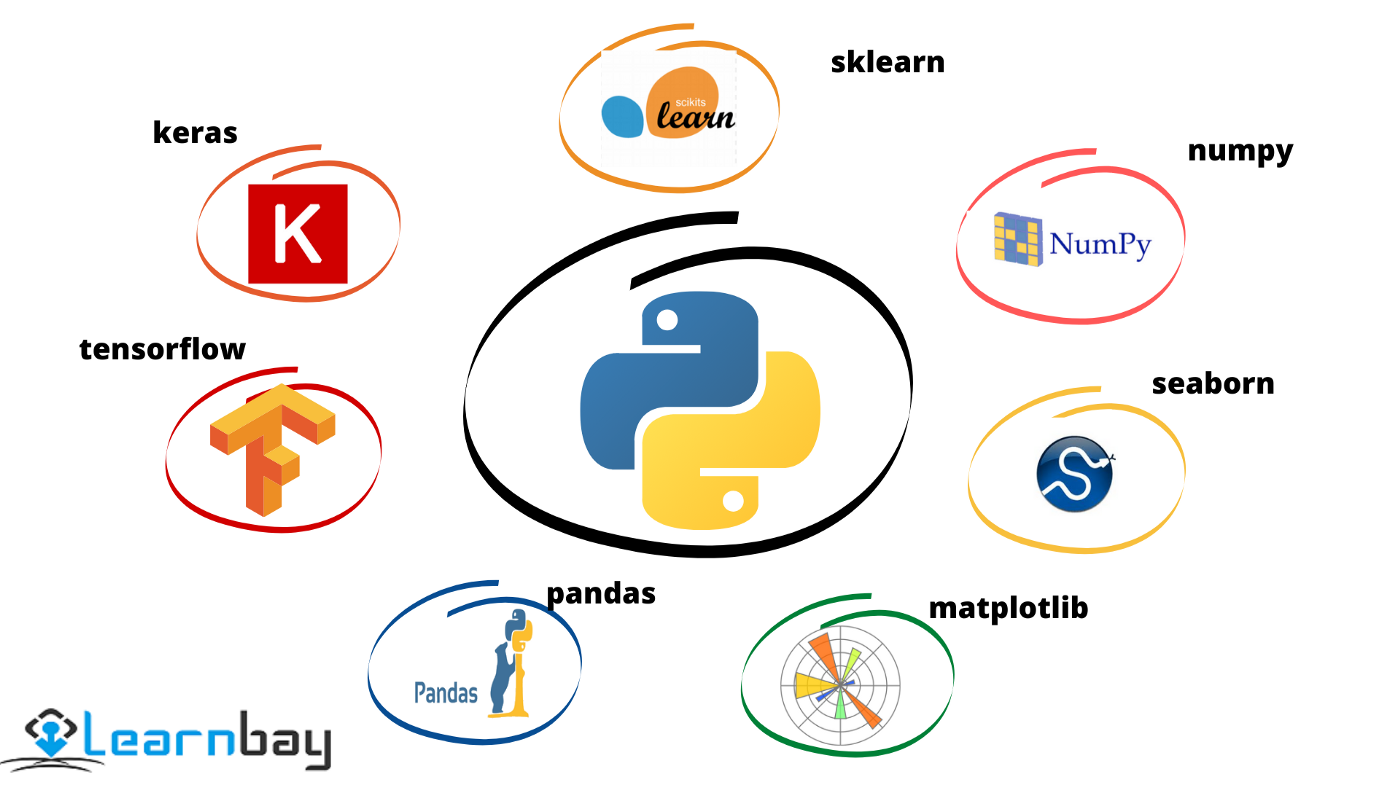


1. *Python*

Python is a popular multi-purpose programming language widely used for its flexibility, as well as its extensive collection of libraries, which are valuable for analytics and complex calculations.

Python’s extensibility means that it has thousands of libraries dedicated to analytics, including the widely used Python Data Analysis Library (also known as Pandas).

For the most part, data analytics libraries in Python are at least somewhat derived from the NumPy library, which includes hundreds of mathematical calculations, operations, and functions.



Additionally, Python’s performance capability is much higher than other popular languages used in data analytics, and its compatibility with a greater array of other languages means that it is simply more convenient in most cases.

Python’s relatively light usage of memory and other processing resources means that it can quickly outstrip languages like MATLAB or R, which are built specifically for statistical analysis.

***How Can I Use Python for Data Analytics?***

## There are several ways you can integrate python data analytics into your existing business intelligence and analytics tools.

**One of the most common uses** for Python is in its ability to create and manage data structures quickly — Pandas, for instance, offers a plethora of tools to manipulate, analyze, and even represent data structures and complex datasets.

This includes time series and more complex data structures such as merging, pivoting, and slicing tables to create new views and perspectives on existing sets.

**Elsewhere**, tools like Scikit-Learn (also known as Sklearn) provides advanced analytics tools combined with complex machine learning capabilities.

This allows you to build more sophisticated models, performing more complex and multivariate regressions, as well as data preprocessing.

Combined with libraries such as iPython and NumPy itself, these tools can form the foundation of a powerful data analytics suite.

1. *Anaconda*

Anaconda is an open-source distribution of the Python and R programming languages for data science that aims to simplify package management and deployment. Package versions in Anaconda are managed by the package management system, conda, which analyzes the current environment before executing an installation to avoid disrupting other frameworks and packages.

The Anaconda distribution comes with over 250 packages automatically installed. Over 7500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI (graphical user interface), Anaconda Navigator, as a graphical alternative to the command line interface. Anaconda Navigator is included in the Anaconda distribution, and allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages, install them in an environment, run the packages and update them.

The big difference between conda and the pip package manager is in how package dependencies are managed, which is a significant challenge for Python data science. When pip installs a package, it automatically installs any dependent Python packages without checking if these conflict with previously installed packages. It will install a package and any of its dependencies regardless of the state of the existing installation.

While Anaconda supports some functionality you find in a data science platform, like [Domino](https://www.dominodatalab.com/product/), it provides a subset of that functionality. Domino and other platforms not only support package management, but they also support capabilities like collaboration, reproducibility, scalable compute, and [model monitoring](https://www.dominodatalab.com/resources/model-monitoring-best-practices/). Conda can be used within the Domino environment.

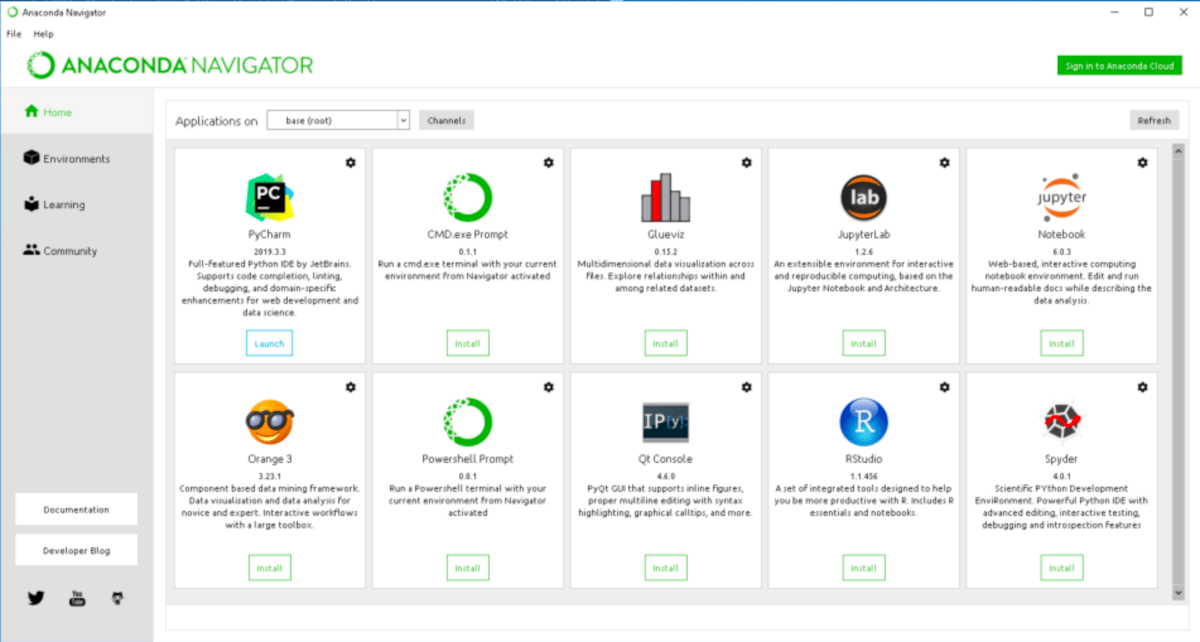


Figure 2.1

1. *Jupiter Notebook*

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The “notebook” term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context.

According to the official website of [Jupyter](https://jupyter.org/), Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

Jupyter Book is an open-source project for building books and documents from computational material. It allows the user to construct the content in a mixture of Markdown, an extended version of Markdown called MyST, Maths & Equations using Math Jax, Jupyter Notebooks, reStructuredText, the output of running Jupyter Notebooks at build time. Multiple output formats can be produced (currently single files, multipage HTML web pages and PDF files).

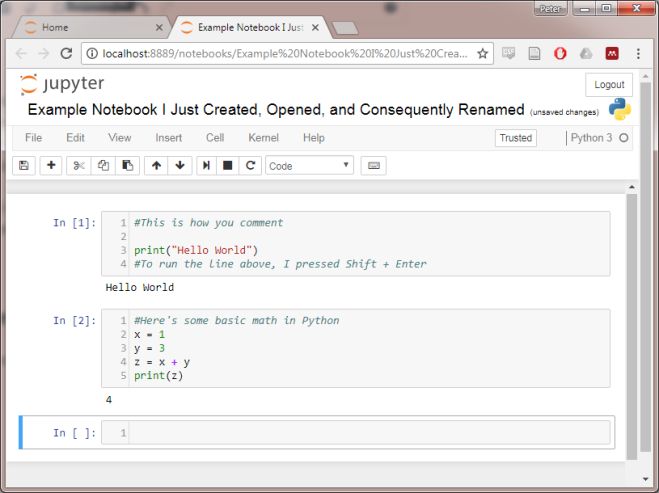


Figure 2.2

1. *Libraries*

* ***Numpy***

NumPy (**Numerical Python**) is an open-source Python library that’s used in almost every field of science and engineering. It’s the universal standard for working with numerical data in Python, and it’s at the core of the scientific Python and PyData ecosystems. NumPy users include everyone from beginning coders to experienced researchers doing state-of-the-art scientific and industrial research and development. The NumPy API is used extensively in Pandas, SciPy, Matplotlib, scikit-learn, scikit-image and most other data science and scientific Python packages.

The NumPy library contains multidimensional array and matrix data structures. It provides **Nd array**, a homogeneous n-dimensional array object, with methods to efficiently operate on it. NumPy can be used to perform a wide variety of mathematical operations on arrays. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.



* ***Pandas***

Pandas is a [Python](https://www.python.org/) package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open-source data analysis/manipulation tool available in any language. It is already well on its way toward this goal.

Pandas is well suited for many different kinds of data:

* Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
* Ordered and unordered (not necessarily fixed-frequency) time series data.
* Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
* Any other form of observational / statistical data sets. The data need not be labeled at all to be placed into a Pandas data structure

The two primary data structures of pandas, [Series](https://pandas.pydata.org/docs/reference/api/pandas.Series.html#pandas.Series) (1-dimensional) and [Data Frame](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html#pandas.DataFrame) (2-dimensional), handle the vast majority of typical use cases in finance, statistics, social science, and many areas of engineering. For R users, [Data Frame](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html#pandas.DataFrame) provides everything that R’s data frame provides and much more. pandas is built on top of [NumPy](https://numpy.org/) and is intended to integrate well within a scientific computing environment with many other 3rd party libraries.

Here are just a few of the things that pandas does well:

* Easy handling of missing data (represented as NaN) in floating point as well as non-floating point data
* Size mutability: columns can be inserted and deleted from Data Frame and higher dimensional objects
* Automatic and explicit data alignment: objects can be explicitly aligned to a set of labels, or the user can simply ignore the labels and let Series, Data Frame, etc. automatically align the data for you in computations
* Powerful, flexible group by functionality to perform split-apply-combine operations on data sets, for both aggregating and transforming data
* Make it easy to convert ragged, differently-indexed data in other Python and NumPy data structures into Data Frame objects
* Intelligent label-based slicing, fancy indexing, and sub setting of large data sets
* Intuitive merging and joining data sets
* Flexible reshaping and pivoting of data sets
* Hierarchical labeling of axes (possible to have multiple labels per tick)
* Robust IO tools for loading data from flat files (CSV and delimited), Excel files, databases, and saving / loading data from the ultrafast HDF5 format
* Time series-specific functionality: date range generation and frequency conversion, moving window statistics, date shifting, and lagging.

Many of these principles are here to address the shortcomings frequently experienced using other languages / scientific research environments. For data scientists, working with data is typically divided into multiple stages: munging and cleaning data, analyzing / modeling it, then organizing the results of the analysis into a form suitable for plotting or tabular display. pandas is the ideal tool for all of these tasks.



* ***Matplotlib***

**Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib’s APIs (Application Programming Interfaces) to embed plots in GUI applications.**

A Python matplotlib script is structured so that a few lines of code are all that is required in most instances to generate a visual data plot. The matplotlib scripting layer overlays two APIs:

* The pyplot API is a hierarchy of Python code objects topped by matplotlib.pyplot
* An OO (Object-Oriented) API collection of objects that can be assembled with greater flexibility than pyplot. This API provides direct access to Matplotlib’s backend layers.

### *Matplotlib and Pyplot in Python*

The pyplot API has a convenient MATLAB-style stateful interface. In fact, matplotlib was originally written as an open-source alternative for MATLAB. The OO API and its interface is more customizable and powerful than pyplot, but considered more difficult to use. As a result, the pyplot interface is more commonly used, and is referred to by default in this article.

Understanding matplotlib’s pyplot API is key to understanding how to work with plots:

* **matplotlib.pyplot.figure: Figure** is the top-level container. It includes everything visualized in a plot including one or more **Axes**.
* **matplotlib.pyplot.axes**:**Axes** contain most of the elements in a plot**: Axis, Tick, Line2D, Text,**etc., and sets the coordinates. It is the area in which data is plotted. Axes include the X-Axis, Y-Axis, and possibly a Z-Axis, as well.

### *Matplotlib UI Menu*

When matplotlib is used to create a plot, a User Interface (UI) and menu structure are generated. The UI can be used to customize the plot, as well as to pan/zoom and toggle various elements.

### What is matplotlib figure 1

Figure 2.3

### *Matplotlib and NumPy*

NumPy is a package for scientific computing. NumPy is a required dependency for matplotlib, which uses NumPy functions for numerical data and multi-dimensional arrays as shown in the following code snippet:

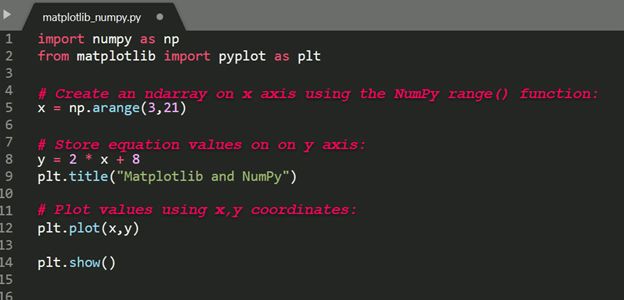


Figure 2.4

### *Matplotlib and Pandas*

Pandas is a library used by matplotlib mainly for data manipulation and analysis. Pandas provides an in-memory 2D data table object called a Data frame. Unlike numpy, pandas is not a required dependency of matplotlib. Pandas and NumPy are often used together, as shown in the following code snippet:

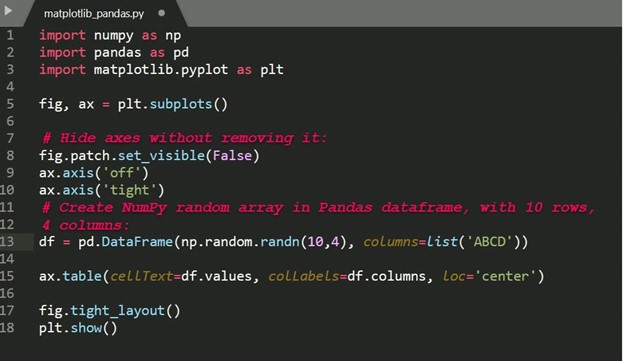
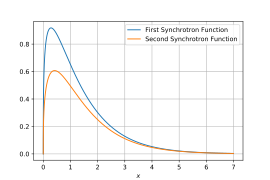
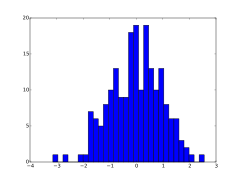


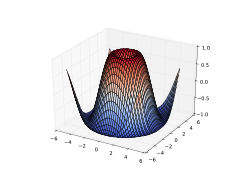
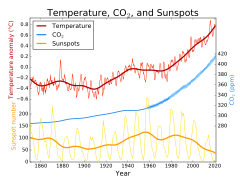
Figure 2.5

Pyplot is a Matplotlib module which provides a MATLAB-like interface. Matplotlib is designed to be as usable as MATLAB, with the ability to use Python, and the advantage of being free and open-source.

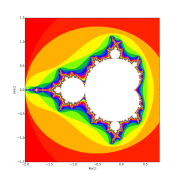
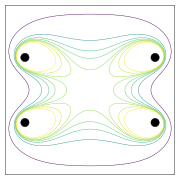
* ***Some Examples***

Line plot Figure 2.6 Histogram Figure 2.7

3D plot Figure 2.8 Line plot Figure 2.9



Contour plot Figure 2.10 Image plot Figure 2.11

* ***Seaborn***

Seaborn is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on the top of [matplotlib](https://www.geeksforgeeks.org/python-introduction-matplotlib/) library and also closely integrated to the data structures from [pandas](https://www.geeksforgeeks.org/introduction-to-pandas-in-python/).

Seaborn aims to make visualization the central part of exploring and understanding data. It provides dataset-oriented APIs, so that we can switch between different visual representations for same variables for better understanding of dataset.

## Different categories of plot in Seaborn

Plots are basically used for visualizing the relationship between variables. Those variables can be either be completely numerical or a category like a group, class or division. Seaborn divides plot into the below categories – 

* **Relational plots:** This plot is used to understand the relation between two variables.
* [**Categorical plots:**](https://www.geeksforgeeks.org/seaborn-categorical-plots/)This plot deals with categorical variables and how they can be visualized.
* [**Distribution plots:**](https://www.geeksforgeeks.org/seaborn-distribution-plots/)This plot is used for examining univariate and bivariate distributions
* [**Regression plots:**](https://www.geeksforgeeks.org/seaborn-regression-plots/)The regression plots in seaborn are primarily intended to add a visual guide that helps to emphasize patterns in a dataset during exploratory data analyses.
* [**Matrix plots:**](https://www.geeksforgeeks.org/ml-matrix-plots-in-seaborn/) A matrix plot is an array of scatterplots.
* **Multi-plot grids:**It is an useful approach is to draw multiple instances of the same plot on different subsets of the dataset.

## API abstraction across visualizations

There is no universally best way to visualize data. Different questions are best answered by different plots. Seaborn makes it easy to switch between different visual representations by using a consistent dataset-oriented API.

The function [relplot()](https://seaborn.pydata.org/generated/seaborn.relplot.html#seaborn.relplot) is named that way because it is designed to visualize many different statistical relationships. While scatter plots are often effective, relationships where one variable represents a measure of time are better represented by a line.

## Statistical estimation and error bars

Often, we are interested in the average value of one variable as a function of other variables. Many seaborn functions will automatically perform the statistical estimation.

When statistical values are estimated, seaborn will use bootstrapping to compute confidence intervals and draw error bars representing the uncertainty of the estimate.

Statistical estimation in seaborn goes beyond descriptive statistics. For example, it is possible to enhance a scatterplot by including a linear regression model (and its uncertainty) using [lmplot()](https://seaborn.pydata.org/generated/seaborn.lmplot.html#seaborn.lmplot).

## Informative distributional summaries

Statistical analyses require knowledge about the distribution of variables in your dataset. The seaborn function [displot()](https://seaborn.pydata.org/generated/seaborn.displot.html#seaborn.displot) supports several approaches to visualizing distributions. These include classic techniques like histograms and computationally-intensive approaches like kernel density estimation.

Seaborn also tries to promote techniques that are powerful but less familiar, such as calculating and plotting the empirical cumulative distribution function of the data.

## Specialized plots for categorical data

Several specialized plot types in seaborn are oriented towards visualizing categorical data. They can be accessed through [**cat plot()**](https://seaborn.pydata.org/generated/seaborn.catplot.html#seaborn.catplot). These plots offer different levels of granularity. At the finest level, you may wish to see every observation by drawing a “swarm” plot: a scatter plot that adjusts the positions of the points along the categorical axis so that they don’t overlap.

Alternately, you could use kernel density estimation to represent the underlying distribution that the points are sampled from.

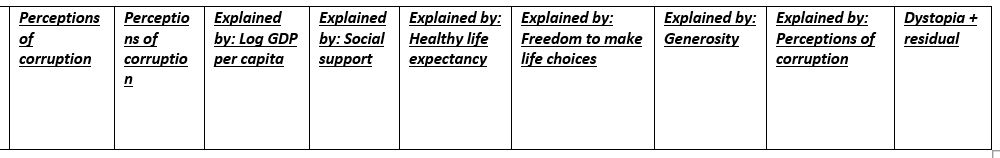
# ***CHAPTER: 3***

***DATA DICTIONARY***

**DATA DICTIONARY**

* **Dataset Elements:**

****

****

**Description of the Data Elements:**

1. ***Country Name:***

The country names used as headings are not official names, but are the commonly known names of the countries in English. *Ethnologic* uses the ISO 3166 standard as a starting point in determining what geopolitical entities to list as countries. As a consequence, some political dependencies are listed in a country section of their own while others are included within the larger country with which they are associated. We generally do not create a separate country section for smaller geopolitical entities which have only one widespread language. The *ethnologic* takes no position on issues of national sovereignty by this arrangement which is intended solely to facilitate the navigation of the published information.

The information elements reported on the main page for each country are as follows:

## Official name

If the name used by the country in its official documents differs from the popular English name as given in the heading, the official name of the country is listed here. There may be more than one official name listed in more than one language. In a few cases, additional or former names used to identify the country are also included.

## Sovereignty status

If the geopolitical entity is not a fully sovereign nation, a comment is given here to describe its status in relation to the sovereign state with which it is associated.

## Population

These figures are taken from the most recent national census data were available or are the current estimated population from the United Nations or another reliable source, which is identified. Country populations from these sources may be estimates based on population trends rather than the results of actual head counts.

## General remarks

The country information may also contain general remarks about the political status, the geography, or the population.

## Principal languages

Languages that have been identified as having a working function at the nation-wide level are listed here, whether this is by statute or is the de facto situation. For a fuller discussion of what we mean by a working function, see [Official recognition](https://www.ethnologue.com/about/language-status#FICLabels).

## Literacy rate

This is an estimate of the percentage of the population in the country that is literate in any language. Data are primarily from UNESCO but may come from various other sources if more recent estimates are available.

## International conventions

We have identified 9 conventions within the body of international law that affirm the language and culture rights of indigenous and minority peoples. This element of the country information lists which of these conventions the country in focus has subscribed to. Knowing this information can be of use to those advocating for indigenous and minority languages within the country.

1. ***Regional indicator:***

The regional indicator symbols are a set of 26 alphabetic [Unicode](https://en.wikipedia.org/wiki/Unicode) characters (A–Z) intended to be used to encode [ISO 3166-1 alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2) two-letter [country codes](https://en.wikipedia.org/wiki/Country_code) in a way that allows optional special treatment.

These were defined in October 2010 as part of the [Unicode 6.0](https://en.wikipedia.org/wiki/Unicode#6.0) support for [emoji](https://en.wikipedia.org/wiki/Emoji), as an alternative to encoding separate characters for each country flag. Although they can be displayed as Roman letters, it is intended that implementations may choose to display them in other ways, such as by using [national flags](https://en.wikipedia.org/wiki/National_flag).[[1]](https://en.wikipedia.org/wiki/Regional_indicator_symbol#cite_note-1)[[2]](https://en.wikipedia.org/wiki/Regional_indicator_symbol#cite_note-n3727-2) The Unicode FAQ indicates that this mechanism should be used and that symbols for national flags will not be directly encoded.

1. ***Ladder score:***

Cantril's Ladder is an overall measure of how someone feels about their life using two simple questions. The first question asks people to rate their lives on a ladder where the bottom in their worst possible life (0) and the top is their best possible life (10). The second question asks them to rate where they think they will be on the ladder in five years. A score of 7 or higher now and 8 or higher in the future on Cantril's ladder puts someone in the "thriving" category, along with half of the US population. A score of 4 or below on both questions puts you in the "suffering" category. Everyone in between is categorized as "struggling."

These two Cantril's ladder questions can be used to identify which populations one might need to focus on. It can also be used to monitor the health of a population over time. It can be used to evaluate improvement efforts. Change in this measure can be seen as soon as 3-6 months and is often used by employers and communities in this way.

* 1. ***GDP:***

GDP per person has a strong relationship with national happiness, explaining 56% of the differences across countries.

But the richer someone is, the smaller the boost in happiness from becoming richer. Happiness plateaus as the average income in society reaches $70,000.

The shape of the relationship between happiness and income is significant for two reasons. Firstly, in the broader context of inequality this supports the case for distributive policies. Secondly, for rich countries, there may be more effective ways for society to boost happiness than by solely pursuing more economic growth.

***v. Perceptions of corruption:***

The perception of corruption data that we use in our work is in fact a measure of institutional trust, which captures both government and business corruption. More specifically, survey respondents are asked whether corruption is widespread in government and business? This also has a good relationship with broader trust in society, where people are asked whether they can trust others.

It is no surprise that emerging markets have higher perceptions of corruption than developed markets. Within developed markets, the Nordic countries have the lowest perception of corruption and highest social trust levels. This helps to explain why these countries consistently rank as the some of the happiest in the world.

* 1. ***Healthy Life:***

The measure we use is *self-reported health* – the share of those reporting “at least good” health. Importantly, this reflects that self-assessed health combines both mental and physical health. Mental health is especially important to consider in advanced economies, where life expectancy and other measures of physical health are already very high.

* 1. ***Social support:***

Considering the power of social support to predict happiness, social support could increase self-confidence, self-disclosure, and self-esteem in an individual, thereby helping him/her achieve goals, satisfaction with life, and, consequently, happiness, and satisfying relationships not only make people happy, but they are also associated with better health and even longer life.

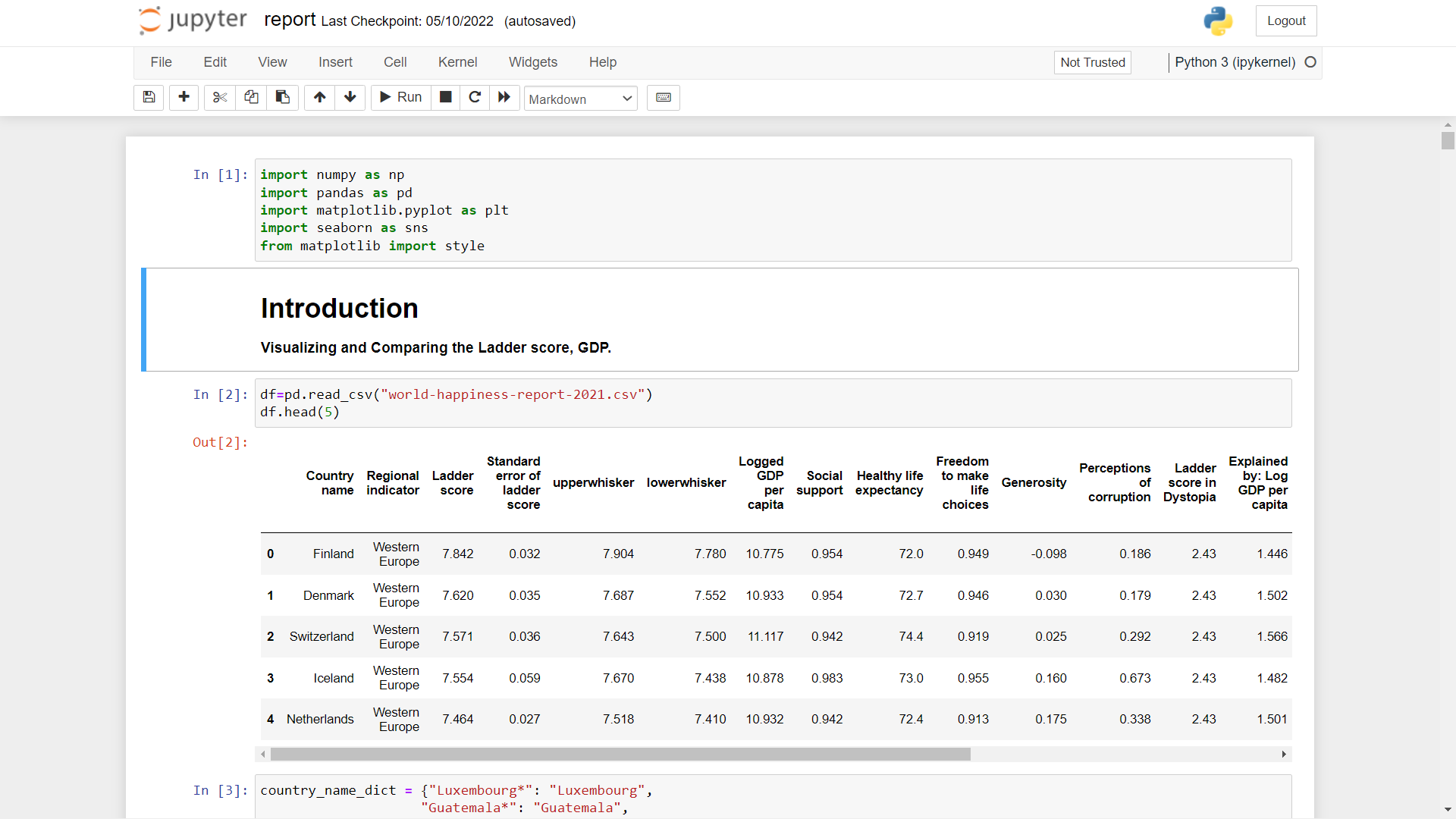
* 1. ***Generosity:***

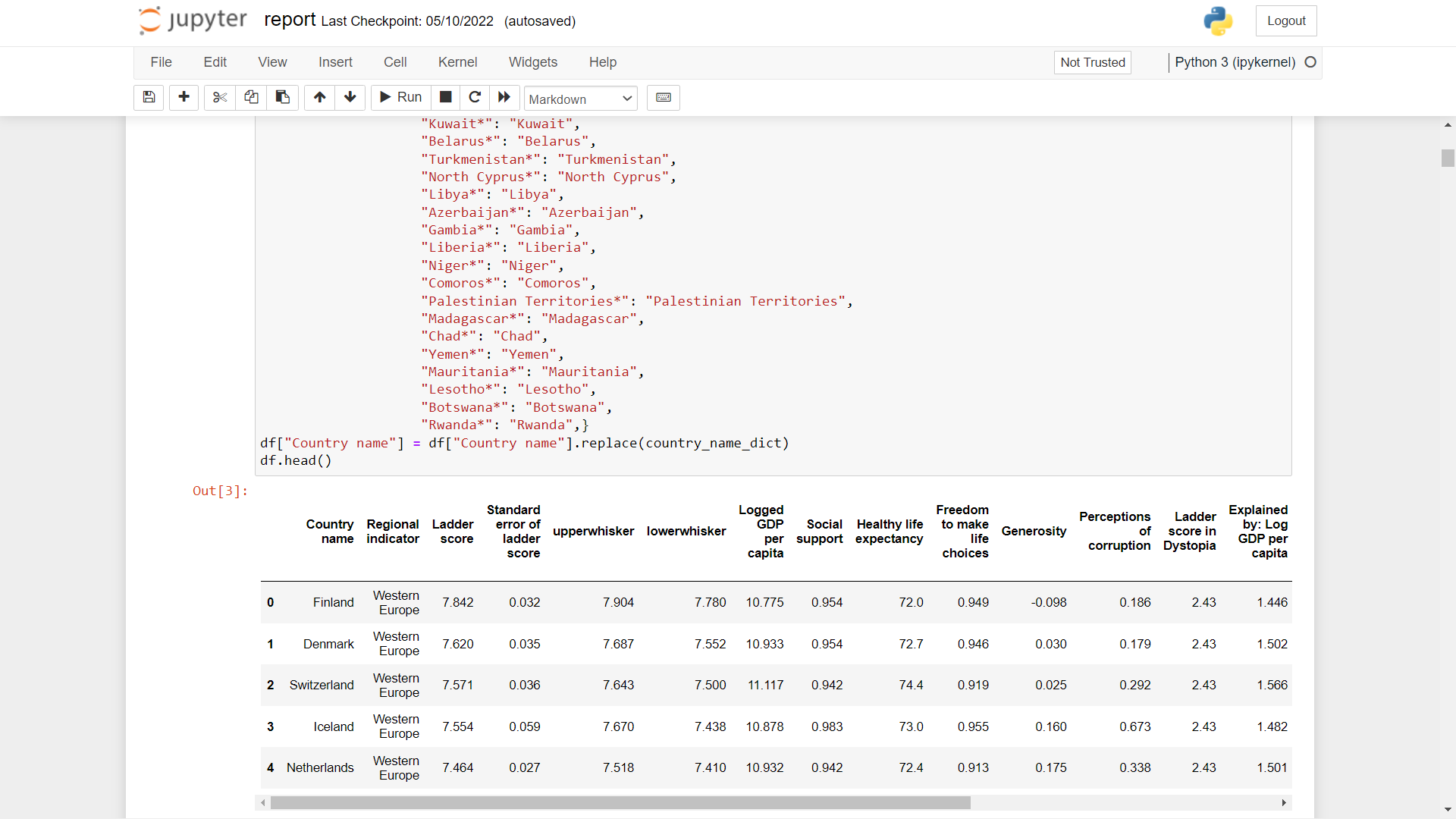
Some studies have found that people are happier when spending money on others than on themselves, and this happiness motivates them to be generous in the future. And even small acts of kindness, like picking up something someone else has dropped, make people feel happy.

# ***CHAPTER: 4***

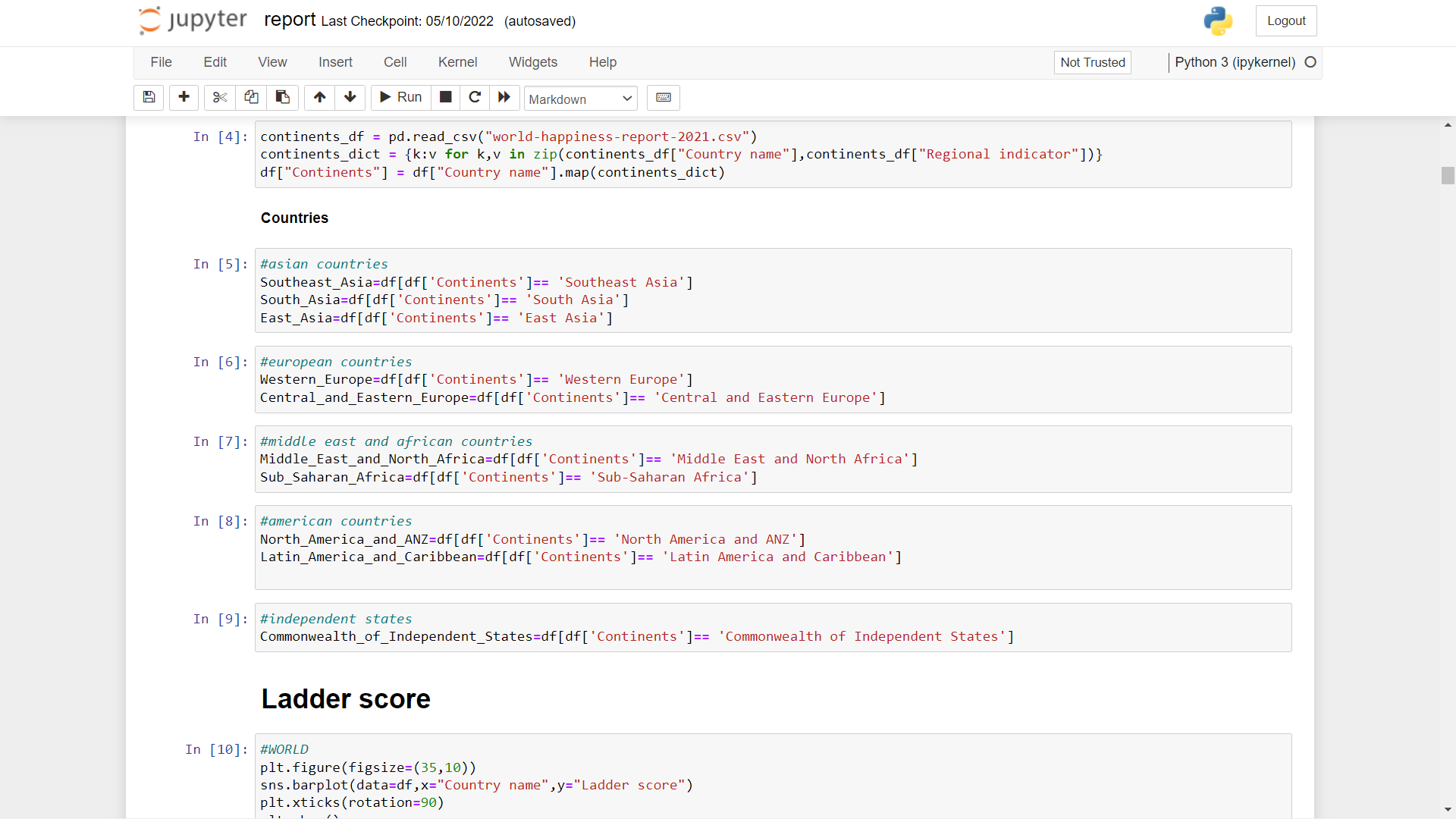
***SCREEN SHOTS***

1. **Importing Dataset –**



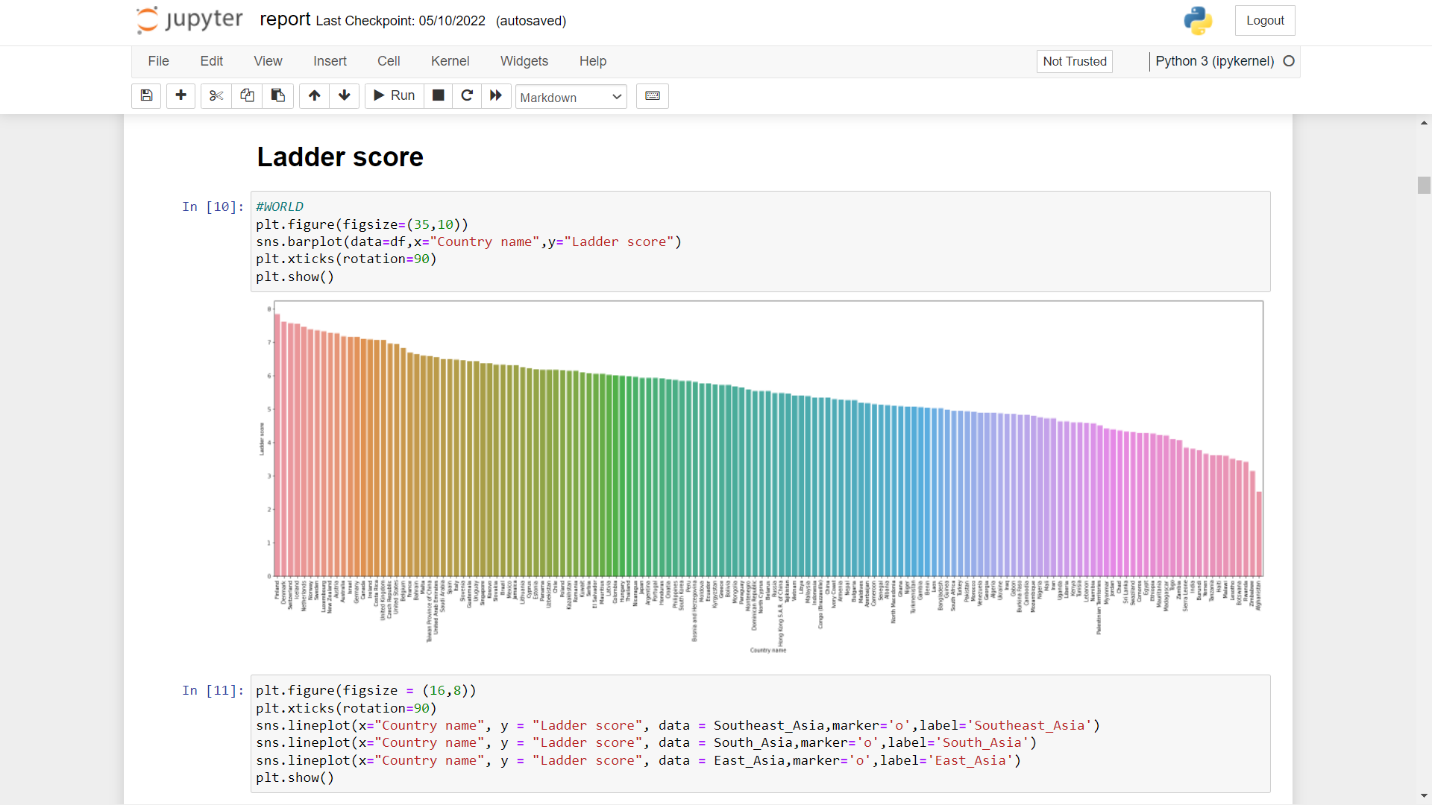


1. **Dividing countries into sections –**

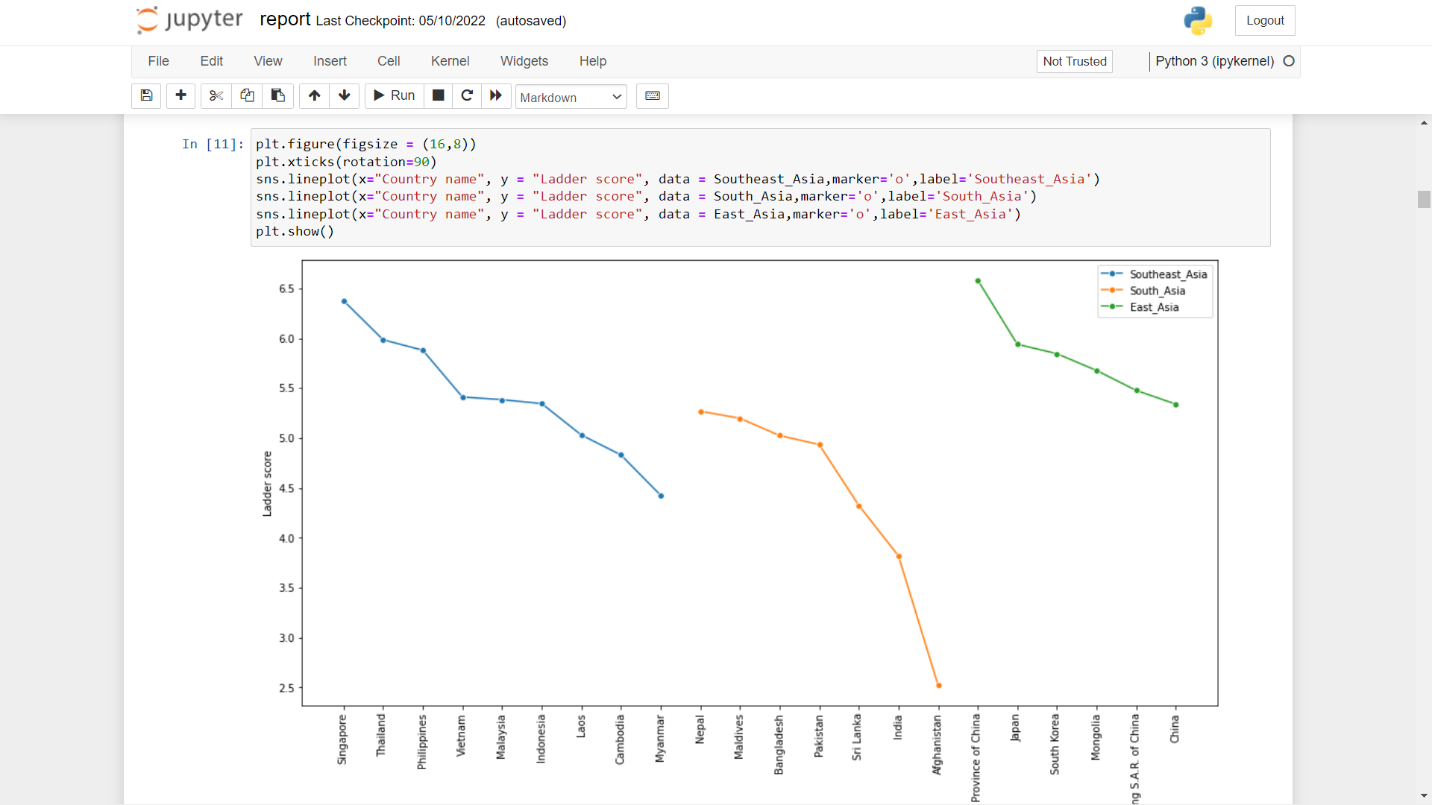


1. **Ladder Score –**

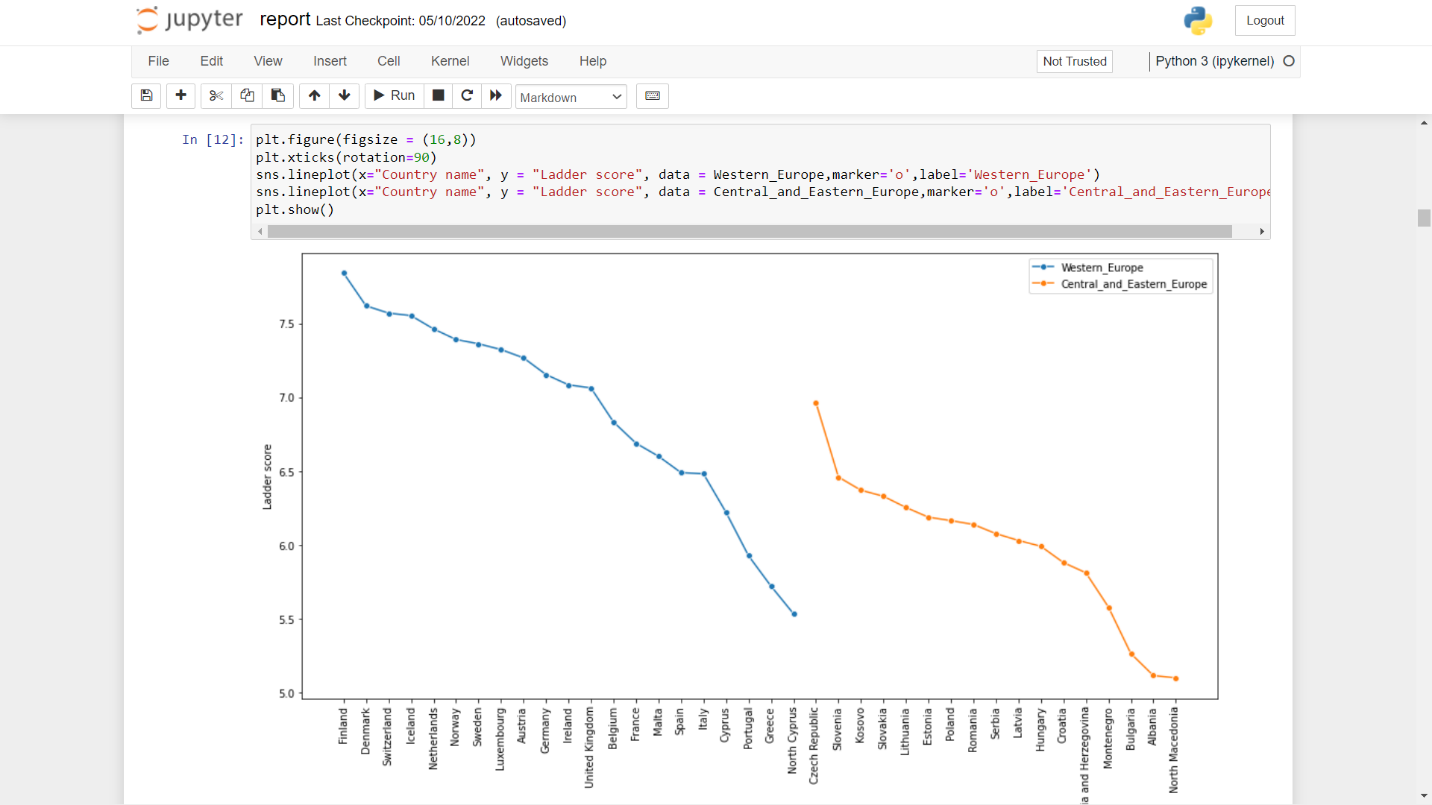
**3.1) All Countries –**



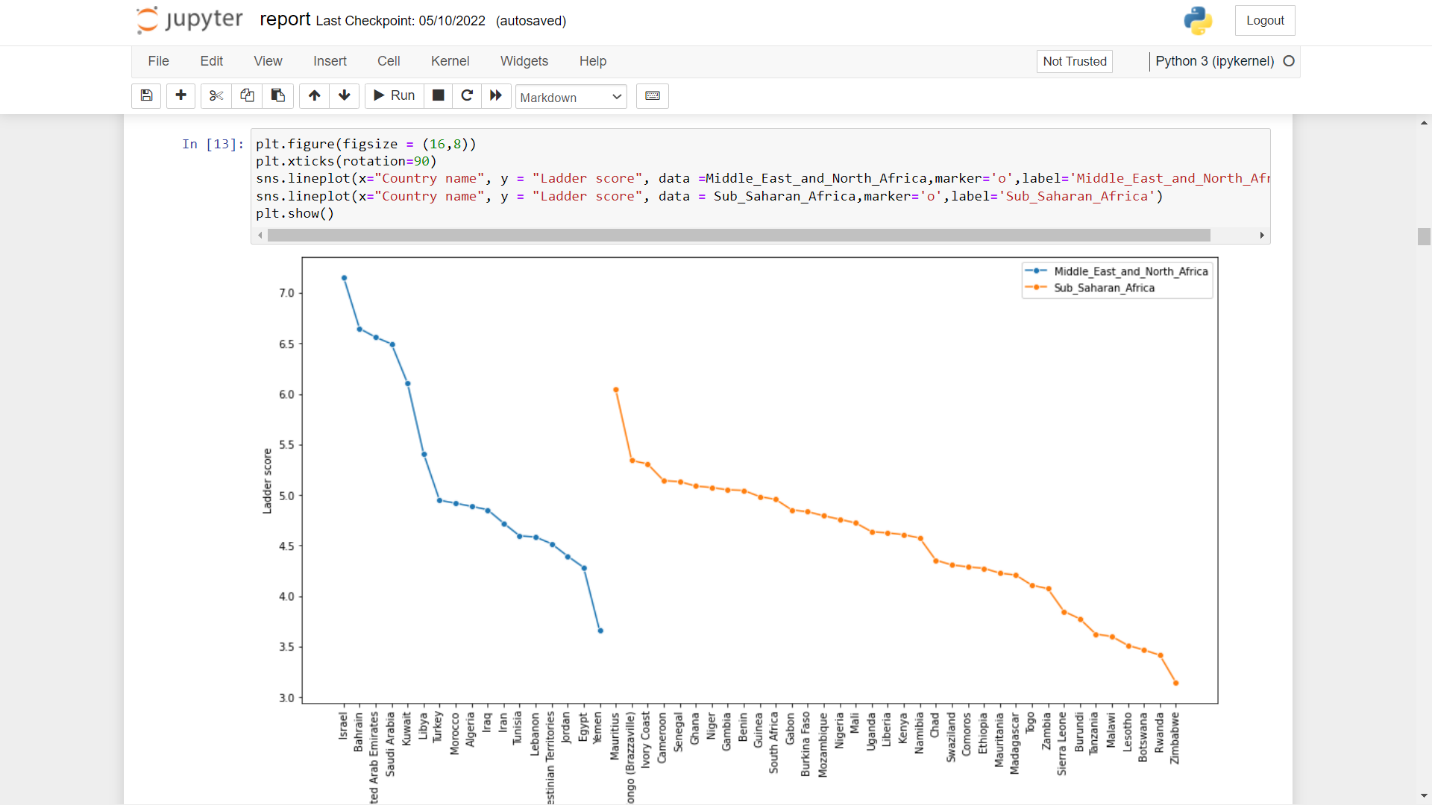
**3.2) Asian Countries –**



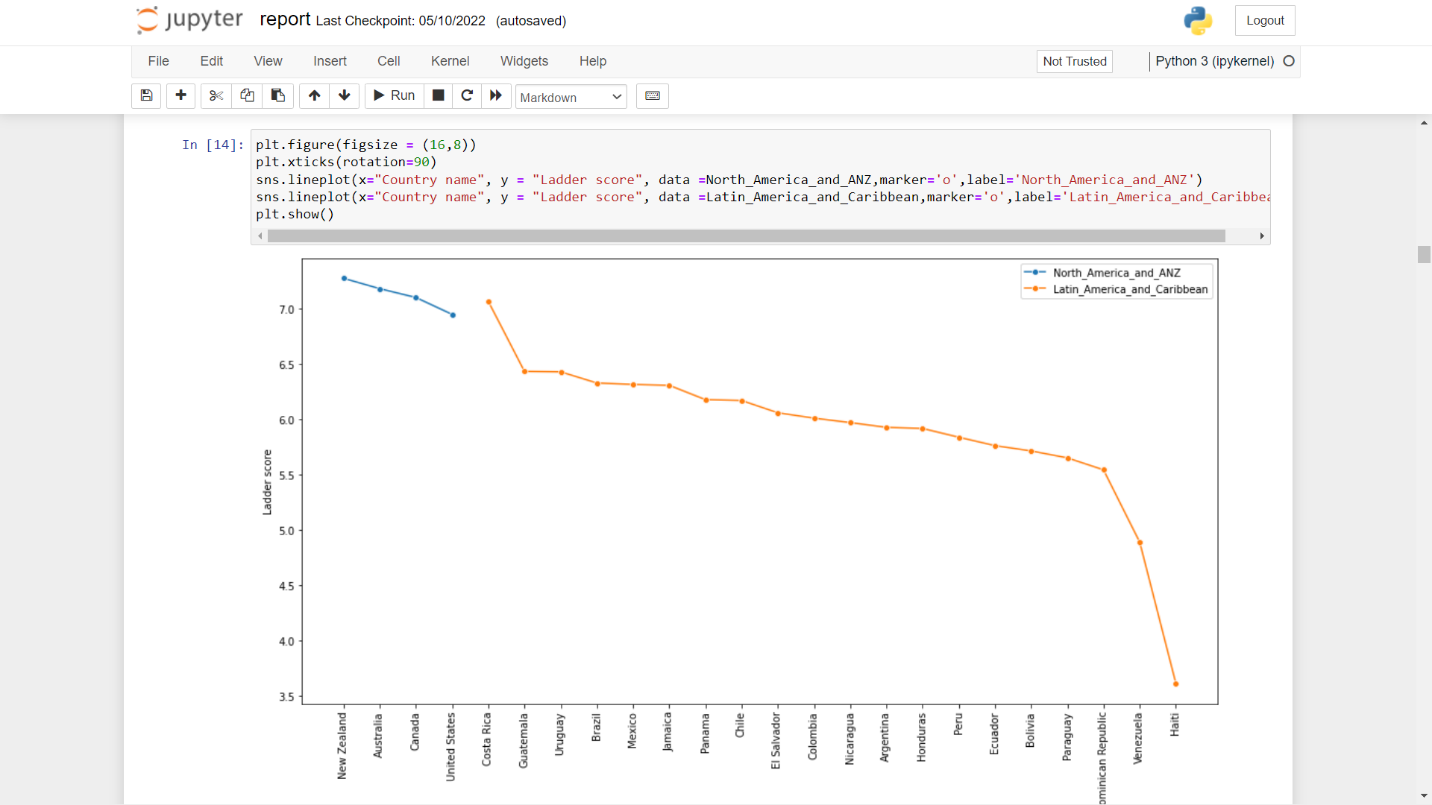
**3.3) European Countries –**



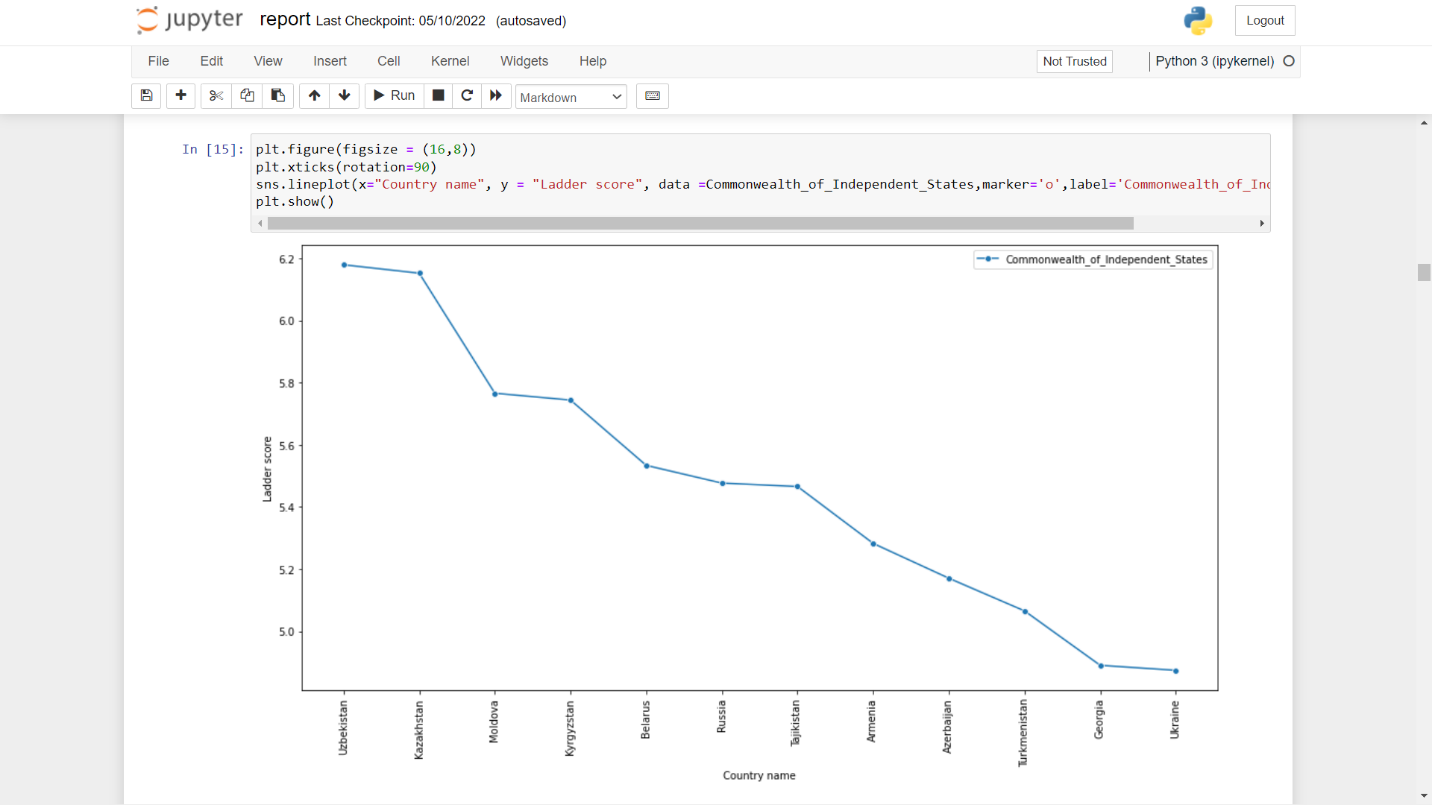
**3.4) Middle East & African Countries –**



**3.5) American Countries –**

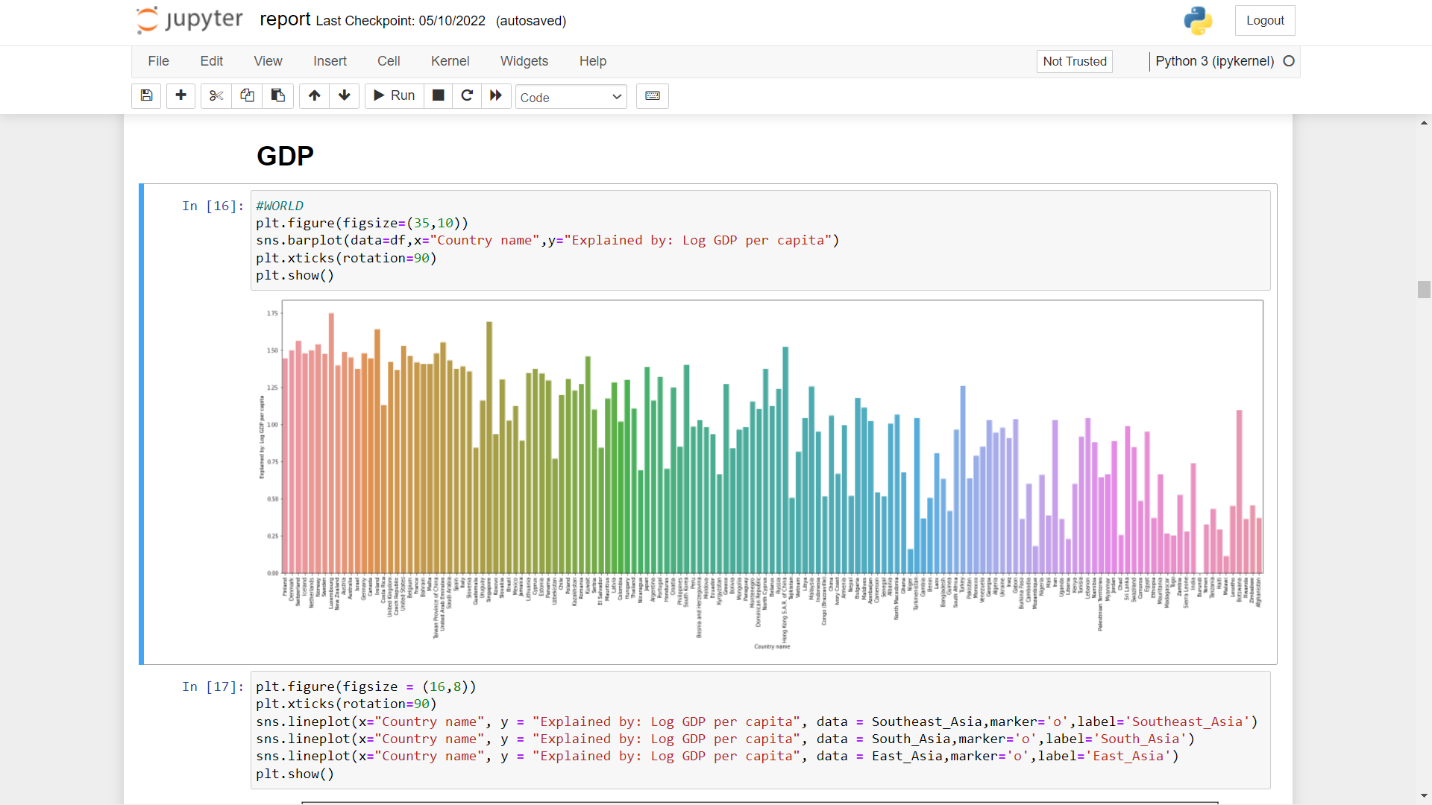


**3.6) Independent Countries –**

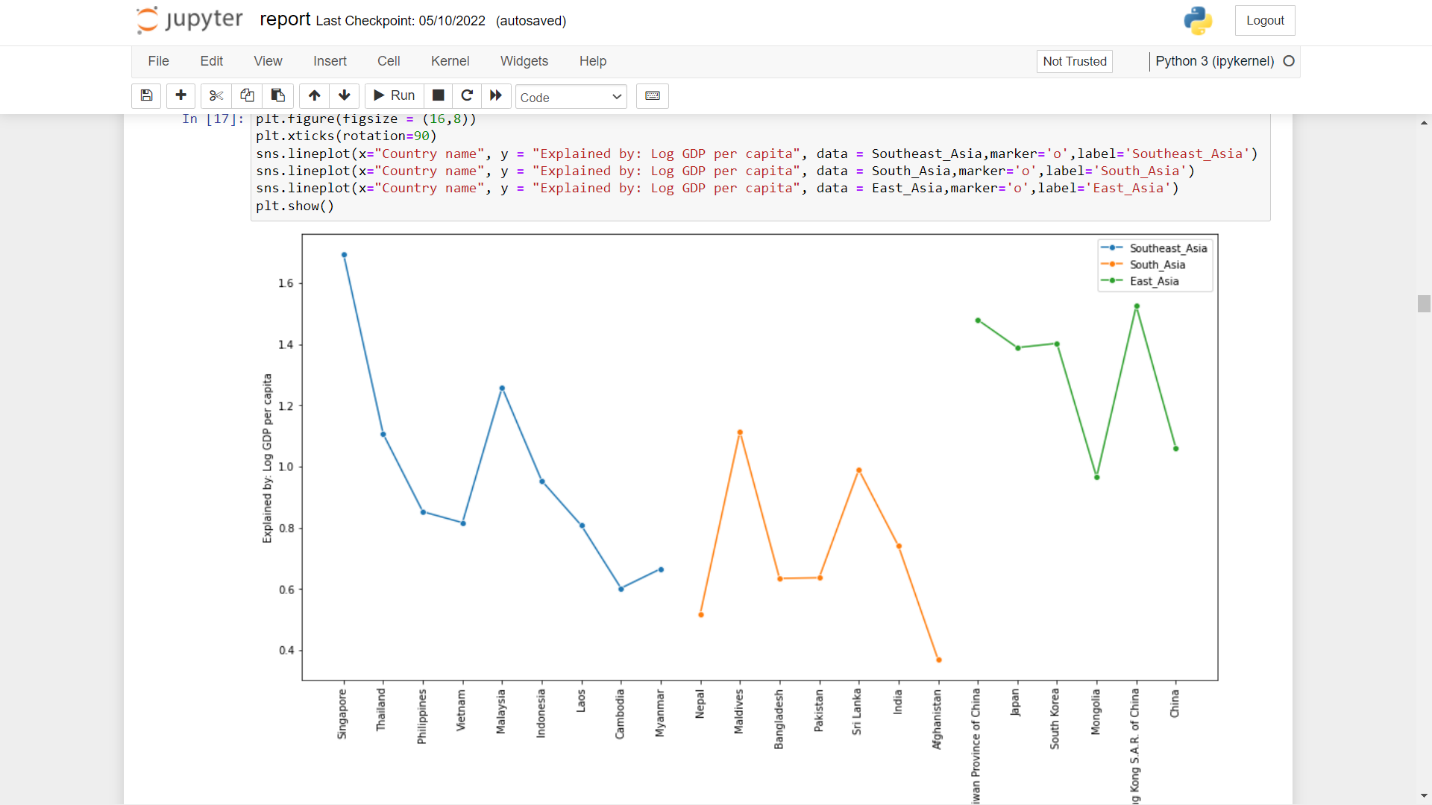


1. **GDP –**

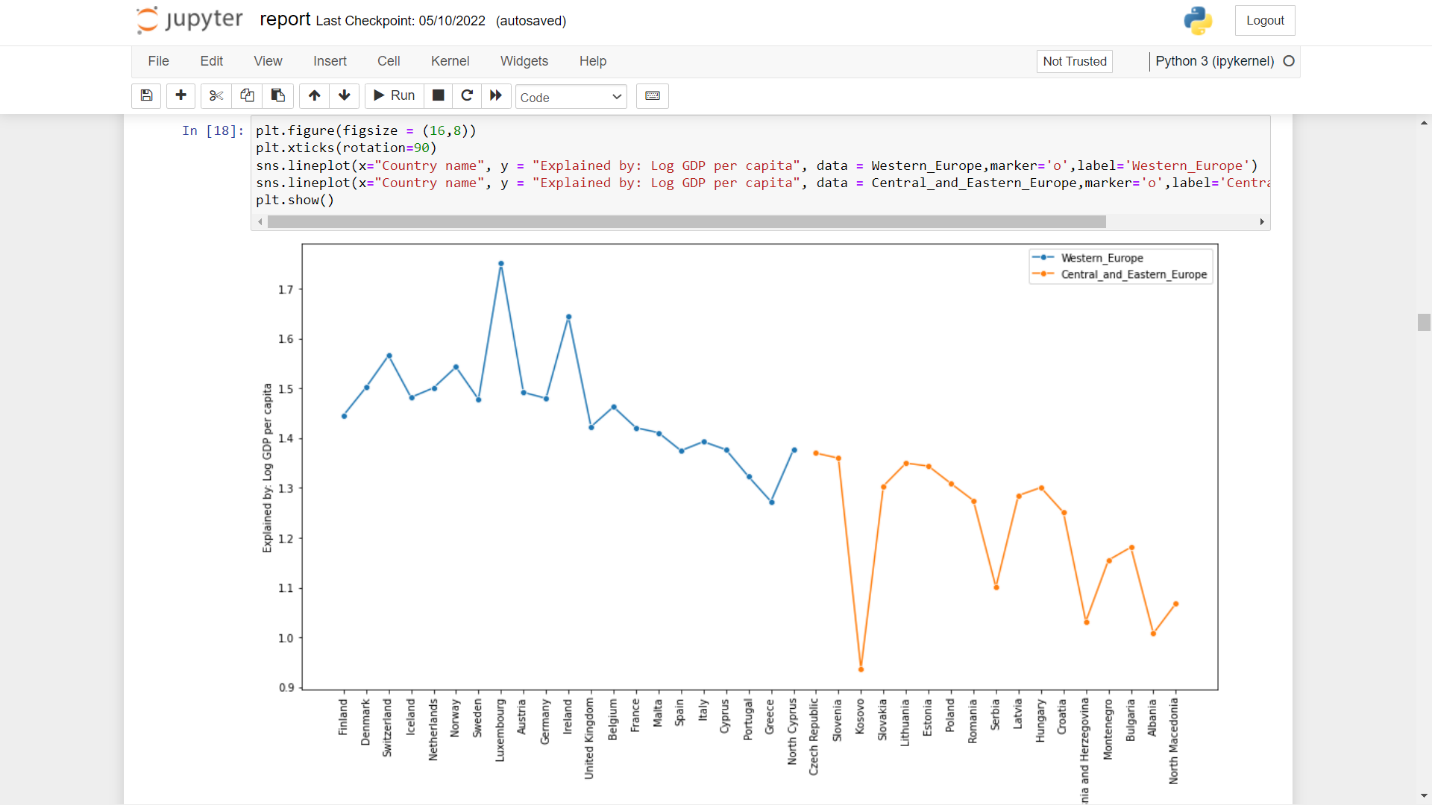
**4.1) All Countries –**



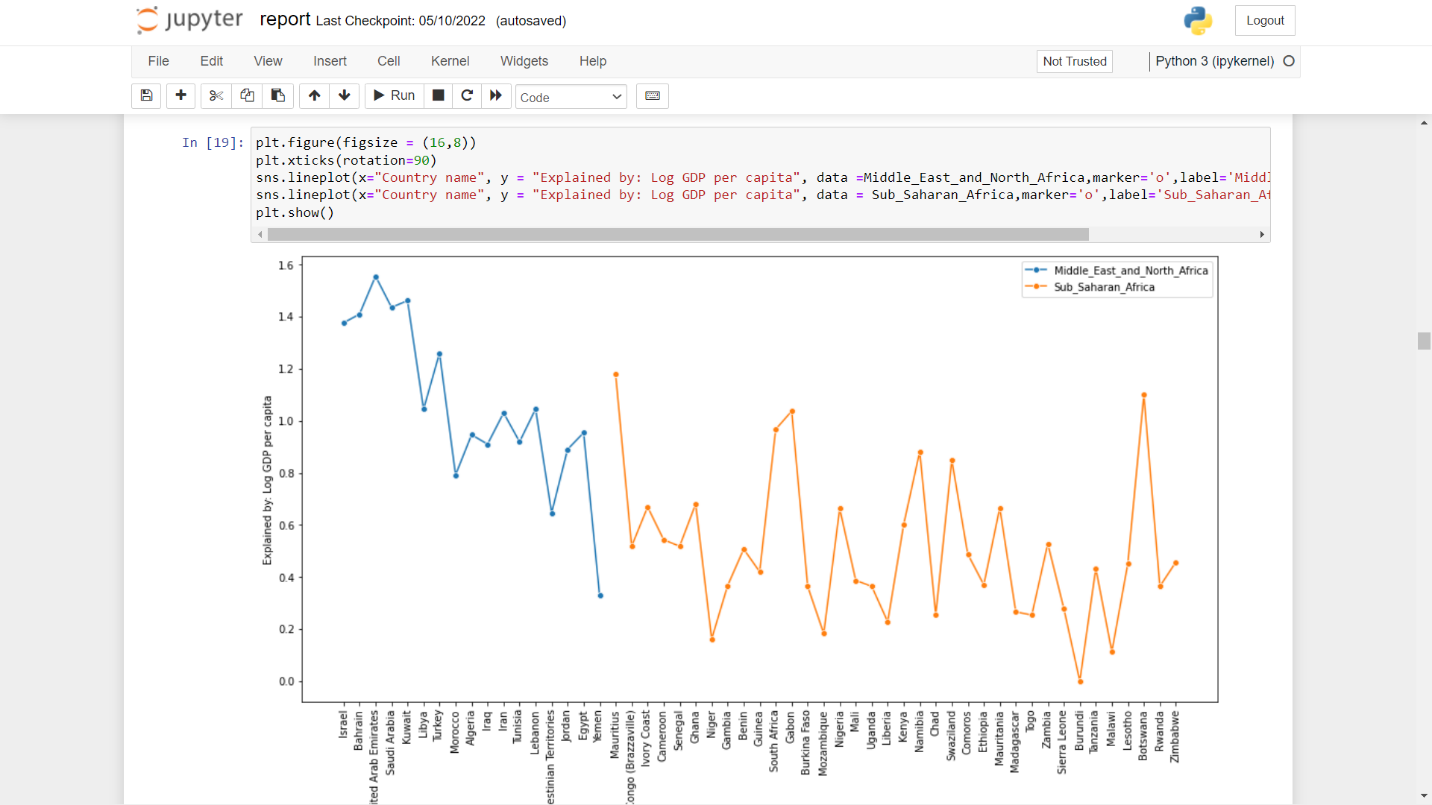
**4.2) Asian Countries –**



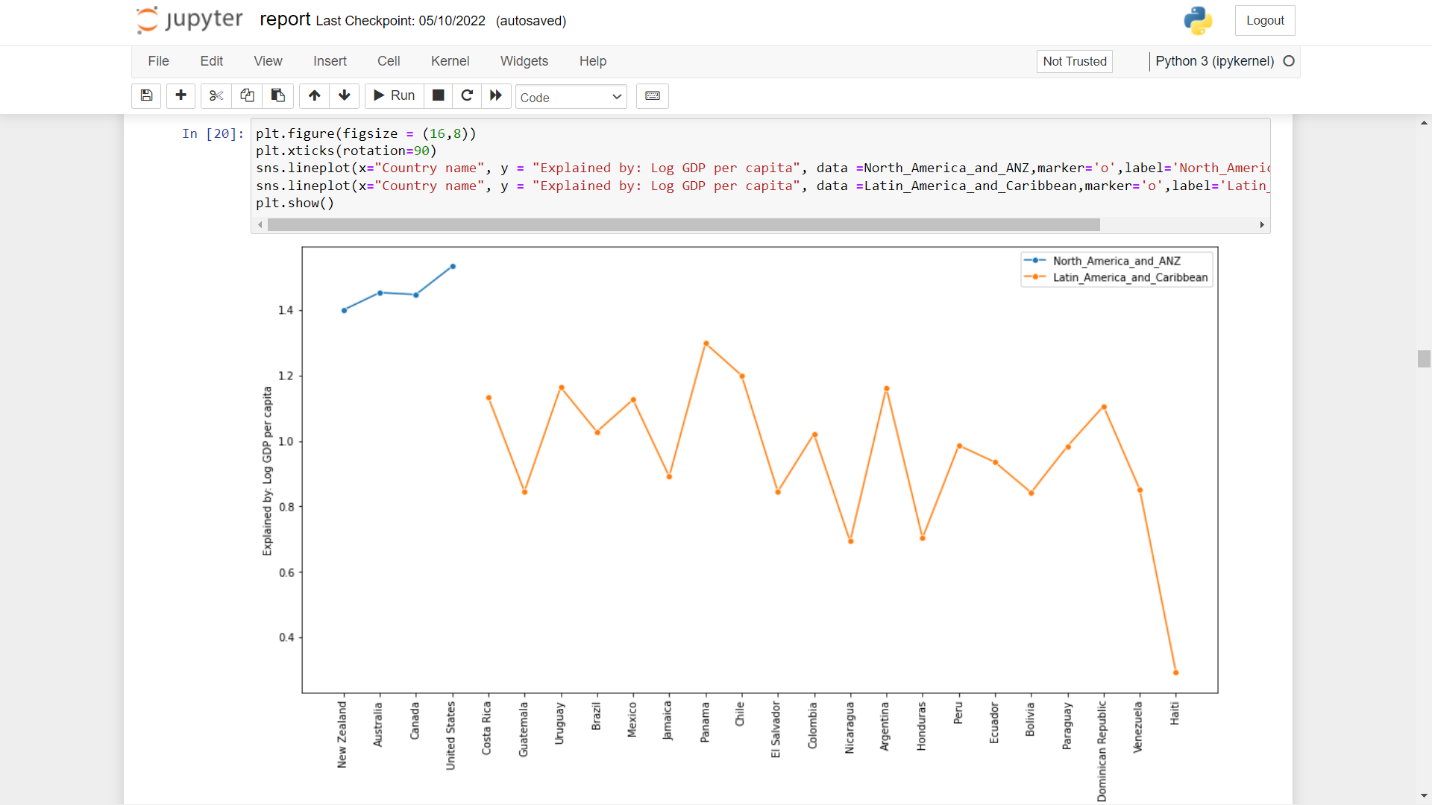
**4.3) European Countries –**



**4.4) Middle East & African Countries –**



**4.5) American Countries –**

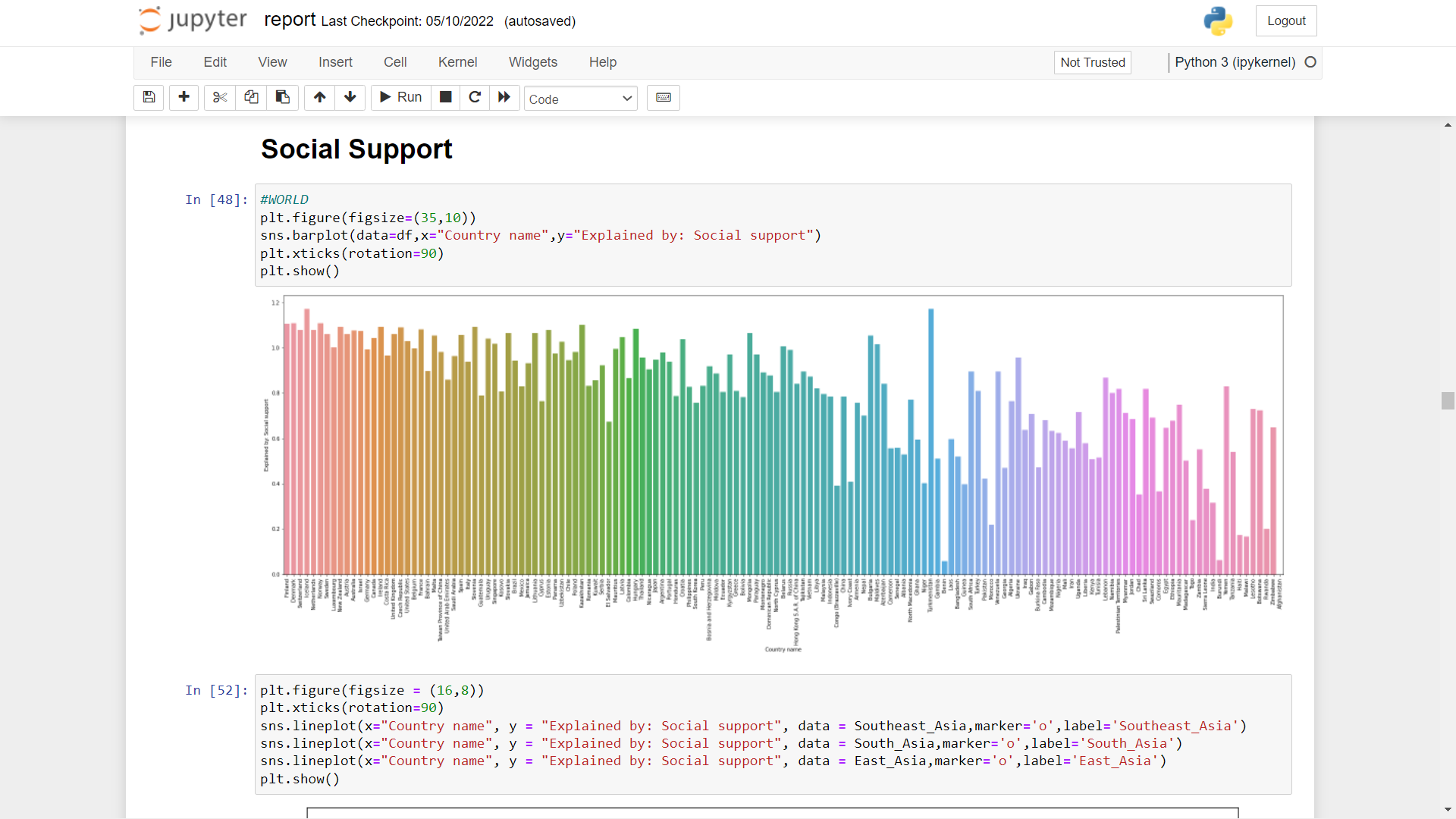


**4.6) Independent Countries –**

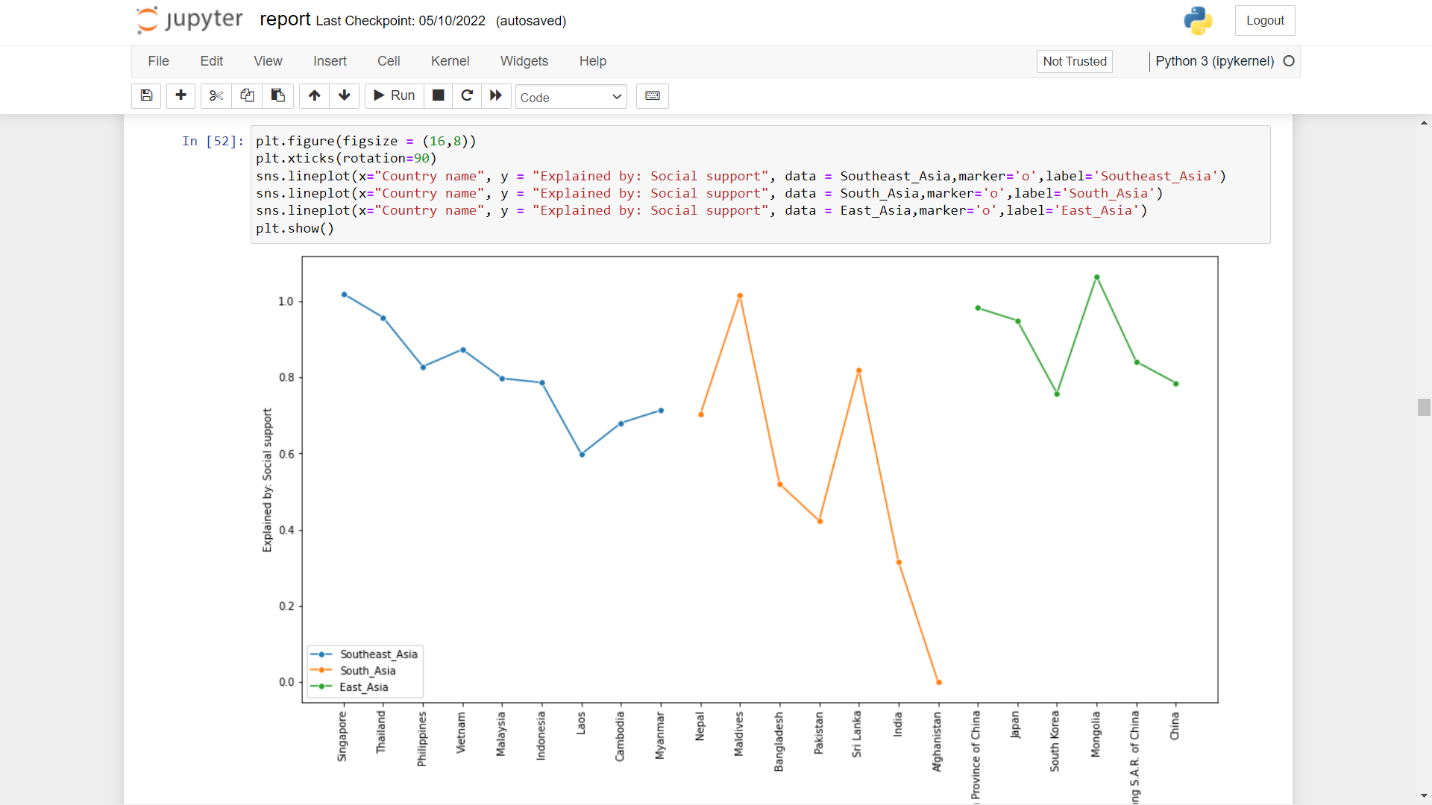


1. **Social Support –**

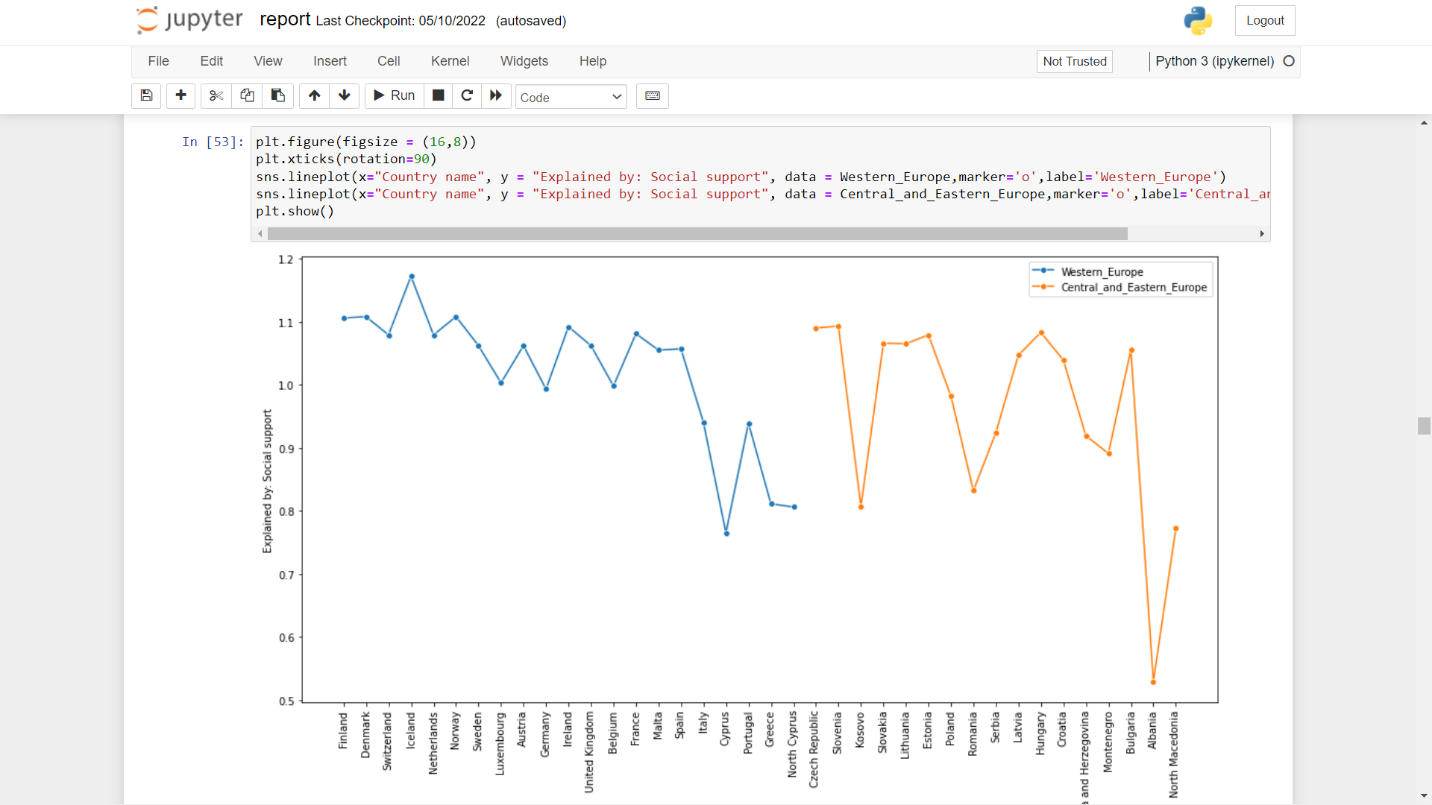
**5.1) All Countries –**



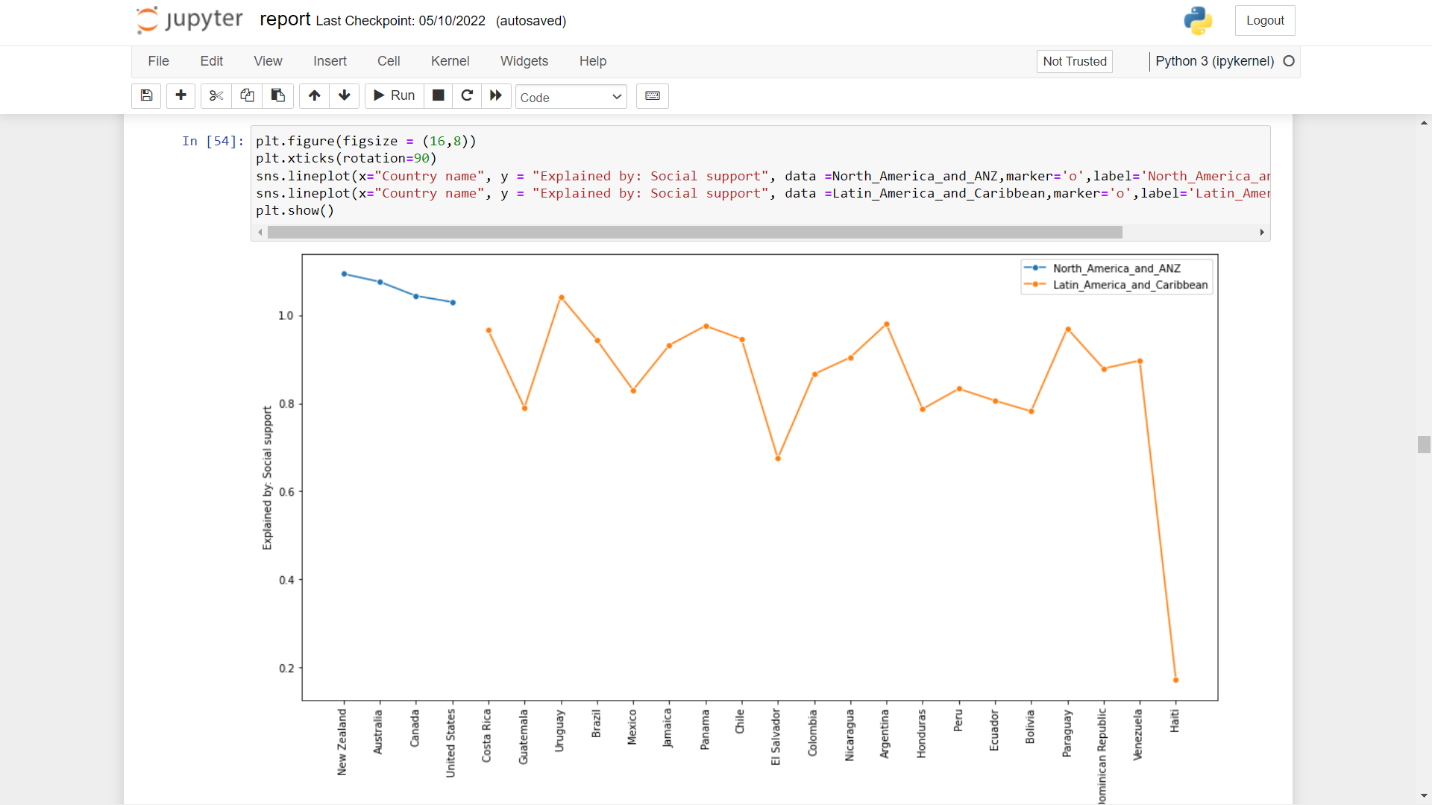
**5.2) Asian Countries –**



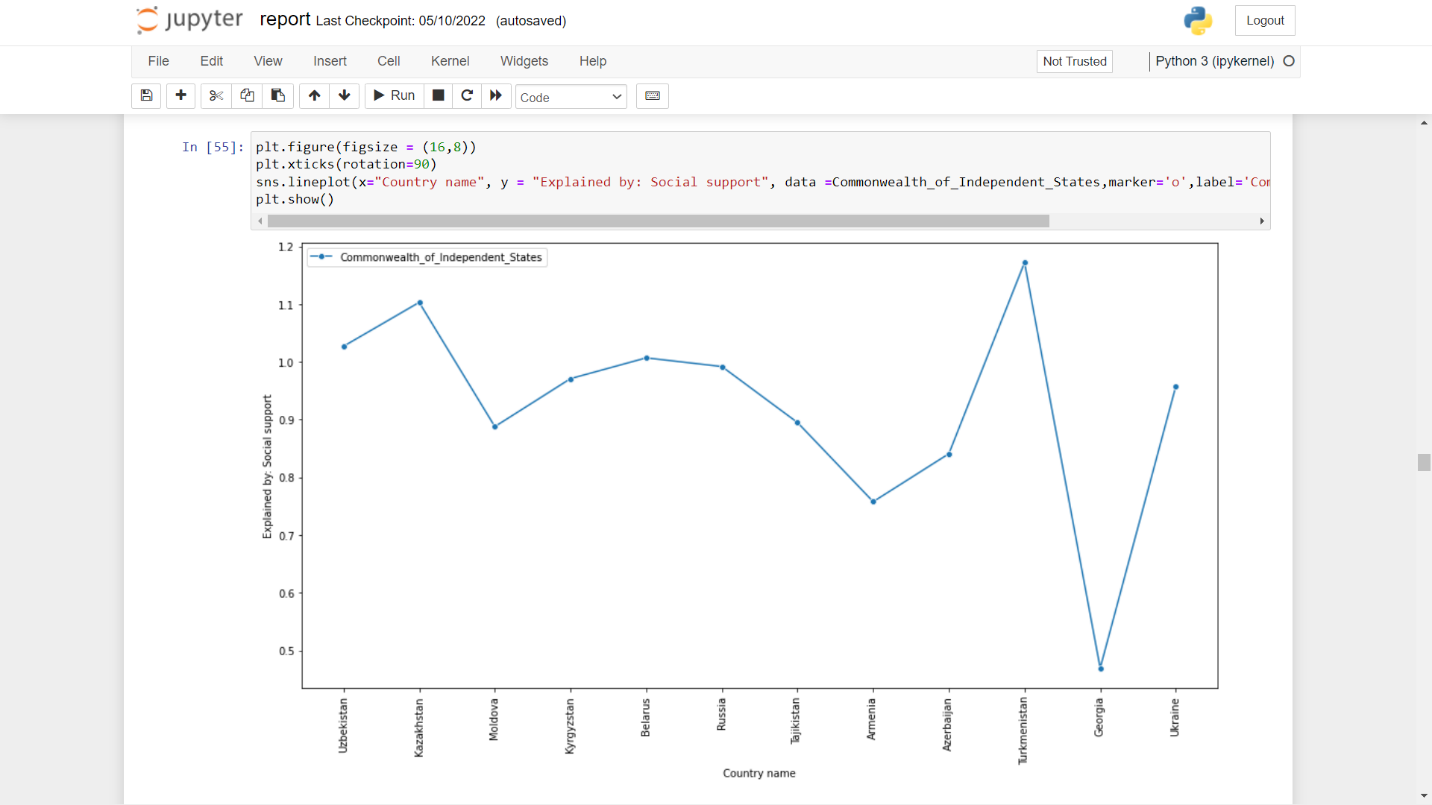
**5.3) European Countries –**



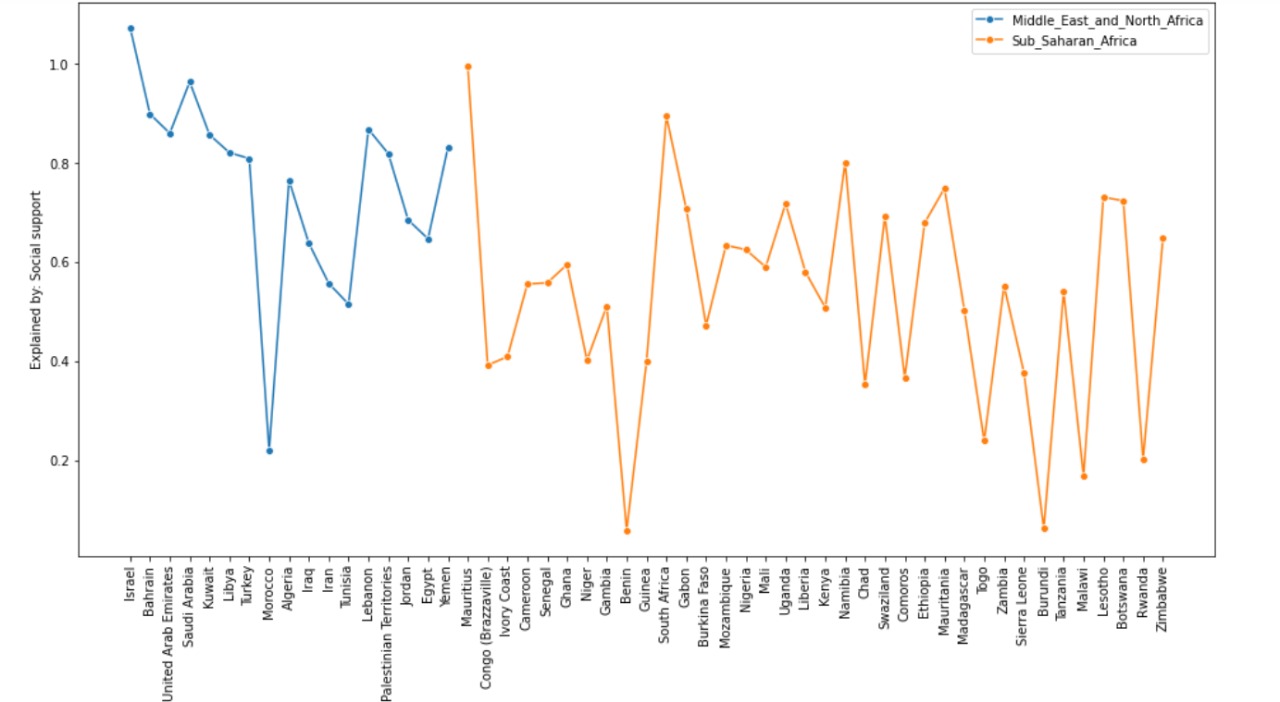
**5.4) American Countries –**



**5.5) Independent Countries –**



**5.6) Middle East & African Countries –**

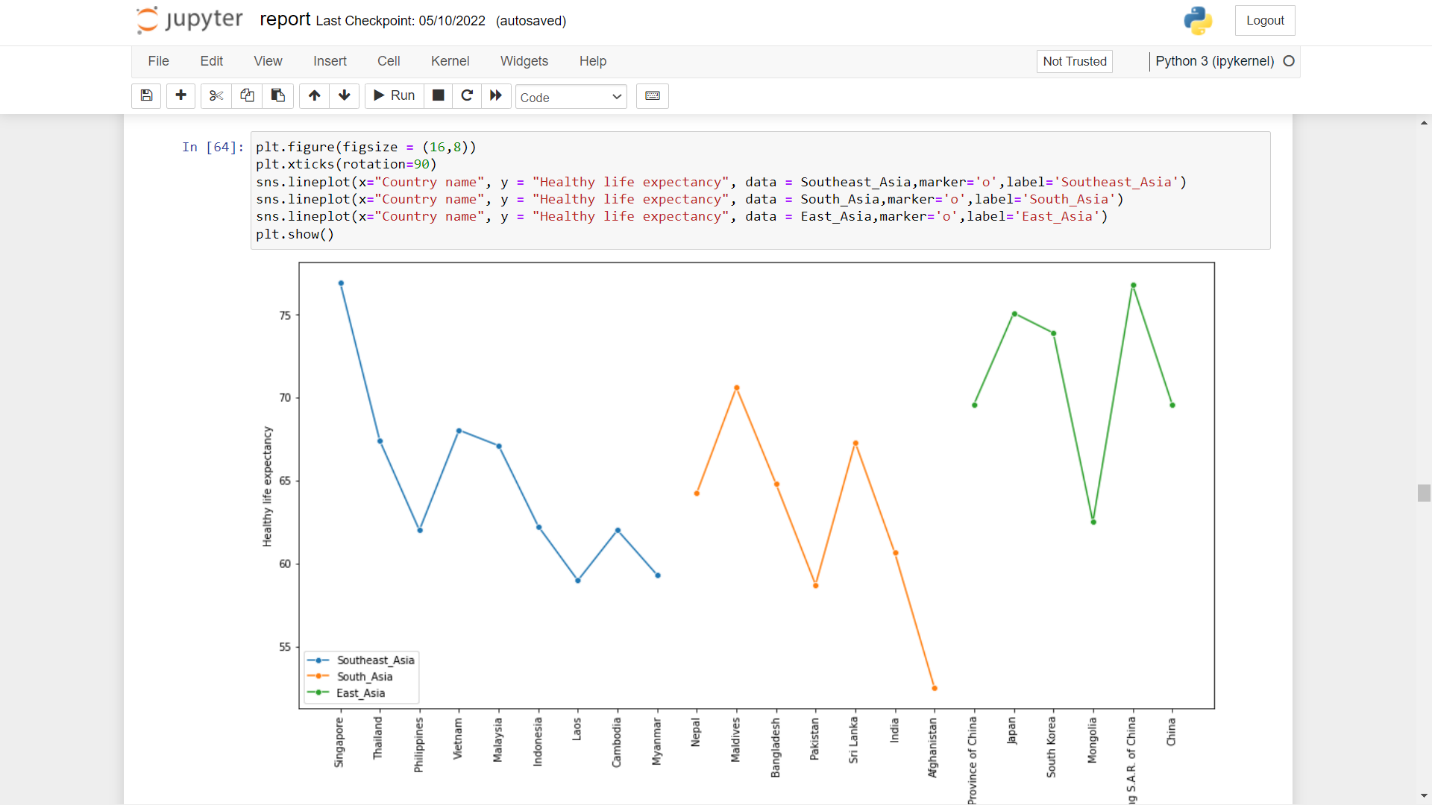
****

1. **Healthy Life Expectancy –**

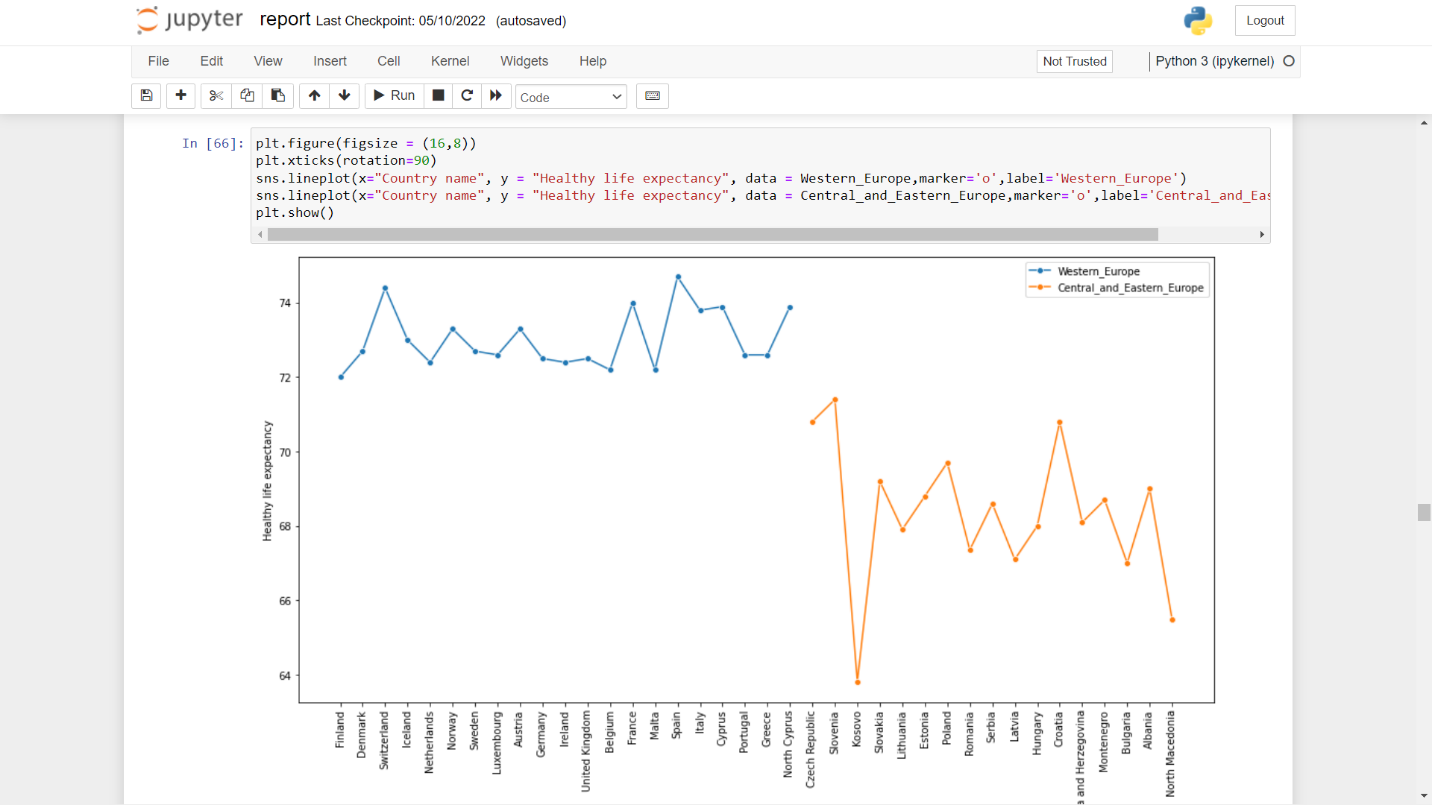
**6.1) All Countries –**



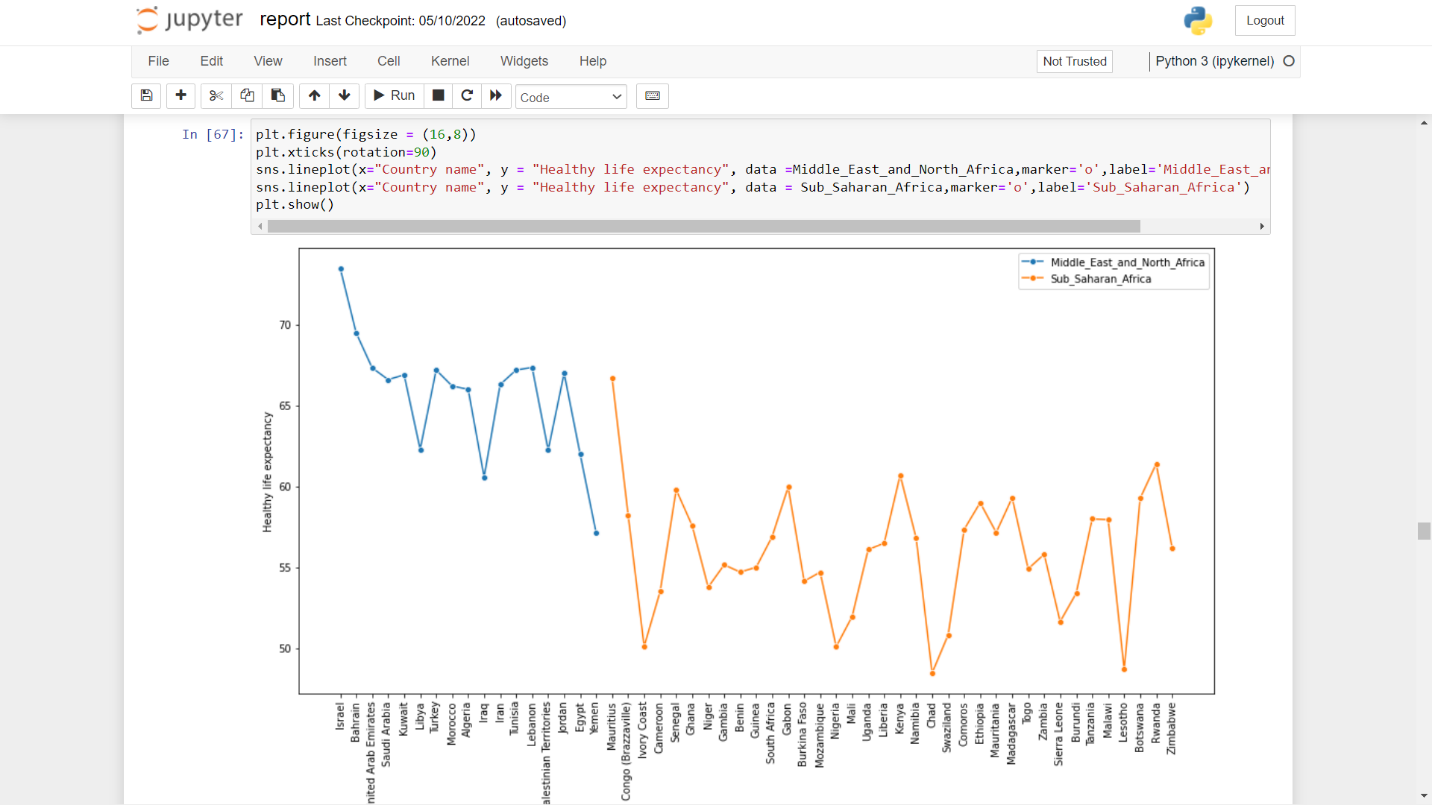
**6.2) Asian Countries –**



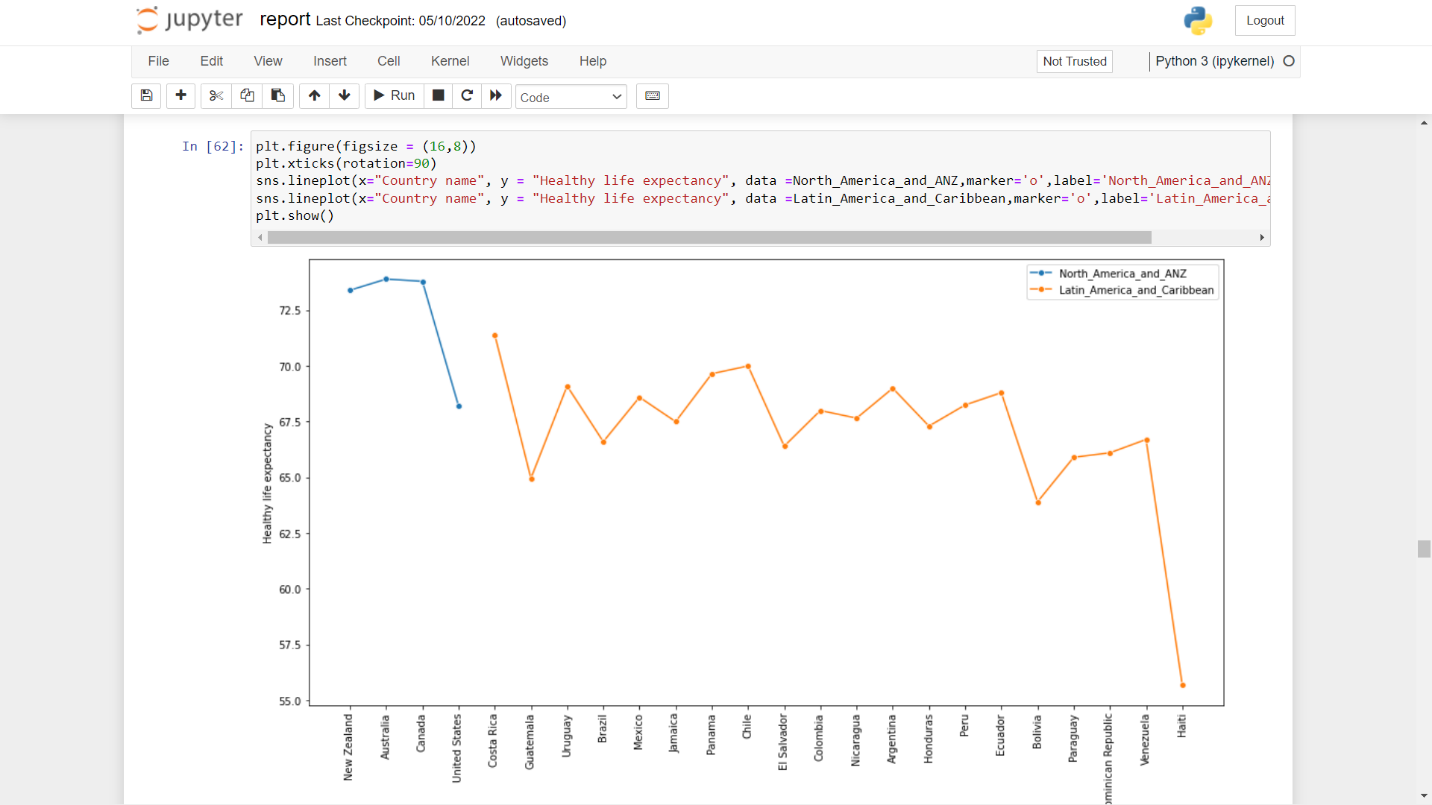
**6.3)European Countries –**



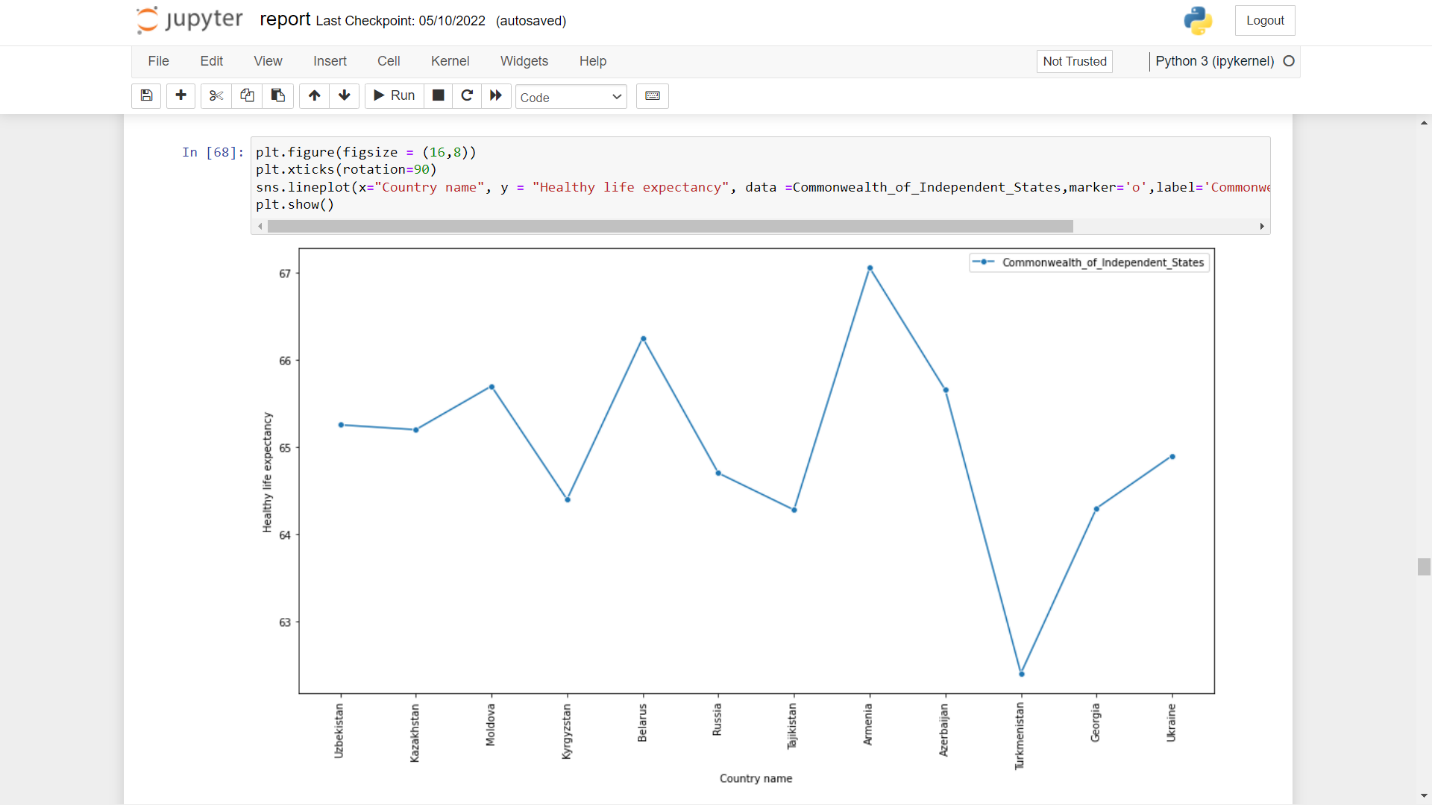
**6.4) Middle East & African Countries –**



**6.5) American Countries –**

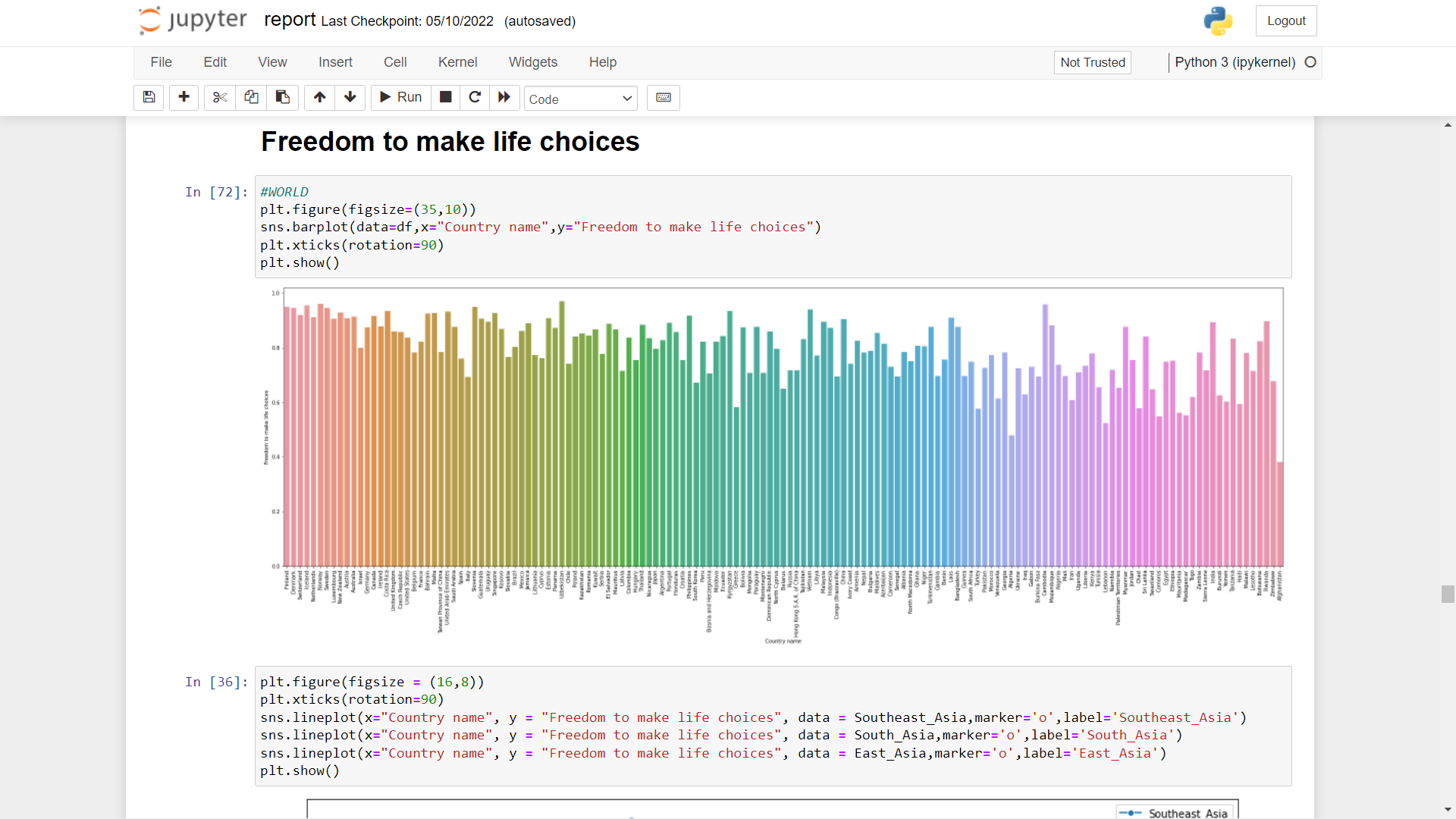


**6.6) Independent Countries –**

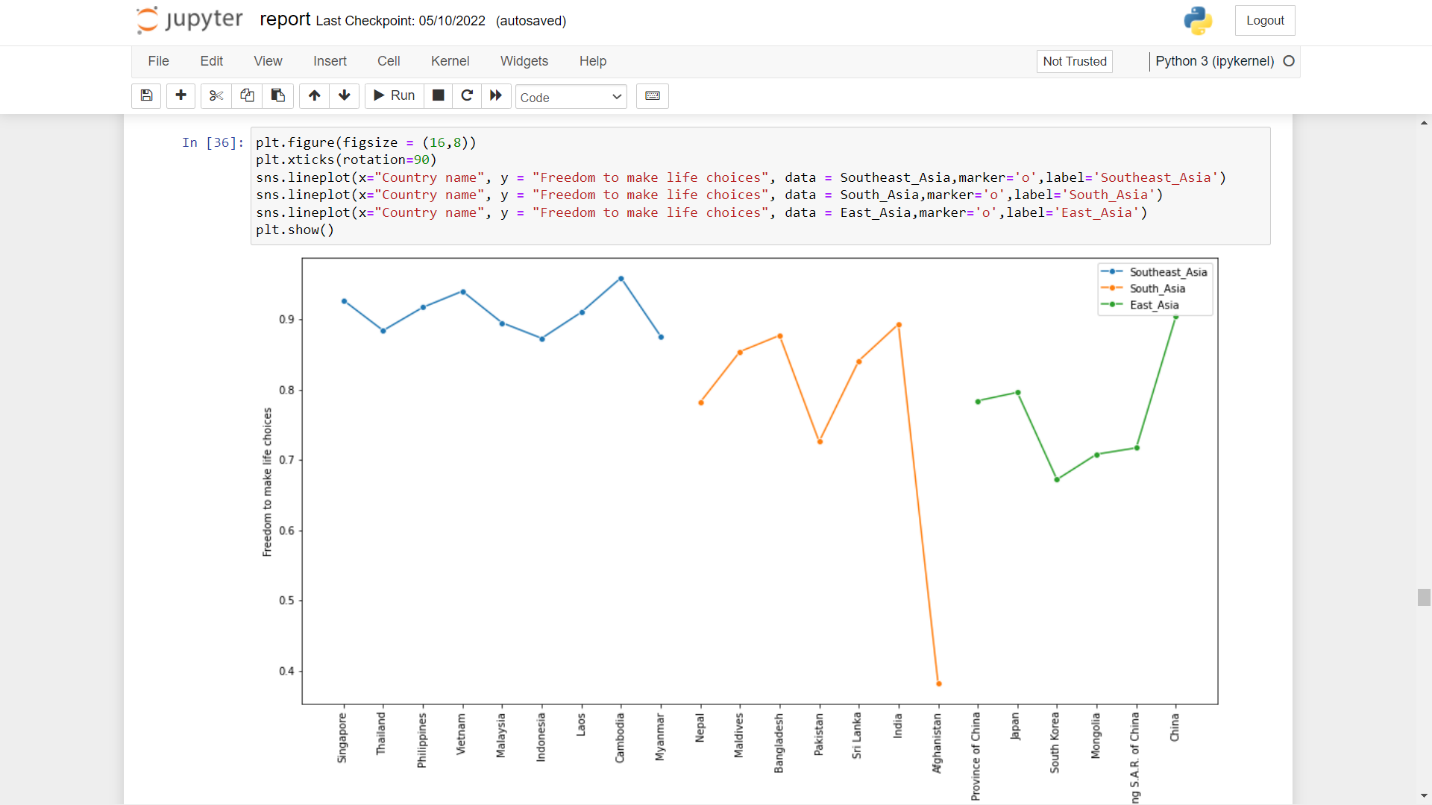


1. **Freedom to make life choices –**

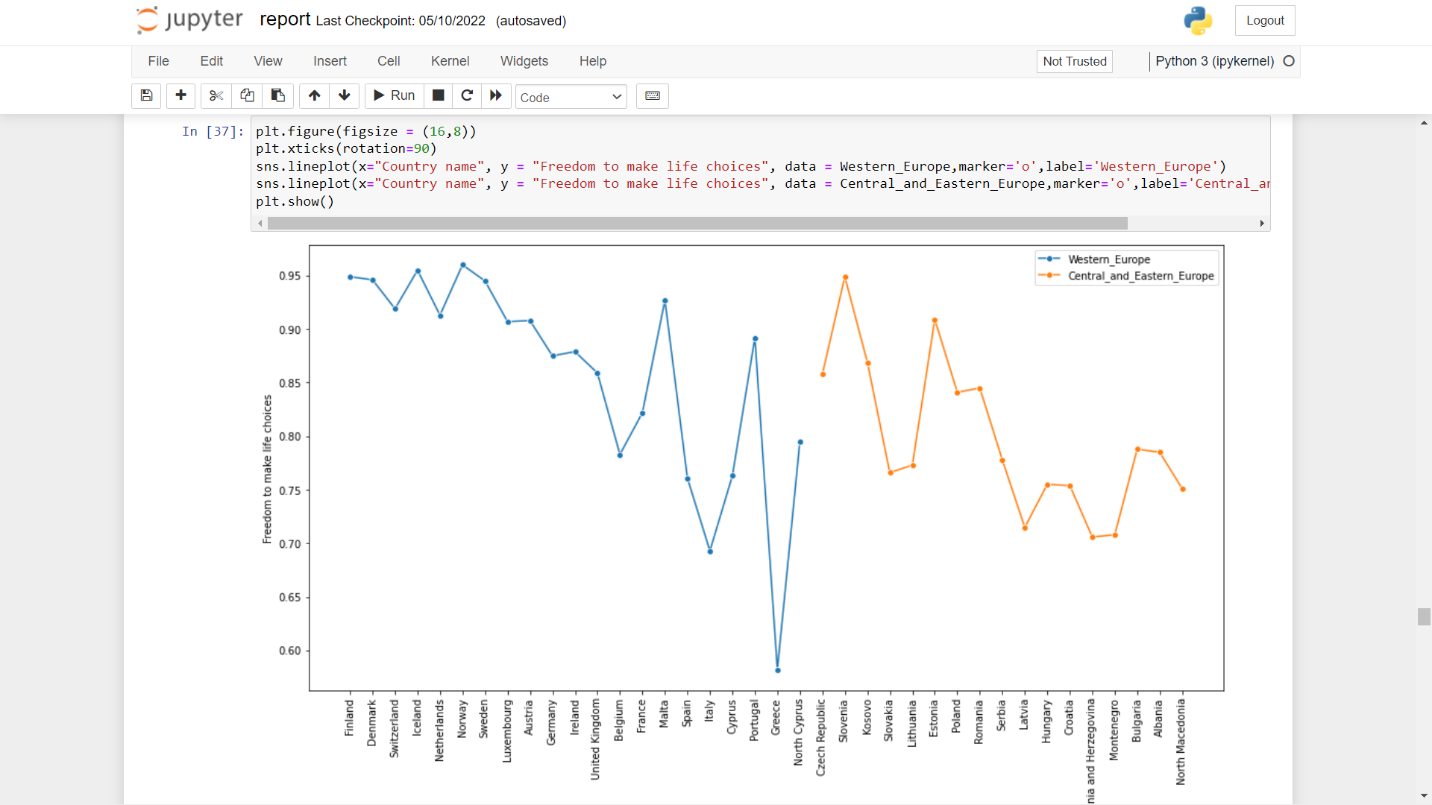
**7.1) All Countries –**



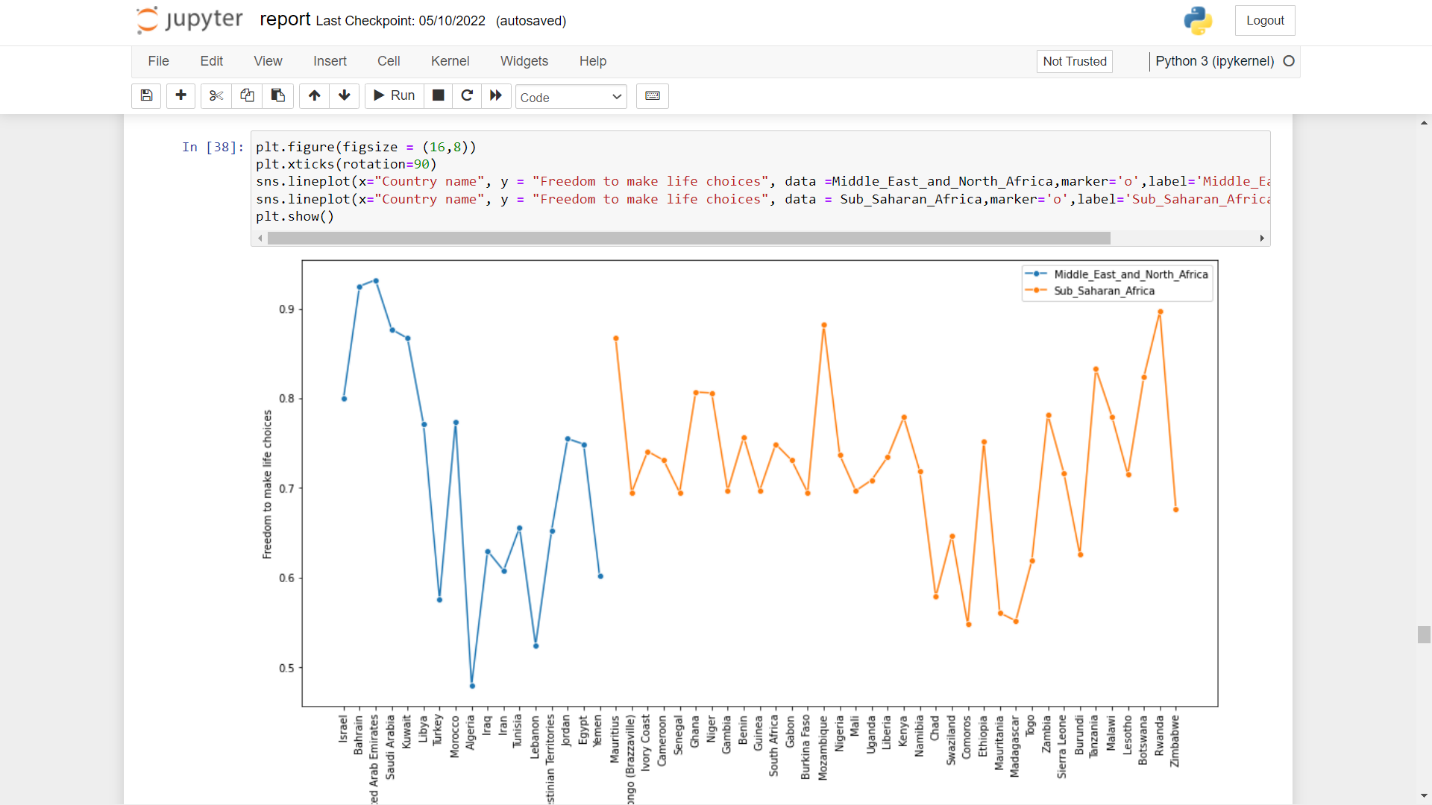
**7.2) Asian Countries –**



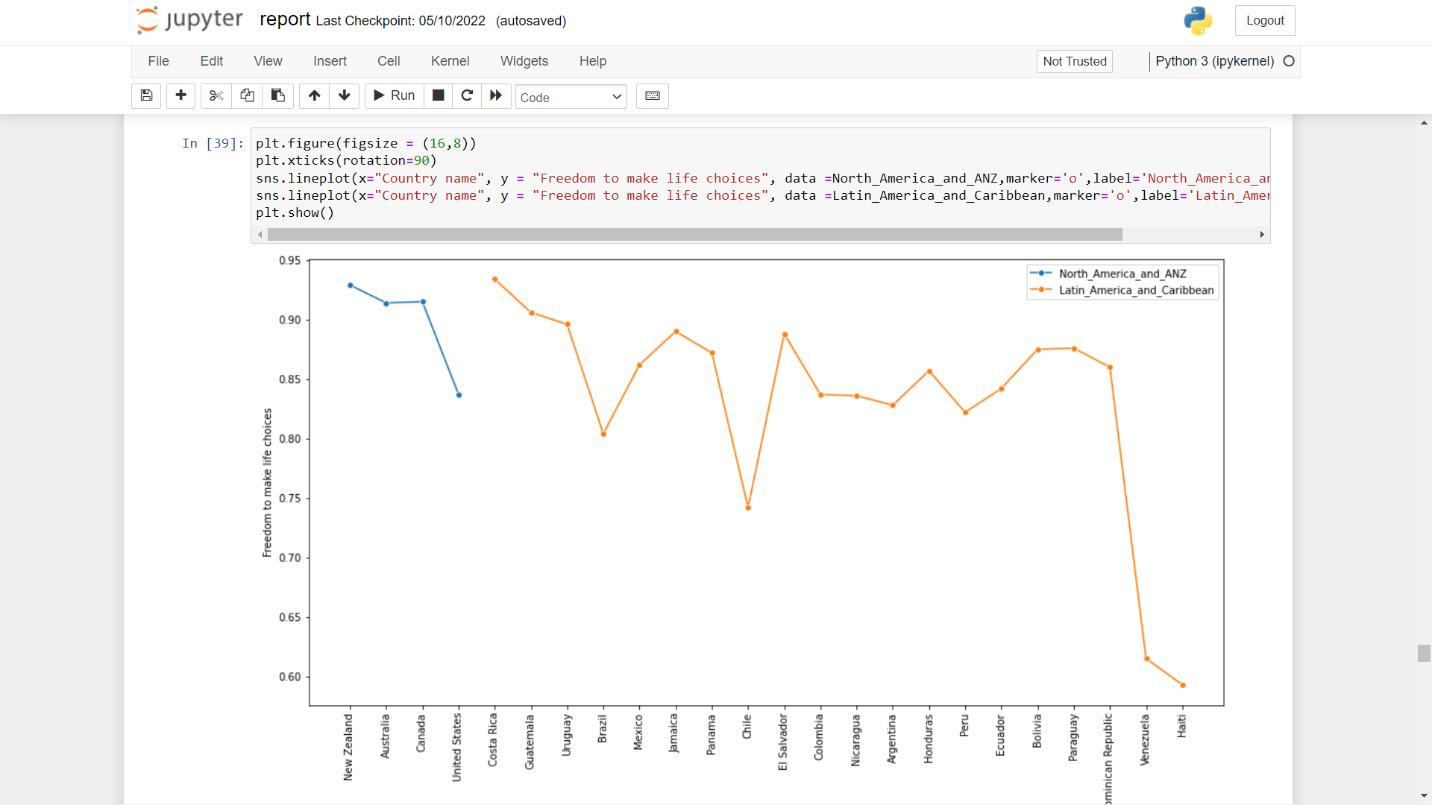
**7.3) European Countries –**



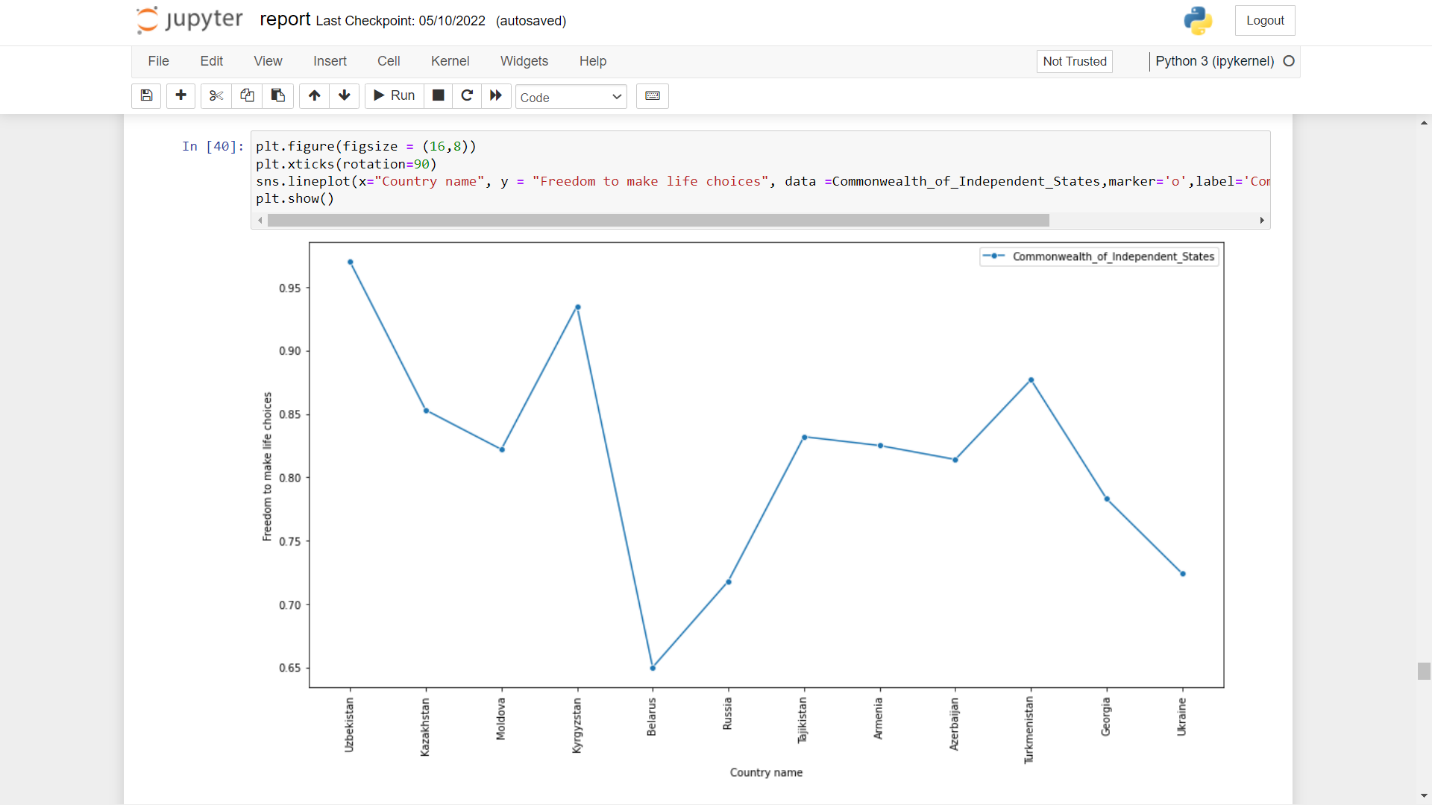
**7.4) Middle East & African Countries –**



**7.5) American Countries –**

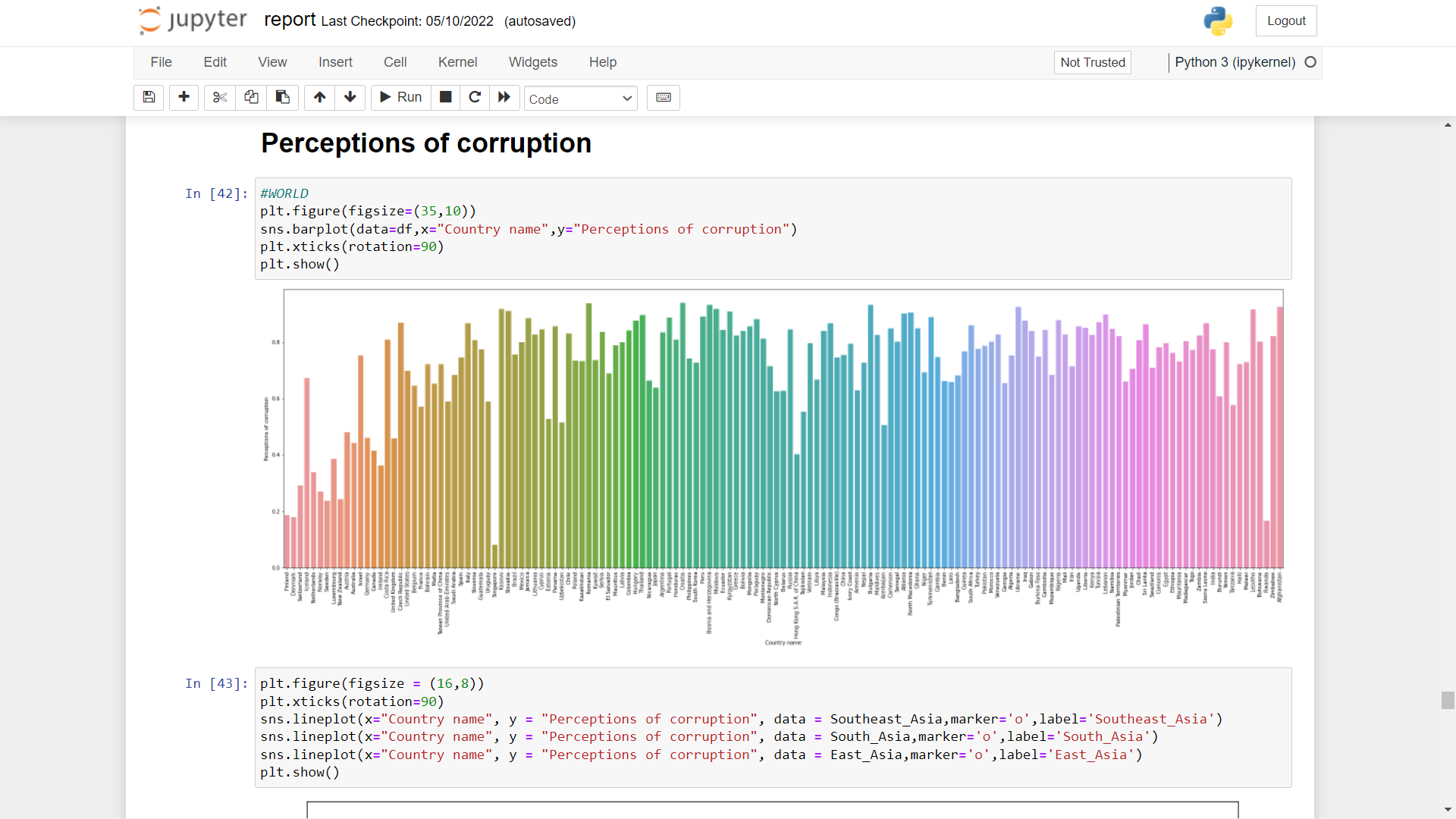


**7.6) Independent Countries –**

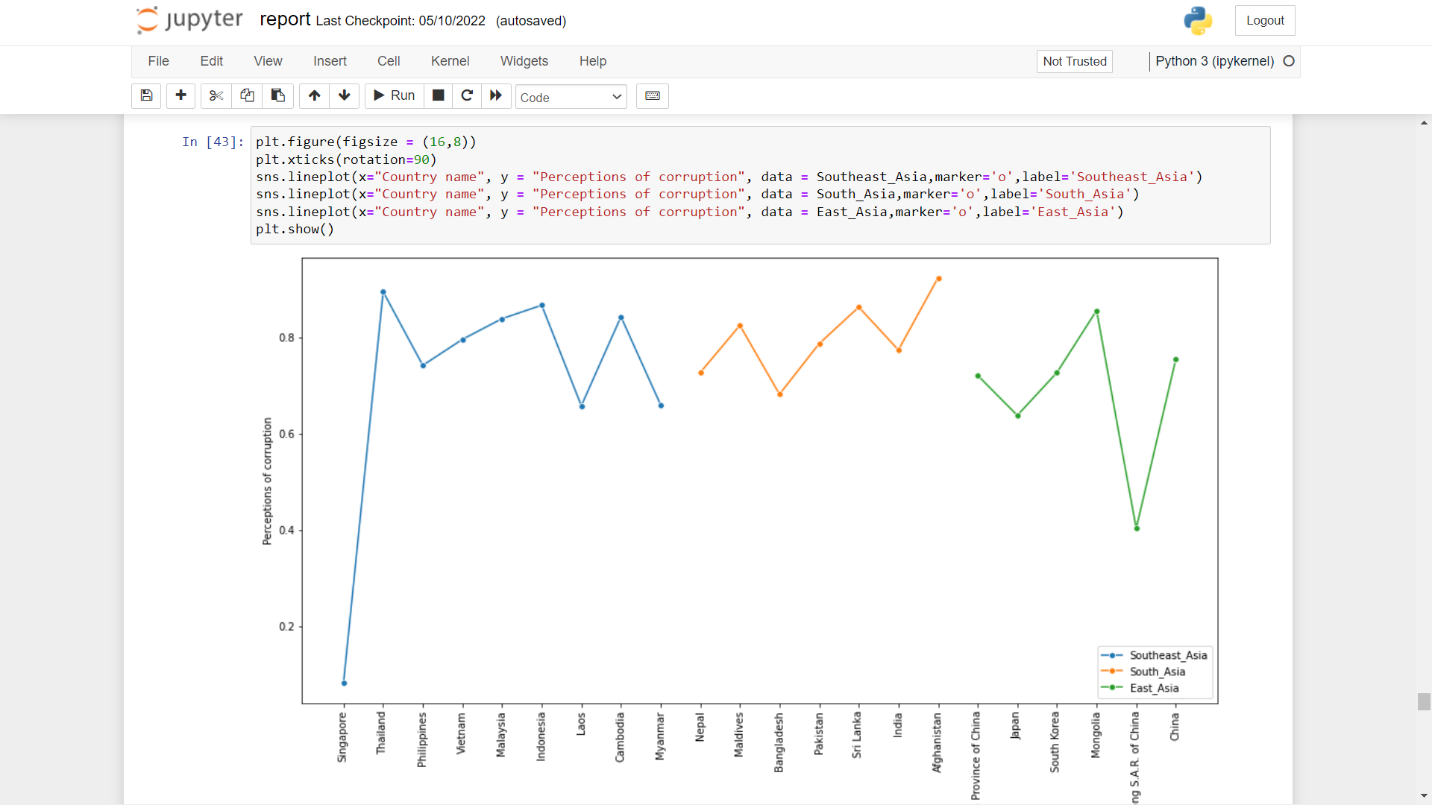


1. **Perception of Corruption –**

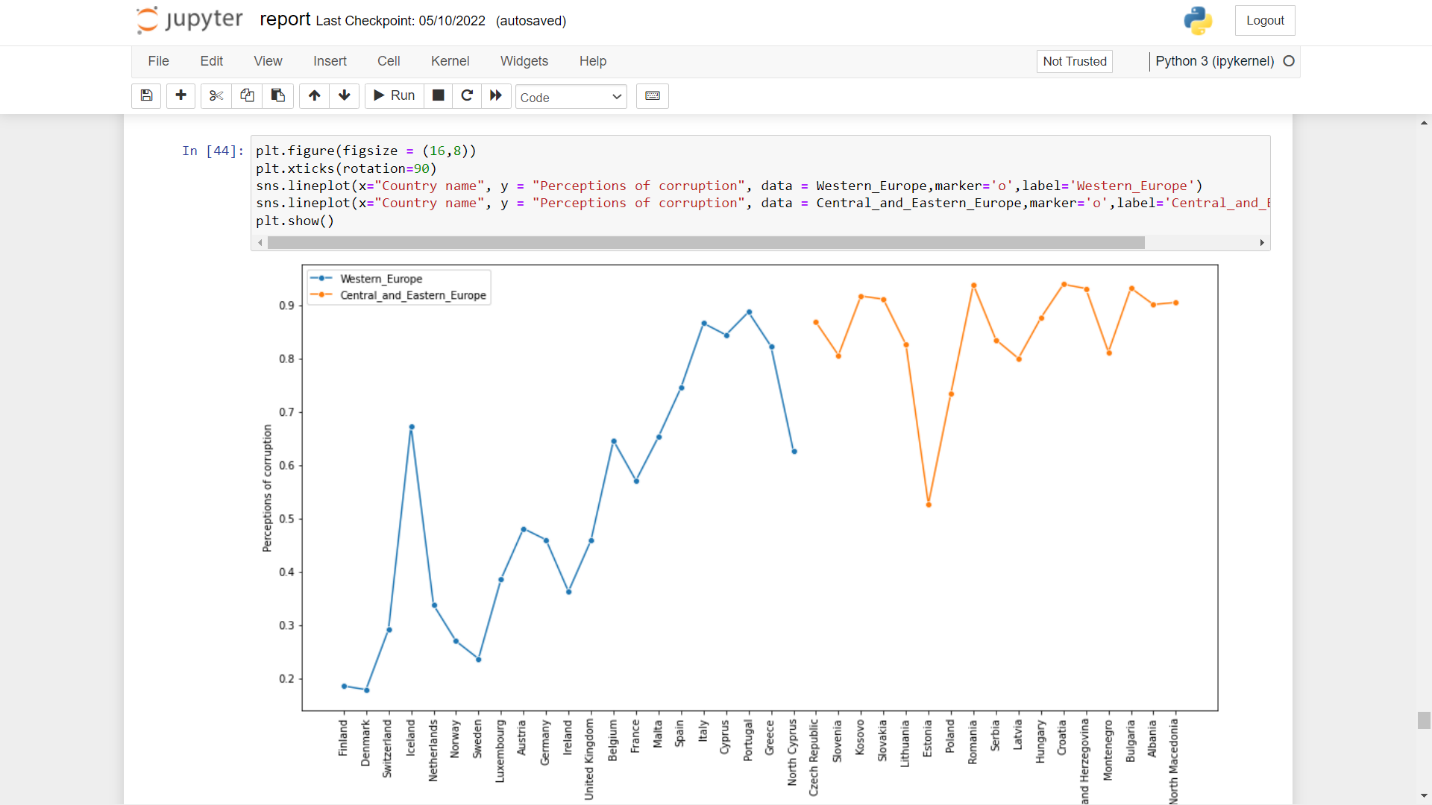
**8.1) All Countries –**



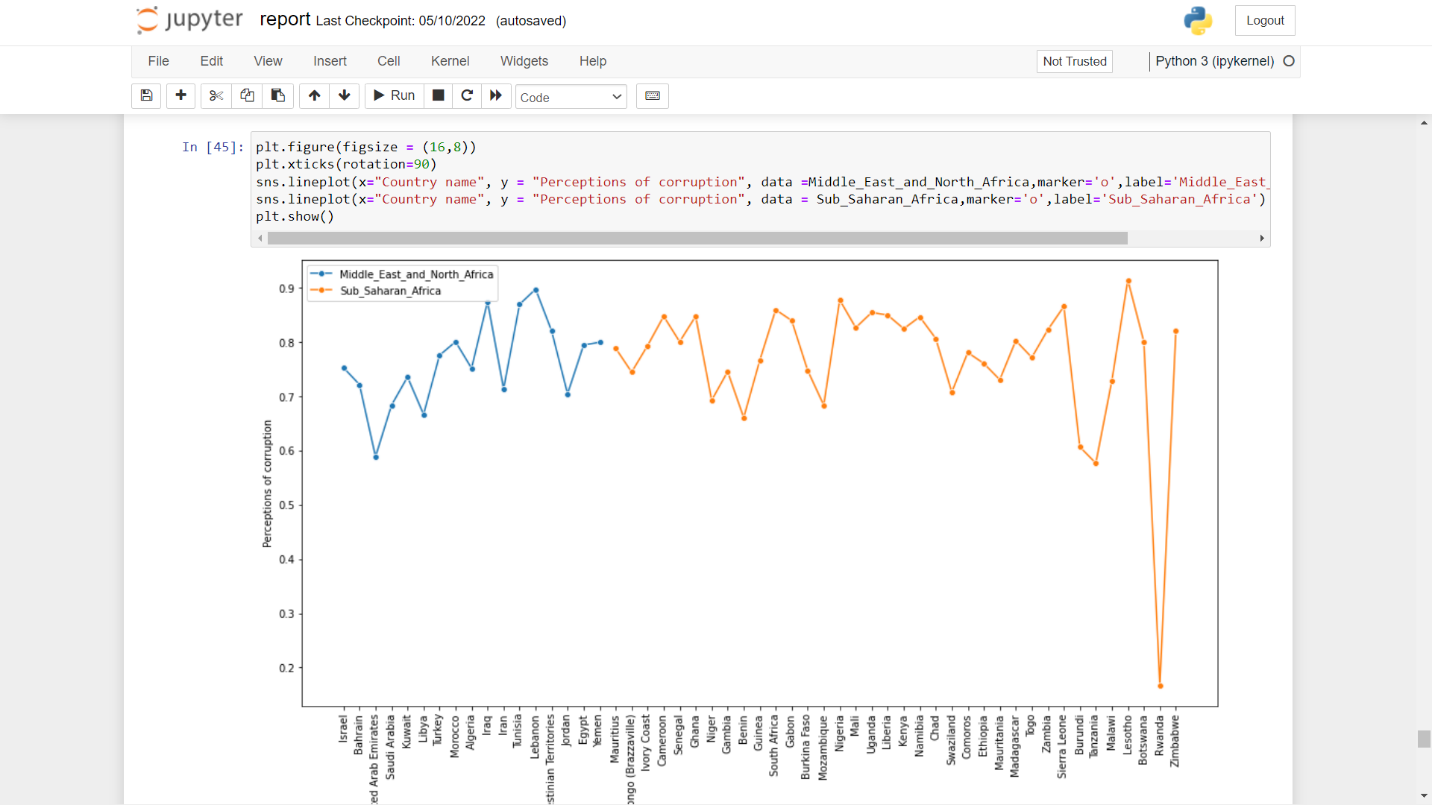
**8.2) Asian Countries –**



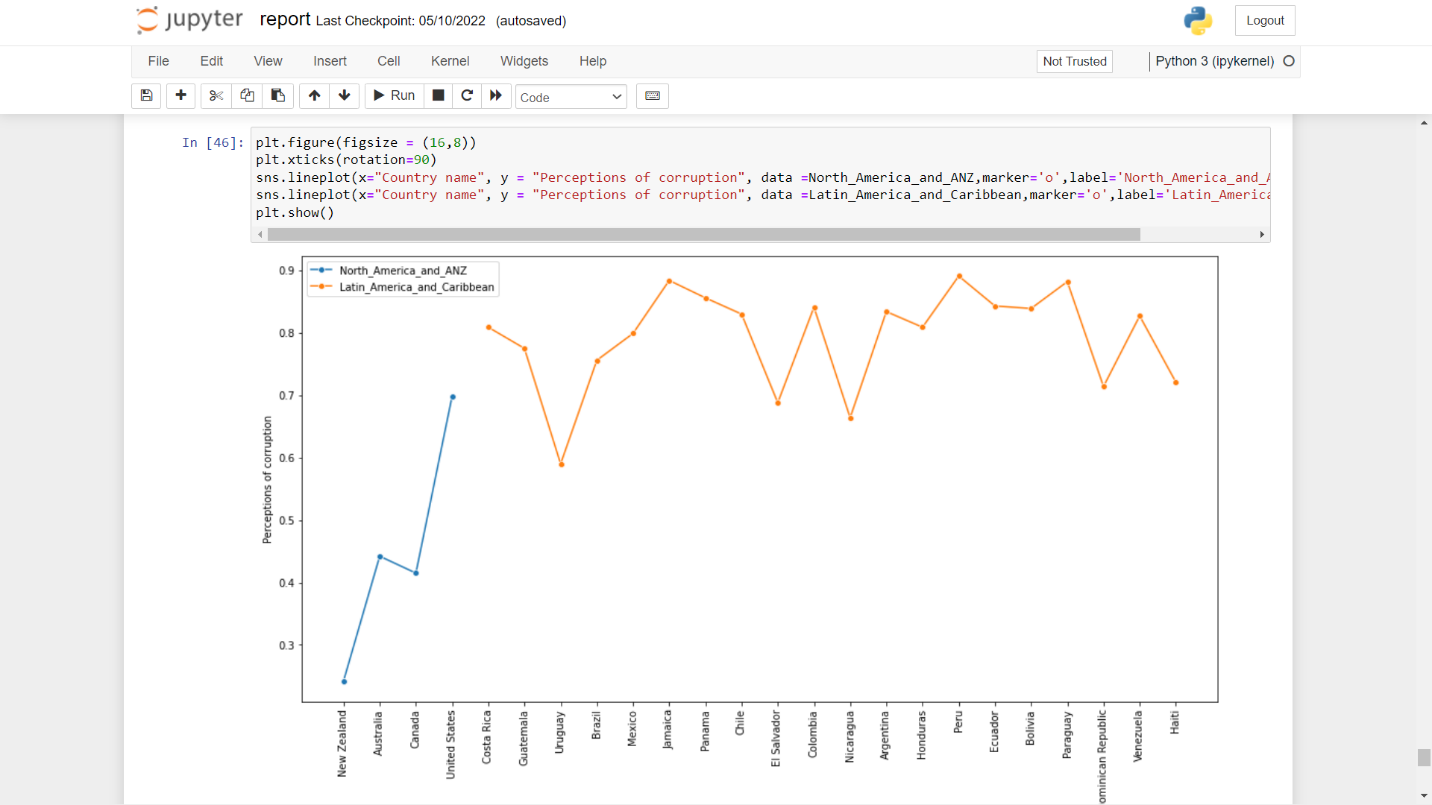
**8.3) European Countries –**



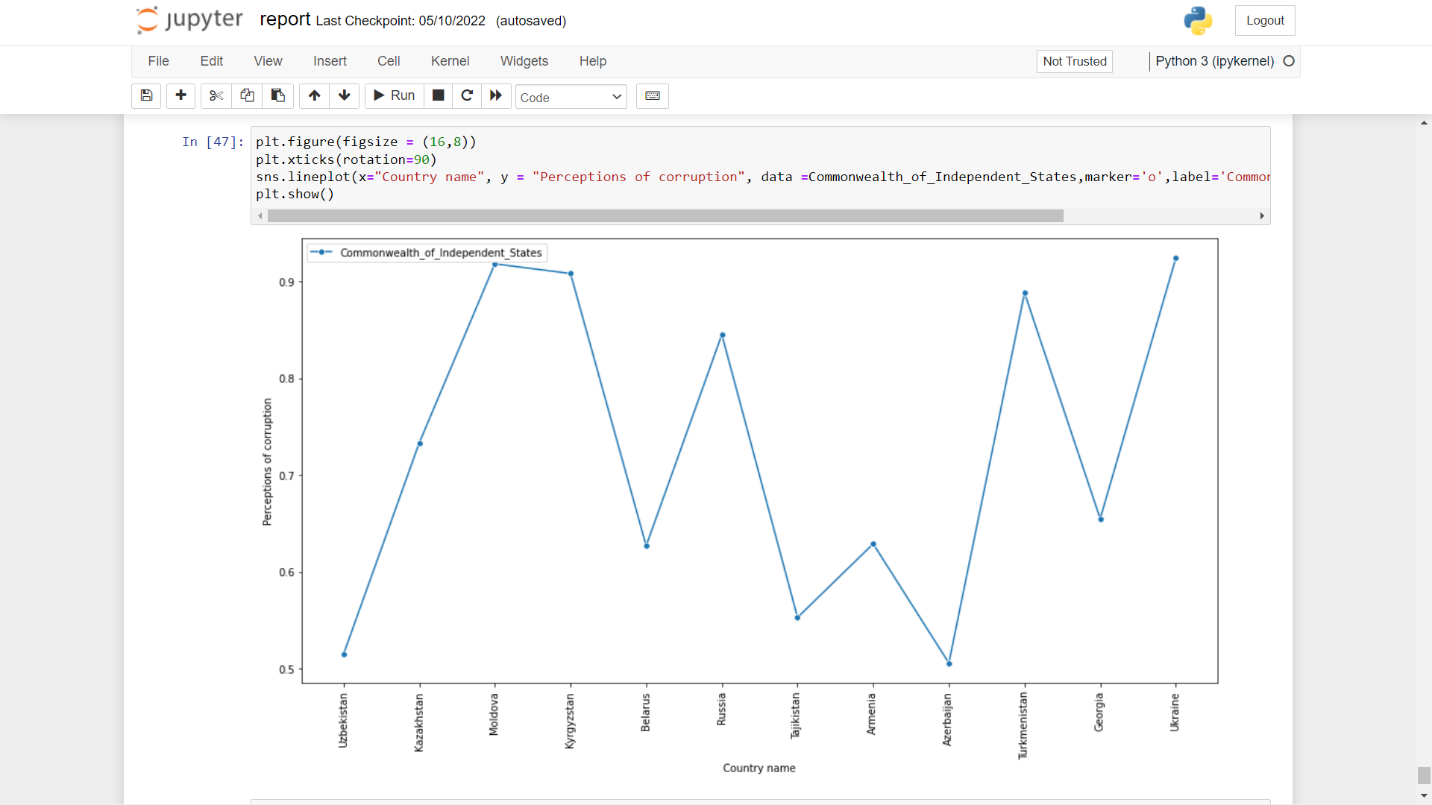
**8.4) Middle East & African –**



**8.5) American Countries –**



**8.6) Independent Countries –**



# ***CHAPTER: 5***

***LIMITATIONS & CONCLUSION***

### Limitations:

* There is huge difference between evaluations and experiences of well-being. For example, Columbia came 37th rank in 2018 world happiness report but 1st by daily emotional experience.
* World Happiness Report are more appropriate for measuring national-level rather than individual-level happiness.

### Conclusion:

The focus on global happiness has finally made it to a big stage, resulting in some interesting statistics and key findings that could potentially further the world's well-being.

In early April, the United Nations held its first conference on happiness with over 600 delegates, including leaders and representatives from nations around the world, attending.

What the U.N. conference produced, among many things, is a robust 160-page [report](http://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf) with pages of references, case studies, and discussions of  happiness metrics and research as well as suggestions for national and world policy goals.

**Six Things To Note:**

* Average happiness in the United States has not risen during times of strong economic growth.
* Political freedom and strong social networks and an absence of corruption are more important than income in regard to what makes people happy.
* Denmark, Norway, Finland and the Netherlands earned the highest scores (7.6 out of 10) for life evaluation.
* Benin, Central African Republic, Togo and the Sierra Leone had an average life evaluation score of 3.4.
* The conference aimed to draw attention to the use of well-being as an alternative metric to Gross National Product for assessing national growth, wellness, and success.
* Conference attendees made a commitment to discuss happiness and well-being at the upcoming [Rio+20 United Nations Conference](http://www.uncsd2012.org/rio20/about.html) on Sustainable Development.

**Other Interesting Highlights and Conclusions:**

**Highlight #1:** There are three major happiness evaluation tools currently in use -- the Gallup World Poll (GWP), the World Values Survey (WVS), and the European Social Survey (ESS). These surveys focus on the responses to any combination of the following questions: "How happy are you now?" "How happy were you yesterday?" and, "How happy are you with your life as a whole these days?"

**Conclusion:** Regular collection of happiness data on a large scale can inform policy-making and help us identify what "deliverables" should be created to foster well-being.

**Highlight #2:**External (e.g., income, work, community, governance) and personal (e.g., health, family, education, age) causes of happiness and misery based on 30 years of research were discussed.

**Conclusion:** Absolute income is important in poor countries, but rich countries tend to place more importance on comparative income. Other factors with a strong impact on happiness include quality of work, social trust, freedom of choice, and political participation.

**Highlight #3:** Research suggests once baseline happiness has been met, happiness varies more with the quality of human relationships rather than with income.

**Conclusion:** Create goals that work toward strong communities with high degrees of trust, high employment and quality of employment, improved mental and physical health, support of family life, and accessible, quality education. The foundation for better policy-making is tied to explaining happiness, measuring and analyzing happiness, and translating research on well-being into actionable resources.

# ***CHAPTER: 6***

***BIBLIOGRAPHY***

### Bibliography

* [**www.w3schools.com**](http://www.w3schools.com)

### <https://worldhappiness.report/>

### <https://en.wikipedia.org/wiki/World_Happiness_Report>

### <https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world>

### https://www.javatpoint.com/