

Mini Project Report

on

Data visualization bot

Submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

Session 2020-21
in

Information Technology

By

Ishank

1900320130076

Ms. Nandita Goyal

Assistant Professor

**DEPARTMENT OF INFORMATION TECHNOLOGY
ABES ENGINEERING COLLEGE, GHAZIABAD**



**AFFILIATED TO
DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW
(Formerly UPTU)**

Student's Declaration

I / We hereby declare that the work being presented in this report entitled "**Data Visualization Bot**" is an authentic record of my / our own work carried out under the supervision of **Ms.Nandita Goyal ,Assistant Professor, Information Technology.**

The matter embodied in this report has not been submitted by me / us for the award of any other degree.

Date:

Signature of student

(Name: Ishank)

(Roll No. 1900320130076)

Department: Information Technology

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of HOD

Signature of Coordinator

**(Name: Prof. Amit Sinha)
Nandita Goyal)**

(Name:Ms.

(Information technology)

**(Assistant
professor)
(Information technology)**

Date:.....

Acknowledgement

Presentation inspiration and motivation have always played a key role in the success of any venture.

I express my sincere thanks to my project coordinator Ms. **Nandita Goyal** to encourage me to the highest peak and to provide me this opportunity to prepare this project. I extend my hearty thanks for giving me the proper guidance even in this time , when everything is continuing on just online platforms .

I am highly indebted to Ms. **Nandita Goyal** for the constant supervision , for providing all the necessary information and support in completing the project.

Finally, I am sincerely thankful to all those people who are directly or indirectly involved in the successful completion of this project work.

Signature of student
(Name: **ISHANK**)
(Roll No. **1900320130076**)

Table of Contents

S. No.	Contents	Page No.
	Student's Declaration	i
	Acknowledgment	ii
Chapter 1 :	Introduction	1
Chapter 2 :	Problem Statement	2
Chapter 3 :	Project Objective	3
Chapter 4 :	Project Methodology	4
Chapter 5 :	Details of Project Work	5-7
5.1 :	Data science	5
5.2 :	Need of visualization	6
5.3 :	Code details	6
5.4 :	Bot role	6-7
Chapter 6 :	Results and Discussion	8-15
Chapter 7 :	Conclusion and Future Scope	16
	References	17
	Appendix-I (Coding)	18-20

INTRODUCTION

My project name is “ Data Visualization Bot ” the basic idea behind my project is to automate the process of visualizing data through some basic programming, in this project I used python language.

Python is interpreted programming language apart from its basic syntax its only about its various functions/packages/libraries which can be widely used in different fields i.e. in web development, data analysis, visualization, machine learning, deep learning, artificial intelligence, data science and so on.

In my project I used basically matplotlib to plot various graphs on given data basically numeric data,

Apart from that I used some other some analysis and scientific calculation packages also i.e. pandas and numpy.

Problem statement

As its a very hactic task to read all the data as industry based data is in millions of rows and is not possible to read all of them and to create useful insights from that and if someone does, it is not possible to convey someone those insights directly through the given or driven data so there is a need of visualize those insights graphically so that a person can understand that easily and in no time.

And it's not very common that the person having analytics skills should also have visualizing skills so “ we have to made a project which can help these people to visualize their data without having any in depth knowledge of it ”.

Project Objective

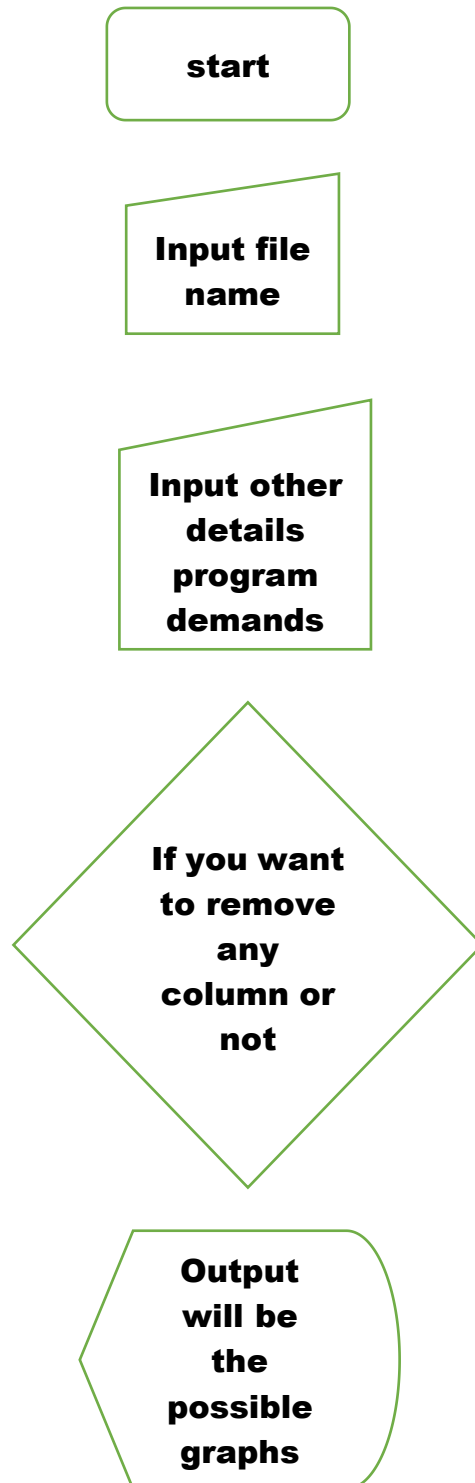
Objective of this project is try to automate the process of visualizing data which can help those people who don't have any in depth knowledge in this field

Now the targeted people can be a school/college going students or faculties who is working on his presentation and can add some visuals and graphs into it and can make it more attractive and meaningful too.

Till now my project can only visualize only numeric values without any nan/none/zeros values.

Project Methodology

the basic methodology behind my project is as follows :



Details of Project Work

I pick my project idea from the domain of data science,

Now data science is itself a huge domain to talk about we will take a brief look about this

Data science :- data science as the term itself describes the science which can be done to any kind of data available now a days

If we see it in more structured way data science is divided into several branches/steps as follows :-

1.Data collection

2.Data processing

3.Data cleaning and analysis

4.Data visualization

5.Creating and deploying ML models

Now in them I choose data visualization and try to make mini project in it I tried to automate the process for limited types of graphs and visuals for only cleaned numeric data (here cleaned refers to the data free from nan/zero/duplicated values)

Need of visualization :-

.Ease of understanding and analysing millions of data in minutes

.As you can read some small volume of data consisting few columns and several rows,But still it is very hacktic and boring task.

Code details :

i have coded in python programming language in this project and used several libraries of it including :

.matplotlib : for plotting graphs

.pandas : for importing data as data frames

.numpy : for performing mathametical calculations

Bot role :-

The bot role is just to make sure user can easily access the programme but in my project it can just speak out some printed lines on the console ,the library used for this is :-

.dispatch function from win32com.client library : to make computer speak out the phrases shown on console window for

for the ease of user.

The voice over of bot we used here is the default one of the library i.e. "SAPI.SpVoice".

Apart from these in the rest part i have used simple python loops, variables and data structures like list and dictionaries in my project to automate the process with the help of in built visualization functions and libraries.

My project is of total hundred line codes with comments in between the code for better readability and understanding of the code for the reader, developer and for me also in later time .

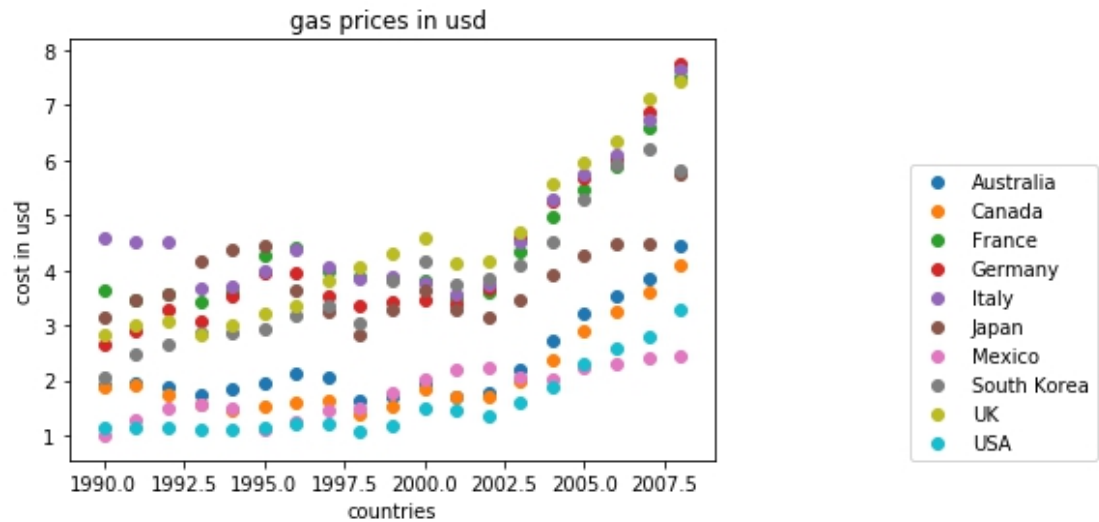
Sample input :-

	A	B	C	D	E	F	G	H	I	J	K
1	Year	Australia	Canada	France	Germany	Italy	Japan	Mexico	South Korea	UK	USA
2	1990	1.91	1.87	3.63	2.65	4.59	3.16	1	2.05	2.82	1.16
3	1991	1.96	1.92	3.45	2.9	4.5	3.46	1.3	2.49	3.01	1.14
4	1992	1.89	1.73	3.56	3.27	4.53	3.58	1.5	2.65	3.06	1.13
5	1993	1.73	1.57	3.41	3.07	3.68	4.16	1.56	2.88	2.84	1.11
6	1994	1.84	1.45	3.59	3.52	3.7	4.36	1.48	2.87	2.99	1.11
7	1995	1.95	1.53	4.26	3.96	4	4.43	1.11	2.94	3.21	1.15
8	1996	2.12	1.61	4.41	3.94	4.39	3.64	1.25	3.18	3.34	1.23
9	1997	2.05	1.62	4	3.53	4.07	3.26	1.47	3.34	3.83	1.23
10	1998	1.63	1.38	3.87	3.34	3.84	2.82	1.49	3.04	4.06	1.06
11	1999	1.72	1.52	3.85	3.42	3.87	3.27	1.79	3.8	4.29	1.17
12	2000	1.94	1.86	3.8	3.45	3.77	3.65	2.01	4.18	4.58	1.51
13	2001	1.71	1.72	3.51	3.4	3.57	3.27	2.2	3.76	4.13	1.46
14	2002	1.76	1.69	3.62	3.67	3.74	3.15	2.24	3.84	4.16	1.36
15	2003	2.19	1.99	4.35	4.59	4.53	3.47	2.04	4.11	4.7	1.59
16	2004	2.72	2.37	4.99	5.24	5.29	3.93	2.03	4.51	5.56	1.88
17	2005	3.23	2.89	5.46	5.66	5.74	4.28	2.22	5.28	5.97	2.3
18	2006	3.54	3.26	5.88	6.03	6.1	4.47	2.31	5.92	6.36	2.59
19	2007	3.85	3.59	6.6	6.88	6.73	4.49	2.4	6.21	7.13	2.8
20	2008	4.45	4.08	7.51	7.75	7.63	5.74	2.45	5.83	7.42	3.27

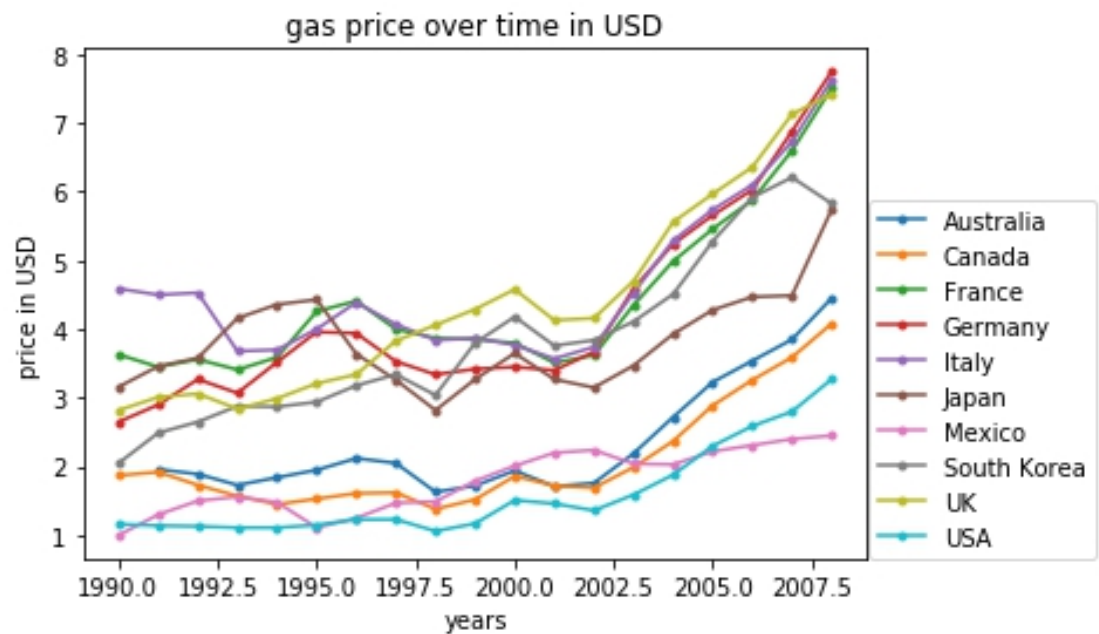
Results and Discussion

Results:-

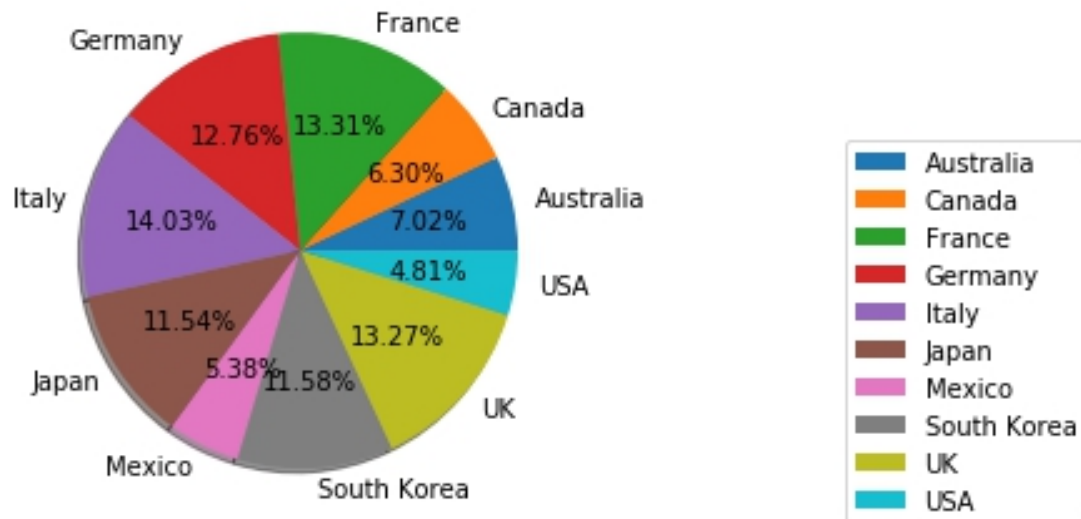
Scatter plot:



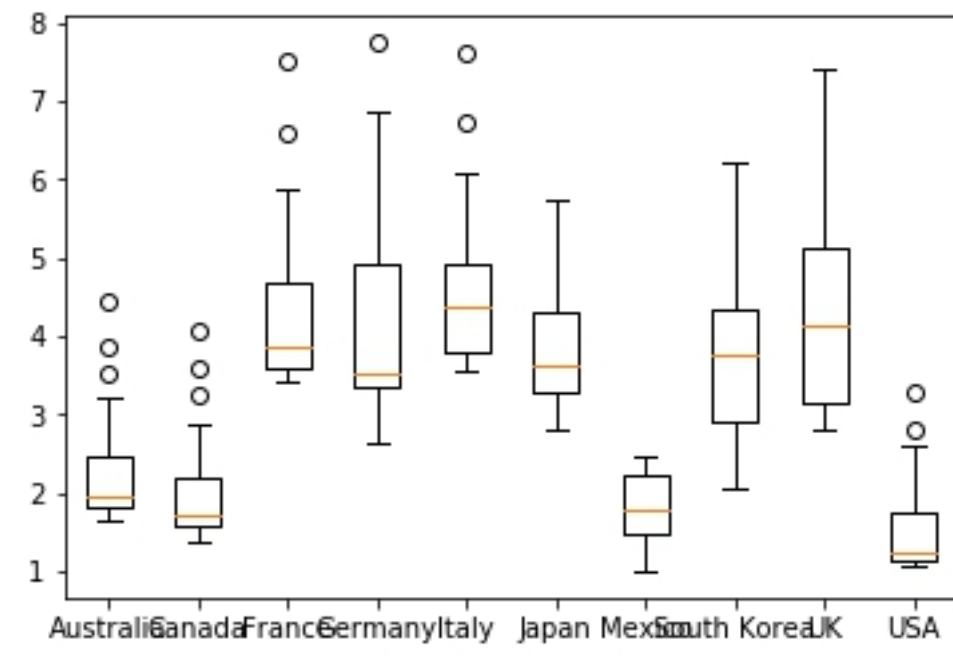
Line graph:



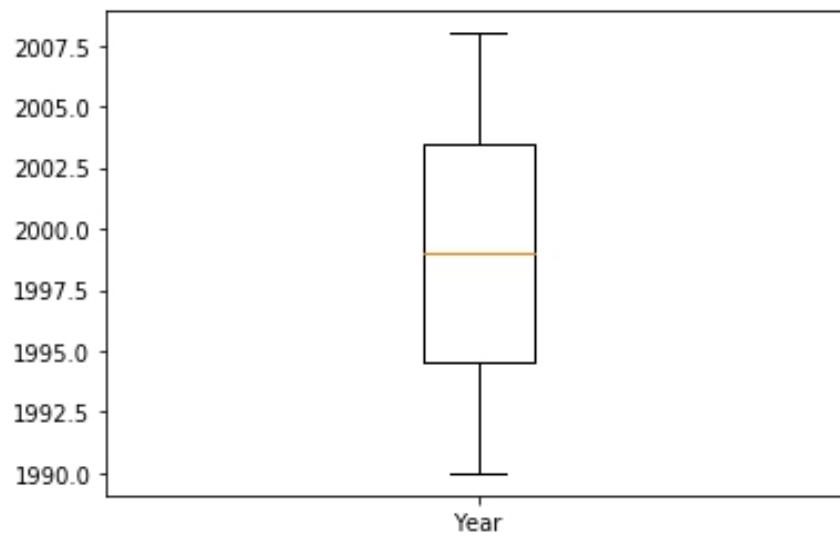
Pie graph:



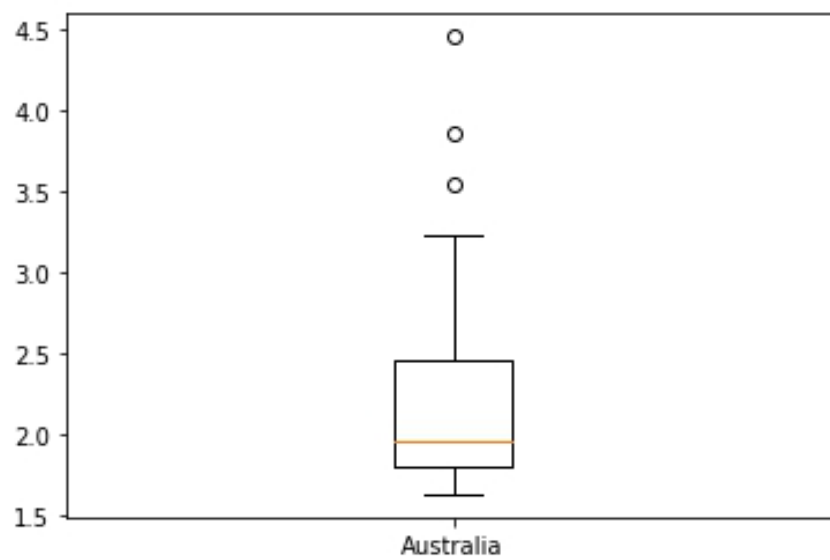
Box plot:



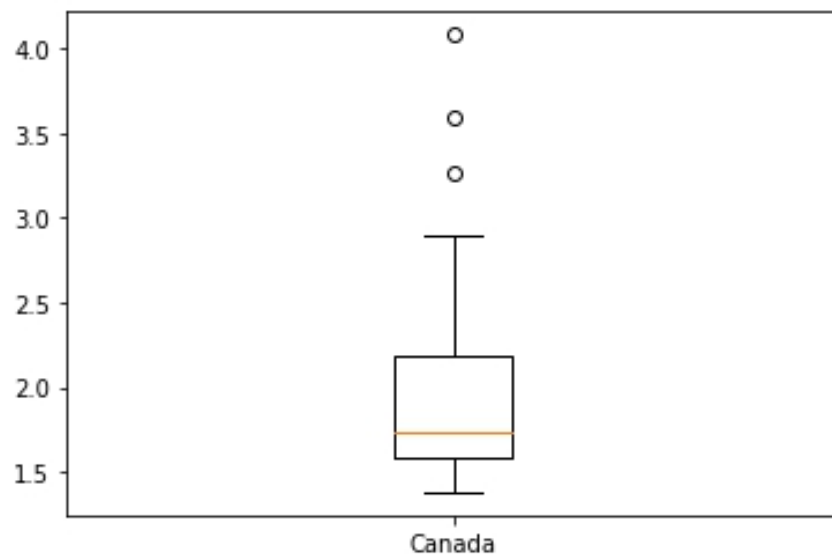
Box plot of every column



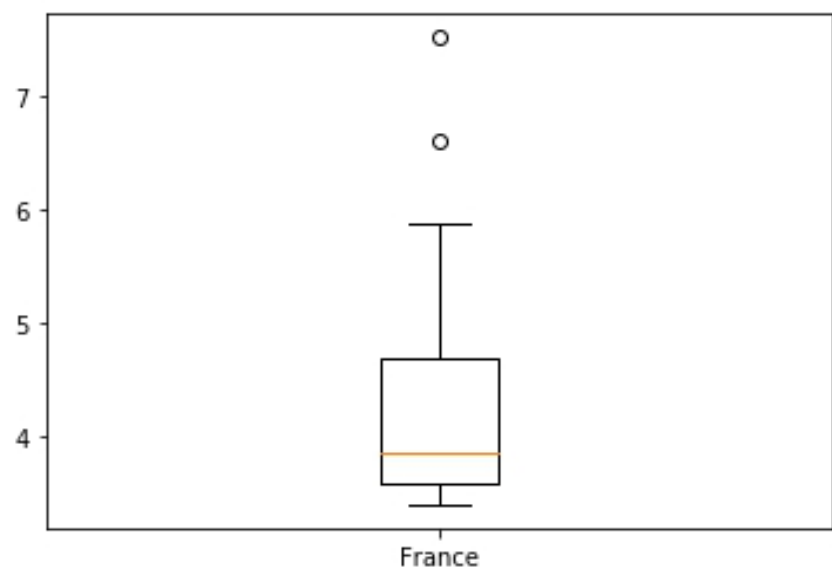
□



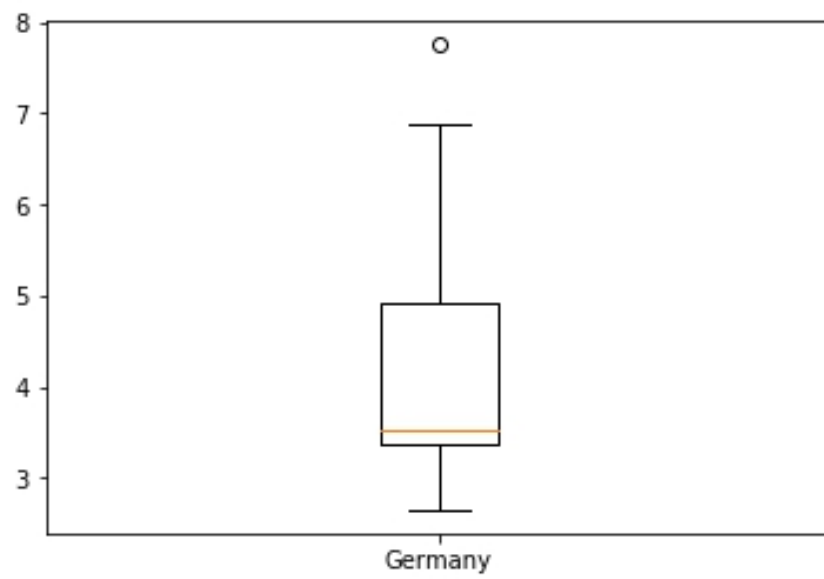
□



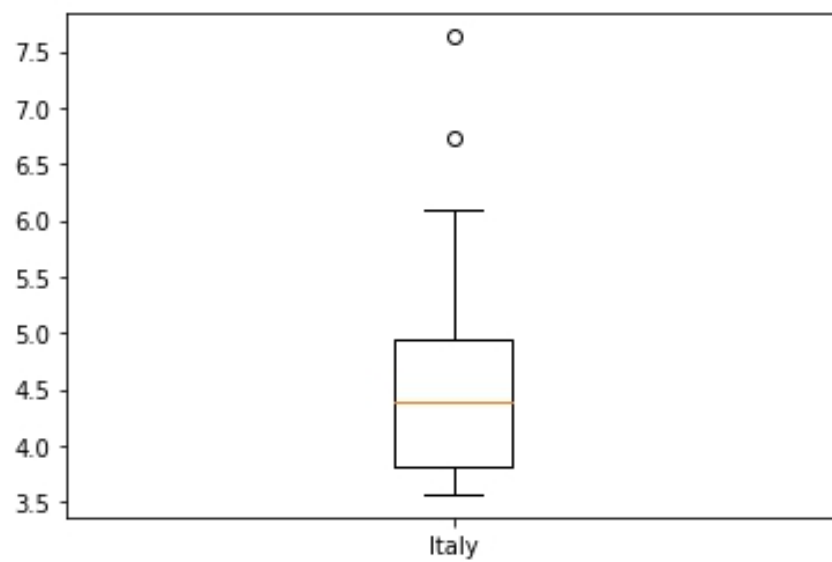
□



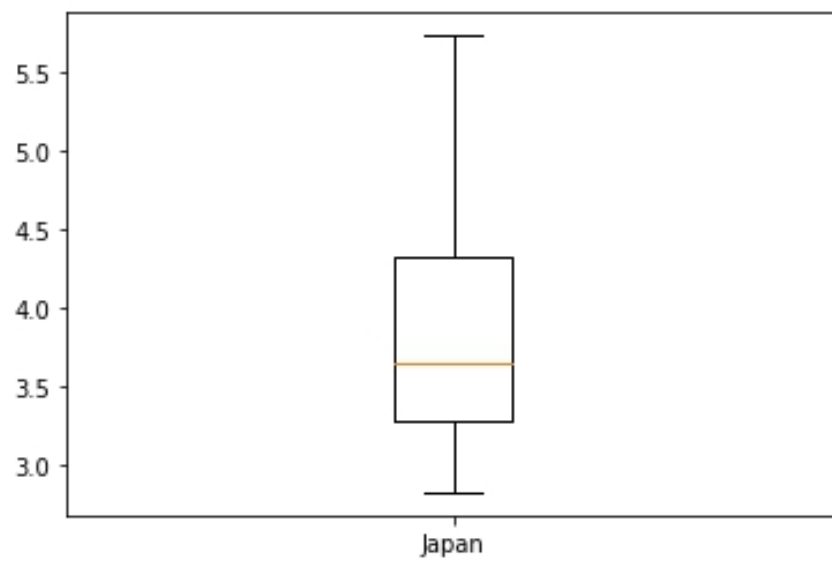
□



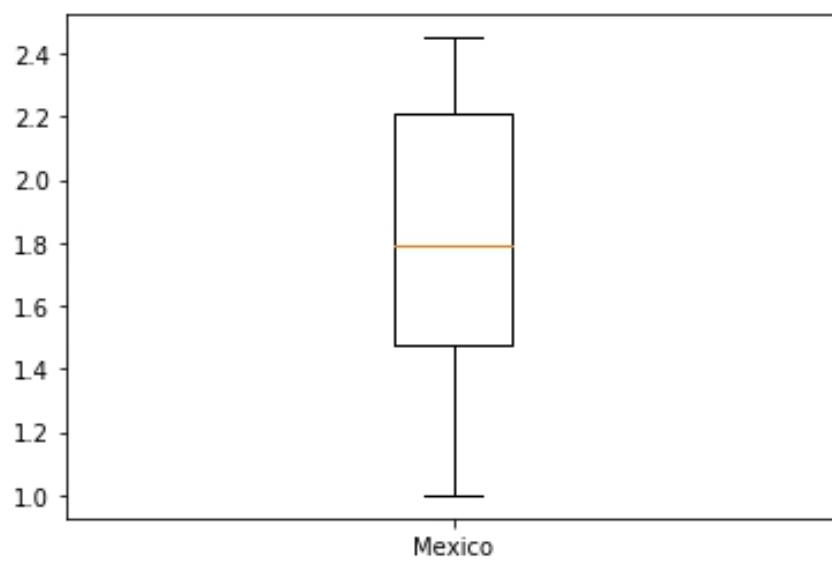
□



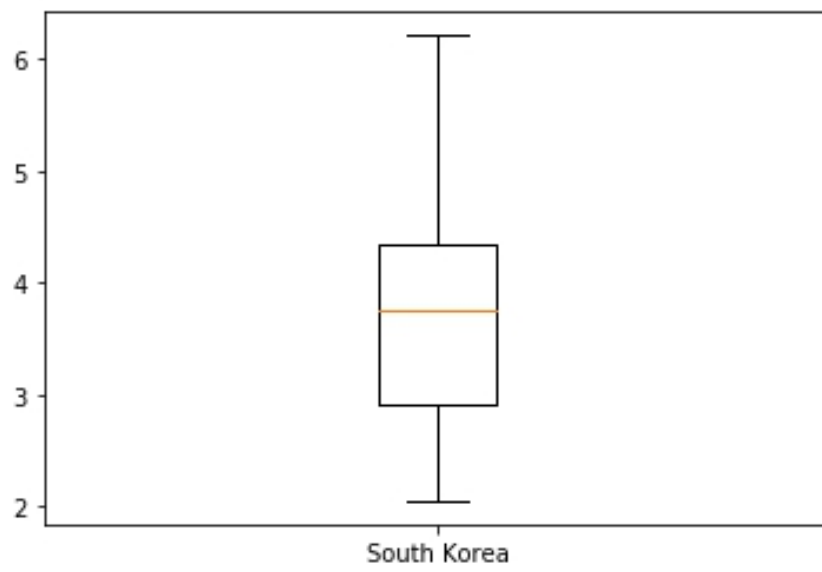
□



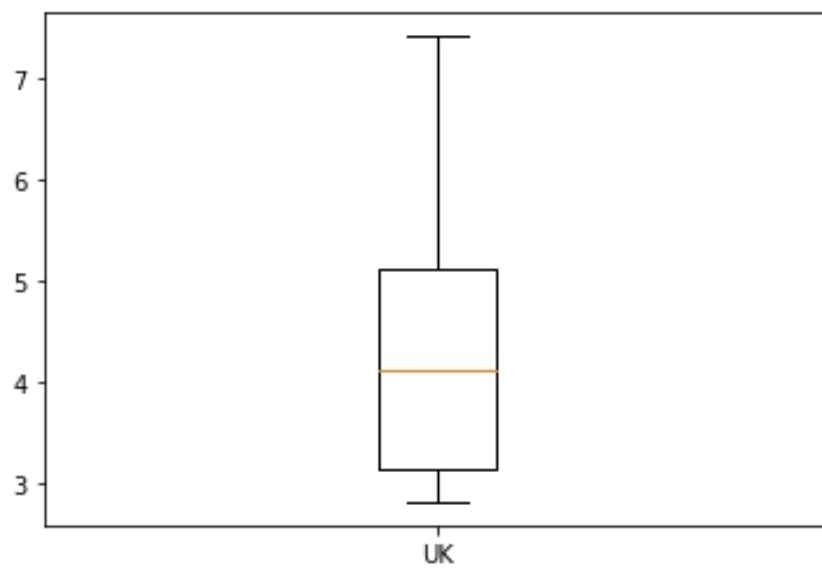
□



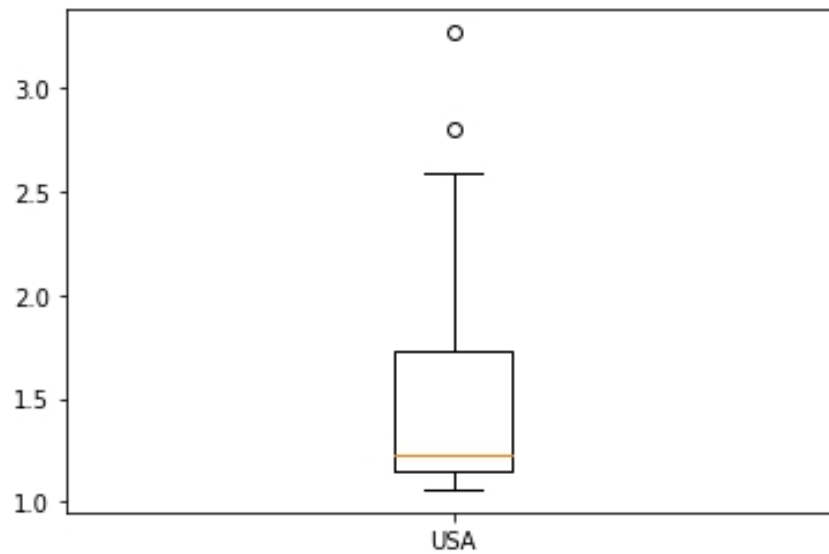
□



☐



☐



Limitations :-

- .Uncleaned data or the data containing zeros/nan/duplicate values can't be process/visualized**
- .data file should be present in the same directory or the folder in which program file is present**
- .data file should be in csv format**
- .file should only contain numeric values**

Conclusion and Future Scope

Future scope :-

This project can be merged into some app or website for better user interface and ease of user and can widely be used by any person in this world.

In future times it can be updated with more but basic analytics functions like removing duplicates and zero/null values from the data so it will provide more access to the user to do analytics as well in it of basics level not complex sql queries.

and further on a chat bot can also be added to it so that the new user can do their desired task without any difficulty and in lesser time.

References

1. <https://www.youtube.com/channel/UCq6XkhO5SZ66N04IcPbqNcw>
2. <https://www.youtube.com/user/edurekaIN>
3. <https://stackoverflow.com/>

Appendix 1

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from win32com.client import Dispatch
s = Dispatch("SAPI.SpVoice")
print("write EXACT file name with format as saved in your
computer and make sure it is present in the same directory in
which programme file is present")
s.Speak("write EXACT file name with format as saved in your
computer and make sure it is present in the same directory in
which programe file is present")
file = input()#input file name
mydata = pd.read_csv(file)
print("enter the X label")
s.Speak("enter the X label")
plt.xlabel(input())
print("enter the Y label")
s.Speak("enter the y label")
plt.ylabel(input())
#plt.plot(mydata.Year, mydata.Australia, )
#plt.plot(mydata.Year, mydata.Canada, )
print("enter the graph title")
s.Speak("enter the graph title")
plt.title(input())
print("enter the EXACT COLOUMN NAME you want to choose at
Y AXIS")
s.Speak("enter the EXACT COLOUMN NAME you want to choose
at Y AXIS")
t= input()
#so that i don't have to input all the columes of the data it will
iterate its coloums
def linegraph():
    for i in mydata:
        if i != t:
```

```

        print(i)
        plt.plot(mydata[t],mydata[i],label = i,marker='.')
        plt.legend(loc = (1.1,0))#to put legend outside the
                                graph
def scatter():
    for i in mydata:
        if i != t:
            # print(i)
            plt.scatter(mydata[t], mydata[i],label = i )
            plt.legend(loc = (1.1,0))
def histogram():
    for i in mydata:
        if i != t:
            # print(i)
            plt.hist(mydata[i],label = i )
            plt.legend(loc = (1.1,0))
def piegraph():
    val = []
    print("enter the coloumn you want to remove from pie graph
note, it will only be removed for pie graph")
    s.Speak("enter the coloumn you want to remove from pie
graph note, it will only be removed for pie graph")
    f=input()
    for i in mydata.columns:
        if i != f:
            z = mydata[i].tolist()
            m = np.mean(z)
            val.append(m)
    print(val)
    t= mydata.columns.tolist()
    t.remove(f)
    print(t)
    plt.pie(val,labels = t,autopct = '%0.3f%%',shadow =
True,radius = 1.3)
    plt.legend(loc = (1.3,0))
def boxplot():
    val = []
    print("enter the coloumn you want to remove or just write no
then enter")
    s.Speak("enter the coloumn you want to remove or just

```

```

    write no then enter")
    f=input()
    for i in mydata.columns:
        if i != f:
            z = mydata[i].tolist()
            m = np.mean(z)
            val.append(m)
    print(val)
    print(np.mean(val))
    t= mydata.columns.tolist()
    if f!="no":
        t.remove(f)
    print(t)
    plt.figure()
    plt.boxplot(val)
    box = []
    label = []
    for j in mydata.columns:
        if j != f:
            z = mydata[j].tolist()#convert all the row entries into
a list of that coloumn
            box.append(z)
            label.append(j)
            plt.figure()
            plt.boxplot(z,labels = [j])
    plt.figure()
    print("enter the title of main graph")
    s.Speak("enter the title of main graph")
    plt.title(input())
    plt.boxplot(box,labels = label)
plt.figure(1)
scatter()
plt.figure(2)
linegraph()
plt.figure(3)
histogram()
plt.figure(4)
piegraph()
plt.figure(5)
boxplot()

```