CLUSTERING ASSIGNMENT

ANKIT BHARDWAJ

Step to procedure in the Assignment

- Problem Statement
- Data Understanding
- Performing Clustering
- Identifying the Countries
- Recommendations and Conclusion

Problem Statement – Identify top countries that are in direst need of aid. Categorization will be based on socio-economic and health factors that determine the development of the country.

Data Understanding

```
In [284]: # Importing the libraries
          import pandas as pd
          import numpy as np

          import matplotlib.pyplot as plt
          import seaborn as sns

          from sklearn.cluster import KMeans
```
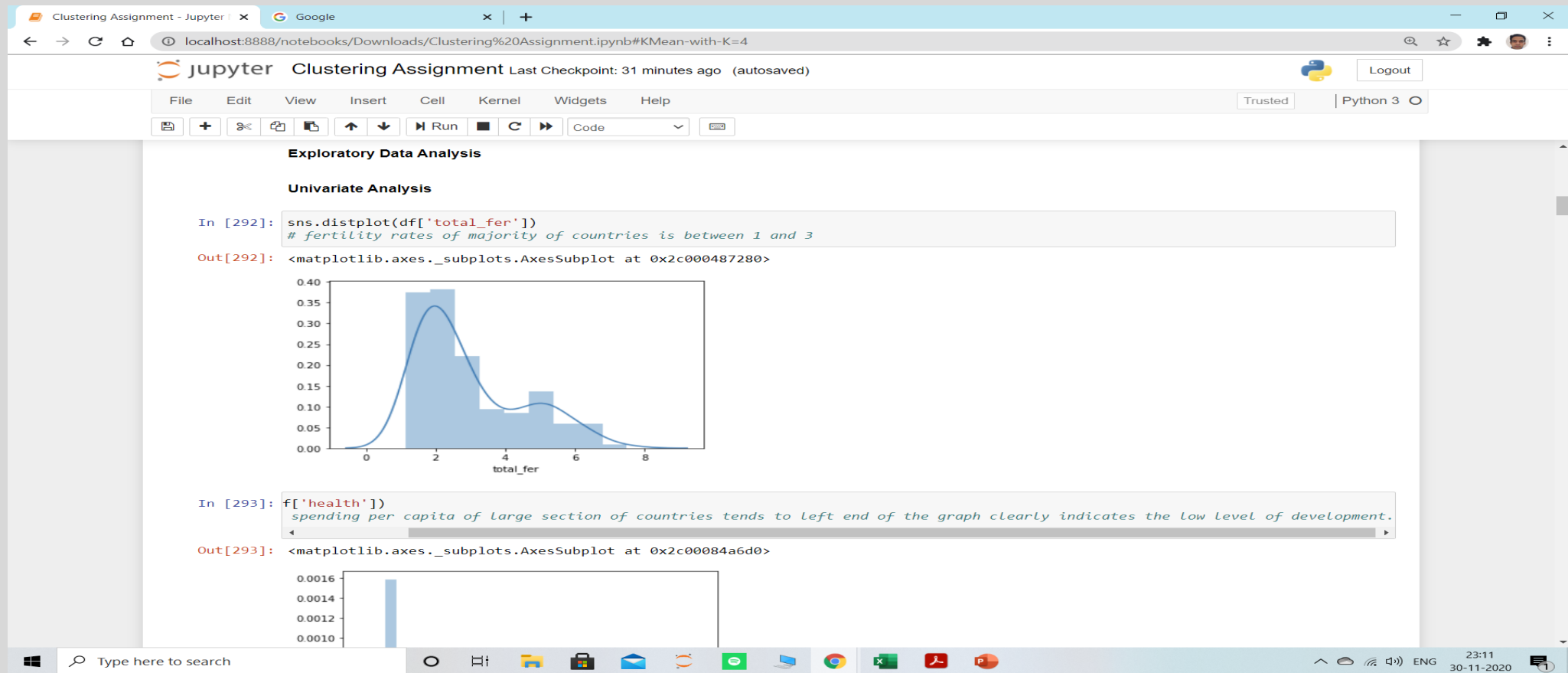
```
In [285]: import warnings
          warnings.filterwarnings('ignore')
```

```
In [286]: df = pd.read_csv("Country-data.csv")
          df.head()
```
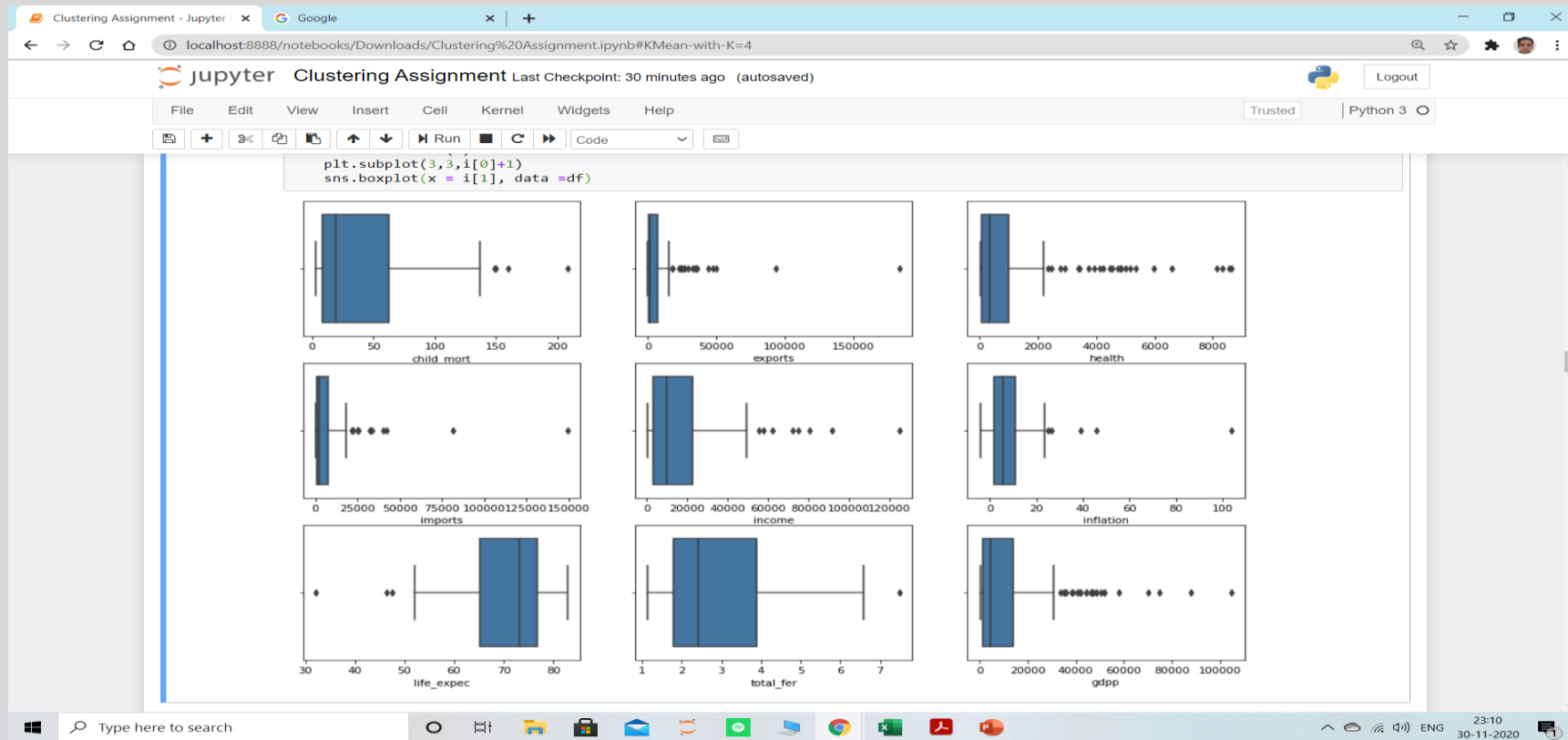
Out[286]:

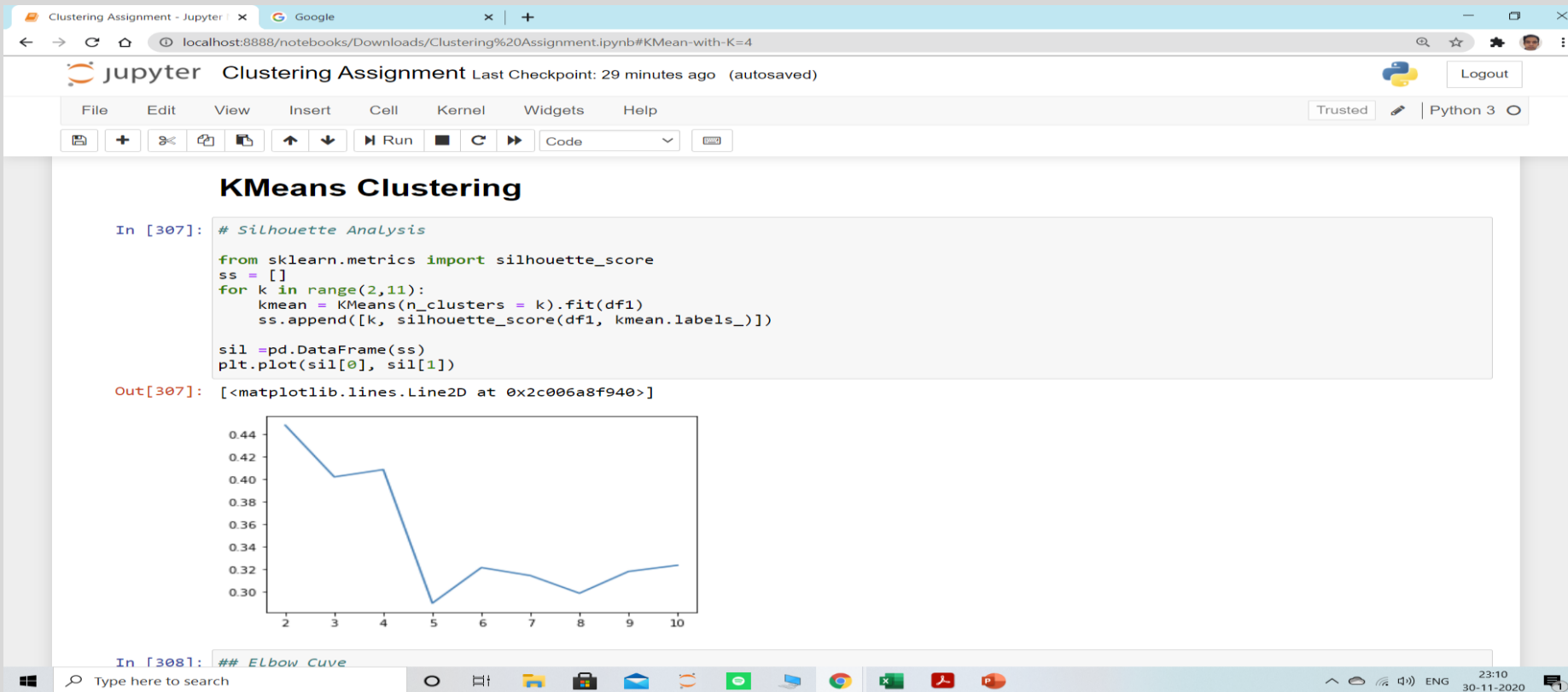|	country	child_mort	exports	health	imports	income	inflation	life_expec	total_fer	gdpp
0	Afghanistan	90.2	10.0	7.58	44.9	1610	9.44	56.2	5.82	553
1	Albania	16.6	28.0	6.55	48.6	9930	4.49	76.3	1.65	4090
2	Algeria	27.3	38.4	4.17	31.4	12900	16.10	76.5	2.89	4460
3	Angola	119.0	62.3	2.85	42.9	5900	22.40	60.1	6.16	3530
4	Antigua and Barbuda	10.3	45.5	6.03	58.9	19100	1.44	76.8	2.13	12200

Data Understanding – Various steps will be involved such as Exploratory Data Analysis including uni and bivariate analysis
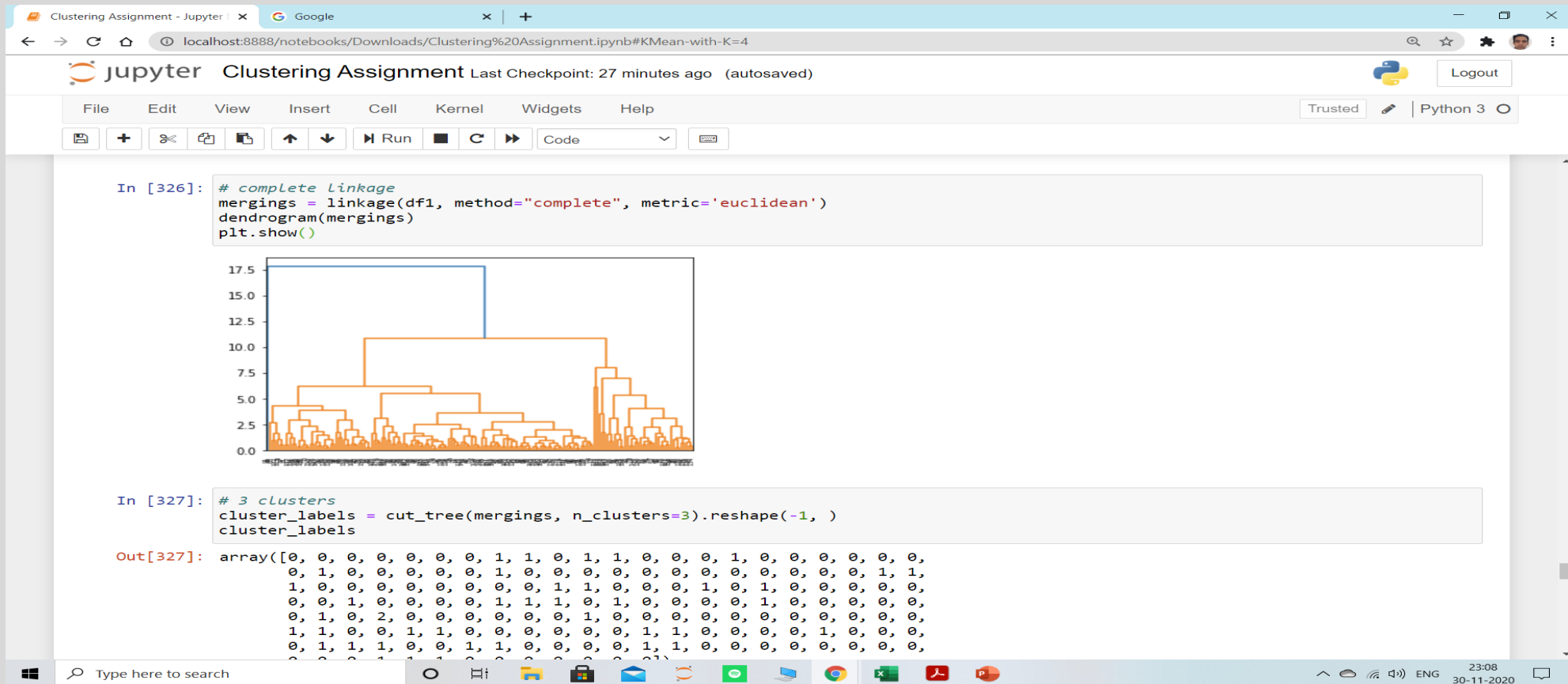
Data Understanding – EDA will be followed by outlier analysis

Performing Clustering – Both kind will be used K-Means and Hierarchical Clustering

Both kind of Hierarchical Clustering will be used single and complete linkage.

Identifying the top five countries in dire need for aid

Recommendations and Conclusion

- The top five countries were selected on the basis of three key variables that is GDPP (GDP Per Capita), Child Mortality rate and Income of the country.
- The socio-economic and health factors determine are key indicators of a country to receive any kind of financial aid.

THE END