**DATA VISUALIZATION**

**Name: Ismail Usmangani Sarigat**

**Student Id: 21088719**

**Module Name: Applied Data Science – 1**

**Rework-1**

**ADS REWORK – 1**

**Q1. Produce a line plot showing multiple lines with proper labels and legend. Describe what conclusions you can draw from this plot.**

* **CODE:**

# Importing libraries

import pandas as pd

import matplotlib.pyplot as plt

# Read csv file

agriculture = pd.read\_csv('agri\_land.csv')

print(agriculture)

plt.figure(figsize=(8,8))

plt.hist(agriculture["2000"], label="2000", alpha=0.8, bins=4)

plt.hist(agriculture["2010"], label="2010", alpha=0.5, bins=4)

plt.legend()

plt.xlabel("Agricultural Land (%)")

plt.ylabel("Frequency")

plt.show()

* **EXPLANATION:**

In this dataset, the details shown is of agricultural land used in different regions around the world, in different years. The prominent reason to select this graph was to note that out of differences of how frequently the agriculture land is use in two year(2000, 2010,) within ten years period. As a result of which, it can be seen from the result that use of land is shows fluctuations. But it also keep decreasing over the period of time as compare to 2000.

Dataset:<https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?end=2020&start=1961&view=chart>

* **OUTPUT:**

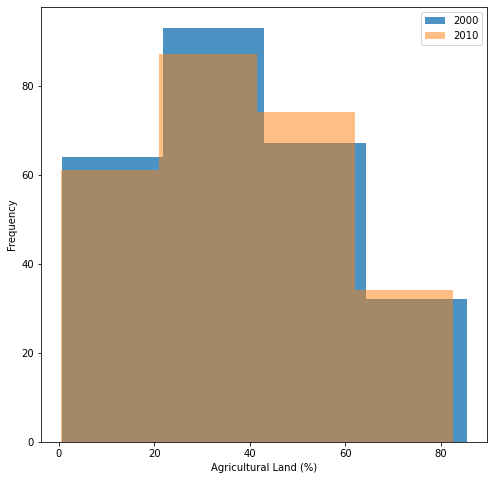


Figure 1 Plot

**Q2. Produce graphs using two other visualisation methods. Explain why you picked this type of graph and describe what conclusions you can draw.**

1. **Line Graph –**

* **CODE:**

# Read csv file

forest = pd.read\_csv('forest\_area.csv')

print(forest)

plt.figure(figsize=(12,9))

plt.plot(forest["1990"], color="Brown", label="1990")

plt.plot(forest["2000"], color="Blue", label="2000")

plt.plot(forest["2010"], color="Orange", label="2010")

plt.legend()

plt.xlabel("N")

plt.ylabel("Forest Area (%)")

plt.show()

* **EXPLANATION:**

The program depicts the Forest area around the world during different years. Here, I have three-line plots to showcase percentage of forest area in different years such as 1990, 2000 and 2010. The land reach to nearly 100% in year 1990 at one point, in 2010 it keep fluctuating through out the years addition to this, total forest area from 2000 keep on decreasing little and was 2nd in the position compare to 2010.

Datalink: <https://data.worldbank.org/indicator/AG.LND.FRST.ZS?end=2020&start=1961&view=chart>

* **OUTPUT:**

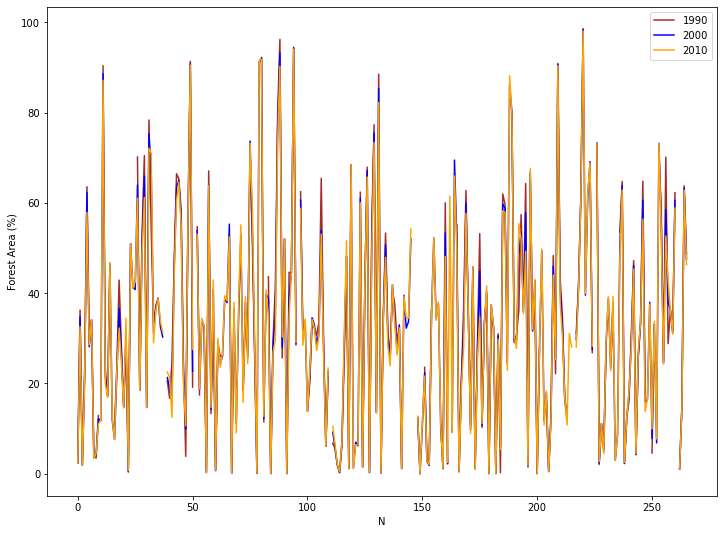


Figure 2: Line Plot

1. **Pie Plot –**

* **CODE:**

# Read csv file

region = pd.read\_csv('revenue.csv')

print(region)

name = ["North", "East", "South", "West"]

plt.figure()

plt.pie(region["TOTAL REVENUE"], labels=name, autopct='%1.1f%%')

plt.title("Total Revenue in Every Region")

plt.show()

* **EXPLANATION:**

The dataset contain revenue of every regions and the graph best suited to depict the proportion is pie plot. The graph shows that north has the highest revenue of all the regions at 42.1% and the least has the least revenue at just 11.%.

**Dataset Link:** https://data.worldbank.org/indicator/revenue

* **OUTPUT:**

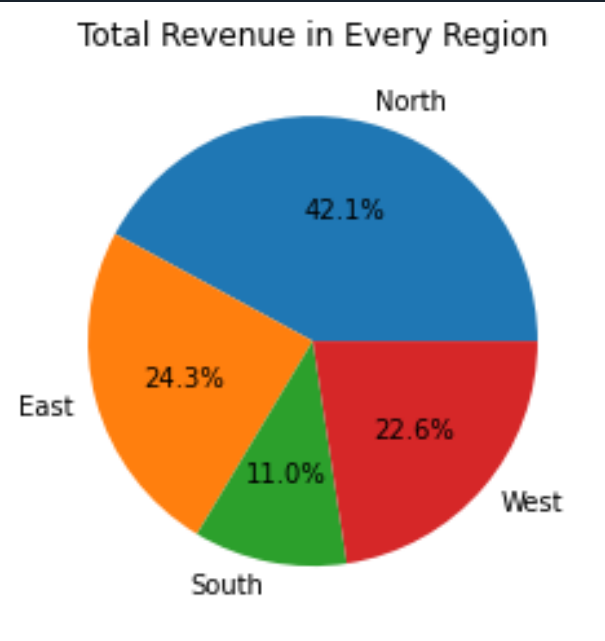


Figure 3: Pie Plot

**Link of my Repo –** [**https://github.com/is2295/ADS\_REWORK-1/**](https://github.com/is2295/ADS_REWORK-1/)