Applied Data Science

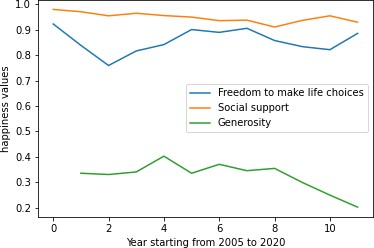
ASSIGNMENT 1

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

* CODE:

|  |  |
| --- | --- |
|  |  |
|  | #line graph  #reading data from csv file using pandas libraries  happinesDataReport = pd.read\_csv("world-happiness-reportmain.csv") |
|  |  |
|  | #plotting the line graph |
|  | plt.figure() |
|  | plt.plot(happinesDataReport["Freedom to make life choices"], label="Freedom to make life choices") |
|  | plt.plot(happinesDataReport["Social support"], label="Social support") |
|  | plt.plot(happinesDataReport["Generosity"], label="Generosity") |
|  | plt.xlabel("Year starting from 2005 to 2020") |
|  | plt.ylabel("happiness values") |
|  | plt.legend() |
|  | plt.savefig("linegraph.png") |
|  | plt.show() |

* OUTPUT:



* EXPLANATION:

The above combination of line graph is happiness report of United Kingdom starting from 2005 to 2020. There are three criteria such as generosity, social support, and freedom to make life choices. The major reason to choose this graph for happiness data set record is comparison can be easily seen and understandable. social support was always on the pick on the people’s life. In 2005 generosity was approximately 0.39 but then it was decreased continuously and reached to 0.2. However, freedom to make life decision stayed same during this time with some fluctuations. As well as there was little decrement in social support.

* DATASET LINK:

<https://www.kaggle.com/datasets>

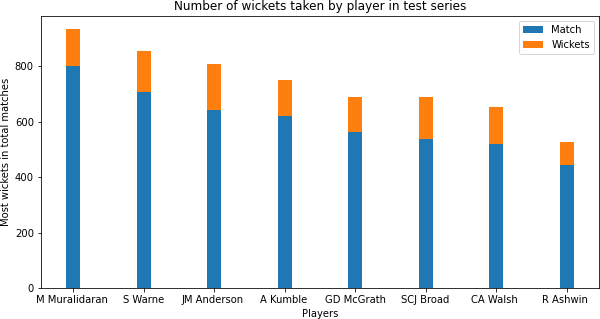
# Q2. Produce graphs using two other visualization methods. Explain why you picked this type of graph and describe what conclusion you can draw.

1. **Bar Graph**

* CODE:

|  |  |
| --- | --- |
|  | #import libraries  import pandas as pd  #bar graph |
|  | #reading data from csv file using pandas libraries |
|  | testCricketData=pd.read\_csv("Most Wickets in Test Cricket .csv") |
|  |  |
|  | #plotting graph using library function |
|  |  |
|  | plt.figure(figsize=(10,5)) |
|  | plt.bar(testCricketData["Player "], testCricketData["Wickets "], width=0.2, label="Match") |
|  | plt.bar(testCricketData["Player "], testCricketData["Matches"], bottom =testCricketData["Wickets "], width=0.2, label="Wickets") |
|  | plt.title("Number of wickets taken by player in test series") |
|  | plt.xlabel("Players") |
|  | plt.ylabel("Most wickets in total matches") |
|  | plt.legend() |
|  | plt.savefig("bargraph") |
|  | plt.show() |

* OUTPUT:



* EXPLANATION:

The above combination of bar graph is the Most Test wickets taken by 8 different players during specific span. The major reason to choose this bar graph for this data set is, using bar graph we can easily differentiate who has taken more wickets and who has taken least wickets. From graph, it can be observed that Highest wicket taker is M Muralidaran. He has taken more than approximate 1000 wickets in 800 match.

* DATASET LINK:

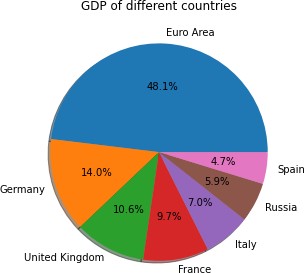
<https://www.kaggle.com/datasets>

# Pie Graph

* + CODE:

|  |  |
| --- | --- |
|  | #pie chart  #reading csv file using pandas library |
|  | economyIndicatorData = pd.read\_csv("Economy\_Indicators.csv") |
|  |  |
|  | #creating array to give label |
|  | countryName = ["Euro Area", "Germany", "United Kingdom", "France", "Italy", "Russia", "Spain"] |
|  |  |
|  | #plotting graph using pandas library |
|  | plt.figure(figsize=(20,5)) |
|  | plt.pie(economyIndicatorData["GDP"], labels=countryName, autopct='%1.1f%%',shadow=True) |
|  |  |
|  | plt.title("GDP of different countries") |
|  | plt.savefig("piegraph.png") |
|  | plt.show() |

* + OUTPUT:



* + EXPLANATION:

The above pie chart shows the GDP of different counties. Because of the specific value of GDP, we can use it to plot a pie chart. The countries are Germany, Russia, United Kingdom, etc. The highest GDP is covered by Euro area with 48.1% and lowest GDP is of Spain with 4.7%.

* + DATALINK:

[https://data.world](https://data.world/)

* GitHub Link: <https://github.com/Harvi05/AppliedDataScience>