**Data Analysis and Visualisation in Python**

In this project, we used Python to extract and analyse specific data from the 'GDP (Nominal) per Capita.csv' dataset. Our first step was to load the data from the CSV file into a DataFrame named "df" using Colab notebook. To quickly explore the dataset, we printed the first 10 rows and the last 5 rows, providing a snapshot of the data. Additionally, we identified the country with the highest UN\_Estimate value and examined essential columns, including 'Country/Territory' and 'UN\_Region', which helped us better understand the data’s structure and attributes.

Throughout the project, we made use of various Python functionalities. For instance, quotes in Python (single, double, or triple quotes) were used to define string values, while escape characters were used to handle special characters within strings. We used print() commands to display output, with display() being useful for a more formatted output in Colab, while print() and show() provided simpler forms of output to the console.

We also explored important positional arguments to define the order of parameters in functions, which helped streamline our functions. Furthermore, we utilised algorithms and arithmetic operators for data manipulation and analysis, along with shortcut operators for efficient coding. Conditional statements allowed us to perform operations based on specific conditions, such as filtering countries with MF\_Estimate values below the average.

To assist with the visual analysis of the data, we employed several visualisation techniques, including bar plots, scatter plots, and box plots. These visualisations were instrumental in uncovering patterns, distributions, and relationships within the data, allowing us to gain deeper insights.

Additionally, we used Python’s correlation function to explore the relationships between the estimates provided by the World Bank, UN, and IMF. By analysing these correlations, we were able to understand how the estimates from these organisations align or differ, providing valuable context for the dataset.

**Please download the ‘GDP (nominal) per Capita.csv’ dataset** [**here**](https://justit831-my.sharepoint.com/:x:/g/personal/danpe_justit_co_uk/EV1Xzb5eNENHmOVMDssxyoMBqTCVcLg18U4qOLUDZZHSkw?e=PAbKfN)**.**

# Read and save the ‘GDP (nominal) per Capita’ data to a data frame called “df” in Colab notebook.

df = pd.read\_csv('GDP (nominal) per Capita.csv')

df.to\_csv('GDP (nominal) per Capita.csv', index=False)

#Print the first 10 rows

df.head(10)

A screenshot of a computer screen

AI-generated content may be incorrect.

#Print the last 5 rows

df.tail()

A screenshot of a computer

AI-generated content may be incorrect.

#Country has highest UN\_Estimate

gdp=df[df["UN\_Estimate"]==df["UN\_Estimate"].max()]

gdp['Country/Territory']

A screenshot of a computer

AI-generated content may be incorrect.

#Country has highest Worlbank Estimate

gdp=df[df["WorldBank\_Estimate"]==df["WorldBank\_Estimate"].max()]

gdp['Country/Territory']

A close up of a text

AI-generated content may be incorrect.

#Country has highest IMF Estimate

gdp=df[df["IMF\_Estimate"]==df["IMF\_Estimate"].max()]

gdp['Country/Territory']

A screenshot of a computer

AI-generated content may be incorrect.

# Calculate the average of 'Worldbank\_Estimate' and 'UN\_Estimate' columns

avg\_worldbank\_UN = df[['WorldBank\_Estimate', 'UN\_Estimate']].mean()

avg\_worldbank\_UN

A close up of words

AI-generated content may be incorrect.

# Minimum and maximum of 'Worldbank\_Estimate'

df["WorldBank\_Estimate"].agg(["min","max"])

A screenshot of a computer

AI-generated content may be incorrect.

# Print ‘Country/Territory’ and ‘UN\_Region’ columns

df[['Country/Territory', 'UN\_Region']]

A screenshot of a cellphone

AI-generated content may be incorrect.

# Countries below average by IMF\_Estimate

gdp=df[df["IMF\_Estimate"]<df["IMF\_Estimate"].mean()]

gdp.iloc[:,1:4]

A screenshot of a computer

AI-generated content may be incorrect.

#Histogram

df.hist(figsize=(10,8))

plt.show()

A group of graphs with numbers and numbers

AI-generated content may be incorrect.

df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].hist(figsize=(12,9))

plt.show()

A graph of different numbers

AI-generated content may be incorrect.

df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].hist(bins=5, figsize=(12,9))

plt.show()

A graph of different sizes and numbers

AI-generated content may be incorrect.

df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].hist(bins=3, figsize=(12,9))

plt.show()

A graph of different types of data

AI-generated content may be incorrect.

# Correlation

df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].corr()

A screenshot of a graph

AI-generated content may be incorrect.

corr = df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].corr()

plt.figure(figsize=(9,6))

sns.heatmap(corr)

plt.show()

A screenshot of a color scheme

AI-generated content may be incorrect.

corr = df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].corr()

plt.figure(figsize=(9,6))

sns.heatmap(corr, annot=True)

plt.show()

A screenshot of a color chart

AI-generated content may be incorrect.

corr = df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].corr()

plt.figure(figsize=(9,6))

sns.heatmap(corr, annot=True, fmt=".2f", cmap = 'GnBu', annot\_kws={"size": 12})

plt.show()

A screenshot of a graph

AI-generated content may be incorrect.

corr = df[["IMF\_Estimate", "UN\_Estimate", "WorldBank\_Estimate"]].corr()

plt.figure(figsize=(9,6))

sns.heatmap(corr, annot=True, cmap = 'Purples')

plt.title("Correlation Map")

plt.show()

A screenshot of a computer screen

AI-generated content may be incorrect.

corr = df.select\_dtypes(include=[int, float]).corr()

plt.figure(figsize=(9,6))

sns.heatmap(corr, annot=True, cmap = 'Purples')

plt.show()

A screenshot of a graph

AI-generated content may be incorrect.

#Print the first 5 rows

df.head()

A screenshot of a computer

AI-generated content may be incorrect.

sns.barplot(x="UN\_Region", y="WorldBank\_Estimate", data=df, errorbar=None)

plt.show()

A graph of blue rectangular bars

AI-generated content may be incorrect.

sns.barplot(x="WorldBank\_Estimate", y="UN\_Region", data=df, errorbar=None)

plt.show()

A graph of a number of countries/regions

AI-generated content may be incorrect.

fig = plt.figure(figsize = (8,5))

ax = sns.barplot(x = "IMF\_Estimate",  y = "UN\_Region",

data = df, errorbar = None)

ax.bar\_label(ax.containers[0])

plt.show()

A graph with numbers and a bar

AI-generated content may be incorrect.

fig = plt.figure(figsize = (8,5))

ax = sns.barplot(x = "UN\_Region",  y = "IMF\_Estimate",

                 data = df, errorbar = None)

ax.bar\_label(ax.containers[0])

ax.set\_title("Regions by IMF Estimate")

plt.show()

A graph of blue bars with white text

AI-generated content may be incorrect.

#Scatter Plot

df.plot(x='UN\_Region', y='UN\_Estimate', kind='scatter',

        figsize=(10,6),

        title="Scatter Plot")

plt.show()

A graph with blue dots

AI-generated content may be incorrect.

#Boxplot

sns.boxplot(x=df["UN\_Estimate"])

plt.show()

A graph with a bar graph

AI-generated content may be incorrect.

df[df["UN\_Estimate"]>50000].head()

A screenshot of a computer

AI-generated content may be incorrect.

sns.boxplot(x=df["WorldBank\_Estimate"])

plt.show()

A graph of a bar graph

AI-generated content may be incorrect.

sns.boxplot(x=df["IMF\_Estimate"])

plt.show()

A graph with a bar and numbers

AI-generated content may be incorrect.

df[df["UN\_Estimate"]>100000]

A screenshot of a computer

AI-generated content may be incorrect.

#Create another dataframe called data excluding 5 countries with highest UN estimate

data = df[-(df["UN\_Estimate"]>100000)]

data.head()

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

df\_filtered = df[(df["UN\_Estimate"] < upper\_boundary) & (df["UN\_Estimate"] > lower\_boundary)]

df\_filtered.head()

A screenshot of a computer

AI-generated content may be incorrect.