**Data Analysis with Python**

**Duration: 2 hours**

**Data Inspection and Exploration**

1. **Load the Data:**

* Load the CSV file into a Pandas DataFrame.
* Data file : GDP (nominal) per Capita

1. **Display the First and Last 3 Rows:**

* Use the head() and tail() functions to display the first and last 3 rows of the DataFrame.

1. **Display the dimensions of the data set:**

* Use appropriate function to find out number rows and columns in the dataset

1. **Display Information about the DataFrame:**

* Use the info() function to display concise information about the DataFrame, including the data types and non-null values , record your observations

1. **Subset Using loc:**

* Use the loc indexer to create a subset of the DataFrame. Select specific rows and columns based on labels.

1. **Subset Using iloc:**

* Use the **iloc** indexer to create another subset of the DataFrame. Select specific rows and columns based on integer positions ( you choose any column you like )

1. **Conditional Subsetting:**

* Use conditional statements to create a subset of the DataFrame.
* **Subsetting based on specific conditions for numerical columns:**
  + Select rows where the IMF\_Estimate is greater than 100.
  + Select rows where the WorldBank\_Estimate is less than 50.
  + Select rows where the UN\_Estimate is between 75 and 150.

1. **Summary Statistics:**

* Generate summary statistics for numerical columns (IMF\_Estimate, WorldBank\_Estimate, UN\_Estimate).

1. **Analyse based on nonnumeric column:**

* Count the number of countries in each UN region.
* Create subsets for the following UN regions
  + Africa – Call this subset GDP\_Africa
  + Europe – Call this subset GDP\_Europe
  + Create summary statistics for the above two regions and compare. Write short paragraphs to describe your findings. Feel free to make any assumptions.

1. **Conclusion:**

Write a brief conclusion based on your observations. Mention any interesting patterns or trends you notice in the data. Discuss any further steps or questions that arose during the analysis.

**ANSWERS (Scripts attached)**

**Task 1:**

import pandas as pd

gdpdata =pd.read\_csv("GDP.csv")

gdpdata **# Task 1: load data**

**Task 2a:**

gdpdata.head(3) **# Task 2a: Display first 3 rows of the dataframe**

**Task 2b:**

gdpdata.tail(3) **# Task 2b: Display last 3 rows of the dataframe**

**Task 3:**

gdpdata.shape **# Task 3: Display dimensions of data set**

**Task 4:**

gdpdata.info() **# Task 4: Information of DataFrame**

**Task 5:**

topcountry = gdpdata.loc[1:10,'Country/Territory'] **# Task 5: Subset of top 10 Countries**

topcountry

**Task 6:**

topregion = gdpdata.iloc[1:10,2**] # Task 6: Subset of top 10 UN\_Region**

topregion

**Task 7a:**

imf=gdpdata[gdpdata['IMF\_Estimate']>100] **# Task 7a: IMF Estimate > 100**

imf

**Task 7b:**

wb=gdpdata[gdpdata['WorldBank\_Estimate']<50] **# Task 7b: WorldBank Estimate < 50**

wb

**Task 7c:**

un=gdpdata[(gdpdata['UN\_Estimate']>=750) & (gdpdata['UN\_Estimate']<=1000)] **# Task 7c: UN Estimate is between 75 and 100**

un

**Task 8a:**

gdpdata['IMF\_Estimate'].describe() **# Task 8a: Summary Statistic for IMF\_Estimate**

**Task 8b:**

gdpdata['WorldBank\_Estimate'].describe() **# Task 8b: Summary Statistic for WorldBank\_Estimate**

**Task 8c:**

gdpdata['UN\_Estimate'].describe() **# Task 8c: Summary Statistic for UN\_Estimate**

**Task 9a:**

countryperregion = gdpdata.groupby('UN\_Region')['Country/Territory'].nunique() **# Task 9a: Number of countries in each UN Region**

countryperregion

**Task 9b:**

africa = gdpdata[gdpdata['UN\_Region']=='Africa'] **# Task 9b: Create a subset Africa**

Africa

**Task 9c:**

europe = gdpdata[gdpdata['UN\_Region']=='Europe'] **# Task 9c: Create a subset Europe**

Europe

**Task 9d:**

africa.describe() **# Task 9d: Summary Statistic for Africa**

**Task 9e:**

europe.describe() **# Task 9e: Summary Statistic for Europe**

**Task 9f: Comments**

* + - Observation: 55 African Countries VS 48 European Countries
    - But the average GBP in Europe is roughly 12.3 times / 18.3 times / 16.8 times higher than in Africa according to IMF, WorldBank and UN Estimate.

**Task 10:**

1. **Find the statistic for America, Asia, Oceania to compare:**

america = gdpdata[gdpdata['UN\_Region']=='Americas']

america.describe()

asia = gdpdata[gdpdata['UN\_Region']=='Asia']

asia.describe()

oceania = gdpdata[gdpdata['UN\_Region']=='Oceania']

oceania.describe()

1. **Conclusion:**
   * + Based on the statistic, we have the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **IMF Estimate** | **WorldBank Estimate** | **UN Estimate** | **Number of Countries** |
| **Africa** | 2,802.35 | 2,470.84 | 2,417.93 | 55 |
| **Oceania** | 9,133.15 | 15,113.65 | 12,613.75 | 20 |
| **Americans** | 11,871.04 | 18,565.13 | 18,703.75 | 48 |
| **Asia** | 16,665.25 | 13,921.31 | 14,069.02 | 51 |
| **Europe** | 34,446.75 | 45,193.69 | 40,610.79 | 48 |