

Day_33_281123

January 23, 2024

1 Pandas

```
[97]: import pandas as pd
```

```
[98]: !gdown 1M82e0z0MJgV1ISQe3auBFnglYvMyZyBh
```

Downloading...

From: <https://drive.google.com/uc?id=1M82e0z0MJgV1ISQe3auBFnglYvMyZyBh>

To: C:\Data\Data_science\Data Science RIA\3 Python\Pandas\Codes\mckinsey (1).csv

```
0%|          | 0.00/83.8k [00:00<?, ?B/s]
100%|#####| 83.8k/83.8k [00:00<00:00, 41.6MB/s]
```

- Heterogenous data
- Visualization
- Manipulate the dataframe
- Complex analysis
- Easy

2 Importing data

```
[99]: df = pd.read_csv("mckinsey (1).csv")
df
```

```
[99]:
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

[1704 rows x 6 columns]

```
[100]: type(df)
```

```
[100]: pandas.core.frame.DataFrame
```

```
[101]: type(df[['country']])
```

```
[101]: pandas.core.frame.DataFrame
```

```
[102]: type(df['country'])
```

```
[102]: pandas.core.series.Series
```

```
[103]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   country         1704 non-null   object
1   year            1704 non-null   int64
2   population       1704 non-null   int64
3   continent        1704 non-null   object
4   life_exp        1704 non-null   float64
5   gdp_cap         1704 non-null   float64
dtypes: float64(2), int64(2), object(2)
memory usage: 80.0+ KB
```

Memory	
1 KB	1024B
1 MB	1024 KB
1 GB	1024 MB
1 TB	1024 GB

3 Getting top n records

```
[104]: df.head(4)
```

```
[104]:
```

	country	year	population	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138

4 Getting bottom n records

```
[105]: df.tail(4)
```

```
[105]:
```

	country	year	population	continent	life_exp	gdp_cap
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

```
[106]: df.shape
```

```
[106]: (1704, 6)
```

5 Creating a dataframe

Using dictionaries

```
[107]: new_df = pd.DataFrame(  
    {  
        "Name": ["Sai", "Sharan", "Bunny", "Shiva", "Sagar"],  
        "Age": [22, 27, 20, 29, 25],  
        "City": ["HYD", "BEN", "KNL", "BEN", "BEN"],  
        "Phno": [72880, 94954, 98981, 88771, 90890]  
    }  
)
```

```
[108]: new_df
```

```
[108]:
```

	Name	Age	City	Phno
0	Sai	22	HYD	72880
1	Sharan	27	BEN	94954
2	Bunny	20	KNL	98981
3	Shiva	29	BEN	88771
4	Sagar	25	BEN	90890

```
[109]: new_df.shape
```

```
[109]: (5, 4)
```

```
[110]: new_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 5 entries, 0 to 4  
Data columns (total 4 columns):  
#   Column  Non-Null Count  Dtype  
---  ---  
0   Name    5 non-null      object  
1   Age     5 non-null      int64
```

```

2   City      5 non-null      object
3   Phno      5 non-null      int64
dtypes: int64(2), object(2)
memory usage: 292.0+ bytes

```

5.0.1 Using Lists

```
[111]: new_df2 = pd.DataFrame([['Sai',22,'HYD',9900],
                               ['Ria',23,'MAR',7788],
                               ['Mahesh',45,'HYD',4005]],
                              columns=['Name','Age','City','Phno'])
```

```
[112]: new_df2
```

```
[112]:
```

	Name	Age	City	Phno
0	Sai	22	HYD	9900
1	Ria	23	MAR	7788
2	Mahesh	45	HYD	4005

```
[113]: new_df2[['City','Phno','Name','Age']]
```

```
[113]:
```

	City	Phno	Name	Age
0	HYD	9900	Sai	22
1	MAR	7788	Ria	23
2	HYD	4005	Mahesh	45

5.1 Unique values

```
[114]: df.country.unique()
```

```
[114]: array(['Afghanistan', 'Albania', 'Algeria', 'Angola', 'Argentina',
              'Australia', 'Austria', 'Bahrain', 'Bangladesh', 'Belgium',
              'Benin', 'Bolivia', 'Bosnia and Herzegovina', 'Botswana', 'Brazil',
              'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon',
              'Canada', 'Central African Republic', 'Chad', 'Chile', 'China',
              'Colombia', 'Comoros', 'Congo, Dem. Rep.', 'Congo, Rep.',
              'Costa Rica', 'Cote d'Ivoire', 'Croatia', 'Cuba', 'Czech Republic',
              'Denmark', 'Djibouti', 'Dominican Republic', 'Ecuador', 'Egypt',
              'El Salvador', 'Equatorial Guinea', 'Eritrea', 'Ethiopia',
              'Finland', 'France', 'Gabon', 'Gambia', 'Germany', 'Ghana',
              'Greece', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Haiti',
              'Honduras', 'Hong Kong, China', 'Hungary', 'Iceland', 'India',
              'Indonesia', 'Iran', 'Iraq', 'Ireland', 'Israel', 'Italy',
              'Jamaica', 'Japan', 'Jordan', 'Kenya', 'Korea, Dem. Rep.',
              'Korea, Rep.', 'Kuwait', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
              'Madagascar', 'Malawi', 'Malaysia', 'Mali', 'Mauritania',
              'Mauritius', 'Mexico', 'Mongolia', 'Montenegro', 'Morocco',
              'Mozambique', 'Myanmar', 'Namibia', 'Nepal', 'Netherlands',
```

```
'New Zealand', 'Nicaragua', 'Niger', 'Nigeria', 'Norway', 'Oman',
'Pakistan', 'Panama', 'Paraguay', 'Peru', 'Philippines', 'Poland',
'Portugal', 'Puerto Rico', 'Reunion', 'Romania', 'Rwanda',
'Sao Tome and Principe', 'Saudi Arabia', 'Senegal', 'Serbia',
'Sierra Leone', 'Singapore', 'Slovak Republic', 'Slovenia',
'Somalia', 'South Africa', 'Spain', 'Sri Lanka', 'Sudan',
'Swaziland', 'Sweden', 'Switzerland', 'Syria', 'Taiwan',
'Tanzania', 'Thailand', 'Togo', 'Trinidad and Tobago', 'Tunisia',
'Turkey', 'Uganda', 'United Kingdom', 'United States', 'Uruguay',
'Venezuela', 'Vietnam', 'West Bank and Gaza', 'Yemen, Rep.',
'Zambia', 'Zimbabwe'], dtype=object)
```

5.2 Get the count

```
[115]: df['continent'].value_counts()
```

```
[115]: continent
Africa      624
Asia        396
Europe      360
Americas    300
Oceania      24
Name: count, dtype: int64
```

5.3 Rename the column

```
[116]: df.rename(
    {
        'country': 'COUNTRY',
        'population': 'POPULATION'
    }, axis=1, inplace=True #Inplace will save the changes to original dataframe
)
```

```
[117]: df
```

```
[117]:
```

	COUNTRY	year	POPULATION	continent	life_exp	gdp_cap
0	Afghanistan	1952	8425333	Asia	28.801	779.445314
1	Afghanistan	1957	9240934	Asia	30.332	820.853030
2	Afghanistan	1962	10267083	Asia	31.997	853.100710
3	Afghanistan	1967	11537966	Asia	34.020	836.197138
4	Afghanistan	1972	13079460	Asia	36.088	739.981106
...
1699	Zimbabwe	1987	9216418	Africa	62.351	706.157306
1700	Zimbabwe	1992	10704340	Africa	60.377	693.420786
1701	Zimbabwe	1997	11404948	Africa	46.809	792.449960
1702	Zimbabwe	2002	11926563	Africa	39.989	672.038623
1703	Zimbabwe	2007	12311143	Africa	43.487	469.709298

[1704 rows x 6 columns]

```
[118]: df_1 = df.T
```

```
[119]: df_1.head()
```

```
[119]:
```

	0	1	2	3	4	\
COUNTRY	Afghanistan	Afghanistan	Afghanistan	Afghanistan	Afghanistan	
year	1952	1957	1962	1967	1972	
POPULATION	8425333	9240934	10267083	11537966	13079460	
continent	Asia	Asia	Asia	Asia	Asia	
life_exp	28.801	30.332	31.997	34.02	36.088	

	5	6	7	8	9	\
COUNTRY	Afghanistan	Afghanistan	Afghanistan	Afghanistan	Afghanistan	
year	1977	1982	1987	1992	1997	
POPULATION	14880372	12881816	13867957	16317921	22227415	
continent	Asia	Asia	Asia	Asia	Asia	
life_exp	38.438	39.854	40.822	41.674	41.763	

	...	1694	1695	1696	1697	1698	1699	\
COUNTRY	...	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	
year	...	1962	1967	1972	1977	1982	1987	
POPULATION	...	4277736	4995432	5861135	6642107	7636524	9216418	
continent	...	Africa	Africa	Africa	Africa	Africa	Africa	
life_exp	...	52.358	53.995	55.635	57.674	60.363	62.351	

	1700	1701	1702	1703
COUNTRY	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe
year	1992	1997	2002	2007
POPULATION	10704340	11404948	11926563	12311143
continent	Africa	Africa	Africa	Africa
life_exp	60.377	46.809	39.989	43.487

[5 rows x 1704 columns]

```
[120]: df_1.rename(  
    {  
        'country': 'COUNTRY',  
        'population': 'POPULATION'  
    }  
)
```

```
[120]:
```

	0	1	2	3	4	\
COUNTRY	Afghanistan	Afghanistan	Afghanistan	Afghanistan	Afghanistan	
year	1952	1957	1962	1967	1972	

POPULATION	8425333	9240934	10267083	11537966	13079460
continent	Asia	Asia	Asia	Asia	Asia
life_exp	28.801	30.332	31.997	34.02	36.088
gdp_cap	779.445314	820.85303	853.10071	836.197138	739.981106

	5	6	7	8	9	\
COUNTRY	Afghanistan	Afghanistan	Afghanistan	Afghanistan	Afghanistan	
year	1977	1982	1987	1992	1997	
POPULATION	14880372	12881816	13867957	16317921	22227415	
continent	Asia	Asia	Asia	Asia	Asia	
life_exp	38.438	39.854	40.822	41.674	41.763	
gdp_cap	786.11336	978.011439	852.395945	649.341395	635.341351	

	...	1694	1695	1696	1697	1698	\
COUNTRY	...	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	
year	...	1962	1967	1972	1977	1982	
POPULATION	...	4277736	4995432	5861135	6642107	7636524	
continent	...	Africa	Africa	Africa	Africa	Africa	
life_exp	...	52.358	53.995	55.635	57.674	60.363	
gdp_cap	...	527.272182	569.795071	799.362176	685.587682	788.855041	

		1699	1700	1701	1702	1703
COUNTRY		Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe	Zimbabwe
year		1987	1992	1997	2002	2007
POPULATION		9216418	10704340	11404948	11926563	12311143
continent		Africa	Africa	Africa	Africa	Africa
life_exp		62.351	60.377	46.809	39.989	43.487
gdp_cap		706.157306	693.420786	792.44996	672.038623	469.709298

[6 rows x 1704 columns]

5.4 Deleting a column

```
[121]: df.drop('continent',axis=1,inplace=True)
```

```
[122]: df
```

```
[122]:
```

	COUNTRY	year	POPULATION	life_exp	gdp_cap
0	Afghanistan	1952	8425333	28.801	779.445314
1	Afghanistan	1957	9240934	30.332	820.853030
2	Afghanistan	1962	10267083	31.997	853.100710
3	Afghanistan	1967	11537966	34.020	836.197138
4	Afghanistan	1972	13079460	36.088	739.981106
...
1699	Zimbabwe	1987	9216418	62.351	706.157306
1700	Zimbabwe	1992	10704340	60.377	693.420786
1701	Zimbabwe	1997	11404948	46.809	792.449960

1702	Zimbabwe	2002	11926563	39.989	672.038623
1703	Zimbabwe	2007	12311143	43.487	469.709298

[1704 rows x 5 columns]

```
[123]: df.drop(columns=['year', 'life_exp'], inplace=True)
```

```
[124]: df
```

```
[124]:
```

	COUNTRY	POPULATION	gdp_cap
0	Afghanistan	8425333	779.445314
1	Afghanistan	9240934	820.853030
2	Afghanistan	10267083	853.100710
3	Afghanistan	11537966	836.197138
4	Afghanistan	13079460	739.981106
...
1699	Zimbabwe	9216418	706.157306
1700	Zimbabwe	10704340	693.420786
1701	Zimbabwe	11404948	792.449960
1702	Zimbabwe	11926563	672.038623
1703	Zimbabwe	12311143	469.709298

[1704 rows x 3 columns]

5.5 Adding a new column into data frame

```
[125]: df['gdp_cap_per'] = df['gdp_cap']/100
```

```
[126]: df
```

```
[126]:
```

	COUNTRY	POPULATION	gdp_cap	gdp_cap_per
0	Afghanistan	8425333	779.445314	7.794453
1	Afghanistan	9240934	820.853030	8.208530
2	Afghanistan	10267083	853.100710	8.531007
3	Afghanistan	11537966	836.197138	8.361971
4	Afghanistan	13079460	739.981106	7.399811
...
1699	Zimbabwe	9216418	706.157306	7.061573
1700	Zimbabwe	10704340	693.420786	6.934208
1701	Zimbabwe	11404948	792.449960	7.924500
1702	Zimbabwe	11926563	672.038623	6.720386
1703	Zimbabwe	12311143	469.709298	4.697093

[1704 rows x 4 columns]