

session-3

September 2, 2024

1 Session_3

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

```
[3]: regions = pd.read_csv('Regions.csv')
```

```
[4]: regions.head(10)
```

```
[4]:
```

	Country	Region
0	Afghanistan	Southern Asia
1	Albania	Central and Eastern Europe
2	Algeria	Middle East and Northern Africa
3	Angola	Sub-Saharan Africa
4	Argentina	Latin America and Caribbean
5	Armenia	Central and Eastern Europe
6	Australia	Australia and New Zealand
7	Austria	Western Europe
8	Azerbaijan	Central and Eastern Europe
9	Bahrain	Middle East and Northern Africa

```
[5]: part_1_2015 = pd.read_csv('World_Happiness_2015_Part1.csv')
part_2_2015 = pd.read_csv('World_Happiness_2015_Part2.csv')
part_3_2015 = pd.read_csv('World_Happiness_2015_Part3.csv')
```

```
[6]: part_1_2015.shape
```

```
[6]: (50, 11)
```

2 Concatenation

```
[7]: 'yuaan' + 'hussain'
```

```
[7]: 'yuaanhussain'
```

```
[8]: part_2_2015.shape
```

[8]: (50, 11)

```
[9]: part_3_2015.shape
```

[9]: (58, 11)

```
[10]: #no of coloumns and rows and column name must similar at a time of concatenation
pd.concat([part_1_2015, part_2_2015])
```

```
[10]:
```

	Country	Happiness Rank	Happiness Score	Standard Error	\
0	Switzerland	1	7.587	0.03411	
1	Iceland	2	7.561	0.04884	
2	Denmark	3	7.527	0.03328	
3	Norway	4	7.522	0.03880	
4	Canada	5	7.427	0.03553	
..	
45	Bosnia and Herzegovina	96	4.949	0.06913	
46	Lesotho	97	4.898	0.09438	
47	Dominican Republic	98	4.885	0.07446	
48	Laos	99	4.876	0.06698	
49	Mongolia	100	4.874	0.03313	

	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	\
0	1.39651	1.34951	0.94143	0.66557	
1	1.30232	1.40223	0.94784	0.62877	
2	1.32548	1.36058	0.87464	0.64938	
3	1.45900	1.33095	0.88521	0.66973	
4	1.32629	1.32261	0.90563	0.63297	
..	
45	0.83223	0.91916	0.79081	0.09245	
46	0.37545	1.04103	0.07612	0.31767	
47	0.89537	1.17202	0.66825	0.57672	
48	0.59066	0.73803	0.54909	0.59591	
49	0.82819	1.30060	0.60268	0.43626	

	Trust (Government Corruption)	Generosity	Dystopia Residual
0	0.41978	0.29678	2.51738
1	0.14145	0.43630	2.70201
2	0.48357	0.34139	2.49204
3	0.36503	0.34699	2.46531
4	0.32957	0.45811	2.45176
..
45	0.00227	0.24808	2.06367
46	0.12504	0.16388	2.79832
47	0.14234	0.21684	1.21305
48	0.24249	0.42192	1.73799
49	0.02666	0.33230	1.34759

[100 rows x 11 columns]

```
[11]: pd.concat([part_1_2015, part_2_2015], keys=['part_1', 'part_2'])
```

```
[11]:
```

		Country	Happiness Rank	Happiness Score	\
part_1	0	Switzerland	1	7.587	
	1	Iceland	2	7.561	
	2	Denmark	3	7.527	
	3	Norway	4	7.522	
	4	Canada	5	7.427	
...		
part_2	45	Bosnia and Herzegovina	96	4.949	
	46	Lesotho	97	4.898	
	47	Dominican Republic	98	4.885	
	48	Laos	99	4.876	
	49	Mongolia	100	4.874	

		Standard Error	Economy (GDP per Capita)	Family	\
part_1	0	0.03411	1.39651	1.34951	
	1	0.04884	1.30232	1.40223	
	2	0.03328	1.32548	1.36058	
	3	0.03880	1.45900	1.33095	
	4	0.03553	1.32629	1.32261	
...		
part_2	45	0.06913	0.83223	0.91916	
	46	0.09438	0.37545	1.04103	
	47	0.07446	0.89537	1.17202	
	48	0.06698	0.59066	0.73803	
	49	0.03313	0.82819	1.30060	

		Health (Life Expectancy)	Freedom	Trust (Government Corruption)	\
part_1	0	0.94143	0.66557		0.41978
	1	0.94784	0.62877		0.14145
	2	0.87464	0.64938		0.48357
	3	0.88521	0.66973		0.36503
	4	0.90563	0.63297		0.32957
...		
part_2	45	0.79081	0.09245		0.00227
	46	0.07612	0.31767		0.12504
	47	0.66825	0.57672		0.14234
	48	0.54909	0.59591		0.24249
	49	0.60268	0.43626		0.02666

		Generosity	Dystopia Residual
part_1	0	0.29678	2.51738
	1	0.43630	2.70201

	2	0.34139	2.49204
	3	0.34699	2.46531
	4	0.45811	2.45176
...	
part_2	45	0.24808	2.06367
	46	0.16388	2.79832
	47	0.21684	1.21305
	48	0.42192	1.73799
	49	0.33230	1.34759

[100 rows x 11 columns]

```
[12]: pd.concat([part_1_2015, part_2_2015], ignore_index=True)
```

```
[12]:
```

	Country	Happiness Rank	Happiness Score	Standard Error	\
0	Switzerland	1	7.587	0.03411	
1	Iceland	2	7.561	0.04884	
2	Denmark	3	7.527	0.03328	
3	Norway	4	7.522	0.03880	
4	Canada	5	7.427	0.03553	
..	
95	Bosnia and Herzegovina	96	4.949	0.06913	
96	Lesotho	97	4.898	0.09438	
97	Dominican Republic	98	4.885	0.07446	
98	Laos	99	4.876	0.06698	
99	Mongolia	100	4.874	0.03313	

	Economy (GDP per Capita)	Family Health (Life Expectancy)	Freedom	\
0	1.39651	1.34951	0.94143	0.66557
1	1.30232	1.40223	0.94784	0.62877
2	1.32548	1.36058	0.87464	0.64938
3	1.45900	1.33095	0.88521	0.66973
4	1.32629	1.32261	0.90563	0.63297
..
95	0.83223	0.91916	0.79081	0.09245
96	0.37545	1.04103	0.07612	0.31767
97	0.89537	1.17202	0.66825	0.57672
98	0.59066	0.73803	0.54909	0.59591
99	0.82819	1.30060	0.60268	0.43626

	Trust (Government Corruption)	Generosity	Dystopia Residual
0	0.41978	0.29678	2.51738
1	0.14145	0.43630	2.70201
2	0.48357	0.34139	2.49204
3	0.36503	0.34699	2.46531
4	0.32957	0.45811	2.45176
..

95	0.00227	0.24808	2.06367
96	0.12504	0.16388	2.79832
97	0.14234	0.21684	1.21305
98	0.24249	0.42192	1.73799
99	0.02666	0.33230	1.34759

[100 rows x 11 columns]

3 Skill Test - 1

Concatenate all three files together

```
[13]: dataset = pd.concat([part_1_2015, part_2_2015, part_3_2015], ignore_index=True)
```

4 Skill Test - 2 - Data Analysis

Which regions has the maximum average Happiness Score

```
[14]: regions.head()
```

```
[14]:
```

	Country	Region
0	Afghanistan	Southern Asia
1	Albania	Central and Eastern Europe
2	Algeria	Middle East and Northern Africa
3	Angola	Sub-Saharan Africa
4	Argentina	Latin America and Caribbean

```
[15]: dataset.head()
```

```
[15]:
```

	Country	Happiness Rank	Happiness Score	Standard Error	\
0	Switzerland	1	7.587	0.03411	
1	Iceland	2	7.561	0.04884	
2	Denmark	3	7.527	0.03328	
3	Norway	4	7.522	0.03880	
4	Canada	5	7.427	0.03553	

	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	\
0	1.39651	1.34951		0.94143	0.66557
1	1.30232	1.40223		0.94784	0.62877
2	1.32548	1.36058		0.87464	0.64938
3	1.45900	1.33095		0.88521	0.66973
4	1.32629	1.32261		0.90563	0.63297

	Trust (Government Corruption)	Generosity	Dystopia	Residual
0	0.41978	0.29678		2.51738
1	0.14145	0.43630		2.70201

2	0.48357	0.34139	2.49204
3	0.36503	0.34699	2.46531
4	0.32957	0.45811	2.45176

```
[16]: #axis = 0 -> by row
#axis = 1 -> by column
pd.concat([regions, dataset], axis=1)
```

```
[16]:      Country      Region      Country \
0  Afghanistan  Southern Asia  Switzerland
1    Albania  Central and Eastern Europe  Iceland
2    Algeria  Middle East and Northern Africa  Denmark
3    Angola  Sub-Saharan Africa  Norway
4  Argentina  Latin America and Caribbean  Canada
..      ...
153  Vietnam  Southeastern Asia  Rwanda
154    Yemen  Middle East and Northern Africa  Benin
155    Zambia  Sub-Saharan Africa  Syria
156  Zimbabwe  Sub-Saharan Africa  Burundi
157      NaN  NaN  Togo
```

	Happiness Rank	Happiness Score	Standard Error	\
0	1	7.587	0.03411	
1	2	7.561	0.04884	
2	3	7.527	0.03328	
3	4	7.522	0.03880	
4	5	7.427	0.03553	
..	
153	154	3.465	0.03464	
154	155	3.340	0.03656	
155	156	3.006	0.05015	
156	157	2.905	0.08658	
157	158	2.839	0.06727	

	Economy (GDP per Capita)	Family Health (Life Expectancy)	Freedom	\
0	1.39651	1.34951	0.94143	0.66557
1	1.30232	1.40223	0.94784	0.62877
2	1.32548	1.36058	0.87464	0.64938
3	1.45900	1.33095	0.88521	0.66973
4	1.32629	1.32261	0.90563	0.63297
..
153	0.22208	0.77370	0.42864	0.59201
154	0.28665	0.35386	0.31910	0.48450
155	0.66320	0.47489	0.72193	0.15684
156	0.01530	0.41587	0.22396	0.11850
157	0.20868	0.13995	0.28443	0.36453

	Trust (Government Corruption)	Generosity	Dystopia Residual
0	0.41978	0.29678	2.51738
1	0.14145	0.43630	2.70201
2	0.48357	0.34139	2.49204
3	0.36503	0.34699	2.46531
4	0.32957	0.45811	2.45176
..
153	0.55191	0.22628	0.67042
154	0.08010	0.18260	1.63328
155	0.18906	0.47179	0.32858
156	0.10062	0.19727	1.83302
157	0.10731	0.16681	1.56726

[158 rows x 13 columns]

5 Data Merging

```
[17]: inner_join = regions.merge(dataset, on='Country', how='inner')
```

```
[18]: inner_join.head()
```

```
[18]:
```

	Country	Region	Happiness Rank	\
0	Afghanistan	Southern Asia	153	
1	Albania	Central and Eastern Europe	95	
2	Algeria	Middle East and Northern Africa	68	
3	Angola	Sub-Saharan Africa	137	
4	Argentina	Latin America and Caribbean	30	

	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	\
0	3.575	0.03084	0.31982	0.30285	
1	4.959	0.05013	0.87867	0.80434	
2	5.605	0.05099	0.93929	1.07772	
3	4.033	0.04758	0.75778	0.86040	
4	6.574	0.04612	1.05351	1.24823	

	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	\
0	0.30335	0.23414	0.09719	
1	0.81325	0.35733	0.06413	
2	0.61766	0.28579	0.17383	
3	0.16683	0.10384	0.07122	
4	0.78723	0.44974	0.08484	

	Generosity	Dystopia Residual
0	0.36510	1.95210
1	0.14272	1.89894
2	0.07822	2.43209

```
3      0.12344      1.94939
4      0.11451      2.83600
```

```
[19]: inner_join.shape
```

```
[19]: (151, 12)
```

```
[20]: regions.describe()
```

```
[20]:
```

	Country	Region
count	157	157
unique	157	10
top	Afghanistan	Sub-Saharan Africa
freq	1	38

```
[34]: inner_join[inner_join.Region == 'Southern Asia']['Happiness Score'].mean()
```

```
[34]: 4.580857142857142
```

6 Group by

```
[36]: groupby_region = inner_join.groupby('Region')
```

```
[23]: groupby_region.get_group('Southern Asia')['Happiness Score'].mean()
```

```
[23]: 4.580857142857142
```

```
[24]: groupby_region['Happiness Score'].mean()
```

```
[24]: Region
Australia and New Zealand      7.285000
Central and Eastern Europe      5.332931
Eastern Asia                    5.626167
Latin America and Caribbean     6.144682
Middle East and Northern Africa  5.330789
North America                   7.273000
Southeastern Asia               5.317444
Southern Asia                    4.580857
Sub-Saharan Africa              4.125647
Western Europe                  6.689619
Name: Happiness Score, dtype: float64
```

```
[25]: max(groupby_region['Happiness Score'].mean())
```

```
[25]: 7.285
```



```
[26]: min(groupby_region['Happiness Score'].mean())
```

```
[26]: 4.125647058823529
```

```
[27]: groupby_region['Happiness Score'].mean().sort_values()
```

```
[27]: Region
Sub-Saharan Africa          4.125647
Southern Asia               4.580857
Southeastern Asia          5.317444
Middle East and Northern Africa 5.330789
Central and Eastern Europe  5.332931
Eastern Asia                5.626167
Latin America and Caribbean 6.144682
Western Europe              6.689619
North America               7.273000
Australia and New Zealand   7.285000
Name: Happiness Score, dtype: float64
```

```
[28]: groupby_region['Happiness Score'].mean().sort_values(ascending=False)
```

```
[28]: Region
Australia and New Zealand   7.285000
North America               7.273000
Western Europe              6.689619
Latin America and Caribbean 6.144682
Eastern Asia                5.626167
Central and Eastern Europe  5.332931
Middle East and Northern Africa 5.330789
Southeastern Asia          5.317444
Southern Asia               4.580857
Sub-Saharan Africa          4.125647
Name: Happiness Score, dtype: float64
```

```
[29]: groupby_region['Happiness Score'].mean().sort_values(ascending=False).head(1)
↳ #Ans of Skill Test-2
```

```
[29]: Region
Australia and New Zealand   7.285
Name: Happiness Score, dtype: float64
```

```
[30]: dataset.describe()
```

```
[30]:
```

	Happiness Rank	Happiness Score	Standard Error	\
count	158.000000	158.000000	158.000000	
mean	79.493671	5.375734	0.047885	
std	45.754363	1.145010	0.017146	

min	1.000000	2.839000	0.018480
25%	40.250000	4.526000	0.037268
50%	79.500000	5.232500	0.043940
75%	118.750000	6.243750	0.052300
max	158.000000	7.587000	0.136930

	Economy (GDP per Capita)	Family	Health (Life Expectancy) \
count	158.000000	158.000000	158.000000
mean	0.846137	0.991046	0.630259
std	0.403121	0.272369	0.247078
min	0.000000	0.000000	0.000000
25%	0.545808	0.856823	0.439185
50%	0.910245	1.029510	0.696705
75%	1.158448	1.214405	0.811013
max	1.690420	1.402230	1.025250

	Freedom	Trust (Government Corruption)	Generosity \
count	158.000000	158.000000	158.000000
mean	0.428615	0.143422	0.237296
std	0.150693	0.120034	0.126685
min	0.000000	0.000000	0.000000
25%	0.328330	0.061675	0.150553
50%	0.435515	0.107220	0.216130
75%	0.549092	0.180255	0.309883
max	0.669730	0.551910	0.795880

	Dystopia Residual
count	158.000000
mean	2.098977
std	0.553550
min	0.328580
25%	1.759410
50%	2.095415
75%	2.462415
max	3.602140

```
[31]: groupby_region['Happiness Score'].agg(['mean', 'sum', 'count'])
```

	mean	sum	count
Region			
Australia and New Zealand	7.285000	14.570	2
Central and Eastern Europe	5.332931	154.655	29
Eastern Asia	5.626167	33.757	6
Latin America and Caribbean	6.144682	135.183	22
Middle East and Northern Africa	5.330789	101.285	19
North America	7.273000	14.546	2
Southeastern Asia	5.317444	47.857	9

Southern Asia	4.580857	32.066	7
Sub-Saharan Africa	4.125647	140.272	34
Western Europe	6.689619	140.482	21

```
[32]: inner_join.Region.unique()
```

```
[32]: array(['Southern Asia', 'Central and Eastern Europe',
        'Middle East and Northern Africa', 'Sub-Saharan Africa',
        'Latin America and Caribbean', 'Australia and New Zealand',
        'Western Europe', 'Southeastern Asia', 'North America',
        'Eastern Asia'], dtype=object)
```

```
[37]: left_join = regions.merge(dataset, on = 'Country', how = 'left')
```

```
[38]: left_join.shape
```

```
[38]: (157, 12)
```

```
[39]: regions.shape
```

```
[39]: (157, 2)
```

```
[40]: left_join[left_join['Happiness Score'].isnull()]
```

```
[40]:
```

	Country	Region	Happiness	Rank \
13	Belize	Latin America and Caribbean		NaN
95	Namibia	Sub-Saharan Africa		NaN
112	Puerto Rico	Latin America and Caribbean		NaN
124	Somalia	Sub-Saharan Africa		NaN
125	Somaliland Region	Sub-Saharan Africa		NaN
128	South Sudan	Sub-Saharan Africa		NaN

	Happiness Score	Standard Error	Economy (GDP per Capita)	Family \
13	NaN	NaN	NaN	NaN
95	NaN	NaN	NaN	NaN
112	NaN	NaN	NaN	NaN
124	NaN	NaN	NaN	NaN
125	NaN	NaN	NaN	NaN
128	NaN	NaN	NaN	NaN

	Health (Life Expectancy)	Freedom	Trust (Government Corruption) \
13	NaN	NaN	NaN
95	NaN	NaN	NaN
112	NaN	NaN	NaN
124	NaN	NaN	NaN
125	NaN	NaN	NaN
128	NaN	NaN	NaN

	Generosity	Dystopia	Residual
13	NaN		NaN
95	NaN		NaN
112	NaN		NaN
124	NaN		NaN
125	NaN		NaN
128	NaN		NaN

```
[41]: right_join = regions.merge(dataset, on = 'Country', how = 'right')
```

```
[42]: right_join.shape
```

```
[42]: (158, 12)
```

```
[43]: dataset.shape
```

```
[43]: (158, 11)
```

```
[44]: right_join[right_join['Region'].isnull()]
```

```
[44]:
```

	Country	Region	Happiness Rank	Happiness Score	\
21	Oman	NaN	22	6.853	
90	Somaliland region	NaN	91	5.057	
93	Mozambique	NaN	94	4.971	
96	Lesotho	NaN	97	4.898	
100	Swaziland	NaN	101	4.867	
125	Djibouti	NaN	126	4.369	
147	Central African Republic	NaN	148	3.678	

	Standard Error	Economy (GDP per Capita)	Family	\
21	0.05335	1.36011	1.08182	
90	0.06161	0.18847	0.95152	
93	0.07896	0.08308	1.02626	
96	0.09438	0.37545	1.04103	
100	0.08742	0.71206	1.07284	
125	0.08096	0.44025	0.59207	
147	0.06112	0.07850	0.00000	

	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	\
21	0.76276	0.63274		0.32524
90	0.43873	0.46582		0.39928
93	0.09131	0.34037		0.15603
96	0.07612	0.31767		0.12504
100	0.07566	0.30658		0.03060
125	0.36291	0.46074		0.28105
147	0.06699	0.48879		0.08289

	Generosity	Dystopia	Residual
21	0.21542		2.47489
90	0.50318		2.11032
93	0.22269		3.05137
96	0.16388		2.79832
100	0.18259		2.48676
125	0.18093		2.05125
147	0.23835		2.72230