Basic Analysis using Numpy and Pandas

2015 Dataset

To import library

In [1]:
```
import numpy as np
```

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

To import dataset

In [3]:
```
data=pd.read_csv(r"C:\Users\user\Downloads\2015.csv")
data
```

Out[3]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fre
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.6
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.6
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.6
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.6
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.6
...	...	...	...	...	...	...	...	...	
153	Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.5
154	Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.4
155	Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193	0.1
156	Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396	0.1
157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.3

158 rows × 12 columns

To get Top 20 record

In [4]:
```
data.head(20)
```

Out[4]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fr
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0
5	Finland	Western Europe	6	7.406	0.03140	1.29025	1.31826	0.88911	0
6	Netherlands	Western Europe	7	7.378	0.02799	1.32944	1.28017	0.89284	0
7	Sweden	Western Europe	8	7.364	0.03157	1.33171	1.28907	0.91087	0
8	New Zealand	Australia and New Zealand	9	7.286	0.03371	1.25018	1.31967	0.90837	0
9	Australia	Australia and New Zealand	10	7.284	0.04083	1.33358	1.30923	0.93156	0
10	Israel	Middle East and Northern Africa	11	7.278	0.03470	1.22857	1.22393	0.91387	0
11	Costa Rica	Latin America and Caribbean	12	7.226	0.04454	0.95578	1.23788	0.86027	0
12	Austria	Western Europe	13	7.200	0.03751	1.33723	1.29704	0.89042	0
13	Mexico	Latin America and Caribbean	14	7.187	0.04176	1.02054	0.91451	0.81444	0
14	United States	North America	15	7.119	0.03839	1.39451	1.24711	0.86179	0
15	Brazil	Latin America and Caribbean	16	6.983	0.04076	0.98124	1.23287	0.69702	0
16	Luxembourg	Western Europe	17	6.946	0.03499	1.56391	1.21963	0.91894	0
17	Ireland	Western Europe	18	6.940	0.03676	1.33596	1.36948	0.89533	0
18	Belgium	Western Europe	19	6.937	0.03595	1.30782	1.28566	0.89667	0

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fr
19	United Arab Emirates	Middle East and Northern Africa	20	6.901	0.03729	1.42727	1.12575	0.80925	0

To get last 20 record

In [5]:
```
data.tail(20)
```

Out[5]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)
138	Congo (Brazzaville)	Sub-Saharan Africa	139	3.989	0.06682	0.67866	0.66290	0.31051
139	Comoros	Sub-Saharan Africa	140	3.956	0.04797	0.23906	0.79273	0.36315
140	Uganda	Sub-Saharan Africa	141	3.931	0.04317	0.21102	1.13299	0.33861
141	Senegal	Sub-Saharan Africa	142	3.904	0.03608	0.36498	0.97619	0.43540
142	Gabon	Sub-Saharan Africa	143	3.896	0.04547	1.06024	0.90528	0.43372
143	Niger	Sub-Saharan Africa	144	3.845	0.03602	0.06940	0.77265	0.29707
144	Cambodia	Southeastern Asia	145	3.819	0.05069	0.46038	0.62736	0.61114
145	Tanzania	Sub-Saharan Africa	146	3.781	0.05061	0.28520	1.00268	0.38215
146	Madagascar	Sub-Saharan Africa	147	3.681	0.03633	0.20824	0.66801	0.46721
147	Central African Republic	Sub-Saharan Africa	148	3.678	0.06112	0.07850	0.00000	0.06699
148	Chad	Sub-Saharan Africa	149	3.667	0.03830	0.34193	0.76062	0.15010
149	Guinea	Sub-Saharan Africa	150	3.656	0.03590	0.17417	0.46475	0.24009
150	Ivory Coast	Sub-Saharan Africa	151	3.655	0.05141	0.46534	0.77115	0.15185
151	Burkina Faso	Sub-Saharan Africa	152	3.587	0.04324	0.25812	0.85188	0.27125
152	Afghanistan	Southern Asia	153	3.575	0.03084	0.31982	0.30285	0.30335
153	Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864
154	Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910
155	Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193
156	Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396
157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443

Statistical Analysis

In [6]:
```
data.describe()
```

Out[6]:

	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	(Go C
count	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	158.000000	1
mean	79.493671	5.375734	0.047885	0.846137	0.991046	0.630259	0.428615	
std	45.754363	1.145010	0.017146	0.403121	0.272369	0.247078	0.150693	
min	1.000000	2.839000	0.018480	0.000000	0.000000	0.000000	0.000000	
25%	40.250000	4.526000	0.037268	0.545808	0.856823	0.439185	0.328330	
50%	79.500000	5.232500	0.043940	0.910245	1.029510	0.696705	0.435515	
75%	118.750000	6.243750	0.052300	1.158448	1.214405	0.811013	0.549092	
max	158.000000	7.587000	0.136930	1.690420	1.402230	1.025250	0.669730	

To get row and column

In [7]:
```
print(np.shape(data))
```

```
(158, 12)
```

Find Number of Elements

In [8]:
```
np.size(data)
```

Out[8]: 1896

Find Missing values

In [9]:
```
data.isna()
```

Out[9]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	..
153	False	False	False	False	False	False	False	False	False
154	False	False	False	False	False	False	False	False	False
155	False	False	False	False	False	False	False	False	False
156	False	False	False	False	False	False	False	False	False
157	False	False	False	False	False	False	False	False	False

158 rows × 12 columns

To drop the missing values

In [10]:
```
data.dropna(axis=1,how="any")
```

Out[10]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Fre
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.6
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.6
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.6
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.6
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.6
...	...	...	...	...	...	...	...	...	
153	Rwanda	Sub-Saharan Africa	154	3.465	0.03464	0.22208	0.77370	0.42864	0.5
154	Benin	Sub-Saharan Africa	155	3.340	0.03656	0.28665	0.35386	0.31910	0.4
155	Syria	Middle East and Northern Africa	156	3.006	0.05015	0.66320	0.47489	0.72193	0.1
156	Burundi	Sub-Saharan Africa	157	2.905	0.08658	0.01530	0.41587	0.22396	0.1
157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.3

158 rows × 12 columns

In [11]:
```
data["Country"]
```

Out[11]:
```
0      Switzerland
1          Iceland
2          Denmark
3           Norway
4           Canada
           ...
153         Rwanda
154          Benin
155          Syria
156        Burundi
157           Togo
Name: Country, Length: 158, dtype: object
```

In [12]:
```
data1=data[['Happiness Rank','Happiness Score']]
data1
```

Out[12]:

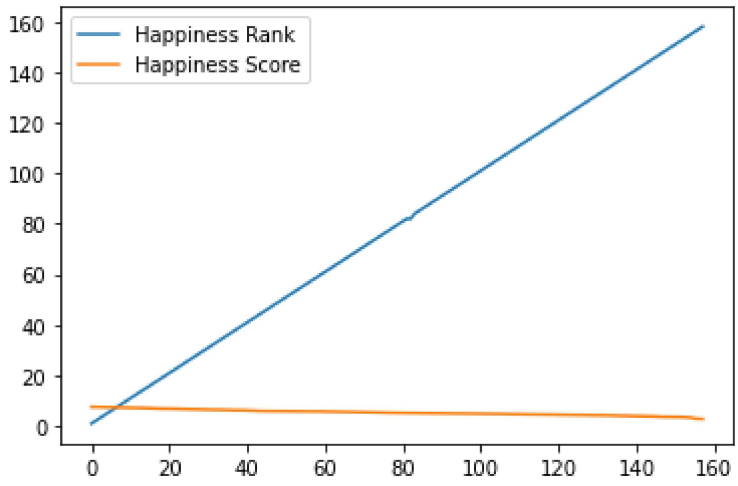
	Happiness Rank	Happiness Score
0	1	7.587
1	2	7.561
2	3	7.527
3	4	7.522
4	5	7.427
...	...	...
153	154	3.465
154	155	3.340
155	156	3.006
156	157	2.905
157	158	2.839

158 rows × 2 columns

In [13]:
```
import matplotlib as mp
```

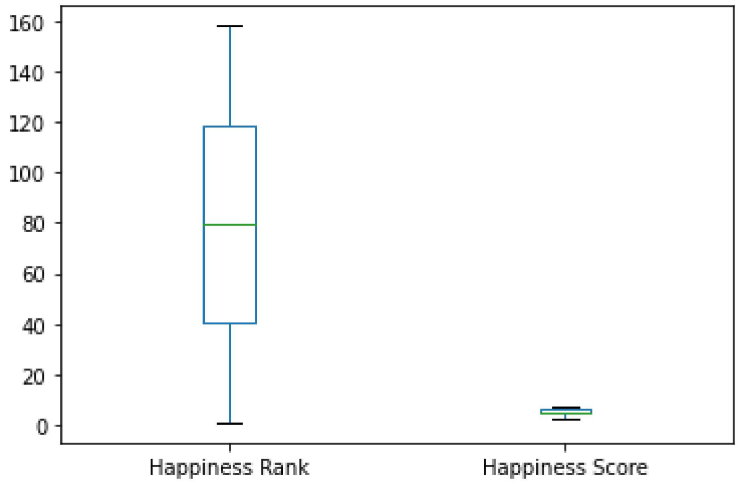
In [14]:
```
data1.plot.line()
```

Out[14]: <AxesSubplot:>



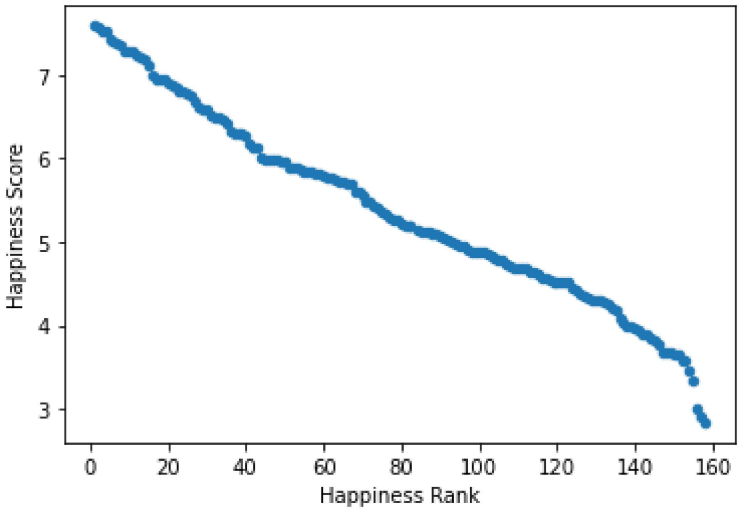
In [15]:
```
data1.plot.box()
```

Out[15]: <AxesSubplot:>



In [17]:
```
data1.plot.scatter(x='Happiness Rank',y='Happiness Score')
```

Out[17]: <AxesSubplot:xlabel='Happiness Rank', ylabel='Happiness Score'>

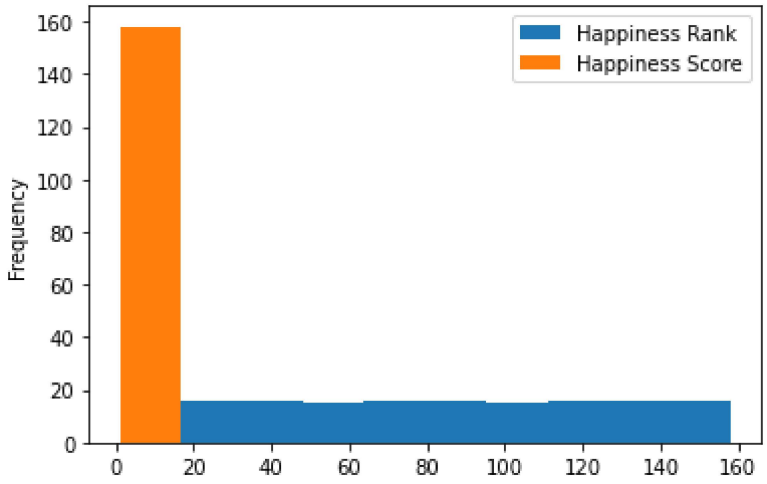


In [18]:
```
data1.plot.area()
```

Out[18]: <AxesSubplot:>

In [19]:
```
data1.plot.hist()
```

Out[19]: <AxesSubplot:ylabel='Frequency'>



In [ ]: