Reading in Files

```
import pandas as pd

df=pd.read_csv(r"C:\Users\User\Documents\FileSorter\companies.csv")
#df=pd.read_table(r"C:\Users\User\Documents\FileSorter\companies.txt")
for text file
#df=pd.read_table(r"C:\Users\User\Documents\FileSorter\
companies.csv",sep=',')
#df=pd.read_csv(r"C:\Users\User\Documents\FileSorter\
companies.txt",sep='\t')
df
```

```
    Rank                   Name                       Industry  \
0      1                Walmart                         Retail
1      2                 Amazon  Retail and cloud computing
2      3                  Apple         Electronics industry
3      4    UnitedHealth Group                     Healthcare
4      5    Berkshire Hathaway                   Conglomerate
..   ...                    ...                            ...
95    96                   TIAA                     Financials
96    97                    CHS       Agriculture cooperative
97    98  Bristol-Myers Squibb     Pharmaceutical industry
98    99  Dow Chemical Company           Chemical industry
99   100               Best Buy                         Retail

   Revenue (USD millions) Revenue growth  Employees  \
0                 648,125           6.0%  2,100,000
1                 574,785          11.9%  1,525,000
2                 383,482          -2.8%    161,000
3                 371,622          14.6%    440,000
4                 364,482          20.7%    396,500
..                    ...            ...        ...
95                 45,735          11.8%     16,023
96                 45,590          -4.6%     10,609
97                 45,006          -2.5%     34,100
98                 44,622         -21.6%     35,900
99                 43,452          -6.1%     85,000

                    Headquarters
0           Bentonville, Arkansas
1             Seattle, Washington
2           Cupertino, California
3           Minnetonka, Minnesota
4                 Omaha, Nebraska
..                            ...
95          New York City, New York
96  Inver Grove Heights, Minnesota
```

```
97         New York City, New York
98               Midland, Michigan
99            Richfield, Minnesota

[100 rows x 7 columns]
```

```python
#JSON Files
#df=pd.read_json(r"C:\Users\User\Documents\FileSorter\companies.json")
#df=pd.read_excel(r"C:\Users\User\Documents\FileSorter\
companies.xlsx") read excel file
#df=pd.read_excel(r"C:\Users\User\Documents\FileSorter\
companies.xlsx",sheet_name='sheet1') read specific sheet

pd.set_option('display.max.rows',50)# how many rows u want to display

df.info()# to get info about the df which you imported
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 7 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Rank                    100 non-null    int64
 1   Name                    100 non-null    object
 2   Industry                100 non-null    object
 3   Revenue (USD millions)  100 non-null    object
 4   Revenue growth          100 non-null    object
 5   Employees               100 non-null    object
 6   Headquarters            100 non-null    object
dtypes: int64(1), object(6)
memory usage: 5.6+ KB
```

```python
df.shape # gives rows*cols)
```

```
(100, 7)
```

```python
df.head(10) # to get first 10 rows of dataframe
```

```
   Rank                  Name                        Industry  \
0     1               Walmart                          Retail
1     2                Amazon      Retail and cloud computing
2     3                 Apple            Electronics industry
3     4    UnitedHealth Group                      Healthcare
4     5    Berkshire Hathaway                    Conglomerate
5     6            CVS Health                      Healthcare
6     7            ExxonMobil              Petroleum industry
7     8              Alphabet  Technology and cloud computing
8     9  McKesson Corporation                          Health
9    10               Cencora              Pharmacy wholesale

  Revenue (USD millions) Revenue growth  Employees
```

```
Headquarters
0                 648,125           6.0%  2,100,000       Bentonville,
Arkansas
1                 574,785          11.9%  1,525,000          Seattle,
Washington
2                 383,482          -2.8%    161,000        Cupertino,
California
3                 371,622          14.6%    440,000       Minnetonka,
Minnesota
4                 364,482          20.7%    396,500             Omaha,
Nebraska
5                 357,776          10.9%    259,500    Woonsocket,
Rhode Island
6                 344,582         -16.7%     61,500
Spring, Texas
7                 307,394           8.7%    182,502   Mountain View,
California
8                 276,711           4.8%     48,000
Irving, Texas
9                 262,173           9.9%     44,000  Conshohocken,
Pennsylvania
```

```
df.tail(10) #to get last 10 rows of dataframe
```

```
    Rank                          Name                 Industry  \
90    91         Capital One Financial               Financials
91    92  Plains All American Pipeline       Petroleum industry
92    93      World Kinect Corporation          Energy trading
93    94                           AIG                Insurance
94    95                     Coca-Cola                 Beverage
95    96                          TIAA               Financials
96    97                           CHS  Agriculture cooperative
97    98          Bristol-Myers Squibb  Pharmaceutical industry
98    99          Dow Chemical Company        Chemical industry
99   100                      Best Buy                   Retail

   Revenue (USD millions) Revenue growth Employees  \
90                 49,484          29.0%    51,987
91                 48,712         -15.1%     4,200
92                 47,711         -19.2%     5,289
93                 46,802         -17.1%    25,200
94                 45,754           6.4%    79,100
95                 45,735          11.8%    16,023
96                 45,590          -4.6%    10,609
97                 45,006          -2.5%    34,100
98                 44,622         -21.6%    35,900
99                 43,452          -6.1%    85,000

                Headquarters
90            Richmond, Virginia
```

```
91                    Houston, Texas
92                    Doral, Florida
93           New York City, New York
94                  Atlanta, Georgia
95           New York City, New York
96    Inver Grove Heights, Minnesota
97           New York City, New York
98                 Midland, Michigan
99              Richfield, Minnesota
```

```python
df['Rank']
```

```
0       1
1       2
2       3
3       4
4       5
     ...
95     96
96     97
97     98
98     99
99    100
Name: Rank, Length: 100, dtype: int64
```

```python
df.loc[6] # to get the details of the loc of the index
```

```
Rank                                           7
Name                                  ExxonMobil
Industry                      Petroleum industry
Revenue (USD millions)                   344,582
Revenue growth                            -16.7%
Employees                                 61,500
Headquarters                       Spring, Texas
Name: 6, dtype: object
```

```python
df.iloc[6] # to get the details of the loc of the index even though u changed the index
```

```
Rank                                           7
Name                                  ExxonMobil
Industry                      Petroleum industry
Revenue (USD millions)                   344,582
Revenue growth                            -16.7%
Employees                                 61,500
Headquarters                       Spring, Texas
Name: 6, dtype: object
```

Filtering and Ordering

```
df[df['Rank']<=10] #to get all Companies whose rank below 10

   Rank                   Name                            Industry  \
0     1                Walmart                              Retail
1     2                 Amazon      Retail and cloud computing
2     3                  Apple             Electronics industry
3     4     UnitedHealth Group                       Healthcare
4     5     Berkshire Hathaway                     Conglomerate
5     6             CVS Health                       Healthcare
6     7             ExxonMobil                Petroleum industry
7     8               Alphabet  Technology and cloud computing
8     9    McKesson Corporation                           Health
9    10                Cencora                Pharmacy wholesale

  Revenue (USD millions) Revenue growth  Employees
Headquarters
0                648,125           6.0%  2,100,000       Bentonville,
Arkansas
1                574,785          11.9%  1,525,000         Seattle,
Washington
2                383,482          -2.8%    161,000       Cupertino,
California
3                371,622          14.6%    440,000       Minnetonka,
Minnesota
4                364,482          20.7%    396,500             Omaha,
Nebraska
5                357,776          10.9%    259,500    Woonsocket,
Rhode Island
6                344,582         -16.7%     61,500
Spring, Texas
7                307,394           8.7%    182,502   Mountain View,
California
8                276,711           4.8%     48,000
Irving, Texas
9                262,173           9.9%     44,000  Conshohocken,
Pennsylvania

specific_company=['Alphabet','Berkshire Hathaway']
df[df['Name'].isin(specific_company)]
#to get all Companies whose Name is mentioned

   Rank                 Name                          Industry  \
4     5   Berkshire Hathaway                      Conglomerate
7     8             Alphabet  Technology and cloud computing

  Revenue (USD millions) Revenue growth Employees
Headquarters
4                364,482          20.7%   396,500             Omaha,
```

```
Nebraska
7                 307,394           8.7%   182,502  Mountain View,
California

df[df['Name'].str.contains('Wal')]
#you get all the Company Names which has Wal name in it

    Rank                       Name                Industry  \
0      1                    Walmart                  Retail
27    28  Walgreens Boots Alliance  Pharmaceutical industry
46    47   The Walt Disney Company                    Media

   Revenue (USD millions) Revenue growth  Employees
Headquarters
0                 648,125           6.0%  2,100,000  Bentonville,
Arkansas
27                139,081           4.8%    268,500    Deerfield,
Illinois
46                 88,898           7.5%    199,125    Burbank,
California

df2=df.set_index('Name')
df2

                      Rank                 Industry Revenue (USD
millions)  \
Name

Walmart                  1                   Retail
648,125
Amazon                   2  Retail and cloud computing
574,785
Apple                    3       Electronics industry
383,482
UnitedHealth Group       4               Healthcare
371,622
Berkshire Hathaway       5             Conglomerate
364,482
...                    ...                      ...
...
TIAA                    96               Financials
45,735
CHS                     97    Agriculture cooperative
45,590
Bristol-Myers Squibb    98    Pharmaceutical industry
45,006
Dow Chemical Company    99         Chemical industry
44,622
Best Buy               100                   Retail
43,452
```

```
                     Revenue growth  Employees
Headquarters
Name

Walmart                        6.0%  2,100,000           Bentonville,
Arkansas
Amazon                        11.9%  1,525,000               Seattle,
Washington
Apple                         -2.8%    161,000             Cupertino,
California
UnitedHealth Group            14.6%    440,000             Minnetonka,
Minnesota
Berkshire Hathaway            20.7%    396,500                  Omaha,
Nebraska
...                             ...        ...
...
TIAA                          11.8%     16,023         New York City,
New York
CHS                           -4.6%     10,609  Inver Grove Heights,
Minnesota
Bristol-Myers Squibb          -2.5%     34,100         New York City,
New York
Dow Chemical Company         -21.6%     35,900                Midland,
Michigan
Best Buy                      -6.1%     85,000             Richfield,
Minnesota

[100 rows x 6 columns]
```

```
df2.filter(items=['Name','Industry'])
#to get only name and industry removing all cols
```

```
                                         Industry
Name
Walmart                                    Retail
Amazon                  Retail and cloud computing
Apple                         Electronics industry
UnitedHealth Group                      Healthcare
Berkshire Hathaway                    Conglomerate
...                                            ...
TIAA                                    Financials
CHS                        Agriculture cooperative
Bristol-Myers Squibb        Pharmaceutical industry
Dow Chemical Company              Chemical industry
Best Buy                                    Retail

[100 rows x 1 columns]
```

```
df2.filter(like='Wal',axis=0)
#to get all names which has Wal in it

                           Rank                 Industry  \
Name
Walmart                       1                   Retail
Walgreens Boots Alliance     28  Pharmaceutical industry
The Walt Disney Company      47                    Media

                         Revenue (USD millions) Revenue growth
Employees  \
Name

Walmart                                 648,125           6.0%
2,100,000
Walgreens Boots Alliance                139,081           4.8%
268,500
The Walt Disney Company                  88,898           7.5%
199,125

                                    Headquarters
Name
Walmart                    Bentonville, Arkansas
Walgreens Boots Alliance     Deerfield, Illinois
The Walt Disney Company      Burbank, California

df2.loc['Walmart']
# to get details of Walmart

Rank                                              1
Industry                                     Retail
Revenue (USD millions)                      648,125
Revenue growth                                 6.0%
Employees                                 2,100,000
Headquarters                  Bentonville, Arkansas
Name: Walmart, dtype: object

df2.iloc[2]
# to get details of index 2

Rank                                              3
Industry                       Electronics industry
Revenue (USD millions)                      383,482
Revenue growth                                -2.8%
Employees                                   161,000
Headquarters                  Cupertino, California
Name: Apple, dtype: object

df
```

```
    Rank                   Name                   Industry  \
0      1                Walmart                     Retail
1      2                 Amazon  Retail and cloud computing
2      3                  Apple        Electronics industry
3      4    UnitedHealth Group                  Healthcare
4      5    Berkshire Hathaway                Conglomerate
..   ...                    ...                        ...
95    96                   TIAA                 Financials
96    97                    CHS    Agriculture cooperative
97    98  Bristol-Myers Squibb     Pharmaceutical industry
98    99  Dow Chemical Company           Chemical industry
99   100               Best Buy                     Retail

   Revenue (USD millions) Revenue growth  Employees  \
0                 648,125           6.0%  2,100,000
1                 574,785          11.9%  1,525,000
2                 383,482          -2.8%    161,000
3                 371,622          14.6%    440,000
4                 364,482          20.7%    396,500
..                    ...            ...        ...
95                 45,735          11.8%     16,023
96                 45,590          -4.6%     10,609
97                 45,006          -2.5%     34,100
98                 44,622         -21.6%     35,900
99                 43,452          -6.1%     85,000

                      Headquarters
0           Bentonville, Arkansas
1             Seattle, Washington
2           Cupertino, California
3           Minnetonka, Minnesota
4                 Omaha, Nebraska
..                             ...
95         New York City, New York
96  Inver Grove Heights, Minnesota
97         New York City, New York
98               Midland, Michigan
99            Richfield, Minnesota

[100 rows x 7 columns]
```

```python
df[df['Rank']<10].sort_values(by='Rank',ascending=False)
#to sort the values
```

```
   Rank                  Name                          Industry  \
8     9  McKesson Corporation                            Health
7     8              Alphabet  Technology and cloud computing
6     7            ExxonMobil               Petroleum industry
5     6            CVS Health                       Healthcare
4     5    Berkshire Hathaway                     Conglomerate
```

```
3     4    UnitedHealth Group                        Healthcare
2     3                 Apple            Electronics industry
1     2                Amazon      Retail and cloud computing
0     1               Walmart                          Retail

  Revenue (USD millions) Revenue growth  Employees
Headquarters
8                 276,711           4.8%     48,000
Irving, Texas
7                 307,394           8.7%    182,502  Mountain View,
California
6                 344,582         -16.7%     61,500
Spring, Texas
5                 357,776          10.9%    259,500   Woonsocket, Rhode
Island
4                 364,482          20.7%    396,500            Omaha,
Nebraska
3                 371,622          14.6%    440,000      Minnetonka,
Minnesota
2                 383,482          -2.8%    161,000      Cupertino,
California
1                 574,785          11.9%  1,525,000        Seattle,
Washington
0                 648,125           6.0%  2,100,000      Bentonville,
Arkansas
```

```
df[df['Rank']<10].sort_values(by=['Revenue (USD
millions)','Rank'],ascending=False)
#to sort the values but here it is based on revenue
```

```
   Rank                  Name                        Industry  \
0     1               Walmart                          Retail
1     2                Amazon      Retail and cloud computing
2     3                 Apple            Electronics industry
3     4    UnitedHealth Group                        Healthcare
4     5    Berkshire Hathaway                      Conglomerate
5     6            CVS Health                        Healthcare
6     7             ExxonMobil             Petroleum industry
7     8              Alphabet  Technology and cloud computing
8     9  McKesson Corporation                          Health

  Revenue (USD millions) Revenue growth  Employees
Headquarters
0                 648,125           6.0%  2,100,000      Bentonville,
Arkansas
1                 574,785          11.9%  1,525,000        Seattle,
Washington
2                 383,482          -2.8%    161,000      Cupertino,
California
3                 371,622          14.6%    440,000      Minnetonka,
```

```
Minnesota
4                 364,482           20.7%    396,500            Omaha,
Nebraska
5                 357,776           10.9%    259,500   Woonsocket, Rhode
Island
6                 344,582          -16.7%     61,500
Spring, Texas
7                 307,394            8.7%    182,502  Mountain View,
California
8                 276,711            4.8%     48,000
Irving, Texas
```

```
df[df['Rank']<10].sort_values(by=['Revenue (USD
millions)','Rank'],ascending=[False,True])
#to sort the values but here it is based on revenue but ascending u
can choose as per ur wish
```

```
----------------------------------------------------------------------
-----
NameError                                 Traceback (most recent call
last)
Cell In[1], line 1
----> 1 df[df['Rank']<10].sort_values(by=['Revenue (USD
millions)','Rank'],ascending=[False,True])

NameError: name 'df' is not defined
```

Indexing

```
import pandas as pd

df = pd.read_csv(r"C:\Users\User\Downloads\world_population.csv")
df
```

```
     Rank CCA3            Country           Capital Continent  \
0      36  AFG        Afghanistan             Kabul      Asia
1     138  ALB            Albania            Tirana    Europe
2      34  DZA            Algeria           Algiers    Africa
3     213  ASM     American Samoa         Pago Pago   Oceania
4     203  AND            Andorra  Andorra la Vella    Europe
..    ...  ...                ...               ...       ...
229   226  WLF  Wallis and Futuna          Mata-Utu   Oceania
230   172  ESH     Western Sahara          El Aaiún    Africa
231    46  YEM              Yemen             Sanaa      Asia
232    63  ZMB             Zambia            Lusaka    Africa
233    74  ZWE           Zimbabwe            Harare    Africa

     2022 Population  2020 Population  2015 Population  2010
Population  \
0         41128771.0       38972230.0       33753499.0
28189672.0
1          2842321.0        2866849.0        2882481.0
2913399.0
2         44903225.0       43451666.0       39543154.0
35856344.0
3            44273.0          46189.0          51368.0
54849.0
4            79824.0          77700.0          71746.0
71519.0
..               ...              ...              ...              ..
.
229          11572.0          11655.0          12182.0
13142.0
230         575986.0         556048.0         491824.0
413296.0
231       33696614.0       32284046.0       28516545.0
24743946.0
232       20017675.0       18927715.0              NaN
13792086.0
233       16320537.0       15669666.0       14154937.0
12839771.0

     2000 Population  1990 Population  1980 Population  1970
Population  \
0         19542982.0       10694796.0       12486631.0
```

```
10752971.0
1         3182021.0        3295066.0        2941651.0
2324731.0
2        30774621.0       25518074.0       18739378.0
13795915.0
3           58230.0          47818.0          32886.0
27075.0
4           66097.0          53569.0          35611.0
19860.0
..              ...              ...              ...              ..
.
229         14723.0          13454.0          11315.0
9377.0
230        270375.0         178529.0         116775.0
76371.0
231      18628700.0       13375121.0        9204938.0
6843607.0
232       9891136.0        7686401.0        5720438.0
4281671.0
233      11834676.0       10113893.0        7049926.0
5202918.0

     Area (km²)  Density (per km²)  Growth Rate  World Population
Percentage
0      652230.0            63.0587       1.0257
0.52
1       28748.0            98.8702       0.9957
0.04
2     2381741.0            18.8531       1.0164
0.56
3         199.0           222.4774       0.9831
0.00
4         468.0           170.5641       1.0100
0.00
..          ...                ...          ...
...
229       142.0            81.4930       0.9953
0.00
230    266000.0             2.1654       1.0184
0.01
231    527968.0            63.8232       1.0217
0.42
232    752612.0            26.5976       1.0280
0.25
233    390757.0            41.7665       1.0204
0.20

[234 rows x 17 columns]
```

```
df = pd.read_csv(r"C:\Users\User\Downloads\
world_population.csv",index_col="Country")
df

                   Rank CCA3           Capital Continent  2022
Population  \
Country

Afghanistan          36  AFG             Kabul      Asia
41128771.0
Albania             138  ALB            Tirana    Europe
2842321.0
Algeria              34  DZA           Algiers    Africa
44903225.0
American Samoa      213  ASM         Pago Pago   Oceania
44273.0
Andorra             203  AND  Andorra la Vella    Europe
79824.0
...                 ...  ...               ...       ...
...
Wallis and Futuna   226  WLF          Mata-Utu   Oceania
11572.0
Western Sahara      172  ESH          El Aaiún    Africa
575986.0
Yemen                46  YEM             Sanaa      Asia
33696614.0
Zambia               63  ZMB            Lusaka    Africa
20017675.0
Zimbabwe             74  ZWE            Harare    Africa
16320537.0

                   2020 Population  2015 Population  2010
Population  \
Country

Afghanistan             38972230.0       33753499.0       28189672.0

Albania                  2866849.0        2882481.0        2913399.0

Algeria                 43451666.0       39543154.0       35856344.0

American Samoa             46189.0          51368.0          54849.0

Andorra                    77700.0          71746.0          71519.0

...                            ...              ...              ...

Wallis and Futuna          11655.0          12182.0          13142.0

Western Sahara            556048.0         491824.0         413296.0
```

```
Yemen                   32284046.0        28516545.0        24743946.0

Zambia                  18927715.0               NaN        13792086.0

Zimbabwe                15669666.0        14154937.0        12839771.0


                   2000 Population  1990 Population  1980
Population  \
Country

Afghanistan             19542982.0        10694796.0        12486631.0

Albania                  3182021.0         3295066.0         2941651.0

Algeria                 30774621.0        25518074.0        18739378.0

American Samoa             58230.0           47818.0           32886.0

Andorra                    66097.0           53569.0           35611.0

...                            ...               ...               ...

Wallis and Futuna          14723.0           13454.0           11315.0

Western Sahara            270375.0          178529.0          116775.0

Yemen                   18628700.0        13375121.0         9204938.0

Zambia                   9891136.0         7686401.0         5720438.0

Zimbabwe                11834676.0        10113893.0         7049926.0


                   1970 Population  Area (km²)  Density (per km²)  \
Country
Afghanistan             10752971.0    652230.0            63.0587
Albania                  2324731.0     28748.0            98.8702
Algeria                 13795915.0   2381741.0            18.8531
American Samoa             27075.0       199.0           222.4774
Andorra                    19860.0       468.0           170.5641
...                            ...         ...                ...
Wallis and Futuna           9377.0       142.0            81.4930
Western Sahara             76371.0    266000.0             2.1654
Yemen                    6843607.0    527968.0            63.8232
Zambia                   4281671.0    752612.0            26.5976
Zimbabwe                 5202918.0    390757.0            41.7665

                   Growth Rate  World Population Percentage
Country
Afghanistan             1.0257                         0.52
```

```
Albania                     0.9957                        0.04
Algeria                     1.0164                        0.56
American Samoa              0.9831                        0.00
Andorra                     1.0100                        0.00
...                            ...                         ...
Wallis and Futuna           0.9953                        0.00
Western Sahara              1.0184                        0.01
Yemen                       1.0217                        0.42
Zambia                      1.0280                        0.25
Zimbabwe                    1.0204                        0.20

[234 rows x 16 columns]
```

```python
df.reset_index(inplace=True)
df
```

```
               Country  Rank CCA3           Capital Continent  \
0          Afghanistan    36  AFG             Kabul      Asia
1              Albania   138  ALB            Tirana    Europe
2              Algeria    34  DZA           Algiers    Africa
3       American Samoa   213  ASM         Pago Pago   Oceania
4              Andorra   203  AND  Andorra la Vella    Europe
..                 ...   ...  ...               ...       ...
229  Wallis and Futuna   226  WLF          Mata-Utu   Oceania
230     Western Sahara   172  ESH          El Aaiún    Africa
231              Yemen    46  YEM             Sanaa      Asia
232             Zambia    63  ZMB            Lusaka    Africa
233           Zimbabwe    74  ZWE            Harare    Africa

     2022 Population  2020 Population  2015 Population  2010
Population  \
0         41128771.0       38972230.0       33753499.0
28189672.0
1          2842321.0        2866849.0        2882481.0
2913399.0
2         44903225.0       43451666.0       39543154.0
35856344.0
3            44273.0          46189.0          51368.0
54849.0
4            79824.0          77700.0          71746.0
71519.0
..               ...              ...              ...              ..
.
229          11572.0          11655.0          12182.0
13142.0
230         575986.0         556048.0         491824.0
413296.0
231       33696614.0       32284046.0       28516545.0
24743946.0
232       20017675.0       18927715.0              NaN
```

```
13792086.0
233       16320537.0       15669666.0       14154937.0
12839771.0

     2000 Population  1990 Population  1980 Population  1970
Population  \
0         19542982.0       10694796.0       12486631.0
10752971.0
1          3182021.0        3295066.0        2941651.0
2324731.0
2         30774621.0       25518074.0       18739378.0
13795915.0
3            58230.0          47818.0          32886.0
27075.0
4            66097.0          53569.0          35611.0
19860.0
..               ...              ...              ...              ..
.
229          14723.0          13454.0          11315.0
9377.0
230         270375.0         178529.0         116775.0
76371.0
231       18628700.0       13375121.0        9204938.0
6843607.0
232        9891136.0        7686401.0        5720438.0
4281671.0
233       11834676.0       10113893.0        7049926.0
5202918.0

     Area (km²)  Density (per km²)  Growth Rate  World Population
Percentage
0      652230.0            63.0587       1.0257
0.52
1       28748.0            98.8702       0.9957
0.04
2     2381741.0            18.8531       1.0164
0.56
3         199.0           222.4774       0.9831
0.00
4         468.0           170.5641       1.0100
0.00
..          ...                ...          ...
...
229       142.0            81.4930       0.9953
0.00
230    266000.0             2.1654       1.0184
0.01
231    527968.0            63.8232       1.0217
0.42
```

```
232    752612.0           26.5976       1.0280
0.25
233    390757.0           41.7665       1.0204
0.20

[234 rows x 17 columns]

df.set_index('Country', inplace=True)
df

                   Rank CCA3           Capital Continent  2022
Population  \
Country

Afghanistan          36  AFG             Kabul      Asia
41128771.0
Albania             138  ALB            Tirana    Europe
2842321.0
Algeria              34  DZA           Algiers    Africa
44903225.0
American Samoa      213  ASM         Pago Pago   Oceania
44273.0
Andorra             203  AND  Andorra la Vella    Europe
79824.0
...                 ...  ...               ...       ...
...
Wallis and Futuna   226  WLF          Mata-Utu   Oceania
11572.0
Western Sahara      172  ESH          El Aaiún    Africa
575986.0
Yemen                46  YEM             Sanaa      Asia
33696614.0
Zambia               63  ZMB            Lusaka    Africa
20017675.0
Zimbabwe             74  ZWE            Harare    Africa
16320537.0

                   2020 Population  2015 Population  2010
Population  \
Country

Afghanistan             38972230.0       33753499.0       28189672.0

Albania                  2866849.0        2882481.0        2913399.0

Algeria                 43451666.0       39543154.0       35856344.0

American Samoa             46189.0          51368.0          54849.0

Andorra                    77700.0          71746.0          71519.0
```

```
...                        ...          ...          ...
Wallis and Futuna      11655.0      12182.0      13142.0
Western Sahara        556048.0     491824.0     413296.0
Yemen               32284046.0   28516545.0   24743946.0
Zambia              18927715.0          NaN   13792086.0
Zimbabwe            15669666.0   14154937.0   12839771.0

                   2000 Population  1990 Population  1980
Population  \
Country
Afghanistan         19542982.0   10694796.0   12486631.0
Albania              3182021.0    3295066.0    2941651.0
Algeria             30774621.0   25518074.0   18739378.0
American Samoa         58230.0      47818.0      32886.0
Andorra                66097.0      53569.0      35611.0
...                        ...          ...          ...
Wallis and Futuna      14723.0      13454.0      11315.0
Western Sahara        270375.0     178529.0     116775.0
Yemen               18628700.0   13375121.0    9204938.0
Zambia               9891136.0    7686401.0    5720438.0
Zimbabwe            11834676.0   10113893.0    7049926.0

                   1970 Population  Area (km²)  Density (per km²)  \
Country
Afghanistan         10752971.0    652230.0            63.0587
Albania              2324731.0     28748.0            98.8702
Algeria             13795915.0   2381741.0            18.8531
American Samoa         27075.0       199.0           222.4774
Andorra                19860.0       468.0           170.5641
...                        ...         ...                ...
Wallis and Futuna       9377.0       142.0            81.4930
Western Sahara         76371.0    266000.0             2.1654
Yemen                6843607.0    527968.0            63.8232
```

```
Zambia                          4281671.0    752612.0            26.5976
Zimbabwe                        5202918.0    390757.0            41.7665

                   Growth Rate  World Population Percentage
Country
Afghanistan             1.0257                         0.52
Albania                 0.9957                         0.04
Algeria                 1.0164                         0.56
American Samoa          0.9831                         0.00
Andorra                 1.0100                         0.00
...                        ...                          ...
Wallis and Futuna       0.9953                         0.00
Western Sahara          1.0184                         0.01
Yemen                   1.0217                         0.42
Zambia                  1.0280                         0.25
Zimbabwe                1.0204                         0.20

[234 rows x 16 columns]
```

```python
df.loc['Western Sahara']#location
```

```
Rank                                  172
CCA3                                  ESH
Capital                         El Aaiún
Continent                          Africa
2022 Population                  575986.0
2020 Population                  556048.0
2015 Population                  491824.0
2010 Population                  413296.0
2000 Population                  270375.0
1990 Population                  178529.0
1980 Population                  116775.0
1970 Population                   76371.0
Area (km²)                       266000.0
Density (per km²)                  2.1654
Growth Rate                        1.0184
World Population Percentage          0.01
Name: Western Sahara, dtype: object
```

```python
df.iloc[1]#Integer location
```

```
Rank                                  138
CCA3                                  ALB
Capital                            Tirana
Continent                          Europe
2022 Population                 2842321.0
2020 Population                 2866849.0
2015 Population                 2882481.0
2010 Population                 2913399.0
2000 Population                 3182021.0
```

```
1990 Population                       3295066.0
1980 Population                       2941651.0
1970 Population                       2324731.0
Area (km²)                              28748.0
Density (per km²)                       98.8702
Growth Rate                              0.9957
World Population Percentage                0.04
Name: Albania, dtype: object
```

```
df.reset_index(inplace=True)

df.set_index(['Continent','Country'], inplace=True) #Multi index
df
```

```
                                 Rank CCA3           Capital  2022
Population  \
Continent Country

Asia      Afghanistan              36  AFG             Kabul
41128771.0
Europe    Albania                 138  ALB            Tirana
2842321.0
Africa    Algeria                  34  DZA           Algiers
44903225.0
Oceania   American Samoa          213  ASM         Pago Pago
44273.0
Europe    Andorra                 203  AND  Andorra la Vella
79824.0
...                               ...  ...               ...
...
Oceania   Wallis and Futuna       226  WLF          Mata-Utu
11572.0
Africa    Western Sahara          172  ESH          El Aaiún
575986.0
Asia      Yemen                    46  YEM             Sanaa
33696614.0
Africa    Zambia                   63  ZMB            Lusaka
20017675.0
          Zimbabwe                 74  ZWE            Harare
16320537.0

                                 2020 Population  2015 Population  \
Continent Country
Asia      Afghanistan                 38972230.0       33753499.0
Europe    Albania                      2866849.0        2882481.0
Africa    Algeria                     43451666.0       39543154.0
Oceania   American Samoa                 46189.0          51368.0
Europe    Andorra                        77700.0          71746.0
...                                          ...              ...
Oceania   Wallis and Futuna              11655.0          12182.0
```

```
Africa    Western Sahara               556048.0         491824.0
Asia      Yemen                      32284046.0       28516545.0
Africa    Zambia                     18927715.0              NaN
          Zimbabwe                   15669666.0       14154937.0

                             2010 Population  2000 Population  \
Continent Country
Asia      Afghanistan                28189672.0       19542982.0
Europe    Albania                     2913399.0        3182021.0
Africa    Algeria                    35856344.0       30774621.0
Oceania   American Samoa                54849.0          58230.0
Europe    Andorra                       71519.0          66097.0
...                                         ...              ...
Oceania   Wallis and Futuna             13142.0          14723.0
Africa    Western Sahara               413296.0         270375.0
Asia      Yemen                      24743946.0       18628700.0
Africa    Zambia                     13792086.0        9891136.0
          Zimbabwe                   12839771.0       11834676.0

                             1990 Population  1980 Population  \
Continent Country
Asia      Afghanistan                10694796.0       12486631.0
Europe    Albania                     3295066.0        2941651.0
Africa    Algeria                    25518074.0       18739378.0
Oceania   American Samoa                47818.0          32886.0
Europe    Andorra                       53569.0          35611.0
...                                         ...              ...
Oceania   Wallis and Futuna             13454.0          11315.0
Africa    Western Sahara               178529.0         116775.0
Asia      Yemen                      13375121.0        9204938.0
Africa    Zambia                      7686401.0        5720438.0
          Zimbabwe                   10113893.0        7049926.0

                             1970 Population  Area (km²)  Density (per
km²)  \
Continent Country

Asia      Afghanistan                10752971.0    652230.0
63.0587
Europe    Albania                     2324731.0     28748.0
98.8702
Africa    Algeria                    13795915.0   2381741.0
18.8531
Oceania   American Samoa                27075.0       199.0
222.4774
Europe    Andorra                       19860.0       468.0
170.5641
...                                         ...         ...
...
Oceania   Wallis and Futuna              9377.0       142.0
```

```
81.4930
Africa    Western Sahara          76371.0    266000.0
2.1654
Asia      Yemen                 6843607.0    527968.0
63.8232
Africa    Zambia                4281671.0    752612.0
26.5976
          Zimbabwe              5202918.0    390757.0
41.7665

                             Growth Rate  World Population Percentage

Continent Country

Asia      Afghanistan             1.0257                         0.52

Europe    Albania                 0.9957                         0.04

Africa    Algeria                 1.0164                         0.56

Oceania   American Samoa          0.9831                         0.00

Europe    Andorra                 1.0100                         0.00

...                                  ...                          ...

Oceania   Wallis and Futuna       0.9953                         0.00

Africa    Western Sahara          1.0184                         0.01

Asia      Yemen                   1.0217                         0.42

Africa    Zambia                  1.0280                         0.25

          Zimbabwe                1.0204                         0.20


[234 rows x 15 columns]
```

```
df.sort_index(ascending=False)
```

```
                             Rank CCA3      Capital  2022 Population  \
Continent     Country
South America Venezuela        51  VEN      Caracas       28301696.0
              Uruguay         133  URY   Montevideo        3422794.0
              Suriname        170  SUR   Paramaribo         618040.0
              Peru             44  PER         Lima       34049588.0
              Paraguay        109  PRY     Asunción        6780744.0
...                           ...  ...          ...              ...
Africa        Burkina Faso     58  BFA  Ouagadougou       22673762.0
              Botswana        144  BWA     Gaborone        2630296.0
```

```
              Benin           77  BEN   Porto-Novo       13352864.0
              Angola          42  AGO       Luanda       35588987.0
              Algeria         34  DZA      Algiers       44903225.0

                            2020 Population  2015 Population  2010
Population  \
Continent     Country

South America Venezuela          28490453.0       30529716.0
28715022.0
              Uruguay             3429086.0        3402818.0
3352651.0
              Suriname             607065.0         575475.0
546080.0
              Peru               33304756.0       30711863.0
29229572.0
              Paraguay            6618695.0        6177950.0
5768613.0
...                                     ...              ...
...
Africa        Burkina Faso       21522626.0       18718019.0
16116845.0
              Botswana            2546402.0        2305171.0
2091664.0
              Benin              12643123.0       10932783.0
9445710.0
              Angola             33428485.0       28127721.0
23364185.0
              Algeria            43451666.0       39543154.0
35856344.0

                            2000 Population  1990 Population  1980
Population  \
Continent     Country

South America Venezuela                 NaN       19750579.0
15210443.0
              Uruguay             3292224.0        3117012.0
2953750.0
              Suriname             478998.0         412756.0
375112.0
              Peru               26654439.0       22109099.0
17492406.0
              Paraguay            5123819.0        4059195.0
3078912.0
...                                     ...              ...
...
Africa        Burkina Faso       11882888.0        9131361.0
6932967.0
              Botswana            1726985.0        1341474.0
```

```
938578.0
               Benin                 6998023.0       5133419.0
3833939.0
               Angola               16394062.0      11828638.0
8330047.0
               Algeria              30774621.0      25518074.0
18739378.0

                            1970 Population  Area (km²)  Density (per
km²)  \
Continent      Country

South America Venezuela             11355475.0         NaN
30.8820
               Uruguay               2790265.0    181034.0
18.9069
               Suriname               379918.0    163820.0
3.7727
               Peru                 13562371.0   1285216.0
26.4933
               Paraguay              2408787.0    406752.0
16.6705
...                                        ...         ...
...
Africa         Burkina Faso          5611666.0    272967.0
83.0641
               Botswana               592244.0    582000.0
4.5194
               Benin                 3023443.0    112622.0
118.5635
               Angola                6029700.0   1246700.0
28.5466
               Algeria              13795915.0   2381741.0
18.8531

                            Growth Rate  World Population Percentage
Continent      Country
South America Venezuela          1.0036                         0.35
               Uruguay            0.9990                         0.04
               Suriname           1.0082                         0.01
               Peru               1.0099                         0.43
               Paraguay           1.0115                         0.09
...                                 ...                          ...
Africa         Burkina Faso       1.0259                         0.28
               Botswana           1.0162                         0.03
               Benin              1.0274                         0.17
               Angola             1.0315                         0.45
               Algeria            1.0164                         0.56

[234 rows x 15 columns]
```

```python
df.loc['Africa','Angola'] #Multi indexing works here
```

```
Rank                                     42
CCA3                                    AGO
Capital                              Luanda
2022 Population                  35588987.0
2020 Population                  33428485.0
2015 Population                  28127721.0
2010 Population                  23364185.0
2000 Population                  16394062.0
1990 Population                  11828638.0
1980 Population                   8330047.0
1970 Population                   6029700.0
Area (km²)                        1246700.0
Density (per km²)                   28.5466
Growth Rate                          1.0315
World Population Percentage            0.45
Name: (Africa, Angola), dtype: object
```

```python
df.iloc[1] #doesnt go based on Multi it goes same with earlier one
```

```
Rank                                    138
CCA3                                    ALB
Capital                              Tirana
2022 Population                   2842321.0
2020 Population                   2866849.0
2015 Population                   2882481.0
2010 Population                   2913399.0
2000 Population                   3182021.0
1990 Population                   3295066.0
1980 Population                   2941651.0
1970 Population                   2324731.0
Area (km²)                          28748.0
Density (per km²)                   98.8702
Growth Rate                          0.9957
World Population Percentage            0.04
Name: (Europe, Albania), dtype: object
```

```
import pandas as pd

df= pd.read_csv(r"C:\Users\User\Downloads\Flavors.csv")
df
```

```
                      Flavor Base Flavor Liked  Flavor Rating  Texture 
Rating  \
0     Mint Chocolate Chip     Vanilla   Yes           10.0
8.0
1               Chocolate   Chocolate   Yes            8.8
7.6
2                 Vanilla     Vanilla    No            4.7
5.0
3            Cookie Dough     Vanilla   Yes            6.9
6.5
4              Rocky Road   Chocolate   Yes            8.2
7.0
5               Pistachio     Vanilla    No            2.3
3.4
6             Cake Batter     Vanilla   Yes            6.5
6.0
7              Neapolitan     Vanilla    No            3.8
5.0
8  Chocolte Fudge Brownie   Chocolate   Yes            8.2
7.1

   Total Rating
0          18.0
1          16.6
2           9.7
3          13.4
4          15.2
5           5.7
6          12.5
7           8.8
8          15.3
```

```
#group=df.groupby('Base Flavor')
#group.mean()
#group.count()
#group.min()
#group.max()
#group.sum()
df.groupby('Base Flavor').agg({'Flavor Rating':
['mean','max','count','sum'],'Texture Rating':
['mean','max','count','sum']})
```

```
            Flavor Rating                      Texture Rating

                     mean   max count   sum            mean   max count
```

```
sum
Base Flavor

Chocolate              8.4   8.8     3  25.2       7.233333  7.6     3
21.7
Vanilla                5.7  10.0     6  34.2       5.650000  8.0     6
33.9
```

```
df.groupby(['Base Flavor','Liked']).mean()
```

```
----------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:1942, in GroupBy._agg_py_fallback(self, how, values, ndim,
alt)
   1941 try:
-> 1942     res_values = self._grouper.agg_series(ser, alt,
preserve_dtype=True)
   1943 except Exception as err:

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\ops.py:864, in
BaseGrouper.agg_series(self, obj, func, preserve_dtype)
    862     preserve_dtype = True
--> 864 result = self._aggregate_series_pure_python(obj, func)
    866 npvalues = lib.maybe_convert_objects(result, try_float=False)

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\ops.py:885, in
BaseGrouper._aggregate_series_pure_python(self, obj, func)
    884 for i, group in enumerate(splitter):
--> 885     res = func(group)
    886     res = extract_result(res)

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:2454, in GroupBy.mean.<locals>.<lambda>(x)
   2451 else:
   2452     result = self._cython_agg_general(
   2453         "mean",
-> 2454         alt=lambda x: Series(x,
copy=False).mean(numeric_only=numeric_only),
   2455         numeric_only=numeric_only,
   2456     )
   2457     return result.__finalize__(self.obj, method="groupby")

File ~\anaconda3\Lib\site-packages\pandas\core\series.py:6549, in
Series.mean(self, axis, skipna, numeric_only, **kwargs)
   6541 @doc(make_doc("mean", ndim=1))
   6542 def mean(
   6543     self,
```

```
   (...)
   6547     **kwargs,
   6548 ):
-> 6549     return NDFrame.mean(self, axis, skipna, numeric_only,
**kwargs)

File ~\anaconda3\Lib\site-packages\pandas\core\generic.py:12420, in
NDFrame.mean(self, axis, skipna, numeric_only, **kwargs)
  12413 def mean(
  12414     self,
  12415     axis: Axis | None = 0,
   (...)
  12418     **kwargs,
  12419 ) -> Series | float:
> 12420     return self._stat_function(
  12421         "mean", nanops.nanmean, axis, skipna, numeric_only,
**kwargs
  12422     )

File ~\anaconda3\Lib\site-packages\pandas\core\generic.py:12377, in
NDFrame._stat_function(self, name, func, axis, skipna, numeric_only,
**kwargs)
  12375 validate_bool_kwarg(skipna, "skipna", none_allowed=False)
> 12377 return self._reduce(
  12378     func, name=name, axis=axis, skipna=skipna,
numeric_only=numeric_only
  12379 )

File ~\anaconda3\Lib\site-packages\pandas\core\series.py:6457, in
Series._reduce(self, op, name, axis, skipna, numeric_only,
filter_type, **kwds)
   6453     raise TypeError(
   6454         f"Series.{name} does not allow
{kwd_name}={numeric_only} "
   6455         "with non-numeric dtypes."
   6456     )
-> 6457 return op(delegate, skipna=skipna, **kwds)

File ~\anaconda3\Lib\site-packages\pandas\core\nanops.py:147, in
bottleneck_switch.__call__.<locals>.f(values, axis, skipna, **kwds)
    146 else:
--> 147     result = alt(values, axis=axis, skipna=skipna, **kwds)
    149 return result

File ~\anaconda3\Lib\site-packages\pandas\core\nanops.py:404, in
_datetimelike_compat.<locals>.new_func(values, axis, skipna, mask,
**kwargs)
    402     mask = isna(values)
--> 404 result = func(values, axis=axis, skipna=skipna, mask=mask,
**kwargs)
```

```
    406 if datetimelike:

File ~\anaconda3\Lib\site-packages\pandas\core\nanops.py:720, in
nanmean(values, axis, skipna, mask)
    719 the_sum = values.sum(axis, dtype=dtype_sum)
--> 720 the_sum = _ensure_numeric(the_sum)
    722 if axis is not None and getattr(the_sum, "ndim", False):

File ~\anaconda3\Lib\site-packages\pandas\core\nanops.py:1701, in
_ensure_numeric(x)
   1699 if isinstance(x, str):
   1700     # GH#44008, GH#36703 avoid casting e.g. strings to numeric
-> 1701     raise TypeError(f"Could not convert string '{x}' to
numeric")
   1702 try:

TypeError: Could not convert string 'ChocolateRocky RoadChocolte Fudge
Brownie' to numeric

The above exception was the direct cause of the following exception:

TypeError                                 Traceback (most recent call
last)
Cell In[41], line 1
----> 1 df.groupby(['Base Flavor','Liked']).mean()

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:2452, in GroupBy.mean(self, numeric_only, engine,
engine_kwargs)
   2445     return self._numba_agg_general(
   2446         grouped_mean,
   2447         executor.float_dtype_mapping,
   2448         engine_kwargs,
   2449         min_periods=0,
   2450     )
   2451 else:
-> 2452     result = self._cython_agg_general(
   2453         "mean",
   2454         alt=lambda x: Series(x,
copy=False).mean(numeric_only=numeric_only),
   2455         numeric_only=numeric_only,
   2456     )
   2457     return result.__finalize__(self.obj, method="groupby")

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:1998, in GroupBy._cython_agg_general(self, how, alt,
numeric_only, min_count, **kwargs)
   1995     result = self._agg_py_fallback(how, values,
ndim=data.ndim, alt=alt)
   1996     return result
```

```
-> 1998 new_mgr = data.grouped_reduce(array_func)
   1999 res = self._wrap_agged_manager(new_mgr)
   2000 if how in ["idxmin", "idxmax"]:

File ~\anaconda3\Lib\site-packages\pandas\core\internals\
managers.py:1469, in BlockManager.grouped_reduce(self, func)
   1465 if blk.is_object:
   1466     # split on object-dtype blocks bc some columns may raise
   1467     #  while others do not.
   1468     for sb in blk._split():
-> 1469         applied = sb.apply(func)
   1470         result_blocks = extend_blocks(applied, result_blocks)
   1471 else:

File ~\anaconda3\Lib\site-packages\pandas\core\internals\
blocks.py:393, in Block.apply(self, func, **kwargs)
    387 @final
    388 def apply(self, func, **kwargs) -> list[Block]:
    389     """
    390     apply the function to my values; return a block if we are
not
    391     one
    392     """
--> 393     result = func(self.values, **kwargs)
    395     result = maybe_coerce_values(result)
    396     return self._split_op_result(result)

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:1995, in
GroupBy._cython_agg_general.<locals>.array_func(values)
   1992     return result
   1994 assert alt is not None
-> 1995 result = self._agg_py_fallback(how, values, ndim=data.ndim,
alt=alt)
   1996 return result

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\
groupby.py:1946, in GroupBy._agg_py_fallback(self, how, values, ndim,
alt)
   1944     msg = f"agg function failed [how->{how},dtype-
>{ser.dtype}]"
   1945     # preserve the kind of exception that raised
-> 1946     raise type(err)(msg) from err
   1948 if ser.dtype == object:
   1949     res_values = res_values.astype(object, copy=False)

TypeError: agg function failed [how->mean,dtype->object]

df.groupby('Base Flavor').describe()
```

```
            Flavor
Rating                                                  \
                    count mean       std  min    25%  50%  75%   max

Base Flavor

Chocolate             3.0  8.4  0.346410  8.2  8.200  8.2  8.5   8.8

Vanilla               6.0  5.7  2.710719  2.3  4.025  5.6  6.8  10.0


            Texture Rating            ...            Total Rating
\
                     count      mean  ...    75%  max         count
mean
Base Flavor                           ...

Chocolate              3.0  7.233333  ...  7.350  7.6           3.0
15.70
Vanilla                6.0  5.650000  ...  6.375  8.0           6.0
11.35


                  std   min     25%   50%     75%   max
Base Flavor
Chocolate    0.781025  15.2  15.250  15.3  15.950  16.6
Vanilla      4.263684   5.7   9.025  11.1  13.175  18.0

[2 rows x 24 columns]
```

Merge Join Concatenate

```
import pandas as pd

df1=pd.read_csv(r"C:\Users\User\Downloads\LOTR.csv")
df2=pd.read_csv(r"C:\Users\User\Downloads\LOTR 2.csv")

df1
```

```
   FellowshipID FirstName     Skills
0          1001     Frodo     Hiding
1          1002   Samwise  Gardening
2          1003   Gandalf     Spells
3          1004    Pippin  Fireworks
```

```
df2
```

```
   FellowshipID FirstName   Age
0          1001     Frodo    50
1          1002   Samwise    39
2          1006   Legolas  2931
3          1007    Elrond  6520
4          1008  Barromir    51
```

```
df1.merge(df2, how='inner') #does inner join
```

```
   FellowshipID FirstName     Skills  Age
0          1001     Frodo     Hiding   50
1          1002   Samwise  Gardening   39
```

```
df1.merge(df2, how='inner', on='FellowshipID')
```

```
   FellowshipID FirstName_x     Skills FirstName_y  Age
0          1001       Frodo     Hiding       Frodo   50
1          1002     Samwise  Gardening     Samwise   39
```

```
df1.merge(df2, how='outer') #does outer join returns everything
```

```
   FellowshipID FirstName     Skills     Age
0          1001     Frodo     Hiding    50.0
1          1002   Samwise  Gardening    39.0
2          1003   Gandalf     Spells     NaN
3          1004    Pippin  Fireworks     NaN
4          1006   Legolas        NaN  2931.0
5          1007    Elrond        NaN  6520.0
6          1008  Barromir        NaN    51.0
```

```
df1.merge(df2, how='left') #does left join all left it takes and
overlap of right
```

```
   FellowshipID FirstName     Skills   Age
0          1001     Frodo     Hiding  50.0
1          1002   Samwise  Gardening  39.0
2          1003   Gandalf     Spells   NaN
3          1004    Pippin  Fireworks   NaN
```

```
df1.merge(df2, how='right') #does right join everything right overlap
with left
```

```
   FellowshipID FirstName     Skills   Age
0          1001     Frodo     Hiding    50
1          1002   Samwise  Gardening    39
2          1006   Legolas        NaN  2931
3          1007    Elrond        NaN  6520
4          1008  Barromir        NaN    51
```

```
df1.merge(df2, how='cross') # each val in left dataframe with each
value in right dataframe
```

```
    FellowshipID_x FirstName_x     Skills  FellowshipID_y FirstName_y
Age
0             1001       Frodo     Hiding            1001       Frodo
50
1             1001       Frodo     Hiding            1002     Samwise
39
2             1001       Frodo     Hiding            1006     Legolas
2931
3             1001       Frodo     Hiding            1007      Elrond
6520
4             1001       Frodo     Hiding            1008    Barromir
51
5             1002     Samwise  Gardening            1001       Frodo
50
6             1002     Samwise  Gardening            1002     Samwise
39
7             1002     Samwise  Gardening            1006     Legolas
2931
8             1002     Samwise  Gardening            1007      Elrond
6520
9             1002     Samwise  Gardening            1008    Barromir
51
10            1003     Gandalf     Spells            1001       Frodo
50
11            1003     Gandalf     Spells            1002     Samwise
39
12            1003     Gandalf     Spells            1006     Legolas
2931
13            1003     Gandalf     Spells            1007      Elrond
6520
14            1003     Gandalf     Spells            1008    Barromir
```

```
51
15              1004      Pippin  Fireworks                1001        Frodo
50
16              1004      Pippin  Fireworks                1002      Samwise
39
17              1004      Pippin  Fireworks                1006      Legolas
2931
18              1004      Pippin  Fireworks                1007       Elrond
6520
19              1004      Pippin  Fireworks                1008     Barromir
51
```

```
df1.join(df2,on='FellowshipID',how='outer',lsuffix='_left',rsuffix='_right')
```

```
     FellowshipID  FellowshipID_left FirstName_left     Skills  \
NaN             0                NaN            NaN        NaN
NaN             1                NaN            NaN        NaN
NaN             2                NaN            NaN        NaN
NaN             3                NaN            NaN        NaN
NaN             4                NaN            NaN        NaN
0.0          1001             1001.0          Frodo     Hiding
1.0          1002             1002.0        Samwise  Gardening
2.0          1003             1003.0        Gandalf     Spells
3.0          1004             1004.0         Pippin  Fireworks

     FellowshipID_right FirstName_right     Age
NaN              1001.0           Frodo    50.0
NaN              1002.0         Samwise    39.0
NaN              1006.0         Legolas  2931.0
NaN              1007.0          Elrond  6520.0
NaN              1008.0        Barromir    51.0
0.0                 NaN             NaN     NaN
1.0                 NaN             NaN     NaN
2.0                 NaN             NaN     NaN
3.0                 NaN             NaN     NaN
```

```
df4=df1.set_index('FellowshipID').join(df2.set_index('FellowshipID'),lsuffix='_left',rsuffix='_right',how='outer')
df4
```

```
             FirstName_left     Skills FirstName_right   Age
FellowshipID
1001                  Frodo     Hiding           Frodo  50.0
1002                Samwise  Gardening         Samwise  39.0
1003                Gandalf     Spells             NaN   NaN
1004                 Pippin  Fireworks             NaN   NaN
```

Concat

```
pd.concat([df1,df2])
```

	FellowshipID	FirstName	Skills	Age
0	1001	Frodo	Hiding	NaN
1	1002	Samwise	Gardening	NaN
2	1003	Gandalf	Spells	NaN
3	1004	Pippin	Fireworks	NaN
0	1001	Frodo	NaN	50.0
1	1002	Samwise	NaN	39.0
2	1006	Legolas	NaN	2931.0
3	1007	Elrond	NaN	6520.0
4	1008	Barromir	NaN	51.0

```
pd.concat([df1,df2],join='inner')# joining cols which are the same
```

	FellowshipID	FirstName
0	1001	Frodo
1	1002	Samwise
2	1003	Gandalf
3	1004	Pippin
0	1001	Frodo
1	1002	Samwise
2	1006	Legolas
3	1007	Elrond
4	1008	Barromir

```
pd.concat([df1,df2],join='outer')#takes all
```

	FellowshipID	FirstName	Skills	Age
0	1001	Frodo	Hiding	NaN
1	1002	Samwise	Gardening	NaN
2	1003	Gandalf	Spells	NaN
3	1004	Pippin	Fireworks	NaN
0	1001	Frodo	NaN	50.0
1	1002	Samwise	NaN	39.0
2	1006	Legolas	NaN	2931.0
3	1007	Elrond	NaN	6520.0
4	1008	Barromir	NaN	51.0

```
pd.concat([df1,df2],join='outer', axis=1)#joining based on index
```

	FellowshipID	FirstName	Skills	FellowshipID	FirstName	Age
0	1001.0	Frodo	Hiding	1001	Frodo	50
1	1002.0	Samwise	Gardening	1002	Samwise	39
2	1003.0	Gandalf	Spells	1006	Legolas	2931
3	1004.0	Pippin	Fireworks	1007	Elrond	6520
4	NaN	NaN	NaN	1008	Barromir	51

```
df1.append(df2)
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_10484\3062608662.py in ?()
----> 1 df1.append(df2)

~\anaconda3\Lib\site-packages\pandas\core\generic.py in ?(self, name)
   6295             and name not in self._accessors
   6296             and self._info_axis._can_hold_identifiers_and_holds_name(name)
   6297         ):
   6298             return self[name]
-> 6299         return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'append'
```

Pandas Visualization

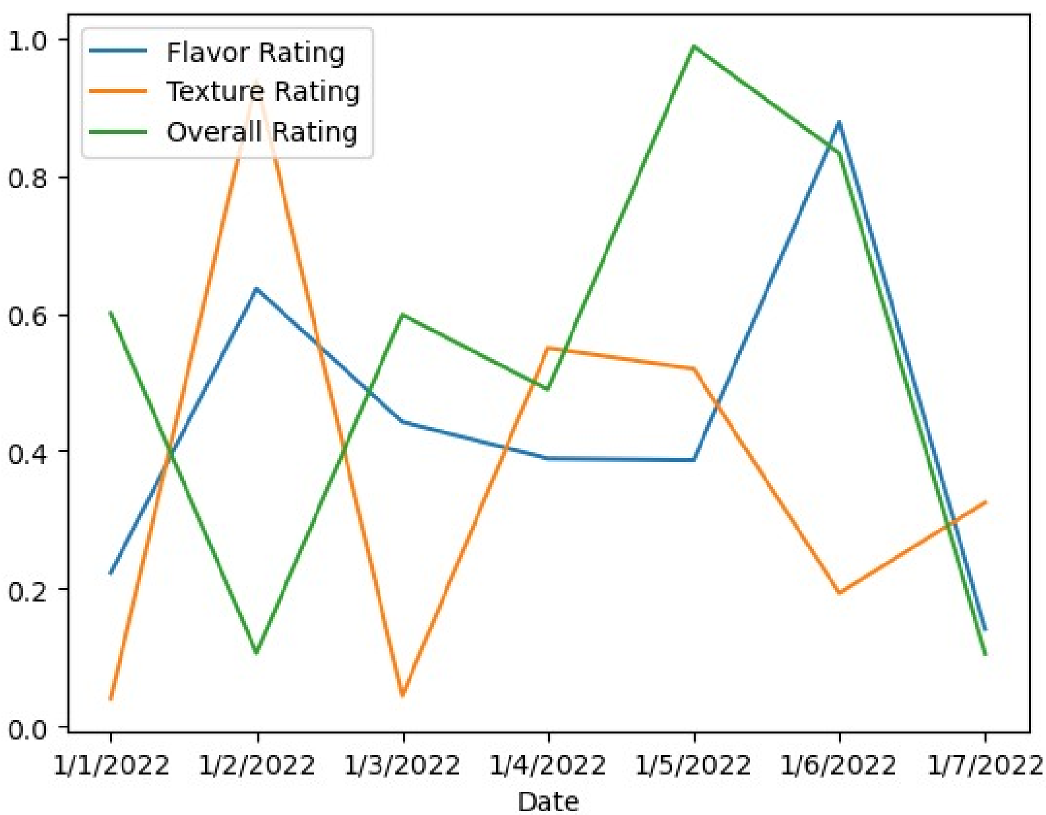
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df= pd.read_csv(r"C:\Users\User\Downloads\Ice Cream Ratings.csv")
df=df.set_index('Date')
df
```

	Flavor Rating	Texture Rating	Overall Rating
Date			
1/1/2022	0.223090	0.040220	0.600129
1/2/2022	0.635886	0.938476	0.106264
1/3/2022	0.442323	0.044154	0.598112
1/4/2022	0.389128	0.549676	0.489353
1/5/2022	0.386887	0.519439	0.988280
1/6/2022	0.877984	0.193588	0.832827
1/7/2022	0.140995	0.325110	0.105147

```
df.plot(kind='line')
```

```
<Axes: xlabel='Date'>
```

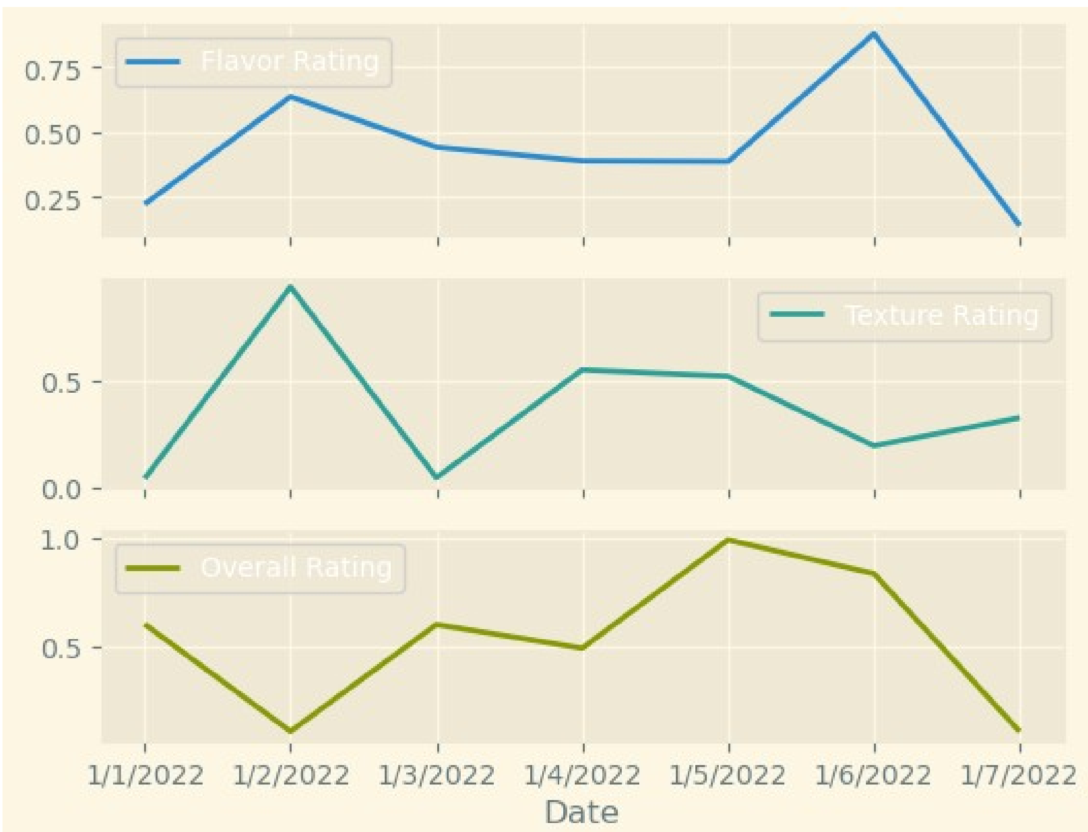


```
print(plt.style.available)
plt.style.use('Solarize_Light2')
```

```
['Solarize_Light2', '_classic_test_patch', '_mpl-gallery', '_mpl-gallery-nogrid', 'bmh', 'classic', 'dark_background', 'fast', 'fivethirtyeight', 'ggplot', 'grayscale', 'seaborn-v0_8', 'seaborn-v0_8-bright', 'seaborn-v0_8-colorblind', 'seaborn-v0_8-dark', 'seaborn-v0_8-dark-palette', 'seaborn-v0_8-darkgrid', 'seaborn-v0_8-deep', 'seaborn-v0_8-muted', 'seaborn-v0_8-notebook', 'seaborn-v0_8-paper', 'seaborn-v0_8-pastel', 'seaborn-v0_8-poster', 'seaborn-v0_8-talk', 'seaborn-v0_8-ticks', 'seaborn-v0_8-white', 'seaborn-v0_8-whitegrid', 'tableau-colorblind10']
```

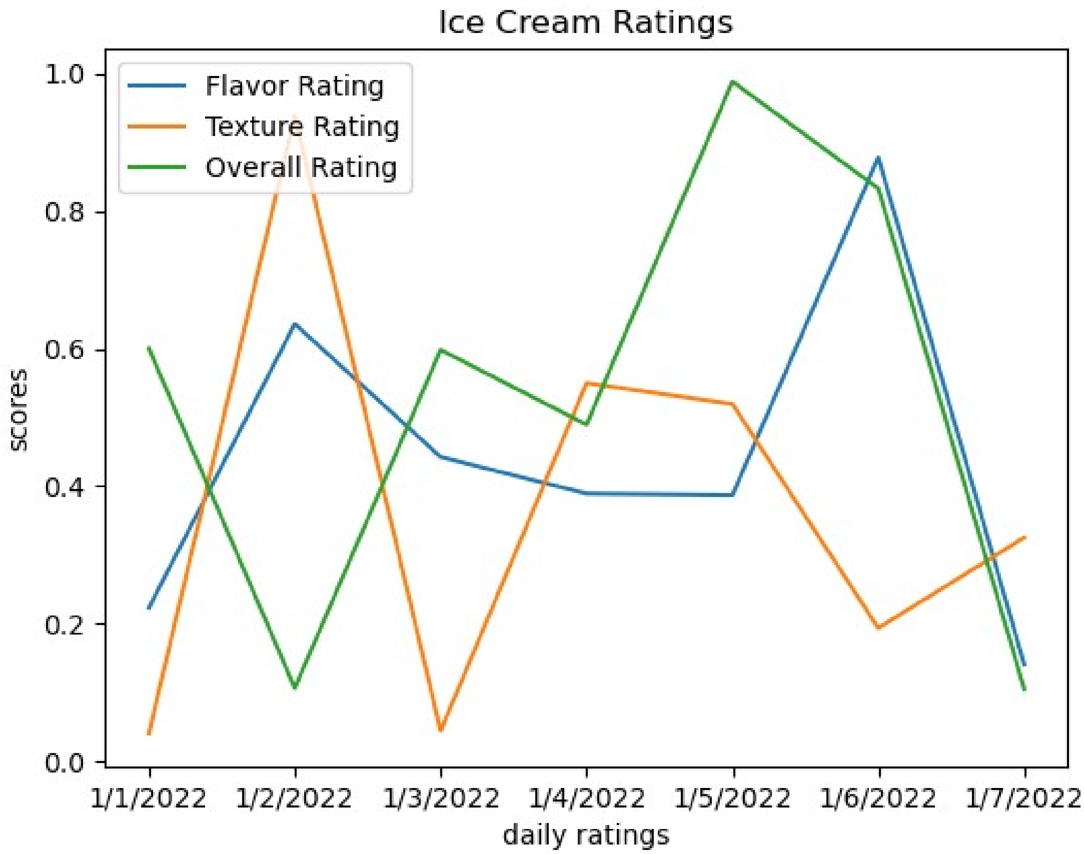
```
df.plot(kind='line',subplots=True)
```

```
array([<Axes: xlabel='Date'>, <Axes: xlabel='Date'>,
       <Axes: xlabel='Date'>], dtype=object)
```



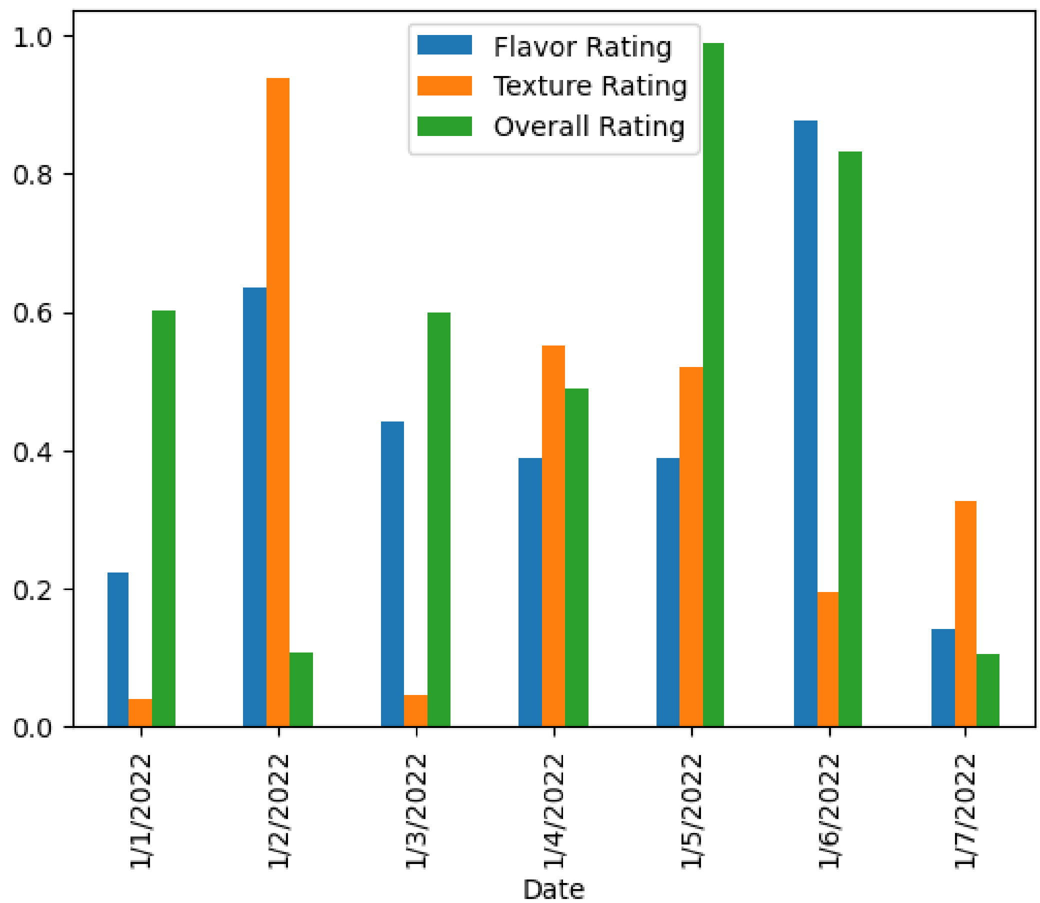

```
df.plot(kind='line',title='Ice Cream Ratings', xlabel='daily ratings',ylabel='scores')
```

```
<Axes: title={'center': 'Ice Cream Ratings'}, xlabel='daily ratings', ylabel='scores'>
```



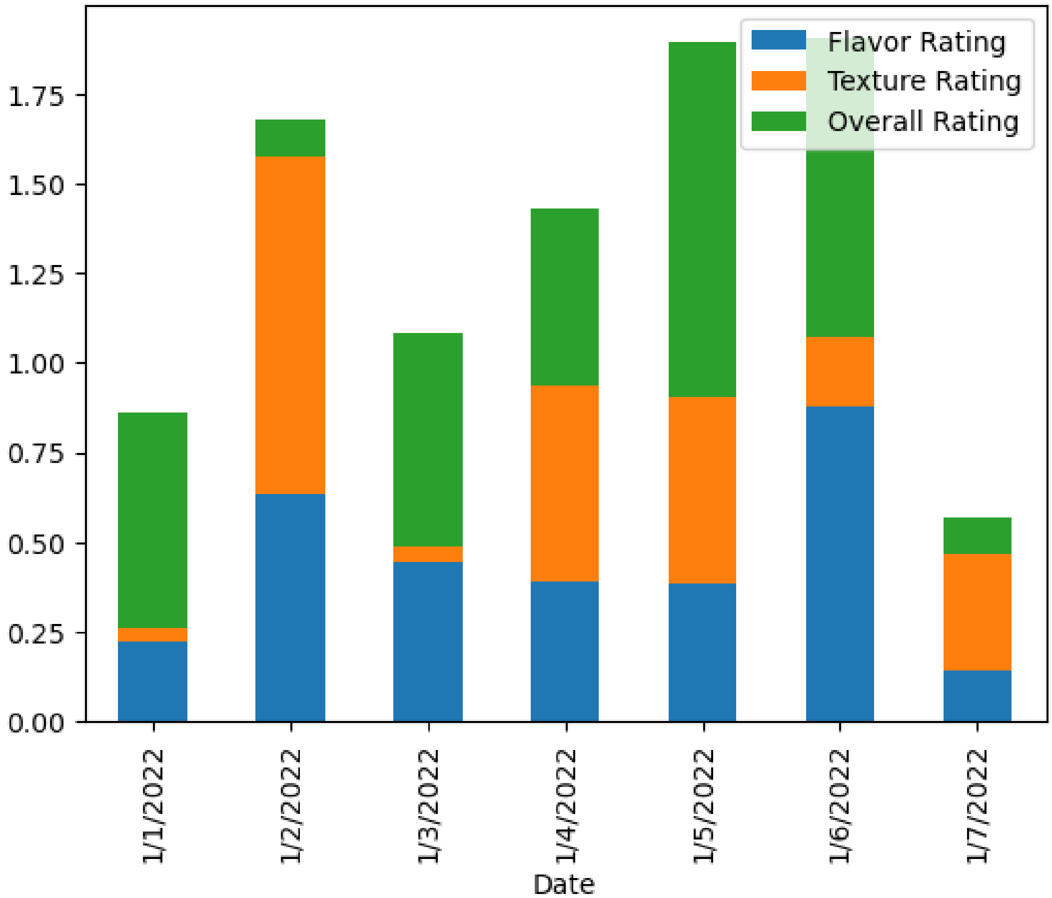

```
df.plot(kind='bar')
```

```
<Axes: xlabel='Date'>
```



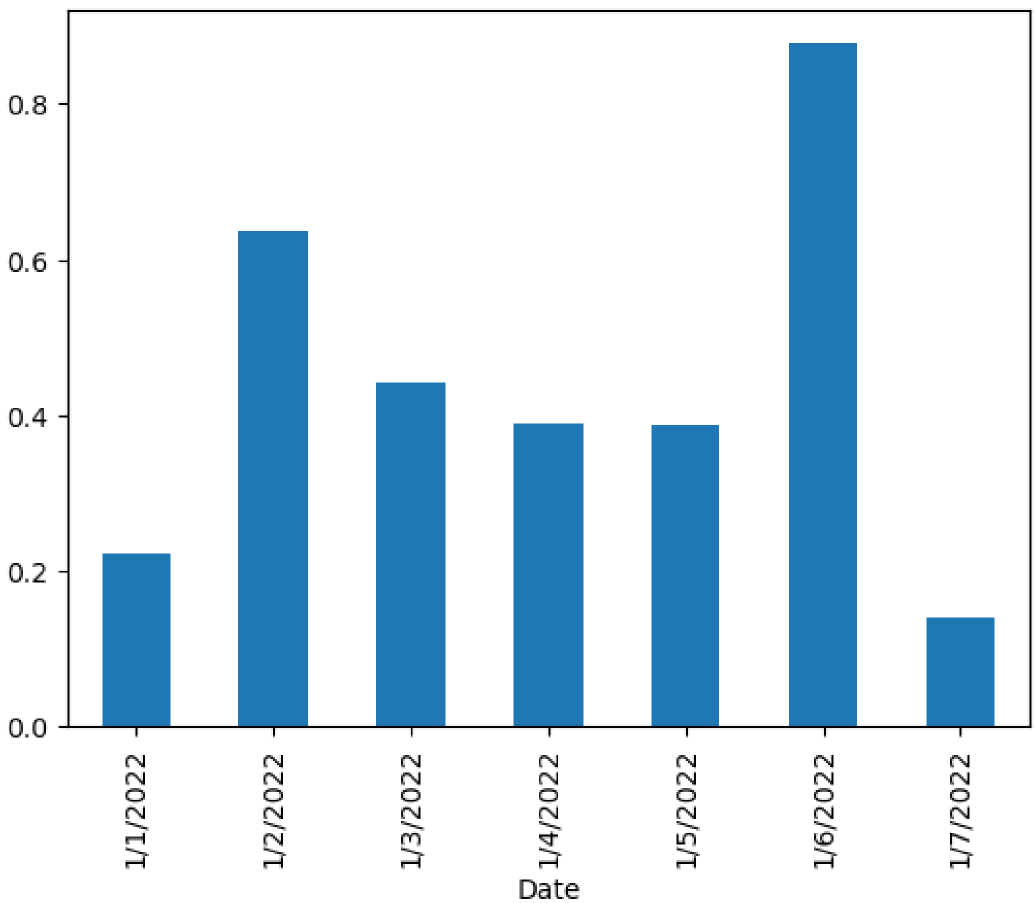

```
df.plot(kind='bar', stacked= True)
```

```
<Axes: xlabel='Date'>
```



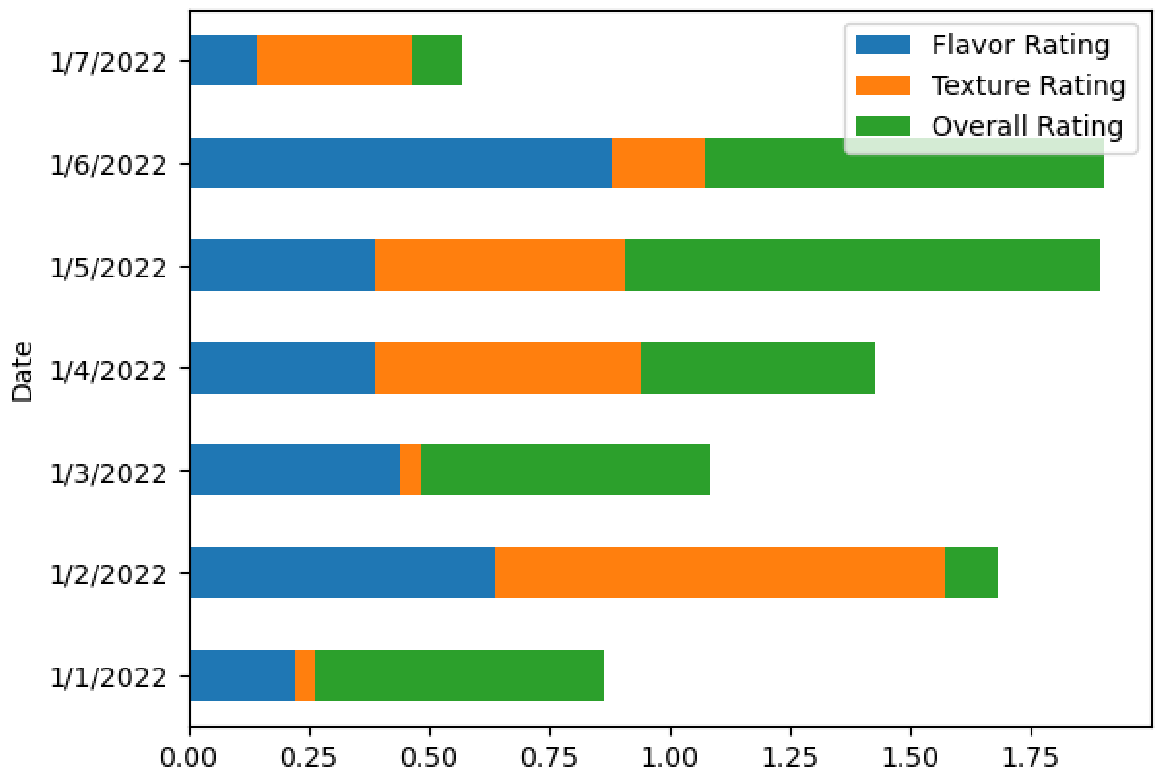

```
df['Flavor Rating'].plot(kind='bar', stacked= True)
```

```
<Axes: xlabel='Date'>
```



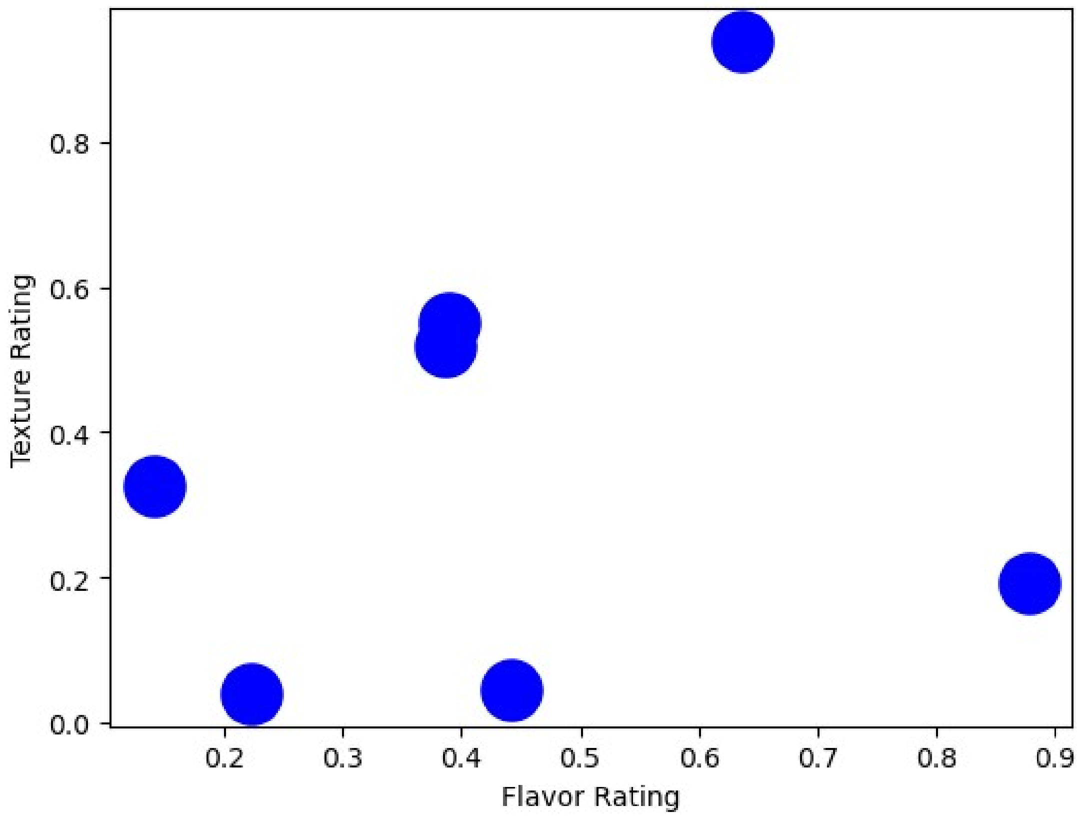

```
df.plot.barh(stacked= True)
```

```
<Axes: ylabel='Date'>
```



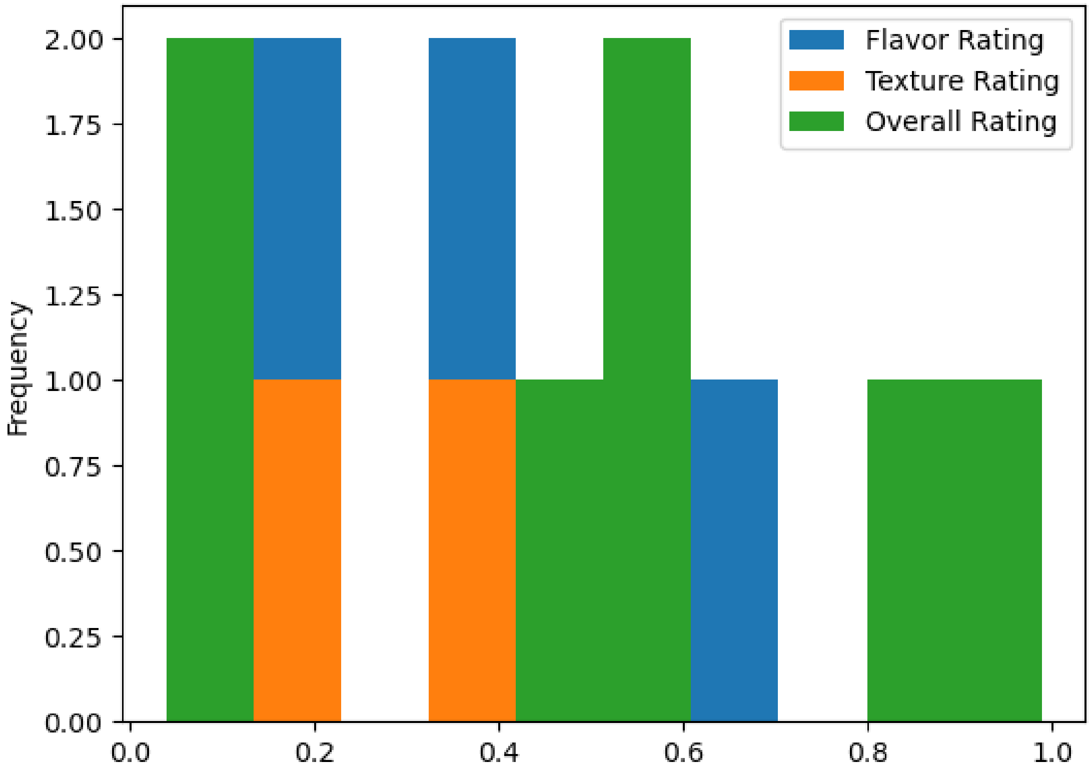

```
df.plot.scatter(x='Flavor Rating',y='Texture Rating',s=500,c='BLue')

<Axes: xlabel='Flavor Rating', ylabel='Texture Rating'>
```



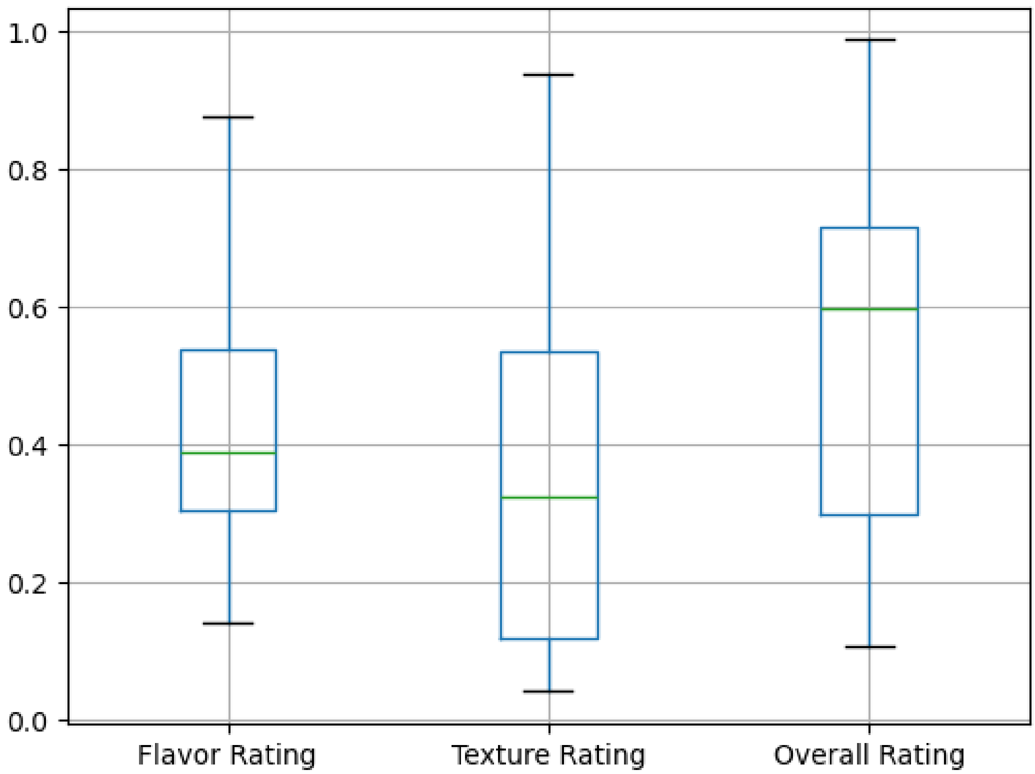

```
df.plot.hist()
```

```
<Axes: ylabel='Frequency'>
```



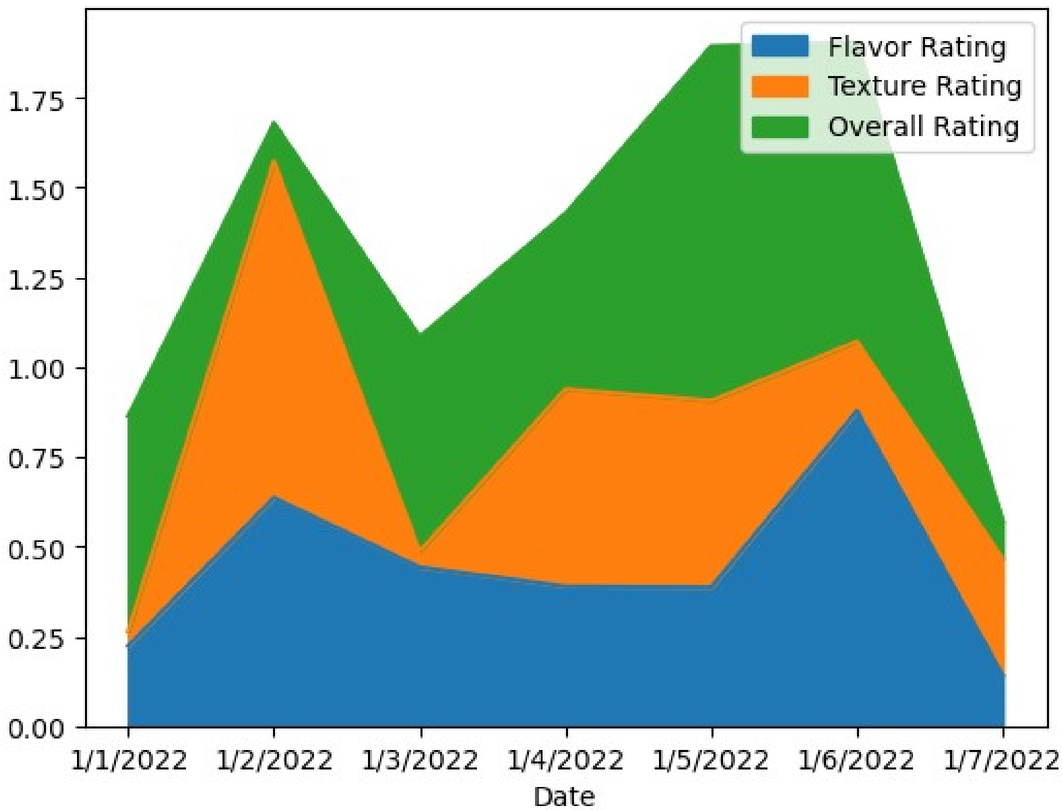

```
df.boxplot()
```

```
<Axes: >
```



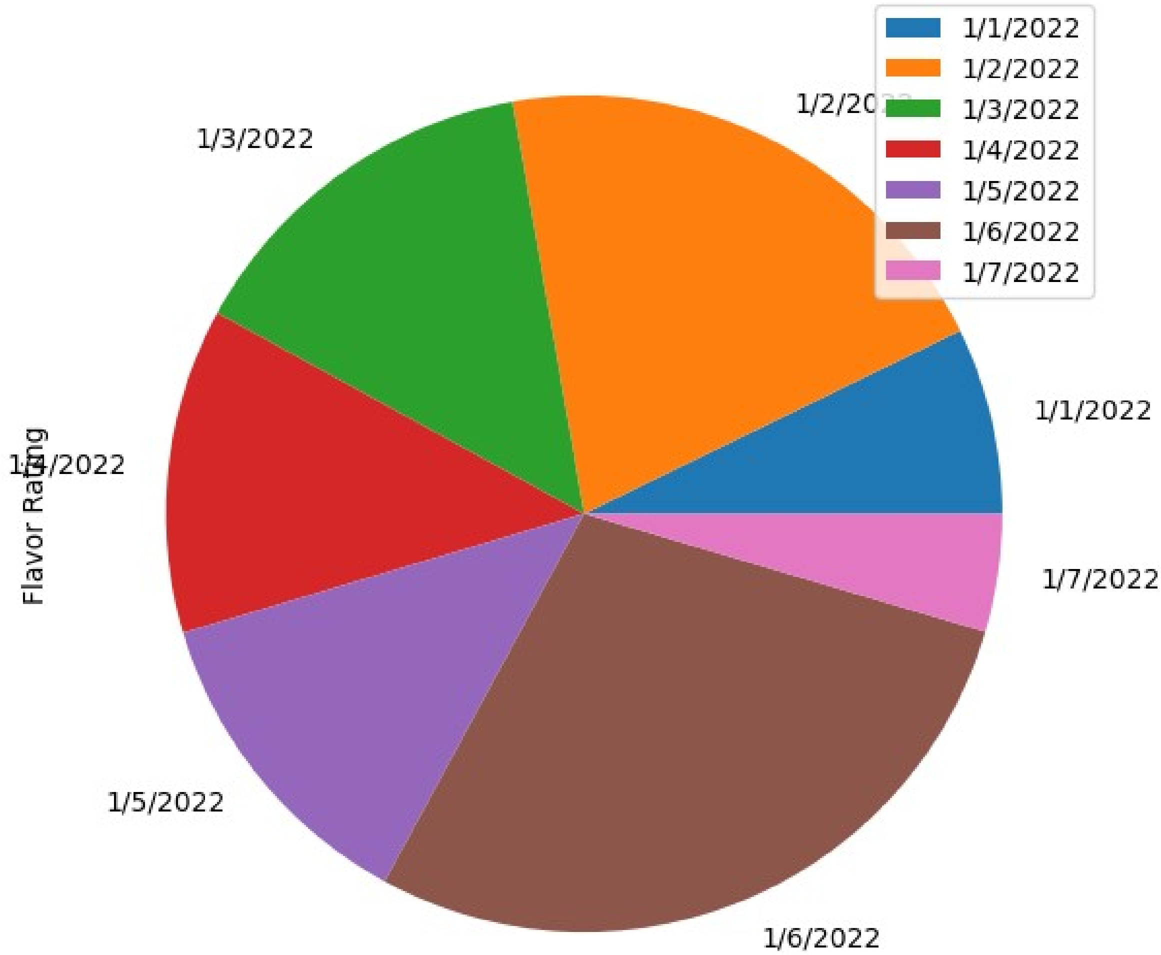

```
df.plot.area()
```

```
<Axes: xlabel='Date'>
```




```
df.plot.pie(y='Flavor Rating',figsize=(10,7))
```

```
<Axes: ylabel='Flavor Rating'>
```


1/1/2022
1/2/2022
1/3/2022
1/4/2022
1/5/2022
1/6/2022
1/7/2022
Flavor Rating