PandasTraining

October 7, 2020

1 Pandas Training

1.1 Loading Data into Pandas

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

```
[2]: df = pd.read_csv('matches.csv')
```

```
[3]: # df
     df.head()
     # df.head(3)
     # df.tail()
```

[3]:
```
    id  season       city        date                        team1  \
 0   1    2017   Hyderabad  2017-04-05          Sunrisers Hyderabad
 1   2    2017        Pune  2017-04-06               Mumbai Indians
 2   3    2017      Rajkot  2017-04-07                Gujarat Lions
 3   4    2017      Indore  2017-04-08        Rising Pune Supergiant
 4   5    2017   Bangalore  2017-04-08  Royal Challengers Bangalore

                           team2                  toss_winner toss_decision  \
 0  Royal Challengers Bangalore  Royal Challengers Bangalore         field
 1       Rising Pune Supergiant       Rising Pune Supergiant         field
 2        Kolkata Knight Riders        Kolkata Knight Riders         field
 3              Kings XI Punjab              Kings XI Punjab         field
 4              Delhi Daredevils  Royal Challengers Bangalore           bat

    result  dl_applied                       winner  win_by_runs  \
 0  normal           0          Sunrisers Hyderabad           35
 1  normal           0       Rising Pune Supergiant            0
 2  normal           0        Kolkata Knight Riders            0
 3  normal           0              Kings XI Punjab            0
 4  normal           0  Royal Challengers Bangalore           15

    win_by_wickets player_of_match                                        venue  \
 0               0    Yuvraj Singh  Rajiv Gandhi International Stadium, Uppal
 1               7       SPD Smith    Maharashtra Cricket Association Stadium
 2              10         CA Lynn     Saurashtra Cricket Association Stadium
```

```
3                6        GJ Maxwell                Holkar Cricket Stadium
4                0        KM Jadhav               M Chinnaswamy Stadium

           umpire1           umpire2 umpire3
0     AY Dandekar          NJ Llong     NaN
1  A Nand Kishore           S Ravi     NaN
2     Nitin Menon        CK Nandan     NaN
3   AK Chaudhary  C Shamshuddin     NaN
4             NaN              NaN     NaN
```

```
[4]: df = pd.read_csv('pokemon_data.csv')
     df.head()
```

```
[4]:    #          Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
     0  1     Bulbasaur  Grass  Poison  45      49       49       65
     1  2       Ivysaur  Grass  Poison  60      62       63       80
     2  3      Venusaur  Grass  Poison  80      82       83      100
     3  3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
     4  4    Charmander   Fire     NaN  39      52       43       60

        Sp. Def  Speed  Generation  Legendary
     0       65     45           1      False
     1       80     60           1      False
     2      100     80           1      False
     3      120     80           1      False
     4       50     65           1      False
```

For Reading Excel Files Need to Install a Third Party Package called xlrd

You can make use of 'pip install xlrd' from scripts directory otherwise in Jupyter Notebook as shown below.,

```
[5]: !pip install xlrd
```

```
Requirement already satisfied: xlrd in
/Users/rajath/.virtualenvs/ai/lib/python3.8/site-packages (1.2.0)
```

```
[6]: df_xlsx = pd.read_excel('pokemon_data.xlsx')
     df_xlsx.head()
```

```
[6]:    #          Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
     0  1     Bulbasaur  Grass  Poison  45      49       49       65
     1  2       Ivysaur  Grass  Poison  60      62       63       80
     2  3      Venusaur  Grass  Poison  80      82       83      100
     3  3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
     4  4    Charmander   Fire     NaN  39      52       43       60

        Sp. Def  Speed  Generation  Legendary
```

```
        0          65      45              1       False
        1          80      60              1       False
        2         100      80              1       False
        3         120      80              1       False
        4          50      65              1       False
```

```
[7]: df_tab = pd.read_csv('pokemon_data.txt')
     df_tab.head()
```

```
[7]:    #\tName\tType 1\tType 2\tHP\tAttack\tDefense\tSp. Atk\tSp.
     Def\tSpeed\tGeneration\tLegendary
     0  1\tBulbasaur\tGrass\tPoison\t45\t49\t49\t65\t6...
     1  2\tIvysaur\tGrass\tPoison\t60\t62\t63\t80\t80\...
     2  3\tVenusaur\tGrass\tPoison\t80\t82\t83\t100\t1...
     3  3\tVenusaurMega Venusaur\tGrass\tPoison\t80\t1...
     4  4\tCharmander\tFire\t\t39\t52\t43\t60\t50\t65\...
```

```
[8]: df_tab = pd.read_csv('pokemon_data.txt', delimiter='\t')
     df_tab.head()
```

```
[8]:    #                  Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
     0  1             Bulbasaur  Grass  Poison  45      49       49       65
     1  2               Ivysaur  Grass  Poison  60      62       63       80
     2  3               Venusaur  Grass  Poison  80      82       83      100
     3  3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
     4  4             Charmander   Fire     NaN  39      52       43       60

        Sp. Def  Speed  Generation  Legendary
     0       65     45           1      False
     1       80     60           1      False
     2      100     80           1      False
     3      120     80           1      False
     4       50     65           1      False
```

1.1.1 Reading Data in Pandas

```
[9]: df.columns
```

```
[9]: Index(['#', 'Name', 'Type 1', 'Type 2', 'HP', 'Attack', 'Defense', 'Sp. Atk',
            'Sp. Def', 'Speed', 'Generation', 'Legendary'],
           dtype='object')
```

Reading a Particular Column

```
[10]: # df['Name']
      df['Name'][0:5]
      # df.Name
```

```
[10]: 0                Bulbasaur
      1                  Ivysaur
      2                 Venusaur
      3    VenusaurMega Venusaur
      4               Charmander
      Name: Name, dtype: object
```

Reading Multiple Columns

```
[11]: print(df[['Name','Speed','Generation']][0:5])
```

```
                       Name  Speed  Generation
0                 Bulbasaur     45           1
1                   Ivysaur     60           1
2                  Venusaur     80           1
3     VenusaurMega Venusaur     80           1
4                Charmander     65           1
```

Reading Each Row

```
[12]: # Reading a Particular Row
      print(df.iloc[1])
```

```
#                    2
Name           Ivysaur
Type 1           Grass
Type 2          Poison
HP                  60
Attack              62
Defense             63
Sp. Atk             80
Sp. Def             80
Speed               60
Generation           1
Legendary        False
Name: 1, dtype: object
```

```
[13]: # Reading a Set of Rows
      print(df.iloc[0:4])
```

```
    #                   Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
0   1              Bulbasaur  Grass  Poison  45      49       49       65
1   2                Ivysaur  Grass  Poison  60      62       63       80
2   3               Venusaur  Grass  Poison  80      82       83      100
3   3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122

   Sp. Def  Speed  Generation  Legendary
0       65     45           1      False
1       80     60           1      False
```

```
2         100       80                1         False
3         120       80                1         False
```

Iterating over the data frame

```
[14]:  # for index, row in df.iterrows():
       #     print(index, row['Name'])
       #     print(index, row)
       df.loc[df['Type 1'] == 'Fire'][0:5]
```

```
[14]:     #                     Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
       4  4               Charmander   Fire     NaN  39      52       43       60
       5  5               Charmeleon   Fire     NaN  58      64       58       80
       6  6                Charizard   Fire  Flying  78      84       78      109
       7  6  CharizardMega Charizard X   Fire   Dragon  78     130      111      130
       8  6  CharizardMega Charizard Y   Fire   Flying  78     104       78      159

          Sp. Def  Speed  Generation  Legendary
       4       50     65           1      False
       5       65     80           1      False
       6       85    100           1      False
       7       85    100           1      False
       8      115    100           1      False
```

1.1.2   Sorting / Describing Data

```
[15]:  df.describe()
```

```
[15]:               #          HP      Attack     Defense     Sp. Atk     Sp. Def  \
       count  800.000000  800.000000  800.000000  800.000000  800.000000  800.000000
       mean   362.813750   69.258750   79.001250   73.842500   72.820000   71.902500
       std    208.343798   25.534669   32.457366   31.183501   32.722294   27.828916
       min      1.000000    1.000000    5.000000    5.000000   10.000000   20.000000
       25%    184.750000   50.000000   55.000000   50.000000   49.750000   50.000000
       50%    364.500000   65.000000   75.000000   70.000000   65.000000   70.000000
       75%    539.250000   80.000000  100.000000   90.000000   95.000000   90.000000
       max    721.000000  255.000000  190.000000  230.000000  194.000000  230.000000

                  Speed   Generation
       count  800.000000   800.00000
       mean    68.277500     3.32375
       std     29.060474     1.66129
       min      5.000000     1.00000
       25%     45.000000     2.00000
       50%     65.000000     3.00000
       75%     90.000000     5.00000
       max    180.000000     6.00000
```

```python
[16]:  # df.sort_values('Name')[0:5]
       # df.sort_values('Name', ascending=False)[0:5]
       df.sort_values(['Type 1', 'HP'], ascending=[True, False])[0:5]
```

```
[16]:          #                 Name Type 1   Type 2  HP  Attack  Defense  \
       520    469              Yanmega    Bug   Flying  86      76       86
       698    637             Volcarona    Bug     Fire  85      60       65
       231    214             Heracross    Bug  Fighting  80     125       75
       232    214  HeracrossMega Heracross    Bug  Fighting  80     185      115
       678    617             Accelgor    Bug      NaN  80      70       40

            Sp. Atk  Sp. Def  Speed  Generation  Legendary
       520      116       56     95           4      False
       698      135      105    100           5      False
       231       40       95     85           2      False
       232       40      105     75           2      False
       678      100       60    145           5      False
```

Making Changes to the Data

```python
[17]:  df['Total'] = df['HP']+df['Attack']
```

```python
[18]:  df.head(5)
```

```
[18]:     #               Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
       0   1          Bulbasaur  Grass  Poison  45      49       49       65
       1   2            Ivysaur  Grass  Poison  60      62       63       80
       2   3           Venusaur  Grass  Poison  80      82       83      100
       3   3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
       4   4         Charmander   Fire     NaN  39      52       43       60

          Sp. Def  Speed  Generation  Legendary  Total
       0       65     45           1      False     94
       1       80     60           1      False    122
       2      100     80           1      False    162
       3      120     80           1      False    180
       4       50     65           1      False     91
```

```python
[19]:  df = df.drop(columns=['Total'])
```

```python
[20]:  df.head(5)
```

```
[20]:     #               Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
       0   1          Bulbasaur  Grass  Poison  45      49       49       65
       1   2            Ivysaur  Grass  Poison  60      62       63       80
       2   3           Venusaur  Grass  Poison  80      82       83      100
       3   3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
```

```
4  4             Charmander    Fire      NaN  39      52          43           60
```

```
   Sp. Def  Speed  Generation  Legendary
0       65     45           1      False
1       80     60           1      False
2      100     80           1      False
3      120     80           1      False
4       50     65           1      False
```

[21]: ```python
df['Total'] = df.iloc[:,4:6].sum(axis=1)
```

[22]: ```python
df.head(5)
```

[22]:
```
   #              Name Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
0  1         Bulbasaur  Grass  Poison  45      49       49       65
1  2           Ivysaur  Grass  Poison  60      62       63       80
2  3          Venusaur  Grass  Poison  80      82       83      100
3  3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
4  4        Charmander   Fire     NaN  39      52       43       60
```

```
   Sp. Def  Speed  Generation  Legendary  Total
0       65     45           1      False     94
1       80     60           1      False    122
2      100     80           1      False    162
3      120     80           1      False    180
4       50     65           1      False     91
```

[23]: ```python
cols = df.columns.values
```

[24]: ```python
cols
```

[24]: ```python
array(['#', 'Name', 'Type 1', 'Type 2', 'HP', 'Attack', 'Defense',
       'Sp. Atk', 'Sp. Def', 'Speed', 'Generation', 'Legendary', 'Total'],
      dtype=object)
```

[25]: ```python
cols = list(df.columns)
```

[26]: ```python
cols
```

[26]: ```python
['#',
 'Name',
 'Type 1',
 'Type 2',
 'HP',
 'Attack',
 'Defense',
 'Sp. Atk',
```

```
        'Sp. Def',
        'Speed',
        'Generation',
        'Legendary',
        'Total']
```

[27]:
```
df = df[cols[0:4] + [cols[-1]] + cols[4:12]]
```

[28]:
```
df.head()
```

[28]:
```
   #            Name Type 1 Type 2  Total  HP  Attack  Defense  \
0  1        Bulbasaur  Grass  Poison     94  45      49       49
1  2          Ivysaur  Grass  Poison    122  60      62       63
2  3         Venusaur  Grass  Poison    162  80      82       83
3  3  VenusaurMega Venusaur  Grass  Poison    180  80     100      123
4  4       Charmander   Fire     NaN     91  39      52       43

   Sp. Atk  Sp. Def  Speed  Generation  Legendary
0       65       65     45           1      False
1       80       80     60           1      False
2      100      100     80           1      False
3      122      120     80           1      False
4       60       50     65           1      False
```

[29]:
```
df.to_csv('modified.csv', index=False)
```

To Save File in Excel need to install third party library called 'openpyxl'

[30]:
```
!pip install openpyxl
```

```
Requirement already satisfied: openpyxl in
/Users/rajath/.virtualenvs/ai/lib/python3.8/site-packages (3.0.5)
Requirement already satisfied: jdcal in
/Users/rajath/.virtualenvs/ai/lib/python3.8/site-packages (from openpyxl)
(1.4.1)
Requirement already satisfied: et-xmlfile in
/Users/rajath/.virtualenvs/ai/lib/python3.8/site-packages (from openpyxl)
(1.0.1)
```

[31]:
```
df.to_excel('modified.xlsx', index=False)
```

[32]:
```
df.to_csv('modified.txt', index=False, sep='\t')
```

Filtering of Data

[33]:
```
df.loc[(df['Type 1'] == 'Grass') & (df['Type 2'] == 'Poison')][0:10]
df.loc[(df['Type 1'] == 'Grass') | (df['Type 2'] == 'Poison')][0:10]
```

```
new_df = df.loc[(df['Type 1'] == 'Grass') & (df['Type 2'] == 'Poison') &
    (df['HP'] > 70)][0:10]
```

[34]: `new_df`

[34]:
```
       #               Name Type 1 Type 2  Total   HP  Attack  Defense  \
2      3            Venusaur  Grass  Poison    162   80      82       83
3      3   VenusaurMega Venusaur  Grass  Poison    180   80     100      123
50    45           Vileplume  Grass  Poison    155   75      80       85
77    71          Victreebel  Grass  Poison    185   80     105       65
652  591           Amoonguss  Grass  Poison    199  114      85       70

     Sp. Atk  Sp. Def  Speed  Generation  Legendary
2        100      100     80           1      False
3        122      120     80           1      False
50       110       90     50           1      False
77       100       70     70           1      False
652       85       80     30           5      False
```

[35]: `# new_df = new_df.reset_index()`

[36]: `# new_df`

[37]: `new_df.reset_index(drop=True, inplace=True)`

[38]: `new_df`

[38]:
```
     #               Name Type 1 Type 2  Total   HP  Attack  Defense  \
0    3            Venusaur  Grass  Poison    162   80      82       83
1    3   VenusaurMega Venusaur  Grass  Poison    180   80     100      123
2   45           Vileplume  Grass  Poison    155   75      80       85
3   71          Victreebel  Grass  Poison    185   80     105       65
4  591           Amoonguss  Grass  Poison    199  114      85       70

   Sp. Atk  Sp. Def  Speed  Generation  Legendary
0      100      100     80           1      False
1      122      120     80           1      False
2      110       90     50           1      False
3      100       70     70           1      False
4       85       80     30           5      False
```

[39]: `df.loc[~ df['Name'].str.contains('Mega')][0:5]`

[39]:
```
   #        Name Type 1  Type 2  Total  HP  Attack  Defense  Sp. Atk  Sp. Def  \
0  1   Bulbasaur  Grass  Poison     94  45      49       49       65       65
1  2     Ivysaur  Grass  Poison    122  60      62       63       80       80
2  3    Venusaur  Grass  Poison    162  80      82       83      100      100
```

9

```
4  4  Charmander   Fire    NaN   91  39       52       43       60       50
5  5  Charmeleon   Fire    NaN  122  58       64       58       80       65

   Speed  Generation  Legendary
0     45           1      False
1     60           1      False
2     80           1      False
4     65           1      False
5     80           1      False
```

[40]:
```python
import re

df.loc[df['Type 1'].str.contains('Fire|Grass', regex=True)][0:5]
```

[40]:
```
   #                 Name Type 1  Type 2  Total  HP  Attack  Defense  \
0  1            Bulbasaur  Grass  Poison     94  45      49       49
1  2              Ivysaur  Grass  Poison    122  60      62       63
2  3              Venusaur  Grass  Poison   162  80      82       83
3  3  VenusaurMega Venusaur  Grass  Poison  180  80     100      123
4  4           Charmander   Fire     NaN     91  39      52       43

   Sp. Atk  Sp. Def  Speed  Generation  Legendary
0       65       65     45           1      False
1       80       80     60           1      False
2      100      100     80           1      False
3      122      120     80           1      False
4       60       50     65           1      False
```

[41]:
```python
df.loc[df['Type 1'].str.contains('Fire|Grass', flags=re.I, regex=True)][0:5]
```

[41]:
```
   #                 Name Type 1  Type 2  Total  HP  Attack  Defense  \
0  1            Bulbasaur  Grass  Poison     94  45      49       49
1  2              Ivysaur  Grass  Poison    122  60      62       63
2  3              Venusaur  Grass  Poison   162  80      82       83
3  3  VenusaurMega Venusaur  Grass  Poison  180  80     100      123
4  4           Charmander   Fire     NaN     91  39      52       43

   Sp. Atk  Sp. Def  Speed  Generation  Legendary
0       65       65     45           1      False
1       80       80     60           1      False
2      100      100     80           1      False
3      122      120     80           1      False
4       60       50     65           1      False
```

[42]:
```python
df.loc[df['Name'].str.contains('^pi[a-z]*', flags=re.I, regex=True)][0:5]
```

```
[42]:        #                 Name    Type 1  Type 2  Total  HP  Attack  Defense  \
      20    16               Pidgey    Normal  Flying     85  40      45       40
      21    17            Pidgeotto    Normal  Flying    123  63      60       55
      22    18              Pidgeot    Normal  Flying    163  83      80       75
      23    18  PidgeotMega Pidgeot    Normal  Flying    163  83      80       80
      30    25              Pikachu  Electric     NaN     90  35      55       40

          Sp. Atk  Sp. Def  Speed  Generation  Legendary
      20       35       35     56           1      False
      21       50       50     71           1      False
      22       70       70    101           1      False
      23      135       80    121           1      False
      30       50       50     90           1      False
```

Conditional Changes

```
[43]: df.loc[df['Type 1'] == 'Fire', 'Type 1'] = 'Flamer'
```

```
[44]: df.head()
```

```
[44]:     #                  Name   Type 1  Type 2  Total  HP  Attack  Defense  \
      0   1             Bulbasaur    Grass  Poison     94  45      49       49
      1   2               Ivysaur    Grass  Poison    122  60      62       63
      2   3              Venusaur    Grass  Poison    162  80      82       83
      3   3  VenusaurMega Venusaur    Grass  Poison    180  80     100      123
      4   4            Charmander   Flamer     NaN     91  39      52       43

         Sp. Atk  Sp. Def  Speed  Generation  Legendary
      0       65       65     45           1      False
      1       80       80     60           1      False
      2      100      100     80           1      False
      3      122      120     80           1      False
      4       60       50     65           1      False
```

```
[45]: df.loc[df['Type 1'] == 'Grass', 'Legendary'] = True
```

```
[46]: df.head()
```

```
[46]:     #                  Name   Type 1  Type 2  Total  HP  Attack  Defense  \
      0   1             Bulbasaur    Grass  Poison     94  45      49       49
      1   2               Ivysaur    Grass  Poison    122  60      62       63
      2   3              Venusaur    Grass  Poison    162  80      82       83
      3   3  VenusaurMega Venusaur    Grass  Poison    180  80     100      123
      4   4            Charmander   Flamer     NaN     91  39      52       43

         Sp. Atk  Sp. Def  Speed  Generation  Legendary
      0       65       65     45           1       True
```

```
   1          80          80      60            1       True
   2         100         100      80            1       True
   3         122         120      80            1       True
   4          60          50      65            1      False
```

[47]: `df.loc[df['Speed'] > 10, ['Generation', 'Legendary']] = ['Test 1', 'Test 2']`

[48]: `df.head()`

[48]:
```
   #               Name   Type 1  Type 2  Total  HP  Attack  Defense  \
0  1          Bulbasaur    Grass  Poison     94  45      49       49
1  2            Ivysaur    Grass  Poison    122  60      62       63
2  3           Venusaur    Grass  Poison    162  80      82       83
3  3  VenusaurMega Venusaur  Grass  Poison    180  80     100      123
4  4         Charmander   Flamer     NaN     91  39      52       43

   Sp. Atk  Sp. Def  Speed  Generation  Legendary
0       65       65     45      Test 1     Test 2
1       80       80     60      Test 1     Test 2
2      100      100     80      Test 1     Test 2
3      122      120     80      Test 1     Test 2
4       60       50     65      Test 1     Test 2
```

[49]:
```
df = pd.read_csv('pokemon_data.csv')
df.head()
```

[49]:
```
   #               Name  Type 1  Type 2  HP  Attack  Defense  Sp. Atk  \
0  1          Bulbasaur   Grass  Poison  45      49       49       65
1  2            Ivysaur   Grass  Poison  60      62       63       80
2  3           Venusaur   Grass  Poison  80      82       83      100
3  3  VenusaurMega Venusaur  Grass  Poison  80     100      123      122
4  4         Charmander    Fire     NaN  39      52       43       60

   Sp. Def  Speed  Generation  Legendary
0       65     45           1      False
1       80     60           1      False
2      100     80           1      False
3      120     80           1      False
4       50     65           1      False
```

Aggregate Statistics (GroupBy)

[50]: `df.groupby(['Type 1']).mean().sort_values('Defense', ascending=False)`

[50]:
```
                #          HP      Attack     Defense     Sp. Atk     Sp. Def  \
Type 1
Steel   442.851852  65.222222  92.703704  126.370370  67.518519  80.629630
```
```
12
```

Type 1	#	HP	Attack	Defense	Sp. Atk	Sp. Def
Rock	392.727273	65.363636	92.863636	100.795455	63.340909	75.477273
Dragon	474.375000	83.312500	112.125000	86.375000	96.843750	88.843750
Ground	356.281250	73.781250	95.750000	84.843750	56.468750	62.750000
Ghost	486.500000	64.437500	73.781250	81.187500	79.343750	76.468750
Water	303.089286	72.062500	74.151786	72.946429	74.812500	70.517857
Ice	423.541667	72.000000	72.750000	71.416667	77.541667	76.291667
Grass	344.871429	67.271429	73.214286	70.800000	77.500000	70.428571
Bug	334.492754	56.884058	70.971014	70.724638	53.869565	64.797101
Dark	461.354839	66.806452	88.387097	70.225806	74.645161	69.516129
Poison	251.785714	67.250000	74.678571	68.821429	60.428571	64.392857
Fire	327.403846	69.903846	84.769231	67.769231	88.980769	72.211538
Psychic	380.807018	70.631579	71.456140	67.684211	98.403509	86.280702
Electric	363.500000	59.795455	69.090909	66.295455	90.022727	73.704545
Flying	677.750000	70.750000	78.750000	66.250000	94.250000	72.500000
Fighting	363.851852	69.851852	96.777778	65.925926	53.111111	64.703704
Fairy	449.529412	74.117647	61.529412	65.705882	78.529412	84.705882
Normal	319.173469	77.275510	73.469388	59.846939	55.816327	63.724490

Type 1	Speed	Generation	Legendary
Steel	55.259259	3.851852	0.148148
Rock	55.909091	3.454545	0.090909
Dragon	83.031250	3.875000	0.375000
Ground	63.906250	3.156250	0.125000
Ghost	64.343750	4.187500	0.062500
Water	65.964286	2.857143	0.035714
Ice	63.458333	3.541667	0.083333
Grass	61.928571	3.357143	0.042857
Bug	61.681159	3.217391	0.000000
Dark	76.161290	4.032258	0.064516
Poison	63.571429	2.535714	0.000000
Fire	74.442308	3.211538	0.096154
Psychic	81.491228	3.385965	0.245614
Electric	84.500000	3.272727	0.090909
Flying	102.500000	5.500000	0.500000
Fighting	66.074074	3.370370	0.000000
Fairy	48.588235	4.117647	0.058824
Normal	71.551020	3.051020	0.020408

```python
[51]: df.groupby(['Type 1']).mean().sort_values('Speed', ascending=False)
```

[51]:

Type 1	#	HP	Attack	Defense	Sp. Atk	Sp. Def	\
Flying	677.750000	70.750000	78.750000	66.250000	94.250000	72.500000	
Electric	363.500000	59.795455	69.090909	66.295455	90.022727	73.704545	
Dragon	474.375000	83.312500	112.125000	86.375000	96.843750	88.843750	
Psychic	380.807018	70.631579	71.456140	67.684211	98.403509	86.280702	

```
Dark      461.354839  66.806452  88.387097   70.225806   74.645161  69.516129
Fire      327.403846  69.903846  84.769231   67.769231   88.980769  72.211538
Normal    319.173469  77.275510  73.469388   59.846939   55.816327  63.724490
Fighting  363.851852  69.851852  96.777778   65.925926   53.111111  64.703704
Water     303.089286  72.062500  74.151786   72.946429   74.812500  70.517857
Ghost     486.500000  64.437500  73.781250   81.187500   79.343750  76.468750
Ground    356.281250  73.781250  95.750000   84.843750   56.468750  62.750000
Poison    251.785714  67.250000  74.678571   68.821429   60.428571  64.392857
Ice       423.541667  72.000000  72.750000   71.416667   77.541667  76.291667
Grass     344.871429  67.271429  73.214286   70.800000   77.500000  70.428571
Bug       334.492754  56.884058  70.971014   70.724638   53.869565  64.797101
Rock      392.727273  65.363636  92.863636  100.795455   63.340909  75.477273
Steel     442.851852  65.222222  92.703704  126.370370   67.518519  80.629630
Fairy     449.529412  74.117647  61.529412   65.705882   78.529412  84.705882

              Speed  Generation  Legendary
Type 1
Flying   102.500000    5.500000   0.500000
Electric  84.500000    3.272727   0.090909
Dragon    83.031250    3.875000   0.375000
Psychic   81.491228    3.385965   0.245614
Dark      76.161290    4.032258   0.064516
Fire      74.442308    3.211538   0.096154
Normal    71.551020    3.051020   0.020408
Fighting  66.074074    3.370370   0.000000
Water     65.964286    2.857143   0.035714
Ghost     64.343750    4.187500   0.062500
Ground    63.906250    3.156250   0.125000
Poison    63.571429    2.535714   0.000000
Ice       63.458333    3.541667   0.083333
Grass     61.928571    3.357143   0.042857
Bug       61.681159    3.217391   0.000000
Rock      55.909091    3.454545   0.090909
Steel     55.259259    3.851852   0.148148
Fairy     48.588235    4.117647   0.058824
```

```python
df.groupby(['Type 1']).sum()
df.groupby(['Type 1']).count()
```

```
           #  Name  Type 2  HP  Attack  Defense  Sp. Atk  Sp. Def  Speed  \
Type 1
Bug       69    69      52  69      69       69       69       69     69
Dark      31    31      21  31      31       31       31       31     31
Dragon    32    32      21  32      32       32       32       32     32
Electric  44    44      17  44      44       44       44       44     44
Fairy     17    17       2  17      17       17       17       17     17
Fighting  27    27       7  27      27       27       27       27     27
```

Fire	52	52	24	52	52	52	52	52	52
Flying	4	4	2	4	4	4	4	4	4
Ghost	32	32	22	32	32	32	32	32	32
Grass	70	70	37	70	70	70	70	70	70
Ground	32	32	19	32	32	32	32	32	32
Ice	24	24	11	24	24	24	24	24	24
Normal	98	98	37	98	98	98	98	98	98
Poison	28	28	13	28	28	28	28	28	28
Psychic	57	57	19	57	57	57	57	57	57
Rock	44	44	35	44	44	44	44	44	44
Steel	27	27	22	27	27	27	27	27	27
Water	112	112	53	112	112	112	112	112	112

	Generation	Legendary
Type 1		
Bug	69	69
Dark	31	31
Dragon	32	32
Electric	44	44
Fairy	17	17
Fighting	27	27
Fire	52	52
Flying	4	4
Ghost	32	32
Grass	70	70
Ground	32	32
Ice	24	24
Normal	98	98
Poison	28	28
Psychic	57	57
Rock	44	44
Steel	27	27
Water	112	112

Working with Large Amounts of Data

```
[53]: # for df in pd.read_csv('pokemon_data.csv', chunksize=5):
      #     print("CHUNK DF")
      #     print(df)
```

```
[ ]:
```