Pandas Tutorial

May 8, 2021

1 Pandas Tutorial

1.0.1 Uploading DataSet

```
[1]: from google.colab import files
     uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving pokemon_data.csv to pokemon_data.csv
```

1.0.2 Importing Library

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

1.0.3 Loading data into Pandas

```
[3]: df = pd.read_csv('pokemon_data.csv')
     print(df.head(10))
```

```
     #                        Name  Type 1  ...  Speed  Generation  Legendary
0    1                   Bulbasaur   Grass  ...     45           1      False
1    2                     Ivysaur   Grass  ...     60           1      False
2    3                    Venusaur   Grass  ...     80           1      False
3    3        VenusaurMega Venusaur   Grass  ...     80           1      False
4    4                  Charmander    Fire  ...     65           1      False
5    5                  Charmeleon    Fire  ...     80           1      False
6    6                   Charizard    Fire  ...    100           1      False
7    6   CharizardMega Charizard X    Fire  ...    100           1      False
8    6   CharizardMega Charizard Y    Fire  ...    100           1      False
9    7                    Squirtle   Water  ...     43           1      False

[10 rows x 12 columns]
```

1.0.4 Reading data in Pandas

```
[4]: ## Read Headers
     # df.columns

     ## Read each Column
     # df[['Name', 'Type 1', 'HP']][0:10]

     ## Read each Row
     # print(df.iloc[0:4])

     # for index, row in df.iterrows():
     #     print(index, row['Name'])

     df.loc[df['Type 1'] == "Fire"]

     ## Read a specific Location(R,C)
     # print(df.iloc[2, 1])
```

```
[4]:          #                   Name    Type 1   ...   Speed   Generation   Legendary
     4         4              Charmander    Fire    ...     65            1       False
     5         5              Charmeleon    Fire    ...     80            1       False
     6         6               Charizard    Fire    ...    100            1       False
     7         6    CharizardMega Charizard X   Fire   ...    100            1       False
     8         6    CharizardMega Charizard Y   Fire   ...    100            1       False
     42       37                  Vulpix    Fire    ...     65            1       False
     43       38                Ninetales    Fire    ...    100            1       False
     63       58                Growlithe    Fire    ...     60            1       False
     64       59                 Arcanine    Fire    ...     95            1       False
     83       77                   Ponyta    Fire    ...     90            1       False
     84       78                 Rapidash    Fire    ...    105            1       False
     135     126                   Magmar    Fire    ...     93            1       False
     147     136                  Flareon    Fire    ...     65            1       False
     158     146                  Moltres    Fire    ...     90            1        True
     169     155                Cyndaquil    Fire    ...     65            2       False
     170     156                  Quilava    Fire    ...     80            2       False
     171     157               Typhlosion    Fire    ...    100            2       False
     236     218                   Slugma    Fire    ...     20            2       False
     237     219                 Magcargo    Fire    ...     30            2       False
     259     240                    Magby    Fire    ...     83            2       False
     263     244                    Entei    Fire    ...    100            2        True
     270     250                    Ho-oh    Fire    ...     90            2        True
     276     255                   Torchic    Fire    ...     45            3       False
     277     256                 Combusken    Fire    ...     55            3       False
     278     257                  Blaziken    Fire    ...     80            3       False
     279     257    BlazikenMega Blaziken    Fire    ...    100            3       False
     352     322                    Numel    Fire    ...     35            3       False
     353     323                  Camerupt    Fire    ...     40            3       False
```

```
354  323        CameruptMega Camerupt    Fire  ...    20      3    False
355  324                      Torkoal    Fire  ...    20      3    False
435  390                      Chimchar   Fire  ...    61      4    False
436  391                      Monferno   Fire  ...    81      4    False
437  392                      Infernape  Fire  ...   108      4    False
518  467                      Magmortar  Fire  ...    83      4    False
542  485                      Heatran    Fire  ...    77      4     True
557  498                        Tepig    Fire  ...    45      5    False
558  499                      Pignite    Fire  ...    55      5    False
559  500                       Emboar    Fire  ...    65      5    False
572  513                      Pansear    Fire  ...    64      5    False
573  514                     Simisear   Fire  ...   101      5    False
614  554                     Darumaka    Fire  ...    50      5    False
615  555  DarmanitanStandard Mode        Fire  ...    95      5    False
616  555      DarmanitanZen Mode         Fire  ...    55      5    False
692  631                      Heatmor    Fire  ...    65      5    False
721  653                     Fennekin   Fire  ...    60      6    False
722  654                      Braixen    Fire  ...    73      6    False
723  655                      Delphox    Fire  ...   104      6    False
730  662                   Fletchinder   Fire  ...    84      6    False
731  663                    Talonflame   Fire  ...   126      6    False
735  667                       Litleo    Fire  ...    72      6    False
736  668                       Pyroar    Fire  ...   106      6    False
799  721                     Volcanion   Fire  ...    70      6     True

[52 rows x 12 columns]
```

1.0.5 Sorting/Describing Data

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

```
[6]:                 #          HP       Attack  ...       Sp. Def       Speed
     Generation
     count  800.000000  800.000000  800.000000  ...  800.000000  800.000000
     800.00000
     mean   362.813750   69.258750   79.001250  ...   71.902500   68.277500
     3.32375
     std    208.343798   25.534669   32.457366  ...   27.828916   29.060474
     1.66129
     min      1.000000    1.000000    5.000000  ...   20.000000    5.000000
     1.00000
     25%    184.750000   50.000000   55.000000  ...   50.000000   45.000000
     2.00000
     50%    364.500000   65.000000   75.000000  ...   70.000000   65.000000
     3.00000
     75%    539.250000   80.000000  100.000000  ...   90.000000   90.000000
     5.00000
```

```
max    721.000000  255.000000  190.000000  ...  230.000000  180.000000
6.00000
```

```
[8 rows x 8 columns]
```

[7]: `df.sort_values(['Type 1', 'HP'], ascending=[1, 0])`

[7]:
```
         #                      Name  Type 1  ...  Speed  Generation  Legendary
520    469               Yanmega       Bug  ...     95           4      False
698    637               Volcarona      Bug  ...    100           5      False
231    214               Heracross      Bug  ...     85           2      False
232    214  HeracrossMega Heracross    Bug  ...     75           2      False
678    617               Accelgor      Bug  ...    145           5      False
..     ...                    ...      ...   ...    ...         ...        ...
106     98                Krabby     Water  ...     50           1      False
125    116                Horsea     Water  ...     60           1      False
129    120                Staryu     Water  ...     85           1      False
139    129                Magikarp   Water  ...     80           1      False
381    349                Feebas     Water  ...     80           3      False
```

```
[800 rows x 12 columns]
```

1.0.6 Making changes to the data

[8]:
```python
# df['Total'] = df['HP'] + df['Attack'] + df['Defense'] + df['Sp. Atk'] +
  →df['Sp. Def'] + df['Speed']
# df = df.drop(columns=['Total'])

df['Total'] = df.iloc[:, 4:10].sum(axis=1)
cols = list(df.columns)
df = df[cols[0:4] + [cols[-1]] + cols[4:12]]

df.head(5)
```

[8]:
```
   #          Name  Type 1  ...  Speed  Generation  Legendary
0  1       Bulbasaur  Grass  ...     45           1      False
1  2        Ivysaur  Grass  ...     60           1      False
2  3       Venusaur  Grass  ...     80           1      False
3  3  VenusaurMega Venusaur  Grass  ...     80           1      False
4  4      Charmander   Fire  ...     65           1      False
```

```
[5 rows x 13 columns]
```

1.0.7 Saving our Data

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

1.0.8 Filtering Data

```
[10]: new_df = df.loc[(df['Type 1'] == 'Grass') & (df['Type 2'] == 'Poison') &
      →(df['HP'] > 70)]
      # new_df.to_csv('filtered.csv')

      new_df.reset_index(drop = True, inplace = True)
      new_df
```

```
[10]:        #                    Name Type 1  ... Speed  Generation  Legendary
      0       3                Venusaur  Grass  ...    80           1      False
      1       3  VenusaurMega Venusaur  Grass  ...    80           1      False
      2      45               Vileplume  Grass  ...    50           1      False
      3      71              Victreebel  Grass  ...    70           1      False
      4     591               Amoonguss  Grass  ...    30           5      False

      [5 rows x 13 columns]
```

```
[11]: df.loc[~df['Name'].str.contains('Mega')]
```

```
[11]:        #                   Name   Type 1  ... Speed  Generation  Legendary
      0        1             Bulbasaur    Grass  ...    45           1      False
      1        2               Ivysaur    Grass  ...    60           1      False
      2        3              Venusaur    Grass  ...    80           1      False
      4        4            Charmander     Fire  ...    65           1      False
      5        5            Charmeleon     Fire  ...    80           1      False
      ..     ...                   ...      ...  ...   ...         ...        ...
      794    718        Zygarde50% Forme   Dragon  ...    95           6       True
      795    719               Diancie     Rock  ...    50           6       True
      797    720   HoopaHoopa Confined  Psychic  ...    70           6       True
      798    720    HoopaHoopa Unbound  Psychic  ...    80           6       True
      799    721              Volcanion     Fire  ...    70           6       True

      [751 rows x 13 columns]
```

```
[12]: import re
      df.loc[df['Type 1'].str.contains('Fire|grass', flags=re.IGNORECASE, regex=True)]
```

```
[12]:        #                    Name Type 1  ... Speed  Generation  Legendary
      0        1             Bulbasaur  Grass  ...    45           1      False
      1        2               Ivysaur  Grass  ...    60           1      False
      2        3              Venusaur  Grass  ...    80           1      False
      3        3  VenusaurMega Venusaur  Grass  ...    80           1      False
      4        4            Charmander   Fire  ...    65           1      False
      ..     ...                   ...    ...  ...   ...         ...        ...
      735    667               Litleo   Fire  ...    72           6      False
      736    668               Pyroar   Fire  ...   106           6      False
      740    672               Skiddo  Grass  ...    52           6      False
      741    673               Gogoat  Grass  ...    68           6      False
      799    721             Volcanion   Fire  ...    70           6       True
```

```
[122 rows x 13 columns]
```

```
[13]: df.loc[df['Name'].str.contains('^pi[a-z]*', flags=re.IGNORECASE, regex=True)]
```

```
[13]:          #                       Name    Type 1   ...  Speed   Generation   Legendary
      20      16               Pidgey     Normal   ...     56            1       False
      21      17            Pidgeotto     Normal   ...     71            1       False
      22      18              Pidgeot     Normal   ...    101            1       False
      23      18    PidgeotMega Pidgeot     Normal   ...    121            1       False
      30      25              Pikachu   Electric   ...     90            1       False
      136    127               Pinsir        Bug   ...     85            1       False
      137    127      PinsirMega Pinsir        Bug   ...    105            1       False
      186    172                Pichu   Electric   ...     60            2       False
      219    204               Pineco        Bug   ...     15            2       False
      239    221             Piloswine        Ice   ...     50            2       False
      438    393               Piplup      Water   ...     40            4       False
      558    499              Pignite       Fire   ...     55            5       False
      578    519               Pidove     Normal   ...     43            5       False

      [13 rows x 13 columns]
```

1.0.9 Conditional Changes

```
[14]: df.loc[df['Type 1'] == 'Fire', 'Type 1'] = 'Flammer'
      df
```

```
[14]:         #                         Name    Type 1   ...  Speed   Generation   Legendary
      0       1                 Bulbasaur      Grass   ...     45            1       False
      1       2                   Ivysaur      Grass   ...     60            1       False
      2       3                  Venusaur      Grass   ...     80            1       False
      3       3       VenusaurMega Venusaur      Grass   ...     80            1       False
      4       4                Charmander    Flammer   ...     65            1       False
      ..     ...                       ...        ...   ...    ...          ...         ...
      795    719                   Diancie       Rock   ...     50            6        True
      796    719          DiancieMega Diancie       Rock   ...    110            6        True
      797    720        HoopaHoopa Confined    Psychic   ...     70            6        True
      798    720         HoopaHoopa Unbound    Psychic   ...     80            6        True
      799    721                 Volcanion    Flammer   ...     70            6        True

      [800 rows x 13 columns]
```

```
[15]: df.loc[df['Type 1'] == 'Flammer', 'Type 1'] = 'Fire'
      df
```

```
[15]:         #                       Name    Type 1   ...  Speed   Generation   Legendary
      0       1               Bulbasaur      Grass   ...     45            1       False
      1       2                 Ivysaur      Grass   ...     60            1       False
      2       3                Venusaur      Grass   ...     80            1       False
```

```
3       3  VenusaurMega Venusaur      Grass  ...      80            1      False
4       4              Charmander      Fire  ...      65            1      False
..    ...                     ...       ...  ...     ...          ...        ...
795   719               Diancie      Rock  ...      50            6       True
796   719     DiancieMega Diancie      Rock  ...     110            6       True
797   720    HoopaHoopa Confined   Psychic  ...      70            6       True
798   720    HoopaHoopa Unbound    Psychic  ...      80            6       True
799   721              Volcanion      Fire  ...      70            6       True

[800 rows x 13 columns]
```

1.0.10 Aggregate Statistics(Groupby)

```
[16]: df = pd.read_csv('modified.csv')
      df.groupby(['Type 1']).mean().sort_values('Attack', ascending=False)
```

```
[16]:                     #        Total         HP  ...       Speed   Generation
      Legendary
      Type 1                                          ...
      Dragon       474.375000   550.531250  83.312500  ...   83.031250     3.875000
      0.375000
      Fighting     363.851852   416.444444  69.851852  ...   66.074074     3.370370
      0.000000
      Ground       356.281250   437.500000  73.781250  ...   63.906250     3.156250
      0.125000
      Rock         392.727273   453.750000  65.363636  ...   55.909091     3.454545
      0.090909
      Steel        442.851852   487.703704  65.222222  ...   55.259259     3.851852
      0.148148
      Dark         461.354839   445.741935  66.806452  ...   76.161290     4.032258
      0.064516
      Fire         327.403846   458.076923  69.903846  ...   74.442308     3.211538
      0.096154
      Flying       677.750000   485.000000  70.750000  ...  102.500000     5.500000
      0.500000
      Poison       251.785714   399.142857  67.250000  ...   63.571429     2.535714
      0.000000
      Water        303.089286   430.455357  72.062500  ...   65.964286     2.857143
      0.035714
      Ghost        486.500000   439.562500  64.437500  ...   64.343750     4.187500
      0.062500
      Normal       319.173469   401.683673  77.275510  ...   71.551020     3.051020
      0.020408
      Grass        344.871429   421.142857  67.271429  ...   61.928571     3.357143
      0.042857
      Ice          423.541667   433.458333  72.000000  ...   63.458333     3.541667
      0.083333
```

```
Psychic    380.807018   475.947368   70.631579   ...    81.491228     3.385965
           0.245614
Bug        334.492754   378.927536   56.884058   ...    61.681159     3.217391
           0.000000
Electric   363.500000   443.409091   59.795455   ...    84.500000     3.272727
           0.090909
Fairy      449.529412   413.176471   74.117647   ...    48.588235     4.117647
           0.058824

[18 rows x 10 columns]
```

```
[17]: df['Count'] = 1
      count_df = df.groupby(['Type 1', 'Type 2']).count()['Count']
      count_df.to_csv('Count.csv')
      count_df
```

```
[17]: Type 1  Type 2
      Bug     Electric     2
              Fighting     2
              Fire         2
              Flying      14
              Ghost        1
                          ..
      Water   Ice          3
              Poison       3
              Psychic      5
              Rock         4
              Steel        1
      Name: Count, Length: 136, dtype: int64
```

1.0.11 Working with large amounts of data

```
[18]: for df in pd.read_csv('modified.csv', chunksize=5):
          print("**********")
          print(df)
```

```
**********
   #                 Name Type 1  ... Speed  Generation  Legendary
0  1            Bulbasaur  Grass  ...    45           1      False
1  2              Ivysaur  Grass  ...    60           1      False
2  3             Venusaur  Grass  ...    80           1      False
3  3  VenusaurMega Venusaur  Grass  ...    80           1      False
4  4           Charmander   Fire  ...    65           1      False

[5 rows x 13 columns]
**********
   #                 Name Type 1  ... Speed  Generation  Legendary
5  5           Charmeleon   Fire  ...    80           1      False
```

8

```
6   6                     Charizard   Fire  ...    100             1      False
7   6   CharizardMega Charizard X   Fire  ...    100             1      False
8   6   CharizardMega Charizard Y   Fire  ...    100             1      False
9   7                     Squirtle  Water  ...     43             1      False

[5 rows x 13 columns]
**********
     #                     Name Type 1  ...  Speed  Generation  Legendary
10   8                Wartortle  Water  ...     58           1      False
11   9                Blastoise  Water  ...     78           1      False
12   9   BlastoiseMega Blastoise  Water  ...     78           1      False
13  10                 Caterpie    Bug  ...     45           1      False
14  11                  Metapod    Bug  ...     30           1      False

[5 rows x 13 columns]
**********
     #                     Name Type 1  ... Speed  Generation  Legendary
15  12                Butterfree    Bug  ...    70           1      False
16  13                   Weedle    Bug  ...    50           1      False
17  14                   Kakuna    Bug  ...    35           1      False
18  15                  Beedrill    Bug  ...    75           1      False
19  15   BeedrillMega Beedrill    Bug  ...   145           1      False

[5 rows x 13 columns]
**********
     #                     Name  Type 1  ... Speed  Generation  Legendary
20  16                   Pidgey  Normal  ...    56           1      False
21  17                 Pidgeotto  Normal  ...    71           1      False
22  18                  Pidgeot  Normal  ...   101           1      False
23  18   PidgeotMega Pidgeot  Normal  ...   121           1      False
24  19                  Rattata  Normal  ...    72           1      False

[5 rows x 13 columns]
**********
     #        Name  Type 1  Type 2  ...  Sp. Def  Speed  Generation  Legendary
25  20   Raticate  Normal     NaN  ...       70     97           1      False
26  21    Spearow  Normal  Flying  ...       31     70           1      False
27  22     Fearow  Normal  Flying  ...       61    100           1      False
28  23      Ekans  Poison     NaN  ...       54     55           1      False
29  24      Arbok  Poison     NaN  ...       79     80           1      False

[5 rows x 13 columns]
**********
     #            Name    Type 1  ...  Speed  Generation  Legendary
30  25          Pikachu  Electric  ...     90           1      False
31  26           Raichu  Electric  ...    110           1      False
32  27        Sandshrew    Ground  ...     40           1      False
33  28        Sandslash    Ground  ...     65           1      False
```

```python
[19]: new_df1 = pd.DataFrame(columns = df.columns)

      for df in pd.read_csv('modified.csv', chunksize=5):
          results = df.groupby(['Type 1']).count()

          new_df1 = pd.concat([new_df1, results])
```

```python
[20]: new_df1
```

```
[20]:           # Name Type 1 Type 2 Total  ... Sp. Atk Sp. Def Speed Generation
      Legendary
      Fire     1    1    NaN     0     1 ...       1       1     1          1
      1
      Grass    4    4    NaN     4     4 ...       4       4     4          4
      4
      Fire     4    4    NaN     3     4 ...       4       4     4          4
      4
      Water    1    1    NaN     0     1 ...       1       1     1          1
      1
      Bug      2    2    NaN     0     2 ...       2       2     2          2
      2
      ...      ..  ...   ...    ...   ... ...     ...     ...   ...        ...
      ...
      Fairy    1    1    NaN     0     1 ...       1       1     1          1
      1
      Flying   2    2    NaN     2     2 ...       2       2     2          2
      2
      Fire     1    1    NaN     1     1 ...       1       1     1          1
      1
      Psychic  2    2    NaN     2     2 ...       2       2     2          2
      2
      Rock     2    2    NaN     2     2 ...       2       2     2          2
      2

      [433 rows x 13 columns]
```

```python
[ ]: !wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py
     from colab_pdf import colab_pdf
     colab_pdf('Pandas Tutorial.ipynb')
```

```
--2021-05-08 10:29:12--  https://raw.githubusercontent.com/brpy/colab-
pdf/master/colab_pdf.py
Resolving raw.githubusercontent.com (raw.githubusercontent.com)...
185.199.108.133, 185.199.109.133, 185.199.110.133, ...
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.108.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1864 (1.8K) [text/plain]
Saving to: colab_pdf.py
```

```
colab_pdf.py          100%[===================>]   1.82K  --.-KB/s    in 0s
```

```
2021-05-08 10:29:12 (41.8 MB/s) - colab_pdf.py saved [1864/1864]
```

```
WARNING: apt does not have a stable CLI interface. Use with caution in scripts.
```

```
WARNING: apt does not have a stable CLI interface. Use with caution in scripts.
```

```
Extracting templates from packages: 100%
```