TheDataLytics

https://www.youtube.com/channel/UC-kt49G2u5YeR42wLHorKKw

ANUURAG EDLABADKAR

SUBSCRIBE

LIKE SUBSCRIBE PRESS

python™

ANUURAG EDLABADKAR

pandas

```
import pandas as pd
import numpy as np
```

```
url =
'https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipot
le.tsv'
df_chipotle = pd.read_csv(url, sep = '\t')
df_chipotle
```

```
      order_id  ...  item_price
0            1  ...      $2.39
1            1  ...      $3.39
2            1  ...      $3.39
3            1  ...      $2.39
4            2  ...     $16.98
...        ...  ...        ...
4617      1833  ...     $11.75
4618      1833  ...     $11.75
4619      1834  ...     $11.25
4620      1834  ...      $8.75
4621      1834  ...      $8.75

[4622 rows x 5 columns]
```

```
# clean the item_price column and transform it in a float
prices = [float(value[1 : -1]) for value in df_chipotle.item_price]

# reassign the column with the cleaned prices
df_chipotle.item_price = prices

# make the comparison
chipo10 = df_chipotle[df_chipotle['item_price'] > 10.00]
chipo10.head()

len(chipo10)
```

```
1130
```

```
# delete the duplicates in item_name and quantity
chipo_filtered = df_chipotle.drop_duplicates(['item_name','quantity'])

# select only the products with quantity equals to 1
chipo_one_prod = chipo_filtered[chipo_filtered.quantity == 1]

# select only the item_name and item_price columns
price_per_item = chipo_one_prod[['item_name', 'item_price']]

# sort the values from the most to less expensive
price_per_item.sort_values(by = "item_price", ascending = False)
```

```
                                  item_name  item_price
606                        Steak Salad Bowl       11.89
1229                    Barbacoa Salad Bowl       11.89
1132                    Carnitas Salad Bowl       11.89
```

```
7                                  Steak Burrito       11.75
168                        Barbacoa Crispy Tacos       11.75
39                                 Barbacoa Bowl       11.75
738                            Veggie Soft Tacos       11.25
186                            Veggie Salad Bowl       11.25
62                                   Veggie Bowl       11.25
57                                Veggie Burrito       11.25
250                                Chicken Salad       10.98
5                                   Chicken Bowl       10.98
8                               Steak Soft Tacos        9.25
554                        Carnitas Crispy Tacos        9.25
237                          Carnitas Soft Tacos        9.25
56                           Barbacoa Soft Tacos        9.25
92                             Steak Crispy Tacos       9.25
664                                  Steak Salad        8.99
54                                    Steak Bowl        8.99
3750                              Carnitas Salad        8.99
21                              Barbacoa Burrito        8.99
27                              Carnitas Burrito        8.99
33                                 Carnitas Bowl        8.99
11                          Chicken Crispy Tacos        8.75
12                            Chicken Soft Tacos        8.75
44                            Chicken Salad Bowl        8.75
1653                         Veggie Crispy Tacos        8.49
16                               Chicken Burrito        8.49
1694                                Veggie Salad        8.49
1414                                       Salad        7.40
510                                      Burrito        7.40
520                                 Crispy Tacos        7.40
673                                         Bowl        7.40
298                            6 Pack Soft Drink        6.49
10                           Chips and Guacamole        4.45
1                                           Izze        3.39
2                               Nantucket Nectar        3.39
674        Chips and Mild Fresh Tomato Salsa            3.00
111      Chips and Tomatillo Red Chili Salsa            2.95
233       Chips and Roasted Chili Corn Salsa            2.95
38     Chips and Tomatillo Green Chili Salsa            2.95
3      Chips and Tomatillo-Green Chili Salsa            2.39
300      Chips and Tomatillo-Red Chili Salsa            2.39
191       Chips and Roasted Chili-Corn Salsa            2.39
0               Chips and Fresh Tomato Salsa            2.39
40                                         Chips        2.15
6                                  Side of Chips        1.69
263                            Canned Soft Drink        1.25
28                                   Canned Soda        1.09
34                                 Bottled Water        1.09
```

```
df_chipotle.item_name.sort_values()
```

```
3389    6 Pack Soft Drink
341     6 Pack Soft Drink
1849    6 Pack Soft Drink
1860    6 Pack Soft Drink
2713    6 Pack Soft Drink
              ...
2384    Veggie Soft Tacos
781     Veggie Soft Tacos
2851    Veggie Soft Tacos
1699    Veggie Soft Tacos
1395    Veggie Soft Tacos
Name: item_name, Length: 4622, dtype: object
```

```
df_chipotle.sort_values(by = "item_name")
```

```
      order_id  ...  item_price
3389      1360  ...       12.98
341        148  ...        6.49
1849       749  ...        6.49
1860       754  ...        6.49
2713      1076  ...        6.49
...        ...  ...         ...
2384       948  ...        8.75
781        322  ...        8.75
2851      1132  ...        8.49
1699       688  ...       11.25
1395       567  ...        8.49

[4622 rows x 5 columns]
```

```
df_chipotle.sort_values(by = "item_price", ascending = False).head(1)
```

```
      order_id  quantity  ... choice_description item_price
3598      1443        15  ...                NaN      44.25

[1 rows x 5 columns]
```

```
chipo_salad = df_chipotle[df_chipotle.item_name == "Veggie Salad
Bowl"]
chipo_salad
```

```
      order_id  ...  item_price
186         83  ...       11.25
295        128  ...       11.25
455        195  ...       11.25
496        207  ...       11.25
960        394  ...        8.75
1316       536  ...        8.75
1884       760  ...       11.25
2156       869  ...       11.25
2223       896  ...        8.75
```

```
2269       913  ...        8.75
2683      1066  ...        8.75
3223      1289  ...       11.25
3293      1321  ...        8.75
4109      1646  ...       11.25
4201      1677  ...       11.25
4261      1700  ...       11.25
4541      1805  ...        8.75
4573      1818  ...        8.75

[18 rows x 5 columns]
```

```
chipo_drink_steak_bowl = df_chipotle[(df_chipotle.item_name == "Canned
Soda") & (df_chipotle.quantity > 1)]
len(chipo_drink_steak_bowl)
```

```
20
```

```
# Create an example dataframe about a fictional army
raw_data = {'regiment': ['Nighthawks', 'Nighthawks', 'Nighthawks',
'Nighthawks', 'Dragoons', 'Dragoons', 'Dragoons', 'Dragoons',
'Scouts', 'Scouts', 'Scouts', 'Scouts'],
            'company': ['1st', '1st', '2nd', '2nd', '1st', '1st',
'2nd', '2nd','1st', '1st', '2nd', '2nd'],
            'deaths': [523, 52, 25, 616, 43, 234, 523, 62, 62, 73, 37,
35],
            'battles': [5, 42, 2, 2, 4, 7, 8, 3, 4, 7, 8, 9],
            'size': [1045, 957, 1099, 1400, 1592, 1006, 987, 849, 973,
1005, 1099, 1523],
            'veterans': [1, 5, 62, 26, 73, 37, 949, 48, 48, 435, 63,
345],
            'readiness': [1, 2, 3, 3, 2, 1, 2, 3, 2, 1, 2, 3],
            'armored': [1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1],
            'deserters': [4, 24, 31, 2, 3, 4, 24, 31, 2, 3, 2, 3],
            'origin': ['Arizona', 'California', 'Texas', 'Florida',
'Maine', 'Iowa', 'Alaska', 'Washington', 'Oregon', 'Wyoming',
'Louisana', 'Georgia']}

army = pd.DataFrame(raw_data, columns = ['regiment', 'company',
'deaths', 'battles', 'size', 'veterans', 'readiness', 'armored',
'deserters', 'origin'])

army
```

```
     regiment company  deaths  ...  armored  deserters      origin
0  Nighthawks     1st     523  ...        1          4     Arizona
1  Nighthawks     1st      52  ...        0         24  California
2  Nighthawks     2nd      25  ...        1         31       Texas
3  Nighthawks     2nd     616  ...        1          2     Florida
4    Dragoons     1st      43  ...        0          3       Maine
5    Dragoons     1st     234  ...        1          4        Iowa
```

```
6     Dragoons     2nd     523  ...        0        24      Alaska
7     Dragoons     2nd      62  ...        1        31  Washington
8       Scouts     1st      62  ...        0         2      Oregon
9       Scouts     1st      73  ...        0         3     Wyoming
10      Scouts     2nd      37  ...        1         2    Louisana
11      Scouts     2nd      35  ...        1         3     Georgia

[12 rows x 10 columns]
```

```
army = army.set_index('origin')
army
```

```
              regiment company  deaths  ...  readiness  armored
deserters
origin                                  ...

Arizona     Nighthawks     1st     523  ...          1        1
4
California  Nighthawks     1st      52  ...          2        0
24
Texas       Nighthawks     2nd      25  ...          3        1
31
Florida     Nighthawks     2nd     616  ...          3        1
2
Maine         Dragoons     1st      43  ...          2        0
3
Iowa          Dragoons     1st     234  ...          1        1
4
Alaska        Dragoons     2nd     523  ...          2        0
24
Washington    Dragoons     2nd      62  ...          3        1
31
Oregon          Scouts     1st      62  ...          2        0
2
Wyoming         Scouts     1st      73  ...          1        0
3
Louisana        Scouts     2nd      37  ...          2        1
2
Georgia         Scouts     2nd      35  ...          3        1
3

[12 rows x 9 columns]
```

```
army['veterans']
```

```
origin
Arizona        1
California     5
Texas         62
Florida       26
Maine         73
```

```
Iowa           37
Alaska        949
Washington     48
Oregon         48
Wyoming       435
Louisana       63
Georgia       345
Name: veterans, dtype: int64
```

```
army[['veterans', 'deaths']]
```

```
            veterans  deaths
origin
Arizona            1     523
California         5      52
Texas             62      25
Florida           26     616
Maine             73      43
Iowa              37     234
Alaska           949     523
Washington        48      62
Oregon            48      62
Wyoming          435      73
Louisana          63      37
Georgia          345      35
```

```
army.columns
```

```
Index(['regiment', 'company', 'deaths', 'battles', 'size', 'veterans',
       'readiness', 'armored', 'deserters'],
      dtype='object')
```

```
army.loc[['Maine','Alaska'] , ["deaths","size","deserters"]]
```

```
        deaths  size  deserters
origin
Maine       43  1592          3
Alaska     523   987         24
```

```
army.iloc[3:7, 3:6]
```

```
         battles  size  veterans
origin
Florida        2  1400        26
Maine          4  1592        73
Iowa           7  1006        37
Alaska         8   987       949
```

```
army.iloc[3:]
```

```
              regiment company  deaths  ...  readiness  armored
deserters
```

```
origin                                   ...

Florida     Nighthawks     2nd     616   ...          3        1
2
Maine         Dragoons     1st      43   ...          2        0
3
Iowa          Dragoons     1st     234   ...          1        1
4
Alaska        Dragoons     2nd     523   ...          2        0
24
Washington    Dragoons     2nd      62   ...          3        1
31
Oregon          Scouts     1st      62   ...          2        0
2
Wyoming         Scouts     1st      73   ...          1        0
3
Louisana        Scouts     2nd      37   ...          2        1
2
Georgia         Scouts     2nd      35   ...          3        1
3

[9 rows x 9 columns]
```

```
army.iloc[:3]
```

```
              regiment company  deaths   ...  readiness  armored
deserters
origin                                   ...

Arizona     Nighthawks     1st     523   ...          1        1
4
California  Nighthawks     1st      52   ...          2        0
24
Texas       Nighthawks     2nd      25   ...          3        1
31

[3 rows x 9 columns]
```

```
army.iloc[: , 4:7]
```

```
army[army['deaths'] > 50]
```

```
              regiment company  deaths   ...  readiness  armored
deserters
origin                                   ...

Arizona     Nighthawks     1st     523   ...          1        1
4
California  Nighthawks     1st      52   ...          2        0
24
Florida     Nighthawks     2nd     616   ...          3        1
```

```
2
Iowa          Dragoons     1st      234  ...          1         1
4
Alaska        Dragoons     2nd      523  ...          2         0
24
Washington    Dragoons     2nd       62  ...          3         1
31
Oregon          Scouts     1st       62  ...          2         0
2
Wyoming         Scouts     1st       73  ...          1         0
3

[8 rows x 9 columns]
```

```
army[(army['deaths'] > 500) | (army['deaths'] < 50)]
```

```
            regiment company  deaths  ...  readiness  armored
deserters
origin                                ...

Arizona   Nighthawks     1st     523  ...          1        1
4
Texas     Nighthawks     2nd      25  ...          3        1
31
Florida   Nighthawks     2nd     616  ...          3        1
2
Maine       Dragoons     1st      43  ...          2        0
3
Alaska      Dragoons     2nd     523  ...          2        0
24
Louisana      Scouts     2nd      37  ...          2        1
2
Georgia       Scouts     2nd      35  ...          3        1
3

[7 rows x 9 columns]
```

```
army[(army['regiment'] != 'Dragoons')]
```

```
              regiment company  deaths  ...  readiness  armored
deserters
origin                                  ...

Arizona     Nighthawks     1st     523  ...          1        1
4
California  Nighthawks     1st      52  ...          2        0
24
Texas       Nighthawks     2nd      25  ...          3        1
31
Florida     Nighthawks     2nd     616  ...          3        1
2
```

```
Oregon          Scouts     1st      62  ...         2        0
2
Wyoming         Scouts     1st      73  ...         1        0
3
Louisana        Scouts     2nd      37  ...         2        1
2
Georgia         Scouts     2nd      35  ...         3        1
3

[8 rows x 9 columns]
```

```
army.loc['Arizona':'Florida',:]
```

```
              regiment company  deaths  ...  readiness  armored
deserters
origin                                  ...

Arizona     Nighthawks     1st     523  ...          1        1
4
California  Nighthawks     1st      52  ...          2        0
24
Texas       Nighthawks     2nd      25  ...          3        1
31
Florida     Nighthawks     2nd     616  ...          3        1
2

[4 rows x 9 columns]
```

```
army.loc[['Arizona', 'Texas']]
```

```
           regiment company  deaths  ...  readiness  armored
deserters
origin                               ...

Arizona  Nighthawks     1st     523  ...          1        1
4
Texas    Nighthawks     2nd      25  ...          3        1
31

[2 rows x 9 columns]
```

```
army.loc['Arizona', 'deaths']
```

```
523
```

```
army.loc['Arizona', 'size']
```

```
1045
```

```
euro12 = pd.read_csv('Euro2012TEAM.csv')
euro12
```

```
                   Team  Goals  ...  Subs off  Players Used
0               Croatia      4  ...         9            16
1        Czech Republic      4  ...        11            19
2               Denmark      4  ...         7            15
3               England      5  ...        11            16
4                France      3  ...        11            19
5               Germany     10  ...        15            17
6                Greece      5  ...        12            20
7                 Italy      6  ...        18            19
8           Netherlands      2  ...         7            15
9                Poland      2  ...         7            17
10             Portugal      6  ...        14            16
11  Republic of Ireland      1  ...        10            17
12               Russia      5  ...         7            16
13                Spain     12  ...        17            18
14               Sweden      5  ...         9            18
15              Ukraine      2  ...         9            18

[16 rows x 35 columns]
```

```
euro12.Goals
```

```
0      4
1      4
2      4
3      5
4      3
5     10
6      5
7      6
8      2
9      2
10     6
11     1
12     5
13    12
14     5
15     2
Name: Goals, dtype: int64
```

```
euro12.shape[0], euro12.shape[1]
```

```
(16, 35)
```

```
euro12.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 35 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
```

```
 0   Team                        16 non-null     object 
 1   Goals                       16 non-null     int64  
 2   Shots on target             16 non-null     int64  
 3   Shots off target            16 non-null     int64  
 4   Shooting Accuracy           16 non-null     object 
 5   % Goals-to-shots            16 non-null     object 
 6   Total shots (inc. Blocked)  16 non-null     int64  
 7   Hit Woodwork                16 non-null     int64  
 8   Penalty goals               16 non-null     int64  
 9   Penalties not scored        16 non-null     int64  
 10  Headed goals                16 non-null     int64  
 11  Passes                      16 non-null     int64  
 12  Passes completed            16 non-null     int64  
 13  Passing Accuracy            16 non-null     object 
 14  Touches                     16 non-null     int64  
 15  Crosses                     16 non-null     int64  
 16  Dribbles                    16 non-null     int64  
 17  Corners Taken               16 non-null     int64  
 18  Tackles                     16 non-null     int64  
 19  Clearances                  16 non-null     int64  
 20  Interceptions               16 non-null     int64  
 21  Clearances off line         15 non-null     float64
 22  Clean Sheets                16 non-null     int64  
 23  Blocks                      16 non-null     int64  
 24  Goals conceded              16 non-null     int64  
 25  Saves made                  16 non-null     int64  
 26  Saves-to-shots ratio        16 non-null     object 
 27  Fouls Won                   16 non-null     int64  
 28  Fouls Conceded              16 non-null     int64  
 29  Offsides                    16 non-null     int64  
 30  Yellow Cards                16 non-null     int64  
 31  Red Cards                   16 non-null     int64  
 32  Subs on                     16 non-null     int64  
 33  Subs off                    16 non-null     int64  
 34  Players Used                16 non-null     int64  
dtypes: float64(1), int64(29), object(5)
memory usage: 4.5+ KB
```

```
euro12.describe()
```

```
           Goals  Shots on target  ...  Subs off  Players Used
count  16.000000        16.000000  ...  16.00000     16.000000
mean    4.750000        17.125000  ...  10.87500     17.250000
std     2.886751        10.582218  ...   3.53789      1.527525
min     1.000000         7.000000  ...   7.00000     15.000000
25%     2.750000         9.750000  ...   8.50000     16.000000
50%     4.500000        13.000000  ...  10.50000     17.000000
75%     5.250000        22.000000  ...  12.50000     18.250000
max    12.000000        42.000000  ...  18.00000     20.000000
```

```
[8 rows x 30 columns]

# filter only giving the column names

discipline = euro12[['Team', 'Yellow Cards', 'Red Cards']]
discipline

                   Team  Yellow Cards  Red Cards
0               Croatia             9          0
1        Czech Republic             7          0
2               Denmark             4          0
3               England             5          0
4                France             6          0
5               Germany             4          0
6                Greece             9          1
7                 Italy            16          0
8           Netherlands             5          0
9                Poland             7          1
10             Portugal            12          0
11  Republic of Ireland             6          1
12               Russia             6          0
13                Spain            11          0
14               Sweden             7          0
15              Ukraine             5          0

discipline.sort_values(['Red Cards', 'Yellow Cards'], ascending =
False)

                   Team  Yellow Cards  Red Cards
6                Greece             9          1
9                Poland             7          1
11  Republic of Ireland             6          1
7                 Italy            16          0
10             Portugal            12          0
13                Spain            11          0
0               Croatia             9          0
1        Czech Republic             7          0
14               Sweden             7          0
4                France             6          0
12               Russia             6          0
3               England             5          0
8           Netherlands             5          0
15              Ukraine             5          0
2               Denmark             4          0
5               Germany             4          0

round(discipline['Yellow Cards'].mean())

7
```

```
euro12[euro12.Goals > 6]

       Team  Goals  Shots on target  ...  Subs on Subs off Players
Used
5   Germany     10               32  ...       15       15
17
13    Spain     12               42  ...       17       17
18

[2 rows x 35 columns]

euro12[euro12.Team.str.startswith('G')]

      Team  Goals  Shots on target  ...  Subs on Subs off Players Used
5  Germany     10               32  ...       15       15           17
6   Greece      5                8  ...       12       12           20

[2 rows x 35 columns]

euro12.iloc[: , 0:7]

                   Team  Goals  ...  % Goals-to-shots  Total shots
(inc. Blocked)
0               Croatia      4  ...             16.0%
32
1        Czech Republic      4  ...             12.9%
39
2               Denmark      4  ...             20.0%
27
3               England      5  ...             17.2%
40
4                France      3  ...              6.5%
65
5               Germany     10  ...             15.6%
80
6                Greece      5  ...             19.2%
32
7                 Italy      6  ...              7.5%
110
8           Netherlands      2  ...              4.1%
60
9                Poland      2  ...              5.2%
48
10             Portugal      6  ...              9.3%
82
11  Republic of Ireland      1  ...              5.2%
28
12               Russia      5  ...             12.5%
59
13                Spain     12  ...             16.0%
100
```

```
14                Sweden      5  ...             13.8%
39
15               Ukraine      2  ...              6.0%
38

[16 rows x 7 columns]
```

```
euro12.iloc[: , :-3]
```

```
                   Team  Goals  ...  Yellow Cards  Red Cards
0               Croatia      4  ...             9          0
1        Czech Republic      4  ...             7          0
2               Denmark      4  ...             4          0
3               England      5  ...             5          0
4                France      3  ...             6          0
5               Germany     10  ...             4          0
6                Greece      5  ...             9          1
7                 Italy      6  ...            16          0
8           Netherlands      2  ...             5          0
9                Poland      2  ...             7          1
10             Portugal      6  ...            12          0
11  Republic of Ireland      1  ...             6          1
12               Russia      5  ...             6          0
13                Spain     12  ...            11          0
14               Sweden      5  ...             7          0
15              Ukraine      2  ...             5          0

[16 rows x 32 columns]
```

```
euro12.loc[euro12.Team.isin(['England', 'Italy', 'Russia']),
['Team','Shooting Accuracy']]
```

```
       Team Shooting Accuracy
3   England             50.0%
7     Italy             43.0%
12   Russia             22.5%
```