HAM_panda

March 30, 2025

1 PSMDSRC103 - March 23 2025

1.1 pandas

[2]:
```
# !pip install pandas
import pandas as pd
```

[6]:
```
meteorites_df = pd.read_csv('Meteorite_Landings.csv')
```

[10]:
```
meteorites_df.name
```

[10]:
```
0             Aachen
1             Aarhus
2               Abee
3           Acapulco
4            Achiras
             …
45711     Zillah 002
45712         Zinder
45713           Zlin
45714      Zubkovsky
45715     Zulu Queen
Name: name, Length: 45716, dtype: object
```

[12]:
```
for i in list(meteorites_df.columns):
    print(i)
```

```
name
id
nametype
recclass
mass (g)
fall
year
reclat
reclong
GeoLocation
```

[14]:
```
meteorites_df.year
```

[14]:
```
0        01/01/1880 12:00:00 AM
1        01/01/1951 12:00:00 AM
2        01/01/1952 12:00:00 AM
3        01/01/1976 12:00:00 AM
4        01/01/1902 12:00:00 AM
                  …
45711    01/01/1990 12:00:00 AM
45712    01/01/1999 12:00:00 AM
45713    01/01/1939 12:00:00 AM
45714    01/01/2003 12:00:00 AM
45715    01/01/1976 12:00:00 AM
Name: year, Length: 45716, dtype: object
```

[16]:
```
count(meteorites_df.year)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[16], line 1
----> 1 count(meteorites_df.year)

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

[18]:
```
print(meteorites_df.year)
```

```
0        01/01/1880 12:00:00 AM
1        01/01/1951 12:00:00 AM
2        01/01/1952 12:00:00 AM
3        01/01/1976 12:00:00 AM
4        01/01/1902 12:00:00 AM
                  …
45711    01/01/1990 12:00:00 AM
45712    01/01/1999 12:00:00 AM
45713    01/01/1939 12:00:00 AM
45714    01/01/2003 12:00:00 AM
45715    01/01/1976 12:00:00 AM
Name: year, Length: 45716, dtype: object
```

[20]:
```
import requests
response = requests.get('https://data.nasa.gov/resource/gh4g-9sfh.json', params␣
 ↪= {'$limit':50_000})
if response.ok:
    payload = response.json()
else:
    print(f'Request was not successful and return code: {response.status_code}.
 ↪')
    payload = None
```

[26]:

```
#payload
```

[30]:

```
meteorites_df = pd.DataFrame(payload)
meteorites_df
```

[30]:

```
             name     id nametype              recclass    mass   fall  \
0          Aachen      1    Valid                    L5      21   Fell
1          Aarhus      2    Valid                    H6     720   Fell
2            Abee      6    Valid                   EH4  107000   Fell
3        Acapulco     10    Valid           Acapulcoite    1914   Fell
4         Achiras    370    Valid                    L6     780   Fell
…           …    …      …                   …      …    …
45711  Zillah 002  31356    Valid               Eucrite     172  Found
45712      Zinder  30409    Valid  Pallasite, ungrouped      46  Found
45713        Zlin  30410    Valid                    H4     3.3  Found
45714   Zubkovsky  31357    Valid                    L6    2167  Found
45715  Zulu Queen  30414    Valid                  L3.7     200  Found

                          year      reclat      reclong  \
0      1880-01-01T00:00:00.000   50.775000     6.083330
1      1951-01-01T00:00:00.000   56.183330    10.233330
2      1952-01-01T00:00:00.000   54.216670  -113.000000
3      1976-01-01T00:00:00.000   16.883330   -99.900000
4      1902-01-01T00:00:00.000  -33.166670   -64.950000
…                          …         …          …
45711  1990-01-01T00:00:00.000   29.037000    17.018500
45712  1999-01-01T00:00:00.000   13.783330     8.966670
45713  1939-01-01T00:00:00.000   49.250000    17.666670
45714  2003-01-01T00:00:00.000   49.789170    41.504600
45715  1976-01-01T00:00:00.000   33.983330  -115.683330

                                             geolocation  \
0         {'latitude': '50.775', 'longitude': '6.08333'}
1      {'latitude': '56.18333', 'longitude': '10.23333'}
2        {'latitude': '54.21667', 'longitude': '-113.0'}
3         {'latitude': '16.88333', 'longitude': '-99.9'}
4       {'latitude': '-33.16667', 'longitude': '-64.95'}
…                                                    …
45711     {'latitude': '29.037', 'longitude': '17.0185'}
45712   {'latitude': '13.78333', 'longitude': '8.96667'}
45713     {'latitude': '49.25', 'longitude': '17.66667'}
45714   {'latitude': '49.78917', 'longitude': '41.5046'}
45715  {'latitude': '33.98333', 'longitude': '-115.68…

      :@computed_region_cbhk_fwbd :@computed_region_nnqa_25f4
0                             NaN                         NaN
1                             NaN                         NaN
```

```
2                             NaN                         NaN
3                             NaN                         NaN
4                             NaN                         NaN
…                           …                         …
45711                         NaN                         NaN
45712                         NaN                         NaN
45713                         NaN                         NaN
45714                         NaN                         NaN
45715                           8                        1177

[45716 rows x 12 columns]
```

[32]:

```
meteorites_df.columns
```

[32]:

```
Index(['name', 'id', 'nametype', 'recclass', 'mass', 'fall', 'year', 'reclat',
       'reclong', 'geolocation', ':@computed_region_cbhk_fwbd',
       ':@computed_region_nnqa_25f4'],
      dtype='object')
```

[34]:

```
#ganito best practice for getting flat/csv files
filepath = 'Meteorite_Landings.csv'
meteorites_df = pd.read_csv(filepath)
meteorites_df
```

[34]:

```
             name     id nametype              recclass  mass (g)  fall  \
0          Aachen      1    Valid                    L5      21.0  Fell
1          Aarhus      2    Valid                    H6     720.0  Fell
2            Abee      6    Valid                   EH4  107000.0  Fell
3        Acapulco     10    Valid           Acapulcoite    1914.0  Fell
4         Achiras    370    Valid                    L6     780.0  Fell
…           …    …      …                   …       …    …
45711  Zillah 002  31356    Valid               Eucrite     172.0  Found
45712      Zinder  30409    Valid  Pallasite, ungrouped      46.0  Found
45713        Zlin  30410    Valid                    H4       3.3  Found
45714   Zubkovsky  31357    Valid                    L6    2167.0  Found
45715  Zulu Queen  30414    Valid                  L3.7     200.0  Found

                         year    reclat    reclong             GeoLocation
0      01/01/1880 12:00:00 AM  50.77500    6.08333       (50.775, 6.08333)
1      01/01/1951 12:00:00 AM  56.18333   10.23333    (56.18333, 10.23333)
2      01/01/1952 12:00:00 AM  54.21667 -113.00000      (54.21667, -113.0)
3      01/01/1976 12:00:00 AM  16.88333  -99.90000       (16.88333, -99.9)
4      01/01/1902 12:00:00 AM -33.16667  -64.95000     (-33.16667, -64.95)
…                       …       …        …                     …
45711  01/01/1990 12:00:00 AM  29.03700   17.01850       (29.037, 17.0185)
45712  01/01/1999 12:00:00 AM  13.78333    8.96667     (13.78333, 8.96667)
45713  01/01/1939 12:00:00 AM  49.25000   17.66667       (49.25, 17.66667)
```

```
45714  01/01/2003 12:00:00 AM  49.78917   41.50460     (49.78917, 41.5046)
45715  01/01/1976 12:00:00 AM  33.98333 -115.68333  (33.98333, -115.68333)

[45716 rows x 10 columns]
```

[38]:
```
meteorites_df.shape #output is (count rows, count columns)
```

[38]:
```
(45716, 10)
```

[40]:
```
meteorites_df.dtypes # ang output ay mga data types?
```

[40]:
```
name            object
id               int64
nametype        object
recclass        object
mass (g)       float64
fall            object
year            object
reclat         float64
reclong        float64
GeoLocation     object
dtype: object
```

[44]:
```
meteorites_df.head() # default ay first 5 rows kapag walang number sa loob ng␣
↪par
```

[44]:
```
       name   id nametype     recclass  mass (g)  fall  \
0    Aachen    1    Valid           L5      21.0  Fell
1    Aarhus    2    Valid           H6     720.0  Fell
2      Abee    6    Valid          EH4  107000.0  Fell
3  Acapulco   10    Valid  Acapulcoite    1914.0  Fell
4   Achiras  370    Valid           L6     780.0  Fell

                     year    reclat    reclong           GeoLocation
0  01/01/1880 12:00:00 AM  50.77500    6.08333     (50.775, 6.08333)
1  01/01/1951 12:00:00 AM  56.18333   10.23333  (56.18333, 10.23333)
2  01/01/1952 12:00:00 AM  54.21667 -113.00000    (54.21667, -113.0)
3  01/01/1976 12:00:00 AM  16.88333  -99.90000     (16.88333, -99.9)
4  01/01/1902 12:00:00 AM -33.16667  -64.95000   (-33.16667, -64.95)
```

[46]:
```
meteorites_df.tail()# default ay last 5 rows kapag walang number sa loob
```

[46]:
```
             name     id nametype              recclass  mass (g)   fall  \
45711  Zillah 002  31356    Valid               Eucrite     172.0  Found
45712      Zinder  30409    Valid  Pallasite, ungrouped      46.0  Found
45713        Zlin  30410    Valid                    H4       3.3  Found
45714   Zubkovsky  31357    Valid                    L6    2167.0  Found
45715  Zulu Queen  30414    Valid                  L3.7     200.0  Found
```

```
                          year    reclat    reclong           GeoLocation
45711  01/01/1990 12:00:00 AM  29.03700   17.01850      (29.037, 17.0185)
45712  01/01/1999 12:00:00 AM  13.78333    8.96667    (13.78333, 8.96667)
45713  01/01/1939 12:00:00 AM  49.25000   17.66667      (49.25, 17.66667)
45714  01/01/2003 12:00:00 AM  49.78917   41.50460    (49.78917, 41.5046)
45715  01/01/1976 12:00:00 AM  33.98333 -115.68333  (33.98333, -115.68333)
```

[50]:

```
meteorites_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45716 entries, 0 to 45715
Data columns (total 10 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   name         45716 non-null  object
 1   id           45716 non-null  int64
 2   nametype     45716 non-null  object
 3   recclass     45716 non-null  object
 4   mass (g)     45585 non-null  float64
 5   fall         45716 non-null  object
 6   year         45425 non-null  object
 7   reclat       38401 non-null  float64
 8   reclong      38401 non-null  float64
 9   GeoLocation  38401 non-null  object
dtypes: float64(3), int64(1), object(6)
memory usage: 3.5+ MB
```

[70]:

```
meteorites_df[['recclass','name']]
```

[70]:

```
                    recclass        name
0                         L5      Aachen
1                         H6      Aarhus
2                        EH4        Abee
3                Acapulcoite    Acapulco
4                         L6     Achiras
…                          …         …
45711                Eucrite  Zillah 002
45712  Pallasite, ungrouped      Zinder
45713                     H4        Zlin
45714                     L6   Zubkovsky
45715                   L3.7  Zulu Queen

[45716 rows x 2 columns]
```

[94]:

```
meteorites_df.iloc[100:104, [0,3,4,6]]
meteorites_df.iloc[:,[0,3,4,6]]
```

```
[94]:              name              recclass  mass (g)                    year
      0          Aachen                    L5      21.0  01/01/1880 12:00:00 AM
      1          Aarhus                    H6     720.0  01/01/1951 12:00:00 AM
      2            Abee                   EH4  107000.0  01/01/1952 12:00:00 AM
      3        Acapulco           Acapulcoite    1914.0  01/01/1976 12:00:00 AM
      4         Achiras                    L6     780.0  01/01/1902 12:00:00 AM
      …            …                    …       …                  …
      45711  Zillah 002               Eucrite     172.0  01/01/1990 12:00:00 AM
      45712      Zinder  Pallasite, ungrouped      46.0  01/01/1999 12:00:00 AM
      45713        Zlin                    H4       3.3  01/01/1939 12:00:00 AM
      45714   Zubkovsky                    L6    2167.0  01/01/2003 12:00:00 AM
      45715  Zulu Queen                  L3.7     200.0  01/01/1976 12:00:00 AM

      [45716 rows x 4 columns]
```

```
[102]: meteorites_df.loc[:,'mass (g)':'year']
```

```
[102]:        mass (g)   fall                    year
       0          21.0   Fell  01/01/1880 12:00:00 AM
       1         720.0   Fell  01/01/1951 12:00:00 AM
       2      107000.0   Fell  01/01/1952 12:00:00 AM
       3        1914.0   Fell  01/01/1976 12:00:00 AM
       4         780.0   Fell  01/01/1902 12:00:00 AM
       …          …    …                     …
       45711     172.0  Found  01/01/1990 12:00:00 AM
       45712      46.0  Found  01/01/1999 12:00:00 AM
       45713       3.3  Found  01/01/1939 12:00:00 AM
       45714    2167.0  Found  01/01/2003 12:00:00 AM
       45715     200.0  Found  01/01/1976 12:00:00 AM

       [45716 rows x 3 columns]
```

```
[152]: #boolean masks: Boolean mask is an array-like structure of Boolean values to␣
        ↪specify columns/rows to select (True) or not (False).
       meteorites_df[meteorites_df['mass (g)']>50]
```

```
[152]:              name     id nametype     recclass  mass (g)   fall  \
       1          Aarhus      2    Valid           H6     720.0   Fell
       2            Abee      6    Valid          EH4  107000.0   Fell
       3        Acapulco     10    Valid  Acapulcoite    1914.0   Fell
       4         Achiras    370    Valid           L6     780.0   Fell
       5        Adhi Kot    379    Valid          EH4    4239.0   Fell
       …            …    …      …          …       …    …
       45709  Zhongxiang  30406    Valid         Iron  100000.0  Found
       45710  Zillah 001  31355    Valid           L6    1475.0  Found
       45711  Zillah 002  31356    Valid      Eucrite     172.0  Found
       45714   Zubkovsky  31357    Valid           L6    2167.0  Found
```

```
45715  Zulu Queen  30414    Valid         L3.7     200.0  Found

                         year    reclat    reclong             GeoLocation
1      01/01/1951 12:00:00 AM  56.18333   10.23333    (56.18333, 10.23333)
2      01/01/1952 12:00:00 AM  54.21667 -113.00000      (54.21667, -113.0)
3      01/01/1976 12:00:00 AM  16.88333  -99.90000       (16.88333, -99.9)
4      01/01/1902 12:00:00 AM -33.16667  -64.95000     (-33.16667, -64.95)
5      01/01/1919 12:00:00 AM  32.10000   71.80000            (32.1, 71.8)
…                        …       …        …                       …
45709  01/01/1981 12:00:00 AM  31.20000  112.50000           (31.2, 112.5)
45710  01/01/1990 12:00:00 AM  29.03700   17.01850       (29.037, 17.0185)
45711  01/01/1990 12:00:00 AM  29.03700   17.01850       (29.037, 17.0185)
45714  01/01/2003 12:00:00 AM  49.78917   41.50460     (49.78917, 41.5046)
45715  01/01/1976 12:00:00 AM  33.98333 -115.68333  (33.98333, -115.68333)

[19874 rows x 10 columns]
```

```
[136]: meteorites_df[meteorites_df.fall == 'Found']
```

```
[136]:
                        name     id nametype              recclass  mass (g)  \
37     Northwest Africa 5815  50693    Valid                    L5     256.8
520      Cumulus Hills 04075  32531    Valid             Pallasite       9.6
757     Dominion Range 03239  32591    Valid                    L6      69.5
804     Dominion Range 03240  32592    Valid                   LL5     290.9
1111                   Abajo      4    Valid                    H5     331.0
…                      …    …      …                    …       …
45711             Zillah 002  31356    Valid               Eucrite     172.0
45712                 Zinder  30409    Valid  Pallasite, ungrouped      46.0
45713                   Zlin  30410    Valid                    H4       3.3
45714              Zubkovsky  31357    Valid                    L6    2167.0
45715             Zulu Queen  30414    Valid                  L3.7     200.0

        fall                    year    reclat    reclong  \
37     Found                     NaN  0.00000    0.00000
520    Found  01/01/2003 12:00:00 AM       NaN        NaN
757    Found  01/01/2002 12:00:00 AM       NaN        NaN
804    Found  01/01/2002 12:00:00 AM       NaN        NaN
1111   Found  01/01/1982 12:00:00 AM  26.80000 -105.41667
…      …                     …       …        …
45711  Found  01/01/1990 12:00:00 AM  29.03700   17.01850
45712  Found  01/01/1999 12:00:00 AM  13.78333    8.96667
45713  Found  01/01/1939 12:00:00 AM  49.25000   17.66667
45714  Found  01/01/2003 12:00:00 AM  49.78917   41.50460
45715  Found  01/01/1976 12:00:00 AM  33.98333 -115.68333

                  GeoLocation
37                 (0.0, 0.0)
```

```
520                        NaN
757                        NaN
804                        NaN
1111       (26.8, -105.41667)
…                          …
45711        (29.037, 17.0185)
45712      (13.78333, 8.96667)
45713        (49.25, 17.66667)
45714      (49.78917, 41.5046)
45715   (33.98333, -115.68333)

[44609 rows x 10 columns]
```

```
[142]: meteorites_df[(meteorites_df['mass (g)']>50) & (meteorites_df.fall == 'Found')]
```

```
[142]:                          name     id nametype recclass   mass (g)   fall  \
       37     Northwest Africa 5815  50693    Valid       L5     256.80  Found
       757      Dominion Range 03239  32591    Valid       L6      69.50  Found
       804      Dominion Range 03240  32592    Valid      LL5     290.90  Found
       1111                    Abajo      4    Valid       H5     331.00  Found
       1112          Abar al' Uj 001  51399    Valid     H3.8     194.34  Found
       …                          …    …      …      …        …      …
       45709              Zhongxiang  30406    Valid     Iron  100000.00  Found
       45710              Zillah 001  31355    Valid       L6    1475.00  Found
       45711              Zillah 002  31356    Valid  Eucrite     172.00  Found
       45714               Zubkovsky  31357    Valid       L6    2167.00  Found
       45715              Zulu Queen  30414    Valid     L3.7     200.00  Found

                               year    reclat    reclong              GeoLocation
       37                       NaN   0.00000    0.00000               (0.0, 0.0)
       757    01/01/2002 12:00:00 AM       NaN        NaN                      NaN
       804    01/01/2002 12:00:00 AM       NaN        NaN                      NaN
       1111   01/01/1982 12:00:00 AM  26.80000 -105.41667       (26.8, -105.41667)
       1112   01/01/2008 12:00:00 AM  22.72192   48.95937    (22.72192, 48.95937)
       …                          …       …         …                     …
       45709  01/01/1981 12:00:00 AM  31.20000  112.50000          (31.2, 112.5)
       45710  01/01/1990 12:00:00 AM  29.03700   17.01850        (29.037, 17.0185)
       45711  01/01/1990 12:00:00 AM  29.03700   17.01850        (29.037, 17.0185)
       45714  01/01/2003 12:00:00 AM  49.78917   41.50460      (49.78917, 41.5046)
       45715  01/01/1976 12:00:00 AM  33.98333 -115.68333   (33.98333, -115.68333)

       [18854 rows x 10 columns]
```

```
[150]: meteorites_df.query("`mass (g)` > 1e6 and fall == 'Fell'")
```

```
[150]:              name     id nametype     recclass    mass (g)  fall  \
       29         Allende   2278    Valid          CV3   2000000.0  Fell
```

```
419          Jilin  12171    Valid          H5   4000000.0  Fell
506  Kunya-Urgench  12379    Valid          H5   1100000.0  Fell
707  Norton County  17922    Valid     Aubrite   1100000.0  Fell
920   Sikhote-Alin  23593    Valid  Iron, IIAB  23000000.0  Fell

                       year    reclat    reclong             GeoLocation
29   01/01/1969 12:00:00 AM  26.96667 -105.31667  (26.96667, -105.31667)
419  01/01/1976 12:00:00 AM  44.05000  126.16667      (44.05, 126.16667)
506  01/01/1998 12:00:00 AM  42.25000   59.20000           (42.25, 59.2)
707  01/01/1948 12:00:00 AM  39.68333  -99.86667   (39.68333, -99.86667)
920  01/01/1947 12:00:00 AM  46.16000  134.65333      (46.16, 134.65333)
```

[162]:

```
#calculating summary statistics
meteorites_df.fall.value_counts()
```

[162]:

```
fall
Found    44609
Fell      1107
Name: count, dtype: int64
```

[166]:

```
meteorites_df.value_counts(subset=['nametype','fall'],normalize=True)
```

[166]:

```
nametype  fall
Valid     Found    0.974145
          Fell     0.024215
Relict    Found    0.001641
Name: proportion, dtype: float64
```

[168]:

```
print(meteorites_df['mass (g)'].mean())
```

```
13278.078548601512
```

[170]:

```
print(meteorites_df['mass (g)'].median())
```

```
32.6
```

[172]:

```
print(meteorites_df['mass (g)'].mode())
```

```
0    1.3
Name: mass (g), dtype: float64
```

[174]:

```
meteorites_df['mass (g)'].quantile([0.25,0.5,0.75])
```

[174]:

```
0.25      7.2
0.50     32.6
0.75    202.6
Name: mass (g), dtype: float64
```

[176]:

```
meteorites_df['mass (g)'].min()
meteorites_df['mass (g)'].max()
```

[176]:

```
60000000.0
```

[178]:

```
meteorites_df.iloc[meteorites_df['mass (g)'].idxmax()]
```

[178]:

```
name                                Hoba
id                                 11890
nametype                           Valid
recclass                       Iron, IVB
mass (g)                      60000000.0
fall                               Found
year              01/01/1920 12:00:00 AM
reclat                         -19.58333
reclong                         17.91667
GeoLocation         (-19.58333, 17.91667)
Name: 16392, dtype: object
```

[180]:

```
meteorites_df.iloc[meteorites_df['mass (g)'].idxmin()]
```

[180]:

```
name                                Gove
id                                 52859
nametype                          Relict
recclass                     Relict iron
mass (g)                             0.0
fall                               Found
year              01/01/1979 12:00:00 AM
reclat                         -12.26333
reclong                        136.83833
GeoLocation        (-12.26333, 136.83833)
Name: 12640, dtype: object
```

[182]:

```
meteorites_df.describe()
```

[182]:

```
                 id      mass (g)        reclat       reclong
count  45716.000000  4.558500e+04  38401.000000  38401.000000
mean   26889.735104  1.327808e+04    -39.122580     61.074319
std    16860.683030  5.749889e+05     46.378511     80.647298
min        1.000000  0.000000e+00    -87.366670   -165.433330
25%    12688.750000  7.200000e+00    -76.714240      0.000000
50%    24261.500000  3.260000e+01    -71.500000     35.666670
75%    40656.750000  2.026000e+02      0.000000    157.166670
max    57458.000000  6.000000e+07     81.166670    354.473330
```

[184]:

```
meteorites_df.describe(include='all')
```

[184]:

```
          name            id nametype recclass      mass (g)   fall  \
count    45716  45716.000000    45716    45716  4.558500e+04  45716   
unique   45716           NaN        2      466           NaN      2   
top     Aachen           NaN    Valid       L6           NaN  Found   
freq         1           NaN    45641     8285           NaN  44609   
mean       NaN  26889.735104      NaN      NaN  1.327808e+04    NaN   
std        NaN  16860.683030      NaN      NaN  5.749889e+05    NaN   
min        NaN      1.000000      NaN      NaN  0.000000e+00    NaN   
25%        NaN  12688.750000      NaN      NaN  7.200000e+00    NaN   
50%        NaN  24261.500000      NaN      NaN  3.260000e+01    NaN   
75%        NaN  40656.750000      NaN      NaN  2.026000e+02    NaN   
max        NaN  57458.000000      NaN      NaN  6.000000e+07    NaN   

                          year        reclat       reclong GeoLocation
count                    45425  38401.000000  38401.000000       38401
unique                     266           NaN           NaN       17100
top     01/01/2003 12:00:00 AM           NaN           NaN  (0.0, 0.0)
freq                      3323           NaN           NaN        6214
mean                       NaN    -39.122580     61.074319         NaN
std                        NaN     46.378511     80.647298         NaN
min                        NaN    -87.366670   -165.433330         NaN
25%                        NaN    -76.714240      0.000000         NaN
50%                        NaN    -71.500000     35.666670         NaN
75%                        NaN      0.000000    157.166670         NaN
max                        NaN     81.166670    354.473330         NaN
```

[192]:

```
meteorites_df.recclass.nunique()
```

[192]:

```
466
```

2 Seatwork 6.1 Programming Exercise: Getting Started with Pandas!

2.1 Using the 2019_Yellow_Taxi_Trip_Data.csv dataset, accomplish the following items and submit a PDF of the notebook:

2.2 (25 marks) Create a DataFrame by reading in the 2019_Yellow_Taxi_Trip_Data.csv file. Examine the first 5 rows.

2.3 (25 marks) Find the dimensions (number of rows and number of columns) in the data.

2.4 (25 marks) Using the data in the 2019_Yellow_Taxi_Trip_Data.csv file, calculate summary statistics for the fare_amount, tip_amount, tolls_amount, and total_amount columns.

2.5 (25 marks) Isolate the fare_amount, tip_amount, tolls_amount, and total_amount for the longest trip by distance (trip_distance).

2.6 Submit the whole notebook as PDF, with the exercise section at the end of the notebook. Include a link to your Github repo in the comment section.

```
[204]: # create a dataframe
       filename2 = '2019_Yellow_Taxi_Trip_Data.csv'
       ttdata = pd.read_csv(filename2)

       #examine the first 5 rows
       ttdata.head() #default is 5 if you do not put a number
```

```
[204]:    vendorid     tpep_pickup_datetime    tpep_dropoff_datetime  \
       0         2  2019-10-23T16:39:42.000  2019-10-23T17:14:10.000
       1         1  2019-10-23T16:32:08.000  2019-10-23T16:45:26.000
       2         2  2019-10-23T16:08:44.000  2019-10-23T16:21:11.000
       3         2  2019-10-23T16:22:44.000  2019-10-23T16:43:26.000
       4         2  2019-10-23T16:45:11.000  2019-10-23T16:58:49.000

          passenger_count  trip_distance  ratecodeid store_and_fwd_flag  \
       0                1           7.93           1                  N
       1                1           2.00           1                  N
       2                1           1.36           1                  N
       3                1           1.00           1                  N
       4                1           1.96           1                  N

          pulocationid  dolocationid  payment_type  fare_amount  extra  mta_tax  \
       0           138           170             1         29.5    1.0      0.5
       1            11            26             1         10.5    1.0      0.5
       2           163           162             1          9.5    1.0      0.5
       3           170           163             1         13.0    1.0      0.5
       4           163           236             1         10.5    1.0      0.5
```

```
   tip_amount  tolls_amount  improvement_surcharge  total_amount  \
0        7.98          6.12                    0.3         47.90
1        0.00          0.00                    0.3         12.30
2        2.00          0.00                    0.3         15.80
3        4.32          0.00                    0.3         21.62
4        0.50          0.00                    0.3         15.30

   congestion_surcharge
0                   2.5
1                   0.0
2                   2.5
3                   2.5
4                   2.5
```

[206]:

```
#find the dimensions (number of rows and number of columns) in the data
ttdata.shape
```

[206]:

```
(10000, 18)
```

[234]:

```
#calculate summary statistics for the fare_amount, tip_amount, tolls_amount,␣
 ↪and total_amount columns
subttdata = ttdata.loc[:
 ↪,['fare_amount','tip_amount','tolls_amount','total_amount']]
subttdata.describe()
```

[234]:

```
        fare_amount    tip_amount  tolls_amount  total_amount
count  10000.000000  10000.000000  10000.000000  10000.000000
mean      15.106313      2.634494      0.623447     22.564659
std       13.954762      3.409800      6.437507     19.209255
min      -52.000000      0.000000     -6.120000    -65.920000
25%        7.000000      0.000000      0.000000     12.375000
50%       10.000000      2.000000      0.000000     16.300000
75%       16.000000      3.250000      0.000000     22.880000
max      176.000000     43.000000    612.000000    671.800000
```

[220]:

```
#Isolate the fare_amount, tip_amount, tolls_amount, and total_amount for the␣
 ↪longest trip by distance (trip_distance).
#ttdata.loc[ttdata['trip_distance'].idxmax()]
ttdata.loc[ttdata['trip_distance'].idxmax(), ['fare_amount',␣
 ↪'tip_amount','tolls_amount','total_amount']]
```

[220]:

```
fare_amount      176.0
tip_amount       18.29
tolls_amount      6.12
total_amount    201.21
Name: 8338, dtype: object
```

[ ]:

```
```