20260429_dataIO

April 29, 2026

1 Working with data in Python: Data import, export, and cleaning

1.1 About me

- Background in Data Science
- PhD candidate at UGent Crime Lab, Ghent University (Belgium)
- Research topics:
  - Social aspects of criminal behavior
  - Criminal networks
  - Formal models of co-offending
- Tools:
  - Computational modeling
  - Agent-based modeling
  - Social network analysis

1.2 Objectives

- Learn how to **import** and **export** data in Python
  - text files
  - CSV/Excel files
  - JSON files
  - API requests
- Learn how to clean and manipulate data in Python

[2]:
```
import os
```

[3]:
```
print(os.listdir())
```

```
['20260429_dataIO.ipynb', 'my_data.pkl', '3_5_Data_Cleaning_EDA_Lecture.ipynb',
'abstract.docx', 'denver_crime', 'swedish_crime',
'3_4_Internet_Data_Lecture.ipynb', 'state_crime.json', '.virtual_documents',
'.ipynb_checkpoints', 'key.txt', '3_1_Working_with_Files_Lecture.ipynb']
```

[5]:
```
if os.path.exists("20260429_dataIO.ipynb"):
    doAnalysis()
else:
    raise Error
```

[5]:
```
True
```

[6]:
```
os.chdir('..')
```

[7]:
```
print(os.listdir())
```

```
['20260304_CrimNetsPy', '.idea', '.virtual_documents', '.ipynb_checkpoints',
'dataViz', '20260429_dataIO', 'ABM', 'ideas']
```

[8]:
```
os.chdir('20260429_dataIO')
```

[9]:
```
print(os.listdir())
```

```
['20260429_dataIO.ipynb', 'my_data.pkl', '3_5_Data_Cleaning_EDA_Lecture.ipynb',
'abstract.docx', 'denver_crime', 'swedish_crime',
'3_4_Internet_Data_Lecture.ipynb', 'state_crime.json', '.virtual_documents',
'.ipynb_checkpoints', 'key.txt', '3_1_Working_with_Files_Lecture.ipynb']
```

[10]:
```
# create a new folder
os.makedirs("IACA_test", exist_ok=True)
```

[ ]:

[ ]:
```
os.makedirs()
```

[ ]:

[ ]:

[ ]:

1.3 Text files

You can acces text file in Python with the help of built-in `open()` function. `mode` options are:

- `r`: Opens the file in read-only mode. Starts reading from the beginning of the file and is the default mode for the open() function.
- `rb`: Opens the file as read-only in binary format and starts reading from the beginning of the file. While binary format can be used for different purposes, it is usually used when dealing with things like images, videos, etc.
- `r+`: Opens a file for reading and writing, placing the pointer at the beginning of the file.
- `w`: Opens in write-only mode. The pointer is placed at the beginning of the file and this will overwrite any existing file with the same name. It will create a new file if one with the same name doesn't exist.
- `wb`: Opens a write-only file in binary mode.
- `w+`: Opens a file for writing and reading.
- `wb+`: Opens a file for writing and reading in binary mode.
- `a`: Opens a file for appending new information to it. The pointer is placed at the end of the file. A new file is created if one with the same name doesn't exist.
- `ab`: Opens a file for appending in binary mode.

- a+: Opens a file for both appending and reading.
- ab+: Opens a file for both appending and reading in binary mode.

[19]:
```
key = open(file='key.txt', mode='r')
output = key.read()
print(output)
```

```
nskfnsdkknfk4kn2nknnlU9u2390+sef29
```

[14]:
```
key.closed
```

[14]:
```
False
```

[16]:
```
key.close()
```

[17]:
```
key.closed
```

[17]:
```
True
```

[20]:
```
with open(file='key.txt', mode='r') as file:
    output = file.read()

print(output)
```

```
nskfnsdkknfk4kn2nknnlU9u2390+sef29
```

[21]:
```
with open(file='key.txt', mode='r') as file:
    output = file.read()

print(output)
```

```
nskfnsdkknfk4kn2nknnlU9u2390+sef29
jkewnfjnewkfnwj
32r92jj2f3
sdkvnknwve
```

[24]:
```
data_from_file = output.split('\n')
```

[26]:
```
data_from_file[2]
```

[26]:
```
'32r92jj2f3'
```

[ ]:
```

```

[27]:
```
with open("key.txt", mode="r") as file:
    output = file.readlines()

print(output)
```

```
['nskfnsdkknfk4kn2nknnlU9u2390+sef29\n', 'jkewnfjnewkfnwj\n', '32r92jj2f3\n',
'sdkvnknwve']
```

[29]:
```
output_new = []
for item in output:
    temp = item.replace('\n', '')
    output_new.append(temp)
print(output_new)
```

```
['nskfnsdkknfk4kn2nknnlU9u2390+sef29', 'jkewnfjnewkfnwj', '32r92jj2f3',
'sdkvnknwve']
```

[15]:
```
with open("key.txt", mode="r") as file:
    key = file.read()

print(key)
```

```
nskfnsdkknfk4kn2nknnlU9u2390+sef29
efnkn32k=*@hknnfklw
Mkfneknw09(KIn
Kn!Kn21kn31[l
```

[18]:
```
with open("key.txt", mode="r") as file:
    # save each line as an element in a list
    key = file.readlines()
print(key)
```

```
['nskfnsdkknfk4kn2nknnlU9u2390+sef29\n', 'efnkn32k=*@hknnfklw\n',
'Mkfneknw09(KIn\n', 'Kn!Kn21kn31[l']
```

[30]:
```
# append
with open("key.txt", mode="a") as file:
    file.write("\nThis is a new line.")
```

[ ]:

[ ]:

[31]:
```
# rewrite
if not os.path.exists('key.txt'):
    with open("key.txt", mode="w") as file:
        file.write("Oh no...")
else:
    with open("key.txt", mode="a") as file:
        file.write("\nOh no...")
```

[ ]:

[ ]:

[ ]:

[ ]:

1.4 CSV files

A **comma-separated values (CSV)** file is a delimited text file that uses a comma to separate values. A CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.

Denver Crime Data: https://www.kaggle.com/datasets/paultimothymooney/denver-crime-data

[32]:
```
import pandas as pd
```

[33]:
```
swedish_crimes = pd.read_csv("swedish_crime/Swedish_crime_rates.csv", sep=',')
```

[34]:
```
swedish_crimes.head()
```

[34]:

	Year	crimes.total	crimes.penal.code	crimes.person	murder	assault	\
0	1950	2784	2306	120	1	105	
1	1951	3284	2754	125	1	109	
2	1952	3160	2608	119	1	104	
3	1953	2909	2689	119	1	105	
4	1954	3028	2791	126	1	107	

	sexual.offenses	rape	stealing.general	burglary	…	vehicle.theft	\
0	40	5	1578	295	…	NaN	
1	45	6	1899	342	…	NaN	
2	39	4	1846	372	…	NaN	
3	45	5	1929	361	…	NaN	
4	41	5	1981	393	…	NaN	

	out.of.vehicle.theft	shop.theft	robbery	fraud	criminal.damage	\
0	NaN	NaN	3	209	72	
1	NaN	NaN	3	310	73	
2	NaN	NaN	3	217	82	
3	NaN	NaN	4	209	88	
4	NaN	NaN	4	236	101	

	other.penal.crimes	narcotics	drunk.driving	population
0	477	0.0	49	7014000
1	530	0.0	66	7073000
2	553	0.0	78	7125000
3	220	0.0	91	7171000
4	237	0.0	103	7213000

```
[5 rows x 21 columns]
```

[36]:
```
swedish_crimes['crimes.total'].mean()
```

[36]:
```
10084.257575757576
```

[39]:
```
swedish_crimes_copy = pd.read_csv(
    "swedish_crime/Swedish_crime_rates_copy.csv",
    sep='\t')
swedish_crimes_copy
```

[39]:

	Year	crimes.total	crimes.penal.code	crimes.person	murder	assault	\
0	1950	2784	2306	120	1	105	
1	1951	3284	2754	125	1	109	
2	1952	3160	2608	119	1	104	
3	1953	2909	2689	119	1	105	
4	1954	3028	2791	126	1	107	
..	…	…	…	…	…	…	
61	2011	14988	12267	1005	2	947	
62	2012	14734	12148	974	2	914	
63	2013	14603	12072	896	3	837	
64	2014	14890	12305	923	3	859	
65	2015	15342	12803	932	3	868	

	sexual.offenses	rape	stealing.general	burglary	…	vehicle.theft	\
0	40	5	1578	295	…	NaN	
1	45	6	1899	342	…	NaN	
2	39	4	1846	372	…	NaN	
3	45	5	1929	361	…	NaN	
4	41	5	1981	393	…	NaN	
..	…	…	…	…	…	…	
61	181	69	5783	995	…	213.0	
62	178	66	5613	922	…	169.0	
63	184	63	5552	888	…	159.0	
64	210	69	5572	913	…	147.0	
65	184	60	5404	917	…	135.0	

	out.of.vehicle.theft	shop.theft	robbery	fraud	criminal.damage	\
0	NaN	NaN	3	209	72	
1	NaN	NaN	3	310	73	
2	NaN	NaN	3	217	82	
3	NaN	NaN	4	209	88	
4	NaN	NaN	4	236	101	
..	…	…	…	…	…	
61	706.0	663.0	103	1220	1774	
62	655.0	635.0	97	1356	1600	
63	543.0	509.0	87	1545	1470	

```
64                542.0       505.0       86   1610             1559
65                552.0       515.0       86   1896             1989

    other.penal.crimes  narcotics  drunk.driving  population
0                  477        0.0             49     7014000
1                  530        0.0             66     7073000
2                  553        0.0             78     7125000
3                  220        0.0             91     7171000
4                  237        0.0            103     7213000
..                 …          …              …          …
61                2721      946.0            314     9449000
62                2586      994.0            286     9519000
63                2532     1002.0            279     9644000
64                2585      983.0            271     9747000
65                2539      960.0            267     9851000

[66 rows x 21 columns]
```

[41]:
```
denver_crimes = pd.read_csv(
    "denver_crime/crime.csv",
    sep=",",
    encoding_errors="replace")
```

[42]:
```
denver_crimes
```

[42]:
```
        incident_id        offense_id  offense_code  offense_code_extension  \
0         202268791   202268791299900          2999                       0
1        2021387586  2021387586299900          2999                       0
2        2020641486  2020641486299900          2999                       0
3        2018612468  2018612468299900          2999                       0
4        2020293614  2020293614299900          2999                       0
…              …                 …             …                       …
386860   2023534182  2023534182240400          2404                       0
386861   2023534279  2023534279260900          2609                       0
386862   2023533509  2023533509131506          1315                       6
386863   2023533048  2023533048131506          1315                       6
386864   2023532323  2023532323131506          1315                       6

                 offense_type_id offense_category_id  first_occurrence_date  \
0        criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
1        criminal-mischief-other     public-disorder    7/7/2021 9:02:00 PM
2        criminal-mischief-other     public-disorder  10/29/2020 1:30:00 AM
3        criminal-mischief-other     public-disorder    9/6/2018 5:00:00 PM
4        criminal-mischief-other     public-disorder    5/8/2020 5:00:00 AM
…                            …                   …                      …
386860    theft-of-motor-vehicle          auto-theft  10/3/2023 11:30:00 AM
386861  fraud-by-use-of-computer  white-collar-crime  9/27/2023 11:00:00 AM
```

```
386862            agg-aslt-shoot  aggravated-assault  10/3/2023 10:44:00 AM
386863            agg-aslt-shoot  aggravated-assault   10/3/2023 5:30:00 AM
386864            agg-aslt-shoot  aggravated-assault   10/2/2023 5:47:00 PM

        last_occurrence_date          reported_date          incident_address  \
0                        NaN   2/10/2022 3:16:00 AM        1107 N SANTA FE DR
1                        NaN   7/8/2021 12:55:00 AM               815 16TH ST
2                        NaN  10/29/2020 4:31:00 AM       4745 N FEDERAL BLVD
3       9/6/2018 11:00:00 PM    9/7/2018 9:58:00 AM         65 S FEDERAL BLVD
4        5/8/2020 6:30:00 PM  5/13/2020 10:00:00 AM        12295 E ALBROOK DR
…                        …                   …                   …
386860  10/3/2023 3:00:00 PM   10/3/2023 5:40:00 PM        8501 E ALAMEDA AVE
386861                   NaN   10/3/2023 6:34:00 PM          517 E BAYAUD AVE
386862                   NaN  10/3/2023 12:03:00 PM  N JOLIET ST / E 55TH AVE
386863                   NaN   10/3/2023 9:47:00 AM           585 S ALTON WAY
386864                   NaN   10/2/2023 6:16:00 PM      1198 N SHERIDAN BLVD

            geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0       3140929.0  1692612.0 -104.998910  39.733957           1          123
1       3142470.0  1697098.0 -104.993342  39.746248           6          611
2       3133352.0  1710396.0 -105.025520  39.782888           1          111
3       3133534.0  1685797.0 -105.025330  39.715357           4          411
4       3184065.0  1710782.0 -104.845074  39.783082           5          521
…             …          …          …         …          …           …
386860  3171584.0  1684914.0 -104.890099  39.712310           3          321
386861  3146000.0  1685678.0 -104.981021  39.714843           3          311
386862  3179460.0  1715407.0 -104.861343  39.795869           5          511
386863  3173405.0  1682706.0 -104.883680  39.706214           3          321
386864  3125724.0  1692953.0 -105.052966  39.735110           1          122

       neighborhood_id  is_crime  is_traffic  victim_count
0         lincoln-park         1           0             1
1                  cbd         1           0             1
2             berkeley         1           0             1
3               barnum         1           0             1
4            montbello         1           0             1
…                    …       …           …             …
386860     lowry-field         1           0             1
386861           speer         1           0             1
386862       montbello         1           0             1
386863         windsor         1           0             1
386864     west-colfax         1           0             1

[386865 rows x 20 columns]
```

[44]:
```
denver_crimes.isna().sum()
```

[44]:
```
incident_id                 0
offense_id                  0
offense_code                0
offense_code_extension      0
offense_type_id             0
offense_category_id         0
first_occurrence_date       0
last_occurrence_date       175556
reported_date               0
incident_address            15503
geo_x                       15503
geo_y                       15503
geo_lon                     15769
geo_lat                     15769
district_id                 57
precinct_id                 0
neighborhood_id             689
is_crime                    0
is_traffic                  0
victim_count                0
dtype: int64
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:
```
swedish_crimes_copy = pd.read_csv("Swedish_crime_rates_copy.csv", sep='\t')
```

[ ]:
```
swedish_crimes_copy.head()
```

[51]:
```
denver_crimes = pd.read_csv("denver_crime/crime.csv", sep=",",␣
↪encoding_errors="replace")
```

[37]:
```
denver_crimes
```

[37]:

	incident_id	offense_id	offense_code	offense_code_extension	\
0	202268791	202268791299900	2999	0	
1	2021387586	2021387586299900	2999	0	
2	2020641486	2020641486299900	2999	0	
3	2018612468	2018612468299900	2999	0	
4	2020293614	2020293614299900	2999	0	
…	…	…	…	…	

```
386860   2023534182  2023534182240400          2404                      0
386861   2023534279  2023534279260900          2609                      0
386862   2023533509  2023533509131506          1315                      6
386863   2023533048  2023533048131506          1315                      6
386864   2023532323  2023532323131506          1315                      6

                offense_type_id offense_category_id  first_occurrence_date  \
0        criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
1        criminal-mischief-other     public-disorder    7/7/2021 9:02:00 PM
2        criminal-mischief-other     public-disorder  10/29/2020 1:30:00 AM
3        criminal-mischief-other     public-disorder    9/6/2018 5:00:00 PM
4        criminal-mischief-other     public-disorder    5/8/2020 5:00:00 AM
…                            …                   …                    …
386860    theft-of-motor-vehicle          auto-theft  10/3/2023 11:30:00 AM
386861  fraud-by-use-of-computer  white-collar-crime  9/27/2023 11:00:00 AM
386862           agg-aslt-shoot  aggravated-assault  10/3/2023 10:44:00 AM
386863           agg-aslt-shoot  aggravated-assault   10/3/2023 5:30:00 AM
386864           agg-aslt-shoot  aggravated-assault   10/2/2023 5:47:00 PM

       last_occurrence_date          reported_date          incident_address  \
0                       NaN   2/10/2022 3:16:00 AM        1107 N SANTA FE DR
1                       NaN   7/8/2021 12:55:00 AM               815 16TH ST
2                       NaN  10/29/2020 4:31:00 AM       4745 N FEDERAL BLVD
3       9/6/2018 11:00:00 PM    9/7/2018 9:58:00 AM         65 S FEDERAL BLVD
4        5/8/2020 6:30:00 PM  5/13/2020 10:00:00 AM        12295 E ALBROOK DR
…                        …                   …                    …
386860  10/3/2023 3:00:00 PM   10/3/2023 5:40:00 PM        8501 E ALAMEDA AVE
386861                  NaN   10/3/2023 6:34:00 PM          517 E BAYAUD AVE
386862                  NaN  10/3/2023 12:03:00 PM  N JOLIET ST / E 55TH AVE
386863                  NaN   10/3/2023 9:47:00 AM           585 S ALTON WAY
386864                  NaN   10/2/2023 6:16:00 PM      1198 N SHERIDAN BLVD

           geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0       3140929.0  1692612.0 -104.998910  39.733957           1          123
1       3142470.0  1697098.0 -104.993342  39.746248           6          611
2       3133352.0  1710396.0 -105.025520  39.782888           1          111
3       3133534.0  1685797.0 -105.025330  39.715357           4          411
4       3184065.0  1710782.0 -104.845074  39.783082           5          521
…             …          …          …         …          …           …
386860  3171584.0  1684914.0 -104.890099  39.712310           3          321
386861  3146000.0  1685678.0 -104.981021  39.714843           3          311
386862  3179460.0  1715407.0 -104.861343  39.795869           5          511
386863  3173405.0  1682706.0 -104.883680  39.706214           3          321
386864  3125724.0  1692953.0 -105.052966  39.735110           1          122

       neighborhood_id  is_crime  is_traffic  victim_count
0         lincoln-park         1           0             1
```

```
1                  cbd         1           0             1
2             berkeley         1           0             1
3               barnum         1           0             1
4            montbello         1           0             1
…                   …       …           …             …
386860     lowry-field         1           0             1
386861           speer         1           0             1
386862       montbello         1           0             1
386863         windsor         1           0             1
386864     west-colfax         1           0             1

[386865 rows x 20 columns]
```

[ ]:
```
denver_crimes.isna().sum()
```

[49]:
```
denver_crimes.dropna(how='any', axis='columns', inplace=True)
```

[50]:
```
denver_crimes
```

[50]:
```
        incident_id        offense_id  offense_code  offense_code_extension  \
0         202268791   202268791299900          2999                       0
1        2021387586  2021387586299900          2999                       0
2        2020641486  2020641486299900          2999                       0
3        2018612468  2018612468299900          2999                       0
4        2020293614  2020293614299900          2999                       0
…               …                 …             …                       …
386860   2023534182  2023534182240400          2404                       0
386861   2023534279  2023534279260900          2609                       0
386862   2023533509  2023533509131506          1315                       6
386863   2023533048  2023533048131506          1315                       6
386864   2023532323  2023532323131506          1315                       6

                offense_type_id offense_category_id  first_occurrence_date  \
0        criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
1        criminal-mischief-other     public-disorder    7/7/2021 9:02:00 PM
2        criminal-mischief-other     public-disorder  10/29/2020 1:30:00 AM
3        criminal-mischief-other     public-disorder    9/6/2018 5:00:00 PM
4        criminal-mischief-other     public-disorder    5/8/2020 5:00:00 AM
…                            …                   …                      …
386860    theft-of-motor-vehicle          auto-theft  10/3/2023 11:30:00 AM
386861  fraud-by-use-of-computer  white-collar-crime  9/27/2023 11:00:00 AM
386862           agg-aslt-shoot  aggravated-assault  10/3/2023 10:44:00 AM
386863           agg-aslt-shoot  aggravated-assault   10/3/2023 5:30:00 AM
386864           agg-aslt-shoot  aggravated-assault   10/2/2023 5:47:00 PM

              reported_date  precinct_id  is_crime  is_traffic  victim_count
0        2/10/2022 3:16:00 AM          123         1           0             1
```

```
1        7/8/2021 12:55:00 AM          611         1           0             1
2       10/29/2020 4:31:00 AM          111         1           0             1
3         9/7/2018 9:58:00 AM          411         1           0             1
4       5/13/2020 10:00:00 AM          521         1           0             1
…                          …          …       …           …            …
386860   10/3/2023 5:40:00 PM          321         1           0             1
386861   10/3/2023 6:34:00 PM          311         1           0             1
386862  10/3/2023 12:03:00 PM          511         1           0             1
386863   10/3/2023 9:47:00 AM          321         1           0             1
386864   10/2/2023 6:16:00 PM          122         1           0             1

[386865 rows x 12 columns]
```

[ ]:

[52]:
```
denver_crimes = pd.read_csv("denver_crime/crime.csv", sep=",",␣
↪encoding_errors="replace")

denver_crimes.isna().sum()
```

[52]:
```
incident_id                  0
offense_id                   0
offense_code                 0
offense_code_extension       0
offense_type_id              0
offense_category_id          0
first_occurrence_date        0
last_occurrence_date    175556
reported_date                0
incident_address         15503
geo_x                    15503
geo_y                    15503
geo_lon                  15769
geo_lat                  15769
district_id                 57
precinct_id                  0
neighborhood_id            689
is_crime                     0
is_traffic                   0
victim_count                 0
dtype: int64
```

[55]:
```
denver_crimes['last_occurrence_date'].isna()
```

[55]:
```
0         True
1         True
2         True
```

```
3         False
4         False
          …
386860    False
386861     True
386862     True
386863     True
386864     True
Name: last_occurrence_date, Length: 386865, dtype: bool
```

[56]:
```
denver_crimes.loc[denver_crimes['last_occurrence_date'].isna(),:]
```

[56]:

	incident_id	offense_id	offense_code	offense_code_extension	\
0	202268791	202268791299900	2999	0	
1	2021387586	2021387586299900	2999	0	
2	2020641486	2020641486299900	2999	0	
5	2020600882	2020600882299900	2999	0	
6	2020233739	2020233739299900	2999	0	
…	…	…	…	…	
386859	2023531878	2023531878240400	2404	0	
386861	2023534279	2023534279260900	2609	0	
386862	2023533509	2023533509131506	1315	6	
386863	2023533048	2023533048131506	1315	6	
386864	2023532323	2023532323131506	1315	6	

	offense_type_id	offense_category_id	first_occurrence_date	\
0	criminal-mischief-other	public-disorder	2/10/2022 2:50:00 AM	
1	criminal-mischief-other	public-disorder	7/7/2021 9:02:00 PM	
2	criminal-mischief-other	public-disorder	10/29/2020 1:30:00 AM	
5	criminal-mischief-other	public-disorder	10/8/2020 2:26:00 AM	
6	criminal-mischief-other	public-disorder	4/18/2020 6:30:00 AM	
…	…	…	…	
386859	theft-of-motor-vehicle	auto-theft	8/18/2023 12:01:00 AM	
386861	fraud-by-use-of-computer	white-collar-crime	9/27/2023 11:00:00 AM	
386862	agg-aslt-shoot	aggravated-assault	10/3/2023 10:44:00 AM	
386863	agg-aslt-shoot	aggravated-assault	10/3/2023 5:30:00 AM	
386864	agg-aslt-shoot	aggravated-assault	10/2/2023 5:47:00 PM	

	last_occurrence_date	reported_date	incident_address	\
0	NaN	2/10/2022 3:16:00 AM	1107 N SANTA FE DR	
1	NaN	7/8/2021 12:55:00 AM	815 16TH ST	
2	NaN	10/29/2020 4:31:00 AM	4745 N FEDERAL BLVD	
5	NaN	10/8/2020 3:43:00 AM	2100 W 30TH AVE	
6	NaN	4/18/2020 6:30:00 AM	1272 N PENNSYLVANIA ST	
…	…	…	…	
386859	NaN	10/2/2023 7:10:00 PM	2500 BLK N GILPIN ST	
386861	NaN	10/3/2023 6:34:00 PM	517 E BAYAUD AVE	

```
386862                NaN  10/3/2023 12:03:00 PM  N JOLIET ST / E 55TH AVE
386863                NaN   10/3/2023 9:47:00 AM           585 S ALTON WAY
386864                NaN   10/2/2023 6:16:00 PM      1198 N SHERIDAN BLVD

            geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0       3140929.0  1692612.0 -104.998910  39.733957           1          123
1       3142470.0  1697098.0 -104.993342  39.746248           6          611
2       3133352.0  1710396.0 -105.025520  39.782888           1          111
5       3137149.0  1701888.0 -105.012173  39.759477           1          113
6       3146039.0  1693575.0 -104.980722  39.736521           6          623
…             …          …          …         …          …            …
386859  3149905.0  1700074.0 -104.966840  39.754301           2          211
386861  3146000.0  1685678.0 -104.981021  39.714843           3          311
386862  3179460.0  1715407.0 -104.861343  39.795869           5          511
386863  3173405.0  1682706.0 -104.883680  39.706214           3          321
386864  3125724.0  1692953.0 -105.052966  39.735110           1          122

       neighborhood_id  is_crime  is_traffic  victim_count
0         lincoln-park         1           0             1
1                  cbd         1           0             1
2             berkeley         1           0             1
5             highland         1           0             1
6         capitol-hill         1           0             1
…                    …       …           …             …
386859        whittier         1           0             1
386861           speer         1           0             1
386862       montbello         1           0             1
386863         windsor         1           0             1
386864     west-colfax         1           0             1

[175556 rows x 20 columns]
```

[57]:
```
denver_crimes['first_occurrence_date']
```

[57]:
```
0          2/10/2022 2:50:00 AM
1           7/7/2021 9:02:00 PM
2         10/29/2020 1:30:00 AM
3           9/6/2018 5:00:00 PM
4           5/8/2020 5:00:00 AM
                  …
386860    10/3/2023 11:30:00 AM
386861    9/27/2023 11:00:00 AM
386862    10/3/2023 10:44:00 AM
386863     10/3/2023 5:30:00 AM
386864     10/2/2023 5:47:00 PM
Name: first_occurrence_date, Length: 386865, dtype: object
```

[59]:
```
denver_crimes['first_occurrence_date'] = pd.to_datetime(
    denver_crimes['first_occurrence_date'],
    format="%m/%d/%Y %I:%M:%S %p")
```

[64]:
```
denver_crimes.to_csv(
    "denver_crime/crime_v2.csv",
    index=False,
    sep=',')
```

[67]:
```
link = "https://gist.githubusercontent.com/mkulakowski2/4289746/raw/
 ↪671d60a704b57b6d591b02a7b2d48f7b8ec17e1a/crime.csv"
online_df = pd.read_csv(link)
online_df
```

[67]:

	Year	Population	Murder	Rape	Robbery	Assault	Burglary	CarTheft
0	1965	18073000	836	2320	28182	27464	183443	58452
1	1966	18258000	882	2439	30098	29142	196127	64368
2	1967	18336000	996	2665	40202	31261	219157	83775
3	1968	18113000	1185	2527	59857	34946	250918	104877
4	1969	18321000	1324	2902	64754	36890	248477	115400
5	1970	18190740	1444	2875	81149	39145	267474	125674
6	1971	18391000	1823	3225	97682	42318	273704	127658
7	1972	18366000	2026	4199	86391	45926	239886	105081
8	1973	18265000	2040	4852	80795	47781	246246	112328
9	1974	18111000	1919	5240	86814	51454	271824	104095
10	1975	18120000	1996	5099	93499	54593	301996	116274
11	1976	18084000	1969	4663	95718	54638	318919	133504
12	1977	17924000	1919	5272	84703	57193	309735	133669
13	1978	17748000	1820	5168	83785	58484	292956	119264
14	1979	17649000	2092	5394	93471	60949	308302	124343
15	1980	17506690	2228	5405	112273	60329	360925	133041
16	1981	17594000	2166	5479	120344	60189	350422	136849
17	1982	17659000	2013	5159	107843	59818	295245	137880
18	1983	17667000	1958	5296	94783	59452	249115	127861
19	1984	17735000	1786	5599	89900	64872	222956	115392
20	1985	17783000	1683	5706	89706	68270	219633	106537
21	1986	17772000	1907	5415	91360	76528	217010	113247
22	1987	17825000	2016	5537	89721	82417	216826	125329
23	1988	17898000	2244	5479	97434	91239	218060	153898
24	1989	17950000	2246	5242	103983	91571	211130	171007
25	1990	17990455	2605	5368	112380	92105	208813	187591
26	1991	18058000	2571	5085	112342	90186	204499	181287
27	1992	18119000	2397	5152	108154	87608	193548	168922
28	1993	18197000	2420	5008	102122	85802	181709	151949
29	1994	18169000	2016	4700	86617	82100	164650	128873
30	1995	18136000	1550	4290	72492	74351	146562	102596
31	1996	18185000	1353	4174	61822	64857	129828	89900

32	1997	18137000	1093	4075	56094	63628	118306	79697
33	1998	18175000	924	3843	49125	62023	104821	68171
34	1999	18196601	903	3563	43821	58860	93217	58261
35	2000	18976457	952	3530	40539	60090	87946	54231
36	2001	19084350	960	3546	36555	56961	80400	48287
37	2002	19134293	909	3885	36653	53583	76700	47366
38	2003	19212425	934	3775	35790	48987	75453	45204
39	2004	19280727	889	3608	33506	46911	70696	41002
40	2005	19315721	874	3636	35179	46150	68034	35736
41	2006	19306183	921	3169	34489	45387	68565	32134
42	2007	19297729	801	2926	31094	45094	64857	28030
43	2008	19467789	836	2799	31789	42122	65537	25096
44	2009	19541453	781	2582	28141	43606	62769	21871
45	2010	19395206	868	2797	28630	44197	65839	20639
46	2011	19465197	774	2752	28396	45568	65397	19311

[66]:
```
pd.read_clipboard(sep=',')
```

[66]:

	Year	Population	Murder	Rape	Robbery	Assault	Burglary	CarTheft
0	1965	18073000	836	2320	28182	27464	183443	58452
1	1966	18258000	882	2439	30098	29142	196127	64368
2	1967	18336000	996	2665	40202	31261	219157	83775
3	1968	18113000	1185	2527	59857	34946	250918	104877
4	1969	18321000	1324	2902	64754	36890	248477	115400
5	1970	18190740	1444	2875	81149	39145	267474	125674
6	1971	18391000	1823	3225	97682	42318	273704	127658
7	1972	18366000	2026	4199	86391	45926	239886	105081
8	1973	18265000	2040	4852	80795	47781	246246	112328
9	1974	18111000	1919	5240	86814	51454	271824	104095
10	1975	18120000	1996	5099	93499	54593	301996	116274
11	1976	18084000	1969	4663	95718	54638	318919	133504
12	1977	17924000	1919	5272	84703	57193	309735	133669
13	1978	17748000	1820	5168	83785	58484	292956	119264
14	1979	17649000	2092	5394	93471	60949	308302	124343
15	1980	17506690	2228	5405	112273	60329	360925	133041
16	1981	17594000	2166	5479	120344	60189	350422	136849
17	1982	17659000	2013	5159	107843	59818	295245	137880
18	1983	17667000	1958	5296	94783	59452	249115	127861
19	1984	17735000	1786	5599	89900	64872	222956	115392
20	1985	17783000	1683	5706	89706	68270	219633	106537
21	1986	17772000	1907	5415	91360	76528	217010	113247
22	1987	17825000	2016	5537	89721	82417	216826	125329
23	1988	17898000	2244	5479	97434	91239	218060	153898
24	1989	17950000	2246	5242	103983	91571	211130	171007
25	1990	17990455	2605	5368	112380	92105	208813	187591
26	1991	18058000	2571	5085	112342	90186	204499	181287
27	1992	18119000	2397	5152	108154	87608	193548	168922

```
28  1993    18197000    2420  5008   102122    85802    181709    151949
29  1994    18169000    2016  4700    86617    82100    164650    128873
30  1995    18136000    1550  4290    72492    74351    146562    102596
31  1996    18185000    1353  4174    61822    64857    129828     89900
32  1997    18137000    1093  4075    56094    63628    118306     79697
33  1998    18175000     924  3843    49125    62023    104821     68171
34  1999    18196601     903  3563    43821    58860     93217     58261
35  2000    18976457     952  3530    40539    60090     87946     54231
36  2001    19084350     960  3546    36555    56961     80400     48287
37  2002    19134293     909  3885    36653    53583     76700     47366
38  2003    19212425     934  3775    35790    48987     75453     45204
39  2004    19280727     889  3608    33506    46911     70696     41002
40  2005    19315721     874  3636    35179    46150     68034     35736
41  2006    19306183     921  3169    34489    45387     68565     32134
42  2007    19297729     801  2926    31094    45094     64857     28030
43  2008    19467789     836  2799    31789    42122     65537     25096
44  2009    19541453     781  2582    28141    43606     62769     21871
45  2010    19395206     868  2797    28630    44197     65839     20639
46  2011    19465197     774  2752    28396    45568     65397     19311
```

[ ]:

[60]:
```
denver_crimes['offense_category_id'].apply(lambda x: x.replace("-", " ").
↪capitalize())
```

[60]:
```
0            Public disorder
1            Public disorder
2            Public disorder
3            Public disorder
4            Public disorder
                ...
386860           Auto theft
386861    White collar crime
386862    Aggravated assault
386863    Aggravated assault
386864    Aggravated assault
Name: offense_category_id, Length: 386865, dtype: object
```

[69]:
```
denver_crimes.head()
```

[69]:
```
   incident_id        offense_id  offense_code  offense_code_extension  \
0    202268791   202268791299900          2999                       0
1   2021387586  2021387586299900          2999                       0
2   2020641486  2020641486299900          2999                       0
3   2018612468  2018612468299900          2999                       0
4   2020293614  2020293614299900          2999                       0
```

```
          offense_type_id offense_category_id  first_occurrence_date  \
0  criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
1  criminal-mischief-other     public-disorder    7/7/2021 9:02:00 PM
2  criminal-mischief-other     public-disorder  10/29/2020 1:30:00 AM
3  criminal-mischief-other     public-disorder    9/6/2018 5:00:00 PM
4  criminal-mischief-other     public-disorder    5/8/2020 5:00:00 AM

   last_occurrence_date          reported_date     incident_address  \
0                   NaN   2/10/2022 3:16:00 AM   1107 N SANTA FE DR
1                   NaN   7/8/2021 12:55:00 AM          815 16TH ST
2                   NaN  10/29/2020 4:31:00 AM  4745 N FEDERAL BLVD
3  9/6/2018 11:00:00 PM    9/7/2018 9:58:00 AM    65 S FEDERAL BLVD
4   5/8/2020 6:30:00 PM  5/13/2020 10:00:00 AM   12295 E ALBROOK DR

       geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0  3140929.0  1692612.0 -104.998910  39.733957           1          123
1  3142470.0  1697098.0 -104.993342  39.746248           6          611
2  3133352.0  1710396.0 -105.025520  39.782888           1          111
3  3133534.0  1685797.0 -105.025330  39.715357           4          411
4  3184065.0  1710782.0 -104.845074  39.783082           5          521

  neighborhood_id  is_crime  is_traffic  victim_count
0    lincoln-park         1           0             1
1             cbd         1           0             1
2        berkeley         1           0             1
3          barnum         1           0             1
4       montbello         1           0             1
```

[83]:
```
ct = pd.crosstab(
    denver_crimes['neighborhood_id'],
    denver_crimes['offense_category_id'])
#
# ct["Total"] = ct.sum(axis=1)
# ct = ct.sort_values(by="Total", ascending=False)
#
# ct.style.background_gradient(cmap='Greens')
```

[87]:
```
filtered_df = denver_crimes.loc[denver_crimes['neighborhood_id'] ==␣
 ↪'lincoln-park', :]
filtered_df
```

[87]:
```
        incident_id         offense_id  offense_code  offense_code_extension  \
0         202268791    202268791299900          2999                       0
12       2021415767   2021415767299900          2999                       0
21      20206010014  20206010014299900          2999                       0
98       2021334418   2021334418299900          2999                       0
192      2022106657   2022106657299900          2999                       0
```

```
…             …                   …          …                   …
386576   2023531958   2023531958240400        2404                    0
386583   2023527452   2023527452240400        2404                    0
386603   2023531526   2023531526240400        2404                    0
386658   2023533998   2023533998359900        3599                    0
386824   2023533150   2023533150240400        2404                    0

                offense_type_id offense_category_id  first_occurrence_date  \
0       criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
12      criminal-mischief-other     public-disorder   7/21/2021 4:41:00 PM
21      criminal-mischief-other     public-disorder    8/1/2020 1:30:00 PM
98      criminal-mischief-other     public-disorder   6/13/2021 2:30:00 AM
192     criminal-mischief-other     public-disorder    3/2/2022 7:00:00 PM
…                             …                   …                      …
386576   theft-of-motor-vehicle          auto-theft  9/28/2023 12:00:00 PM
386583   theft-of-motor-vehicle          auto-theft  9/29/2023 11:00:00 PM
386603   theft-of-motor-vehicle          auto-theft   9/29/2023 9:00:00 PM
386658      drug-pcs-other-drug        drug-alcohol   10/3/2023 3:13:00 PM
386824   theft-of-motor-vehicle          auto-theft  10/2/2023 10:00:00 PM

        last_occurrence_date         reported_date          incident_address  \
0                        NaN  2/10/2022 3:16:00 AM        1107 N SANTA FE DR
12                       NaN  7/21/2021 4:41:00 PM     5TH ST / W COLFAX AVE
21       8/1/2020 2:00:00 PM  8/5/2020 11:56:00 AM           1346 N OSAGE ST
98                       NaN  6/13/2021 2:33:00 AM         1331 N SPEER BLVD
192      3/3/2022 8:46:00 AM   3/3/2022 8:46:00 AM           1040 W 14TH AVE
…                          …                     …                         …
386576  10/2/2023 2:00:00 AM  10/2/2023 3:03:00 PM        1150 N GALAPAGO ST
386583  9/30/2023 6:02:00 AM  9/30/2023 6:02:00 AM        1200 N GALAPAGO ST
386603  9/30/2023 2:00:00 AM  10/2/2023 4:30:00 PM          777 N BANNOCK ST
386658                   NaN  10/3/2023 7:27:00 PM  1400 BLOCK N MARIPOSA ST
386824  10/3/2023 6:30:00 AM  10/3/2023 7:41:00 AM    1000 BLK N KALAMATH ST

            geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0       3140929.0  1692612.0 -104.998910  39.733957           1          123
12      3138188.0  1694755.0 -105.008615  39.739881           1          123
21      3139376.0  1694039.0 -105.004404  39.737897           1          123
98      3141191.0  1693918.0 -104.997953  39.737538           1          123
192     3140295.0  1694196.0 -105.001133  39.738315           1          123
…               …          …           …         …          …
386576  3141864.0  1692904.0 -104.995580  39.734744           1          123
386583  3141672.0  1693222.0 -104.996256  39.735620           1          123
386603  3143312.0  1690483.0 -104.990480  39.728076           1          123
386658  3139899.0  1694303.0 -105.002539  39.738614           1          123
386824  3140705.0  1692201.0 -104.999715  39.732832           1          123

       neighborhood_id  is_crime  is_traffic  victim_count
```

```
0         lincoln-park         1           0             1
12        lincoln-park         1           0             1
21        lincoln-park         1           0             1
98        lincoln-park         1           0             1
192       lincoln-park         1           0             1
…                  …       …         …           …
386576    lincoln-park         1           0             1
386583    lincoln-park         1           0             1
386603    lincoln-park         1           0             1
386658    lincoln-park         1           0             1
386824    lincoln-park         1           0             1

[8656 rows x 20 columns]
```

[89]:
```
filtered_df\
    .query("neighborhood_id == 'lincoln-park'")
```

[89]:
```
         incident_id         offense_id  offense_code  offense_code_extension  \
0          202268791    202268791299900          2999                       0
12        2021415767   2021415767299900          2999                       0
21       20206010014  20206010014299900          2999                       0
98        2021334418   2021334418299900          2999                       0
192       2022106657   2022106657299900          2999                       0
…               …                …             …                       …
386576   2023531958   2023531958240400          2404                       0
386583   2023527452   2023527452240400          2404                       0
386603   2023531526   2023531526240400          2404                       0
386658   2023533998   2023533998359900          3599                       0
386824   2023533150   2023533150240400          2404                       0

                offense_type_id offense_category_id  first_occurrence_date  \
0       criminal-mischief-other     public-disorder   2/10/2022 2:50:00 AM
12      criminal-mischief-other     public-disorder   7/21/2021 4:41:00 PM
21      criminal-mischief-other     public-disorder    8/1/2020 1:30:00 PM
98      criminal-mischief-other     public-disorder   6/13/2021 2:30:00 AM
192     criminal-mischief-other     public-disorder    3/2/2022 7:00:00 PM
…                           …                   …                   …
386576   theft-of-motor-vehicle          auto-theft  9/28/2023 12:00:00 PM
386583   theft-of-motor-vehicle          auto-theft  9/29/2023 11:00:00 PM
386603   theft-of-motor-vehicle          auto-theft   9/29/2023 9:00:00 PM
386658      drug-pcs-other-drug        drug-alcohol   10/3/2023 3:13:00 PM
386824   theft-of-motor-vehicle          auto-theft  10/2/2023 10:00:00 PM

        last_occurrence_date         reported_date          incident_address  \
0                        NaN  2/10/2022 3:16:00 AM        1107 N SANTA FE DR
12                       NaN  7/21/2021 4:41:00 PM     5TH ST / W COLFAX AVE
21       8/1/2020 2:00:00 PM  8/5/2020 11:56:00 AM           1346 N OSAGE ST
```

```
98                       NaN  6/13/2021 2:33:00 AM         1331 N SPEER BLVD
192      3/3/2022 8:46:00 AM   3/3/2022 8:46:00 AM           1040 W 14TH AVE
…                        …                   …                         …
386576  10/2/2023 2:00:00 AM  10/2/2023 3:03:00 PM        1150 N GALAPAGO ST
386583  9/30/2023 6:02:00 AM  9/30/2023 6:02:00 AM        1200 N GALAPAGO ST
386603  9/30/2023 2:00:00 AM  10/2/2023 4:30:00 PM          777 N BANNOCK ST
386658                   NaN  10/3/2023 7:27:00 PM  1400 BLOCK N MARIPOSA ST
386824  10/3/2023 6:30:00 AM  10/3/2023 7:41:00 AM    1000 BLK N KALAMATH ST

            geo_x      geo_y     geo_lon    geo_lat district_id  precinct_id  \
0       3140929.0  1692612.0 -104.998910  39.733957           1          123
12      3138188.0  1694755.0 -105.008615  39.739881           1          123
21      3139376.0  1694039.0 -105.004404  39.737897           1          123
98      3141191.0  1693918.0 -104.997953  39.737538           1          123
192     3140295.0  1694196.0 -105.001133  39.738315           1          123
…             …          …          …         …          …            …
386576  3141864.0  1692904.0 -104.995580  39.734744           1          123
386583  3141672.0  1693222.0 -104.996256  39.735620           1          123
386603  3143312.0  1690483.0 -104.990480  39.728076           1          123
386658  3139899.0  1694303.0 -105.002539  39.738614           1          123
386824  3140705.0  1692201.0 -104.999715  39.732832           1          123

       neighborhood_id  is_crime  is_traffic  victim_count
0         lincoln-park         1           0             1
12        lincoln-park         1           0             1
21        lincoln-park         1           0             1
98        lincoln-park         1           0             1
192       lincoln-park         1           0             1
…                  …         …           …             …
386576    lincoln-park         1           0             1
386583    lincoln-park         1           0             1
386603    lincoln-park         1           0             1
386658    lincoln-park         1           0             1
386824    lincoln-park         1           0             1

[8656 rows x 20 columns]
```

[96]:
```
url = "https://gist.githubusercontent.com/mkulakowski2/4289746/raw/
↪671d60a704b57b6d591b02a7b2d48f7b8ec17e1a/crime.csv"
```

1.5 Excel files

[68]:
```
df = pd.read_excel(
    "denver_crime/denver_crimes.xlsx",
    sheet_name="offense_code")
```

[69]:
```
df.to_excel('path_to_save', sheet_name='fasfas')
```

[69]:

	OBJECTID	OFFENSE_CODE	OFFENSE_CODE_EXTENSION	OFFENSE_TYPE_ID	OFFENSE_TYPE_NAME	OFFENSE_CATEGORY_ID	OFFENSE_CATEGORY_NAME	IS_CRIME	IS_TRAFFIC
0	1	2804	1	stolen-property-possession	Possession of stolen property	all-other-crimes	All Other Crimes	1	0
1	2	2804	2	fraud-possess-financial-device	Possession of a financial device	all-other-crimes	All Other Crimes	1	0
2	3	2901	0	damaged-prop-bus	Damaged business property	public-disorder	Public Disorder	1	0
3	4	2902	0	criminal-mischief-private	Criminal mischief to private property	public-disorder	Public Disorder	1	0
4	5	2903	0	criminal-mischief-public	Criminal mischief to public property	public-disorder	Public Disorder	1	0
..	…	…	…	…	…	…	…	…	…
295	296	2799	0	theft-embezzle	Embezzlement by an employee	white-collar-crime	White Collar Crime	1	0
296	297	2801	0	stolen-property-sale-of	Sale of stolen property	all-other-crimes	All Other Crimes	1	0
297	298	2803	0	stolen-property-buy-sell-rec	Buy, sell or receive stolen property	all-other-crimes	All Other Crimes	1	0
298	299	2804	0	outside-steal-recovered-veh	Recovered vehicle stolen outside Denver	all-other-crimes	All Other Crimes	1	0
299	300	1315	6	agg-aslt-shoot	Assault causing serious bodily injury involvin…	aggravated-assault	Public Disorder	1	0

```
296      All Other Crimes         1           0
297      All Other Crimes         1           0
298      All Other Crimes         1           0
299    Aggravated Assault         1           0

[300 rows x 9 columns]
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

1.6 JSON files

A **JSON (JavaScript Object Notation)** file is a lightweight, text-based format for storing and exchanging structured data using key–value pairs and lists. It's widely used for APIs and data sharing because it's easy for both humans to read and programs (including Python) to parse.

https://corgis-edu.github.io/corgis/json/state_crime/

[70]:
```
import json
```

[71]:
```
with open('state_crime.json', mode='r') as file:
    state_crime = json.load(file)
```

[78]:
```
state_crime[0]['Data']['Population']
```

[78]:
```
3266740
```

[80]:
```
transformed_df = pd.json_normalize(state_crime)
transformed_df
```

[80]:
```
         State  Year  Data.Population  Data.Rates.Property.All  \
0      Alabama  1960          3266740                   1035.4
1      Alabama  1961          3302000                    985.5
2      Alabama  1962          3358000                   1067.0
3      Alabama  1963          3347000                   1150.9
4      Alabama  1964          3407000                   1358.7
…        …   …              …                        …
3110   Wyoming  2015           586107                   1902.6
3111   Wyoming  2016           585501                   1957.3
```

```
3112  Wyoming  2017           579315                    1830.4
3113  Wyoming  2018           577737                    1785.1
3114  Wyoming  2019           578759                    1571.1

      Data.Rates.Property.Burglary  Data.Rates.Property.Larceny  \
0                            355.9                        592.1
1                            339.3                        569.4
2                            349.1                        634.5
3                            376.9                        683.4
4                            466.6                        784.1
…                            …                            …
3110                         300.6                       1500.9
3111                         302.5                       1518.2
3112                         275.0                       1421.0
3113                         264.0                       1375.9
3114                         241.2                       1206.7

      Data.Rates.Property.Motor  Data.Rates.Violent.All  \
0                          87.3                   186.6
1                          76.8                   168.5
2                          83.4                   157.3
3                          90.6                   182.7
4                         108.0                   213.1
…                         …                       …
3110                      101.0                   222.1
3111                      136.6                   244.2
3112                      134.5                   237.5
3113                      145.2                   212.2
3114                      123.2                   217.4

      Data.Rates.Violent.Assault  Data.Rates.Violent.Murder  …  \
0                          138.1                      12.4  …
1                          128.9                      12.9  …
2                          119.0                       9.4  …
3                          142.1                      10.2  …
4                          163.0                       9.3  …
…                          …                          …  …
3110                       179.8                       2.7  …
3111                       195.7                       3.4  …
3112                       176.4                       2.6  …
3113                       150.6                       2.3  …
3114                       147.6                       2.2  …

      Data.Rates.Violent.Robbery  Data.Totals.Property.All  \
0                           27.5                     33823
1                           19.1                     32541
2                           22.5                     35829
```

```
3                          24.7                      38521
4                          29.1                      46290
…                          …                         …
3110                       10.1                      11151
3111                       10.1                      11460
3112                       13.1                      10604
3113                       17.3                      10313
3114                       11.6                       9093

      Data.Totals.Property.Burglary  Data.Totals.Property.Larceny  \
0                             11626                         19344
1                             11205                         18801
2                             11722                         21306
3                             12614                         22874
4                             15898                         26713
…                             …                             …
3110                           1762                          8797
3111                           1771                          8889
3112                           1593                          8232
3113                           1525                          7949
3114                           1396                          6984

      Data.Totals.Property.Motor  Data.Totals.Violent.All  \
0                           2853                     6097
1                           2535                     5564
2                           2801                     5283
3                           3033                     6115
4                           3679                     7260
…                           …                        …
3110                         592                     1302
3111                         800                     1430
3112                         779                     1376
3113                         839                     1226
3114                         713                     1258

      Data.Totals.Violent.Assault  Data.Totals.Violent.Murder  \
0                            4512                        406
1                            4255                        427
2                            3995                        316
3                            4755                        340
4                            5555                        316
…                            …                           …
3110                         1054                         16
3111                         1146                         20
3112                         1022                         15
3113                          870                         13
3114                          854                         13
```

```
      Data.Totals.Violent.Rape  Data.Totals.Violent.Robbery
0                          281                          898
1                          252                          630
2                          218                          754
3                          192                          828
4                          397                          992
…                          …                            …
3110                       173                           59
3111                       205                           59
3112                       263                           76
3113                       243                          100
3114                       324                           67

[3115 rows x 21 columns]
```

[81]:
```
my_dict = {'webinar': 'IACA_Python_IO'}
my_dict
```

[81]:
```
{'webinar': 'IACA_Python_IO'}
```

[83]:
```
with open('IACA.json', mode='w') as file:
    json.dump(my_dict, fp=file)
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

1.7 API requests

> An **API (Application Programming Interface)** is a set of rules that allows different software applications to communicate and work with each other.

APIs commonly return data in formats like JSON, which can be easily used in Python.

In simple terms, an API is like a **contract between two pieces of software**: if one program sends a request in a specific format, the other program responds with data or functionality. Think of it this way: a graphical user interface (GUI) or command line interface (CLI) lets *humans* interact with software, while an API lets *software interact with other software*.

1.7.1 Basic elements of an API:

- **Access**: determines who is allowed to request data or services (e.g., via API keys or authentication)
- **Request**: the query sent to the API. It typically includes:
  - **Methods**: the type of action you want to perform (e.g., retrieve data, send data)
  - **Parameters**: additional details that refine the request (e.g., date range, location)
- **Response**: the data or result returned by the API

1.7.2 Web APIs

A **web API** allows communication over the internet between a client (like your Python script) and a server. These are widely used in modern applications.

They can operate: - on the **server side** (providing data/services) - on the **client side** (requesting and using them)

Many organizations provide web APIs to access their data or services.

Source: https://www.wallarm.com/what/what-is-an-api-call

Toronto Police Service Break and Enter https://data.tps.ca/datasets/TorontoPS::break-and-enter-open-data/about

[84]:
```
import requests
```

[88]:
```
url = "https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/services/
 ↪Break_and_Enter_Open_Data/FeatureServer/0/query?
 ↪where=1%3D1&outFields=*&outSR=4326&f=json"
response = requests.get(url)
```

[7]:
```
# import geopandas as gpd
```

[86]:
```
response
```

[86]:
```
<Response [200]>
```

- 2xx → success
  - 200 OK → request successful
- 3xx → redirect
  - 301 Moved Permanently → the resource has been moved to a different URL
- 4xx → your fault
  - 400 Bad Request → the server cannot process the request due to a client error (e.g., malformed request syntax, invalid request message framing, or deceptive request routing)
  - 401 Unauthorized → the request requires user authentication
  - 403 Forbidden → the server understood the request but refuses to authorize it
  - 404 Not Found → the server has not found anything matching the request URI

- 5xx → server's fault
  - 500 Internal Server Error → the server encountered an unexpected condition that prevented it from fulfilling the request
  - 502 Bad Gateway → the server, while acting as a gateway or proxy, received an invalid response from the upstream server it accessed in attempting to fulfill the request
  - 503 Service Unavailable → the server is currently unable to handle the request due to temporary overloading or maintenance of the server

[ ]:

[106]:
```
data = response.json()
# data
```

[95]:
```
import geopandas as gpd
```

[107]:
```
break_df = gpd.GeoDataFrame.from_features(data['features'])

break_df.head()
```

[107]:

```
                 geometry  OBJECTID EVENT_UNIQUE_ID    REPORT_DATE  \
0  POINT (-79.413 43.651)         1  GO-20141260521  1388552400000
1  POINT (-79.422 43.792)         2  GO-20141261478  1388552400000
2   POINT (-79.438 43.64)         3  GO-20141261592  1388552400000
3  POINT (-79.309 43.709)         4  GO-20141261949  1388552400000
4  POINT (-79.383 43.647)         5  GO-20141262372  1388552400000

        OCC_DATE REPORT_YEAR REPORT_MONTH REPORT_DAY REPORT_DOY  REPORT_DOW  \
0  1388552400000        2014      January          1          1  Wednesday
1  1388552400000        2014      January          1          1  Wednesday
2  1388552400000        2014      January          1          1  Wednesday
3  1388552400000        2014      January          1          1  Wednesday
4  1388552400000        2014      January          1          1  Wednesday

   … UCR_CODE UCR_EXT       OFFENCE     CSI_CATEGORY HOOD_158  \
0  …     2120     200           B&E  Break and Enter      081
1  …     2120     200           B&E  Break and Enter      036
2  …     2120     220  B&E W'Intent  Break and Enter      085
3  …     2120     220  B&E W'Intent  Break and Enter      054
4  …     2120     200           B&E  Break and Enter      170

          NEIGHBOURHOOD_158 HOOD_140         NEIGHBOURHOOD_140 LONG_WGS84  \
0    Trinity-Bellwoods (81)      081    Trinity-Bellwoods (81) -79.412798
1     Newtonbrook West (36)      036     Newtonbrook West (36) -79.421663
2       South Parkdale (85)      085       South Parkdale (85) -79.437897
3    O'Connor-Parkview (54)      054    O'Connor-Parkview (54) -79.309002
4  Yonge-Bay Corridor (170)      076  Bay Street Corridor (76) -79.383017

   LAT_WGS84
```

```
0  43.650928
1  43.792329
2  43.640341
3  43.708708
4  43.646603

[5 rows x 30 columns]
```

[ ]:
```
pd.to_datetime(break_df['REPORT_YEAR'])
```

[99]:
```
params = {
    'outFields': '*',
    'where': '1=1', # True
    'f': 'geojson'
}
```

[100]:
```
url = f"https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/services/
 ↪Break_and_Enter_Open_Data/FeatureServer/0/query"
```

[101]:
```
response = requests.get(url, params=params)
response.url
```

[101]:
```
'https://services.arcgis.com/S9th0jAJ7bqgIRjw/arcgis/rest/services/Break_and_Ent
er_Open_Data/FeatureServer/0/query?outFields=%2A&where=1%3D1&f=geojson'
```

[108]:
```
break_df = break_df \
    .set_crs(epsg=4326) \
    .to_crs(epsg=3857)

# As the columns LONG_WGS84 and LAT_WGS84 suggest, the location coordinates are␣
 ↪based on the WGS 84 geographic coordinate system (EPSG:4326). We will␣
 ↪convert those geographic coordinates into the Web Mercator projection (EPSG:
 ↪3857). This projected coordinate system represents the Earth on a flat,␣
 ↪two-dimensional plane using meters as units, the same system used by most␣
 ↪online mapping services (like OpenStreetMap, Google Maps, or CartoDB).
```

[ ]:
```
# break_df.loc[brea.ina(),:]
```

[113]:
```
break_df_v2 = break_df.query(
    "(LONG_WGS84 != 0) | (LAT_WGS84 != 0)"
)
```

[114]:
```
break_df_v2.explore()
```

[114]:
```
<folium.folium.Map at 0x71956712ee40>
```

1.8 Pickle

A pickle file is a binary file used in Python to store and retrieve objects, such as lists, dictionaries, or trained models. It allows you to save complex Python data structures and load them later, but it's not secure or portable outside Python.

You should never load a pickle file from an untrusted source, as it can execute arbitrary code during unpickling. Only load pickle files from sources you trust.

[115]:
```
l = [1, 2, 2]
```

[117]:
```
import pickle
```

[118]:
```
with open('my_data.pkl', 'wb') as f:
    pickle.dump(l, f)
```

[119]:
```
with open('my_data.pkl', 'rb') as f:
    loaded_data = pickle.load(f)
```

[120]:
```
loaded_data
```

[120]:
```
[1, 2, 2]
```

Thank you!

Ruslan Klymentiev

PhD Candidate

UGent Crime Lab

e-mail: ruslan.klymentiev@ugent.be

web: https://rklymentiev.com

[ ]:
```

```