

# Data Types for Data Science in Python Project

## Counting Crime (Part 1)

### Data Source:

Chicago Data Portal – Crimes – 2001 to Present(Updated Jan 4, 2022)

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>

Downloaded and saved data as “crime.csv”

**Data Set Overview:** The data file contains

ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description, Arrest, Domestic, Beat, District, Ward, Community Area, FBI Code, X Coordinate, Y Coordinate, Year, Updated On, Latitude, Longitude, Location

**Purposes:** To find an answer for the below questions:

- Question #1: What are the top 3 months with the highest number of crimes?
- Question #2: Find the most common crimes by location type by month in 2021
- Question #3: Determine the Arrest by District by Year

We don’t need to use the whole dataset to answer those questions. The first step before performing coding, we will read the questions carefully to determine which columns we’re going to need for our analysis.

- Question 1 wants us to find the top 3 months with the highest number of crimes so we only need column “Date” and “Primary Type”
- For question 2, we will need the “Location Description” column
- “Arrest” and “District” columns will be needed for question 3

In part 1, we will work on Question #1 and Question #2 first. So we will pick the following columns for our project:

- Date
- Primary Type
- Location Description
- Arrest

Also, we will need to anticipate which module we will import for our analysis

- The dataset is .csv so we will have to import csv module
- We also work on date so import datetime module is necessary
- Moreover, Counter and defaultdict from collections will be needed for our analysis

We can either import all modules that we're going to use in the beginning or we can import each module based on each step of our analysis.

### #1: Find the top 3 months with the highest number of crimes

```
# Import the csv module
import csv

# Create the file object: csvfile
csvfile = open('crime.csv', 'r')

# Create an empty list: crime_data
crime_data = []

# Loop over the csv reader on the file object
for row in csv.reader(csvfile):
    # Append the date, type of crime, location description and arrest
    crime_data.append((row[2], row[5], row[7], row[8]))

# Remove the first element from crime_data
crime_data.pop(0)

# Print the first 10 records
print(crime_data[:10])
```

<script.py> output:

```
[('09/05/2015 01:30:00 PM', 'BATTERY', 'RESIDENCE', 'false'), ('09/04/2015
11:30:00 AM', 'THEFT', 'CTA BUS', 'false'), ('09/01/2018 12:01:00 AM', 'THEFT',
'RESIDENCE', 'false'), ('09/05/2015 12:45:00 PM', 'NARCOTICS', 'SIDEWALK',
'true'), ('09/05/2015 01:00:00 PM', 'ASSAULT', 'APARTMENT', 'false'),
('09/05/2015 10:55:00 AM', 'BURGLARY', 'RESIDENCE', 'false'), ('09/04/2015
06:00:00 PM', 'BURGLARY', 'RESIDENCE-GARAGE', 'false'), ('09/05/2015 01:00:00
PM', 'THEFT', 'GROCERY FOOD STORE', 'true'), ('09/05/2015 11:30:00 AM',
'ROBBERY', 'STREET', 'false'), ('05/01/2016 12:25:00 AM', 'DECEPTIVE PRACTICE',
'', 'false')]
```

We're going to find how many crimes committed each month

```
# Import necessary modules
from collections import Counter
from datetime import datetime

# Create a Counter Object: crimes_by_month
crimes_by_month = Counter()

# Loop over the crime_data list
for row in crime_data:
    # Convert the first element of each item into a Python Datetime Object: date
    date = datetime.strptime(row[0], '%m/%d/%Y %H:%M:%S %p')
    # Increment the counter for the month of the row by one
    crimes_by_month[date.month] += 1

# Print the 3 most common months for crime
print(crimes_by_month.most_common(3))

<script.py> output:

[(7, 694471), (8, 687627), (5, 662317)]
```

We can see that in Chicago, the top three months that have highest crimes are May, July, Aug:

- July: 694,471 cases
- Aug: 687,627 cases
- May: 662,317 cases

## #2: Find the most common crimes by location type by month in 2021

We will convert crime\_data list into a dictionary keyed by month with a list of location values for each month and filtered to the record for the year of 2021

```
# Import necessary modules
from collections import defaultdict
from datetime import datetime

# Create a dictionary that defaults to a list: locations_by_month
```

```

locations_by_month = defaultdict(list)

# Loop over the crime_data:
for row in crime_data:
    # Convert the first element to a date object
    date = datetime.strptime(row[0], '%m/%d/%Y %H:%M:%S %p')
    # If the year is 2021
    if date.year == 2021:
        # Set the dictionary key to the month and append the location
        locations_by_month[date.month].append(row[2])

# Print the dictionary
print(locations_by_month)

<script.py> output:

defaultdict(<class 'list'>, {1: ['APARTMENT', 'RESIDENCE', 'APARTMENT', 'STREET',
'STREET', 'CHA APARTMENT', 'APARTMENT', 'STREET', 'RESIDENCE - GARAGE', 'STREET',
'APARTMENT', 'SMALL RETAIL STORE', 'RESIDENCE - GARAGE', 'STREET', 'BARBERSHOP',
'APARTMENT', 'RESIDENCE', 'CAR WASH', ...]}

```

Now we will determine common crimes by month and location type

```

# Import Counter from collections
from collections import Counter

# Loop over the items from locations_by_month using tuple expansion of the month
for month, locations in locations_by_month.items():
    # Make a Counter of the locations
    location_count = Counter(locations)
    # Print the month
    print(month)
    # Print the most common location
    print(location_count.most_common(5))

<script.py> output:

1
[('APARTMENT', 3727), ('STREET', 3677), ('RESIDENCE', 3478), ('SIDEWALK', 663),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 388)]
2

```

```

[('STREET', 3154), ('APARTMENT', 3138), ('RESIDENCE', 2401), ('SIDEWALK', 512),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 328)]
3
[('STREET', 3919), ('APARTMENT', 3594), ('RESIDENCE', 2760), ('SIDEWALK', 860),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 496)]
4
[('STREET', 3874), ('APARTMENT', 3561), ('RESIDENCE', 2403), ('SIDEWALK', 805),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 452)]
5
[('STREET', 4509), ('APARTMENT', 3830), ('RESIDENCE', 2738), ('SIDEWALK', 1115),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 520)]
6
[('STREET', 4834), ('APARTMENT', 3978), ('RESIDENCE', 2870), ('SIDEWALK', 1192),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 561)]
7
[('STREET', 4845), ('APARTMENT', 3795), ('RESIDENCE', 2759), ('SIDEWALK', 1226),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 598)]
8
[('STREET', 4641), ('APARTMENT', 3823), ('RESIDENCE', 2489), ('SIDEWALK', 1246),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 608)]
9
[('STREET', 4981), ('APARTMENT', 3704), ('RESIDENCE', 2570), ('SIDEWALK', 1269),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 625)]
10
[('STREET', 4967), ('APARTMENT', 3667), ('RESIDENCE', 2466), ('SIDEWALK', 1208),
('PARKING LOT / GARAGE (NON RESIDENTIAL)', 695)]
11
[('STREET', 4377), ('APARTMENT', 3411), ('RESIDENCE', 2190), ('SIDEWALK', 873),
('SMALL RETAIL STORE', 614)]
12
[('STREET', 3532), ('APARTMENT', 3025), ('RESIDENCE', 1957), ('SIDEWALK', 718),
('SMALL RETAIL STORE', 502)]

```

We can see total counts of crimes based on the five common locations for each month from Jan to Dec.