DATA TYPES FOR DATA SCIENCE

Case Study - Counting Crimes

Jason Myers
Instructor

Data Set Overview

```
Date,Block,Primary Type,Description,
Location Description,Arrest,Domestic, District

05/23/2016 05:35:00 PM,024XX W DIVISION ST,ASSAULT,SIMPLE,
STREET,false,true,14

03/26/2016 08:20:00 PM,019XX W HOWARD ST,BURGLARY,FORCIBLE ENTRY,
SMALL RETAIL STORE,false,false,24
```

- Chicago Open Data Portal https://data.cityofchicago.org/

Part 1 - Step 1

- Read data from CSV

```
In [1]: import csv

In [2]: csvfile = open('ART_GALLERY.csv', 'r')

In [3]: for row in csv.reader(csvfile):
   ...:     print(row)
```

Part 1 - Step 2

- Create and use a Counter with a slight twist

```
In [1]: from collections import Counter

In [2]: nyc_eatery_count_by_types = Counter(nyc_eatery_types)
```

- Use date parts for Grouping like in Chapter 4

```
In [1]: daily_violations = defaultdict(int)

In [2]: for violation in parking_violations:
   ...:     violation_date = datetime.strptime(violation[4], '%m/%d/%Y')
   ...:     daily_violations[violation_date.day] += 1
```

Part 1 - Step 3

- Group data by Month
- The date components we learned about earlier.

```
In [1]: from collections import defaultdict

In [2]: eateries_by_park = defaultdict(list)

In [3]: for park_id, name in nyc_eateries_parks:
   ...:     eateries_by_park[park_id].append(name)
```

Part 1 - Final

- Find 5 most common locations for crime each month.

```
In [1]: print(nyc_eatery_count_by_types.most_common(3))
[('Mobile Food Truck', 114), ('Food Cart', 74), ('Snack Bar', 24)]
```

DATA TYPES FOR DATA SCIENCE

Let's practice!

DATA TYPES FOR DATA SCIENCE

Case Study - Crimes by District and Differences by Block

Jason Myers
Instructor

Part 2 - Step 1

- Read in the CSV data as a dictionary

```
In [1]: import csv

In [2]: csvfile = open('ART_GALLERY.csv', 'r')

In [3]: for row in csv.DictReader(csvfile):
   ...:     print(row)
```

- Pop out the key and store the remaining dict

```
In [1]: galleries_10310 = art_galleries.pop('10310')
```

Part 2 - Step 2

- Pythonically iterate over the Dictionary

```
In [1]: for zip_code, galleries in art_galleries.items():
   ...:     print(zip_code)
   ...:     print(galleries)
```

Wrapping Up

- Use sets for uniqueness

```
In [1]: cookies_eaten_today = ['chocolate chip', 'peanut butter',
   ...: 'chocolate chip', 'oatmeal cream', 'chocolate chip']

In [2]: types_of_cookies_eaten = set(cookies_eaten_today)

In [3]: print(types_of_cookies_eaten)
set(['chocolate chip', 'oatmeal cream', 'peanut butter'])
```

- difference() set method as at the end of Chapter 1

```
In [1]: cookies_jason_ate.difference(cookies_hugo_ate)
set(['oatmeal cream', 'peanut butter'])
```

DATA TYPES FOR DATA SCIENCE

Let's practice!

DATA TYPES FOR DATA SCIENCE

Final thoughts

Jason Myers
Instructor