Notes_Udacity_Intro_to_Data_Science

November 26, 2016

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

In [1]:
```
# Data Acquisition
# Accessing an API
# Scraping a web page
# Combine data from different formats
```

In [2]:
```
# CSV: Comma Separated Values
```

In [3]:
```
## Representing a CSV as a list of rows

# Option 1: Each row is a list
csv = [['A1', 'A2', 'A3'],
       ['B2', 'B2', 'B3']]
```

In [4]:
```
csv
```

Out[4]:
```
[['A1', 'A2', 'A3'], ['B2', 'B2', 'B3']]
```

In [5]:
```
# Option 2: Each row is a dictionary
# Works well if your csv has a header
# Keys can be column names, fields can be values

# Overall structure would be a list of dictionaries
```

In [6]:
```
'''
import unicodecsv

enrollments = [] => create list of enrollments
f - open('enrollments.csv', 'rb') => open the file => b flag changes doc
reader = unicodecsv.DictReader(f) => dict since has header row

for row in reader:
    enrollments.append(row) => iterator used for loop to access each elemer

    => you can only access iterator once

f.close() => close file
```

```
enrollments[0] => output
'''
```

Out[6]: "\nimport unicodecsv\n\nenrollments = []\nf - open('enrollments.csv', 'rb')

In [7]:
```
'''
More succinct version

import unicodecsv

enrollments = []
with open('enrollments.csv', 'rb') as f:
    reader = unicodecsv.DictReader(f)

    for row in reader:
        enrollments.append(row)

enrolmments[0]


'''
```

Out[7]: "\nMore succinct version\n\nimport unicodecsv\n\nenrollments = []\nwith ope

In [ ]:
```
'''
=> converting iterator to a list


import unicodecsv

with open('enrollments.csv', 'rb') as f:
    reader = unicodecsv.DictReader(f)
    enrollments = list(reader)

enrollments[0]

'''
```