0100_itemgetter_and_groupby

December 7, 2018

1 Sorting and grouping dictionary items with itemgetter and groupby

Often in data science we might use Pandas to store mixed text and numbers (Pandas then allows easy sorting and grouping), but sometimes you may want to stick to pure Python and use lists of dictionary items. Sorting and sub-grouping of lists of dictionary items may be performed in Python using itemgetter (to help sort) and groupby to sub-group.

1.1 Sorting lists of dictionary items with itemgetter

```python
In [1]: # Set up a list of dictionary items and add content

        people = []
        people.append({'born':1966, 'gender':'male', 'name':'Bob'})
        people.append({'born':1966, 'gender':'female', 'name':'Anne'})
        people.append({'born':1966, 'gender':'male', 'name':'Adam'})
        people.append({'born':1970, 'gender':'male', 'name':'John'})
        people.append({'born':1970, 'gender':'female', 'name':'Daisy'})
        people.append({'born': 1968, 'gender':'male', 'name':'Steve'})

        # import methodsfrom operator import itemgetter
        from operator import itemgetter

        # Sort by 'born' and 'name'
        people.sort(key=itemgetter('born','name'))

        # Print out sorted list
        for item in people:
            print(item)

{'born': 1966, 'gender': 'male', 'name': 'Adam'}
{'born': 1966, 'gender': 'female', 'name': 'Anne'}
{'born': 1966, 'gender': 'male', 'name': 'Bob'}
{'born': 1968, 'gender': 'male', 'name': 'Steve'}
{'born': 1970, 'gender': 'female', 'name': 'Daisy'}
{'born': 1970, 'gender': 'male', 'name': 'John'}
```

1.2 Grouping the sorted list of dictionary items with groupby

Note: We must always sort our list of dictionary items in the required order before sub-grouping them.

```
In [2]: # Set up a list of dictionary items and add content

        people = []
        people.append({'born':1966, 'gender':'male', 'name':'Bob'})
        people.append({'born':1966, 'gender':'female', 'name':'Anne'})
        people.append({'born':1966, 'gender':'male', 'name':'Adam'})
        people.append({'born':1970, 'gender':'male', 'name':'John'})
        people.append({'born':1970, 'gender':'female', 'name':'Daisy'})
        people.append({'born': 1968, 'gender':'male', 'name':'Steve'})

        # import methods

        from operator import itemgetter
        from itertools import groupby

        # First sort by required field
        # Groupby only finds groups that are collected consecutively
        people.sort(key=itemgetter('born'))

        # Now iterate through groups (here we will group by the year born)
        for born, items in groupby(people, key=itemgetter('born')):
                print (born)
                for i in items:
                        print(' ', i)

1966
  {'born': 1966, 'gender': 'male', 'name': 'Bob'}
  {'born': 1966, 'gender': 'female', 'name': 'Anne'}
  {'born': 1966, 'gender': 'male', 'name': 'Adam'}
1968
  {'born': 1968, 'gender': 'male', 'name': 'Steve'}
1970
  {'born': 1970, 'gender': 'male', 'name': 'John'}
  {'born': 1970, 'gender': 'female', 'name': 'Daisy'}
```