CompSci 190: Pivots, Joins & Review

Jeff Forbes
October 1, 2018
Challenge Question

Which NBA teams spent the most on their starters in 2016?

- Each team has one starter per position
- Assume the starter for a team & position is the player with the highest salary on that team in that position

How do you generate a table of the names of the starters for each team?

<table>
<thead>
<tr>
<th>PLAYER</th>
<th>POSITION</th>
<th>TEAM</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Millsap</td>
<td>PF</td>
<td>Atlanta Hawks</td>
<td>18.6717</td>
</tr>
<tr>
<td>Al Horford</td>
<td>C</td>
<td>Atlanta Hawks</td>
<td>12</td>
</tr>
<tr>
<td>Tiago Splitter</td>
<td>C</td>
<td>Atlanta Hawks</td>
<td>9.75625</td>
</tr>
</tbody>
</table>
Pivot

- Cross-classifies according to two categorical variables
- Produces a grid of counts or aggregated values
- Two required arguments:
  - First: variable that forms column labels of grid
  - Second: variable that forms row labels of grid
- Two optional arguments (include both or neither)
  - `values='column_label_to_aggregate'`
  - `collect=function_with_which_to_aggregate`

(Demo)
Joining Two Tables

Keep all rows in the table that have a match ...

... for the value in this column ...

... somewhere in this other table's ...

... column that contains matching values.

drinks.join('Cafe', discounts, 'Location')

<table>
<thead>
<tr>
<th>Drink</th>
<th>Cafe</th>
<th>Price</th>
<th>Coupon</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk Tea</td>
<td>Tea One</td>
<td>4</td>
<td>25%</td>
<td>Tea One</td>
</tr>
<tr>
<td>Espresso</td>
<td>Nefeli</td>
<td>2</td>
<td>50%</td>
<td>Nefeli</td>
</tr>
<tr>
<td>Latte</td>
<td>Nefeli</td>
<td>3</td>
<td>5%</td>
<td>Tea One</td>
</tr>
<tr>
<td>Espresso</td>
<td>Abe's</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The joined column is sorted automatically.
Test 1

• Topics
  ○ Causality
  ○ Python
    ■ Data (names, values & types)
    ■ Expressions (numbers, strings, arrays, & tables)
    ■ Functions
  ○ Visualization
    ■ Charts & Histograms

• Do the review questions
• Bring 2 sheets of notes