



Lecture 10

Joins and Table Examples

Slides created by John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)
Contributions by Fahad Kamran (fhdkmrn@berkeley.edu) and Vinitra Swamy (vinitra@berkeley.edu)

Announcements

Project 1: World Progress

Video: <https://www.gapminder.org/videos/dont-panic-the-facts-about-population/>

Pivot

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)

Joins

Joining Two Tables

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

Match rows in this table ...

... using values in this column ...

... with rows in that table ...

... using values in that column.

Columns from both tables

drinks

Drink	Cafe	Price
Milk Tea	Tea One	4
Espresso	Nefeli	2
Latte	Nefeli	3
Espresso	Abe's	2

discounts

Coupon	Location
25%	Tea One
50%	Nefeli
5%	Tea One

The joined column is sorted automatically

Cafe	Drink	Price	Coupon
Nefeli	Espresso	2	50%
Nefeli	Latte	3	50%
Tea One	Milk Tea	4	25%
Tea One	Milk Tea	4	5%

(Demo)

Bikes

(Demo)

Shortest Trips

(Demo)

Maps

(Demo)

Maps

A table containing columns of latitude and longitude values can be used to generate a map of markers

 .**map_table**(table, ...)

Either **Marker**
or **Circle**

Column 0: latitudes
Column 1: longitudes
Column 2: labels
Column 3: colors
Column 4: sizes

Applies to all
features:
color='blue'
size=200

Break

Combining Table Methods

Important Table Methods

`t.select(column, ...)` or `t.drop(column, ...)`

`t.take([row, ...])` or `t.exclude([row, ...])`

`t.sort(column, descending=False, distinct=False)`

`t.where(column, are.condition(...))`

`t.apply(function, column, ...)`

`t.group(column)` or `t.group(column, function)`

`t.group([column, ...])` or `t.group([column, ...], function)`

`t.pivot(cols, rows)` or `t.pivot(cols, rows, vals, function)`

`t.join(column, other_table, other_table_column)`

<http://data8.org/datascience/tables.html>

Discussion Question

Generate a table with one row per cafe that has the name and discounted price of its cheapest discounted drink

drinks

Drink	Cafe	Price
Milk Tea	Tea One	4
Espresso	Nefeli	2
Coffee	Nefeli	3
Espresso	Abe's	2

discounts

Coupon	Location
5%	Tea One
50%	Nefeli
25%	Tea One

cheapest

Cafe	Drink	Discounted Price
Nefeli	Espresso	1
Tea One	Milk Tea	3

(Demo)

Spring 2016 Midterm, Q2(b)

- (b) (8 pt) Each row of the `trip` table from lecture describes a single bicycle rental in the San Francisco area. Durations are integers representing times in seconds. The first three rows out of 338343 appear below.

Start	End	Duration
Ferry Building	SF Caltrain	765
San Antonio Shopping Center	Mountain View City Hall	1036
Post at Kearny	2nd at South Park	307

Write a Python expression below each of the following descriptions that computes its value. The first one is provided for you. You *may* use up to two lines and introduce variables.

- The average duration of a rental.

```
total_duration = sum(trip.column(2))
total_duration / trip.num_rows
```

- The name of the station where the most rentals ended (assume no ties).
- The number of stations for which the average duration ending at that station was more than 10 minutes.

(Demo)

Take-Home Question

Generate a table of the names of the starters for each team

Team	C	PF	PG	SF	SG
Atlanta Hawks	Miles Plumlee	Mike Muscala	Dennis Schroder	Taurean Prince	Kent Bazemore
Boston Celtics	Aron Baynes	Al Horford	Kyrie Irving	Gordon Hayward	Jaylen Brown
Brooklyn Nets	Timofey Mozgov	Luis Scola	Jeremy Lin	DeMarre Carroll	Allen Crabbe
Charlotte Hornets	Dwight Howard	Marvin Williams	Kemba Walker	Michael Kidd-Gilchrist	Nicolas Batum
Chicago Bulls	Robin Lopez	Nikola Mirotic	Kris Dunn	Denzel Valentine	Justin Holiday
Cleveland Cavaliers	Kevin Love	Ante Zizic	George Hill	LeBron James	JR Smith
Dallas Mavericks	Dwight Powell	Josh McRoberts	J.J. Barea	Harrison Barnes	Wesley Matthews
Denver Nuggets	Mason Plumlee	Paul Millsap	Devin Harris	Wilson Chandler	Mike Miller
Detroit Pistons	Andre Drummond	Blake Griffin	Reggie Jackson	Stanley Johnson	Langston Galloway
Golden State Warriors	Zaza Pachulia	Draymond Green	Stephen Curry	Kevin Durant	Klay Thompson